

УТВЕРЖДЕН
RU.СКСХД.02.05.- 03 33 01

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

«BAUM STORAGE AI v2»

Руководство администратора

RU.СКСХД.02.05.- 03 33 01

Листов 445

2023

АННОТАЦИЯ

В данном документе описано руководство администратора программного обеспечения BAUM STORAGE AI v2.

СОДЕРЖАНИЕ

1. Введение	20
1.1. Область применения	20
1.1.1. Рабочее место пользователя	20
1.1.2. АРМ администратора	21
1.2. Краткое описание возможностей	22
1.3. Уровень подготовки пользователей	23
1.4. Перечень эксплуатационной документации, с которой необходимо ознакомиться пользователю	24
2. Подготовка к работе	25
2.1. Функционал РСХД (Аппаратное обеспечение шестиконтроллерной СХД)	25
2.1.1. Предоставление ресурсов пользователям	25
2.1.2. Управление хранилищем данных	25
2.1.3. Режимы работы системы	26
2.1.4. Первоначальная настройка	27
2.1.5. Контроль выполнения операций.	28
2.1.6. Вход пользователя в интерфейс управления	29
2.1.7. Выход из интерфейса управления	29
2.1.8. Организация интерфейса управления	30
2.1.9. Назначение меню интерфейса управления	30
2.1.9.1. Меню «Система»	30
2.1.9.2. Меню «Сервис S3»	33
2.1.9.3. Меню «Сетевые настройки»	34
2.1.9.4. Меню «Клиентские подключения»	35
2.1.10. Настройка доступа к ресурсу по протоколу S3	35
2.1.10.1. Создание пользователей S3	36
2.1.10.2. Удаление пользователя S3	37
2.1.10.3. Создание бакета S3	37
2.1.10.4. Удаление бакета S3	37
2.1.11. Настройка доступа к ресурсу по протоколу NFS	38
2.1.11.1. Создание и публикация тома NFS	38
2.1.11.2. Удаление тома NFS	39
2.1.11.3. Редактирование доступа к тому NFS	40
2.1.11.4. Изменение режима «Только чтение» для тома NFS	40
2.1.12. Настройка клиентов и групп доступа	41
2.1.12.1. Создание клиента	41
2.1.12.2. Редактирование параметров клиента	42
2.1.12.3. Удаление клиента	43
2.1.12.4. Создание группы доступа	43
2.1.12.5. Редактирование списка группы	44
2.1.12.6. Удаление группы доступа	44

2.1.13. Настройка Сети	45
2.1.13.1. Создание агрегированного интерфейса	45
2.1.13.2. Редактирование агрегированного интерфейса	46
2.1.13.3. Удаление агрегированного интерфейса	47
2.1.13.4. Создание виртуального интерфейса	48
2.1.13.5. Редактирование виртуального интерфейса	49
2.1.13.6. Удаление виртуального интерфейса	50
2.1.13.7. Подключение к серверу доменных имен и серверам времени	50
2.1.13.8. Создание кластерной группы портов	51
2.1.13.9. Удаление кластерной группы портов	52
2.2. Функционал СХД (Аппаратное обеспечение двухконтроллерной СХД)	53
2.2.1. Первоначальная настройка	53
2.2.1.1. Вход в систему	53
2.2.1.2. Описание структуры интерфейса управления	54
2.2.1.3. Изменение IP-адреса интерфейса управления контроллера	57
2.2.1.4. Включение мониторинга	58
2.2.1.5. Включение и выключение служб протоколов	59
2.2.1.6. Создание и удаление клиента	60
2.2.1.7. Системный рейд	61
2.2.1.8. Выход из системы управления	62
2.2.1.9. Уведомления об ошибках	62
2.2.1.10. Настройка сети	63
2.2.1.10.1. Изменение имени контроллера	66
2.2.1.10.2. Изменение имени кластера	66
2.2.1.10.3. Назначение IP-адреса интерфейсу	66
2.2.1.10.4. Удаление IP-адреса	67
2.2.1.10.5. Создание агрегированного интерфейса	68
2.2.1.10.6. Удаление агрегированного интерфейса	69
2.2.1.10.7. Создание виртуального интерфейса	69
2.2.1.10.8. Удаление виртуального интерфейса	70
2.2.1.10.9. Маршрутизация	70
2.2.1.10.10. Настройка адресов DNS и NTP серверов	72
2.2.1.10.11. Настройка параметров Active Directory и LDAP	73
2.2.1.10.12. Настройка управляющего интерфейса	74
2.2.2. Управление ресурсами	76
2.2.2.1. Дисковое пространство	76
2.2.2.1.1. Работа с дисками	76
2.2.2.1.2. Управление подсветкой дисков	77
2.2.2.1.3. Очистка меток	77
2.2.2.1.4. S.M.A.R.T	78
2.2.2.1.5. Запасные диски (Hot spare)	79

2.2.2.1.6. Замена диска после его извлечения	79
2.2.2.1.7. Удаление внешней дисковой полки	81
2.2.2.1.8. Подключение внешней дисковой полки	82
2.2.2.1.9. Очистка дисков	82
2.2.2.2. Работа с дисковыми пулами	83
2.2.2.2.1. Виды пулов	83
2.2.2.2.1.1. Обычные пулы.	83
2.2.2.2.1.2. Быстрые пулы.	84
2.2.2.2.2. Создание пула	84
2.2.2.2.3. Создание пула с автоматическим выбором дисков	86
2.2.2.2.4. Изменение пула	88
2.2.2.2.4.1. Добавление в пул запасного диска	88
2.2.2.2.4.2. Замена диска в пуле на запасной	89
2.2.2.2.4.3. Удаление из пула запасного диска	90
2.2.2.2.4.4. Включение/отключение подсветки диска в полке	91
2.2.2.2.4.5. Замена диска на свободный диск	92
2.2.2.2.4.6. Экспорт пула	93
2.2.2.2.4.7. Импорт пула	93
2.2.2.2.4.8. Добавление дисков в пул (расширение пула)	94
2.2.2.2.4.9. Изменение резервирования свободного места на пуле	96
2.2.2.2.5. Удаление пула	96
2.2.2.2.6. Статусы пула	97
2.2.2.2.7. Миграция ресурсов	98
2.2.2.2.8. Защита данных от повреждений	100
2.2.2.3. Работа с кэш-памятью (Данный раздел относится только к настройке обычных пулов)	100
2.2.2.3.1. Настройки кэша чтения.	100
2.2.2.3.1.1. Добавление к пулу кэша чтения.	100
2.2.2.3.1.2. Удаление из пула кэша чтения.	101
2.2.2.3.2. Настройки кэша записи.	102
2.2.2.3.2.1. Добавление к пулу NVDIMM кэша	102
2.2.2.3.2.2. Удаление из пула NVDIMM кэша	103
2.2.2.3.2.3. Подготовка SSD дисков для кэша записи.	103
2.2.2.3.2.4. Добавление в пул SSD кэша	104
2.2.2.3.2.5. Удаление из пула SSD кэша	105
2.2.2.3.2.6. Удаление разметки SSD дисков, помеченных как кэш на запись.	106
2.2.2.4. Работа с томами	107
2.2.2.4.1. Создание тома	107
2.2.2.4.2. Модификация тома/файловой системы.	108
2.2.2.4.3. Удаление тома.	109
2.2.2.4.4. Работа с томом по протоколу Fibre Channel (FC)	110
2.2.2.4.4.1. Создание FC LUN.	110

2.2.2.4.4.2. Подключение клиентов к FC LUN.	111
2.2.2.4.4.3. Групповое подключение клиентов к FC LUN	112
2.2.2.4.4.4. Отключение клиентов от FC LUN	113
2.2.2.4.4.5. Удаление FC LUN.	113
2.2.2.5. Работа с файловыми системами	114
2.2.2.5.1. Создание файловой системы.	115
2.2.2.5.2. Изменение параметров файловой системы.	116
2.2.2.5.3. Удаление файловой системы.	116
2.2.2.5.3.1. Включение и настройка службы LDAP	117
2.2.2.5.3.2. Подключение клиентов по протоколу NFS.	117
2.2.2.5.3.3. Групповое подключение клиентов по протоколу NFS	119
2.2.2.5.3.4. Отключение клиентов от NFS папки	120
2.2.2.5.3.5. Удаление папки NFS	121
2.2.2.5.4. Работа с файловой системой по протоколу SMB.	121
2.2.2.5.4.1. Интеграция с службой AD.	121
2.2.2.5.4.2. Создание и настройка SMB ресурса.	123
2.2.2.5.4.3. Защита изменения записанных файлов (WORM)	125
2.2.2.5.4.4. Групповое предоставление доступа к файловым системам	126
2.2.2.5.4.5. Создание теневых копий	127
2.2.2.5.5. Работа с файловой системой по другим протоколам (FTP, AFP)	129
2.2.2.5.5.1. Создание FTP ресурса	129
2.2.2.5.5.2. Удаление FTP ресурса	130
2.2.2.5.5.3. Создание AFP ресурса	130
2.2.2.5.5.4. Удаление AFP ресурса	131
2.2.2.6. Управление доступом к ресурсам	131
2.2.2.6.1. Создание клиента.	131
2.2.2.6.2. Удаление клиента	134
2.2.2.6.3. Создание клиента для файлового доступа (NFS).	135
2.2.2.6.4. Создание клиента для блочного доступа (FC)	136
2.2.2.6.5. Настройка групп клиентов.	138
2.2.2.6.5.1. Создание группы клиентов	138
2.2.2.6.5.2. Добавление клиентов в группу	139
2.2.2.6.5.3. Удаление группы	139
2.2.2.7. Управление учетными записями пользователей	140
2.2.2.7.1. Создание нового пользователя	140
2.2.2.7.2. Роли пользователей	141
2.2.2.7.3. Редактирование пользователя	142
2.2.2.7.4. Удаление пользователя	142
2.2.2.8. Работа с мгновенными снимками и клонами	142
2.2.2.8.1. Создание мгновенного снимка тома или файловой системы.	142
2.2.2.8.2. Восстановление данных из мгновенного снимка.	145

2.2.2.8.3. Клонирование мгновенного снимка.	145
2.2.2.8.4. Привязка LUN к созданному клону	147
2.2.2.8.5. Удаление клонов	148
2.2.2.8.6. Создание мгновенных снимков по расписанию.	148
2.2.2.8.7. Удаление задачи создания мгновенных снимков по расписанию.	150
2.2.2.8.8. Удаление мгновенных снимков тома или файловой системы.	150
2.2.2.9. Асинхронная репликация	151
2.2.2.9.1. Создание шаблона расписания	152
2.2.2.9.2. Создание шаблона цели	154
2.2.2.9.3. Задачи приема	155
2.2.2.9.4. Разовые задача репликации	156
2.2.2.9.5. Периодические задачи	158
2.2.2.9.6. Пример настройки асинхронной репликации на вторую СХД	159
2.2.2.10. Синхронная репликация	161
2.2.2.10.1. Задача синхронной репликации	162
2.2.2.10.2. Перемещение тома между пулами одного контроллера	164
2.2.2.10.3. Синхронная репликация по протоколу FC	165
2.2.2.10.4. Пример настройки синхронной репликация на удаленную СХД	167
2.2.2.11. Подключение к файловым ресурсам	168
2.2.2.11.1. Подключение к файловым ресурсам по протоколу NFS	168
2.2.2.11.1.1. Подключение к файловым ресурсам в среде Linux (Ubuntu)	168
2.2.2.11.1.2. Подключение к файловым ресурсам в среде VMWare.	168
2.2.2.11.2. Подключение к файловым ресурсам по протоколу SMB.	172
2.2.2.12. Подключение к блочным ресурсам	174
2.2.2.12.1. Подключение к блочным ресурсам по протоколу FC.	174
2.2.2.12.1.1. Подключение к блочным ресурсам в среде Windows Server 2012 по протоколу FC.	174
2.2.2.12.1.2. Подключение к блочным ресурсам в среде Linux по протоколу FC.	176
2.2.2.12.1.3. Подключение к блочным ресурсам в среде VMWare по протоколу FC.	178
2.2.2.13. Логирование событий	181
2.2.2.13.1. Журналы событий	181
2.2.2.13.2. Выгрузка системных записей	185
2.2.2.13.3. Удаленное логирование	185
2.2.2.13.4. Функция мониторинга и управления через Радиошлюз	185
2.2.2.14. Работа системы в многоконтроллерном режиме	188
2.3. Функционал BAUM STORAGE AI v2 для решения задач ИИ и машинного обучения	190
2.3.1. Регистрация и авторизация пользователя в Системе	190
2.3.2. Личный кабинет пользователя	191
2.3.3. Интерфейс	193
2.3.3.1. Меню интерфейса	193
2.3.3.2. Окно построения модели ИИ	196
2.3.3.3. Настройка внешнего вида интерфейса	199

2.3.3.4. Встроенные функции	200
2.3.3.4.1. Функции элемента «Источник данных»	200
2.3.3.4.2. Функции элемента «Процесс»	201
2.3.4. Загрузка данных в систему	205
2.3.4.1. Создание новой папки	206
2.3.4.2. Загрузка файлов	207
2.3.4.3. Предпросмотр данных	208
2.3.4.4. Взаимодействие с данными	210
2.3.4.5. Создание датасетов и загрузка данных для решения задач классификации	210
2.3.5. Создание модели ИИ	213
2.3.5.1. Создание новой и открытие сохраненной рабочей области	213
2.3.5.1.1. Создание новой рабочей области	213
2.3.5.1.2. Открытие сохраненной рабочей области	214
2.3.5.2. Построение блок-схемы	215
2.3.5.2.1. Блок «Запуск»	215
2.3.5.2.2. Блок «Источник данных»	216
2.3.5.2.3. Блок «Процесс»	220
2.3.5.3. Запуск блок-схемы на рабочей области	222
2.3.6. Сохранение модели ИИ	223
2.3.7. Графическое представление информации на рабочей области	225
2.3.7.1. Графики	225
2.3.7.2. Таблицы	228
2.3.7.3. Изображения	228
2.3.7.4. Описание модели	229
2.3.8. Работа с Дашбордами. Раздел «Визуализация».	230
2.3.8.1. Таблица	231
2.3.8.2. Видео	233
2.3.9. Создание отчета с результатами анализа данных	234
2.3.10. Конвейер приложений	235
2.3.11. Работа с проектом	237
2.3.11.1. Создание нового проекта	237
2.3.11.2. Редактирование проекта	238
2.3.11.3. Наполнение проекта	239
2.3.12. Настройка подключения к источникам данных	244
2.3.12.1. Типы коннекторов	244
2.3.12.2. Порядок работы с коннекторами	249
2.3.12.2.1. Создание коннектора	249
2.3.12.2.2. Запуск коннектора	250
2.3.12.2.3. Подключение к коннектору на дашборде	251
2.3.12.3. Настройка подключения на примере ClickHouse	251
2.3.12.4. Получение данных с камеры видеонаблюдения	257

2.3.13. Примеры работы с BAUM STORAGE AI v2	261
2.3.13.1. Обучение модели прогнозирования температуры воды и газов в котле	261
2.3.13.2. Создание блок схемы для работы с данными в режиме реального времени	275
2.3.13.3. Классификация изображений	286
2.3.13.3.1. Загрузка изображений	286
2.3.13.3.2. Создание модели классификации	290
2.3.13.4. Классификация текстов	296
2.3.13.5. Кластеризация Spark	305
2.3.13.6. Классификация текстовых данных с использованием слоя нейронной сети LSTM	317
2.3.13.7. Заполнение и работа с пропусками в табличных данных	337
2.3.14. Администрирование	345
2.3.14.1. Пользователи и группы	345
2.3.14.2. Настройка отправки уведомлений	348
2.3.15. Дополнительные возможности	352
2.3.15.1. Обращение в службу поддержки	352
2.3.15.2. История изменений	353

СОКРАЩЕНИЯ И ОПРЕДЕЛЕНИЯ

СХД	Система хранения данных
РСХД	Распределенная система хранения данных
ИИ	Искусственный интеллект
АРМ	Автоматизированное рабочее место
ПК	Персональный компьютер
ПАК	Программно-аппаратный комплекс
Apache Kafka	Брокер сообщений, реализующий паттерн Producer-Consumer. Данные из одного и того же топика могут считываться множеством консьюмер-групп одновременно.
Apache Spark	Фреймворк с открытым исходным кодом для реализации распределенной обработки неструктурированных и слабоструктурированных данных.
BPMN	С англ. Business Process Model and Notation. Нотация определяющая способ визуализации процессов в виде диаграмм с определенным набором блоков и взаимосвязей.
YOLO	С англ. You Only Look Once. Архитектура нейронных сетей, предназначенная для детекции объектов на изображении. Отличительной особенностью является подход к решению задачи детекции: исходное изображение сжимается таким образом, чтобы получить квадратную матрицу размером 13 на 13, в каждой клетке которой записана информация о наличии объекта и классе этого объекта на соответствующей части картинке. Таким образом, YOLO просматривает картинку один раз, что существенно увеличивает скорость обработки.
YOLOv5	Усовершенствованная пятая версия YOLO, реализованная на фреймворке PyTorch.
Временной ряд	Совокупность наблюдений, собранных за определенный временной интервал. Этот тип данных используется для поиска долгосрочного тренда, прогнозирования будущего и прочих видов анализа. Анализ временных рядов позволяет обнаруживать тенденции и закономерности в исследуемых процессах, строить прогнозы и предсказывать будущие изменения в бизнесе, на производстве, и в других областях.
Выборка	Случайное подмножество генеральной совокупности.
Датасет	С англ. Data set, набор данных. Коллекция из логических записей, хранящихся в виде кортежа. Набор данных можно сравнить с файлом, но в отличие от файла он является

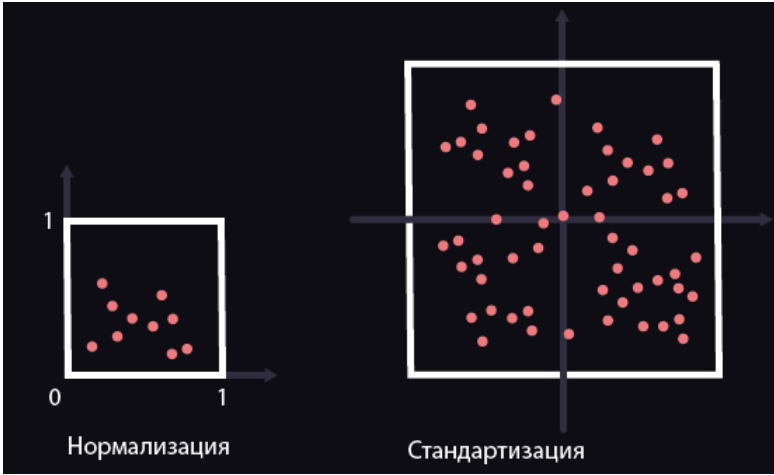
	<p>одновременно и каталогом, и файлом файловой системы, и не может содержать в себе другие наборы.</p> <p>Файловая система ориентирована на хранение записей, которые являются неделимыми единицами хранения. Множества записей объединяются в группы, которые и называются наборами данных. Записи в наборах данных используются приложениями, например, как входные данные. Так, записями набора данных могут быть как текстовые данные, так и изображения.</p> <p>К набору данных можно обратиться, указав точное место его хранения, или, если ранее для набора было зазервировано имя в файловой системе, по имени (второй вариант не реализован).</p> <p>Также датасетом называются данные, которые пользователь загрузил с локального компьютера, а Система при загрузке выполнила их предварительную обработку. Например, в систему загружается временной ряд в формате csv, и при загрузке он преобразуется в структурированные данные или по-другому датасет. Такой датасет пригоден для использования в моделях машинного обучения.</p> <p>В качестве датасета в настоящем документе также выступает набор изображений или файлов, которые используются для решения одной задачи.</p>
<p>Дерево решений</p>	<p>Инструмент прогнозного моделирования. Строится с помощью алгоритмического подхода, который разделяет набор данных на основе различных условий. Относится к классу обучения с учителем, используется для задач классификации и регрессии. Цель в том, чтобы создать модель, которая предсказывает значение целевой переменной, изучая правила принятия решений, выведенные из характеристик данных.</p> <p>Алгоритм:</p> <ol style="list-style-type: none"> 1. Выбрать лучший атрибут, который разделяет наблюдения на группы. 2. Задать соответствующий вопрос. 3. Следовать по путям ответов. <p>Вернуться к шагу 1.</p>
<p>Дисперсия</p>	<p>Мера удаленности того или иного значения выборки от среднего значения.</p> <p>Рассчитывается по формуле:</p> $\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{N}$ <p>где σ^2 – дисперсия случайной величины, x_i – i-ый элемент выборки, \bar{x} – среднее арифметическое, N – количество элементов выборки.</p> <p>Стандартное отклонение рассчитывается как квадратный корень из полученной цифры.</p>

<p>Интеллектуальный анализ текстов</p>	<p>Направление в искусственном интеллекте, целью которого является получение информации из коллекции текстовых документов, основываясь на применении эффективных в практическом плане методов машинного обучения и обработки естественного языка.</p> <p>Ключевыми группами задач ИАТ являются: категоризация текстов, извлечение информации и информационный поиск, изменение информации в коллекциях текстов.</p> <p>Категоризация документов заключается в отнесении документов из коллекции к одной или нескольким группам (классам, кластерам) схожих между собой текстов. Категоризация с участием человека называется классификацией документов, система ИАТ должна отнести тексты к уже определенным классам. Для этого производится обучение с учителем, для чего пользователь предоставляет системе ИАТ как множество классов, так и образцы документов, принадлежащих этим классам.</p> <p>Второй случай категоризации называется кластеризацией документов. При этом система ИАТ сама определяет множество кластеров, по которым могут быть распределены тексты (производится обучение без учителя). В этом случае пользователь сообщает системе ИАТ количество кластеров, на которое ему хотелось бы разбить обрабатываемую коллекцию (в алгоритме программы уже заложена процедура выбора признаков).</p>
<p>Категориальная переменная (качественные данные)</p>	<p>Это данные с ограниченным числом уникальных значений или категорий (например, пол, страна проживания, номер группы, категория товаров, и т.п.). Категориальные поля могут быть как текстовыми, так и числовыми, в которых категории закодированы числовыми кодами (например, 0=женский, а 1=мужской).</p> <p>Номинальные поля, порядковые поля и флаги являются категориальными полями.</p> <p>-Набор (номинальная переменная). Поле, значения которого представляют категории без естественного упорядочивания (например, подразделение компании, в котором работает сотрудник).</p> <p>-Упорядоченный набор (порядковая переменная). Поле, значения которого представляют категории с некоторым естественным для них упорядочением (например, оценки, представляющие степень удовлетворенности или уверенности, или баллы, оценивающие предпочтение).</p> <p>-Флаг. Поле или переменная с двумя отдельными значениями, например Да и Нет.</p>
<p>Классификация</p>	<p>Задача машинного обучения, которая ставит своей целью назначить метку класса наблюдениям из предметной области.</p> <p>Основные типы классификации:</p> <ul style="list-style-type: none"> • бинарная классификация;

	<ul style="list-style-type: none"> • мультиклассовая классификация; • классификация по нескольким меткам; <p>несбалансированная классификация.</p>
Кластеризация	<p>Техника обучения без учителя, которая включает в себя группирование или кластеризацию точек данных. Чаще всего она используется для сегментации потребителей, выявления мошенничества и классификации документов.</p> <p>Кластеризация (или кластерный анализ) – это задача разбиения множества объектов на группы, называемые кластерами. Главное отличие от классификации состоит в том, что перечень групп четко не задан и определяется в процессе работы алгоритма.</p>
Машинное обучение (ML – Machine learning)	<p>Тренировка математической модели на исторических данных для того, чтобы прогнозировать какое-то событие или явление на новых данных. То есть попытка заставить алгоритмы программ совершать действия на основе предыдущего опыта, а не только на основе имеющихся данных.</p> <p>Для обучения нужны исторические данные (обучающая выборка) и значение целевой переменной (то, что прогнозируем), которое соответствует заданным историческим данным. Модель наблюдает и находит зависимости между данными и целевой переменной. Эти зависимости используются моделью для нового набора данных, чтобы прогнозировать целевую переменную, которая неизвестна.</p> <p>Машинное обучение включает в себя целый набор методов и алгоритмов, которые могут предсказать какой-то результат по входным данным.</p> <p>Алгоритмов машинного обучения большое множество: одни эффективны для решения одного типа задач, вторые – для другого.</p> <p>Суть технологии машинного обучения</p> <p>Говоря в общем, машинное обучение – это обучение компьютерной программы или алгоритма постепенному улучшению исполнения поставленной задачи.</p> <p>Машинное обучение обозначает множество математических, статистических и вычислительных методов для разработки алгоритмов, способных решить задачу не прямым способом, а на основе поиска закономерностей в разнообразных входных данных. Решение вычисляется не по четкой формуле, а по установленной зависимости результатов от конкретного набора признаков и их значений. Например, если каждый день в течение недели земля покрыта снегом и температура воздуха существенно ниже нуля, то вероятнее всего, наступила зима. Поэтому машинное обучение применяется для диагностики, прогнозирования, распознавания и принятия решений в</p>

	различных прикладных сферах: от медицины до банковской деятельности.
Мониторинг	Процесс наблюдения и регистрации данных о каком-либо объекте на неразрывно примыкающих друг к другу интервалах времени, в течение которых значения данных существенно не изменяются.
Мониторинг состояния	Наблюдение за состоянием объекта для определения и предсказания момента перехода в предельное состояние. Результат мониторинга состояния объекта представляет собой совокупность диагнозов составляющих его субъектов, получаемых на неразрывно примыкающих друг к другу интервалах времени, в течение которых состояние объекта существенно не изменяется. Принципиальное отличие от мониторинга параметров является наличие интерпретатора измеренных параметров в терминах состояния – экспертной системы поддержки принятия решений о состоянии объекта и дальнейшем управлении.
Наблюдение (строка, запись, точка, сущность)	<p>Ценные данные, собираемые во время исследования или эксперимента. Вместе с масштабом анализа определяет совокупность.</p> <p>Эмпирические исследования – практические эксперименты с результатами на основе реального опыта, а не теории или убеждений. Основопологающим принципом Науки о данных является приоритет наблюдения над предположением.</p> <p>Типы наблюдений:</p> <ul style="list-style-type: none"> ● Числовой: целые (integer), вещественные (real number), числа с плавающей точкой (float). ● Булевый (boolean) – принимает значения 1/0 (да/нет). ● Категориальный. Например, жанры кино: комедия, ужасы, мелодрама. ● Текстовый. <p>Вектор.</p>
Нейронная сеть (или Искусственная нейронная сеть)	<p>Представляет собой математическую модель, а также её программное или аппаратное воплощение, построенную по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма.</p> <p>Нейронные сети решают задачу: по точкам находят функцию. Происходит это путем минимизации ошибки – сводится к минимуму «расстояние» между значениями, предсказываемыми нейронной сетью, и значениями, которые наблюдаются.</p> <p>Под архитектурой нейронной сети понимается ее устройство – последовательность нейронов и связей между ними.</p>
Нормализация	Техника преобразования значений признака, масштабирующая значения таким образом, что они расположены в диапазоне от 0 до 1.

Обучение учителем	<p>с</p> <p>Контролируемое обучение – метод машинного обучения, при котором модель обучается на размеченных данных. Например, исследовав опухоли, установив их размер, плотность и другие метрики, мы передаем эти данные модели с обязательной пометкой, какое наблюдение к какому строению (доброкачественному или злокачественному) относится.</p> <p>Алгоритмы контролируемого обучения подразделяются на следующие модели: классификация, регрессия.</p>
Пайплайн	<p>Последовательность стадий, внутри которых расположены задачи. Расположены они таким образом, что выход каждого элемента является входом следующего.</p>
Признак	<p>Объективная характеристика, характерная черта или свойство, которое может быть определено или измерено.</p> <p>В статистике независимые переменные X используются для предсказания зависимого признака Y</p>
Регрессия (в математической статистике)	<p>Математическое выражение, отражающее связь между зависимой переменной y и независимыми переменными x.</p> <p>Алгоритмы регрессии используются для контролируемого обучения моделей искусственного интеллекта. Модели обучают прогнозировать числовые значения целевых переменных.</p>
Стандартизация	<p>Техника преобразования значений признака, адаптирующая признаки с разными диапазонами значений к моделям машинного обучения, использующих дистанцию для прогнозирования. Это разновидность нормализации с использованием стандартизированной оценки преобразует значения так, что из каждого наблюдения каждого признака вычитается среднее значение и результат делится на стандартное отклонение этого признака:</p> $x_{i, \text{станд.}} = \frac{x_i - \mu}{\sigma},$ <p>где $x_{i, \text{станд.}}$ – стандартизированный элемент признака, x_i – исходный элемент, μ – среднее арифметическое, σ – стандартное отклонение.</p> <p>Преобразование необходимо, поскольку признаки датасета могут иметь большие различия между своими диапазонами, и для моделей машинного обучения это спровоцирует искаженное восприятие данных.</p> <p>Стандартизация, в отличие от нормализации, не имеет ограничивающего диапазона:</p>

	
Стандартизованная оценка	Метрика, характеризующая удаленность наблюдения от среднего значения совокупности данных.
Стандартное отклонение	Мера разброса в наборе числовых данных. Показывает, насколько далеко от среднего арифметического находятся точки данных. Чем меньше стандартное отклонение, тем более сгруппированы данные вокруг центра (среднего). Чем отклонение больше, тем больше разброс значений.
Тренировочные (обучающие) данные	Часть датасета, обучающая основа модели машинного обучения. Является одной из составляющих разделенного набора данных наряду с <i>тестовыми</i> и <i>валидационными</i> данными.
Валидационные (тестовые) данные	Часть датасета, основа для проверки работоспособности модели машинного обучения.
Числовая переменная	Numeric variable - переменная, выраженная различными видами чисел.
Целевая (зависимая) переменная	Target variable – признак датасета, который предстоит предсказывать модели машинного обучения. Зависимой ее называют, поскольку в ходе разведочного анализа данных выявляется корреляция между одной или несколькими переменными-предикторами и рассматриваемым целевым признаком.

1. Введение

1.1. Область применения

Данное руководство применимо для пользователя программного обеспечения BAUM STORAGE AI v2.

1.1.1. Рабочее место пользователя

Рабочее место пользователя BAUM STORAGE AI v2 представляет из себя ПК (см. рис. 1.1.1-1). Работа пользователя с BAUM STORAGE AI v2 осуществляется через браузер. В адресной строке необходимо ввести адрес, который указывается отдельно при конкретной поставке.

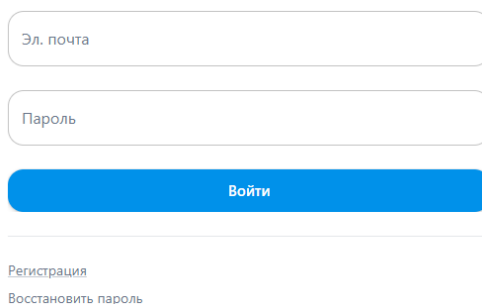


Рисунок 1.1.1-1 – Страница авторизации пользователя в Системе

После успешной авторизации пользователь может начать работать на BAUM STORAGE AI v2. На рисунке 1.1.1-2 представлен типовой пример собранного пайплайна.



Рисунок 1.1.1-2 - Пример типowego пайплайна и визуализации

Пример главного рабочего окна представлен на рисунке 1.1.1-3. У пользователя есть возможность подключать коннекторы, собирать пайплайны, обучать модели, смотреть отчёты и собирать проекты

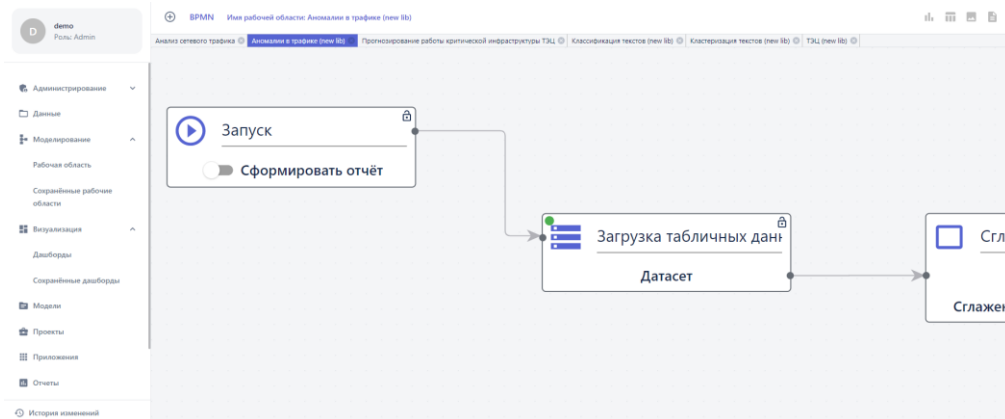


Рисунок 1.1.1-3 - Пример рабочего окна пользователя

1.1.2. АРМ администратора

АРМ администратора позволяет администратору настраивать систему хранения данных. На рисунке 1.1.1-4 представлен пример рабочего окна. У администратора есть возможность производить настройку, производить мониторинг состояния и управлять ресурсами.

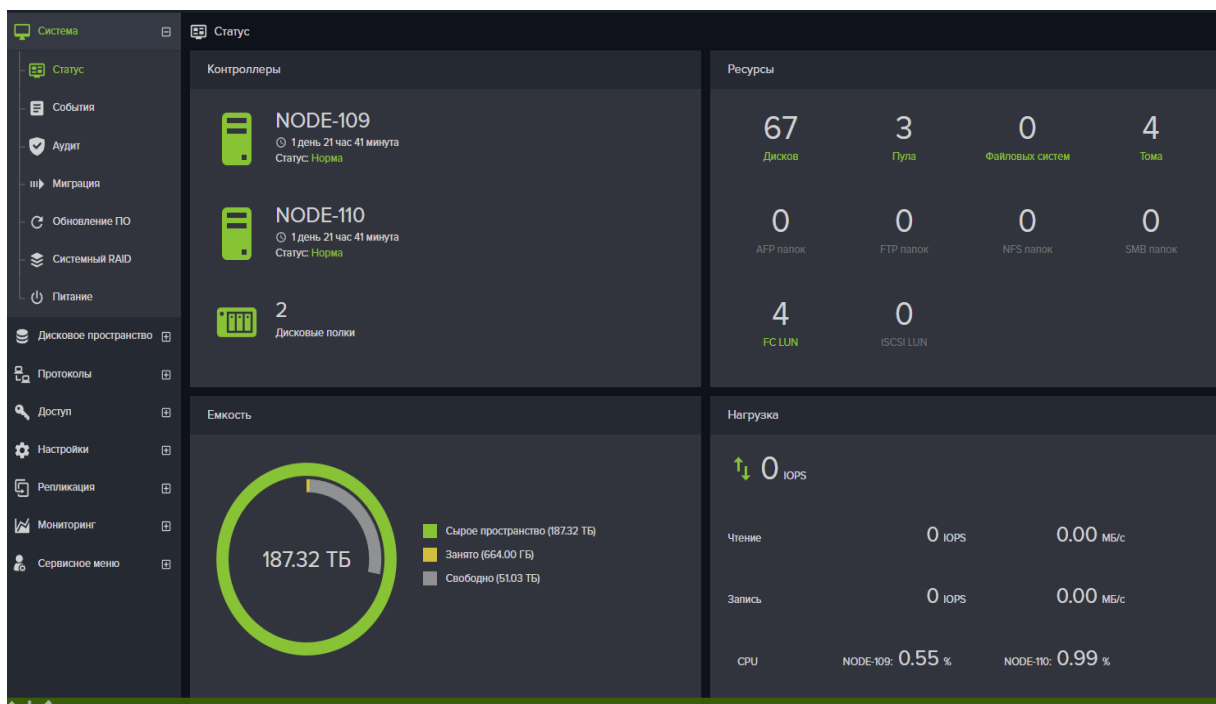


Рисунок 1.1.1-4 - Пример рабочего окна АРМ администратора

1.2. Краткое описание возможностей

Данное руководство содержит материал и ссылки, необходимые для работы с BAUM STORAGE AI v2.

Функционал СХД программного обеспечения BAUM STORAGE AI v2 предназначено для управления аппаратными ресурсами системы хранения данных на всех уровнях, от дисков до файловой системы и предоставлением универсального доступа к этим ресурсам внешним клиентам.

СХД должно состоять из двух контроллеров, собранных в отказоустойчивый кластер, и общего для них дискового массива. В таком случае на каждом из контроллеров можно создавать ресурсы для хранения данных и предоставлять к ним доступ внешним системам, по файловым либо блочным протоколам. В зависимости от типа созданного ресурса (файлового или блочного), он предоставляется клиентам в виде виртуального жесткого диска либо сетевой папки.

Для организации надежного хранения данных, диски из состава общего дискового массива собираются в пулы, которые могут иметь различные типы RAID: 1, 10, 5, 50, 6, 60, а также с тремя дисками под четность – ВЗ. На пулах создаются разделы для хранения данных: тома либо файловые системы. Созданные разделы, в свою очередь, предоставляются внешним системам (клиентам) по одному из файловых либо блочных протоколов: FC, ISER, CIFS(SMB), FTP. Есть возможность создания «быстрых» пулов, обладающих улучшенной производительностью.

Для защиты данных, записанных на тома, применяются снэпшоты (снимки) и репликация томов на другую СХД. Восстановление данных из снимка возможно в двух режимах: восстановление исходного тома на момент создания снимка и клонирование снимка с созданием нового тома. Репликация данных на другую СХД выполняется синхронным либо асинхронным способом.

Управление доступом к созданным ресурсам выполняется при помощи фильтрации адресов клиентов. Имеется возможность расширенного управления доступом к файловым ресурсам через интеграцию со службами каталогов Active Directory (для протокола SMB).

Управление СХД осуществляется через веб-интерфейс или интерфейс командной строки, который работает на специально выделенном для этой цели сетевом интерфейсе на каждом из контроллеров. Для управления кластером можно подключаться к управляющему интерфейсу любого контроллера.

Отказоустойчивость СХД реализована за счет кластеризации ресурсов. Контроллеры работают в режиме Active ALUA. При отказе одного из контроллеров, обслуживание его ресурсов берет на себя рабочий контроллер. Этот процесс называется миграцией ресурсов.

Все действия пользователя в интерфейсе управления, а также системные события от ПО или оборудования СХД записываются в системном журнале (логе), из которого можно выгрузить историю событий за указанный период. Система следит за переполнением раздела на котором хранятся файлы логов и при необходимости удаляет старые файлы.

Синхронизация времени на узлах СХД выполняется при помощи внешнего NTP сервера. Перед созданием и выделением ресурсов рекомендуется указать NTP серверы в меню сетевых настроек и выбрать временную зону.

Функционал РСХД BAUM STORAGE AI v2 предназначен для управления распределенным аппаратным комплексом, служащим для передачи и отказоустойчивого хранения больших объемов пользовательских данных.

Данное ПО разработано для использования в составе программно-аппаратного комплекса хранения данных, предназначено для автоматизации процессов передачи и хранения больших объемов данных в компаниях с географически разнесенными структурными подразделениями. В состав кластера РСХД для поддержания отказоустойчивости должно входить не менее двух модулей управления и не менее двух модулей хранения.

Схема типового подключения РСХД показана на рисунке ниже.

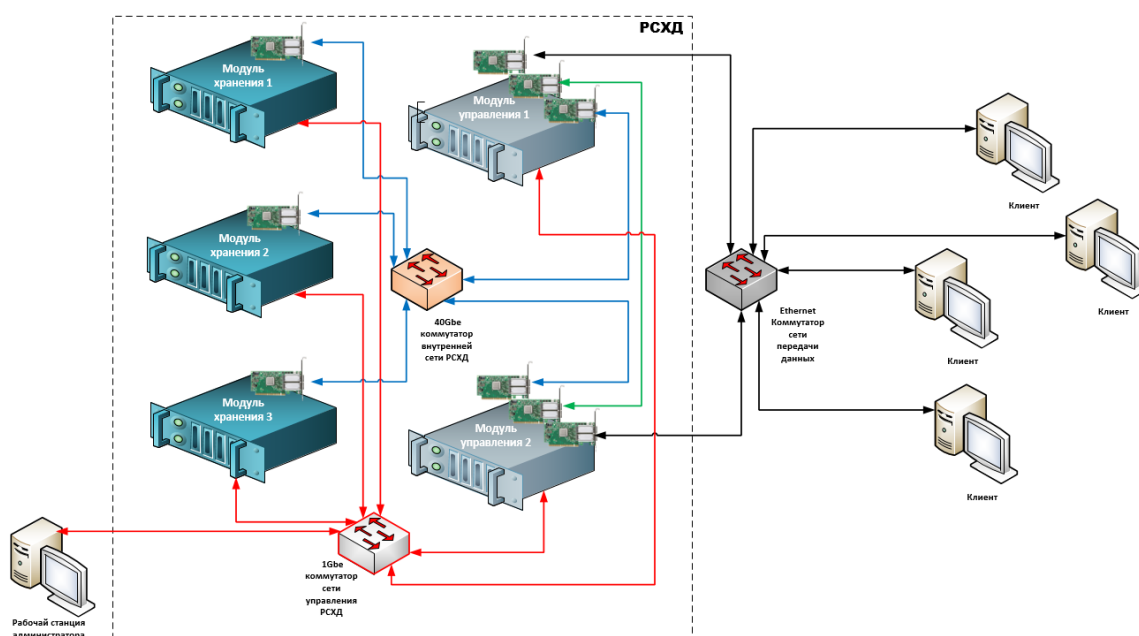


Рисунок 1.2. Типовая схема подключения модулей РСХД

Функционал BAUM STORAGE AI v2 для решения задач ИИ и машинного обучения предназначен для обработки структурированных и неструктурированных массивов данных и обучения моделей искусственного интеллекта для создания баз знаний, предиктивной аналитики в промышленности и медицине. BAUM STORAGE AI v2 позволяет выполнять полный цикл data science.

1.3. Уровень подготовки пользователей

Пользователи должны иметь навыки работы на ПК и с интернет-браузерами и изучить порядок выполнения действий и функций, описанный в данном руководстве.

В Системе предполагается наличие ролей пользователей – аналитик, моделирующий с помощью конструктора искусственного интеллекта работу бизнес-процесса, и оператор, обладающий возможностью только просматривать результаты запуска конструктора. BAUM STORAGE AI v2 позволяет настраивать уровни доступа к разделам системы для разных пользователей.

1.4. Перечень эксплуатационной документации, с которой необходимо ознакомиться пользователю

До начала работы Пользователь должен ознакомиться со следующими документами:

- 1) Общее описание системы
- 2) Руководство пользователя (данный документ)

2. Подготовка к работе

2.1. Функционал РСХД (Аппаратное обеспечение шестиконтроллерной СХД)

2.1.1. Предоставление ресурсов пользователям

Система передается заказчику в предварительно настроенном состоянии, готовая к работе. На первом этапе администратору необходимо задать ip-адреса интерфейсу управления в соответствии с адресным пространством своей сети. Также должны быть указаны адреса DNS и NTP серверов и настроен почтовый клиент для отправки уведомлений на электронную почту (отправка уведомлений на почту в текущей версии не реализована).

Второй этап - настройка кластерных групп портов, по которым будут розданы ресурсы СХД. Для этого на каждом из модулей управления выбираются порты, по которым будет работать протокол S3 и на них создаются виртуальные интерфейсы. Созданные виртуальные интерфейсы с первого модуля управления и со второго модуля управления нужно объединить попарно в кластерную группу портов. Виртуальным интерфейсам должны быть назначены IP-адреса в соответствии с адресным пространством сети.

Третий этап – создание списков доступа к ресурсам СХД. Для каждого ресурса создается запись – клиент, в которую заносятся ip-адреса либо IQN компьютеров, которым будет разрешен доступ к ресурсу. Клиенты собираются в группы доступа, которые будут привязаны к создаваемым ресурсам.

Для протокола S3 создаются пользователи, для доступа которых система генерирует ключ доступа.

Четвертый этап – выделение ресурса клиенту по конкретному протоколу. Администратор выбирает протокол, по которому будет подключаться клиент и создает раздел для хранения данных, котируя его размер. Настраивается процент заполнения выделенного раздела, по достижению которого администратору будет отправлено оповещение о заканчивающемся свободном месте (оповещение о переполнении раздела в текущей версии не реализовано). К создаваемому разделу подключается группа доступа со списком разрешенных адресов хостов.

2.1.2. Управление хранилищем данных

Хранилище данных СХД физически распределено по всем модулям хранения системы. Свободное пространство хранилища может быть использовано всеми клиентами СХД по мере записи ими данных. Защита от бесконтрольного использования клиентами дискового объема решена при помощи дисковых квот.

При добавлении нового модуля хранения, общее пространство хранения системы увеличится в соответствии с объемом добавленного модуля, после чего по команде администратора система автоматически перераспределяет реплики записанных в системе данных по всем имеющимся модулям хранения для обеспечения избыточности хранимых данных.

При отключении или отказе модуля хранения, система автоматически запустит процесс перемещения блоков данных между оставшимися модулями хранения и по команде администратора выведет отказавший модуль из состава кластера.

Повторное включение выведенных из состава кластера модулей выполняется администратором.


При выходе из строя жесткого диска в модуле хранения, система автоматически реплицирует блоки на модулях хранения таким образом, чтобы имелось достаточное количество реплик для поддержания избыточности данных. Администратор должен подтвердить добавление нового жесткого диска в состав модуля хранения.

Модули управления работают в отказоустойчивых парах, реплицируя друг на друга кэши записи и метаданные. При отключении одного из модулей управления пары, его клиенты автоматически переключатся на другой модуль управления. Включение в работу отказавшего модуля управления выполняет администратор.

Пространство хранения при необходимости может быть дефрагментировано в фоновом режиме при появлении неиспользуемых участков памяти на дисках в фоновом режиме.

2.1.3. Режимы работы системы

Система может работать в трёх режимах: нормальный режим, режим ограниченной функциональности и режим ограниченной производительности.

Информацию о состоянии кластера можно увидеть, нажав на значок  «Состояние», который находится на панели уведомлений интерфейса управления.

Нормальный режим – штатный режим работы системы, при котором все компоненты системы полностью работоспособны.

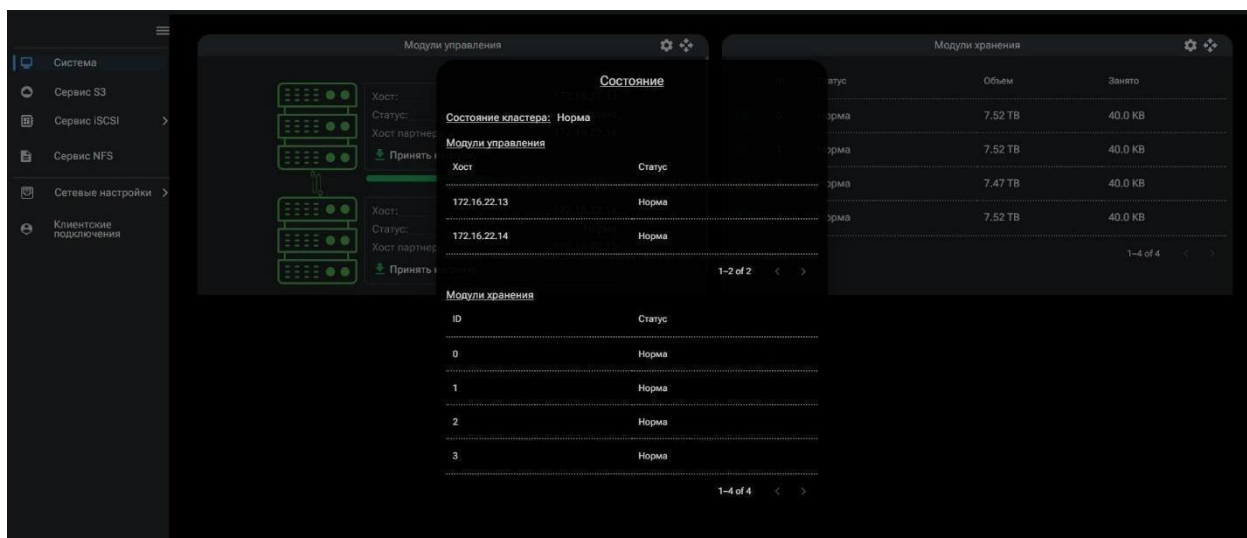


Рисунок 2. Пример окна меню «Состояние» при работе СХД в нормальном режиме

Режим ограниченной функциональности – режим работы после отказа одного или нескольких модулей хранения (зависит от начальной конфигурации кластера) в результате чего хранилище данных системы перешло в режим только чтение.

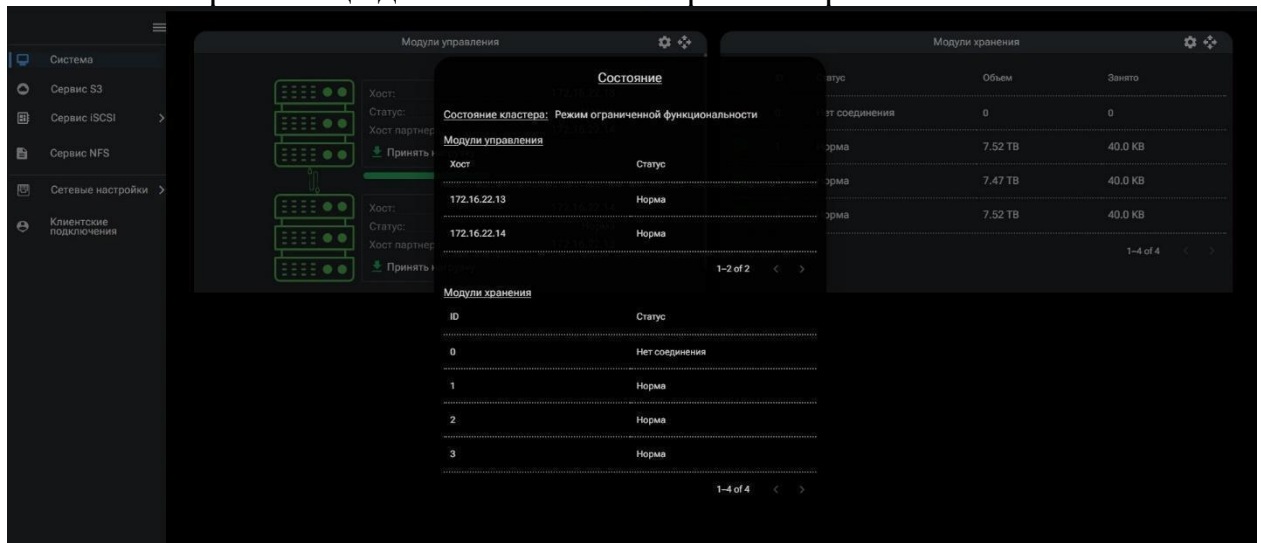


Рисунок 3. Пример окна меню «Состояние» при переходе СХД в режим ограниченной функциональности

Режим ограниченной производительности – режим работы после отказа модуля управления в одной или нескольких отказоустойчивых парах. Обслуживание клиентов принимает на себя оставшийся работоспособным модуль управления, вследствие чего снижается общая производительность системы.

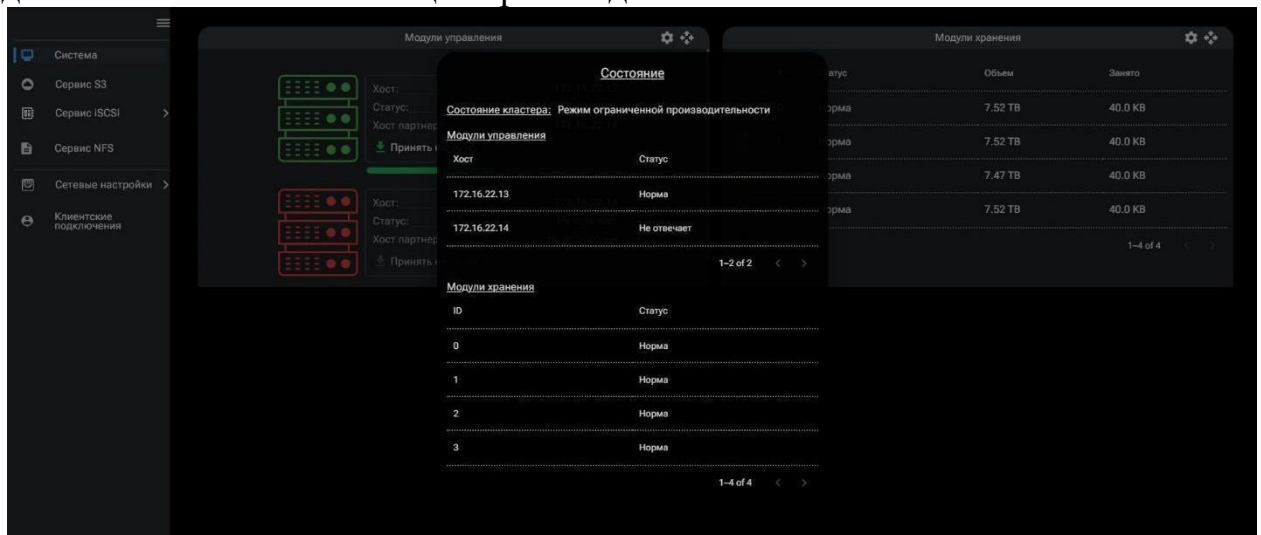


Рисунок 4. Пример окна меню «Состояние» при переходе СХД в режим ограниченной производительности

2.1.4. Первоначальная настройка

До начала работ по созданию ресурсов и подключению пользователей, необходимо выполнить ряд действий по первичному конфигурированию системы. Начальное конфигурирование распределенной системы хранения данных включает в себя:

1. Настройку ip-адреса интерфейса управления;

2. Настройку соединения с DNS и NTP серверами;
3. Настройка ip-адресов сетевых интерфейсов.

2.1.4.1. Настройка сетевых интерфейсов

Интеграция системы хранения данных в сетевую инфраструктуру требует настройки управляющих сетевых интерфейсов (для администрирования системы) и нагрузочных сетевых интерфейсов (для подключения клиентов). До использования, сетевым интерфейсам должны быть присвоены IP-адреса. Настройки IP-адресов выполняются на вкладке меню «Интерфейсы». Описание инструментария настройки сетевых интерфейсов приведено в разделе 4.5 «Настройка сети» настоящего руководства.

2.1.4.2. Настройка соединения с серверами доменных имен и серверами времени

Для синхронизации работы системы, необходимо подключение к серверу имен (DNS) и серверу времени (NTP), доступ к которым должен выполняться из сети управления. Настройка адресов основного и резервного DNS и NTP серверов выполняется во вкладке меню «DNS & NTP», в области «Сетевые настройки».

2.1.5. Контроль выполнения операций.

Для контроля исполнения команд пользователя, в графическом интерфейсе управления предусмотрены уведомления. Они представляют собой специальные окна, отображаемые в правом нижнем углу экрана. Уведомления сообщают о успешном либо не успешном выполнении операции. Пример уведомления представлен на Рисунок 19. При успешном выполнении команды, система сообщит об удачном выполнении операции.

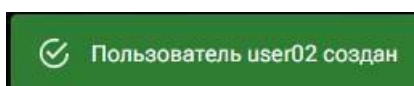


Рисунок 19. Пример уведомления о успешно выполненной операции

В случае ошибки, окно уведомления будет иметь красный цвет. Дополнительную информацию об ошибке можно скопировать в буфер обмена компьютера, нажав на ссылку «Копировать данные в буфер» См. Рисунок 20.

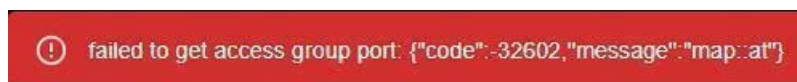


Рисунок 20. Пример уведомления об ошибке

2.1.6. Вход пользователя в интерфейс управления

Компьютер, с которого осуществляется администрирование системой должен быть подключен к сети управления, в которую подключены управляющие интерфейсы модулей (серверов) РСХД.

Для выполнения входа в интерфейс управления в адресной строке интернет-браузера необходимо ввести IP-адрес управляющего интерфейса системы, заданный при её инсталляции.

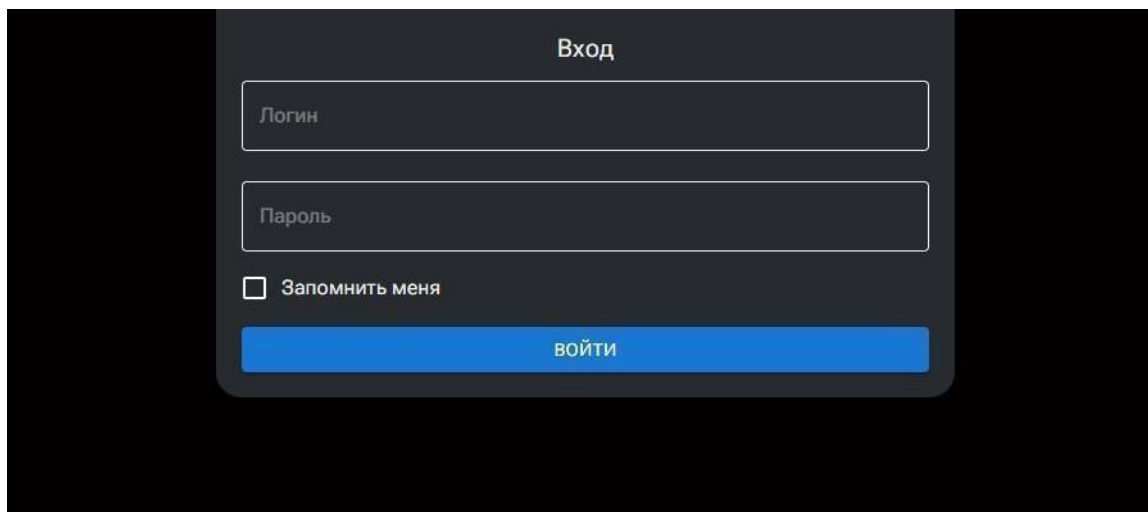


Рисунок 5. Окно входа в систему

После подключения, появится окно входа в систему (см. Рисунок 5), в котором необходимо ввести логин и пароль, после чего нажать на кнопку «Войти» (при первом входе в систему, используйте логин и пароль указанные ниже, если вы уже изменили пароль, используйте его):

- Логин: **admin**;

- Пароль: **123456**.

Если вход успешно выполнен, на экране появится меню «Система»

2.1.7. Выход из интерфейса управления


Для выхода из интерфейса управления используется кнопка  расположенная в верхней части окна панели управления.



Рисунок 6. Панель уведомлений интерфейса управления

2.1.8. Организация интерфейса управления

В левой части главного окна интерфейса управления расположены заголовки разделов основного меню управления системы.

В правой части окна отражены сведения о просматриваемых разделах, а также размещены элементы управления функционалом системы.

Вверху окна горизонтально расположена панель уведомлений, в которой отображаются меню «Состояние», меню «Операции с кластером», кнопка выхода из системы и кнопка смены языка, Рисунок 6.

В меню «Операции с кластером» расположены кнопки управления функционалом кластера:

Кнопка «**Обновить кластер**» - служит для включения обнаруженных (новых) модулей в состав кластера.

2.1.9. Назначение меню интерфейса управления

2.1.9.1. Меню «Система»

На вкладке меню «Система» расположена информация о составе кластера, статус работы кластера и информация о размере информационного хранилища системы и его утилизации. В правой части окна выведен список подключенных модулей хранения, из которых состоит информационное хранилище. Окно вкладки меню «Система» показано на Рисунок 7.

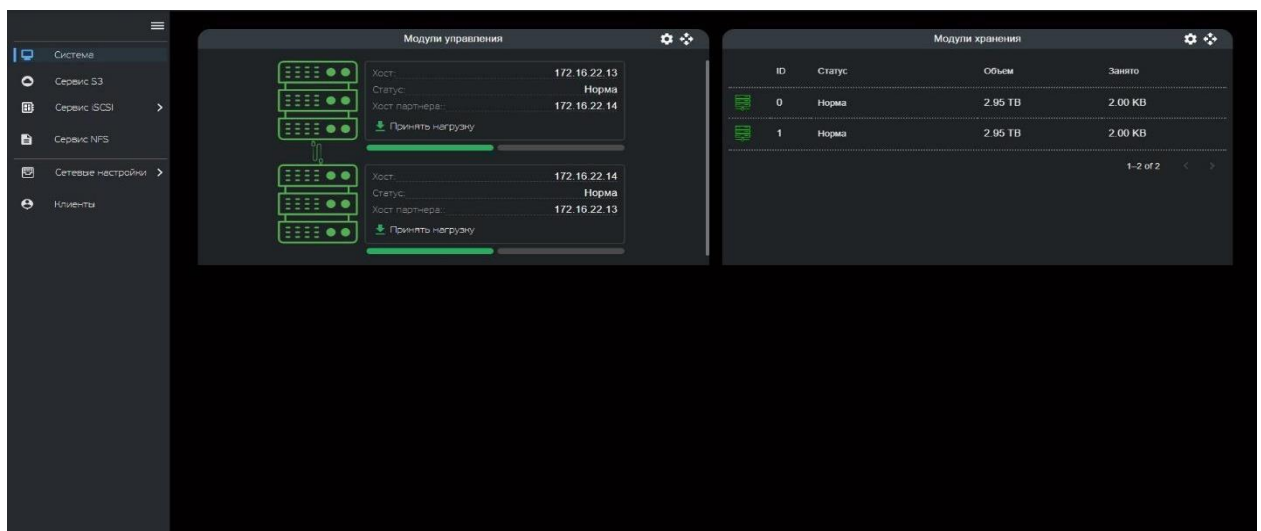


Рисунок 7. Вкладка меню «Система»

Список «**Модули управления**» содержит перечень модулей управления (МУ) кластера.

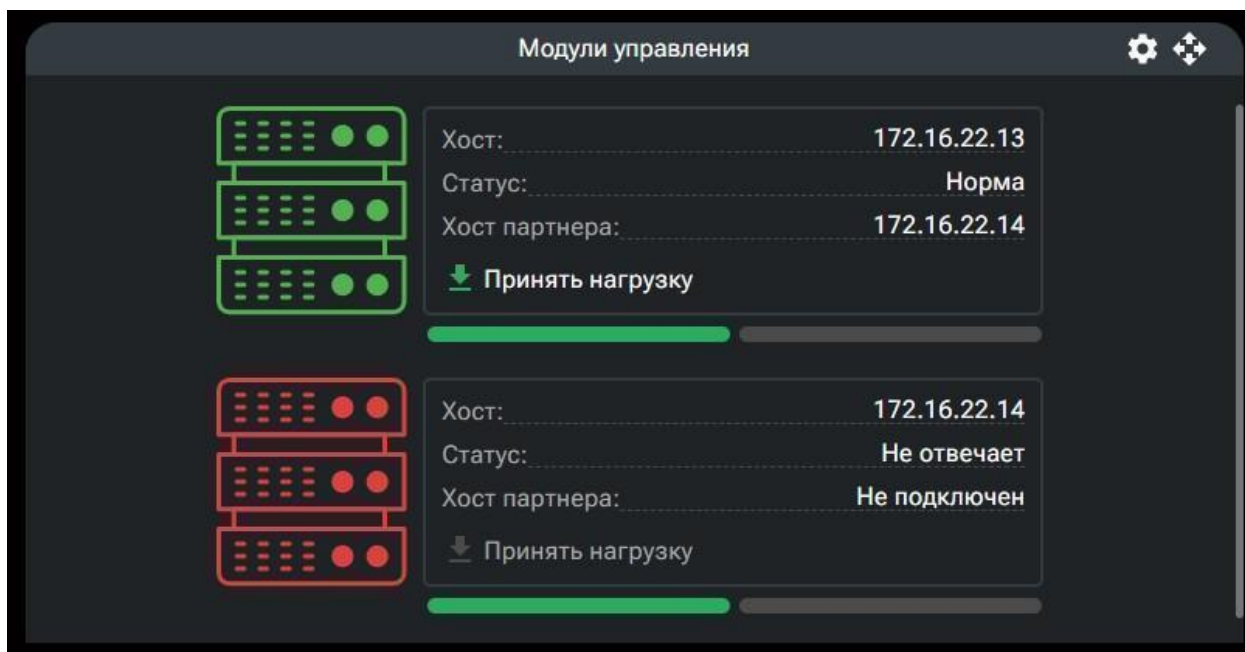


Рисунок 8. Список модулей управления кластера СХД

Слева отображается индикатор статуса МУ (зеленый - в сети, красный - не в сети)

В столбце «Хост» выводится IP-адрес модуля управления.

В столбце «Статус» выводится статус соединения:

- **Норма** – модуль отвечает на запросы других модулей кластера;
- **Не отвечает** – связь с модулем отсутствует.

В столбце «Хост партнера» выводится IP-адрес парного модуля управления (модули управления в кластере собираются в отказоустойчивые пары), который работает в паре.

Информацию по кэшу записи выбранного модуля управления может быть получена по щелчку ЛК мыши на зеленый таб.

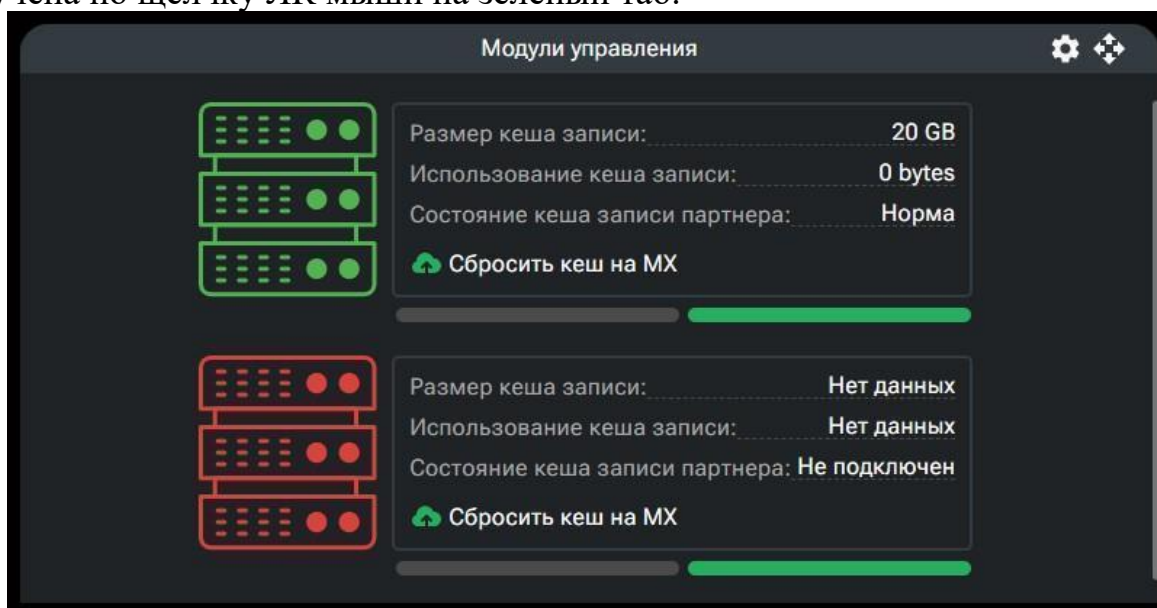


Рисунок 9. Окно информации о кеше записи.

В окне информации по кэшу записи выводится:

Размер кэша записи;

Использование кэша записи;

Состояние кэша записи партнера.

Список «Модули хранения» содержит видимые в кластере модули хранения.

Индикатор статуса МХ отображает доступность модуля:

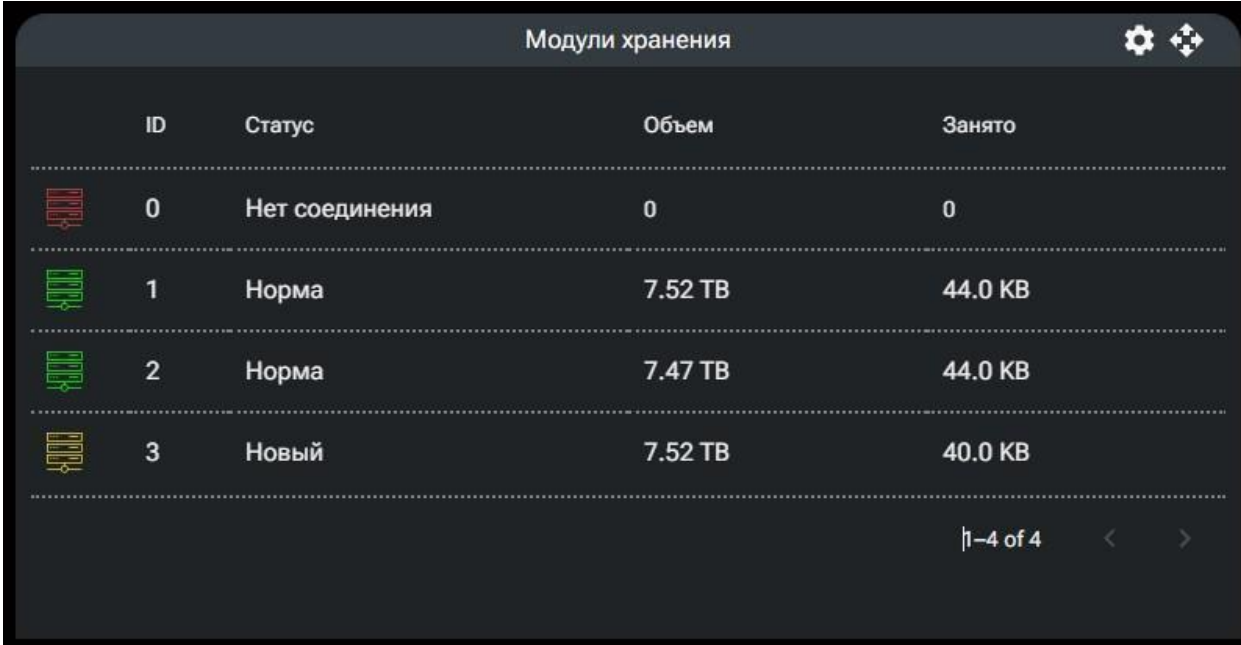
- **зеленый** – доступен;
- **красный** – не отвечает;
- **желтый** – новый (можно подключить к кластеру).

Столбец «ID» показывает идентификатор модуля хранения;

Столбец «Статус» показывает статус работы модуля:

- **Норма** – модуль работает в штатном режиме;
- **Новый** – модуль не включен в состав кластера;
- **Нет соединения** – модуль отключен;
- **Обновление кластера** – выполняется процесс перераспределения

блоков данных между модулями хранения.



ID	Статус	Объем	Занято
0	Нет соединения	0	0
1	Норма	7.52 TB	44.0 KB
2	Норма	7.47 TB	44.0 KB
3	Новый	7.52 TB	40.0 KB

Рисунок 10. Список модулей хранения кластера СХД

В столбцах «Объем» и «Занято» выводится информация о максимальном объеме хранения и занятом пространстве на модулях.

Дополнительная информация по конкретному модулю хранения может быть получена в окне свойств модуля, для чего нужно нажать ПК мыши на строке выбранного модуля хранения (в текущей версии не реализовано).

2.1.9.2. Меню «Сервис S3»

На вкладке меню «Сервис S3» собрана информация и инструментарий управления ресурсами хранения, предоставляемыми пользователям по протоколу S3. Окно вкладки меню представлено на Рисунок 11.

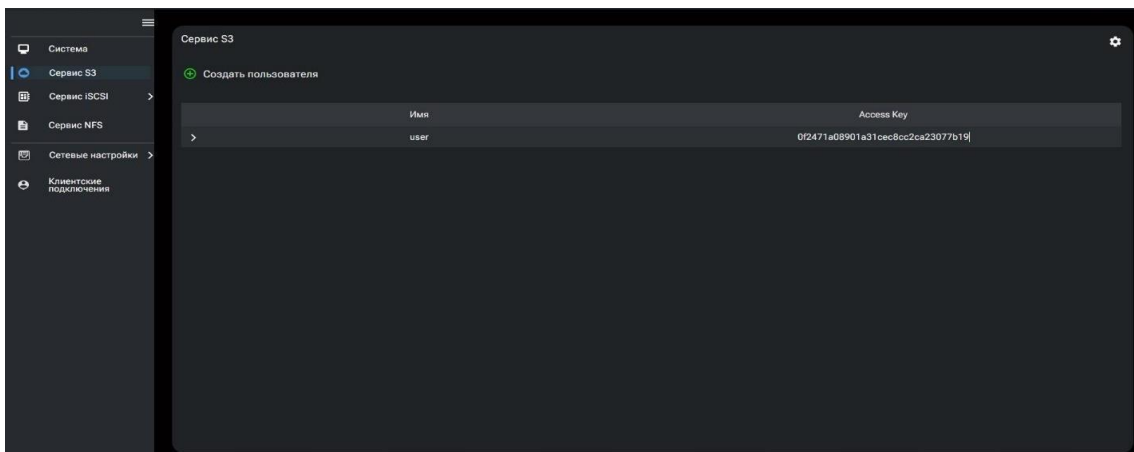


Рисунок 11. Вкладка меню «Сервис S3»

В таблице, размещен список пользователей, которым предоставлен доступ к ресурсам СХД по протоколу S3.

Для получения дополнительной информации о пользователе или его удаления, необходимо развернуть свойства нужного пользователя в списке. В открывшемся окне представлена информация, необходимая для настройки подключения по S3 на стороне клиента и список бакетов, которые созданы для этого пользователя.

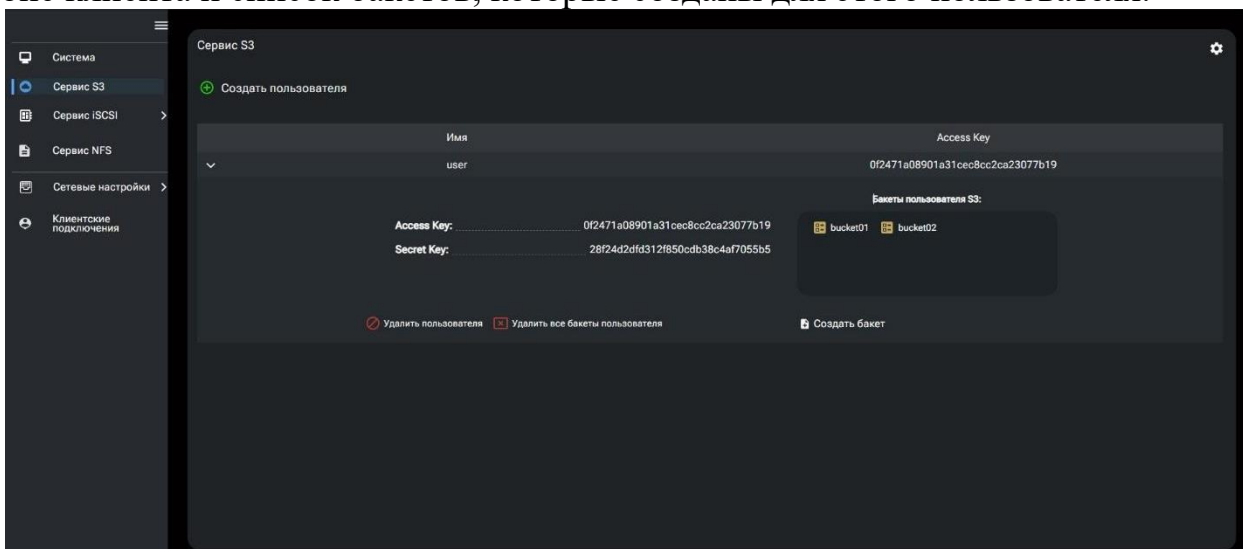


Рисунок 12. Окно свойств пользователя S3

2.1.9.3. Меню «Сетевые настройки»

На вкладке меню «Сетевые настройки» расположены инструменты для работы с сетевыми интерфейсами и настройки ip DNS и NTP сервера.

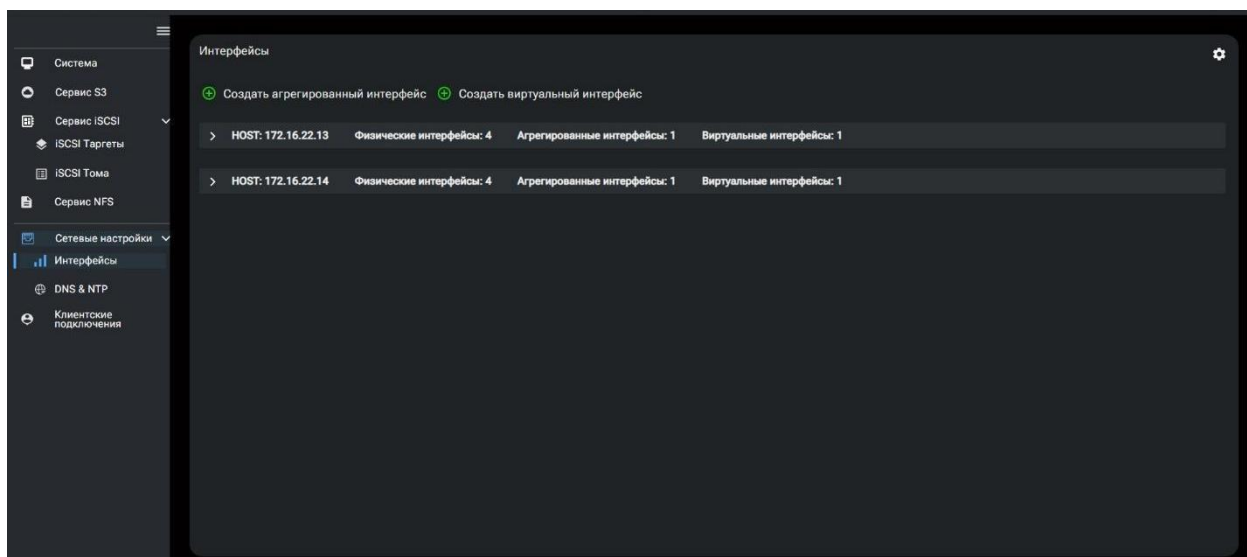


Рисунок 13. Вкладка меню «Сетевые настройки»

В меню «Интерфейсы» расположен список Модулей управления. Для просмотра информации об интерфейсах данного МУ необходимо развернуть вкладку используя вертикальный скролл-бар, как показано на Рисунок 14.

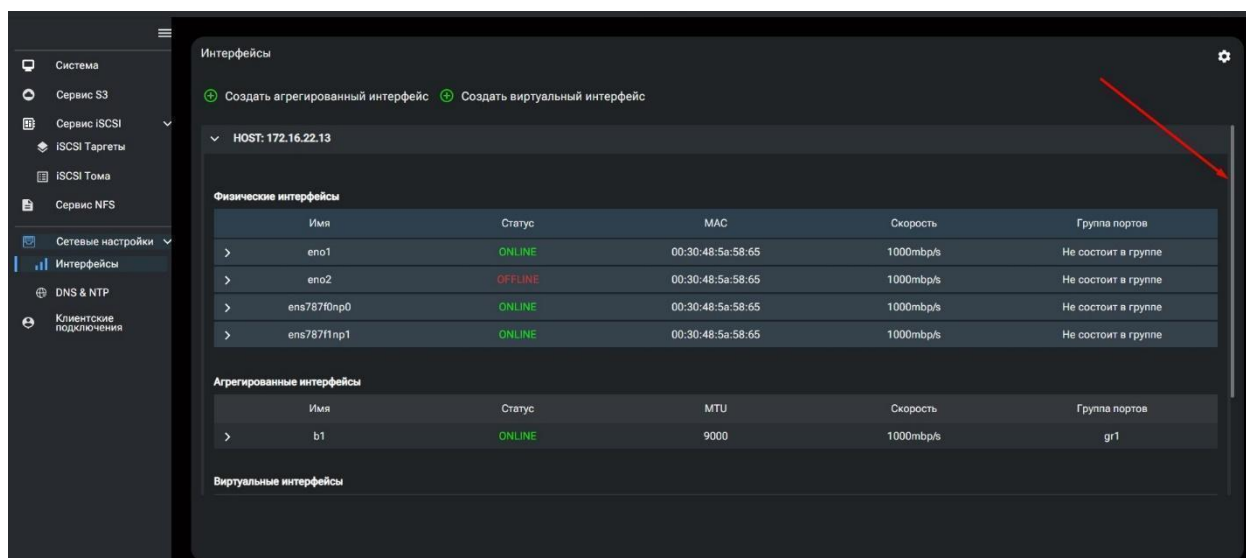


Рисунок 14. Вкладка меню «Интерфейсы»

В списке «Кластерные группы портов» выводятся созданные группы портов. Для редактирования свойств группы портов, сделайте щелчок ЛК мыши на имени группы в списке. В открывшемся окне можно изменить состав группы.

2.1.9.4. Меню «Клиентские подключения»

На вкладке меню «Клиенты» собраны инструменты для управления клиентами и группами доступа к ресурсам. Перед настройкой ресурсов необходимо создать клиентов со списками IP-адресов или IQN имен и добавить их в группы доступа, которые затем будут привязаны к созданным ресурсам, Рисунок 15.

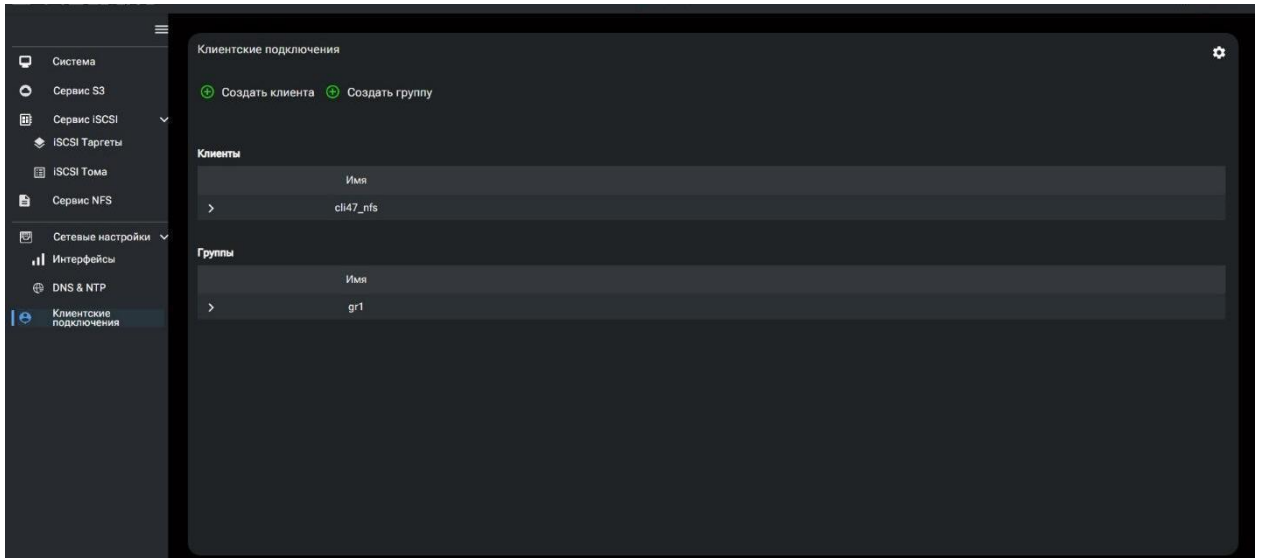


Рисунок 15. Вкладка меню «Клиенты»

2.1.10. Настройка доступа к ресурсу по протоколу S3

S3-протокол — протокол, совместимый с объектным хранилищем Amazon S3 (Simple Storage Service). Протокол позволяет просто хранить объекты, почти так же, как с помощью файловой системы. Основным отличием S3 является возможность хранить большой объём данных без разбиения их на папки.

2.1.10.1. Создание пользователей S3

Настройка доступа по протоколу S3 не требует создания клиента и привязки группы доступа.

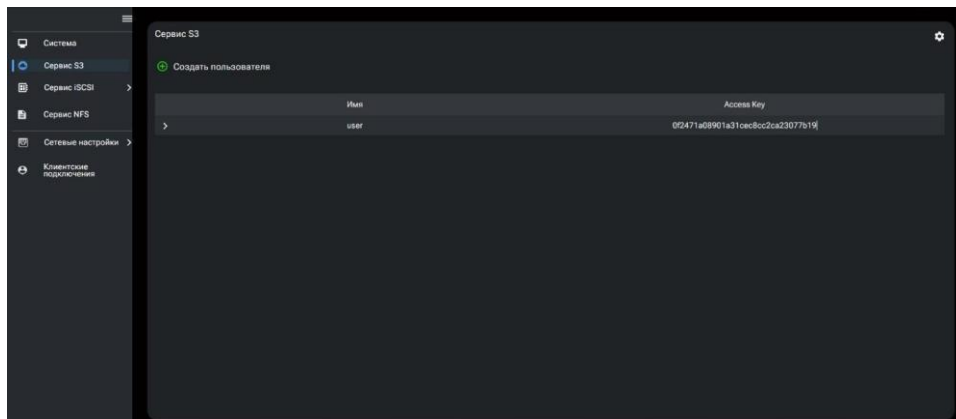


Рисунок 16. Окно вкладки меню «Сервис S3»

Для доступа хранилищу по протоколу S3, необходимо создать пользователя. Создание пользователя выполняется следующим способом:

1. Перейдите на вкладку меню «Сервис S3»;
2. Нажмите на кнопку «Создать пользователя», см. Рисунок 16;
3. В открывшемся окне введите имя пользователя, см. Рисунок 17;

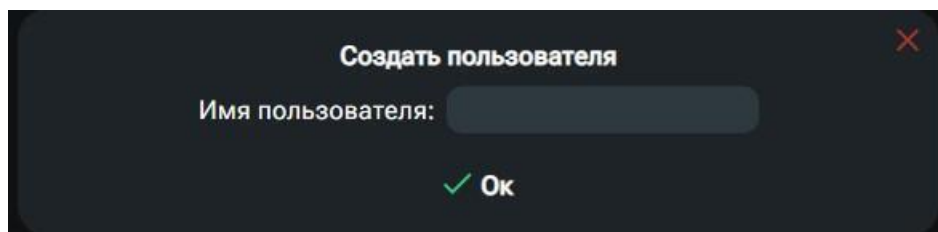


Рисунок 17. Окно создания пользователя S3

4. Нажмите на кнопку «Ок». В списке пользователей добавится имя созданного пользователя.

5. Разверните созданного пользователя в списке. Откроется окно свойств пользователя S3, в полях которой будут выведены автоматически сгенерированные значения «ID доступа» и «Секретный ключ доступа», которые необходимы для настройки клиента, см. Рисунок 18.

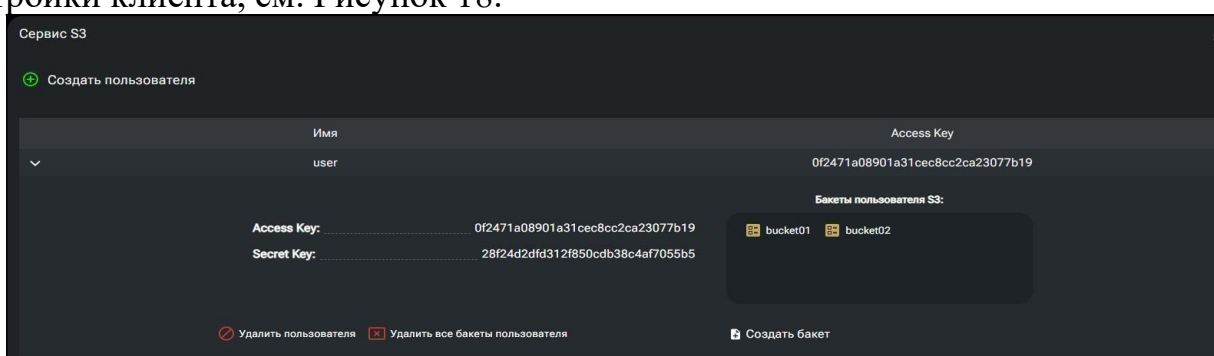


Рисунок 18. Окно редактирования клиента

2.1.10.2. Удаление пользователя S3

Для удаления пользователя выполните следующие действия:

1. На вкладке меню «Сервис S3», откройте окно свойств выбранного пользователя, развернув свойства пользователя в списке пользователей S3;
2. В открывшемся окне свойств выбранного пользователя нажмите на кнопку «Удалить пользователя», см. Рисунок 23. Запись пользователя будет удалена из списка, а в правом нижнем углу экрана появится уведомление об удачном завершении операции.

2.1.10.3. Создание бакета S3

В протоколе S3 объектами хранения являются корзины (buckets). Для создания бакета S3 выполните следующие действия:

1. Перейдите на вкладку меню «Сервис S3»;
2. В списке пользователей разверните свойства нужного пользователя;
3. В Открывшемся окне свойств пользователя нажмите на кнопку «Создать бакет», см. Рисунок 19;

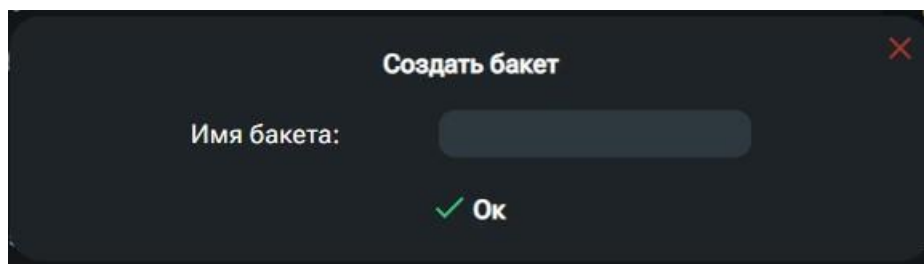


Рисунок 19. Окно создания нового бакета

4. В текстовое поле введите имя создаваемого бакета;
5. Нажмите на кнопку «Ок».

В списке каталогов появится название созданного бакета, а в правом нижнем углу экрана появится уведомление об удачном завершении операции.

2.1.10.4. Удаление бакета S3

Для удаления бакета S3 выполните следующие действия:

1. Перейдите на вкладку меню «Сервис S3»;
2. В списке пользователей выберите пользователя, чей бакет вы планируете удалить и откройте окно свойств;
3. В открывшемся окне выберите в списке нужный бакет и нажмите на нем ЛК мыши;
4. В открывшемся окне свойств S3 бакета нажмите на кнопку «Удалить бакет», см. Рисунок 20;
5. После удаления запись бакета исчезнет из списка, в правом нижнем углу экрана появится уведомление о успешно выполненной операции.

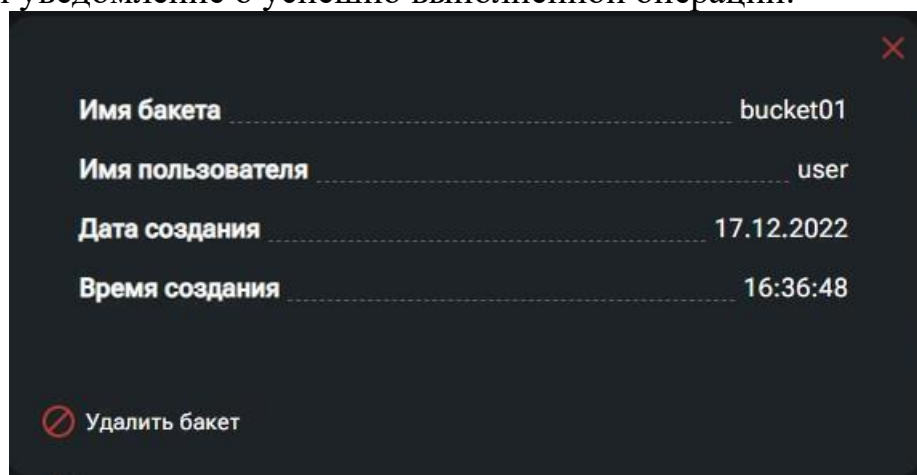


Рисунок 20. Окно свойств бакета S3

2.1.11. Настройка доступа к ресурсу по протоколу NFS

Network File System (NFS) — протокол сетевого доступа к файловым системам, позволяет монтировать удалённые файловые системы через сеть, обеспечивает пользователям доступ к файлам, позволяет работать с этими файлами точно так же, как и с локальными.

Для презентации ресурсов внешним клиентам по протоколу NFS, необходимо следующее:

- создать группу доступа, в которой указать IP адрес (адреса) компьютеров, которым разрешено подключаться к созданному тому;
- создать том NFS;
- привязать группу доступа к созданному тому.

Опубликованный на СХД том NFS после подключения к клиентскому компьютеру будет отображаться как сетевая папка.

2.1.11.1. Создание и публикация тома NFS

Для создания тома NFS выполните следующие действия:

1. Перейдите на вкладку меню «Сервис NFS»;
2. Нажмите на кнопку «Создать NFS том», откроется окно создания NFS тома, см. Рисунок 21;

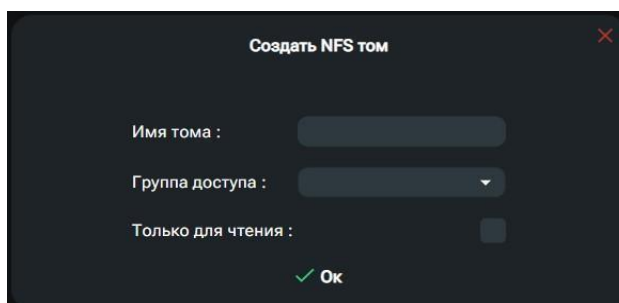


Рисунок 21. Окно создания тома данных NFS

3. Введите имя тома в поле «Имя тома»;
4. Выберите, созданную ранее, группу доступа из списка «Группа доступа» (привязку группы доступа позже можно изменить);
5. Отметьте, если это необходимо, флажок «Только чтение» (можно изменить позже);
6. Нажмите на кнопку «Ок».

После выполнения процедуры в списке опубликованных томов отобразится имя созданного тома, и система сообщит об удачном завершении операции, см. Рисунок 22.

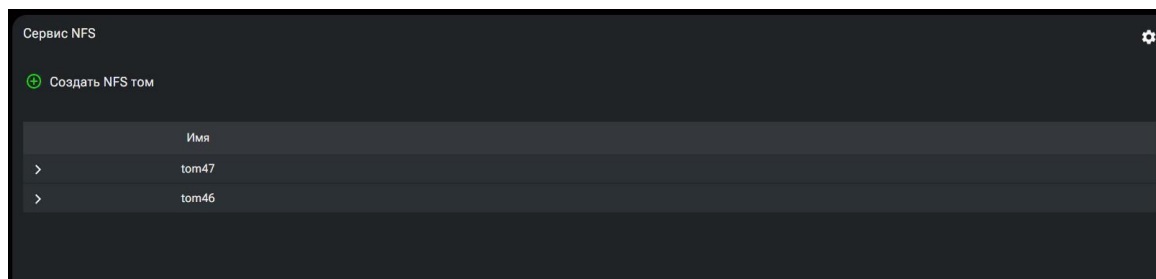


Рисунок 22. Список созданных томов NFS

2.1.11.2. Удаление тома NFS

Для удаления тома NFS выполните следующие действия:

1. Перейдите на вкладку меню «Сервис NFS»;
2. Разверните свойства выбранного тома для его удаления в списке опубликованных томов NFS;
3. В открывшемся окне свойств тома нажмите на кнопку «Удалить том» и подтвердите удаление нажав кнопку «Да», см. Рисунок 23.

После выполнения процедуры в списке опубликованных томов исчезнет имя удаленного тома, и система сообщит об удачном завершении операции.

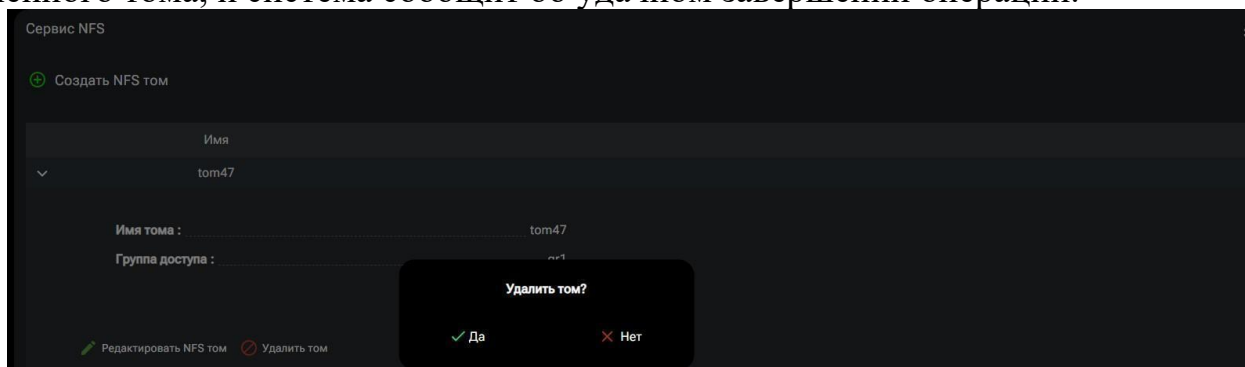


Рисунок 23. Удаление тома NFS

2.1.11.3. Редактирование доступа к тому NFS

Для изменения привязки группы доступа к тому NFS выполните следующие действия:

1. Перейдите на вкладку меню «Сервис NFS»;
2. Разверните выбранный том в списке опубликованных томов NFS для его редактирования;
3. В открывшемся окне свойств тома, нажмите на «редактировать NFS том» и выберите нужную «Группу доступа»;
4. Нажмите на кнопку «Ок», см. Рисунок 24.

После выполнения процедуры привязка группы доступа будет изменена на выбранную, и система сообщит об удачном завершении операции.

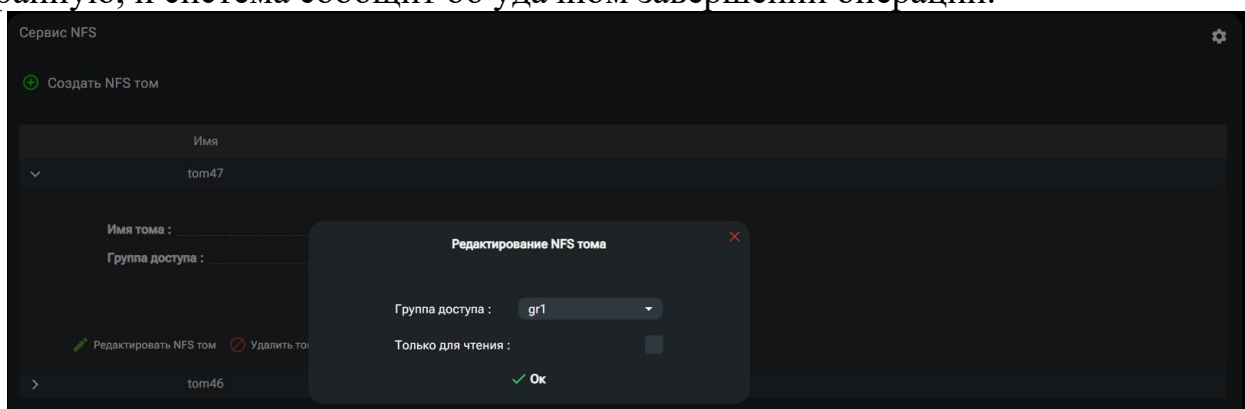


Рисунок 24. Редактирование NFS тома

2.1.11.4. Изменение режима «Только чтение» для тома NFS

Изменение режима чтение-запись на только чтение или обратно, выполняется следующим способом:

1. Перейдите на вкладку меню «Сервис NFS»;
2. В списке опубликованных томов NFS разверните выбранный том и нажмите «Редактировать NFS том»;
3. В открывшемся окне свойств тома, измените состояние флажка «Только для чтения» и нажмите на кнопку «Ок», см. Рисунок 25.

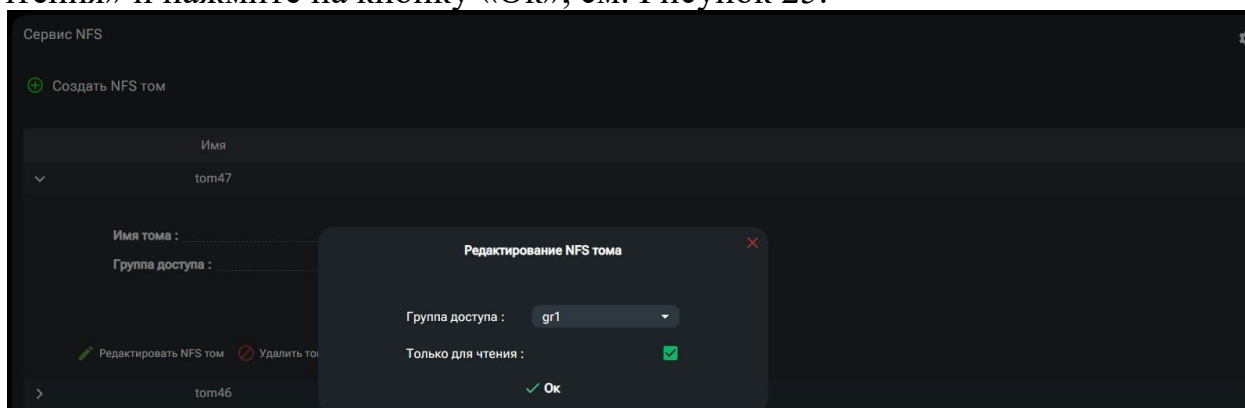


Рисунок 25. Флажок изменение режима чтение-запись тома NFS

После выполнения процедуры режим чтения-записи будет изменен, и система сообщит об удачном завершении операции.

2.1.12. Настройка клиентов и групп доступа

Настройка доступа к ресурсам СХД требует создания IP-списков адресов компьютеров, которым будет разрешен доступ к опубликованным ресурсам. Списки доступа к ресурсам реализованы следующим способом: для каждого компьютера-клиента создается запись на СХД – так называемый клиент.

Клиент может хранить информацию о IP-адресе компьютера.

Клиенты собираются в группы доступа, которые привязываются к ресурсам.

2.1.12.1. Создание клиента

Для создания нового клиента для доступа к ресурсам выполните следующие действия:

1. Перейдите на вкладку меню «Клиентские подключения», см. Рисунок 37;
2. Нажмите на кнопку «Создать клиента»;
3. В открывшемся окне введите имя клиента, IP-адреса компьютеров, которым будет разрешен доступ и таргет IQN (добавление IP-адреса и IQN производится с помощью «+»), см. Рисунок 26;
4. Нажмите на кнопку «Создать».

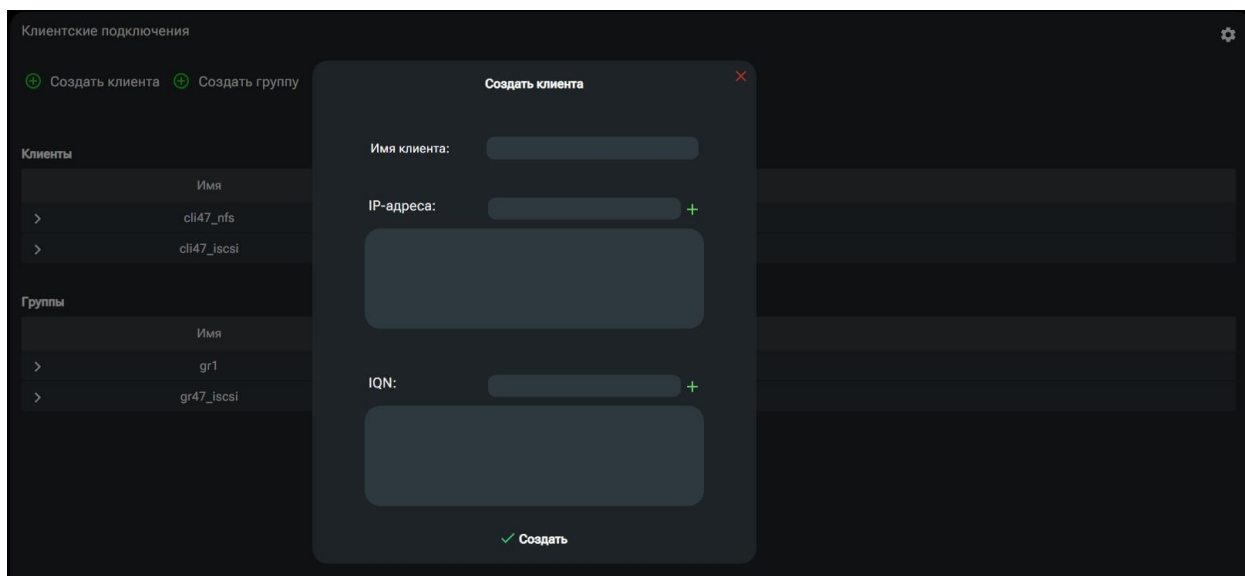


Рисунок 26. Окно создания нового клиента

После успешного выполнения процедуры создания клиента система сообщит об удачном завершении операции.

Разрешено вводить как несколько IP-адресов, так и несколько имен IQN собирая в одном клиенте данные о нескольких подключениях. Введенные данные будут отражены в окне в виде списка.

2.1.12.2. Редактирование параметров клиента

Для изменения списка доступа, записанного в клиенте необходимо выполнить редактирование его параметров. Для этого выполните следующие действия:

1. Перейдите на вкладку меню «Клиентские подключения»;
2. В списке «Клиенты», разверните свойства нужного клиента;
3. В свойствах данного клиента нажмите кнопку «Редактирование клиента»
4. В открывшемся окне, добавьте (+) либо удалите (x) IP-адреса компьютеров, которым будет разрешен доступ.
5. Нажмите на кнопку «Редактировать клиента».

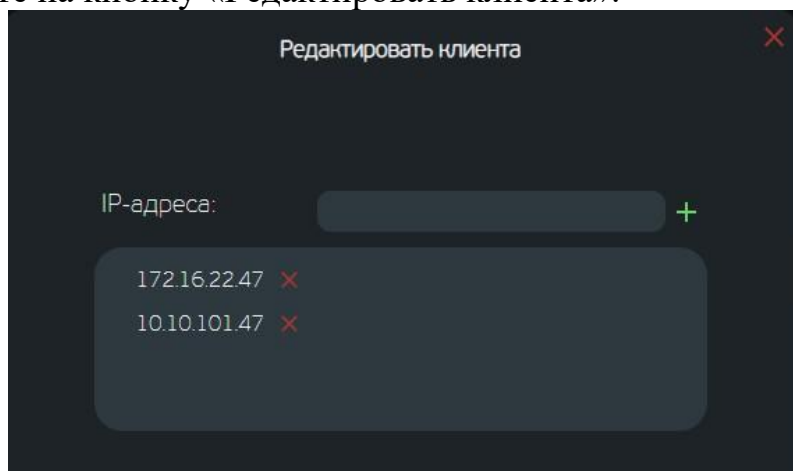


Рисунок 27. Окно редактирования клиента.

После успешного выполнения процедуры редактирования параметров клиента система сообщит об удачном завершении операции.

2.1.12.3. Удаление клиента

Удаление клиента прекратит доступ к ресурсу внешних хостов, которые были указаны в списке доступа этого клиента. Для удаления клиента выполните следующие действия:

1. Перейдите на вкладку меню «Клиентские подключения»;
2. В списке «Клиенты», выберите нужного клиента и разверните его свойства;
3. В открывшемся окне свойств клиента нажмите на кнопку «Удалить клиента», см. Рисунок 28;

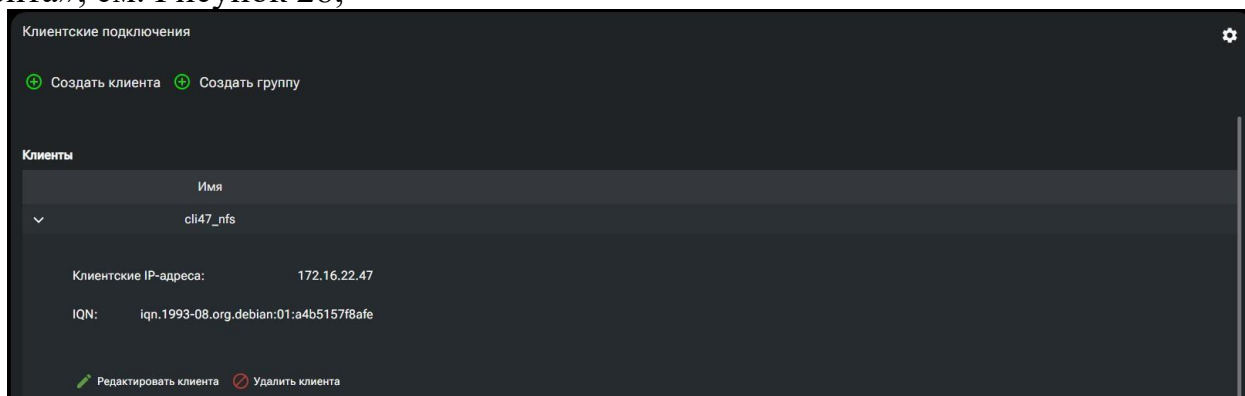


Рисунок 28. Окно свойств клиента.

После успешного выполнения процедуры удаления клиента, имя клиента исчезнет из списка, и система сообщит об удачном завершении операции.

2.1.12.4. Создание группы доступа

Группы доступа предназначены для группировки клиентов и удобного управления разрешениями на доступ к ресурсам СХД.

Для создания группы доступа выполните следующие действия:

1. Перейдите на вкладку меню «Клиентские подключения».
2. Нажмите на кнопку «Создать группу».
3. В открывшемся окне введите имя группы в поле «Имя»;
4. Отметьте флажками тех клиентов в списке, которых вы хотите включить в состав группы, см. Рисунок 29;
5. Нажмите на кнопку «Создать».

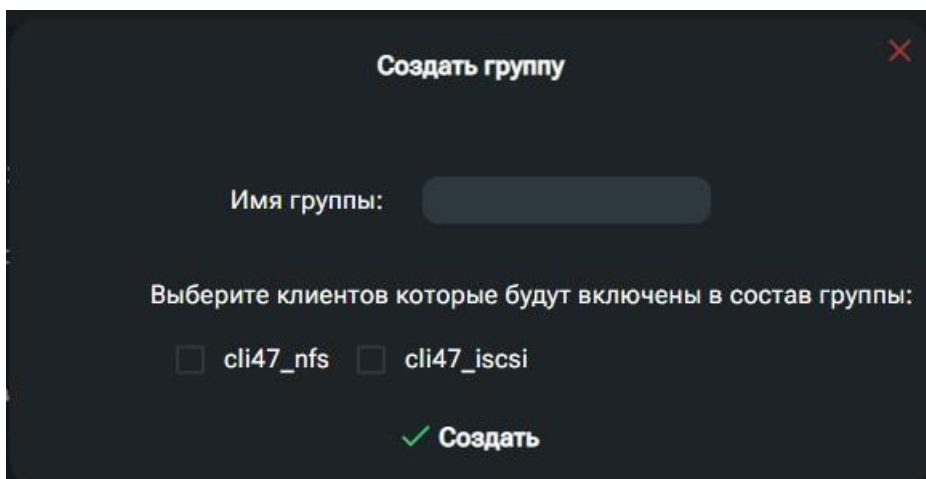


Рисунок 29. Окно создания группы доступа

После успешного выполнения процедуры создания группы доступа имя появится в списке групп, и система сообщит об удачном завершении операции.

2.1.12.5. Редактирование списка группы

Изменять состав группы доступа можно и после создания группы. Для этого выполните следующее:

1. Перейдите на вкладку меню «Клиентские подключения»;
2. В области «Группы доступа», выберите в списке нужную группу доступа и разверните её свойства;
3. Нажмите на кнопку «Редактировать группу»
4. В открывшемся окне отметьте флажками только тех клиентов, которых вы хотите включить в состав группы, см. Рисунок 30;
5. Нажмите на кнопку «Ок».

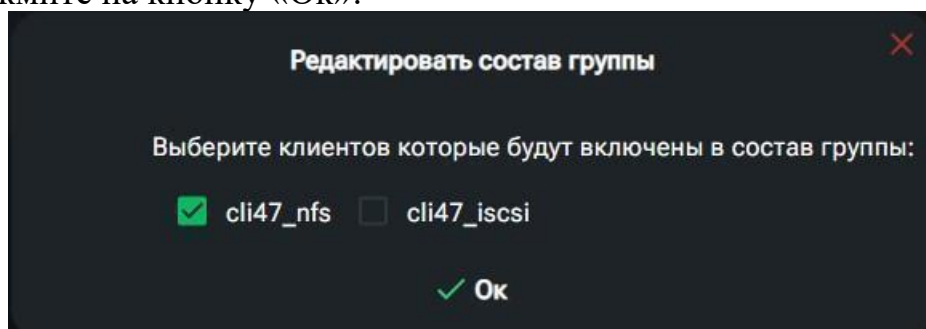


Рисунок 30. Редактирование состава группы доступа

После успешного выполнения процедуры будет сохранен новый состав группы доступа, и система сообщит об удачном завершении операции.

2.1.12.6. Удаление группы доступа

Удаление группы доступа не повлечет за собой удаление клиентов. Для удаления группы доступа выполните следующие действия:

1. Перейдите на вкладку меню «Клиентские подключения»;
2. В списке «Группы доступа», выберите нужную группу и разверните ее

свойства;

3. В открывшемся окне свойств группы нажмите на кнопку «Удалить группу».

После успешного выполнения процедуры, группа доступа будет удалена и исчезнет из списка, а система сообщит об удачном завершении операции.

2.1.13. Настройка Сети

Настройки сетевых интерфейсов модулей управления выполняются в меню «Сетевые настройки» - «Интерфейсы». см. Рисунок 31.

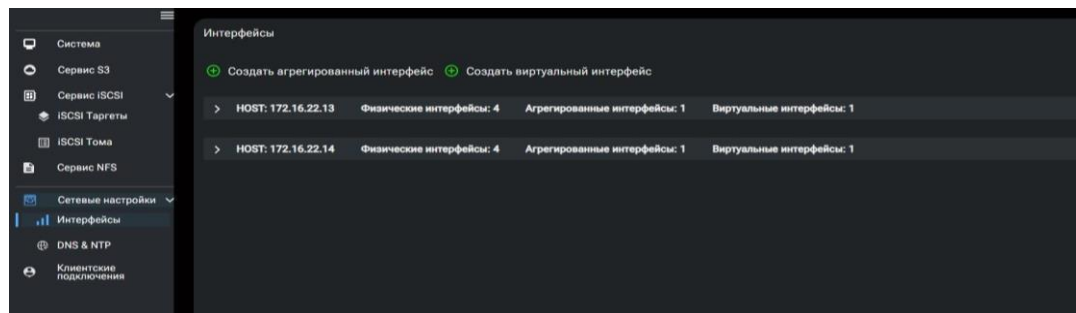


Рисунок 31. Меню «Интерфейсы»

2.1.13.1. Создание агрегированного интерфейса

Создание агрегированного интерфейса, необходимо для увеличения пропускной способности канала передачи данных либо повышения его отказоустойчивости. Для этого используется протокол управления агрегацией каналов (LACP), предоставляющий методы управления объединением нескольких физических портов в единый логический канал. LACP позволяет сетевому устройству согласовывать автоматическое объединение ссылок, отправляя пакеты LACP одноранговому узлу (напрямую подключенному устройству, которое также реализует LACP).

Для создания нового агрегированного интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Нажмите на кнопку «Создать агрегированный интерфейс»;
3. В открывшемся окне добавления агрегированного интерфейса (см.

Рисунок 32), введите:

- имя создаваемого группового интерфейса;
- значение MTU;
- выберите хост
- тип балансировки

4. Отметьте флажками физические сетевые интерфейсы, выбранные для работы в составе агрегированного интерфейса;

5. Нажмите на кнопку «Создать»

После успешного выполнения процедуры, в списке выбранного хоста появится новый агрегированный интерфейс, и система сообщит об удачном завершении операции.

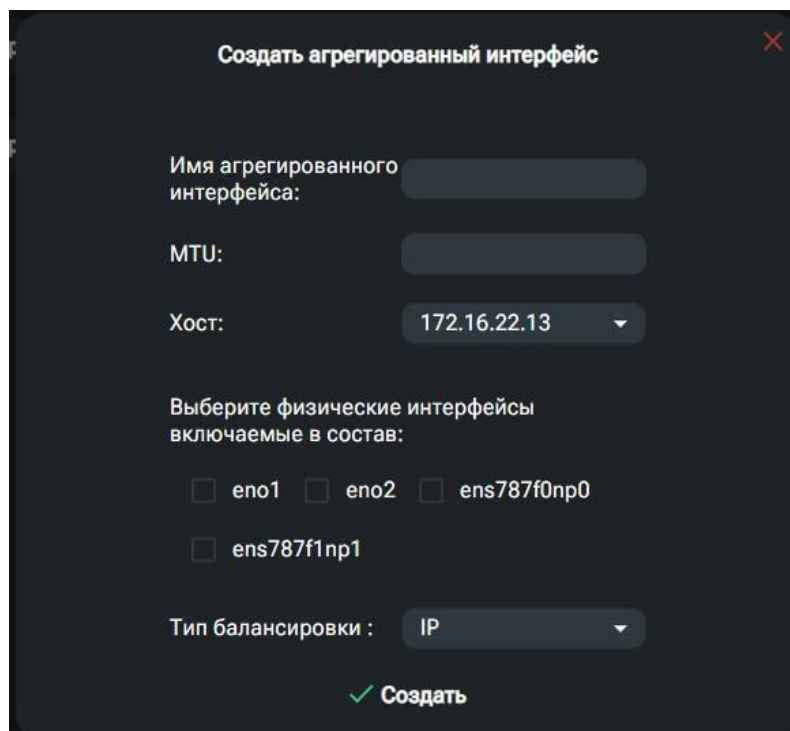


Рисунок 32. Окно добавления агрегированного интерфейса

2.1.13.2. Редактирование агрегированного интерфейса

Состав созданного агрегированного интерфейса, MTU и тип балансировки нагрузки можно впоследствии изменить. Для изменения свойств группового интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Разверните свойства хоста, на котором создан агрегированный интерфейс;
3. В таблице выберите созданный агрегированный интерфейс и разверните его свойства;
4. Нажмите на кнопку «Редактировать интерфейс»; см. Рисунок 33
5. Измените требуемые параметры и нажмите на кнопку «Ок». см. Рисунок

34

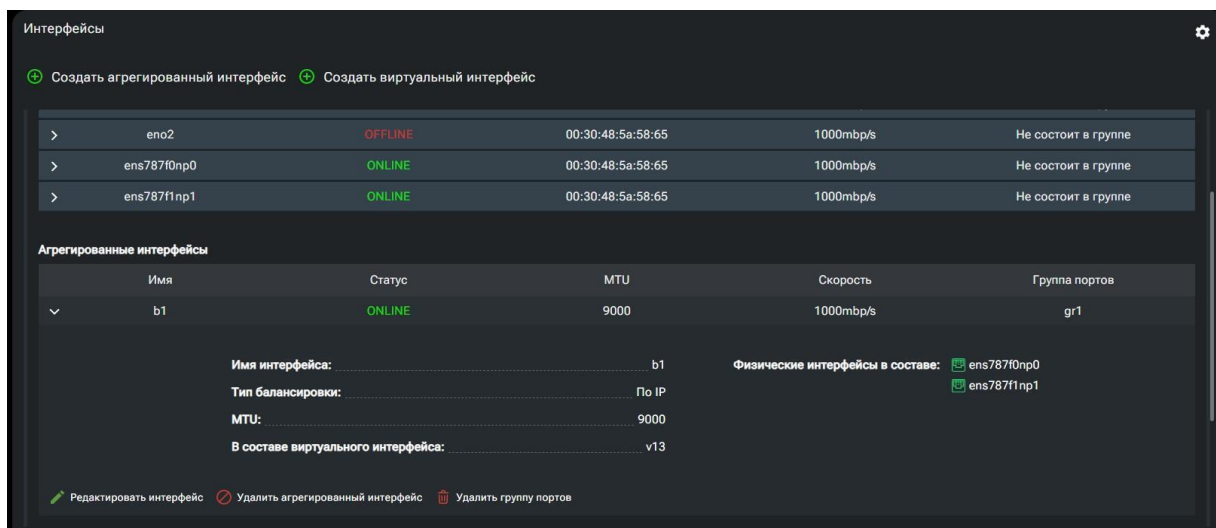


Рисунок 33. Окно свойств агрегированного интерфейса

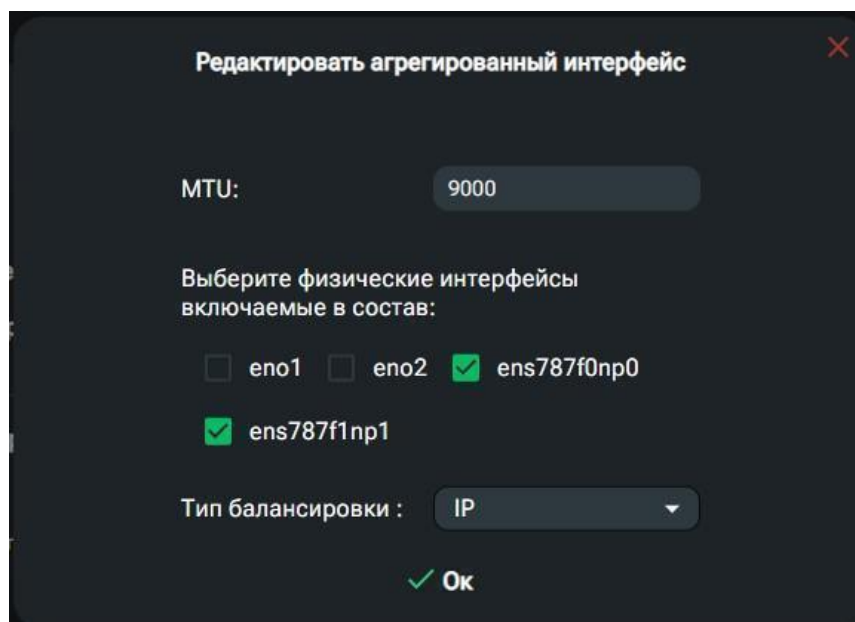


Рисунок 34. Окно редактирования агрегированного интерфейса

После успешного выполнения процедуры, свойства агрегированного интерфейса изменятся на выбранные и система сообщит об удачном завершении операции.

2.1.13.3. Удаление агрегированного интерфейса

Для удаления группового интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Разверните свойства необходимого хоста;
3. В таблице агрегированных интерфейсов выберите нужный и разверните его свойства;
4. В открывшемся окне свойств агрегированного интерфейса нажмите «Удалить агрегированный интерфейс», см. Рисунок 33,
5. Подтвердите удаление нажав кнопку «Да»

После успешного выполнения процедуры в списке исчезнет групповой сетевой интерфейс, и система сообщит об удачном завершении операции.

2.1.13.4. Создание виртуального интерфейса

Виртуальный интерфейс — это сетевой интерфейс, который создается поверх физического интерфейса и предоставляет те же возможности, что и физический интерфейс.

Для создание виртуального интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Нажмите на кнопку «Создать виртуальный интерфейс», см. Рисунок 31
3. В открывшемся окне добавления виртуального интерфейса (см. Рисунок 35), введите:

- имя создаваемого виртуального интерфейса;
- значение MTU;
- выберите хост
- интерфейс, на котором будет создан виртуальный интерфейс;
- IP-адрес;
- маска подсети;
- роутер (шлюз) по умолчанию;
- номер VLAN.

4. Нажмите на кнопку «Создать»

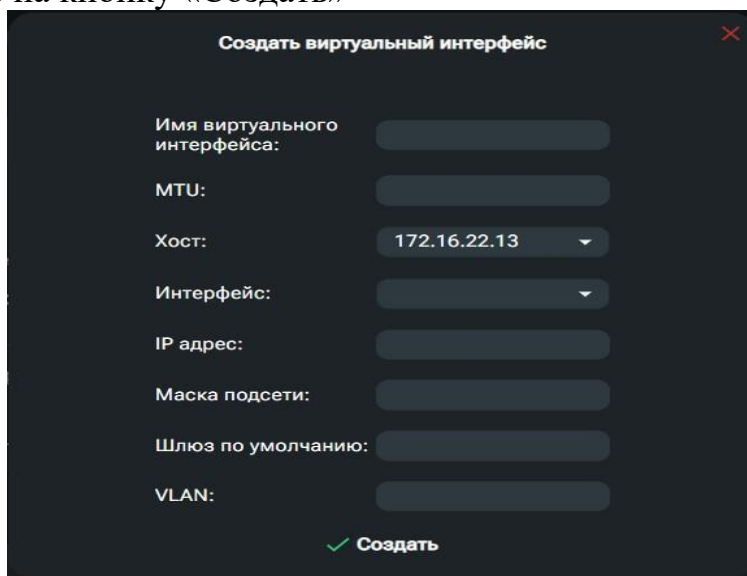


Рисунок 35. Окно создания виртуального сетевого интерфейса.

После успешного выполнения процедуры система сообщит об удачном завершении операции новый интерфейс отобразится в списке виртуальных интерфейсов.

2.1.13.5. Редактирование виртуального интерфейса

Состав созданного виртуального интерфейса можно в последствии изменить. Для изменения свойств виртуального интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Разверните свойства хоста, на котором создан виртуальный интерфейс;
3. В таблице выберите созданный виртуальный интерфейс и разверните его свойства;
4. Нажмите на кнопку «Редактировать интерфейс»; см. Рисунок 37
5. В открывшемся окне редактирования виртуального интерфейса, см. Рисунок 36, разрешено изменять следующие поля:

- значение MTU;
- интерфейс
- IP-адрес;
- маска подсети;
- роутер (шлюз) по умолчанию;
- номер VLAN.

6. После изменения значений нужных полей, нажмите на кнопку «Ок».

Рисунок 36. Редактирование свойств виртуального интерфейса

Имя	Статус	MTU	Скорость	IP
v13	ONLINE	9000	1000mb/s	10.10.101.13

Рисунок 37. Окно свойств виртуального интерфейса

После успешного выполнения процедуры, новые настройки виртуального интерфейса сохраняются, и система сообщит об удачном завершении операции.

2.1.13.6. Удаление виртуального интерфейса

Для удаления виртуального интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Разверните свойства нужного хоста;
3. В открывшемся окне выберите нужный виртуальный интерфейс и разверните его свойства;
4. В открывшемся окне нажмите кнопку «Удалить виртуальный интерфейс», см. Рисунок 37;
5. Подтвердите удаление нажав кнопку «Да».

После успешного выполнения процедуры, в списке исчезнет виртуальный сетевой интерфейс, и система сообщит об удачном завершении операции.

2.1.13.7. Подключение к серверу доменных имен и серверам времени

Для правильной работы кластера СХД в составе сетевой инфраструктуры, используется технология DNS - Domain Name System, при помощи которой имена сетевых узлов сопоставляются IP-адресам этих узлов. При настройке СХД, требуется указать используемые в сети адреса основного и резервного сервера доменных имен (DNS сервера).

Для синхронизации времени узлов СХД как между собой, так и с клиентскими хостами, требуется внешний сервер времени (NTP сервер), адрес которого необходимо указать при начальной настройке СХД.

Настройка подключения к серверам DNS и NTP выполняется следующим способом:

1. Перейдите на вкладку меню «Сетевые настройки» - «DNS & NTP», см. Рисунок 38;
2. В открывшемся окне настроек, введите IP-адреса основного и резервного серверов доменных имен и NTP в соответствующие;
3. Нажмите на кнопку «Сохранить настройки».

После успешного выполнения процедуры и система сообщит об удачном завершении операции.

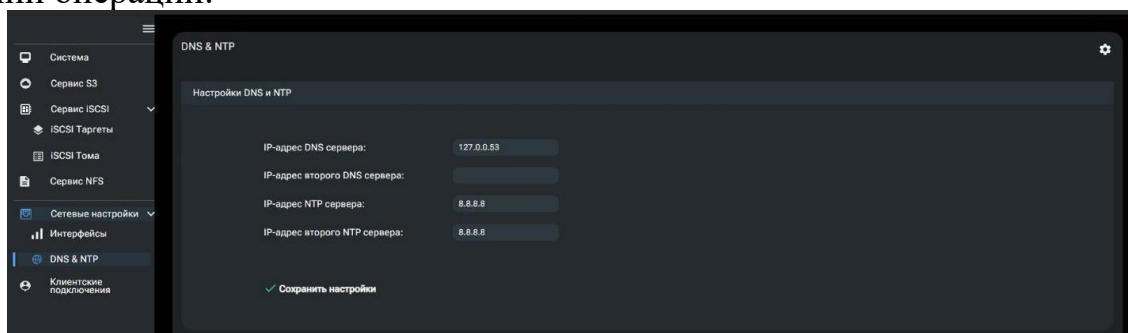


Рисунок 38. Настройка соединения с серверами DNS и NTP

2.1.13.8. Создание кластерной группы портов

Для поддержания отказоустойчивого доступа к ресурсам, в СХД реализован алгоритм переключения сетевых адресов отказавшего модуля управления (МУ) на модуль, состоящий с ним в паре. Для того, чтобы система знала на какой сетевой порт нужно перенести ip-адрес с отказавшего порта, эти порты нужно сгруппировать. Порты группируются попарно – порт с первого МУ, порт со второго МУ, образуя кластерную группу портов.

Для создания кластерной группы портов выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Разверните свойства нужного хоста;
3. Выберите нужный интерфейс
4. Нажмите на кнопку «Создать группу портов», см. Рисунок 39
5. В открывшемся окне введите имя группы и выберите порты со второго контроллера, которые нужно объединить в кластерную группу, см. Рисунок 40
6. Нажмите на кнопку «Создать».

После успешного выполнения процедуры и система сообщит об удачном завершении операции, и в колонке Группа портов на интерфейсе добавится имя созданной группы.

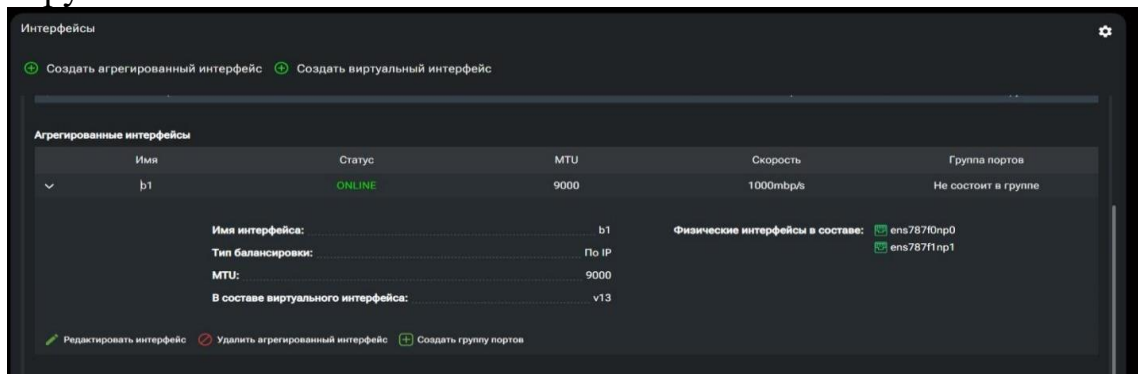


Рисунок 39. Окно группы портов

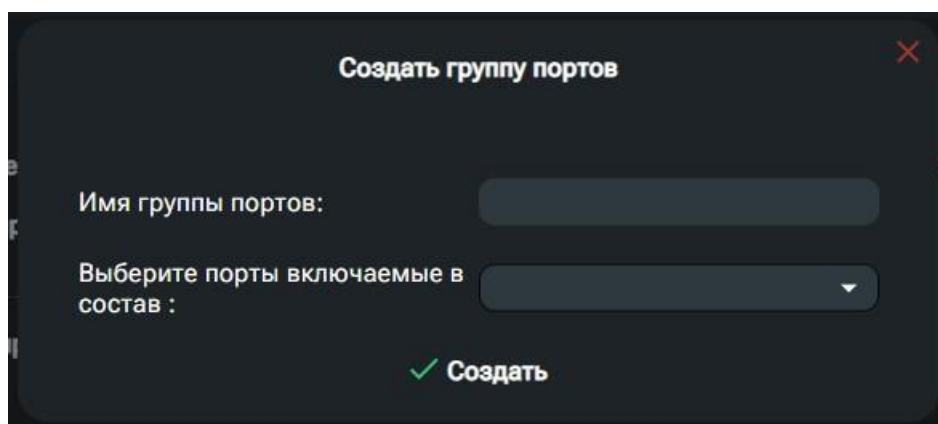


Рисунок 40. Окно создания группы портов

2.1.13.9. Удаление кластерной группы портов

Для удаления группы портов, выполните следующие действия:

1. Перейдите на вкладку меню «Сетевые настройки» - «Интерфейсы»;
2. Разверните свойства нужного хоста;
3. Выберите и разверните нужный интерфейс
4. Нажмите на кнопку «Удалить группу портов».

После успешного выполнения процедуры и система сообщит об удачном завершении операции.

2.2. Функционал СХД (Аппаратное обеспечение двухконтроллерной СХД)

2.2.1. Первоначальная настройка

2.2.1.1. Вход в систему

Для входа в систему в адресной строке браузера, запущенного на компьютере администратора, необходимо ввести IP-адрес управляющего интерфейса СХД, заданный при инсталляции системы.

Для шифрования трафика между компьютером администратора и web интерфейсом используется самоподписанный ssl сертификат. В текущей версии ПО у пользователя нет возможности самому заменить ssl сертификат. Замена сертификата выполняется сервисным инженером после обращения в службу поддержки.

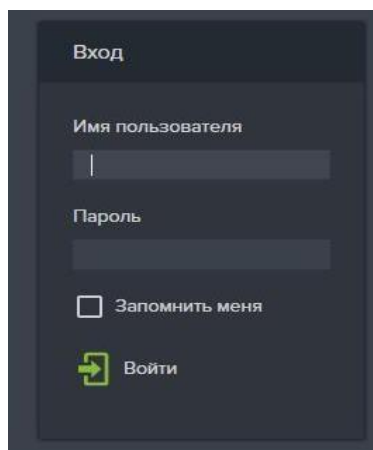


Рисунок 41. Окно входа в систему.

После подключения появится окно входа в систему (см. Рисунок 41), в котором требуется ввести логин и пароль.

Заводские настройки:

- Логин: **admin**

- Пароль: **123456**

По требованиям безопасности, требуется сменить пароль.

Если вход успешно выполнен, на экране появится главная страница интерфейса управления. (См. Рисунок 42)



Внимание! Все настройки ПО BAUM STORAGE AI v2 синхронизируются в кластере. При отказе одного из серверов кластера работа комплекса не прерывается.

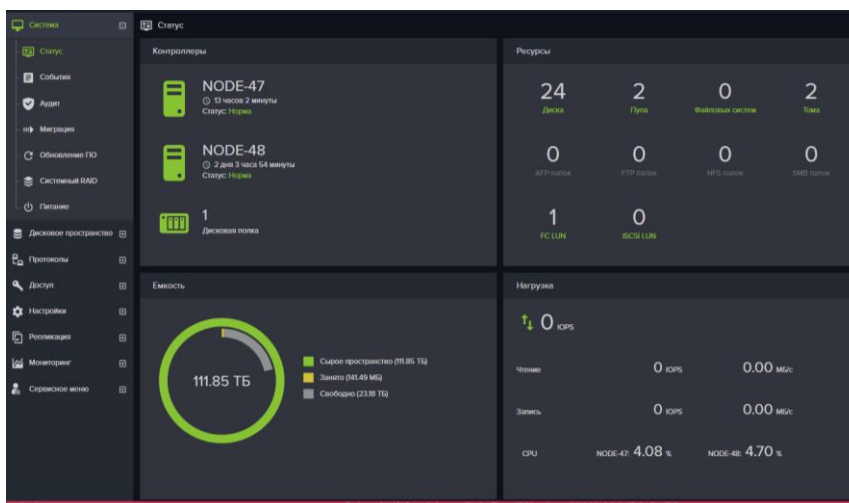


Рисунок 42. Главная страница интерфейса управления

2.2.1.2. Описание структуры интерфейса управления

В левой части главного окна интерфейса управления расположены заголовки разделов основного меню (см. Рисунок 42).

В правой части окна отражены сведения о просматриваемых разделах, разделенные на соответствующие поля информации и настроек.

Вверху окна горизонтально расположена панель уведомлений, в которой отображаются значки индикаторов статуса работы компонентов системы: контроллеров, логирования, системного RAID, дисков, аппаратного обеспечения. Там же, при наступлении соответствующих событий, появляются значки уведомлений о запуске заданий репликации и перестроения пулов (см. Рисунок 43).



Рисунок 43. Панель уведомлений

В рабочей системе значки индикаторов отображаются зеленым цветом. Изменение цвета на желтый, сигнализирует о ошибках в работе данного компонента. Красный цвет сигнализирует о серьезной проблеме. Цифра под значком показывает количество проблем.

Щелчок мышкой на значке индикатора отобразит информационную панель, на которой будет указана причина ошибки.

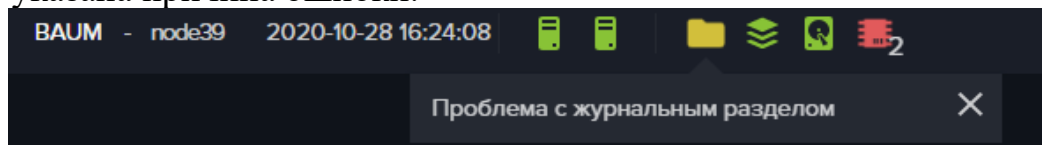


Рисунок 44. Информационная панель

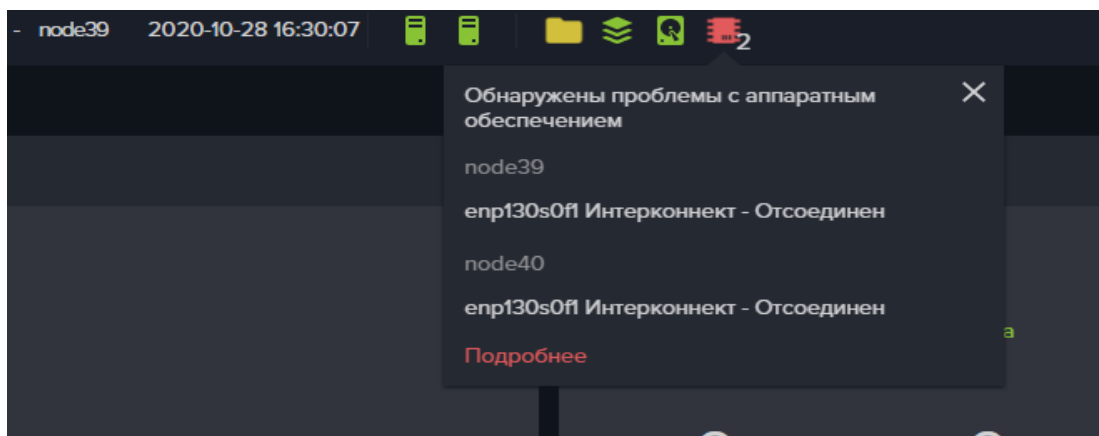


Рисунок 45. Пример сообщений в информационной панели

Нажатие на кнопку «**Подробнее**», которая расположена внизу информационной панели, открывает меню мониторинга, в котором можно увидеть дополнительную информацию.

Справа на панели уведомлений находятся значки входа в меню выбора языка (русский, английский) и настроек системы, в которых указывается текущий пользователь, его уровень прав, и значок выхода из системы. Нажатие на пункт меню «**О программе**» откроет окно с информацией о версии ПО.

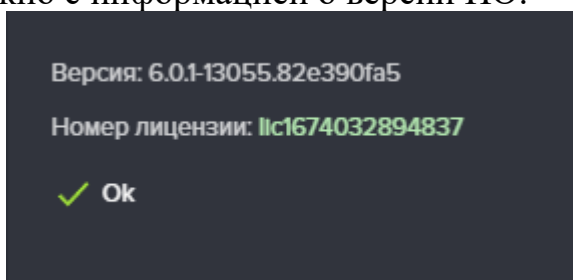



Рисунок 46. Информация о версии ПО

В нижней части окна, расположена панель оперативного журнала, по умолчанию свернутая в полосу. Слева на ней размещены стрелки для её раскрытия и закрытия , нажав на которые можно раскрыть или свернуть панель, и изображение ластика, нажатие на который стирает сообщение из заголовка окна. В зависимости от наличия важных сообщений, цвет панели меняется с зеленой на красную. Текст сообщения выводится в заголовок панели.

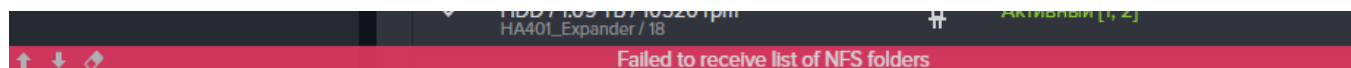


Рисунок 47. Панель оперативного журнала

Центральная часть окна «Статус» разделена на области.

В области «**Контроллеры**» отображаются имена контроллеров (узлов кластера), статусы работы контроллеров), время работы каждого контроллера и количество дисковых полок.

При смене статуса изменяется и цвет значка контроллера. Возможны следующие статусы работы контроллеров:

- **Норма** (зеленый цвет);
- **Недоступен** (красный цвет);
- **Отдал ресурсы** (желтый цвет);
- **Состояние неизвестно** (серый цвет).

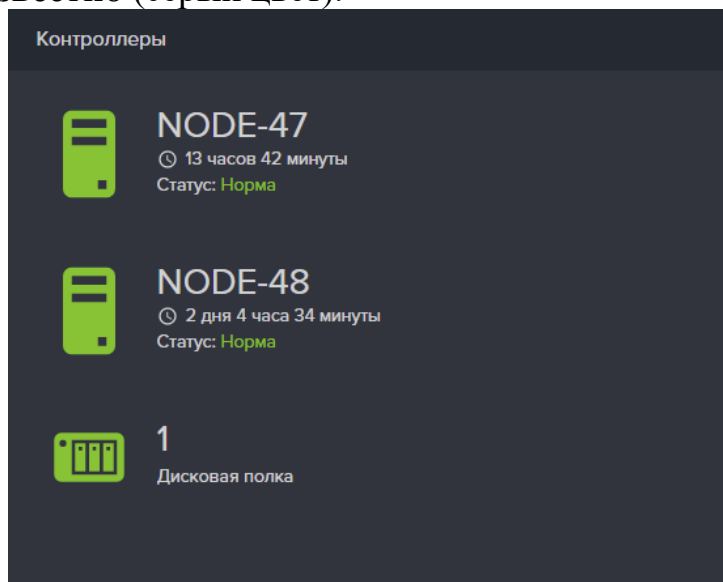


Рисунок 48. Область «Контроллеры»

В области «**Ресурсы**» отображаются сведения о ресурсах кластера.

Сюда выводится информация о количестве дисков, установленных в системе, количестве созданных дисковых пулов, томов, файловых систем и ресурсов, розданных по различным протоколам.

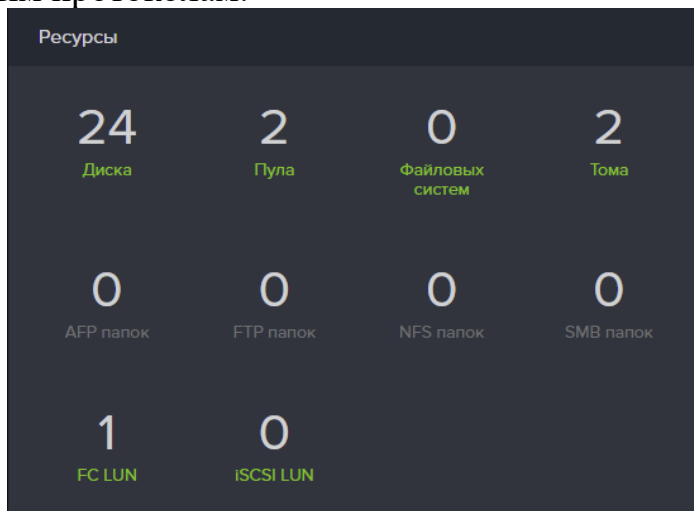


Рисунок 49. Область «Ресурсы»

В области «**Емкость**» представлены сведения о суммарной (сырой) емкости всех установленных в системе дисков, занятого и свободного объёма.

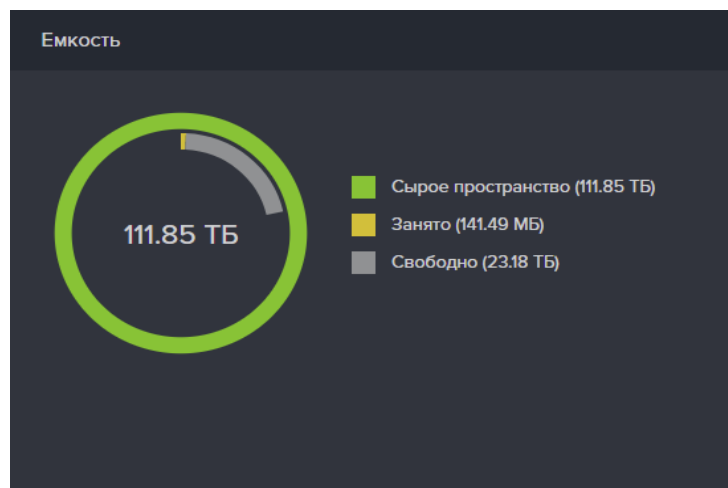


Рисунок 50. Область «Емкость»

В области «**Нагрузка**» выводится информация о суммарной нагрузке на систему (в iops), а также отдельно для операций чтения и записи (в iops и МБ/с. Информация представлена только для блочных протоколов.

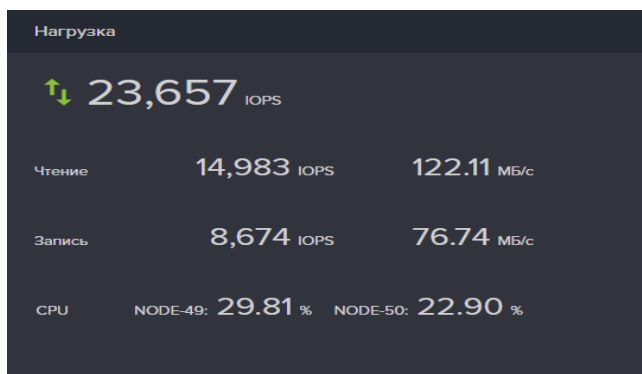


Рисунок 51. Область «Нагрузка»

2.2.1.3. Изменение IP-адреса интерфейса управления контроллера

Иногда в процессе эксплуатации системы возникает необходимость изменить IP-адрес на управляющем интерфейсе. Для назначения управляющему интерфейсу нового IP-адреса потребуется второй свободный интерфейс. Если в системе отсутствует второй сетевой интерфейс. Все действия должны выполняться из интерфейса соседнего контроллера, во избежание потери доступа к интерфейсу управления после смены адреса.

Выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Сетевые интерфейсы**»
2. Выберите на контроллере, к которому вы подключились, любой свободный интерфейс. Если его нет, необходимо освободить интерфейс от привязанных служб. Для этого откройте панель свойств выбранного интерфейса, нажмите на кнопку «**Редактировать адреса**» и отвяжите все службы, убрав все

флажки в окне **«Привязать службы»**.

3. Создаете (или измените через удаление старого) IP-адрес выбранного интерфейса. Выберите адрес, который будете временно использовать для управления. При добавлении нового адреса отметьте флажок **«Добавить шлюз по умолчанию»** и введите адрес и метрику шлюза.

4. В окне редактирования адреса нажмите на кнопку **«Пометить как управляющий»**, тем самым создав второй управляющий интерфейс.

5. Подключитесь к веб-интерфейсу управления по адресу нового интерфейса.

6. Откройте окно свойств «старого» управляющего интерфейса и нажмите на кнопку **«Редактировать IP адреса»**.

7. Сделайте интерфейс не управляющим, нажав на кнопку **«Пометить как не управляющий»**. Сохраните результат, нажав на кнопку **«Ок»**.

8. Измените, через удаление старого, IP-адрес выбранного интерфейса на нужный вам IP-адрес. При добавлении нового адреса отметьте флажок **«Добавить шлюз по умолчанию»** и введите адрес и метрику шлюза.

9. Снова сделайте интерфейс управляющим, как было описано ранее.

10. Подключитесь к веб-интерфейсу управления по новому адресу.

11. Сделайте второй интерфейс не управляющим, и удалите с него временный IP-адрес.

Описанным способом можно выполнять и смену управляющих интерфейсов (пункты 1-6).

2.2.1.4. Включение мониторинга

Служба SNMP необходима для мониторинга СХД внешними программами мониторинга, например, Zabbix. Также её необходимо включить для отображение данных о нагрузке в панели **«Нагрузка»**, расположенной на вкладке меню **«Система»-«Статус»**.

Для включения службы SNMP предназначенной для вывода статистики работы функционала СХД выполните следующие действия:

- Перейдите на вкладку меню **«Настройки»** → **«Прочие настройки»**;

- Включите службу SNMP, нажав кнопку **«Включить службу SNMP»** и подтвердите выбор в открывшемся окне. Для отключения службы SNMP нажмите **«Отключить службу SNMP»**;

На этой же вкладке меню можно скачать MIB-файл с описанием параметров для мониторинга (см. Рисунок 52).

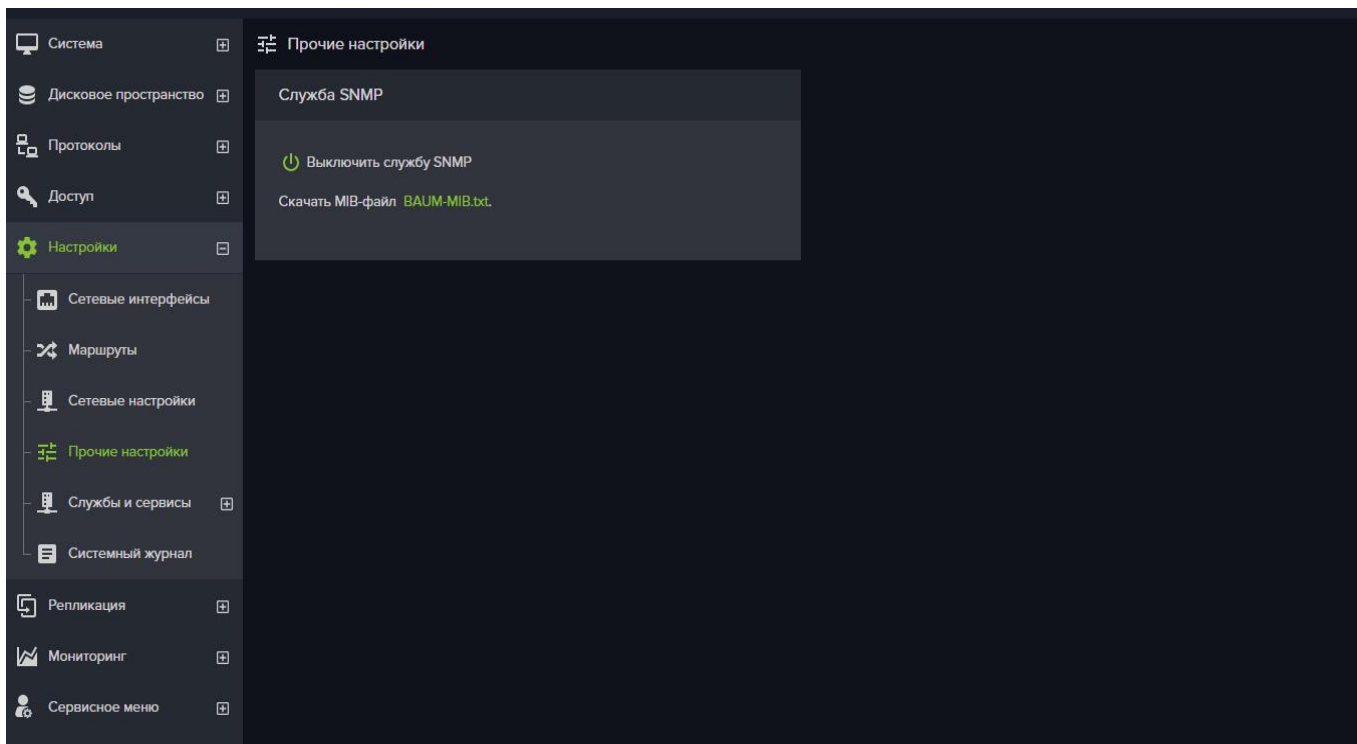


Рисунок 52. Настройка SNMP уведомлений

2.2.1.5. Включение и выключение служб протоколов

ПО BAUM STORAGE AI v2 имеет встроенные службы для создания виртуальных дисков для протоколов Fibre Channel, SMB, FTP. Каждая служба отвечает за создание диска по соответствующему протоколу:

- Fibre Channel - протокол широко применяется для создания сетей хранения данных. Имеет высокую скорости передачи данных, малую задержку и расширяемость;

- FTP - стандартный протокол, предназначенный для передачи файлов по TCP-сетям (например, Интернет). Использует 21-й порт. FTP часто используется для загрузки сетевых страниц и других документов с частного устройства разработки на открытые сервера хостинга.

- SMB - протокол сетевого доступа к файловым системам. Основан на протоколе вызова удалённых процедур. Позволяет подключать (монтировать) удалённые файловые системы через сеть.

Также реализована служба для SNMP (SNMP v.2) протокола (стандартный интернет-протокол для управления устройствами в IP-сетях на основе архитектур TCP/UDP), используемая для мониторинга работы ПО Платформы СХД.

Для включения или отключения служб протоколов выполните следующие действия:

12. Разверните содержимое пункта меню «**Протоколы**» и выберите нужный протокол из списка. В основной области окна появится название выбранной службы

протокола и информация о текущих настройках;

13. Для включения службы нажмите кнопку **«Включить службу»** и подтвердите своё действие в открывшемся окне;

14. Выключение службы выполняется нажатием на кнопку **«Выключить службу»**.

В результате появится уведомление о том, что операция выполнена успешно и статус службы был изменен.

2.2.1.6. Создание и удаление клиента

Для доступа к ресурсам СХД необходимо внести адреса подключаемых хостов в список доступа. Это реализовано при помощи создания так называемого клиента.

Создания клиента со всеми адресами WWPN, IQN и IP используется для удобного подключения виртуальных дисков и файловых систем.

Для создания клиента выполните следующие действия:

1. Перейдите во вкладку меню **«Доступ»** → **«Клиенты и группы»**;
2. В области **«Клиенты»** нажмите кнопку **«Создать нового клиента»**.

Откроется одноименное окно;

3. Впишите имя клиента;
4. В область WWPN нажмите **«Добавить WWPN»** впишите идентификаторы для Fibre Channel;
5. В области IP нажмите **«Добавить IP адрес»** вписать адреса для NFS и SMB;
6. Нажмите на кнопку **«Создать»**.

В результате выполненных действий, списке **«Клиенты»** появится новая строка с именем клиента.

Чтобы получить информацию о настройках клиента нажмите на стрелку перед именем клиента. При этом под строкой списка откроется панель свойств с информацией о имеющихся настройках и кнопки **«Редактировать»** и **«Удалить»**. При необходимости изменения настроек созданного клиента или его удаления выберите соответственно кнопки **«Редактировать»** или **«Удалить»**.

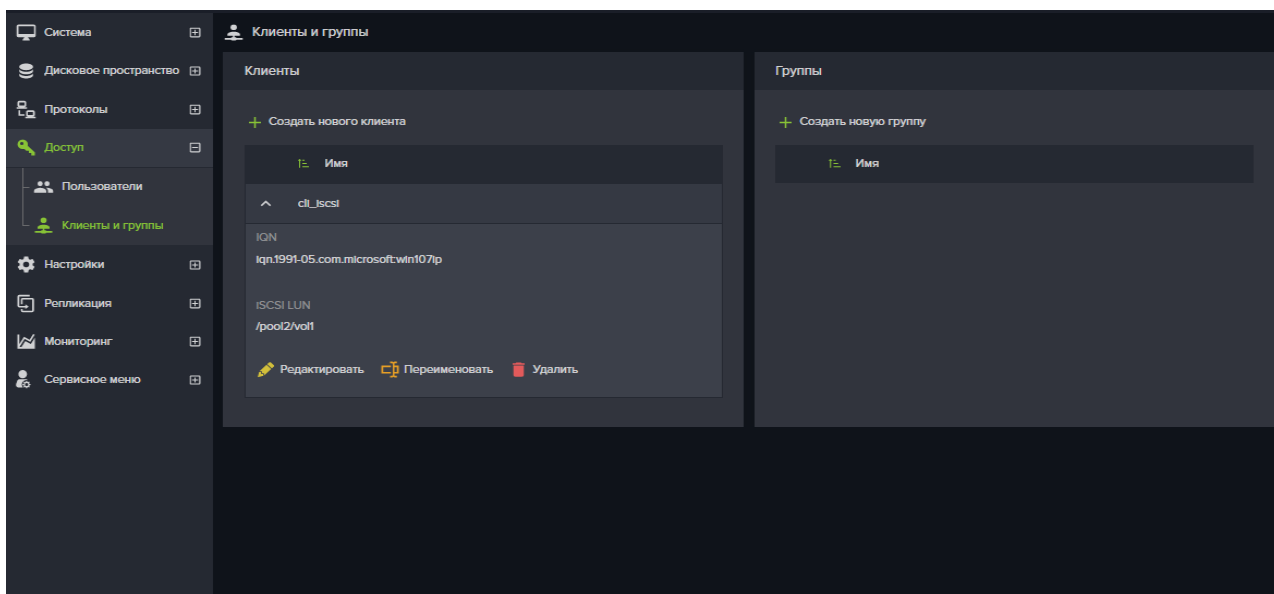


Рисунок 53. Вкладка меню «Доступ»

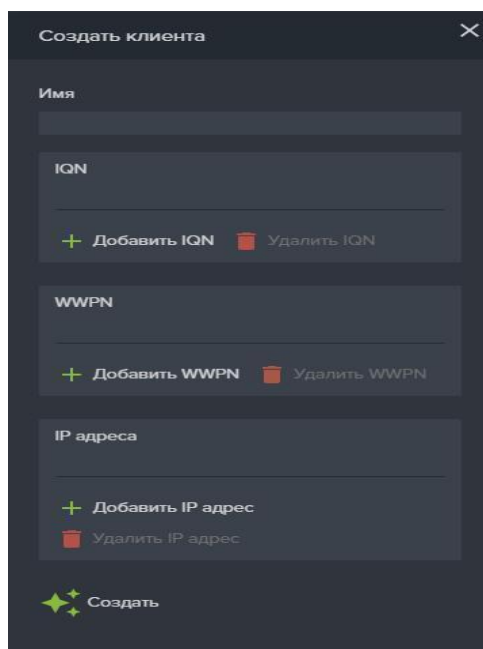


Рисунок 54. Окно создания клиента

2.2.1.7. Системный рейд

Программное обеспечение СХД установлено на отдельных SATA дисках для отказоустойчивости, объединенных в зеркало, которые установлены внутри каждого контроллера. При нормальной работе системы, в меню «Система» → «Системный рейд» отображаются четыре диска: два для первого контроллера и два для второго, как показано в примере на Рисунок 55. При неисправности диска, система удалит его из

списка дисков. В списке дисков, серийные номера дисков выводятся справа от номеров слотов.

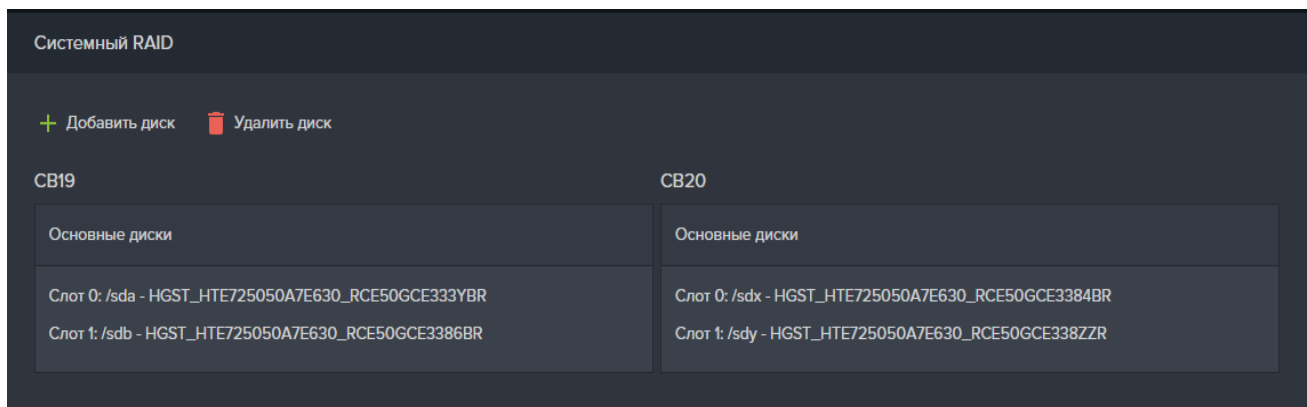




Рисунок 55. Меню управления системным RAID

2.2.1.8. Выход из системы управления

Для выхода из системы управления выполните следующие действия:

1. В верхнем правом углу интерфейса нажмите на кнопку 
2. Нажмите на значок выхода рядом с именем оператора 

В результате будет выполнено перенаправление на страницу входа.

2.2.1.9. Уведомления об ошибках

В интерфейсе управления системой предусмотрен механизм уведомлений, который выводит сообщения о удачно выполненных командах либо ошибках при выполнении команд. Данные сообщения появляются в правом нижнем углу экрана и исчезают через 10 секунд.

Окно подтверждения выполнения команды показано на Рисунок 56:

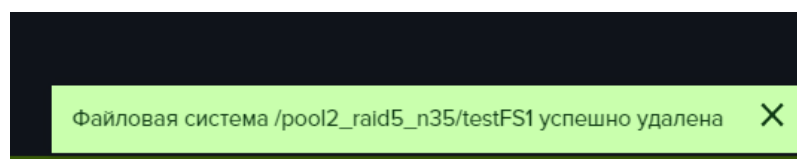


Рисунок 56. Окно подтверждения выполнения команды

В случае ошибочных действий администратора или неготовности системы выполнить команду возвращается сообщение об ошибке. Пример окна сообщения об ошибке показан на Рисунок 57:

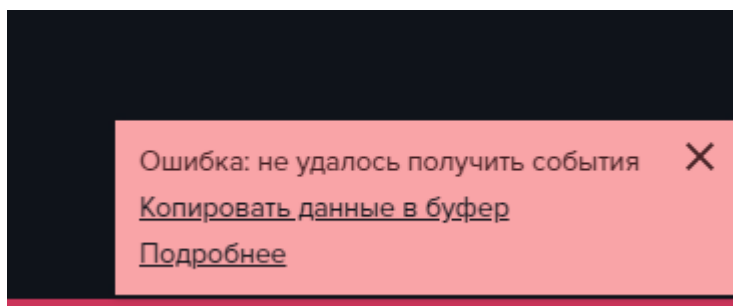


Рисунок 57. Окно сообщения об ошибке

Для получения подробной информации об ошибке в окне уведомления предусмотрена ссылка «Подробнее», нажатие на которую приведет к открытию окна с детальным описанием ошибки, пример которого представлен на Рисунок 58. Краткая информация об ошибке, содержащая её номер и описание, может быть скопирована в буфер обмена при нажатии на ссылку «Копировать данные в буфер».

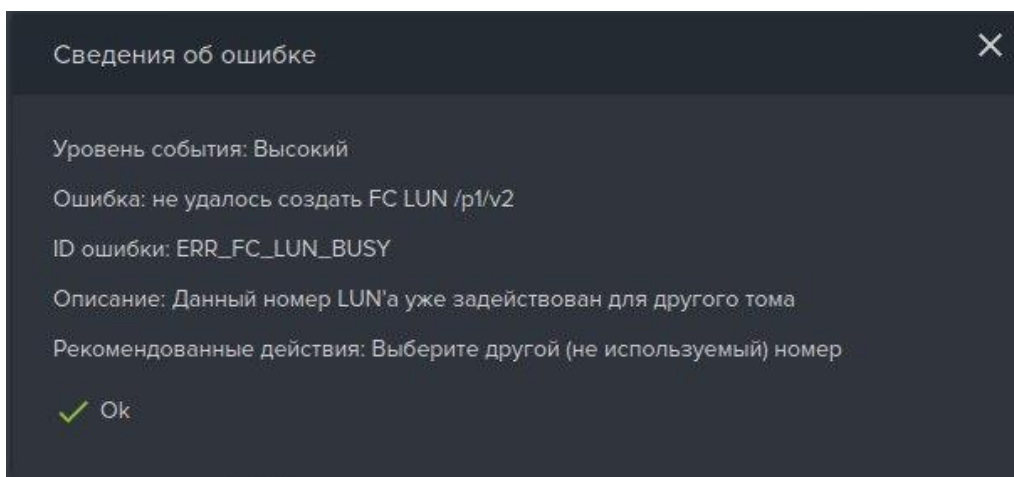


Рисунок 58. Окно с подробной информацией об ошибке

2.2.1.10. Настройка сети

Для изменения сетевых настроек выполните следующие действия:

Перейдите на вкладку меню **«Настройки»** → **«Сетевые интерфейсы»** (см. Рисунок 59). На ней отображены все сетевые интерфейсы изделия с их текущими настройками и состояниями:

- Имя сетевого интерфейса;
- Физический или виртуальный;
- Включен или выключен;
- MAC адрес сетевого интерфейса.

После щелчка мыши на стрелку вниз перед именем сетевого интерфейса, откроется его панель свойств. В нижней части панели можно выбрать необходимое действие для выбранного сетевого интерфейса:

- Редактировать (выбрать параметр MTU);
- Удалить (агрегированный или виртуальный интерфейс);
- Привязать интерфейсы к группам;
- Выключить интерфейсы;
- Редактировать адрес (добавить или удалить ip адрес).

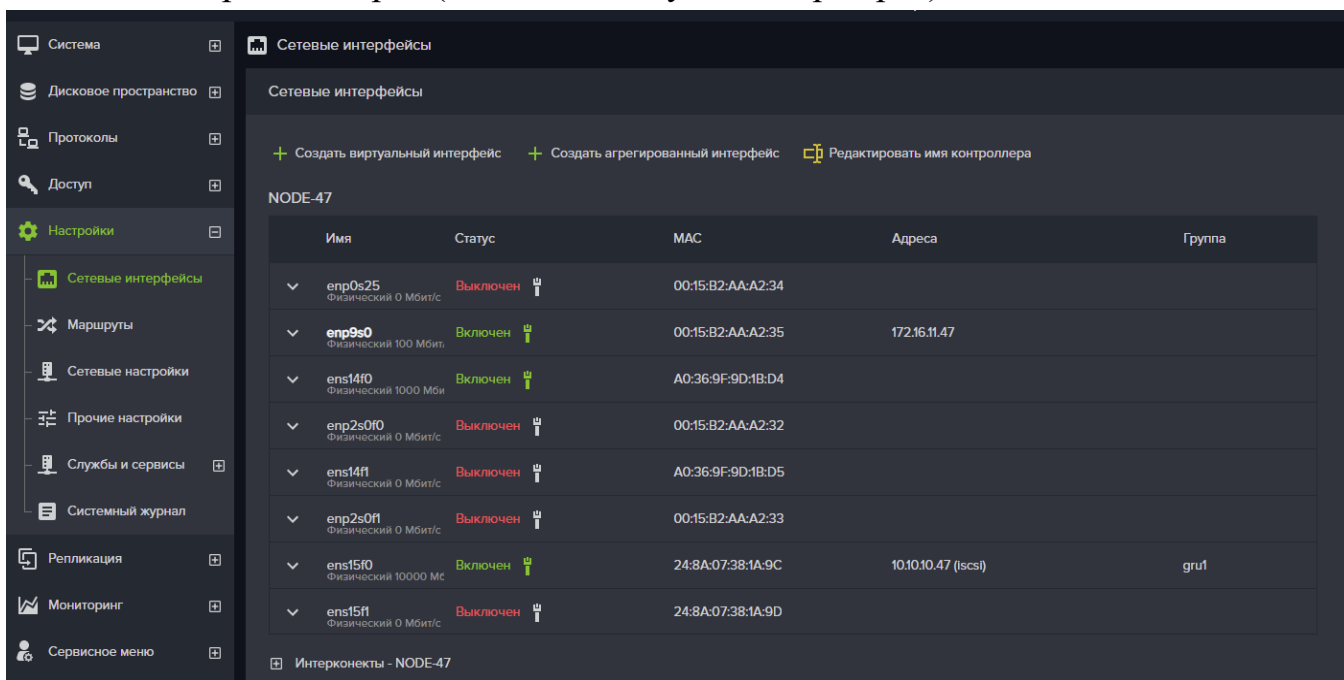


Рисунок 59. Вкладка меню «Сетевые интерфейсы».

В графе «Имя», управляющий интерфейс выделен жирным шрифтом. Ниже имени интерфейса, указана его скорость соединения. Информация о максимальной скорости порта указана в панели свойств интерфейса (см. Рисунок 60).

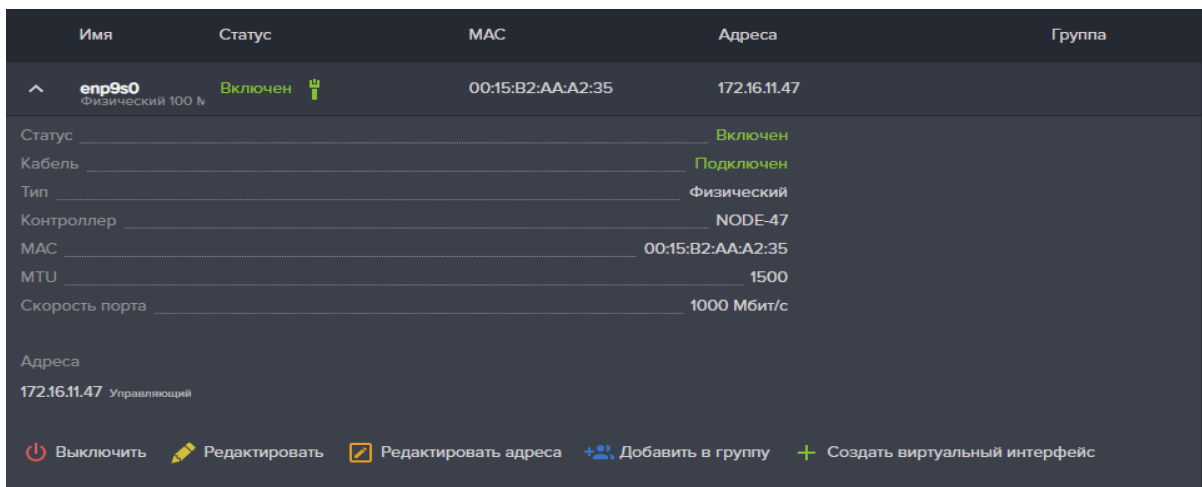


Рисунок 60. Панель свойств сетевого интерфейса

В отдельном списке на вкладке меню «Сетевые интерфейсы» сгруппированы интерфейсы, выполняющие роль между узлового интерконнекта и агрегированный интерфейс, собранный на их основе.

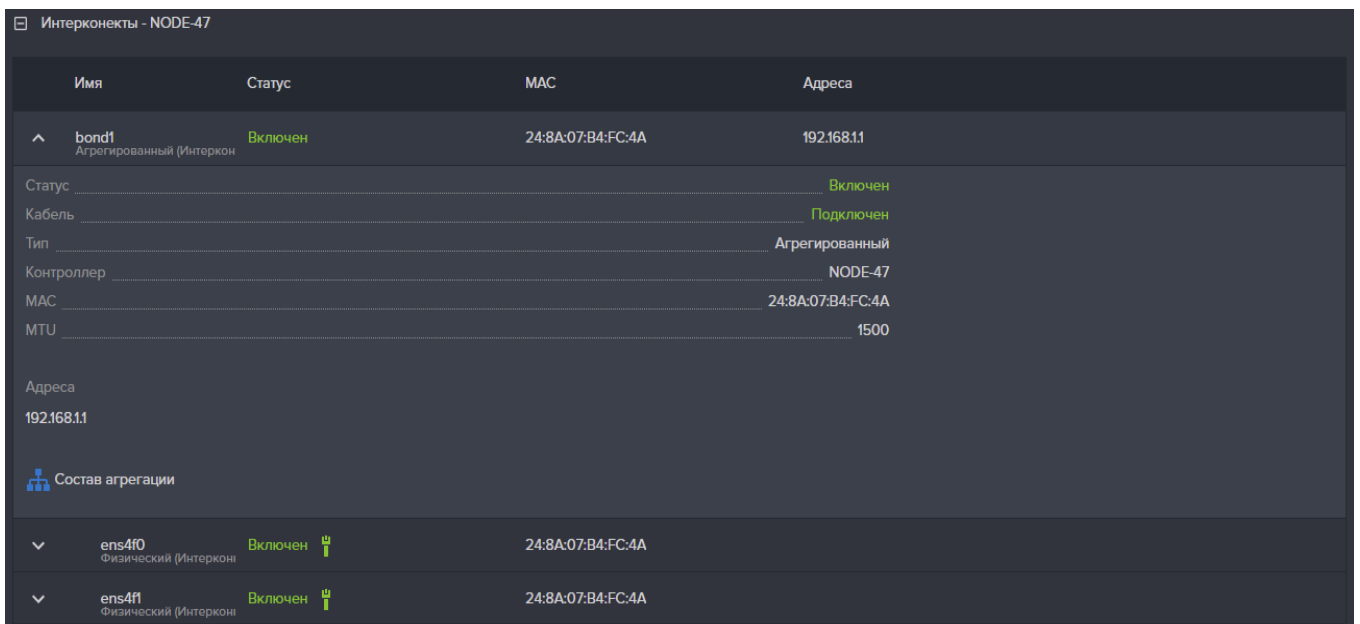


Рисунок 61. Панель свойств интерконнектов

В панели свойств интерконнектов выведена информация о их свойствах (см. Рисунок 61). Система не предоставляет инструментов для управления интерфейсами из группы интерконнектов, но позволяет изменить состав агрегации.

Для этого нажмите на кнопку «Состав агрегации» в панели свойств агрегированного интерфейса с именем bond 1 (или 2). Откроется окно «Состав агрегированного интерфейса», в котором нужно выбрать новый состав агрегации (см. Рисунок 62)

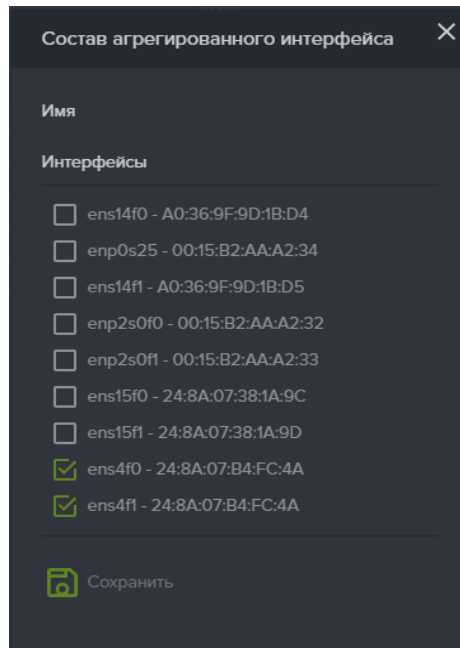


Рисунок 62. Окно изменения состава агрегации

2.2.1.10.1. Изменение имени контроллера

При заводской настройке, контроллерам присваиваются имена, получаемые из номера кластера и порядкового номера контроллера. Например, если при установке кластера был задан номер 13, то контроллеры будут иметь имена: NODE-13 и NODE-14. В последствии эти имена могут быть изменены.

Для изменения имени контроллера выполните следующие действия.

1. Перейдите на вкладку меню «**Настройки**» → «**Сетевые интерфейсы**».
2. Нажмите на кнопку «**Редактировать имя контроллера**». Откроется одноименное окно (см. Рисунок 63);
3. Выберите контроллер из списка;
4. Введите новое имя контроллера;
5. Нажмите кнопку «**Сохранить**».

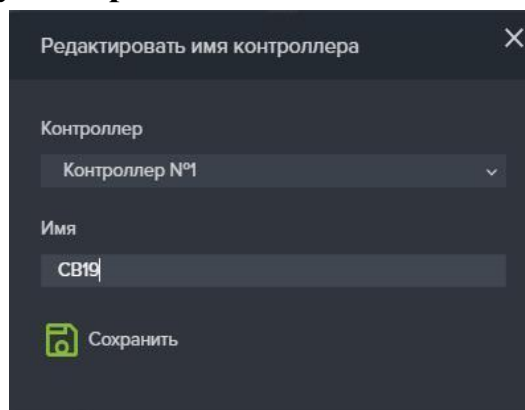


Рисунок 63. Изменение имени контроллера

2.2.1.10.2. Изменение имени кластера

Чтобы изменить имя кластера нажмите на имя кластера, расположенное на панели уведомлений в верхней части экрана (подчеркнуто на рисунке). В открывшемся окне измените имя кластера и нажмите «Сохранить».

2.2.1.10.3. Назначение IP-адреса интерфейсу

Для задания IP-адреса сетевому интерфейсу выполните следующие действия.

1. Перейдите на вкладку меню «**Настройки**» → «**Сетевые интерфейсы**».

На ней отображены все сетевые интерфейсы изделия с их текущими настройками и состояниями, для просмотра подробных параметров разверните панель свойств, нажав на стрелку перед названием нужного интерфейса;

2. Выберите интерфейс;

3. Нажмите кнопку «**Редактировать адреса**», откроется окно «Редактировать IP адреса» (см. Рисунок 64);

4. Нажмите кнопку «**Добавить**», откроется окно ввода адреса;

5. Впишите в открывшемся окне Адрес/Маску;

6. При необходимости отметьте флажок «**Добавить шлюз по умолчанию**» и введите адрес шлюза и метрику;

7. Нажмите кнопку «**Добавить**».

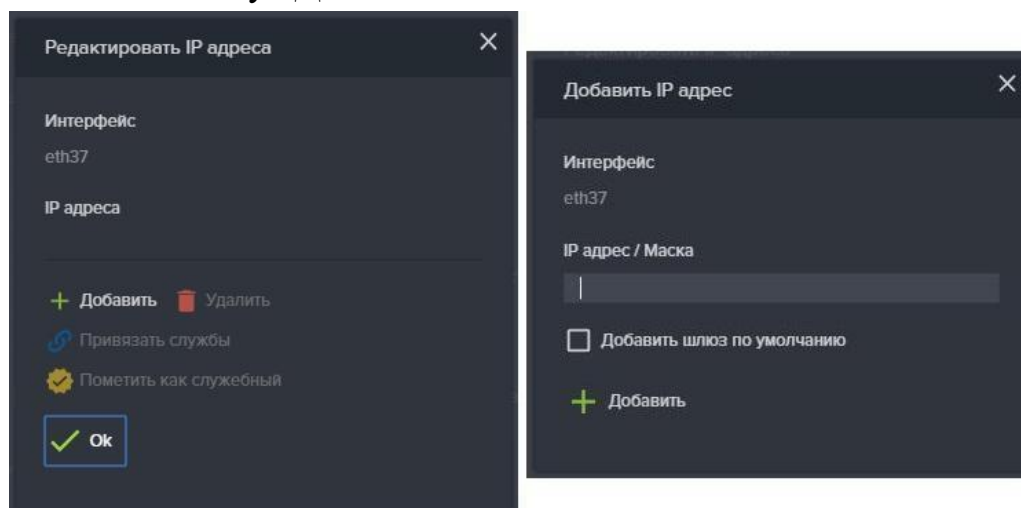


Рисунок 64. Окно редактирования IP адреса.

2.2.1.10.4. Удаление IP-адреса

Для удаления IP-адреса сетевого интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Сетевые интерфейсы**».

На ней отображены все сетевые интерфейсы изделия с их текущими настройками и состояниями.

2. Развернув панель свойств нужного сетевого интерфейса, выберите пункт **«Редактировать адреса»**.
3. В открывшемся окне выберите из списка нужный адрес и нажмите кнопку **«Удалить»** (см. Рисунок 65).
4. Подтвердите удаление в появившемся окне, введя слово **ok** и нажмите **«Удалить»**.

Если адрес помечен как служебный (управляющий), система не позволит удалить его. Для удаления управляющего адреса, сначала выберите адрес в списке и нажмите **«Пометить как не служебный»**. После этого удалите адрес как это делалось в предыдущих пунктах.

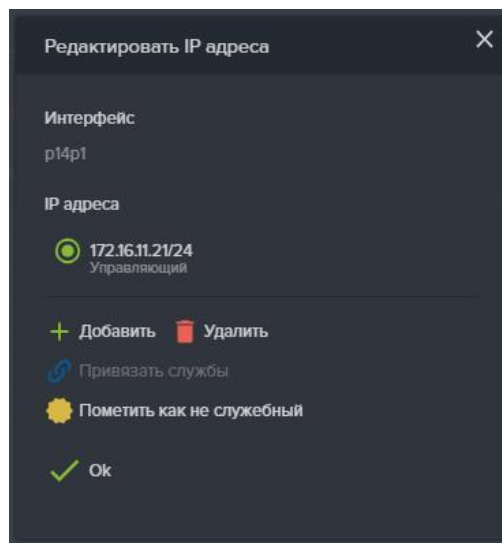


Рисунок 65. Удаление IP-адреса.

2.2.1.10.5. Создание агрегированного интерфейса

Для создания агрегированного интерфейса выполните следующие действия:

1. Перейдите на вкладку меню **«Настройки»** → **«Сетевые интерфейсы»** (см. Рисунок 66);
2. Нажмите на кнопку **«Создать агрегированный интерфейс»**;
3. Выберите нужный контроллер из выпадающего списка;
4. Введите имя создаваемого агрегированного интерфейса;
5. Выберите нужное значение MTU из списка;
6. Выберите из списка доступные интерфейсы для агрегации;
7. Нажмите на кнопку **«Создать»**;
8. Далее, для создания IP-адреса агрегированного канала выполните следующие действия:
 - a. Выберите агрегированный интерфейс, развернув его панель свойств;

- b. Нажмите **«Редактировать адреса»**;
- c. Нажмите **«Добавить»** и появившемся окне введите Адрес/Маску;
- d. Нажмите на кнопку **«Добавить»**.

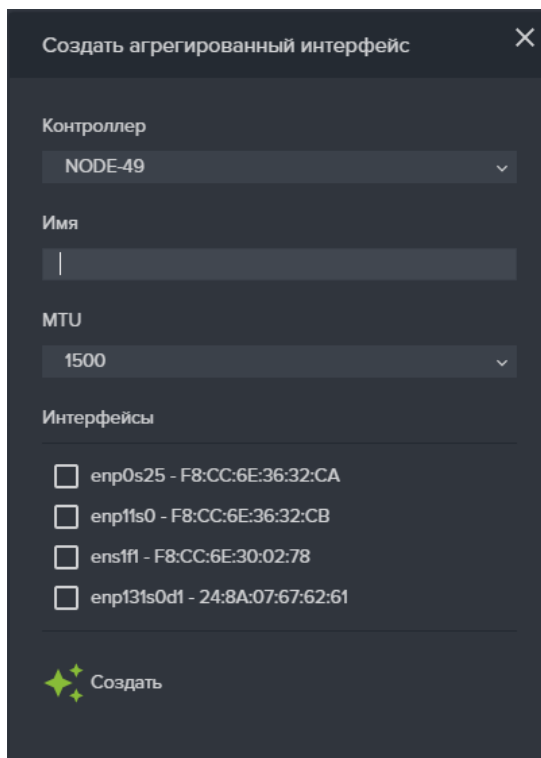


Рисунок 66. Создание агрегированного интерфейса.

2.2.1.10.6. Удаление агрегированного интерфейса

Для удаления агрегированного интерфейса выполните следующие действия.

1. Перейдите на вкладку **«Настройки»** → **«Сетевые интерфейсы»**;
2. Выберите нужный агрегированный интерфейс, развернув его панель свойств;
3. Нажмите на кнопку **«Удалить»**;
4. Подтвердить удаление в появившемся окне, введя слово **ok** и нажмите **«Удалить»**.

2.2.1.10.7. Создание виртуального интерфейса

Для создания виртуального интерфейса выполните следующие действия.

1. Перейдите на вкладку меню **«Настройки»** → **«Сетевые интерфейсы»**;
2. Нажмите на кнопку **«Создать виртуальный интерфейс»**, откроется одноименное окно (см. Рисунок 67);
3. Выберите нужный контроллер из выпадающего списка;

4. Выберите из списка нужный интерфейс;
5. Введите имя создаваемого интерфейса;
6. Введите идентификатор VLAN ID;
7. Нажмите на кнопку «Создать».

Далее, для создания IP адреса VLAN интерфейса выполните следующие действия:

1. Выберите ранее созданный виртуальный интерфейс, развернув его панель свойств;
2. Нажмите кнопку «**Редактировать адреса**», откроется окно редактирования IP адреса;
3. Нажмите кнопку «**Добавить**»;
4. В открывшемся окне введите адрес/маску;
5. Нажмите кнопку «**Добавить**».

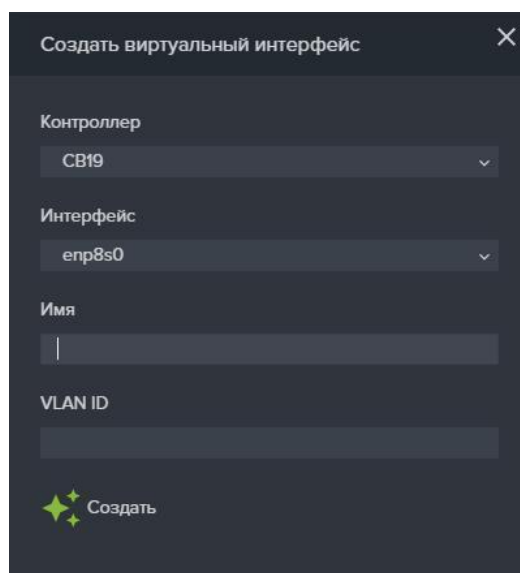


Рисунок 67. Окна создания виртуального интерфейса.

2.2.1.10.8. Удаление виртуального интерфейса

Для удаления виртуального интерфейса выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Сетевые интерфейсы**»;
2. Выберите виртуальный интерфейс, раскрыв его панель свойств;
3. Нажмите на кнопку «**Удалить**»;
4. Введите в открывшемся окне слово **ok** и нажмите на кнопку «**Удалить**».

2.2.1.10.9. Маршрутизация

Настройка маршрутов для сетевых интерфейсов выполняется на вкладке меню «Настройки»-«Маршруты».

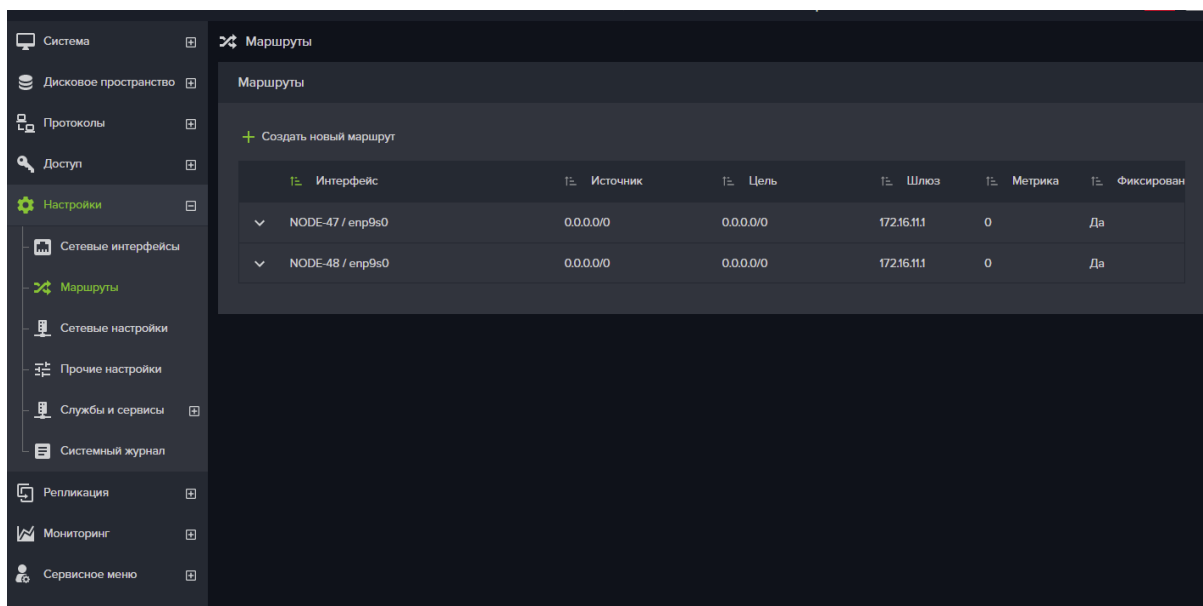


Рисунок 68. Вкладка меню «Маршруты»

Для создания нового маршрута выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Маршруты**».
2. В таблице маршрутизации нажмите на кнопку «**Создать новый маршрут**». Откроется одноименное окно (см. Рисунок 68).
3. Выберите нужный контроллер из выпадающего списка;
4. Выберите из списка нужный интерфейс;
5. Введите Адрес/Маску источника;
6. Введите Адрес/Маску цели;
7. Укажите шлюз;
8. Укажите метрику маршрута;
9. При необходимости отметьте флажок «**Фиксированный**»;
10. Нажмите на кнопку «**Создать**».

В таблице маршрутизации появится новый маршрут.

Для изменения параметров маршрута предназначена кнопка «Редактировать».

Изменить можно только метрику и состояние флажка «Фиксированный».

Изменение других параметров не предусмотрено, для их изменения удалите и снова создайте маршрут с нужными параметрами.

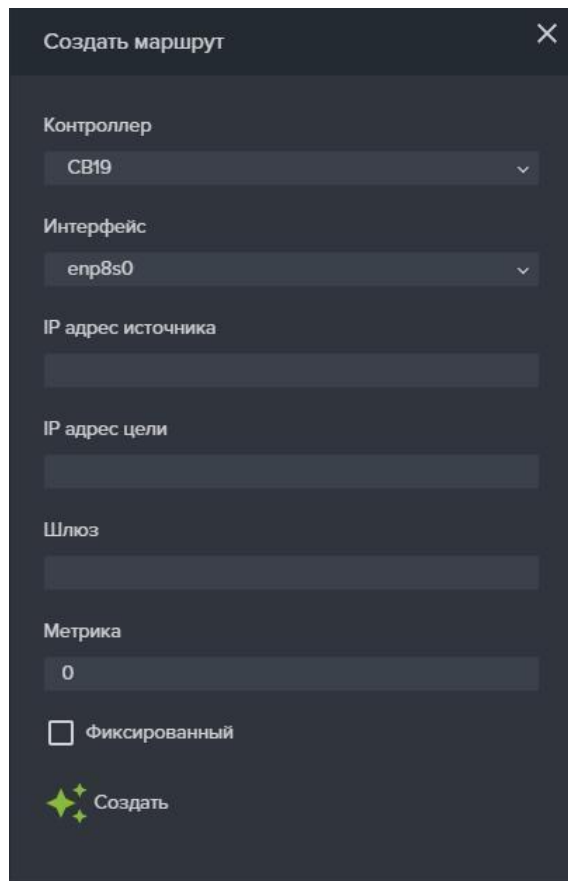


Рисунок 69. Создание маршрута.

2.2.1.10.10. Настройка адресов DNS и NTP серверов

Для настройки адресов DNS и NTP серверов, которые будет использовать СХД, выполните следующие действия:

1. Перейдите на вкладку меню **«Настройки»** → **«Сетевые настройки»** (см. Рисунок 70).

2. Введите в соответствующие поля параметры DNS серверов:

- a. Введите адрес DNS сервера;
- b. Введите адрес резервного DNS сервера;
- c. Введите адрес второго резервного DNS сервера;
- d. Нажмите на кнопку **«Сохранить»**.

3. Введите в соответствующие поля параметры NTP серверов:

- a. Введите адрес NTP сервера;
- b. Введите адрес резервного NTP сервера;
- c. Введите адрес второго резервного NTP сервера;
- d. Нажмите на кнопку **«Сохранить»**.

4. Выберите нужную Временную зону из списка и нажмите на кнопку **«Сохранить»**.

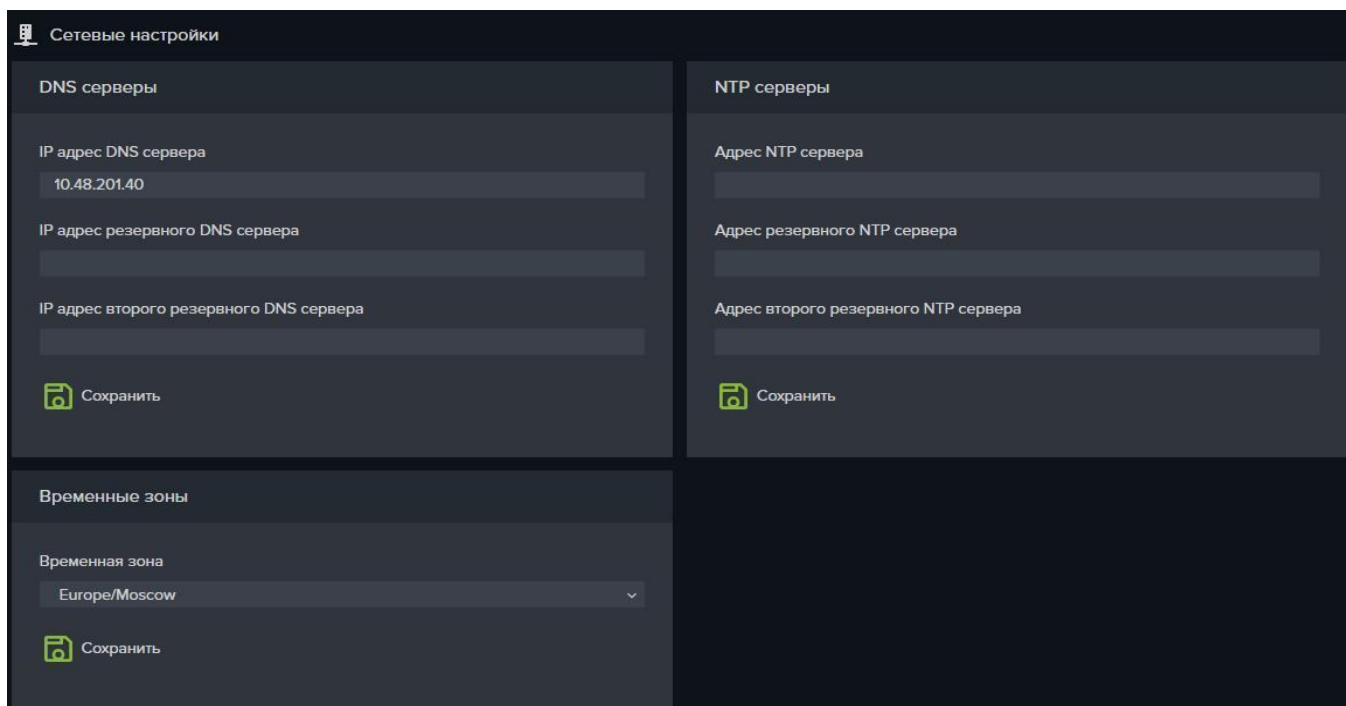


Рисунок 70. Настройки DNS и NTP серверов.

2.2.1.10.11. Настройка параметров Active Directory и LDAP

Для настройки интеграции СХД с каталогами AD и LDAP выполните следующие действия:

Перейдите на вкладку меню «Настройки» → «Службы и сервисы» → «AD/LDAP» (см. Рисунок 71).

Для ввода контроллеров в домен AD, выполните следующее:

- a. Введите имя домена;
- b. Нажмите на кнопку «Сохранить»;
- c. Введите имя пользователя домена и его пароль;
- d. Нажмите на кнопку «Ввести в домен».

Для вывода контроллеров из домена введите имя администратора домена и его пароль, после чего нажмите на кнопку «Вывести из домена».

Для подключения контроллеров к серверу LDAP, выполните следующее:

- a. Введите **URI** и **BASE** LDAP сервера;
- b. Нажмите на кнопку «Сохранить»;
- c. Нажмите кнопку «Включить службу LDAP».

Для изменения настроек: выключите службу, измените настройки и нажмите на кнопку «Сохранить». После чего включите службу повторно.

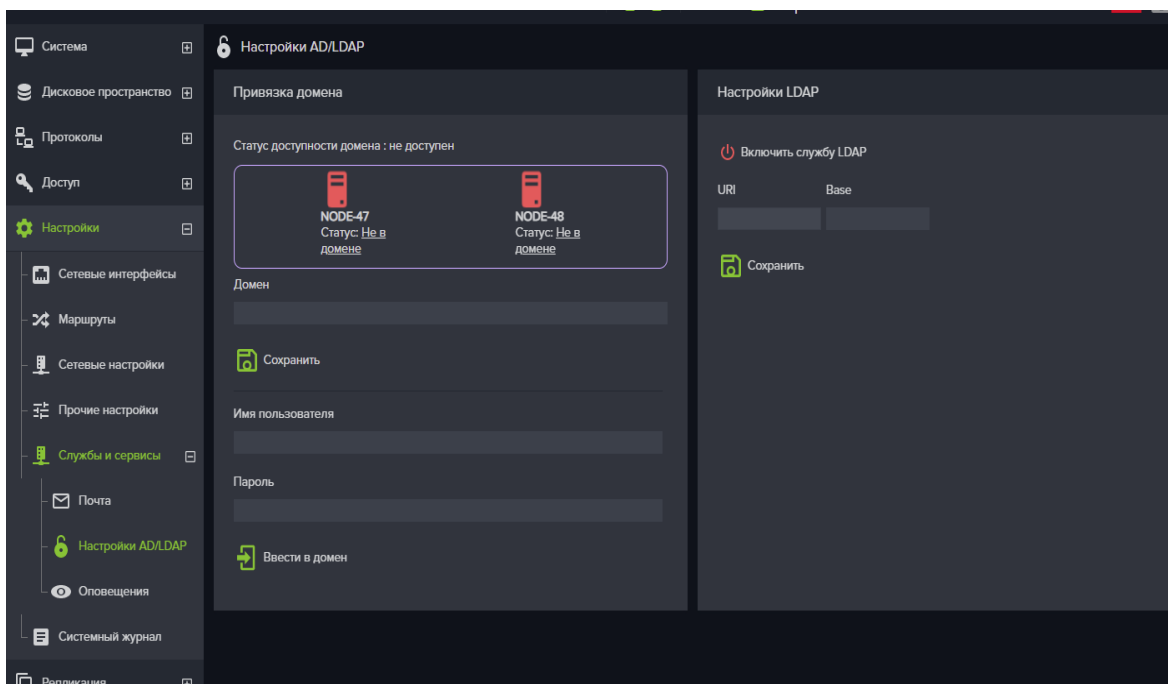


Рисунок 71. Настройки параметров AD и LDAP

2.2.1.10.12. Настройка управляющего интерфейса

Управляющий интерфейс служит только для подключения к интерфейсу управления СХД. Управляющий интерфейс не предназначен для передачи данных по файловым протоколам, за исключением протокола FTP.

Управляющий интерфейс должен быть настроен для каждого контроллера СХД. Во время новой установки системы контроллерам назначаются IP-адреса управляющих интерфейсов, которые впоследствии можно изменить. Также имеется возможность сделать управляющим другой интерфейс СХД.

Во вкладке меню «Сетевые интерфейсы» имя управляющего интерфейса выделено жирным шрифтом, как показано на Рисунок 73.

Для того чтобы сменить управляющий интерфейс выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**»-«**Сетевые интерфейсы**»;
2. Выберите интерфейс, который хотите сделать управляющим, развернув его панель свойств;
3. Нажмите на кнопку «**Редактировать адреса**»;
4. Если интерфейсу не назначен IP-адрес, нажмите на кнопку «**Добавить**» и введите адрес и маску подсети;
5. Если адрес уже назначен, выберите его и нажмите на кнопку «**Назначить управляющим**», как показано на Рисунок 72;
6. В том случае, когда к интерфейсу привязана какая-либо служба

протокола, её нужно сначала отвязать. Для этого нажмите на кнопку «**Привязать службы**» и в открывшемся окне снимите все флажки перед именами служб, после чего нажмите кнопку «**Ок**»;

7. Подключитесь к созданному управляющему интерфейсу, войдите в систему введя логин и пароль администратора;

8. Снимите статус управляющего интерфейса у первого интерфейса. Для этого откройте его панель свойств и нажмите на кнопку «**Редактировать адреса**»;

9. Выберите IP-адрес и нажмите на кнопку «**Пометить как не управляющий**», после чего подтвердите действие.

10. Выполните эти же действия для второго контроллера.

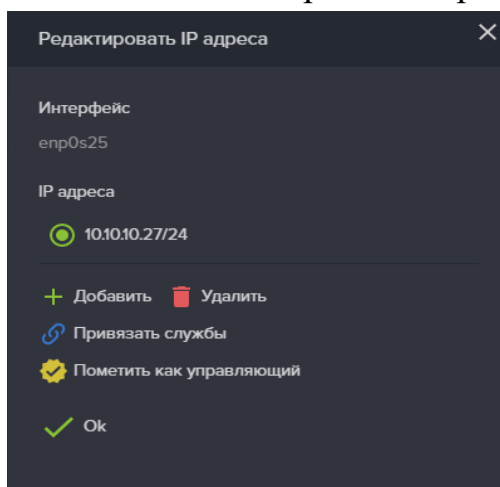


Рисунок 72. Назначение управляющего интерфейса

Сетевые интерфейсы

+ Создать виртуальный интерфейс + Создать агрегированный интерфейс Редактировать имя контроллера

NODE-47

Имя	Статус	MAC	Адреса	Группа
enp0s25 Физический 0 Мбит/с	Выключен	00:15:B2:AA:A2:34		
enp9s0 Физический 100 Мбит/с	Включен	00:15:B2:AA:A2:35	172.16.11.47	
ens14f0 Физический 1000 Мби	Включен	A0:36:9F:9D:1B:D4		
enp2s0f0 Физический 0 Мбит/с	Выключен	00:15:B2:AA:A2:32		
ens14f1 Физический 0 Мбит/с	Выключен	A0:36:9F:9D:1B:D5		
enp2s0f1 Физический 0 Мбит/с	Выключен	00:15:B2:AA:A2:33		
ens15f0 Физический 10000 Мс	Включен	24:8A:07:38:1A:9C	10.10.10.47 (iscsi)	gru1
ens15f1 Физический 0 Мбит/с	Выключен	24:8A:07:38:1A:9D		

Интерконекты - NODE-47

NODE-48

Рисунок 73. Управляющий интерфейс в списке сетевых интерфейсов

2.2.2. Управление ресурсами

2.2.2.1. Дисковое пространство

2.2.2.1.1. Работа с дисками

Для просмотра сведений о дисках, а также для управления дисковой подсистемой необходимо перейти в раздел меню «**Дисковое пространство**» → «**Диски**». В этом разделе выводится список всех дисковых накопителей, присутствующих в системе. В этом списке отображаются данные о типе диска, его размере, скорости (для шпиндельных дисков), наименование полки, номер слота в полке, принадлежность пулу, наличие метки пула, статус и индикатор подсветки диска.

В интерфейс программы включена возможность фильтрации списка дисков по их типу, размеру и принадлежности пулу и сортировка по дисковой полке. Для этого нужно нажать на название столбца.

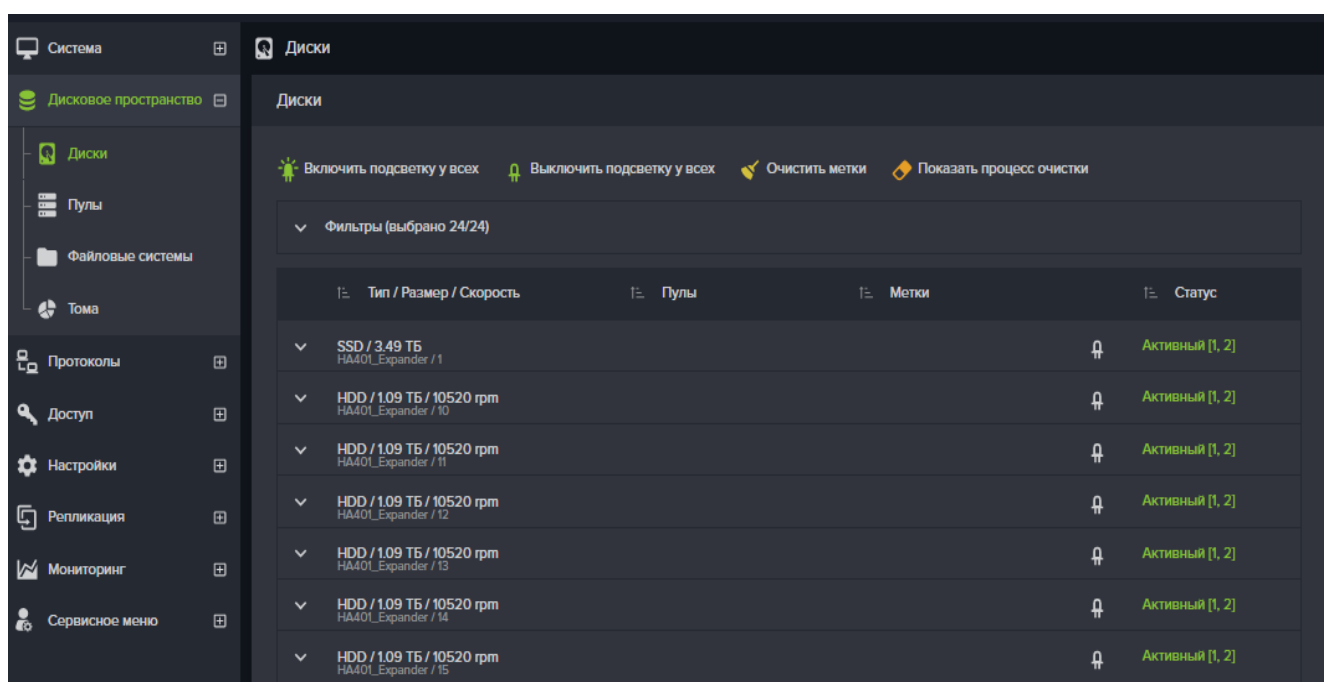



Рисунок 74. Вкладка меню Дисковое пространство -> Диски.

2.2.2.1.2. Управление подсветкой дисков

Управлять подсветкой дисков из данного раздела можно тремя различными путями:

1. Включить подсветку у всех дисков, нажав на кнопку сверху «**Включить/Выключить подсветку у всех**»
2. Включить подсветку у всех дисков, предварительно отфильтрованных по определенным параметрам (тип, размер, принадлежность пулам) в любом соотношении. Для этого необходимо сначала выполнить фильтрацию и нажать на

кнопку сверху **«Включить/Выключить подсветку у всех»**

3. Включить подсветку у одного конкретного диска. Для этого нужно найти интересующий диск в списке и нажать на **кнопку-индикатор подсветки**, расположенной справа от метки и слева от статуса диска. 

2.2.2.1.3. Очистка меток

На диски, собранные в пул пишется специальная метка пула, уникальная для каждого пула в системе.

Очистка метки с диска требуется тогда, когда в систему вставили диск из другой системы. Для того, чтобы добавить такой диск в пул, метку с диска необходимо удалить.



Внимание! Не используйте данную опцию на дисках, входящих в пул с нужными данными, существует вероятность потери данных с пула!

Для очистки меток выполните следующие действия:

1. **Перейдите в раздел меню «Дисковое пространство» → «Диски».**
2. Нажмите на кнопку **«Очистить метки»** сверху окна.
3. Выберите диск, с которого требуется очистить метку.
4. Нажмите кнопку **«Очистить»**.
5. В некоторых случаях потребуется подтвердить действие, введя **«ok»** и нажав на **«Очистить»**

Процесс очистки произойдет мгновенно, диск окажется в системе без метки.

2.2.2.1.4. S.M.A.R.T

Для диагностики и проактивного мониторинга дисковой подсистемы в программный комплекс включена функция отслеживания изменения количества ошибок, получаемых из S.M.A.R.T каждого из дисков. Пороговые значения параметров устанавливаются в соответствии с рекомендациями производителя, при появлении хотя бы одной невосстановимой ошибки диска (чтении/записи) система выдаст предупреждение пользователю о необходимости замены жесткого диска.

Для SSD дисков так же установлено пороговое значение на «износ». Когда значение этого параметра в S.M.A.R.T достигнет 80% система предложит заменить диск.

Для просмотра данных S.M.A.R.T диска перейдите в меню **«Дисковое пространство» → «Диски»**, выберите диск из списка, развернув его дополнительные свойства, и нажмите на кнопку **«Данные SMART»**. На экране появится следующая

информация по состоянию жесткого диска: серийный номер, количество запусков (для шпиндельных дисков), время работы, температура, износ (для SSD дисков), ошибки чтения, ошибки записи, ошибки верификации, ошибки восстановления, Код ошибки ASC/ASCQ. См. Рисунок 75.

Ошибка S.M.A.R.T не приведет к автоматической замене диска на запасной (hot spare) диск.

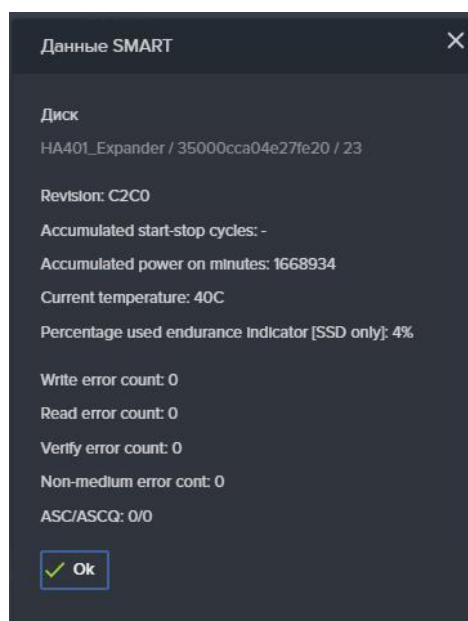


Рисунок 75. Окно вывода параметров SMART

Причину, по которой диск вышел из строя, как правило, показывает код ASC/ASCQ. Значение кодов можно посмотреть в документе по ссылке: https://www.t10.org/lists/asc-num.htm#ASC_0D

2.2.2.1.5. Запасные диски (Hot spare)

Запасные (Hot spare) диски используются для автоматической замены отказавшего диска в пуле. Для каждого пула используются свои запасные диски (глобальный Hot spare не реализован). В случае выхода из строя диска в пуле, произойдет автоматическая замена отказавшего диска на запасной диск и пул перейдет в режим RESILVERING (восстановление целостности) и после его завершения в рабочее состояние.

Если к пулу не были добавлены запасные диски и произошел отказ диска в пуле, то система не позволит добавить запасные диски к поврежденному пулу (пул в статусе DEGRADED). Замена диска в поврежденном пуле выполняется вручную. После замены диска автоматически запустится процесс восстановления целостности пула (RESILVERING).

2.2.2.1.6. Замена диска после его извлечения

Диск отображаемый в окне управления дисками («Дисковое пространство»-«Пулы», панель свойств) могут иметь следующие статусы:

ONLINE – нормальное состояние, диск работоспособен;

UNAVAIL или **UNKNOWN** – диск отключен (извлечен);

FAULTED – ошибка контрольных сумм.

При выходе из строя накопителя, или его извлечении из слота корзины, в меню аппаратного мониторинга (на панели дисковых полок) изображение дискового слота, соответствующего удаленному накопителю будет подсвечено красным цветом (Рисунок 40).

Для обычных и быстрых пулов алгоритм замены диска различается.

Замена диска на новый в быстром пуле:

1. Извлеките нужный диск из слота корзины, предварительно подсветив его (кнопка «Подсветка диска»);
2. Подождите 10 секунд, изображение слота станет красным;
3. Щелкните мышкой на слот и в открывшемся окне подтвердите удаление диска.
4. Перейдите в меню «Дисковое пространство»-«Пулы»;
5. Выберите пул из которого извлекли диск и в панели свойств нажмите на кнопку «Управление дисками»;
6. Извлеченный диск будет отображен в секции "Требуют замены" в окне "Управление дисками". Все диски, которые отображаются в секции "Требуют замены" необходимо заменить, прежде чем выполнять какие-либо операции.

Если просто вернуть диск в полку, диск НЕ вернется в состав пула!

Замена диска на новый в обычном пуле:

1. Извлеките нужный диск из слота корзины, предварительно подсветив его (кнопка «Подсветка диска»);
2. Подождите 10 секунд, изображение слота станет красным;
3. Щелкните мышкой на слот и в открывшемся окне подтвердите удаление диска.
4. Перейдите в меню «Дисковое пространство»-«Пулы»;
5. Выберите пул из которого извлекли диск и в панели свойств нажмите на кнопку «Управление дисками»;
6. Извлеченный диск будет отображен в секции основных дисков, со статусом UNAVAIL;
7. Замените диск на другой свободный диск такой же модели, для чего отметьте диск в списке и нажмите на кнопку «Заменить». В открывшемся окне выберите из списка диск для замены и нажмите на кнопку «Заменить».

Если просто вернуть извлеченный диск в полку, то диск вернется в состав пула! (в меню мониторинга, изображение слота дисковой полки в который вставили диск, станет зеленым, и значок диска на панели уведомлений перестанет мигать и тоже станет зеленым).

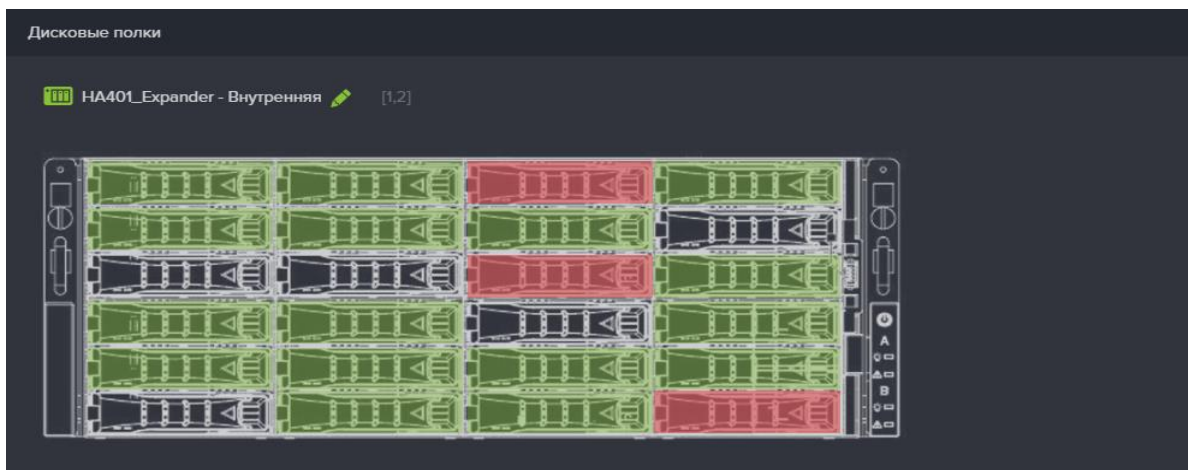


Рисунок 76. Отображение удаленных или неисправных дисков в меню аппаратного мониторинга

Извлечение (удаление) диска выполняется следующим образом:

1. Извлеките нужный диск из слота корзины;
2. Подождите 10 секунд, пока изображение дискового слота в меню аппаратного мониторинга не станет красным;
3. Щелкните мышью на изображении дискового слота и в открывшемся окне нажмите **«Подтвердить удаление диска»**;
4. Убедитесь в том, что изображение слота изменило цвет на серый, а значок диска на панели уведомлений стал зеленым.

2.2.2.1.7. Удаление внешней дисковой полки

Во избежание потери данных, перед отключением и последующим удалением внешней дисковой полки убедитесь в том, что диски этой полки не используются в пулах. В случае если диск состоит в каком-либо пуле, необходимо выполнить его замену на запасной диск, не находящийся в отключаемой полке. Замена диска выполняется из меню **«Дисковое пространство»-«Пулы»**, кнопка **«Управление дисками»**.

Для удаления полки из системы выполните следующие действия:

1. Выключите питание дисковой полки;
2. Перейдите в меню **«Мониторинг»-«Аппаратное обеспечение»**;
3. Перейдите в область дисковых полок. Отображение отключенной

дисковой полки исчезнет, рядом с названием полки появятся надпись «Отсоединена» и кнопка «**Подтвердить отключение**» см. Рисунок 77;

4. После нажатия на кнопку, введите «**ОК**» в появившемся диалоговом окне;
5. Удаленная полка пропадет из мониторинга.

Вместе с полкой будут автоматически отключены и все диски, установленные в отключенной полке.

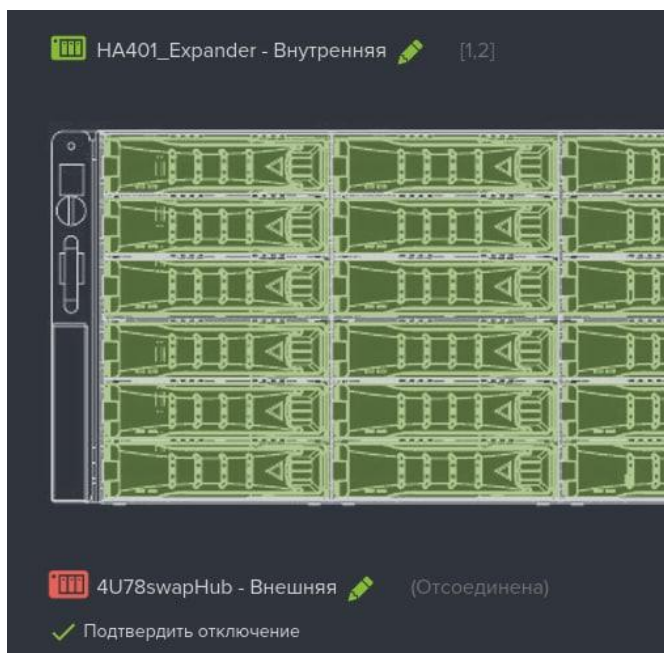


Рисунок 77. Экран меню мониторинга после отключения питания полки

2.2.2.1.8. Подключение внешней дисковой полки

Подключите дисковую полку к контроллерам СХД. Включите питание полки. Система автоматически «увидит» новую полку и все установленные в ней диски.

2.2.2.1.9. Очистка дисков

Для удаления информации с диска перед его утилизацией или передачей для использования другими лицами в системе предусмотрена функция очистки диска, гарантирующая удаление информации с накопителя. Для того, чтобы очистить диск необходимо выполнить следующие действия:

1. Удалить пул, или удалить метку с диска (система не позволит очистить диск, входящий в состав пула).
2. Выбрать диск, развернув его дополнительные свойства.
3. Нажать на кнопку «**Очистить**».
4. Подтвердить операцию, введя «**ок**» в открывшемся окне и нажать

«Очистить».

В результате выполненных действий начнется процедура трехкратной перезаписи диска, для удаления информации и затруднения ее восстановления. Данный процесс занимает длительное время, посмотреть статус прохождения процедуры по всем дискам можно с помощью нажатия кнопки **«Показать процесс очистки»**. В верхней части окна будет выведен список всех дисков, процесс очистки по которым находится в активном состоянии и индикация прогресса в процентах.

2.2.2.2. Работа с дисковыми пулами

Дисковый пул — устройство хранения, состоящее из нескольких жестких дисков. Дисковые пулы являются основными элементами хранения данных.

Для получения максимальной производительности рекомендуется на каждом контроллере создавать не более двух пулов и не более шести lun, равномерно распределив их по созданным пулам.

2.2.2.2.1. Виды пулов

В системе имеется два режима работы пулов: обычный и быстрый. У уже созданного пула изменить режим работы невозможно, поэтому перед созданием пула необходимо иметь в виду особенности его работы, описанные далее в этом разделе.

2.2.2.2.1.1. Обычные пулы.

Пулы, созданные в обычном режиме наиболее универсальны. Обычные пулы хорошо подходят для длительного хранения данных с использованием шпиндельных дисков (HDD) большого объёма. Для увеличения быстродействия обычных пулов используется дисковый кэш (быстрые SSD диски). Использование SSD накопителей существенно увеличит производительность пула, но не сможет раскрыть весь потенциал накопителей.

Обычные пулы поддерживают следующие типы избыточности: RAID 0,1, 5,6 В3, 10, 50, 60, В30. На пулах этого вида можно создавать как блочные, так и файловые ресурсы.

При создании обычного пула есть возможность указать количество рейд-групп. Рейд-группа – составная часть пула, имеющая определенный тип отказоустойчивости (типы RAID). В Пуле все рейд-группы должны состоять из одинакового количества дисков идентичной модели. СХД самостоятельно отбирает диски, из которых можно составить пул.

В пуле может быть создана как одна рейд-группа, так и несколько. В рейд группу рекомендуется собирать не более 14 дисков. Если вы планируете собрать пул из большого кол-ва дисков, рекомендуем распределить их по нескольким рейд-группам так, чтобы в каждой группе было от 6 до 14 дисков. Увеличение количества рейд-групп в пуле ведет к увеличению производительности пула.

Для удобного создания рейд-групп при создании пула используйте опцию «**Выбрать автоматически**», поручив СХД самой выполнить расчет количества дисков в рейд группе в соответствии с типом RAID и общим количеством дисков, из которых будет создан пул. Процесс автоматического подбора дисков описан в разделе «Создание пула с автоматическим выбором дисков».

Такие функции, как: клонирование ресурсов и репликация данных, в текущей версии ПО реализованы только для обычных пулов.

2.2.2.2.1.2. Быстрые пулы.

Пулы, созданные в «быстром» режиме, обладают значительно большей производительностью, но меньшей функциональностью. Рекомендуется использовать в быстрых пулах только SSD накопители. Быстрые пулы будут хорошим выбором для работы систем виртуализации и баз данных. Быстрые пулы позволяют создавать на них только блочные ресурсы. Пулы этого вида, не нуждаются в подключении дискового кэша. Данный вид пула позволяет использовать следующие типы избыточности: RAID 0, 1, 5, 10. При создании «быстрого» пула уровня RAID10 требуется указать количество дисковых групп. В Пуле все дисковые группы должны состоять из одинакового количества дисков идентичной модели. СХД самостоятельно отбирает диски, из которых можно составить пул.

Для лучшей производительности, рекомендуется собирать в пул 6 - 10 накопителей.

2.2.2.2.2. Создание пула

Для создания нового пула выполнить следующие действия:

1. Перейдите на вкладку «**Дисковое пространство**» → «**Пулы**» (см.

Рисунок 42);

2. Нажмите кнопку вверху окна «**Создать пул**»;

3. Введите имя создаваемого пула;

4. Выберите режим работы пула;

5. Выберите из выпадающего списка контроллер, который будет владельцем пула;

6. Выберите тип отказоустойчивости пула из списка, где:

- **RAID0** — дисковый пул повышенной производительности с чередованием, без отказоустойчивости;

- **RAID1** — зеркальный дисковый пул;

- **RAID5** — дисковый пул с чередованием и «невыведенным диском чётности»; допускает выход из строя только одного диска;

- **RAID6** — дисковый пул с чередованием и двумя «невыделенным дисками чётности», использующий две контрольные суммы, вычисляемые двумя независимыми способами, допускает выход из строя не более чем двух дисков. Данный тип применим только к «обычным» пулам;

- **RAIDB3** - дисковый пул с чередованием и тремя «невыделенным дисками чётности», допускает выход из строя не более чем трех дисков. Данный тип применим только к «обычным» пулам;

7. Отметьте диски, из которых будет создаваться пул;
8. Добавьте SSD кэш (только для обычных пулов);
9. Задайте размер зарезервированной области, в процентах от размера пула (только для обычных пулов).
10. Нажмите на кнопку «Создать». В списке пулов появится созданный пул.

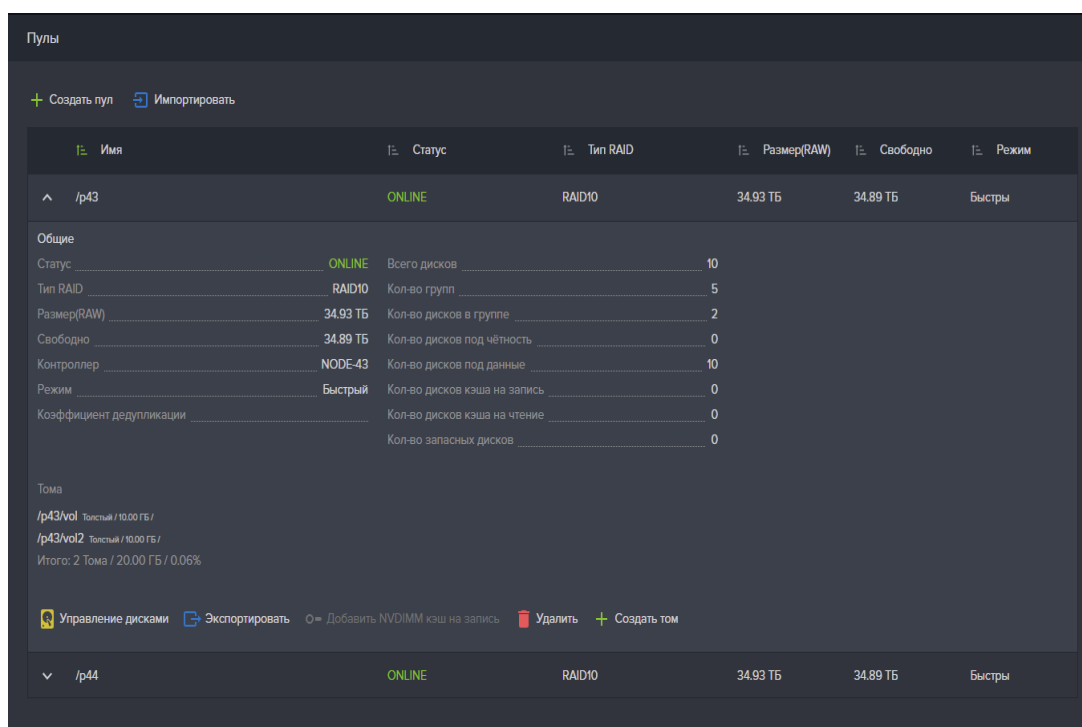


Рисунок 78. Вкладка меню «Пулы» с открытой панелью свойств пула.

В свойствах пула отображается полный размер пула - RAW, рассчитанный как сумма объёмов всех дисков из которых состоит пул (за исключением запасных и кэшей) и размер свободного места для записи данных (с учетом объема требуемого для поддержания избыточности хранения данных).



Внимание! Имя пула не может начинаться с буквы «с» (si) и не должно содержать символ нижнего подчеркивания. Пул с таким именем создать не удастся!

Создать пул

Имя
турpool

Режим
Обычный

Контроллер
NODE-43

Тип RAID
RAID0
RAID0
RAID1
RAID5
RAID6
RAIDB3

Выбрать автоматически

35002538a67b01f90
SSD / 3.49 ТБ / HA401_Expander / 20

35002538a67b021c0
SSD / 3.49 ТБ / HA401_Expander / 3

35002538a67b022b0
SSD / 3.49 ТБ / HA401_Expander / 14

35002538a67b012e0
SSD / 3.49 ТБ / HA401_Expander / 2

35002538a67b01270
SSD / 3.49 ТБ / HA401_Expander / 11

35002538a67b01ec0
SSD / 3.49 ТБ / HA401_Expander / 4

Выбрано 0 дисков

Резервирование свободного места на пуле, %:
10

Создать

Создать пул

Имя
fastpool

Режим
Быстрый

Контроллер
NODE-43

Тип RAID
RAID0
RAID0
RAID1
RAID5

Диски

Выбрать автоматически

35002538a67b01f90
SSD / 3.49 ТБ / HA401_Expander / 20

35002538a67b021c0
SSD / 3.49 ТБ / HA401_Expander / 3

35002538a67b022b0
SSD / 3.49 ТБ / HA401_Expander / 14

35002538a67b012e0
SSD / 3.49 ТБ / HA401_Expander / 2

35002538a67b01270
SSD / 3.49 ТБ / HA401_Expander / 11

35002538a67b01ec0
SSD / 3.49 ТБ / HA401_Expander / 4

Выбрано 0 дисков

Создать

Рисунок 79. Настройки создаваемого пула.

2.2.2.2.3. Создание пула с автоматическим выбором дисков

Опция автоматического выбора дисков позволяет автоматизировать расчет количества возможных рейд групп исходя из заданного количества дисков и типа пула. При помощи этой опции можно создать комбинированные типы RAID, например, RAID10.

Для создания нового пула с автоматическим выбором дисков выполните следующие действия:

1. Перейдите на вкладку «Дисковое пространство» → «Пулы» (см. Рисунок 42);
 2. Нажмите кнопку «Создать пул», она находится в верхней части окна;
 3. Введите имя пула;
 4. Выберите контроллер, который будет владельцем пула;
 5. Выберите тип отказоустойчивости пула из списка;
 6. Выберите пункт «**Выбрать автоматически**». При этом окно создания пула примет вид как на Рисунок 80;
 7. Выберите тип и объем дисков из списка;
 8. Выберите количество дисков в пуле;
 9. Выберите количество дисков в группе;
 10. Задайте размер зарезервированной области (в процентах от размера пула).
 11. Нажмите на кнопку «Создать».
- В результате выполненных действий в списке пулов появится новый пул.

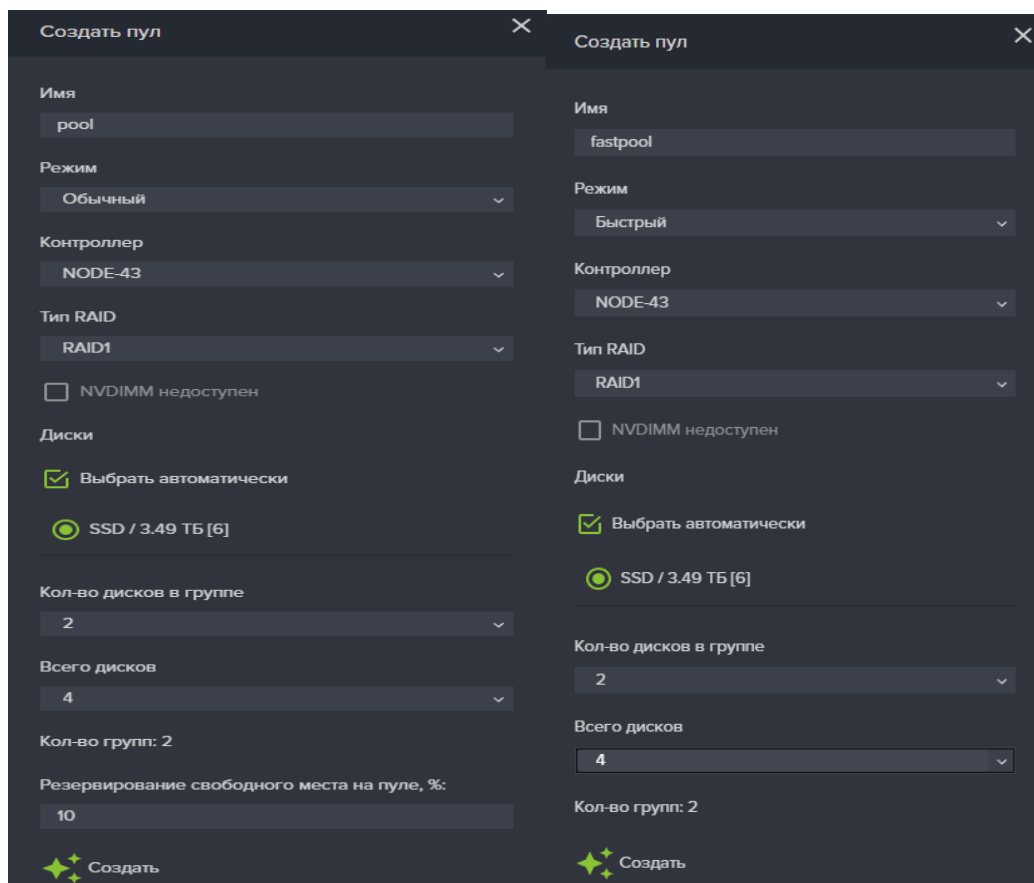


Рисунок 80. Настройки автоматически создаваемого пула.

Пулы уровней RAID10, 50, 60 создаются при помощи добавления рейд групп, как показано на Рисунок 44.

Резервирование свободного места на пуле.

Резервирование свободного места на пуле используется только для обычных пулов. Резервирование предотвращает заметное падение скорости записи при заполнении всего пула. Этот резерв используется файловой системой при записи. Может быть задан резерв от 1% до 20%. Для магнитных дисков рекомендуемый размер резерва – 10%. Для SSD дисков – 3%. Размер резерва может быть изменен в процессе эксплуатации системы.

2.2.2.2.4. Изменение пула

2.2.2.2.4.1. Добавление в пул запасного диска

При создании пула, рекомендуется подключить к нему запасной (hot spare) диск. При выходе из строя одного из дисков пула, запасной диск автоматически подключается и запускается процесс восстановления.

Добавление запасного диска можно выполнить только при исправном пуле. Если пул изменит статус на Degraded, возможность добавления запасных дисков будет заблокирована.

Для добавления в пул запасного диска выполните следующие действия:

1. Перейдите на вкладку «Дисковое пространство» → «Пулы»;
2. В окне «Пулы» разверните поле свойств выбранного пула;
3. Нажмите на кнопку «Управление дисками», откроется окно управления дисками пула;
4. Нажмите на кнопку «Добавить запасные диски», откроется окно добавления запасных дисков (см. Рисунок 81);
5. Выберите нужное количество дисков из списка и нажмите кнопку «Добавить».

При повторном появлении в системе, автоматически замененного диска, система будет считать его «свободным».

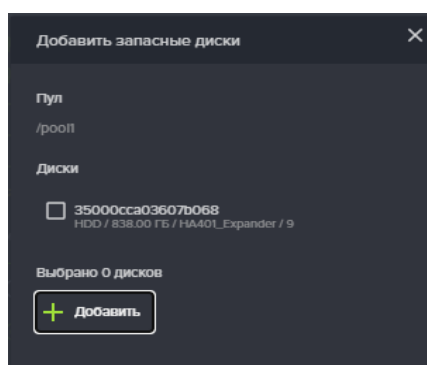


Рисунок 81. Добавление в пул запасного диска.



Внимание! Подключенными в качестве запасных могут быть только диски аналогичные используемым в пуле.



Внимание! Система не позволит добавить запасной диск к поврежденному пулу (статус *DEGRADED*). Подключайте запасные диски заранее!

2.2.2.2.4.2. Замена диска в пуле на запасной

Если в состав пула добавлены запасные диски, то появляется возможность замены любого из дисков пула на запасной диск. Для замены диска выполните следующие действия:

1. Перейдите на вкладку «Дисковое пространство» → «Пулы»;
2. В окне «Пулы» разверните поле свойств выбранного пула, в котором собираетесь заменить диск;
3. Нажмите на кнопку «Управление дисками», откроется окно управления дисками пула (см. Рисунок 82);

4. Выберите заменяемый диск из списка;
 5. Нажмите на кнопку «**Заменить**»;
 6. В появившемся окне выберите из списка диск для замены, если такой имеется в системе. Если список пуст, то это указывает на отсутствие свободного диска *аналогичного* заменяемому;
 7. Нажмите на кнопку «**Заменить**».
- При просмотре свойства пула будет показан **новый** диск.

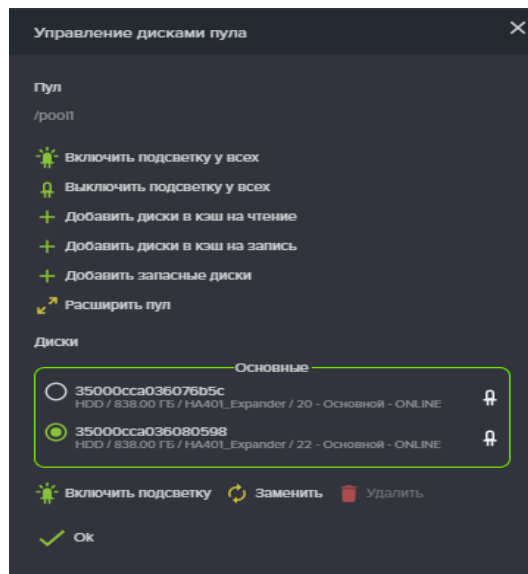


Рисунок 82. Окно управления дисками пула.

При необходимости замены нескольких дисков, необходимо повторить процесс последовательно для каждого диска, дожидаясь окончания восстановления целостности пула после выполнения каждой замены диска.



Внимание! Замена диска производится только на *аналогичный* диск.

2.2.2.2.4.3. Удаление из пула запасного диска

Для удаления из пула запасного диска выполните следующие действия:

1. Перейдите на вкладку «**Дисковое пространство**» → «**Пулы**»;
2. В окне «**Пулы**» разверните поле свойств выбранного дискового пула;
3. Нажмите на кнопку «**Управление дисками**», откроется окно управления дисками пула;
4. В списке выберите запасной диск, для его удаления (см. Рисунок 83);
5. Нажмите на кнопку «**Удалить**»;
6. Подтвердите удаление введя слова «**ok**» в появившемся окне.

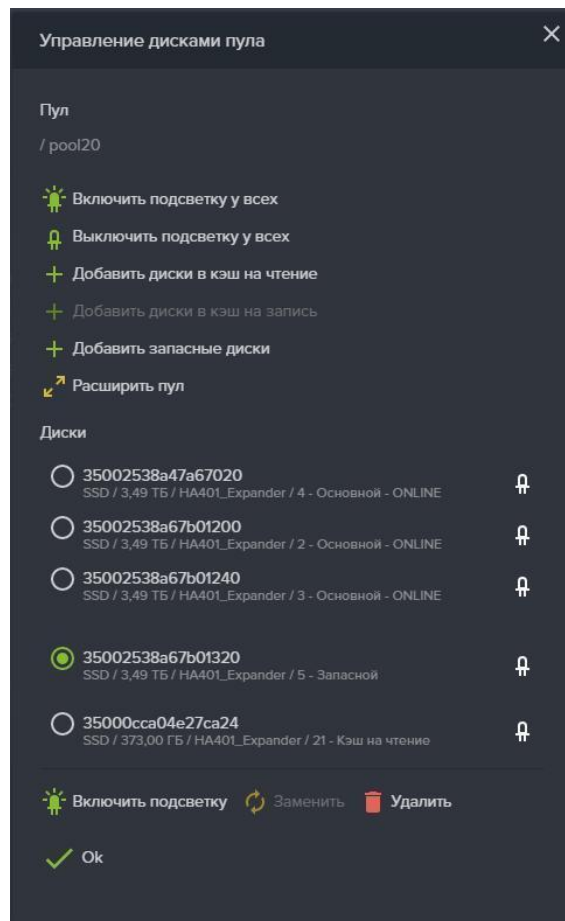



Рисунок 83. Удаление из пула запасного диска.

В результате в области уведомления появится новая запись о успешно выполненной операции («Из пула «/имя пула» успешно удален запасной диск»), а при просмотре информации о массиве удаленный диск исчезнет из списка.

2.2.2.2.4.4. Включение/отключение подсветки диска в полке

Для управления подсветкой дисков в полке выполните следующие действия:

1. Перейдите на вкладку «Дисковое пространство» → «Диски»;
2. В окне «Диски» разверните поле свойств диска, на котором нужно включить индикацию. Так же можно включить индикацию как на дисках в пуле и свободных дисках, так и на дисках в кэше на запись и на чтение;
3. Нажмите на кнопку «**Включить подсветку**» (см. Рисунок 84);
4. Также включение подсветки диска произойдет  при нажатии на значок, расположенный на каждой строке в списке дисков. Если подсветка диска включена, цвет значка изменится на зелёный. Повторное нажатие отключает подсветку.

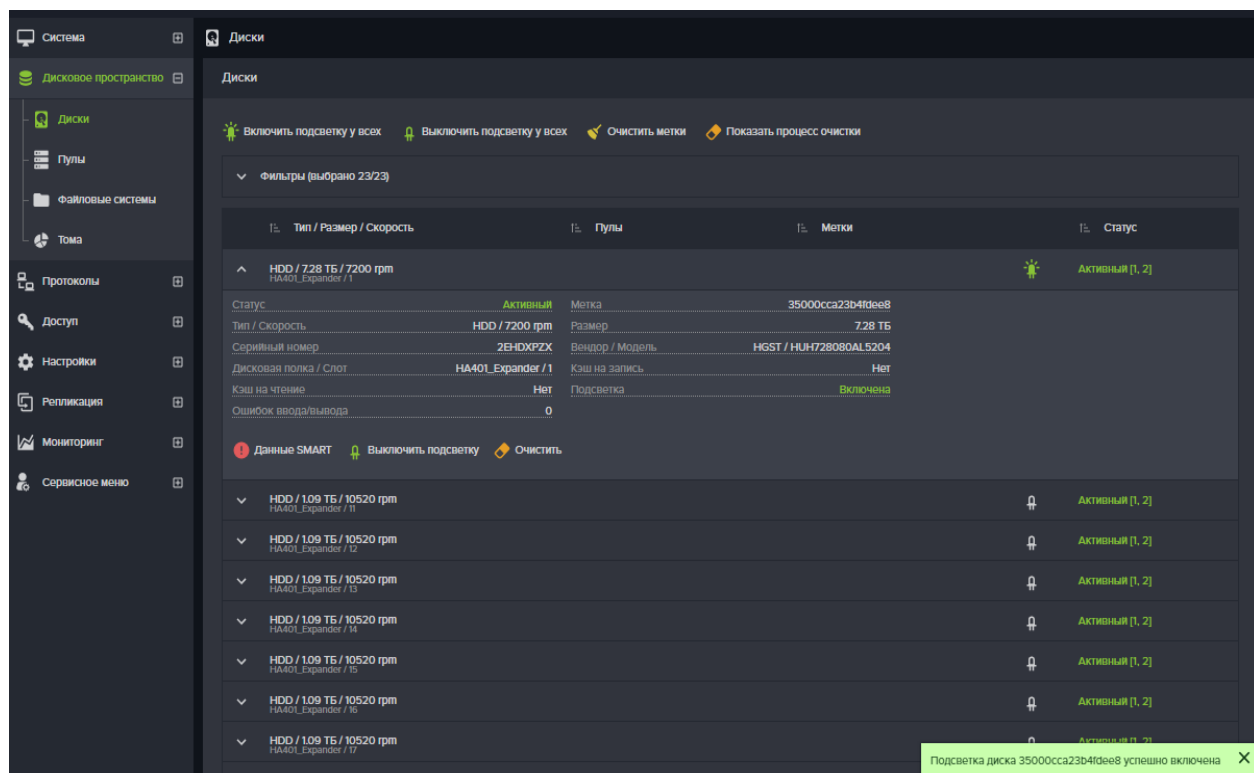


Рисунок 84. Включение/отключение подсветки в полке.

Для включения подсветки всех дисков нажмите на кнопку «**Включить подсветку у всех**», расположенную вверху вкладки меню «**Диски**».

Индикацию диска можно также включить из вкладки «**Дисковое пространство**» → «**Пулы**». Для чего выполните следующие действия:

1. Разверните в окне «**Пулы**» поле свойств пула, в котором необходимо подсветить диск;
2. Нажмите на кнопку «**Управление дисками**»;
3. Выберите нужный диск;
4. Нажмите на кнопку «**Включить подсветку**» внизу списка, или «**Включить подсветку у всех**» сверху. В первом случае включится подсветка у одного диска пула, во втором – подсветятся все диски пула.

Проверить статус индикации можно во вкладке меню «**Диски**».

2.2.2.2.4.5. Замена диска на свободный диск

Имеется возможность заменить диск в составе пула на аналогичный свободный диск (не входящий в состав других пулов). Функция замены диска требуется в следующих случаях:

- Необходимо заменить один исправный основной диск на другой исправный диск;

- Отказ диска (или извлечение диска) и к пулу не был подключен запасной диск (hot spare);

- Ошибка S.M.A.R.T (при таких ошибках система уведомит Администратора о необходимости замены диска, но автоматической замены на запасной диск не произойдет).

Для выполнения замены диска во всех этих случаях потребуется свободный диск аналогичной модели.

Выполните следующие действия:

1. Перейдите в меню «Дисковое пространство»-«Пулы»;
2. Выберите нужный пул и в панели свойств нажмите на кнопку «Управление дисками»;
3. В окне дисков пула выберите требующий замены диск и нажмите на кнопку «Заменить»;
4. В открывшемся окне выберите из списка диск для замены и нажмите на кнопку «Заменить». Произойдет замена диска.

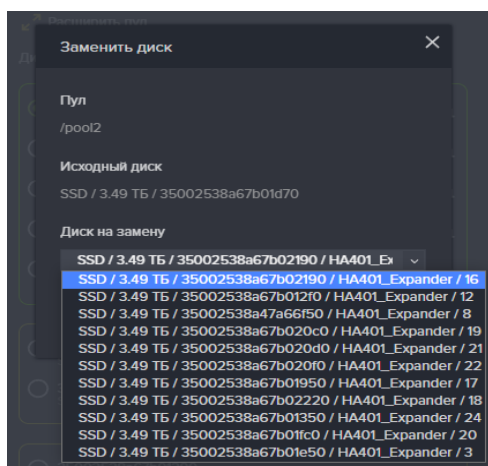


Рисунок 85. Замена диска на свободный диск.

2.2.2.2.4.6. Экспорт пула

Экспорт пула может понадобиться для физического переноса дисков составляющих пул на другую систему. Диски экспортированного пула сохраняют метки пула, по которым возможен дальнейший импорт этого пула.

Для экспорта имеющегося пула выполнить следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Пулы»;
2. В окне «Пулы» разверните поле свойств выбранного дискового пула;
3. Нажмите на кнопку «Экспортировать»;
4. Введите в появившемся окне слово **ok** и нажмите «Подтвердить».

После операции экспортирования, пул исчезнет из списка пулов, но все данные на нем останутся.

2.2.2.2.4.7. Импорт пула

Для импорта пула выполнить следующие действия:

1. Перейдите на вкладку меню «**Дисковое пространство**» → «**Пулы**»;
2. Нажать на кнопку «**Импортировать пул**»;
3. В появившемся окне (см. Рисунок 86) выберите из списка нужный пул и задайте имя под которым он будет отображаться в системе.
4. Выбрать контроллер, который будет управлять пулом;
5. Нажмите на кнопку «**Импорт**».

После завершения операции импортированный пул появится в списке пулов.

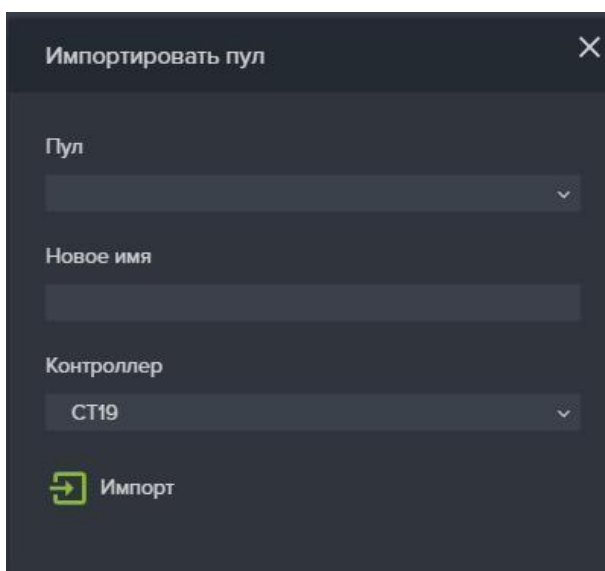


Рисунок 86. Импортировать пул.

2.2.2.2.4.8. Добавление дисков в пул (расширение пула)

В панели свойств пула можно увидеть список созданных томов/файловых систем, в котором будет название тома/файловой системы, тип, объем в абсолютной величине и % от общего объема пула. В случае, если суммарный объем томов и файловых систем будет превышать объем пула при последующем создании тома или файловой системы будет выдаваться предупреждение о возможности переполнения пула. Что бы избежать подобной ситуации имеется возможность расширить объем пула за счет добавления новых дисков.

Для расширения имеющегося дискового пула выполните следующие действия:

1. Перейдите на вкладку меню «**Дисковое пространство**» → «**Пулы**»;
2. Выберите пул, который должен быть расширен, открыв его поле свойств;

3. Нажмите на кнопку «**Управление дисками**», откроется окно управления дисками пула;
4. Нажать на кнопку «**Расширить пул**», откроется одноименное окно (см. Рисунок 87);
5. Выберите диски, которые необходимо добавить в пул. Для удобства выбора необходимого для расширения пула количества дисков, используйте флажок «**Выбрать автоматически**»;
6. Нажмите на кнопку «**Расширить**».

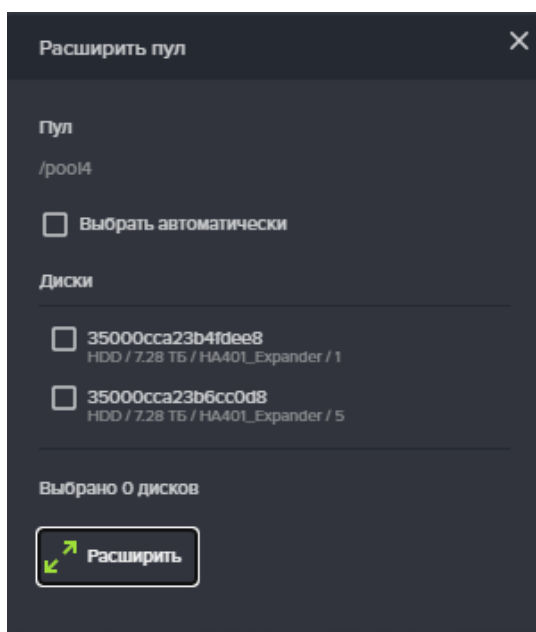


Рисунок 87. Добавление дисков в пул.

В результате в области уведомления появится новая запись о успешно выполненной операции («Пул </имя пула> успешно расширен»).

При просмотре свойств пула будут показаны добавленные физические диски, имеющий тип «основной».



Внимание! Данная операция невозможна для пулов уровня RAID0.



Внимание! Обратная операция невозможна. Для освобождения дисков массив должен быть удалён. Данные будут потеряны.

Расширение **обычного пула** происходит не на одиночный диск, а группами. Количество дисков в дополнительной группе должно равняться исходному количеству дисков в исходной группе. Диски должны быть идентичны дискам в пуле. В дальнейшем, полученное пространство пула может быть использовано либо для создания новых томов/ файловых систем, либо для расширения уже существующих.

Расширение **быстрого пула** может быть выполнено на произвольное количество дисков, но только в том случае, если на полученном дополнительном пространстве планируется создание новых томов/файловых систем. В случае, если планируется расширение уже имеющихся томов/ФС, то для расширения пула должно использоваться количество дисков, равное имеющемуся в пуле.

Для расширения быстрого пула уровня RAID10 требуется такое же количество дисков, какое изначально используется в пуле.

2.2.2.2.4.9. Изменение резервирования свободного места на пуле

Резервирование используется только для обычных пулов. Резервирование свободного места на пуле служит для предотвращения падения скорости записи, когда пул практически заполнен. Заполнение пула может произойти неожиданно, например, после создания очередного снимка и быстрого изменения (удаления) данных на томе. В таком случае, благодаря заданному резерву свободного места, у файловой системы СХД всегда будут пустые блоки для записи новых данных и не придется тратить время на поиск освобожденных блоков.

Изначально, при создании пула, задается резерв в 10% от его объема. Это значение можно изменить в свойствах пула, нажав на кнопку **«Резервирование свободного места на пуле»**. Окно редактирования размера резерва представлено на рисунке ниже.

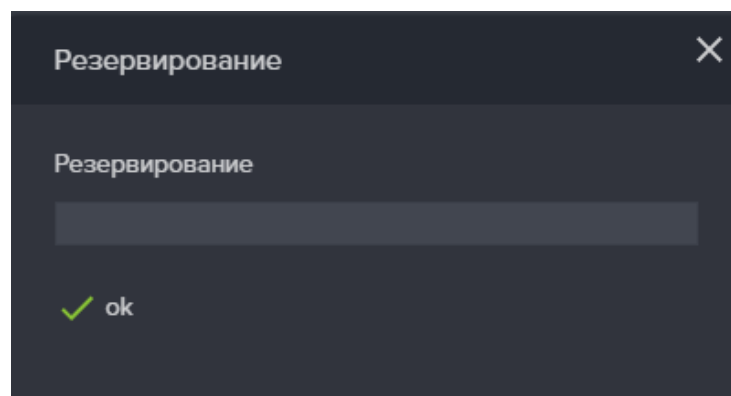


Рисунок 88. Окно редактирование размера резерва

Изменить размер резерва можно в пределах от 1% до 20%. Для магнитных дисков рекомендуемый размер резерва – 10%. Для SSD и NVME дисков – 3%.

Если на пуле недостаточно свободного места, увеличить резерв не получится.

2.2.2.2.5. Удаление пула

Перед удалением пула убедитесь, что на нём нет томов или файловых систем, которые будут препятствовать его удалению. Посмотреть какие тома или файловые системы созданы на пуле можно в панели его свойств.

Для удаления пула выполните следующие действия:

перейдите на вкладку меню «Дисковое пространство» → «Пулы» (см. Рисунок 89):

1. Разверните поле свойств выбранного пула;
2. Нажмите на кнопку «Удалить»;
3. Подтвердите удаление, введя в открывшемся окне слова «ok» и нажав кнопку «Удалить».

При успешном удалении пула, в области уведомления появится сообщение «пул /имя_пула удален» и удаленный пул исчезнет из списка пулов.



Внимание! Удаление пула невозможно, если на нем имеются созданные тома или файловые системы.

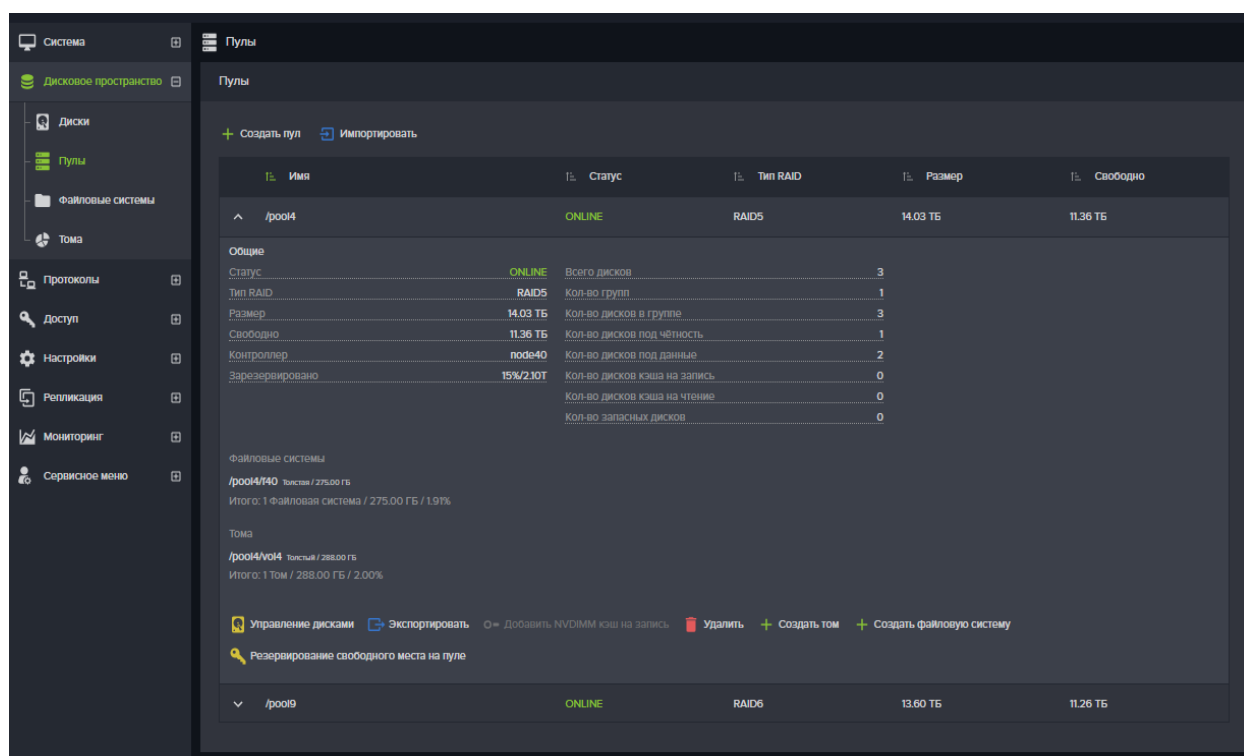


Рисунок 89. Вкладка меню «Пулы».

2.2.2.2.6. Статусы пула

Статус пула отображает его состояние. Параметр «Статус» может принимать следующие значения:

- **ONLINE** – пул подключен и работоспособен;
- **DEGRADED** – пул потерял избыточность, но по-прежнему работоспособен;
- **LOST** – пул не доступен;
- **RESILVIRING** – пул восстанавливает избыточность;
- **REPAIRING** – пул проверяет целостность данных;
- **FAULTED/ UNAVAIL** – пул не исправен;
- **FREEING** – выполняется фоновая очистка после удаления тома или ФС.

Статус **ONLINE** сигнализирует о нормальной работе пула. В рабочей системе все пулы должны иметь такой статус.

Статус **DEGRADED** сигнализирует о потере избыточности данных пула, из-за отказа одного или нескольких накопителей. При наличии у пула запасных дисков, отказавший диск будет автоматически выведен из состава пула и заменен запасным. После чего включится процесс восстановления избыточности пула.

Статус **LOST** сигнализирует о потере системой пула. Это может возникнуть при физической недоступности дисков пула (например, при отключении дисковой полки) либо при разрушении пула в следствии отказа большего числа накопителей, чем выдерживает заданный уровень избыточности пула.

Статусы **RESILVIRING** и **REPAIRING** сигнализируют о начавшемся процессе проверки целостности данных и восстановлении избыточности пула. В зависимости от нагрузки и конфигурации пула, этот режим может занять разное по длительности время. После удачного восстановления статус пула должен измениться на **ONLINE**.

Статус **FAULTED** или **UNAVAIL** сигнализирует о разрушении пула и его неработоспособности. Необходимо предпринять действия по восстановлению информации и работоспособности пула. Для этого необходимо обратиться в службу технической поддержки.

Статус **FREEING** сигнализирует о выполнении фонового процесса очистки данных на пуле. Это происходит автоматически после удаления с пула тома или файловой системы. Процесс фоновой очистки может занимать длительное время, зависящее от величины нагрузки на пул. По мере выполнения фоновой очистки на пуле, будет увеличиваться размер свободного места. Выполняющийся процесс фоновой очистки не предполагает каких-либо ограничений в использовании пула.

2.2.2.2.7. Миграция ресурсов

Вкладка «Миграция» используется для перемещения ресурсов с одного контроллера на другой при возврате ресурсов после автоматически отработавшей

миграции либо с целью проведения сервисных работ на каком-либо из контроллеров (см. Рисунок 90).

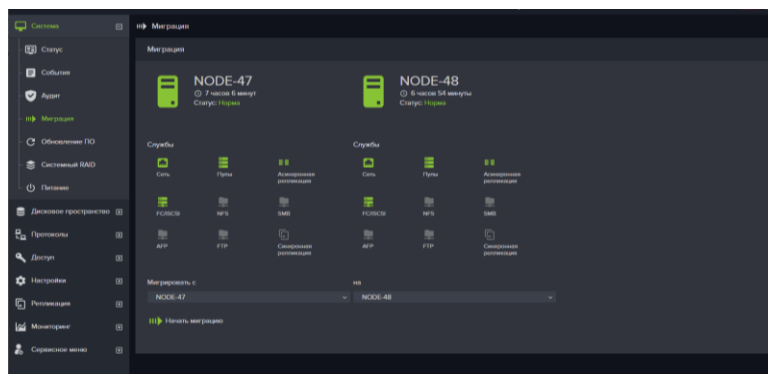


Рисунок 90. Вкладка меню Миграция



Внимание! При миграции происходит временная потеря связи клиентов с СХД (если МРЮ на клиенте не применим или не настроен).

Для запуска процесса миграции выполните следующие действия:

1. Перейдите на вкладку меню «Система» → «Миграция»;
2. В выпадающих списках «Мигрировать с» и «на» задайте направление миграции;
3. Нажмите на кнопку «Начать миграцию»;

При этом интерфейс блокируется на время выполнения операции, которая обычно занимает 1-3 минут.

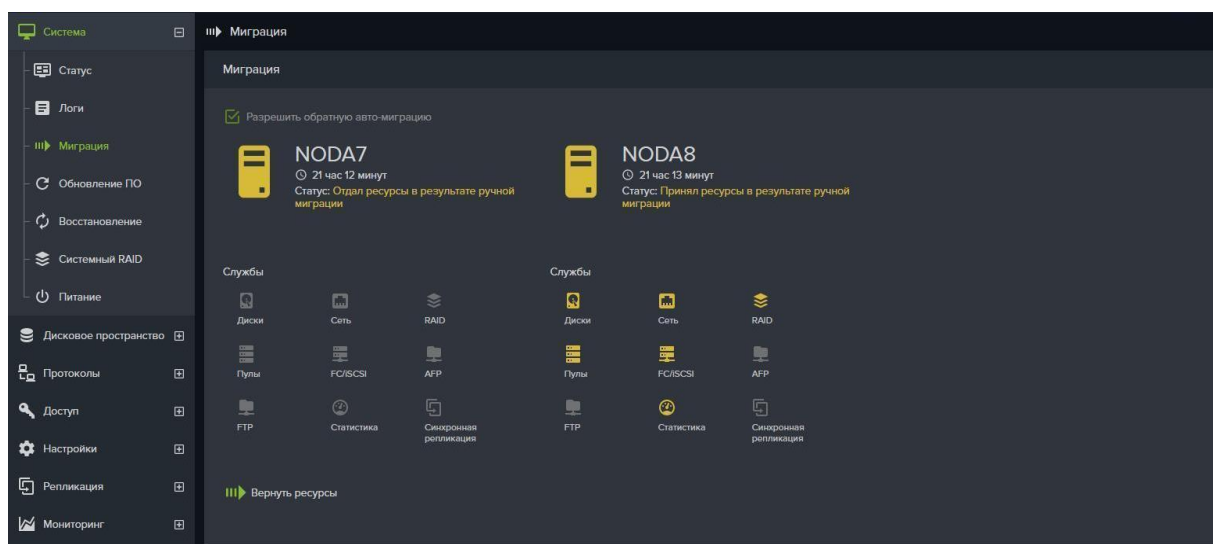


Рисунок 91. Окончание миграции ресурсов.

По окончании процесса миграции на вкладке отображаются новые статусы контроллеров: «Отдал ресурсы» и «Принял ресурсы», в зависимости от того, на какой из контроллеров были перенесены ресурсы (см. Рисунок 91).

В редких случаях процесс миграции может занимать более длительное время. Это связано с работой служб, которым нужно некоторое время, для того чтобы корректно отключить ресурсы.

Для возврата ресурсов следует нажать кнопку «Вернуть ресурсы».

После выполнения миграции на системе блокируется создание ресурсов, при этом некоторые пункты меню становятся неактивны. После ручного возврата ресурсов в первоначальное состояние, эти пункты меню вновь станут активны.

При отключении или перезагрузке одного из контроллеров, миграция его ресурсов на работающий контроллер начнется автоматически. После восстановления работоспособности контроллера, его ресурсы возвращаются вручную. Для этого следует нажать на кнопку «Вернуть ресурсы» (см. Рисунок 91).



Внимание! Если в системе выполняется перестроение пула (*rebuild, resilvering*), время проведения ручной миграции может занять значительное время. При этом с высокой вероятностью произойдет отключение ресурсов.

Рекомендуется дождаться окончания перестроения пула перед выполнением ручной миграции ресурсов, чтобы исключить её долгое выполнение.

2.2.2.2.8. Защита данных от повреждений

Для защиты записанных на пуле данных от повреждений, средствами файловой системы осуществляется подсчет и хранение контрольных сумм блоков данных. При чтении блоков данных, их контрольная сумма вычисляется и сверяется с сохраненной. При её несовпадении контрольных сумм осуществляется автоматическое восстановление блока на основе избыточных данных.

2.2.2.3. Работа с кэш-памятью (Данный раздел относится только к настройке обычных пулов)

2.2.2.3.1. Настройки кэша чтения.

2.2.2.3.1.1. Добавление к пулу кэша чтения.

Одноуровневое кэширование операций чтения осуществляется на уровне блоков файловой системы. В качестве контейнера для кэша используется свободная оперативная память контроллера, для всех пулов системы используется общий кэш. Данный функционал используется по умолчанию, и не может быть изменен.

Двухуровневое кэширование осуществляется на уровне блоков файловой системы. Двухуровневое кэширование в качестве контейнера для кэша использует как оперативную память, так и SSD диски, которые подключаются к конкретному пулу.

Двухуровневое кэширование применимо только к обычным пулам. Для быстрых пулов дисковый кэш не применяется.

Для подключения к пулу дисков под кэш чтения второго уровня выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Пулы»;

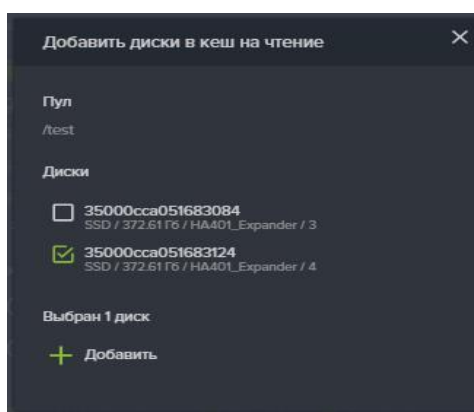


Рисунок 92. Добавление в пул кэш на чтение.

2. Разверните панель свойств пула, к которому будет добавлен диск в кэш на чтение;
3. Нажмите на кнопку «Управление дисками»;
4. Нажмите на кнопку «Добавить диски в кэш на чтение», откроется одноименное окно (см. Рисунок 92);
5. Выберите **SSD диски**, которые нужно добавить в кэш;
6. Нажмите кнопку «Добавить».

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции («В пул «/имя пула» успешно добавлены диски кэша для чтения»).

2.2.2.3.1.2. Удаление из пула кэша чтения.

Для удаления из пула кэша на чтения выполните следующие действия:

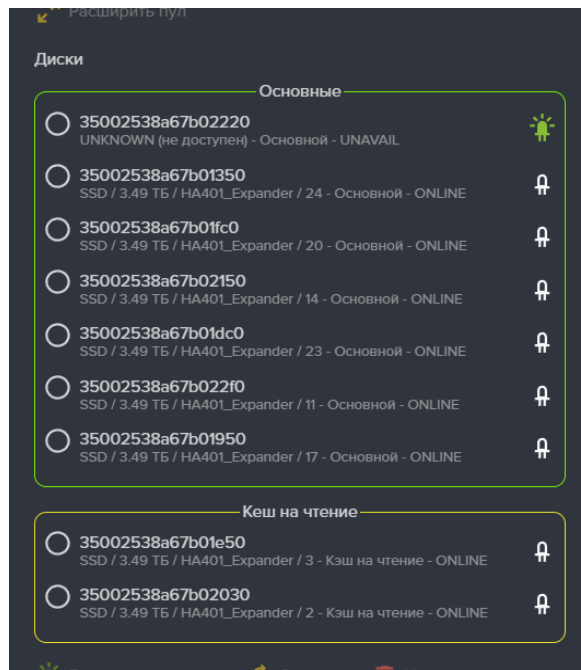


Рисунок 93. Удаление из пула кэша на чтение.

1. Перейдите на вкладку меню «Дисковое пространство» → «Пулы»;
2. Разверните в окне «Пулы» панель свойств пула, из которого будет удален кэш на чтение;
3. Нажмите на кнопку «Управление дисками», откроется окно управления дисками пула (см. Рисунок 93);
4. В разделе «Диски» выберите диск, помеченный как «Кэш для чтения»;
5. Нажать на кнопку «Удалить»;
6. Подтвердите удаление, введя в открывшемся окне слово «ok» и нажав кнопку «Удалить».

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции («Из пула </имя пула> успешно удален диск кэша для чтения»).

2.2.2.3.2. Настройки кэша записи.

Для кэширования операций записи может использоваться энергонезависимая память NVRAM. Если в системе установлен модуль NVDIMM, появляется возможность подключения к пулу NVDIMM кэша, вместо SSD кэша. При этом для разных пулов можно использовать различные типы кэширования: SSD или NVDIMM.

Рекомендуется для ускорения записи всегда добавлять к обычным пулам кэш на запись!

2.2.2.3.2.1. Добавление к пулу NVDIMM кэша

Для подключения NVDIMM кэша к выбранному пулу выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Пулы» (см. Рисунок 94);
2. В окне «Пулы» разверните панель свойств пула, к которому будет добавлен кэш на запись;
3. Нажмите кнопку «Добавить NVDIMM кэш на запись». В списке параметров пула, параметр «Количество дисков кэша на запись» изменит значение на «NVDIMM кэш подключен», а название кнопки изменится на «Отключить NVDIMM кэш на запись»;

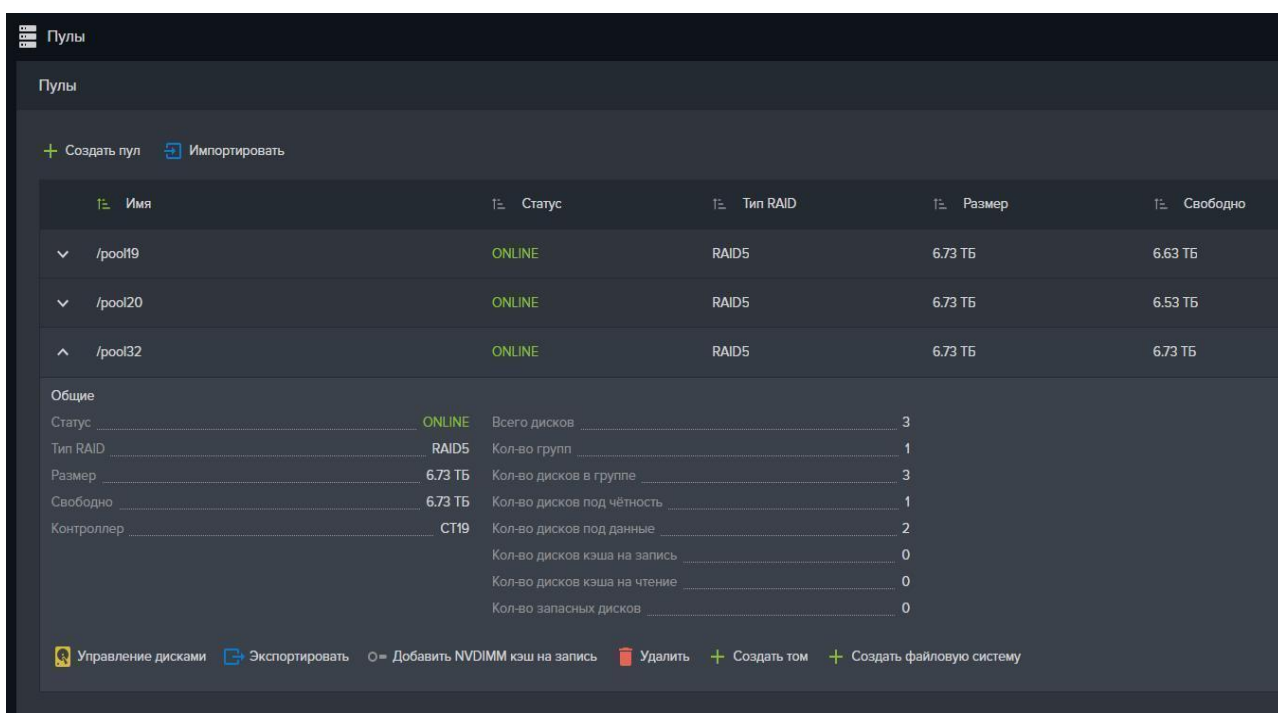


Рисунок 94. Вкладка меню «Пулы».

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции

2.2.2.3.2.2. Удаление из пула NVDIMM кэша

Для удаления из пула кэша на запись выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Пулы»;
2. В окне «Пулы» разверните панель свойств пула, в котором будет отключен кэш на запись;

3. Нажмите на кнопку «Отключить NVDIMM кэш на запись», надпись «NVDIMM кэш подключен» исчезнет, а название кнопки изменится на «Включить кэш на запись»;

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции.

2.2.2.3.2.3. Подготовка SSD дисков для кэша записи.

Перед тем, как подключать кэш записи к пулам, необходимо предварительно выполнить разметку SSD дисков как кэш на запись. При этом на дисках будут созданы несколько разделов, количество которых зависит от размера дисков, но не более 16.

Для разметки дисков под кэш для записи выполните следующие действия:

1. Перейдите на вкладку «Дисковое пространство» → «Диски»;
2. Разверните панель свойств диска, который необходимо пометить под кэш на запись;
3. Нажмите на кнопку «Пометить как кэш для записи» (см. Рисунок 95).

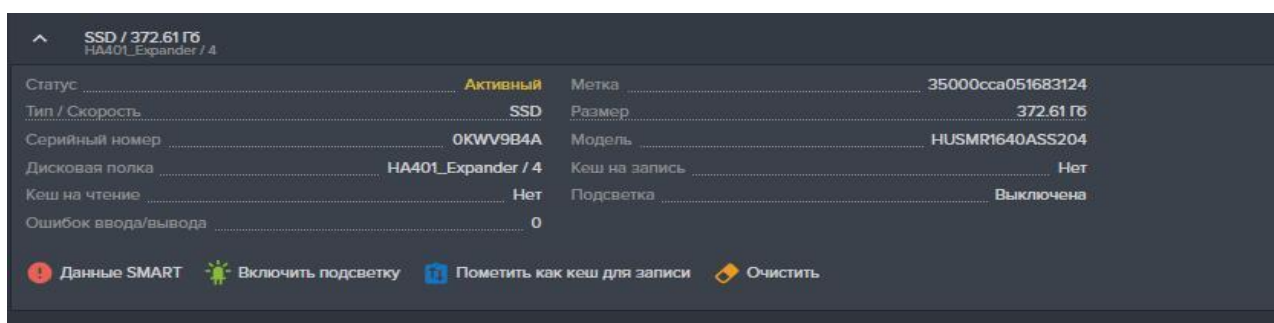


Рисунок 95. Панель свойств диска.

В результате в области уведомления появится новая запись о успешно выполненной операции («Диск <имя диска> помечен как кэш для записи»), а в колонке «Кэш для записи» появится пометка «Да».



Внимание! Диск можно добавить к пулу как кэш для записи, только если он размечен как «Кэш на запись».

2.2.2.3.2.4. Добавление в пул SSD кэша

Для подключения к пулу кэша на запись необходимо два физических диска для обеспечения его защищенности. Добавлять диски разрешается только парами. Для лучшей производительности рекомендуется не разделять диски кэша записи между несколькими пулами, к каждому пулу подключайте отдельную пару дисков.

Для добавления кэша на запись к определённому пулу выполните следующие действия:

4. Перейдите на вкладку меню «Дисковое пространство» → «Пулы»;
5. В окне «Пулы» разверните панель свойств пула, к которому будет добавлен кэш на запись;
6. Нажмите кнопку «Управление дисками», откроется одноименное окно;
7. Нажмите на кнопку «Добавить диски в кэш на запись». В этом окне отображаться только диски, заранее помеченные как кэш на запись;
8. Выберите 2 диска;
9. Нажмите на кнопку «Добавить»(см. Рисунок 96).

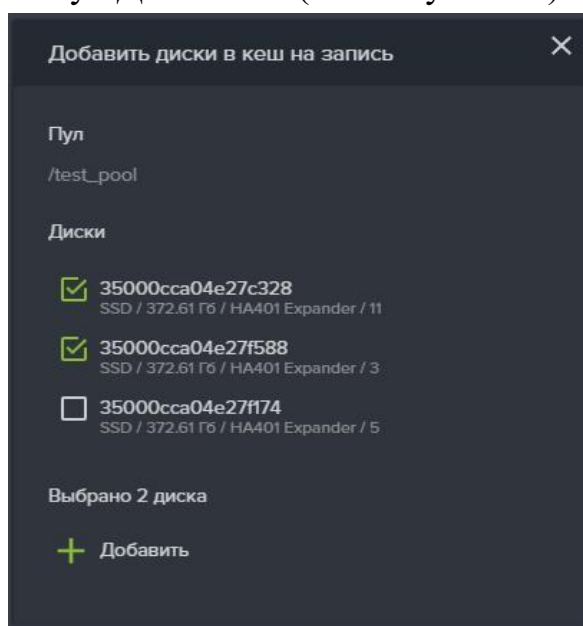


Рисунок 96. Добавление в пул кэша на запись.

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции («В пул </имя пула> успешно добавлены диски кэша для записи»).



Внимание! При разметке диска в качестве кэша записи, он разделяется на несколько равных разделов. Для надежного хранения данных, кэш записи использует пару разделов для каждого пула. При подключении к пулу, кэш записи использует по одному разделу с первого и второго диска, создавая из них зеркала RAID1, т.е. к пулу должны быть подключены два разных диска! Допускается использование для кэша записи трёх дисков, подключенных к разным пулам, где один диск общий для различных кэшей записи.



Внимание! Для кэша требуется использовать только SSD диски, при этом пара дисков размеченная как кэш на запись может быть назначена нескольким пулам, но не более чем четырём пулам. Для получения лучшей производительности рекомендуется для каждого пула подключать отдельную пару дисков.

2.2.2.3.2.5. Удаление из пула SSD кэша

Для удаления из пула кэша на запись выполните следующие действия:

4. Перейдите на вкладку меню «**Дисковое пространство**» → «**Пулы**»;
5. В окне «**Пулы**» разверните панель свойств пула, в котором будет отключен кэш на запись;
6. Нажмите на кнопку «**Управление дисками**», откроется одноименное окно;
7. Выберите любой из SSD дисков помеченных как «**Кэш для записи**» (см. Рисунок 97);
8. Нажмите на кнопку «**Удалить**»;
9. Подтвердите удаление, введя в открывшемся окне слова «**ok**» и нажмите кнопку «**Удалить**».

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции («Из пула </имя пула> успешно удалены диски кэша для записи»).



Рисунок 97. Удаление из пула кэша на запись

При просмотре свойств пула не будет показано дисков, имеющих тип «кэш на запись».

2.2.2.3.2.6. Удаление разметки SSD дисков, помеченных как кэш на запись.

Если диски размечены под кэш записи, то их нельзя использовать в пуле как кэш для чтения, или для хранения на них данных. Для освобождения SSD дисков выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Диски»;
2. В окне «Диски» разверните панель свойств диска, на котором необходимо удалить метку «кэш на запись»;
3. Нажмите на кнопку «Пометить как свободный» (см. Рисунок 98).

Если в панели свойств диска кнопка «Пометить как свободный» отсутствует - это означает, что диск подключен к какому-либо пулу. После отключения диска от всех привязанных к нему пулов кнопка должна появиться.

В результате выполненных действий, в области уведомления появится новая запись о успешно выполненной операции («Диск <имя диска> помечен как свободный»). В свойствах диска, в строке «Кэш на запись» появится пометка «Нет».

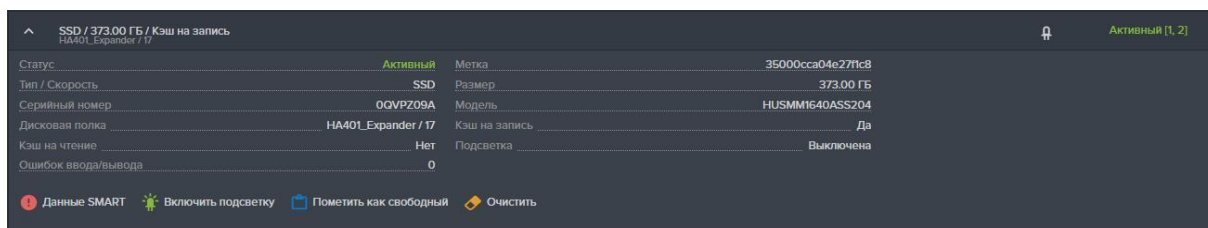


Рисунок 98. Панель свойств диска



Внимание! Диск, используемый как кэш на запись нельзя пометить как свободный, пока он подключен к другим пулам. Перед освобождением используемых как кэш на запись дисков, удалите их из всех пулов.

2.2.2.4. Работа с томами

При создании тома, необходимо иметь в виду, что созданные впоследствии снимки и клоны этого тома также будут размещены на том же пуле. По мере разрастания размера снимка, он может со временем, теоретически, увеличиться до размера тома.

При создании томов и файловых систем, рекомендуется оставлять свободным некоторое количество свободного места на пуле для будущих снимков и клонов.

В этой версии ПО, тома созданные на быстрых пулах не поддерживают создания снимков, эта возможность будет реализована в следующих версиях.



Внимание! Имя тома не может начинаться с буквы «с» (си). Том с таким именем создать не получится.

2.2.2.4.1. Создание тома

Том создается на существующем пуле. Том может быть создан «толстым», занимающий фиксированный объем на пуле и «тонким», увеличивающий свой размер по мере заполнения данными.

В этой версии ПО создание тонких томов на быстрых пулах не поддерживается!

Для создания тома выполните следующие действия:

1. Перейдите на вкладку меню «Дисконное пространство» → «Тома» или «Дисконное пространство» → «Пулы» и разверните панель свойств выбранного пула.
2. Нажмите на кнопку «Создать том», откроется одноименное окно (см. Рисунок 99).
3. Введите параметры тома:

- a. Введите имя тома;
 - b. Выберите пул из списка;
 - c. Задайте размер тома;
 - d. Выберите размер блока из выпадающего меню;
 - e. Выбрать «тонкий» или «толстый» тип тома;
4. Нажать кнопку «Создать».

После удачного выполнения операции в списке томов появится новый том.

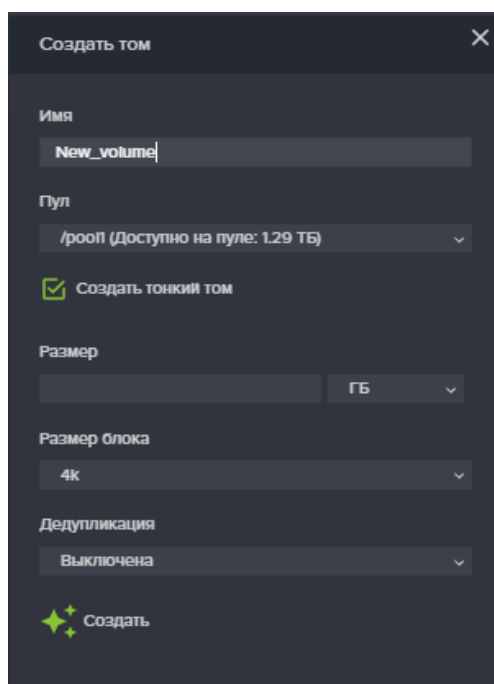


Рисунок 99. Создание тома.

2.2.2.4.2. Модификация тома/файловой системы.

Модификация тома или файловой системы для всех протоколов выполняется следующим образом:

1. Выберите вкладку меню «Дисковое пространство» → «Томы (Файловые системы)»
2. Разверните поле свойств тома и выберите пункт «Редактировать», откроется окно «Редактировать том».
3. Редактированию подлежат следующие параметры:
 - Размер.
4. Измените параметры и нажмите на кнопку «Сохранить».

2.2.2.4.3. Удаление тома.

Перед удалением тома удостоверьтесь, что у него отсутствуют зависимые снимки и клоны. Для этого в панели свойств тома, служат параметры «Количество

снимков» и «Количество клонов». Если у тома имеются снимки и клоны их необходимо удалить перед удалением тома.

Для удаления тома выполните следующие действия:

1. Выберите вкладку меню («Дисковое пространство» → «Том»);
2. Разверните панель свойств выбранного тома и нажмите на кнопку «Удалить»;
3. Подтвердите в открывшемся окне свои действия введя слова «ок» и нажмите кнопку «Удалить».

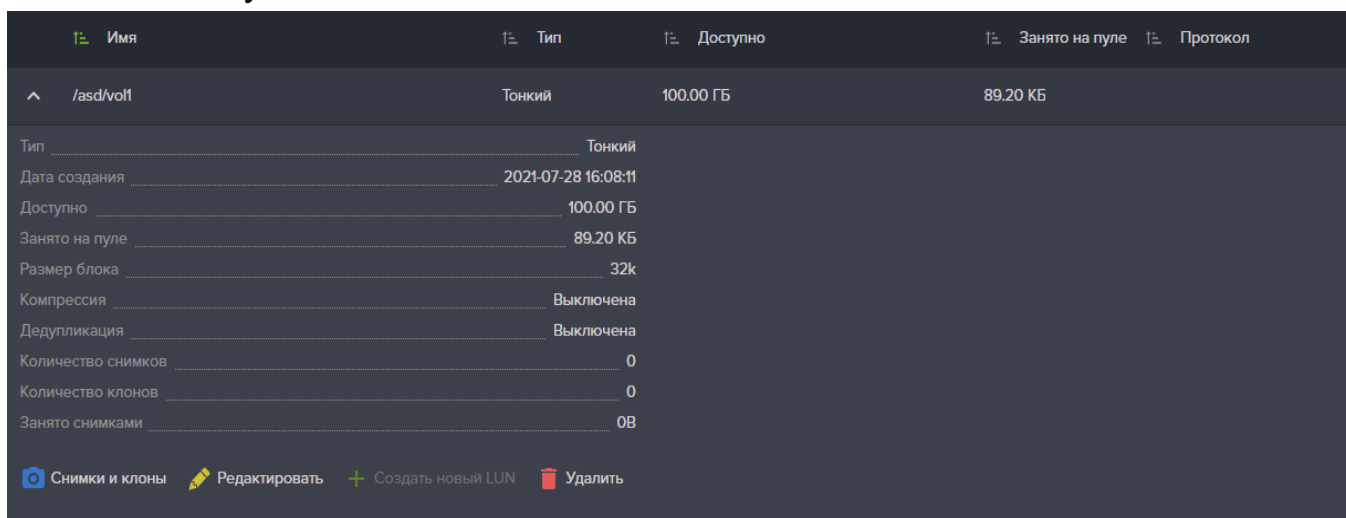


Рисунок 100. Панель свойств тома.

После удаления тома автоматически запустится процесс очистки данных на пуле, которые относились к удаленному тому. Это может занять некоторое время. О ходе очистки данных сигнализирует появления на верхней панели значка очистки пулов, как показано на рисунке ниже.

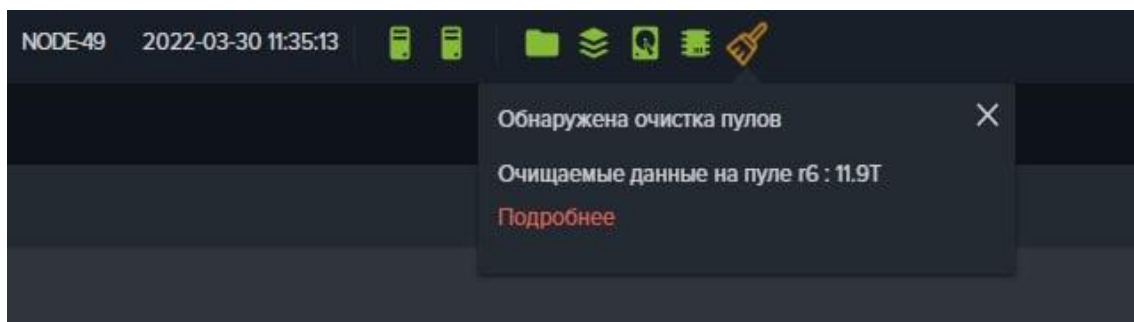


Рисунок 101. Значок очистки пулов



Внимание! Удаление тома или файловой системы возможно, только при отсутствие зависимых от него снимков/клонов.

2.2.2.4.4. Работа с томом по протоколу Fibre Channel (FC)

Инструментарий для настройки FC ресурсов находится на вкладке меню «Протоколы» → «FC».

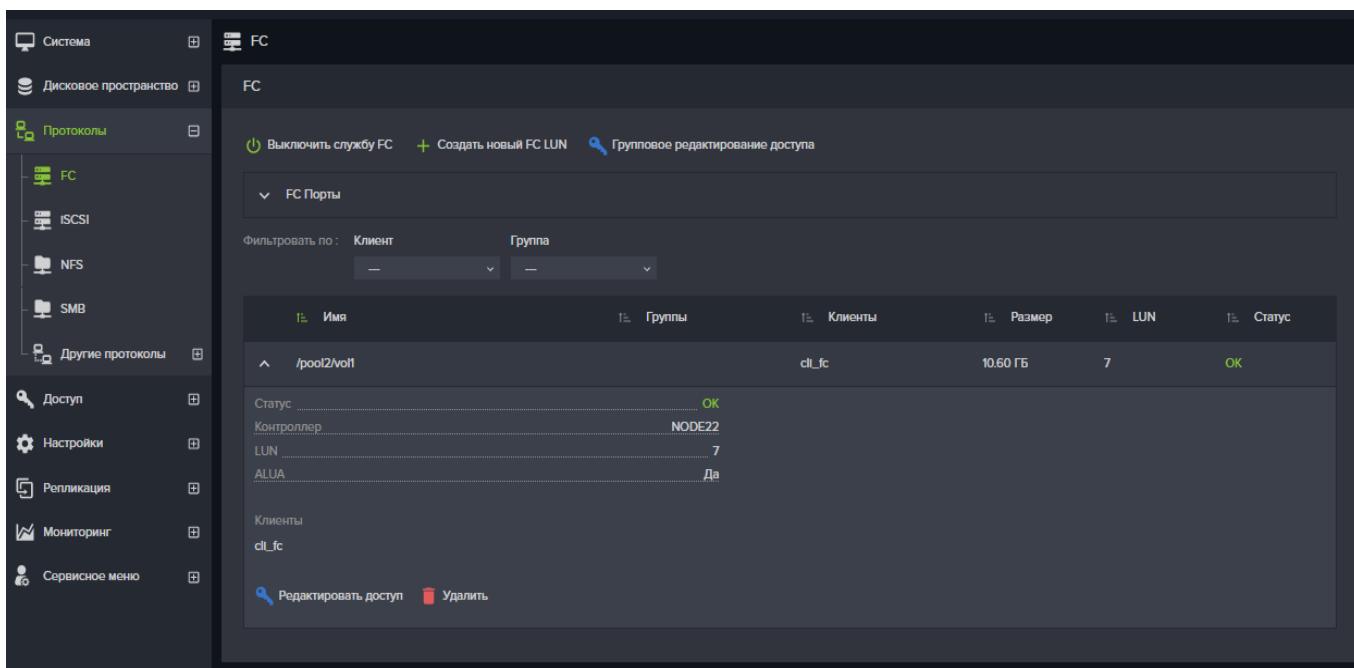


Рисунок 102. Окно настройки протокола FC

2.2.2.4.4.1. Создание FC LUN.

Перед созданием FC LUNов, необходимо включить службу FC. При выключенной службе кнопка «Создать новый FC LUN» будет недоступна!

Для создания FC луна выполните следующие действия:

1. Перейдите на вкладку меню «Протоколы» → «FC»;
2. Нажмите на кнопку «Создать новый FC LUN», откроется одноименное окно;

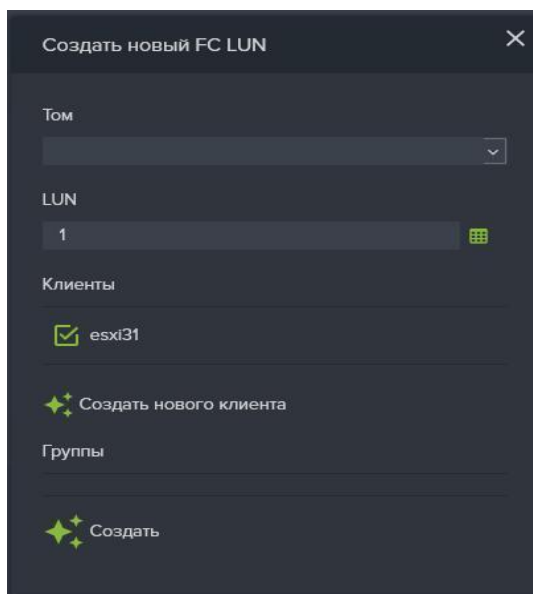



Рисунок 103. Создание луна Fibre Channel.

3. Выберите нужный том из выпадающего списка (если список пуст, то перейти к пункту «Создание тома»);

4. Введите номер LUN или выберите нужный номер из предлагаемых системой, нажав на кнопку, расположенную справа от поля ввода номера LUN;

5. Нажмите на кнопку **«Создать»**.

В результате выполненных действий, в списке лунов появится новый FC LUN.

Номер LUN можно задать в диапазоне от 1 до 1024. Номера после 254 могут некорректно работать на Windows системах. В окне быстрого выбора LUN (открывается по нажатию кнопки ) отображаются номера с 1 по 254.

2.2.2.4.4.2. Подключение клиентов к FC LUN.

Для предоставления доступа клиентам к FC LUN выполните следующие действия:

1. Перейдите на вкладку меню **«Протоколы»** → **«FC»**;

2. Выберите нужный FC LUN из списка, развернув его панель свойств;

3. Нажмите на кнопку **«Редактировать доступ»**, откроется одноименное окно (см. Рисунок 104);

4. Выберите нужных клиентов, отметив флажки перед именем в списке или создайте их нажав кнопку **«Создать нового клиента»**, для предоставления им доступа;

5. Нажмите на кнопку **«Сохранить»**

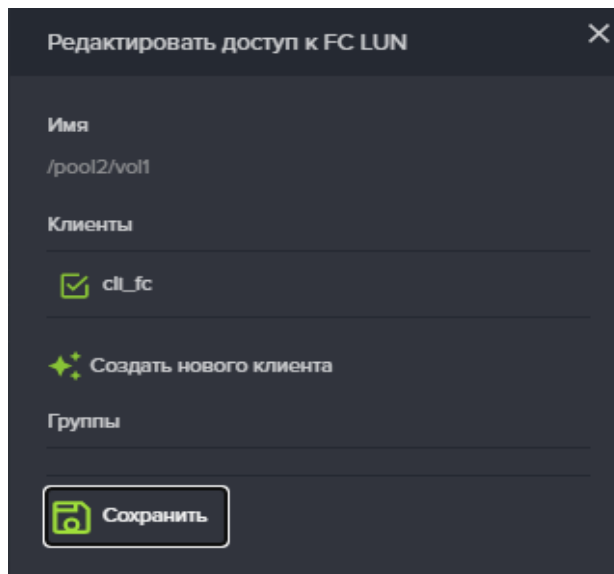


Рисунок 104. Редактирование доступа к FC луны

При удачном завершении операции в секции «Клиенты» панели свойств редактируемого луна появится имя клиента, которому был предоставлен доступ.

2.2.2.4.4.3. Групповое подключение клиентов к FC LUN

Для предоставления доступа к нескольким FC LUN выполните следующее:

1. На вкладке меню «**Протоколы**» → «**FC**» нажмите кнопку «**Групповое редактирование доступа**».
2. Откроется окно, представленное на Рисунок 104;
3. В панели **Том** отметьте нужные ресурсы, при этом в поле **LUN** будут перечислены номера лунов отмеченных ресурсов;
4. В поле **Клиент** выберите нужного клиента, которому будет открыт доступ к отмеченным выше лунам;

Также можно выбрать группу. При настройке необходимо убедиться, что один и тот же ресурс не отдан нескольким разным клиентам (если они не состоят в кластере);

5. Нажмите «**Сохранить**».
6. Доступ будет настроен сразу всем выбранным ресурсам.

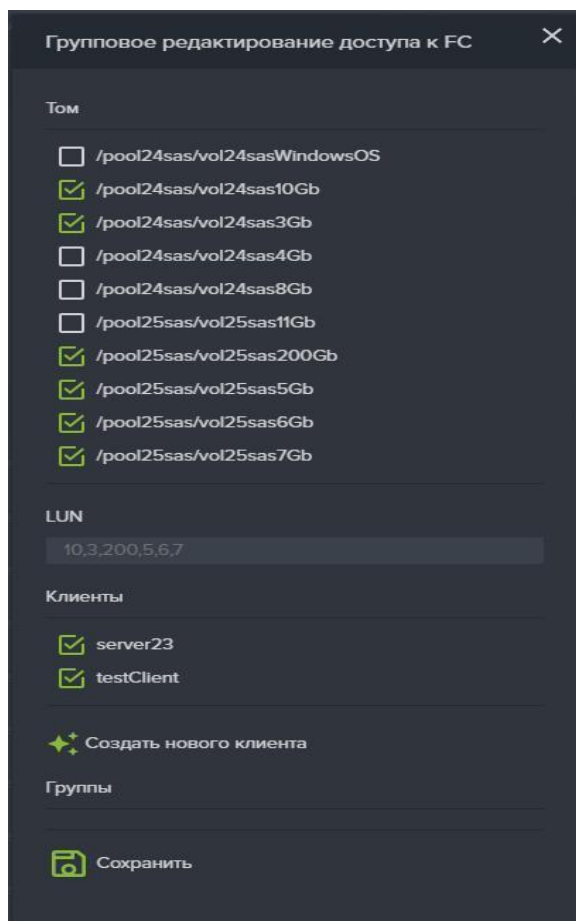


Рисунок 105. Групповое управление доступом

2.2.2.4.4.4. Отключение клиентов от FC LUN

Для того что бы запретить доступ клиентам к LUN выполните следующие действия:

1. Перейдите на вкладку меню «**Протоколы**» → «**FC**»;
2. Выберите нужный FC LUN из списка, развернув его панель свойств;
3. Нажмите на кнопку «**Редактировать доступ**», откроется одноименное окно (см. Рисунок 104);
4. Уберите флажки перед именем клиентов, которым будет отключен доступ;
5. Нажать на кнопку «**Сохранить**», откроется окно с предупреждением о возможной потере доступа к луну;
6. Подтвердить удаление введя слово «**ок**» и нажмите «**Удалить**».

2.2.2.4.4.5. Удаление FC LUN.

Для удаления LUN выполните следующие действия

1. Перейдите на вкладку меню «**Протоколы**» → «**FC**»;

2. Выберите нужный FC LUN из списка, развернув его панель свойств;
3. Нажмите на кнопку «Удалить», откроется окно с предупреждением о потере доступа к луну (см. Рисунок 106);
4. Подтвердите удаление введя слово «ок» и нажмите «Удалить».

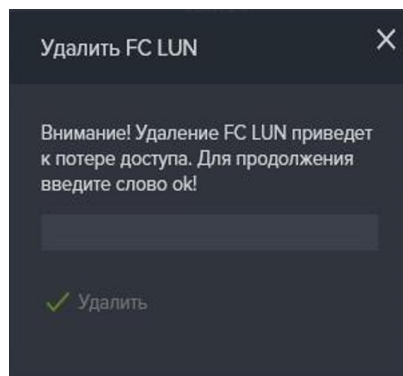


Рисунок 106. Окно подтверждения удаления LUN.

2.2.2.5. Работа с файловыми системами

Перед настройкой доступа пользователей к файловым системам (при создании общих сетевых папок) необходимо выполнить ряд настроек, которые позволят выполнить автоматическое переключение сетевых папок между контроллерами в случае миграции ресурсов внутри кластера. Без этих настроек невозможно включить службы файловых протоколов и создать сетевые папки.

Порядок действий для создания сетевых папок:

1. Создайте файловые системы на пуле (пулах) СХД;
2. Выберите сетевые интерфейсы, через которые будут подключены клиенты файловых ресурсов и назначьте им IP-адреса. Интерфейсы должны быть выбраны как на первом, так и на втором контроллере.
3. Привяжите назначенные IP-адреса к службам файловых протоколов, на первом и втором контроллере;
4. Добавьте в группу выбранные интерфейсы;
5. Для папки SMB, введите СХД в домен Active Directory и настройте доступ пользователям домена;
6. Для папки NFS, FTP, создайте клиента, чтобы открыть к ней доступ хостам;
7. Включите службу нужного файлового протокола, создайте общие папки для доступа пользователей;
8. Привяжите клиента к созданным папкам (этот пункт не относится к службе SMB).

В дальнейшем создавать сетевые папки на настроенном интерфейсе можно уже без дополнительной настройки.



Внимание! *Имя файловой системы не может начинаться с буквы «с» (cu).*

2.2.2.5.1. Создание файловой системы.

Внимание! Создание файловых систем возможно только на обычных пулах!

Для создания файловой системы выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Файловые системы»;
2. Нажмите на кнопку «Создать файловую систему», откроется одноименное окно (см. Рисунок 107);
3. Введите в соответствующих полях имя файловой системы и её размер;
4. Выберите из выпадающего меню пул, в котором будет создана файловая система и размер блока;
5. Отметьте флажок «Резервировать место» для создания «толстой» файловой системы;
6. Нажмите кнопку «Создать».

В списке файловых систем появится новая файловая система.

Рисунок 107. Окно создание файловой системы.

2.2.2.5.2. Изменение параметров файловой системы.

Некоторые параметры созданной файловой системы можно редактировать. Для этого откройте панель свойств выбранной файловой системы и нажмите кнопку **«Редактировать»**. В открывшемся окне имеется возможность изменить размер файловой системы (см. Рисунок 108). Для сохранения сделанных изменений нажмите на кнопку **«Сохранить»**.

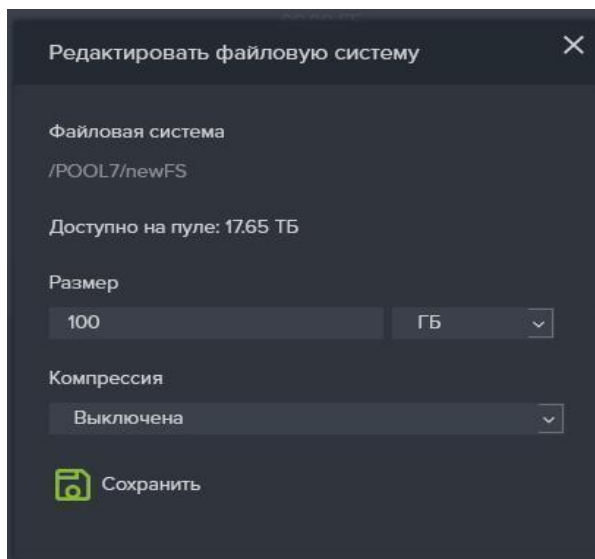


Рисунок 108. Окно изменения параметров файловой системы.

2.2.2.5.3. Удаление файловой системы.

Для удаления файловой системы выполните следующие действия:

1. Перейдите на вкладку меню **«Дисковое пространство»** → **«Файловые системы»**;
2. Выберите файловую систему из списка, развернув её панель свойств;
3. Нажмите на кнопку **«Удалить»** (см. Рисунок 109);
4. Подтвердите удаление, введя слово **«ок»** и нажмите кнопку **«Удалить»**.

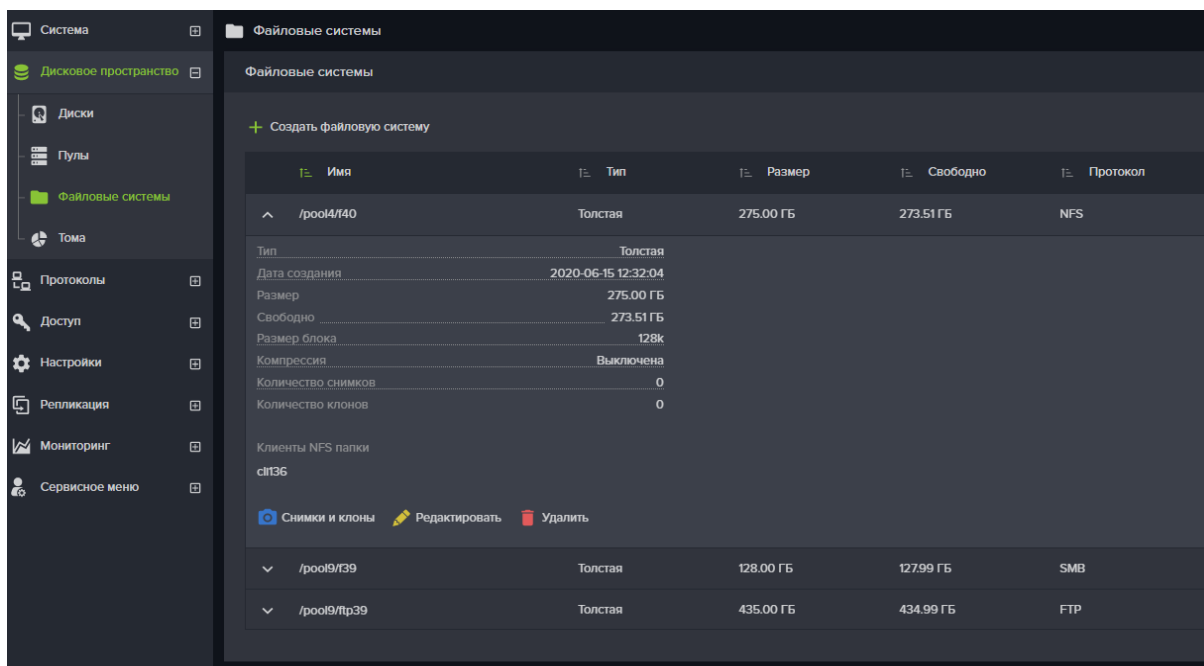


Рисунок 109. Панель свойств файловой системы

2.2.2.5.3.1. Включение и настройка службы LDAP

Для ограничения доступа к NFS ресурсу, выдавая разрешения пользователям используется интеграция с каталогом LDAP. Для этого необходимо наличие в сети сервера LDAP с анонимным доступом, а также включенная и настроенная служба LDAP на СХД.

Для настройки службы LDAP выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Службы и сервисы**» → «**Настройки AD/LDAP**»;
2. В области настроек LDAP введите URI вида «**ldap://hostname**» и Base вида «**ou=organisation_unit,dc=domain,dc=domain**», например: «**ldap://ldap.baum.local; ou=ldap_users,dc=baum,dc=local**»;
3. Нажмите на кнопку «**Сохранить**» (см. Рисунок 112);
4. После сохранения параметров нажмите кнопку «**Включить службу LDAP**».

После успешного включения службы появится возможность назначать права для ресурсов пользователям и группам LDAP.

2.2.2.5.3.2. Подключение клиентов по протоколу NFS.

Для предоставления клиентам доступа по протоколу NFS, выполните следующие действия:

1. Перейдите на вкладку меню «**Протоколы**» → «**NFS**»;

2. Выберите NFS папку, развернув её панель свойств;
3. Нажмите на кнопку «**Редактировать доступ**», откроется соответствующее окно (см. Рисунок 110);

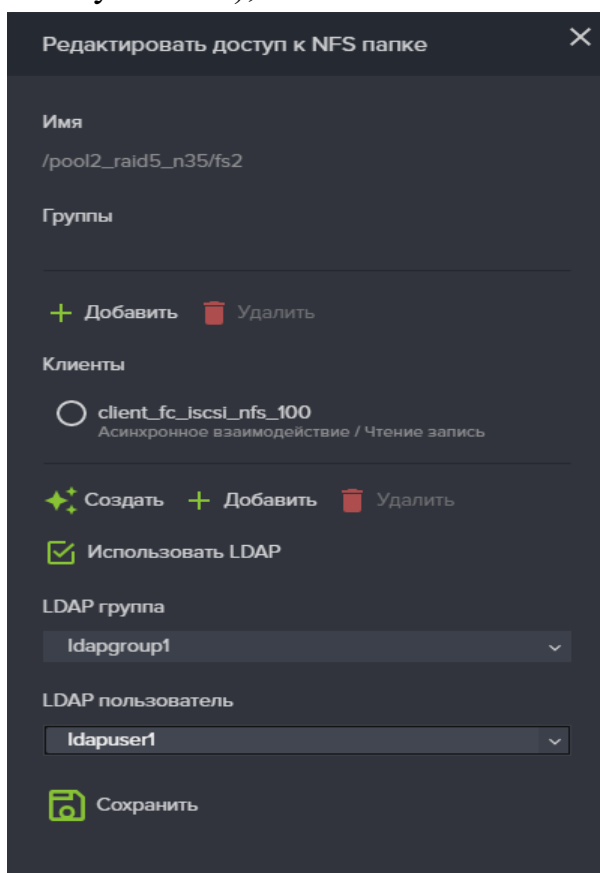


Рисунок 110. Окно редактирование доступа к NFS папке.

4. Нажмите на кнопку «**Добавить**» клиента или «**Добавить**» группу, см. Рисунок 87, (предполагается, что вы уже настроили клиента в меню "Доступ");

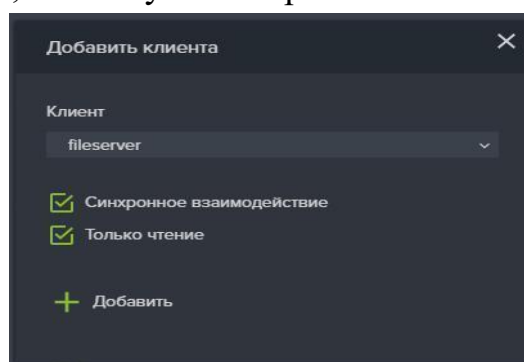


Рисунок 111. Добавить клиента к NFS папке.

5. Выберите клиента (или группу, если вы выбрали добавление группы) которому необходимо дать доступ к NFS папке, из выпадающего списка.
6. Снимите при необходимости флажки «**Синхронное взаимодействие**» и «**Только чтение**».

7. Нажмите на кнопку «**Добавить**», имя клиента или имя группы появится в списке «**Клиенты**» или «**Группы**».

8. Если вы хотите дать доступ пользователям или группам LDAP, отметьте флажок «**Использовать LDAP**» и укажите в списках нужную группу или пользователя.

9. Нажмите на кнопку «**Сохранить**», клиенты и группы будут добавлены к NFS папке и отобразятся в области свойств (см. Рисунок 112).

10. При необходимости изменения настроек NFS папки нажмите на кнопку «**Редактировать доступ**».

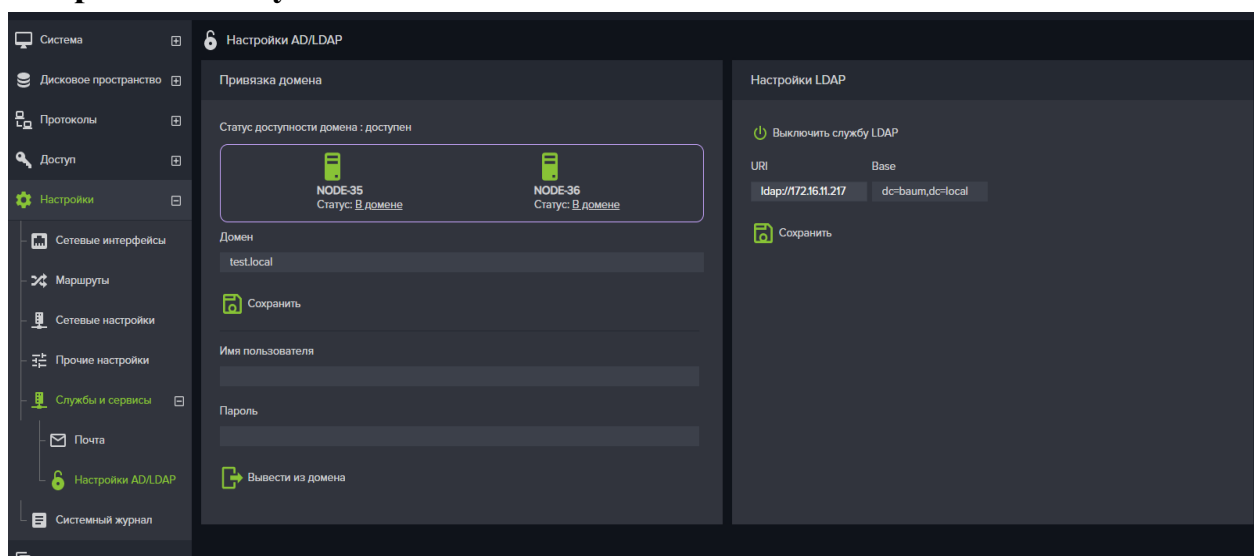


Рисунок 112. Вкладка меню настройки службы LDAP

У созданной NFS папки могут быть несколько статусов, которые отображаются в графе статус окна настройки протокола NFS.

NOT SHARED - не привязан клиент (при этом пользователь ldap может быть привязан);

NOT SET PERMISSION - клиент привязан, но не заданы пользователь или группа ldap;

ОК – привязан клиент и настроены права ldap.

2.2.2.5.3.3. Групповое подключение клиентов по протоколу NFS

Для предоставления доступа клиентам или группам сразу к нескольким NFS ресурсам используется групповое редактирование доступа. Для предоставления доступа выполните следующие действия:

1. Перейдите на вкладку меню «**Протоколы**» → «**NFS**»;
2. Нажмите кнопку «**Групповое редактирование доступа**»;
3. Откроется одноименное окно, изображенное на рисунке ниже (см.

Рисунок 113);

4. Отметьте флажками нужные файловые системы;
 5. Добавьте клиентов или при необходимости группы, которым будет предоставлен доступ к отмеченным NFS ресурсам;
 6. Если вы даете доступ группам и пользователям LDAP, отметьте флажок «Использовать LDAP» и выберите группу или пользователя;
 7. Нажмите кнопку «**Сохранить**».
- В результате ко всем выбранным файловым ресурсам будет назначен доступ.

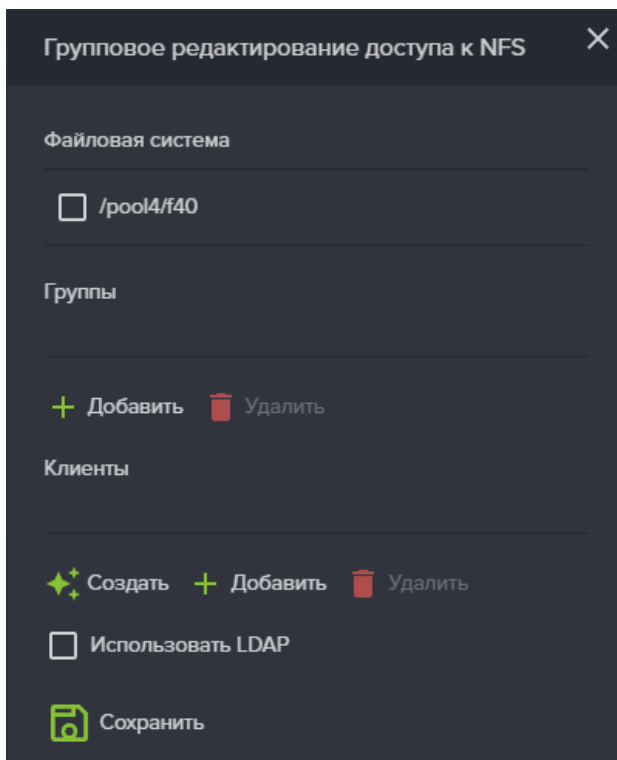


Рисунок 113. Групповое предоставление доступа

2.2.2.5.3.4. Отключение клиентов от NFS папки

Перед отключением клиента от NFS папки, необходимо выполнить отключение папки на стороне хоста, к которому подключена NFS папка. В противном случае существует вероятность зависания хоста.

Для того что бы запретить доступ клиента или группы к NFS папке выполните следующие действия:

1. Отключите NFS папку на стороне хоста к которому подключена папка;
2. Перейдите на вкладку меню «**Протоколы**» → «**NFS**»;
3. Выберите NFS папку, развернув её панель свойств;
4. Нажмите на кнопку «**Редактировать доступ**»;
5. В открывшемся окне выберите клиента или группу, которую

необходимо удалить;

6. Нажмите на кнопку **«Удалить»** клиента или **«Удалить»** группу для удаления из списка клиента или группы соответственно;

7. Нажать на кнопку **«Сохранить»**.

2.2.2.5.3.5. Удаление папки NFS

Для удаления папки NFS выполните следующие действия:

1. Перейдите на вкладку меню **«Протоколы»** → **«NFS»**;

2. Выберите NFS папку которую необходимо удалить, развернув её панель свойств;

3. Нажмите на кнопку **«Удалить»**, откроется окно с предупреждением о потере доступа к NFS папке;

4. Подтвердите удаление введя слово **«ok»** и нажмите **«Удалить»**.

В результате в области уведомления появится новая запись о успешно выполненной операции: (NFS папка /пул/файловая_система успешно удалена). NFS папка исчезнет из списка папок.

2.2.2.5.4. Работа с файловой системой по протоколу SMB.

2.2.2.5.4.1. Интеграция с службой AD.

Домен – основная административная единица в сетевой инфраструктуре, в которую входят все сетевые объекты.

Для того чтобы ввести систему в домен, выполните следующие действия:

1. Перейдите на вкладку меню **«Настройки»** → **«Сетевые настройки»**;

2. Укажите в поле **«IP-адрес DNS сервера»** адрес контроллера домена и нажмите на кнопку **«Сохранить»**;

3. Перейдите на вкладку меню **«Настройки»** → **«Сетевые интерфейсы»**;

4. Нажмите на кнопку **«Редактировать имя контроллера»**, откроется одноименное окно;

5. Задайте уникальные имена для **«Контроллер №1»** и **«Контроллер №2»**;

6. Нажмите кнопку **«Сохранить»**;

7. Выберите на первом контроллере Ethernet интерфейс, через который предполагается раздавать SMB ресурс, открыв его панель свойств. (Интерфейс **не должен** использоваться для управления системой);

8. Нажмите на кнопку **«Редактировать адреса»**;

9. В открывшемся окне, нажмите **«Добавить»**, откроется окно для ввода IP адреса;

10. Введите IP адрес, по которому будет осуществляется доступ к SMB папке;
 11. Нажмите на кнопку «Добавить»;
 12. Выберите созданный адрес в окне редактирования IP адреса;
 13. Нажмите на кнопку «Привязать службы», откроется одноименное окно;
 14. В открывшемся окне поставьте флажок smb;
 15. Нажмите на кнопку «Привязать» (Привязку адресов также можно выполнить из окна настройки протокола SMB, нажав на кнопку «Привязать адреса»);
 16. Повторить те же операции по созданию адреса для свободного интерфейса на втором контроллере;
 17. Перейдите на вкладку меню «Настройки» → «Службы и сервисы» → «AD/LDAP»;
 18. В области «Привязка домена», введите имя домена и имя и пароль пользователя домена, обладающего правом ввода компьютеров в домен;
 19. Нажмите на кнопку «Ввести в домен»;
- При успешном вводе в домен, в области уведомления появится сообщение о вводе в домен и статус «*Не в домене*» сменится на «*В домене*».

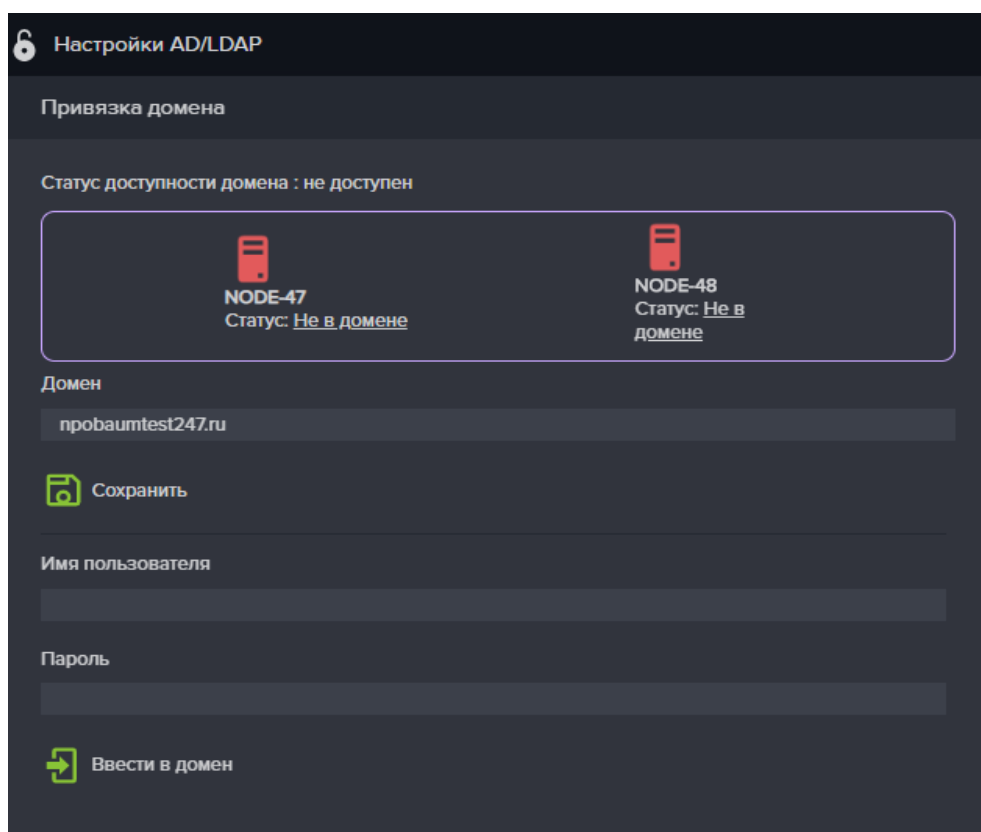


Рисунок 114. Меню настройки интеграции с Active Directory



Внимание! Если контроллеры не вводятся в домен убедитесь, что время на обоих контроллерах СХД идентично с контроллером AD, в качестве DNS сервера стоит тот же сервер, который указан в AD, и имеется доступ к контроллеру домена с обоих контроллеров СХД.

2.2.2.5.4.2. Создание и настройка SMB ресурса.

Доступ к SMB папке предоставляется только членам домена Windows. Перед созданием SMB папок введите СХД в домен. Разграничение доступа к папке SMB осуществляется контроллером домена Active Directory.

Перед созданием папок убедитесь в том, что к службе SMB привязаны IP-адреса на обоих контроллерах, а сама служба включена.

Для создания SMB папки выполните следующие действия:

1. Перейдите во вкладку «**Протоколы**» → «**SMB**»;
2. Нажмите на кнопку «**Создать новую SMB папку**», откроется одноименное окно;
3. Выберите из выпадающего списка созданную на предыдущем этапе файловую систему;
4. Нажмите кнопку «**Создать**»;
5. Выберите SMB папку в списке, открыв её панель свойств;

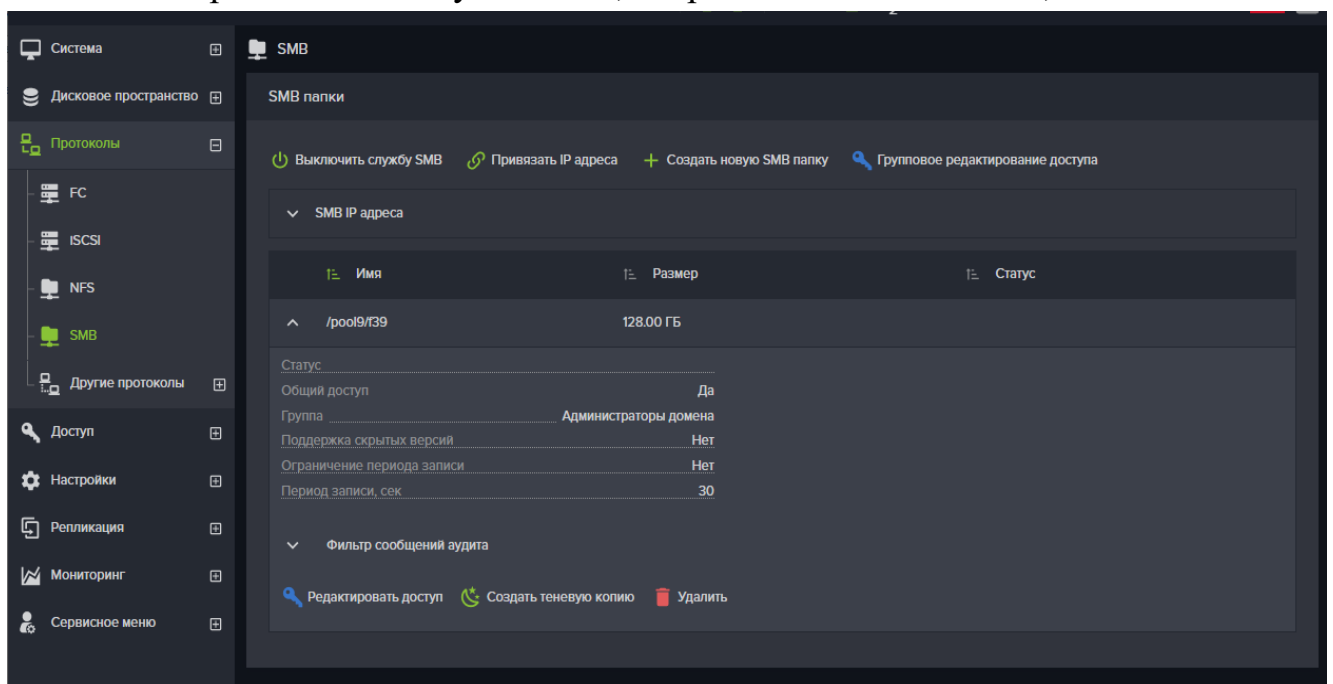


Рисунок 115. Меню настройки службы протокола SMB

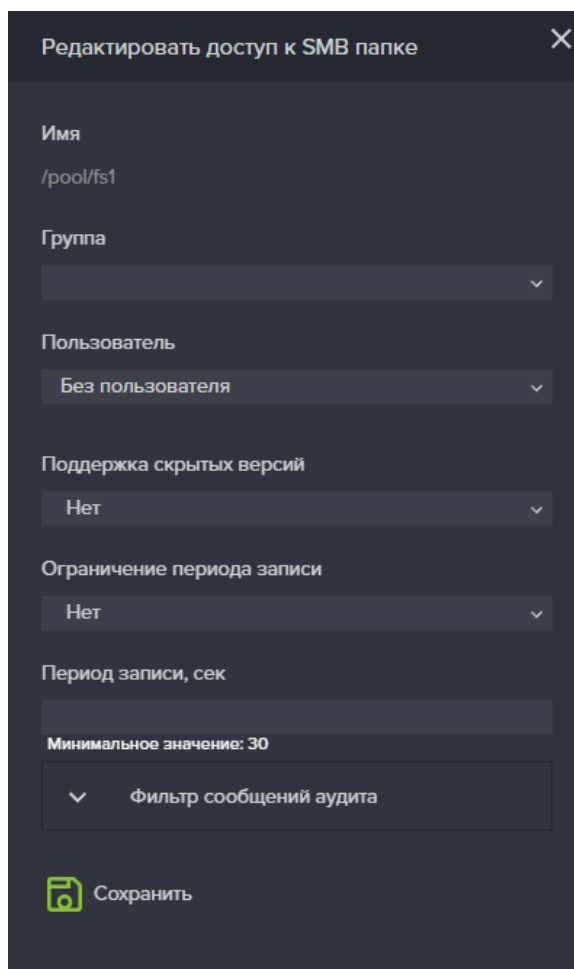


Рисунок 116. Окно редактирования доступа к SMB папке

6. Нажмите на кнопку **«Редактировать доступ»**, откроется одноименное окно (см. Рисунок 116);
7. Выберите в поле **«Группа»** имя группы созданной в AD домена;
8. Выберите доменного пользователя из списка;
9. Нажмите на кнопку **«Сохранить»**.

При большом количестве групп и пользователей в домене, при открытии списка выбора отобразятся только первые 2000 объектов. В этом случае, для выбора группы или пользователя (даже если они не видны в списке), введите объекта имя в строку поиска.

В строке поиска введите либо полное имя нужного объекта, либо начальные символы его имени и символ **«*»**. Используйте только строчные буквы в поисковой строке.

Найденные объекты отобразятся в списке, из которого можно выбрать нужный, см. Рисунок 117.

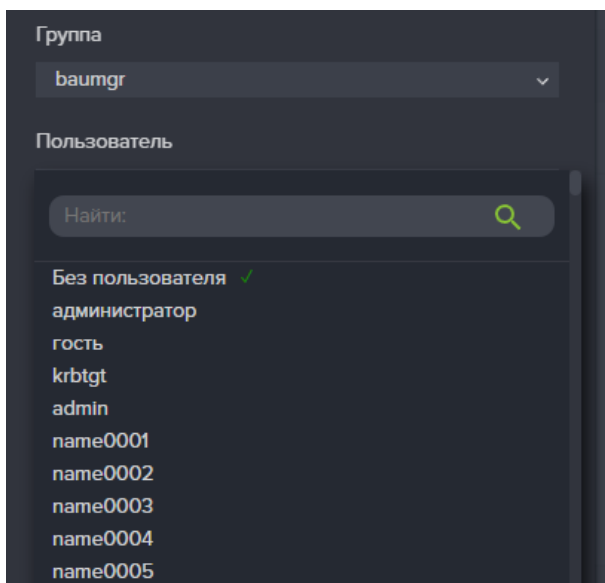


Рисунок 117. Список пользователей домена

2.2.2.5.4.3. Защита изменения записанных файлов (WORM)

Для блокировки изменения записанных файлов и папок имеется возможность включить защиту записанных файлов WORM (Write Only Read Many). После записи в защищенную папку файлов, они будут доступны для изменения или удаления в течении заданного времени (минимальное время – 30 секунд), после чего будут заблокированы для изменений. Доступ к таким файлам будет возможен только на чтение.

Если на момент включения блокировки изменений в папке уже были файлы, блокировка распространится также и на них.

Если была включена блокировка и записаны файлы, то после отключения блокировки все файлы в папке вновь могут быть изменены либо удалены.

Если после включения блокировки в поле «Период записи, сек» будет введено значение времени большее установленного ранее, то после сохранения значений, файлы в папке будут вновь доступны для изменений на период времени равный разнице нового и старого значений.

Для включения блокировки от изменений записанных файлов выполните следующее:

1. Войдите в меню «**Протоколы**» → «**SMB**»;
2. Разверните панель свойств созданной SMB папки;
3. Нажмите на кнопку «**Редактировать доступ**», Рисунок 118;
4. В открывшемся меню выберите в раскрывающемся списке «**Ограничение периода записи**» - «**Да**»;

5. В поле «**Период записи, сек.**» введите время, после которого будет включаться блокировка файлов.

6. Нажмите кнопку «**Сохранить**».

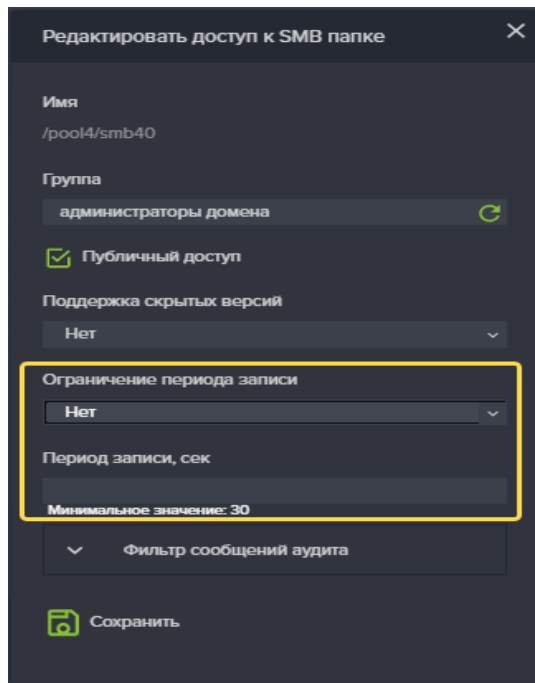


Рисунок 118. Установка блокировки изменений записанных файлов

2.2.2.5.4.4. Групповое предоставление доступа к файловым системам

Для одновременной настройки доступа к нескольким SMB папкам выполните следующее:

1. Перейдите во вкладку меню «**Протоколы**» → «**SMB**»;
2. Создайте нужное количество SMB папок, к которым нужно настроить одинаковые права доступа;
3. Нажмите кнопку «**Групповое редактирование доступа**» см. Рисунок 119;

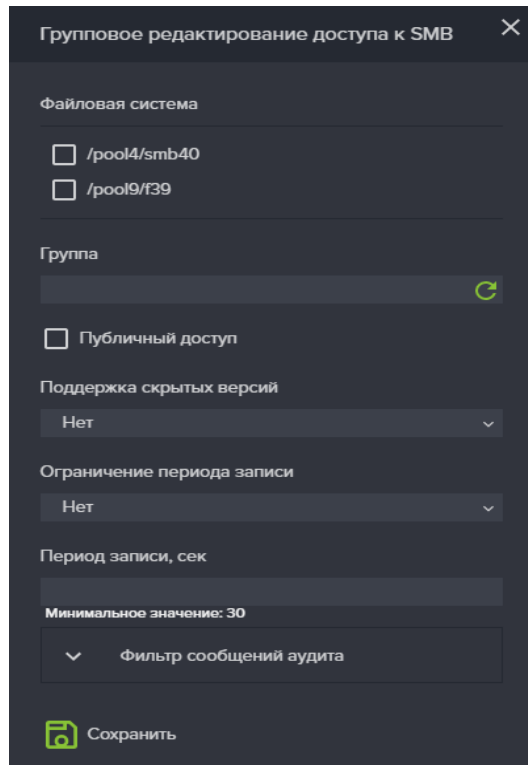


Рисунок 119. Окно группового предоставления доступа к SMB папкам

4. Выберите нужные файловые системы (SMB папки);
5. Введите в поле «**Группа**» имя группы, созданной в АД домена;
6. Установите флажки нужных прав доступа;
7. При необходимости настройте параметры поддержки скрытых версий и ограничение периода записи;
8. Нажмите «**Создать**».

2.2.2.5.4.5. Создание теневых копий

В ПО Платформе СХД имеется возможность создания копий SMB папки, которые отображаются клиентом Windows как теневые копии (скрытые копии).

Для того что бы создать теневую копию необходимо выполнить следующие действия.

1. Перейдите во вкладку меню «**Протоколы**» → «**SMB**»;
2. Выберите в списке нужную папку SMB, развернув панель свойств;
3. Нажмите на кнопку «**Редактировать доступ**»;
4. Выберите значение «**Да**» в выпадающем списке «**Поддержка скрытых версий**»;
5. Нажмите на кнопку «**Сохранить**»;
6. Нажмите на кнопку «**Создать теневую копию**» в панели свойств папки

SMB.

Существует также альтернативный способ создания теневой копии:

1. Перейдите на вкладку меню «Дисковое пространство» → «Файловые системы»;
2. Выберите нужную файловую систему из списка, открыв её панель свойств;

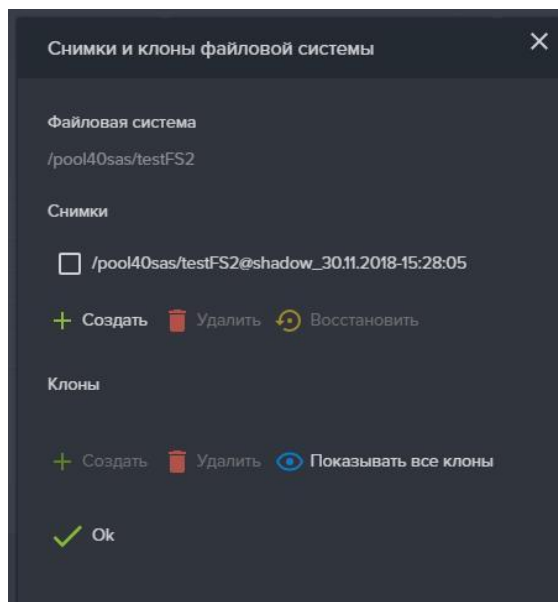


Рисунок 120. Редактирование SMB папки

3. Нажмите на кнопку «Снимки и клоны», откроется окно «Снимки и клоны файловой системы» (см. Рисунок 120);

4. В области снимки нажмите на кнопку «Создать», откроется окно «Создать снимок»;

5. Введите имя снимка, по шаблону: shadow_06.07.2017-14:45:00 (dd.mm.yyyy-hh:mm:ss);

6. Нажмите кнопку «Создать».

Для проверки созданной теневой копии, на компьютере под управлением ОС Windows выполните следующие:

1. Выберите в списке дисков подключенный сетевой диск и откройте его свойства;

2. Перейдите на вкладку «Предыдущие версии».

Созданные снимки будут присутствовать в списке «Версии папки».

2.2.2.5.5. Работа с файловой системой по другим протоколам (FTP, AFP)

2.2.2.5.5.1. Создание FTP ресурса

Протокол FTP не требует привязки к интерфейсу, поскольку работает через управляющий интерфейс.

Для создания FTP ресурса выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Файловые системы»;
2. Нажмите на кнопку «Создать файловую систему», откроется одноименное окно;
3. Введите в соответствующих полях имя файловой системы и её объем;
4. Выберите из выпадающего списка пул, в котором будет создана файловая система и размер блока;
5. Отметьте, при необходимости, флажок «Резервировать место» для создания «толстой» файловой системы;
6. Нажмите кнопку «Создать»;
7. Перейдите на вкладку меню «Протоколы» → «Другие протоколы» → «FTP»;
8. Нажмите на кнопку «Включить службу FTP», если служба выключена;
9. Нажмите на кнопку «Создать новую FTP папку»;
10. Выберите из списка свободную файловую систему, которую собираемся выдавать по FTP;
11. Нажмите на кнопку «Создать»;
12. В списке появится новая FTP папка.

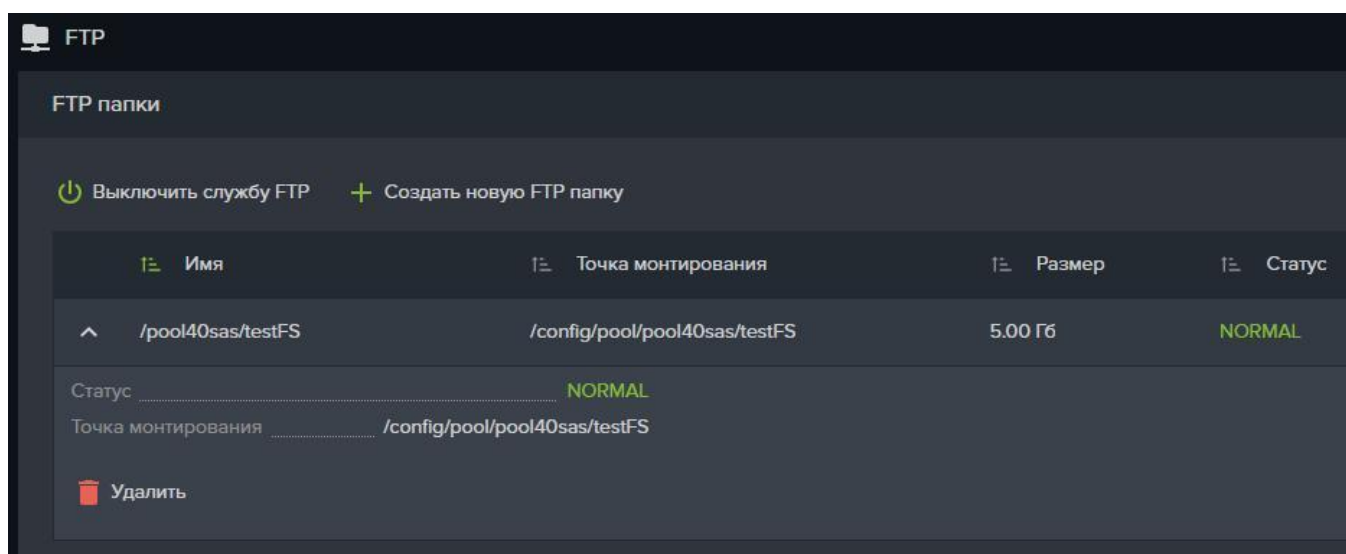


Рисунок 121. Окно настройки FTP ресурса

2.2.2.5.5.2. Удаление FTP ресурса

Для удаления FTP ресурса выполните следующие действия:

1. Перейдите на вкладку меню «**Протоколы**» → «**Другие протоколы**» → «**FTP**»;
2. Выберите в списке нужный FTP ресурс, развернув его панель свойств;
3. Нажмите на кнопку «**Удалить**»;
4. Подтвердите удаление набрав «**ok**» в открывшемся окне и нажмите на кнопку «**Удалить**»;

После успешного выполнения ресурс будет удален, все клиенты, использующие данный ресурс будут отключены.

2.2.2.5.5.3. Создание AFP ресурса

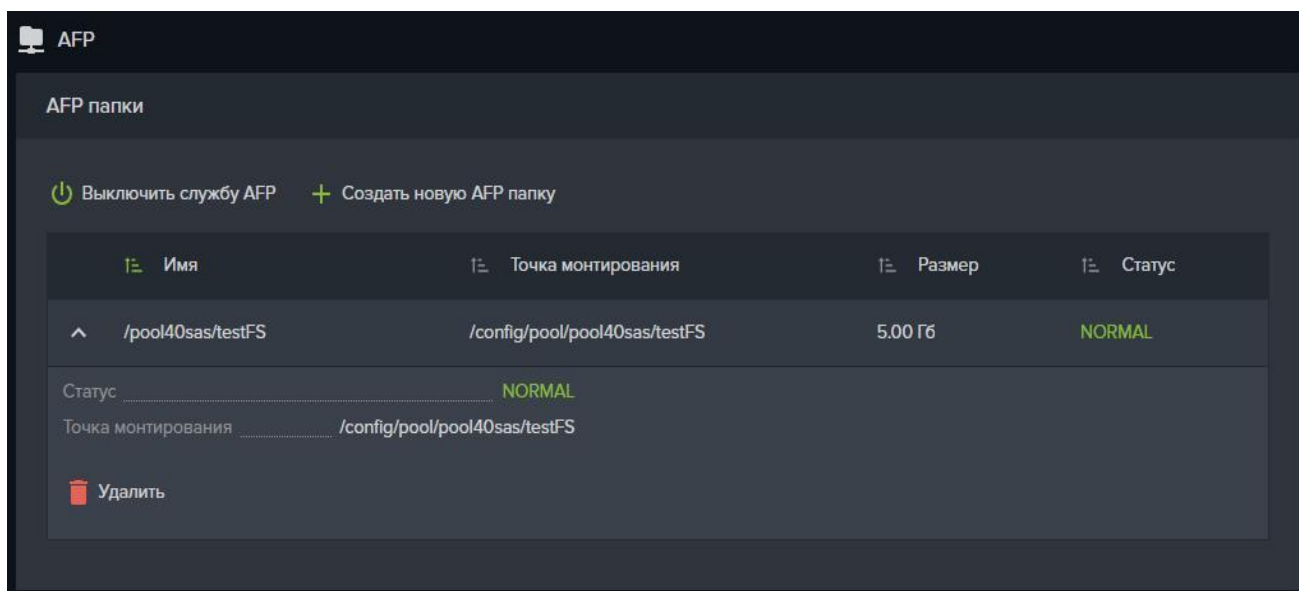


Рисунок 122. Окно создания AFP ресурса

1. Перейдите на вкладку меню «**Дисковое пространство**» → «**Файловые системы**»;
2. Нажмите на кнопку «**Создать файловую систему**», откроется одноименное окно;
3. Введите в соответствующих полях имя файловой системы и её объем;
4. Выберите из выпадающего списка пул, в котором будет создана файловая система и размер блока;
5. Отметьте, при необходимости, флажок «**Резервировать место**» для создания «толстой» файловой системы;
6. Нажмите кнопку «**Создать**»;

7. Перейти во вкладку меню **«Протоколы»** → **«Другие протоколы»** → **«AFP»**;
8. Нажмите на кнопку **«Включить службу AFP»**, см. Рисунок 122;
9. Нажмите на кнопку **«Создать новую AFP папку»**;
10. Выберите из списка свободную файловую систему, которую собираемся выдавать по AFP;
11. Нажмите на кнопку **«Создать»**.
12. В списке появится новая AFP папка

2.2.2.5.5.4. Удаление AFP ресурса

Для удаления AFP ресурса выполните следующие действия:

1. Перейдите на вкладку **«Протоколы»** → **«Другие протоколы»** → **«AFP»**
2. Выберите в списке нужный AFP ресурс, развернув его панель свойств.
3. Нажмите на кнопку **«Удалить»**.
4. Подтвердить удаление набрав **«ok»** в открывшемся окне и нажмите на кнопку **«Удалить»**.

После успешного выполнения ресурс будет удален, все клиенты, использующие данный ресурс будут отключены.

2.2.2.6. Управление доступом к ресурсам

Для разрешения доступа к ресурсам СХД определенным хостам, необходимо прописать их адреса в списке доступа. В ПО BAUM STORAGE AI v2 для этой цели используются так называемые клиенты.

2.2.2.6.1. Создание клиента.

Для удобной работы с хостами клиенты создаются сразу и для блочного и для файлового доступа.

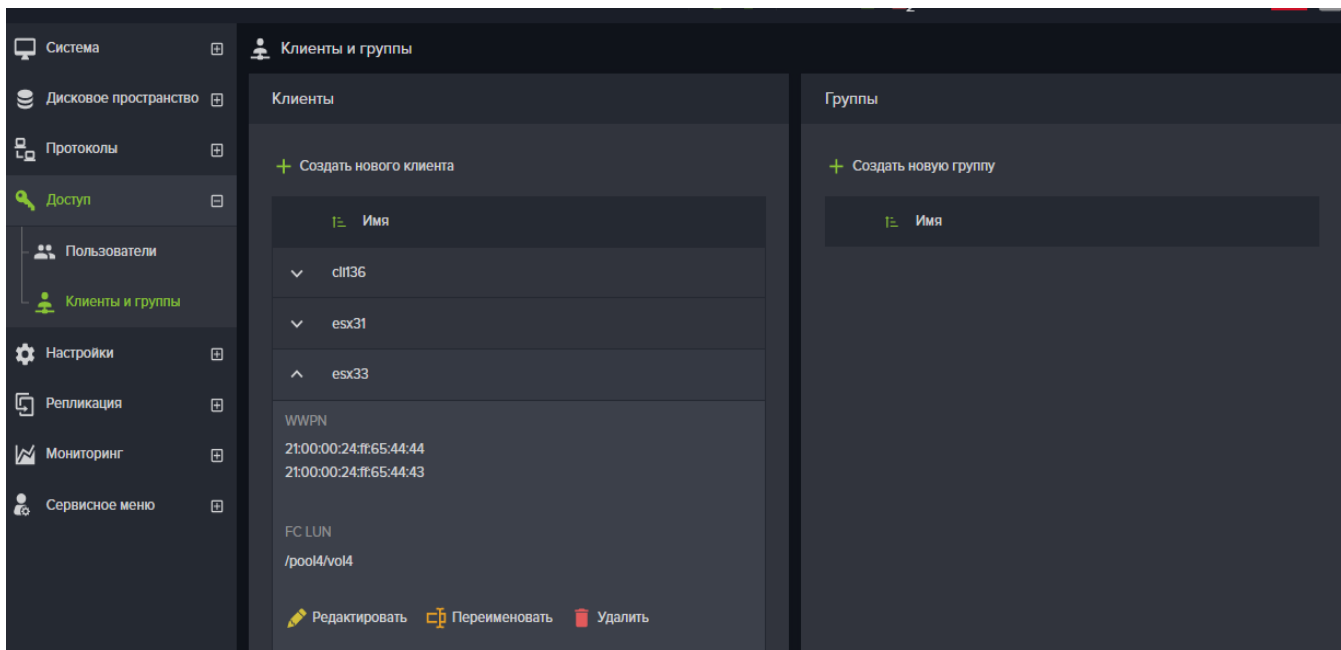


Рисунок 123. Окно вкладки меню «Клиенты и группы».

У созданного клиента в последствии можно изменить имя и списки хостов. Для этого развернув панель свойств выбранного клиента нажмите кнопку «Переименовать» или «Редактировать» (см. Рисунок 123).

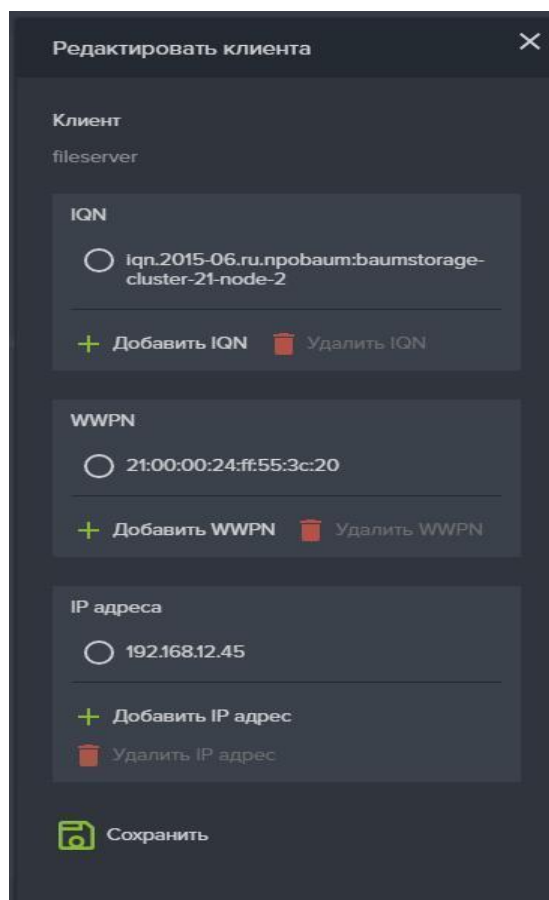


Рисунок 124. Создание клиента

Для создания нового клиента выполните следующие действия.

1. Перейдите на вкладку меню «Доступ» → «Клиенты и группы»;
2. Нажмите на кнопку «Создать нового клиента», откроется окно создания нового клиента (см. Рисунок 124);
3. Введите имя создаваемого клиента в соответствующее поле; Далее введите адреса хостов, для которых вы настраиваете доступ;
4. Нажмите на кнопку «Добавить IQN» и введите IQN в открывшемся окне;
5. Нажмите на кнопку «Добавить». И в списке появится введенный адрес;

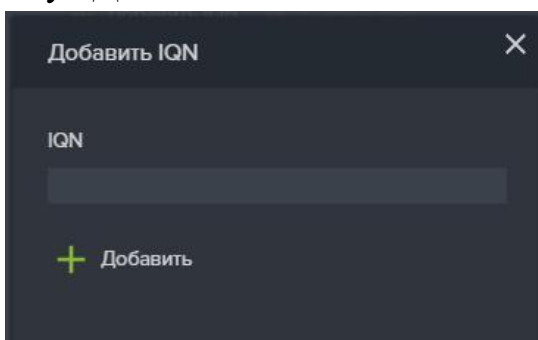


Рисунок 125. Окно добавления IQN

6. Нажмите на кнопку «Добавить WWPN», откроется одноименное окно;
7. Поставьте флажок «Выбрать из клиентов, доступных на фабрике»;
8. Выберите из выпадающего списка нужный wwpn клиента (см. Рисунок 126);
9. Нажмите на кнопку «Добавить».
10. Таким способом добавьте все нужные wwpn, они добавятся к списку в окне клиента;

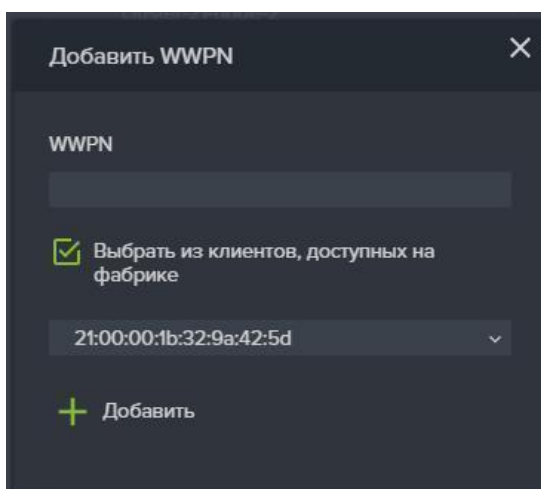


Рисунок 126. Окно добавления WWPN

11. Нажмите на кнопку «**Добавить IP адрес**»;
12. Введите в открывшемся окне IP адрес. (см. Рисунок 127);
13. Нажмите на кнопку «**Добавить**», адрес добавится в список;
14. В окне настройки клиента нажмите на кнопку «**Сохранить**».

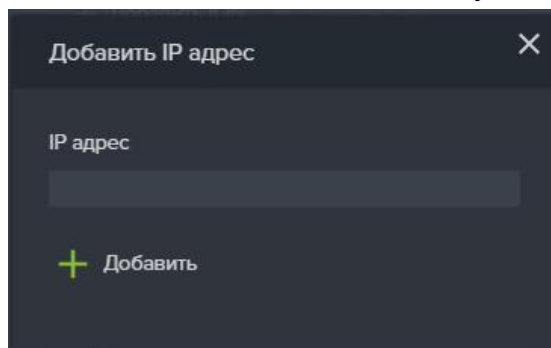


Рисунок 127. Окно добавления IP адреса

В результате выполненных действий в списке «Клиенты» появится имя созданного клиента, а в области уведомления появится сообщение «Клиент имя_клиента успешно создан».

2.2.2.6.2. Удаление клиента

Для удаления клиента, выполните следующие действия:

1. Перейдите на вкладку меню «**Доступ**» → «**Клиенты и группы**»;
2. Выберите нужного клиента из списка «**Клиенты**», развернув его панель свойств, как показано на рисунке ниже;
3. Нажмите на кнопку «**Удалить**»;
4. Подтвердите удаление введя слово «**ok**», в открывшемся окне и нажмите кнопку «**Удалить**»;

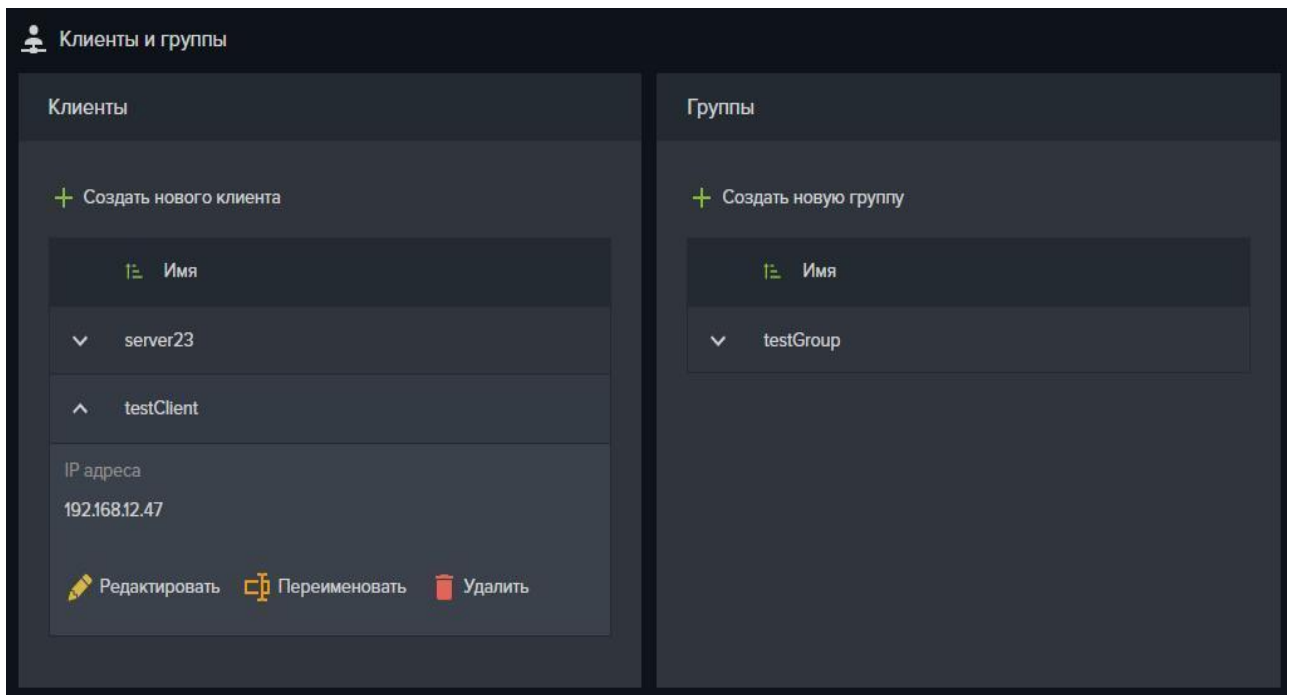


Рисунок 128. Окно настроек клиента

В результате выполненных действий имя клиента исчезнет из списка «Клиенты» и в области уведомления появится сообщение «клиент имя_клиента успешно удален».

2.2.2.6.3. Создание клиента для файлового доступа (NFS).

Для создания нового клиента для файлового доступа выполните следующие действия:

1. Перейдите на вкладку меню «**Протоколы**» → «**NFS**»
2. Выберите нужную NFS папку, развернув его панель свойств.
3. Нажмите на кнопку «**Редактировать доступ**».
4. Нажмите на кнопку «**Создать**» в области «**Клиенты**», откроется окно «**Создать клиента**».
5. Введите имя клиента.
6. В области «**IP адреса**» нажмите кнопку «**Добавить**»
7. Введите IP-адрес клиента и нажмите кнопку «**Добавить**».
8. Нажмите кнопку «**Создать**».
9. В области «**Клиенты**» нажать на кнопку «**Добавить**», откроется окно «**Добавить клиента**» (см. Рисунок 129).
10. При необходимости отметьте флажками пункты «**Синхронное взаимодействие**» и «**Только чтение**».
11. Нажмите кнопку «**Добавить**».

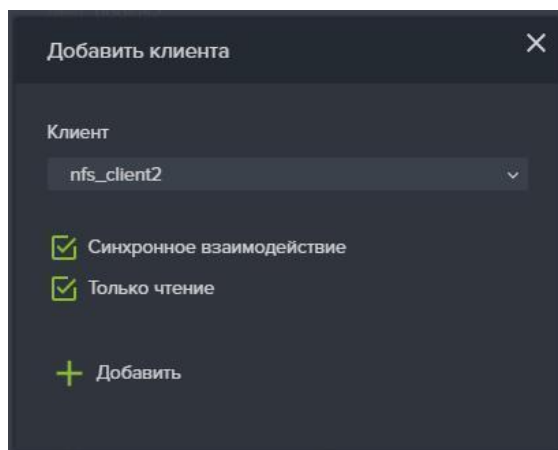


Рисунок 129. Добавление NFS клиента

2.2.2.6.4. Создание клиента для блочного доступа (FC)

Для создания нового клиента для блочного доступа выполнить следующие действия.

1. Перейдите на вкладку меню **«Протоколы»** → **«FC»**
2. Выберите нужный FC LUN из списка, развернув его панель свойств.
3. Нажмите на кнопку **«Редактировать доступ»**, откроется окно:

Редактировать доступ к FC LUN. (см. Рисунок 130)

4. Нажмите на кнопку **«Создать нового клиента»**, откроется окно:

Создание клиента (см. Рисунок 131).

5. Впишите в поле **«Имя»** имя создаваемого клиента.
6. В области WWPN нажмите на кнопку **«Добавить WWPN»**, откроется одноименное окно.

7. Введите WWPN вручную в поле WWPN, или выберите из выпадающего списка нужные WWPN для чего отметьте флажок **«Выбрать из клиентов, доступных на фабрике»** (см. Рисунок 132)

8. Нажмите на кнопку **«Добавить»**.
9. Нажать на кнопку **«Создать»**.

В результате имя клиента появится в списке **«Клиенты»** в меню **«Редактировать доступ к FC луноу»**

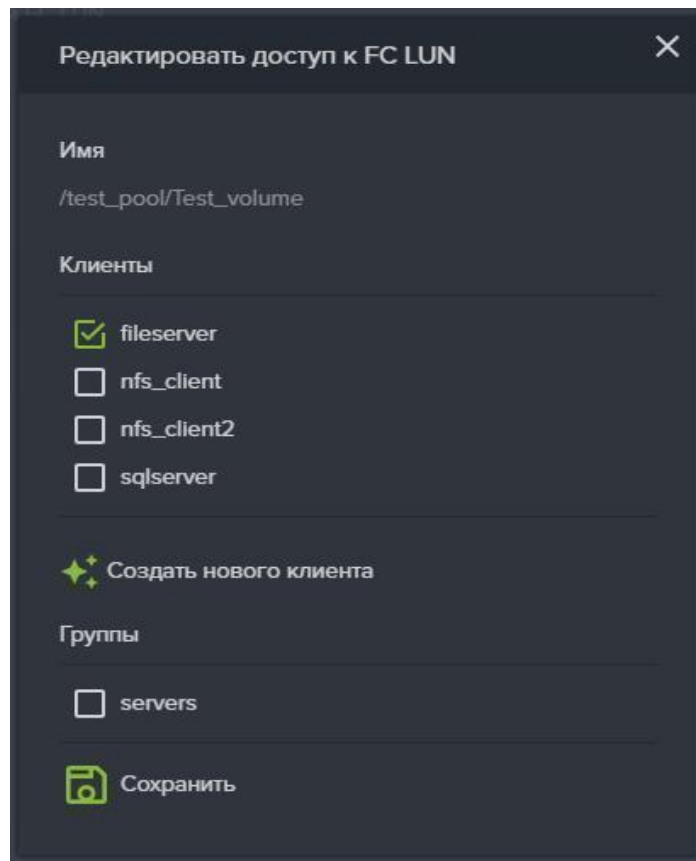


Рисунок 130. Окно «Редактировать доступ к FC луну».

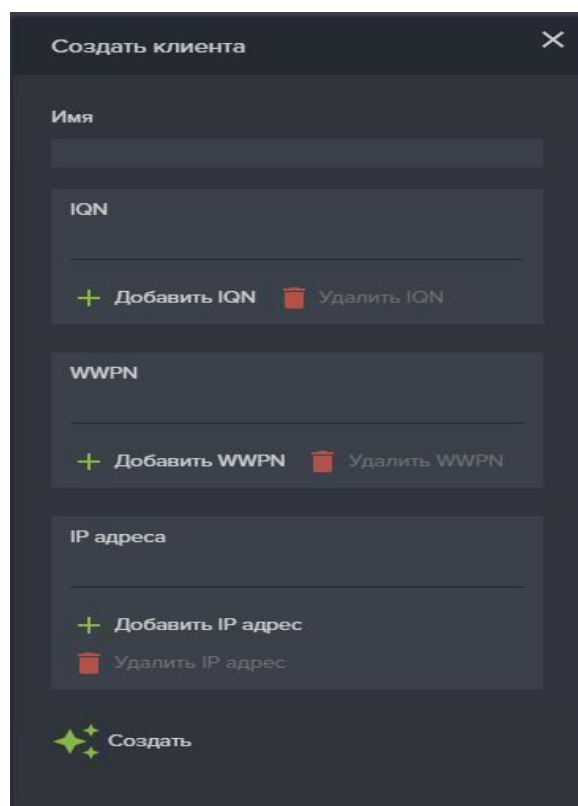


Рисунок 131. Создать нового FC клиента

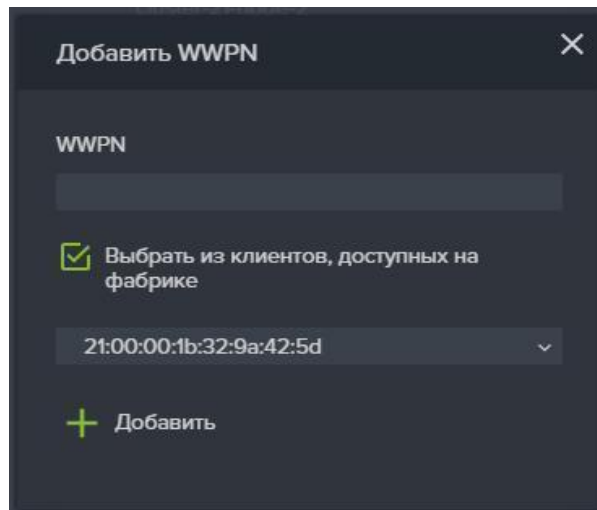


Рисунок 132. Добавление WWN клиенту

2.2.2.6.5. Настройка групп клиентов.

2.2.2.6.5.1. Создание группы клиентов

Для создания группы клиентов выполните следующие действия:

1. Перейдите на вкладку меню «Доступ» → «Клиенты и группы»;
2. Нажмите на кнопку «Создать новую группу», откроется окно создания группы (см. Рисунок 133);
3. Введите имя группы в поле «Имя»;
4. Нажать на кнопку «Создать».

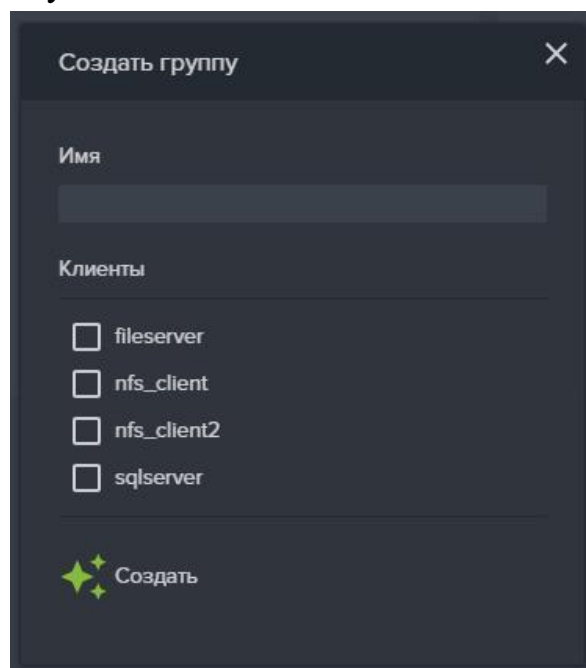


Рисунок 133. Окно создания группы клиентов

В результате в списке «Имя группы» появится имя новой группы и в области уведомлений появится сообщение «Группа **имя_группы** успешно создана».

2.2.2.6.5.2. Добавление клиентов в группу

Для добавления клиентов в группу выполните следующие действия:

1. Перейдите на вкладку меню «Доступ» → «Клиенты и группы».
2. Выберите нужную группу в списке «Группы», развернув её панель свойств.
3. Нажмите на кнопку «Редактировать» (см. Рисунок 134)
4. Выберите клиентов, которых необходимо добавить в группу
5. Нажать на кнопку «Сохранить»

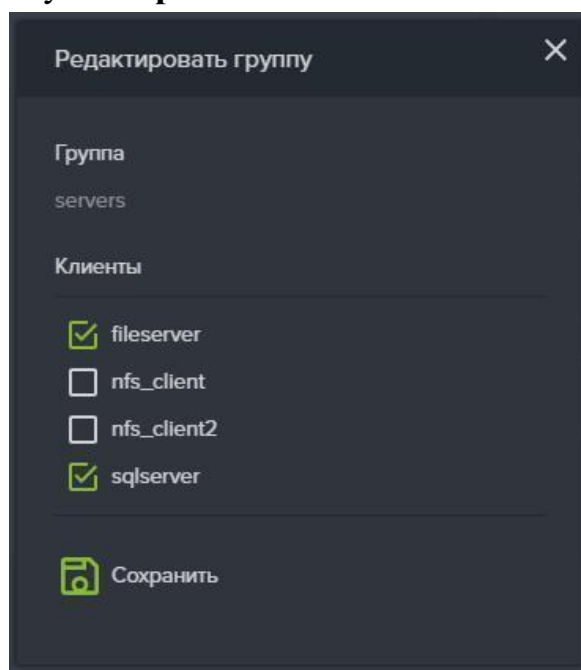


Рисунок 134. Добавление клиентов в группу

В результате действий в списке «Клиенты» появится список клиентов, принадлежащих данной группе и в области уведомлений появится сообщение «Группа **имя_группы** успешно отредактирована».

2.2.2.6.5.3. Удаление группы

Для удаления группы выполните следующие действия:

1. Перейдите на вкладку меню «Доступ» → «Клиенты и группы».
2. Выберите нужную группу в списке «Группы», развернув её панель свойств.
3. Нажмите на кнопку «Удалить».

4. Подтвердите удаление введя слово **«ok»**, в открывшемся окне и нажмите кнопку **«Удалить»**.

В результате выполненных действий в списке **«Имя группы»** исчезнет группа и в области уведомлений появится сообщение **«Группа имя_группы успешно удалена»**.

2.2.2.7. Управление учетными записями пользователей

2.2.2.7.1. Создание нового пользователя

Для администрирования системы хранения данных (СХД) используются учетные записи пользователей, создаваемые и хранимые локально в СХД. По умолчанию, в системе создана только одна учетная запись администратора с логином Admin, которая обладает всеми полномочиями на конфигурирование СХД. При создании новых пользователей, имеется возможность ограничить их права, присвоив им определенные роли в системе. Учетную запись Admin удалить нельзя. В случае утери пароля пользователя **«admin»**, обратитесь в службу технической поддержки для сброса пароля.

Для создания нового пользователя для администрирования СХД требуется выполнить следующие действия.

1. Перейдите на вкладку меню **«Доступ»** → **«Пользователи»**.
2. Нажать на кнопку **«Создать нового пользователя»**.
3. Введите имя нового пользователя.
4. Выберите Роль пользователя из трех возможных: **«Пользователь (только чтение)»**, **«Пользователь»** и **«Администратор»** (особенности ролей смотрите далее раздел Роли пользователей)
5. Введите срок действия пароля;
6. Введите пароль и подтверждение пароля.
7. Нажмите кнопку **«Создать»**.

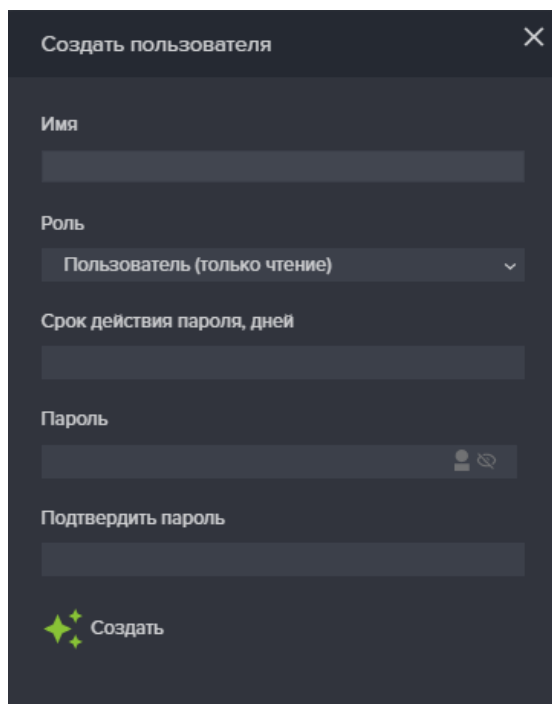

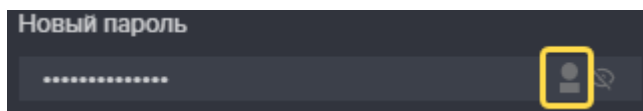


Рисунок 135. Создание нового пользователя

Для просмотра подсказки по требованиям к сложности пароля, нажмите на значок  в правой стороне строки ввода пароля:



Требования к сложности пароля:

- минимальная длина пароля должна составлять восемь символов;
- пароль должен содержать символы обоих регистров;
- пароль должен содержать хотя бы одну цифру;
- пароль должен содержать хотя бы один спецсимвол [~#\$*!@&()]

2.2.2.7.2. Роли пользователей

В системе хранения данных (СХД) предусмотрены три роли пользователей:

1. **«Пользователь (только чтение)»** - роль создана для операторов, следящих за состоянием системы, но с недостаточными компетенциями для ее настройки или переконфигурирования. Пользователь сможет получить доступ к просмотру сведения всех разделов, а также сможет выгрузить лог-файлы для их дальнейшего анализа.

2. **«Пользователь»** - роль создана для администраторов, работающих с СХД на уровне ресурсов и протоколов. Пользователь сможет получить доступ к управлению созданию и удалению ресурсов, перераспределению прав доступа на

ресурсы, создавать и удалять пулы/тома/файловые системы.

3. **«Администратор»** - роль создана для администраторов, работающих с СХД на уровне архитектора. Пользователь сможет получить доступ к глобальным настройкам, влияющим на функционирование всей системы, таким как миграция, включение/отключение служб протоколов, настройки сетевых интерфейсов и маршрутов, ввод и вывод из домена, управление репликациями, обновление микрокода.

2.2.2.7.3. Редактирование пользователя

Для редактирования параметров доступа пользователей выполните следующие действия:

1. Перейдите на вкладку меню **«Доступ»** → **«Пользователи»**;
2. Найдите в списке нужного пользователя и разверните его панель свойств. Редактированию подлежат: роль пользователя и его пароль;
3. Для редактирования пароля нажмите на кнопку **«Редактировать пароль»**, введите старый пароль, срок действия пароля и новый пароль два раза;
4. Для редактирования роли нажмите на кнопку **«Редактировать роль»**, выберите требуемую роль из списка и нажмите на кнопку **«Сохранить»**.



Внимание! *Изначально созданный Admin может управлять паролями всех пользователей и администраторов, в последующем создаваемые администраторы – только своим паролем и пароль пользователей, пользователи – только своим паролем.*

2.2.2.7.4. Удаление пользователя

Для удаления пользователей выполните следующие действия:

1. Перейдите на вкладку меню **«Доступ»** → **«Пользователи»**;
2. Найдите в списке нужного пользователя и разверните его панель свойств;
3. Нажмите на кнопку **«Удалить»**;
4. Подтвердите удаление введя слово **«ok»** в открывшемся окне и нажмите кнопку **«Удалить»**.

После успешного выполнения операции пользователь будет удален.

2.2.2.8. Работа с мгновенными снимками и клонами

2.2.2.8.1. Создание мгновенного снимка тома или файловой системы.

Мгновенный снимок (snapshot) представляет собой копию файловой системы или тома, доступную только для чтения. Снимки размещаются на том же пуле, что и тома /файловые системы, с которых они сделаны. Снимок, созданный на тонком томе, в момент своего создания практически не занимает места, но по мере того, как данные на томе будут изменяться, он будет увеличиваться в объёме. Снимок, созданный на толстом томе, в момент своего создания резервирует для себя место равное записанному на том объёму данных. Это правило действует только для первого снимка толстого тома. Все последующие снимки в момент создания не занимают место. Теоретически снимок со временем может вырасти до размера тома. Чем быстрее будет изменяться информация на томе, тем быстрее будет расти снимок. При использовании снимков необходимо оставить для них некоторый объём свободного места на пуле.

Рекомендуется удалять снимки сразу, как только они станут не нужны.

В текущей версии ПО, создание мгновенных снимков и клонов не поддерживается для томов, созданных на быстрых пулах.

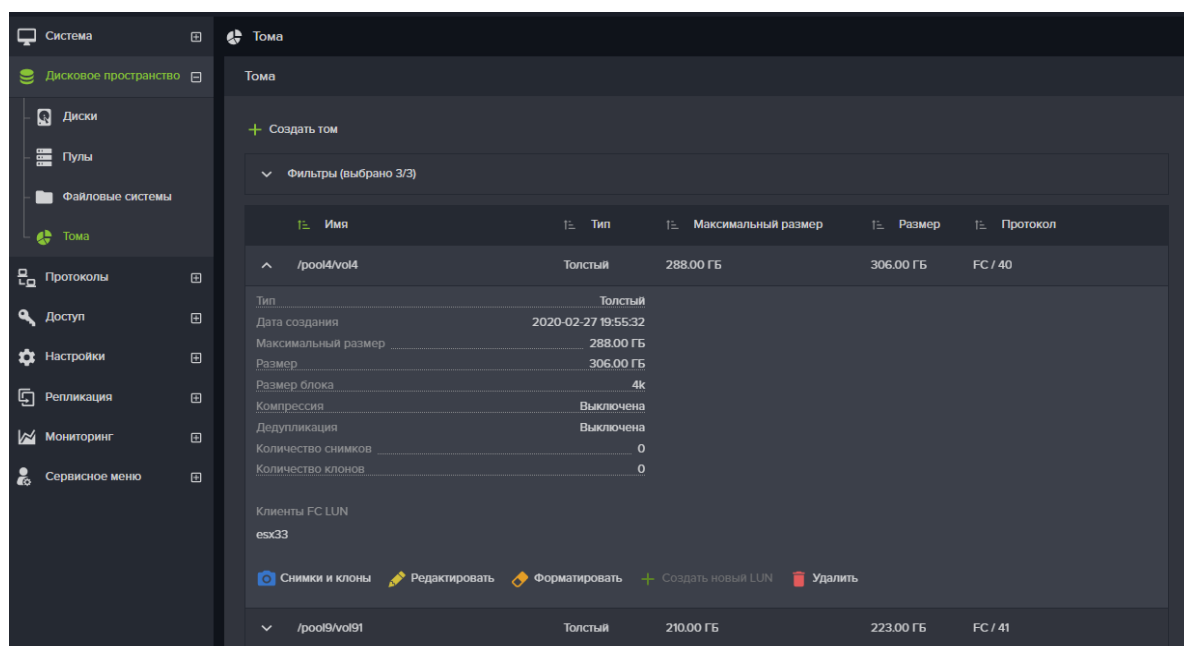


Рисунок 136. Свойства Тома

Снимки создаются в рамках быстрой и простой процедуры:

1. Перейдите на вкладку меню «Дисковое пространство» → «Тома» либо «Файловые системы» (см. Рисунок 136);

2. Выберите нужный том в списке, развернув его панель свойств и нажмите на кнопку «Снимки и клоны»;
3. В открывшемся окне нажмите кнопку «Создать» в разделе «Снимки» (см. Рисунок 137);
4. В открывшемся окне введите имя снимка и нажмите кнопку «Создать».



Внимание! Имя снимка должно начинаться с буквы.

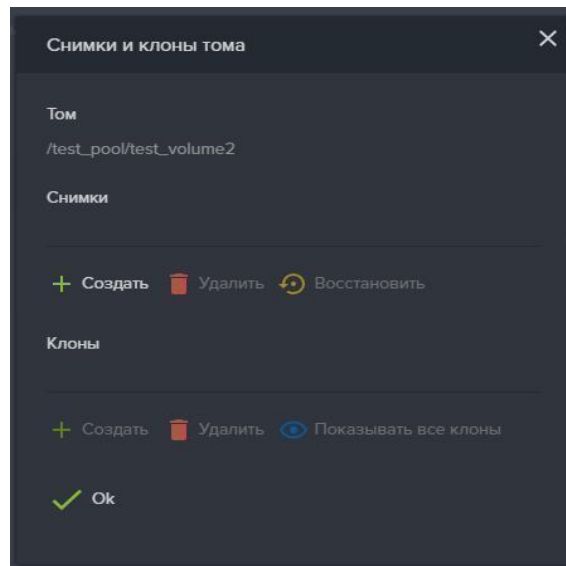


Рисунок 137. Создание снимка Тома

Созданный снимок отобразится в области «Снимки» в формате /имя_пула/имя_тома@имя_снимка (см. Рисунок 138).

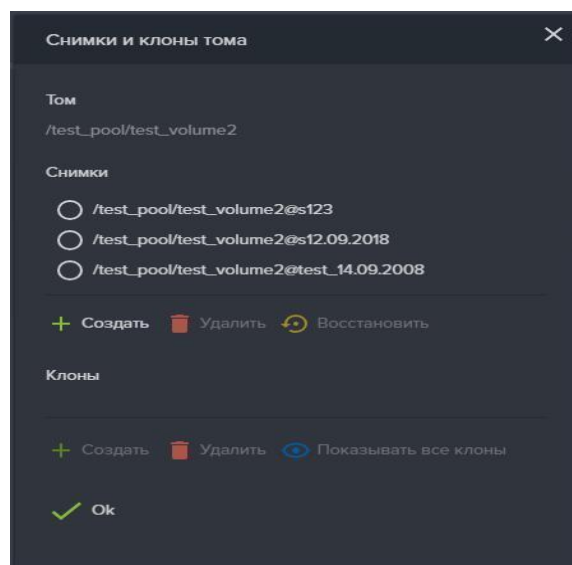


Рисунок 138. Список снимков Тома.

При необходимости может быть создано несколько снимков. При этом каждый снимок будет содержать изменения данных относительно предыдущего снимка.

2.2.2.8.2. Восстановление данных из мгновенного снимка.



Внимание! Во избежание возможной потери данных, перед выполнением восстановления данных из снимка, необходимо отключить нагрузку от восстанавливаемого ресурса.

Алгоритм восстановления тома такой:

1. останавливаем нагрузку и отключаем виртуальный диск на клиенте;
2. на СХД восстанавливаем раздел из снимка, дожидаясь завершения операции;
3. на клиенте повторно подключаем диск, проверяем что нужные данные были восстановлены, возобновляем нагрузку.

Внимание! При восстановлении данных тома или файловой системы при помощи снимков имеется одна особенность. Если восстанавливается один из последовательности нескольких снимков, то все снимки, сделанные позднее него, будут автоматически удалены.

Для того чтобы восстановить данные на момент создания любого снимка и при этом не потерять остальные, более поздние снимки, необходимо создать клон на основе выбранного снимка. К созданному клону может быть предоставлен доступ как к полноценному тому или файловой системе, однако он будет привязан к родительскому тому и снимку.

Для восстановления данных тома или файловой системы из снимка, выполните следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство»-«Тома» (или Файловые Системы);
2. Выберите из списка томов/ФС нужный том, содержимое которого хотите восстановить из снимка;
3. На панели свойств тома нажмите на кнопку «Снимки и клоны», откроется окно «Снимки и клоны»;
4. В списке имеющихся снимков отметьте нужный снимок, см. Рисунок 138;
5. Нажмите на кнопку «Восстановить»;
6. Подтвердите операцию введя «ОК» в открывшемся диалоговом окне;
7. Дождитесь завершения процедуры восстановления данных;

2.2.2.8.3. Клонирование мгновенного снимка.

Для работы с содержимом снимка, этот снимок необходимо клонировать.

Для клонирования снимка, выполните следующие действия.

8. Перейдите на вкладку меню «Дисковое пространство» → «Тома» либо «Файловые системы»;
9. Выберите нужный том в списке, развернув его панель свойств и нажмите на кнопку «Снимки и клоны» (см. Рисунок 136);
10. В открывшемся окне выберите ранее созданный снимок;
11. Нажмите в разделе «Клоны» на кнопку «Создать» (см. Рисунок 139);
12. Введите в открывшемся окне имя клона.
13. Нажмите на кнопку «Создать».



Внимание! Имена клонов не могут начинаться с буквы «с» (*си*) или цифр, и не должны содержать точку и другие спецсимволы.

В результате выполненных действий в списке «Клоны» появится клон с именем /имя_пула/имя_клона, а в области уведомления появится сообщение «Клон /имя_пула/имя_клона успешно создан».

Для просмотра всех клонов, без учета их принадлежности конкретному снимку, нажмите на кнопку «Показывать все клоны» (см. Рисунок 140).

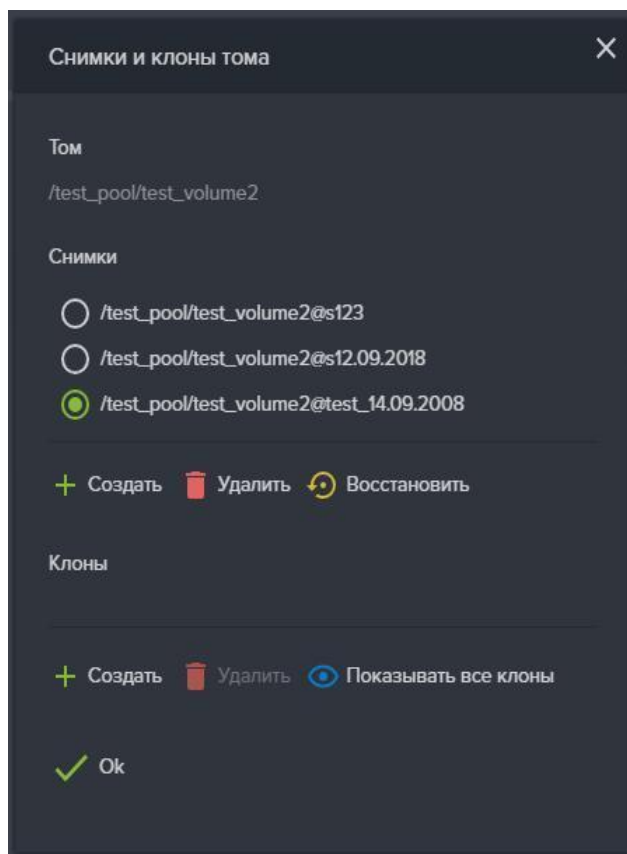


Рисунок 139. Окно Снимки и клоны тома.

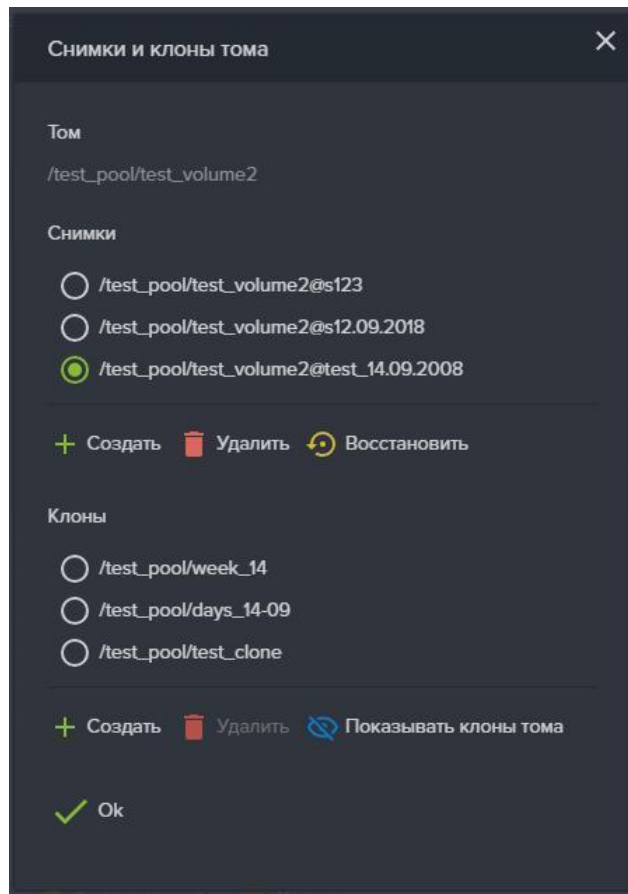


Рисунок 140. Окно снимков и клонов со списком сделанных клонов

2.2.2.8.4. Привязка LUN к созданному клону

После создания клона, его можно презентовать клиенту. Для этого необходимо привязать LUN.

Для привязки LUN к созданному клону выполните следующие действия:

1. На вкладке меню «**Томы**» откройте панель свойств созданного клона;
2. Нажмите на кнопку «**Создать новый LUN**»;
3. В открывшемся окне выберите протокол (FC) для доступа к создаваемому ресурсу;
4. Выберите свободный номер LUN;
5. Отметьте флажком клиентов или группу, которым будет разрешен доступ к создаваемому ресурсу;
6. Нажмите кнопку «**Создать**».

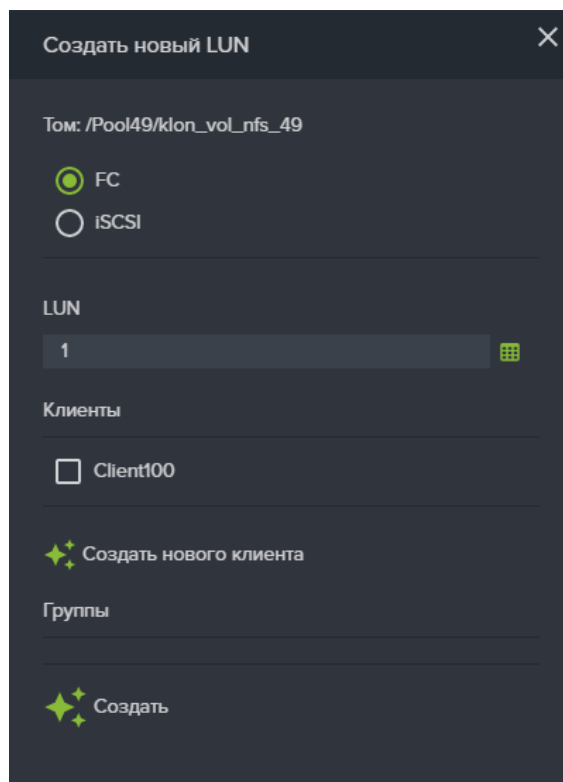


Рисунок 141. Окно привязки луна к клону

2.2.2.8.5. Удаление клонов

Перед удалением клона, необходимо удалить привязанный к нему LUN (если таковой имеется). Для удаления LUN перейдите на вкладку меню «Протоколы» - «FC» и удалите LUN.

Для удаления клона, выполните следующие действия:

7. Перейдите на вкладку меню «Дисковое пространство» → «Тома» либо «Файловые системы»;
8. Выберите нужный том или файловую систему в списке, развернув его панель свойств и нажмите на кнопку «Снимки и клоны» (см. Рисунок 136);
9. В открывшемся окне выберите ранее созданный клон;
10. В разделе «Клоны» нажмите на кнопку «Удалить»;
11. Подтвердите удаление введя в открывшемся окне слова «**ok**» и нажмите «Удалить».

В результате выполненных действий имя клона исчезнет из списка «Клоны» и в области уведомления появляется сообщение «Клон /имя_пула/имя_клона успешно удален.»

2.2.2.8.6. Создание мгновенных снимков по расписанию.

Для создания мгновенных снимков (файловой системы или тома) по расписанию, выполните следующие действия:

1. Перейдите на вкладку меню «Репликация» → «Периодические задачи»;
2. Нажмите на кнопку «Создать новую задачу», откроется окно создания задач (см. Рисунок 142);
3. Выберите из списка файловую систему или том, с которого хотите сделать снимок;
4. Выберите из списка созданный заранее шаблон расписания;
5. Задайте количество копий, по достижению которого следующие снимки будут перезаписывать ранее созданные;
6. Нажмите на кнопку «Добавить расписание», внизу в области «Реплики(цель/расписание)», появится выбранное расписание и кол-во его запусков.
7. Нажмите на кнопку «Создать».
8. В списке периодических задач должна появиться новая задача.

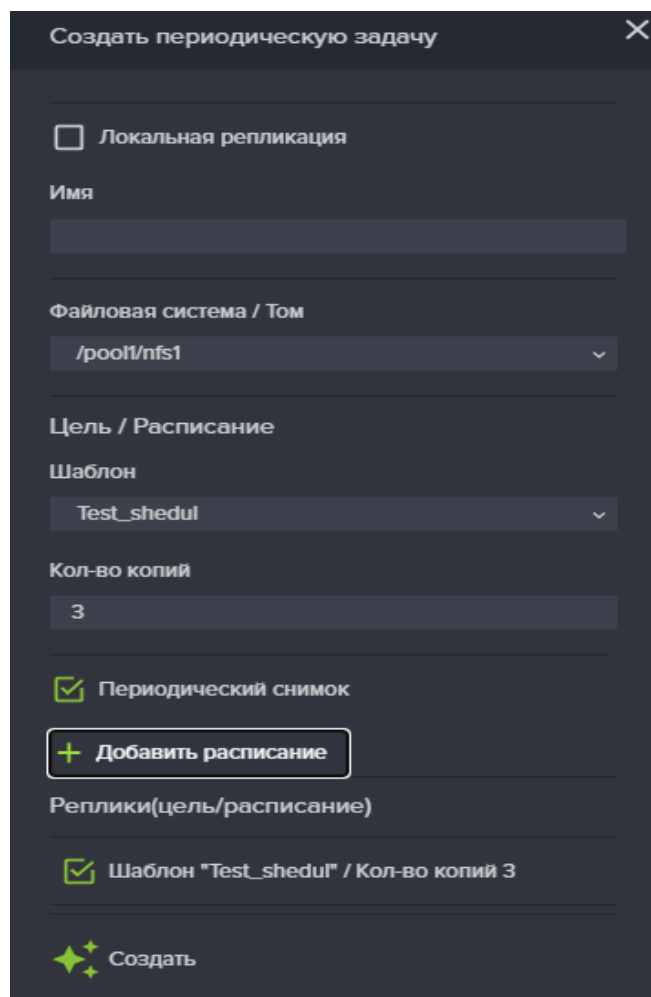


Рисунок 142. Окно создания расписания снимков

При необходимости запущенное задание на выполнение снимков по расписанию можно приостановить.

Для этого раскройте панель свойств задачи и нажмите кнопку «Отменить». В последствии создание снимков можно продолжить, нажав кнопку «Перезапустить» см. Рисунок 143.

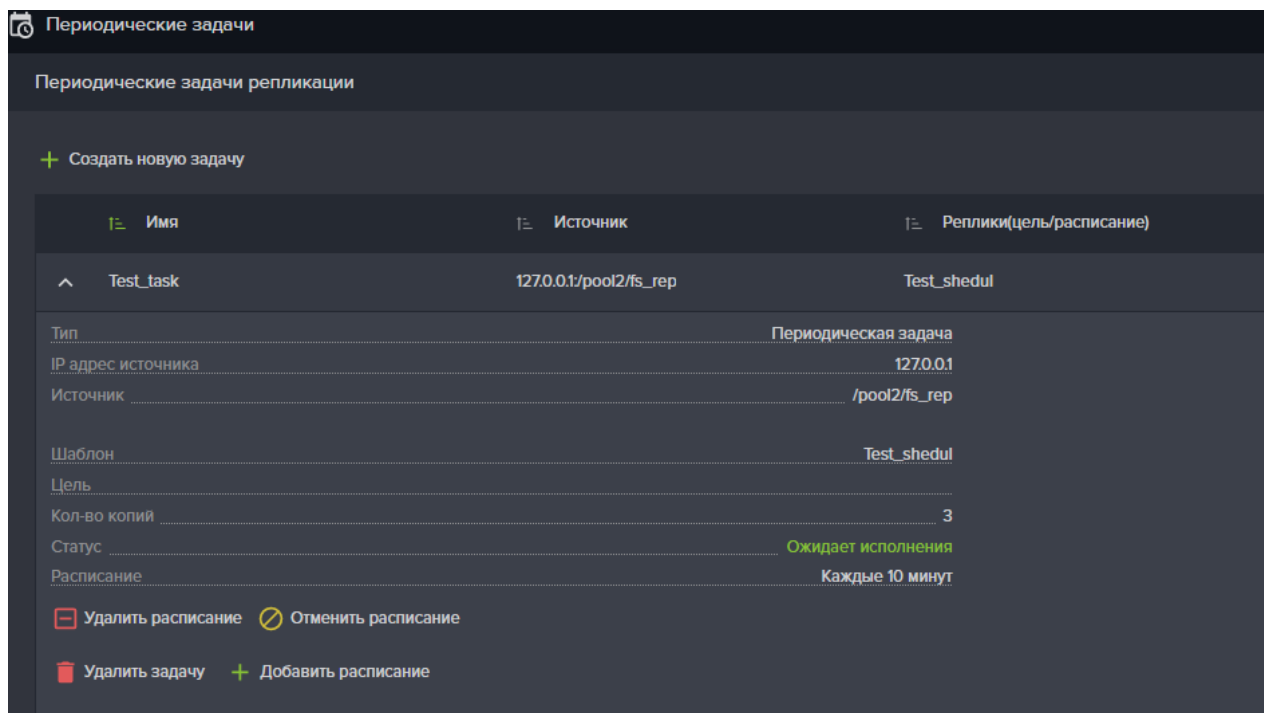


Рисунок 143. Окно свойств периодических снимков

Созданные снимки будут появляться в Свойствах выбранной файловой системы или тома.

2.2.2.8.7. Удаление задачи создания мгновенных снимков по расписанию.

Для удаления задачи выполните следующие действия:

1. Перейдите на вкладку меню «Репликация» → «Периодические задачи»;
2. Выберите нужную задачу, развернув её панель свойств;
3. Нажать на кнопку «Удалить»;
4. Подтвердите удаление, введя в открывшемся окне слова «ok» и нажмите «Удалить».

При успешном удалении расписания снимка, в области уведомления появится сообщение «Задача репликации **имя_задачи** успешно удалена» и имя задачи исчезнет из списка.

2.2.2.8.8. Удаление мгновенных снимков тома или файловой системы.

Удаление снимка доступно только для снимков, из которых не созданы клоны. Если требуется удалить снимок, из которого сделан клон – нужно сначала удалить клон, и только после этого удалить снимок.

Для удаления снимка тома или файловой системы выполнить следующие действия:

1. Перейдите на вкладку меню «Дисковое пространство» → «Тома» либо «Файловые системы»;
2. Выберите нужный том или файловую систему в списке, развернув его панель свойств и нажмите на кнопку «Снимки и клоны», откроется одноименное окно;
3. В области «Снимки» выберите нужный снимок и нажмите на кнопку «Удалить» (см. Рисунок 144).

Чтобы удалить несколько снимков отметьте их флажками, для удаления сразу всех снимков, отметьте флажок «Выбрать все»;

4. Подтвердите удаление, введя в открывшемся окне слова «ок» и нажмите «Удалить».

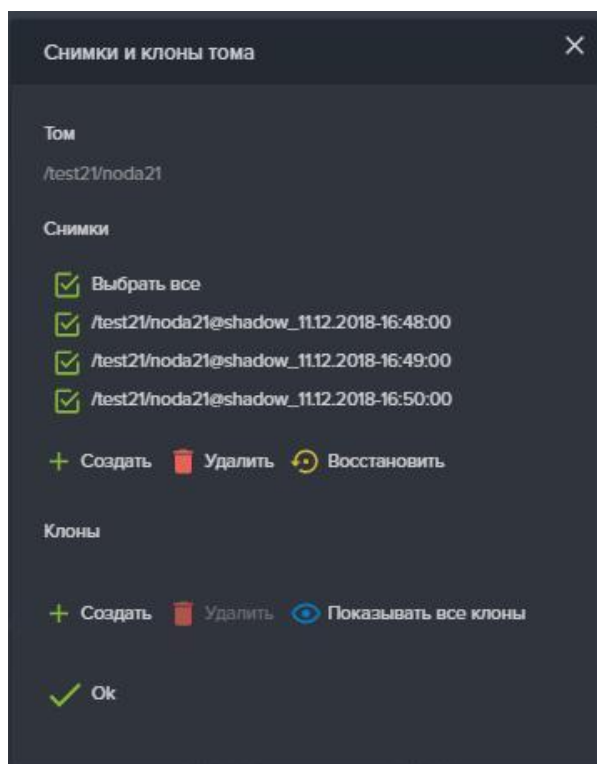


Рисунок 144. Окно снимков и клонов

При успешном удалении снимка, в области уведомления появится сообщение «Снимок /пул/том@снимок успешно удален» и в области «Снимки» удалится строки выбранных снимков, над которыми выполнялись операции.

2.2.2.9. Асинхронная репликация

В текущей версии ПО, репликация томов, созданных на быстрых пулах, не поддерживается.

Как и службы файловых протоколов, служба асинхронной репликации требует привязки к интерфейсу, через который будет выполняться взаимодействие со второй СХД. Для обеспечения отказоустойчивости, служба асинхронной репликации должна быть привязана к интерфейсу как на первом, так и на втором контроллере. Эти интерфейсы должны быть объединены в группу. Это позволит продолжить репликацию даже в случае отключения одного из контроллеров.

Репликация возможна только между томами одного вида: обычный том реплицируется на обычный том, быстрый том реплицируется на быстрый том. Могут быть реплицированы как тома так и файловые системы.

Перед началом настройки репликации необходимо создать шаблон цели. Если начало репликации должно быть выполнено в определенное время, понадобится создать шаблон расписания.

До начала настройки репликации убедитесь в том, что в сетевых настройках указан адрес NTP сервера, и время на целевой и удаленной системах, между которыми будет настроена репликация, совпадает.

2.2.2.9.1. Создание шаблона расписания

Для создания шаблона расписания войдите на вкладку меню «Репликация» → «Шаблоны» → «Расписания».

1. Нажмите на кнопку «**Добавить новый шаблон расписания**»;
2. В открывшемся окне, введите имя расписания;
3. Выберите тип запуска: «**Основное**» или «**Интервал**»;
4. Выберите периодичность запуска: при выборе опции «**Ежедневно**», запуск будет выполняться каждый день в установленное время. Выбор опции «**Выбранные дни**» позволяет указать конкретные дни, в которые будет осуществлен запуск задачи;

Выбор опции «**Интервал**» позволяет задать промежуток между запусками в днях, часах и минутах;

5. Нажмите кнопку «**Создать**», для создания нового расписания. Новый шаблон должен появиться в списке шаблонов расписаний.

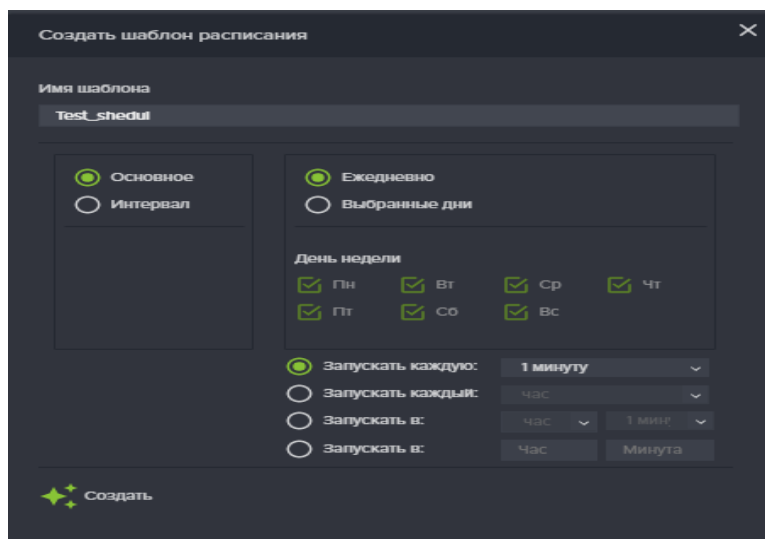


Рисунок 145. Окно шаблона расписания

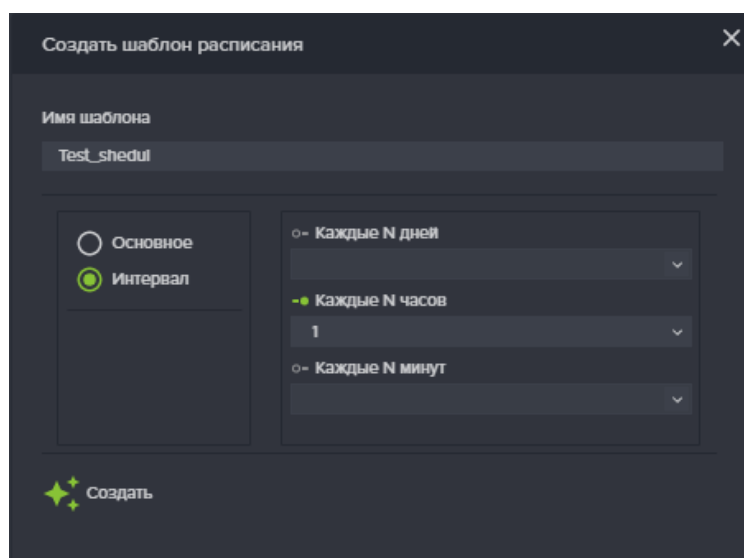


Рисунок 146. Окно расписания после выбора опции «Интервал»

После создания расписания его можно изменить, нажав на кнопку «Изменить расписание» в панели свойств созданного расписания.

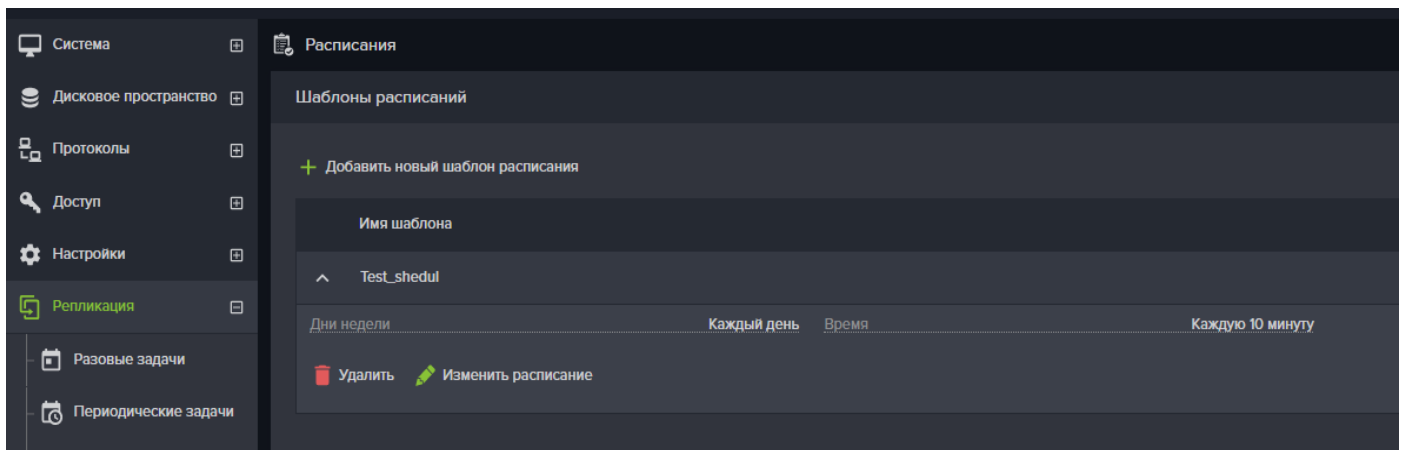


Рисунок 147. Окно меню «Шаблоны расписаний»

2.2.2.9.2. Создание шаблона цели

Для создания шаблона цели войдите на вкладку меню «Репликация» → «Шаблоны» → «Цели».

1. Нажмите на кнопку «**Добавить новую цель**»;
2. В открывшемся окне введите имя для новой цели;
3. При репликации на пул того же контроллера, выберите опцию «**Локальная репликация**»;
4. Задайте IP-адрес источника, откуда будет выполняться репликация;
5. Задайте IP-адрес цели, и нажмите на кнопку «**Получить GUID-ы**»;
6. В появившемся списке выберите пул, на котором будет создана реплика;
7. В поле «**Новое имя файловой системы/ Тома**» введите имя;
8. Нажмите на кнопку «**Создать**», новый шаблон должен появиться в списке шаблонов целей.

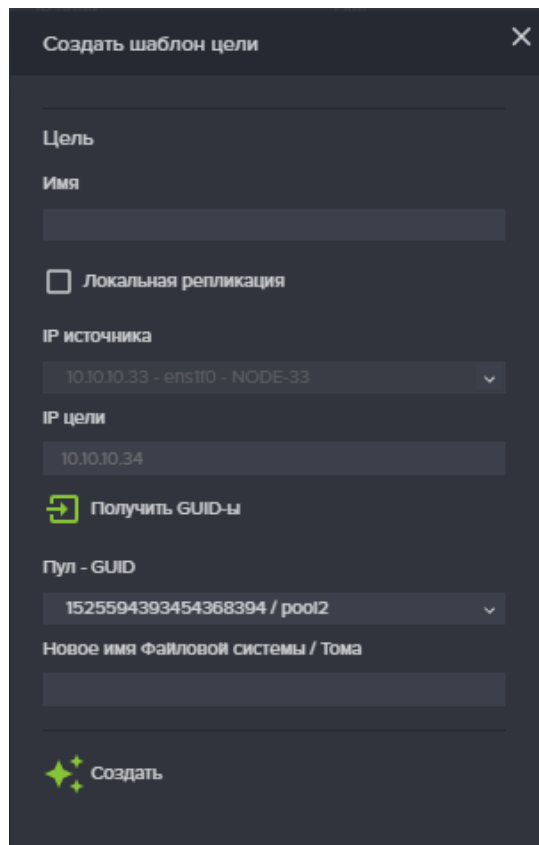


Рисунок 148. Создание шаблона цели

2.2.2.9.3. Задачи приема

Для создания задач приема репликации на удаленном СХД необходимо выполнить следующие действия:

1. Перейдите на вкладку меню **«Репликация»** → **«Задачи приема»**;
2. Нажмите на кнопку **«Создать новую задачу»** (см. Рисунок 149);
3. Введите название задачи в поле **«Имя»**;
4. Выберите из выпадающего меню **«Цель»** пул, на который будет осуществляться репликация;
5. Нажмите на кнопку **«Добавить IP адрес»**, откроется окно добавления IP адресов источников;
6. Введите IP адрес в соответствующее поле и нажмите на кнопку **«Добавить»** (см. Рисунок 150);
7. Добавьте другие IP адреса, при необходимости принимать репликации с нескольких СХД;
8. Отметьте флажок **«Автоматическое удаление»** при необходимости удаления задачи после запуска;
9. Нажмите на кнопку **«Создать»**;

10. В списке задач приема должна появиться новая задача.

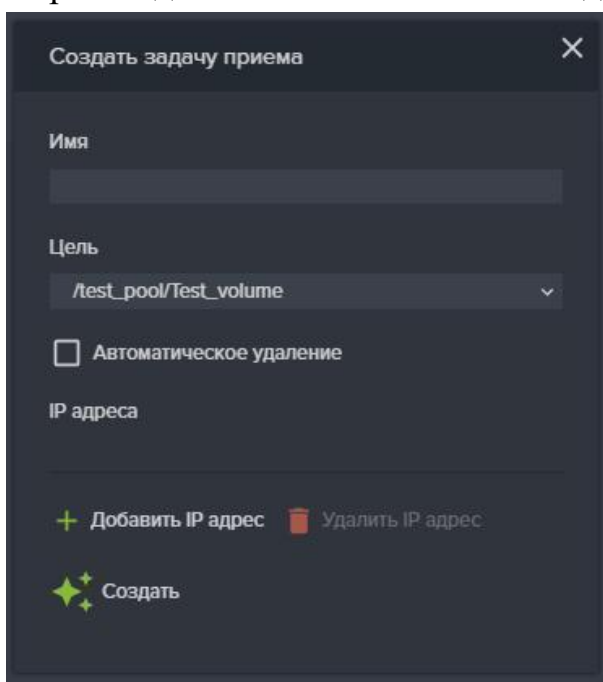


Рисунок 149. Задача приема репликации

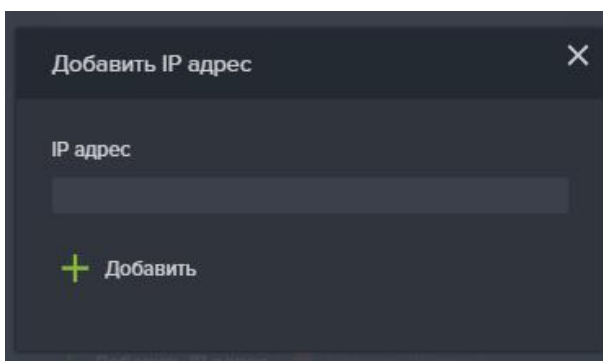


Рисунок 150. Добавление IP адреса

При успешном создании задачи приема, в области уведомления появится сообщение «Задача репликации **Название_задачи** успешно создана» и имя созданной задачи появится в списке.

2.2.2.9.4. Разовые задача репликации

Для создания задачи репликации тома или файловой системы на другой пул или удаленное СХД необходимо выполнить следующие действия.

Для репликации на другую СХД:

1. Перейдите на вкладку меню «**Репликация**» → «**Разовые задачи**»;
2. Нажмите на кнопку «**Создать новую задачу**» (см. Рисунок 151);
3. Введите название задачи в поле «**Имя**»;
4. В списке «**Файловая система / Том**» области «**Источник**» выберите

файловую систему либо том, с которого будет осуществляться репликация;

5. В области «**Цель/Расписание**» выберите из списка заранее созданный шаблон цели, а если нужно отложить выполнение задачи, нажмите кнопку «**Отложенная задача**» и выберите заранее созданный шаблон расписания.

6. Для автоматического удаления созданной задачи после её выполнения, отметьте опцию «**Автоматическое удаление**»

7. Нажмите на кнопку «**Создать**».

Для настройки локальной репликации:

1. Установите флажок «**Локальная репликация**»;

2. В области «**Источник**» выберите из списка имеющуюся файловую систему или том для репликации.

3. В списке «**Цель**» выберите заранее созданный шаблон цели;

4. При необходимости отложить выполнение, нажмите кнопку «**Отложенная задача**» и выберите созданный шаблон расписания, либо добавьте новый шаблон расписания

5. Нажмите «**Создать**». В списке разовых задач репликации должна появиться новая задача.

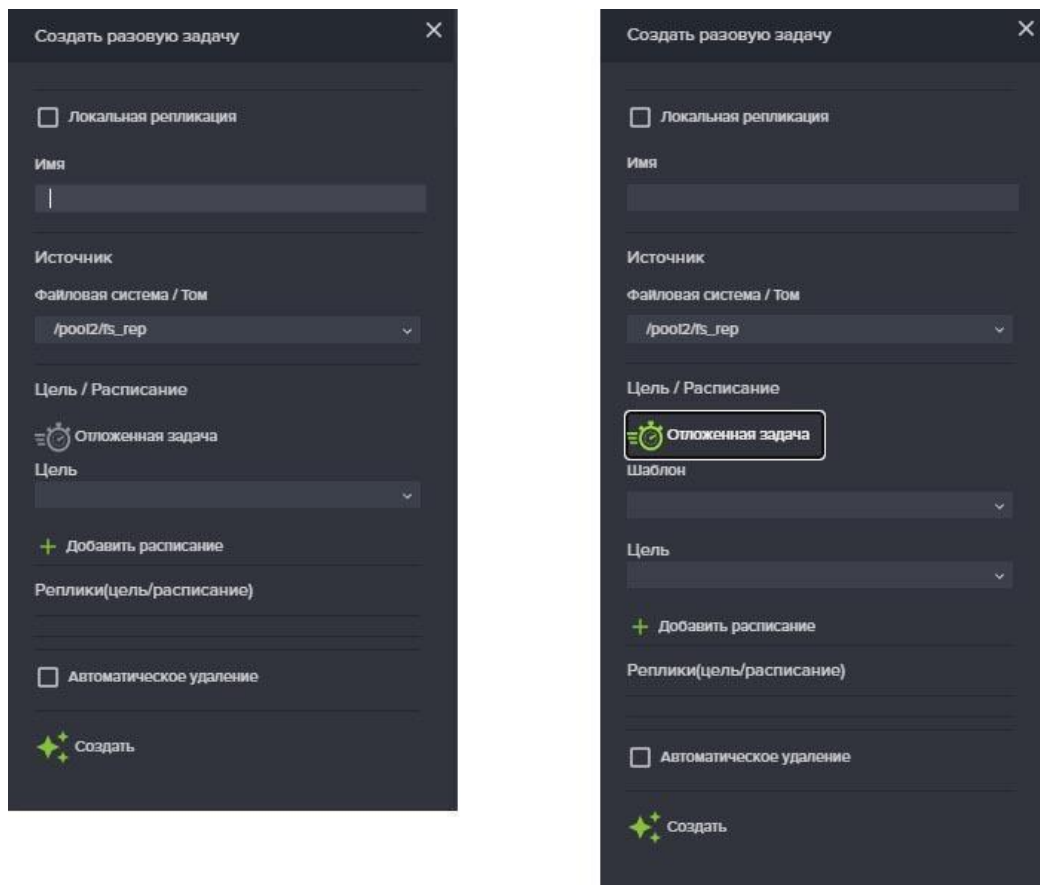


Рисунок 151. Создание разовой задачи репликации

При успешном создании задачи репликации, в области уведомления появится сообщение «Задача репликации Название задачи успешно создана» и в списке разовых задач появится имя задачи. В свойствах задачи появится статус «Выполняется». После выполнения задачи репликации статус изменится на «Выполнено». На другом пуле или на удаленной СХД появится реплика файловой системы или тома.

2.2.2.9.5. Периодические задачи

Для создания периодических задач репликации, которые должны выполняться по расписанию, выполните следующие действия:

Для репликации на другую СХД.

1. Перейдите на вкладку меню «Репликация» → «Периодические задачи»;
2. Нажмите на кнопку «Создать новую задачу», откроется окно создания задач (см.Рисунок 152);

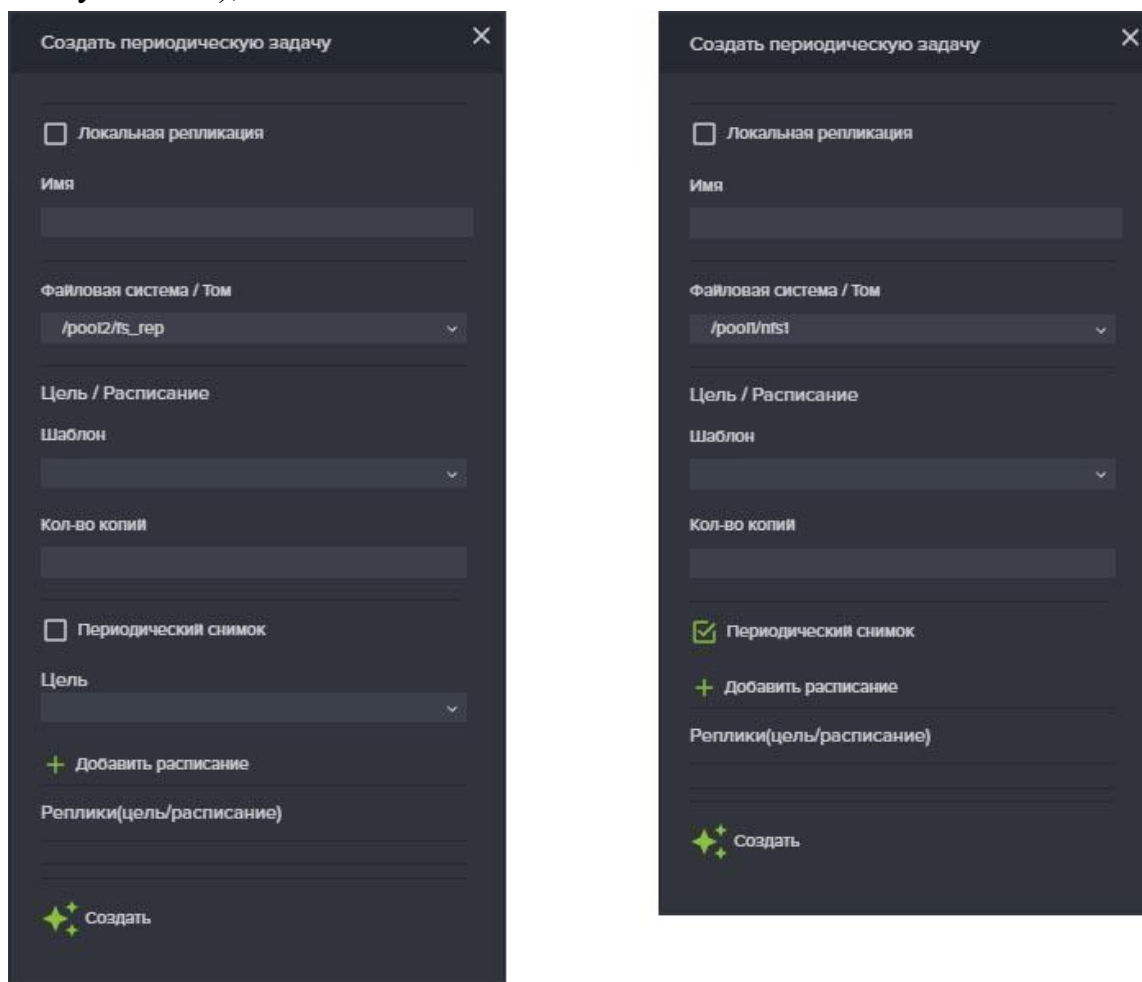


Рисунок 152. Расписание периодической задачи репликации

3. Задайте Имя создаваемой задачи;

4. Выберите из списка файловую систему либо том, которую будете реплицировать;
5. Выберите из списка заранее созданный шаблон расписания, к которому задан интервал репликации;
6. Укажите количество копий (снимков, которые будут храниться на принимающей стороне);
7. Выберите из списка заранее созданный шаблон цели;
8. Нажмите на кнопку «**Добавить расписание**». В поле «**Реплики (цель/расписание)**» появится название выбранного шаблона расписания и кол-во копий.
9. Нажмите на кнопку «**Создать**».
10. В списке периодических задач репликации должна появиться новая задача.

Для локальной репликации.

Локальная репликация выполняется в пределах одной ноды.

Шаблон цели, который будет использоваться для локальной репликации, должен быть создан с опцией «**Локальная репликация**».

1. Поставьте флажок «**Локальная репликация**»;
2. Задайте Имя для создаваемой задачи;
3. Выберите из списка файловую систему либо том, которую будете реплицировать;
4. Выберите из списка заранее созданный шаблон расписания, к которому задан интервал репликации;
5. Укажите количество копий (снимков, которые будут храниться на принимающей стороне);
6. Выберите из списка заранее созданный шаблон цели;
7. Нажмите на кнопку «**Добавить расписание**». В поле «**Реплики (цель/расписание)**» появится название выбранного шаблона расписания и кол-во копий.
8. Нажмите на кнопку «**Создать**».
9. В списке периодических задач репликации должна появиться новая задача.

2.2.2.9.6. Пример настройки асинхронной репликации на вторую СХД

Имеется 2 СХД, каждая состоит из двух контроллеров (1/2 и 3/4).

На одном контроллере каждой СХД создано по одному обычному пулу с именами "rep1" на первой СХД и "rep2" на второй СХД. На пуле "rep1", той СХД с которой будет произведена репликация, создан том с именем "vol1". На пуле "rep2", той СХД куда будет произведена репликация, создан том с именем "new_vol1".

На каждой СХД есть группа сетевых интерфейсов для репликации, с настроенными IP адресами. К IP адресам должна быть привязана служба ASR. Сетевые интерфейсы должны быть объединены в группу.

1. Подключитесь к одному из контроллеров СХД2.
2. Перейдите на вкладку меню «Репликация» – «Задачи приема». Нажмите на кнопку «Создать новую задачу».
3. Введите имя для задачи – "get_rep".
4. В поле «Цель», выберите "/rep2/new_vol1". Это том, на который будет произведена репликация.
5. Добавьте IP-адрес контроллера СХД1, откуда будет проводиться репликация (IP-адрес интерфейса для репликации со стороны СХД1).
6. Нажмите на кнопку «Создать».
7. Подключитесь к контроллеру СХД1 и перейдите на вкладку меню "Шаблоны" - "Цели".
8. Нажмите на кнопку "Добавить новую цель".
9. Введите имя "target"
10. Укажите IP адрес источника. Это IP сетевого интерфейса контроллера, с которого будет идти репликация.
11. Укажите IP цели. Это IP сетевого интерфейса контроллера, на который будет идти репликация.
12. Нажмите на кнопку "Получить GUID-ы".
13. Из выпадающего списка выберите пул СХД2, куда будет произведена репликация "/rep2".
14. В поле "Новое имя Файловой систем/ Тома" укажите том "new_vol1". (Это имя уже созданного тома, куда будет выполняться репликация).
15. Нажмите на кнопку "Создать".
16. Перейдите на вкладку меню "Репликация" - "Разовые задачи". Нажмите на кнопку "Создать новую задачу".
17. Введите имя "task".
18. В выпадающем списке "Источник" - "Файловая система / Том" выберите том, который будете реплицировать - "/rep1/vol1".

19. В выпадающем списке "Цель", выберите шаблон нашей цели - "target".

20. Нажмите на кнопку "Добавить расписание".

21. Нажмите на кнопку "Создать".

Проверяем репликацию.

1. Подключаем том /rep1/vol1 СХД1 к любому клиенту и записываем на него любой файл.

2. Уже созданным заданием обновляем реплицированный том (нажимаем на кнопку «перезапустить»).

3. Подключаем том /rep2/new_vol1 СХД2 к любому клиенту, проверяем, что файл появился.

2.2.2.10. Синхронная репликация

В текущей версии ПО можно одновременно выполнять только одну задачу синхронной репликации. Репликация может выполняться только в одну сторону.

Внимание! Имена томов под метаданные не должны оканчиваться цифрой!

Система позволяет выполнять синхронную репликацию тома на другой пул текущего кластера, либо на удаленный кластер по протоколу FC. Репликация может выполняться между блочными устройствами (томами), размещенными на быстрых либо обычных пулах в любой комбинации. Размер тома под данные на принимающей стороне (цель), должен быть на 1 ГБ больше размера тома на отдающей стороне (источник).

До начала настройки репликации убедитесь в том, что в сетевых настройках указан адрес NTP сервера, и время на целевой и удаленной системах, между которыми будет настроена репликация, совпадает.

Вкладка меню синхронной репликации показана на Рисунок 153.

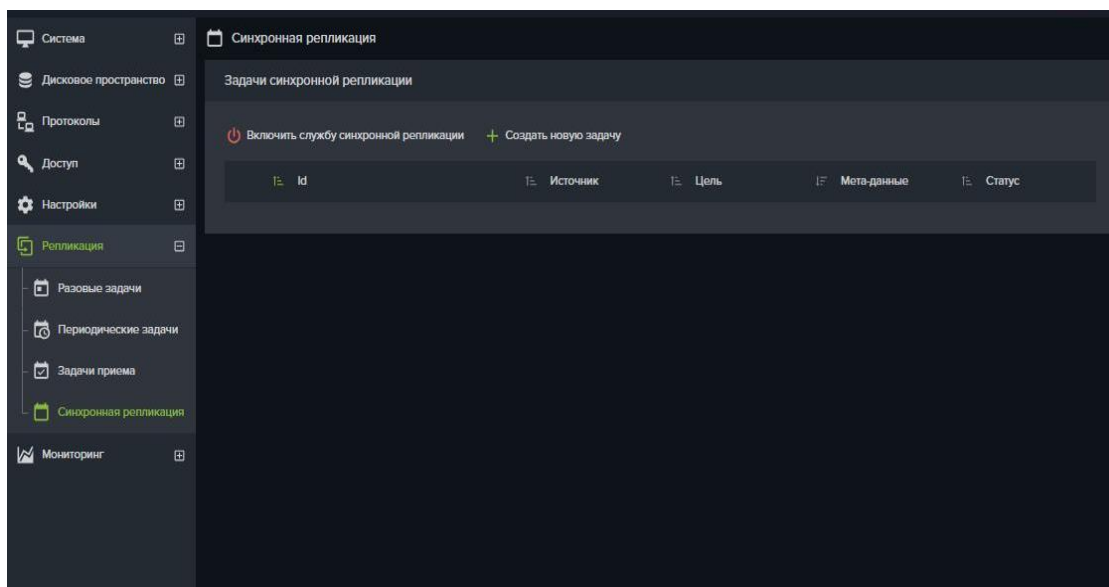



Рисунок 153. Окно меню синхронной репликации

2.2.2.10.1. Задача синхронной репликации

Для создания задачи синхронной репликации выполните следующие действия:

1. Перейдите на вкладку меню «**Репликация**» → «**Синхронная репликация**»;
2. Нажмите «**Включить службу синхронной репликации**», если служба выключена;
3. Нажмите на кнопку «**Создать новую задачу**», откроется окно задачи см. Рисунок 154;
4. В поле «**Тип**» выберите «**Перемещение**» или «**Репликация**»;
5. В поле «**Источник**» выберите том который будет реплицирован или перемещен;
6. В поле «**Цель**» выберите том для приёма  реплики (обновите поле если не отобразится целевой том);
7. В поле «**Мета-данные**» укажите том под мета-данные;
8. Задайте ограничение скорости;
9. Нажмите «**Создать**» для запуска задачи.

Созданная задача появится в списке задач синхронной репликации. В поле свойств задачи, в строке «Завершено» будут показаны проценты выполнения синхронизации данных, а после завершения синхронизации в строке «Статус репликации» появится слово «Норма» см. Рисунок 156.

Выполняющуюся синхронную репликацию можно приостановить или отменить. Для этого предусмотрены кнопки «**Приостановить**» и «**Отменить**».

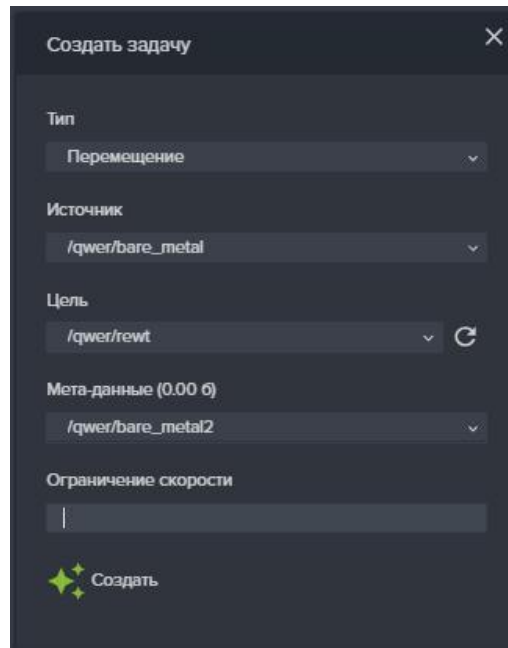


Рисунок 154. Окно задачи синхронной репликации

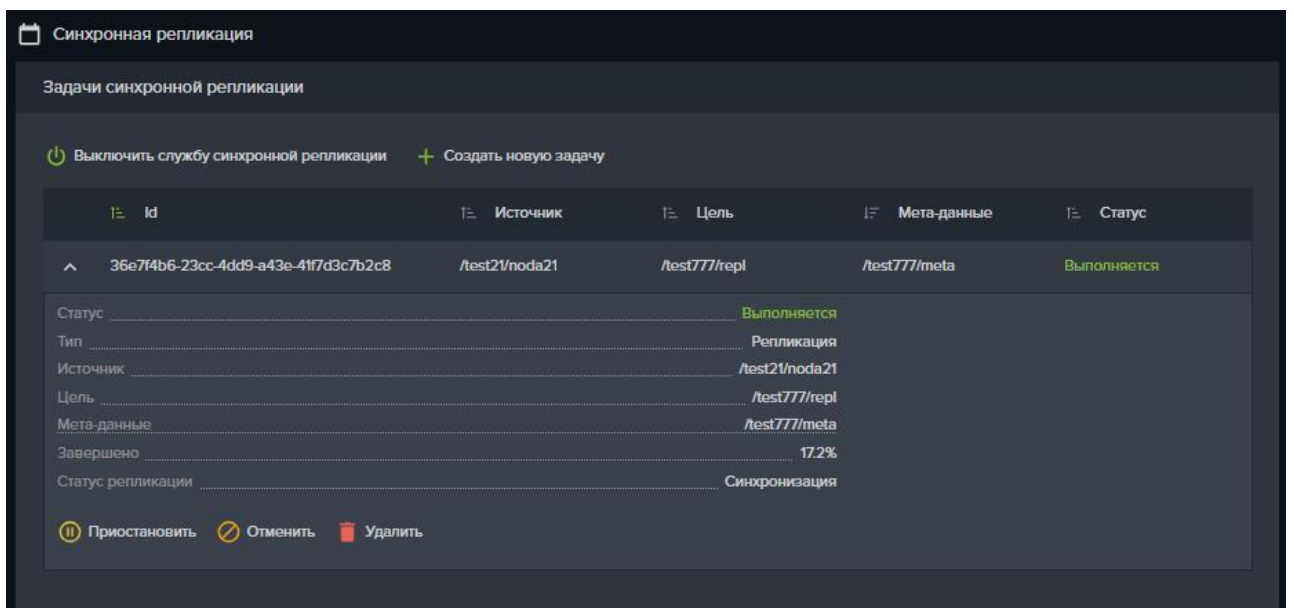


Рисунок 155. Окно свойств задачи синхронной репликации. Синхронизация данных

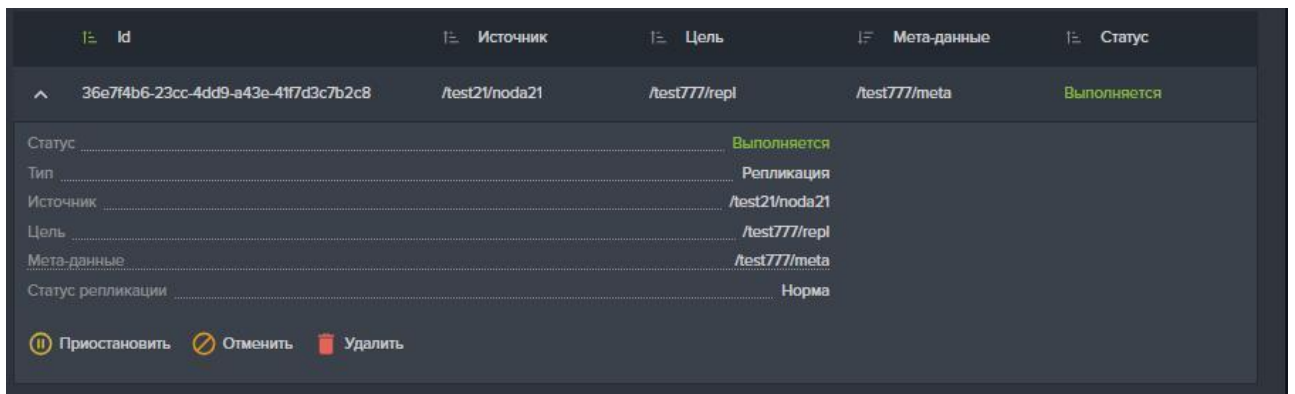


Рисунок 156. Завершение синхронизации данных при синхронной репликации

2.2.2.10.2. Перемещение тома между пулами одного контроллера

Перенос тома является видом синхронной репликации. Данный функционал работает только с «толстыми» томами.

Для выполнения перемещения тома на другой пул выполните следующие действия:

1. Выберите или создайте пул, на который планируете перенос тома;
2. Создайте на пуле толстый том объёмом больше на 1 ГБ, чем том который будет перемещён;
3. Создайте на этом же пуле толстый том размером 2-3 Гб под мета-данные;
4. Перейдите на вкладку меню «**Репликация**» → «**Синхронная репликация**»;
5. Нажмите «**Включить службу синхронной репликации**», если служба выключена;
6. Нажмите на кнопку «**Создать новую задачу**», откроется окно задачи см. Рисунок 132;
7. В поле «**Тип**» выберите «**Перемещение**»;
8. В поле «**Источник**» выберите том который будет реплицирован или перемещён;
9. В поле «**Цель**» выберите том для приёма реплики;
10. В поле «**Мета-данные**» укажите том под мета-данные;
11. Задайте ограничение скорости (если требуется снизить нагрузку от репликации на исходный том);
12. Нажмите «**Создать**» для запуска задачи.

Созданная задача появится в списке задач синхронной репликации. В поле свойств задачи, в строке «Завершено» будут показаны проценты выполнения

синхронизации данных, а после завершения синхронизации в строке «Статус репликации» появится слово «Норма» см. Рисунок 156. Если том источник был привязан к LUN, то после завершения задачи, выполнится автоматическое переключению к LUN целевого тома.

Процесс перемещения тома можно приостановить или отменить. Для этого предусмотрены кнопки «**Приостановить**» и «**Отменить**».

2.2.2.10.3. Синхронная репликация по протоколу FC

Для синхронной репликации на удаленную СХД по протоколу FC требуется определить какие из FC портов на отдающей СХД будут использоваться в качестве инициаторов и заранее отключить от них внешних клиентов.

Для синхронной репликации должны использоваться только «толстые» тома.

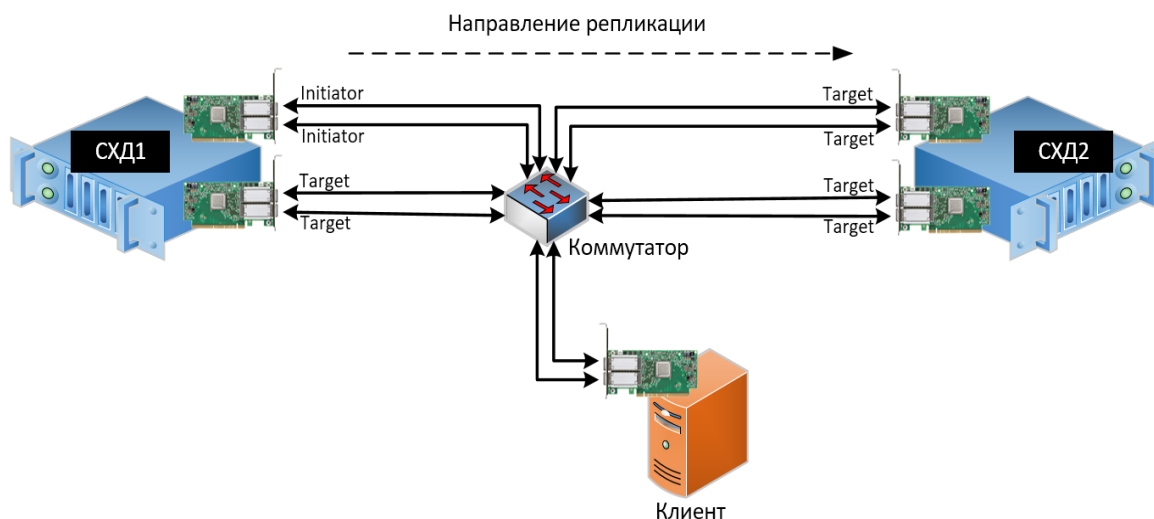


Рисунок 157. Схема коммутации портов СХД для настройки синхронной репликации

Синхронная репликация выполняется в два этапа: перенос данных на удаленный том - синхронизация и синхронная запись на оба тома.

Настройка, выполняемая на отдающем кластере:

1. На отдающем кластере переключите по одному порту каждого контроллера в режим инициатора, для чего перейдите на вкладку меню «**Протоколы**» → «**FC**»;
2. Разверните область «**FC порты**» см. Рисунок 158;
3. На каждом контроллере отметьте порт, который будет инициатором и нажмите кнопку «**Пометить как инициатор**»;
4. Уточните какой контроллер является владельцем тома, который будет реплицирован на другой кластер;

5. Перейдите в меню «**Дисковое пространство**» → «**Тома**»,
6. Создайте «толстый» том под мета-данные размером 2 Гб. Том должен быть создан на том же контроллере кластера, что и том выбранный для репликации.

Настройка, выполняемая на принимающем кластере:

1. Перейдите в меню «**Дисковое пространство**» → «**Тома**»;
2. Создайте «толстый» том на 1ГБ больше, чем том который будет реплицирован. Этот том будет использован для приёма реплики;
3. Перейдите в меню «**Протоколы**» → «**ФС**»;
4. Создайте LUN на созданном ранее томе;
5. Создайте клиента с wwrp адресами портов отдающей СХД (инициаторы).

Настройка задачи синхронной репликации на принимающей СХД:

1. Прейдите в меню «**Репликация**» - «**Синхронная репликация**»;
2. Включите службу синхронной репликации нажав «**Включить службу синхронной репликации**»;
3. Нажмите «**Создать задачу**», откроется окно задачи;
4. В поле «**Тип**» выберите «**Репликация**»;
5. В поле «**Источник**» выберите том, который будет реплицирован на другую СХД;
6. В поле «**Цель**» выберите том подготовленный для приёма реплики (lun);
7. В поле «**Мета-данные**» выберите том под мета-данные;
8. Нажмите «**Создать**» для запуска репликации.

Созданная задача появится в списке задач синхронной репликации. В процессе переноса данных будут показаны проценты выполнения, а после завершения процесса строка статус должна показать слово «**Норма**».



Внимание! Для доступа к реплике тома, сначала отключите клиента от тома – источника и приостановите (или удалите) задачу репликации.

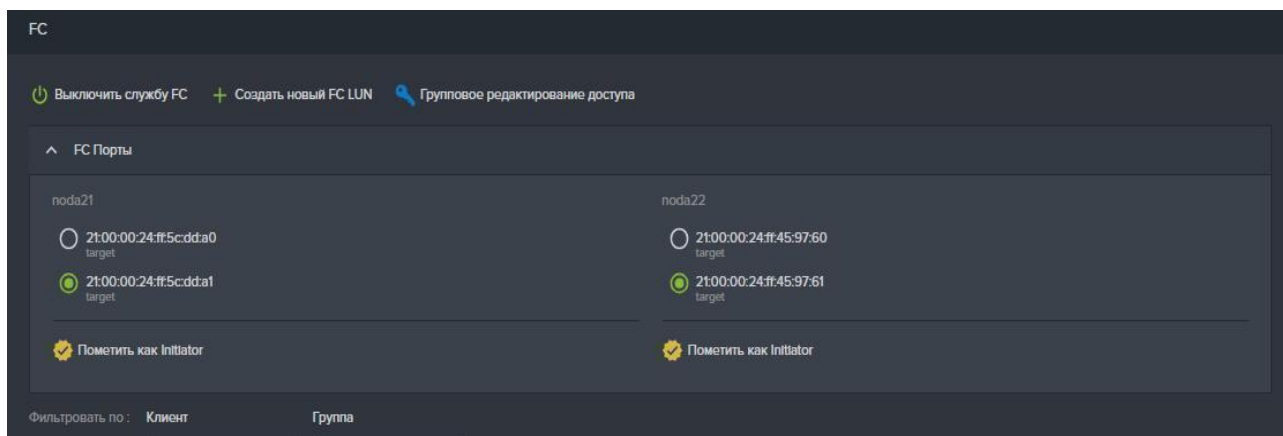


Рисунок 158. Панель FC порты на вкладке меню «Протоколы»-«FC»

2.2.2.10.4. Пример настройки синхронной репликация на удаленную СХД

Рабочая нагрузка на том первого кластера СХД выполняется виртуальным клиентом, запущенным на сервере VMware ESXI.

1. Настраиваем первый кластер СХД: Создаём на контроллере №1 пул с именем her1 и толстый том с именем Vola, размером 50 Гб. Этот том мы будем реплицировать.

2. Создаём на этом же контроллере толстый том под мета-данные, размером 2 Гб.

3. На каждом из контроллеров первого, «отдающего» кластера переводим два порта в режим инициатора, предварительно убедившись в том, что они не отданы клиенту.

4. Настраиваем клиента, в настройки которого вносим wwrp адреса сервера ESXI.

5. Создаём лун на томе Vola и отдаём его клиенту ESXI.

6. Настраиваем второй кластер СХД: Создаём на любом контроллере кластера толстый том с именем Volb, размером 51 Гб. Объем тома приемника должен быть на 1Гб больше чем у тома источника

7. Создаём и настраиваем клиента, указав те wwrp адреса портов первого кластера, которые мы перевели в режим инициатора.

8. Создаём лун на созданном томе Volb, и отдаём его первому кластеру.

9. Выполняем настройку задачи синхронной репликации на первом кластере:

10. Переходим на первый кластер, настраиваем задачу синхронной репликации, предварительно запустив соответствующую службу.

11. Нажмите «Создать новую задачу»

12. В окне задачи вводим:

Тип –Репликация;

Источник – том Vola, созданный на первом кластере;

Цель – для поиска луна соседней ноды, нажимаем на кнопку – сканирование дисков, в списке должен отобразиться том, созданный на втором кластере;

Мета-данные – том под метаданные, созданный на первом кластере.

17. Нажимаем кнопку «Создать»

18. При открытии задачи статус репликации должен показывать проценты выполнения, и после окончания процесса покажет слово «Норма». А статус задачи будет «Выполняется»

19. Отключите том, с которого проводилась репликация от клиента;

20. Отмените либо приостановите запущенную задачу репликации;

21. Подключите к хосту среплицированный том и проверьте доступность файлов.

2.2.2.11. Подключение к файловым ресурсам

Файловый ресурс предоставляется клиентам в виде каталога доступного пользователю в локальной сети, при этом файловую систему организует СХД.

Подключение к клиентским хостам выполняется как напрямую (DAS), так и с использованием сети передачи данных.

2.2.2.11.1. Подключение к файловым ресурсам по протоколу NFS

2.2.2.11.1.1. Подключение к файловым ресурсам в среде Linux (Ubuntu)



Внимание! Все команды выполняются с правами суперпользователя (*root*).

Для подключения ресурса по протоколу NFS выполните следующие действия:

1. Установите NFS-клиент: **apt-get install nfs-common**.
2. Создайте каталог в /mnt: **mkdir /mnt/nfs** (Есть возможность создать каталог в другом месте, но рекомендуется здесь).
3. Смонтируйте NFS ресурс к клиенту: **mount -t nfs ip:/config/массив/ресурс /mnt/nfs** (монтируйте туда, где был создан каталог).

2.2.2.11.1.2. Подключение к файловым ресурсам в среде VMWare.

Для подключения datastore по протоколу NFS выполните следующие действия:

1. Запустите web клиент VMWare vSphere.
2. Перейдите на вкладку «Datastores».
3. Нажать на кнопку «**New Datastore**» (см. Рисунок 159).

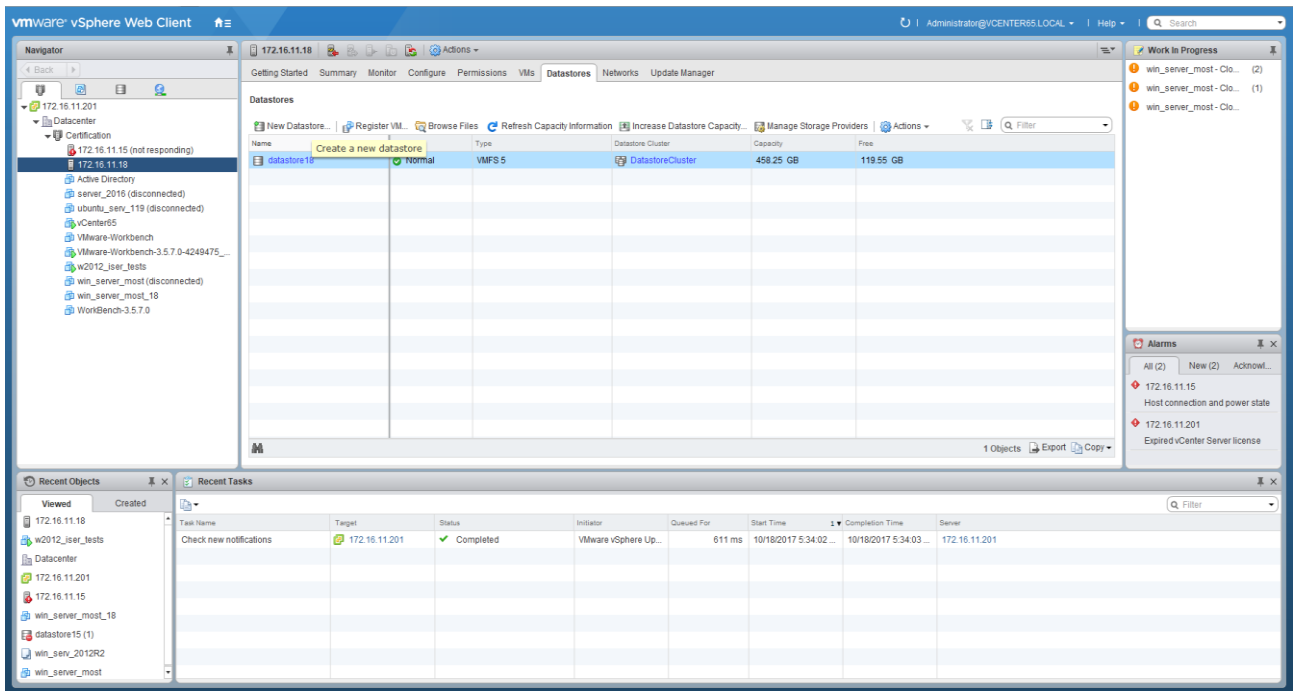


Рисунок 159. ESXi Datastore.

4. Выберите тип datastore - «NFS» (см. Рисунок 159)

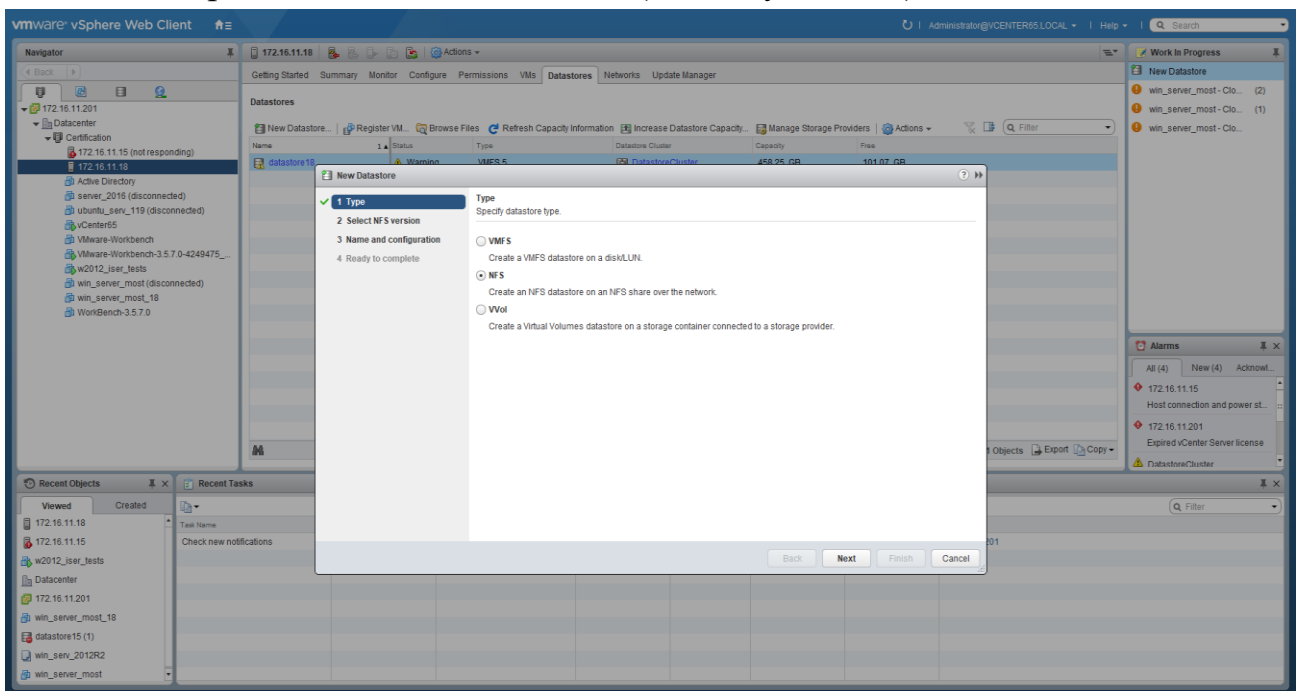


Рисунок 160. Окно выбора типа Datastore

5. Нажмите на кнопку «Next».
6. Выберите версию NFS – NFS 3 (см. Рисунок 160).

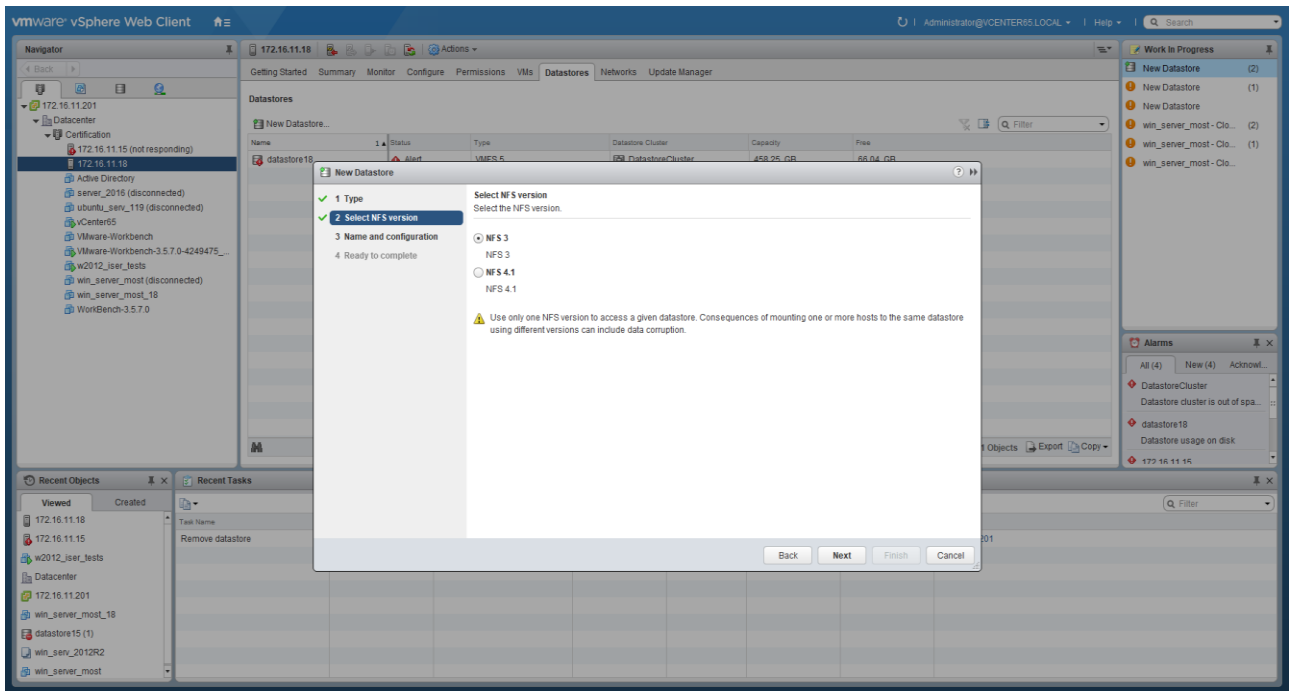


Рисунок 161. Версия NFS

7. Нажмите на кнопку «Next».
8. Введите в соответствующие поля требуемые данные (см. Рисунок 162):
 - В поле «*Datastore name*» введите имя datastore;
 - В поле «*Folder*» введите точку монтирования, которую можно найти в свойствах NFS папки во вкладке «Протоколы» — «NFS»;
 - В поле «*Server*» введите IP адрес контроллера владельца NFS ресурса.

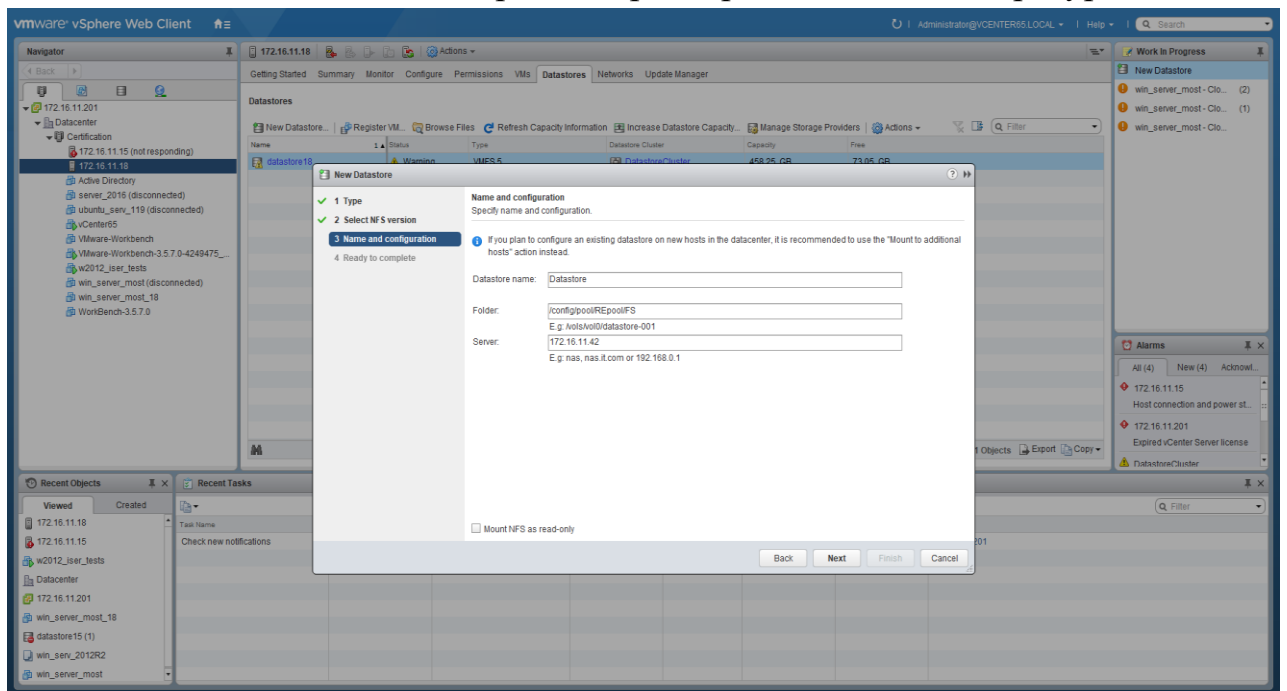


Рисунок 162. Данные NFS папки

9. Нажмите на кнопку «Next».
10. Подтвердите введенные данные нажатием кнопки «Finish» (см. Рисунок 163).

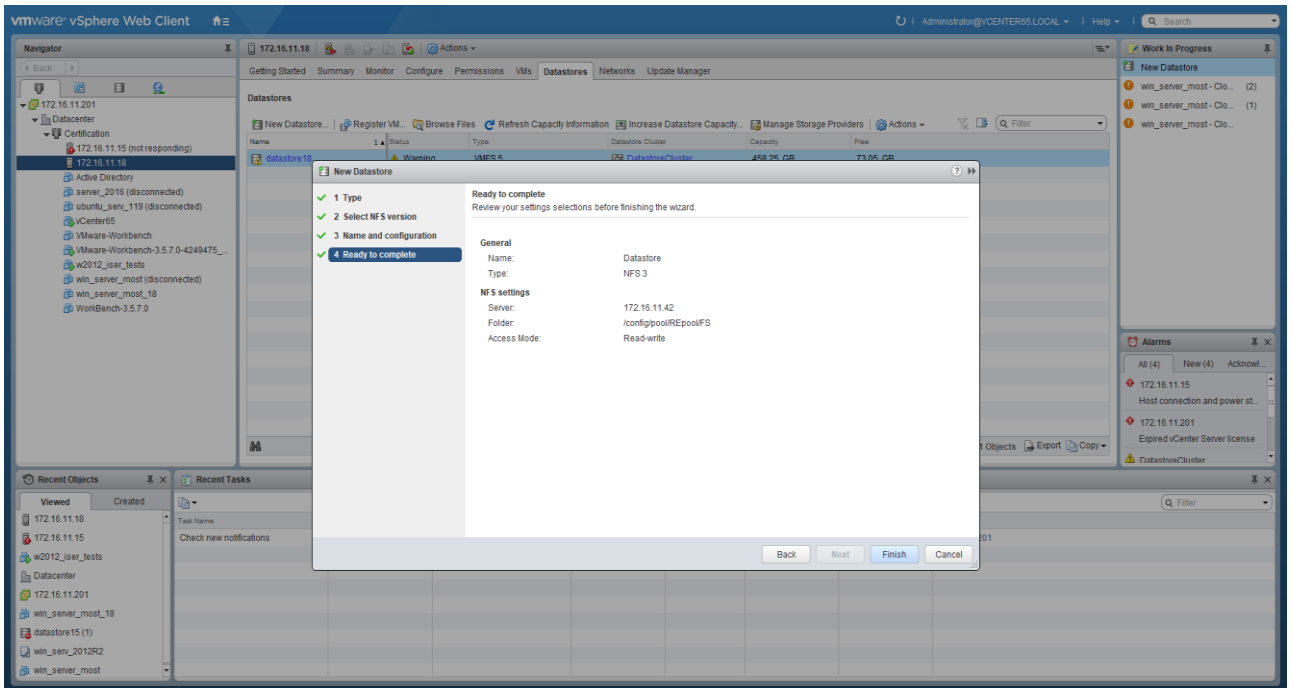


Рисунок 163. Подтверждение создания Datastore

В результате выполненных действий имя подключенного datastore по протоколу NFS появится в списке «Datastores» (см. Рисунок 164).

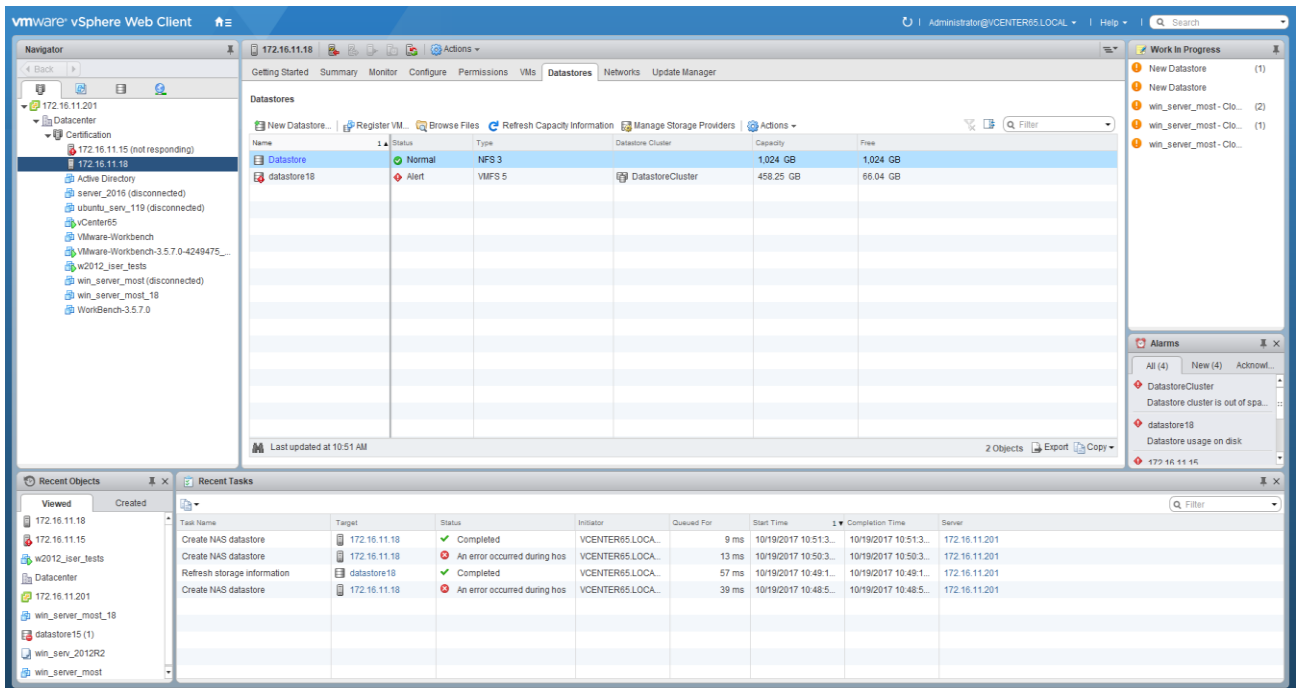


Рисунок 164. Datastores

2.2.2.11.2. Подключение к файловым ресурсам по протоколу SMB.

Для подключения ресурса по протоколу SMB выполните следующие действия:

1. В меню «Пуск» нажать на правую клавишу мыши на «Компьютер» (см. Рисунок 165).

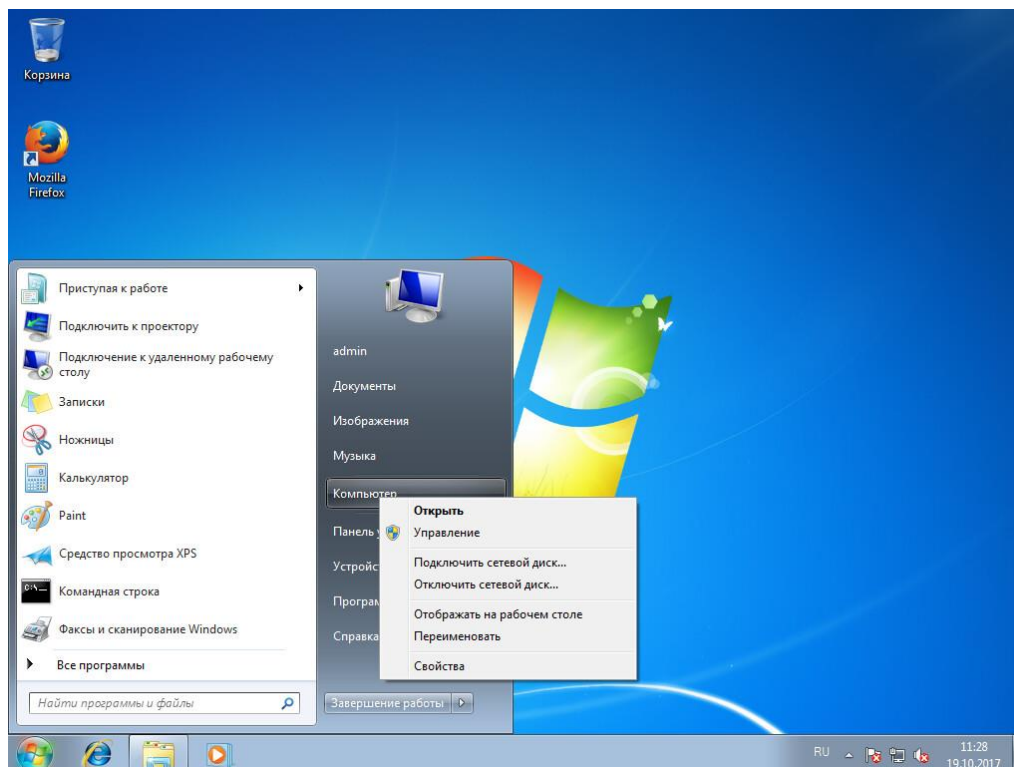


Рисунок 165. Подключение сетевого диска в Windows 7

2. Нажмите на кнопку «Подключить сетевой диск...»
3. Введите необходимые данные (см. Рисунок 166).
 - В поле «Диск» выбрать букву диска для подключения сетевого диска;
 - В поле «Папка» выбрать необходимый ресурс по шаблону:
«\\имя_контроллера\имя_пула_имя_файловой_системы»;
4. Нажмите на кнопку «Готово».

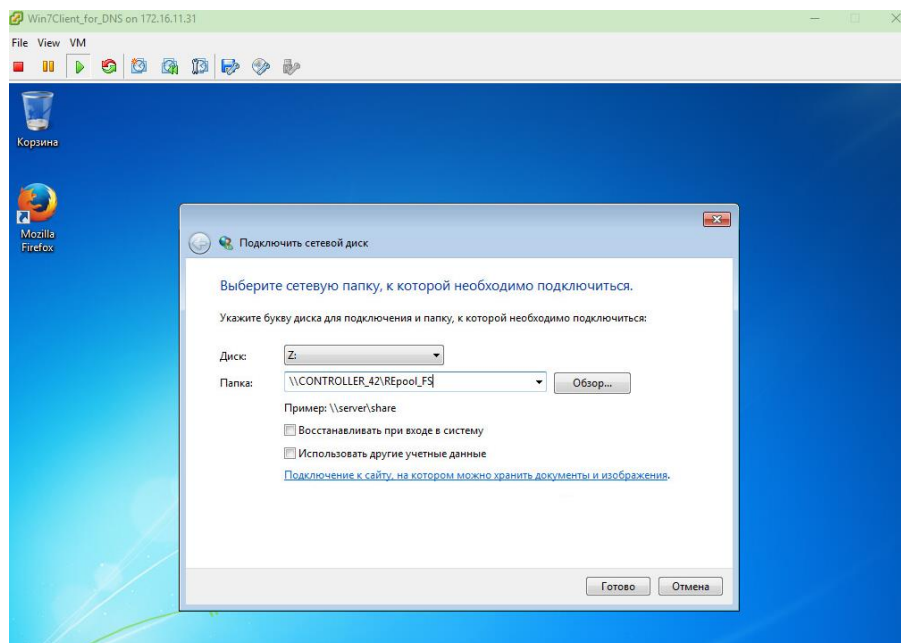


Рисунок 166. Окно подключения сетевого диска.

В результате выполненных действий в «Компьютер» в области «Сетевое размещение» появится сетевой диск (см. Рисунок 165).

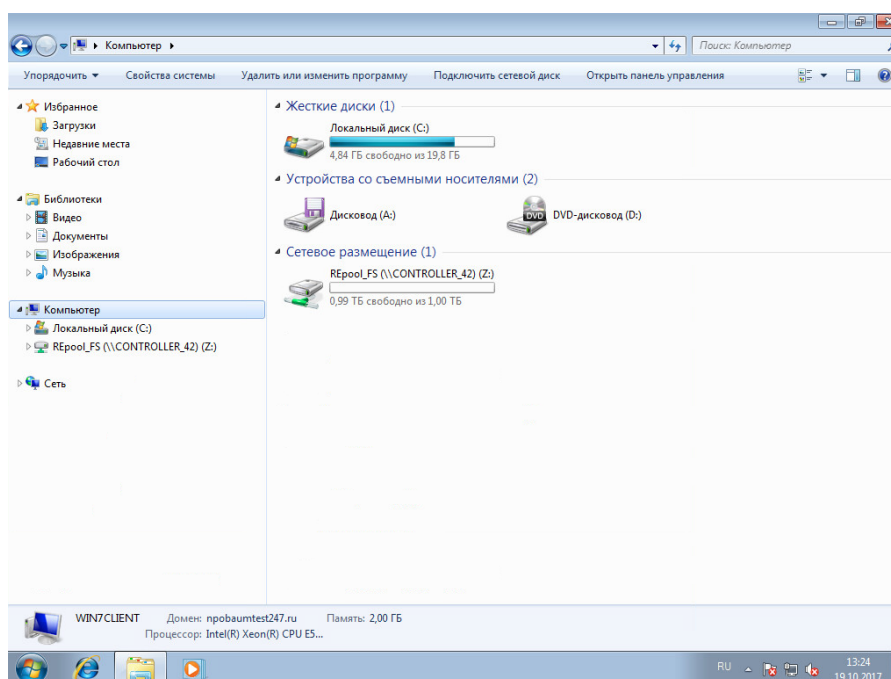


Рисунок 165. Сетевой диск.

2.2.2.12. Подключение к блочным ресурсам

Виртуальный том (протокол FC) – ресурс, выделенный в определенном пуле, представляется в виде блочного устройства. Для клиента блочное устройство представляется как обычный жесткий диск компьютера. На виртуальном диске пользователь может создать необходимую ему файловую систему и работать с ним как с обычным диском компьютера. При этом подключение к хосту может быть как

прямое (DAS), так и через сеть хранения данных (SAN). Настройка драйверов MPIO и DSM для протокола Fibre Channel осуществляется одинаково.

Адреса WWN или IQN target обоих контроллеров можно посмотреть в разделе соответствующего протокола (Протоколы – FC) развернув пункт FC Targets.

2.2.2.12.1. Подключение к блочным ресурсам по протоколу FC.

2.2.2.12.1.1. Подключение к блочным ресурсам в среде Windows Server 2012 по протоколу FC.

Для управления настройками и просмотра WWN порта можно использовать специализированное ПО от производителя Fibre Channel адаптера. Для этого:

1. Перейдите в панель управления компьютером и выберите раздел «MPIO»;
2. На вкладке «Обнаружение многопутевых устройств» нажмите кнопку «Добавить». Система выдаст сообщение о необходимости перезагрузки системы. После перезагрузки в разделе «Управление дисками» панели «Управление компьютером» можно убедиться, что диск будет доступен по нескольким маршрутам (см. Рисунок 166);

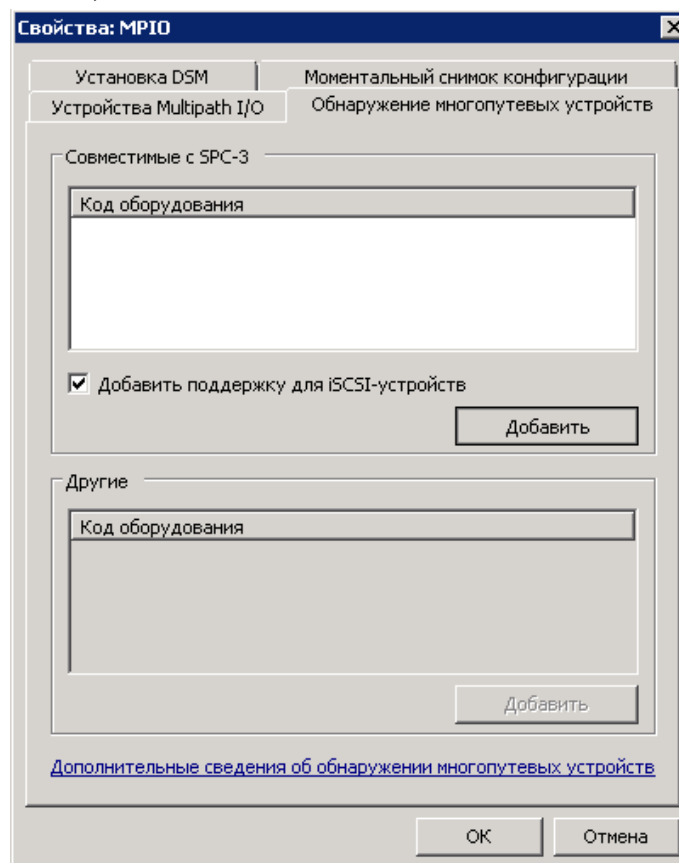


Рисунок 166. Добавление устройств

3. В разделе «Управление дисками» панели «Управление компьютером»

нажмите правой кнопкой мыши по созданному диску. В окне «Свойства: наименование диска» на вкладке «Многопутевой ввод-вывод» выбрать политику МPIO «**Хотя бы глубина очереди**» (см. Рисунок 167). Затем перейти на вкладку «Драйвер» и нажать кнопку «**Подробнее**». В окне «Подробные сведения о DSM» задать рекомендуемые параметры, которые должны совпадать с ПО BAUMSTORAGE AI. На этом настройки завершены. Рекомендуемые параметры для DSM модуля показаны на следующем рисунке (см. Рисунок 167).

Для оптимальной производительности рекомендуется использовать Jumbo- frame на всей цепочке BAUMSTORAGE AI – ОС хоста. В ПО BAUMSTORAGE AI настраиваются на вкладке меню «Сетевые интерфейсы» с помощью параметра «MTU».

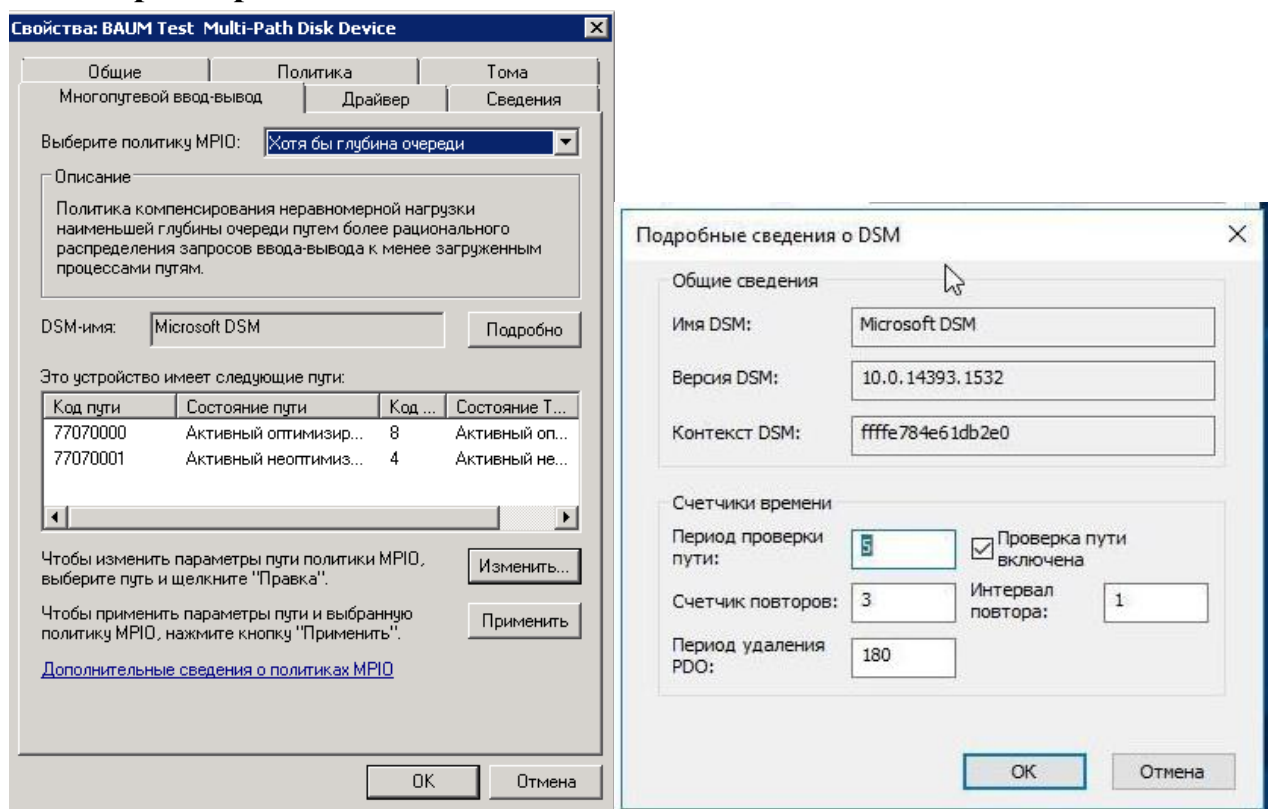


Рисунок 167. Настройка МPIO

2.2.2.12.1.2. Подключение к блочным ресурсам в среде Linux по протоколу FC.

Настройка для протокола Fibre Channel:

Для просмотра WWN портов по протоколу Fibre Channel можно использовать пакет **sysfsutils**.

1. Установку набора утилит можно выполнить с помощью команд: **aptitude install sysfsutils** (для Ubuntu/Debian)

yum install sysfsutils (для RHEL/CentOS)

2. Получить информацию о WWN номере порта можно, выполнив команду:

```
systool -c fc_host -v
```

Либо вывести список WWN портов:

```
cat /sys/class/fc_host/host*/port_name Полученная информация будет задана в port_name.
```

3. Пересканирование ресурсов FC:

```
for host in `ls /sys/class/scsi_host/`; do echo "- - -" > /sys/class/scsi_host/${host}/scan; done
```

4. Удалить блочное устройство:

```
echo 1 > /sys/block/sdX/device/delete
```

5. Установить пакет DM-Multipath (MPIO):

```
yum install device-mapper-multipath (для RHEL/CentOS)  
aptitude install multipath-tools (для Debian/Ubuntu Linux).
```

6. Создайте файл **/etc/multipath.conf** и внесите в него секцию **devices** для правильного обнаружения блочных устройств экспортируемых ПО
BAUMSTORAGE AI:

```
device {  
    vendor "BAUM"  
    product ".*"  
    detect_prio yes  
    dev_loss_tmo "infinity"  
    no_path_retry queue  
    prio "alua"  
    path_selector "queue-length 0"  
    path_grouping_policy "group_by_prio"  
    path_checker "directio"  
    failback "immediate"  
    rr_weight "uniform"  
    user_friendly_names yes  
}
```

7. выполните для применения настроек, сделанных в файле

/etc/multipath.conf:

multipath -k

> reconfigure

```
multipath -k reconfigure\\
multipathd -k \\
>>reconfigure\\
multipath -ll
```

Посмотреть состояние МPIO устройств можно с помощью команды **multipath**

```
mpatha (23030303030303031) dm-11 BAUM ,Test
size=1000G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=30 status=active
  |- 13:0:0:1 sdb 68:80 active ready running
  `-- 14:0:0:1 sdb 68:96 active ready running
```

Настройка завершена.

2.2.2.12.1.3. Подключение к блочным ресурсам в среде VMWare по протоколу FC.

Для подключения к блочным ресурсам по протоколу FC в среде VMWare выполните следующие действия:

1. Перейдите на вкладку «*Configuration*» нужного хоста ESXi.
2. Выберите вкладку «*Storage Adapters*».
3. Выберите порт Fibre Channel.
4. Нажмите на правую клавишу мыши на девайсе и выберите «**Manage**

Paths...» (см. Рисунок 168).

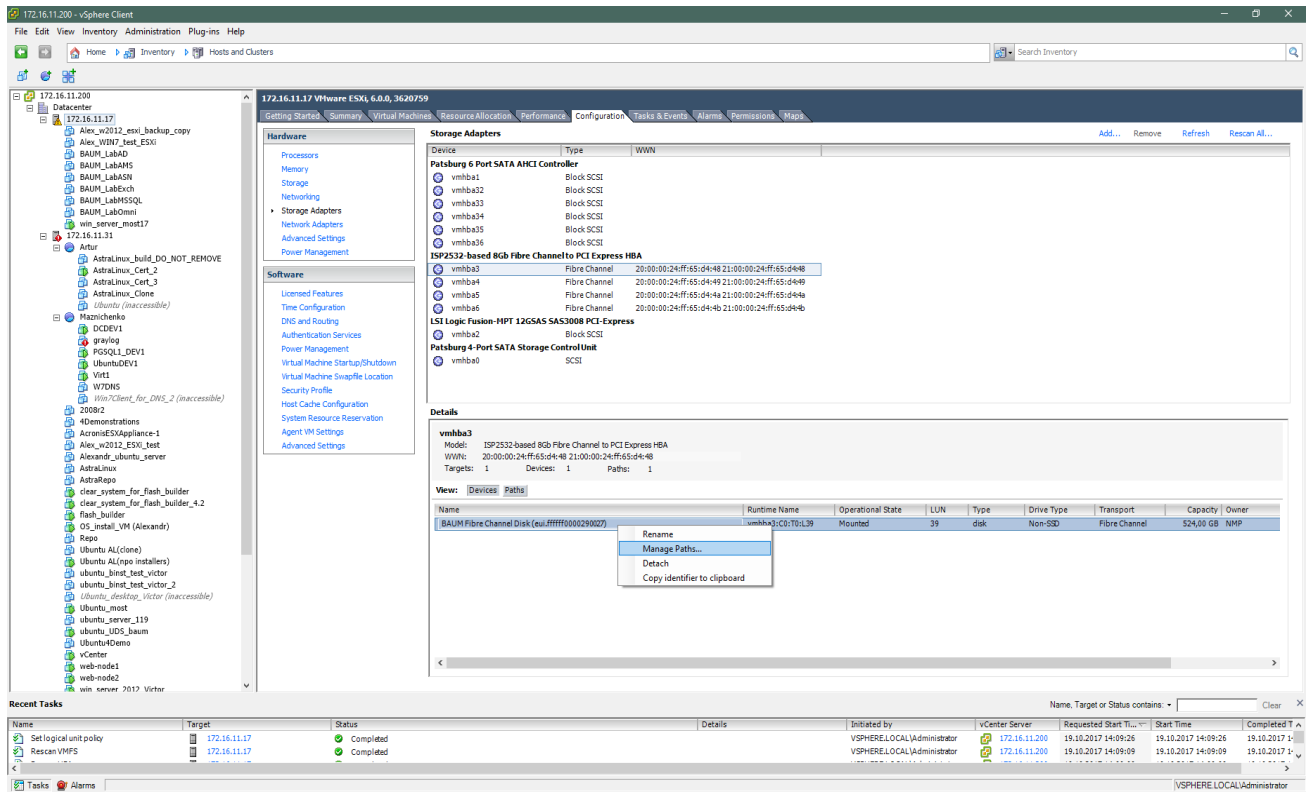


Рисунок 168. FC Configuration

5. В пункте Path Selection выберите **Round Robin** (см. Рисунок 169).
6. **Нажмите** на кнопку «**Change**»

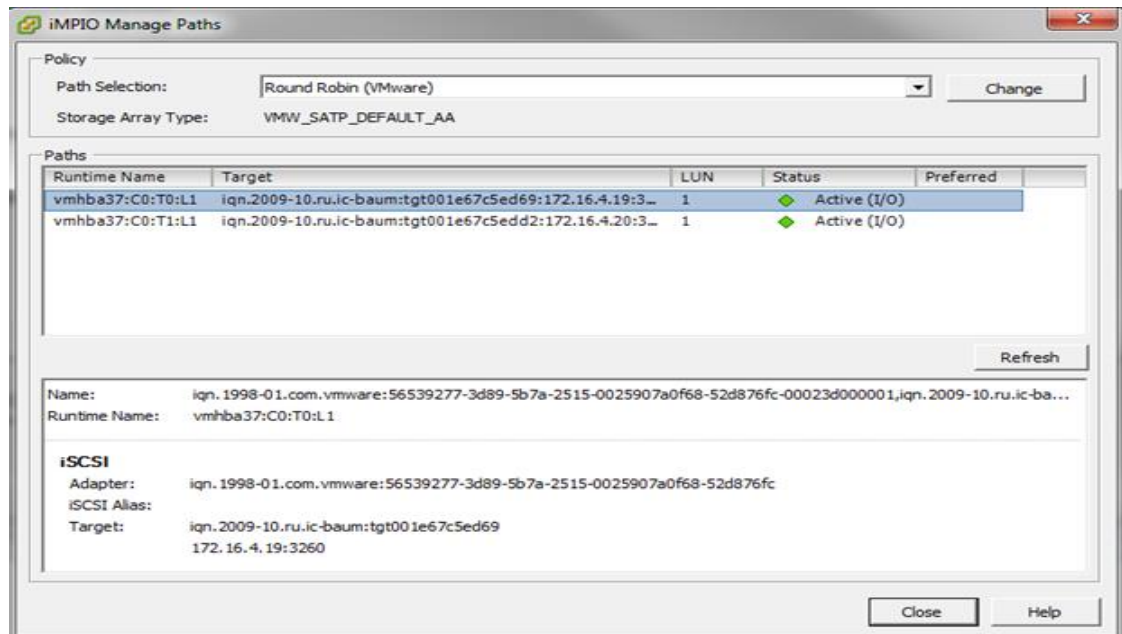


Рисунок 169. Manage Paths ESXI

7. Перейдите на вкладку «**Configuration**» - «**Storage**»
8. **Нажмите** на кнопку «**Add Storage**» (см. Рисунок 170)

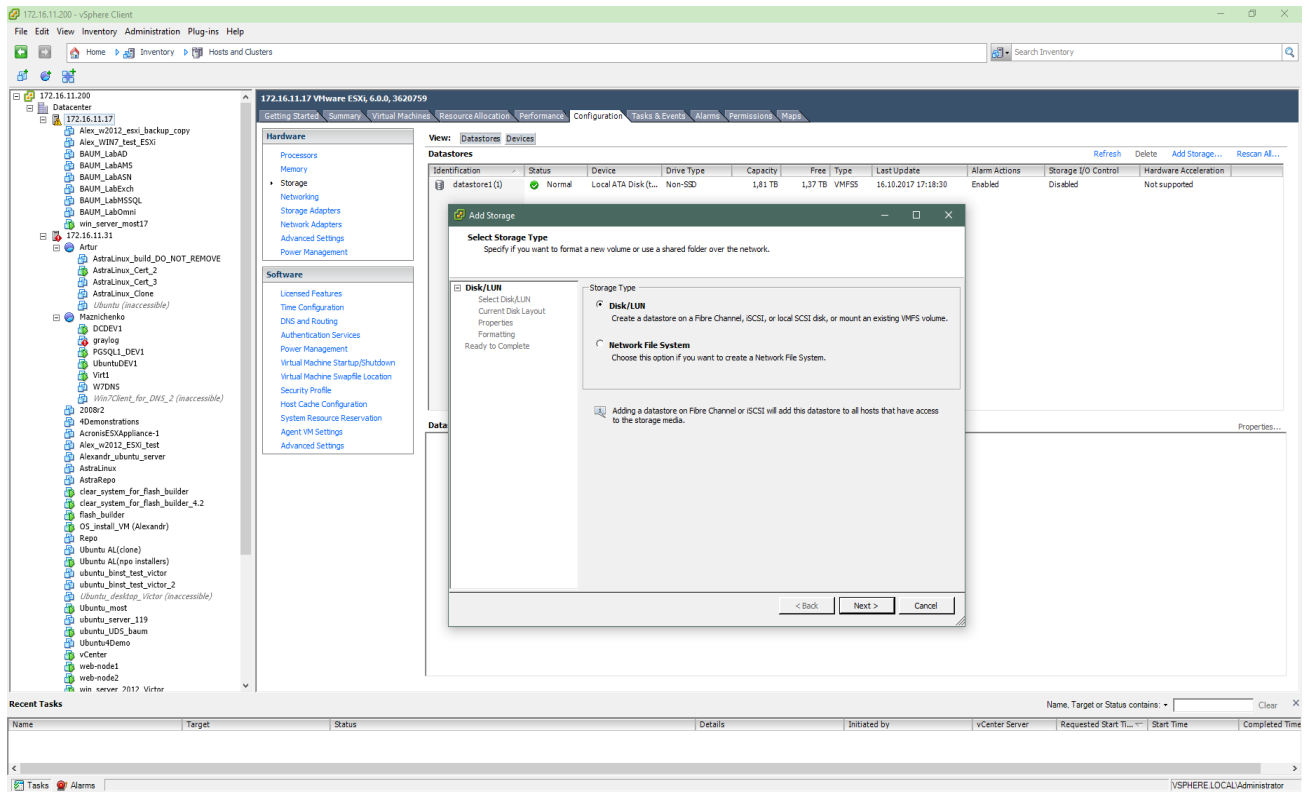


Рисунок 170. Добавление Storage

9. Выберите Storage тип **Disk/Lun**.

10. Нажмите на кнопку «Next» (см. Рисунок 171).

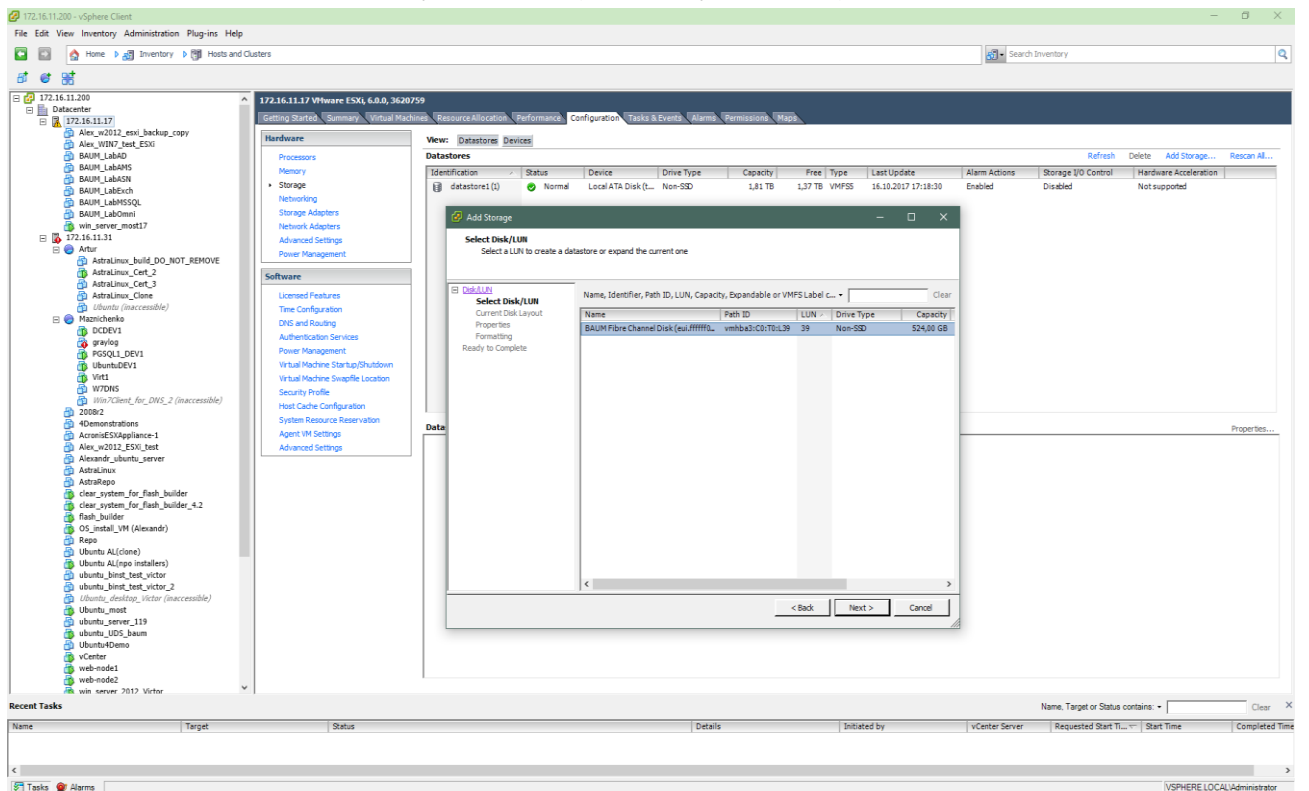


Рисунок 171. Выбор луна Storage

11. Выберите нужный лун.
12. Нажмите на кнопку «Next»
13. Введите имя луна в поле «Enter a datastore name» (см. Рисунок 172).

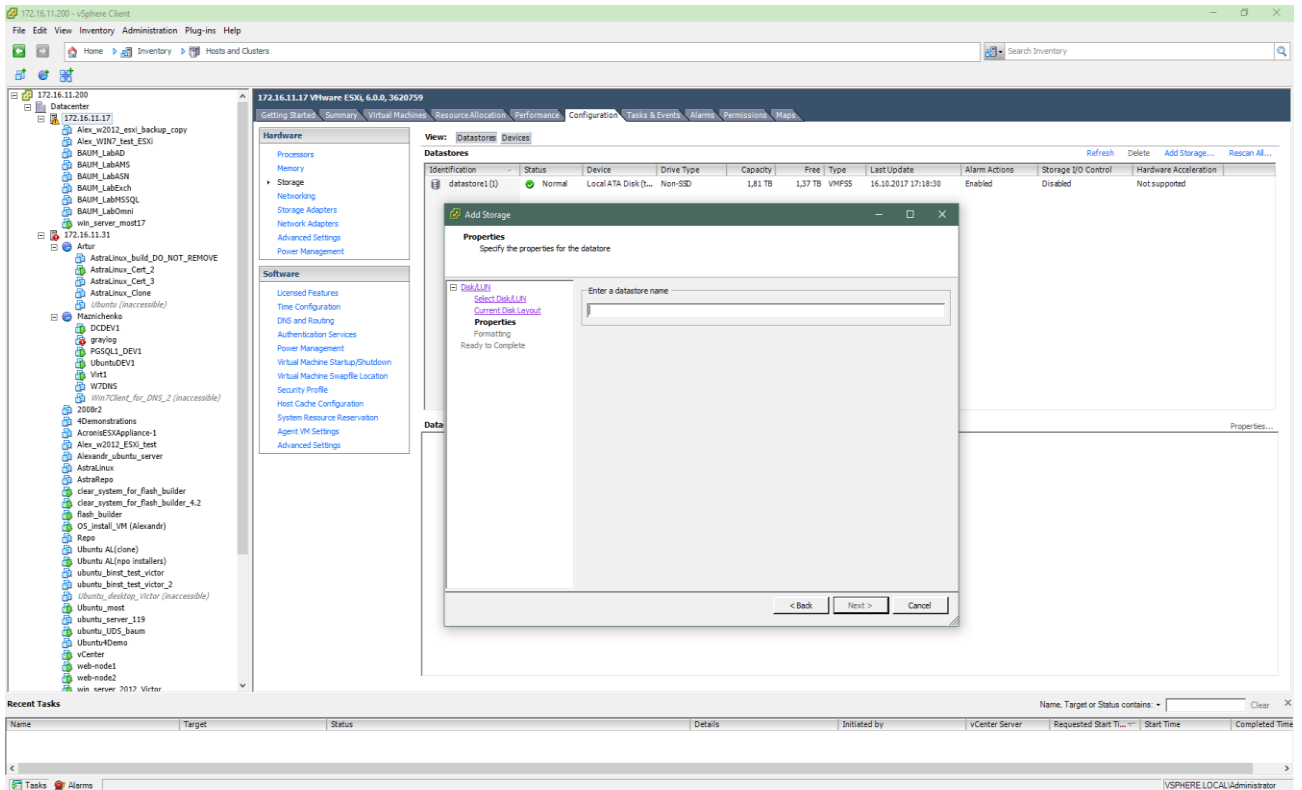


Рисунок 172. Имя Datastore

14. Выберите размер datastore.
15. Нажмите на кнопку «Next» (см. Рисунок 173).

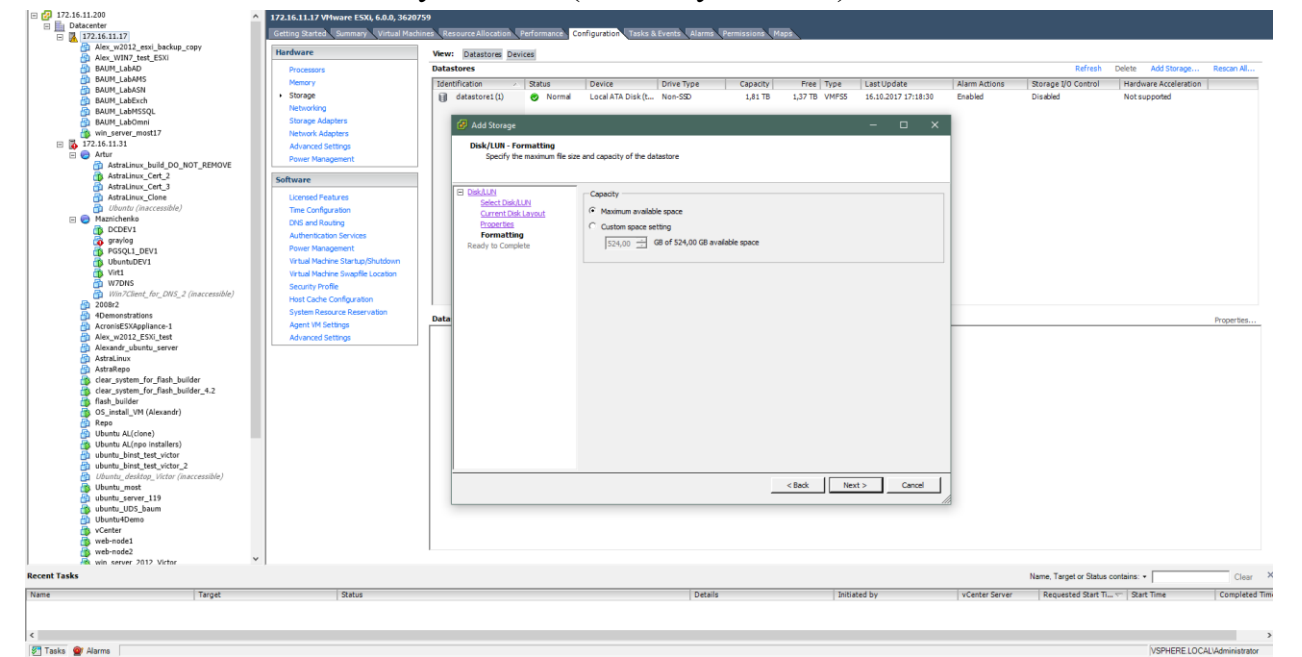


Рисунок 173. Размер Datastore

16. Подтвердите создание нажатием на кнопку «**Finish**»
В результате этих действий в списке появится новый datastore.

2.2.2.13. Логирование событий

2.2.2.13.1. Журналы событий

В процессе работы СХД события записываются в системный журнал. Часть этих событий, важные для администратора системы дублируются в оперативном журнале и журнале аудита безопасности.

В журнале аудита безопасности логируются только события входа и выхода пользователей. Вкладка меню «Аудит» представлена на рисунке ниже (Рисунок 174).

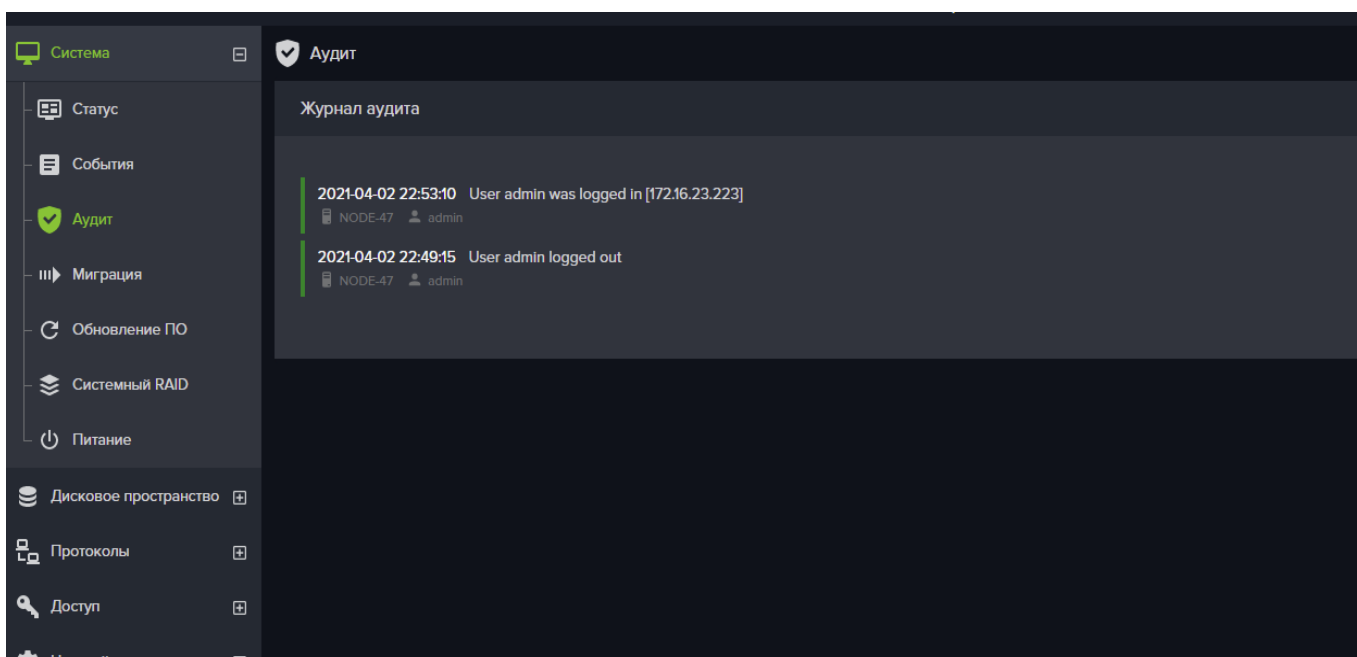


Рисунок 174. Вкладка меню аудита

На вкладке меню «События», в панели «Оперативный журнал», выводятся записи событий.

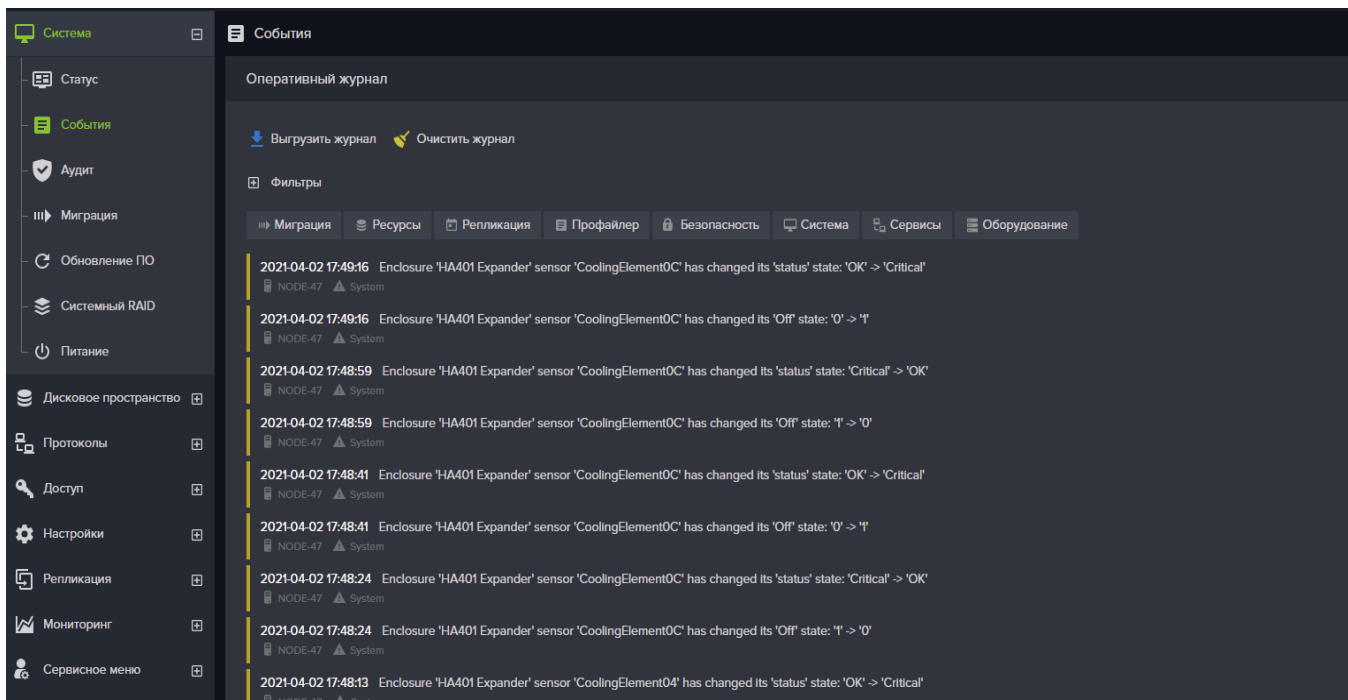


Рисунок 175. Вкладка меню оперативного журнала

Имеется возможность выборки событий за указанный интервал времени, а также использования фильтров для отображения следующих типов событий:

- авторизация;
- миграция;
- репликация;
- ресурсы;
- профайлер
- безопасность;
- система;
- сервисы;
- оборудование.

Фильтр работает по принципу включения или исключения показа выбираемых пунктов.

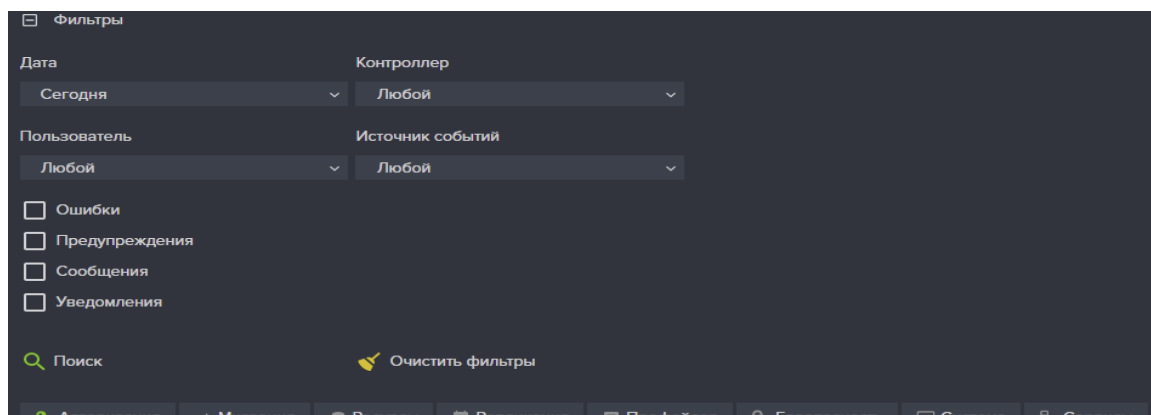


Рисунок 176. Развернутая панель фильтров

При этом фильтры, имеющие вид кнопок (в нижней части панели фильтров) применяются к списку уже найденных событий.

Также имеется возможность выгрузки журнала и его очистки. Для этой цели служат кнопки «**Выгрузить журнал**» и «**Очистить журнал**».

Записи системного журнала не отображаются в интерфейсе управления, вместо этого на вкладке «**Настройки**»-«**Системный журнал**», в панели «**Системный журнал**» имеется возможность создать архив за нужное количество недель и выборочно выгрузить журналы событий нажатием на нужные архивы в списке. Системные журналы создаются для каждого контроллера в отдельности из интерфейса этого контроллера.

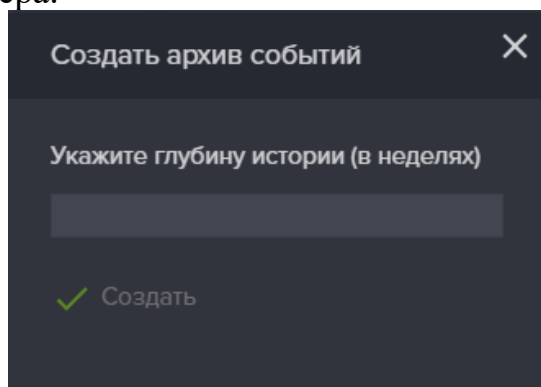


Рисунок 177. Окно создания архива системных событий

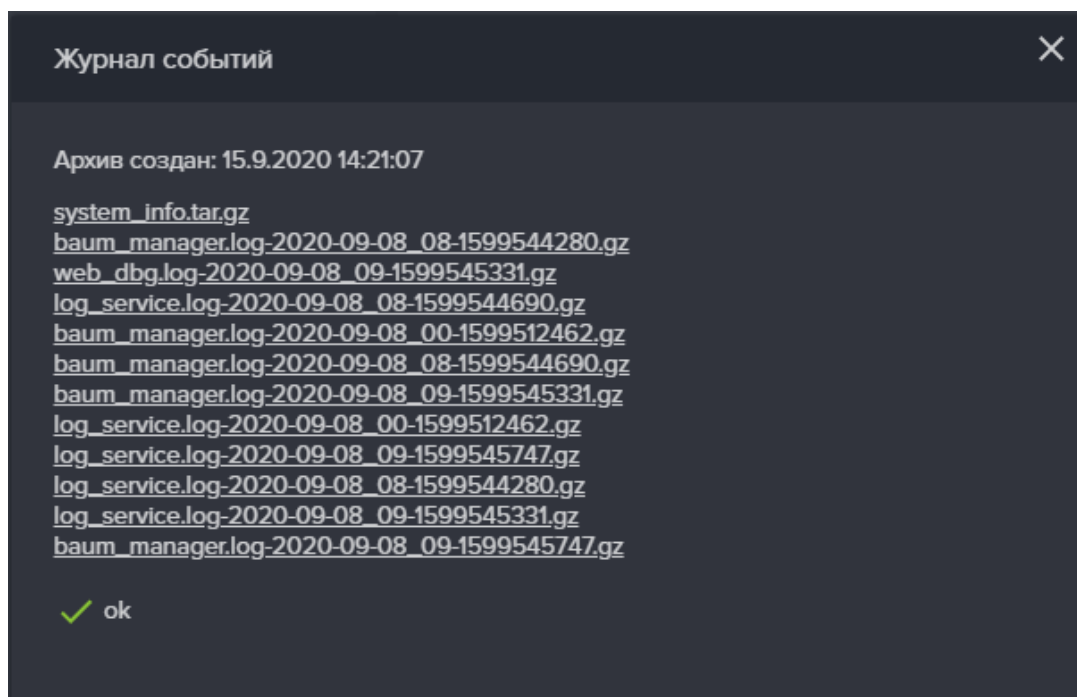


Рисунок 178. Окно с ссылками для выгрузки журналов событий

Настройка удаленного логирования выполняется на панели «**Удаленное логирование**». Перед включением удаленного логирования необходимо указать IP адрес и порт удаленного сервера, а также выбрать протокол.

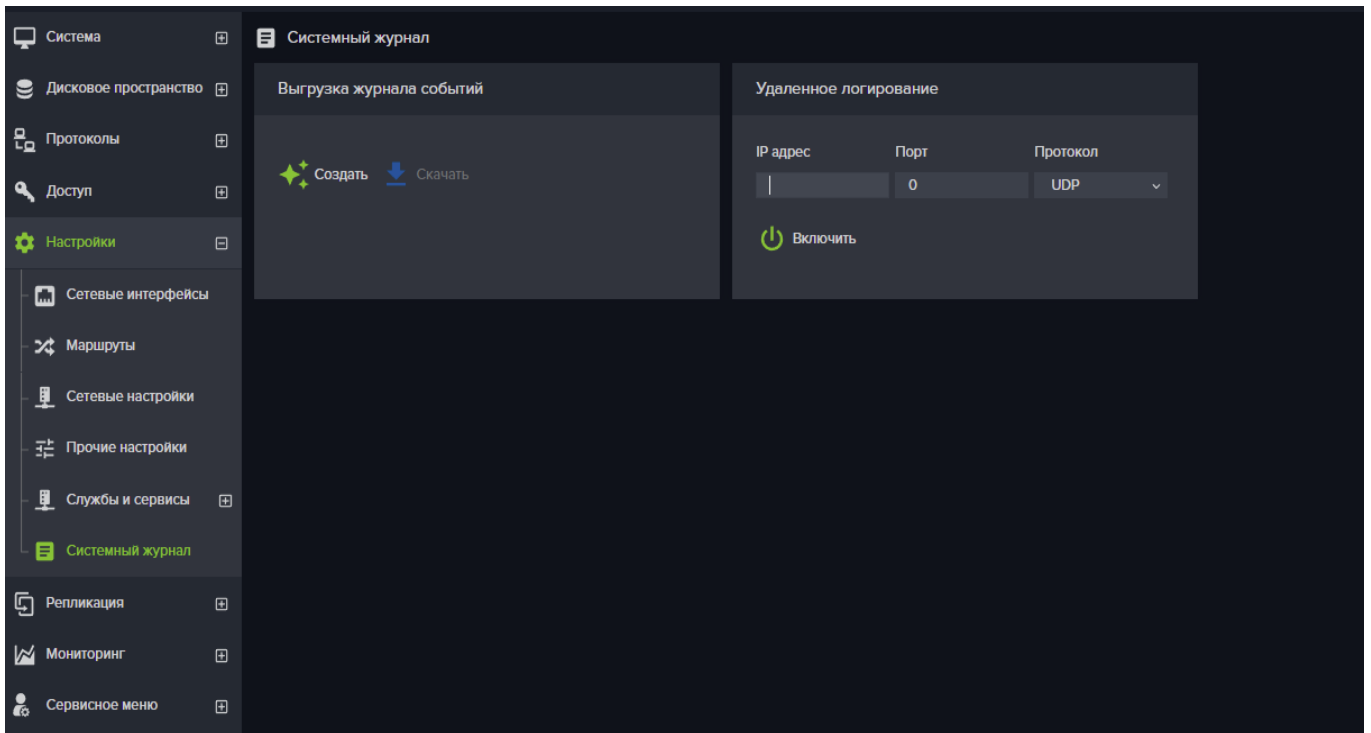


Рисунок 179. Системный журнал событий

2.2.2.13.2. Выгрузка системных записей

Инструментарий для выгрузки журналов событий расположен в области «Выгрузка журнала событий». Журналы создаются скачиваются с каждого контроллера в отдельности.

Для подготовки архива событий и их выгрузки выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Системный журнал**» выбранного контроллера (см. Рисунок 179);
2. В области «Выгрузка журнала событий» нажать на кнопку «**Создать**»;
3. После создания архива, появится его название размер;
4. Нажмите на кнопку «**Скачать**», для загрузки архива на жесткий диск вашего компьютера.

2.2.2.13.3. Удаленное логирование

Инструментарий настройки удаленного логирования расположен в области «Удаленное логирование».

Выбор протокол для удаленного логирования зависит от их поддержки хостом, которому СХД будет отправлять логи.

СХД поддерживает три протокола для отправки сообщений журнала: UDP, TCP и RELP.

Протокол UDP не обеспечивает надежную доставку, что может привести к потери некоторых сообщений журнала. Его использование не рекомендуется.

Протокол TCP обеспечивает большую надежность передачи сообщений журнала, выбор его предпочтительнее нежели протокола UDP.

Протокол RELP — Reliable Event Logging Protocol надёжнее TCP, не теряет сообщения при разрыве соединения. Решает проблему с многострочными сообщениями.

Для настройки удаленного логирования выполните следующие действия:

1. Перейдите на вкладку меню «**Настройки**» → «**Системный журнал**» (см. Рисунок 179);
2. Введите в области «Удаленное логирование» IP адрес, порт и выберите нужный протокол;
3. Нажмите на кнопку «**Включить**»;

После успешного включения службы данные логирования будут отправляться на удаленный сервер по указанному протоколу на указанный порт.

2.2.2.13.4. Функция мониторинга и управления через Радиощлюз

В состав серверной платформы СХД входит устройство сбора логов, мониторинга и управления параметрами серверной платформы СХД через радиоканал (далее Радиощлюз)

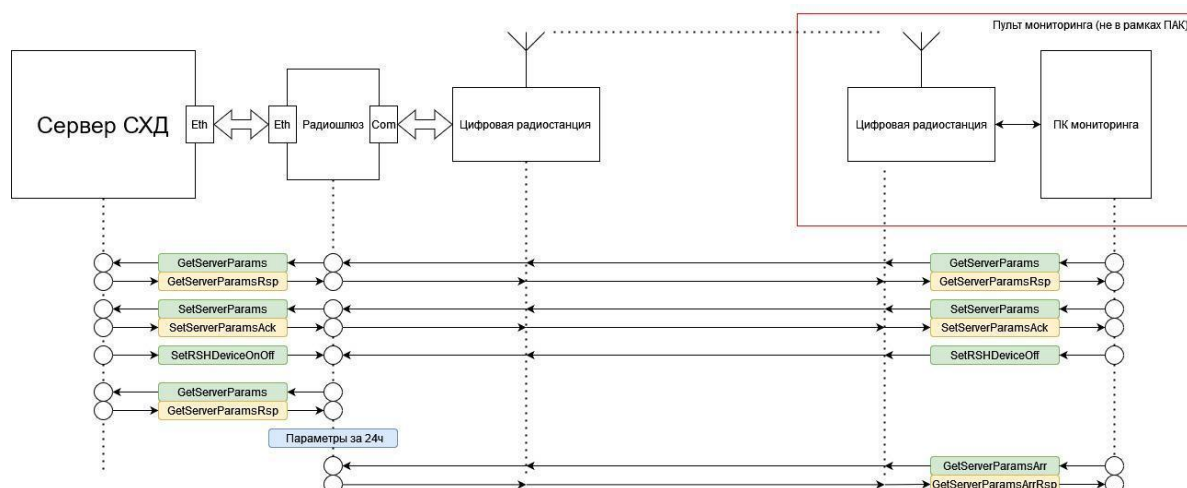


Рисунок 180 - Функциональная схема Радиощлюза в составе серверной платформы СХД

Основные свойства BAUM STORAGE AI v2 (в части применения устройства сбора логов, мониторинга и управления параметрами серверной платформы СХД через радиоканал):

1. Надёжность

- Передача данных по COM порту. Радиощлюз обеспечивает надёжную обработку данных по COM порту с использованием бита чёткости и 2 стоповых бита

- Передача данных по сетевому протоколу. Радиошлюз обеспечивает надежную обработку данных по сетевому протоколу, используя TCP протокол или UDP протокол с контролем получения пакетов

2. Функциональность

- Функция мониторинга основных параметров СХД и сбора логов. В процессе своей работы серверная платформа СХД записывает логи об операциях чтения, записи, удаления и обновления данных. Также работа серверной платформы СХД описывается рядом основных и вспомогательных параметров. Может возникнуть необходимость предоставления возможности получения этой информации в том числе по радиоканалу для последующего анализа, сохранения и учёта. Информация выдаётся с серверной платформы СХД по запросу. Удаленный пульт мониторинга по цифровому радиоканалу посылает в эфир команду на считывание параметров GetServerParams. Данная команда принимается цифровой радиостанцией на стороне серверной платформы и транслируется в Радиошлюз по COM порту. В радиошлюзе происходит валидация команды и упаковка её в сетевой пакет для последующей передачи в серверную платформу СХД. Серверная платформа СХД принимает команду и выдаёт обратно ответ GetServerParamRsp в течение 5 секунд об основных параметрах СХД (общая загрузка, наличие ошибок, производительность и тд) в виде сетевого пакета. Пакет, содержащий ответ СХД, принимается Радиошлюзом. Происходит подготовка полученных данных для передачи в цифровую радиостанцию по COM порту. Происходит отправка данных в цифровую радиостанцию, в которой данные восстанавливаются, проверяются и производится подготовка к отправке данных в цифровой радиоканал на удаленный пульт управления. Производится отправка данных по цифровому радиоканалу на удаленный пульт, где информация сохраняется для дальнейшего использования.

- Функция получения от цифровой радиостанции управляющих команд для установки значений основных параметров СХД для последующей их передачи в СХД. Может существовать необходимость удаленного задания определенных параметров серверной платформы СХД (уровень предельного заполнения, настройки RAID и тд). Для этого с удаленного радиопулльта подаётся управляющая команда с параметрами SetServerParams, которая передается по цифровому радиоканалу на Радиошлюз. Данная команда принимается цифровой радиостанцией на стороне серверной платформы и транслируется в Радиошлюз по COM порту. В радиошлюзе происходит валидация команды и упаковка её в сетевой пакет для последующей передачи в серверную платформу СХД. Серверная платформа СХД принимает команду, выполняет её и выдаёт обратно подтверждение получения команды SetServerParamAck в течение 5 секунд в виде сетевого пакета. Пакет, содержащий подтверждение об обработке СХД команды,

принимается Радиошлюзом. Происходит подготовка полученных данных для передачи в цифровую радиостанцию по СОМ порту. Происходит отправка данных в цифровую радиостанцию, в которой данные восстанавливаются, проверяются и производится подготовка к отправке данных в цифровой радиоканал на удаленный пульт управления. Производится отправка данных по цифровому радиоканалу на удаленный пульт, где происходит информирование оператора радиопульта.

- Функция получения от СХД управляющей команды на включение/выключение устройства. Оператор СХД имеет возможность включения и выключения Радиошлюза. Для этого со стороны серверной платформы СХД формируется команда SetRSHDeviceOnOff, упаковывается в сетевой пакет и передается в Радиошлюз. При получении команды SetRSHDeviceOnOff с параметром 1 Радиошлюз включается, при получении команды SetRSHDeviceOnOff с параметром 0 Радиошлюз выключается.

- Функция получения от радиопульта управляющей команды на выключение устройства. Оператор радиопульта имеет возможность выключения Радиошлюза. Для этого он может послать команду SetRSHDeviceOff. При получении команды SetRSHDeviceOff Радиошлюз выключается.

- Функция сохранения основных параметров СХД за последние 24 часа. Может существовать необходимость удаленного получения динамики изменения основных параметров СХД за последние 24 часа. Для этого Радиошлюз самостоятельно раз в 15 минут выдаёт команду GetServerParams по сетевому протоколу. Серверная платформа СХД получает эту команду и в течение 5 сек возвращает ответ GetServerParamRsp по сетевому протоколу. Ответ сохраняется в энергонезависимую память устройства. Массив ответов за 24 часа доступен для считывания со стороны удаленного радиопультата.

- Функция считывания основных параметров СХД за последние 24 часа. Для этого с радиопульта может быть подана команда GetServerParamsArr. В ответ Радиошлюз выдает параметры в сообщении GetServerParamsArrRsp

2.2.2.14. Работа системы в многоконтроллерном режиме

При объединении нескольких пар контроллеров в многоконтроллерную систему, управление ими осуществляется из единого веб интерфейса. Для этого каждой паре контроллеров задаются уникальные имена (имя кластера). Подключившись к общему для всех кластеров системы интерфейсу управления администратор системы может выбрать нужный кластер, и затем управлять его ресурсами. Для общего управления необходимо, чтобы управляющие интерфейсы всех контроллеров были в одной подсети и имели физическое соединение между собой. Схема объединения контроллеров представлена на рисунке ниже.

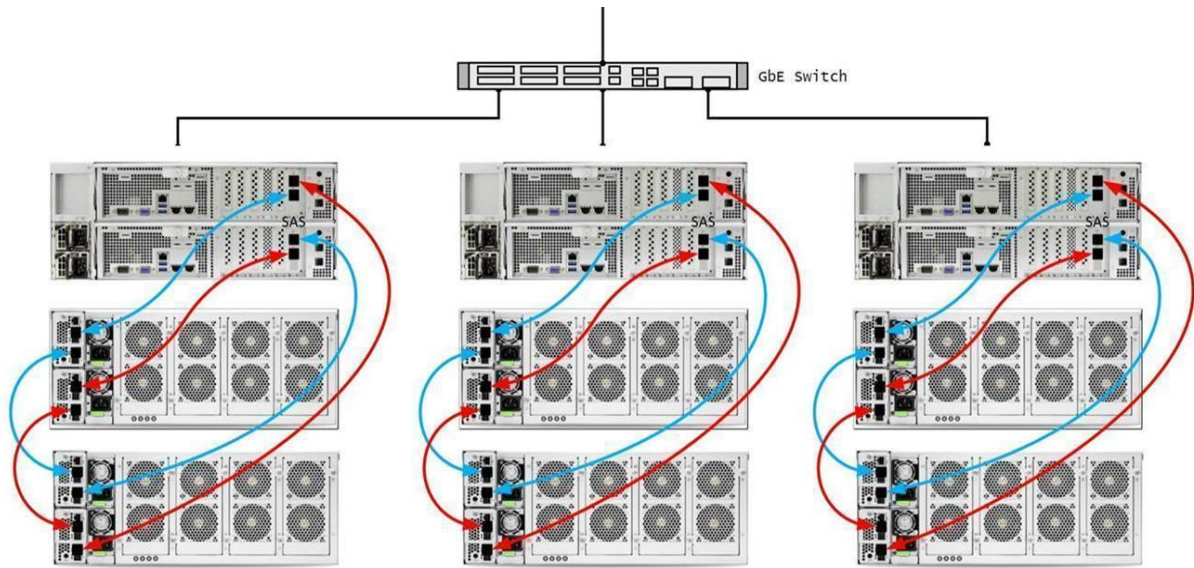


Рисунок 181. Соединение СХД для работы в многоконтроллерном режиме

2.3. Функционал BAUM STORAGE AI v2 для решения задач ИИ и машинного обучения

2.3.1. Регистрация и авторизация пользователя в Системе

Для начала работы с Системой необходимо пройти процедуру регистрации, для этого:

1. Перейдите на страницу регистрации:

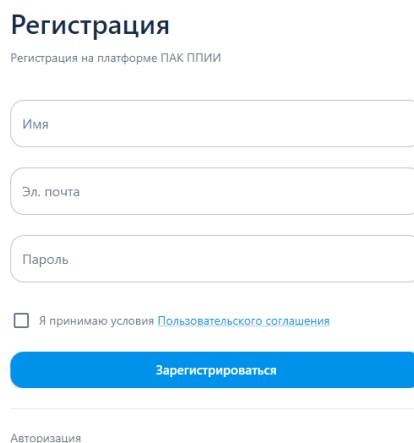


Рисунок 182 – Страница регистрации пользователя в Системе

По ссылке «Авторизация» доступен переход на страницу авторизации.

2. Заполните следующие поля:
 - Имя пользователя;
 - Адрес электронной почты;
 - Пароль для входа в Систему.
3. Ознакомьтесь с условиями пользовательского соглашения.

Регистрация возможна только при условии их принятия.

4. Нажмите кнопку «Зарегистрироваться».

После этого шага выполняется регистрация пользователя в Системе – в БД **MongoDB** создается новая запись с уникальным идентификатором пользователя.

5. Вы можете сохранить связку логин и пароль для автоматического заполнения при последующей авторизации на текущем устройстве.

В случае, если регистрация была пройдена ранее, при входе в систему осуществляется процедура **авторизации**. Для этого:

1. Перейдите на страницу авторизации:

Авторизация

Вход на платформу ПАК ППИИ

[Регистрация](#)
[Восстановить пароль](#)

Рисунок 183 – Страница авторизации пользователя в Системе

По ссылке «Регистрация» доступен переход на страницу регистрации.

По ссылке «Восстановить пароль» доступен переход на страницу восстановления пароля.

2. Введите следующую информацию:

- Адрес электронной почты;
- Пароль для входа в Систему.



3. Нажмите кнопку «Войти».

2.3.2. Личный кабинет пользователя

Личный кабинет – это персональная страница, доступ к которой есть только у одного пользователя (его владельца).

Как работать с личным кабинетом:

1. Перейдите на страницу личного кабинета одним из способов:

1.1. На панели инструментов нажмите на кнопку с инициалами/аватаркой пользователя (кнопка ) -> кнопку «Профиль» (кнопка ):

Большова Анна
Роль: Пользователь

 Профиль

 Настройки

Выйти

Рисунок 184 – Переход в личный кабинет

1.2. В левом верхнем углу Главного окна в разделе с информацией о пользователе нажмите кнопку с инициалами/аватаркой пользователя:

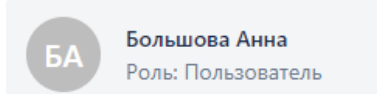


Рисунок 185 – Отображение данных пользователя

2. Откроется страница «Профиль» по умолчанию на вкладке

«Главная»:

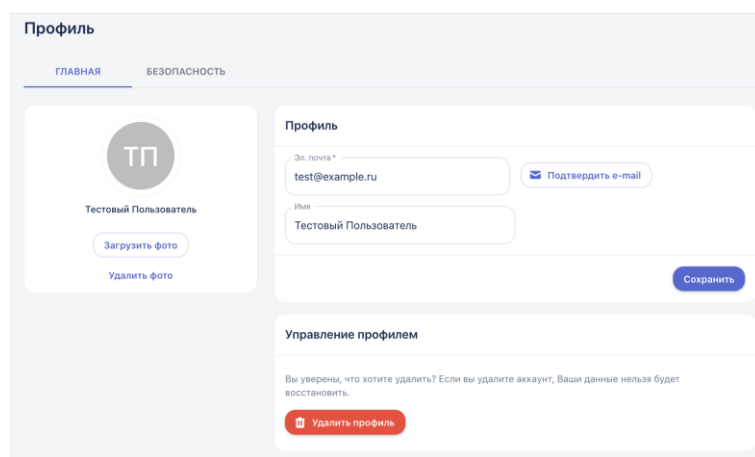


Рисунок 186 – Страница «Профиль»

Вкладка «Главная» состоит из:

- аватарки пользователя (по умолчанию это инициалы имени пользователя, загрузка фото не реализована в текущей версии);
- блока для ввода личной информации о пользователе – имени пользователя (логина), адреса электронной почты.
- блока «Управление профилем», в котором реализована возможность удаления своей учетной записи. Перед попыткой выполнить это действие, система выдает предупреждение о последствиях.

Чтобы изменить адрес электронной почты в поле «Электронная почта» укажите новый адрес и нажмите кнопку «Сохранить» (далее сохранение новых настроек в профиле предполагается по умолчанию). Подтверждение e-mail не является обязательным действием при смене адреса.

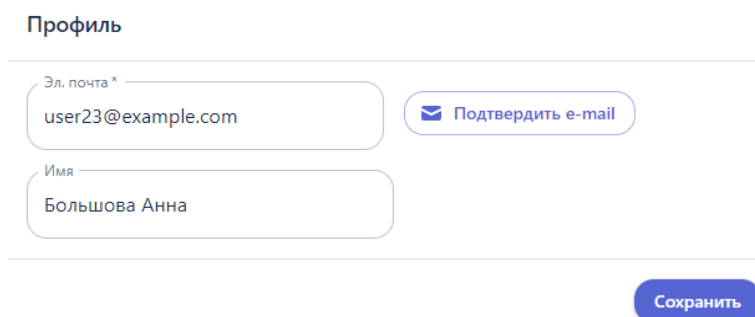


Рисунок 187 – Изменение адреса электронной почты в настройках профиля

Чтобы изменить имя пользователя, в поле «Имя» укажите свой новый логин. Тогда следующая авторизация будет выполняться уже с новым логином.

2.3.3. Интерфейс

2.3.3.1. Меню интерфейса

Пункты меню имеют древовидную структуру и представлены в виде вложенных папок:

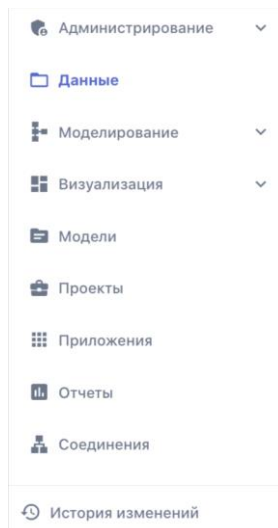







Рисунок 188 – Пункты меню BAUM STORAGE AI v2





Состав:

Таблица 2 – Описание пунктов меню

	Администрирование	<p>Данный раздел позволяет создавать ролевые модели пользователей - <i>группы</i>, согласно которым разделяются уровни доступа к разным модулям системы. Тут же осуществляется управление и назначение ролей всем пользователям системы.</p> <p>В разделе также реализована возможность настройки <i>каналов уведомлений</i> (это может быть email или telegram), которые могут быть использованы для получения автоматических оповещений от системы в случае выполнения заданных условий.</p>
	Данные	<p>Данный блок предназначен для загрузки в Систему <i>входных данных</i>, это могут быть файлы в различных форматах:</p> <ul style="list-style-type: none">● Таблицы в форматах csv, txt, xlsx, xls - табличные данные для создания интеллектуальной базы знаний, в том числе временные ряды. <i>Временной ряд</i> – это собранный в разные моменты времени статистический материал о значении каких-либо параметров исследуемого процесса. Возникают

		<p>временные ряды в результате измерения некоторого показателя. Это могут быть как показатели технических систем, так и показатели природных (погодные условия), социальных, экономических и других систем. Пример временного ряда – показания датчиков на производстве, анализ которых позволяет прогнозировать выход за критические отметки целевых переменных, принять меры и предотвратить возможную поломку оборудования или аварию.</p> <p>Уже загруженные в систему файлы с данными, преобразованные в датасеты, отображаются в формате таблицы (реализовано только для файлов с расширением <i>.csv</i>). Эти датасеты используются в качестве входных данных для обучения модели искусственного интеллекта.</p> <ul style="list-style-type: none"> ● Изображения в форматах: jpeg, jpg, png. ● Видео в форматах: avi, mp4. ● Текстовые файлы в форматах: txt, doc, docx. <p>Примеры задач для работы с текстом – классификация текста (например, определение авторства), распознавание и оцифровка рукописного текста, чтение и анализ новостного фона, определение тенденций в науке по научным статьям «Скопус», и т.д.</p>
	<p>Моделирование</p>	<p>Сердце программы, которое позволяет описывать бизнес-процессы и выполнять целевые действия.</p> <ul style="list-style-type: none"> ● Рабочая область. Для описания бизнес-процессов заказчика на рабочей области создаются <i>элементные блок-схемы</i>, которые позволяют выстраивать цепочки, взаимосвязи, условия и т.д. Блок-схема отвечает на вопрос «Что делает процесс». ● Сохраненные рабочие области. Рабочие области с ранее созданными блок-схемами. Есть возможность в любое время вернуться и продолжить работу над блок-схемой.

	Визуализация	<ul style="list-style-type: none"> ● Дашборды. Динамически настраиваемые дашборды. Содержат загруженные и преобразованные массивы данных, представленные в виде таблиц, графиков, гистограмм. ● Сохранённые дашборды. Дашборды, созданные пользователем или группой пользователей.
	Модели	<p><i>Модель на основе машинного обучения</i> – это абстракция, которая обучена распознаванию определенного типа закономерностей. Хранится в виде файла.</p> <p>Для обучения модели нужны исторические данные (обучающая выборка) и значение целевой переменной (то, что прогнозируем), которое соответствует заданным историческим данным. С помощью алгоритмов машинного обучения модель наблюдает и находит зависимости между данными и целевой переменной. Эти зависимости используются моделью для нового набора данных (тестовой выборки), чтобы прогнозировать целевую переменную, которая неизвестна.</p> <p>Все модели разделяются на обучение с учителем и без учителя. <i>Обучение с учителем</i> подразделяется на две подкатегории: регрессия и классификация.</p> <p>В <i>регрессионных моделях</i> вывод является непрерывным. Наиболее распространенные типы регрессионных моделей – линейная регрессия, деревья решений, случайный лес, нейронная сеть.</p> <p>В <i>классификационных моделях</i> вывод является дискретным. Наиболее распространенные типы классификационных моделей – логистическая регрессия, метод опорных векторов.</p> <p>В отличие от обучения с учителем, <i>обучение без учителя</i> используется для того, чтобы сделать выводы из входных данных без отсылок на отмеченные результаты. Два основных метода, используемых в обучении без учителя, включают <i>кластеризацию</i> и <i>снижение размерности</i>.</p>

		Сохранённые модели – обученные модели, которые используются в моделировании бизнес-процессов с данными в режиме реального времени.
	Проекты	Сущность «Проект» объединяет в себе: загруженные в Систему файлы, созданные рабочие области, обученные модели ИИ, визуализацию результатов на дашбордах и сформированные отчеты. Так Система позволяет консолидировать всю информацию по проекту в одном разделе для систематизации и удобного доступа пользователей.
	Приложения	Приложение содержит «упакованную» обученную модель ИИ – без необходимости разработки отдельного кода для создания приложения с моделью. Такое приложение можно скачать и развернуть за пределами программы, интегрировать с внешними системами, настроить получение входных данных и выполнить прогнозы.
	Отчеты	Подготовленные по установленной форме отчеты. Могут храниться результаты ИИ исследований, результаты спроектированного бизнес-процесса.
	Соединения	Раздел «Соединения» предназначен для настройки подключения к внешним источникам данных с целью получения данных из других систем. В этом разделе пользователь может создавать новые и просматривать уже созданные коннекторы. <i>Коннектор</i> – это сущность, которая объединяют в себе источник подключения и запрос на получение данных из него.

2.3.3.2. Окно построения модели ИИ

На примере страницы BAUM STORAGE AI v2, на которой строится конструктор модели искусственного интеллекта, рассматривается, из каких частей состоит интерфейс (см. рис. 189)

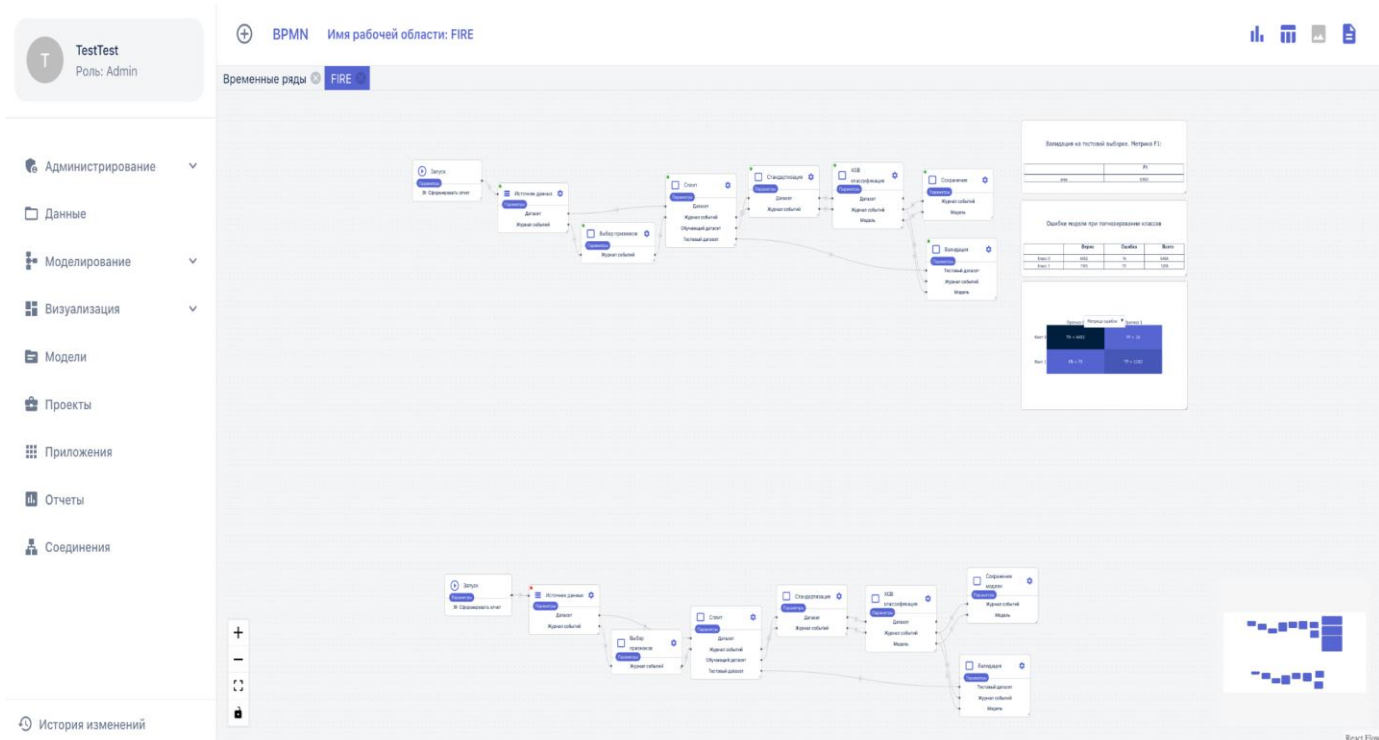

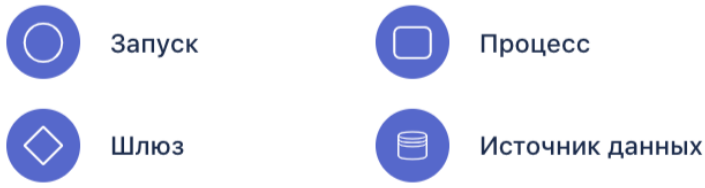






Рисунок 189 – Интерфейс BAUM STORAGE AI v2

На верхней панели инструментов рабочей области представлены следующие кнопки:

Таблица 4 – Кнопки панели инструментов рабочей области

	<p>Создание рабочей области</p>	<p>При нажатии на кнопку открывается форма, в которой нужно указать название рабочей области:</p> <div style="text-align: center;"> <p>Введите имя рабочей области</p> <input data-bbox="829 1344 1165 1400" type="text" value="Имя рабочей области"/> <input data-bbox="829 1422 1165 1478" type="button" value="Создать"/> </div> <p>Одна рабочая область может содержать несколько блок-схем. Реализована возможность запуска как отдельной блок-схемы, так и всех блок-схем на рабочей области.</p> <p>Рабочую область можно добавить в проект, таким образом организовав доступ к рабочей области для всех пользователей этого проекта.</p>
---	--	--

BPMN	<p>Добавить элемент</p>	<p>При нажатии на кнопку « BPMN » открывается блок, содержащий графические элементы нотации BPMN 2.0:</p> <div data-bbox="630 309 1375 539" style="border: 1px solid #ccc; padding: 10px; margin: 10px auto; width: fit-content;">  <p style="text-align: center;"> Запуск Процесс Шлюз Источник данных </p> </div> <p>Чтобы добавить элемент на блок-схему достаточно нажать на кнопку с элементом.</p> <p>Типы элементов нотации BPMN 2.0:</p> <table border="1" data-bbox="528 689 1469 1646"> <tr> <td data-bbox="528 689 730 824">Запуск</td> <td data-bbox="730 689 1469 824"><i>Иницирующее событие</i> – главный элемент, обозначающий начало блок-схемы.</td> </tr> <tr> <td data-bbox="528 824 730 958">Процесс</td> <td data-bbox="730 824 1469 958">Действие, выполняемое в ходе бизнес-процесса. Является основным элементом.</td> </tr> <tr> <td data-bbox="528 958 730 1346">Шлюз</td> <td data-bbox="730 958 1469 1346">Действие предназначено для разветвления алгоритма по веткам – прохождение сценария по каждой из веток выполняется при определенных условиях. Когда выполняется раздвоение потока операций, прописывается одно условие, и при его выполнении сценарий проходит по одной из веток.</td> </tr> <tr> <td data-bbox="528 1346 730 1646">Источник данных</td> <td data-bbox="730 1346 1469 1646"><i>Объект данных</i> – это информационный объект (датасет, файл, модель и т.д.), который обрабатывается и передается в ходе выполнения бизнес-процесса. Действие предназначено для выбора уже загруженной и/или сохраненной в Системе сущности.</td> </tr> </table>	Запуск	<i>Иницирующее событие</i> – главный элемент, обозначающий начало блок-схемы.	Процесс	Действие, выполняемое в ходе бизнес-процесса. Является основным элементом.	Шлюз	Действие предназначено для разветвления алгоритма по веткам – прохождение сценария по каждой из веток выполняется при определенных условиях. Когда выполняется раздвоение потока операций, прописывается одно условие, и при его выполнении сценарий проходит по одной из веток.	Источник данных	<i>Объект данных</i> – это информационный объект (датасет, файл, модель и т.д.), который обрабатывается и передается в ходе выполнения бизнес-процесса. Действие предназначено для выбора уже загруженной и/или сохраненной в Системе сущности.
Запуск	<i>Иницирующее событие</i> – главный элемент, обозначающий начало блок-схемы.									
Процесс	Действие, выполняемое в ходе бизнес-процесса. Является основным элементом.									
Шлюз	Действие предназначено для разветвления алгоритма по веткам – прохождение сценария по каждой из веток выполняется при определенных условиях. Когда выполняется раздвоение потока операций, прописывается одно условие, и при его выполнении сценарий проходит по одной из веток.									
Источник данных	<i>Объект данных</i> – это информационный объект (датасет, файл, модель и т.д.), который обрабатывается и передается в ходе выполнения бизнес-процесса. Действие предназначено для выбора уже загруженной и/или сохраненной в Системе сущности.									
	<p>Графики</p>	<p>После успешной отработки пайплайна здесь будет отображаться список доступных для визуализации графиков.</p>								
	<p>Таблицы</p>	<p>После успешной отработки пайплайна здесь будет отображаться список доступных для визуализации таблиц.</p>								

	Изображения	После успешной отработки пайплайна здесь будет отображаться список доступных для визуализации изображений.
	Описание	Отображается описание обученной модели ИИ.

- **Рабочая область.** Предназначена для построения графической бизнес-модели. С помощью блок-схемы выполняется настройка последовательных операций обработки данных, обучение моделей, выстраивается «pipeline» (цепочка процессов преобразования).

В правом нижнем углу рабочей области отображается миниатюра бизнес-процесса:



Рисунок 190 – Миниатюра моделируемого бизнес-процесса

- **Карточка элемента блок-схемы.** В карточке прописываются условия обработки данных – откуда берутся данные, какие операции над ними выполняются. Для перехода в карточку нажмите кнопку «Настройки» на элементе. Чтобы скрыть карточку элемента, используйте кнопку «Свернуть».

Карточка содержит *программный код* с основными блоками:

- входные данные (источник входных данных);
- преобразование данных;
- выходные данные – куда и в каком виде передаются

преобразованные данные.

- **Вкладки с рабочими областями** - вы можете работать с несколькими рабочими областями одновременно, для удобного перехода между ними используйте вкладки.

2.3.3.3. Настройка внешнего вида интерфейса

Чтобы изменить настройки интерфейса, кликните на свою аватарку и перейдите в пункт меню «Настройки». Откроется окно, в котором можно изменить следующее:

- **Тема** – светлая или темная.
- **Адаптивные размеры шрифтов** – адаптация размеров шрифтов программы для работы с ней в мобильных устройствах.
- **Компактный вид** - фиксированная ширина на некоторых экранах.

- **Закругленные углы.** Если выбрать настройку, то окна в программе будут иметь закругленные углы. Иначе окна будут иметь прямые углы.

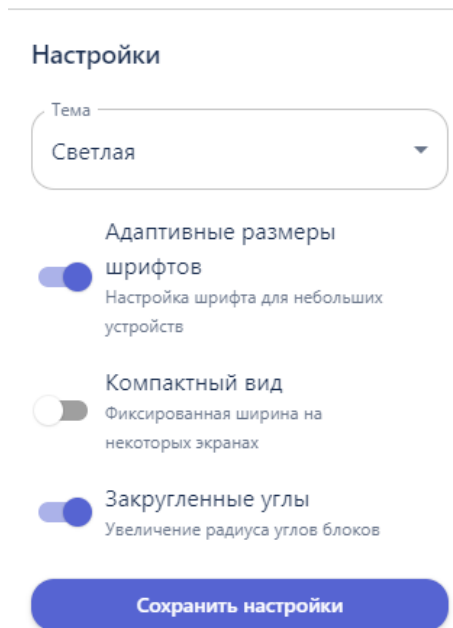


Рисунок 191 – Настройки внешнего вида интерфейса программы

Чтобы применить выбранные настройки к интерфейсу нажмите кнопку «Сохранить настройки».

2.3.3.4. Встроенные функции

Каждая функция относится к одному из модулей Системы: *модулю препроцессинга входных данных, модулю машинного обучения, модулю нейронных сетей/глубокого обучения, модулю анализа данных или модулю визуализации.*

Описание всех функций представлено в Приложении 1 настоящего документа.

За каждым типом блока закреплен определенный набор функций:

2.3.3.4.1. Функции элемента «Источник данных»

Таблица 5 – Набор функций элемента «Источник данных»

Группа	Функция
Загрузка данных	1. Загрузка изображений для object detection
	2. Загрузка табличных данных из коннектора
	3. Загрузка табличных данных

Группа	Функция
	4. Загрузка изображений для классификации
	5. Загрузка текстовых файлов для классификации
	6. Загрузка текстовых файлов для кластеризации
	7. Загрузка модели
	8. Загрузка графа
Spark	9. Загрузка табличных данных из файла CSV (Spark)
	10. Загрузка табличных данных из папки CSV (Spark)
	11. Загрузка модели
	12. Загрузка табличных данных из коннектора (Spark)

2.3.3.4.2. Функции элемента «Процесс»

Таблица 6 – Набор функций элемента «Процесс»

Группа	Подгруппа	Функция
Анализ данных	Загрузка данных	1. Преобразование данных во временной ряд
	Препроцессинг	2. Стабилизация дисперсии
		3. Стандартизация
		4. Дифференцирование временного ряда
		5. One-Hot Encoding
		6. Создание признаков для временного ряда
		7. Кодирование целевого признака

Группа	Подгруппа	Функция
		8. Порядковое кодирование категориальных признаков
	Тесты на нормальность распределения	9. Коэффициент асимметрии Skewness
	Тесты на стационарность временного ряда	10. Тест Дики-Фуллера
	-	11. Выбор признаков и целевых признаков
		12. Матрица корреляции
		13. Косинусное расстояние
		14. Поиск пропущенных значений
		15. Анализ временных рядов
		16. Визуализация Real Time
		17. Запись в датасет логирования
Машинное обучение	Классификация	18. Логистическая регрессия
		19. Модель XGBClassifier
		20. Дерево решений для классификации
		21. Случайный лес для классификации
		22. Categorical Naive Bayes
		23. Multinomial Naive Bayes

Группа	Подгруппа	Функция
		24. Complement Naive Bayes
		25. Gaussian Naive Bayes
		26. Bernoulli Naive Bayes
	Обучение учителя без	27. Кластеризация DBSCAN
		28. Метод локтя K-Means
		29. Кластеризация K-Means
		30. Агломеративная кластеризация иерархическая
		31. Изоляционный лес
	Регрессия	32. Линейная регрессия
		33. Полиномиальная регрессия
		34. Дерево решений для регрессии
		35. Случайный лес для регрессии
		36. Метод опорных векторов для регрессии
		37. Байесовская гребневая регрессия
38. Метод k-ближайших соседей для регрессии		
Авторегрессия	39. ARIMA/SARIMAX	
Работа с текстами	40. Автореферирование	
-	41. Валидация модели	

Группа	Подгруппа	Функция
		42. Прогноз модели
		43. Разделение датасета на обучающую и тестовую выборки
Глубокое обучение	Классификация	44. Классификация изображений (Валидация)
		45. Классификация (табличные данные)
	Регрессия	46. Регрессия (табличные данные)
	Обнаружение объектов	47. YOLOv5
	-	48. Валидация модели классификации изображений
Предобработка данных	-	49. Заполнение пропусков
		50. Сглаживание временного ряда
		51. Срез временного ряда по индексу
		52. Фильтрация текстового шума
		53. Лемматизация текста
		54. Векторизация текста
Управление моделями	Классификация	55. Сохранение модели классификации изображений
	Обнаружение объектов	56. Сохранение модели YoloV5

Группа	Подгруппа	Функция
		57. Сохранение модели Spark
		58. Сохранение модели
Отправка уведомлений	-	59. Отправка уведомлений
Spark	-	60. Выбор признаков и целевых признаков
		61. Разделение датасета на обучающую и тестовую выборки
		62. Валидация модели
		63. Сохранение датасета Spark в CSV
		64. Прогноз модели
		65. Косинусное расстояние
	Преобработка	66. Порядковое кодирование признаков
		67. Нормализация признаков
	Классификация	68. Модель градиентного бустинга Spark для бинарной классификации
	Кластеризация	69. Кластеризация Spark DBSCAN

“–” в таблице означает, что у функции нет подгруппы, и она напрямую относится к группе функций.

2.3.4. Загрузка данных в систему

Сразу после авторизации в Системе открывается её начальная страница – раздел «Данные».

Раздел «Данные» имеет внешний вид аналогичный проводнику файлов в операционной системе компьютера. Пользователи имеют возможность

создавать папки, загружать файлы и делать структуры, удобные для личного использования. По умолчанию для нового пользователя раздел Данные выглядит следующим образом:

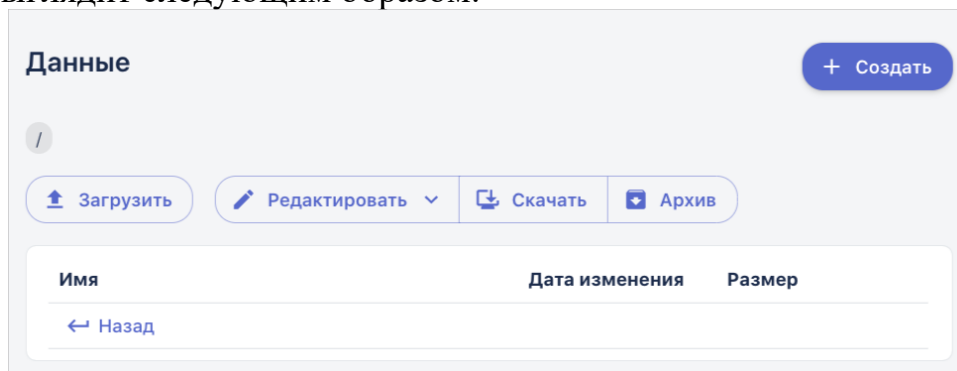


Рисунок 192 – Вид раздела Данные

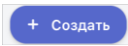
Основные кнопки для работы с разделом:

1. «Создать» - позволяет создавать папки и разметки.
2. «Загрузить» - осуществление непосредственно загрузки файлов в систему
3. «Редактировать» - не работает в данной версии системы
4. «Скачать» - не работает в данной версии системы
5. «Архив» - не работает в данной версии системы

В данном разделе вы можете создавать папки, добавлять в них файлы, создавая удобную и понятную структуру. Также в разделе осуществляется создание разметки для видео и изображений, и добавление папок классификации для дальнейшего использования в обучении искусственного интеллекта.

Вы можете загружать в систему файлы разных форматов и типов. Данные можно загружать напрямую в основную директорию, или создавая новые папки внутри.

2.3.4.1. Создание новой папки

Для создания новой папки нажмите на кнопку «Создать»  и в открывшемся окне в поле «Тип» выберите «Категория» (так в системе называются папки), а в поле «Название» введите название будущей папки, например, «Табличные данные»:

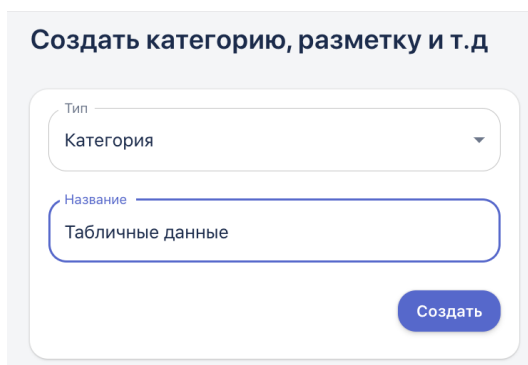


Рисунок 193 – Создание новой папки в разделе Данные

После этого папка появится в разделе Данные:

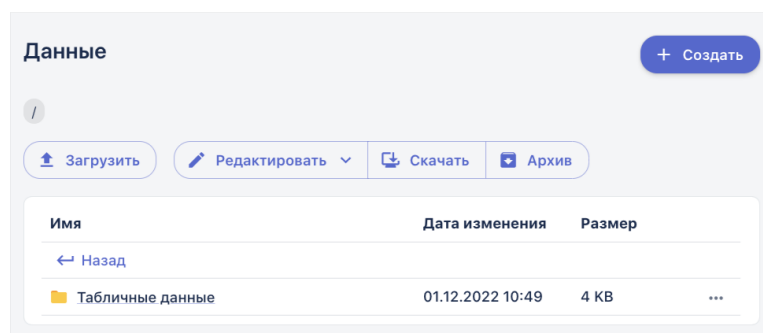


Рисунок 194 – Новая папка в разделе Данные

Обратите внимание, что папка будет добавлена в том разделе, из которого вы нажали кнопку «Создать». Т.е. далее вы можете перейти в папку «Табличные данные» и создать внутри еще одну категорию-папку.

2.3.4.2. Загрузка файлов

Для того чтобы загрузить файлы в папку, кликните на неё и перейдите в ее содержимое. Далее нажмите кнопку «Загрузить»:

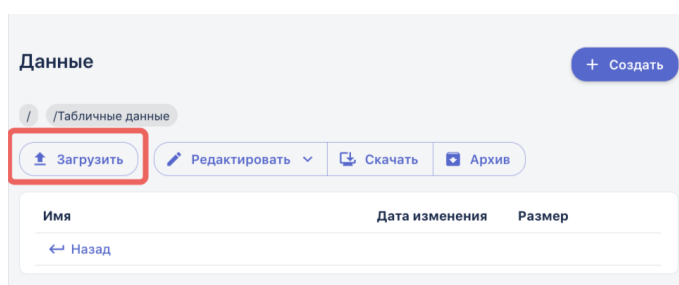


Рисунок 195 – Загрузка файлов в папку

В открывшемся окне левой кнопкой мыши нажмите на ссылку выбора файла. Указать путь к файлу для загрузки на вашем ПК. Второй вариант – перенести файлы с локального компьютера в этот раздел по технологии «drag n drop».

Выбранные файлы отобразятся в нижней части окна загрузки:

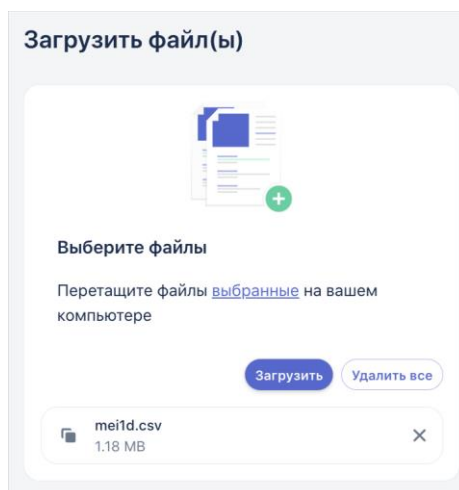


Рисунок 196 – Отображение выбранного файла

При необходимости выбранные файлы можно удалить по одному, нажав на крестик, или все вместе, нажав кнопку «Удалить все».

Для того чтобы загрузить выбранные файлы, нажмите на кнопку «Загрузить». Файлы отобразятся в папке:

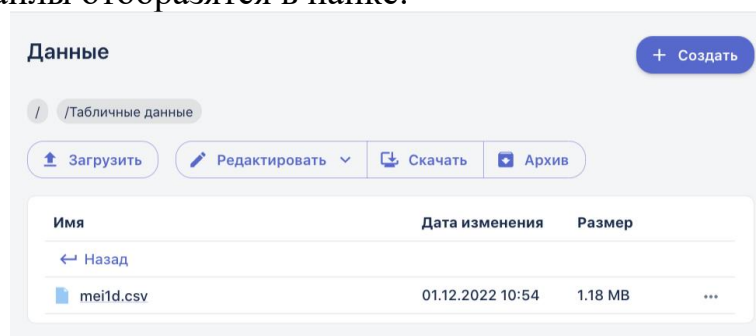


Рисунок 197 – Загруженный файл в разделе Данные

2.3.4.3. Предпросмотр данных

В системе реализована функция предпросмотра загруженных файлов следующих форматов:

- табличные данные формата .csv
- изображения в форматах: jpeg, jpg, png.
- видео в форматах: avi, mp4.

Если кликнуть на файл с изображением или видео - его предпросмотр начнется прямо в окне, чтобы выйти из режима предпросмотра - просто кликните в любую область экрана за пределами медиафайла.

При предпросмотре табличных данных в формате csv, в нижней части экрана отобразится окно с тремя вкладками, содержащими описательно статистический анализ.

На вкладке «Подробнее» (показано на примере файла с временным рядом) пользователь имеет возможность увидеть:

- размер датасета (количество строк, столбцов)
- гистограммы распределения каждого признака (числового)



Рисунок 198– Вкладка «Подробнее» окна отображения данных о датасете

На вкладке «Компактно» отображается состав строк и столбцов файлов:

area	T	P	U	Ff	Td	RRR	DD_Ветер_с_вос	DD_Ветер_с_вос	DD_Ветер_с_вос	DD_Ветер_с_зап	DD_Ветер_с_зап
0	-0.59	763.32	84.87	4.37	-2.92	0.13	0	0	0	0.37	0
0	-3	770.51	74.12	1.12	-7.15	0	0.12	0	0	0.25	0.12
0	-2.95	769.77	89.62	2.62	-4.52	0	0.87	0	0.12	0	0
0	1.85	766.77	85.37	3.5	-0.45	0	0.75	0	0.25	0	0
0	1.48	762.78	83.5	3	-1.12	0.25	0.75	0	0.12	0	0
0	0.8	772.26	89.75	1.5	-0.82	0.51	0	0	0	0.25	0
0	-1.32	776.48	84.62	2.25	-3.7	0	0.12	0.37	0	0	0

Рисунок 199 – Вкладка «Компактно» окна отображения данных о датасете

На вкладке «Столбцы» для каждого признака отображается:

- количество пропусков (Missing) в абсолютном и в процентном выражении, под пропуском понимается пустая ячейка в таблице
- среднее значение (Mean)
- стандартное отклонение (Std. Deviation)
- квантили (quantiles):



Рисунок 200– Вкладка «Столбцы» окна отображения данных о датасете

2.3.4.4. Взаимодействие с данными

Загруженный файл или созданную папку можно скачать, скачать архивом, переименовать или удалить. Для этого нажмите на три точки в правой части раздела и выберите соответствующую кнопку:

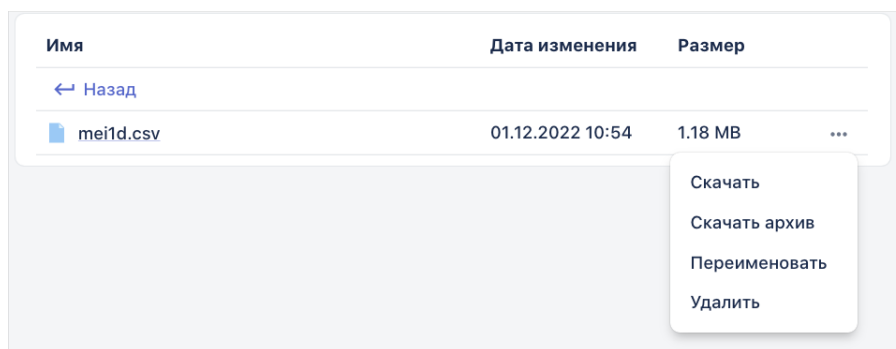



Рисунок 201 – Удаление или скачивание файла из раздела Данные
Вы можете удалять файлы по отдельности или сразу целую папку.

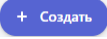
Для навигации между файлами и папками можно воспользоваться кнопкой «Назад»:  .

2.3.4.5. Создание датасетов и загрузка данных для решения задач классификации

Задача классификации относится к классу задач «обучение с учителем», которое предполагает наличие набора *размеченных* данных для тренировки модели. Наличие размеченного датасета означает, что каждому примеру в обучающем наборе соответствует ответ, который должен получить алгоритм. С помощью BAUM STORAGE AI v2 можно решать задачи классификации следующих типов данных:

- изображений
- текстов
- табличных данных

В задачах классификации алгоритм предсказывает *дискретные значения*, соответствующие номерам классов, к которым принадлежат объекты. В обучающем датасете каждый объект будет иметь соответствующую метку. Пользователь должен подготовить данные перед подачей их в алгоритм машинного обучения. В данном разделе рассмотрим, как создаются группы и классы в разделе «Данные».

1. Перейдите в пункт меню «Данные».
2. Чтобы создать новую группу нажмите кнопку , откроется форма создания новой группы:

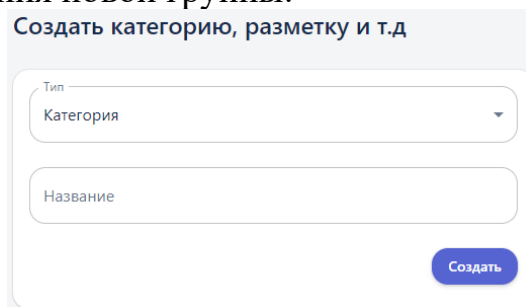


Рисунок 202 – Создание новой группы

3. В открывшейся форме в поле «Тип» выберите значение «Категория», в поле «Название» введите название новой группы, и нажмите кнопку «Создать».

Примечание – здесь под группой имеется в виду папка, в которую будут загружаться данные

для решения задачи. Пользователь создает такие папки самостоятельно,

4. Далее в папке, созданной в шаге 2, создайте две *подпапки* – «Train» и «Test» (в описании приведены примеры названий, отражающие их суть, нет необходимости давать папкам аналогичные название, главное, чтобы вам было понятно назначение каждой из них). В группу «Train» будут загружаться файлы для обучения будущей модели машинного обучения, а в группу «Test» – файлы для проверки ‘качества’ уже обученной модели. Качество алгоритма оценивается тем, насколько точно он может правильно классифицировать объекты из валидационной выборки.

5. Следующим этапом в папке «Train» создаются еще подпапки, отражающие *классы*. При этом количество классов равно двум, если решается задача *бинарной классификации*, и больше двух – для *многоклассовой классификации*. В каждую из этих папок загружаются объекты – файлы, содержащие данные соответствующего класса. Например, создается папка-класс «самолеты» и туда загружаются изображения самолетов. Количество данных для обучения модели должно

быть достаточным, оно определяется аналитиком самостоятельно, и зависит от направленности и специфики решаемой задачи.

6. Повторите действия из шага 4 для папки «Test». **Важно** – Названия создаваемых классов в тестовой выборке должны абсолютно совпадать с названиями классов в обучающей выборке. Как правило, данные для валидации составляют 20% от общего объема выборки.

Итоговая структура группы:

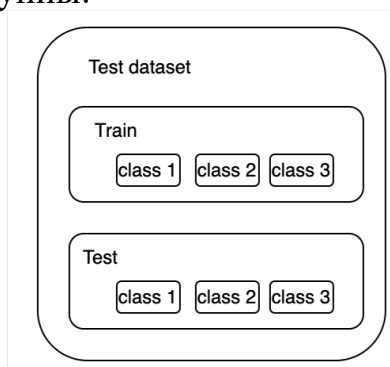


Рисунок 203 – Принцип создания групп и классов для решения задач классификации

7. Последний шаг – это присвоение метки «Классификация» созданным группам. Такая метка назначается папкам «Train» и «Test» (примеры названий). Это действие позволяет сформировать датасеты. Для этого нажмите на три точки в строке с названием папки и выберите действие «Классификация»:

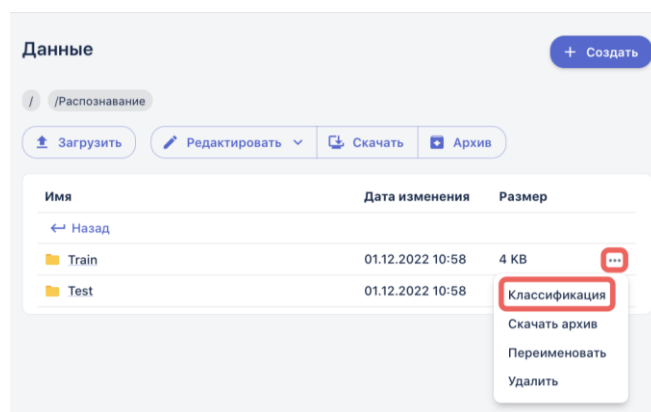


Рисунок 204 – Присвоение метки ‘Классификация’

На этом этап подготовки данных для решения задачи классификации завершается. Когда вы будете строить блок-схему и выбирать исходные данные, папки с меткой ‘Классификация’ будут доступны для выбора.

Для удаления ошибочно созданных или неактуальных папок используется действие «Удалить», а для сохранения папки на локальном компьютере в виде архива – действие «Скачать архив».

2.3.5. Создание модели ИИ

В разделе «Моделирование» осуществляется процесс построения блок-схем - соединение последовательных функции процессов анализа, обработки и преобразования исходных данных для построения и обучения моделей искусственного интеллекта. Построение таких блок-схем осуществляется на рабочих областях.

2.3.5.1. Создание новой и открытие сохраненной рабочей области

2.3.5.1.1. Создание новой рабочей области

1. Перейдите в пункт меню системы Моделирование -> Рабочая область:

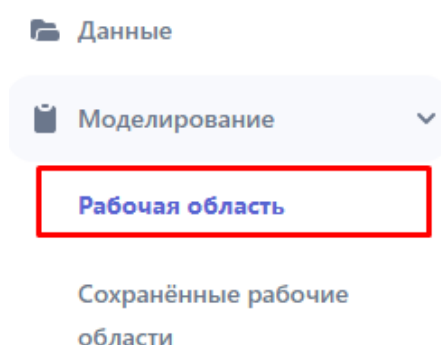


Рисунок 205 – Переход на рабочую область

Откроется страница с пустой рабочей областью:

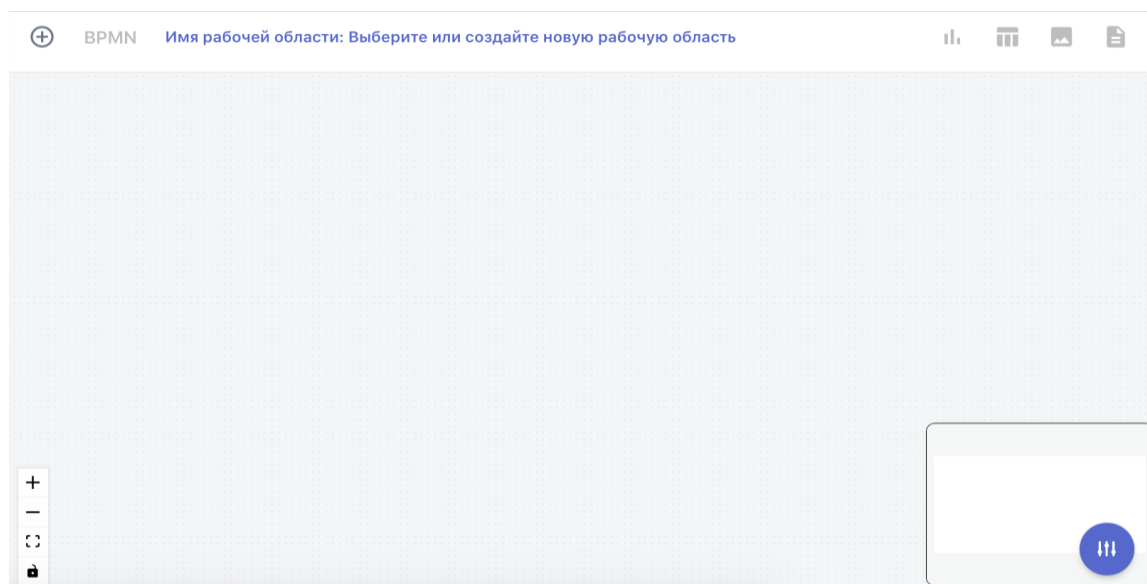



Рисунок 206 – Пустая рабочая область

2. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка ).

В открывшейся форме введите название новой рабочей области и нажмите кнопку «Создать»:

Введите имя рабочей области

Рисунок 207 – Создание новой рабочей области

На панели инструментов отобразится название созданной рабочей области:

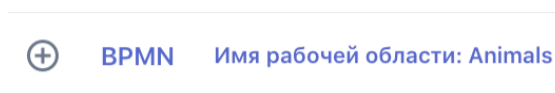


Рисунок 208 – Отображение названия рабочей области

После присвоения названия рабочая область автоматически сохраняется в раздел «Сохраненные рабочие области».

2.3.5.1.2. Открытие сохраненной рабочей области

Для того чтобы открыть ранее созданную рабочую область, перейдите в пункт меню системы Моделирование -> Сохраненные рабочие области. Откроется страница «Рабочие области»:

Рабочие области		
<input type="text" value="Поиск"/>		
Название	Создан	
Spark_Traffic	21.09.2022 10:11	
img classification test	14.09.2022 08:58	
TestTestTEstTEst	08.09.2022 11:28	
NB	04.09.2022 16:07	

Рисунок 209 – Страница со списком созданных в Системе рабочих областей

На странице рабочие области отображаются в порядке их создания – последняя созданная рабочая область отображается вверху списка.

Чтобы открыть сохраненную рабочую область кликните на её название. Чтобы удалить ненужную рабочую область - кликните кнопку удалить в правой части строки с названием области.

2.3.5.2. Построение блок схемы

Предварительным условием для построения блок схемы является:

- 1) Загруженные в систему данные (файлы или датасеты)
- 2) Созданная рабочая область (на одной рабочей области можно создать неограниченное количество блок-схем).

Построение блок-схемы осуществляется путем добавления на рабочую область элементов (блоков) и соединение их между собой.

2.3.5.2.1. Блок «Запуск»

Блок «Запуск» обозначает начало блок-схемы, и всегда является её первым элементом. Так как на рабочей области может быть несколько блок-схем, именно по блоку «Запуск» определяется их количество, и идентифицируется принадлежность блоков к той или иной блок-схеме.

Для добавления на блок-схему элемента «Запуск» нажмите на кнопку «Добавить элемент» (кнопка **BPMN**) на панели инструментов:



Откроется меню выбора блоков – библиотека графических элементов нотации BPMN 2.0:

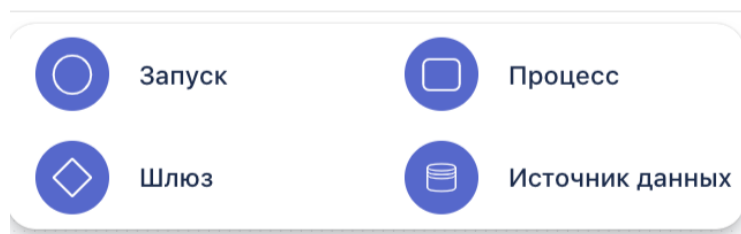
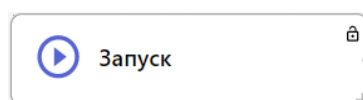



Рисунок 210 – Меню блоков


Выберите элемент «Запуск». Выбранный элемент будет добавлен на рабочую область:



Кнопка  на элементе предназначена для запуска текущей блок-схемы после её сборки. У элемента «Запуск», как первого элемента блок-схемы, есть только одна точка выхода, предназначенная для соединения с последующими элементами блок-схемы.

При необходимости размер блока можно увеличить или уменьшить, потянув за уголок в правой нижней части элемента:

Обратите внимание, что в правом нижнем углу рабочей области отображается уменьшенное изображение создаваемой блок-схемы, эта функция упрощает навигацию по полотну.

После того, как блок схема будет собрана и готова к запуску, нужно будет нажать на кнопку , после этого вид блока изменится и появится возможность создать отчет по результатам отработки:

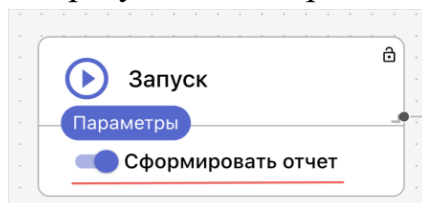



Рисунок 211 – Опция формирования отчета на блоке «Запуск»

2.3.5.2.2. Блок «Источник данных»

Следующим элементом в блок схеме после «Запуска» всегда является «Источник данных» - блок, который определяет какие данные будут использоваться в сценарии.

Для того чтобы добавить блок на рабочую область, откройте меню блоков и выберите элемент «Источник данных» (кнопка ). Выбранный элемент появится на рабочей области конструктора. Вы можете передвинуть элемент на любую часть рабочей области, нажав на него. Для понятной визуализации процесса, рекомендуется расположить «Источник данных» правее элемента «Запуск».

Для объединения элементов в блок-схему, их требуется соединить между собой. Для этого нажмите на точку выхода блока, которая отображается в виде круглой точки на правой грани блока, и перетащите мышью появившуюся стрелку в сторону нужного блока, как показано на рисунке:

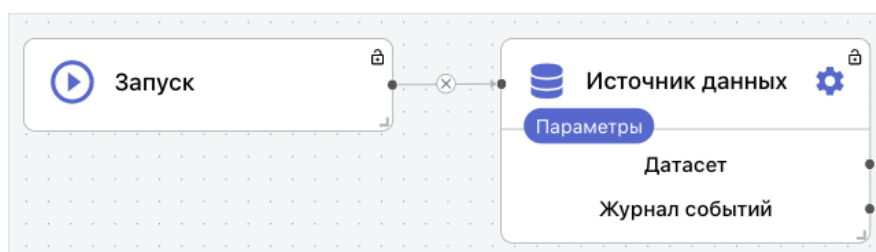




Рисунок 212 – Соединение блоков между собой

Соедините стрелкой элементы «Запуск» и «Источник данных» (далее соединение нового элемента с предыдущим является действием по умолчанию). Элемент «Источник данных» уже имеет несколько точек

выхода, которые соединяются с одноименными точками последующих элементов. Для удаления соединения необходимо нажать на кнопку  на линии соединения.

На блоке «Источник данных» отображаются два компонента:

- 1) «Датасет» — это непосредственно сами данные.
- 2) «Журнал событий» содержит информацию обо всех преобразованиях с данными, которые выполняются в текущем блоке пайплайна. Ведение журнала позволяет сохранить историю преобразований над данными, и при необходимости выполнить обратное преобразование.

Чтобы открыть настройки элемента нажмите на нем кнопку «Параметры»  (по умолчанию параметры элемента развернуты):

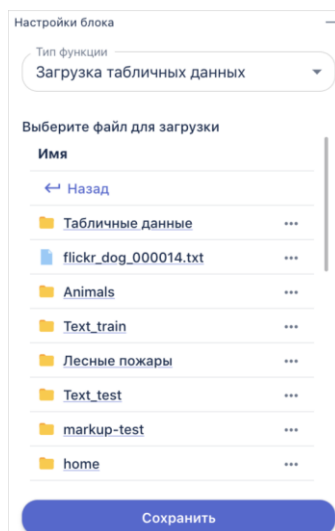


Рисунок 213– Настройка параметров блока «Источник данных»

** Дальнейшие действия по настройке блоков указаны для примера. При работе с Системой пользователь должен выбрать необходимые параметры исходя из своей задачи и загруженных данных.*

После добавления на рабочую область, для элемента «Источник данных» по умолчанию выбрана функция: тип функции «Загрузка данных» -> функция «Загрузка табличных данных». Это можно увидеть в верхней части окна настройки параметров элемента:

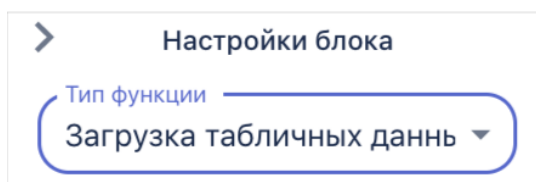


Рисунок 214 – Отображение типа функции в параметрах блока

Для того чтобы посмотреть другие доступные функции, нужно нажать на выпадающий список. Весь список доступных функций и их описания доступны в Таблице 18.1 – Перечень автоматизированных функций элемента «Источник данных».

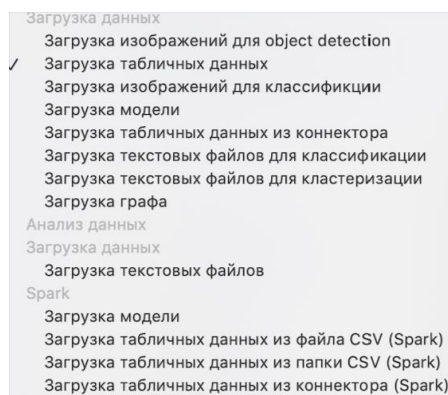


Рисунок 215 – Список возможных функций элемента «Источник данных»

В разделе «Выберите файл» отображается структура папок из разделе «Данные», чтобы выбрать файл достаточно перейти в нужную папку и кликнуть на три точки в правой части строки с названием файла и нажать «Выбрать»:

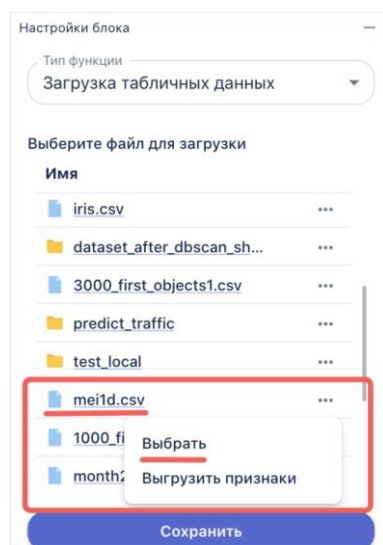


Рисунок 216 – Отображение папок и файлов из раздела «Данные» в параметрах блока «Источник данных»

После этого в нижней части окна отобразится его название:

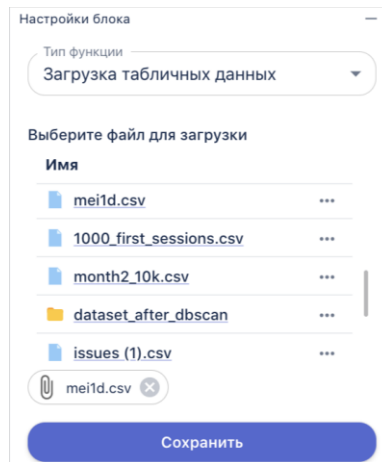


Рисунок 217 – Отображение выбранного файла, предназначенного для загрузки в блок-схему

Кнопка «Выгрузить признаки» используется для других блоков, где в настройках необходимо указать целевые признаки для конкретной функции.

Если кликнуть на название файла на рабочей области отобразится его предпросмотр:

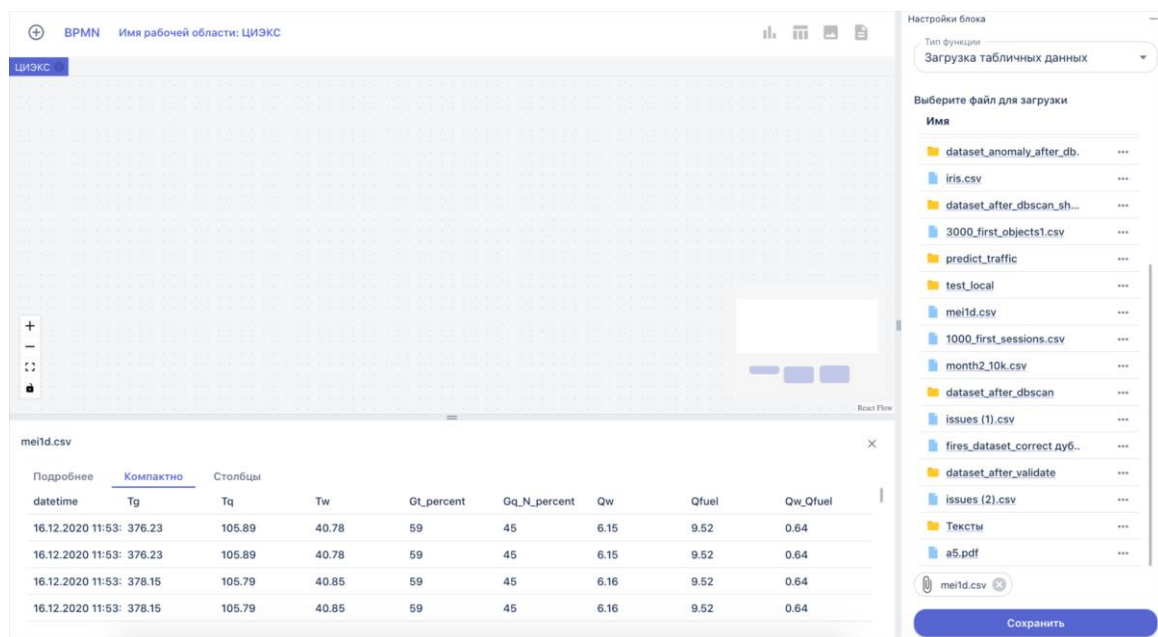


Рисунок 218 – Предпросмотр файла на рабочей области

Для сохранения выбранных настроек нажмите на панели параметров кнопку «Сохранить» (далее сохранение настроек элемента предполагается по умолчанию). Для удаления неверно добавленного на рабочую область элемента предусмотрена кнопка «Удалить блок».

Любой блок можно переименовать, чтобы дать ему понятное название, отображающее суть происходящего процесса. Для этого дважды щелкните левой кнопкой мыши на текущее название элемента в рабочей области и измените его. Чтобы новое название сохранилось достаточно щелкнуть мышью в любом месте на рабочей области.

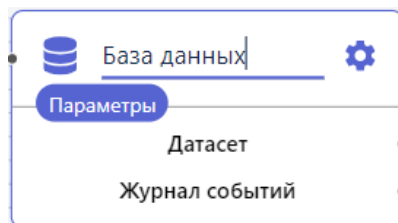


Рисунок 219 – Ввод названия элемента

Для того чтобы удалить блок, кликните по нему правой кнопкой мыши и нажмите «Удалить».

2.3.5.2.3. Блок «Процесс»

Блок «Процесс» предназначен для выполнения операций над данными. Блок-схема может содержать несколько элементов «Процесс», настроенных пользователем для выполнения определенных задач. Весь список доступных функций и их описания доступны в Таблице 18.2 – Перечень автоматизированных функций элемента «Процесс».

Аналогично тому, как было описано в предыдущих пунктах, откройте меню блоков и выберите элемент «Процесс» (☐). Выбранный элемент появится на рабочей области.

Далее будет показан принцип настройки свойств блока на примере одной функции. Выбор функции определяется типом решаемой задачи.

Для примера выберите для элемента функцию: раздел «Машинное обучение» -> функция «Разделение датасета на обучающую и тестовую выборки»:

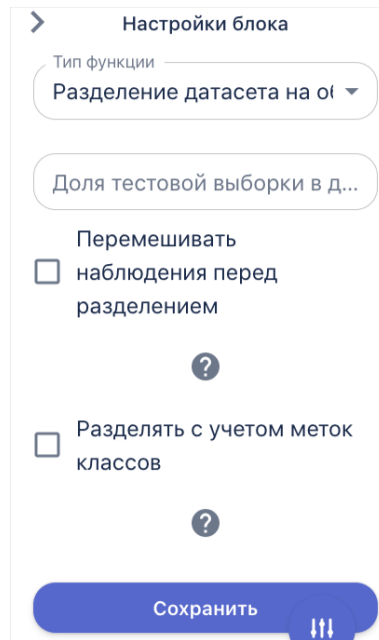


Рисунок 220 – Панель свойств блока «Процесс»

Далее осуществляется настройка параметров следующим образом:

- В разделе «Параметры» → в поле «Доля тестовой выборки в датасете» введите значение 0.2;
- Оставьте пустым поле «Перемешивать наблюдения перед разделением». Рядом с полем есть подсказка, что не рекомендуется перемешивать наблюдения во временных рядах (выбирайте действие в зависимости от типа входных данных);
- Установите галочку в поле «Разделять с учетом меток классов» – применяется для задач классификации (выбирайте действие в зависимости от решаемой задачи).

Измените название элемента на «Сплит датасета»:

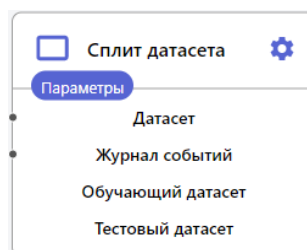


Рисунок 221 – Отображение блока на рабочей области после настройки его параметров и ввода названия

Обратите внимание, что соединение элемента «Процесс» с другими элементами блок-схемы выполняется только после настройки и сохранения его параметров. Это связано с тем, что каждая функция имеет свой набор

компонентов, который отображается на элементе после сохранения его настроек.

Соединить элемент «Процесс» с предыдущими элементами блок-схемы можно следующим образом:

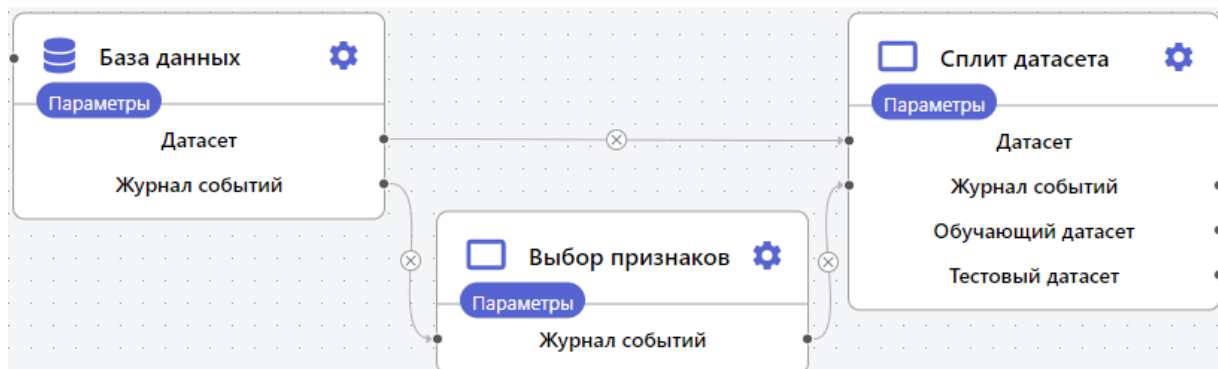



Рисунок 222 – Соединения между несколькими блоками (организация связей/пробрасывание данных)

Обратите внимание, что соединять можно только идентичные (одноименные) компоненты блоков.

2.3.5.3. Запуск блок-схемы на рабочей области

Чтобы запустить блок схему нужно нажать на кнопку  на элементе «Запуск». В результате элементы пайплайна начинают последовательно запускаться. При обработке блока он загорается оранжевым цветом, а после успешного завершения – зеленым:

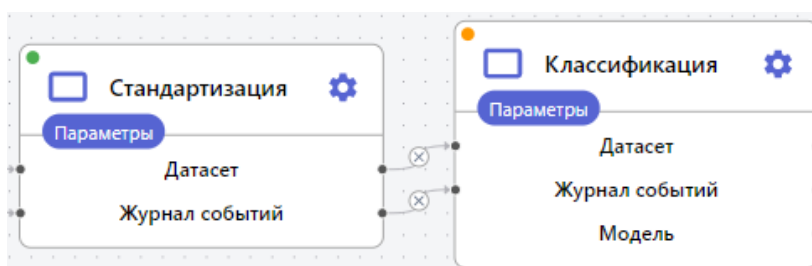


Рисунок 223 – Последовательная обработка блоков пайплайна

В случае, если блок не отработал (например, вследствие того, что был неправильно настроен или из-за неверных входных данных), на нем появится индикатор красного цвета. Такой блок необходимо проверить, изменить настройки и запустить заново:

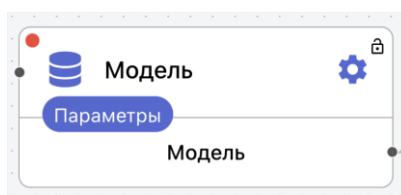


Рисунок 224 – Индикатор ошибки при обработке блока

2.3.6. Сохранение модели ИИ

Для сохранения модели ИИ, обученной выполнению какой-либо задачи, на блок-схему необходимо добавить специальный блок. Этот блок настраивается для элемента «Процесс» с помощью функции «Управление моделями» → функция «Сохранение модели»:

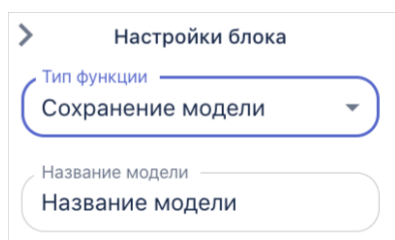
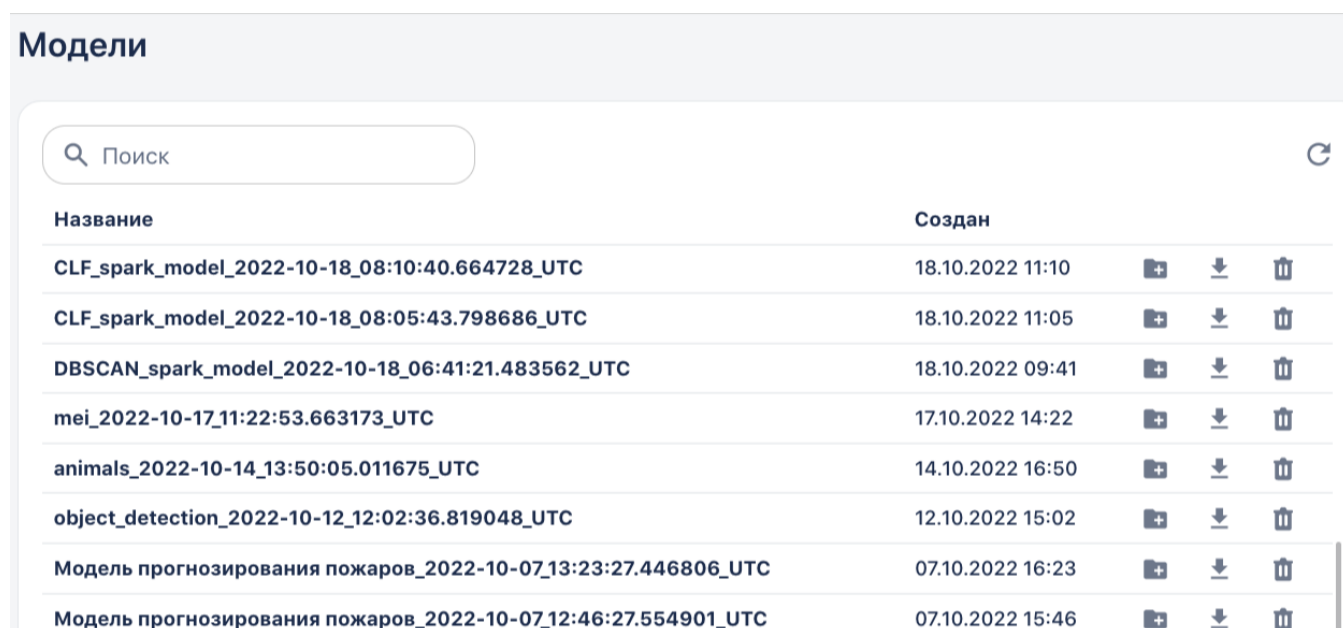


Рисунок 225 – Настройка блока для сохранения обученной модели ИИ

После успешной отработки блок-схемы, которая содержит такой блок, название модели будет добавлено в список сохраненных моделей в меню «Модели»:





Название	Создан			
CLF_spark_model_2022-10-18_08:10:40.664728.UTC	18.10.2022 11:10	+	↓	🗑️
CLF_spark_model_2022-10-18_08:05:43.798686.UTC	18.10.2022 11:05	+	↓	🗑️
DBSCAN_spark_model_2022-10-18_06:41:21.483562.UTC	18.10.2022 09:41	+	↓	🗑️
mei_2022-10-17_11:22:53.663173.UTC	17.10.2022 14:22	+	↓	🗑️
animals_2022-10-14_13:50:05.011675.UTC	14.10.2022 16:50	+	↓	🗑️
object_detection_2022-10-12_12:02:36.819048.UTC	12.10.2022 15:02	+	↓	🗑️
Модель прогнозирования пожаров_2022-10-07_13:23:27.446806.UTC	07.10.2022 16:23	+	↓	🗑️
Модель прогнозирования пожаров_2022-10-07_12:46:27.554901.UTC	07.10.2022 15:46	+	↓	🗑️

Рисунок 226 – Вкладка меню «Модели»

Обратите внимание, что модель будет сохраняться столько раз, сколько будет запущена блок-схема с элементом «Сохранение модели», при этом в разделе будет меняться временная отметка создания записи.

Модель можно использовать следующими способами:

1. Создать на её основе приложение, которое предназначено для интеграции с внешними системами. Для этого нажмите на значок «Новое приложение»  в строке с названием модели. Новое приложение появится в соответствующем разделе системы.

2. Скачать. Для этого нажмите на значок «Скачать»  в строке с названием модели. После этого на ваш компьютер будет сохранен архив с тремя файлами:

- **model.pkl** - сама модель.
- **vars_dict.pkl** - словарь преобразований. Преобразования необходимо сохранять, чтобы при анализе новой порции данных над ними выполнялись все те же преобразования, что и при обучении модели.

- **info.json** - служебный файл, куда прописывается тип модели.









3. Использовать при построении новой блок-схемы в качестве источника данных (например, для целей прогнозирования). Пример можно посмотреть в разделе Работа с данными в режиме реального времени.

4. Создать коннектора с обученной моделью (например, для распознавания объектов на видео или изображениях). Это позволит проверить обучение модели на новой порции данных. Подробнее описано в разделе Проверка обученной модели на локальных данных.

2.3.7. Графическое представление информации на рабочей области


После *сборки и успешного* запуска блок-схемы на рабочей области есть возможность посмотреть результаты обучения созданной модели в виде графиков, таблиц и изображений, которые можно вывести прямо на рабочую область.

Примечание: под сборкой имеется в виду, что на рабочую область добавлены все элементы блок-схемы, и они последовательно соединены между собой. А успешным считается запуск блок-схемы, когда все ее элементы отработали с «зеленым»

После успешного запуска в верхней части панели инструментов рабочей области станут активными следующие кнопки:    , где  – графики,  – таблицы,  – изображения,  – описание модели.

Вы можете нажать на те кнопки визуализации, которые подсвечиваются фиолетовым цветом. При этом голубым подсвечиваются только те иконки, которые актуальны для запущенной схемы, т.к. блоки могут содержать разные функции, имеющие разное графическое представление.

2.3.7.1. Графики

Чтобы отобразить результаты работы блок схемы в виде графиков необходимо нажать на кнопку  .

Полный список доступных в Системе графиков с объяснением интерпретации результатов доступен в **Базе знаний**. За каждым типом блока закреплен определенный набор графиков, например, для блока Анализ данных -> Анализ временных данных доступны следующие графики: Линейный график, ACF/PACF, Декомпозиция, Свечной график, Time profile, Extended, Bollinger Bands Stochastic Oscillator.

Для того чтобы добавить график на рабочую область из выпадающего списка выберите нужное название. Например, «Time profile временного ряда» в анализе временных рядов:

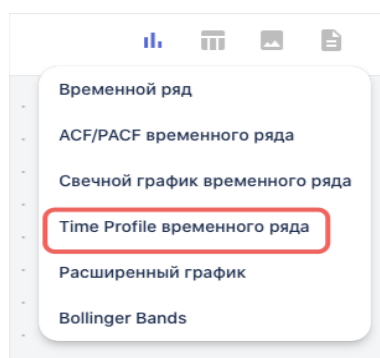


Рисунок 227 – Список доступных графиков для анализа временных рядов

В результате на рабочую область будет добавлен график:



Рисунок 229 – Time profile временного ряда на рабочей области

Для ряда графиков доступен выбор из выпадающего списка признака, для которого составляется визуализация. Например, можно два раза выбрать график Time profile и для одного указать признак Tq, а для другого Tw и сопоставлять их значения одновременно на рабочей области:



Рисунок 230 – Отображение одного типа графика для разных признаков

Также в правом углу рамки графической визуализации, при наведении на неё курсора, отображаются следующие кнопки:



Рисунок 231 – Кнопки для работы с визуализацией

где:









	- скачать график, как рисунок в формате png
	- при нажатии лупы, можно увеличить любой участок графика
	- при нажатии данной кнопки, можно передвигаться по плоскости графика в любую сторону
	11. приблизить, увеличить
	- уменьшить
	- автоматически подогнать размер под границы
	- вернуться к исходному виду

Таблица 232 - Функционал кнопок работы с визуализацией

В программе также реализована возможность создания графиков с данными, получаемыми в режиме реального времени. Для таких графиков существует отдельный блок, находящийся в разделе «Анализ данных» -> «Визуализация Real Time». Главное отличие от стандартного блока «Визуализация» в том, что для каждого графика необходимо задавать число периодов в окне и период окна - параметры, которые определяют интервал, который будет отображаться на графике в рабочей области.

2.3.7.2. Таблицы

Чтобы отобразить результаты работы модели в виде таблиц нажмите кнопку .

Например:

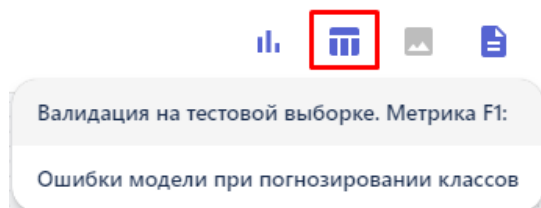
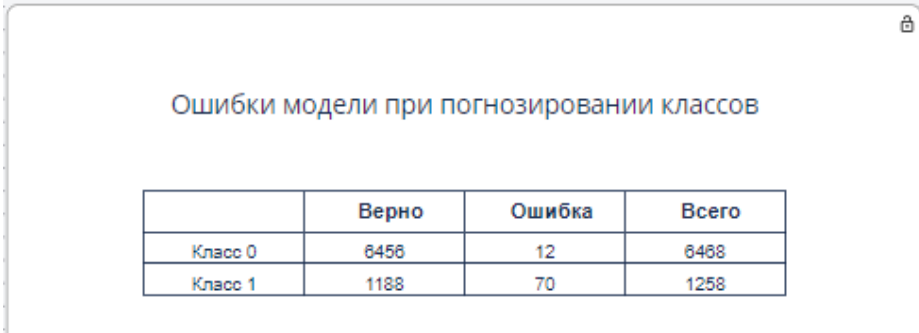


Рисунок 233 – Список таблиц для визуализации


Чтобы отобразить таблицу на рабочей области, нажмите на её название в выпадающем списке и она появится на экране, например:



	Верно	Ошибка	Всего
Класс 0	6458	12	6468
Класс 1	1188	70	1258

Рисунок 234 – Блок визуализации «Ошибки модели при прогнозировании классов» (показан пример для сценария 14.1 «Прогнозирование лесных пожаров»)

2.3.7.3. Изображения

Чтобы отобразить результаты работы модели в виде таблиц нажмите кнопку  .

Например:

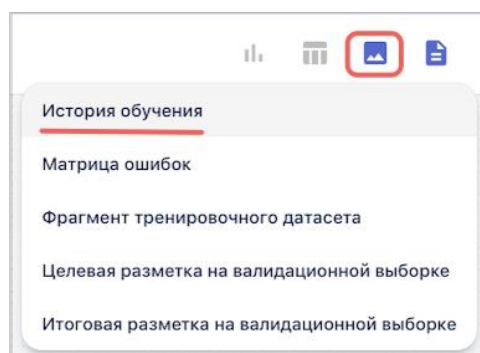


Рисунок 235 – Список изображений для визуализации

Далее выберите изображение из списка и визуализация появится на рабочей области:

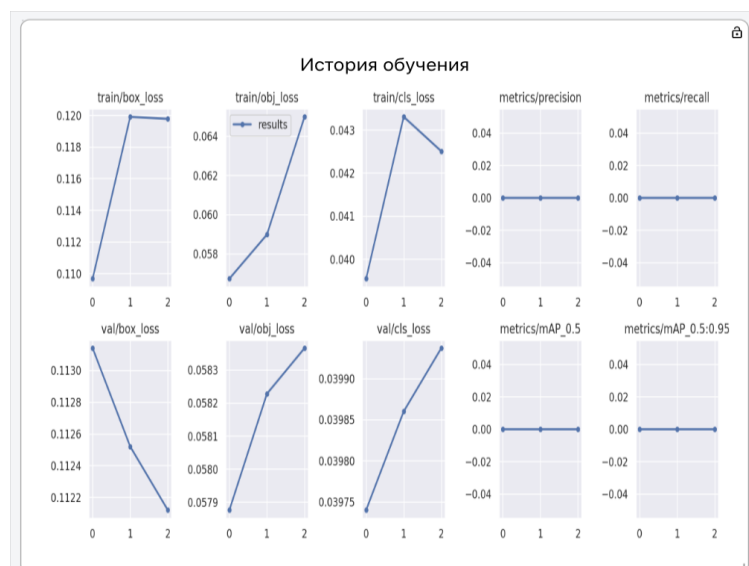


Рисунок 236 – Блок визуализации «История обучения (показан пример для сценария 14.2 «Обучение модели ИИ распознаванию объектов на изображениях»)

2.3.7.4. Описание модели

Чтобы посмотреть описание модели нажмите кнопку  .

Описание варьируется в зависимости от функций, которые были применены в модели. Например, для блок-схемы, где присутствовали элементы «Стандартизация», «Валидация» и «XGB классификации» описание будет выглядеть следующим образом:

Лучшие гиперпараметры при кросс-валидации:

max_depth	5
n_estimators	50

Лучшая метрика F1 при кросс-валидации:
0.959

Время обучения полной обучающей выборки в сек:
13.875

Модель

```
XGBClassifier(base_score=0.5, booster='gbtree',
colsample_bylevel=1, colsample_bynode=1,
colsample_bytree=1, enable_categorical=False,
gamma=0, gpu_id=-1, importance_type=None,
interaction_constraints="", learning_rate=0.300000012,
max_delta_step=0, max_depth=5, min_child_weight=1,
missing=nan, monotone_constraints='()'),
n_estimators=50, n_jobs=40, num_parallel_tree=1,
predictor='auto', random_state=42, reg_alpha=0,
reg_lambda=1, scale_pos_weight=1, subsample=1,
tree_method='exact', validate_parameters=1,
verbosity=None)
```

Список преобразований целевых признаков:
без преобразований.

Список преобразований признаков:
Стандартизация.

Рисунок 237 – Вариант описания модели

Если на рабочей области размещены несколько блок схем, при нажатии на описание вы увидите информацию по каждой из них.

2.3.8. Работа с Дашбордами. Раздел «Визуализация».

Дашборд – это интерактивная рабочая область, которая наглядно представляет, визуализирует, объясняет и анализирует данные. На рабочую область пользователь может добавлять графики, таблицы, диаграммы, визуализацию пайплайнов для последующей работы с ними.

Работа с дашбордами осуществляется в разделе «Визуализация» → «Дашборды». Для создания нового дашборда на панели инструментов нажмите кнопку « ». В открывшемся окне введите название создаваемого

дашборда, например «Таблица», и нажмите кнопку «Создать»:

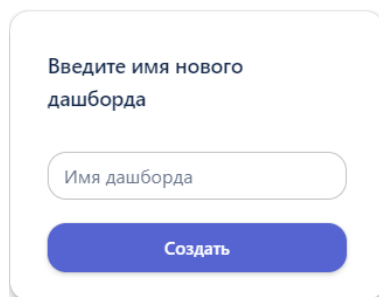


Рисунок 238 – Создание нового дашборда

В результате отобразится название дашборда (это текущий дашборд, с которым работает пользователь):

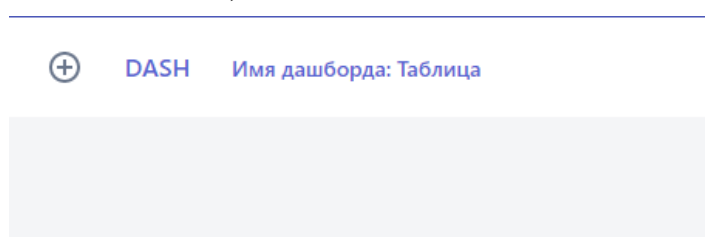


Рисунок 239 – Отображение наименования дашборда

Новый дашборд создается с пустой рабочей областью. Чтобы его наполнить добавляются интерактивные блоки. Для этого нажмите кнопку **DASH** и выберите нужный тип интерактивного блока:

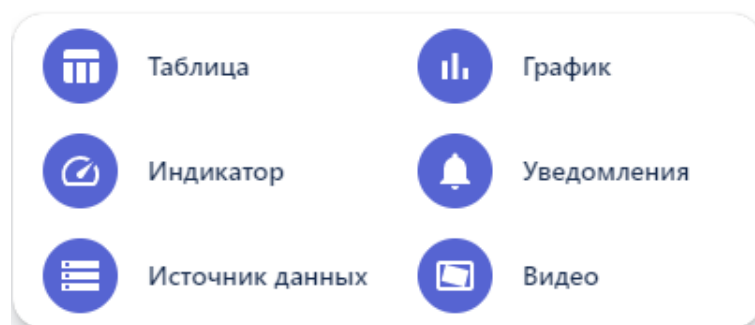


Рисунок 240 – Добавление интерактивного блока

В текущей версии Системы реализована работа с блоками: «Таблица», «График» и «Видео». Предварительным условием для добавления любого из типов блоков является создание коннектора в разделе «Соединения», в котором настроено получение данных. При этом это могут быть как данные, получаемые из внешних источников, так и сгенерированные внутри Системы. Подробнее о создании коннекторов вы можете посмотреть в разделе Работа со всеми типами коннекторов.

2.3.8.1. Таблица

Интерактивный блок «Таблица» предназначен для:


- Отображения подключения к внешним базам данных, в виде табличных данных, обновляющихся в режиме реального времени. В таком случае настраивается подключение к коннекторам с названиями типов баз данных (clickhouse, postgresql, mongo);

- Записи и сохранения в Системе информации, получаемой из внешних баз данных. Используется коннектор с типом «save_table»;

- Прогнозирования целевых событий, когда данные для анализа поступают из внешних баз данных. Используется коннектор с типом «table_app»;

- Отображение таблиц, полученных в результате отработки блок-схем, которые содержат блоки, имеющие в качестве выходной информации визуализации в виде таблиц. Используется коннектор с типом «constructor». Такой тип коннектора создается автоматически, после успешного запуска блок-схемы.

Например, чтобы настроить дашборд с подключением к коннектору с типом «clickhouse»:

1. Создайте новый дашборд.
2. На дашборд добавьте интерактивный блок с типом «Таблица».
3. Чтобы подгрузить данные в таблицу в правом верхнем углу нажмите кнопку . Отобразится список коннекторов, созданных в Системе:

Connectors












Название	Создан	Статус	
 Коннектор ClickHouse	12.08.2022 14:02	Stopped	  
set_features_targets_(TRAFFIC)	03.08.2022 16:21	Started	  
load_data_(TRAFFIC)	03.08.2022 15:45	Started	  

Рисунок 241 – Список коннекторов

В открывшемся окне для коннекторов отображаются их состояния (Запущен (Started) / Остановлен(Stopped)). Если статус коннектора «остановлен», данные из внешних источников не поступают в Систему. Из этого окна можно запустить/остановить коннектор, выполнить его редактирование при необходимости, или удалить его.

4. Запустите коннектор, нажатием кнопки  в строке с коннектором (по умолчанию, коннекторы с типом «clickhouse» создаются в статусе «Stopped»).

5. Выбрать коннектор из списка нажатием на него левой кнопкой мыши.

6. На дашборде отобразится результат подключения к БД «ClickHouse» в виде таблицы с данными:



Data	Y5401	Y5402	Y5707	Y5708	Y5403	Y5404	Y5705	Y5706	P596
2021-09-01	2.701100111	1.108100056	2.03839993	5.59660005	6.69929981	0.02500000	0.66269999	4.93319988	0.889
2021-09-01	2.67519998	1.13929998	2.01699995	5.46570014	6.62709999	0.02390000	0.63190001	4.85069990	0.889
2021-09-01	2.73690009	1.15919995	2.01990008	5.43620014	6.67539978	0.02390000	0.79119998	5.02390003	0.889
2021-09-01	2.36220002	1.17149996	1.86570000	5.13219976	6.21269989	0.02380000	0.74739998	4.57070016	0.879
2021-09-01	2.05369997	1.26590001	1.77119994	4.87739992	5.67309999	0.02360000	0.63870000	4.03620004	0.870
2021-09-01	2.01929998	1.27520000	1.75940001	4.91300010	5.60090017	0.02380000	0.65689998	4.08449983	0.870
2021-09-01	2.00040006	1.31459999	1.75559997	4.91750001	5.59810018	0.02170000	0.67479997	4.09439992	0.860
2021-09-01	2.06110000	1.32720005	1.77139997	4.89720010	5.77670001	0.02170000	0.74140000	4.28620004	0.860
2021-09-01	2.07389998	1.28970003	1.76119995	4.87970018	5.73740005	0.02170000	0.71109998	4.22100019	0.860
2021-09-01	2.10010004	1.27939999	1.77030003	4.88380002	5.83239984	0.02170000	0.69459998	4.13929986	0.870

Рисунок 242 – Отображение табличных данных из коннектора clickhouse


Коннектор можно остановить или запустить прямо на дашборде, для этого нажмите кнопку «остановить»/«запустить». Для того чтобы удалить дашборд - нажмите на крестик в правом верхнем углу блока.

2.3.8.2. Видео

Интерактивный блок «Видео» предназначен для:

- Отображения видеопотока данных в режиме реального времени, с удаленной камеры видеонаблюдения;
- Записи и сохранения полученного видеопотока;
- Классификации изображений с локального компьютера.

Рассмотрим настройку блока на примере подключения к коннектору с типом «video_stream».

1. Создайте новый дашборд, например, «Видеопоток».
2. Добавьте интерактивный блок «Видео» на дашборд.
3. Выберите для блока предварительно созданный коннектор с типом «video_stream».
4. Для начала получения данных с коннектора запустите его нажатием кнопки  (при выборе коннектора для блока, или на самом блоке).

Ниже представлены примеры подключения к камерам с трансляцией видео потока с улицы и из магазина:

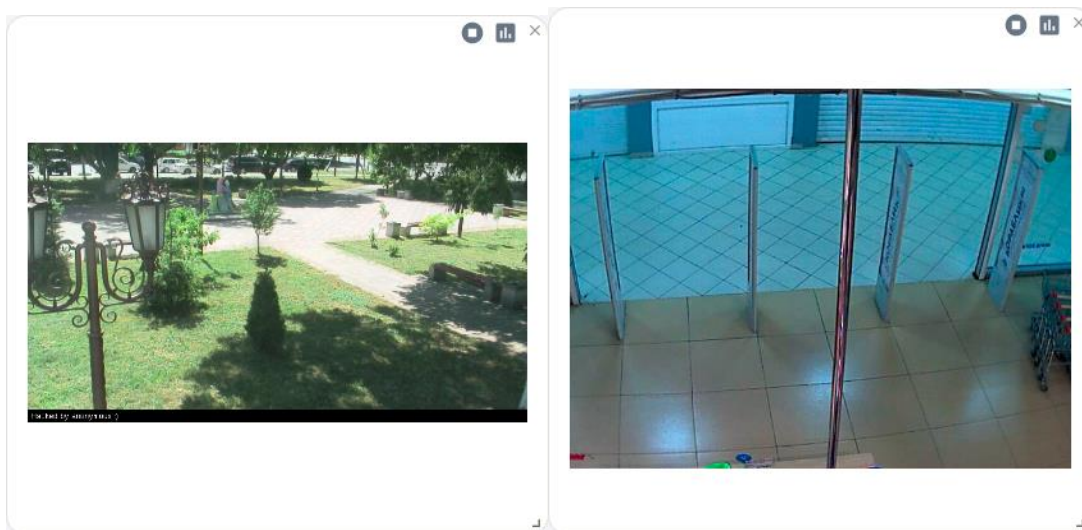

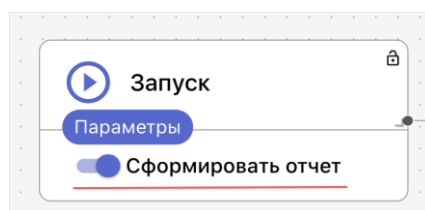


Рисунок 243 – Дашборды в видео потоком

Чтобы остановить визуализацию используется кнопка . При этом если в коннекторе настроено сохранение видеопотока, после нажатия этой кнопки файл с видео с камеры сохранится в разделе «Данные» в папке «video».

2.3.9. Создание отчета с результатами анализа данных

Более подробную информацию с результатами обучения модели можно просмотреть в отчете, который создается также после сборки блок-схемы. Для формирования отчета на первом блоке необходимо перевести бегунок вправо:



В названии отчета будет указано название блок-схемы, по которой формируется отчет и временная отметка его формирования. Для просмотра сформированного отчета после запуска блок-схемы нужно перейти в пункт меню «Отчеты» и выбрать из списка отчет, нажав на его название:

Отчеты

Поиск

Название	Создан	
ПМИ тесты_2022-09-24_14:48:53.045715.UTC	24.09.2022 17:48	🗑️
ПМИ тесты_2022-09-24_14:48:27.691319.UTC	24.09.2022 17:48	🗑️
ПМИ тесты_2022-09-23_05:24:12.891941.UTC	23.09.2022 08:24	🗑️
NB_2022-09-22_15:12:52.825178.UTC	22.09.2022 18:12	🗑️
NB_2022-09-22_15:11:59.062263.UTC	22.09.2022 18:12	🗑️
NB_2022-09-22_14:10:34.206449.UTC	22.09.2022 17:10	🗑️
NB_2022-09-22_14:10:34.126945.UTC	22.09.2022 17:10	🗑️
NB_2022-09-22_14:10:34.049589.UTC	22.09.2022 17:10	🗑️
NB_2022-09-22_14:10:33.609967.UTC	22.09.2022 17:10	🗑️
NB_2022-09-22_14:10:33.851399.UTC	22.09.2022 17:10	🗑️

Рисунок 244 – Список автоматически сформированных отчетов

После нажатия на название отчета открывается отдельная вкладка с отчетом. В отчете кроме результатов обучения созданной модели (см. «Визуализацию») отображаются также: входные данные, отдельно выборки – обучающая и тестовая, датасет после стандартизации признаков, и т.д. Состав отчета отличается в зависимости от метода, который использовался в решении задачи ИИ.

2.3.10. Конвейер приложений

Конвейер приложений позволяет создать приложение на основе обученной модели. В таком приложении заложен шаблон, умеющий предсказывать наступление интересующих событий. Приложение можно развернуть отдельно за пределами системы, интегрировать с внешними системами, настроить получение входных данных и выполнять прогнозы.


Для того чтобы сформировать приложение необходимо перейти в раздел “Модели”:

Модели

Поиск

Название	Создан			
mei_2022-08-22_09:16:01.656355.UTC	22.08.2022 12:16	+	↓	🗑️
quotes_2022-08-17_14:54:44.696451.UTC	17.08.2022 17:54	+	↓	🗑️
fires_2022-08-15_07:56:09.705763.UTC	15.08.2022 10:56	+	↓	🗑️
fires_2022-08-15_06:48:48.599366.UTC	15.08.2022 09:48	+	↓	🗑️
fires_2022-08-15_09:48:48.256081.UTC	15.08.2022 09:48	+	↓	🗑️
quotes_2022-08-12_12:05:01.714450.UTC	12.08.2022 15:05	+	↓	🗑️
quotes_2022-08-12_11:40:48.733868.UTC	12.08.2022 14:40	+	↓	🗑️
fires_2022-08-12_09:11:56.817140.UTC	12.08.2022 09:11	+	↓	🗑️
quotes_2022-08-11_11:12:58.506885.UTC	11.08.2022 14:12	+	↓	🗑️

Рисунок 245 – Список сохраненных моделей

Для формирования приложения нажать  рядом с названием обученной модели, после этого оно появится в разделе «Приложения» в папке apps:








Имя	Дата изменения	Размер
← Назад		
 app_Модель прогнозирования пожаров_2022-09-29_15:41:47.492927_UTC_2022-10-05_08:01:21.776636_UTC.zip	05.10.2022 11:04	2.05 GB ...
 app_mei_2022-09-28_08:46:11.744358_UTC_2022-10-05_07:39:11.734478_UTC.zip	05.10.2022 10:45	2.05 GB ...
 app_animals_2022-09-29_11:13:59.053051_UTC_2022-10-05_13:05:22.917782_UTC.zip	05.10.2022 16:14	2.05 GB ...
 app_traffic_cluster_2022-09-29_15:52:20.319485_UTC_2022-10-05_13:05:28.186870_UTC.zip	05.10.2022 16:23	2.05 GB ...
 app_kmn_regress_2022-09-30_07:39:55.734590_UTC_2022-10-05_09:10:45.509929_UTC.zip	05.10.2022 12:17	2.05 GB ...
 app_dbscan_2022-09-30_07:39:55.424130_UTC_2022-10-07_13:08:23.272750_UTC.zip	07.10.2022 16:14	2.05 GB ...
 app_animals_2022-09-29_11:13:59.053051_UTC_2022-10-05_09:10:35.997092_UTC.zip	05.10.2022 12:14	2.05 GB ...

Рисунок 246 – Страница «Приложения»

Приложение упаковано в docker-контейнер и доступно для скачивания. Для того чтобы скачать приложение - нажмите на три точки в строке с названием приложения и скачайте его.

Комплектность приложения после создания и скачивания:

- **app.py** - файл приложения, которое принимает данные, вычисляет и возвращает результат прогнозирования на основе обученной модели
- **const.py** - переменные в приложении.
- **methods.py** - методы, используемые в модели
- **model.py** - загружает модель из файлов `model.pkl`, `model_vars_dict.pkl`, в которых хранится модель и параметры.
- **preprocess_and_predict.py** - подготовка и вычисление переменных для прогнозирования.
- **run.sh** - командный интерпретатор для Linux.
- **run.bat** - командный интерпретатор для Windows.
- **requirements.txt** - зависимости для сборки приложения.
- **dockerfile** - файл конфигурации, в котором расписано пошаговое создание среды для работы приложения.
- **docker-compose.yml** - файл с командами для запуска среды приложения.

Приложение позволяет решать задачи предиктивной аналитики для новых данных с использованием обученной модели. Предназначено для использования в сторонних системах.

2.3.11. Работа с проектом

Сущность «Проект» реализована с целью объединить *группу пользователей* для работы над одним проектом. При этом проект может объединять в себе такие сущности как: «модель», «рабочая область», «дашборд», «отчет», «файл», «коннектор». О том, как добавить каждую из этих сущностей в проект, написано в рамках данного раздела.

2.3.11.1. Создание нового проекта

1. Перейдите в пункт меню «Проекты»:

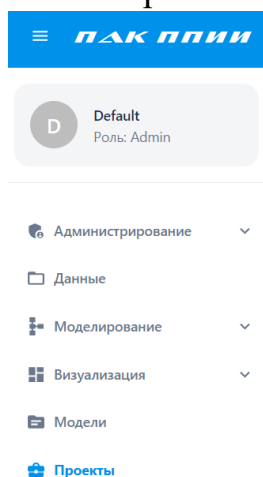


Рисунок 247 –Пункт меню «Проекты»

2. Откроется страница «Проекты», на которой отображаются все *проекты, созданные пользователем*:

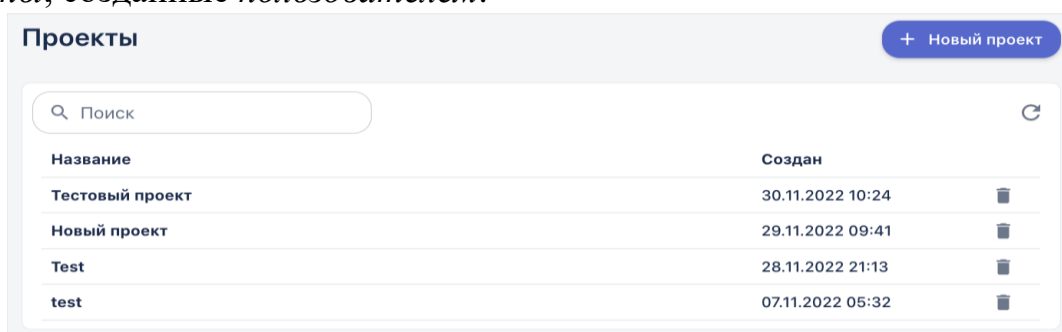


Рисунок 248 – Страница с проектами, доступными пользователю

3. Нажмите на кнопку «Новый проект», откроется окно создания нового проекта:

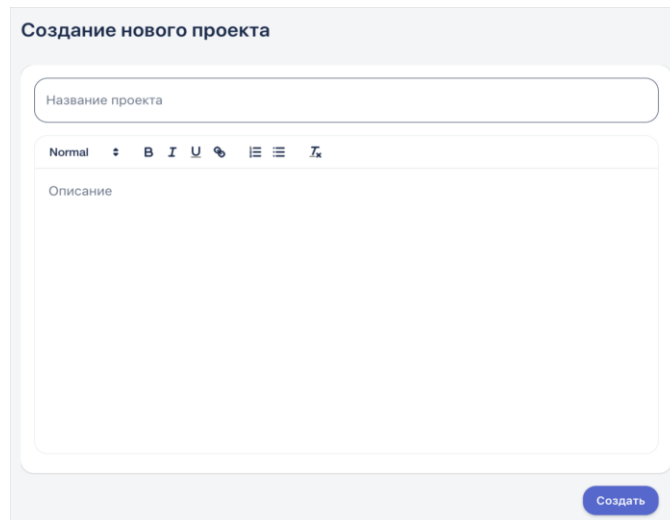


Рисунок 249 – Создание проекта

4. Задайте название проекта, например, «Проект тест» (обязательное поле)
5. Задайте описание проекта (необязательное поле), вы можете использовать средства форматирования текста:
 - Заголовки и подзаголовки
 - Жирный текст
 - Курсив
 - Подчеркивание
 - Вставка ссылки
 - Список
 - Нумерованный список
 - Кнопка очистки форматирования
6. Нажмите кнопку «Создать».
7. На страницу «Проекты» добавится новый проект.

2.3.11.2. Редактирование проекта

Пользователь имеет возможность отредактировать название проекта и его описание:

1. На странице «Проекты» перейдите в проект, который требует редактирования
2. В открывшемся окне в верхнем правом углу нажмите кнопку «Редактировать» :

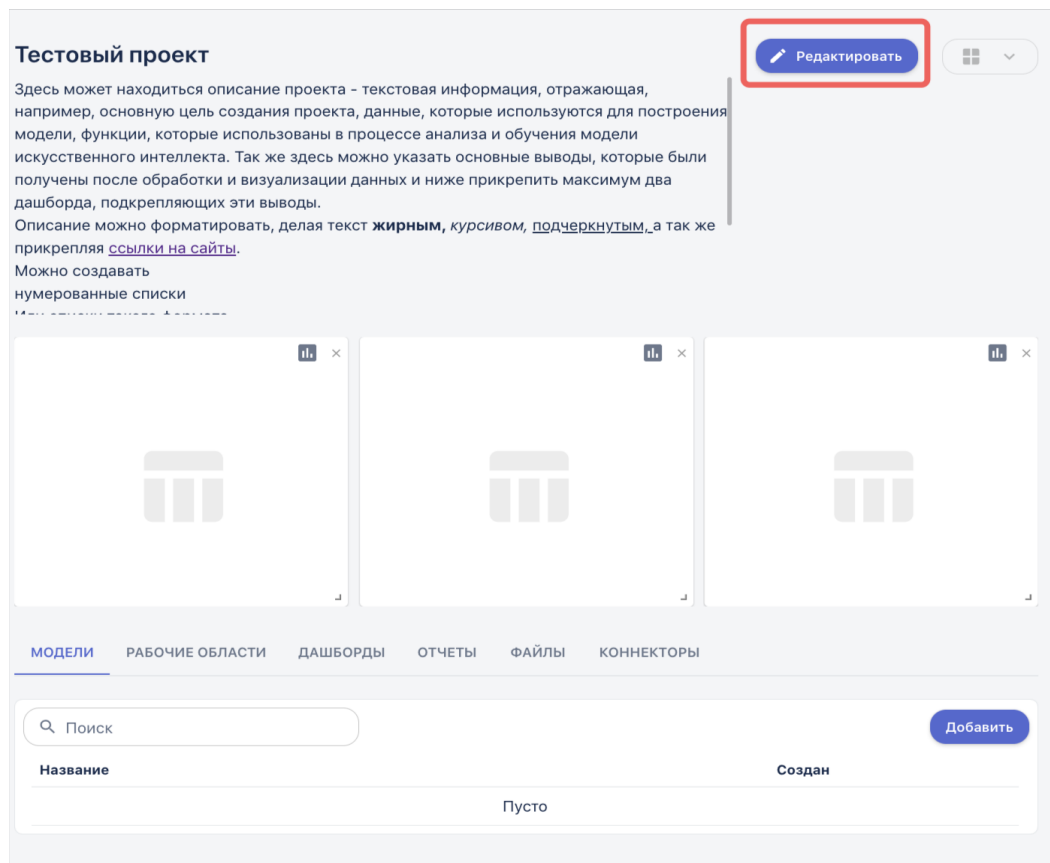


Рисунок 250 – Переход к редактированию существующего проекта

3. Задайте новое название проекта и его описание, затем нажмите кнопку «Сохранить»
4. Система вернется на страницу со списком проектов, для того чтобы посмотреть обновленное описание, перейдите в проект, нажав на его наименование.

2.3.11.3. Наполнение проекта

После того, как проект создан, его можно наполнить *содержимым* – теми сущностями, над которыми предстоит совместно работать группе пользователей. Для этого:

1. На странице «Проекты» перейдите по ссылке с названием созданного проекта, кликнув на его название:

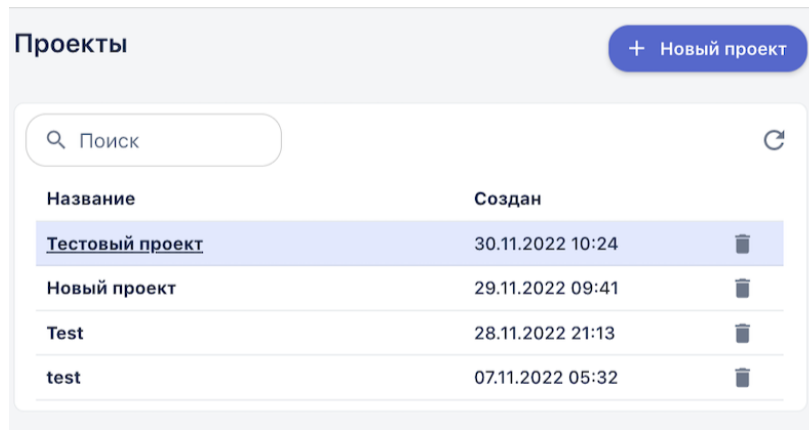


Рисунок 251 – Переход на страницу проекта для его просмотра и редактирования

2. Откроется *страница проекта* на первой вкладке «Модели»:

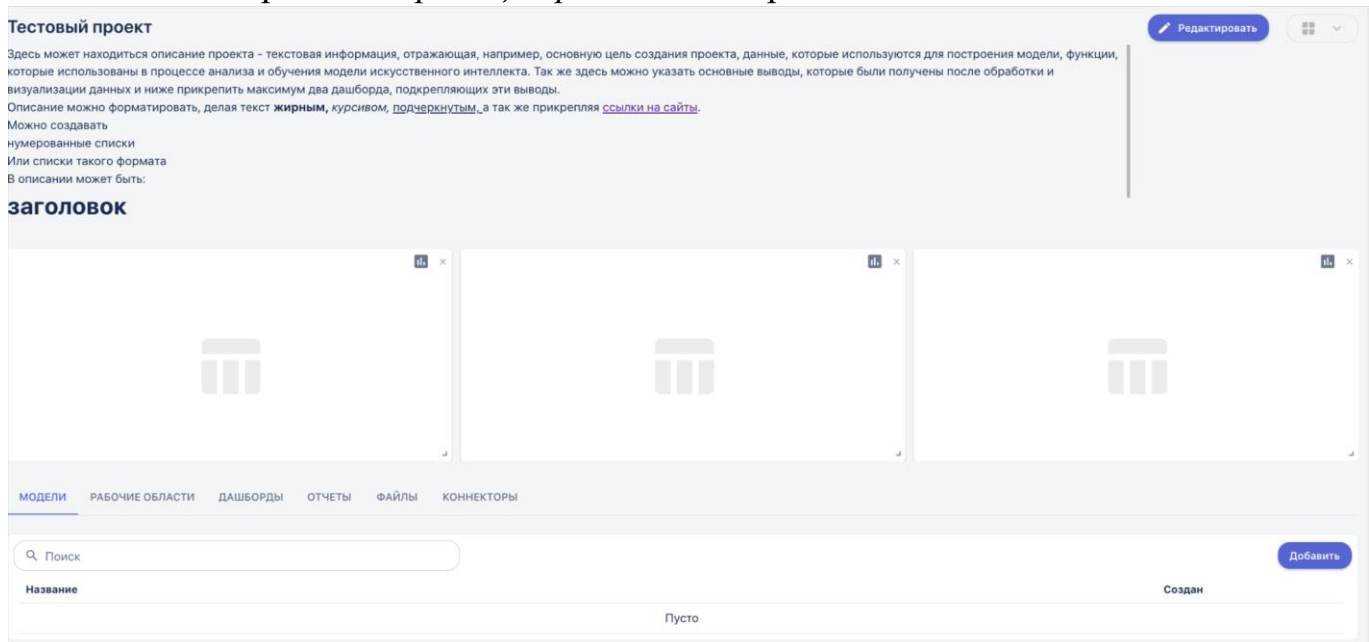


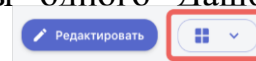
Рисунок 252 – Страница проекта

Целиком страница проекта состоит из вкладок с одноименными сущностями: модели, рабочие области, дашборды, отчеты, файлы, коннекторы.

3. **Работа с дашбордами.** Под описанием по умолчанию находятся три пустых дашборда. Вы можете сделать следующее в данном разделе:

- а. Удалить ненужные дашборды, для этого нажмите “х” в правом верхнем углу дашборда.

Обратите внимание! Вы можете добавить максимум три дашборда в данный раздел проекта. После удаления хотя бы одного Дашборда активируется кнопка добавления новой визуализации:



в. Добавить новый дашборд. Для этого нажмите на кнопку «Добавить дашборд» и в выпадающем списке выберите нужный тип визуализации:

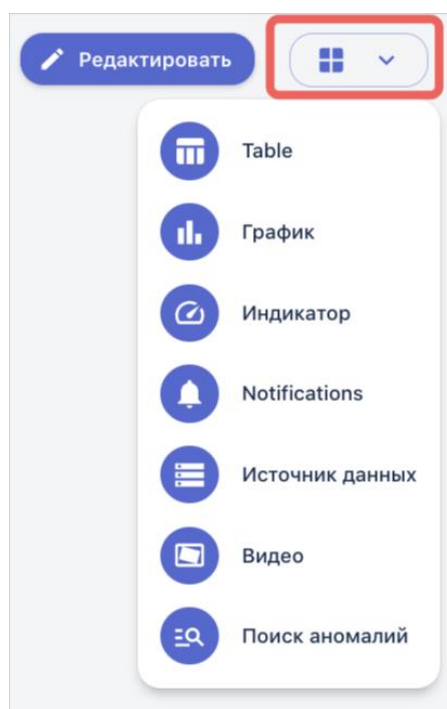



Рисунок 253 – Выбор типа визуализации

После этого под описание проекта добавится новый пустой дашборд выбранного типа

с. Выбрать коннектор на отображения информации на дашборде. Для этого нажмите на кнопку коннектор в правом углу дашборда (). Отобразится список всех созданных коннекторов в системе соответствующего типа. Выберите необходимый коннектор из списка, кликнув на его наименование:

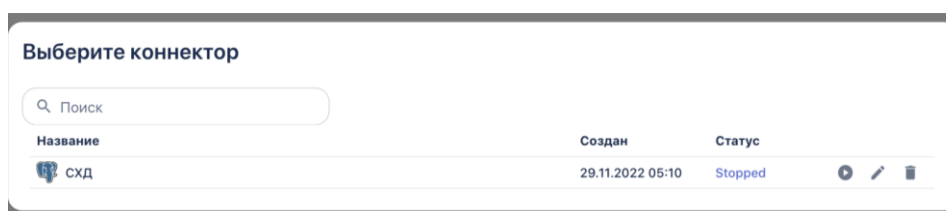


Рисунок 254 – Выбор коннектора

После этого визуализация отобразится на дашборде.

4. **Добавление сущности «Модель».** На вкладке «Модели» пользователь нажимает кнопку «Добавить», и открывается выпадающий список со всеми моделями ИИ, созданными в Системе. Выбирается модель и добавляется в проект.

Модель доступна для скачивания по кнопке  :

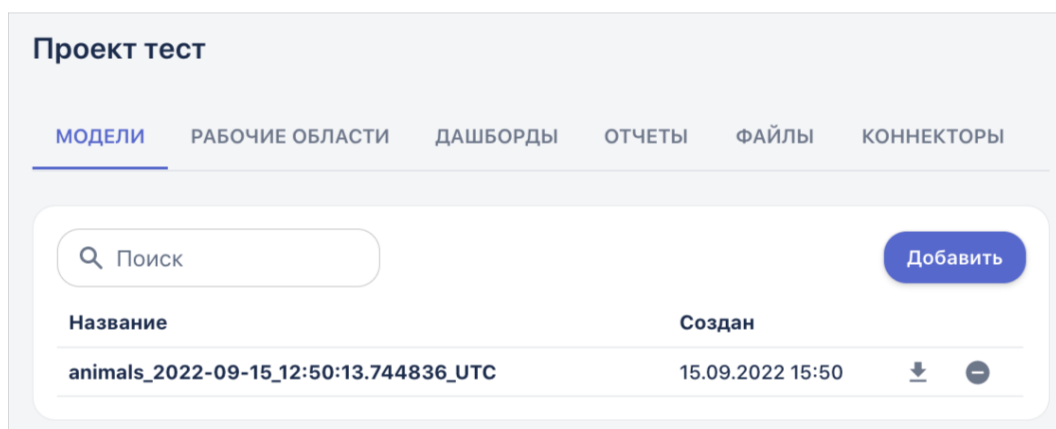


Рисунок 255 – Загруженная в проект модель

5. **Добавление сущности «Рабочая область».** Перейдите на вкладку «Рабочие области» и нажмите кнопку «Добавить». Из выпадающего списка выберите рабочую область для добавления в проект:

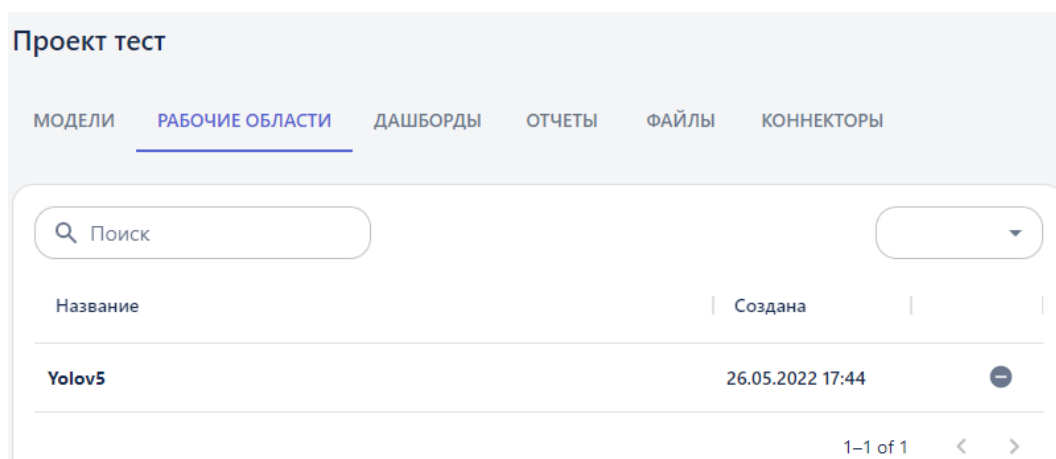


Рисунок 256 – Вкладка «Рабочие области»

6. **Добавление сущности «Дашборд».** Перейдите на вкладку «Дашборды» и нажмите кнопку «Добавить», после чего выберите дашборд из выпадающего списка:

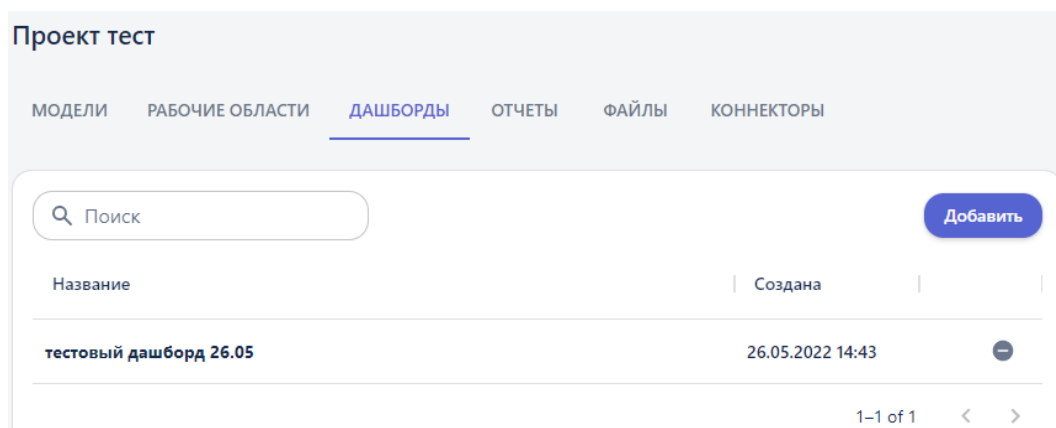


Рисунок 257 – Вкладка «Дашборды»

7. **Добавление сущности «Отчет».** Перейдите на вкладку «Отчеты» и нажмите кнопку «Добавить». Из выпадающего списка выберите отчет для добавления в проект.

8. **Добавление сущности «Файл».** Перейдите на вкладку «Файлы» и нажмите кнопку «Добавить». Из выпадающего списка выберите файл для добавления в проект:

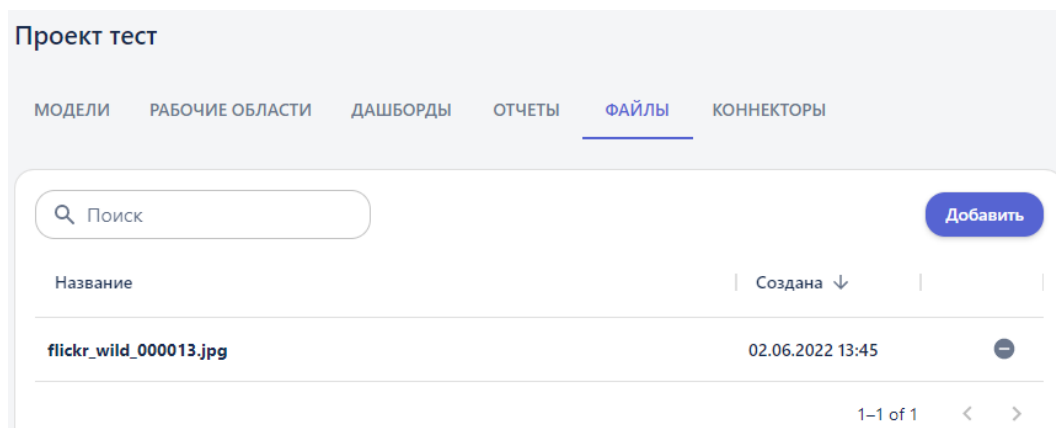


Рисунок 258 – Вкладка «Файлы»

9. **Добавление сущности «Коннектор».** Перейдите на вкладку «Коннекторы» и нажмите кнопку «Добавить». Из выпадающего списка выберите коннектор для добавления в проект:

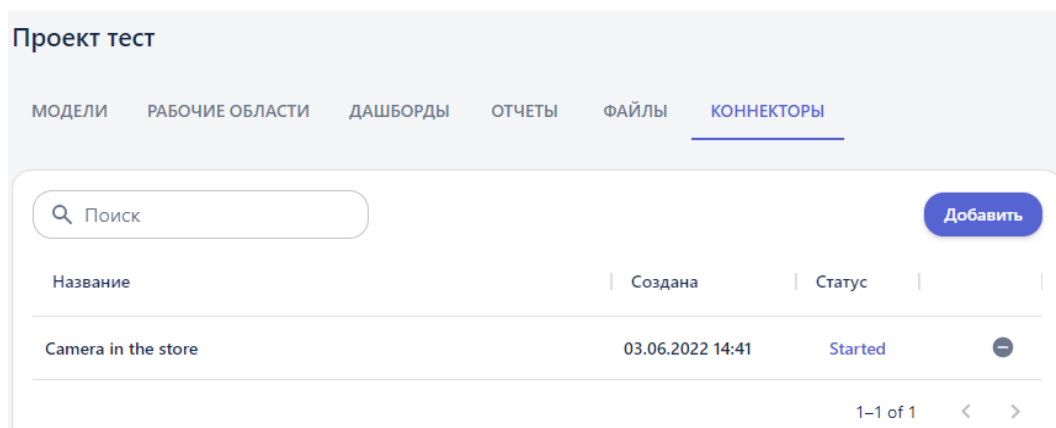


Рисунок 259 – Вкладка «Коннекторы»

2.3.12. Настройка подключения к источникам данных

В BAUM STORAGE AI v2 реализована возможность подключения к *внешним системам*, выступающим в качестве *источников данных*. При этом данные в режиме реального времени могут поступать из следующих источников: БД (ClickHouse, PostgreSQL, MongoDB) и камеры видеонаблюдения. В данном разделе рассказывается, как настроить в Системе такие подключения. Подробно о визуализации информации из коннекторов на дашбордах написано в разделе «Работа с дашбордами»

Все подключения настраиваются в пункте меню «Соединения», где создаются сущности «Коннектор». *Коннекторы* объединяют в себе источник подключения и запрос на получение данных из него. В данном разделе описаны все типы коннекторов и сценарии работы с ними на примерах с тестовыми данными.

2.3.12.1. Типы коннекторов

1. «ClickHouse»

Данный коннектор предназначен для подключения к БД «ClickHouse». Настройка подключения описана выше в разделе Настройка подключения на примере ClickHouse.

2. «PostgreSQL»

Данный коннектор предназначен для подключения к БД «PostgreSQL». Настраивается по аналогии с коннектором «ClickHouse».

3. «Mongo»

Данный коннектор предназначен для подключения к БД «MongoDB». Настраивается по аналогии с коннектором «ClickHouse».

4. «Table_app»

На вход коннектора поступают *табличные данные* в онлайн режиме, например из БД «PostgreSQL». Чтобы анализировать входные данные используется обученная в системе *модель*. Рассмотрим коннектор для задачи прогнозирования лесных пожаров, которая на вход принимает данные о погодных условиях в онлайн-режиме (эти табличные данные аналогичны тем, на которых обучалась модель):

The screenshot shows a configuration form for a connector. It has two main sections: '1 Основная информация' and '2 Параметры'. In the first section, there is a 'Название' field with the value 'Fire_forecast app' and an empty 'Описание' field. The second section contains several dropdown menus: 'Источник данных' (set to 'ETL'), 'Тип коннектора' (set to 'table_app'), 'Коннектор' (set to 'Fire_forecast dat...'), and 'Модель' (set to 'fires_2022-06-13_11:44:11.614557.UTC'). There is also a 'Медиафайлы' dropdown menu and a checked checkbox for 'Постоянное обновление'.

Рисунок 260 – Настройка коннектора «Table_app»

В поле «Коннектор» выбирается *коннектор*, который подключен к таблице с погодными условиями. А в поле «Модель» выбирается обученная модель прогнозирования лесных пожаров. Устанавливается галочка в поле «Постоянное обновление» – активирование режима ожидания новых данных на входе настраиваемого коннектора.

5. «Save_table»

Данный коннектор предназначен для сохранения в Системе в виде файлов (на данный момент реализовано сохранение файлов в формате csv) табличных данных, поступающих из сторонних систем. Директория для сохранения файлов в Системе – это раздел «Данные».

В поле «Коннектор» указывается коннектор для подключения к таблице внешнего источника. Устанавливается галочка в поле «Постоянное обновление».

Для такого типа коннектора обязательно указать на выбор:

- Количество строк данных - количество строк из таблицы, которые будут сохранены.
- Интервал - промежуток времени в секундах, в течение которого коннектор должен собирать информацию из таблицы, по истечению этого времени загрузка прекратится и файл будет сохранен.

Рисунок 261 – Настройка коннектора «save_table»

6. «Detection_app»

Данный коннектор используется в задаче распознавания объектов (Object detection) на видеопотоке данных в онлайн-режиме. Пример настройки коннектора:

Рисунок 262 – Настройка коннектора «Detection_app»

В поле «Коннектор» выбирается камера для подключения. А в поле «Модель» выбирается обученная модель распознавания изображений, которая умеет распознавать определенные объекты на изображениях/видео. При этом модель умеет определять именно те объекты, распознавать

которые она обучалась, и предполагается, что на видео с камеры также будут эти объекты. Устанавливается галочка в поле «Постоянное обновление», для получения видео с камеры в режиме реального времени.

Если не устанавливать галочку в поле «Постоянное обновление», то выполняется автоматическое разбиение полученного видеопотока на эпизоды (*раскадровка*). Такие эпизоды партициями передаются модулю «Apache kafka», который в Системе отвечает за передачу данных. Внешние системы могут подключиться к брокеру сообщений, и получить доступ к раскадрованному видео – на просмотр и обработку.

7. «Images_detection_app»

Данный коннектор также используется в задаче распознавания объектов, только входными данными для анализа являются на выбор – готовый видеофайл, или серия из нескольких изображений. Пример настройки коннектора:

The screenshot shows a configuration form for the 'Images_detection_app' connector. It is divided into two main sections: '1 Основная информация' (Basic Information) and '2 Параметры' (Parameters). In the 'Основная информация' section, there is a 'Название' (Name) field with the value 'Detection_dogs_video' and an empty 'Описание' (Description) field. In the 'Параметры' section, there are several dropdown menus: 'Источник данных' (Data Source) is set to 'ETL', 'Тип коннектора' (Connector Type) is 'images_detectio...', 'Модель' (Model) is 'YOLOV5_DEMO_2022-06-13_11:36:28.225911_UTC', and 'Медиафайлы' (Media Files) is 'tovarischa.mp4'. At the bottom of the 'Параметры' section, there is a checkbox labeled 'Постоянное обновление' (Continuous Update) which is currently unchecked.

Рисунок 263 – Настройка коннектора «Images_detection_app»

В поле «Модель» выбирается обученная модель «Yolov5». В поле «Медиафайлы» выбирается либо один видео файл, загруженный в раздел «Данные», либо несколько изображений. В поле «Постоянное обновление» галочка не устанавливается. После запуска коннектора с данным типом автоматически выполняется сохранение размеченного видеофайла/серии изображений в разделе «Данные».

***Важно:** Чтобы модель могла распознавать объекты, обучение должно пройти минимум на трехстах эпохах.

8. «Classification_app»

Данный коннектор используется в задаче *классификации изображений*, где на вход коннектора для анализа подается серия из нескольких изображений. Пример настройки коннектора:

The screenshot shows a web interface titled "Редактировать коннектор" (Edit connector) with a "Назад" (Back) button. It is divided into two sections: "1 Основная информация" (Main information) and "2 Параметры" (Parameters). In the "Основная информация" section, there is a "Название" (Name) field containing "Classification" and an empty "Описание" (Description) field. The "Параметры" section includes dropdown menus for "Источник данных" (Data source) set to "ETL", "Тип коннектора" (Connector type) set to "classification_app", and "Коннектор" (Connector) set to "Коннектор". The "Модель" (Model) dropdown is set to "ANIMALS_DEMO_2022-06-13_10:45:26.270270.UTC". Under "Медиафайлы" (Media files), there are six image thumbnails: flickr_cat_000057.jpg, flickr_cat_000058.jpg, flickr_cat_000055.jpg, flickr_dog_000057.jpg, flickr_dog_000055.jpg, and flickr_dog_000056.jpg. At the bottom, there is an unchecked checkbox for "Постоянное обновление" (Constant update) and a blue "Сохранить" (Save) button.

Рисунок 264 – Настройка коннектора «Classification_app»

В поле «Модель» выбирается обученная модель классификации изображений. В поле «Медиафайлы» – серия изображений, которые необходимо классифицировать с использованием обученной модели. *Для данного типа коннектора не нужно устанавливать признак «Постоянное обновление», так как анализируются данные, загружаемые с локального устройства.

9. «Constructor» (автоматически создаваемый)

Условием создания коннектора является следующее: создается блок-схема, где один из элементов имеет на выходе *визуализацию* – выходным параметром элемента является *таблица*, *график* и т.д. Пользователь запускает такую блок-схему, и после успешной отработки элемента с визуализацией в Системе создается коннектор. Число создаваемых коннекторов при запуске блок-схемы соответствует числу элементов с визуализацией на этой блок-схеме.

Название коннектора формируется из названия элемента и названия рабочей области, и должно являться уникальным в рамках Системы.

Пример – *yolov5_train_(yolov5_noses_eyes)*, где название рабочей области указано в скобках.

Такой коннектор автоматически создается в статусе «Started» – пользователь не должен запускать коннектор, и может сразу же перейти к просмотру данных коннектора в окне дашборда.

2.3.12.2. Порядок работы с коннекторами

Предварительно коннектор должен быть создан и запущен (за исключением служебных коннекторов, которые создаются и запускаются автоматически). Только после этого выполняется подключение к нему через окно дашборда.

2.3.12.2.1. Создание коннектора

Создание коннекторов осуществляется в разделе меню «Соединения». Для коннекторов, в которых настраивается подключение к внешним источникам данных (к внешним базам данных, к камере видеонаблюдения), дополнительно создаются сущности – *источник данных*, и *ETL*. Для остальных типов, эти сущности не создаются, а сразу создается сущность «Коннектор».

Для того чтобы создать новый коннектор, перейдите на вкладку «Коннекторы» и нажмите на кнопку «Создать коннектор»:

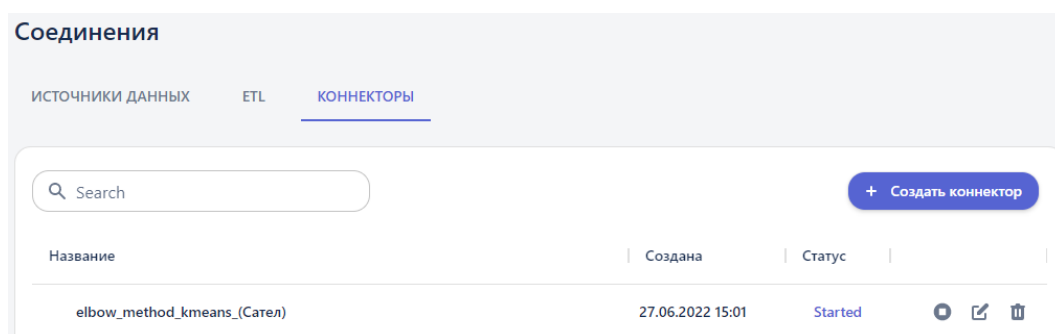


Рисунок 265 – Переход к созданию коннектора на вкладке «Коннекторы»

В открывшейся форме выберите тип коннектора в одноименном поле «Тип коннектора», и заполните поля:

Создать новый коннектор ← Назад

1 Основная информация

Название

Описание

2 Параметры

Источник данных

Тип коннектора

Модель

Медиафайлы

Постоянное обновление

Создать

Рисунок 266 – Форма создания нового коннектора

После заполнения формы нажмите кнопку «Создать».

2.3.12.2.2. Запуск коннектора

Сразу после создания коннектору присваивается статус «Stopped». Далее коннектор запускается для того, чтобы активировать его работу (начать получать данные из внешних источников, запустить алгоритм обработки данных). Для этого нажмите кнопку «▶» в строке с коннектором:

Соединения

ИСТОЧНИКИ ДАННЫХ ETL КОННЕКТОРЫ


Search + Создать коннектор

Название	Создана	Статус	
validation_(TEXT)	07.06.2022 16:58	Started	▶ ✎ 🗑
validation_(FIRE)	07.06.2022 13:54	Started	▶ ✎ 🗑
xgb_classifier_(FIRE)	07.06.2022 13:54	Stopped	▶ ✎ 🗑

Рисунок 267 – Запуск коннектора

В результате коннектору присваивается статус «Started», и он готов к визуализации на дашборде.

2.3.12.2.3. Подключение к коннектору на дашборде

Для подключения коннектора на дашборде нажимается кнопка DASH, чтобы добавить интерактивный блок на рабочую область. После добавления на интерактивного блоке нужно нажать  в правом углу.

Откроется модальное окно со списком коннекторов для подключения.

2.3.12.3. Настройка подключения на примере ClickHouse

Действия:

1. Перейдите в пункт меню «Соединения»:

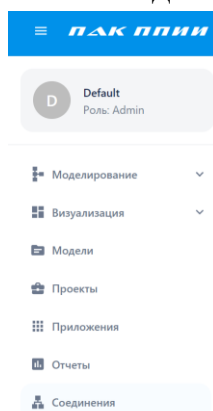


Рисунок 268 – Пункт меню Соединения

Откроется страница «Соединения» на первой вкладке «Источники данных», на которой отображаются все ранее созданные источники:

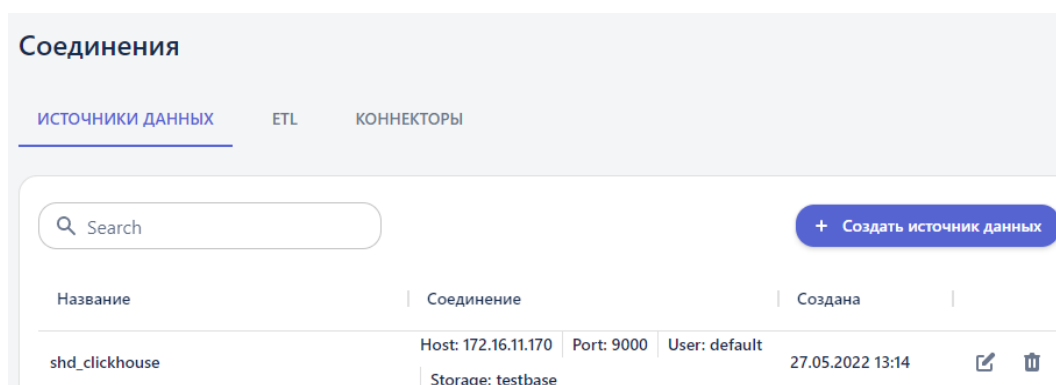


Рисунок 269 – Вкладка «Источники данных»

2. Создание нового источника. Нажмите на кнопку «Создать источник данных». Откроется окно «Создание нового источника данных»:

Создание нового источника данных

← Назад

1 Основная Информация

2 Параметры

Название

Хост

Порт

Описание

Имя хранилища

Тип хранилища...

Имя пользователя

Пароль

Создать

Рисунок 270 – Окно настройки источника данных

Заполните поля:

→ *Название.* Пользователь задает название источника «TEST DATA SOURCE (clickhouse)», к которому будет настраиваться подключение.

→ *Хост.* Указывается хост протокола TCP/IP, т.е. сетевой интерфейс устройства, предоставляющего сервис формата «клиент-сервер», где сервером выступает БД **ClickHouse**, а клиентом – **BAUM STORAGE AI v2**. По сути это IP-адрес подключаемой БД. Необходимо указать «172.16.11.116».

→ *Порт* – номер порта, по которому устанавливается соединение с сервером, на котором установлена БД **ClickHouse**. Указать «12366».

→ *Имя хранилища* – название базы данных, которое указано на подключаемом сервере. Указать «default».

→ *Тип хранилища.* Из выпадающего списка необходимо выбрать тип «clickhouse»:

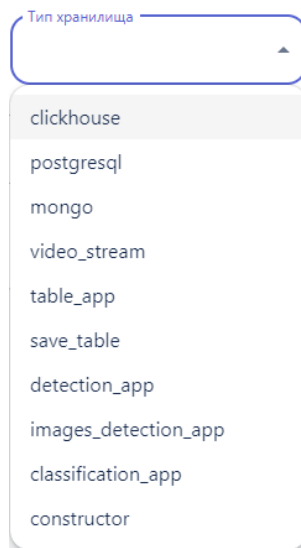


Рисунок 271 – Список возможных типов хранилища

Описание всех типов коннекторов представлено в разделе Классификация коннекторов

→ *Имя пользователя, пароль* – параметры учетной записи администратора внешнего сервера для разрешения доступа к данным. Указать пользователя «clickhouse_operator», и пароль для него «clickhouse_operator_password».

→ *Описание.* Вводится дополнительная информация по источнику, необязательное поле.

Для регистрации в Системе источника нажмите кнопку «Создать».

3. Создание нового ETL. Сущность «ETL» (дословно Extract, Transform, Load – с англ. извлечение, преобразование загрузка) содержит в себе sql запрос для извлечения данных из источника. То есть в шаге 2 создается источник и прописывается запрос для извлечения данных из него.

3.1. Перейдите на вкладку «ETL»:

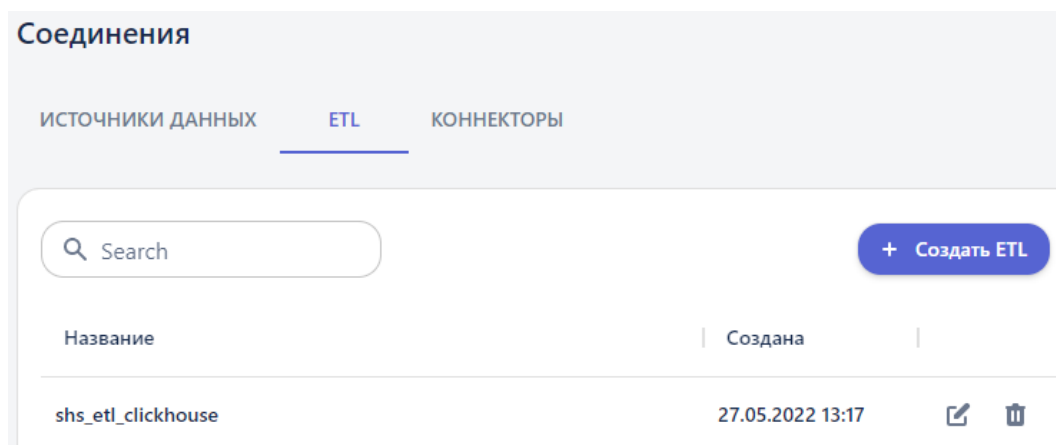


Рисунок 272 – Вкладка ETL

3.2. Нажмите на кнопку «Создать ETL». Откроется окно «Создать новый ETL»:

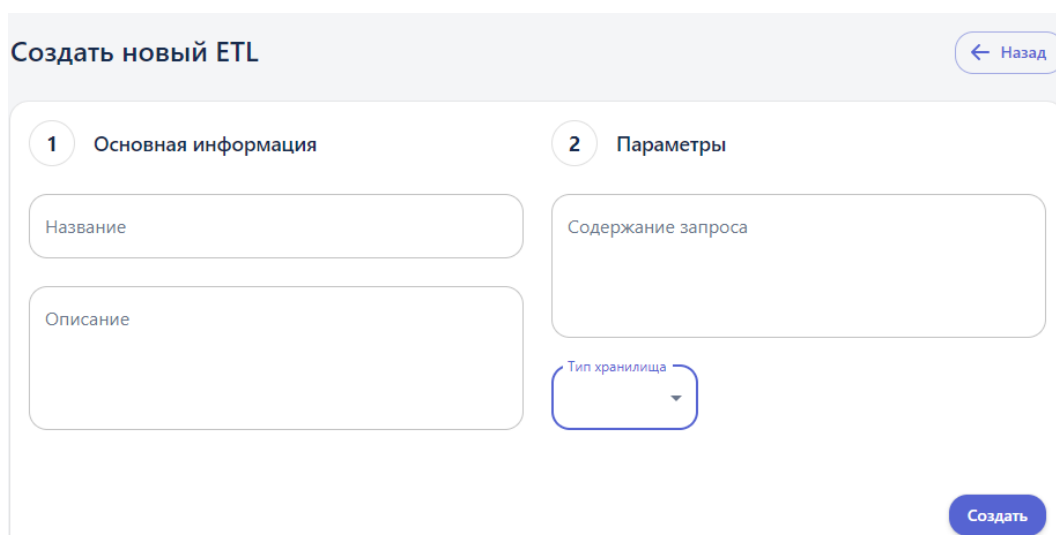


Рисунок 273 – Окно настройки ETL

3.3. Заполните поля:

→ *Название.* Пользователь вручную задает название ETL «TEST ETL (clickhouse)» – запрос на извлечение данных.

→ *Содержание запроса.* Прописывается непосредственно sql запрос для извлечения данных из внешнего сервера. При этом указывается название таблицы, из которой данные извлекаются, в запросе «SELECT * FROM stock». Чтобы извлечь данные только из первых ста строк этой таблицы используется запрос «SELECT * FROM stock LIMIT 100».

→ *Тип хранилища.* Выбирается тип «clickhouse».

→ *Описание* (необязательное поле).

3.4. Нажмите кнопку «Создать».

4. Создание нового коннектора:

4.1. Перейдите на вкладку «Коннекторы»:

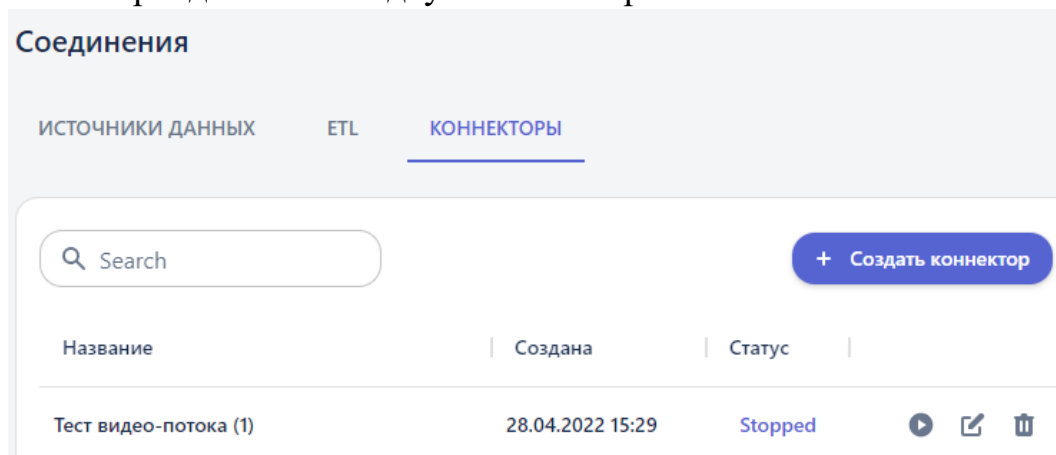


Рисунок 274 – Вкладка коннекторов

4.2. Нажмите на кнопку «Создать коннектор». Откроется окно «Создать новый коннектор»:

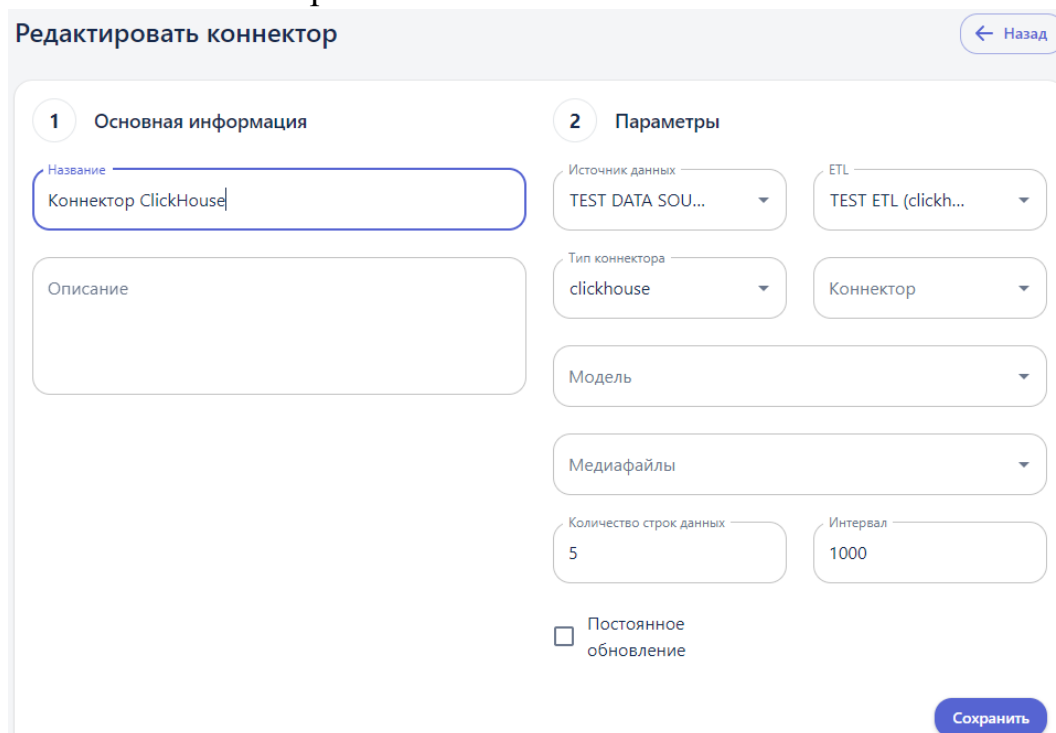


Рисунок 275 – Окно настройки коннектора

4.3. Заполните поля:

→ *Название.* Пользователь вручную задает название создаваемого коннектора «Коннектор ClickHouse».

→ *Источник данных.* Из списка выбирается источник «TEST DATA SOURCE (clickhouse)», созданный в шаге 2.

→ *ETL*. Из списка выбирается ETL «TEST ETL (clickhouse)», созданный в шаге 3.

→ *Тип коннектора*. По умолчанию выбирается первый тип из списка – clickhouse, оставить выбранное значение.

→ *Количество строк данных*. Данные из внешней БД поступают порциями, по указанному или меньшему количеству строк за раз.

→ *Интервал* – периодичность, с которой выполняются запросы во внешнюю БД. Заполняется числовым значением (в миллисекундах), или значением в формате «Дата» (*второй вариант не реализован в текущей версии*). Если задать интервал 1000 мс, и указать количество строк пять, каждую секунду будет запрашиваться пять записей.

→ *Постоянное обновление*. Признак устанавливается, когда данные ожидаются бесконечно. Если признак не установить, запрос данных завершится, при получении их в полном объеме.

→ *Описание*.

Остальные поля на форме создания коннектора для текущего сценария не заполняются, они используются при создании других типов коннекторов. Нажмите кнопку «Создать».

4.4. Сразу после создания коннектору присваивается статус «Stopped»:

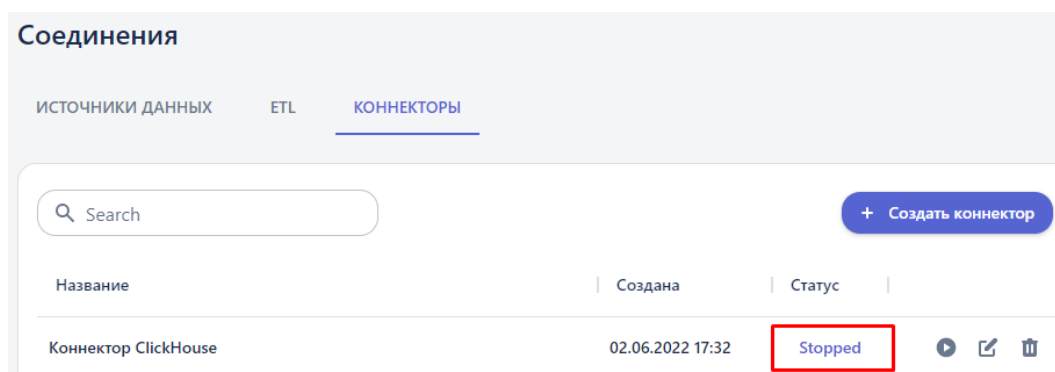


Рисунок 276 – Статус запуска коннектора

5. **Запуск коннектора.** Чтобы данные из источника начали поступать в Систему необходимо запустить коннектор. Для этого нажмите на кнопку «▶» в строке с коннектором. В результате статус меняется на значение «Started».

Вы так же можете отображать табличные данные ClickHouse в режиме реального времени, используя для визуализации сущность «Дашборд» и подключаясь к запущенному коннектору.

2.3.12.4. Получение данных с камеры видеонаблюдения

Подключение к камере наблюдения настраивается по аналогии с подключением к БД ClickHouse. Отличие заключается в выборе канала подключения, а сам порядок действий повторяется:

1. **Создание нового источника.** На вкладке «Источники данных» нажмите кнопку «Создать источник данных». Заполните поля:

Создание нового ресурса данных

← Назад

1 Основная Информация

2 Параметры

Название

Хост

Порт

Описание

Имя хранилища

Тип хранилища
video_str...

Имя пользователя

Пароль

Создать

Рисунок 277 – Создание источника данных типа video_stream

→ *Название.* Пользователь задает название источника «Камера магазина».

→ *Хост.* Пользователь вводит сетевой адрес камеры (в данном примере: «http://158.58.130.148:80/mjpg/video.mjpg»). Все сведения, вводимые начиная с этого поля, необходимо уточнять непосредственно у владельцев смежной системы, с которой настраивается соединение.

→ *Порт.* Пользователь вводит номер порта для подключения к камере «0» (обязательно числовое значение).

→ *Имя хранилища.* Пользователь вводит название камеры «pass».

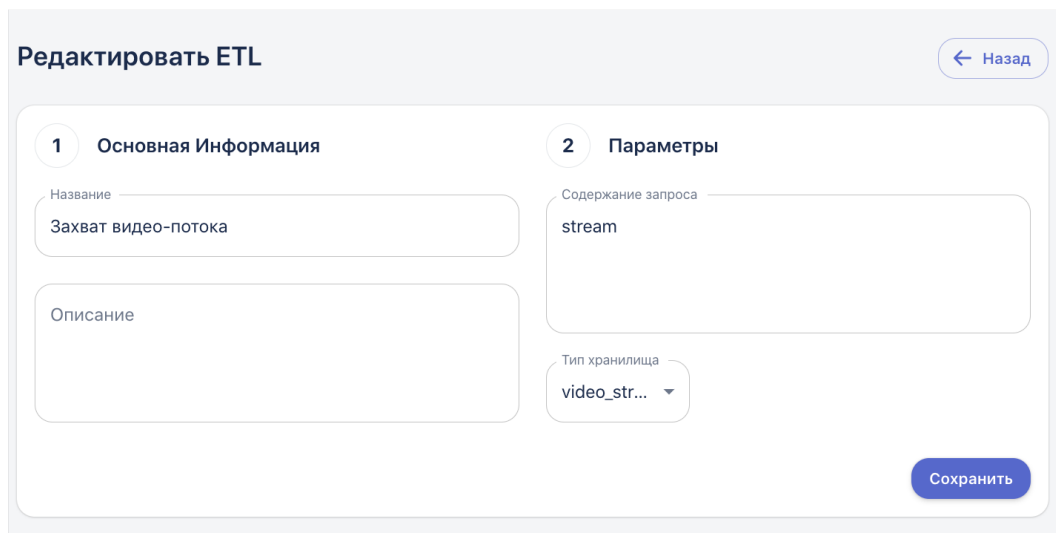
→ *Тип хранилища.* Из выпадающего списка выбрать тип «video_stream».

→ *Имя пользователя, пароль.* Пользователь вводит параметры учетной записи администратора, имеющего доступ к видеокамере. Указываются соответственно пользователь «pass» и пароль «pass».

→ *Описание.*

Для регистрации в Системе источника нажмите кнопку «Создать» (далее действие по сохранению введенных настроек предполагаются по умолчанию).

2. Создание ETL. На вкладке ETL нажмите на кнопку «Создать ETL» и заполните поля следующим образом:



The screenshot shows a web interface for editing an ETL connector. The title is 'Редактировать ETL' with a 'Назад' button. There are two tabs: '1 Основная Информация' and '2 Параметры'. Under '1 Основная Информация', there is a 'Название' field containing 'Захват видео-потока' and an empty 'Описание' field. Under '2 Параметры', there is a 'Содержание запроса' field containing 'stream' and a 'Тип хранилища' dropdown menu with 'video_str...' selected. A 'Сохранить' button is located at the bottom right of the form.

Рисунок 278 – Параметры ETL типа video stream

В случае если коннектор создаётся для захвата видеопотока (его отображения в режиме он-лайн на дашборде), тогда в поле «Содержание запроса» укажите «stream». Для коннекторов, которые подразумевают сохранение видео, запрос должен содержать «save».

3. Создание нового коннектора. Перейдите на вкладку «Коннекторы», нажать на кнопку «Создать коннектор» и в открывшемся окне заполните поля:

Рисунок 279 – Параметры коннектора типа video stream

→ *Название.* Пользователь задает название создаваемого коннектора «Camera in the store».

→ *Источник данных.* Из списка выбирается источник, созданный в шаге 1.

→ *ETL.* Из списка выбирается ETL «Захват видео-потока» – запрос на получение онлайн видео потока данных с камеры (запрос «stream»).

→ Для коннектора с типом «video_stream» доступен для выбора еще один ETL «Сохранение видеопотока» – запрос на сохранение полученных данных с камеры в раздел «Данные» → «Видео» (запрос «save»). Убедиться, что данный тип ETL доступен для выбора. Продолжение сценария, как сохранить видеопоток, см. в разделе 4.13.

→ *Тип коннектора.* Выбирается тип «video_stream».

→ *Описание* (необязательное поле).

Остальные поля на форме для данного типа коннектора остаются незаполненными.

4. **Запуск коннектора.** На вкладке «Коннекторы» в строке с коннектором «Camera in the store» нажмите кнопку запуска «▶». Коннектору присваивается статус «Started» – с этого момента в Систему начинают поступать данные из внешнего источника:

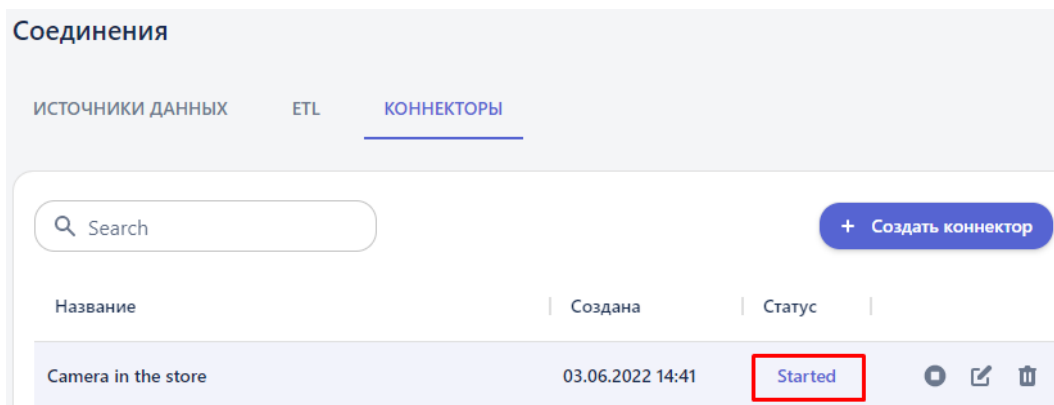


Рисунок 280 – Статус запуска коннектора

Для остановки коннектора в строке с ним нажмите кнопку «⏸️».

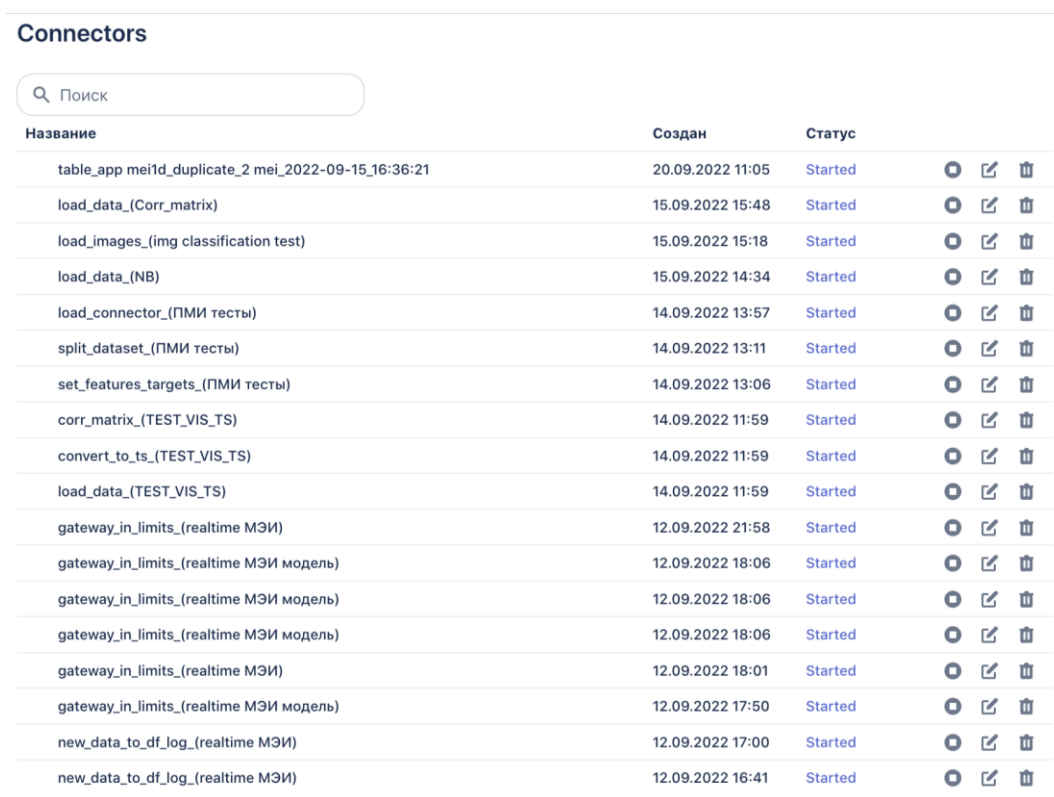


Рисунок 281 – Выбор коннектора для отображения в окне дашборда

Для выбора и подключения коннектора он выбирается из списка. Данные, получаемые от коннектора, отобразятся на интерактивном блоке после подключения:

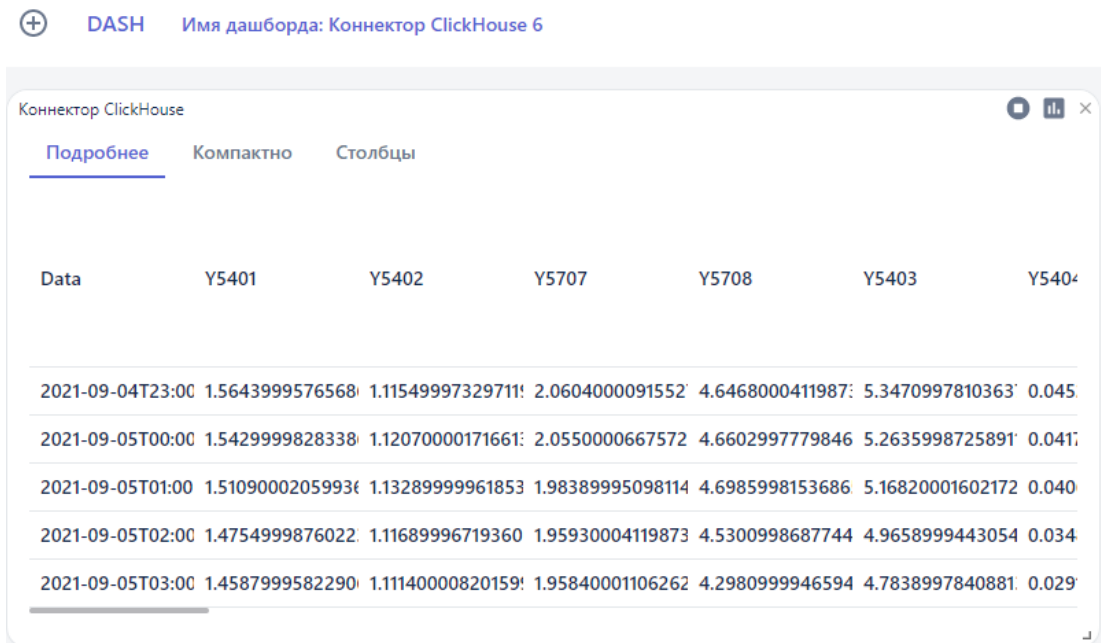


Рисунок 282 – Дашборд с запущенным коннектором

2.3.13. Примеры работы с VAUM STORAGE AI v2

2.3.13.1. Обучение модели прогнозирования температуры воды и газов в котле

Загрузка входных данных:

1. В левой части главного окна на панели вкладок Системы откройте вкладку «Данные».
2. На открывшейся странице нажмите кнопку
3. В открывшемся окне в качестве Типа выберите «Категория», в поле Название введите вручную «МЭИ» и нажмите на кнопку «Создать»:

Создать категорию, разметку и т.д.

Тип
Категория

Название
МЭИ

Создать

Рисунок 283 - Создание новой папки в разделе Данные

4. В разделе «Данные» появится папка «МЭИ», для загрузки файлов перейдите в неё и нажмите на кнопку «Загрузить» на верхней панели:

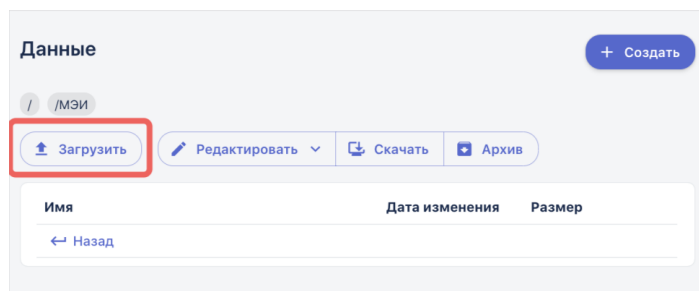


Рисунок 284 - Загрузка файлов в папку

5. В открывшемся окне нажмите на кнопку «Выбрать файлы» и укажите путь к заранее подготовленному файлу **med1d.csv**, в котором содержатся данные о температуре газов и воды в котле за определенный промежуток времени. Второй вариант – перенести файлы в этот раздел по технологии «drag n drop».

Выбранные файлы отобразятся в нижней части окна загрузки:

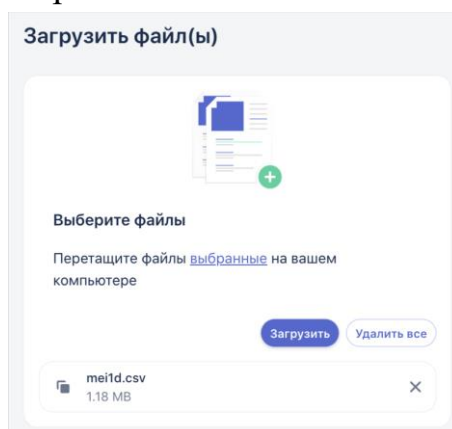


Рисунок 285 – Отображение выбранного файла

6. Нажмите на кнопку «Загрузить». Файл с входными данными отобразится в папке:

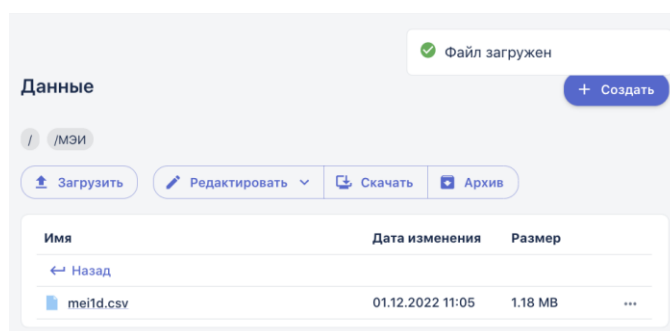



Рисунок 286 – Загрузка выбранного файла

Создание новой рабочей области

Полностью блок схема представлена в [Таблице 16.7.](#)

7. Перейдите в пункт меню системы **Моделирование** → **Рабочая область**. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка ):

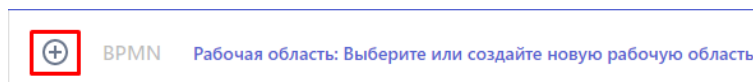


Рисунок 287 - Создание новой рабочей области

8. В открывшейся форме введите название новой рабочей области «МЭИ» и нажмите кнопку «Создать»:

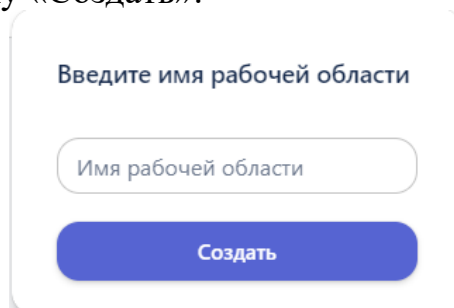


Рисунок 288 - Ввод имени рабочей области

9. На панели инструментов отобразится название созданной рабочей области.

Добавление элемента «Запуск»:

1. На панели инструментов блок-схемы нажмите кнопку «Добавить элемент» (кнопка **BPMN**)

2. В открывшейся библиотеке графических элементов выберите элемент «Запуск» :

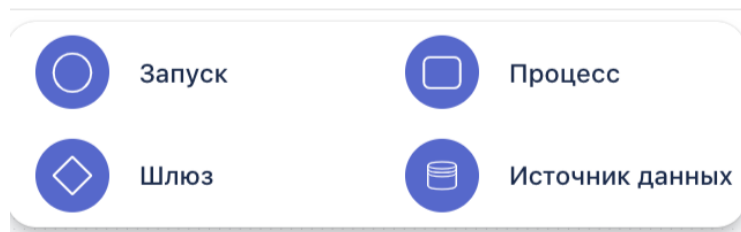


Рисунок 289 - Возможные элементы блок схемы

3. На рабочую область добавится элемент «Запуск»:

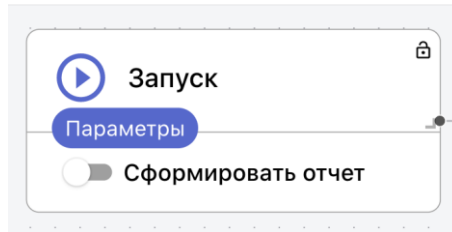


Рисунок 290 - Блок Запуск

Добавление и настройка элемента «Источник данных».

1. Добавьте на рабочую область элемент «Источник данных»:

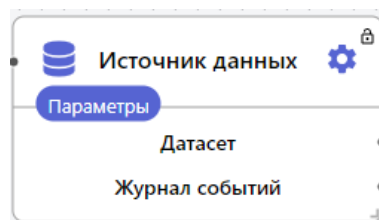



Рисунок 291 - Блок Источник данных

2. **Открытие настроек элемента.** На элементе «Источник данных» нажмите на кнопку . При настройке текущего элемента он выделяется синей рамкой. Справа откроется панель настроек элемента, где будут отображаться созданные в разделе папки и файлы с табличными данными.

3. **Выбор данных для загрузки в блок-схему.** Для того чтобы найти нужный файл, кликните на папку и перейдите в нее, выберите из списка файл, загруженный в Систему в шаге 1 «me1d.csv», нажмите на три точки в строке с названием файла и кликните «Выбрать». Внизу отобразится название выбранного файла:

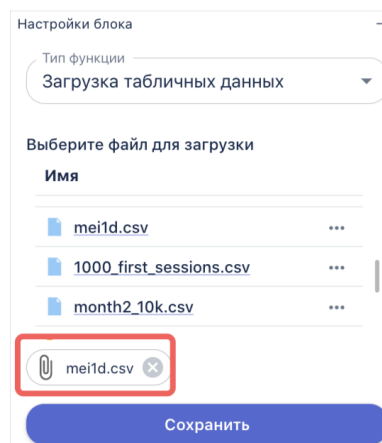
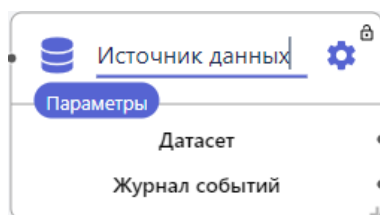


Рисунок 292 - Выбор файла

4. **Сохранение настроек элемента.** На панели настроек элемента нажмите на кнопку «Сохранить» (далее сохранение настроек элемента предполагается по умолчанию).

5. **Ввод названия элемента.** Чтобы задать название элемента нужно дважды щелкнуть левой кнопкой мыши на название элемента в рабочей области, и ввести нужное название в поле с названием, доступным для редактирования:



6. **Установка соединений.** Соедините выходную точку элемента «Запуск» с входной точкой элемента «Источник данных» с помощью левой кнопки мыши:

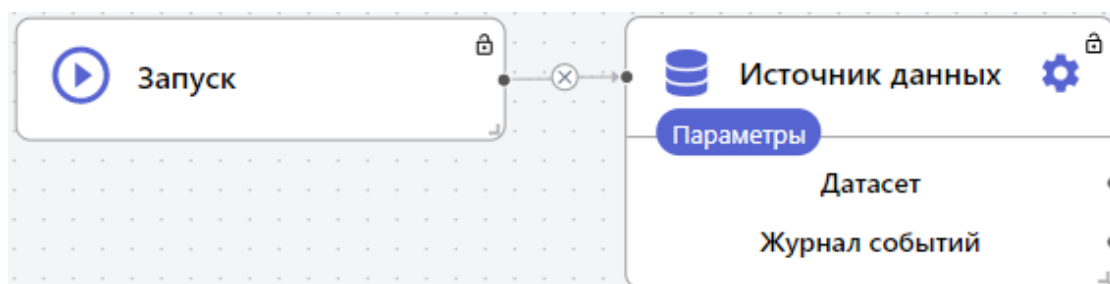


Рисунок 15.1.11 - Соединение элементов Запуск и Источник данных

Добавление и настройка элемента «Процесс». Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

1. На панели свойств элемента выбрать из списка функцию: тип функции «Загрузка данных» -> функция «Преобразование данных во временной ряд».

В разделе «Параметры» отобразятся поля:

Рисунок 293 - Параметры функции «Преобразование данных во временной ряд»

2. В поле «Шаг ресемплирования» введите 5
3. В поле «Частота ресемплирования» выберите 7.Секунды
4. В поле «Агрегирующая функция» выберите 7.Медиана
5. В поле «Столбец с временной меткой» введите datetime

При помощи функции «Преобразование данных во временной ряд» можно привести данные к другой дискретности/периодичности, например в нашем случае - 5 секунд, а в качестве значения выбирается медиана - значение посередине.

6. На панели настроек элемента нажмите на кнопку «Сохранить».
7. Измените название элемента на «во временной ряд».
8. Соедините с элементы:

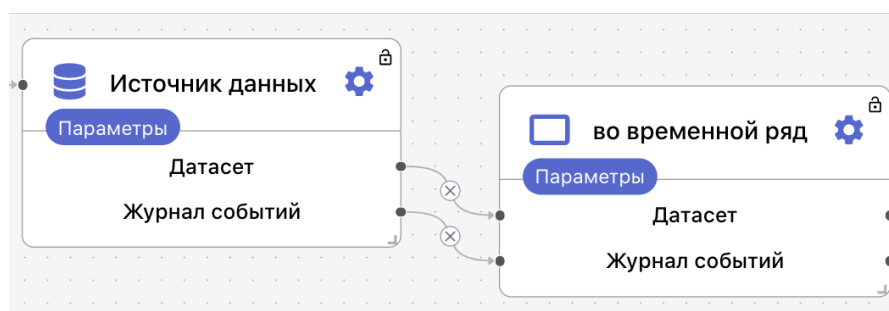


Рисунок 294 - Соединение элементов Источник данных и Во временной ряд

Срез временного ряда. Добавьте на рабочую область элемент «Процесс» и настройте его:

1. Выберите из списка функцию: раздел «Предобработка данных» -> функция «Срез временного ряда по индексу»:

2. В разделе «Параметры» в поле «Дата начала» введите значение 2020-12-16 11:00:00, а в поле «Дата окончания» - 2020-12-16 15:00.
3. Измените название элемента на «срез времен ряда».
4. Соедините элементы:

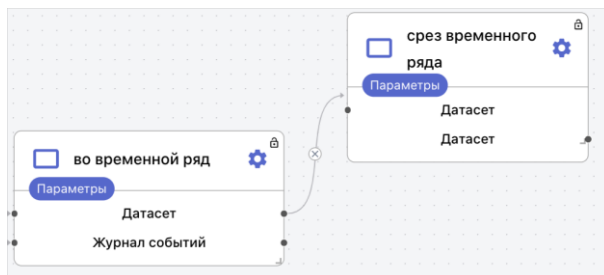


Рисунок 295 - Соединение элементов Во временной ряд и Срез временного ряда

Выбор признаков и целевых показателей. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Анализ данных» -> функция «Выбор признаков и целевых признаков»
2. В разделе «Параметры» -> в поле «Признаки» оставьте пустым, а в поле «Целевые признаки» вы можете либо поочередно ввести сначала признак Tq и нажать Enter, затем Tw и нажать Enter или вы можете найти в списке файлов me1d.csv, нажать на три точки рядом с его названием и выбрать «Выгрузить признаки», тогда в поле автоматически подтянутся все признаки, который есть в файле. Вам необходимо удалить ненужные и оставить только Tq и Tw.

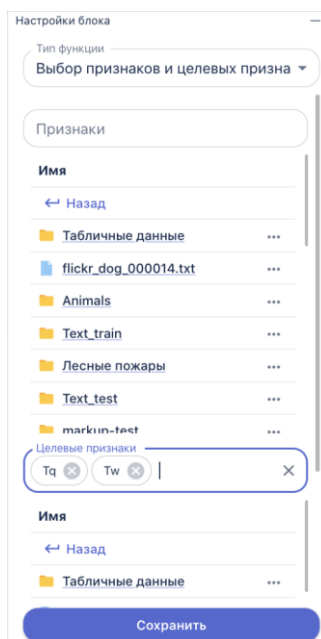


Рисунок 296 - Параметры функции «Выбор признаков и целевых признаков»

3. Измените название элемента на «Установить признаки».
4. Соедините элементы:

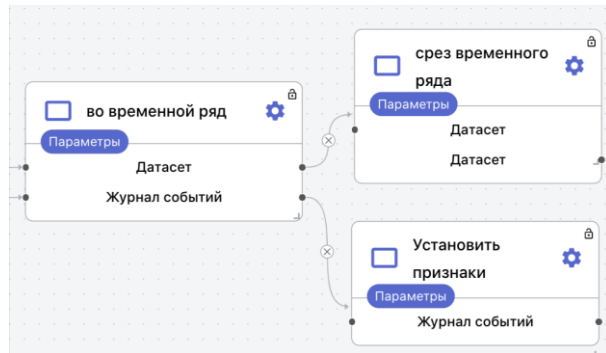


Рисунок 297 - Соединение элементов Во временной ряд Установить признаки

Разделение датасета на обучающую и тестовую выборки. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Машинное обучение» - > функция «Разделение датасета на обучающую и тестовую выборки».
2. Укажите долю тестовой выборки – 0.2. Так 80% данных будут использованы для обучения модели, и 20% – для тестирования.
3. Не нужно ставить галочки в полях «Перемешивать наблюдения перед разделением» и «Разделять с учетом меток классов» (параметр активируется только для решения задач классификации)
4. Измените название элемента на «Сплит».
5. Соедините элементы:

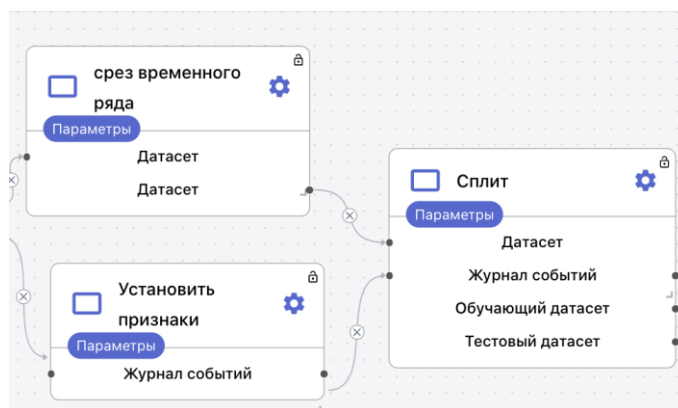


Рисунок 298 - Соединение элементов Установить признаки и Срез временного ряда с блоком Сплит

Стабилизация дисперсии. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Препроцессинг» -> функция «Стабилизация дисперсии».
2. В качестве метода выберите 2. yeo-johnson
3. Установите галочку в поле «Замена значений столбцов».
4. Установите галочку в поле «Стандартизация».
5. В поле «Флаг признака» выбрать из списка значение «1. Признаки»

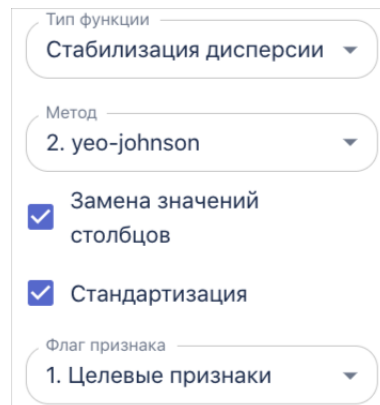


Рисунок 299 - Параметры функции «Стабилизация дисперсии»

6. Введите название для элемента: «Стабилизация».
7. Соедините элементы:

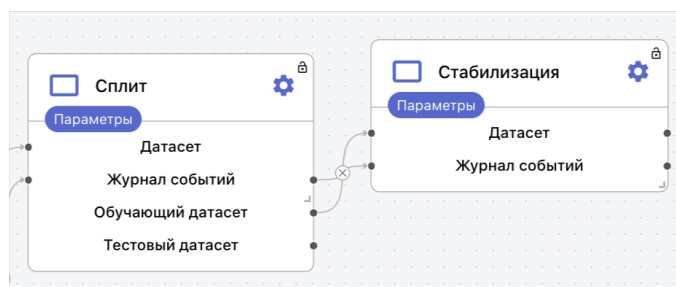


Рисунок 300 - Соединение элементов Сплит и Стабилизация

Skew и Дики-Фуллер для Стабилизации. Данный этап не является обязательным, но позволяет оценить стал ли ряд стационарным после всех преобразований.

1. Добавьте на рабочую область и настройте два элемента «Процесс»:
2. Для первого блока в карточке элемента выберите из списка функцию: раздел «Тест на нормальность распределения» -> функция «Коэффициент асимметрии Skewness». В поле признаки введите ['Tq', 'Tw'] и сохраните. Наименование блока можно изменить на «Skew».
3. Для второго блока в карточке элемента выберите из списка функцию: «Тест на стационарность временного ряда» -> функция «Тест

Дики-Фуллера». В поле «пороговое значение alpha» введите [0.05].
Наименование элемента можно изменить на «Дики-Фуллер».

4. Соедините элементы:

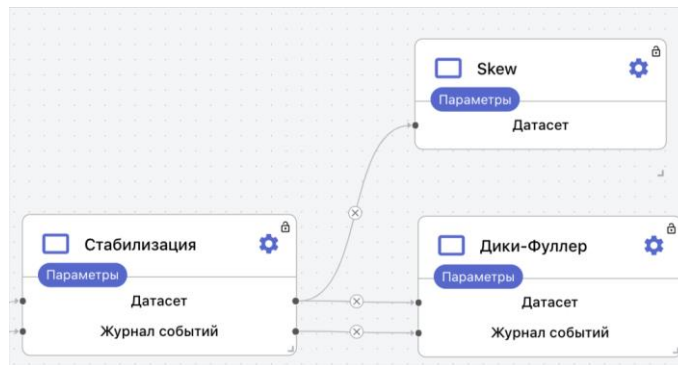


Рисунок 301 - Соединение элементов Стабилизация с Skew и Дики-Фуллер

Дифференцирование временного ряда. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Препроцессинг» -> функция «Дифференцирование временного ряда».

2. В поле «Шаг дифференцирования введите: [1,1] (вместе с квадратными скобками).

Шаг сдвига - это параметр, который определяет, на сколько наблюдений мы сдвигаем временной ряд, чтобы получить разницу в значениях.

3. Измените название элемента на «Дифференцирование».

4. Соедините элементы:

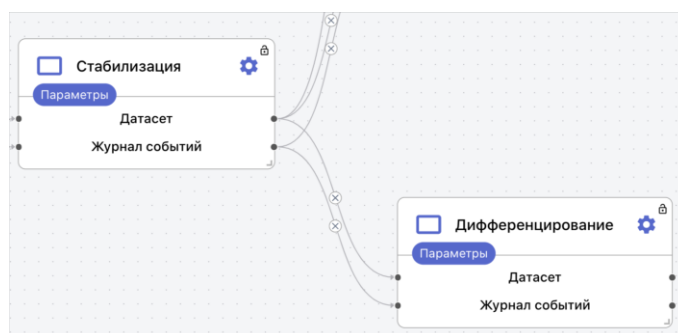


Рисунок 302 - Соединение элементов Стабилизация и Дифференцирование

Для блока «Дифференцирование» опционально можно провести **тест Дики-Фуллера**. Для этого необходимо:

1. Добавить на рабочую область и настроить элемент «Процесс» и выбрать из списка функцию: «Тест на стационарность временного ряда» -> функция «Тест Дики-Фуллера». В поле «пороговое значение alpha» вводим [0.05]. Наименование элемента можно изменить на «Дики-Фуллер».

2. Соединить элементы:

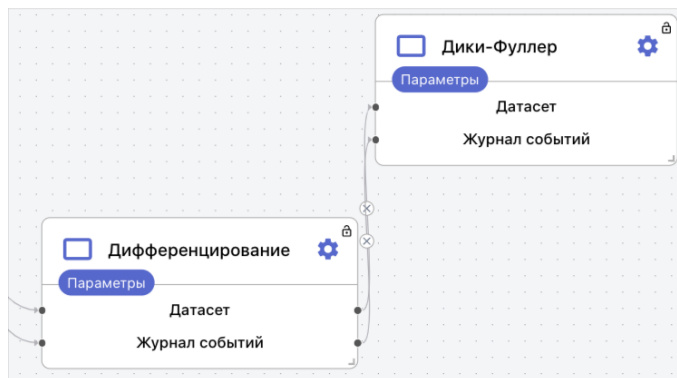


Рисунок 303 - Соединение элементов Дифференцирование и Дики-Фуллер

Создание признаков для временного ряда. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Препроцессинг» -> функция «Создание признаков для временного ряда».

2. В поле «Максимальное количество лагов» укажите: [1].

3. Измените название элемента на «Создание признаков».

4. Соедините элементы:

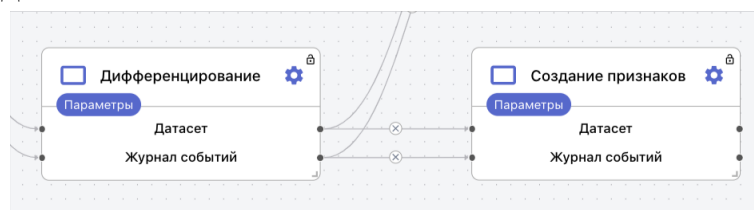


Рисунок 304 - Соединение элементов Дифференцирование и Создание признаков

Метод k-ближайших соседей для регрессии. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Регрессия» -> функция «Метод k-ближайших соседей для регрессии».

2. В поле «Количество ближайших соседей» укажите [2,3,5,10] (вместе с квадратными скобками).

3. Установите галочку в поле «Оптимизация гиперпараметров»

4. В поле «Тип веса для соседей» выберите и «Единый» и «По расстоянию»
5. В поле «Метрика расстояния» выберите и «Евклидово», «Косинусное» и «Манхэттенское»
6. В поле «Метрика для оптимизации» выберите и «1. RMSE» и в поле «Количество фолдов для оптимизации» указать [3]

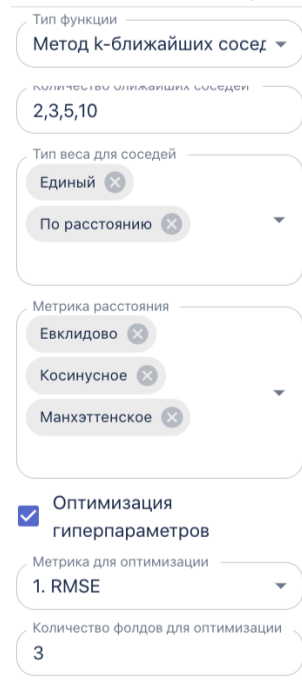


Рисунок 305 - Параметры функции «Метод k-ближайших соседей для регрессии»

7. Измените название элемента на «KNN».
8. Соедините элементы:

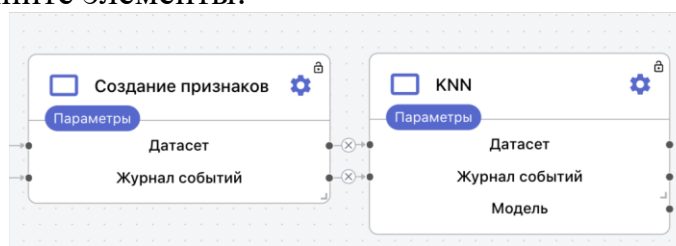


Рисунок 306 - Соединение элементов Создание признаков и KNN
Валидация модели. Добавьте на рабочую область и настройте элемент «Процесс»:

1. Выберите из списка функцию: раздел «Машинное обучение» - > функция «Валидация модели»:

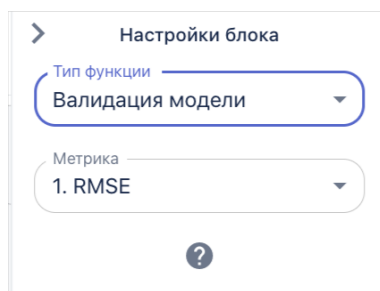


Рисунок 307 - Параметры функции «Валидация модели»

2. В разделе «Параметры» -> в поле «Метрика» выберите значение «5. F1» – метрика для валидации. Анализируется связь между выбранными погодными условиями и целевым признаком. Так оценивается вероятность возникновения лесного пожара по всем показателям. Функция возвращает величину вероятности в виде значения от 0 до 1.

3. Введите название элемента «Валидация».

4. Установите соединения с предыдущим элементом «Классификация» и элементом «Сплит датасета», как показано в Приложении 7. **Обратите внимание!** Элементы в блок-схеме могут соединяться не только последовательно.

Сохранение модели. Добавьте на рабочую область и настройте элемент «Процесс»:

1. В карточке элемента выберите из списка функцию: раздел «Управление моделями» -> функция «Сохранение модели».

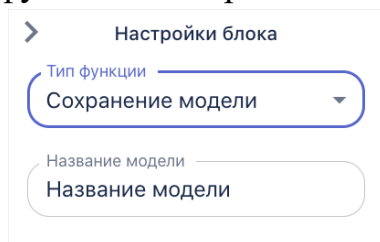


Рисунок 308 - Параметры функции «Сохранение модели»

2. В разделе «Параметры» введите название модели, с которым она будет сохранена, например «mei».

3. Измените название элемента на «Сохранение модели».

4. Соедините элементы:

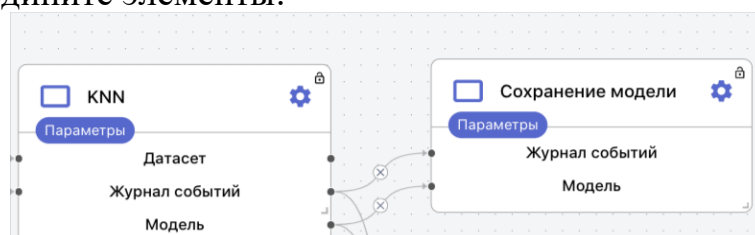



Рисунок 309 - Соединение элементов KNN и Сохранение модели

Запуск пайплайна. Чтобы запустить сборку пайплайна нажмите на кнопку  на первом элементе «Запуск» собранной блок-схемы. При этом отображение элемента «Запуск» изменится и появится опция Сформировать отчет:

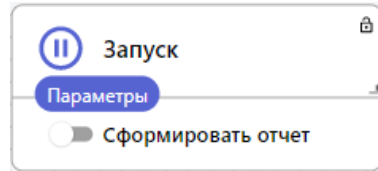
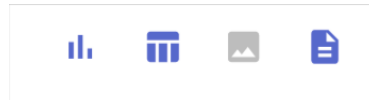


Рисунок 310 - Соединение элементов KNN и Сохранение модели

Если активировать параметр «Сформировать отчет», в результате запуска пайплайна будет создан отчет.

Визуализация результатов. После того как все элементы схемы будут успешно обработаны, на панели инструментов появляются кнопки:



2.3.13.2. Создание блок схемы для работы с данными в режиме реального времени

BAUM STORAGE AI v2 позволяет обрабатывать не только статичную информацию, но также данные, получаемые в режиме реального времени. Возможность прогнозировать значения признаков и линий их трендов создает базу для оперативного и своевременного принятия решений, позволяющих избежать критических ситуаций. BAUM STORAGE AI v2 также позволяет создавать визуализации данных в реальном времени при помощи различных графиков, что делает информацию более наглядной и обеспечивает широту ее применения.

В данном разделе будет рассмотрен пример работы с BAUM STORAGE AI v2 для прогнозирования значений в режиме реального времени. Для выполнения этого сценария, необходимо построить и запустить пайплайн Обучение модели прогнозирования температуры воды и газов в котле для обучения модели ИИ на аналогичных данных в виде статичной таблицы. Вторым этапом созданная модель применяется для обработки данных и моментального прогнозирования значений.

Т.о. второй этап работы с данными в режиме реального времени заключается в настройке коннектора и построении второго пайплайна, который будет получать информацию из соединения и обрабатывать её с помощью ранее созданной и обученной модели. В рамках примера также

будет показано, как создавать и интерпретировать графики, на которых отображается информация в режиме реального времени.

1. Создание коннектора.

1.1. Перейдите в пункт меню «Соединения». Перейти в пункт меню «Соединения»:

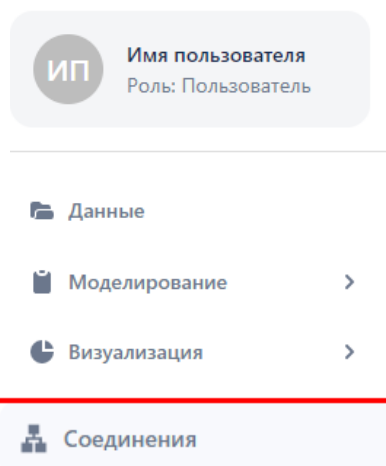


Рисунок 311 - Пункт меню Соединения

Откроется страница «Соединения» на первой вкладке «Источники данных», на которой отображаются все ранее созданные источники:

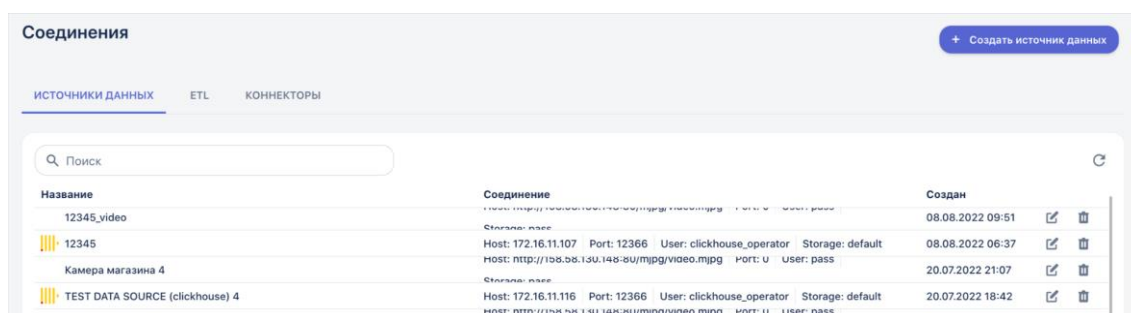


Рисунок 312 - Список созданных источников данных

1.2. Для создания нового источника, нажмите кнопку «Создать источник данных» в верхнем правом углу. Откроется окно «Создание нового ресурса данных»:

Рисунок 313 - Создание нового источника данных

1.3. Заполните поля следующим образом:

- *Название.* Задайте название источника «postgres dataset 116».
- *Хост.* Указывается хост протокола TCP/IP, т.е IP-адрес подключаемой БД, например: «172.16.11.116».
- *Порт.* Номер порта, по которому устанавливается соединение с сервером, на котором установлена БД **postgresql**. Указать «9999».
- *Имя хранилища.* Название базы данных, которое указано на подключаемом сервере. Указать «dataset».
- *Тип хранилища.* Из выпадающего списка выберите тип «postgresql»:
- *Имя пользователя, пароль.* Параметры учетной записи администратора внешнего сервера для разрешения доступа к данным. Указать пользователя «postgres», и пароль «example».
- *Описание.* Вводится дополнительная информация по источнику, необязательное поле.
- Для регистрации в Системе источника нажмите кнопку «Создать».

1.4. Созданный коннектор отобразится в общем списке. При необходимости изменить настройки используйте кнопку «Редактировать»:

Название	Соединение	Создан
postgres dataset 116	Host: 172.16.11.116 Port: 9999 User: postgres	19.08.2022 09:29
12345_video	Storage: dataset Host: http://158.58.130.148:80/mjpg/video.mjpg	08.08.2022 09:51
	Port: 0 User: root Storage: none	

Рисунок 314 - Отображение созданного источника в списке
1.5. **Создание ETL.** Создание нового ETL осуществляется на вкладке «ETL» в разделе Соединения:

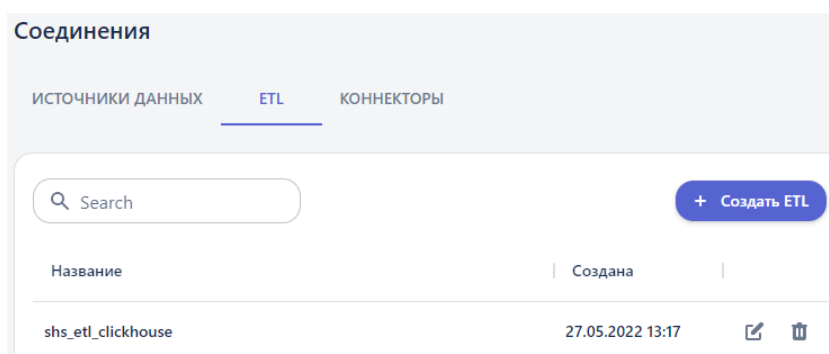


Рисунок 315 - Список созданных ETL

1.5.1. Нажмите кнопку «Создать ETL». Откроется окно «Создать новый ETL»:

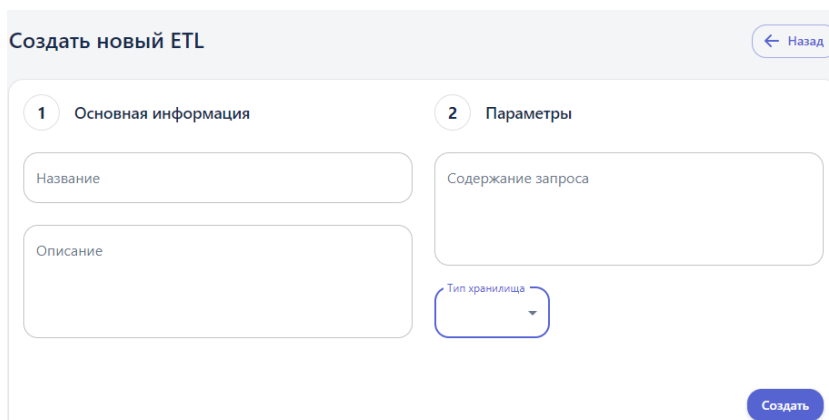


Рисунок 316 - Создание нового ETL

1.5.2. Заполните поля следующим образом:

→ *Название.* Пользователь вручную задает название ETL «me1d_duplicate_2» – запроса на извлечение данных.

→ *Содержание запроса.* Прописывается sql запрос для извлечения данных из внешнего сервера: `select "Tq", "Tw" from me1d_duplicate_2`

→ *Тип хранилища.* Выбирается тип «postgresql».

→ Нажмите на кнопку «Создать».

1.5.3. Созданный ETL можно менять и редактировать по аналогии с источником данных.

1.6. **Создание Коннектора.** Перейдите на вкладку «Коннекторы»:

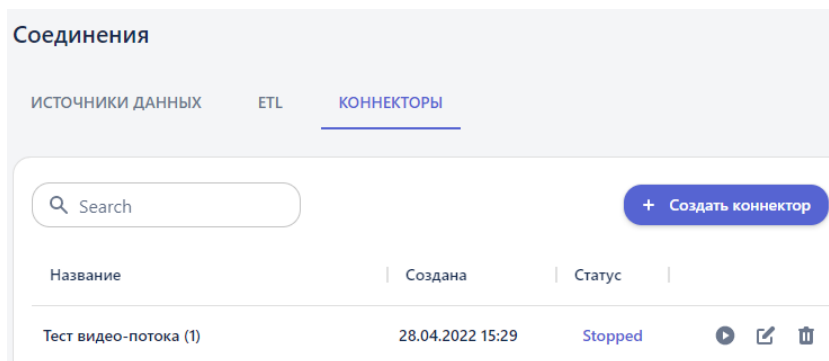


Рисунок 317 - Список созданных коннекторов

1.6.1. Нажмите на кнопку «Создать коннектор». Откроется окно «Создать новый коннектор»:

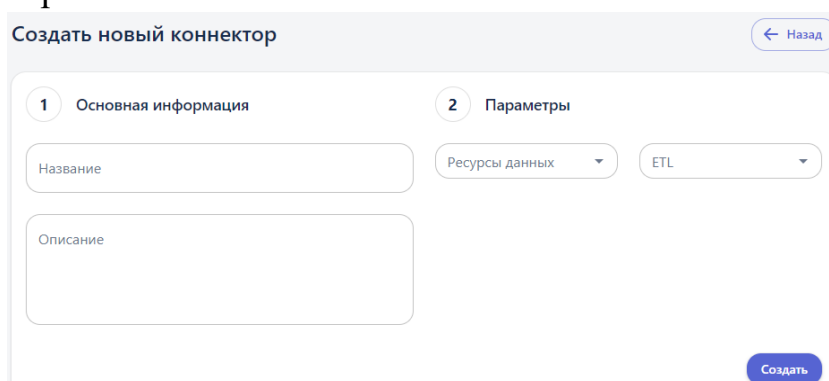


Рисунок 318 - Создание нового коннектора

1.6.2. Заполните поля:

→ *Название*. Пользователь вручную задает название создаваемого коннектора, например «me1d_duplicate_2».

→ *Ресурсы данных*. Из списка выбирается источник «postgres dataset 116», созданный в шаге 1

→ *ETL*. Из списка выбирается ETL «me1d_duplicate_2», созданный в шаге 2.

→ *Описание*.

1.6.3. Нажмите кнопку «Создать».

1.6.4. Сразу после создания коннектору присваивается статус «Stopped»:

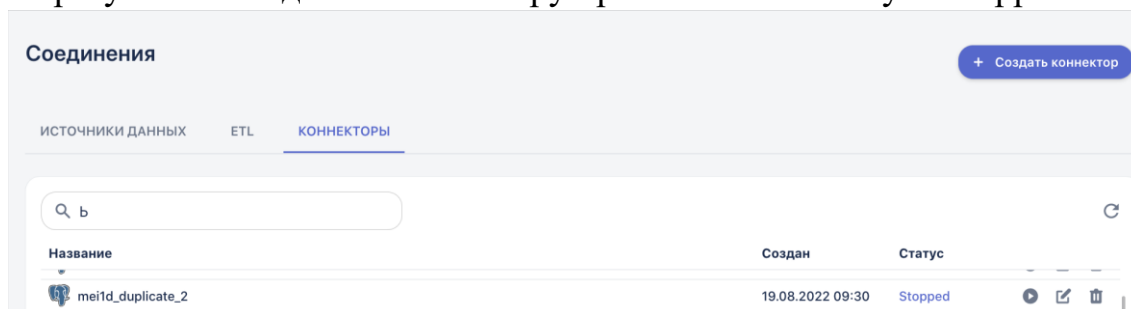




Рисунок 319 - Отображение нового коннектора в списке

1.7. Запуск коннектора. Чтобы данные из источника начали поступать в Систему необходимо запустить коннектор. Для этого нажмите на кнопку «» в строке с коннектором. В результате статус меняется на значение «Started». Теперь коннектор можно использовать в качестве источника данных при построении пайплайна.

2. Создание блок схемы.

Следующим этапом мы переходим к созданию второго пайплайна, где будет использована ранее обученная модель и созданный коннектор. Т.о. будет происходить обработка данных, получаемых в режиме реального время для получения прогноза температуры воды и котла.

2.1. Создание новой рабочей области.

2.1.1. Перейдите в пункт меню системы **Моделирование** → **Рабочая область**. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка )

2.1.2. В открывшейся форме введите название новой рабочей области «МЭИ realtime» и нажмите кнопку «Создать»:

2.1.3. На панели инструментов отобразится название созданной рабочей области.

2.2. Добавление первого элемента «Источник данных». В данном примере мы будем использовать два источника данных: коннектор и модель. Для загрузки в пайплайн данных из коннектора добавьте элемент «Источник данных» на рабочую область и настройте элемент:

2.2.1. На элементе нажмите на кнопку . Откроется панель настроек элемента.

2.2.2. На панели настроек элемента выберите из списка функцию: раздел «Загрузка данных» → функция «Загрузка табличных данных из коннектора».

2.2.3. В поле «Выберите файл» из списка всех созданных коннекторов выберите «me1d_duplicate_2».

2.2.4. Установите галочку в поле «Онлайн данные»:

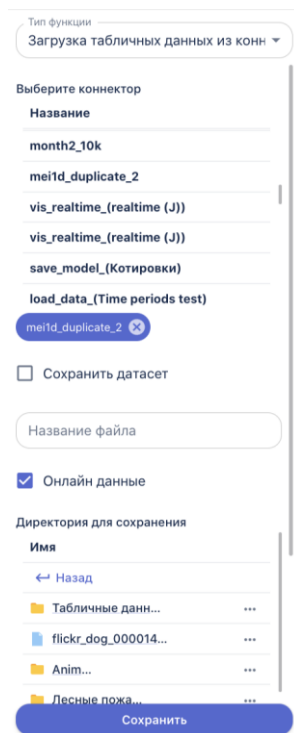


Рисунок 320 - Выбор файла для загрузки

- 2.2.5. Нажмите на кнопку «Сохранить».
- 2.2.6. Переименуйте блок в «Данные МЭИ»
- 2.2.7. Соедините элементы:

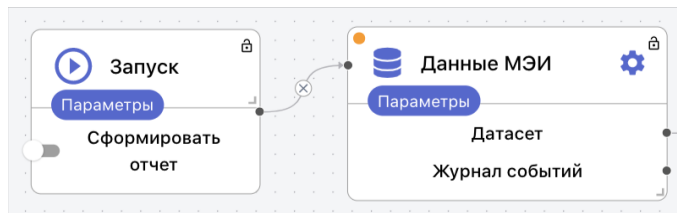




Рисунок 321 - Соединение элементов Запуск и Данные МЭИ

Полностью схема пайплайна представлена в [Таблице 16.8](#)

2.3. **Добавление второго элемента «Источник данных».** Для того чтобы включить в пайплайн ранее созданную обученную модель, нужно добавить еще один элемент «Источник данных» на рабочую область (кнопка ). Чтобы настроить элемент:

- 2.3.1. На элементе нажмите на кнопку . Откроется панель настроек элемента.
- 2.3.2. На панели настроек элемента выберите из списка функцию: раздел «Машинное обучение» -> функция «Загрузка модели».
- 2.3.3. В списке моделей выберите «mei»
- 2.3.4. Переименуйте блок в «Модель»

2.4. Запись в датасет логирования. В данном блоке будет осуществляться логирование новой поступающей новой информации в датасет, для этого на рабочую область добавляется элемент «Процесс»:

2.4.1. В карточке элемента выберите из списка функцию: раздел «Анализ данных» -> функция «Запись в датасет логирования».

2.4.2. Переименуйте блок в «Логирование»

2.4.3. Соедините элементы:

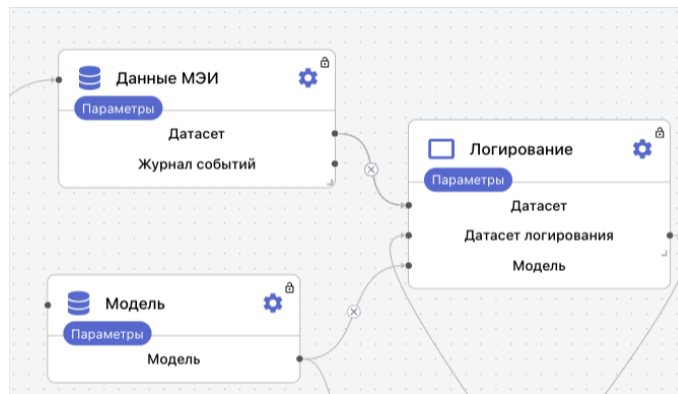


Рисунок 322 - Соединение элементов Данные МЭИ, Модель и Логирование

2.5. Прогноз. Добавьте на рабочую область и настройте элемент «Процесс»:

2.5.1. В карточке элемента выбрать из списка функцию: раздел «Машинное обучение» -> функция «Прогноз модели».

2.5.2. Переименуйте блок в «Прогноз»

2.5.3. Соедините элементы:

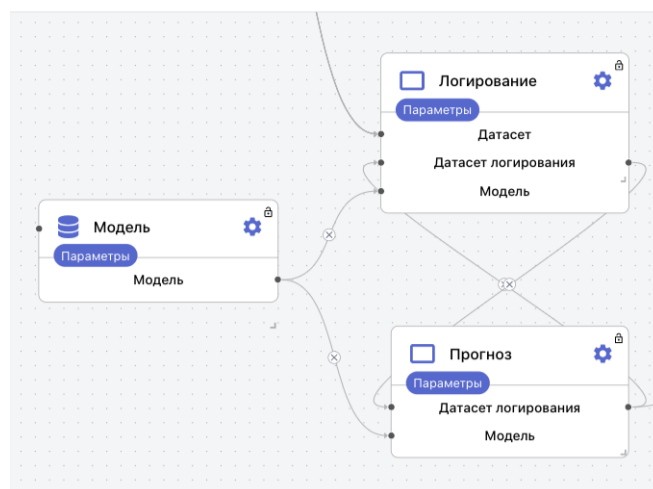


Рисунок 323 - Соединение элементов Модель, Логирование и Прогноз

2.6. **Визуализация.** Добавьте на рабочую область и настройте элемент «Процесс»:

2.6.1. В карточке элемента выберите из списка функцию: раздел «Анализ данных» -> функция «Визуализация Real-Time».

2.6.2. В параметрах блока выберите все графики и установите для них следующие параметры:

- Линейный график: Число периодов в окне: 10; Период окна: 5.Минуты

- Свечной график: Число периодов: 1; Период: 5.Минуты; Число периодов в окне: 10; Период окна: 5.Минуты

2.6.3. Переименуйте блок в «Визуализация»

2.6.4. Соедините элементы:

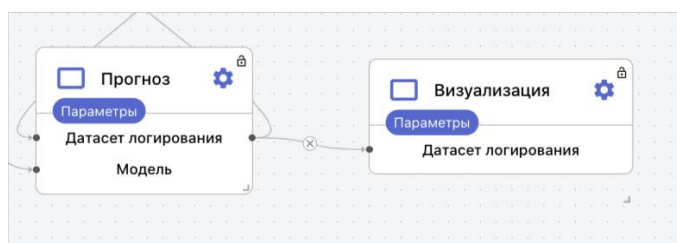



Рисунок 324 - Соединение элементов Прогноз и Визуализация

2.7. **Запуск блок-схемы.** Для запуска блок-схемы нажмите на кнопку  на первом элементе «Запуск». Все элементы блок-схемы должны отработать с зелеными индикаторами.

При запуске real time блок схем элементы будут отрабатывать снова и снова и индикаторы на пайплайне будут постоянно менять цвета. Если необходимо остановить обработку данных - нажмите кнопку запуск повторно.

3. Визуализация результатов прогнозирования температуры на графиках

1. После того как все элементы схемы будут успешно обработаны, на панели инструментов активизируется кнопка «Графики»:



Графики будут доступны в зависимости от выбранных в рамках блока «Визуализация». В текущем примере это Линейный и Свечной графики:

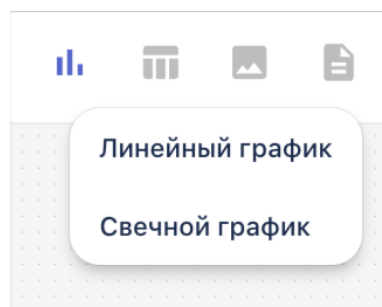


Рисунок 325 - Список доступных графиков для отображения на рабочей области

Для отображения графиков на рабочей области нужно выбрать необходимые, кликнув на их названия.

Обратите внимание, что блок схема работает в режиме реального времени и данные на графиках будут постоянно обновляться в зависимости от заданного лага в графиках.

1.1. Линейный график:

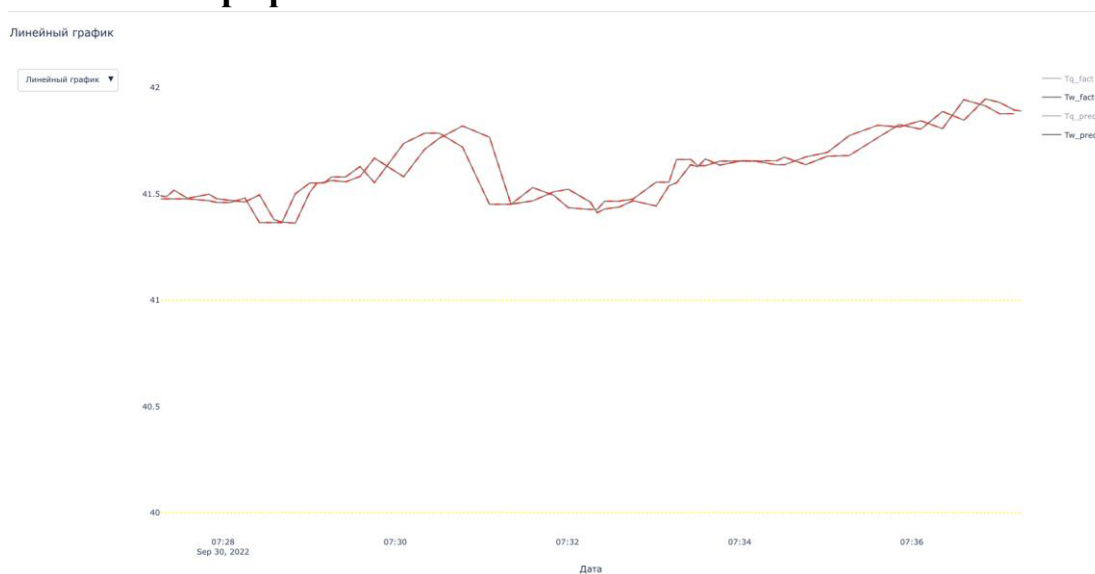


Рисунок 326 - Линейный график временного ряда в режиме реального времени

График показывает изменения значений целевых признаков (Tq и Tw) в течение заданного периода времени, который устанавливается в параметрах блока «Визуализация». В нашем примере мы выбрали период равный 10 минутам:

Линейный график

Число периодов в окне

Период окна

Рисунок 327 - Параметры настройки линейного графика

Соответственно на графике отображаются все значения признаков за последние 10 минут. Каждый раз, когда блок схема будет обрабатывать - временной отрезок ниже будет сдвигаться вперед, но значения всегда будут в пределах 10 минут. Т.к. на пайплайне присутствует блок прогнозирования, на линейном графике мы также видим прогноз значений параметров. Горизонт планирования равен шагу ресемплирования из пайплайна обучения, это значение записывается моделью.

1.2. Свечной график:

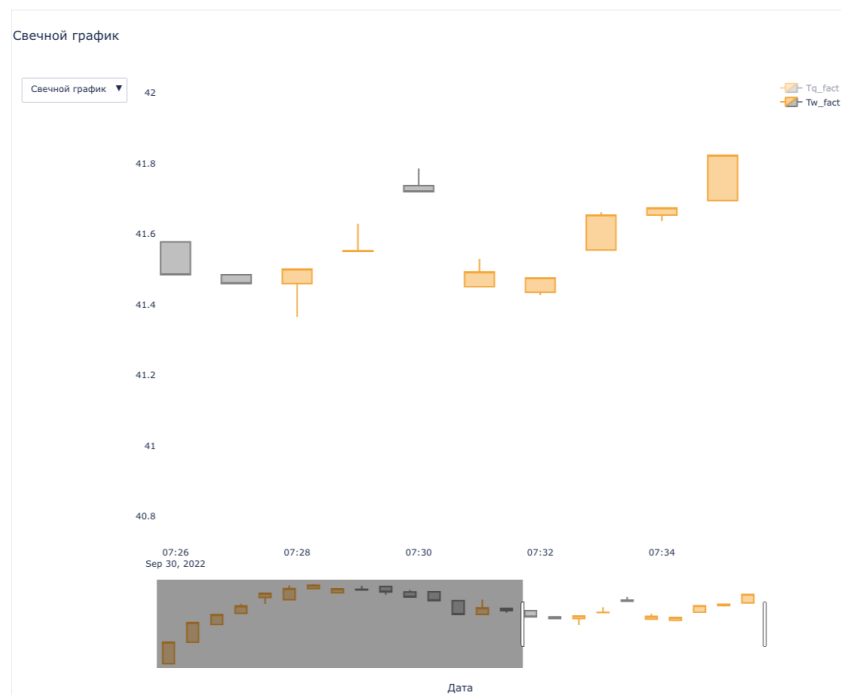


Рисунок 328 - Свечной график в режиме реального времени

График также показывает изменения значений параметров в течение заданного периода времени. При этом задается период, в рамках которого будет сформирована т.н. свеча (столбик на графике); и число периодов в окне. В нашем примере - это одна минута, и мы можем видеть 10 свечей на графике, т.к. задали число периодов в окне равное 10 минутам:

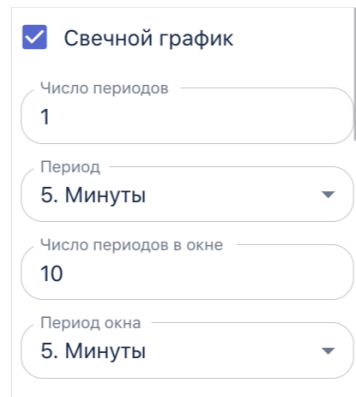


Рисунок 329 - Параметры настройки свечного графика

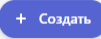
На горизонтальной оси отображается время, для которого было зафиксировано значение. На вертикальной - само значение признака. Прямоугольники (свечи) на графике отображают разницу между значением параметра на начало периода и на конец. Линии, исходящие из свечей, показывают максимальное и минимальное значение параметра за период времени. Если значение параметра на конец периода выше, чем на начало – то свеча окрасится в оранжевый цвет; если значение на конец периода ниже, чем на начало – в серый.

Созданные графики можно использовать для создания индивидуальных дашбордов.

2.3.13.3. Классификация изображений

Классификация изображений позволяет отнести изображение к определенному классу. Например, из набора изображений определить какое изображение относится к классу “Кошка” или “Собака”. Сначала в системе создаются папки, которые будут далее пополнены изображениями, классификации которых будет обучаться модель ИИ. Для загрузки в систему могут быть использованы файлы следующих форматов: 'jpg', 'jpeg', 'png'. Для каждого класса будет создана отдельная папка, данных из которой будут использованы при построении пайплайна для обучения модели опознавать определенные объекты на изображении.

2.3.13.3.1. Загрузка изображений

1. Первым шагом откройте раздел «Данные»
2. Для создания новой папки нажмите , после чего откроется форма создания нового типа данных:

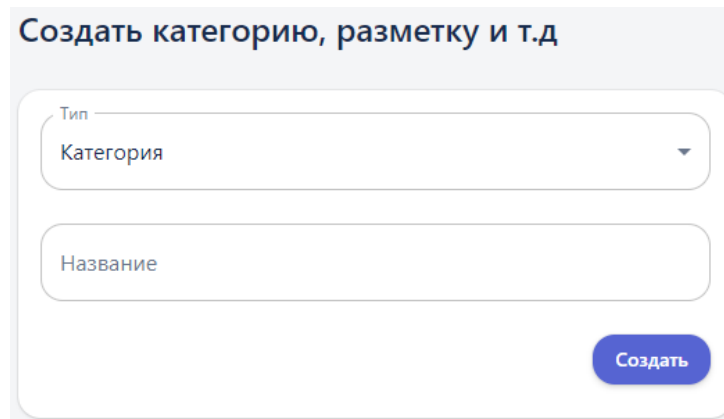


Рисунок 330 – Создание новой папки для классификации

3. В открывшейся в поле «Тип» выберите значение «Категория», в поле «Название» введите название новой группы, например «Данные для классификации изображений», и нажмите кнопку «Создать». **Примечание** – *Здесь под категорией имеется в виду папка, в которую будут складываться данные для решения задачи.*

4. Перейдите в созданную папку «Данные для классификации изображений» и создайте две новые папки внутри – «Animals Train» и «Animals Test». В группу «Animals Train» будут складываться данные для обучения будущей модели машинного обучения, а в группу «Animals Test» – данные для валидации или проверки ‘качества’ уже обученной модели. При этом для обучения модели необходимо использовать большее количество файлов, в нашем примере пропорция составляет 4 к 1.

5. В папке «Animals Train» создайте еще две папки, которые и будут определять классы, – «Dogs» и «Cats». Количество классов равно двум, так как в данном сценарии решается задача *бинарной классификации* (для многоклассовой классификации создавалось бы больше двух классов). В класс «Dogs» загружаются изображения собак, а в класс «Cats» загружаются изображения кошек.

6. Для того чтобы загрузить файлы, перейдите в нужную папку и нажмите кнопку «Загрузить»:

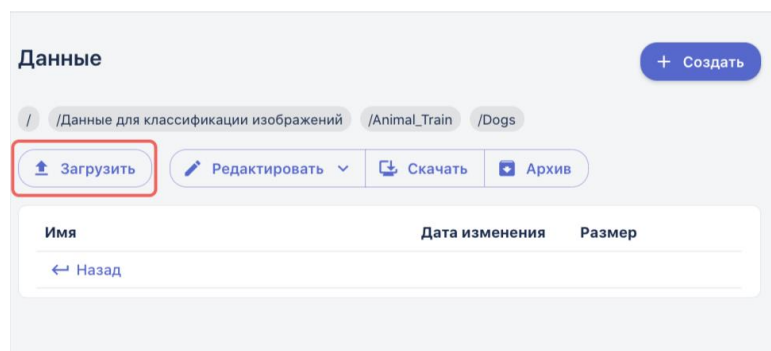


Рисунок 331 – Кнопка загрузки файлов для классификации

7. Откроется окно загрузки файлов:

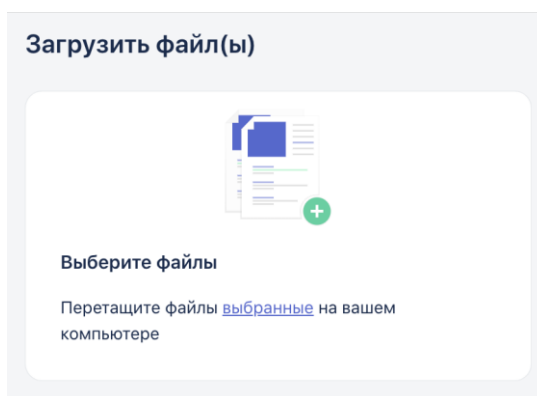


Рисунок 332 – Окно выбора файлов

8. Для выбора файлов кликните в область окна «Выберите файлы» или перетащите их по технологии drag and drop (из окна папки на вашем ПК в окно браузера).

Обратите внимание: за раз можно добавить максимум 10 файлов. Соответственно, если нужно загрузить больше файлов, нужно повторить выбор несколько раз.

После того, как все файлы выбраны, при необходимости вы можете удалить ненужные файлы, нажав на крестик в правой части строки с файлом, или нажать «Удалить все», если это требуется.

9. Когда все файлы выбраны и готовы к загрузке, нажмите на кнопку «Загрузить»:

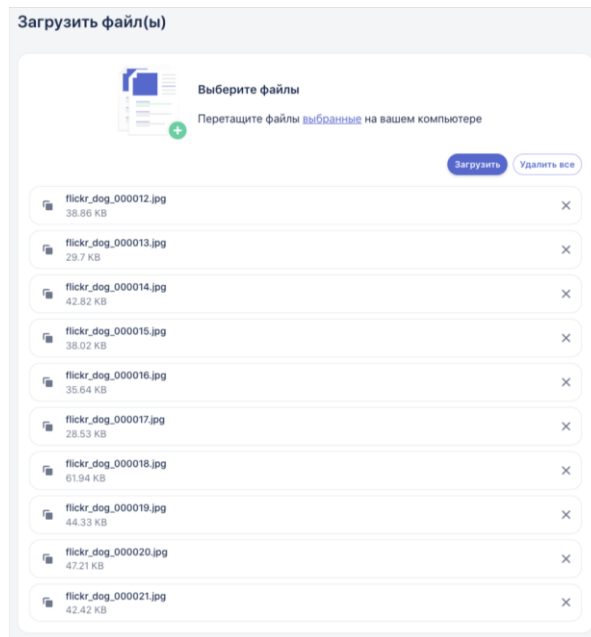


Рисунок 333 – Список выбранных файлов

10. В результате загруженные файлы отобразятся в папке:

Имя	Дата изменения	Размер
<input type="checkbox"/> flickr_dog_000023.jpg	15.09.2022 15:12	44 KB
<input type="checkbox"/> flickr_dog_000035.jpg	15.09.2022 15:12	42.88 KB
<input type="checkbox"/> flickr_dog_000053.jpg	15.09.2022 15:12	33.57 KB
<input type="checkbox"/> flickr_dog_000037.jpg	15.09.2022 15:12	55.95 KB
<input type="checkbox"/> flickr_dog_000049.jpg	15.09.2022 15:12	29.93 KB
<input type="checkbox"/> flickr_dog_000046.jpg	15.09.2022 15:12	22.78 KB
<input type="checkbox"/> flickr_dog_000013.jpg	15.09.2022 15:12	29.7 KB
<input type="checkbox"/> flickr_dog_000021.jpg	15.09.2022 15:12	42.42 KB
<input type="checkbox"/> flickr_dog_000022.jpg	15.09.2022 15:12	36.35 KB
<input type="checkbox"/> flickr_dog_000020.jpg	15.09.2022 15:12	47.21 KB
<input type="checkbox"/> flickr_dog_000034.jpg	15.09.2022 15:12	46.5 KB
<input type="checkbox"/> flickr_dog_000029.jpg	15.09.2022 15:12	42.47 KB
<input type="checkbox"/> flickr_dog_000033.jpg	15.09.2022 15:12	28.1 KB
<input type="checkbox"/> flickr_dog_000012.jpg	15.09.2022 15:12	38.86 KB

Рисунок 334 – Загруженные файлы в папке

11. Вышеописанные действия повторяются для папки «Animals Train» -> «Cats».

12. Далее по аналогии создаются и заполняются папки «Animals Test» -> «Dogs» и «Cats»,туда загружаются файлы для валидации модели.

13. Чтобы удалить группу/класс достаточно удалить соответствующую папку в разделе Данные, нажав на три точки в строке с этой папкой.

14. После того, как обучающая и валидационная выборки собраны, для папок «Animals Test» и «Animals Train» добавляется параметр классификация. Для этого в строке с папкой нажмите на три точки и кликните на кнопку «Классификация», после этого содержимое папки будет готово для использования при построении модели:

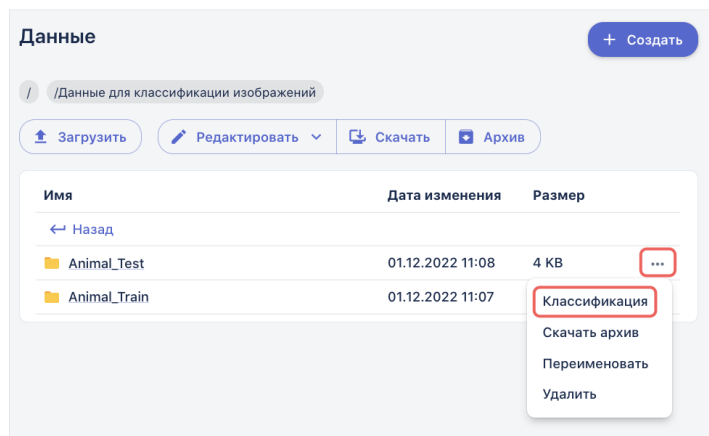



Рисунок 335 – Кнопка присвоения параметра Классификации папке

Обратите внимание: данное действие необходимо выполнить один раз. Даже если позже в папку будут добавлены новые файлы, они будут учтены при построении или запуске модели классификации.

2.3.13.3.2. Создание модели классификации

На левой рабочей панели выберите Моделирование-> Рабочая область.

Нажмите , чтобы добавить новую модель (например, “Animals”) и нажмите кнопку «Создать» (см. рис. 13.3.3). Чтобы воспользоваться существующими моделями выберите Моделирование-> Сохраненная рабочие области. Выберем готовую блок-схему из модели “Animals”. Она состоит из 4 блоков: “Загрузка изображений”, “Классификация”, “Валидация” и “Сохранение” (см. рис. 13.3.4 или приложение 6).

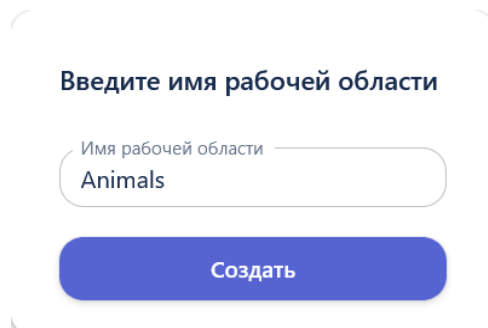


Рисунок 336 - Модель классификации изображений

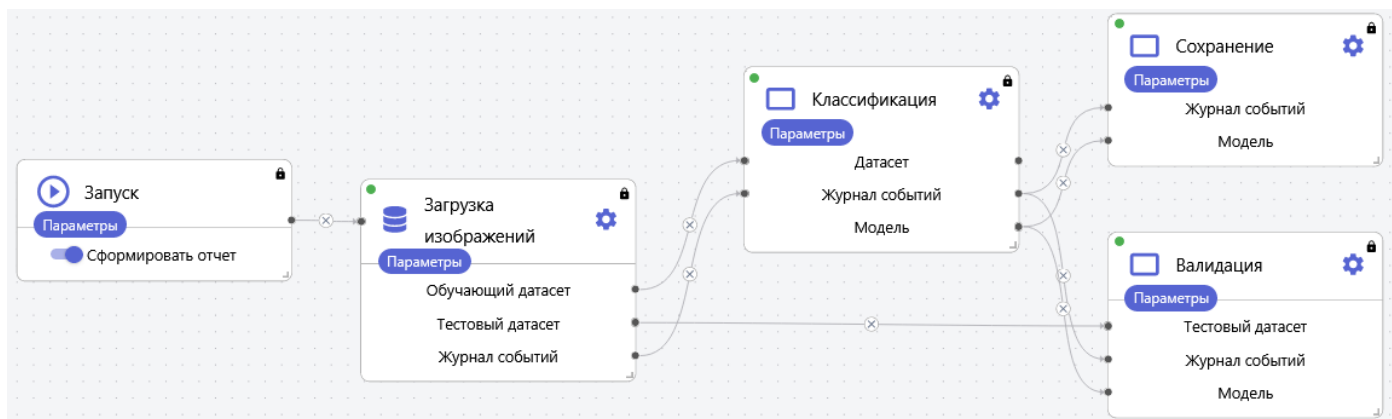


Рисунок 337 - Модель “Animals” для классификации изображений

1. Настройки блока “Запуск”

На рисунке 14.3.7 представлен блок “Запуск”. Нажмите кнопку **Параметры**, чтобы скрыть/раскрыть параметры блока. Используя ползунок **Сформировать отчет** можно опционально сформировать отчет после успешного выполнения модели (см. рис. 13.3.5).

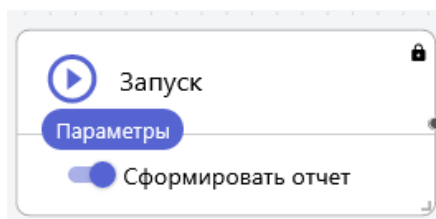



Рисунок 338 - Блок “Запуск”

2. Настройки блока “Загрузка изображений”

На рисунке 13.3.6 представлен блок источник данных - “Загрузка изображений”. Нажмите , чтобы редактировать настройки блока “Загрузка изображений”. В поле “Тип функции” выберете “Анализ данных”. В поле “Список функций” выберете “Загрузка изображений для классификации”.

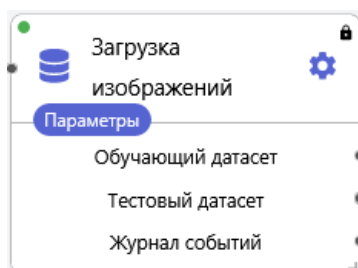


Рисунок 339 - Блок “Загрузка изображений”

В окне настройки параметров блока необходимо выбрать в Группе обучающих изображений выборку `Animals_Train`, а в Группе тестовых - `Animals_Test`:

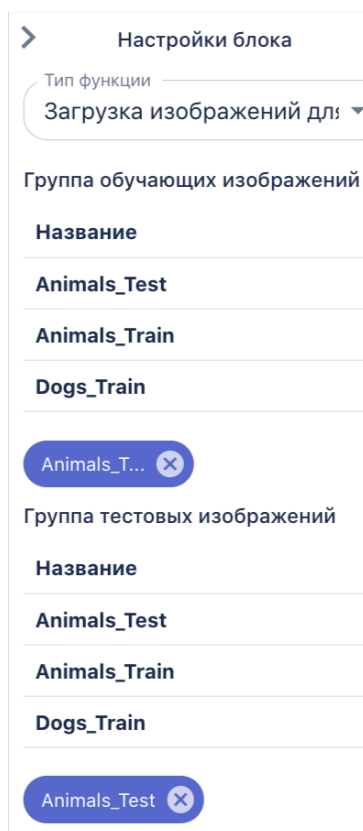



Рисунок 340 - Выбор обучающих и тестовых изображений

В поле “Размер мини-батча” укажите 8 - размер данных (количество изображений), по которым считается функция потерь при градиентном спуске. То есть при обучении на каждом шаге градиентного спуска из всего датасета берется случайным образом 8 объектов. В поле “Высота” и “Ширина” укажите 30, чтобы масштабировать изображение (уменьшить размер). Нажмите кнопку “Сохранить”, чтобы применить настройки.

3. Настройки блока “Классификация”

На рисунке 13.3.8 представлен блок “Классификация”. Нажмите , чтобы редактировать настройки блока.

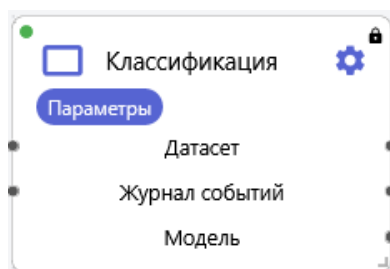


Рисунок 341 - Блок “Классификация”

В поле “Тип функции” выберите “Классификация” - “Классификация изображений”.

В поле “Количество эпох” задайте 60.

В поле “Метрика для обучения” выберите “Accuracy”.

В поле “Алгоритм градиентного спуска” выберите “Adam”.

В поле “Шаг градиентного спуска” укажите null.

В поле “Функция потерь” выберите “binary_crossentropy”.

В поле “Порог классификации” оставьте значение по умолчанию (0,5).

Нажмите “Добавить слой” и заполните значения: Слой: 1.Dense; Число нейронов: 1; Функция активации: 3.sigmoid.

Нажмите “Добавить слой” еще раз и заполните значения: Слой: 2.Flatten.

Нажмите “Добавить слой” еще раз и заполните значения: Слой: 3.Conv2D; Количество фильтров: 4; Размер ядра свертки: 3,3; Размер шага свертки: 1; Толщина отбивки из нулей: Функция активации:

Сохраните настройки блока.

4. Настройки блока “Валидация”

Нажмите  , чтобы редактировать настройки блока.

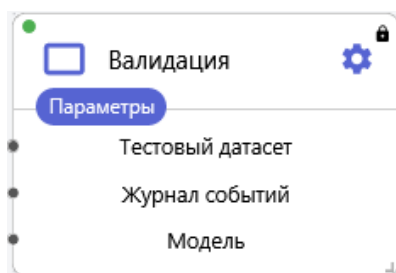


Рисунок 342 - Блок “Валидация”

В поле “Тип функции” выберите “Глубокое обучение”. В поле “Список функций” выберите “Валидация модели классификации изображений”. В поле “Метрика” выберите “F1”.

5. Настройки блока “Сохранение”

Нажмите  , чтобы редактировать настройки блока “Сохранение”.

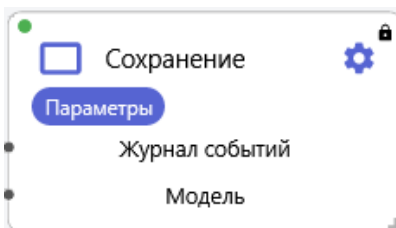



Рисунок 343 - Блок “Сохранение”

В поле “Тип функции” выберите “Управление моделями”. В поле “Список функций” выберите “Сохранение модели классификации изображений”. В поле “Название модели” укажите название модели (например, “animals”).

6. Запуск модели классификации и визуализация результатов

После успешного запуска модели, нажмите , чтобы отобразить результаты выполнения для анализа.

6.1. Графики “История обучения”

Выберите пункт “История обучения”, чтобы оценить качество модели. Мы видим, что с каждым проходом (с каждой эпохой) точность увеличивается (см. рис. Метрика Ассигасу), а количество ошибок уменьшается (см. рис. Функция потерь). Можно видеть, что поле “Количество эпох” равное 60 является оптимальным значением, в то время как 20 было бы недостаточно.

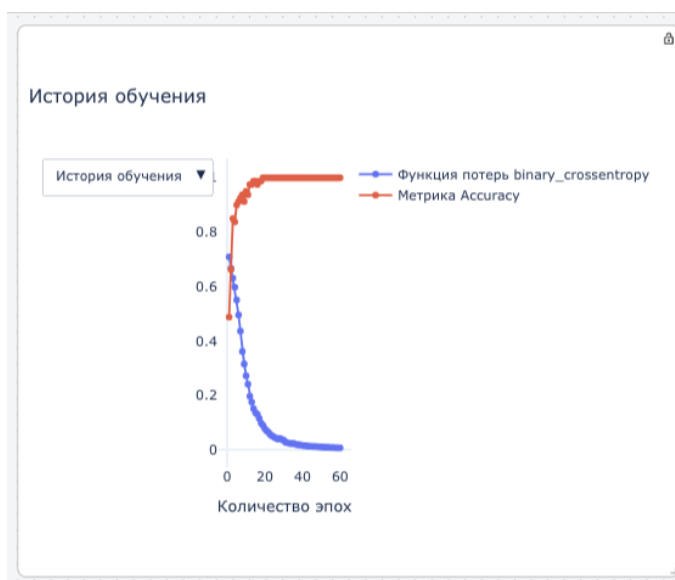


Рисунок 344 - Графики истории обучения


6.2. Матрица ошибок для мультиклассовой классификации

Выберите пункт “Матрица ошибок для мультиклассовой классификации”, чтобы увидеть числовые показатели. На рисунке 345 видно, что модель распознала 10 из 10 кошек, 8 из 10 собак. Для дальнейшего повышения показателей необходимо добавить новые изображения и перезапустить модель.



Рисунок 345 - “Тепловая карта” модели классификации

6.3. Табличные результаты валидации

Нажмите , чтобы ознакомиться табличными результатами валидации. Валидация на тестовой выборке (Метрика F1) и ошибки модели при прогнозировании классов приведены на рисунках ниже:

Валидация на тестовой выборке. Метрика F1:

	F1
1	0.889

Рисунок 346 - Таблица валидации

Ошибки модели при прогнозировании классов

	Верно	Ошибка	Всего
Класс Cats	10	0	10
Класс Dogs	8	2	10

Рисунок 347 - Таблица при прогнозировании

6.4. Архитектура модели

Нажмите  , чтобы посмотреть архитектуру модели:

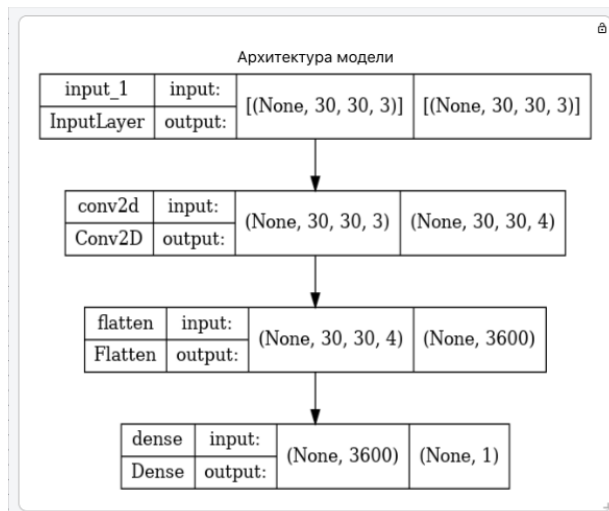


Рисунок 348 - Архитектура модели

2.3.13.4. Классификация текстов

Данный сценарий предполагает решение задачи бинарной классификации произведений по их авторов с использованием в качестве источника текстовых документов в формате .txt.

Для решения задачи выполните следующие действия:

1. Загрузка данных. Набор данных состоит из четырех книг. Половина книг написана Булгаковым, а остальная половина – Клиффордом. Эти книги представляют собой текстовые документы в формате .txt.

- 1.1. Перейдите в раздел данные и нажмите кнопку «Создать»

1.2. В открывшейся в поле «Тип» выберите значение «Категория», в поле «Название» введите название новой группы, например «Данные для классификации текстов», и нажмите кнопку «Создать».

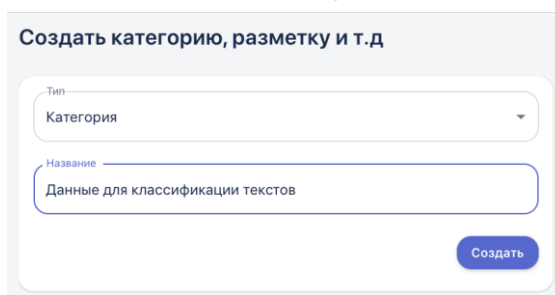


Рисунок 349 - Создание папки в разделе данные

1.3. Перейдите в созданную папку «Данные для классификации текстов» и аналогичным образом создайте две новые папки внутри – «Text Train» и «Text Test». В группу «Text Train» будут складываться данные для обучения будущей модели машинного обучения, а в группу «Text Test» – данные для валидации или проверки ‘качества’ уже обученной модели. При этом для обучения модели необходимо использовать большее количество файлов, в нашем примере в обучающем текстовом файле содержится 5 книг, а в тестовом - 2.

1.4. В папке «Text Train» создайте еще две папки, которые и будут определять классы, – «Clifford» и «Vulgakov». Количество классов равно двум, так как в данном сценарии решается задача бинарной классификации (для многоклассовой классификации создавалось бы больше двух классов). В класс «Clifford» загружается текстовый файл с произведениями Клиффорда, а в класс «Vulgakov» загружаются книги Булгакова.

1.5. Для того чтобы загрузить файлы, перейдите в нужную папку и нажмите кнопку «Загрузить». Откроется окно загрузки файлов:

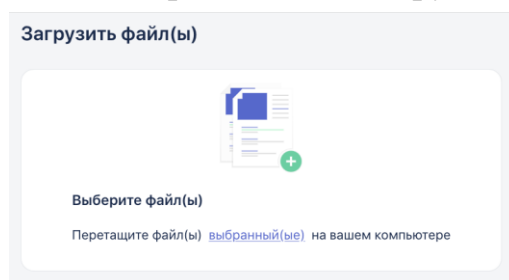


Рисунок 350 - Окно загрузки файлов

1.6. Для выбора файлов кликните в область окна «Выберите файлы» или перетащите их по технологии drag and drop (из окна папки на вашем ПК в окно браузера). Обратите внимание: за раз можно добавить максимум 10 файлов. Соответственно, если нужно загрузить больше файлов, нужно повторить выбор несколько раз.

1.7. После того, как все файлы выбраны, при необходимости вы можете удалить ненужные файлы, нажав на крестик в правой части строки с файлом, или нажать «Удалить все», если это требуется.

1.8. В нашем примере все произведения объединены в один текстовый файл. После его загрузки, нажмите кнопку «Загрузить»:

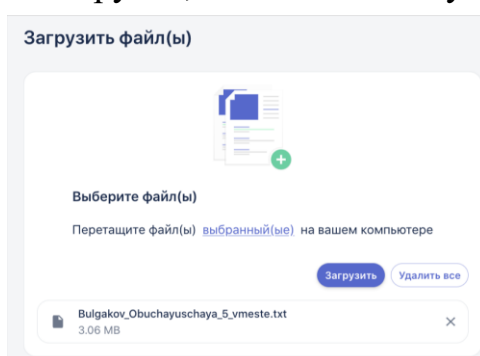


Рисунок 351 - Выбранные файлы в окне загрузки

1.9. В результате загруженные файлы отобразятся в папке:

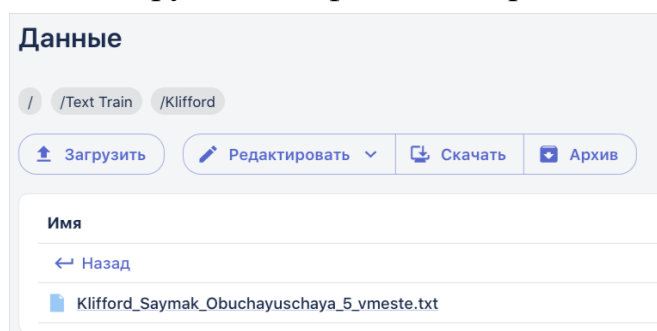


Рисунок 352 - Загруженные файлы в папке

1.10. Вышеописанные действия повторяются для папки «Text Train» -> «Bulgakov».

1.11. Далее по аналогии создаются и заполняются папки «Text Test» -> «Clifford» и «Bulgakov», туда загружаются файлы для валидации модели.

1.12. Чтобы удалить группу/класс достаточно удалить соответствующую папку в разделе Данные, нажав на три точки в строке с этой папкой.

1.13. После того, как обучающая и валидационная выборки собраны, для папок «Text Test» и «Text Train» добавляется параметр классификация. Для этого в строке с папкой нажмите на три точки и кликните на кнопку «Классификация», после этого содержимое папки будет готово для использования при построении модели:

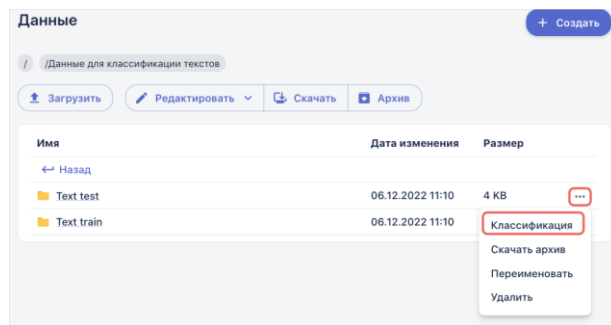


Рисунок 353 - Кнопка присвоения параметра Классификации папке

1.14. Обратите внимание: данное действие необходимо выполнить один раз. Даже если позже в папку будут добавлены новые файлы, они будут учтены при построении или запуске модели классификации.

1. Построение блок-схемы (блок-схему сценария см. в таблице 18.9 «Обучение модели классификации текстов»).

1.1. **Создание новой рабочей области.** Перейдите в раздел «Моделирование» -> Рабочая область. Нажав на \oplus в верхней части экрана, создайте новую рабочую область с названием «Тексты».

1.2. **Блок запуск.** Добавьте на рабочую область элемент «Запуск»:

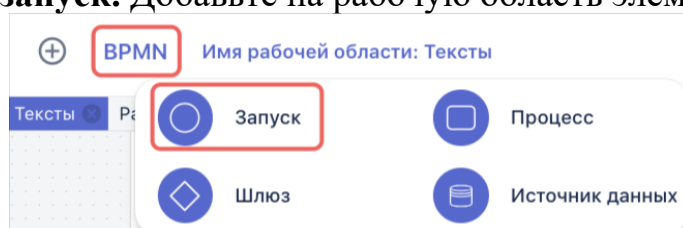
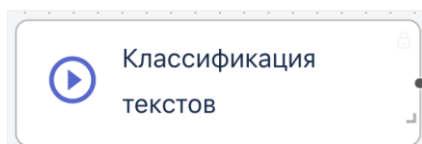


Рисунок 354 - Добавление элемента Запуск на рабочую область

Дважды кликните на название элемента, чтобы переименовать его в «Классификация текстов»:



1.3. **Блок Источник данных.** Добавьте на рабочую область элемент «Источник данных» и в качестве функции выберите «Загрузка текстовых файлов для классификации».

1.1.1. В разделе «Группа обучающих текстов» выберите папку «Text Train», нажав на три точки в строке с ее наименованием и кликнув «Выбрать», в результате папка отобразится в нижней части списка:

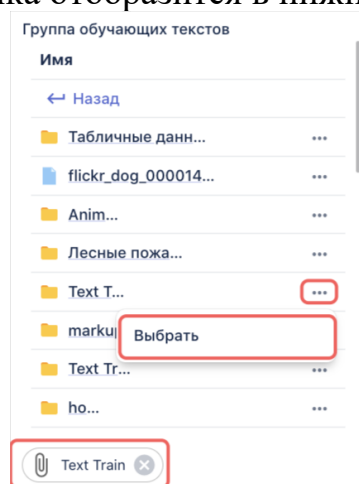
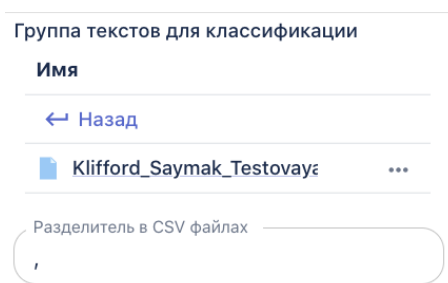


Рисунок 356 - Выбор папок для классификации

1.1.2. Аналогичным образом выберите папку «Text Test» в разделе ниже «Группа тестовых текстов»

1.1.3. В поле «Группа текстов для классификации» вы можете выбрать файл, который необходимо классифицировать с применением обученной модели. В нашем случае, для классификации будет использован тестовый файл из папки Клиффорд*

*Данное поле не является обязательным



1.1.4. Сохраните настройки блока

1.1.5. Переименуйте блок в «Загрузка текстов»

1.1.6. Соедините элементы блок схемы:

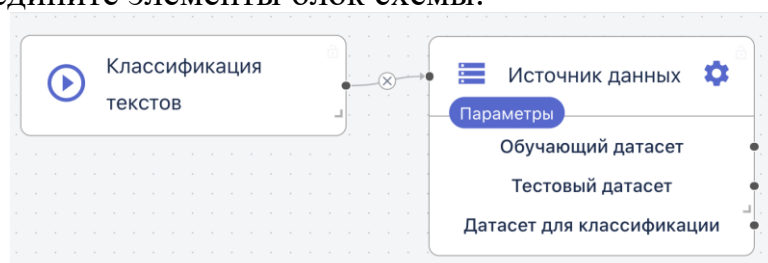


Рисунок 357 - Соединение элементов блок схемы

1.4. **Блок Процесс, Фильтрация шума.** Добавьте на рабочую область три элемента процесс. Для каждого выберите функцию

«Предобработка данных» -> «Фильтрация текстового шума». Не забывайте нажимать «Сохранить» каждый раз после изменения параметров блока процесс.

1.1.7. Переименуйте один блок процесс в «Фильтрация шума train», второй в - «Фильтрация шума test», третий в «Фильтрация шума Клиффорд»

1.1.8. Соедините элементы:

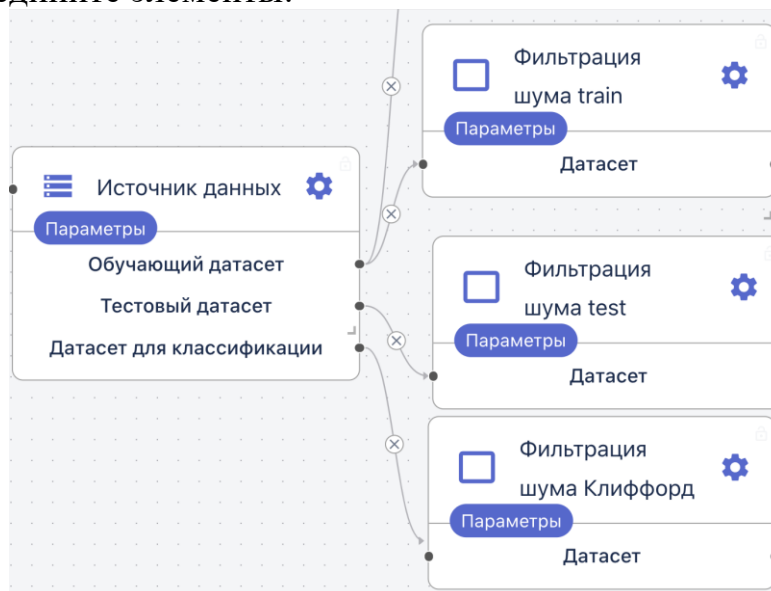


Рисунок 358 - Соединение элементов блок схемы

1.5. **Блок Процесс, Лемматизация.** Добавьте на рабочую область три элемента процесс. Для каждого выберите функцию «Предобработка данных» -> «Лемматизация текста». Не забывайте нажимать «Сохранить» каждый раз после изменения параметров блока процесс.

1.1.9. Переименуйте один блок процесс в «Лемматизация train», второй в - «Лемматизация test», третий - в «Лемматизация Клиффорд»:

1.1.10. Соедините элементы:

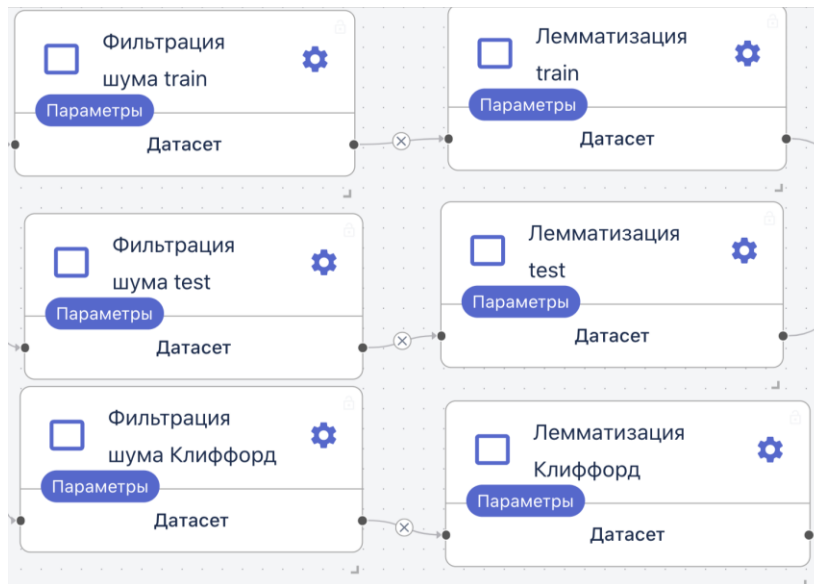


Рисунок 359 - Соединение элементов блок схемы

1.6. **Блок Процесс, Автореферирование.** Добавьте на рабочую область элемент процесс. Выберите функцию «Работа с текстом» -> «Автореферирования текста».

1.1.11. В параметрах блока в поле «Объем автореферата» укажите 200 (это максимальное количество символов, которое отобразится в качестве краткого содержания после применения функции).

1.1.12. Сохраните параметры блока.

1.1.13. Переименуйте блок в «Реферат»

1.1.14. Соедините элементы:

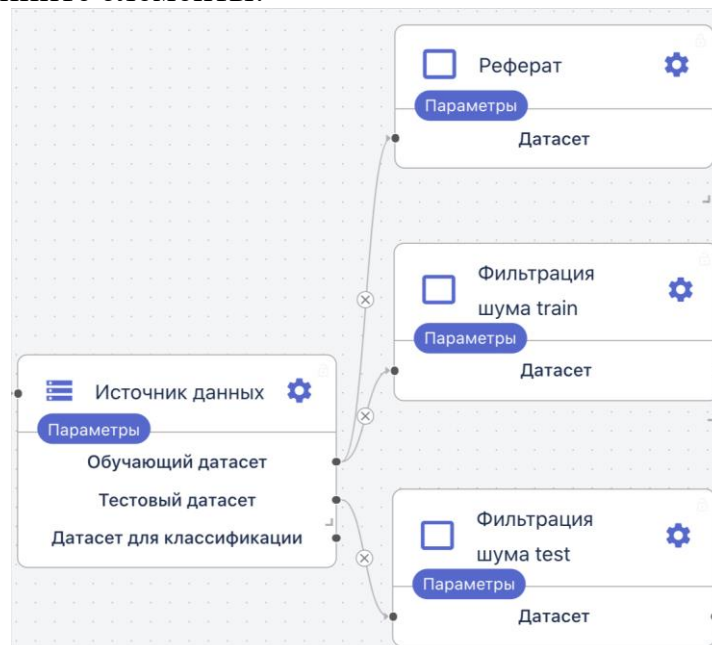


Рисунок 360 - Соединение элементов блок схемы

1.2. **Блок Процесс, Векторизация.** Добавьте на рабочую область элемент процесс. Выберите функцию «Предобработка данных» -> «Векторизация текста».

1.2.1. В параметрах блока выберите «Метод векторизации»: Word to Vec; «Максимальная размерность текста» - 25000; «Количество признаков» - 25:

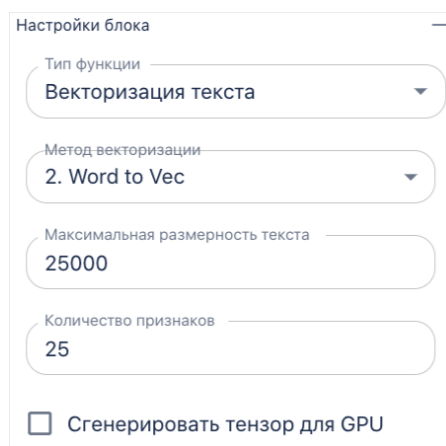


Рисунок 361 - Параметры блока «Векторизация текста»

1.2.2. Сохраните параметры блока

1.2.3. Переименуйте блок в «Векторизация»

1.2.4. Соедините элементы:

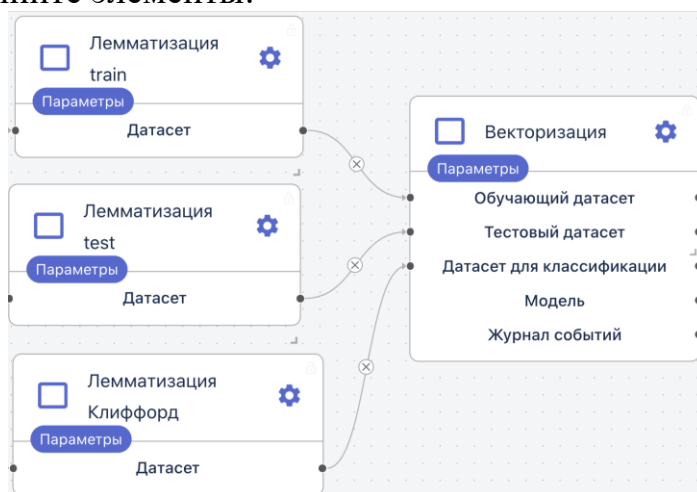


Рисунок 362 - Соединение элементов блок схемы

1.3. **Блок Процесс, Классификация.** Добавьте на рабочую область элемент процесс. Выберите функцию «Классификация» -> «Логистическая регрессия».

1.3.1. В параметрах блока выберите «Коэффициент регуляции»: 1; «Порог классификации» - 0,5; не нужно устанавливать галочки в полях «Флаг возврата вероятности при прогнозе» и «Оптимизация гиперпараметров»:

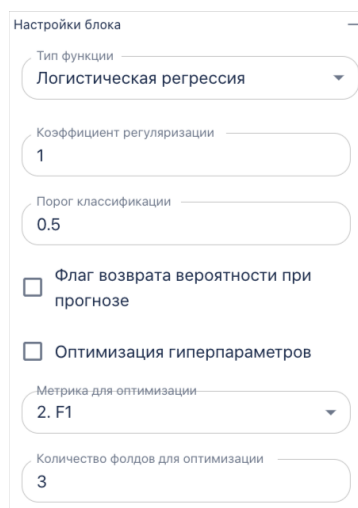


Рисунок 363 - Параметры блока «Логистическая регрессия»

- 1.3.2. Сохраните параметры блока
- 1.3.3. Переименуйте блок в «Классификация»
- 1.3.4. Соедините элементы:

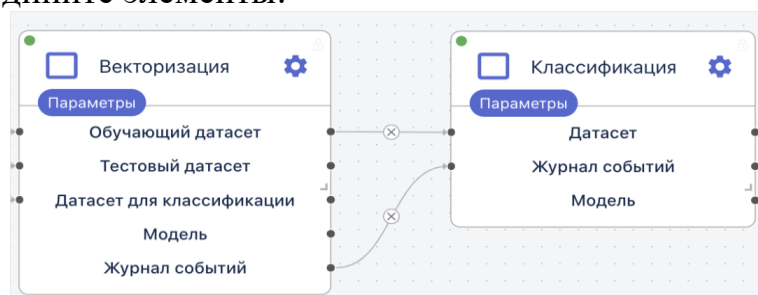
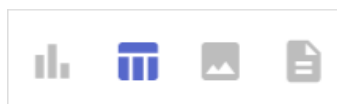


Рисунок 364 - Соединение элементов блок схемы

2. Запуск блок-схемы. Чтобы запустить *собранную блок-схему* нажмите кнопку «▶» на первом элементе блок-схемы – «Запуск». После этого внешний вид блока изменится и появится возможность создания отчета, активируйте настройку при необходимости. Повторно нажмите кнопку «Запуск», блок схема начнет обрабатывать блоки. После того, как все блоки отработают с зелеными индикаторами, процесс запуска считается успешно пройденным.

3. Визуализация результатов. После успешной отработки блок схемы на верхней панели активируются кнопки визуализации:



Нажав на иконку с таблицей, вы сможете выбрать из следующих доступных визуализаций:

- Отфильтрованные тестовые данные для каждого датасета:

Отфильтрованные текстовые данные

text	label
всё живое выехал нашего городишка повернул шоссе позади оказался грузовик	0
белая гвардия посвящается любви евгеньевне белозерской пошел мелкий снег повалил	1

- Лемматизированные текстовые данные для каждого датасета

Лемматизированные данные

text	label
все живой выезжать наш городишко повертывать шоссе позади оказываться грузовик	0
белый гвардия посвящается любовь евгеньевна белозерский пойти мелкий снег повалить	1

- Краткое содержание после применения блока «Автореферирование»:

Краткое содержание

text	label
В Лос-Анджелесе я протаранил грузовик, за рулем которого был Джералд. За рулем был Шкалик Грант, который меня заклинал не допустить, чтобы	0
На русском языке вышла книга «Капитанская дочка» Александра Турби мертвые», написанная в год Рождества Христова 1918 года.	1

2.3.13.5. Кластеризация Spark

В данном разделе рассматривается пример работы программного обеспечения с применением функций Spark. Основная цель - провести кластеризацию (обучение без учителя), а также сформировать разметку, то есть разбить объекты на 2 класса, "0" и "1". В данном примере обрабатываются данные о сетевом трафике, объектами являются сессии. Глобальная цель - обнаружить аномальные сессии, то есть решить задачу бинарной классификации. Цель пайплайна - разделить сессии на кластеры, а затем решить, какие кластеры являются аномальными. Сессии, попавшие в аномальный кластер, получают метку "1", остальные - "0". На выходе имеем размеченные данные, которые далее могут быть использованы для обучения.

1. Загрузка входных данных:
 - 1.1. В левой части главного окна на панели вкладок Системы откройте вкладку «Данные».
 - 1.2. Нажмите на кнопку «Загрузить» на верхней панели.
 - 1.3. В открывшемся окне нажмите на кнопку «Выбрать файлы» и укажите путь к заранее подготовленному файлу **1000_first_sessions.csv**, в

котором содержатся данные о сетевом трафике . Второй вариант – перенести файлы в этот раздел по технологии «drag n drop».

Выбранные файлы отобразятся в нижней части окна загрузки:

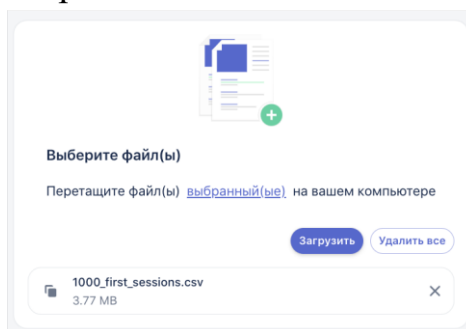


Рисунок 365 – Отображение выбранного файла

1.4. Нажмите на кнопку «Загрузить». Файл с входными данными отобразится в папке.

2. Создание новой рабочей области

Полностью блок схема представлена в Таблице.

2.1. Перейдите в пункт меню системы **Моделирование** → **Рабочая область**. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка ⊕):

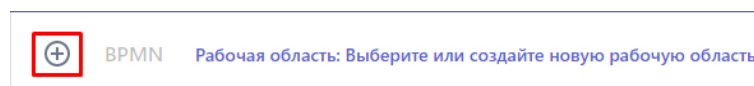


Рисунок 366 - Создание новой рабочей области

2.2. В открывшейся форме введите название новой рабочей области «Spark_Traffic» и нажмите кнопку «Создать»:

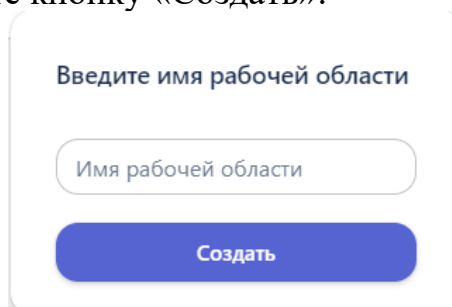


Рисунок 367 - Ввод имени рабочей области

2.3. На панели инструментов отобразится название созданной рабочей области.

3. Добавление элемента «Запуск»:

3.1. На панели инструментов блок-схемы нажмите кнопку «Добавить элемент» (кнопка **BPMN**)

3.2. В открывшейся библиотеке графических элементов выберите элемент «Запуск» :

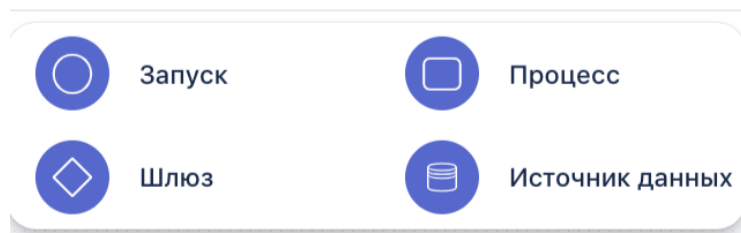


Рисунок 368 - Возможные элементы блок схемы

3.3. На рабочую область добавится элемент «Запуск»:

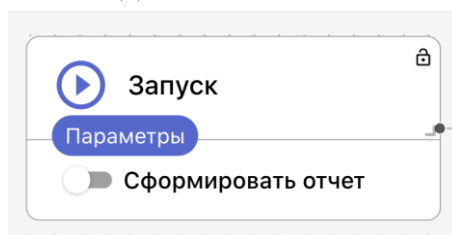


Рисунок 369 - Блок Запуск

3.4. Переименуйте элемент дважды кликнув на слово «Запуск», задайте новое название - «Кластеризация» и кликните в пустое место на рабочей области для сохранения.

4. Добавление и настройка элемента «Источник данных».

4.1. Добавьте на рабочую область элемент «Источник данных»:

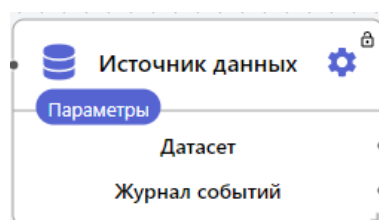



Рисунок 370 - Блок Источник данных

4.2. **Открытие настроек** элемента. На элементе «Источник данных» нажмите на кнопку . Справа откроется панель настроек элемента, где будут отображаться созданные в разделе папки и файлы с табличными данными.

4.3. **Выбор типа загрузки.** Из списка выпадающих функций выберите «Загрузка табличных данных из файла csv Spark»

4.4. **Выбор данных для загрузки в блок-схему.** Для того чтобы найти нужный файл, кликните на папку и перейдите в нее, выберите из списка файл, загруженный в Систему в шаге 1 «1000_first_sessopns.csv»,

нажмите на три точки в строке с ним и кликните «Выбрать». Внизу отобразится название выбранного файла:

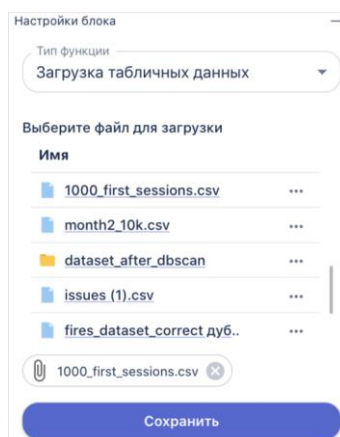


Рисунок 371 - Выбор файла

4.5. **Сохранение настроек элемента.** На панели настроек элемента нажмите на кнопку «Сохранить» (далее сохранение настроек элемента предполагается по умолчанию).

4.6. **Ввод названия элемента.** Чтобы задать название элемента нужно дважды щелкнуть левой кнопкой мыши на название элемента в рабочей области, и ввести нужное название в поле с названием, доступным для редактирования:

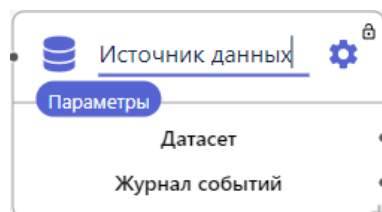


Рисунок 372 - Блок источник данных

4.7. Введите название «Загрузка Spark CSV (file)» и кликните на пустое место на рабочей области для сохранения.

4.8. **Установка соединений.** Соедините выходную точку элемента «Запуск» с входной точкой элемента «Источник данных» с помощью левой кнопка мыши:

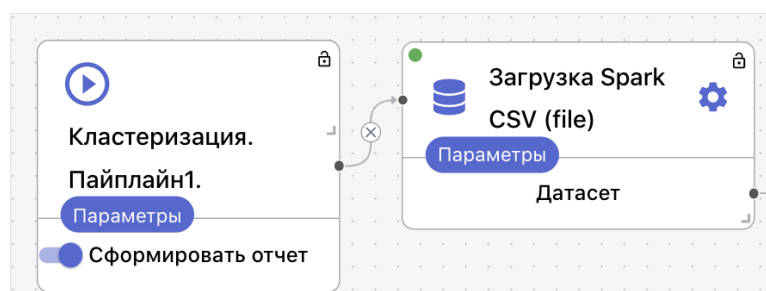


Рисунок 373 - Соединение элементов Запуск и Источник данных

5. Выбор признаков и целевых признаков. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

5.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> функция «Выбор признаков и целевых признаков».

5.2. В поле «Признаки» укажите: поочередно следующие признаки, нажимая Enter после ввода каждого: source_ip, destination_ip, source_port, destination_port, bytes, packages_count. Или вы можете найти в списке исходный файл 1000_first_sessions.csv, нажать на три точки в строке с его названием и кликнуть «Выгрузить признаки», тогда система автоматически заполнит поле «Признаки» всем вариантами из датасета и вам останется только убрать лишние.

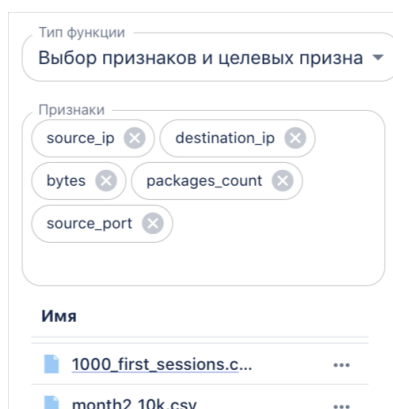


Рисунок 374 - Параметры блока «Выбор признаков и целевых признаков»

5.3. На панели настроек элемента нажмите на кнопку «Сохранить».

5.4. Измените название элемента на «Выбор признаков».

5.5. Соедините с элементы:

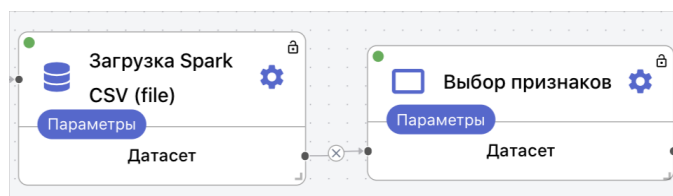


Рисунок 375 - Соединение элементов Источник данных и Выбор признаков

6. Порядковое кодирование признаков. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

6.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> группа «Препроцессинг» -> функция «Порядковое кодирование признаков».

6.2. В поле «Выбранные признаки» вместе с квадратными скобками введите следующие признаки, нажимая Enter после ввода каждого: source_ip, destination_ip, source_port, destination_port. Или вы можете найти в списке исходный файл 1000_first_sessions.csv, нажать на три точки в строке с его названием и кликнуть «Выгрузить признаки», тогда система автоматически заполнит поле «Признаки» всем вариантами из датасета и вам останется только убрать лишние:

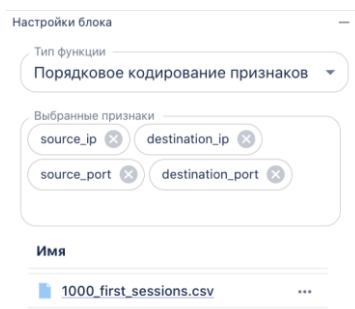


Рисунок 376 - Параметры блока «Порядковое кодирование»

6.3. На панели настроек элемента нажмите на кнопку «Сохранить».

6.4. Измените название элемента на «Label encoder».

6.5. Соедините с элементы:

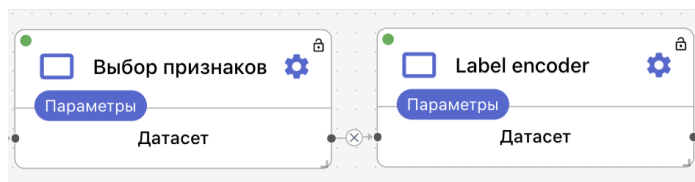


Рисунок 377 - Соединение элементов Выбор признаков и Label encoder

7. Нормализация признаков. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

7.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> группа «Препроцессинг» -> функция «Нормализация признаков».

7.2. На панели настроек элемента нажмите на кнопку «Сохранить».

7.3. Измените название элемента на «Нормализация».

7.4. Соедините с элементы:

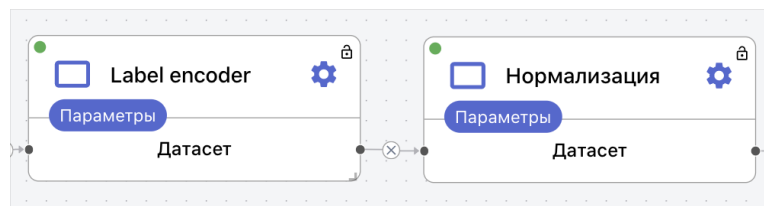


Рисунок 378 - Соединение элементов Label encoder и Нормализация

8. Кластеризация Spark DBSCAN. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

8.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> группа «Кластеризация» -> функция «Кластеризация Spark DBSCAN».

8.2. В поле «Порог для отнесения кластера к аномалиями» укажите: 100.

8.3. В поле «Радиус» укажите: 0,1.

8.4. В поле «Число соседей» укажите: 4.

8.5. В поле «Метрика расстояния» из выпадающего списка выберите: «Евклидово»

8.6. Установите галочку для параметра «Флаг векторизации признаков».

8.7. В поле столбец для группировки перед векторами впишите: `session_id`

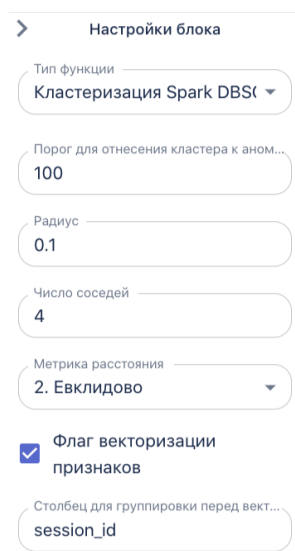


Рисунок 379 - Параметры блока «Кластеризация Spark DBSCAN»

8.8. На панели настроек элемента нажмите на кнопку «Сохранить».

8.9. Измените название элемента на «Кластеризация Spark DBSCAN».

8.10. Соедините с элементами:

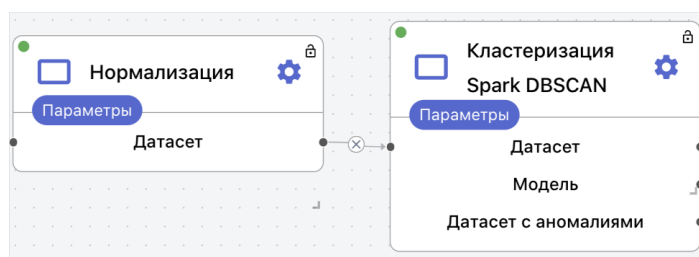


Рисунок 380 - Соединение элементов Нормализация и Кластеризация Spark DBSCAN

9. Сохранение датасета Spark. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

9.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> функция «Сохранение датасета Spark в CSV»

9.2. В поле «Путь до директории для датасета» укажите название папки в разделе «Данные», куда будет сохраняться датасет

9.3. В поле «Название датасета» пропишите вручную необходимое наименование, например, dataset_after_dbscan

9.4. Не ставьте галочку для параметра «Добавить данные к датасету»

9.5. Сохраните настройки

9.6. Измените название блока на «Датасет в CSV»

10. Сохранение модели Spark. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

10.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> функция «Сохранение модели Spark»

10.2. В поле «Название» пропишите вручную необходимое наименование, например, DBSCAN_spark_model:

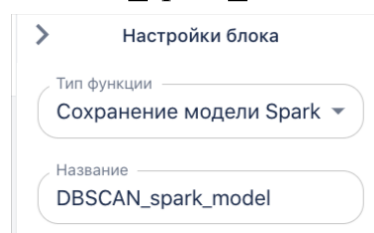


Рисунок 381 - Параметры блока «Сохранение модели Spark»

10.3. Сохраните настройки

10.4. Измените название блока на «Сохранение модели DBSCAN»

11. Сохранение датасета Spark. Чтобы в загруженном датасете выделить признаки и целевые признаки нужно добавить на рабочую область элемент «Процесс» и настроить его:

11.1. На панели свойств элемента выбрать из списка функцию: тип функции «Spark» -> функция «Сохранение датасета Spark в CSV»

11.2. В поле «Путь до директории для датасета» укажите название папки в разделе «Данные», куда будет сохраняться датасет

11.3. В поле «Название датасета» пропишите вручную необходимое наименование, например, dataset_with_anomalies_after_dbscan

11.4. Не ставьте галочку для параметра «Добавить данные к датасету»

11.5. Сохраните настройки

11.6. Измените название блока на «Датасет в CSV (аномалии)»

12. Соединение элементов. Соедините элементы, созданные в пункте 9, 10 и 11 следующим образом:

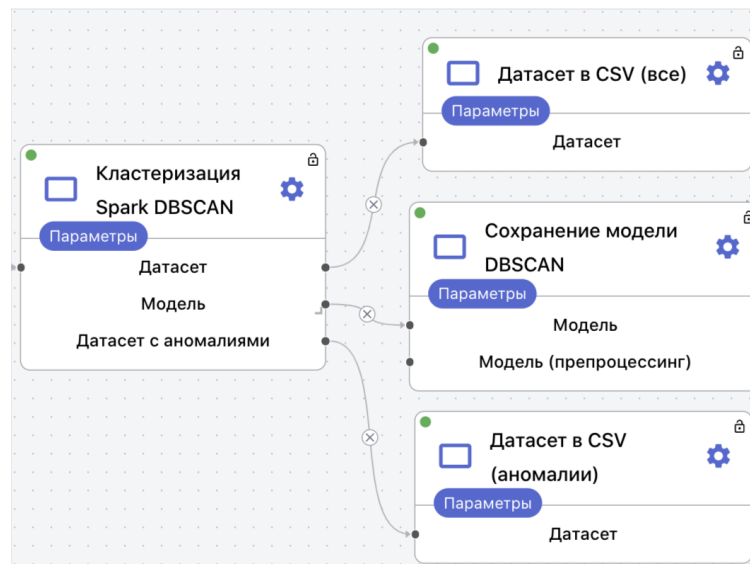



Рисунок 382 - Соединение элементов Классификация Spark DBSCAN и блоков сохранения

13. Запуск пайплайна. Чтобы запустить блок схему нажмите на кнопку  на первом элементе «Запуск» собранной блок-схемы. При этом

отображение элемента «Запуск» изменится и появится опция Сформировать отчет:

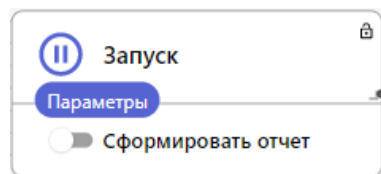
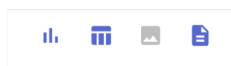


Рисунок 383 - Блок Запуск

Если активировать параметр «Сформировать отчет», в результате запуска пайплайна будет создан отчет.

14. Визуализация результатов. После того как все элементы схемы будут успешно обработаны, на панели инструментов появляются кнопки:



Вы должны увидеть следующие визуализации:

14.1. График «Spark DBSCAN». Этот график позволяет построить диаграмму рассеяния для трех и более пар признаков. Чем больше размер точек или пузырей на диаграмме - тем больше взаимосвязь между признаками. График можно удалять, приближать и т д:

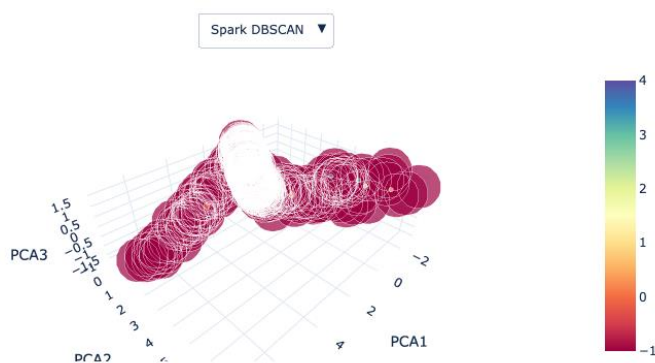


Рисунок 384 - График Spark DBSCAN

14.2. График «Объем кластеров». Круговая диаграмма показывает объем кластеров - т.е. сколько % объектов входит в каждый отдельный кластер. Имена кластеров сортируются в зависимости от их веса в общем проценте. Так в нашем примере большинство объектов попали в кластер -1 и составили 96,1%:

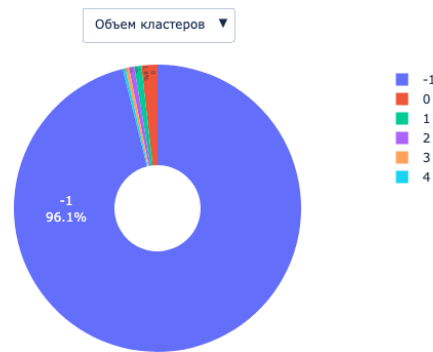


Рисунок 385 - График Объем кластеров

14.3. Таблица «Датасет 1000_first_sessions.csv». Это исходный датасет, загруженный в качестве источника данных:

Датасет 1000_first_sessions.csv

session_id	date_time	source_ip	source_port	destination_ip	destination_port	bytes	packages_count
64210550153181.0	2022-04-20 15:03:27.404	10.174.174.236	52263	10.253.252.11	9400	523	5
64210550153181.0	2022-04-20 15:03:27.405	10.253.252.11	9400	10.174.174.236	52263	1523	3
64210550153181.0	2022-04-20 15:05:07.482	10.174.174.236	52263	10.253.252.11	9400	40	2
64210550153181.0	2022-04-20 15:05:07.482	10.253.252.11	9400	10.174.174.236	52263	20	1
64210550153181.0	2022-04-20 15:05:27.734	10.174.174.236	52302	10.253.252.11	9400	503	4
64210550153181.0	2022-04-20 15:05:27.734	10.253.252.11	9400	10.174.174.236	52302	1523	3
64210550153181.0	2022-04-20 15:07:07.810	10.174.174.236	52302	10.253.252.11	9400	40	2
64210550153181.0	2022-04-20 15:07:07.810	10.253.252.11	9400	10.174.174.236	52302	20	1
64210550153181.0	2022-04-20 15:07:28.110	10.174.174.236	52347	10.253.252.11	9400	503	4
64210550153181.0	2022-04-20 15:07:28.110	10.253.252.11	9400	10.174.174.236	52347	1523	3

Рисунок 386 - Таблица Датасет 1000_first_sessions.csv

14.4. Таблица «Количество объектов в каждом кластере». Данная таблица показывает, сколько объектов попало в разные кластеры:

Количество объектов в каждом кластере

label	volume
4	3
3	4
2	6
1	8
0	18
-1	961

Рисунок 387 - Таблица «Количество объектов в каждом кластере»

14.5. Таблица «Датасет dataset_after_dbscan» отображает преобразованный датасет, где содержатся данные по кластеризации:

Dataser dataset_after_dbscan

label	session_id	concat[0]	concat[1]	concat[2]	concat[2822]	concat[2823]	concat[2824]	concat[2825]	is_anomaly
-1	64210550153181	0.21713441610338304	0.001485884073190391	0.15901444852352142	0	0	0	0	0
-1	76771566702416	0.875923216342926	0.001485884073190391	0.783297061920166	0	0	0	0	0
-1	129323973953981	0.10930576175451279	0.016344724223017693	0.8732436299324036	0	0	0	0	0
-1	136859227321959	0.6174298524856567	0.004457652103155851	0.3076390326023102	0	0	0	0	0
-1	148877861840344	0.0014771048445254564	0.2228526135399574	0.0000989511725199968	0	0	0	0	0
-1	210567685532991	0.005908419378101826	0.005943536292761564	0.000692657835315913	0.9161864188652039	0.0006721075624227524	0.0001346979024674510	0.0002271714988164604	0
-1	217590991923358	0	0.018318449380222912	0	0	0	0	0	0
-1	218100980842663	0.00738552455457928	0.8671619415283203	0.00019790223450399935	0	0	0	0	0
-1	253330641158839	0.005908419378101826	0.35215452313423157	0.000692657835315913	0	0	0	0	0
-1	260730379318740	0.33087149262428284	0.9034175276756287	0.0003988044690079987	0	0	0	0	0
-1	261652820257666	0.31905466318130493	0.2823179622276306	0.017910152673721313	0	0	0	0	0
-1	274134409307370	0	0.037147101014852524	0.00029685336630791426	0	0	0	0	0
-1	347983599388379	0.32348597049713135	0.33878156542778015	0.225509598885120392	0	0	0	0	0
-1	385729102920192	0.18316100537776947	0.7503714561462402	0.00019790223450399935	0	0	0	0	0
-1	432514239859384	0.478581964969635	0.21396730840206146	0.08737383782863617	0	0	0	0	0
-1	461858665095623	0.7223042845726013	0.29420503973960876	0.4494359791276839	0	0	0	0	0
-1	461858665095623	0.7223042845726013	0.29420503973960876	0.4494359791276839	0	0	0	0	0

Рисунок 388 - Таблица «Датасет dataset_after_dbscan»

15. Сохранение файлов в раздел данные. В результате обработки блок схемы в разделе «Данные» должны появиться следующие папки:

15.1. dataset_after_dbscan, в которой отображается сохраненный датасет в формате .csv. Остальные файлы являются системными

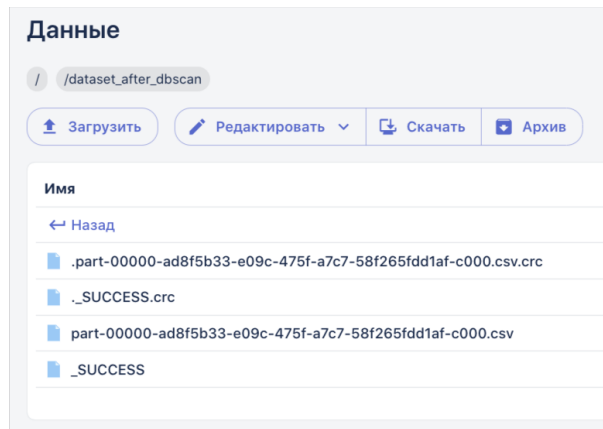


Рисунок 389 - Папка с сохраненным датасетом в разделе Данные

15.2. dataset_with_anomalies_after_dbscan, в которой отобразится сохраненный датасет в формате .csv. Остальные файлы являются системными:

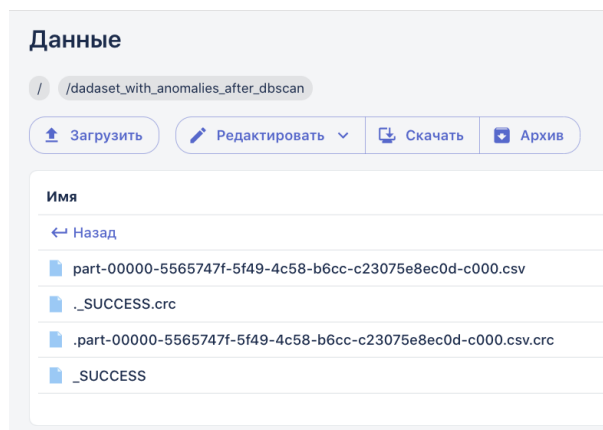


Рисунок 390 - Папка с сохраненным датасетом в разделе Данные

2.3.13.6. Классификация текстовых данных с использованием слоя нейронной сети LSTM

1. Загрузка входных данных:

1.1. В боковом меню главного окна системы выберите пункт «Данные».

1.2. Перейдите в пустую папку и затем нажмите кнопку «Загрузить» на верхней панели.

1.3. В открывшемся окне нажмите кнопку «Выбрать файлы» и укажите путь к родительской папке с дочерними папками, содержащими 2 заранее подготовленных PDF-файла литературных произведений разных авторов.

1.4. Нажмите на кнопку «Загрузить». Выбранная папка будет загружена.

1.5. Перейдите в другую папку и затем нажмите кнопку «Загрузить» на верхней панели.

1.6. В открывшемся окне нажмите кнопку «Выбрать файлы» и укажите путь к другой родительской папке с дочерними папками, содержащими 2 заранее подготовленных PDF-файла других литературных произведений этих же авторов.

1.7. Нажмите на кнопку «Загрузить». Выбранная папка будет загружена.

2. Создание новой рабочей области


2.1. В боковом меню главного окна системы выберите **Моделирование** → **Рабочая область**. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка 

Рисунок 15.6.1 - Создание рабочей области

2.2. В открывшейся форме введите название новой рабочей области «LSTM» и нажмите кнопку «Создать»:

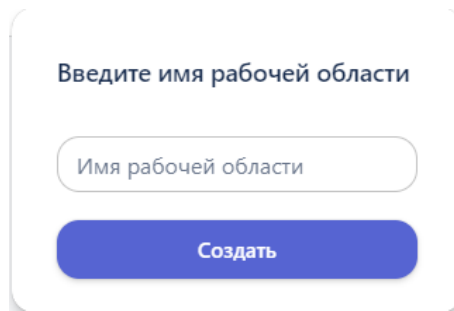


Рисунок 15.6.2 - Ввод имени рабочей области

2.3. На панели инструментов отобразится название созданной рабочей области.

3. Добавление элемента «Запуск»:

3.1. На панели инструментов блок-схемы нажмите кнопку «Добавить элемент» (кнопка **BPMN**).

3.2. В открывшейся библиотеке графических элементов выберите элемент «Запуск»:

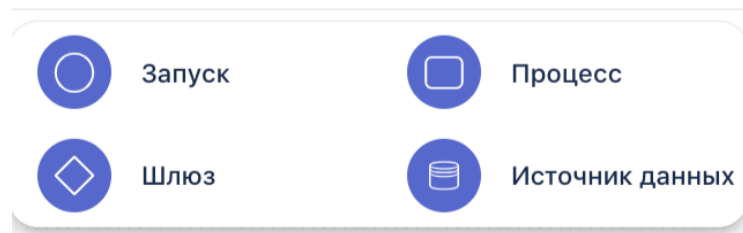


Рисунок 15.6.3 - Возможные элементы блок схемы

3.3. На рабочую область будет добавлен элемент «Запуск»:

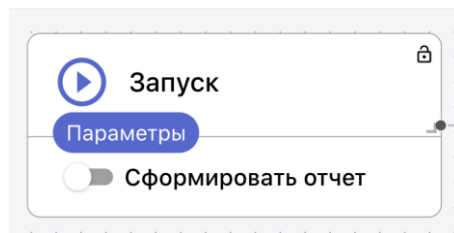


Рисунок 15.6.4 - Блок Запуск

4. Добавление и настройка элемента «Источник данных».

4.1. Добавьте на рабочую область элемент «Источник данных»:

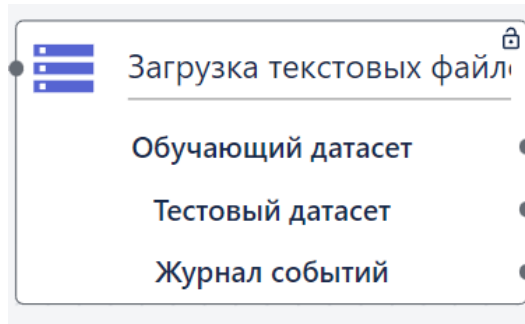



Рисунок 15.6.5 - Блок Источник данных

4.2. **Открытие настроек элемента.** На элементе «Источник данных» нажмите на кнопку . Справа откроется панель настроек элемента, где будут отображаться созданные папки и файлы с табличными данными.

4.3. **Выбор типа загрузки.** Из списка выпадающих функций выберите «Загрузка текстовых файлов для классификации».

4.4. **Выбор данных для загрузки в блок-схему.** В списках «Группа обучающих текстов» и «Группа тестовых текстов» выберите папки с файлами, загруженные в Систему в шаге 1, нажмите на три точки и кликните «Выбрать».

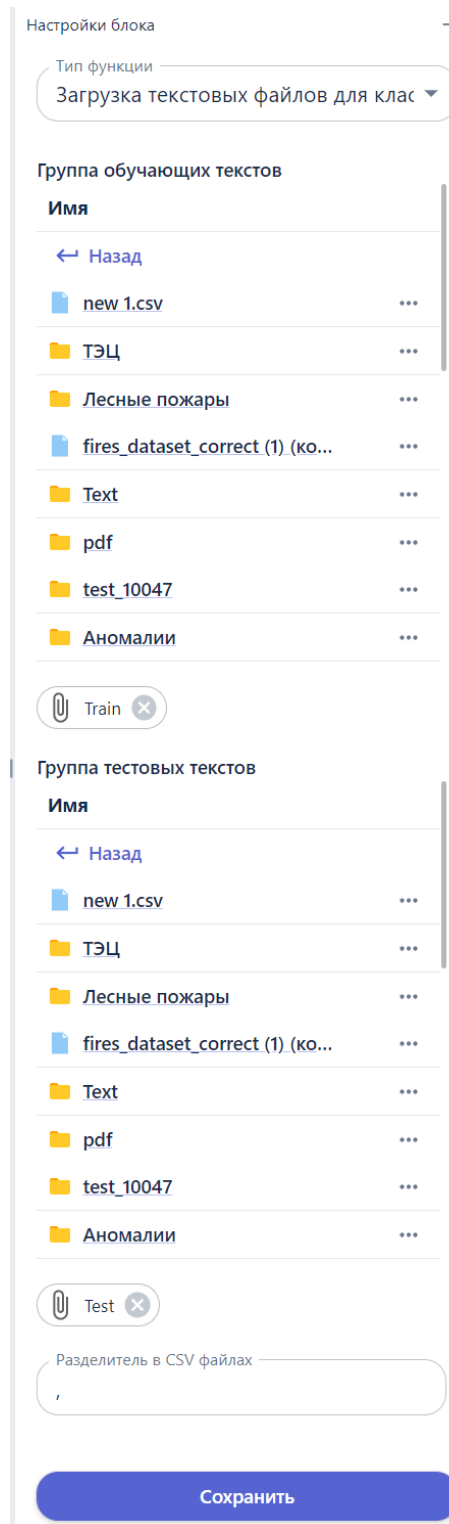


Рисунок 15.6.6 - Выбор файла

4.5. **Сохранение настроек элемента.** На панели настроек элемента нажмите на кнопку «Сохранить» (далее сохранение настроек элемента предполагается по умолчанию).

4.6. **Ввод названия элемента.** Чтобы задать название элемента нужно дважды щелкнуть левой кнопкой мыши на название элемента в рабочей области, и ввести нужное название в поле с названием, доступным для редактирования:

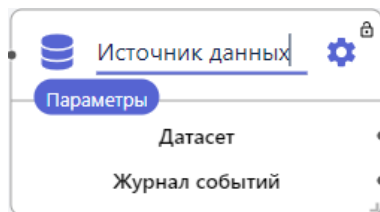



Рисунок 15.6.7 - Блок источник данных

4.7. Введите название «Загрузка текстовых данных» и кликните на пустое место на рабочей области для сохранения.

5. Добавление и настройка элемента «Процесс».

5.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настройте элемент:

5.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Фильтрация текстового шума».

5.1.2. Нажать кнопку «Сохранить».

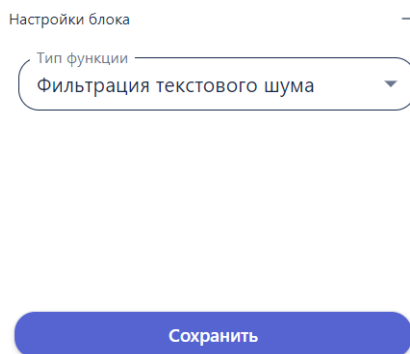



Рисунок 15.6.8 - Настройка блока «Процесс»

6. Добавление и настройка элемента «Процесс».

6.1. Повторите действия, описанные на шаге выше, чтобы продублировать созданный элемент «Процесс».

6.2. Настройте элемент, как указано на шаге выше.

7. Добавление и настройка элемента «Процесс».

7.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настройте элемент:

7.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Лемматизация текста».

7.1.2. Нажать кнопку «Сохранить».

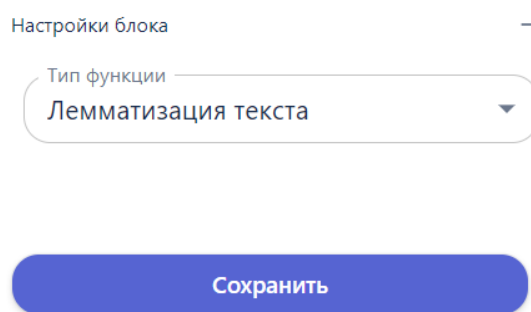



Рисунок 15.6.9 - Блок процесс

8. Добавление и настройка элемента «Процесс».

8.1. Повторите действия, описанные на шаге выше, чтобы продублировать созданный элемент «Процесс».

8.2. Настройте элемент, как указано на шаге выше.

9. Добавление и настройка элемента «Процесс».

9.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

9.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Векторизация текста».

9.1.2. В поле «Максимальная размерность текста» выберите 10000.

9.1.3. В поле «Количество признаков» выберите 10.

9.1.4. Нажать кнопку «Сохранить».

Настройки блока

Тип функции
Векторизация текста

Метод векторизации

Максимальная размерность текста
10000


Количество признаков
10

Сгенерировать тензор для GPU

Сохранить

Рисунок 15.6.10 - Блок процесс

10. Выбор функции и настройка параметров:

- 10.1. Добавить элемент «Процесс» (кнопка ).
- 10.2. В карточке элемента выбрать из списка функцию: раздел «Классификация» -> функция «Классификация (табличные данные)».
- 10.3. В секции «Добавить слой» нажать кнопку + и затем в поле «Слой» выбрать значение «4. LSTM»
- 10.4. В поле «Число нейронов» ввести количество нейронов, соответствующее размерности текста.
- 10.5. В поле «Функция активации» выбрать значение «Sigmoid» (требуемое значение рекомендуется выбирать практическим путем).
- 10.6. В поле «Функция активации (рекурсия)» выбрать значение «Sigmoid» (требуемое значение рекомендуется выбирать практическим путем).
- 10.7. В секции «Добавить слой» нажать кнопку + и затем в поле «Слой» выбрать значение «1. Dense».
- 10.8. В поле «Число нейронов» ввести 1.
- 10.9. В поле «Функция активации» выбрать значение «Sigmoid» (требуемое значение рекомендуется выбирать практическим путем).
- 10.10. В поле «Количество эпох» установить значение 5.
- 10.11. В поле «Размер мини-батча» установить значение 16.
- 10.12. В поле «Алгоритм градиентного спуска» выбрать значение «3. Adam».
- 10.13. В поле «Шаг градиентного спуска» установить значение 0.001.
- 10.14. В поле «Порог классификации» задать значение 0.5.

- 10.15. В поле “Количество фолдов для оптимизации” задать значение
- 3.
- 10.16. Значения остальных параметров оставить без изменения.
 - 10.17. Нажать кнопку “Сохранить”.

Настройки блока

Тип функции
Классификация (табличные данные)

Добавить слой +

Слой
4. LSTM -

Число нейронов
10

Функция активации
2. sigmoid

Функция активации (рекурсия)
2. sigmoid

Доля нейронов для Dropout
0

Доля нейронов для Dropout (рекурсия)
0

Слой
1. Dense -

Число нейронов
1

Функция активации
3. Sigmoid

Количество эпох
5

Размер мини-батча
16

Метрика для обучения

Алгоритм градиентного спуска
3. Adam

Шаг градиентного спуска
0.001

Функция потерь

Перемешивать выборку перед обучением

Порог классификации
0.5

Флаг возврата вероятности при прогнозе

?

Оптимизация гиперпараметров


Метрика для оптимизации

Количество фолдов для оптимизации
3

Сохранить

Рисунок 15.6.11 - Блок процесс

11. Добавление и настройка элемента «Процесс».

11.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

11.1.1. В карточке элемента в списке «Тип функции» в разделе «Машинное обучение» выбрать «Валидация модели (update)».

11.1.2. В поле «Метрика» выбрать «5. Accuracy».

11.1.3. Нажать кнопку «Сохранить».

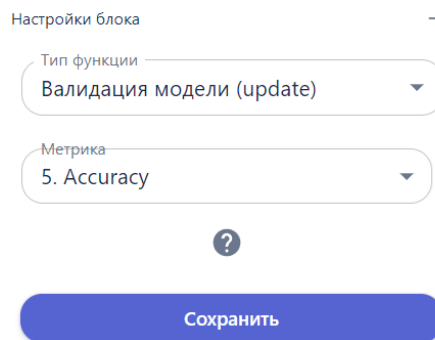


Рисунок 15.6.12 - Блок процесс

12. Установка соединений. Соедините выходные и входные точки элементов, как показано на рисунке ниже.

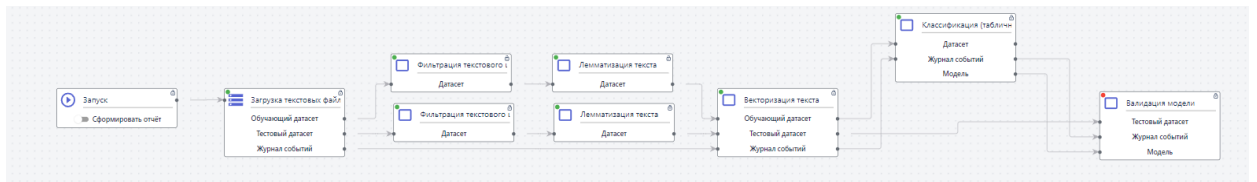



Рисунок 15.6.13 - Установка соединений между блоками

13. **Запуск пайплайна.** Чтобы запустить блок схему нажмите на кнопку  на первом элементе «Запуск» собранной блок-схемы. При этом отображение элемента «Запуск» изменится и появится опция Сформировать отчет:

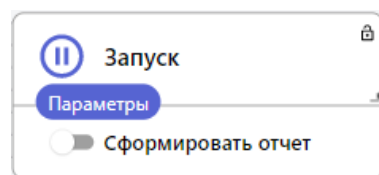


Рисунок 15.6.14 - Блок Запуск

Если активировать параметр «Сформировать отчет», в результате запуска пайплайна будет создан отчет.

14. Визуализация результатов. После того как все элементы схемы будут успешно обработаны, на панели инструментов появляются



кнопки:

Нажмите на вторую слева кнопку и выберите команду «Отфильтрованные текстовые данные». Отобразится таблица с результатами обработки.

15. Извлечение текстового слоя из текстовых данных

Данный сценарий предполагает извлечение текстового слоя из текстовых данных в формате PDF.

Для решения задачи выполните следующие действия:

16. Загрузка входных данных:

16.1. В боковом меню главного окна системы выберите пункт «Данные».

16.2. Перейдите в пустую папку и затем нажмите кнопку «Загрузить» на верхней панели.

16.3. В открывшемся окне нажмите кнопку «Выбрать файлы» и укажите путь к родительской папке с дочерними папками, содержащими 2 заранее подготовленных PDF-файла литературных произведений разных авторов.


16.4. Нажмите на кнопку «Загрузить». Выбранная папка будет загружена.

16.5. Перейдите в другую папку и затем нажмите кнопку «Загрузить» на верхней панели.

16.6. В открывшемся окне нажмите кнопку «Выбрать файлы» и укажите путь к другой родительской папке с дочерними папками, содержащими 2 заранее подготовленных PDF-файла других литературных произведений этих же авторов.

16.7. Нажмите на кнопку «Загрузить». Выбранная папка будет загружена.

17. Создание новой рабочей области

17.1. В боковом меню главного окна системы выберите **Моделирование** → **Рабочая область**. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка ):

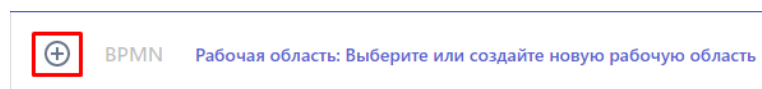


Рисунок 15.7.1 - Создание рабочей области

17.2. В открывшейся форме введите название новой рабочей области «Классификация текстов (new lib)» и нажмите кнопку «Создать»:

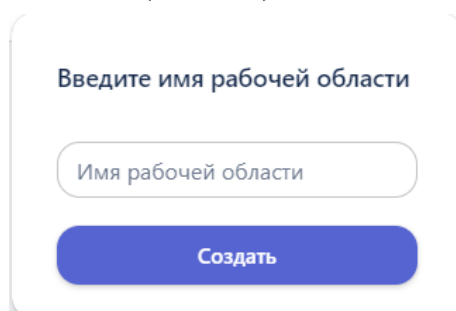


Рисунок 15.7.2 - Ввод имени рабочей области

17.3. На панели инструментов отобразится название созданной рабочей области.

18. Добавление элемента «Запуск»:

18.1. На панели инструментов блок-схемы нажмите кнопку «Добавить элемент» (кнопка **BPMN**).

18.2. В открывшейся библиотеке графических элементов выберите элемент «Запуск»:

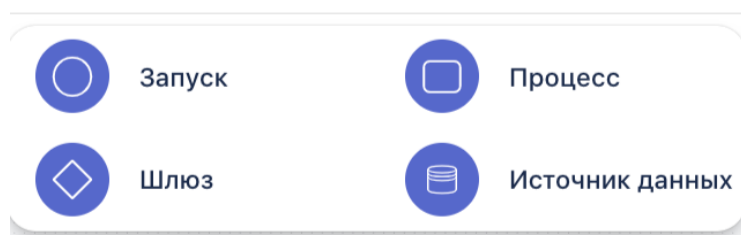


Рисунок 15.7.3 - Возможные элементы блок схемы

18.3. На рабочую область будет добавлен элемент «Запуск»:

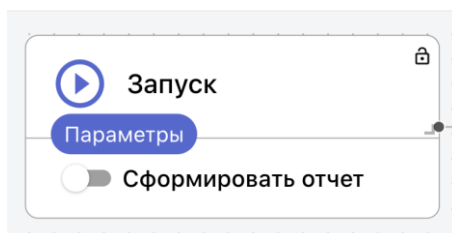


Рисунок 15.7.4 - Блок Запуск

19. Добавление и настройка элемента «Источник данных».

19.1. Добавьте на рабочую область элемент «Источник данных»:

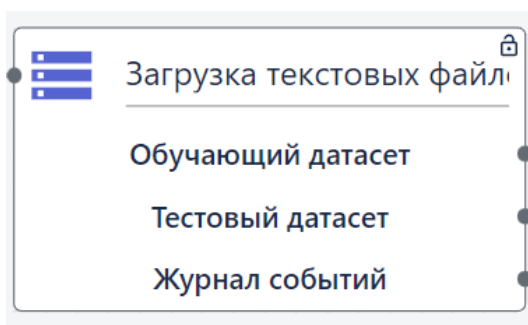



Рисунок 15.7.5 - Блок Источник данных

19.2. **Открытие настроек элемента.** На элементе «Источник данных» нажмите на кнопку . Справа откроется панель настроек элемента, где будут отображаться созданные папки и файлы с табличными данными.

19.3. **Выбор типа загрузки.** Из списка выпадающих функций выберите «Загрузка текстовых файлов для классификации».

19.4. **Выбор данных для загрузки в блок-схему.** В списках «Группа обучающих текстов» и «Группа тестовых текстов» выберите папки с файлами, загруженные в Систему в шаге 1, нажмите на три точки и кликните «Выбрать».

Настройки блока

Тип функции
Загрузка текстовых файлов для клас ▾

Группа обучающих текстов

Имя

← Назад

new 1.csv ...

ТЭЦ ...

Лесные пожары ...

fires_dataset_correct (1) (ко... ...

Text ...

pdf ...

test_10047 ...

Аномалии ...

📎 Train ✕

Группа тестовых текстов

Имя

← Назад

new 1.csv ...

ТЭЦ ...

Лесные пожары ...

fires_dataset_correct (1) (ко... ...

Text ...

pdf ...

test_10047 ...

Аномалии ...

📎 Test ✕

Разделитель в CSV файлах

,

Сохранить

Рисунок 15.7.6 - Выбор файла

19.5. **Сохранение настроек элемента.** На панели настроек элемента нажмите на кнопку «Сохранить» (далее сохранение настроек элемента предполагается по умолчанию).

19.6. **Ввод названия элемента.** Чтобы задать название элемента нужно дважды щелкнуть левой кнопкой мыши на название элемента в рабочей области, и ввести нужное название в поле с названием, доступным для редактирования:

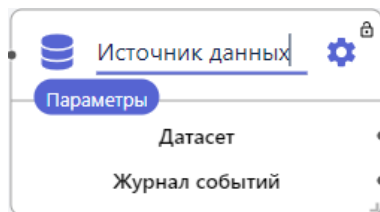



Рисунок 15.8.7 - Блок источник данных

19.7. Введите название «Загрузка текстовых данных» и кликните на пустое место на рабочей области для сохранения.

20. Добавление и настройка элемента «Процесс».

20.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настройте элемент:

20.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Фильтрация текстового шума».

20.1.2. Нажать кнопку «Сохранить».

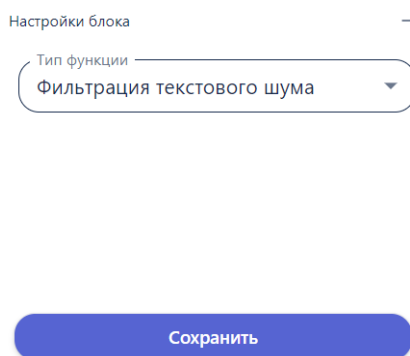



Рисунок 15.8.8 - Настройка блока «Процесс»

21. Добавление и настройка элемента «Процесс».

21.1. Повторите действия, описанные на шаге выше, чтобы продублировать созданный элемент «Процесс».

21.2. Настройте элемент, как указано на шаге выше.

22. Добавление и настройка элемента «Процесс».

22.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настройте элемент:

22.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Лемматизация текста».

22.1.2. Нажать кнопку «Сохранить».

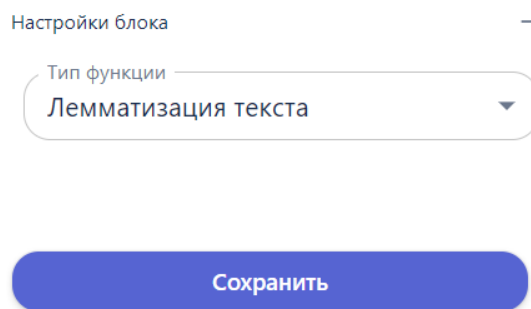



Рисунок 15.8.9 - Блок процесс

23. Добавление и настройка элемента «Процесс».

23.1. Повторите действия, описанные на шаге выше, чтобы продублировать созданный элемент «Процесс».

23.2. Настройте элемент, как указано на шаге выше.

24. Добавление и настройка элемента «Процесс».

24.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

24.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Векторизация текста».

24.1.2. В поле «Максимальная размерность текста» выберите 10000.

24.1.3. В поле «Количество признаков» выберите 10.

24.1.4. Нажать кнопку «Сохранить».

Настройки блока

Тип функции
Векторизация текста

Метод векторизации

Максимальная размерность текста
10000


Количество признаков
10

Сгенерировать тензор для GPU

Сохранить

Рисунок 15.8.10 - Блок процесс

25. Добавление и настройка элемента «Процесс».

25.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

25.1.1. В карточке элемента в списке «Тип функции» в разделе «Классификация» выбрать «Логистическая регрессия».

25.1.2. В поле «Коэффициент регуляризации» выберите 1.

25.1.3. В поле «Порог классификации» выберите 0,5.

25.1.4. В поле «Количество фолдов для оптимизации» выберите 3.

25.1.5. Нажать кнопку «Сохранить».

Настройки блока

Тип функции
Логистическая регрессия

Коэффициент регуляризации
1

Порог классификации
0.5

Флаг возврата вероятности при прогнозе

Оптимизация гиперпараметров


Метрика для оптимизации

Количество фолдов для оптимизации
3

Сохранить

Рисунок 15.8.11 - Блок процесс

26. Добавление и настройка элемента «Процесс».

26.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

26.1.1. В карточке элемента в списке «Тип функции» в разделе «Классификация» выбрать «Модель XGBClassifier».

26.1.2. В поле «Глубина дерева» выберите 2.

26.1.3. В поле «Количество базовых моделей» выберите 100.

26.1.4. В поле «Порог классификации» выберите 0,5.

26.1.5. В поле «Количество фолдов для оптимизации» выберите 3.

26.1.6. Нажать кнопку «Сохранить».

Настройки блока

Тип функции
Модель XGBClassifier

Глубина дерева
2

Количество базовых моделей
100

Порог классификации
0.5

Флаг возврата вероятности при прогнозе

Оптимизация гиперпараметров


Метрика для оптимизации

Количество фолдов для оптимизации
3

Сохранить

Рисунок 15.8.12 - Блок процесс

27. Добавление и настройка элемента «Процесс».

27.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

27.1.1. В карточке элемента в списке «Тип функции» в разделе «Машинное обучение» выбрать «Валидация модели (update)».

27.1.2. В поле «Метрика» выбрать «5. Accuracy».

27.1.3. Нажать кнопку «Сохранить».

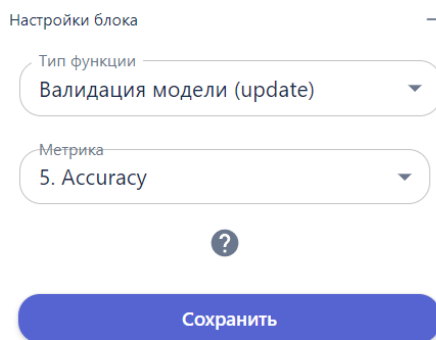


Рисунок 15.8.12 - Блок процесс

28. Установка соединений. Соедините выходные и входные точки элементов, как показано на рисунке ниже.

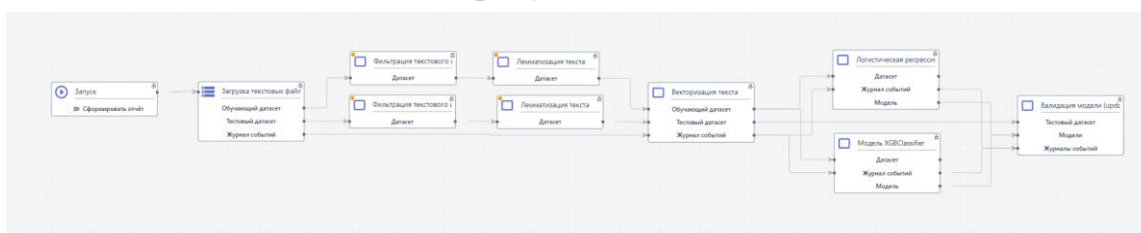



Рисунок 15.8.13 - Установка соединений между блоками

29. Запуск пайплайна. Чтобы запустить блок схему нажмите на кнопку  на первом элементе «Запуск» собранной блок-схемы. При этом отображение элемента «Запуск» изменится и появится опция Сформировать отчет:

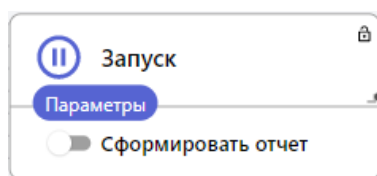


Рисунок 15.8.14 - Блок Запуск

Если активировать параметр «Сформировать отчет», в результате запуска пайплайна будет создан отчет.

30. Визуализация результатов. После того как все элементы схемы будут успешно обработаны, на панели инструментов появляются



кнопки:

Нажмите на вторую слева кнопку и выберите команду «Отфильтрованные текстовые данные». Отобразится таблица с результатами обработки.

2.3.13.7. Заполнение и работа с пропусками в табличных данных

Данный сценарий предполагает работу с пропущенными значениями или пропусками в табличных данных в формате CSV. Система находит пропуски в указанных столбцах и удаляет строки, в которых содержатся пропуски.

Для решения задачи выполните следующие действия:

1. Загрузка входных данных:

1.1. В боковом меню главного окна системы выберите пункт «Данные».

1.2. Нажмите кнопку «Загрузить» на верхней панели.

1.3. В открывшемся окне нажмите кнопку «Выбрать файлы» и укажите путь к заранее подготовленному файлу **train.csv**, в котором содержатся данные с пропусками.

1.4. Нажмите на кнопку «Загрузить». Файл с входными данными отобразится в папке.

2. Создание новой рабочей области

2.1. В боковом меню главного окна системы выберите **Моделирование** → **Рабочая область**. На панели инструментов блок-схемы нажмите кнопку «Создание рабочей области» (кнопка ⊕):

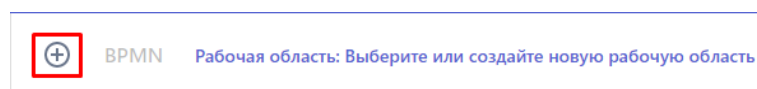


Рисунок 15.8.1 - Создание рабочей области

2.2. В открывшейся форме введите название новой рабочей области «Заполнение пропусков (new lib)» и нажмите кнопку «Создать»:

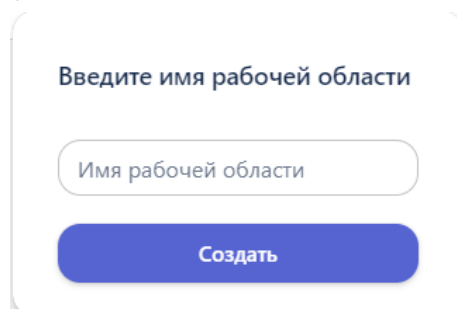


Рисунок 15.8.2 - Ввод имени рабочей области

2.3. На панели инструментов отобразится название созданной рабочей области.

3. Добавление элемента «Запуск»:

3.1. На панели инструментов блок-схемы нажмите кнопку «Добавить элемент» (кнопка **BPMN**).

3.2. В открывшейся библиотеке графических элементов выберите элемент «Запуск»:

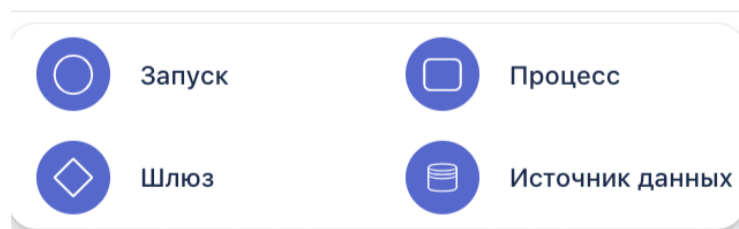


Рисунок 15.8.3 - Возможные элементы блок-схемы

3.3. На рабочую область будет добавлен элемент «Запуск»:

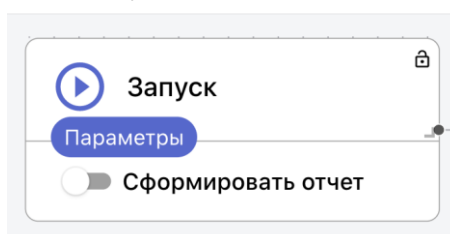


Рисунок 15.8.4 - Блок Запуск

4. Добавление и настройка элемента «Источник данных».

4.1. Добавьте на рабочую область элемент «Источник данных»:

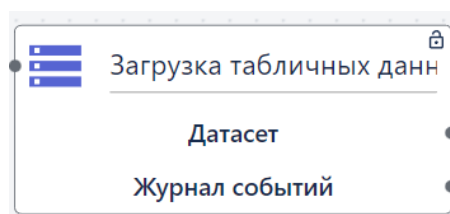



Рисунок 15.8.5 - Блок Источник данных

4.2. **Открытие настроек элемента.** На элементе «Источник данных» нажмите на кнопку . Справа откроется панель настроек элемента, где будут отображаться созданные папки и файлы с табличными данными.

4.3. **Выбор типа загрузки.** Из списка выпадающих функций выберите «Загрузка табличных данных».

4.4. **Выбор данных для загрузки в блок-схему.** Чтобы найти нужный файл, кликните на папку и перейдите в нее, выберите из списка

файл, загруженный в Систему в шаге 1 «train.csv», нажмите на три точки в строке с ним и кликните «Выбрать». Внизу отобразится название выбранного файла:

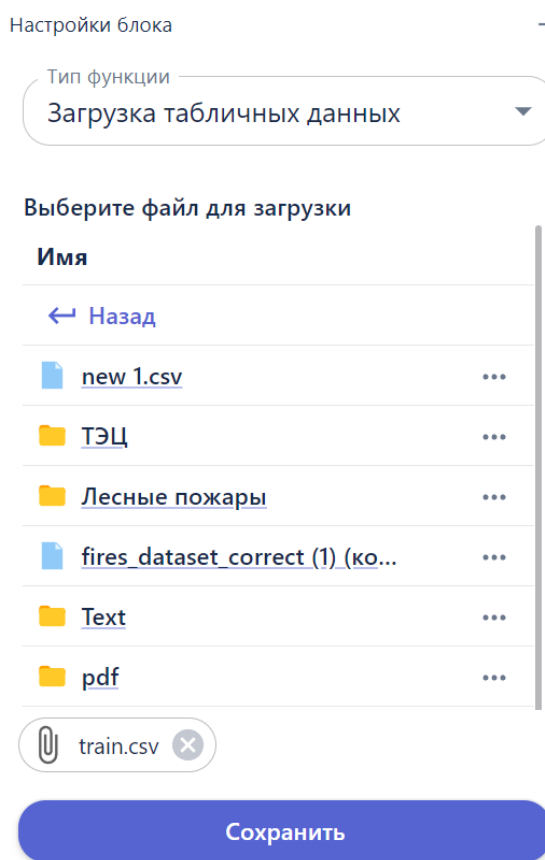


Рисунок 15.8.6 - Выбор файла

4.5. **Сохранение настроек элемента.** На панели настроек элемента нажмите на кнопку «Сохранить» (далее сохранение настроек элемента предполагается по умолчанию).

4.6. **Ввод названия элемента.** Чтобы задать название элемента нужно дважды щелкнуть левой кнопкой мыши на название элемента в рабочей области, и ввести нужное название в поле с названием, доступным для редактирования:

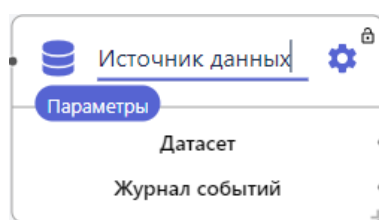



Рисунок 15.8.7 - Блок источник данных

4.7. Введите название «Загрузка табличных данных» и кликните на пустое место на рабочей области для сохранения.

5. Добавление и настройка элемента «Процесс».

5.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настройте элемент:

5.1.1. В карточке элемента в списке «Тип функции» в разделе «Анализ данных» выбрать «Поиск пропущенных значений».

5.1.2. В поле «Признаки» оставить нужные признаки (поля таблицы, в которых будет выполнен поиск пропущенных значений).

5.1.3. Нажать кнопку «Сохранить».

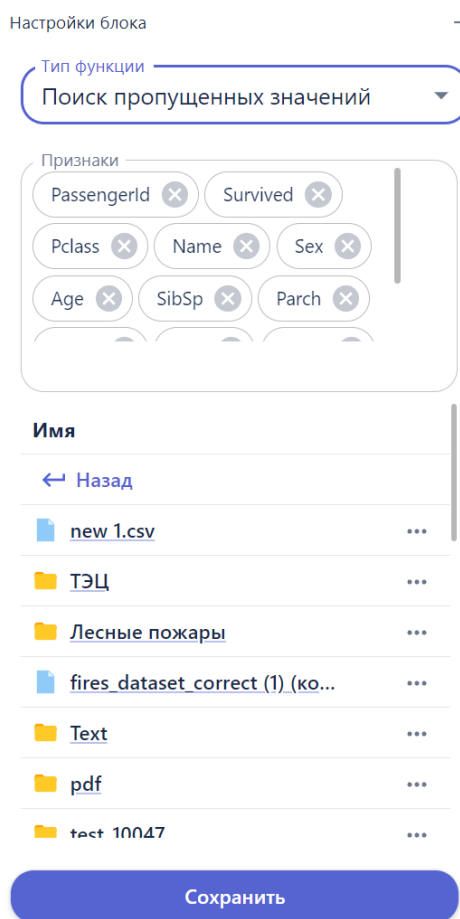



Рисунок 15.8.8 - Блок процесс

6. Добавление и настройка элемента «Процесс».

6.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

15.2.1. В карточке элемента в списке «Тип функции» в разделе «Анализ данных» выбрать «Выбор признаков и целевых признаков».

15.2.2. В поле «Признаки» оставить нужные общие признаки.

15.2.3. В поле «Целевые признаки» оставить нужные целевые признаки .

15.2.4. Нажать кнопку «Сохранить».

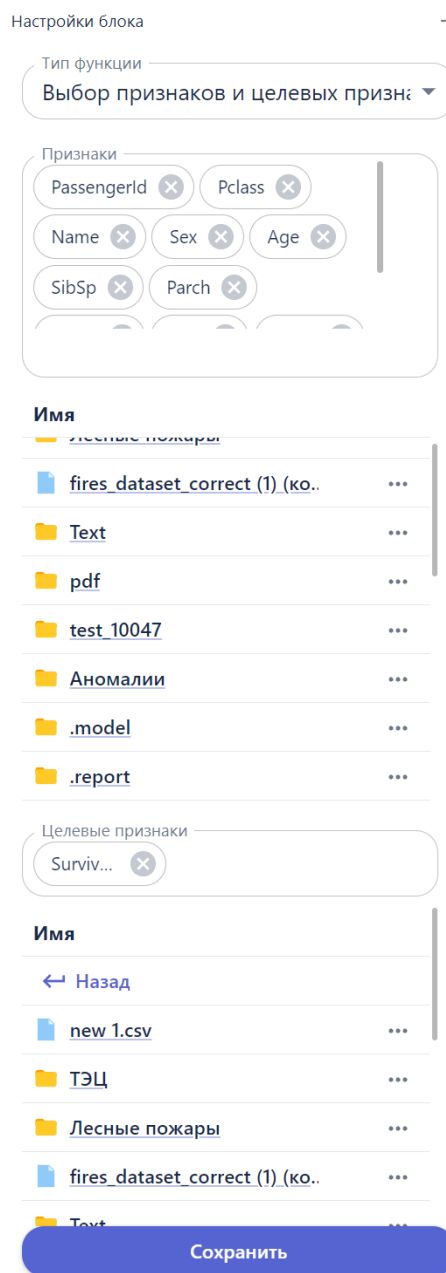



Рисунок 15.8.9 - Блок источник данных

7. Добавление и настройка элемента «Процесс».

7.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

7.1.1. В карточке элемента в списке «Тип функции» в разделе «Предобработка данных» выбрать «Заполнение пропусков».

7.1.2. В разделе «Имя» выбрать файл в формате CSV и выполнить команду «Выгрузить признаки».

7.1.3. В поле «Признак» оставить нужные признаки.

7.1.4. В списке «Метод заполнения пропусков» выбрать следующий метод частичного заполнения пропусков:

- «Удалить строки» - удаление всех строк с пропусками.

7.1.5. Нажать кнопку «Сохранить».

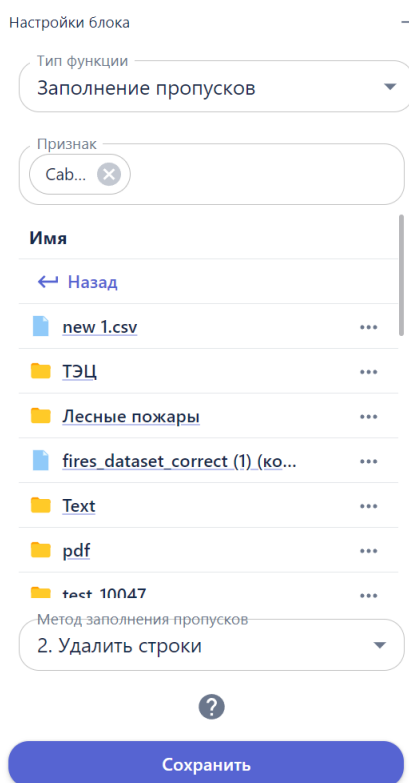



Рисунок 15.8.10 - Блок Процесс

8. Добавление и настройка элемента «Процесс».

8.1. Добавьте на рабочую область элемент «Процесс» (кнопка ). Настроить элемент:

8.2. В карточке элемента в списке «Тип функции» в разделе «Анализ данных» выбрать «Поиск пропущенных значений».

15.2.5. В разделе «Имя» выбрать файл в формате CSV и выполнить команду «Выгрузить признаки».

15.2.6. В поле «Признаки» оставить нужные признаки.

15.2.7. Нажать кнопку «Сохранить».

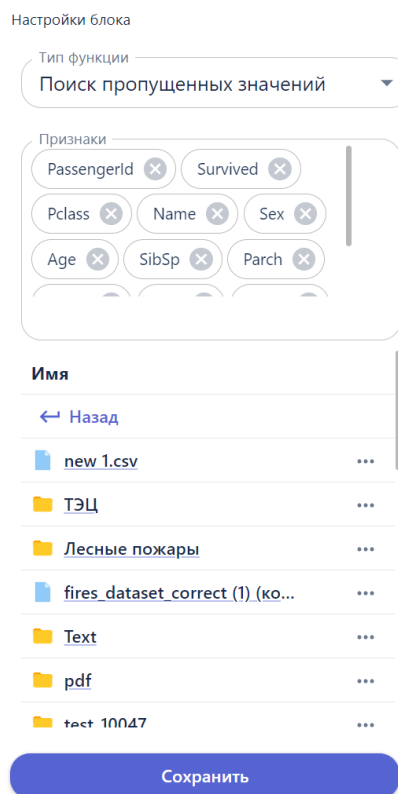


Рисунок 15.8.11 - Блок Процесс

9. Установка соединений. Соедините выходные и входные точки элементов, как показано на рисунке ниже.

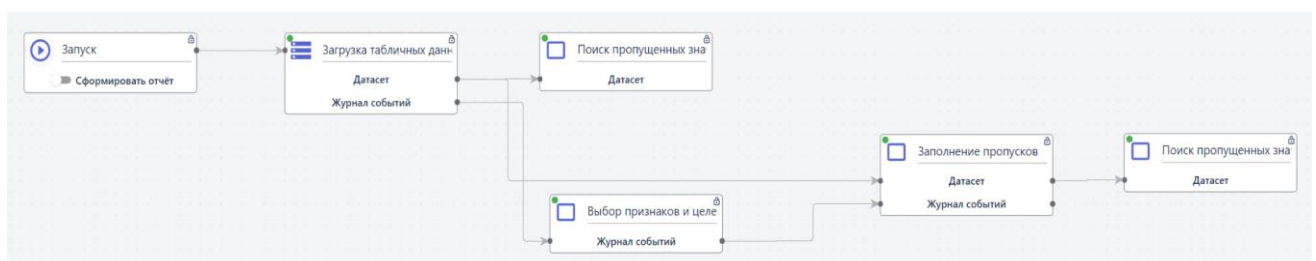



Рисунок 15.8.12 - Установка соединений

10. Запуск пайплайна. Чтобы запустить блок схему нажмите на кнопку  на первом элементе «Запуск» собранной блок-схемы. При этом отображение элемента «Запуск» изменится и появится опция Сформировать отчет:

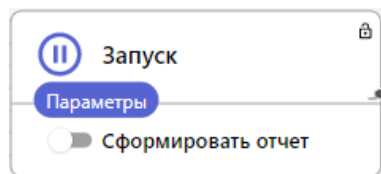
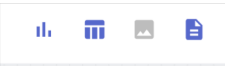


Рисунок 15.8.13 - Блок Запуск

Если активировать параметр «Сформировать отчет», в результате запуска пайплайна будет создан отчет.

11. Визуализация результатов. После того как все элементы схемы будут успешно обработаны, на панели инструментов появляются

кнопки: 

Нажмите на вторую слева кнопку и выберите команду «Количество пропусков». Отобразится таблица с результатами обработки.

2.3.14. Администрирование

2.3.14.1. Пользователи и группы

BAUM STORAGE AI v2 позволяет разделять уровни доступа к разделам меню для разных пользователей в зависимости от требования проекта. Для этого создаются Группы, которые впоследствии назначаются отдельным пользователям.

Создание новой Группы или редактирование уже созданной группы осуществляется в разделе Администрирование -> Группы:

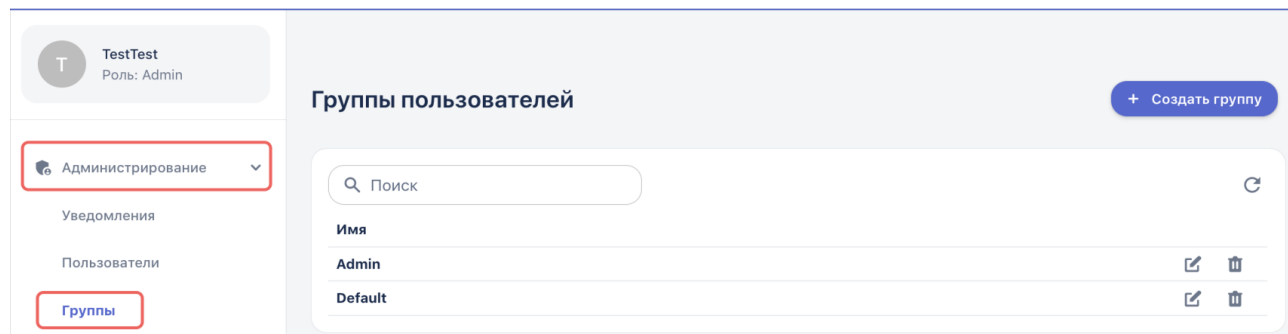


Рисунок 16.1 – Раздел Группы меню Администрирование

В списке отображаются уже созданные группы. Для создания новой группы, нажимается кнопка «Создать группу» в правом верхнем углу, в открывшемся окне отобразятся все доступные для настройки параметров:

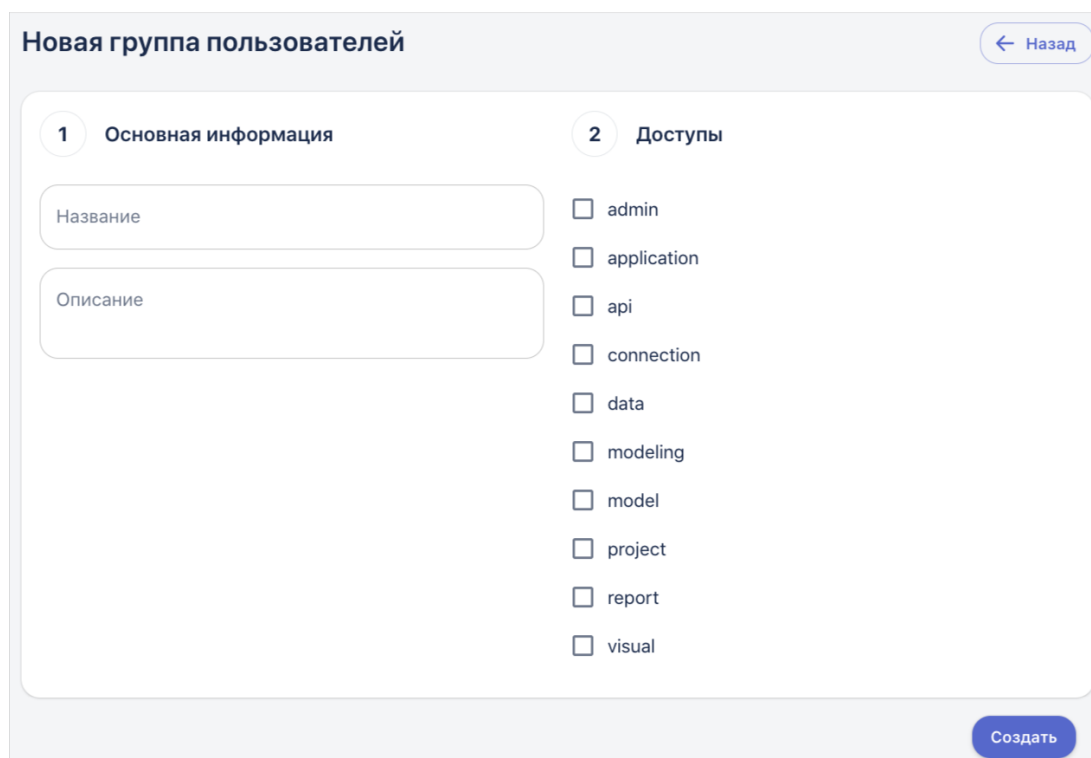


Рисунок 16.2 – Настройка новой группы пользователей

Сначала заполняется основная информация:



- *Название* - задать уникальное название для Группы (обязательное поле)
- *Описание* - внести краткую информацию о применимости данной роли (необязательное поле)

Далее из списка доступов выбираются разделы данных, к которым у пользователей, принадлежащих к этой Группе, должны быть доступны:

- admin - Администрирования
- application - Приложения
- api - API
- connection - Соединения
- data - Данные
- modeling - Моделирование
- model - Модели
- project - Проекты

- report - Отчеты
- visual - Визуализация

После того, как галочки для соответствующих разделов проставлены, нажимается кнопка «Создать». Новая группа появится в списке.

Для того чтобы отредактировать одну из Групп, используется кнопка «Редактировать»  в строке с ее названием. Для того чтобы удалить группу, нажимается кнопка «Удалить» .

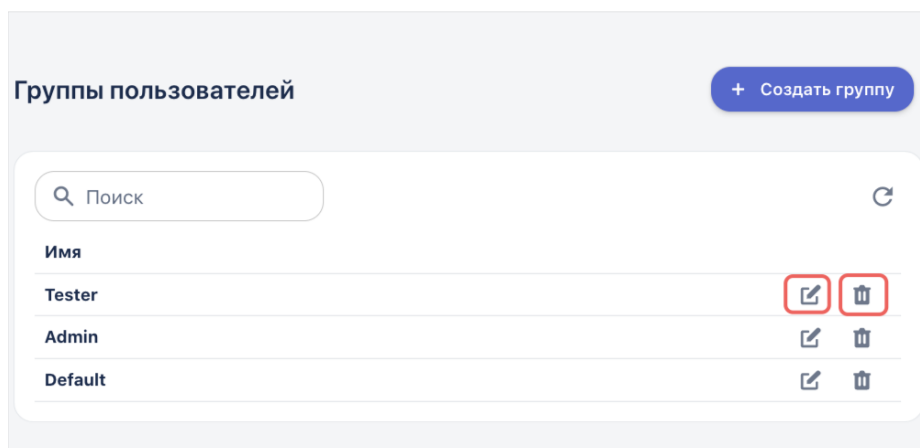


Рисунок 16.3 – Кнопки удаления и редактирования группы пользователей

Список всех пользователей представлен в разделе Администрирование -> Пользователи:

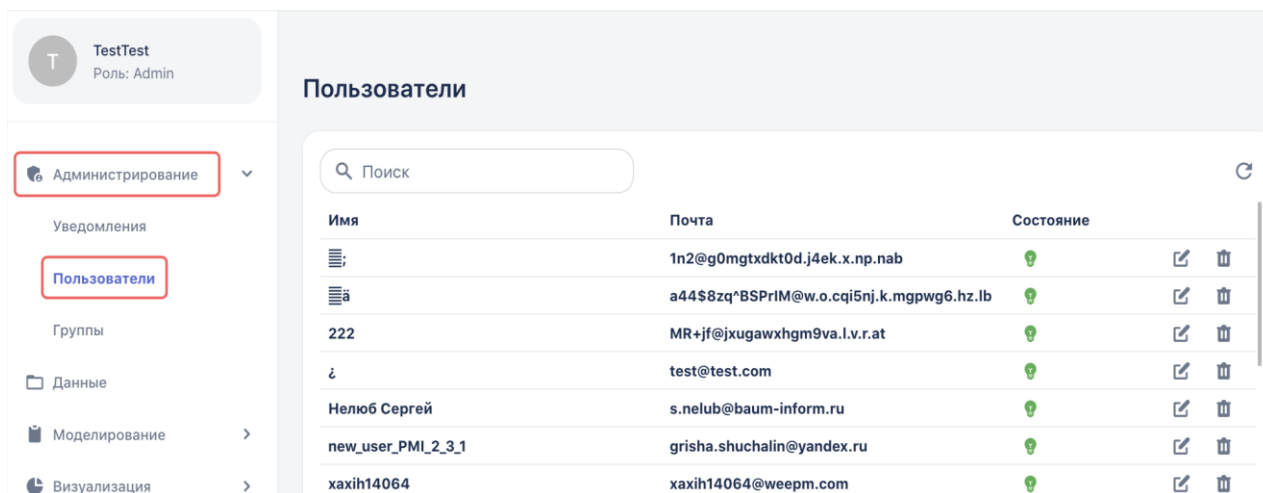



Рисунок 16.4 – Список пользователей в меню Администрирование

Для того чтобы отредактировать настройки пользователя, нажимается кнопка «Редактировать»  в строке с именем:

Пользователи			
<input type="text" value="Поиск"/> ↻			
Имя	Почта	Состояние	
	user@example.com	💡	
Большова Анна	user22@example.com	💡	

Рисунок 16.5 – Редактирование данных пользователя

В открывшемся окне есть возможность изменить имя пользователя, адрес электронной почты, а также назначить Группу:

Редактирование пользователя

Имя

Эл. почта*

Группа пользователей

Рисунок 16.6 – Параметры для редактирования

Обратите внимание, что всем новым пользователям по умолчанию присваивается Группа Default (группа по умолчанию).

Для того чтобы удалить пользователя, используется кнопка «Удалить» в строке с его именем.

Рекомендуется оставлять одного-двух пользователей Администраторов, которые смогут управлять доступами и отвечать за настройки.

2.3.14.2. Настройка отправки уведомлений

В BAUM STORAGE AI v2 реализована возможность отправки уведомлений - сообщений в телеграм или на почту. Этот функционал позволяет пользователям получать автоматические оповещения о результатах работы пайплайнов, при выполнении заданных условий. Создание и настройка каналов осуществляется следующим образом:

1. Перейти в раздел Администрирование -> Уведомления:

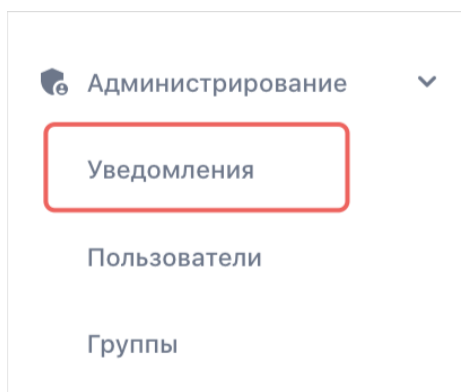


Рисунок 16.6 - Раздел Уведомления

2. В открывшемся окне отобразится список всех существующих каналов уведомлений:

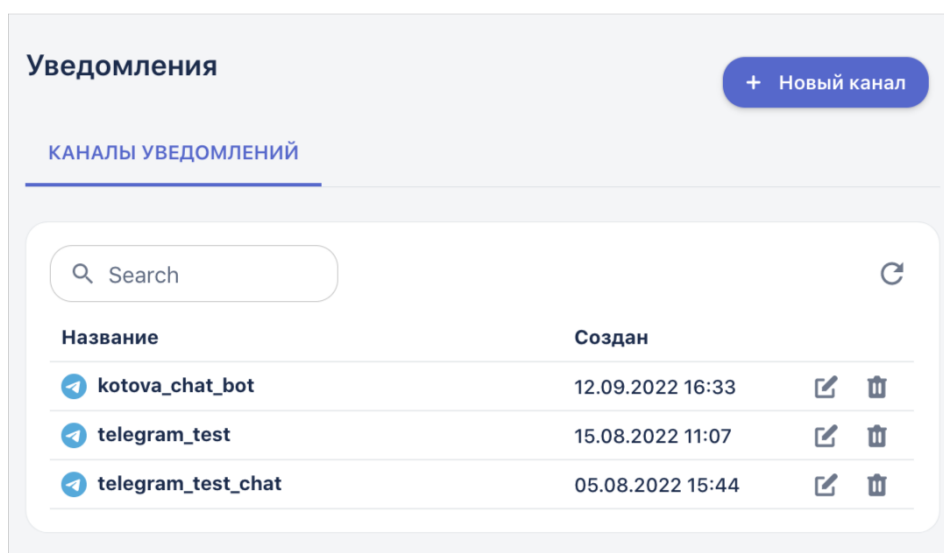


Рисунок 16.6 - Список созданных каналов

3. Для создания нового канала уведомлений нажимается кнопка «Новый канал» в правом верхнем углу

4. Открывается окно создания нового канала уведомлений:

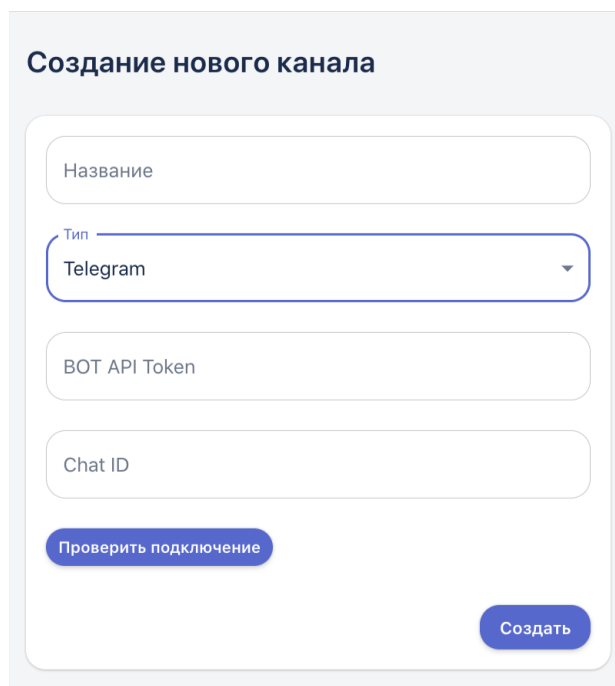


Рисунок 16.7 - Окно создания нового канала

5. Уведомления можно отправлять в телеграм бот или на почту.

5.1. Для создания уведомления с использованием **телеграм бота** поля заполняются следующим образом:

- *Название* - пользователь задаёт название бота из телеграмм
- *Тип* - Telegram
- *BOT API Token* - токен бота, полученный в телеграм при

создании бота

- *Chat ID* - уникальный численный идентификатор чат бота

Примечание: сначала пользователь должен создать чат бота в телеграм или получить его токен и id для настройки канала уведомлений. Создание бота описано в этой статье: <https://tigrm.ru/docs/bots#bofather>

5.2. Для создания уведомления **на почту** поля заполняются следующим образом:

11. *Название* - пользователь задает название канала

12. *Тип* - Почта

13. *Почты* - указываются адреса, на которые должны быть отправлены уведомления. При этом после ввода первого адреса, нужно нажать Enter, потом перейти к указанию следующего адреса и т.д.

Примечание: уведомления на почту не реализованы в текущей версии системы

6. После того, как все поля заполнены, нажимается кнопка «Проверить подключение», если все настройки были указаны верно, система выдаст сообщение об успешности подключения в верхнем правом углу окна. В противном случае, отобразится сообщение «Не удалось подключиться»

7. Завершающим этапом нажимается кнопка «Создать»

8. Новый канал отобразится в списке.

Для того чтобы начать получать уведомления в телеграм, пользователь должен создать пайплайн, который будет содержать блок «Процесс» с функцией «Отправка уведомлений», где указывается необходимый канал:

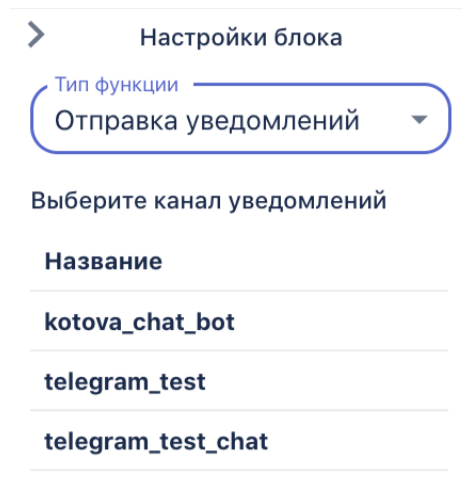


Рисунок 16.8 - Настройка блока «Отправка уведомлений»

Условием для отправки будет служить блок, идущий перед блоком уведомлений. Например, это может быть шлюз, где задаётся параметр, при котором уведомление должно быть отправлено:

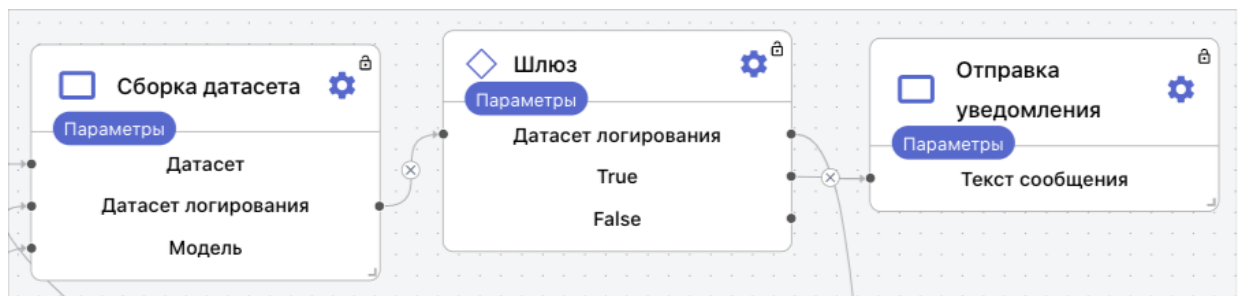


Рисунок 16.9 - Пример построения пайплайна с блоком уведомления

Например, в результате отработки такого пайплайна, пользователи будут получать следующую информацию:



Рисунок 16.10 - Пример уведомлений в телеграм канале

2.3.15. Дополнительные возможности

2.3.15.1. Обращение в службу поддержки

В верхнем правом углу окна системы находится кнопка «Сообщить об ошибке»:

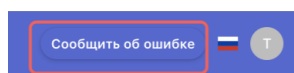


Рисунок 17.1 – Кнопка «Сообщить об ошибке»

В случае, если пользователь сталкивается с какой-то проблемой в работе программного обеспечения, поведением, не описанным в данном руководстве, есть возможность связаться с командой поддержки системы и сообщить о проблеме.

Пользователю необходимо максимально подробно описать проблему и шаги, которые привели к её появлению в текстовом окне сообщения об ошибке, а после нажать кнопку «Отправить»:

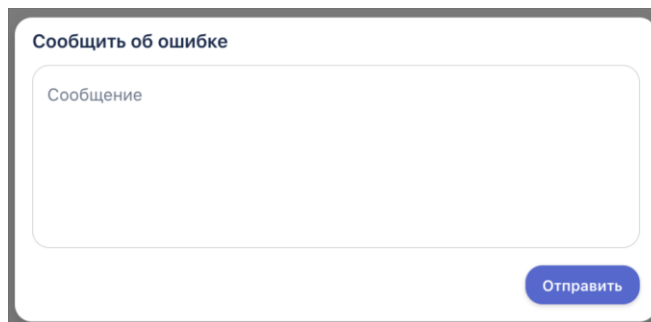


Рисунок 17.2 – Описание найденной в Системе ошибки, и отправка на обработку Разработчику

Ограничение размера сообщения составляет 256 символов.

2.3.15.2. История изменений

История изменений — это раздел системы, отображающий информацию о том в каком блоке и в какое время было совершено какое-либо изменение. Раздел содержит информацию о всех запусках пайплайнов, сформированных отчетах, названия блоков, дату и время каждого изменения.

История изменений	
ResultsReport	20.12.2022 08:55
ResultsReport	20.12.2022 08:55
ResultsReport	20.12.2022 08:55
ResultsReport	20.12.2022 08:55
ResultsReport	20.12.2022 08:55

Рисунок 17.3 – История изменений

Приложение 1. Автоматизированные функции

**Журнал преобразований* – технический объект, который позволяет пробрасывать данные между элементами блок-схемы.

**Выходным параметром для всех функций является *словарь с данными*, который может включать в себя: таблицы, графики, текстовое описание.

Таблица 18.1 – Перечень автоматизированных функций элемента «Источник данных»

Функция	Назначение	Параметры	Выходная информация		
Группа «Загрузка данных»					
Загрузка табличных данных	Преобразует загруженные данные во временной ряд. При этом автоматически определяет: формат файла (csv, txt, xls, xlsx), разделитель, размерность, где <i>разделитель</i> – это символ в датасете с временным рядом, обозначающий деление на ячейки. Для <i>ресемплирования</i> использует агрегирующую функцию	<table border="1"> <tr> <td>Выберите файл</td> <td>Название файла, содержащего временной ряд для анализа</td> </tr> </table>	Выберите файл	Название файла, содержащего временной ряд для анализа	**
Выберите файл	Название файла, содержащего временной ряд для анализа				
Загрузка изображений	<i>* данный функционал находится в разработке, в текущей версии</i>		В БД сохраняются загруженные датасеты,		

<p>для object detection</p>	<p><i>1.0.0 применение функции недоступно.</i></p> <p>Функция предназначена для загрузки в Систему изображений ‘с метками’, с целью дальнейшего решения задачи <i>компьютерного зрения</i> – распознавания отмеченных объектов на новых данных (изображениях/видео).</p>	<p>Группа обучающих изображений</p>	<p>Выбрать папку для <i>обучения</i> нейронной сети с ‘размеченными’ изображениями – на которых с помощью «bounding box» (ограничительной рамки прямоугольной формы) отмечены объекты для распознавания</p>	<p>прошедшие предварительную обработку, и вместе с каждым датасетом дополнительно сохраняется файл <i>label</i> – файл разметки.</p>
		<p>Группа валидационных изображений</p>	<p>Выбрать папку для <i>валидации</i> обученной нейронной сети, который также содержит ‘размеченные’ данные. Здесь разметка будет использоваться для оценки качества модели – путем сравнения предсказания и фактического значения</p>	
<p>Загрузка изображений для классификации</p>	<p>Загружаются изображения с объектами для дальнейшей классификации этих объектов на новых данных. При этом предусмотрено, что изображения с объектами одного</p>	<p>Группа обучающих изображений</p>	<p>Выбрать папку для <i>обучения</i> нейронной сети, который представляет собой каталог с папками, где каждая папка содержит изображения с объектами одного класса. Например – папки с</p>	<p>В БД сохраняются каталоги датасета обучения и валидации, прошедшие предварительную обработку перед загрузкой.</p>

	<p>класса распределены по отдельным папкам. При загрузке выполняется <i>ресайз</i> изображений – изменение (чаще уменьшение) размера изображений до заданного формата. Данные загружаются маленькими порциями, так называемыми <i>мини-батчами</i> (например, за один раз подается два изображения).</p>		<p>изображениями собак, кошек и диких животных (три класса).</p>	
		<p>Группа тестовых изображений</p>	<p>Выбрать папку для <i>валидации</i> обученной нейронной сети, который также представлен в виде каталога с папками отдельных классов.</p>	
		<p>Размер мини-батча</p>	<p>Указывается количество изображений, которое за один раз подается на вход нейронной сети для её обучения.</p>	
		<p>Новая высота</p>	<p>Новая высота изображения</p>	
		<p>Новая ширина</p>	<p>Новая ширина изображения</p>	
<p>Загрузка табличных данных из коннектора</p>	<p>Функция предназначена для подключения к источникам данным в виде баз данных «ClickHouse» или «Postgresql». При этом используется сущность «Коннектор», в</p>			<p>**</p>
		<p>Выберите файл</p>	<p>Из списка выбирается <i>коннектор</i> – источник данных, подключение к которому необходимо выполнить. Источником выступает сторонняя база данных – ClickHouse или Postgresql (соответственно из списка</p>	

<p>которой прописываются настройки для подключения к этим базам данных. Реализована возможность формирования датасета на основании полученных данных. Для этого в настройках функции активируется галочка в поле «Сохранить датасет», и в поле «Название файла» указывается имя для бэкапа таблицы базы данных в настоящий момент времени. В результате в разделе «Данные» сохраняется файл в формате .csv с данными из коннектора.</p> <p>Важно – Подключение выполняется к БД в её состоянии на текущий момент времени.</p>		выбирается коннектор с таким типом).	
	Сохранить датасет	<p>Чтобы сформировать бэкап таблицы внешней БД в поле «Сохранить датасет» устанавливается галочка.</p> <p>Иначе, если не установить галочку в поле «Сохранить датасет», выполняется подключение к внешней БД в её состоянии на текущий момент времени, без дополнительного формирования датасета во внутренней БД.</p>	
	Название файла	Указывается название файла бэкапа таблицы для сохранения во внутренней БД.	
	Онлайн данные	Указывается, осуществляется ли получение данных в режиме онлайн.	

Загрузка модели	<p>Функция предназначена для использования в качестве источника данных ранее обученной модели. При этом система при обработке данных пайплайна применяет ранее полученные знания для построения прогнозов.</p>	<table border="1"> <tr> <td data-bbox="846 213 1133 347"> Модель </td> <td data-bbox="1133 213 1720 347"> Выбор из списка ранее сохраненных моделей </td> </tr> </table>		Модель	Выбор из списка ранее сохраненных моделей			
Модель	Выбор из списка ранее сохраненных моделей							
Загрузка текстовых файлов для классификации	<p>Данная функция предназначена для загрузки текстов, принадлежащих к тем или иным классам, для обучения нейронной сети определять эти классы на новых данных. Функция обязательно используется при решении задач классификации текстов.</p>	<table border="1"> <tr> <td data-bbox="846 748 1133 1225"> Группа обучающих текстов </td> <td data-bbox="1133 748 1720 1225"> Выбор папки для обучения нейронной сети, которая должна содержать в себе подпапки с названиями классов объектов. Данные подпапки содержат тексты, принадлежащие определенному классу. Например, это могут быть: «Пушкин», «Лермонтов», «Толстой». </td> </tr> <tr> <td data-bbox="846 1225 1133 1399"> Группа тестовых текстов </td> <td data-bbox="1133 1225 1720 1399"> Выбор папки для валидации обученной нейронной сети. Папка </td> </tr> </table>		Группа обучающих текстов	Выбор папки для обучения нейронной сети, которая должна содержать в себе подпапки с названиями классов объектов. Данные подпапки содержат тексты, принадлежащие определенному классу. Например, это могут быть: «Пушкин», «Лермонтов», «Толстой».	Группа тестовых текстов	Выбор папки для валидации обученной нейронной сети. Папка	
Группа обучающих текстов	Выбор папки для обучения нейронной сети, которая должна содержать в себе подпапки с названиями классов объектов. Данные подпапки содержат тексты, принадлежащие определенному классу. Например, это могут быть: «Пушкин», «Лермонтов», «Толстой».							
Группа тестовых текстов	Выбор папки для валидации обученной нейронной сети. Папка							

			должна иметь такую же структуру, как и обучающая.	
		Группа текстов для классификации	Здесь можно сразу выбрать файл или папку с файлами, которые необходимо классифицировать с применением обученной модели.	
Загрузка текстовых файлов для кластеризации	Функция обязательно используется при решении задач кластеризации текстов, когда необходимо определить кластеры к которым принадлежат тексты.			
		Группа текстов для кластеризации	Выбор файла, содержащего однотипные данные, подлежащие разделению на кластеры.	
Загрузка графа	Функция предназначена для загрузки и дальнейшего преобразования файлов с форматом .graphml в переменную graph_out с типом данных			
		Выберите файл с графом	Выбирается ранее загруженный в систему файл в формате .graphml.	

	<p>networkx.MultiDiGraph, предназначенных для решения задач с применением теории графов.</p> <p>Граф — это геометрическая фигура, которая состоит из точек и линий, которые их соединяют. Точки называют вершинами графа, а линии — ребрами. Графы имеют очень широкое применение: с их помощью выбирают наиболее выгодное расположение зданий, графами представлены схемы метро, маршруты, схемы игр, блок схемы процессов и т.д.</p>		
Группа «Spark»			
Загрузка табличных данных из	При помощи данной функции осуществляется		

<p>файла (Spark) CSV</p>	<p>загрузка в систему табличных данных с помощью фреймворка для распределенных вычислений Apache Spark, конкретно, с помощью библиотеки PySpark для Python. Датафрейм в PySpark — это таблица, строки которой хранятся в RDD (Отказоустойчивый распределенный набор данных (англ. Resilient Distributed Dataset, RDD) — тип структуры данных, который можно распределить между несколькими узлами в кластере). Работа с датафреймами ведётся по принципу «ленивых вычислений» (англ. lazy evaluations). Это вычисления, которые откладываются до тех пор, пока пользователь</p>	<p>Выберите файл для загрузки</p>	<p>Выбор из списка <i>файла</i> для дальнейшего анализа.</p>	
---------------------------------	---	--	--	--

	<p>не запросит их результат.</p> <p>Данная функция работает только для файлов в формате csv, содержащих big data.</p>				
<p>Загрузка табличных данных из папки CSV (Spark)</p>	<p>Фреймворк «Apache Spark» распределяет хранимые данные по серверам и директориям. Чтобы обратиться к файлам на уровне папки, в которой они хранятся, используется данный метод.</p>	<table border="1"> <tr> <td>Выберите директорию с датасетом для загрузки</td> <td>Выбор из списка <i>папки</i> для дальнейшего анализа.</td> </tr> </table>	Выберите директорию с датасетом для загрузки	Выбор из списка <i>папки</i> для дальнейшего анализа.	
Выберите директорию с датасетом для загрузки	Выбор из списка <i>папки</i> для дальнейшего анализа.				
<p>Загрузка модели</p>	<p>Функция предназначена для использования в качестве источника данных ранее обученной модели ИИ «Spark».</p>	<table border="1"> <tr> <td>Модель</td> <td>Выбор из списка ранее сохраненных моделей Spark (в разработке отдельное <i>API</i> для моделей Spark. Модели будут объединяться в одну группу по семантическому типу).</td> </tr> </table>	Модель	Выбор из списка ранее сохраненных моделей Spark (в разработке отдельное <i>API</i> для моделей Spark. Модели будут объединяться в одну группу по семантическому типу).	
Модель	Выбор из списка ранее сохраненных моделей Spark (в разработке отдельное <i>API</i> для моделей Spark. Модели будут объединяться в одну группу по семантическому типу).				
<p>Загрузка табличных данных из</p>	<p>Функция предназначена для получения табличных данных через</p>	<p>(в разработке)</p>			

коннектора (Spark)	коннектор с типом «ClickHouse», с использованием библиотеки «Spark».		
---------------------------	--	--	--

Таблица 18.2 – Перечень автоматизированных функций элемента «Процесс»

Функция	Назначение	Параметры		Выходная информация				
1.Группа «Анализ данных»								
Выбор признаков и целевых признаков	В датасете выбираются: признаки – измеримые характеристики исследуемого объекта/ процесса, и целевые (зависимые) переменные, значения которых предстоит предсказывать модели.	<table border="1"> <tr> <td>Признаки</td> <td>Характеристики, которые исследуются и выявляется корреляция между ними и рассматриваемым целевым признаком</td> </tr> <tr> <td>Целевые признаки</td> <td>Предсказываемые переменные</td> </tr> </table>	Признаки	Характеристики, которые исследуются и выявляется корреляция между ними и рассматриваемым целевым признаком	Целевые признаки	Предсказываемые переменные		Датасет с размеченными признаками и целевыми признаками
Признаки	Характеристики, которые исследуются и выявляется корреляция между ними и рассматриваемым целевым признаком							
Целевые признаки	Предсказываемые переменные							
Матрица корреляции	1.Алгоритм сначала рассчитывает коэффициенты корреляции по всем признакам (общая матрица корреляции).	<table border="1"> <tr> <td>Топ k-значений для корреляции</td> <td>Количество максимальных значений корреляции (int)</td> </tr> </table>	Топ k-значений для корреляции	Количество максимальных значений корреляции (int)		1. Матрица корреляции с топ-k признаков, имеющих максимальные значения корреляции.		
Топ k-значений для корреляции	Количество максимальных значений корреляции (int)							

	<p>2.Затем в этой матрице отбираются топ-к максимальных (ближе к 1) значений коэффициентов корреляции.</p> <p>3.Строится новая матрица корреляции, состоящая из признаков, для которых найдены максимальные значения коэффициентов.</p>		2.Матрица корреляции по всем признакам.						
Косинусное расстояние	<p>Вычисляется <i>косинусное расстояние</i> между значениями во входном векторе и значениями выбранных столбцов в наблюдениях.</p>	<table border="1"> <tr> <td>Датасет</td> <td>Датасет с исходными данными.</td> </tr> <tr> <td>Признаки</td> <td>Признаки для анализа.</td> </tr> <tr> <td>Вектор</td> <td>Вектор такой же длины, что и количество выбранных в датасете признаков (1D-array).</td> </tr> </table>	Датасет	Датасет с исходными данными.	Признаки	Признаки для анализа.	Вектор	Вектор такой же длины, что и количество выбранных в датасете признаков (1D-array).	Таблица с наблюдениями, наиболее схожими с входным вектором, где первые топ-5 наблюдений выделены жирным шрифтом (по ним рассчитанная косинусная мера имеет значение, наиболее близкое к 0).
Датасет	Датасет с исходными данными.								
Признаки	Признаки для анализа.								
Вектор	Вектор такой же длины, что и количество выбранных в датасете признаков (1D-array).								
Поиск пропущенных значений	<p>Для каждого выбранного признака метод находит пропущенные</p>	<table border="1"> <tr> <td>Признаки</td> <td>Выбираются признаки, в которых необходимо найти пропущенные</td> </tr> </table>	Признаки	Выбираются признаки, в которых необходимо найти пропущенные	Словарь, в котором по каждому выбранному признаку отображается количество				
Признаки	Выбираются признаки, в которых необходимо найти пропущенные								

	значения в наблюдениях.		значения.	пропущенных значений в наблюдениях, и процент пропусков.
Анализ временных рядов	Совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования	Список графиков	Выбираются графики, которые будут использованы для визуализации анализа временных рядов и задаются параметры для них	Линейный график, ACF/PACF, Декомпозиция, Свечной график, Time profile, Extended, Bollinger Bands, Stochastic Oscillator
Запись в датасет логирования	Данная функция применяется при построении пайплайнов в режиме реального времени. В процессе логирования записываются все новые поступающие значения в датасет для дальнейшего использования при валидации и работе с моделью. При этом в процессе логирования записываются:			Датасет с данными

	<ul style="list-style-type: none"> - фактические значения - промежуточные трансформации при препроцессинге до подачи в модель (если в качестве источника данных выбрана модель) - прогнозные значения (если в пайплайне настроен прогноз) 		
Визуализация Real-time	Представление в виде графиков и диаграмм результатов работы пайплайна, содержащего данные, получаемые в режиме реального времени	Список графиков	Выбираются графики, которые будут использованы для визуализации данных в режиме он лайн и задаются параметры для них
			Линейный график, Свечной график, Time Profile, Extended, Bollinger Bands, Stochastic Oscillator
Загрузка данных			
Преобразование данных во временной ряд	Метод редактирует исходные данные, исключая в них	Шаг	<i>Дискретность</i> для временного
			Создается временной ряд, с заданным шагом ресемплирования

	<p>аномалии и искаженные наблюдения, которые могли быть зафиксированы в результате помех. Далее выполняется дискретизация – определяются точки (моменты времени), в которых должны быть произведены выборки значений. Дискретизация производится через равные промежутки времени.</p>	<p>ресемплирован я</p>	<p>ряда – частота фиксирования наблюдений, значения начиная с нано-, микро-, милли-, секунд и заканчивая годами. Указывается оптимальный интервал дискретности.</p>	
		<p>Частота ресемплирован ия</p>	<p>Единица измерения, в которой фиксируются наблюдения.</p>	
		<p>Агрегирующ ая функция</p>	<p>Функция, вычисляющая результат по набору значений группы, где группа – наблюдения в пределах шага ресемплирования. По умолчанию, значение вычисляется функцией медианы.</p>	
		<p>Столбец с временной меткой</p>	<p>Время фиксирования наблюдения. По умолчанию, нулевой столбец в датасете.</p>	
Препроцессинг				
<p>Стабилизация дисперсии</p>	<p>Уменьшает разброс исследуемых данных, чтобы сделать их более компактными и</p>			<p>Преобразованный датасет. При этом преобразования над</p>

	пригодными для работы.	Замена значений столбцов	Преобразование оригинального временного ряда, загруженного в систему. Позволяет заменять трансформируемые столбцы или добавлять новые	целевыми признаками проводятся отдельно.
		Стандартизация	Преобразование значений признака, адаптирующая признаки с разными диапазонами значений к моделям машинного обучения	
		Метод	Выбирается метод, с помощью которого проводится стабилизация дисперсии – приведение данных к нормальному распределению. На выбор два метода – уео-johnson и vox-cox. Метод уео-johnson работает как с отрицательными, так и с положительными значениями, а метод vox-cox только с положительными	
		Флаг признака	Показатели датасета, значения которых предстоит	

			предсказывать модели машинного обучения	
Стандартизация	Чтобы сгладить большие различия между диапазонами признаков датасета и предотвратить искаженное восприятие данных моделью машинного обучения выполняется <i>стандартизация</i> – преобразование и приведение признаков датасета к единому формату	Замена значений столбцов	Подтверждение преобразования оригинального временного ряда (заменой трансформируемых столбцов или добавлением новых)	Преобразованные значения показателей временного ряда, кроме целевого признака
		Флаг признака	Выбрать столбцы для преобразования – все, кроме столбцов с датой и целевым признаком	
Дифференцирование временного ряда	Выполняется дифференцирование целевых признаков (таргетов) временного ряда. При этом временной ряд сдвигается на указанное число шагов в разрезе каждого целевого	Шаг дифференцирования для каждого целевого признака	Есть возможность задать шаг дифференцирования для каждого таргета, в формате [сдвиг для таргета 1, сдвиг для таргета 2, ...], где сдвиг на один шаг применяется для обычного (для избавления от тренда) дифференцирования, сдвиг на несколько шагов – для	К датасету временного ряда добавляются новые столбцы с окончанием '_diff' для каждого указанного таргета. При этом замена колонок не предусмотрена – оригинальные колонки сохраняются для задачи обратного

	<p>признака. Если есть сезонность, сначала проводится сезонное дифференцирование. Желательно дифференцировать ряд как можно меньше раз, потому что с увеличением количества дифференцирований растет дисперсия ошибки прогноза</p>	<table border="1"> <tr> <td data-bbox="853 151 1093 371"></td> <td data-bbox="1093 151 1713 371"> <p>сезонного, а сдвиг, равный нулю означает, что дифференцирование для данного таргета не проводится. Например, [1, 0, 3].</p> </td> </tr> </table>		<p>сезонного, а сдвиг, равный нулю означает, что дифференцирование для данного таргета не проводится. Например, [1, 0, 3].</p>	<p>дифференцирования. Отображаются графики настоящего и сдвинутого временных рядов.</p>		
	<p>сезонного, а сдвиг, равный нулю означает, что дифференцирование для данного таргета не проводится. Например, [1, 0, 3].</p>						
<p>One-Hot Encoding</p>	<p>Метод One Hot Encoding (ONE) применяется, когда в датасете необходимо закодировать категориальные признаки (текстовые), перед подачей в модель. Для кодируемого категориального признака создаются N новых столбцов в датасете, где N – количество уникальных</p>	<table border="1"> <tr> <td data-bbox="853 791 1077 1225"> <p>Флаг удаления первого признака</p> </td> <td data-bbox="1077 791 1691 1225"> <p>Устанавливается, чтобы удалить из итоговой таблицы столбец с признаком, над которым были выполнены преобразования. Так как новые столбцы отражают принадлежность наблюдения к той или иной категории признака, удаление первого признака не повлияет на результат.</p> </td> </tr> <tr> <td data-bbox="853 1225 1077 1385"> <p>Флаг признака</p> </td> <td data-bbox="1077 1225 1691 1385"> <p>Выбираются столбцы, над которыми будут осуществляться преобразования. Значения на</p> </td> </tr> </table>	<p>Флаг удаления первого признака</p>	<p>Устанавливается, чтобы удалить из итоговой таблицы столбец с признаком, над которым были выполнены преобразования. Так как новые столбцы отражают принадлежность наблюдения к той или иной категории признака, удаление первого признака не повлияет на результат.</p>	<p>Флаг признака</p>	<p>Выбираются столбцы, над которыми будут осуществляться преобразования. Значения на</p>	<p>Таблица с новыми столбцами, в которых отражается принадлежность наблюдений к тем или иным категориям преобразованных признаков.</p>
<p>Флаг удаления первого признака</p>	<p>Устанавливается, чтобы удалить из итоговой таблицы столбец с признаком, над которым были выполнены преобразования. Так как новые столбцы отражают принадлежность наблюдения к той или иной категории признака, удаление первого признака не повлияет на результат.</p>						
<p>Флаг признака</p>	<p>Выбираются столбцы, над которыми будут осуществляться преобразования. Значения на</p>						

	<p>категорий. Значения в новых столбцах – 0 или 1, в зависимости от принадлежности к категории. Так каждый новый признак – бинарный характеристический признак категории.</p>	<table border="1" data-bbox="853 154 1688 244"> <tr> <td data-bbox="853 154 1079 244"></td> <td data-bbox="1079 154 1688 244">выбор – признаки или таргеты.</td> </tr> </table>		выбор – признаки или таргеты.	
	выбор – признаки или таргеты.				
<p>Создание признаков для временного ряда</p>	<p>Для временного ряда создаются новые признаки, в которых значения таргетов сдвигаются на указанное число шагов. Например, если для одномерного (с одним таргетом) временного ряда задать сдвиг в один шаг, создается новая колонка y, в которой значение первой строки равно значению второй строки в колонке с таргетом x (категориальный признак), т.е.</p>	<table border="1" data-bbox="853 619 1688 839"> <tr> <td data-bbox="853 619 1079 839">Максимальное количество лагов</td> <td data-bbox="1079 619 1688 839">Указывается, на какое количество шагов может быть сдвинут временной ряд.</td> </tr> </table>	Максимальное количество лагов	Указывается, на какое количество шагов может быть сдвинут временной ряд.	<p>Создается таблица, график со смещенным временным рядом.</p> <p>Процесс создания признаков-лагов сохраняется в журнале преобразований, и далее обрабатывается при препроцессинге.</p>
Максимальное количество лагов	Указывается, на какое количество шагов может быть сдвинут временной ряд.				

	<p>значения сдвигаются на один шаг вперед. Если ряд многомерный – состоящий из нескольких таргетов, то для каждого из них передается общий массив признаков, с учетом лагов всех таргет-рядов.</p> <p>Такое действие является предварительным перед тем, как подавать данные в модель ИИ, чтобы у модели были не только фактические значения таргета, но и прогнозные.</p>				
<p>Преобработка текстовых данных</p>	<p>Алгоритмы машинного обучения не работают с «сырыми данными». Большая часть процесса – это подготовка текста, преобразование ее в</p>	<table border="1" data-bbox="855 1153 1693 1401"> <tr> <td data-bbox="855 1153 1088 1401"> <p>Метод векторизации</p> </td> <td data-bbox="1088 1153 1693 1401"> <p>1. TF-IDF. С англ. TF – term frequency (частота слова), IDF – inverse document frequency (обратная частота документа). Это мера, используемая для оценки</p> </td> </tr> </table>	<p>Метод векторизации</p>	<p>1. TF-IDF. С англ. TF – term frequency (частота слова), IDF – inverse document frequency (обратная частота документа). Это мера, используемая для оценки</p>	<p><i>Числовые векторы, созданные на основе исходной текстовой информации, которые отражают важность использования каждого слова.</i></p>
<p>Метод векторизации</p>	<p>1. TF-IDF. С англ. TF – term frequency (частота слова), IDF – inverse document frequency (обратная частота документа). Это мера, используемая для оценки</p>				

	<p>вид, доступный для восприятия компьютером.</p> <p>В первую очередь выполняется <i>очистка</i> текста. Из текста удаляются бесполезные для машины данные – это большинство знаков пунктуации, особые символы, скобки, теги и т.д.</p> <p>Дальше наступает большой этап предварительной обработки – <i>препроцессинга</i>. Это приведение информации к виду, в котором она более понятна алгоритму. Методы препроцессинга:</p> <ul style="list-style-type: none"> -приведение символов к одному регистру; -<i>токенизация</i> – разбиение текста на 		<p>важности слова в контексте документа, являющегося частью коллекции документов. Вес некоторого слова пропорционален частоте употребления этого слова в документе, и обратно пропорционален частоте употребления слова во всех документах коллекции.</p> <p>Метод TF-IDF используется в задачах <i>анализа текстов</i>, и представляет собой линейный классификатор с разреженными признаками, взвешенными по частоте.</p> <p>Этот метод выбирается по умолчанию.</p> <p>2. Word2vec. Принимает большой <i>текстовый корпус</i> в качестве входных данных, и сопоставляет каждому слову вектор, выдавая координаты слов на выходе. Сначала он генерирует словарь корпуса, а затем вычисляет векторное представление слов, ‘обучаясь’ на входных тестах. Векторное представление</p>	
--	--	--	---	--

	<p>токены. Так называют отдельные компоненты – слова, предложения или фразы;</p> <p>-<i>лемматизация</i> – приведение слов к изначальным словоформам, часто с учетом контекста;</p> <p>-удаление <i>stop-слов</i> – артиклей, междометий и пр.;</p> <p>После подготовки на выходе получается набор подготовленных слов. Но алгоритмы работают с числовыми данными, поэтому из входящей информации создают <i>векторы</i> – представляют ее как набор числовых значений.</p>		<p>основывается на контекстной близости: слова, встречающиеся в тексте с одинаковыми словами (а следовательно, имеющие схожий смысл), будут иметь близкие (по <i>косинусному расстоянию</i>) векторы. Полученные векторные представления слов используются для обработки естественного языка и машинного обучения.</p>	
		<p>Использование GPU и нейронной сети</p>	<p>В данном поле устанавливается галочка, когда обучение модели ИИ происходит с использованием нейронной сети.</p>	

	<p>В ПО BAUM STORAGE AI v2 используются следующие методы векторизации – TF-IDF, и Word2vec.</p>		
<p>Кодирование целевого признака</p>	<p>Данная функция применяется, когда необходимо преобразовать <i>категориальный целевой признак в датасете</i> в числовое значение. Такое преобразование выполняется перед подачей входных данных в алгоритм.</p> <p>Правила перевода категориальный признаков в числовые прописываются в <i>кодировщике</i>. Данная функция представляет собой первый тип кодирования – Label Encoder.</p> <p>Выполняется <i>порядковое</i></p>	<p>Для данной функции не предусмотрен ручной ввод параметров пользователем. На вход функции подается <i>датасет</i> с входными данными, над которыми уже выявлены некоторые преобразования (очистка, выделение целевого признака, и т.д.), и файл <i>vars_dict</i>, в котором содержится информация о выполненных преобразованиях над данными.</p>	<p>Закодированный целевой признак в датасете.</p>

	<p><i>кодирование</i> всех уникальных значений категориального признака: первое (выбранное каким-то образом) уникальное значение кодируется нулем, второе единицей, и так далее, последнее кодируется числом, равным количеству уникальных значений минус единица.</p>		
<p>Порядковое кодирование категориальных признаков</p>	<p>Отличие данной функции в том, что она выполняет преобразование всех <i>категориальных признаков датасета</i> в числовые значения. При этом кодировщик используется тот же, что и в предыдущей функции – Label Encoder, но кодируются признаки. Выполняется <i>порядковое</i></p>	<p>Для данной функции не предусмотрен ручной ввод параметров.</p>	<p>Закодированные категориальные признаки в датасете.</p>

	кодирование каждой категориальной переменной (кроме целевой).		
--	---	--	--

Предобработка данных

Заполнение пропусков	<p>Позволяет заполнять пропущенные значения в датасете одним из следующих способов: среднее, мода, медиана, квантили 0.25, 0.5, 0.75 или по выбору, min, max. Функция применяется для датасетов, созданных из табличных данных. При этом при загрузке файла в систему можно посмотреть количество пропусков (пустых ячеек) в датасете, если для одного из признаков пропусков</p>	<table border="1"> <tr> <td>Индекс столбца</td> <td>столбца</td> </tr> <tr> <td>Способ заполнения пропусков</td> <td>mean, mode, median, max, min, quantile25, quantile50, quantile75 удаление</td> </tr> </table>	Индекс столбца	столбца	Способ заполнения пропусков	mean, mode, median, max, min, quantile25, quantile50, quantile75 удаление	Датасет с заполненными пропусками
		Индекс столбца	столбца				
Способ заполнения пропусков	mean, mode, median, max, min, quantile25, quantile50, quantile75 удаление						

	слишком много, предусмотрена возможность удаления его полностью.						
Сглаживание временного ряда	Позволяет исключить влияние шума в данных и увидеть структуру временного ряда. Для сглаживания применяется метод <i>центрированного скользящего среднего</i> : по временному ряду «скользит окно» определенного размера, в рамках окна значения группируются и по ним рассчитываются средние значения.	<table border="1"> <tr> <td>Список признаков</td> <td>Выбираются признаки для расчета скользящего среднего</td> </tr> <tr> <td>Размер окна для сглаживания</td> <td>Временное окно анализа, определяется количеством входящих в него наблюдений. Например, размер окна три, тогда берутся первые три наблюдения и по ним считаются средние значения признаков</td> </tr> </table>	Список признаков	Выбираются признаки для расчета скользящего среднего	Размер окна для сглаживания	Временное окно анализа, определяется количеством входящих в него наблюдений. Например, размер окна три, тогда берутся первые три наблюдения и по ним считаются средние значения признаков	1. Таблица, содержащая сглаженные значения признаков. 2. График исходных данных и сглаженного временного ряда.
Список признаков	Выбираются признаки для расчета скользящего среднего						
Размер окна для сглаживания	Временное окно анализа, определяется количеством входящих в него наблюдений. Например, размер окна три, тогда берутся первые три наблюдения и по ним считаются средние значения признаков						
Срез временного ряда по индексу	Позволяет создавать выборки данных за период времени, используя временные метки или временные диапазоны.	<table border="1"> <tr> <td>Дата начала</td> <td>Дата начала среза</td> </tr> <tr> <td>Дата окончания</td> <td>Дата окончания среза</td> </tr> </table>	Дата начала	Дата начала среза	Дата окончания	Дата окончания среза	Временной ряд после применения фильтра.
Дата начала	Дата начала среза						
Дата окончания	Дата окончания среза						

<p>Фильтрация текстового шума</p>	<p>Данная функция позволяет очистить текст от шумов: из текста убираются знаки препинания, заглавные буквы (они заменяются на строчные) и стоп-слова (различные служебные части речи - союзы, предлоги, частицы и т.д.)</p>	<p>-</p>	<p>Текст без шумов</p>
<p>Лемматизация текста</p>	<p>Лемматизация - это процесс приведения всех встречающихся форм слова к одной, нормальной словарной форме. В процессе лемматизации используется словарь и морфологический анализ, чтобы привести слово к его канонической форме – т.н. «лемме», в итоге получается текст, состоящий из слов приведенных к</p>	<p>-</p>	<p>Нормализованный текст</p>

	<p>единственному числу, мужскому роду, именительному падежу и инфинитиву (в зависимости от части речи).</p>						
<p>Векторизация текста</p>	<p>Векторизация текста - это процесс преобразования слов в векторы (числа), которые являются «читаемым» форматом для алгоритмов машинного обучения.</p>	<table border="1"> <tr> <td data-bbox="853 491 1088 1241"> <p>Метод векторизации</p> </td> <td data-bbox="1088 491 1688 1241"> <p>1.TD IDF - метод, используемый для оценки важности слова в контексте документа. Вес некоторого слова пропорционален частоте употребления этого слова в документе и обратно пропорционален частоте употребления слова во всех документах коллекции.</p> <p>2.Word to Vec - данный метод использует контекст, чтобы сформировать численные представления слов, в результате слова, используемые в одном и том же контексте, имеют похожие векторы.</p> </td> </tr> <tr> <td data-bbox="853 1241 1088 1430"> <p>Максимальная размерность</p> </td> <td data-bbox="1088 1241 1688 1430"> <p>Указывается примерное количество уникальных слов в тексте.</p> </td> </tr> </table>	<p>Метод векторизации</p>	<p>1.TD IDF - метод, используемый для оценки важности слова в контексте документа. Вес некоторого слова пропорционален частоте употребления этого слова в документе и обратно пропорционален частоте употребления слова во всех документах коллекции.</p> <p>2.Word to Vec - данный метод использует контекст, чтобы сформировать численные представления слов, в результате слова, используемые в одном и том же контексте, имеют похожие векторы.</p>	<p>Максимальная размерность</p>	<p>Указывается примерное количество уникальных слов в тексте.</p>	
<p>Метод векторизации</p>	<p>1.TD IDF - метод, используемый для оценки важности слова в контексте документа. Вес некоторого слова пропорционален частоте употребления этого слова в документе и обратно пропорционален частоте употребления слова во всех документах коллекции.</p> <p>2.Word to Vec - данный метод использует контекст, чтобы сформировать численные представления слов, в результате слова, используемые в одном и том же контексте, имеют похожие векторы.</p>						
<p>Максимальная размерность</p>	<p>Указывается примерное количество уникальных слов в тексте.</p>						

		текста		
		Количество признаков	Указывается количество столбцов таблицы, которая получится в результате преобразования текста в числовой вид	
		Сгенерировать тензор для GPU	Выбирается в случае если предполагается что дальше будет использоваться графический процессор. Тензор - это просто таблица особого вида	

Тесты на нормальность распределения

Коэффициент асимметрии Skewness	Данный метод проверяет выборку на нормальность распределения путем расчета асимметрии данных. Если правый хвост асимметрии длиннее левого, то коэффициент положителен, иначе – отрицателен. Если распределение симметрично (в форме ‘колокола’),			Словарь с данными.
		Признаки	Выбираются все признаки в датасете для расчета коэффициента асимметрии.	

	коэффициент равен нулю.		
Тесты на стационарность временного ряда			
Тест Дики-Фуллера	<p>Проверяется, является ли временной ряд <i>стационарным</i> – не влияют ли на него тренды и сезонность. Для такого ряда суммарные статистические данные согласованы по времени, например, <i>среднее значение</i> и <i>дисперсия наблюдений</i>. Стационарность влияет на легкость моделирования – часто требуется, чтобы временной ряд был стационарным, чтобы быть эффективным.</p>	<p>Пороговое значение alpha</p>	<p>Задается пороговое значение p из теста Дики-Фуллера, с использованием которого интерпретируются результаты гипотез:</p> <ul style="list-style-type: none"> ● <i>Нулевая гипотеза</i> – временной ряд имеет единичный корень, то есть он нестационарный; ● <i>Альтернативная гипотеза</i> – нулевая гипотеза отвергается, и предполагается, что временной ряд не имеет единичного корня, то есть он является стационарным. <p>Значение p ниже порогового значения означает, что отвергается нулевая гипотеза и временной ряд <i>стационарный</i>. Значение p выше порогового значения означает, что подтверждается нулевая гипотеза</p>

			и временной ряд <i>нестационарный</i> . Значение p задается в формате числа с плавающей точкой (float).	
2.Группа «Машинное обучение»				
Валидация модели	На тестовой выборке данных (обычно это 20% датасета) проверяется правильность работы (предсказательная способность) модели ИИ, построенной на основе машинного обучения.	Метрика	Из списка выбирается название метрики для валидации. Для задачи <i>классификации</i> : Accuracy, F1, Precision, Recall, AUC_ROC. Для задачи <i>регрессии</i> : RMSE, MAE, WMAPE.	Таблица со значением выбранной метрики, отражающей количество правильных ответов обученной модели на тестовой выборке данных (максимальное значение метрики равно 1).
		–	Журнал преобразований над данными.	
		–	Обученная модель ИИ.	
Прогноз модели	Выполняется последовательность действий по прогнозированию будущих значений целевых признаков.	-		1.Точность прогноза. 2.Словарь с данными. 3.Датасет логирования.
Разделение датасета на обучающую и	Разделение выборки данных на две категории: для	Доля	Обычно на 80% датасета	1.Отдельно обучающая и тестовая выборки.

тестовую выборки	обучения модели ИИ, и для проверки результатов обучения.	тестовой выборки в датасете	выполняется обучение модели, а на оставшихся 20% – ее валидация. Значение указывается в формате 0.2.	2.Журнал преобразований.
		Перемешивать наблюдения перед разделением	Необходимо установить галочку в поле, чтобы случайным образом поменять местами наблюдения в датасете. Не рекомендуется перемешивать временные ряды, т.к. наблюдения в них упорядочены и зафиксированы последовательно по времени.	
		Разделять с учетом меток классов	Выбирается, учитывать ли долю таргетов при разделении датасета. Используется только для задач классификации, когда объекты распределяются по категориям согласно определенным и заданным заранее признакам.	

Классификация решает задачу разделения множества наблюдений (объектов) на группы, называемые *классами*, на основе анализа их формального описания. При классификации каждое наблюдение относится к определенной группе на основе некоторого качественного свойства. Пусть X – множество описаний объектов, Y – конечное множество номеров/имен/меток классов. Существует неизвестная целевая зависимость отображения $y^*: X \rightarrow Y$, значения которой известны только на объектах обучающей выборки $X^m = (x_1, y_1), \dots, (x_m, y_m)$. Строится алгоритм, способный классифицировать произвольный объект $x \in X$.

Логистическая Регрессия	<p>Используется логистическая функция для моделирования зависимости выходной переменной y от набора входных переменных x, в случае когда первая является бинарной. Например, с помощью логистической регрессии можно оценивать вероятность наступления/или не наступления некоторого события.</p> <p>Предсказывается непрерывная переменная – коэффициент логистической регрессии, принимающий значение от 0 до 1:</p> <ul style="list-style-type: none"> • если значение коэффициента больше 	<p>Коэффициент регуляризации</p> <p>Указывается значение строго больше нуля – положительное вещественное число, с помощью которого добавляется дополнительное ограничение к условию с целью предотвратить переобучение модели.</p>	<p>1. Модель бинарной классификации.</p> <p>2. Словарь с данными.</p> <p>3. Точность модели.</p> <p>4. Журнал преобразований.</p>
		<p>Порог классификации</p> <p>Значение вещественного типа от 0 до 1, определяющее принадлежность объекта к тому или иному классу.</p>	
		<p>Флаг возврата вероятности при прогнозе</p> <p>Возвращает вероятность или метки классов для дальнейшего прогноза после обучения модели. Используется для решения задач бинарной классификации, когда выходная переменная может принимать только два значения – решается вопрос о принадлежности объекта к одному из двух классов.</p>	
		<p>Оптимизация гиперпараметров</p> <p>Флаг подбора гиперпараметров. Флаг активируется, когда</p>	

	<p>порогового значения, то вероятность наступления события равна 1;</p> <ul style="list-style-type: none"> • иначе вероятность наступления события равна 0. 	<p>ов</p>	<p>указывается несколько гиперпараметров.</p>	
		<p>Метрика для оптимизации</p>	<p>Критерий остановки итераций. Настройка, позволяющая определить точность нахождения минимума функции ошибки.</p>	
		<p>Количество фолдов для оптимизации</p>	<p>Датасет делится на фолды – на указанное количество равных частей. При обучении модели каждый фолд становится валидационным один раз, при этом на остальных фолдах выполняется обучение. Каждый раз рассчитывается значение метрики. Затем рассчитывается <i>усредненная метрика</i>, которая характеризует точность модели.</p>	
<p>Модель XGB Classifier</p>	<p>Алгоритм XGBClassifier анализирует связь между признаками и целевым признаком. На обучающей выборке модель обучается соотносить</p>	<p>Глубина дерева</p>	<p>Заданное максимальное число разбиений в ветвях, по достижению которого обучение модели ИИ останавливается.</p>	<p>1. Таблица с матрицей ошибок. 2. Таблица верных и ошибочных прогнозов модели в разрезе классов. 3. Модель бинарной классификации.</p>

	наблюдение к аномалиям, а на тестовой выборке выполняется валидация ответов обученной модели.	Количество базовых моделей	Определяет сколько независимых моделей будет работать над обучением.
		Порог классификации	Значение от 0 до 1, указывающее на верхнюю границу вероятности причисления объекта к классу.
		Флаг возврата вероятности при прогнозе	Возвращает вероятность или метки классов для дальнейшего прогноза после обучения модели. Используется для решения задач бинарной классификации, когда выходная переменная может принимать только два значения – решается вопрос о принадлежности объекта к одному из двух классов.
		Оптимизация гиперпараметров	Флаг подбора гиперпараметров. Флаг активируется, когда указывается несколько гиперпараметров.
		Метрика для оптимизации	Критерий остановки итераций. Настройка, позволяющая определить точность нахождения минимума функции ошибки.

		<p>Количество фолдов для оптимизации.</p>	<p>Датасет делится на фолды – на указанное количество равных частей. При обучении модели каждый фолд становится валидационным один раз, при этом на остальных фолдах выполняется обучение. Каждый раз рассчитывается значение метрики. Затем рассчитывается усредненная метрика, которая характеризует точность модели.</p>	
<p>Дерево решений для классификации</p>	<p>Предсказывает, к какому классу принадлежит объект из обучающего массива данных. Для этого строится дерево решений: древовидная структура, где моменты принятий решений соответствуют узлам, в узлах происходит ветвление процесса на ветки в зависимости от сделанного выбора,</p>	<p>Глубина дерева</p>	<p>Заданное максимальное число разбиений в ветвях, по достижению которого обучение модели ИИ останавливается.</p> <p>Порог классификации</p> <p>Значение порога определяет принадлежность объекта к одному из классов: к положительному – если порог выше указанного значения, к отрицательному – если порог ниже.</p>	<p>1.Датасет с <i>меткой класса</i>, определяющей принадлежность объекта к одному из классов.</p> <p>2.Модель ИИ, обученная классифицировать данные по заданным критериям.</p> <p>3.Журнал преобразований над данными.</p> <p>4.Словарь с переменными (описание модели, таблицы, графики) для</p>

<p>и конечные узлы (листья) – конечные результаты последовательного принятия решений. В узлах, начиная с корневого, выбирается признак, значение которого используется для разбиения всех данных на два класса. Процесс продолжается до тех пор, пока не выполнится <i>критерий остановки</i> – дерево превысило заранее заданный «лимит роста» (достигнута глубина дерева). При этом разбиения выполняются таким образом, чтобы уменьшить выбранный критерий, например <i>энтропию</i> – степень</p>	<p>Флаг возврата вероятности при прогнозе</p>	<p>Возвращает вероятность или метки классов для дальнейшего прогноза после обучения модели.</p>	<p>отображения в интерфейсе Программы.</p>
	<p>Оптимизация гиперпараметров</p>	<p>При выборе флага оптимизации не нужно вручную задавать глубину дерева, или можно задать несколько значений на выбор. Алгоритм подбирает глубину дерева из расчета получения максимального значения метрики.</p>	
	<p>Метрика для оптимизации</p>	<p>Метод, который рассчитывает точность обученной модели. Выбирается один из предлагаемых методов.</p>	
	<p>Количество фолдов для оптимизации</p>	<p>Датасет делится на фолды – на указанное количество равных частей. При обучении модели каждый фолд становится <i>валидационным</i> один раз, при этом на остальных фолдах выполняется обучение. Каждый раз рассчитывается значение метрики. Затем рассчитывается</p>	

	неопределенности в разбиении на классы.		усредненная метрика, которая характеризует точность модели.											
<p>Случайный лес для классификации</p>	<p>Строится множество решающих деревьев, и в задаче классификации принимается решение голосованием по большинству. Все деревья строятся независимо по схеме:</p> <p>1.Выбирается подвыборка обучающей выборки и по ней строится дерево.</p> <p>2.Для построения каждого расщепления в дереве просматривается максимальное количество случайных признаков.</p> <p>3.Выбирается наилучший признак и расщепление по нему (по заранее заданному</p>	<table border="1"> <tr> <td data-bbox="846 352 1106 483">Глубина дерева</td> <td data-bbox="1106 352 1684 483">Максимальная глубина для деревьев решений.</td> </tr> <tr> <td data-bbox="846 483 1106 614">Количество деревьев</td> <td data-bbox="1106 483 1684 614">Число деревьев в «лесу».</td> </tr> <tr> <td data-bbox="846 614 1106 790">Порог классификации</td> <td data-bbox="1106 614 1684 790">—</td> </tr> <tr> <td data-bbox="846 790 1106 1050">Флаг возврата вероятности при прогнозе</td> <td data-bbox="1106 790 1684 1050">—</td> </tr> <tr> <td data-bbox="846 1050 1106 1401">Оптимизация гиперпараметров</td> <td data-bbox="1106 1050 1684 1401">Необходимо активировать галочку в поле, чтобы подобрать гиперпараметры – глубину и количество деревьев. Гиперпараметры подбираются таким образом, чтобы получить максимальное значение метрики.</td> </tr> </table>		Глубина дерева	Максимальная глубина для деревьев решений.	Количество деревьев	Число деревьев в «лесу».	Порог классификации	—	Флаг возврата вероятности при прогнозе	—	Оптимизация гиперпараметров	Необходимо активировать галочку в поле, чтобы подобрать гиперпараметры – глубину и количество деревьев. Гиперпараметры подбираются таким образом, чтобы получить максимальное значение метрики.	**
Глубина дерева	Максимальная глубина для деревьев решений.													
Количество деревьев	Число деревьев в «лесу».													
Порог классификации	—													
Флаг возврата вероятности при прогнозе	—													
Оптимизация гиперпараметров	Необходимо активировать галочку в поле, чтобы подобрать гиперпараметры – глубину и количество деревьев. Гиперпараметры подбираются таким образом, чтобы получить максимальное значение метрики.													

	<p>критерию). Дерево строится, до достижения параметра, ограничивающего его высоту.</p> <p>Таким образом деревья обучаются не только на разных наборах данных, но и используют разные признаки для принятия решений – это создает некоррелированные деревья, которые и защищают друг друга от своих ошибок. Прогноз получается точнее, чем у любого отдельного дерева.</p>	<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>	
<p>Categorical Naive Bayes</p>	<p>Группа байесовских классификаторов позволяет определить к какому классу принадлежит объект на основе теоремы Байеса с допущением о независимости</p>	<p>Параметр сглаживания Лапласа</p>	<p>Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.</p>	<p>1.Датасет с <i>меткой класса</i>, определяющей принадлежность объекта к одному из классов.</p> <p>2.Модель ИИ, обученная классифицировать данные по заданным</p>

	<p>признаков. Категориальный наивный байесовский классификатор применяется для признаков с категориальным распределением.</p>	<p>Априорные вероятности классов</p>	<p>Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.</p>	<p>критериям. 3. Журнал преобразований над данными. 4. Словарь с переменными (описание модели, таблицы, графики) для отображения в интерфейсе Программы.</p>
		<p>Оптимизация гиперпараметров</p>	<p>Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели</p>	
		<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>	
		<p>Количество фолдов для оптимизации</p>	<p>Указывается, на сколько равных частей разбивается входной датасет при обучении модели.</p>	
<p>Multinomial Naive Bayes</p>	<p>Мультиномиальный классификатор применяется для признаков с</p>			<p>1. Датасет с меткой класса, определяющей принадлежность объекта к одному из классов.</p>

	<p>полиномиальным распределением. Пример: классификация текстов, где каждый текст представлен вектором слов (например, мешок слов или tf-idf).</p>	<p>Параметр сглаживания Лапласа</p>	<p>Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.</p>	<p>2. Модель ИИ, обученная классифицировать данные по заданным критериям. 3. Журнал преобразований над данными. 4. Словарь с переменными (описание модели, таблицы, графики) для отображения в интерфейсе Программы.</p>
		<p>Априорные вероятности классов</p>	<p>Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.</p>	
		<p>Оптимизация гиперпараметров</p>	<p>Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели</p>	
		<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>	
		<p>Количество фолдов для оптимизации</p>	<p>Указывается, на сколько равных частей разбивается входной датасет при обучении модели.</p>	

<p>Complement Naive Bayes</p>	<p>Представляет собой вариант адаптации Multinomial Naive Bayes для датасетов с несбалансированным и классами.</p> <p>Вместо вычисления вероятностей принадлежности объекта к конкретному классу для каждого класса вычисляются вероятности того, что объект им не принадлежит.</p> <p>Выбирается наименьшая вероятность "непринадлежности" к классу, так как это означает, что объект с наибольшей вероятностью принадлежит к данному классу.</p>	<table border="1"> <tr> <td data-bbox="846 276 1106 539"> <p>Параметр сглаживания Лапласа</p> </td> <td data-bbox="1106 276 1709 539"> <p>Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.</p> </td> </tr> <tr> <td data-bbox="846 539 1106 842"> <p>Априорные вероятности классов</p> </td> <td data-bbox="1106 539 1709 842"> <p>Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.</p> </td> </tr> <tr> <td data-bbox="846 842 1106 1058"> <p>Оптимизация гиперпараметров</p> </td> <td data-bbox="1106 842 1709 1058"> <p>Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели</p> </td> </tr> <tr> <td data-bbox="846 1058 1106 1281"> <p>Метрика для оптимизации</p> </td> <td data-bbox="1106 1058 1709 1281"> <p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p> </td> </tr> </table>	<p>Параметр сглаживания Лапласа</p>	<p>Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.</p>	<p>Априорные вероятности классов</p>	<p>Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.</p>	<p>Оптимизация гиперпараметров</p>	<p>Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели</p>	<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>	<p>1. Датасет с <i>меткой класса</i>, определяющей принадлежность объекта к одному из классов.</p> <p>2. Модель ИИ, обученная классифицировать данные по заданным критериям.</p> <p>3. Журнал преобразований над данными.</p> <p>4. Словарь с переменными (описание модели, таблицы, графики) для отображения в интерфейсе Программы.</p>
<p>Параметр сглаживания Лапласа</p>	<p>Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.</p>										
<p>Априорные вероятности классов</p>	<p>Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.</p>										
<p>Оптимизация гиперпараметров</p>	<p>Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели</p>										
<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>										

		Количество фолдов для оптимизации	Указывается, на сколько равных частей разбивается входной датасет при обучении модели.	
Gaussian Naive Bayes	Для значений признаков для каждого класса строится распределение Гаусса. В качестве значений правдоподобия для признаков берутся значения функции Гаусса из конкретного распределения (соответствующее признаку и классу).	Параметр сглаживания Лапласа	Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.	<p>1. Датасет с <i>меткой класса</i>, определяющей принадлежность объекта к одному из классов.</p> <p>2. Модель ИИ, обученная классифицировать данные по заданным критериям.</p> <p>3. Журнал преобразований над данными.</p> <p>4. Словарь с переменными (описание модели, таблицы, графики) для отображения в интерфейсе Программы.</p>
		Априорные вероятности классов	Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.	
		Оптимизация гиперпараметров	Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели	

		Метрика для оптимизации	Выбирается одна из предлагаемых метрик для оценки работы модели.	
		Количество фолдов для оптимизации	Указывается, на сколько равных частей разбивается входной датасет при обучении модели.	
Bernoulli Naive Bayes	Применяется для признаков с биномиальным распределением. Пример: классификация текстов, где каждый текст представлен вектором наличия слов из словаря (1 - есть слово, 0 - нет).	Параметр сглаживания Лапласа	Значение для аддитивного сглаживания Лапласа во избежание проблемы нулевой вероятности. Если равен нулю, то сглаживания нет.	1. Датасет с меткой класса, определяющей принадлежность объекта к одному из классов. 2. Модель ИИ, обученная классифицировать данные по заданным критериям. 3. Журнал преобразований над данными. 4. Словарь с переменными (описание модели, таблицы, графики) для
		Априорные вероятности классов	Определяет, будут ли взяты в расчет априорные вероятности классов. Если не активирован, то применяются значения вероятностей для равномерного распределения.	

		Оптимизация гиперпараметров	Определяет, будет ли проводиться оптимизация гиперпараметров при обучении модели	отображения интерфейсе Программы. В
		Метрика для оптимизации	Выбирается одна из предлагаемых метрик для оценки работы модели.	
		Количество фолдов для оптимизации	Указывается, на сколько равных частей разбивается входной датасет при обучении модели.	

Кластеризация – это задача группировки множества объектов на подмножества (кластеры) так, чтобы объекты одного кластера были более похожи друг на друга, чем на объекты других кластеров по какому-либо критерию. Относится к классу задач *обучения без учителя*.

Алгоритм кластеризации DBSCAN	Алгоритм DBScan формирует группы коренных соседей/кластеры, объединяя точки, расположенные рядом. А точки, которые не попадают ни в одну из групп, отмечаются меткой -1 и приравниваются к аномалиям.			1. Модель кластеризации. 2. Выходной датасет, дополненный меткой кластера и/или флагом аномалии. 3. Словарь, содержащий информацию (графики, таблицы, текст) для отображения в пользовательском интерфейсе. 4. Журнал преобразований. <i>**Для алгоритмов кластеризации, регрессии, классификации наборы выходных параметров идентичны, отличие заключается в их содержимом.</i>
		Датасет	Датасет с исходными данными.	
		Журнал преобразований	–	
		Радиус	Радиус в единицах расстояния, в рамках которого выполняется поиск потенциальных соседей (float/list/tuple).	
		Число соседей	Минимальное число ближайших соседей в указанном радиусе для формирования группы коренных соседей (int/list/tuple).	
Метрика расстояния	Метрика расстояния (str/list): расстояние Евклида, косинусное расстояние. По умолчанию «Евклидово расстояние» – используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к кластеру.			

		<p>Оптимизация гиперпараметров</p> <p>Флаг подбора гиперпараметров. При значении «true» выполняется ручной ввод следующих гиперпараметров: радиус, число соседей, метрика расстояния. При значении «false» эти гиперпараметры подбираются автоматически.</p>	
<p>* Параметры: <i>датасет</i> и <i>журнал преобразований</i> являются входными и выходными параметрами для всех алгоритмов.</p>			

<p>Изоляционный лес</p>	<p>Алгоритм поиска аномалий (выбросов) методом «Изоляционный лес»: Изолирует наблюдения, случайным образом выбирая объект, а затем случайным образом выбирая разделения между максимальным и минимальным значениями объекта. Разбиение представлено древовидной структурой, количество разбиений, необходимое для изоляции выборки, равно длине пути от корневого до конечного узла. Эта длина пути является мерой нормальности и функции принятия решений. Когда лес</p>	<table border="1"> <tr> <td data-bbox="846 212 1189 347">Датасет</td> <td data-bbox="1189 212 1704 347">Датасет с исходными данными.</td> </tr> <tr> <td data-bbox="846 347 1189 483">Журнал преобразований</td> <td data-bbox="1189 347 1704 483">—</td> </tr> <tr> <td data-bbox="846 483 1189 651">Количество деревьев</td> <td data-bbox="1189 483 1704 651">По умолчанию устанавливается значение, равное 2</td> </tr> </table>	Датасет	Датасет с исходными данными.	Журнал преобразований	—	Количество деревьев	По умолчанию устанавливается значение, равное 2	<p>**</p>
Датасет	Датасет с исходными данными.								
Журнал преобразований	—								
Количество деревьев	По умолчанию устанавливается значение, равное 2								

	случайных деревьев создает более короткие пути для отдельных объектов, они, скорее всего, являются аномалиями.										
Кластеризация K-Means	<p>Алгоритм кластеризации K-средних:</p> <p>1. Из исходного множества случайным образом выбирается K наблюдений, равное заданному количеству кластеров.</p> <p>2. Для каждого наблюдения определяется ближайший к нему центр кластера (измеряется Евклидово расстояние до центра). Образуются начальные кластеры.</p> <p>3. Вычисляются <i>центры тяжести кластеров</i> – вектора,</p>	<table border="1"> <tr> <td data-bbox="846 534 1131 622">Датасет</td> <td data-bbox="1131 534 1706 622">Датасет с исходными данными.</td> </tr> <tr> <td data-bbox="846 622 1131 798">Журнал преобразований</td> <td data-bbox="1131 622 1706 798">–</td> </tr> <tr> <td data-bbox="846 798 1131 933">Число кластеров</td> <td data-bbox="1131 798 1706 933">По умолчанию устанавливается значение, равное 2.</td> </tr> <tr> <td data-bbox="846 933 1131 1109">Оптимизация гиперпараметров</td> <td data-bbox="1131 933 1706 1109">Флаг подбора гиперпараметров.</td> </tr> </table>	Датасет	Датасет с исходными данными.	Журнал преобразований	–	Число кластеров	По умолчанию устанавливается значение, равное 2.	Оптимизация гиперпараметров	Флаг подбора гиперпараметров.	**
Датасет	Датасет с исходными данными.										
Журнал преобразований	–										
Число кластеров	По умолчанию устанавливается значение, равное 2.										
Оптимизация гиперпараметров	Флаг подбора гиперпараметров.										

	<p>элементы которых представляют собой среднее арифметическое значение признаков кластера.</p> <p>4. Центры кластеров смещаются и объединяют вокруг себя наблюдения, пока центры и границы кластеров не перестанут изменяться.</p>										
<p>Агломеративная иерархическая кластеризация</p>	<p>Последовательно объединяет объекты во все более крупные подмножества, в результате образуется древовидная структура. Отдельные версии иерархии отличаются правилами вычисления расстояния между кластерами. Например, алгоритм средней связи на</p>	<table border="1"> <tr> <td data-bbox="846 834 1162 927">Датасет</td> <td data-bbox="1162 834 1709 927">Датасет с исходными данными.</td> </tr> <tr> <td data-bbox="846 927 1162 1102">Журнал преобразований</td> <td data-bbox="1162 927 1709 1102">–</td> </tr> <tr> <td data-bbox="846 1102 1162 1233">Число кластеров</td> <td data-bbox="1162 1102 1709 1233">Задается оптимальное количество кластеров.</td> </tr> <tr> <td data-bbox="846 1233 1162 1433">Метрика расчета расстояния</td> <td data-bbox="1162 1233 1709 1433">Используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к</td> </tr> </table>	Датасет	Датасет с исходными данными.	Журнал преобразований	–	Число кластеров	Задается оптимальное количество кластеров.	Метрика расчета расстояния	Используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к	<p>**</p>
Датасет	Датасет с исходными данными.										
Журнал преобразований	–										
Число кластеров	Задается оптимальное количество кластеров.										
Метрика расчета расстояния	Используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к										

	<p>каждом шаге объединяет два ближайших кластера, рассчитывая среднюю арифметическую дистанцию между всеми парами объектов.</p>		<p>кластеру. Значения: евклидово расстояние, косинусная мера, расстояние городских кварталов, расстояние Чебышева.</p>	
		<p>Критерий связи для расчета расстояния</p>	<p>Правила вычисления расстояния между кластерами при каждой итерации. Значения:</p> <ul style="list-style-type: none"> -алгоритм средней связи, -алгоритм одиночной связи или ближайшего соседа, -алгоритм полной связи или дальнего соседа, -метод минимума дисперсии Уорда. <p><i>Дисперсия</i> объединяет кластеры с минимальной общей внутрикластерной дисперсией после слияния, в качестве метрики расстояния используется евклидово расстояние. Минимальный использует самые близкие точки в обоих кластерах для</p>	

			расчета расстояния.			
		Оптимизация гиперпараметров	Флаг подбора гиперпараметров.			
Метод локтя K-Means	<p>Метод локтя позволяет вычислить правильное значение k (количество кластеров) и повысить производительность модели. Вычисляется сумма квадратов расстояний между точками, и среднее арифметическое значение (Mean) – сумма элементов датасета, разделенная на их количество. Когда значение k равно 1, сумма квадрата внутри кластера будет большой. По мере увеличения значения k сумма квадратов внутри кластера будет</p>	<table border="1"> <tr> <td>Число кластеров</td> <td>Задается оптимальное количество кластеров.</td> </tr> </table>	Число кластеров	Задается оптимальное количество кластеров.		**
Число кластеров	Задается оптимальное количество кластеров.					

	<p>уменьшаться. Наконец будет построен график между значениями k и суммой квадрата внутри кластера. В момент, когда значение k резко уменьшится будет считаться оптимальным числом кластеров.</p>		
--	---	--	--

Регрессия – математическое выражение, отражающее связь между зависимой переменной y и независимыми переменными x . Алгоритмы регрессии используются для *контролируемого обучения* моделей ИИ – так называемого обучения «с учителем», когда данные размечаются для помощи в прогнозировании. Сопоставляя входные данные и полученные результаты на точность, модель постепенно обучается прогнозировать *числовые значения* целевых переменных.

<p>Линейная регрессия</p>	<p>Прогнозирует целевую переменную Y на основе одной или нескольких независимых переменных X. Для этого между X и Y строится линейная связь.</p>	<table border="1"> <tr> <td data-bbox="835 869 1176 1074">Датасет</td> <td data-bbox="1176 869 1731 1074">Датасет с исходными данными.</td> </tr> <tr> <td data-bbox="835 1074 1176 1284">Журнал преобразований</td> <td data-bbox="1176 1074 1731 1284">–</td> </tr> </table>	Датасет	Датасет с исходными данными.	Журнал преобразований	–	<p>**</p>
Датасет	Датасет с исходными данными.						
Журнал преобразований	–						
<p>Дерево решений для регрессии</p>	<p>Предсказывает значение целевой переменной, изучая</p>	<table border="1"> <tr> <td data-bbox="835 1284 1176 1433">Датасет</td> <td data-bbox="1176 1284 1731 1433">Датасет с исходными данными.</td> </tr> </table>	Датасет	Датасет с исходными данными.	<p>**</p>		
Датасет	Датасет с исходными данными.						

<p>простые правила принятия решений, выведенные из характеристик данных.</p> <p>Представляет собой древовидный граф с узлами, где атрибут – вопрос, ребро – ответ на вопрос, а листья – фактический результат.</p> <p>Наблюдения классифицируются сверху вниз от корня до листьев.</p>	<p>Журнал преобразований</p>	–
	<p>Глубина дерева</p>	Заданное максимальное число разбиений в ветвях, по достижению которого обучение останавливается.
	<p>Оптимизация гиперпараметров</p>	При выборе флага оптимизации не нужно вручную задавать глубину дерева, или можно задать несколько значений на выбор. Алгоритм подбирает глубину дерева из расчета получить максимальное значение метрики.
	<p>Метрика для оптимизации</p>	Метод, который рассчитывает точность обученной модели. Выбирается один из предлагаемых методов.
	<p>Количество фолдов для оптимизации</p>	Датасет делится на фолды – на указанное количество равных частей. При обучении модели регрессии каждый фолд становится валидационным

			<p>один раз, при этом на остальных фолдах выполняется обучение. Каждый раз рассчитывается значение метрики. Затем рассчитывается <i>усредненная метрика</i>, которая характеризует точность модели.</p>	
<p>Случайный лес для регрессии</p>	<p>В отличие от предыдущего алгоритма здесь строится ансамбль решающих деревьев. При этом большое количество некоррелированных моделей (деревьев) превосходит любую из отдельных моделей.</p>			<p>**</p>
		<p>Датасет</p>	<p>Датасет с исходными данными.</p>	
		<p>Журнал преобразований</p>	<p>–</p>	
		<p>Глубина дерева</p>	<p>–</p>	
		<p>Количество деревьев</p>	<p>–</p>	
<p>Оптимизация гиперпараметров</p>	<p>Алгоритм подбирает гиперпараметры: глубину и количество деревьев из расчета получить максимальное значение метрики.</p>			

		<table border="1"> <tr> <td>Метрика для оптимизации</td> <td>Выбирается одна из предлагаемых метрик для оценки работы модели.</td> </tr> <tr> <td>Количество фолдов для оптимизации</td> <td>Указывается, на сколько равных частей разбивается входной датасет при обучении модели.</td> </tr> </table>	Метрика для оптимизации	Выбирается одна из предлагаемых метрик для оценки работы модели.	Количество фолдов для оптимизации	Указывается, на сколько равных частей разбивается входной датасет при обучении модели.			
Метрика для оптимизации	Выбирается одна из предлагаемых метрик для оценки работы модели.								
Количество фолдов для оптимизации	Указывается, на сколько равных частей разбивается входной датасет при обучении модели.								
Полиномиальная регрессия	<p>Метод регрессионного анализа, в которой взаимосвязь между независимыми переменными x и зависимой переменной y моделируется как полином n-ой степени от x. Полиномиальная регрессия соответствует нелинейной зависимости между значением x и соответствующим условным средним y, обозначаемым $E(y x)$.</p>	<table border="1"> <tr> <td>Степень полинома</td> <td>Степень уравнения полиномиальной регрессии, которая определяет линию наилучшего соответствия. При неправильном выборе степени, модель может быть перенасыщена. Значение по умолчанию – 2.</td> </tr> <tr> <td>Только произведение</td> <td>Если установить галочку в поле, то не выполняется возведение в степень, а только перемножение.</td> </tr> <tr> <td>Оптимизация гиперпараметров</td> <td>Нужно активировать галочку в поле, когда выбирается наиболее подходящая степень полинома из нескольких предложенных. А подбирается гиперпараметр так,</td> </tr> </table>	Степень полинома	Степень уравнения полиномиальной регрессии, которая определяет линию наилучшего соответствия. При неправильном выборе степени, модель может быть перенасыщена. Значение по умолчанию – 2.	Только произведение	Если установить галочку в поле, то не выполняется возведение в степень, а только перемножение.	Оптимизация гиперпараметров	Нужно активировать галочку в поле, когда выбирается наиболее подходящая степень полинома из нескольких предложенных. А подбирается гиперпараметр так,	<p>1. Модель полиномиальной регрессии. 2. Словарь с переменными для отображения в интерфейсе. 3. Словарь с преобразованиями данных. 4. Выходной датасет.</p>
Степень полинома	Степень уравнения полиномиальной регрессии, которая определяет линию наилучшего соответствия. При неправильном выборе степени, модель может быть перенасыщена. Значение по умолчанию – 2.								
Только произведение	Если установить галочку в поле, то не выполняется возведение в степень, а только перемножение.								
Оптимизация гиперпараметров	Нужно активировать галочку в поле, когда выбирается наиболее подходящая степень полинома из нескольких предложенных. А подбирается гиперпараметр так,								

	<p>В отличие от линейной регрессии моделирует нелинейно разделенные данные – более гибкая и может моделировать сложные взаимосвязи.</p>		<p>чтобы получить наилучшее значение метрики.</p>	
		<p>Метрика для оптимизации</p>	<p>Значения на выбор: RMSE, MAE, WMAPE, где RMSE – среднеквадратическая ошибка, MAE – средняя абсолютная ошибка, а ошибка – разница между значениями, предсказанными моделью, и фактическими значениями переменной. Эти метрики используются для оценки работы модели регрессии – проверяют точность прогноза и измеряют величину отклонения от фактических значений.</p>	
		<p>Количество фолдов для оптимизации</p>	<p>Указывается, на сколько равных частей разбивается входной датасет при обучении модели.</p>	
<p>Метод опорных векторов для регрессии</p>	<p>В основе регрессии опорных векторов (с англ. SVR – Support Vector Regression) лежит поиск гиперплоскости, при</p>	<p>Тип ядра</p>	<p>Функция ядра (kernel) может принимать значения: {'linear', 'poly', 'rbf', 'sigmoid'}.</p>	<p>**</p>

<p>которой риск в многомерном пространстве будет минимальным. SVR оценивает коэффициенты путем минимизации квадратичных потерь: считается сумма квадратов ошибок (между прогнозом и фактом), и к ней прибавляется штраф в виде произведения <i>коэффициента регуляризации</i> и суммы квадратов весов.</p> <p>*Вместо квадратичной функции используется кусочно-линейная, и задается отступ <i>eps</i> (по умолчанию, равная 0.1): Если разница между прогнозируемым и истинным значением</p>	<p>Степень для ядра полинома</p>	<p>Если в качестве функции ядра используется полиномиальная функция ('poly'), которая является методом нелинейной регрессии, то зависимая переменная связана с независимыми переменными n-ой степени. В поле указывается степень этого ядра.</p>
	<p>Коэффициент регуляризации</p>	<p>Мера степени наказания модели за каждую неверно спрогнозированную точку.</p>
	<p>Оптимизация гиперпараметров</p>	<p>Флаг подбора гиперпараметров.</p>
	<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>
	<p>Количество фолдов для оптимизации</p>	<p>Указывается, на сколько равных частей разбивается входной датасет при обучении модели.</p>

	<p>меньше ϵs (прогнозное значение попадает в пространство гиперплоскости), модель не считает это за ошибку, иначе – берется модуль разницы.</p>										
<p>Байесовская гребневая регрессия</p>	<p>В основе метода лежит формула Байеса, которая дает возможность оценить вероятность событий эмпирическим путем.</p> <p><i>Гребневая</i> регрессия – один из методов снижения размерности. Для гребневой регрессии к функции потерь прибавляется параметр λ, обозначающий размер штрафа. Чем меньше λ, тем выше <i>дисперсия</i> и ниже <i>смещение</i>.</p>	<table border="1"> <tr> <td data-bbox="857 619 1115 842">alpha_1, alpha_2</td> <td data-bbox="1115 619 1697 842">Допустимые максимальные расстояния графика регрессии до верхнего и нижнего доверительного интервала.</td> </tr> <tr> <td data-bbox="857 842 1115 1066">lambda_1, lambda_2</td> <td data-bbox="1115 842 1697 1066">Размеры штрафов при выходе прогнозируемых значений за пределы верхнего и нижнего доверительного интервала.</td> </tr> <tr> <td data-bbox="857 1066 1115 1289">Оптимизация гиперпараметров</td> <td data-bbox="1115 1066 1697 1289">Флаг подбора гиперпараметров.</td> </tr> <tr> <td data-bbox="857 1289 1115 1393">Метрика для</td> <td data-bbox="1115 1289 1697 1393">Выбирается одна из предлагаемых метрик для оценки</td> </tr> </table>	alpha_1, alpha_2	Допустимые максимальные расстояния графика регрессии до верхнего и нижнего доверительного интервала.	lambda_1, lambda_2	Размеры штрафов при выходе прогнозируемых значений за пределы верхнего и нижнего доверительного интервала.	Оптимизация гиперпараметров	Флаг подбора гиперпараметров.	Метрика для	Выбирается одна из предлагаемых метрик для оценки	<p>**</p>
alpha_1, alpha_2	Допустимые максимальные расстояния графика регрессии до верхнего и нижнего доверительного интервала.										
lambda_1, lambda_2	Размеры штрафов при выходе прогнозируемых значений за пределы верхнего и нижнего доверительного интервала.										
Оптимизация гиперпараметров	Флаг подбора гиперпараметров.										
Метрика для	Выбирается одна из предлагаемых метрик для оценки										

	<p>Смещение – это погрешность оценки, возникающая в результате ошибочного предположения в алгоритме обучения. В результате большого смещения алгоритм может пропустить связь между признаками и выводом (недообучение).</p> <p>Дисперсия – это ошибка чувствительности к малым отклонениям в тренировочном наборе. При высокой дисперсии алгоритм может трактовать случайный шум в тренировочном наборе, а не желаемый результат (переобучение).</p>	<table border="1"> <tr> <td data-bbox="853 151 1115 284">оптимизаци и</td> <td data-bbox="1115 151 1691 284">работы модели.</td> </tr> <tr> <td data-bbox="853 284 1115 507">Количество фолдов для оптимизаци и</td> <td data-bbox="1115 284 1691 507">–</td> </tr> </table>	оптимизаци и	работы модели.	Количество фолдов для оптимизаци и	–	
оптимизаци и	работы модели.						
Количество фолдов для оптимизаци и	–						
Метод k-ближайших	Для регрессии объекту		**				

<p>соседей для регрессии</p> <p>присваивается среднее значение по k ближайшим к нему объектам, значения которых уже известны.</p> <p>Алгоритм применяется к выборке с большим количеством атрибутов (многомерной). Для этого перед применением определяется функция расстояния, классический вариант такой функции – <i>евклидова метрика</i>.</p> <p>Разные признаки могут иметь разный диапазон представленных значений в выборке, поэтому выполняется <i>нормализация</i> данных.</p> <p>Некоторые значимые признаки могут быть</p>	<p>Количество ближайших соседей</p> <p>Число k, характеризующее количество соседей в кластере.</p>
	<p>Тип веса для соседей</p> <p>Задается одно из значений: ‘uniform’ (единый – всем признакам присваивается единый вес), или ‘distance’ (по расстоянию).</p> <p>Значение по умолчанию – единый.</p>
	<p>Метрика расстояния</p> <p>Задается одно из значений: ‘chebyshev’ (Чебышева), ‘euclidean’ (Евклидова), ‘cosine’ (Косинусное), ‘cityblock’ (Манхэттенское).</p> <p>Значение по умолчанию – евклидово расстояние, когда вычисляется расстояние между всеми точками попарно.</p>
	<p>Оптимизация гиперпараметров</p> <p>Флаг подбора гиперпараметров.</p>
<p>Метрика для</p> <p>Выбирается одна из предлагаемых метрик для оценки</p>	

	<p>важнее остальных, поэтому для каждого признака задается определенный <i>вес</i>.</p> <p>Алгоритм предполагает, что похожие наблюдения существуют в непосредственной близости:</p> <p>улавливается идея сходства (иногда называемого расстоянием или близостью) благодаря вычислению Евклидова расстояния между точками.</p>	<p>оптимизаци и</p>	<p>работы модели.</p>	
		<p>Количество фолдов для оптимизаци и</p>	<p>Указывается, на сколько равных частей разбивается входной датасет при обучении модели.</p>	
<p>Авторегрессия – модель временного ряда, в которой ее текущее значение <i>линейно</i> зависит от предыдущих (ретроспективных) значений этого же ряда. <i>Линейная зависимость</i> означает, что текущее значение равно взвешенной сумме нескольких предыдущих значений ряда.</p> <p>Зная параметры модели и соответствующие <i>ретроспективные</i> значения временного ряда, можно предсказать его будущие значения. Основное назначение авторегрессионной модели – прогнозирование. Также с ее помощью можно производить анализ временных рядов – выявлять тенденции, сезонность, и другие особенности.</p>				
<p>ARIMA/ SARIMAX</p>	<p>Авторегрессионное интегрированное скользящее среднее (с англ. ARIMA –</p>			<p>**</p>
		<p>Число шагов</p>	<p>Количество шагов, на которые</p>	

<p>autoregressive integrated moving average) используется при работе с временными рядами для более глубокого понимания данных, или предсказания будущих точек ряда. Упоминается как ARIMA (p, d, q), где p, d и q – целые неотрицательные числа, характеризующие порядок для частей модели (соответственно – авторегрессионной, интегрированной и скользящего среднего).</p> <p>Авторегрессия. Модель, использующая зависимую связь между наблюдением и некоторым количеством</p>	<p>для прогноза</p>	<p>модель будет предсказывать.</p>
	<p>Порядок авторегрессии, p</p>	<p>Количество запаздывающих наблюдений, включенных в модель, также называется лаговый порядок. P помогает настроить линию для прогнозирования серии. Чисто авторегрессионные модели напоминают линейную регрессию, где прогностическими переменными являются p числа предыдущих периодов.</p>
	<p>Порядок интегрирования, d</p>	<p>Число обычных дифференцирований – количество раз, когда необработанные наблюдения различаются, также называется степенью различия. В модели ARIMA временные ряды преобразуются в <i>стационарные</i> (серии без тренда и сезонности), используя дифференцирование. Стационарный ряд – это когда среднее значение и дисперсия постоянны во времени.</p>

<p>запаздывающих наблюдений.</p> <p>Интегрированный. Использование разности необработанных наблюдений (например, вычитание наблюдения из наблюдения на предыдущем временном шаге), чтобы сделать временной ряд стационарным.</p> <p>Скользящая средняя. Модель, в которой используется зависимость между наблюдением и остаточной ошибкой из модели скользящего среднего, применяемая к запаздывающим наблюдениям.</p>	<p>Порядок скользящего среднего, q</p>	<p>Размер окна скользящей средней.</p>
	<p>Параметры модели SARIMAX:</p>	
	<p>Порядок авторегрессии</p>	<p>—</p>
	<p>Порядок интегрирования</p>	<p>Число сезонных дифференцирований.</p>
	<p>Порядок скользящего среднего</p>	<p>—</p>
	<p>Сезонный период</p>	<p>Число наблюдений за сезон.</p>
	<p>Оптимизация гиперпараметров</p>	<p>Флаг подбора гиперпараметров.</p>

	<p>Модель SARIMAX используется для временных рядов с учетом сезонности.</p>				
Группа «Работа с текстами»					
<p>Автореферирование текста</p>	<p>Данная функция представляет собой автоматический процесс выделения краткого содержания текста с помощью модели машинного обучения. На выходе получается датасет заданного объема, который можно представить в виде таблицы.</p>	<table border="1"> <tr> <td>Объем автореферата</td> <td>Максимальное количество символов в выходном результате</td> </tr> </table>	Объем автореферата	Максимальное количество символов в выходном результате	<p>Таблица с кратким содержанием</p>
Объем автореферата	Максимальное количество символов в выходном результате				
Группа «Управление моделями»					
<p>Сохранение модели</p>	<p>Сохраняет модель по настроенному в системе пути и названию файла, а также сохраняет словарь с переменными. В этом словаре содержится</p>	<table border="1"> <tr> <td>Название модели</td> <td>Пользователь задает название для обучаемой модели.</td> </tr> </table>	Название модели	Пользователь задает название для обучаемой модели.	<p>Созданная модель сохраняется в пункте меню системы Модели – > Сохранённые модели</p>
Название модели	Пользователь задает название для обучаемой модели.				

	отдельно список независимых переменных и список целевых признаков, с указанием выполненных над ними преобразований. Сохраняет шаг ресемплирования				
Классификация					
Сохранение модели классификации и изображений	Функция предназначена для сохранения в системе модели классификации изображений.	<table border="1"> <tr> <td>Название модели</td> <td>Пользователь задает название для обучаемой модели.</td> </tr> </table>	Название модели	Пользователь задает название для обучаемой модели.	**
Название модели	Пользователь задает название для обучаемой модели.				
Обнаружение объектов					
Сохранение модели YOLO v5	<p>* данный функционал находится в разработке, в текущей версии 1.0.0 применение функции недоступно.</p> <p>Функция предназначена для сохранения в системе модели распознавания изображений «YOLO v5». Сохранив модель, можно создать на ее</p>	<table border="1"> <tr> <td>Название модели</td> <td>Пользователь задает название для обучаемой модели.</td> </tr> </table>	Название модели	Пользователь задает название для обучаемой модели.	**
Название модели	Пользователь задает название для обучаемой модели.				

	<p>основе приложение для последующей интеграции со сторонними системами. Также обученная модель может использоваться повторно для анализа онлайн данных.</p>				
Spark					
<p>Сохранение модели Spark</p>	<p>Функция предназначена для сохранения в системе моделей, собранных с применением технологии Spark.</p>	<table border="1" data-bbox="846 687 1704 863"> <tr> <td data-bbox="846 687 1019 863">Название модели</td> <td data-bbox="1019 687 1704 863">Пользователь задает название для обучаемой модели.</td> </tr> </table>	Название модели	Пользователь задает название для обучаемой модели.	
Название модели	Пользователь задает название для обучаемой модели.				
Группа «Глубокое обучение»					
<p>Валидация модели классификации и изображений</p>	<p>После того, как <i>модель нейронной сети</i> обучена, натренирована и для нее выбраны оптимальные гиперпараметры, необходимо</p>	<table border="1" data-bbox="846 1114 1688 1417"> <tr> <td data-bbox="846 1114 1019 1417">Метрика</td> <td data-bbox="1019 1114 1688 1417"> <p>Метрика, которая оценивает работу обученной модели нейронной сети. Применяется к типу данных – изображения.</p> <p>Чтобы оценить качество модели классификации используются</p> </td> </tr> </table>	Метрика	<p>Метрика, которая оценивает работу обученной модели нейронной сети. Применяется к типу данных – изображения.</p> <p>Чтобы оценить качество модели классификации используются</p>	<p>1. Строится <i>матрица ошибок</i>. Пример для бинарной классификации:</p>
Метрика	<p>Метрика, которая оценивает работу обученной модели нейронной сети. Применяется к типу данных – изображения.</p> <p>Чтобы оценить качество модели классификации используются</p>				

проверить ее точность и адекватность. Для этого выполняется валидация *итоговой модели нейронной сети* на тестовой выборке.

В качестве *входных данных* для функции используются:

- тестовый датасет с изображениями;

- обученная модель;

- словарь с преобразованиями данных;

- выбранная метрика валидации.

Рассчитывается сколько изображений тестовой выборки попадают в каждую ячейку матрицы ошибок.

Оценивается качество классификации.

следующие метрики:

1. **Accuracy** – оценивает долю правильных ответов модели.

2. **F1** (среднее гармоническое) – агрегированная функция, которая позволяет вместо точности и полноты использовать только один параметр качества классификации. Формула:

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

Чем ближе F1 к 1, тем лучше.

*В задачах, в которых точность и полнота не равноценны, применяется взвешенное значение F_β .

3. **Precision**. Метрика, которая оценивает *точность модели*. Рассчитывается по формуле:

$$precision = \frac{TP}{TP + FP}$$

здесь считается точность для класса 1 (для класса 0 считается аналогично).

4. **Recall** – оценивает *полноту модели*:

	Истинный класс	
	0	1
0	TN	FN
1	FP	TP

где столбцы – истинные классы, а строки – предсказанные классы.

Обозначения:

TN – true negative

TP – true positive

FN – false negative

FP – false positive

Например, ячейка **TP** означает, что объект действительно принадлежит классу 1, и для него предсказан класс 1. А **FN** – объект неправильно отнесли к классу 0, хотя он принадлежит к классу 1.

2. Отображается значение выбранной метрики.

$$recall = \frac{TP}{TP + FN}$$

Идеально, чтобы точность и полнота были равны 1 (100%).

5. **AUC_ROC**. Для анализа качества модели применяется ROC-анализ: строится ROC-кривая, которая наиболее часто используется для представления результатов бинарной классификации. Классов два: один называется классом с положительными исходами, второй – с отрицательными исходами. ROC-кривая показывает зависимость количества верно классифицированных положительных примеров от количества неверно классифицированных отрицательных примеров.

Чем выше показатель AUC_ROC, тем качественнее классификатор, при этом значение 0,5 демонстрирует непригодность выбранного метода классификации. AUC_ROC = 0,9-1,0 означает отличное качество модели.

<p>Классификация (табличные данные)</p> <p>Типы решаемых задач: анализ тональности, классификация текста по категориям, распознавание речи и многое другое.</p> <p>Рассмотрим функцию на примере задачи по отнесению документов к определенной категории на основании его содержания. Процесс классификации осуществляется с помощью применения методов машинного обучения, в частности <i>сверточных нейронных сетей</i>.</p> <p>Задача классификации текстов применима в решении следующих задач:</p> <p>борьба с массовой рассылкой рекламы,</p>	<p>Количество эпох</p> <p>Параметр, который показывает сколько раз <i>модель</i> подвергается воздействию обучения.</p>	**
	<p>Размер мини-батча</p> <p>Количество обучающих примеров за одну итерацию. Под примерами имеются в виду <i>наблюдения</i> – строки в табличных данных.</p>	
	<p>Метрика для обучения</p> <p>Метрика «Ассигасу» (точность) показывает долю правильных ответов алгоритма</p>	
	<p>Алгоритм градиентного спуска</p> <p>Метод нахождения минимального значения <i>функции потерь</i>. Алгоритмы на выбор: <i>SGD, Adam</i>, и др.</p>	
	<p>Шаг градиентного спуска</p> <p>Параметр, регулирующий скорость обучения модели. Значения – 0.001 или 0.1.</p>	
	<p>Функция потерь</p> <p>Выбирается одна из функций, в зависимости от задачи:</p>	

<p>распознавание тональности текстов, сортировка документов и т.д.</p> <p>Задача определяется следующим образом: пусть существует конечное множество категорий, на вход алгоритма подается конечное количество документов, и есть целевая функция, которая определяет соответствие для каждой пары (документ, категория). Задача состоит в нахождении этой функции, называемой классификатором.</p> <p>Строится многослойная нейронная сеть, состоящая из слоев:</p>		бинарная классификация или многоклассовая.
	Добавить слой	Задается одно из трех значений: Conv2D, Flatten, Dense (последний – полносвязный слой).
	Перемешивать выборку перед обучением	Необходимо установить галочку в поле, чтобы случайным образом поменять местами наблюдения, соответствующие строкам в таблице
	Порог классификации	Для бинарной классификации значение по умолчанию 0.5, для многоклассовой – параметр не заполняется
	Флаг возврата вероятности при прогнозе	Возвращает вероятность или метки классов для дальнейшего прогноза после обучения модели. Для бинарной классификации может использоваться одно поле с метками 0 и 1, обозначающими

<p>-<i>входной</i>, на который поступают входные признаки;</p> <p>-<i>скрытый</i>, на котором рассчитываются промежуточные результаты;</p> <p>- <i>выходной</i>, на котором выводятся окончательные значения, вычисленные по гипотезе.</p> <p><i>*Сверточными</i> искусственные нейронные сети называются из-за специальной архитектуры, с наличием операций свертки.</p>		<p>принадлежность к тому или иному классу. Для многоклассовой классификации используется несколько полей – каждое поле соответствует отдельному классу (0, 1, 2 и т.д.), и записываются вероятности (от 0 до 1), с которыми наблюдения принадлежат классам</p>	
	<p>Оптимизация гиперпараметров</p>	<p>Алгоритм подбирает гиперпараметры</p>	
	<p>Метрика для оптимизации</p>	<p>Выбирается одна из предлагаемых метрик для оценки работы модели.</p>	
	<p>Количество фолдов для оптимизации</p>	<p>–</p>	

Классификация изображений	<p>Алгоритм решения задачи классификации:</p> <p>1. Берется <i>тренировочная выборка</i> – набор изображений с известными значениями целевого признака Y. Нейронная сеть должна восстановить зависимость между нецелевыми признаками и целевым.</p> <p>2. Задаются основные параметры нейронной сети.</p> <p>3. Выписываются выражения для вероятностей принадлежности наблюдения к тому или иному классу ($Y = 0, 1, 2, 3, \text{ и т.д.}$).</p>	<p>Количество эпох</p> <p>Это гиперпараметр, который определяет сколько раз <i>алгоритм обучения</i> будет обрабатывать весь <i>набор обучающих данных</i>. То есть <i>эпоха</i> – одна итерация в процессе обучения, включающая предъявление всех примеров из обучающего множества</p>	**
		<p>Метрика для обучения</p> <p>Метрика «Accuracy» (точность) показывает долю правильных ответов алгоритма</p>	
		<p>Алгоритм градиентного спуска</p> <p>Метод нахождения минимального значения <i>функции потерь</i>. Минимизация любой функции означает поиск самой глубокой впадины в этой функции. Функция используется, чтобы контролировать ошибку в прогнозах модели. Поиск минимума означает получение наименьшей возможности ошибки или повышение</p>	

	<p>4. По тренировочной выборке составляется функция потерь.</p> <p>5. Функция потерь $L(w)$ содержит вхождения весов нейронной сети. Относительно этих переменных находится <i>точка минимума функции</i> $L(w)$.</p> <p>6. Точка минимума определяет оптимальные веса нейронной сети.</p> <p>7. Весам нейронной сети присваиваются найденные оптимальные значения. Пусть изображение A не принадлежит тренировочной выборке. Объект A прогоняется через нейронную сеть и на выходе получаются</p>	<p>точности модели. Точность увеличивается перебором <i>учебных</i> данных при настройке параметров модели (весов и смещений).</p> <p>Суть алгоритма – процесс получения наименьшего значения ошибки. Аналогично это можно рассматривать как спуск во впадину в попытке найти самое низкое значение ошибки.</p> <p>Можно выбрать один из следующих алгоритмов: <i>SGD</i> (Stochastic gradient descent, с англ. Стохастический градиентный спуск), <i>Adam</i>, и др. Алгоритм Adam является <i>модифицированным</i>, в нем также выполняется минимизация функции потерь. Рассчитываются векторы <i>частных</i> в текущей точке функции, и определяются координаты следующей точки. Частные производные вычисляются, чтобы определить,</p>	
--	--	---	--

<p>вероятности – они и являются предсказаниями для объекта А (по максимальной вероятности определяется принадлежность объекта к классу).</p>		какой был вклад в ошибку по каждому весу.
	Шаг градиентного спуска	<p>Параметр, регулирующий скорость обучения модели – насколько быстро функция потерь спускается к своему минимуму (скорость спуска/поиска). Выбирается значение: 0.001, или 0.1.</p>
	Функция потерь	<p>Функция потерь находится в центре нейронной сети. Она используется для расчета ошибки между <i>реальными</i> и <i>полученными</i> ответами. Главная цель – минимизировать эту ошибку. Или: максимизировать вероятность принадлежности к истинному классу для каждого объекта из тренировочной выборки. Она также может зависеть от таких переменных, как веса и смещения, где <i>смещения</i> – это веса, добавленные к скрытым слоям.</p>

			<p>Выбирается одна из функций, в зависимости от задачи: бинарная классификация или многоклассовая.</p>	
		<p>Добавить слой</p>	<p>Основой алгоритмов распознавания изображений являются сверточные нейронные сети. Для их построения используются три главных типа слоев: сверточный слой, слой подвыборки и полносвязный слой. Соответственно пользователь задает одно из трех значений: Conv2D, Flatten, Dense (последний – полносвязный).</p> <p>В сверточных нейронных сетях одно изображение является одним наблюдением. Таким образом, исходное изображение преобразуется, слой за слоем, от начального значения пикселя до итоговой оценки класса. Слои, идущие до полносвязного, являются средствами предобработки изображения, и используются для выделения различных признаков, которые</p>	

			затем подаются на вход классификатору.	
		Порог классификации	Важно! Параметр заполняется только для задачи бинарной классификации, значение по умолчанию 0.5. Для множественной классификации это поле остается пустым.	
Регрессия				
Регрессия (табличные данные)	Для обучения нейронной сети данные делятся на части меньшего размера, загружают их по очереди и обновляют веса нейросети в конце каждого шага, подстраивая их под данные.	Количество эпох	<p>Указывается количество эпох для обучения модели.</p> <p>Одна эпоха – весь датасет прошел через нейронную сеть в прямом и обратном направлении только один раз. Так как одна эпоха слишком велика для компьютера, датасет делят на партии – <i>батчи</i>.</p> <p>С увеличением числа эпох, веса нейронной сети изменяются все большее количество раз. Кривая с каждым разом лучше подстраивается под данные, переходя последовательно из</p>	**

			<p>плохо обученного состояния в оптимальное. Если вовремя не остановиться, то может произойти переобучение.</p>
		Размер мини-батча	<p>Общее число тренировочных объектов, представленных в одном батче.</p> <p>Нельзя пропустить через нейронную сеть разом весь датасет. Поэтому делим данные на пакеты, сеты или партии.</p> <p><i>Итерации</i> – это число батчей, необходимое для завершения одной эпохи.</p>
		Метрика для обучения	<p>Для регрессии: ['MAE', 'MAPE', 'MSE'].</p>
		Алгоритм градиентного спуска	<p>Алгоритм итеративной оптимизации, используемой в машинном обучении для получения более точного результата. <i>Градиент</i> показывает скорость убывания или возрастания функции. <i>Спуск</i> говорит о том, что мы имеем</p>

			<p>дело с убыванием.</p> <p>Алгоритм итеративный, процедура проводится несколько раз, чтобы добиться оптимального результата. На каждом шаге результат получается лучше.</p>	
		Шаг градиентного спуска	Скорость обучения модели алгоритмом градиентного спуска.	
		Функция потерь	<p>Функция, которая используется для оптимизации алгоритма машинного обучения. Значение, вычисленное такой функцией, называется ‘потерей’.</p> <p>Потери регрессии рассчитываются путем прямого сравнения выходного и истинного значения. Самая популярная функция для регрессионных моделей – это среднеквадратическая ошибка, MSE.</p> <p>Функция потерь определяет, как именно выходные данные</p>	

			<p>связаны с исходными. По сути вычисляется насколько хорошо работает модель – сравнивается то, что модель прогнозирует, с фактическим значением. Сохраняется функция потерь, которая может эффективно наказывать модель, пока та обучается на тренировочных данных.</p>	
		Добавить слой	Из списка выбирается дополнительный слой для нейросети.	
		Перемешивать выборку перед обучением	Необходимо установить галочку в поле, чтобы случайным образом поменять местами наблюдения.	
		Оптимизация гиперпараметров	Если установить галочку в поле, то алгоритм выберет наилучшие гиперпараметры для создания модели из списка предложенных	
		Метрика для оптимизации	Выбирается одна из предлагаемых метрик для оценки работы модели.	

		и					
		Количество фолдов для оптимизации	Указывается, на сколько равных частей разбивается входной датасет при обучении модели.				
Обнаружение объектов							
YOLOv5	<p>* данный функционал находится в разработке, в текущей версии 1.0.0 применение функции недоступно.</p> <p>Датасет должен быть разделен на две папки: <i>train</i> (тренировочная выборка) и <i>val</i> (валидационная). В каждой папке лежат еще две папки: <i>images</i> (изображения) и <i>labels</i> – папка с текстовыми файлами, содержащими метки объектов на этих изображениях в формате YOLO.</p> <p>Для этой функции предварительно</p>	<table border="1"> <tr> <td>Размер мини-батча</td> <td>Указывается количество изображений, которое одновременно подается на вход YOLO. Например, если задать размер 2, за один подход подается два изображения.</td> </tr> <tr> <td>Количество эпох</td> <td>Это гиперпараметр, который определяет сколько раз алгоритм обучения будет обрабатывать весь набор обучающих данных. То есть эпоха – одна итерация в процессе обучения, включающая предъявление всех примеров из обучающего множества. Например, для</td> </tr> </table>	Размер мини-батча	Указывается количество изображений, которое одновременно подается на вход YOLO. Например, если задать размер 2, за один подход подается два изображения.	Количество эпох	Это гиперпараметр, который определяет сколько раз алгоритм обучения будет обрабатывать весь набор обучающих данных. То есть эпоха – одна итерация в процессе обучения, включающая предъявление всех примеров из обучающего множества. Например, для	<p>В результате получаются изображения с обозначением детектированных объектов и значениями <i>confidence</i>. Где <i>confidence</i> – число от 0 до 1, характеризующее ‘уверенность’ модели в том, что детектирован объект или детектирован объект определенного класса. Еще один параметр <i>conf-thres</i> позволяет установить пороговое значение для <i>confidence</i> модели. Все объекты, <i>confidence</i> которых ниже этого</p>
Размер мини-батча	Указывается количество изображений, которое одновременно подается на вход YOLO. Например, если задать размер 2, за один подход подается два изображения.						
Количество эпох	Это гиперпараметр, который определяет сколько раз алгоритм обучения будет обрабатывать весь набор обучающих данных. То есть эпоха – одна итерация в процессе обучения, включающая предъявление всех примеров из обучающего множества. Например, для						

	<p>выполняется разметка изображений на тренировочной и валидационной выборках.</p> <p>Пользователь с помощью 'bounding box' отмечает объекты на изображениях. Алгоритм по точкам объектов находит функцию их обнаружения. На валидационной выборке проверяется точность обученной модели.</p> <p>После запуска функции на выходе получается обученная нейронная сеть, результаты обучения которой сохраняются в БД.</p>	<table border="1" data-bbox="857 156 1691 343"> <tr> <td data-bbox="857 156 1124 343"></td> <td data-bbox="1124 156 1691 343"> <p>выборки в десять изображений и размера мини-батча два, эпоха равна пяти прохождениям.</p> </td> </tr> </table>		<p>выборки в десять изображений и размера мини-батча два, эпоха равна пяти прохождениям.</p>	<p>значения не считаются объектами.</p> <p>Также отображаются: описание модели, графики обучения модели, матрица ошибок на валидационных данных. Где описание модели содержит информацию об оптимизаторе, тренировочном и валидационном датасетах, а также параметры обучения.</p>
	<p>выборки в десять изображений и размера мини-батча два, эпоха равна пяти прохождениям.</p>				
Отправка уведомлений					
<p>Отправка уведомлений</p>	<p>Функция предназначена для</p>		<p>Оповещение в телеграм канал</p>		

	<p>осуществления отправки уведомлений в настроенный канал телеграм или по электронной почте (не реализовано в текущей версии). Данная функция выступает в паре с блоком шлюз, где необходимо задать условия, при которых будет отправлено уведомление.</p>	<p>Канал уведомлений</p>	<p>Выбирается ранее настроенный в разделе Администрирование -> Уведомления канал, на который будет осуществляться отправка сообщений.</p>	
--	--	---------------------------------	--	--

Spark. Группа функций для фреймворка Apache Spark. Названия функций дублируются с теми, что были описаны ранее, разница заключается в использовании другого модуля машинного обучения в Apache Spark *(в следующих версиях планируется сделать все функции универсальными)*.

<p>Сохранение датасета Spark в CSV</p>	<p>Функция распределяет входные данные в несколько файлов в одну директорию. Для этого выбираем: куда сохранить данные, как их назвать, и по необходимости можем подгрузить</p>	<p>Путь до директории для датасета</p>	<p>Выбирается путь до папки, в которую будут сохраняться данные.</p>	<p>Таблица в формате csv с датасетом. Сохраняется в раздел данные</p>
		<p>Название датасета</p>	<p>В этом поле задается название для датасета. По умолчанию датасеты создаются с названиями формата <i>pySpark.csv</i>.</p>	

	<p>новую порцию данных.</p>	<p>Добавить данные к датасету</p>	<p>Если преобразованные данные необходимо сохранять не в виде отдельного файла, а добавить к уже существующему и загруженному, необходимо установить галочку у данного признака. По умолчанию файл перезаписывается.</p>
		<p>Название датасета для валидации</p>	<p>Указывается название, с которым будет сохранен датасет для валидации при активации параметра «Сохранить датасет для валидации»</p>
		<p>Сохранить датасет для валидации</p>	<p>В процессе работы пайплайна, исходный вид набора данных данных теряется, поэтому его нужно передать из блока "Загрузка данных" в конец пайплайна в блок сохранени. Датасет для валидации это и есть нетронутый набор данных в первоначальном виде, к нему только добавляется столбец с результатами.</p>

		Загрузка датасета для валидации в БД	Позволяет загрузить датасет для валидации напрямую в базу данных ClickHouse
Косинусное расстояние	На вход функция получает новые данные для анализа (датасет в формате csv), обученную модель, и числовой вектор. Выполняется поиск объектов, наиболее схожих с заданным вектором, и в качестве меры схожести используется косинусное расстояние - расстояние между значениями во входном векторе и значениями выбранных столбцов в наблюдениях.	—	Таблица «Косинусное расстояние»

Выбор признаков и целевых признаков	Аналогично стандартной функции.				
Разделение датасета на обучающую и тестовую выборки	Аналогично стандартной функции.				
Валидация модели	Аналогично стандартной функции.				
Прогноз модели	Аналогично стандартной функции.				
Порядковое кодирование признаков	Порядковое кодирование - это метод преобразования категориальных данных в цифровой вид. Применяются, когда в датасете существуют НЕ числовые признаки, которые заданы словами и для дальнейшего анализа их нужно преобразовать в числа. Порядковое	<table border="1"> <tr> <td>Выбранные признаки</td> <td>Указываются признаки, над которыми необходимо провести операцию порядкового кодирования.</td> </tr> </table>	Выбранные признаки	Указываются признаки, над которыми необходимо провести операцию порядкового кодирования.	
Выбранные признаки	Указываются признаки, над которыми необходимо провести операцию порядкового кодирования.				

	кодирование позволяет пронумеровать признаки по порядку.		
Нормализация признаков	<p>Нормализация - это приведение числовых признаков к единой шкале. Бывает что числовой признак имеет минимальное и максимальное значение в очень широком диапазоне и это плохо для машинного обучения. Например, есть числовой признак, чье минимальное значение равно 0,001, а максимальное - 100000, нормализация преобразовывает их к диапазону от 0 до 1, то есть 0.001 становится 0, а 100000 становится 1, значения между ними также преобразуются, 50 000 станет примерно</p>	—	—

	равным 0.5. Данная функция позволяет оптимизировать дальнейшие вычисления.				
Модель градиентного бустинга Spark для бинарной классификации	<p>Градиентный бустинг представляет собой ансамбль деревьев решений. В основе данного алгоритма лежит итеративное обучение деревьев решений с целью минимизировать функцию потерь. Благодаря особенностям деревьев решений градиентный бустинг способен работать с категориальными признаками, справляться с нелинейностями. Бустинг – это метод преобразования слабообученных моделей в хорошообученные. В</p>	<table border="1" data-bbox="853 448 1693 708"> <tr> <td data-bbox="853 448 1093 708">Количество базовых моделей</td> <td data-bbox="1093 448 1693 708">Указывается число последовательных итераций по оптимизации модели, то есть количество попыток уменьшить ошибку.</td> </tr> </table>	Количество базовых моделей	Указывается число последовательных итераций по оптимизации модели, то есть количество попыток уменьшить ошибку.	-
Количество базовых моделей	Указывается число последовательных итераций по оптимизации модели, то есть количество попыток уменьшить ошибку.				

	<p>бустинге каждое новое дерево обучается на модифицированной версии исходного датасета.</p>										
<p>Кластеризация Spark DBSCAN</p>	<p>Алгоритм DBScan формирует группы коренных соседей/кластеры, объединяя точки, расположенные рядом. А точки, которые не попадают ни в одну из групп, отмечаются меткой -1 и приравниваются к аномалиям.</p>	<table border="1"> <tr> <td data-bbox="853 491 1093 767"> <p>Порог для отнесения кластера к аномалиям</p> </td> <td data-bbox="1093 491 1688 767"> <p>Указывается граница, которая определяет, когда нужно отнести кластер к аномалии.</p> <p>ество попыток уменьшить ошибку.</p> </td> </tr> <tr> <td data-bbox="853 767 1093 986"> <p>Радиус</p> </td> <td data-bbox="1093 767 1688 986"> <p>Радиус в единицах расстояния, в рамках которого выполняется поиск потенциальных соседей (float/list/tuple).</p> </td> </tr> <tr> <td data-bbox="853 986 1093 1204"> <p>Число соседей</p> </td> <td data-bbox="1093 986 1688 1204"> <p>Минимальное число ближайших соседей в указанном радиусе для формирования группы коренных соседей (int/list/tuple).</p> </td> </tr> <tr> <td data-bbox="853 1204 1093 1404"> <p>Метрика расстояния</p> </td> <td data-bbox="1093 1204 1688 1404"> <p>Метрика расстояния (str/list): расстояние Евклида, косинусное расстояние. По умолчанию «Евклидово расстояние» –</p> </td> </tr> </table>	<p>Порог для отнесения кластера к аномалиям</p>	<p>Указывается граница, которая определяет, когда нужно отнести кластер к аномалии.</p> <p>ество попыток уменьшить ошибку.</p>	<p>Радиус</p>	<p>Радиус в единицах расстояния, в рамках которого выполняется поиск потенциальных соседей (float/list/tuple).</p>	<p>Число соседей</p>	<p>Минимальное число ближайших соседей в указанном радиусе для формирования группы коренных соседей (int/list/tuple).</p>	<p>Метрика расстояния</p>	<p>Метрика расстояния (str/list): расстояние Евклида, косинусное расстояние. По умолчанию «Евклидово расстояние» –</p>	
<p>Порог для отнесения кластера к аномалиям</p>	<p>Указывается граница, которая определяет, когда нужно отнести кластер к аномалии.</p> <p>ество попыток уменьшить ошибку.</p>										
<p>Радиус</p>	<p>Радиус в единицах расстояния, в рамках которого выполняется поиск потенциальных соседей (float/list/tuple).</p>										
<p>Число соседей</p>	<p>Минимальное число ближайших соседей в указанном радиусе для формирования группы коренных соседей (int/list/tuple).</p>										
<p>Метрика расстояния</p>	<p>Метрика расстояния (str/list): расстояние Евклида, косинусное расстояние. По умолчанию «Евклидово расстояние» –</p>										

		<table border="1"> <tr> <td></td> <td>используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к кластеру.</td> </tr> <tr> <td>Флаг векторизации признаков</td> <td>Параметр определяет будет ли проводиться векторизация или нет</td> </tr> <tr> <td>Столбец для группировки перед векторизацией</td> <td>Указывается номер столбца</td> </tr> </table>		используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к кластеру.	Флаг векторизации признаков	Параметр определяет будет ли проводиться векторизация или нет	Столбец для группировки перед векторизацией	Указывается номер столбца	
	используется при кластеризации данных в текущем датасете, а также при отнесении нового объекта к кластеру.								
Флаг векторизации признаков	Параметр определяет будет ли проводиться векторизация или нет								
Столбец для группировки перед векторизацией	Указывается номер столбца								
Дополнительные функции									
Расчет параметров графа	С помощью данной функции осуществляется расчет количества вершин и ребер в выбранном графе.		Таблица «Количество вершин и ребер графа»						

<p>Поиск ближайших вершин</p>	<p>С помощью данной функции осуществляется определение id вершин, ближайших к указанным координатам, и расстояний между ними. Может быть вычислено как для одной точки, так и для нескольких точек.</p>	<table border="1"> <tr> <td data-bbox="853 212 1093 475"> <p>Введите координаты долготы</p> </td> <td data-bbox="1093 212 1691 475"> <p>Координата долготы точки, для которой производится поиск ближайшей вершины в графе: восточная долгота до 180, западная долгота до -180.</p> </td> </tr> <tr> <td data-bbox="853 475 1093 738"> <p>Введите координаты широты</p> </td> <td data-bbox="1093 475 1691 738"> <p>Координата широты точки, для которой производится поиск ближайшей вершины в графе: северная широта до +90, южная широта до -90.</p> </td> </tr> </table>	<p>Введите координаты долготы</p>	<p>Координата долготы точки, для которой производится поиск ближайшей вершины в графе: восточная долгота до 180, западная долгота до -180.</p>	<p>Введите координаты широты</p>	<p>Координата широты точки, для которой производится поиск ближайшей вершины в графе: северная широта до +90, южная широта до -90.</p>	<p>Таблица «Ближайшие вершины»</p>
<p>Введите координаты долготы</p>	<p>Координата долготы точки, для которой производится поиск ближайшей вершины в графе: восточная долгота до 180, западная долгота до -180.</p>						
<p>Введите координаты широты</p>	<p>Координата широты точки, для которой производится поиск ближайшей вершины в графе: северная широта до +90, южная широта до -90.</p>						
<p>Вычисление k кратчайших путей в графе</p>	<p>Определение k кратчайших путей между парой заданных пользователем вершин в графе</p>	<table border="1"> <tr> <td data-bbox="853 914 1093 1177"> <p>Количество базовых моделей</p> </td> <td data-bbox="1093 914 1691 1177"> <p>Указывается число последовательных итераций по оптимизации модели, то есть количество попыток уменьшить ошибку.</p> </td> </tr> <tr> <td data-bbox="853 1177 1093 1396"> <p>Введите номер точки отправления</p> </td> <td data-bbox="1093 1177 1691 1396"> <p>Планируемый номер вершины отправления в графе — индекс вершины графа</p> </td> </tr> </table>	<p>Количество базовых моделей</p>	<p>Указывается число последовательных итераций по оптимизации модели, то есть количество попыток уменьшить ошибку.</p>	<p>Введите номер точки отправления</p>	<p>Планируемый номер вершины отправления в графе — индекс вершины графа</p>	<p>Изображение «Кратчайшие пути в графе»</p>
<p>Количество базовых моделей</p>	<p>Указывается число последовательных итераций по оптимизации модели, то есть количество попыток уменьшить ошибку.</p>						
<p>Введите номер точки отправления</p>	<p>Планируемый номер вершины отправления в графе — индекс вершины графа</p>						

		Введите номер точки прибытия	Планируемый номер вершины прибытия в графе — индекс вершины графа		
--	--	---	---	--	--

Приложение 2. Содержимое файла с погодными условиями (пример табличных данных для подачи в блок-схему)

datetime	area	T	P	U	Ff	Td	RRR	DD_Ветер, дуящий с востока	DD_Ветер, дуящий с востоко- северо- востока	DD_Ветер, дуящий с востоко- юго- востока	DD_Ветер, дуящий с запада	DD_Ветер, дуящий с западо- северо- запада
2018-02-01	0.0	2.8625000000000003	767.675	72.625	2.0	-1.925	0.0	0.0	0.0	0.0	0.25	0.0
2018-02-02	0.0	5.75	765.4625	73.375	2.375	1.0	0.0	0.0	0.0	0.375	0.0	0.0
DD_Ветер, дуящий с западо- юго- запада	DD_Ветер, дуящий с севера	DD_Ветер, дуящий с северо- востока	DD_Ветер, дуящий с северо- запада	DD_Ветер, дуящий с северо- северо- востока	DD_Ветер, дуящий с северо- северо- запада	DD_Ветер, дуящий с юга	DD_Ветер, дуящий с юго- востока	DD_Ветер, дуящий с юго- запада	DD_Ветер, дуящий с юго-юго- востока	DD_Ветер, дуящий с юго-юго- запада		
0.375	0.0	0.0	0.0	0.0	0.0	0.125	0.125	0.125	0.0	0.0		
0.0	0.0	0.0	0.0	0.0	0.0	0.125	0.0	0.0	0.125	0.375		
DD_Штиль, безветрие	N_10% или менее, но не 0	N_100%.	N_20- 30%.	N_40%.	N_60%.	N_70 – 80%.	N_90 или более, но не 100%	N_Небо не видно из-за тумана и/или других метеорологических явлений.	N_Облаков нет.	day	month	thermo_area
0.0	0.0	0.125	0.375	0.125	0.125	0.25	0.0	0.0	0.0	1	2	0.0
0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.125	0.0	0.375	2	2	0.0

Таблица 18.3 – Обучение модели классификации изображений

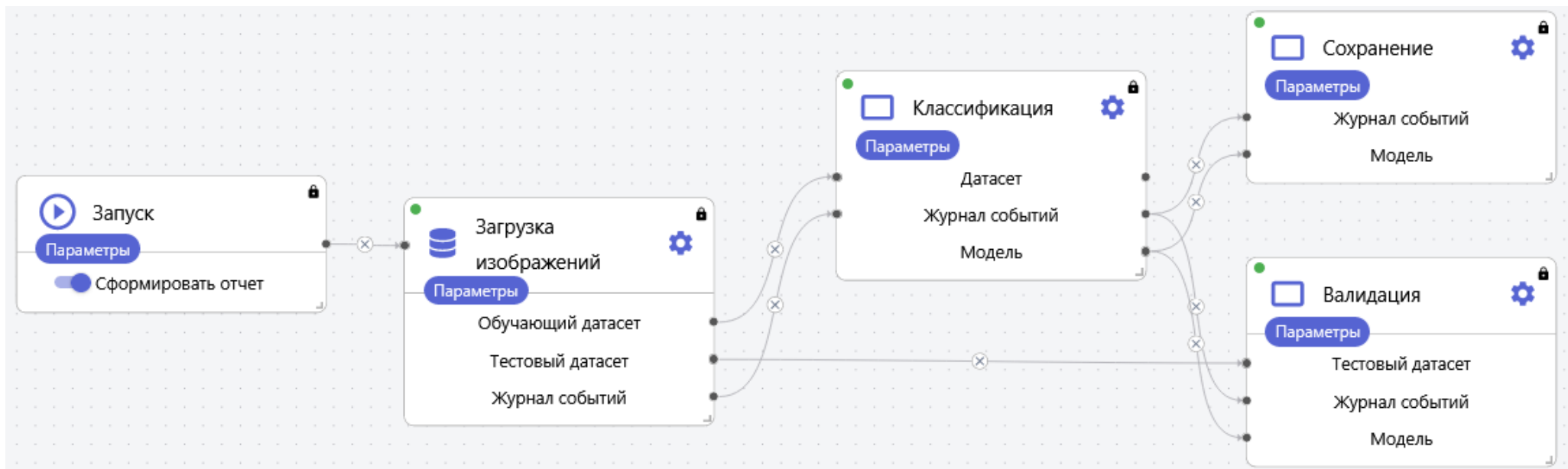


Таблица 18.4 – Обучение модели прогнозирования температуры воды и газов в котле

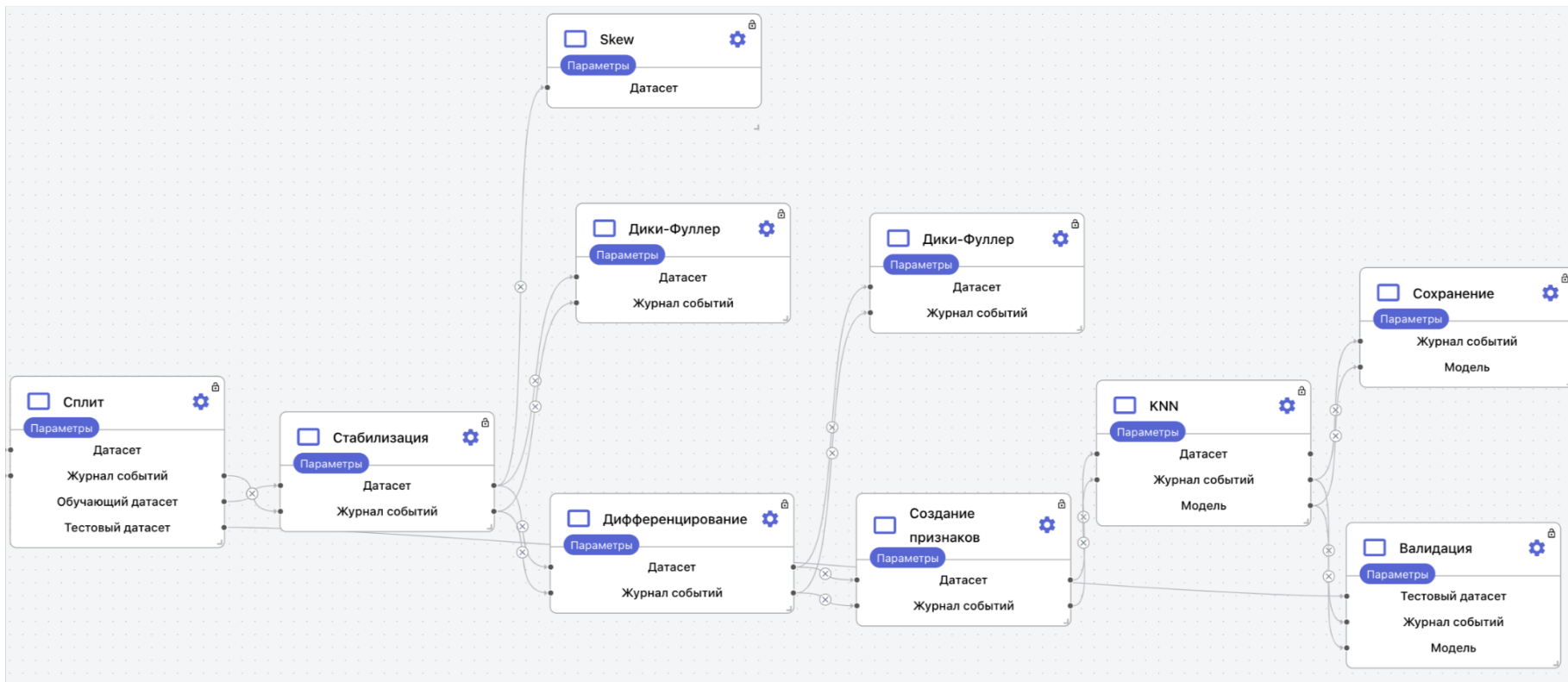
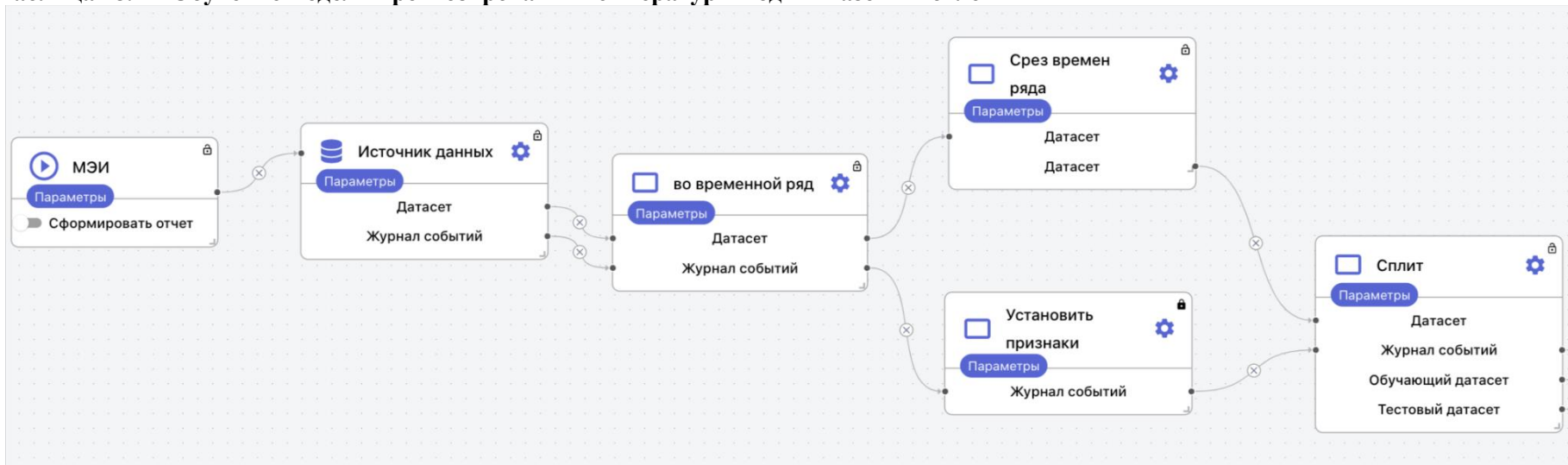


Таблица 18.5 – Пайплайн, обрабатывающий и прогнозирующие данные в режиме реального времени

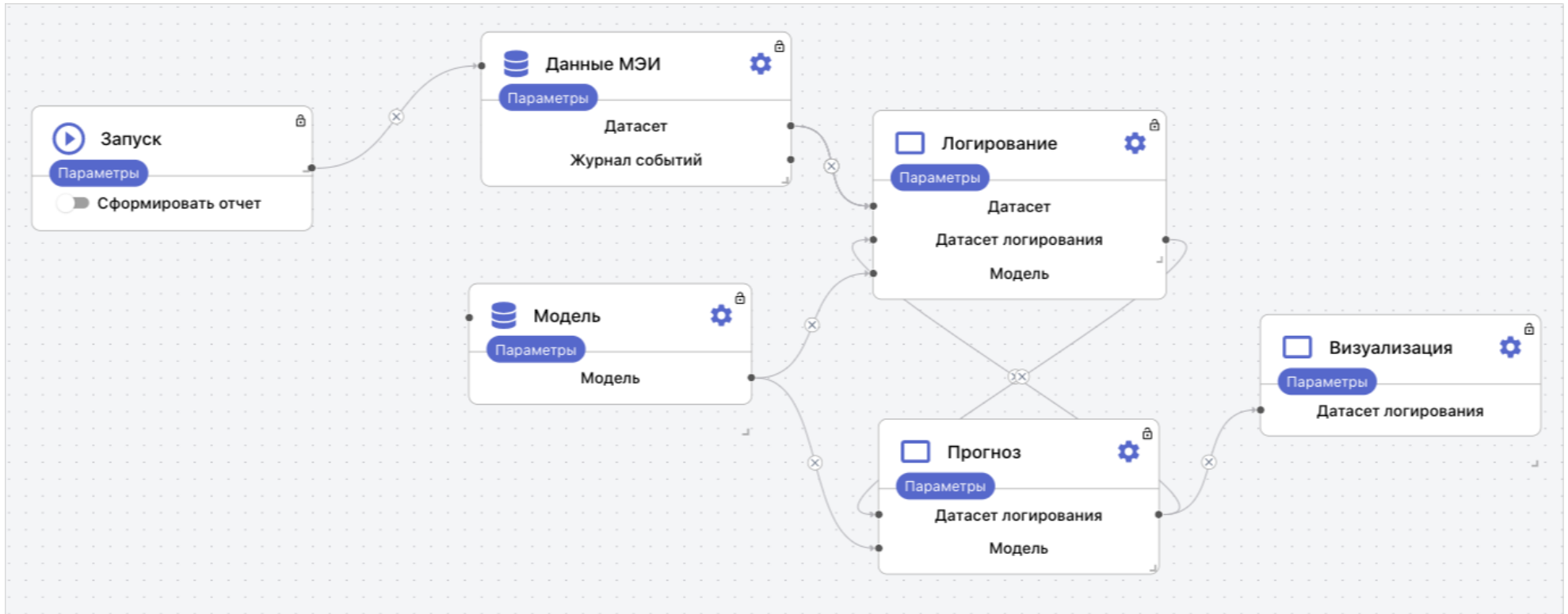


Таблица 18.6 – Обучение модели классификации текстов

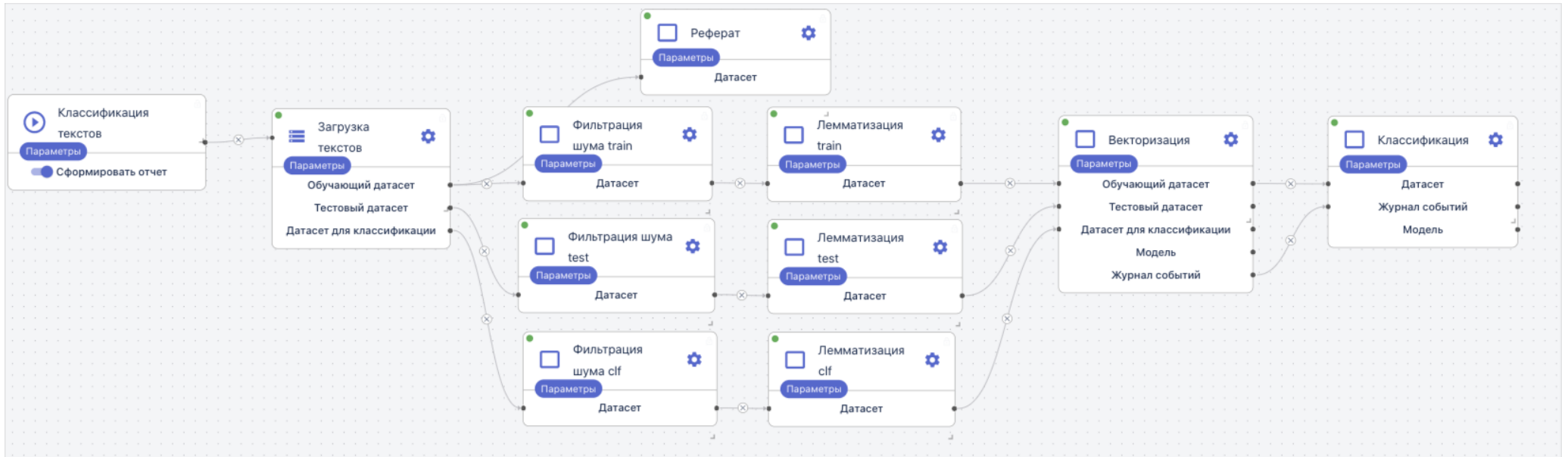
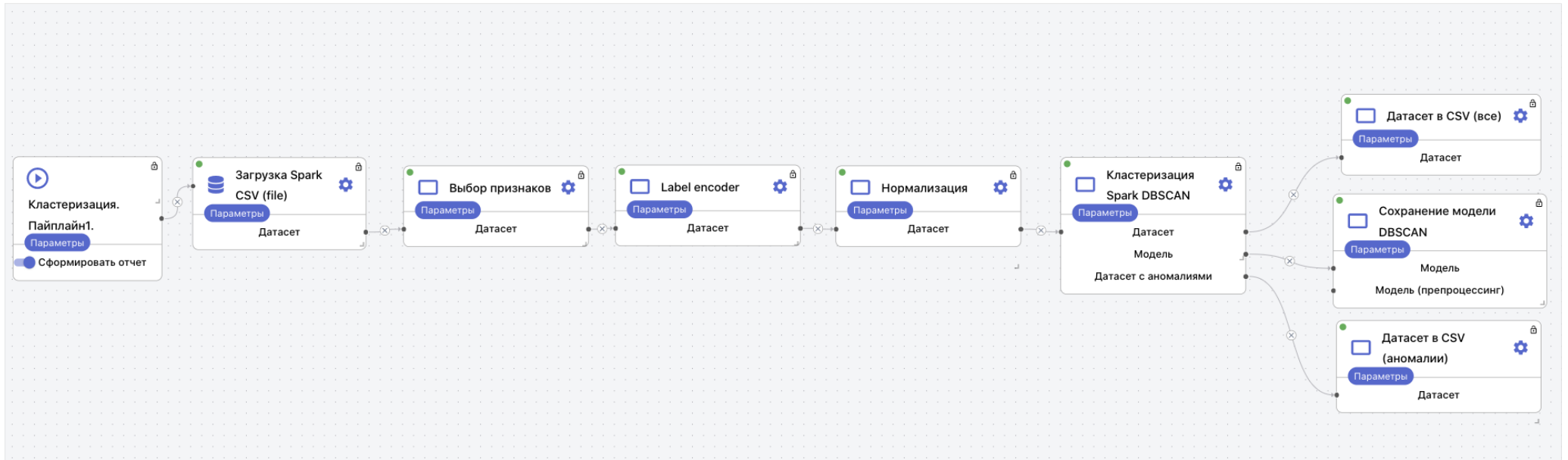


Таблица 18.7 – Обучение модели классификации Spark



Лист изменений

Таблица 19.1 – Лист изменений в версии 1.0.0

Наименование раздела	Содержание изменения	Обоснование
-	-	-