скала^р

Машина больших данных Скала^р МБД.Х

Программно-аппаратный комплекс для обработки больших данных с применением Arenadata Hyperwave (ADH 4)

Технический обзор

версия 3.2 от 01.07.2026

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Уведомление

Документ носит исключительно информационный характер и является актуальным на дату размещения.

Технические характеристики, указанные в документе, носят исключительно справочный характер и не могут служить основанием для претензий.

Технические характеристики произведенных ПАК могут отличаться от приведенных в настоящем документе вследствие модификации ПАК.

Технические характеристики и комплектация ПАК могут быть изменены производителем без уведомления.

Документ не является публичной офертой и не содержит каких-либо обязательств ООО «СКАЛА-Р».

Описание документа

Документ раскрывает, как оптимизированные программно-аппаратные комплексы отвечают современным вызовам, и дает концептуальный и архитектурный обзоры **Машины больших данных Скала^р МБД.Х** (далее **Скала^р МБД.Х**, **Машина**).

Аудитория

Технический обзор предназначен для партнеров **Скала^р** и заказчиков, перед которыми ставятся задачи разработки решения, закупки, управления или эксплуатации **Машины больших данных Скала^р МБД.Х**.

Обратная связь

Скала^р и авторы этого документа открыты для дискуссии и будут благодарны за любые конструктивные замечания. Ваши отзывы и предложения просим направлять по электронной почте на адрес MBD8@skala-r.ru, обязательно указав в теме письма название документа.

ПЕРЕЧЕНЬ ТЕРМИНОВ И СОКРАЩЕНИЙ

Термин, сокращение	Определение
ACID	(англ. atomicity, consistency, isolation, durability) Набор требований к транзакционной системе, обеспечивающий наиболее надежную и предсказуемую ее работу – атомарность, согласованность, изоляцию, устойчивость
Data Lake	«Озеро данных» – централизованное хранилище, в котором можно сохранять структурированные, полуструктурированные и неструктурированные данные в исходном виде
Data Lakehouse	Гибридная архитектура, которая сочетает масштабируемость, гибкость и низкую стоимость хранения данных любого типа, характерные для Data Lake («озера данных»), со структурированностью, надежностью транзакций (ACID) и мощными средствами управления данными, присущими Data Warehouse («хранилищу данных»). Позволяет работать со структурированными и неструктурированными данными в единой системе, поддерживая как BI-аналитику, так и машинное обучение
Data Warehouse	«Хранилище данных» – это предметно-ориентированная, интегрированная, неизменяемая и поддерживающая хронологию совокупность данных, предназначенная для поддержки принятия управленческих решений, выполнения сложных аналитических запросов, агрегирующих большие объемы исторических данных
Erasure coding	Механизм обеспечения отказоустойчивости и сохранности данных, при котором данные разбиваются на фрагменты, кодируются с добавлением избыточной информации и распределяются по различным узлам хранения, что позволяет восстановить исходные данные при потере части фрагментов, обеспечивая более высокую надежность при меньших затратах на хранение по сравнению с репликацией
ETL	(англ. Extract, Transform, Load) Процесс транспортировки данных, при котором информацию из разных мест преобразуют и кладут в новое место
JDBC/ODBC-интерфейс	Интерфейсы программирования приложений (API), которые обеспечивают универсальное взаимодействие приложений (Java – через JDBC, другие языки – через ODBC) с реляционными базами данных, позволяя выполнять SQL-запросы независимо от конкретной СУБД
MLAG	(англ. Multi-Switch Link Aggregation) Технология агрегации каналов, позволяющая одному или нескольким линкам с двух разных сетевых узлов быть объединенными вместе таким образом, что для конечного устройства это выглядит как одиночное соединение
MPP	(англ. Massively Parallel Processing) Архитектура параллельных вычислительных систем, состоящих из множества независимых

Термин, сокращение	Определение
	узлов (серверов), каждый из которых имеет собственную оперативную память, процессоры и диски. Обработка данных в такой системе (например, в аналитических базах данных) происходит путем одновременного выполнения операций на всех узлах, что позволяет очень быстро обрабатывать огромные объемы информации
NFS	(англ. Network File System) Протокол сетевого доступа к файловым системам
NL-SAS	Низкооборотный SAS – тип жестких дисков, в котором сочетаются энергоэффективные и более дешевые магнитные пластины от Nearline-дисков (используемых в больших массивах хранения) с высокопроизводительной и надежной SAS-электроникой. Оптимальное решение для создания экономичных систем хранения больших объемов «холодных» и редко используемых данных, где важны баланс между стоимостью, емкостью и надежностью
ODS	(англ. Operational Data Store) Предметно-ориентированная база данных, предназначенная для интеграции и хранения актуальных операционных данных из различных источников в почти реальном времени. Служит для выполнения оперативных запросов и отчетности, занимая промежуточное положение между транзакционными системами и централизованным хранилищем данных (Data Warehouse)
RAID	(англ. Redundant Array of Independent Disks) Избыточный массив независимых дисков, технология виртуализации данных для объединения нескольких физических дисковых устройств в логический модуль для повышения отказоустойчивости и/или производительности
UPSERT	Операция в базах данных, которая объединяет функциональность обновлений и вставок: позволяет вставить новую запись, если она не существует, или обновить существующую запись, если она уже существует
Ad-hoc-анализ	Гибкий, разовый, исследовательский метод анализа данных, направленный на решение конкретных, непредвиденных задач или поиск ответов на специфические вопросы, которые не предусмотрены в стандартных отчетах
Бакетирование	Метод организации данных внутри партиции (или всей таблицы) для оптимизации запросов, особенно слияний и выборок. Данные распределяются по фиксированному количеству "бакетов" на основе хэш-функции от значений одного или нескольких столбцов. В отличие от партиционирования, которое создает отдельные папки (директории) для разных диапазонов и значений данных, бакетирование группирует данные в файлы внутри одной партиции, что позволяет эффективно отсекать ненужные данные при сканировании и ускорять операции соединения таблиц

Термин, сокращение	Определение
Кластер	Отказоустойчивая архитектура функционала Машины
Машина	Набор аппаратного и программного обеспечения в виде Модулей Скала^р, соединенных вместе для обеспечения определенного метода обработки данных или предоставления ИТ-сервиса с заданными характеристиками. Зарегистрирована в ЕРРРП
Мультитенантность	Предоставление изолированного доступа к общим ресурсам разным арендаторам, то есть тенантам. Это свойство программного обеспечения, которое позволяет нескольким пользователям или организациям использовать одну и ту же программную систему или приложение. Это означает, что каждый пользователь имеет свой собственный набор данных и настроек, но все они работают на одной и той же программной платформе. Это позволяет экономить ресурсы и упрощает управление системой, так как администратор может управлять всеми пользователями и их данными в единой системе
Партиционирование	Разбиение больших таблиц на более мелкие части для ускорения запросов и оптимизации работы базы данных
Шардирование	Разделение баз данных на независимые сегменты, каждый из которых управляется экземпляром СУБД на отдельном узле, обеспечивающее распределенные вычисления. Позволяет улучшить масштабирование и повысить отзывчивость сервиса. Большая база разделяется на несколько частей, называемых шардами, распределяемых по другим узлам и связываемых в одну систему

1. ВВЕДЕНИЕ

Машина больших данных Скала^р МБД.Х – это программно-аппаратный комплекс, предназначенный для создания централизованного и безопасного хранилища данных (Data Lake), консолидирующего информацию любой структуры (структурированную, слабоструктурированную, неструктурированную) в объемах от терабайт до петабайт. ПАК – сертифицированный и готовый к промышленной эксплуатации продукт, построенный на базе отечественной платформы Arenadata Hyperwave (ADH; ранее – Arenadata Hadoop). Архитектура не имеет жестких межсервисных зависимостей, что позволяет развертывать только необходимые компоненты в соответствии с задачами заказчика.

Решение востребовано в крупных организациях, предъявляющих высокие требования к надежности, производительности, безопасности данных и необходимости их комплексного анализа.

Скала^р МБД.Х обеспечивает снижение совокупной стоимости владения (TCO) за счет глубокой интеграции оптимизированного аппаратного обеспечения с программной платформой. Тесная взаимная адаптация компонентов, таких как вычислительные узлы, высокоскоростная сеть 25/100 Гбит/с, а также накопители NVMe SSD и HDD, позволяет не только сократить совокупные расходы, но и достичь высокой производительности.

Ключевые экономические преимущества

- ***Снижение капитальных затрат***: глубокая интеграция исключает необходимость в избыточном запасе по производительности и дополнительном оборудовании для обеспечения совместимости
- ***Сокращение эксплуатационных расходов***: предварительная оптимизация и проверка совместимости всех компонентов значительно упрощают процесс развертывания, администрирования и обслуживания, экономя время и ресурсы специалистов заказчика
- ***Оптимизация затрат на хранение***: использование разных типов накопителей для «горячих» и «холодных» данных, а также интеллектуальное управление их жизненным циклом снижают потребление электроэнергии и нагрузку на инфраструктуру, делая хранение больших объемов данных более экономичным
- ***Экономичная масштабируемость***: модульная архитектура позволяет наращивать вычислительную мощность и объемы хранения поэтапно, добавляя только необходимые модули, без дорогостоящей полной замены оборудования

В результате достигается высокая производительность как для пакетной обработки данных с помощью инструментов Spark и MapReduce, так и для интерактивной аналитики с использованием SQL-движка Impala и интерактивных блокнотов Zeppelin, при оптимизированных совокупных затратах на протяжении всего жизненного цикла **Машины**.

Ключевые технологические преимущества

- ***Единая платформа для всех данных***: консолидация сырых данных из различных источников (ERP, CRM, логи, датчики IoT) в распределенной файловой системе HDFS с возможностью их последующего преобразования, анализа и предоставления через единые интерфейсы (Hive, Impala)
- ***Готовая аналитическая среда***: возможность выполнения сложных запросов к петабайтам данных за секунды (Impala), исследовательского анализа и визуализации (Zeppelin), а также построения воспроизводимых рабочих процессов обработки данных и их интеграции с внешними системами (Sqoop)

- ***Корпоративная безопасность и управление доступом***: встроенные механизмы защиты информации, включая централизованное управление политиками доступа на уровне данных (Ranger), безопасный шлюз для внешних подключений (Knox) и строгую аутентификацию (Kerberos). Это позволяет организовать безопасную мультитенантную работу различных команд с изолированными наборами данных в рамках одного кластера
- ***Промышленная надежность и управляемость***: архитектура обеспечивает отказоустойчивость на всех уровнях – от аппаратного резервирования компонентов до автоматического восстановления данных и сервисов в кластере. Управление и мониторинг жизненного цикла комплекса осуществляются через единые отечественные панели **Скала^р Геном** и **Скала^р Визион**, а также средства управления платформой Arenadata (ADCM)

Программно-аппаратные комплексы **Машины больших данных Скала^р МБД.Х** включены в Единый реестр российской радиоэлектронной продукции и работают на ПО, включенном в реестр Минцифры РФ.

2. ОТЛИЧИТЕЛЬНЫЕ ЧЕРТЫ

Машины Скала^р спроектированы для обеспечения высокой производительности и отказоустойчивости в условиях интенсивных рабочих нагрузок. Архитектура с глубокой интеграцией компонентов позволяет поддерживать согласованные показатели задержки и доступности. Глубокая интеграция и встречная оптимизация компонентов от платформенного ПО до микроконтроллеров в тесном технологическом сотрудничестве с их разработчиками обеспечивают высочайшую устойчивость создаваемых **Машин**.

- Продукты **Скала^р** являются серийно выпускаемыми преднастроенными **Машинами** и позволяют осуществлять быстрое развертывание и ввод в эксплуатацию
- Модульный принцип обеспечивает интеграцию разнородных компонентов ИТ-инфраструктуры в единую платформу предприятий, корпораций и ведомств
- Использование унифицированных узлов позволяет переопределять их роли между различными модулями и гибко перестраивать инфраструктуру при изменении нагрузки
- Единые поддержка и сервисное обслуживание для всех продуктов линейки **Скала^р** от производителя обеспечивают оперативное разрешение инцидентов на стыке технологий
- Безусловное соблюдение принципов отказоустойчивости и масштабируемости на уровне архитектуры соответствует требованиям для использования в критичных и высоконагруженных корпоративных и государственных информационных системах
- Применение **Машин Скала^р** позволяет обеспечить снижение совокупной стоимости владения (ТСО)
- **Машины** обладают широкими возможностями масштабируемости, способны хранить и обрабатывать терабайты данных
- Механизмы распределенной обработки и хранения данных позволяют распределять запросы к данным одновременно на все узлы системы для быстрой подготовки результатов
- Единая программная платформа **Скала^р** обеспечивает управление эксплуатацией и позволяет мониторить работу компонентов **Машины**, а инструменты от компании Arenadata позволяют управлять параметрами СУБД, сервисами и приложениями, входящими в состав ПАК

3. ПОДТВЕРЖДЕННАЯ БЕЗОПАСНОСТЬ

3.1 Операционные системы

Машина больших данных Скала^р МБД.Х поставляется с одной из сертифицированных операционных систем:

- **ОС Альт 8 СП** (сертификат ФСТЭК №3866 от 10.08.2018, действует до 10.08.2028)
- **РЕД ОС** (сертификат ФСТЭК №4060 от 12.01.2019, действует до 12.01.2029)
- **Astra Linux Special Edition** (сертификат ФСТЭК №2557 от 27.01.2012, действует до 27.01.2031)

ОС Альт 8 СП

В соответствии с нормативным правовым актом «Требования безопасности информации к операционным системам» (ФСТЭК России, 2016), ОС предназначена для:

- обеспечения 1, 2, 3 класса защищенности информации, не составляющей государственную тайну, содержащейся в государственных информационных системах
- обеспечения 1, 2, 3 класса защищенности автоматизированной системы управления
- обеспечения безопасности значимых объектов критической информационной инфраструктуры 1, 2, 3 категории значимости
- применения в информационных системах общего пользования 2 класса
- обеспечения выполнения программ в защищенной среде

ОС соответствует требованиям следующих нормативных документов:

- «Требования безопасности информации к операционным системам» (ФСТЭК России, 2016) и «Профиль защиты операционных систем типа А четвертого класса защиты. ИТ.ОС.А4.ПЗ» (ФСТЭК России, 2017) по 4 классу защиты
- «Требования по безопасности информации к средствам контейнеризации» (ФСТЭК России, 2022) по 4 классу защиты
- «Требования по безопасности информации к средствам виртуализации» (ФСТЭК России, 2022) по 4 классу защиты
- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) по 4 уровню доверия
- «Требования по безопасности информации к системам управления базами данных» (ФСТЭК России, 2023) по 4 классу защиты

РЕД ОС

Может применяться для защиты информации в:

- значимых объектах критической информационной инфраструктуры 1 категории
- государственных информационных системах 1 класса защищенности
- автоматизированных системах управления производственными и технологическими процессами 1 класса защищенности
- информационных системах персональных данных при необходимости обеспечения 1 уровня защищенности персональных данных

- информационных системах общего пользования 2 класса

ОС соответствует требованиям следующих нормативных документов:

- «Требования безопасности информации к операционным системам» (ФСТЭК России, 2016) и «Профиль защиты операционных систем типа А четвертого класса защиты. ИТ.ОС.А4.ПЗ» (ФСТЭК России, 2017) по 4 классу защиты
- «Требования по безопасности информации к средствам контейнеризации» (ФСТЭК России, 2022) по 4 классу защиты
- «Требования по безопасности информации к средствам виртуализации» (ФСТЭК России, 2022) по 4 классу защиты
- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) по 4 уровню доверия

Astra Linux Special Edition

Может применяться для защиты информации в:

- информационных системах, не предназначенных для обработки сведений государственной тайны
- государственных информационных системах 1 класса защищенности
- автоматизированных системах управления производственными и технологическими процессами 1 класса защищенности
- информационных системах персональных данных при необходимости обеспечения 1 уровня защищенности персональных данных
- информационных системах общего пользования 2 класса

ОС соответствует требованиям следующих нормативных документов:

- «Требования безопасности информации к операционным системам» (ФСТЭК России, 2016), «Профиль защиты операционных систем типа А первого класса защиты. ИТ.ОС.А1.ПЗ» (ФСТЭК России, 2017) по 1 классу защиты и «Профиль защиты ОС типа А второго класса защиты. ИТ.ОС.А2.ПЗ» по 2 классу защиты
- «Требования по безопасности информации к средствам контейнеризации» (ФСТЭК России, 2022) по 1 классу защиты
- «Требования по безопасности информации к средствам виртуализации» (ФСТЭК России, 2022) по 1 классу защиты
- «Требования по безопасности информации к системам управления базами данных» (ФСТЭК России, 2023) по 1 классу защиты
- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) по 1 уровню доверия

3.2 Платформа

Машина больших данных Скала^р МБД.Х использует сертифицированную платформу **Arenadata Hyperwave** (Сертификат ФСТЭК №4821 от 13.06.2024, действует до 13.06.2029).

Платформа может применяться для защиты информации в:

- государственных информационных системах 1 класса защищенности

- информационных системах персональных данных 1 уровня защищенности
- значимых объектах критической информационной инфраструктуры 1 категории
- автоматизированных системах управления производственными и технологическими процессами 1 класса защищенности
- информационных системах 2 класса общего пользования

Платформа соответствует требованиям нормативного документа «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020, приказ № 76) − по 4 уровню доверия.

3.3 Средства защиты

Обеспечивается полная совместимость **Машины больших данных Скала^р МБД.Х** с наложенными средствами защиты:

- антивирусное средство защиты **Kaspersky Endpoint Security для Linux** (сертификат ФСТЭК №2534 от 27.12.2011, действует до 27.12.2030)
- средство доверенной загрузки ПАК **«Соболь» версия 4** (сертификат ФСТЭК №4043 от 05.12.2018, действует до 05.12.2028)
- система единой аутентификации **Avanpost FAM** (сертификат ФСТЭК №4492 от 13.12.2021, действует до 13.12.2026)

Kaspersky Endpoint Security для Linux

ПО может применяться для защиты информации в:

- государственных информационных системах 1 класса защищенности
- информационных системах персональных данных при необходимости обеспечения 1 уровня защищенности персональных данных
- значимых объектах критической информационной инфраструктуры 1 категории
- автоматизированных системах управления производственными и технологическими процессами 1 класса защищенности
- информационных системах общего пользования 2 класса

ПО соответствует требованиям следующих нормативных документов:

- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) – по 2 уровню доверия
- «Требования к средствам антивирусной защиты» (ФСТЭК России, 2012)
- «Профиль защиты средств антивирусной защиты типа Б 2 класса защиты. ИТ.САВ3.Б2.13» (ФСТЭК России, 2012)
- «Профиль защиты средств антивирусной защиты типа В второго класса защиты. ИТ.САВ3.В2.П3» (ФСТЭК России, 2012)
- «Профиль защиты средств антивирусной защиты типа Г второго класса защиты. ИТ.САВ3.Г2.П3.»
- «Требования к средствам контроля сменных машинных носителей информации» (ФСТЭК России, 2014)
- «Профиль защиты средств контроля подключения сменных машинных носителей информации второго класса защиты. ИТ.СКН.П2.П3» (ФСТЭК России, 2014)

«Соболь» версия 4

ПО может применяться для защиты информации в:

- государственных информационных системах 1 класса защищенности
- информационных системах персональных данных при необходимости обеспечения 1 уровня защищенности персональных данных
- значимых объектах критической информационной инфраструктуры 1 категории
- автоматизированных системах управления производственными и технологическими процессами 1 класса защищенности
- информационных системах общего пользования 2 класса

ПО соответствует требованиям следующих нормативных документов:

- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) – по 2 уровню доверия
- «Требования к средствам доверенной загрузки» (ФСТЭК России, 2013)
- «Профиль защиты средства доверенной загрузки уровня базовой системы ввода-вывода второго класса защиты. ИТ.СДЗ.УБ2.ПЗ» (ФСТЭК России, 2013)

Avanpost FAM

ПО может применяться для защиты информации в:

- государственных информационных системах 1 класса защищенности
- информационных системах персональных данных при необходимости обеспечения 1 и 2 уровня защищенности персональных данных
- значимых объектах критической информационной инфраструктуры 1 категории
- автоматизированных системах управления производственными и технологическими процессами 1 класса защищенности

ПО соответствует требованиям следующих нормативных документов:

- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) – по 4 уровню доверия

4. ПРИНЦИПЫ СОЗДАНИЯ МАШИНЫ

Для обеспечения максимальной надежности и производительности при проектировании **Машины больших данных Скала^р МБД.Х** были заложены принципы, соответствующие лучшим инженерным практикам мирового уровня.

Технологические принципы

- ***Резервирование критических компонентов***: ключевые элементы **Машины**, включая управляющие сервисы, сетевые пути и точки питания, имеют резервные копии. Это гарантирует, что отказ одного элемента не приводит к остановке работы всей **Машины**
- ***Отказоустойчивость на основе модульной архитектуры***: **Машина** построена из функционально законченных и типизированных Модулей. Четкое распределение ролей и стратегическое дублирование функций между Модулями позволяет сохранять предсказуемый уровень производительности и доступного объема хранения, а также продолжать предоставлять сервис, даже если отдельный Модуль или его части выйдут из строя
- ***Производительность за счет оптимизированных компонентов***: используются только полностью совместимые и высокопроизводительные отечественные компоненты, прошедшие всестороннюю проверку и отладку в условиях реальных нагрузок, что обеспечивает максимальную скорость обработки данных
- ***Горизонтальная масштабируемость***: при увеличении объема данных или вычислительной нагрузки система масштабируется путем линейного добавления новых узлов и Модулей, без необходимости полной замены оборудования или сложной реконфигурации

Технические решения

- Конфигурация **Машины** гибко настраивается под конкретные потребности бизнеса за счет использования стандартизированных Модулей, предназначенных для четко определенных функций. Управление и наблюдение за всеми компонентами осуществляется через собственное специализированное ПО **Скала^р**
- Каждый узел и компонент проходит многоуровневые проверки еще на этапе производства, что гарантирует надежность всей **Машины** при вводе в эксплуатацию
- Все аппаратные и программные компоненты глубоко адаптированы друг к другу, что исключает проблемы совместимости и повышает общую эффективность решения

Спрогнозированная нагрузка

- Производительность можно выбирать, подбирая нужное количество **Модулей**
- Все узлы взаимодействуют между собой с одинаковой скоростью

Выделенная сеть внутреннего взаимодействия

- Высокоскоростная и отказоустойчивая сеть внутреннего взаимодействия обеспечивает высокоскоростной обмен данными между узлами
- Параллельная обработка запросов на узлах позволяет суммировать вычислительные мощности всех узлов, участвующих в обработке запроса

4.1 Компоненты ADH для хранения данных

4.1.1 HDFS

Файловая система, предназначенная для хранения файлов больших размеров, поблочно распределенных между узлами вычислительного кластера. Каждый узел данных взаимодействует только со своими внутренними накопителями. Целостность данных в HDFS обеспечивается встроенными механизмами файловой системы. Во время записи для каждого блока данных вычисляется контрольная сумма, которая сохраняется вместе с данными. При последующем чтении происходит проверка контрольных сумм. В случае обнаружения повреждения блока данных система автоматически инициирует его восстановление из соответствующей реплики. Этот процесс осуществляется без участия прикладных приложений.

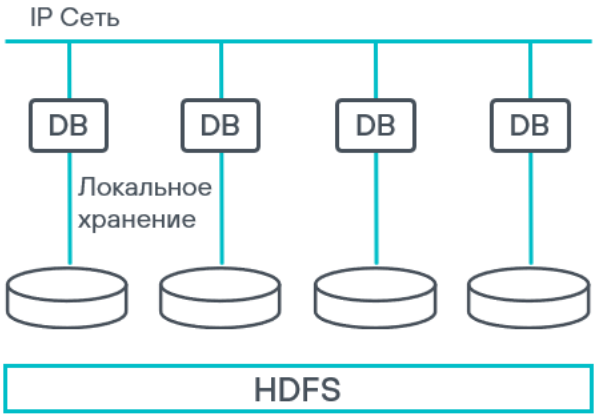

Рисунок 1. Архитектура без общих ресурсов (shared nothing)

Рисунки 1 и 2 отображают архитектуру HDFS. Стойки – логическое объединение узлов данных согласно сетевой топологии кластера.

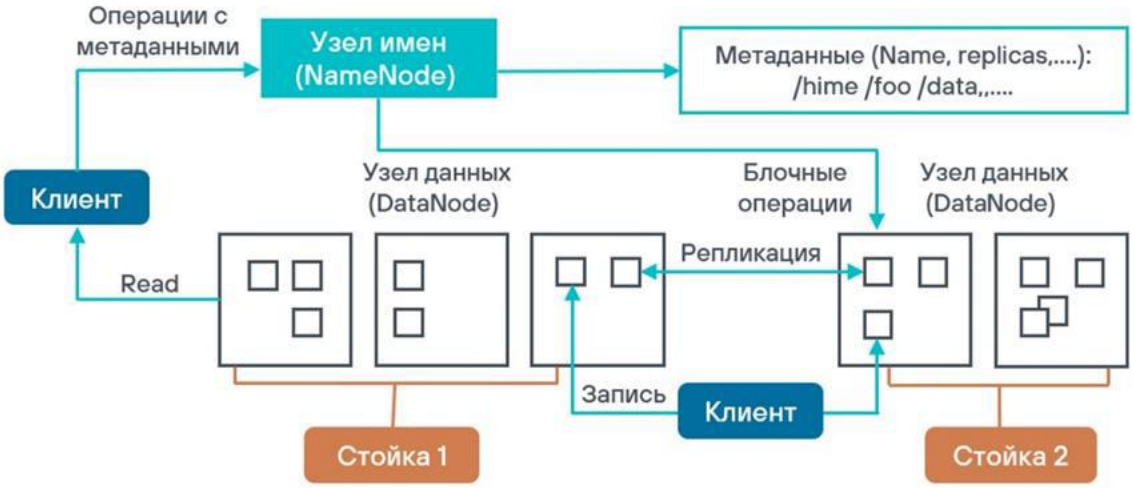

Рисунок 2. Архитектура HDFS

Репликация данных

Репликация позволяет гарантировать защиту от одиночных сбоев. Схема репликации показана ниже (Рисунок 3). Принцип работы репликации:

- узел имен (NameNode) выбирает новые узлы данных для размещения реплик
- сервер имен выполняет балансировку размещения данных по узлам и составляет список узлов для репликации

- стратегия размещения копий данных в кластере строится на принципе распределения копий по разным аппаратным стойкам для защиты от их полного отказа. Первая копия блока данных создается на сервере-источнике записи или, при внешней записи, на случайно выбранном сервере. Вторая копия обязательно размещается на сервере в другой стойке того же дата-центра. Третья копия направляется на сервер в стойке, отличной от двух предыдущих

Критически важным условием для работы этой стратегии является предварительная настройка осведомленности кластера о физической топологии размещения оборудования заказчика. Информация о принадлежности каждого сервера к конкретной стойке и центру обработки данных позволяет системе автоматически выполнять правила размещения, обеспечивая тем самым высочайший уровень отказоустойчивости и эффективное использование сетевой инфраструктуры.

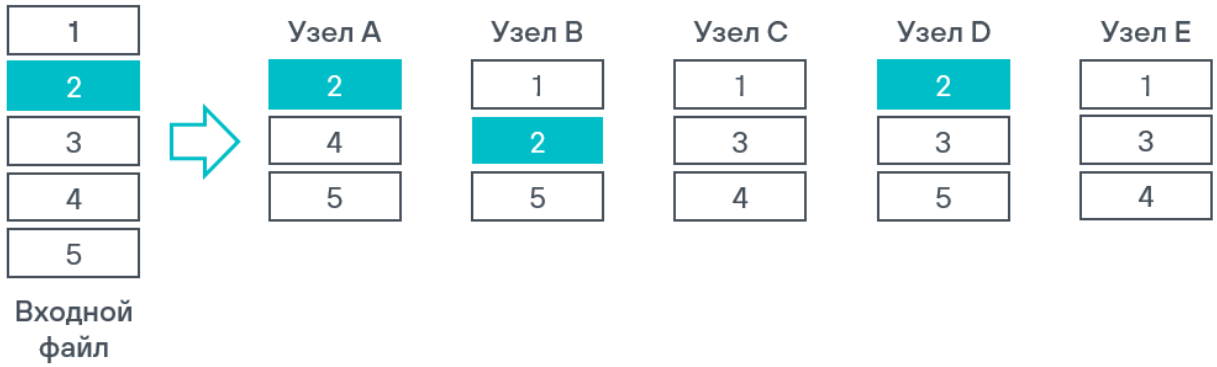

Рисунок 3. Схема распределения данных на HDFS

4.1.2 Iceberg

Это открытый формат таблиц для построения архитектуры Data Lakehouse, который добавляет в озеро данных поддержку ACID-транзакций, возможность выполнения запросов к историческим состояниям данных (путешествие во времени) и безопасную эволюцию схемы таблиц без необходимости переписывать данные. Iceberg решает проблемы отслеживания файлов и жесткого секционирования, которые характерны для традиционных таблиц в Hive. Управляет метаданными и файлами данных, обеспечивая надежность и согласованность. Применяется для построения масштабируемых «озер данных» (Data Lake) с поддержкой операций UPSERT и «временных путешествий» (time travel).

Ключевые свойства

- ACID-семантика
- Снимки (snapshots) и доступ к историческим версиям данных в определенные моменты времени (travel back in time)
- Скрытие партиционирования от пользователя
- Высокая производительность при чтении за счет продвинутых метаданных

4.1.3 HBase

Распределенная, масштабируемая NoSQL база данных, работающая поверх HDFS. Идеально подходит для приложений, требующих быстрого доступа к большим объемам слабоструктурированных данных.

Ключевые свойства

- Хранение данных в виде таблиц с возможностью произвольного доступа по ключу
- Строгая согласованность в пределах одной строки
- Автоматическое шардирование и балансировка нагрузки
- Горизонтальная масштабируемость за счет добавления узлов
- Поддержка версионности данных и временных меток
- Интеграция с Hadoop-экосистемой (MapReduce, Spark)
- Возможность операций чтения/записи в режиме, близком к реальному времени

4.1.4 Ozone

Ozone – это масштабируемое, распределенное и отказоустойчивое объектное хранилище для Hadoop, предназначенное для управления миллиардами объектов данных. Оно дополняет HDFS, предоставляя поддержку объектной парадигмы хранения данных вместе с совместимостью с файловой системой, что делает его идеальным решением для современных "озер данных".

Ключевые свойства

- В отличие от файловой системы HDFS, Ozone разработан для эффективной работы с большим количеством небольших файлов
- Хорошо интегрируется с сервисами Hadoop, такими как Hive, Spark, MapReduce, Kerberos и фреймворком Ranger
- Поддерживает работу с удаленными объектными хранилищами через API Amazon S3
- Может работать с контейнерными средами, такими как Kubernetes и YARN

4.2 Компоненты ADH для работы с данными

4.2.1 Hue

Веб-интерфейс для взаимодействия с кластерами данных Hadoop. Широко используется как стандартный интерфейс во многих дистрибутивах Hadoop (Cloudera, Hortonworks) для демократизации доступа к данным. Позволяет аналитикам и инженерам работать с данными через браузер без использования командной строки, редактировать и выполнять SQL-запросы (к Hive, Impala и др.), просматривать данные, создавать рабочие потоки (workflows) в Oozie.

Ключевые свойства

- Интуитивный интерфейс
- Поддержка множества компонентов (Hive, Impala, Pig, Spark)
- Инструменты для администрирования и отладки

4.2.2 Kyuubi

Отказоустойчивый, многопользовательский JDBC-шлюз корпоративного уровня, построенный на основе Spark. Он предоставляет единый SQL-интерфейс для выполнения аналитических запросов и поддерживает сквозную мультитенантность – изоляцию ресурсов и данных между разными пользователями и рабочими нагрузками в рамках одной платформы. По своей архитектуре является прямой заменой и эволюционным развитием

HiveServer2, но с ключевым отличием – Kyuubi спроектирован для работы в современных облачных и контейнеризированных средах.

Основная задача Kyuubi – предоставить единый интерфейс для выполнения SQL-запросов по стандартным протоколам взаимодействия с базами данных и удаленного вызова процедур. Это позволяет клиентским приложениям (от систем бизнес-аналитики до пользовательских скриптов) работать с данными, не вникая в сложности управления кластером распределенных вычислений.

Ключевые свойства

- Сквозная поддержка доступа нескольких пользователей к данным через единую систему аутентификации и авторизации
- Высокая доступность за счет балансировки нагрузки через Zookeeper
- Поддержка нескольких различных рабочих нагрузок в рамках одной платформы, одной копии данных и одного SQL-интерфейса

4.2.3 Impala

Высокопроизводительный SQL-движок для интерактивной аналитики больших данных, сочетающий скорость коммерческих СУБД с масштабируемостью Hadoop. Часто используется для реализации интерактивных панелей мониторинга и отчетности в реальном времени.

Ключевые свойства

- MPP-архитектура
- Минимальная задержка
- Полная SQL-совместимость и интеграция с экосистемой, что делает Impala идеальным для сценариев, где требуется скорость и простота доступа к данным

4.2.4 Trino

Распределенный SQL-движок для выполнения интерактивных аналитических запросов к большим наборам данных из разнородных источников (HDFS, S3, реляционные БД, NoSQL) без необходимости копирования данных. Позволяет аналитикам использовать стандартный SQL для работы с данными в Hadoop и за его пределами. Часто используется для реализации интерактивной аналитики (ad-hoc-анализа данных).

Ключевые свойства

- Выполнение запросов с низкой задержкой
- Высокая производительность на больших объемах данных
- Поддержка подключения к множеству источников данных с помощью специализированных модулей (коннекторов)
- Возможность федеративных запросов (объединение данных из разных систем)

4.2.5 Zookeeper

Надежный, высокопроизводительный сервис координации для распределенных систем, предоставляющий простой API и гарантии согласованности.

Ключевые свойства

- Отказоустойчивость
- Механизм наблюдения

- Поддержка различных типов узлов и консенсусный алгоритм делают Zookeeper фундаментальным компонентом экосистемы Hadoop

4.2.6 YARN

Универсальный фреймворк, отвечающий за распределение вычислительных ресурсов и выполнение задач обработки данных в кластере. Обеспечивает параллельную работу различных вычислительных движков (например, Spark и Flink), эффективно распределяя нагрузку и изолируя задачи друг от друга. Позволяет оптимально использовать ресурсы кластера и поддерживать стабильную работу системы при высокой нагрузке.

Ключевые свойства

- Отделение управления ресурсами от моделей выполнения, что позволяет запускать разнородные задачи (MapReduce, Spark, Flink и т.п.) в одном кластере
- Масштабируемость до десятков тысяч узлов с эффективным распределением CPU и памяти
- Поддержка мультитенантности и изоляции через Linux Containers
- Отказоустойчивость с автоматическим перезапуском контейнеров и Application Master'ов
- Гибкое планирование на основе очередей (Capacity/Fair Scheduler)
- Интеграции с системами безопасности (Kerberos)
- Мониторинг через REST API и веб-интерфейс

4.2.7 Phoenix

SQL-слой поверх HBase, предоставляющий реляционный интерфейс для работы с данными через стандартный JDBC/ODBC-интерфейс. Позволяет использовать знакомый SQL-синтаксис для выполнения аналитических запросов и взаимодействия с данными в HBase в режиме, близком к реальному времени. Он транслирует SQL-запросы в собственные вызовы HBase и поддерживает ACID-транзакции и вторичные индексы для ускорения поиска.

Ключевые свойства

- Трансляция SQL-запросов в собственные вызовы HBase
- Поддержка ACID-транзакций и вторичных индексов для ускорения поиска
- Низкая задержка при выполнении операций
- Возможность использования стандартных BI-инструментов
- Оптимизация запросов с помощью параллельного выполнения
- Полная совместимость с метаданными HBase

4.2.8 Spark

Высокопроизводительная универсальная платформа для распределенной обработки данных, обеспечивающая выполнение разнородных задач, включая извлечение, преобразование и загрузку данных, обработку потоков информации и решение задач машинного обучения. За счет использования встроенного оптимизатора для составления планов выполнения запросов и специального механизма управления вычислительными ресурсами достигается максимальная скорость обработки при работе с данными в оперативной памяти и сохраняется эффективность при их вытеснении на диск.

Ключевые свойства

- Поддержка разнородных типов задач (пакетная обработка, стриминг, машинное обучение, графовые вычисления) в единой платформе
- Богатые API на Java, Scala, Python и R
- Встроенные библиотеки (Spark SQL, MLlib, Structured Streaming, GraphX)
- Оптимизация запросов через Catalyst Optimizer и Tungsten Engine
- Возможность работы с различными источниками данных (HDFS, S3, Kafka, JDBC)
- Отказоустойчивость и масштабируемость на кластерах из тысяч узлов

4.2.9 Smart Storage Manager

Предоставляет веб-интерфейс, с помощью которого можно создавать правила, запускать действия, проверять статус их выполнения и следить за статистикой кластера. Начиная с версии ADH 4.1.0, Smart Storage Manager поддерживает имперсонацию – выполнение административных действий от имени указанного пользователя, что повышает безопасность и упрощает аудит в мультитенантных средах.

Ключевые свойства

Smart Storage Manager (SSM) добавляет в HDFS следующие возможности:

- перемещение данных между различными типами хранилищ в зависимости от существующих правил
- асинхронная репликация данных между разными кластерами с HDFS или между HDFS и облачным хранилищем. SSM отслеживает операции изменения данных, такие как создание, удаление, добавление и переименование, чтобы обеспечить синхронизацию данных в реальном времени и избежать вычислительных затрат MapReduce
- сжатие небольших файлов в один файл-контейнер, который хранится в HDFS и данные в нем доступны для приложений верхнего уровня
- гибкая настройка включения Erasure coding (EC) или репликации и управление файлами с разными политиками EC с помощью правил
- сжатие данных в HDFS без ограничения доступа к ним для внешних приложений

4.2.10 Hive

Система управления данными в Hadoop, предоставляющая SQL-подобный язык (HiveQL) для запросов к большим наборам данных. Hive, и в особенности его компонент Hive Metastore (HMS), служит центральным каталогом метаданных для таблиц в кластере. Хотя HiveQL может использоваться для пакетных ETL-задач, для интерактивных запросов рекомендуется использование Trino или Spark SQL, которые, как и другие компоненты (Impala, Flink), интегрируются с HMS для доступа к единым метаданным.

Ключевые свойства

- Трансляция HiveQL-запросов в задачи MapReduce, Tez или Spark
- Поддержка структурированных и полуструктурированных данных
- Интеграция с HDFS и Hive Metastore для управления метаданными
- Возможность создания таблиц, партиционирования и бакетирования
- Расширяемость через пользовательские функции (UDF)

4.2.11 Flink

Система для параллельной обработки информации, созданная специально для работы с непрерывными потоками данных с минимально возможной задержкой. Ее ключевой особенностью является единый подход к работе как с потоками, так и с пакетами данных, где обработка потока рассматривается как базовая операция, а пакетная обработка – как ее частный случай. В состав ADH входит Flink History Server, предоставляющий веб-интерфейс для просмотра статистики и отладки завершенных приложений Flink.

Ключевые свойства

- Построения систем, работающих в реальном времени: непрерывного наблюдения за показателями, выявления аномалий и мошеннических схем, анализа сложных цепочек событий в потоке данных и обработки транзакционных операций
- Обработка ровно один раз – система гарантирует, что каждое событие в потоке данных будет учтено и обработано точно один раз, даже в случае сбоев
- Встроенная поддержка временных меток событий – позволяет корректно анализировать данные, поступившие с задержкой, и определять правильный порядок событий на основе времени их возникновения, а не времени поступления в систему
- Обеспечение высокой пропускной способности и устойчивости к сбоям за счет механизма контрольных точек – периодического автоматического сохранения текущего состояния вычислений в надежном хранилище, что позволяет быстро восстановить работу после отказа
- Богатые возможности для сложных преобразований и работы с временными окнами – предоставляет развитые инструменты для агрегации и анализа данных за определенные периоды времени

4.2.12 Solr

Масштабируемая платформа полнотекстового поиска, построенная на Lucene, которая применяется для построения поисковых систем на веб-сайтах, в каталогах продукции и системах работы с контентом, где требуются быстрый поиск и фильтрация.

Ключевые свойства

- Мощный полнотекстовый поиск с поддержкой сложных запросов и ранжирования
- Горизонтальная масштабируемость через шардирование и репликацию индексов
- Функции фасетного поиска – механизма для категоризации и фильтрации результатов поиска по различным атрибутам (фасетам) данных, позволяющего пользователям интерактивно уточнять результаты, с возможностью выделения результатов (highlighting) и автодополнения
- Интеграция с Hadoop (загрузка данных из HDFS)
- REST API и расширяемая архитектура за счет плагинов

4.2.13 Tez

Фреймворк для выполнения высокопроизводительных задач обработки данных с использованием сложных направленных ациклических графов (DAG), который позволяет значительно ускорить выполнение ETL-операций и сложных запросов в экосистеме Hadoop без изменения кода приложений.

Ключевые свойства

- Оптимизированная замена классического MapReduce

- Уменьшение накладных расходов за счет повторного использования контейнеров и динамического планирования задач
- Ускорение выполнения запросов в Hive и Pig
- Поддержка инкрементной обработки данных
- Эффективное управление памятью и ресурсами

4.2.14 Ranger (ADPS)

Фреймворк для централизованного управления безопасностью, который обеспечивает единую точку администрирования политик доступа для различных компонентов, таких как HDFS, Hive и HBase. Ключевой особенностью Ranger является поддержка детализированного контроля доступа на уровне столбцов и строк таблиц, а также возможность динамического маскирования данных, что позволяет гибко настраивать права доступа в зависимости от роли пользователя.

Ключевые свойства

- Единая точка администрирования политик доступа к данным для компонентов (HDFS, Hive, HBase и т.п.)
- Поддержка детализированного контроля (уровень столбцов, строк, маскирование данных)
- Интеграция с LDAP/Active Directory и Kerberos
- Полный аудит всех операций доступа

4.2.15 Knox (ADPS)

Шлюз безопасности (API Gateway) для экосистемы Hadoop, обеспечивающий единую точку входа для внешних приложений и скрывающий внутреннюю топологию кластера. Позволяет безопасно предоставлять доступ к сервисам Hadoop извне периметра кластера, не нарушая его внутреннюю безопасность. Основные функции Knox – это аутентификация пользователей через корпоративные сервисы каталогов, поддержка единого входа и централизованное управление политиками доступа к REST API сервисов платформы.

Ключевые свойства

- Проксирование и агрегация REST/HTTP-API Hadoop-сервисов (HDFS, YARN, HBase и т.п.)
- Аутентификация пользователей через LDAP/Active Directory/SAML
- Поддержка Single Sign-On (SSO)
- Скрытие внутренней топологии кластера от внешнего мира
- Централизованное управление политиками доступа и аудитом

5. ВАРИАТИВНОСТЬМАШИН

Машина больших данных Скала^р МБД.Х проектируется под конкретные задачи заказчика, что позволяет оптимизировать совокупную стоимость владения (TCO). Базовые конфигурации адаптируются под преобладающий тип рабочей нагрузки: высокопроизводительные вычисления, экономичное хранение больших объемов или специализированные сценарии.

Приоритет производительности

Область применения:

- интерактивная аналитика (Ad-hoc) и отчетность на больших данных
- высоконагруженные витрины данных и операционные хранилища (ODS)
- системы, требующие минимальной задержки отклика для бизнес-приложений

Варианты решения:

- увеличенный объем оперативной памяти (RAM) и высокая базовая частота процессоров (CPU) для ускорения обработки в памяти и сложных вычислений
- использование узлов с максимальной CPU-мощностью для быстрого выполнения ETL-задач пакетной и потоковой обработки
- оснащение высокопроизводительными накопителями NVMe SSD для максимальной скорости чтения/записи
- использование RAID 10 для системных накопителей вычислительных узлов для баланса между производительностью и надежностью

Приоритет объема и эффективности хранения

Область применения:

- централизованное хранилище данных (Data Lake)
- долгосрочное хранение данных с датчиков IoT, журналов событий и исторических архивов
- экономичное хранение "холодных" и "теплых" данных с сохранением быстрого доступа

Вариант решения:

- сбалансированная конфигурация с оптимизацией на соотношение "емкость/стоимость"
- масштабируемая конфигурация для создания долгосрочных архивов и моментальных снимков данных
- использование высокоемких SAS/NL-SAS-накопителей. Активное применение Erasure Coding для снижения затрат на хранение при обеспечении отказоустойчивости

Специализированный тюнинг под прикладные задачи

Область применения

- решение уникальных задач, требующих тонкой настройки под специфичную модель данных, паттерны запросов и соглашения об уровне сервиса (SLA)

Подход к реализации проводится совместно архитекторами **Скала^р** и техническими специалистами заказчика и может включать:

- адаптацию конфигурации оборудования (соотношение CPU/RAM/диски) под конкретные движки (Spark, Impala, Trino)
- оптимизацию сетевых настроек и топологии размещения данных в HDFS/Ozone
- кастомизацию политик управления ресурсами YARN и планировщиков для обеспечения предсказуемой производительности мультитенантных нагрузок
- настройку правил миграции данных между типами хранилищ (SSD/HDD) в Smart Storage Manager

6. СОСТАВ МАШИНЫ

Ниже приведены термины, используемые для комплектации **Машины больших данных Скала^р МБД.Х**.

Машина – это набор аппаратного и программного обеспечения в виде **Модулей Скала^р**, соединенных вместе для обеспечения определенного метода обработки данных или предоставления ИТ-сервисов с заданными характеристиками.

Подсистема – логическое объединение компонентов по функциональному признаку, с целью пояснения состава и принципов действия ПАК.

Модуль – это единица поставки **Машин**, выполняющая определенные функции в соответствии с ее назначением. Модуль является единым и неделимым элементом спецификации и содержит набор аппаратных узлов и ПО.

Узел – это элемент, выполняющий определенную задачу в составе **Модуля**.

6.1 Комплекты поставки

Машины больших данных Скала^р МБД.Х поставляются в виде функционально полного набора **Модулей Скала^р** и комплектуются в соответствии с показателями назначения, полученными от заказчика. **Машина** включает в себя базовый комплект и в случае необходимости дополняется комплектом модулей расширения и/или специальными модулями.

Базовый комплект – это набор **Модулей Скала^р**, минимально-необходимый для функционирования всех подсистем, обеспечивающих выполнение основного функционала **Машины**.

Комплект модулей расширения – это набор **Модулей Скала^р**, позволяющий масштабировать ПАК, например, когда не хватает портовой емкости, или есть необходимость увеличить производительность и объем хранения данных. Кроме того, можно добавить специальные **Модули Скала^р**, позволяющие расширить функциональность ПАК.

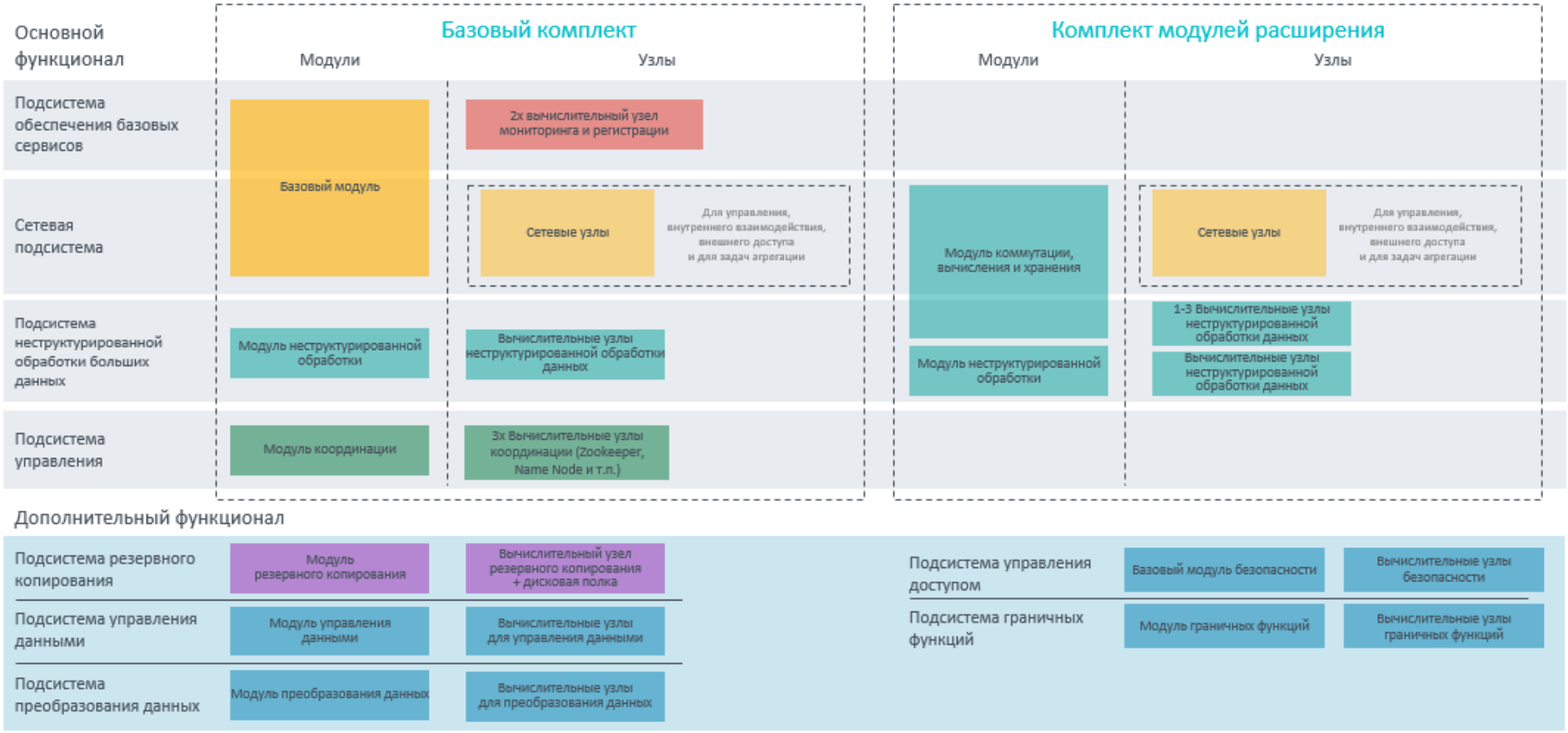

Рисунок 4. Комплектация Машины больших данных Скала^р МБД.Х

6.2 Подсистемы

Функции **Скала^р МБД.Х** логически объединены в подсистемы. Часть подсистем обеспечивают основной функционал и всегда включены в **Машину**, а часть – дополнительный функционал и могут быть добавлены по требованию заказчика.

Основной функционал – это минимальный набор подсистем, необходимых **Машине** для выполнения задач прямого назначения.

Дополнительный функционал – набор подсистем из **Модулей**, обеспечивающих расширение функций **Машины**.

6.2.1 Подсистема обеспечения базовых сервисов и сетевая подсистема

Подсистема обеспечения базовых сервисов отвечает за мониторинг и управление аппаратными и программными компонентами **Машины**. В нее включены вычислительные узлы **Базового модуля** (см. п. 6.3.1), на которых предустановлено сервисное ПО **Скала^р Геном** и **Скала^р Визион**, выполняющее следующие функции:

- сбор, хранение и отображение на дашбордах данных мониторинга
- управление аппаратными компонентами
- управление пользователями и аутентификация (опционально)
- настройка программных компонентов
- настройка интеграции со сторонним ПО

Архитектура подсистемы обеспечения базовых сервисов обеспечивает отказоустойчивый режим работы.

Сетевая подсистема выполняет функций организации сетевой связанности между всеми вычислительными узлами, входящими в состав **Скала^р МБД.Х**, и представляет собой набор сетевых узлов, которые организуют изолированные высокоскоростные сети:

- внутреннего взаимодействия (в зависимости от требований заказчика 25 Гбит/с или 100 Гбит/с) – для организации быстрого функционирования между всеми компонентами ПАК
- внешнего доступа (в зависимости от требований заказчика 25 Гбит/с или 100 Гбит/с) – для организации доступа к данным, что хранятся на узлах, входящих в состав подсистемы аналитической обработки данных
- управления (1 Гбит/с) – для организации передачи сервисной информации с вычислительных узлов, входящих в состав подсистемы аналитической обработки данных, на вычислительные узлы, входящие в состав подсистемы обеспечения базовых сервисов

Стартовый комплект сетевых узлов всегда размещается в **Базовом модуле** (см. п. 6.3.1).

6.2.2 Подсистема неструктурированной обработки больших данных

Основная подсистема, которая выполняет задачи, связанные с хранением и работой с данными различных типов (структурированных, слабоструктурированных, неструктурированных):

- масштабируемая и отказоустойчивая файловая система
- обработка огромных объемов данных из разных источников с использованием распределенных федеративных запросов
- объектное хранилище горячих, теплых и холодных данных на различной инфраструктуре
- хранение аналитических таблиц

- поиск данных

Подсистема реализуется **Модулем неструктурированной обработки** (см. п. 6.3.2).

6.2.3 Подсистема управления

Обеспечивает координацию работы распределенных сервисов кластера, управление метаданными и поддержание согласованности состояния системы. Подсистема является фундаментальной для функционирования высокодоступного и отказоустойчивого кластера.

Подсистема реализуется **Модулем координации** (см. п. 6.3.3).

6.2.4 Подсистема резервного копирования

Подсистема резервного копирования предназначена для создания согласованных копий критических данных платформы и их конфигураций с целью аварийного восстановления. Она включает один или несколько **Модулей резервного копирования** (см. п. 6.3.4), которые предоставляют выделенное отказоустойчивое сетевое хранилище (например, по протоколу сетевой файловой системы, или интегрироваться с внешними объектными хранилищами) в качестве целевого для архивов.

Резервное копирование выполняется с помощью специализированных инструментов платформы Arenadata Hyperwave и может включать:

- создание снимков (snapshots) распределенной файловой системы HDFS для получения согласованной на определенный момент времени копии данных
- копирование данных между кластерами с помощью утилиты DistCp (Distributed Copy)
- экспорт метаданных из Hive Metastore и других систем каталогов
- резервное копирование конфигураций и ключей безопасности с помощью Arenadata Cluster Manager (ADCM)
- политики резервного копирования (периодичность, глубина хранения, набор компонентов) настраиваются в соответствии с требованиями заказчика к точке восстановления (RPO) и времени восстановления (RTO)

6.2.5 Подсистема управления данными

Отвечает за организацию совместной работы с данными, интеграцию метаданных из различных систем обработки и анализа данных, а также предоставляет возможности поиска данных и совместной работы с метаданными, ведения корпоративного бизнес-глоссария и его тесной интеграции с каталогом данных. В основе подсистемы лежит ПО Arenadata Catalog.

Подсистема реализуется **Модулем управления данными** (см. п. 6.3.5).

6.2.6 Подсистема преобразования данных

Состоит из **Модулей преобразования данных** (см. п. 6.3.6), которые необходимы для создания аналитических платформ, а также для интеграции, выгрузки и обработки данных из любых источников. Является основой для выстраивания и оркестрации ETL/ELT-процессов.

6.2.7 Подсистема управления доступом

Включает в себя набор инструментов и механизмов для защиты данных, управления доступом и обеспечения соответствия требованиям безопасности. В основе подсистемы лежит ПО Arenadata Platform Security, которое включает Ranger для централизованного управления политиками доступа и Knox в качестве безопасного шлюза для внешних

подключений. Начиная с версии ADH 4.1.0, платформа предоставляет расширенные возможности для управления доступом к конфиденциальным данным.

Подсистема реализуется **Базовым модулем безопасности** (см. п. 6.3.7).

6.2.8 Подсистема граничных функций

Обеспечивает изолированное исполнение разнородных рабочих нагрузок и предоставление интерфейсов для взаимодействия с платформой. Подсистема отвечает за выполнение задач распределенной обработки данных, включая ETL-пайплайны, потоковую аналитику и обслуживание SQL-запросов от бизнес-приложений и аналитиков.

Подсистема реализуется **Модулем граничных функций** (см. п. 6.3.8)

6.3 Модули

6.3.1 Базовый модуль

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Базовый модуль. Обеспечивает функционирование подсистемы обеспечения базовых сервисов и сетевой подсистемы (см. п. 6.2.1).

Назначение

- Обеспечение сетевой связанности между компонентами
- Организация выделенной сети управления **Машиной**
- Организация подключения к сети заказчика
- Исполнение функций мониторинга и управления компонентами **Машины**

Узлы

- Два вычислительных узла мониторинга и регистрации, которые объединены в зеркальный кластер и используются для служебных функций
- Два сетевых узла 25/100 Гбит/с для организации внутреннего сетевого взаимодействия
- Два сетевых узла 25/100 Гбит/с для организации сети внешнего доступа (опционально)
- Сетевой узел 1 Гбит/с для организации работы сети управления, также может быть выполнен в отказоустойчивом исполнении
- Два сетевых узла 100 Гбит/с для организации агрегации, в случае добавления внутренних портов в крупных конфигурациях ПАК (опционально)

Отказоустойчивость обеспечена

- Резервированием вычислительных узлов, отвечающих за мониторинг и управление компонентами **Машины**
- Технологией RAID для накопителей вычислительных узлов
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)

Применяемое программное обеспечение

- **Скала^р Визион**
- **Скала^р Геном**

- ОС, входящая в состав ПАК
- Сервисное ПО, входящее в состав Arenadata Cluster Manager (ADCM)
- ПО для управления пользователями и аутентификацией (опционально)

6.3.2 Модуль неструктурированной обработки

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Модуль неструктурированной обработки. Обеспечивает функционирование подсистемы неструктурированной обработки больших данных (см. п. 6.2.2), являясь фундаментальным слоем хранения и пакетной обработки данных в комплексе.

Назначение

- Хранение данных любого типа: обеспечивает надежное, отказоустойчивое и масштабируемое хранение структурированных, слабоструктурированных (JSON, XML, лог-файлы) и неструктурированных (изображения, видео, документы) данных
- Поддержка Data Lake/Data Lakehouse: формирует основу для построения централизованного хранилища данных предприятия (Data Lake) и современных гибридных архитектур (Data Lakehouse) благодаря интеграции с Ozone и Iceberg
- Высокопроизводительная пакетная обработка: предоставляет вычислительные ресурсы для выполнения распределенных ETL-задач, сложной аналитики и работ по подготовке данных с использованием таких фреймворков, как Spark и MapReduce
- Обеспечение надежности и эффективности: реализует механизмы репликации и сжатия данных для гарантии их сохранности и снижения затрат на хранение

Узлы

В состав Модуля входят вычислительные узлы, сконфигурированные и распределенные по 2 типам нагрузки в зависимости от преобладающих задач:

- **тип 1** – оптимизирован для рабочих нагрузок, требующих высокой пропускной способности, минимальной задержки доступа к часто используемым («горячим») данным, и для ресурсоемких вычислительных процессов. Оснащается процессорами с большим количеством ядер и высокой тактовой частотой, большим объемом оперативной памяти и высокоскоростными накопителями NVMe SSD
- **тип 2** – сконфигурирован для экономичного хранения больших объемов данных при сохранении производительности. Комплектуется процессорами с оптимальным соотношением цены и производительности, объемной оперативной памятью и высокоемкими накопителями SAS/NL-SAS. Применяется для хранения «холодных» (редко используемых) и «теплых» (среднего спроса) данных, а также для выполнения фоновых заданий, не требующих минимальной задержки отклика

Отказоустойчивость обеспечена

- Репликацией данных в HDFS: данные автоматически реплицируются на разные узлы и серверные стойки. Выход из строя одного или нескольких узлов не приводит к потере данных или неработоспособности кластера. HDFS автоматически восстанавливает фактор репликации на исправных узлах
- Отказоустойчивостью сервисов: критичные сервисы, такие как HDFS DataNode и Ozone Storage Container Manager, развернуты на множестве узлов. Отказ одного узла не влияет на доступность сервиса в целом
- Технологией RAID для дисков вычислительных узлов

- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)

Применяемое программное обеспечение

- Hadoop Distributed File System (HDFS)
- Ozone
- Фреймворки для высокопроизводительной распределенной обработки данных в оперативной памяти
- Компоненты для управления жизненным циклом данных, включая перемещение между типами хранилищ (SSD/HDD), применение Erasure Coding и компрессию файлов
- ОС, входящая в состав ПАК

6.3.3 Модуль координации

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Модуль координации. Обеспечивает функционирование подсистемы управления (см. п. 6.2.3).

Назначение

- Обеспечение консенсуса и координации распределенных сервисов кластера
- Хранение и управление критическими метаданными распределенной файловой системы (HDFS)
- Предоставление сервиса распределенной конфигурации и синхронизации для компонентной экосистемы (Kafka, HBase, Hive)
- Гарантия согласованности состояния кластера и отказоустойчивости ключевых управляющих сервисов

Узлы

В состав Модуля входят три вычислительных узла, которые распределены по 2 типам нагрузки:

- **тип 1** – высокопроизводительный, оптимизирован для узлов с ролью HDFS NameNode и JournalNode, требующих низкой задержки и высокой производительности процессорной подсистемы
- **тип 2** – сбалансированный, используется для развертывания кворума Zookeeper и сервисов метаданных, требующих стабильной работы и надежного хранилища

Отказоустойчивость обеспечена

- Формированием отказоустойчивого кворума (не менее 3 узлов) для сервисов Zookeeper и HDFS NameNode (с использованием JournalNode)
- Автоматическим переключением при отказе основного узла NameNode на резервный узел
- Технологией RAID для накопителей вычислительных узлов
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)

Применяемое программное обеспечение

- Zookeeper
- HDFS NameNode & JournalNode
- ОС, входящая в состав ПАК

6.3.4 Модуль резервного копирования

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Модуль резервного копирования. Обеспечивает функционирование подсистемы резервного копирования (см. п. 6.2.4).

Назначение

- Резервирование и восстановление данных: обеспечивает создание резервных копий данных, хранящихся в основных подсистемах Машины (например, в подсистеме неструктурированной обработки больших данных), и их последующее восстановление в случае необходимости
- Долгосрочное хранение резервных копий: предоставляет выделенное, экономичное хранилище для архивных данных, моментальных снимков состояния системы и других материалов резервного копирования
- Интеграция с экосистемой: функционирует как сетевое хранилище (NFS), обеспечивая совместимость со стандартными механизмами резервного копирования, входящими в состав платформы Arenadata Hyperwave и других компонентов

Узлы

В состав Модуля входит один вычислительный узел, обеспечивающий хранение до 94 Тбайт данных. Хранение осуществляется на высокоемких и экономичных накопителях NL-SAS.

Отказоустойчивость обеспечена

- Аппаратная надежность хранения: технология RAID для накопителей вычислительного узла, защищающая от потери данных при выходе из строя одного или нескольких дисков
- Сетевая доступность: резервирование сетевых коммутаторов (объединение сетевых узлов в MLAG-пару), обеспечивающее отказоустойчивый сетевой путь к хранилищу резервных копий
- Резервирование на уровне Машины: критичность данного Модуля нивелируется общей архитектурой Машины, где основные данные защищены репликацией в HDFS. Модуль резервного копирования предназначен для защиты от логических ошибок, катастрофических сбоев и создания архивных копий

Применяемое программное обеспечение

- Сетевая файловая система (NFS)
- ОС, входящая в состав ПАК

6.3.5 Специализированный модуль (для управления данными)

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Специализированный модуль. Обеспечивает функционирование подсистемы управления данными (см. п. 6.2.5). Интерфейс основного функционального ПО Модуля предоставляет

доступ к каталогу метаданных, бизнес-глоссарию, поиску, профилированию и проверке качества корпоративных данных.

Назначение

- Интеграция метаданных из различных систем обработки
- Анализ данных, поиск данных, совместная работа с метаданными
- Ведение корпоративного бизнес-глоссария и его интеграция с каталогом данных

Узлы

В состав Модуля входит не менее двух вычислительных узлов

Отказоустойчивость обеспечена

- Резервированием вычислительных узлов
- Технологией RAID для накопителей вычислительных узлов
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)

Применяемое программное обеспечение

- Arenadata Catalog (ADC)
- ОС, входящая в состав ПАК

6.3.6 Специализированный модуль (для задач преобразования данных)

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Специализированный модуль. Обеспечивает функционирование подсистемы преобразования данных (см. п. 6.2.6).

Назначение

Используется для решения задач, связанных с интеграцией данных, построения и наполнения хранилищ и витрин данных.

Узлы

В зависимости от модификации, Модуль может состоять из 2× или 3× вычислительных узлов.

Отказоустойчивость обеспечена

- Резервированием вычислительных узлов
- Технологией RAID для накопителей вычислительных узлов
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)

Применяемое программное обеспечение

- ПО для управления ETL-процессами

6.3.7 Базовый модуль безопасности

Базовый модуль безопасности обеспечивает функционирование подсистемы управления доступом (см. п. 6.2.7).

Назначение

- Предоставляет комплексную систему, включающую управление доступом на основе политик, авторизацию и безопасный доступ к платформе и ее сервисам, что помогает защитить конфиденциальные данные и обеспечить соответствие нормативным требованиям.
- В Модуль входят компоненты Arenadata Platform Security (ADPS):
- Ranger Admin & UserSync – серверы управления политиками безопасности и синхронизации пользователей (развертываются на 2-3 узлах для отказоустойчивости)
- Ranger Key Management Service (KMS) – сервис управления ключами шифрования (развертывается на 2 узлах)
- Knox Gateway – шлюз безопасности (развертывается на 2 узлах для балансировки нагрузки и отказоустойчивости)

Узлы

В зависимости от модификации, Модуль может состоять из 3, 5 или 7 вычислительных узлов для обеспечения требуемой производительности и резервирования.

Отказоустойчивость обеспечена

- Кластерной конфигурацией сервисов Ranger и Knox с балансировкой нагрузки и автоматическим переключением при сбое
- Хранением политик Ranger в реплицируемой БД (встроенной или внешней)
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)
- Технологией RAID для системных накопителей вычислительных узлов

Применяемое программное обеспечение

- Arenadata Platform Security (ADPS) на базе Ranger и Knox
- ОС, входящая в состав ПАК

6.3.8 Модуль граничных функций

Название в Едином реестре российской радиоэлектронной продукции – СКАЛА-Р Модуль граничных функций. Обеспечивает функционирование подсистемы граничных функций (см. п. 6.2.8).

Назначение

Предоставляет выделенные вычислительные ресурсы для исполнения разнородных рабочих нагрузок в изолированной среде. Модуль предназначен для запуска и управления задачами распределенной обработки данных, такими как процессы преобразования и перемещения данных (Spark), потоковая обработка (Flink), а также для размещения SQL-шлюзов (Kyuubi), обеспечивая тем самым эластичность, мультитенантность и предсказуемую производительность для бизнес-приложений и аналитиков.

Узлы

Узлы сконфигурированы с акцентом на высокую производительность CPU и большой объем оперативной памяти для эффективной обработки данных в памяти.

Отказоустойчивость обеспечена

- Интеграцией с YARN для управления ресурсами и автоматического перезапуска контейнеров с задачами при сбоях
- Использованием отказоустойчивых режимов исполнения для Spark и Flink
- Развертыванием критичных шлюзов (Kyuubi) в кластерном режиме с балансировкой нагрузки
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)
- Технологией RAID для системных накопителей вычислительных узлов

Применяемое программное обеспечение

- YARN
- Spark (включая Spark SQL, MLlib)
- Flink
- Kyuubi
- ОС, входящая в состав ПАК

7. ПРОГРАММНЫЕ КОМПОНЕНТЫ

7.1 Программная платформа Скала^р Геном

Программная платформа **Скала^р Геном** предназначена для управления жизненным циклом **Машины**.

Основные преимущества **Скала^р Геном**:

- упрощает поддержку **Машин**, установленных у заказчика, включая вопросы обновления компонентов
- снижает требования к квалификации эксплуатирующего **Машину** персонала

7.2 ПО Скала^р Визион (компонент ПО Геном)

Машины поставляются со средствами мониторинга, настроенными на контроль характерных для **Машин Скала^р** параметров с заданными пороговыми значениями, позволяя из единой консоли проводить полноценный мониторинг всех компонентов **Машины**.

Основные функциональные возможности встроенного в **Машину** мониторинга:

- мониторинг серверного оборудования
- мониторинг сетевого оборудования
- мониторинг программно-определяемого хранилища и сервисов объектного хранения
- централизованное хранение и анализ лог-файлов
- система пороговых оповещений и их отправки
- передача данных во внешние системы мониторинга
- возможность централизованного мониторинга нескольких **Машин Скала^р**

8. ГАРАНТИРОВАННОЕ КАЧЕСТВО

Качественные показатели **Машины больших данных Скала^р МБД.Х** обеспечиваются ее соответствием проверенному стандартному варианту, соблюдением установленных норм и требований по формированию, реализацией работ высококвалифицированными специалистами на всех этапах жизненного цикла.

Производство и сборка: без компромиссов

- Используются только **проверенные комплектующие**, отобранные под реальные нагрузки
- Все компоненты собираются строго по регламенту, в соответствии с утвержденной схемой размещения
- Развертывание программного обеспечения и первичная настройка выполняются **автоматизированно**, чтобы исключить «ручные» ошибки
- Каждый комплекс перед отгрузкой проходит функциональное тестирование и проверку на наличие известных уязвимостей
- Отклонения от типового решения **Машины** исключены

Передача в эксплуатацию: все готово к работе

- **Машина** поставляется в полной готовности к работе – готова к эксплуатации сразу после подключения к сети заказчика
- В комплект входят: паспорт изделия и сертификат поддержки, полный пакет документации для прохождения аттестаций и согласований, обучение специалистов заказчика (по запросу)

Техническая поддержка: от производителя, без посредников

- Поддержка входит в поставку (по умолчанию – 1 год, оптимально – 3 или 5 лет)
- Доступны пакеты технической поддержки 9×5 или 24×7, в зависимости от критичности системы заказчика
- Первая и вторая линии поддержки непосредственно от производителя или сертифицированного партнера
- В сложных ситуациях в работу подключаются архитекторы и разработчики самой **Машины** – 3-я линия поддержки в России, без эскалации за рубеж

Сопровождение и развитие под задачи бизнеса

По запросу возможно:

- аппаратное расширение и модернизация
- горизонтальное и вертикальное масштабирование
- адаптация **Машины** под новые задачи

Все доработки выполняются с участием тех, кто проектировал и создавал **Машину больших данных Скала^р МБД.Х**, что снижает риски и обеспечивает стабильность в эксплуатации.

9. РЕАКЦИЯ МАШИНЫ НА ВОЗМОЖНЫЕ ОТКАЗЫ

9.1 Отказы, связанные со стандартными элементами

В **Машине** обеспечена отказоустойчивость ее основных элементов и процессов, в том числе:

- узлов (дублирование процессоров, источников питания и др.)
- подсистемы ввода-вывода (RAID)
- сети внутреннего взаимодействия (дублирование сетевых интерфейсов, MLAG)
- системы резервного копирования

Отказы перечисленных элементов отрабатываются стандартными алгоритмами в соответствии с произведенными настройками. Любой единичный отказ не повлияет на доступность системы в целом, хотя по конкретному сервису возможно небольшое снижение производительности. После устранения неисправности исходная производительность **Машины** также восстанавливается.

9.2 Отказы, связанные с узлами кластера

Аппаратные сбои

Архитектура программного обеспечения, лежащего в основе **Машины больших данных Скала^р МБД.Х**, позволяет построить отказоустойчивый многоузловой кластер.

Архитектура **Машины** построена по принципу распределенной системы с независимыми узлами, не имеющими разделяемых ресурсов, что обеспечивает устойчивость к потере любого из них.

HDFS

Данные защищены репликацией блоков. Выход из строя одного или даже двух узлов данных (в зависимости от расположения реплик) не приводит к потере данных или неработоспособности кластера. HDFS автоматически инициирует процесс восстановления репликации на исправных узлах.

Управляющие сервисы

Критичные сервисы, такие как HDFS NameNode, Zookeeper и компоненты безопасности (Ranger, Knox), развернуты в высокодоступной кластерной конфигурации. Выход из строя одного узла в таком кластере приводит к автоматическому переключению на резервный узел без прерывания обслуживания пользовательских запросов.

Координация

Сервис координации Zookeeper, необходимый для работы многих компонентов, развернут в кворуме из 3 или более узлов. Кластер Zookeeper остается работоспособным при отказе не более чем (N-1)/2 узлов (например, при 3 узлах кластер переживет отказ одного).

Полная потеря кворума Zookeeper является критическим сбоем, который может привести к остановке или нестабильной работе многих ключевых сервисов кластера (HDFS NameNode HA, HBase, Kafka и др.), а не только к переходу в режим «только чтение». Восстановление после такого инцидента требует тщательной процедуры, описанной в документации по аварийному восстановлению, и может включать как автоматические, так и ручные шаги по проверке целостности и согласованности состояния компонентов.

Программные сбои и человеческий фактор

Мониторинг

Встроенная система мониторинга **Скала^р Визион** отслеживает состояние всех сервисов и узлов, генерируя оповещения при деградации или сбое.

Восстановление сервисов

Менеджер кластера Arenadata Cluster Manager (ADCM) позволяет быстро перезапустить отказавшие сервисы или выполнить откат конфигурационных изменений, приведших к нестабильной работе.

Резервное копирование и восстановление

Для защиты от логических ошибок и ошибочных операций реализована система резервного копирования метаданных и самих данных. В зависимости от требований заказчика, может быть настроено резервное копирование в отдельный Модуль резервного копирования или в объектное хранилище.

Детальный порядок обеспечения отказоустойчивости кластера и рекомендации по действиям при его администрировании, включая процедуры аварийного восстановления (Disaster Recovery), могут быть предоставлены по запросу.

10. ТРЕБОВАНИЯ К РАЗМЕЩЕНИЮ МАШИНЫ

Машина больших данных Скала^р МБД.Х представляет собой комплект узлов для размещения в серверный монтажный шкаф 19”, высота 42U и больше, с дальнейшей возможностью модульной расширяемости до 14 стоек (или более).

Монтажный шкаф (стойка) может быть поставлен как опция.

Для подключения шкафа к системе электроснабжения должны быть предусмотрены два независимых входа электропитания.

Расчетная потребляемая мощность шкафа (задается параметрами ЦОД заказчика) определяет топологию размещения модулей и узлов в стойках ЦОД и учитывается при расчете Машины. От этого зависит количество дополнительного коммутационного оборудования в составе Машины.

В месте установки должны быть предусмотрены соответствующие мощности по отводу тепла.

Для подключения к локальной сети заказчика необходим резервированный канал до 4×100 Gigabit Ethernet или до 8×10/25 Gigabit Ethernet. Требуемые трансиверы определяются на этапе формирования спецификации **Машины**.

При развертывании будут выполнены настройки сетевых адресов в соответствии со структурой сети заказчика. Заказчик должен предоставить необходимые данные в соответствии с номенклатурой компонентов **Машины больших данных Скала^р МБД.Х**.

В сети заказчика должны быть настроены соответствующие маршруты и права доступа.

Дальнейшие мероприятия по вводу в эксплуатацию осуществляются заказчиком путем настройки прикладных программных систем.

11. ТЕХНИЧЕСКАЯ ПОДДЕРЖКА

Поставка **Машин больших данных Скала^р МБД.Х** осуществляется с предварительными сборкой, тестированием и настройкой оборудования согласно требованиям заказчика. Качественная поддержка обеспечивается едиными стандартами гарантийного и постгарантийного технического обслуживания:

- пакет услуг по технической поддержке на первый год включен в поставку
- заказчик может выбирать пакет 9×5 или 24×7 (вариант для комплексов критической функциональности)
- срок начально приобретаемой технической поддержки может быть увеличен до 3-х и 5-и лет, также доступна пролонгация поддержки

Состав типовых пакетов услуг по технической поддержке **Скала^р МБД.Х** представлен в таблице 1.

Таблица 1. Пакеты услуг по технической поддержке

Услуги	Пакет «9×5»	Пакет «24×7»
«Режим предоставления услуг 9×5» (в рабочее время по рабочим дням)	+	–
«Режим предоставления услуг 24×7» (круглосуточно)	–	+
Предоставление доступа к системе регистрации запросов/инцидентов Service Desk	+	+
Предоставление доступа к базе знаний по продуктам **Скала^р**	+	+
Предоставление обновлений лицензионного ПО **Скала^р**	+	+
Диагностика, анализ и устранение проблем в работе комплекса **Скала^р**, включая: устранение аппаратных неисправностей техническое сопровождение ПО	+	+
Консультации по работе комплекса **Скала^р**	+	+
«Защита конфиденциальной информации» (неисправные носители информации не возвращаются заказчиком)	Опция	Опция
Замена и ремонт оборудования по месту установки	+	+
Доставка оборудования на замену за счет производителя	+	+
Расширенные параметры обслуживания	–	+

Услуги	Пакет «9×5»	Пакет «24×7»
Времена реагирования и отклика, не более:		
Время регистрации обращений	30 минут, рабочие часы (9×5)	30 минут, круглосуточно (24×7)
Подключение специалиста к решению инцидентов критичного и высокого уровней	В течение 1 рабочего часа (9×5)	В течение 1 часа (24×7)

Примечание к срокам ремонта оборудования

Комплекс **Машина больших данных Скала^р МБД.Х** архитектурно является устойчивым к выходу из строя отдельных компонентов и даже узлов, поэтому нет необходимости в обеспечении дорогостоящего сервиса срочного восстановления оборудования в течение суток и менее. В комплексе предусмотрено как минимум двойное резервирование основных компонентов, позволяющее сохранять данные и работоспособность даже при выходе из строя нескольких дисков и/или серверов (узлов).

12. ПОСТАВКА И ЛИЦЕНЗИРОВАНИЕ ПО

Команда **Скала^р** активно занимается развитием программных продуктов **Машин больших данных Скала^р МБД.8**. Направления развития формируются на основе анализа мирового опыта использования систем подобного класса и пожеланий заказчиков и партнеров. Новые функции реализуются в форме релизов, которые могут выходить несколько раз в год.

Программное обеспечение Arenadata Catalog лицензируется по количеству пользователей с правами администратора или модератора.

Программное обеспечение **Скала^р Геном**, **Скала^р Визион** поставляется исключительно в составе **Машин Скала^р** и лицензируется по метрикам комплекса в соответствии с количеством серверных узлов.

12.1 Политика обновления ПО

Команда **Скала^р** активно занимается развитием собственных программных продуктов. Направления развития формируются на основе анализа мирового опыта использования систем подобного класса и пожеланий заказчиков и партнеров. Новые функции реализуются в форме релизов. Обновления для **Машин**, находящихся в эксплуатации, производятся по согласованию с заказчиком.

О КОМПАНИИ

Скала^р – модульная платформа для построения высоконагруженной ИТ-инфраструктуры (продукт Группы Rubytech).

Программно-аппаратные комплексы (**Машины**) **Скала^р** выпускаются с 2015 года и представляют широкий технологический стек для построения динамических инфраструктур и инфраструктур управления данными высоконагруженных информационных систем.

Продукты **Скала^р** включены в Реестр промышленной продукции, произведенной на территории Российской Федерации, и в Единый реестр российских программ для ЭВМ и БД. Соответствует критериям доверенности и использованию для объектов критической информационной инфраструктуры (КИИ).

Машины Скала^р являются серийно выпускаемыми преднастроенными комплексами, которые быстро развертываются и вводятся в эксплуатацию. Глубокая интеграция технических средств и программного обеспечения в ПАК **Скала^р** позволяет получить расширенные возможности и функциональность, которые недоступны при использовании отдельных компонентов.

Модульный принцип обеспечивает интеграцию разнородных компонентов ИТ-инфраструктуры в единую платформу предприятий, корпораций и ведомств. Единые поддержка и сервисное обслуживание для всех продуктов линейки **Скала^р** от производителя обеспечивают оперативное разрешение инцидентов на стыке технологий.

Дополнительная информация – на сайте www.skala-r.ru.