

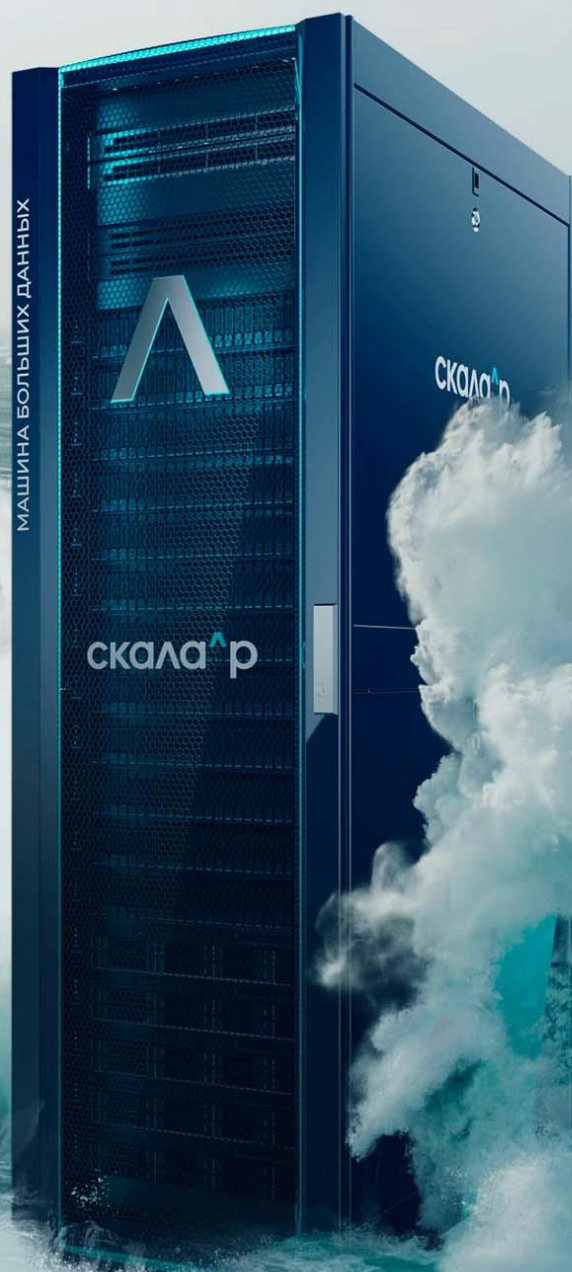


Машина больших данных Скала^р МБД.Х

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

Технический обзор

версия 2.4 от 24.10.2024



ОГЛАВЛЕНИЕ

1. Предисловие.....	4
1.1 Описание документа	4
1.2 Аудитория	4
1.3 Обратная связь	4
2. Введение	5
3. Задачи обработки больших данных	6
4. Отличительные черты	9
5. Подтвержденная безопасность	10
6. Технологические преимущества	12
7. Состав решения	17
8. Специфичные черты	31
9. Гарантированное качество.....	32
10. Реакция Машины на возможные отказы	33
11. Вариативность решения	34
12. Требования к размещению решения.....	35
13. Техническая поддержка	36
14. Поставка и лицензирование ПО	38
Политика обновления ПО.....	38
О Компании	39

Информация, представленная в документе, носит исключительно информационный характер, является актуальной на дату размещения.

Технические характеристики, приведенные в документе — справочные и не могут служить основанием для претензий.

Технические характеристики могут отличаться от приведенных вследствие модификации изделий.

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

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

1. ПРЕДИСЛОВИЕ

1.1 Описание документа

Этот технический обзор дает концептуальный и архитектурный обзоры **Машины больших данных Скала^{Ар} МБД.Х.**

Брошюра описывает то, как оптимизированные программно-аппаратные комплексы отвечают современным вызовам, и фокусируется на **Машине больших данных Скала^{Ар} МБД.Х** как одном из лидирующих решений в этом сегменте.

1.2 Аудитория

Эта брошюра предназначена для сотрудников компании **Скала^{Ар}**, партнеров и Заказчиков, перед которыми ставятся задачи разработки решения, закупки, управления или эксплуатации **Машины больших данных Скала^{Ар} МБД.Х.**

1.3 Обратная связь

Скала^{Ар} и авторы этого документа будут рады обратной связи по нему.

Свяжитесь с командой **Скала^{Ар}** по электронной почте MV@skala-r.ru.

2. ВВЕДЕНИЕ

Машина больших данных Скала^р МБД.Х — программно-аппаратный комплекс для обработки больших данных на базе технологий Arenadata Hadoop (далее — АДН).

Машина Скала^р МБД.Х повышает производительность и отказоустойчивость, снижает затраты за счет проработанной интеграции аппаратного и программного обеспечения, оптимизации алгоритмов для используемых технологий, широкого применения методов обеспечения надежности решения.

Машина Скала^р МБД.Х — адаптированный для корпоративного использования продукт, предназначенный для хранения и обработки слабоструктурированных и неструктурированных данных.

Машина Скала^р МБД.Х — комплексное решение, включающее в себя масштабируемые узлы для проведения вычислений и хранения данных, сверхскоростную сетевую среду, систему интеллектуального управления.

Высокая производительность решения достигается в том числе применением оптимальных по производительности комплектующих и современных стандартов, накопителей SSD, сетевых протоколов 100 Gigabit Ethernet.

Отказоустойчивость обеспечивается применением надежных комплектующих, с резервированием критических компонентов, использованием устойчивых сетевых протоколов.

Машина Скала^р МБД.Х содержит все необходимые элементы для функционирования высоконагруженной системы Hadoop. Подключение к внешним сетям осуществляется с помощью стандартного интерфейса Ethernet.

Машина Скала^р МБД.Х допускает применение различных типов данных, предоставляя возможности для их распределенной обработки и снижения стоимости эксплуатации.

Реализованы функции мониторинга состояния как аппаратных, так и программных компонентов решения, а также необходимые функции управления.

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

3. ЗАДАЧИ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Компании постоянно сталкиваются с ростом объемов получаемой и хранимой информации, эту информацию необходимо оперативно обрабатывать и превращать в результаты для бизнеса.

Данные необходимы современным компаниям, чтобы совершенствовать бизнес-процессы, повышать конкурентоспособность продукции и улучшать качество обслуживания клиентов. Массивы информации о клиентах, сотрудниках, финансовых показателях, транзакциях, операционной деятельности необходимо хранить, обрабатывать, анализировать и постоянно пополнять.

Государственные организации для выполнения возложенных на них функции сталкиваются со следующими задачами:

- Актуализация информации о множестве объектов госорганов
- Оперативная подготовка данных для государственной отчетности
- Предоставления точной и объективной, информации для прогнозирования
- Идентификация фактов, выявление и противодействие мошенничеству
- Сбор, хранение и обработка информации с региональных и местных уровней

Пример: увеличение эффективности лечения пациентов в организациях здравоохранения связано с большими объемами обработки информации в медицинской сфере и неструктурированным набором данных. Анализ медицинских данных позволяет выявлять, факторы риска заболеваний с учетом сезонности, оценивать эффективность превентивных мер и лечения, а также проводить научные исследования.

Банкам требуется постоянное повышение качества обслуживания клиентов для чего необходимо:

- Проводить анализ различных источников данных о доходах и расходах клиентов
- Анализировать транзакции из различных систем для снижения рисков и предотвращение мошенничества
- Собирать и анализировать отзывы клиентов для повышения лояльности

Пример: повышение операционной эффективности в банках и других финансовых организациях зависит от увеличения притока и снижение оттока клиентов и достигается за счет анализа финансовой информации, а также идентификации рисков и угроз. Для этого необходимо проанализировать большие объемы данных о клиентах и транзакциях и выявлять подозрительные действия. На основе этих данных разрабатываются стратегии обеспечения безопасности, прогнозируются количество посетителей в отделениях, остатки средств в банкоматах.

Компании из сферы продаж сталкиваются с задачами постоянного поддержания спроса, что требует:

- Исследования данных покупок по всем номенклатурным позициям для понимания изменений в предпочтениях потребителей
- Анализа данных с множества торговых точек, скорости приобретения товаров клиентами и управление автоматизации процессов пополнения складских запасов
- Выявления причин простоя торгового оборудования и касс путем анализа информации из технических журналов

Пример: увеличение продаж в сфере ритейла и услуг достигается путем разработки персональных предложений и акций программ лояльности, именно для тех товаров,

которые нужны пользователям. Повышение лояльности клиентов достигается за счет анализа данных о продажах и транзакциях, отчетности об ассортименте и складских запасах, а также сбора информации из соцсетей о предпочтениях и статистики поведения покупателей на сайте.

Все данные для выполнения поставленных выше задач зачастую не структурированы или частично структурированы и требуется их преобразование в структурированную форму для последующей обработки, их объем постоянно увеличивается, такие данные называют **большими данными**.

Как правило, при обработке больших данных компании сталкиваются с большим их объемом, чем можно уместить на одной Машине — более нескольких терабайт данных. Для решения подобных задач используются распределенные кластерные технологии.

В таких случаях рекомендуется применять систему управления большими данными Hadoop, это одно из решений для хранения и анализа больших данных на основе продукта ADH, реализованный в виде **Машины больших данных Скала^р МБД.Х**, адаптированный для корпоративного использования, предназначенный для хранения и обработки слабоструктурированных и неструктурированных данных.

Система оптимизирована для работы на виртуальных машинах и удобно масштабируется.

Сценарии применения

Машина больших данных Скала^р МБД.Х на основе ADH применяется в следующих сценариях:

Хранение и обработка больших объемов неструктурированных или слабо структурированных данных

Массивы объемом от одного терабайта сложно и дорого хранить в локальном хранилище. Машина справляется с этой задачей гораздо эффективнее. Также особенно эффективно происходит пакетная обработка, обогащение данных и извлечение информации из внешних источников, ее переработка и очистка с последующей загрузкой в базу данных.

Создание «озер данных»

Машина больших данных Скала^р МБД.Х применяется для формирования «Озер данных» — для сбора информации в виде фалов большого объема в случаях, когда один сервер не справляется и необходима распределенная ее обработка и хранение. Такие данные могут пригодиться в будущем для обучения моделей на основе искусственного интеллекта или для соответствия требованиям законодательства по хранению.

Построение моделей и проверка гипотез

Машина больших данных Скала^р МБД.Х позволяет строить гипотезы и применять их к данным, поступающим из разных источников, в разных форматах. Это позволяет использовать информацию в аналитических системах. При обычном хранении этой информации на локальном хранилище она недоступна для анализа, моделирования, прогнозирования, а значит, бесполезна для компании.

Обработка данных о поведении клиентов

Анализ поведения покупателей позволяет компаниям повысить конверсию страниц, сделать удобнее сайт и корзину интернет-магазина, спрогнозировать, какие товары купят пользователи.

Платформа собирает данные, откуда пользователи приходят на сайт, на какую веб-страницу, по какому поисковому запросу, куда переходят, сколько времени проводят на сайте, что покупают и с каких страниц уходят.

Обработка данных интернета вещей (IoT)

Обработка данных от систем IoT позволяет компаниям сократить расходы, улучшить качество выпускаемой продукции, оптимизировать производство и увеличить продажи.

Фитнес-трекеры или устройства умного дома, передают информацию о местоположении и привычках пользователя, что позволяет подстроить режимы работы и рекомендации под каждого конкретного пользователя.

Анализ данных от IoT в городских экосистемах и промышленности позволяет оптимизировать процесс производства и управления инфраструктурой, мониторинга транспортных средств и грузов, помогает в разработке умных инженерных систем, например, электро-, газо- и водоснабжения.

4. ОТЛИЧИТЕЛЬНЫЕ ЧЕРТЫ

Глубокая интеграция и встречная оптимизация компонентов от платформенного ПО до микроконтроллеров в тесном технологическом сотрудничестве с их разработчиками обеспечивает высочайшую устойчивость создаваемых **Машин Скала^р** позволяет достигать экстремальной производительности и сохранять доступность сервисов и стабильные показатели на предельных нагрузках

- 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 больших данных, которые постоянно пополняются
- Механизмы распределенной обработки и хранения данных позволяют распределять запросы к данным одновременно на все узлы системы для быстрой подготовки результатов
- Управляемость Машин обеспечивается применением единой программной платформы **Скала^р** для управления эксплуатацией и мониторинга, а также использование средств управления от компании Arenadata, позволяет управлять параметрами СУБД, сервисами и приложениями входящими в состав комплекса

5. ПОДТВЕРЖДЕННАЯ БЕЗОПАСНОСТЬ

Машина больших данных Скала^Ар МБД.Х поставляется с сертифицированной ОС Альт СП (сертификат ФСТЭК 3866 от 10.08.2018, действует до 10.08.2028), которая:

Может применяться для защиты информации:

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

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, приказ № 118) по 4 классу защиты
- «Требования по безопасности информации к средствам виртуализации» (ФСТЭК России, 2022, приказ № 187) 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, приказ № 76) по 4 уровню доверия

Протестирована совместимость с наложенными средствами защиты:

Сертифицированное антивирусное средство защиты Kaspersky Endpoint Security для Linux (сертификат ФСТЭК 2534 от 27.12.2011, действует до 27.12.2025):

- «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 второго класса защиты. ИТ.САВЗ.Б2.13» (ФСТЭК России, 2012), «Профиль защиты средств антивирусной защиты типа В второго класса защиты. ИТ.САВЗ.В2.ПЗ» (ФСТЭК России, 2012), «Профиль защиты средств антивирусной защиты типа Г второго класса защиты.

Сертифицированное средство доверенной загрузки ПК «Соболь»
версия 4:

- Подтверждает соответствие требованиям руководящих документов к средствам доверенной загрузки, а также 2 уровню доверия средств технической защиты безопасности и обеспечения безопасности информационных технологий и возможность использования в ИСПДн до УЗ1 включительно, в ГИС до 1-го класса защищенности включительно и в ЗОКИИ до 1 категории включительно.

6. ТЕХНОЛОГИЧЕСКИЕ ПРЕИМУЩЕСТВА

Для обеспечения отказоустойчивости и высокой производительности при проектировании программно-аппаратного комплекса были заложены технологические принципы и применен ряд технических решений, описанных ниже.

Технологические принципы

- Дублирование критичных элементов
- Применение высокопроизводительных компонентов
- Горизонтальное масштабирование вычислительных ресурсов
- Сохранение работоспособности при отказе отдельных элементов системы (в отдельных случаях — со снижением производительности)

Технические решения

- Блочно-модульная архитектура
- Специальное ПО управления и мониторинга
- Глубокая адаптация компонентов для совместной работы в составе продукта
- Многоуровневое тестирование комплекса и его узлов и компонентов при производстве для исключения отказов

Архитектура **Машины больших данных Скала^{Ар} МБД.Х** базируется на следующих принципах:

Спрогнозированная нагрузка

- Производительность можно выбирать, подбирая нужное количество Модулей вычисления и хранения
- Все серверы взаимодействуют между собой с одинаковой скоростью

Выделенная сеть внутреннего взаимодействия

- Высокоскоростная сеть внутреннего взаимодействия, обеспечивающая высокоскоростной обмен данными между узлами, показана на рисунке (Рис. 1)
- Параллельная обработка запросов на узлах обеспечивает суммирование мощностей всех узлов участвующих в обработке запроса

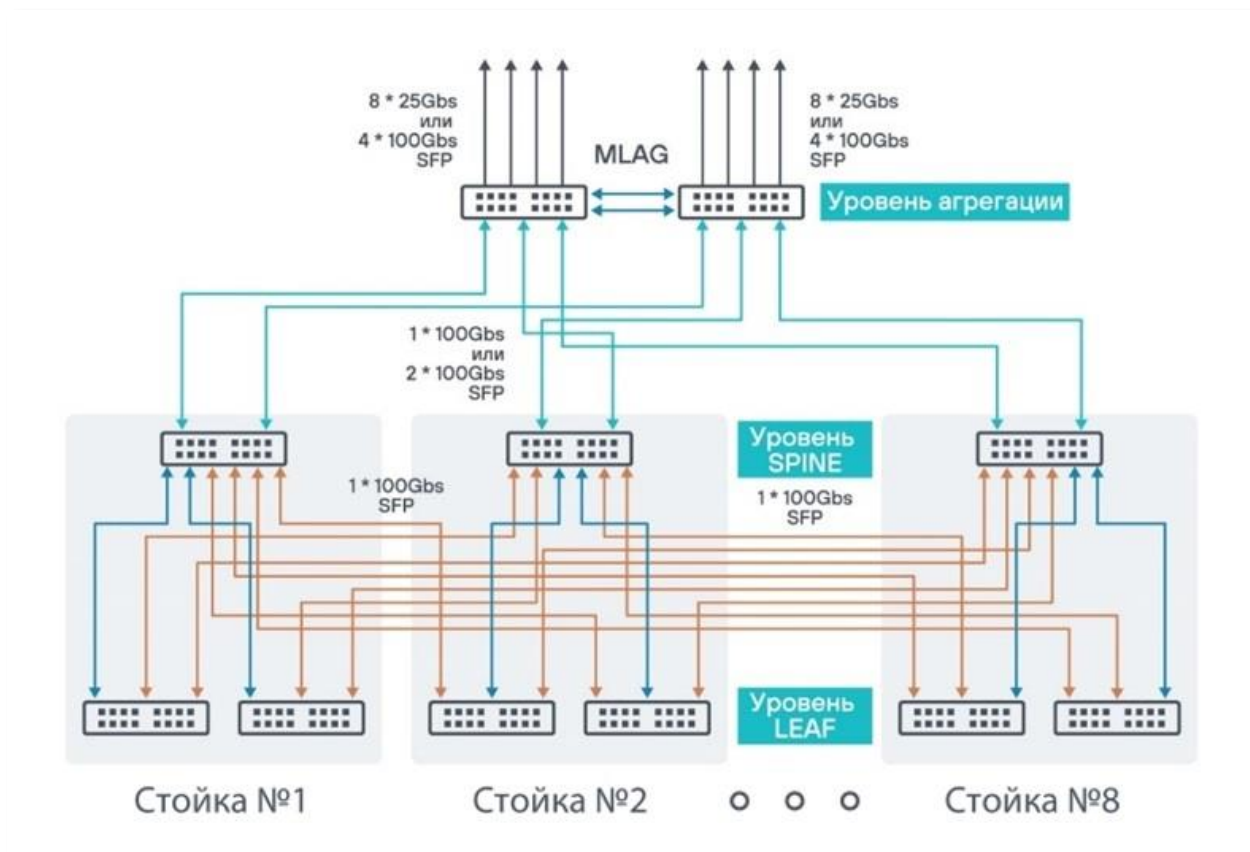


Рис. 1 Схема внутренних соединений Leaf-Spine при горизонтальном масштабировании

Архитектура Машины

Модульный принцип построения **Машины** позволяет адаптироваться под любую нагрузку.

В Базовый модуль входят узлы-коммутаторы (внешнего доступа, внутреннего взаимодействия и управления) и узел мониторинга.

Модуль управления и распределения включает в себя 3 ведущих узла.

Типовой Модуль вычисления и хранения состоит из 2 или 3 узлов данных в зависимости от выбранного типа.

В **Машине** также доступны расширения типовыми Модулями:

- Модули вычисления и хранения
- Модули резервного копирования
- Модули граничных функций

Отказоустойчивость

Построенная на трех узлах отказоустойчивая архитектура (Рис. 2) позволяет гарантировать постоянную доступность сервисов.

Узлы Zookeeper обеспечивают координацию кластера, отслеживая состояние узлов имен, операционных систем и оборудования.

Узлы имен (NameNodes) хранят метаданные файловой системы и метаинформацию о распределении блоков, а также список узлов данных. Они работают в режиме «активный — резервный» (Active NameNode и Standby NameNode).

Узлы данных (DataNodes) отвечают на запросы для хранения информации и объектов. Они непосредственно хранят блоки файлов и обрабатывают операции по чтению данных.

Распределенный журнал изменений ведется на журнальных узлах (JournalNodes) — необходим для синхронизации активного и резервного узлов имен.

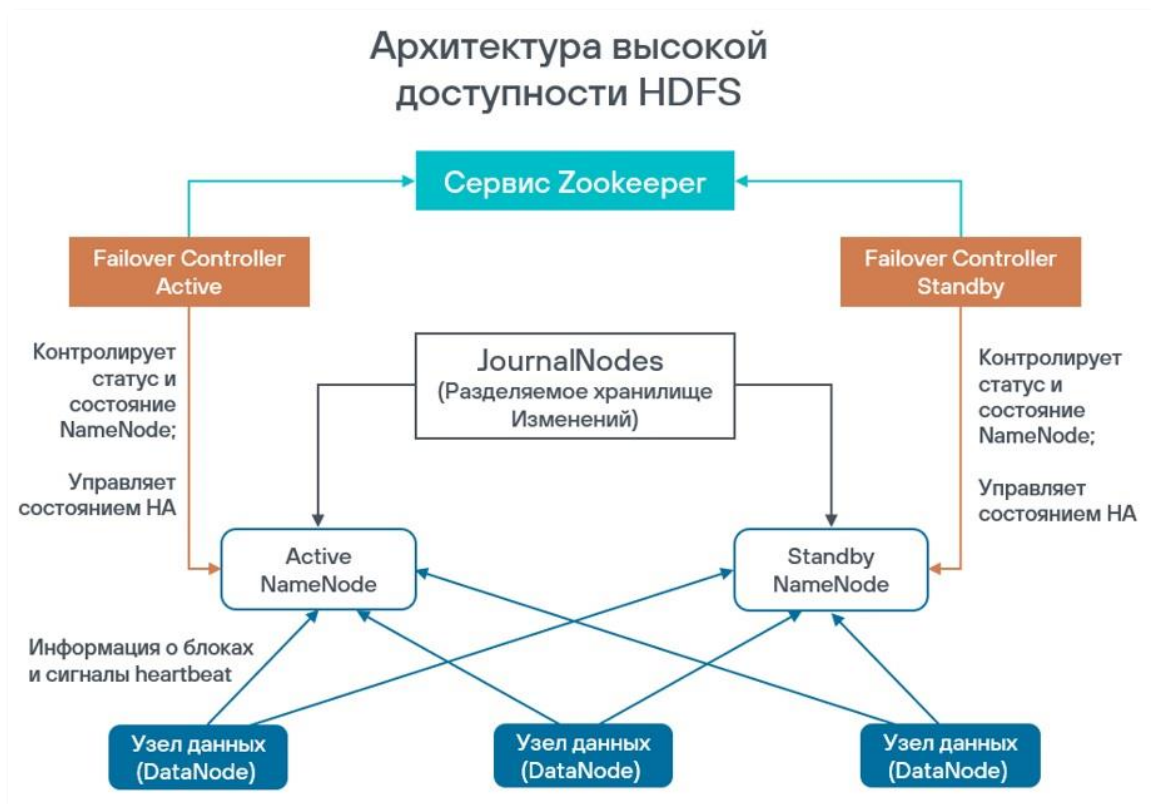


Рис. 2 Схема архитектуры высокой доступности HDFS

Отказоустойчивая архитектура обеспечивает доступность данных за счет нескольких ключевых особенностей. Архитектура построена без общих ресурсов (shared nothing), что также позволяет производить вычисления параллельно для разных блоков на разных узлах (Рис. 3).

HDFS (Hadoop Distributed File System)

Файловая система, предназначенная для хранения файлов больших размеров, поблочно распределенных между узлами вычислительного кластера. Каждый узел данных взаимодействует только со своими внутренними накопителями. HDFS-блоки реплицируются 3 раза, и репликация выполняется в асинхронном режиме. С целью повышения надежности для хранения 2-й и 3-е реплики выбираются те узлы данных, которые расположены в разных серверных стойках. Проверка целостности данных находится в зоне ответственности приложения клиента. При создании файла рекомендуется рассчитывать контрольные суммы каждые 512 байт и сохранять их на узле имен.

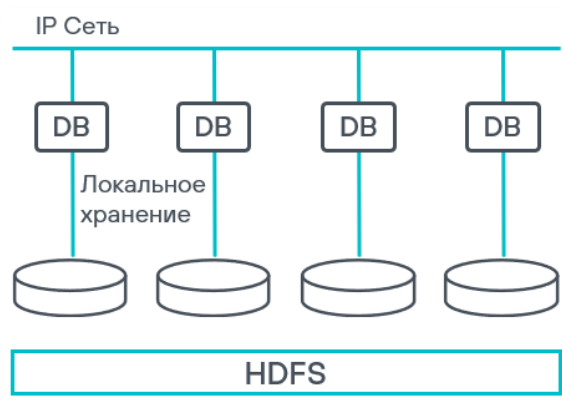


Рис. 3 Архитектура без общих ресурсов (shared nothing)

На Рис. 4 показана архитектура HDFS. Стойки — логическое объединение узлов данных согласно сетевой топологии кластера.

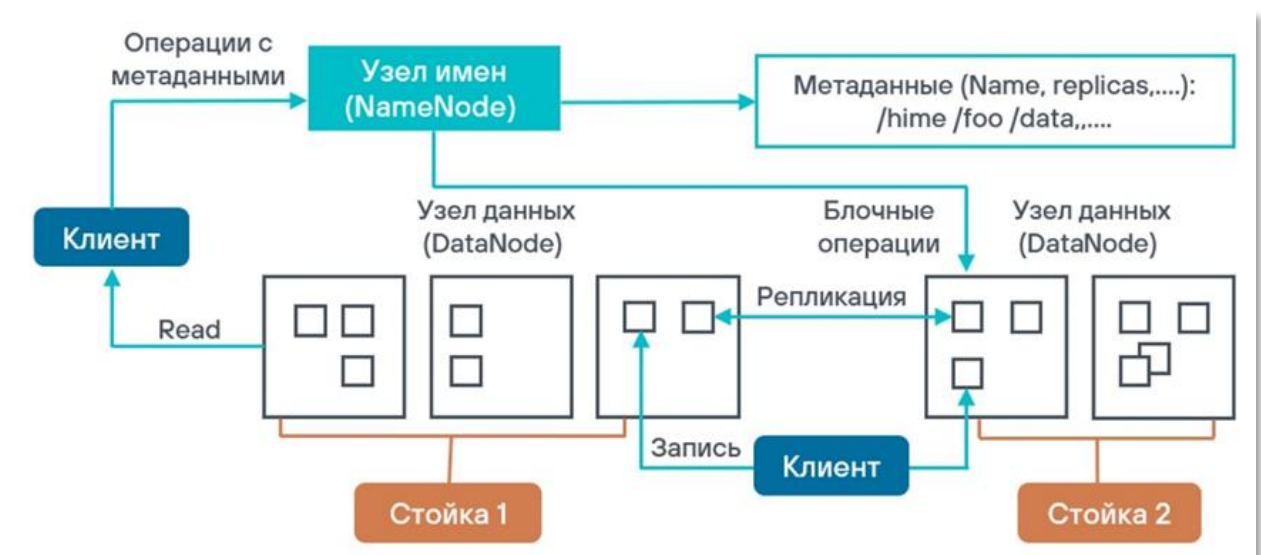


Рис. 4 Архитектура HDFS

Репликация данных

Репликация позволяет гарантировать защиту от одиночных сбоев. Схема репликации показана ниже (Рис. 5). 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
- Первая реплика размещается на первом узле из списка
- Вторая реплика копируется на другой узел в этой же серверной стойке
- Третья реплика записывается на произвольный узел в другой серверной стойке
- Остальные реплики размещаются произвольным способом

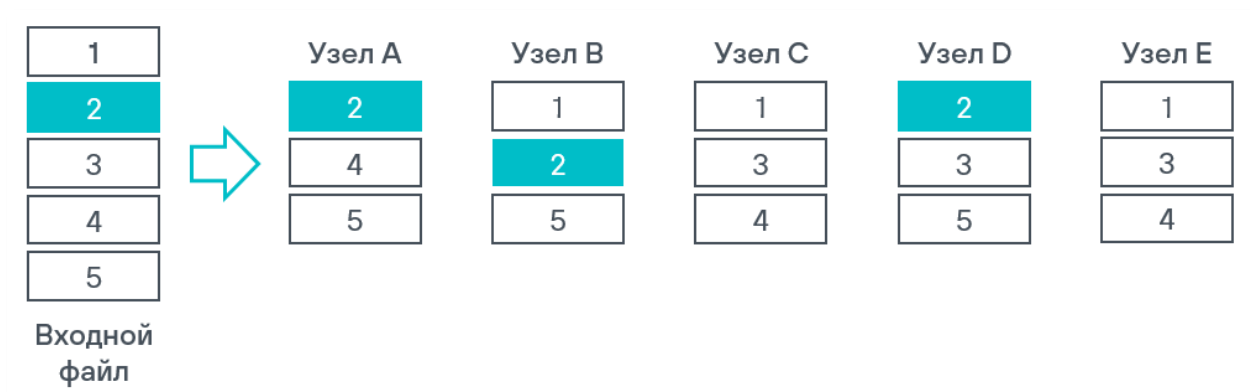


Рис. 5 Схема распределения данных на HDFS

7. СОСТАВ РЕШЕНИЯ

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 сервиса с заданными характеристиками.

Блок — группа Модулей, выполняющих единую функцию в одной или нескольких стойках.

Модуль — это единица поставки **Машин Скала^р**, выполняющих определенные функции в соответствии с их назначением. Он является единым и неделимым элементом спецификации, содержит набор аппаратных узлов и программного обеспечения (ПО).

Узел — это элемент Модуля, выполняющий определенную задачу в составе Модуля.

Секция (Стойка) — набор функциональных блоков модульной архитектуры **Машин Скала^р**, объединенных в один серверный шкаф.

Формирование решения основано на принципе разделения на блоки и Модули. Каждый из блоков комплектуется из набора стандартных модулей. Этим обеспечивается универсальный подход, более высокий уровень технологичности и надежности эксплуатации. Модули, в свою очередь, формируются из одного или нескольких узлов.

Машина Скала^р МБД.Х состоит из следующих блоков (Рис. 6):

Блок коммутации и агрегации

Блок мониторинга и регистрации

Блок управления и распределения

Блок вычисления и хранения

Блок резервного копирования

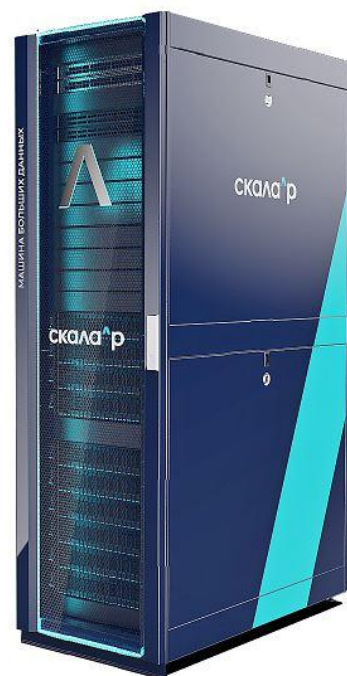


Рис. 6 Блоки в составе Машины больших данных Скала^р МБД.Х

Общая структура Машины

Структура Машины представлена на Рис. 7.

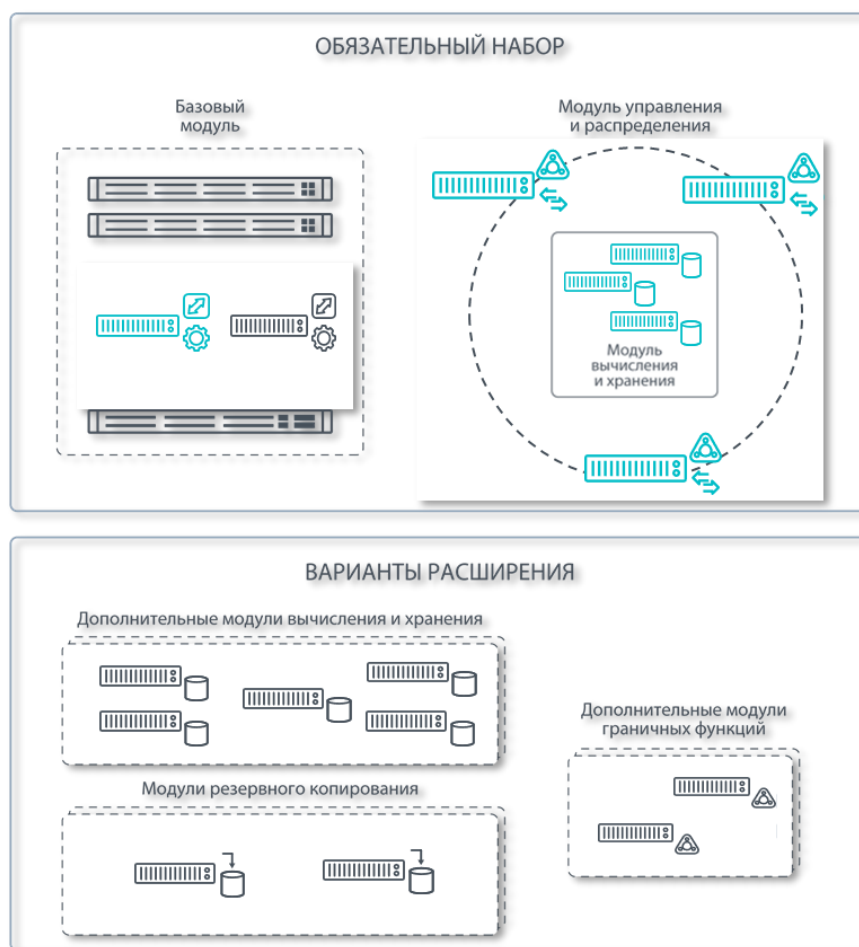


Рис. 7 Принципиальная схема Машины МБД.Х

Обязательный набор состоит из Базового модуля, состоящего из Блока коммутации и агрегации и Блока мониторинга и регистрации, трехузлового Модуля управления и распределения и Модуля вычисления и хранения. Дополнительно Базовый модуль может быть расширен дополнительными Модулями вычисления и хранения, Модулями резервного копирования и Модулями граничных функций.

Базовый модуль

Базовый модуль — основа любой Машины. Может содержать все виды модулей и узлов (Рис. 8), способен доукомплектовываться отдельно стоящими секциями (стойками) расширения. Изначально комплектуется из узлов сетевого взаимодействия, мониторинга и регистрации, узла сетевого управления.

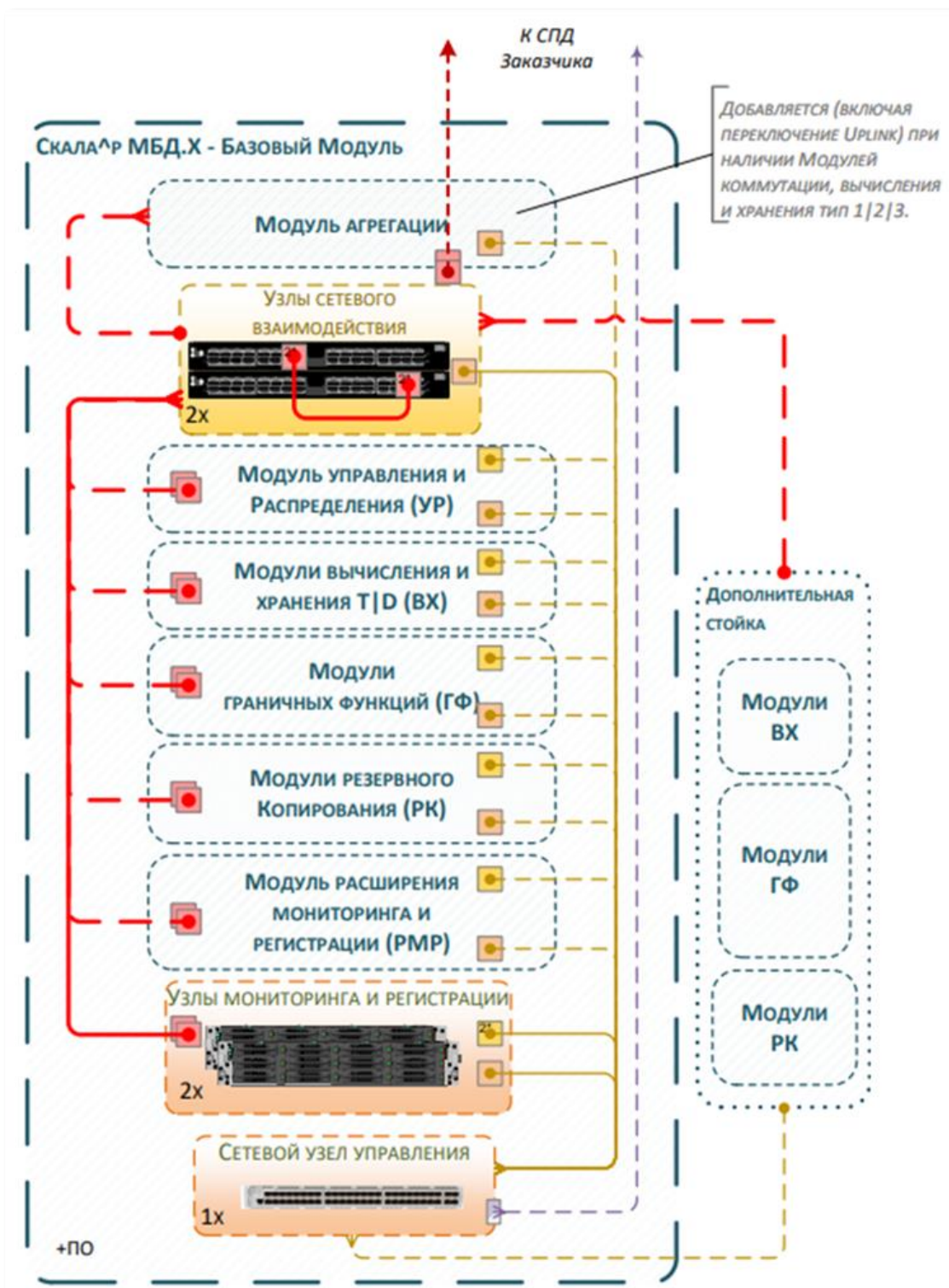


Рис. 8 Базовый модуль

Блок коммутации и агрегации

Обеспечивает передачу данных между элементами **Машины больших данных Скала^{Ар} МБД.Х** (интерконнект) и информационный обмен с внешними сетями, а также низкоуровневое управление узлами Машины. Соответственно эффективное сетевое

взаимодействие является важным фактором для быстрого и надежного функционирования кластера.

Схема сетевого взаимодействия для многоузловых ПАК по архитектуре Leaf-Spine представлена на Рис. 9. Для небольших ПАК может использоваться простая агрегация L2 сетей.

Реализованные подсети:

- Internal VLAN — сеть интерконнекта, для внутреннего взаимодействия между узлами БД, сеть резервного копирования, сеть кластерного взаимодействия на схеме ниже обозначена стрелками оранжевого цвета
- External VLAN — сеть для подключения к сервисам БД внешних пользователей и прикладных систем, подключение к серверу управления на схеме ниже обозначена стрелками черного цвета
- Сети мониторинга и управления — обмен служебными данными, данными для мониторинга и сеть управления узлами Машины на схеме ниже обозначена стрелками синего цвета
- Стрелками зеленого цвета обозначены соединения коммутаторов между стойками с агрегированным трафиком

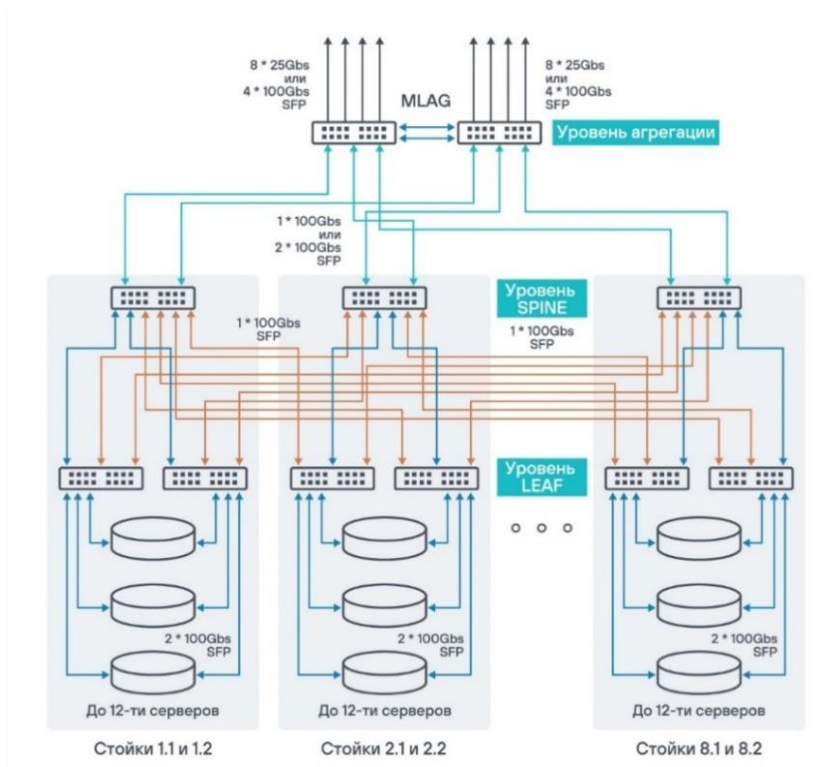


Рис. 9 Общая схема сетевого взаимодействия в случае крупных комплексов и Leaf-Spine

Сети мониторинга и управления Машины:

На рисунке Рис. 10 представлена архитектура сети управления:

- PXE (OS) VLAN — сеть для развертывания операционной системы по PXE, платформы МБД, мониторинга (оранжевого цвета)

- Агрегация (Ring VLAN) — резервная сеть кластерного взаимодействия, доступ к IPMI (зеленого цвета)
- IPMI VLAN — сеть управления оборудованием через интерфейсы удаленного управления (синего цвета)

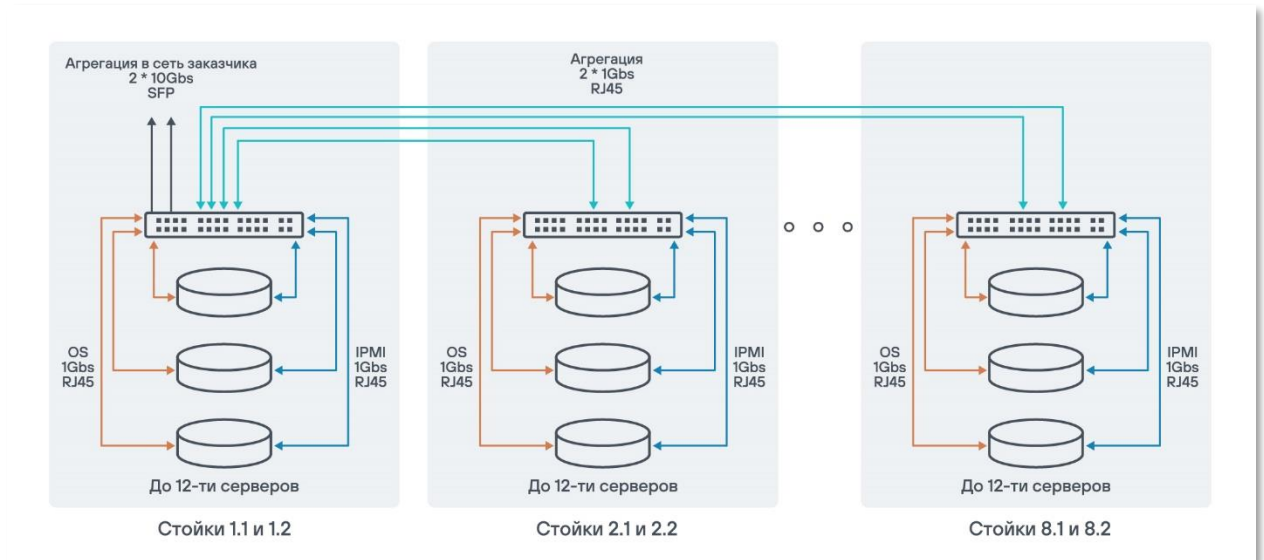


Рис. 10 Общая схема сетей мониторинга и управления

Модули и состав оборудования Блока коммутации и агрегации

Сетевой узел управления

Сетевой узел управления (Рис. 11) состоит из коммутатора для организации сетей мониторинга, управления и служебного обмена.

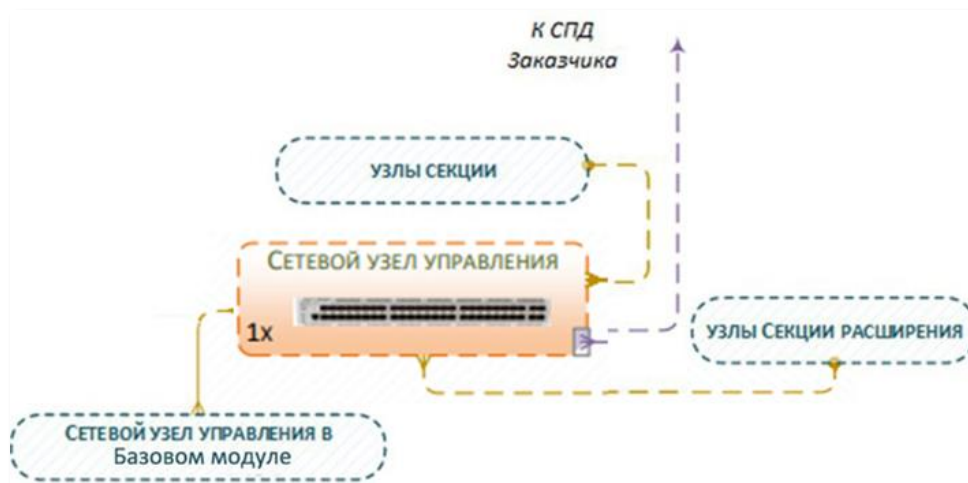


Рис. 11 Сетевой узел управления

Комплект узлов сетевого взаимодействия

Комплект узлов сетевого взаимодействия (Рис. 12) — виртуальный коммутатор, объединенный по технологии MLAG из двух аппаратных коммутаторов для сетевого взаимодействия и клиентского доступа.

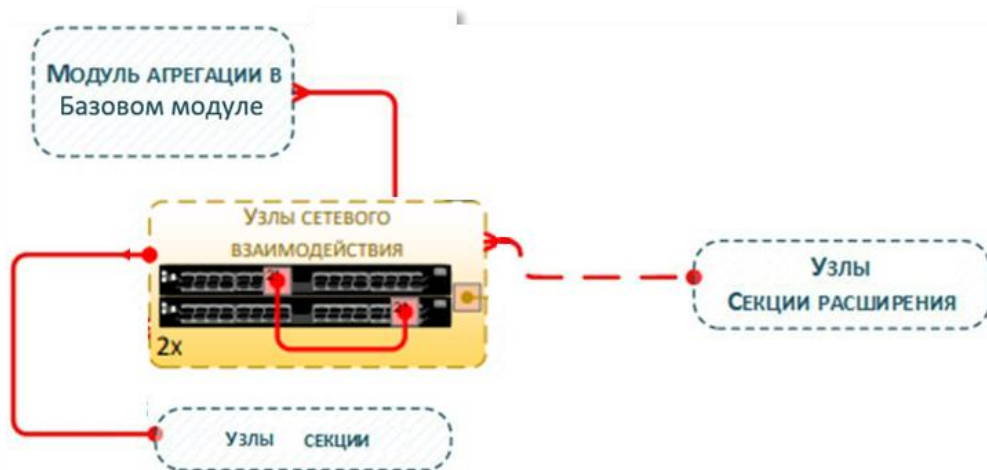


Рис. 12 Модуль сетевого взаимодействия

Модуль агрегации

Модуль агрегации — пара коммутаторов, собранных по технологии MLAG для агрегации коммутаторов интерконнекта.

Модуль агрегации устанавливается в Базовый модуль в тех случаях, когда уже нет доступных портов в базовом комплекте узлов сетевого взаимодействия.

Модуль состоит из узлов сетевого взаимодействия (Рис. 13).



Рис. 13 Модуль агрегации

Блок мониторинга и регистрации

Блок мониторинга и регистрации состоит из двух или трех узлов вычисления и хранения — высокопроизводительных серверов. Он обеспечивает управление программными и аппаратными компонентами.

Основной функционал реализуется с помощью ПО Скала^р Геном и Скала^р Визион:

- Контроль развертывания компонентов **Машины**

- Мониторинг и визуализация работы сети и оборудования, входящего в состав **Машины**
- Накопление данных о функционировании **Машины** для автоматизированной и/или ручной оптимизации настроек аппаратной и программной платформы
- Автоматизированное реагирование на неблагоприятные события и отклонения параметров функционирования **Машины**
- Снижение влияния человеческого фактора — сокращение рисков, связанных с ошибками эксплуатирующего персонала

Блок мониторинга и регистрации (Рис. 14) состоит из двух высокопроизводительных узлов, объединенных в зеркальный кластер, входящих в состав Базового модуля.



Рис. 14 Узлы мониторинга и регистрации

Каждый узел мониторинга и регистрации — специализированный сервер:

- Использует SSD накопители для обеспечения высокой производительности при хранении служебных данных
- Содержит выделенные SSD накопители для загрузки ОС
- Оснащен двухпортовыми сетевыми картами 100 Gigabit Ethernet для интерконнекта в рамках Машины
- Имеет 2 порта 1 Gigabit Ethernet для сетей управления и IPMI
- Содержит два блока питания в режиме резервирования по схеме (1 + 1)
- Имеет два процессора Xeon не ниже 3-го поколения
- Использует от 256 Гбайт ОЗУ

Применяемое программное обеспечение:

- ОС: Альт 8 СП
- Мониторинг и оповещение: **ПО Скала^Ар Визион**
- Управление жизненным циклом и поддержка администрирования комплекса: **ПО Скала^Ар Геном**
- Программное обеспечение Arenadata Cluster Manager (ADCM)

Блок управления и распределения

Модуль управления и распределения

Три узла для создания обязательной обвязки Hadoop: основная и резервная именные ноды, а также нода менеджера заданий. Модуль обеспечивает поддержание отказоустойчивого кластера средствами Zookeeper. Модуль состоит из трех узлов (Рис. 15). Модуль управления и распределения обязателен в составе Машины, но не более одного на **Машину**.



Рис. 15 Модуль управления и распределения

Каждый узел управления и распределения:

- Использует SSD для обеспечения высокой производительности при хранении служебных данных
- Содержит выделенные SSD для загрузки ОС
- Оснащен двухпортовыми сетевыми картами 100 Gigabit Ethernet для интерконнекта в рамках Машины
- Имеет 2 порта 10 Gigabit Ethernet для сетей управления и IPMI
- Содержит два блока питания в режиме резервирования по схеме 1+1
- Имеет два процессора Xeon не ниже 3-го поколения
- Использует не менее 512 Гбайт ОЗУ в узле 1-го типа и от 256 Гбайт ОЗУ в узле 2-го типа

Модуль граничных функций

Модуль является опциональным, состоит из двух узлов (Рис. 16).



Рис. 16 Модуль граничных функций

Модуль обеспечивает функции исполнения программного обеспечения системы **Машины** в отказоустойчивом режиме. Основное его предназначение — обеспечение доступа и запуск специализированных приложений, таких как Spark, Flink и другого ПО из состава ADH.

Модуль коммутации, вычисления и хранения

Модуль комплектуется в отдельную стойку и применяется для горизонтального масштабирования Машины, когда полностью заполнены Базовый модуль и его стойки расширения. В свою очередь этот Модуль также может оснащаться дополнительными стойками расширения. Исходно комплектуется узлами сетевого взаимодействия и управления, и наполняется как минимум одним из следующих Модулей (Рис. 17):

- Модули вычисления и хранения
- Модули резервного копирования
- Модули граничных функций

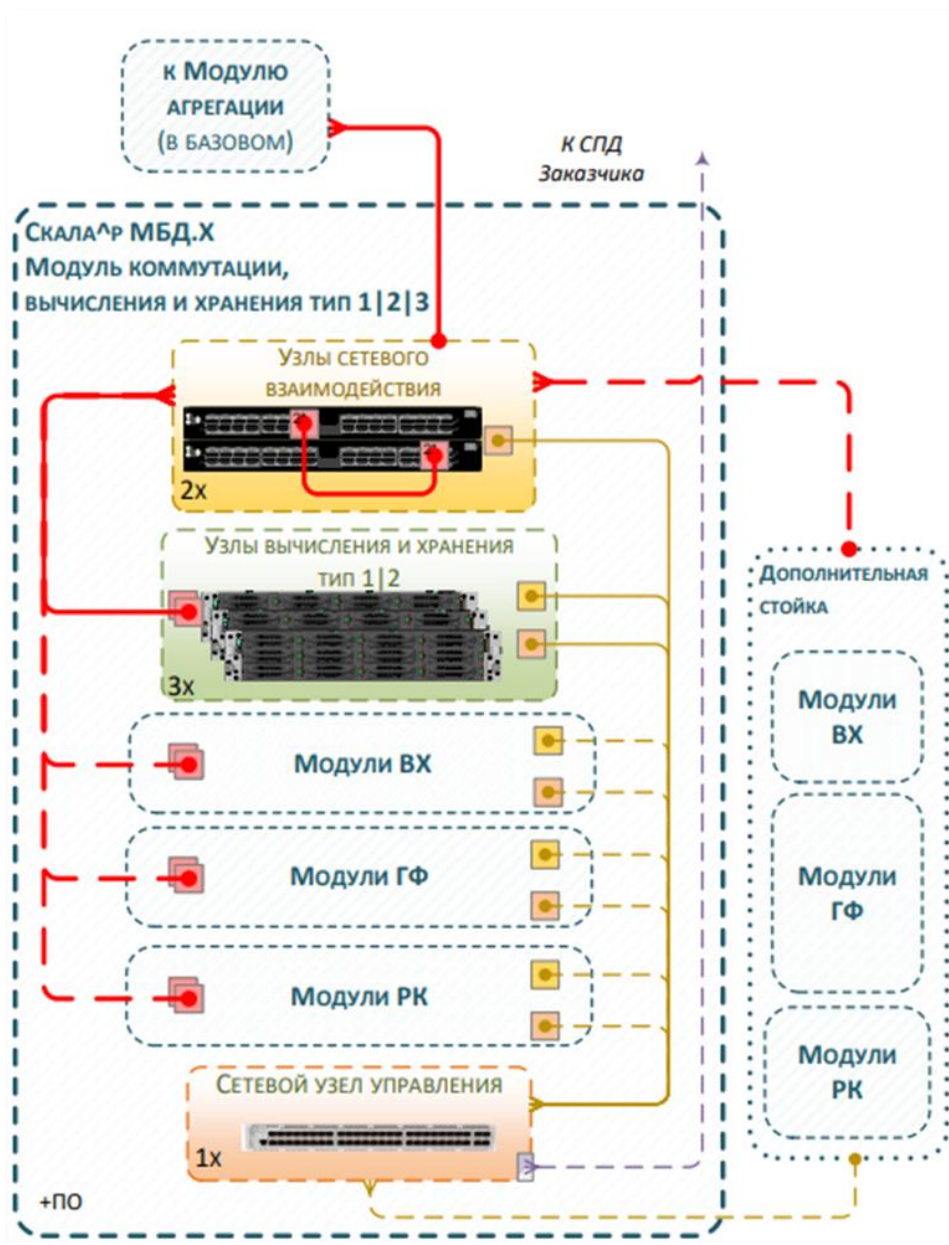


Рис. 17 Модуль коммутации, вычисления и хранения

Секция расширения (отдельно стоящая стойка)

Секция расширения предназначена для размещения дополнительных модулей в дополнение к Базовому модулю и/или Модулю коммутации, вычисления и хранения и содержит следующие виды Модулей:

- Модули вычисления и хранения
- Модули резервного копирования
- Модуль граничных функций

Применение дополнительных стоек обуславливается требованиями и ограничениями инженерной инфраструктуры Заказчика, в том числе — по допустимой потребляемой

мощности и допустимому тепловыделению на отдельный серверный монтажный шкаф (стойку).

К любому основному модулю можно добавить одну или две дополнительных секции (стойки).

Масштабирование Машин больших данных Скала^Ар по этапам в коммутационных стойках

На диаграмме ниже (Рис. 18) представлен пример общих принципов масштабирования.

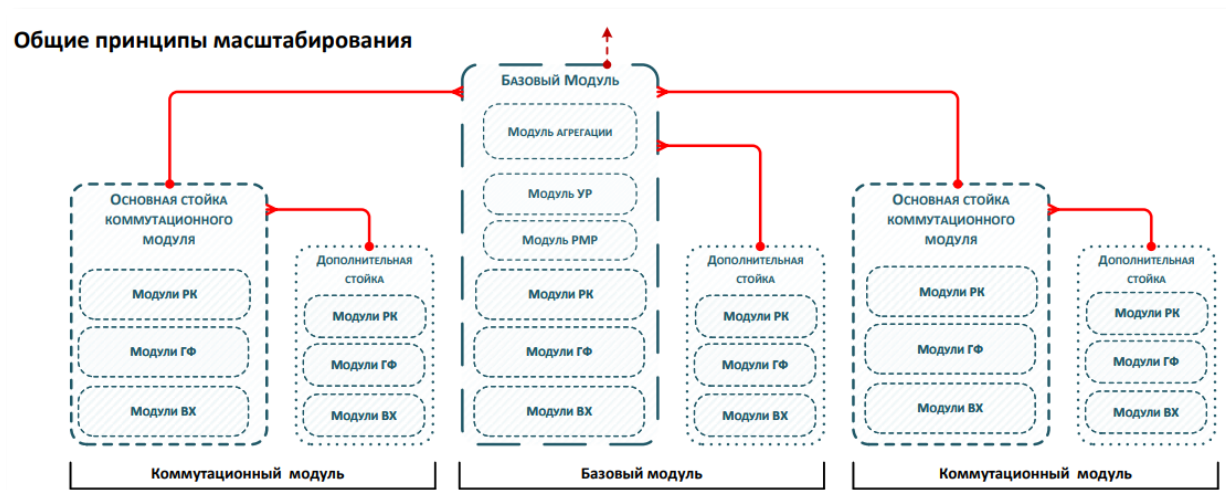


Рис. 18 Пример размещения базовых блоков и вариантов расширения

Горизонтальное масштабирование проводится в несколько этапов:

- Первый этап — это установка модулей в дополнительную стойку после того, как заполнен Базовый модуль. Модули устанавливаются, пока позволяет предельная мощность энергопотребления стойки и имеются порты в узлах сетевого взаимодействия (коммутаторах)
- Установка первой стойки Модуля коммутации, вычисления и хранения выполняется при достижении ограничения количества портов внутреннего взаимодействия в Базовом модуле. Стойка включает в себя набор узлов сетевого взаимодействия и требует установки дополнительного Модуля агрегации сетей в Базовый модуль
- Далее Модуль коммутации может быть расширен дополнительной стойкой, а также **Машина** может расширяться дополнительными Модулями коммутации

Блок вычисления и хранения

Блок вычисления и хранения состоит из типовых Модулей вычисления и хранения на базе аппаратных узлов и программной платформы ADH.

Образуемый узлами модулей кластер ADH не имеет мастер-узла и единой точки входа. Запросы могут отправляться к любому из узлов в кластере. Архитектура кластера представлена на рисунке ниже (Рис. 19). Модуль соединен с Модулем сетевого взаимодействия и сетевым узлом управления.



Рис. 19 Модуль вычисления и хранения

Каждый отдельный узел вычисления и хранения:

- Содержит выделенные накопители SSD для загрузки ОС
- Использует локальные SSD накопители для размещения образов данных и журналов (RAID 10), что обеспечивает повышение производительности (нет необходимости дополнительного внешнего обмена с системой хранения)
- Оснащен двухпортовыми сетевыми картами 100 Gigabit Ethernet для интерконнекта в рамках **Машины**
- Имеет дублированные интерфейсы данных (стандарт IEEE 802.3ad LACP) — повышение производительности, отказоустойчивость (в случае отказа одного из интерфейсов возможно снижение производительности)
- Содержит два блока питания в режиме резервирования по схеме (1 + 1)
- Имеет два процессора Xeon не ниже 3-го поколения
- Использует не менее 512 Гбайт ОЗУ в узле 1-го типа и от 384 Гбайт ОЗУ в узле 2-го типа

Применяемое программное обеспечение:

- ОС: Альт 8 СП
- Arenadata Hadoop (ADH)

Блок резервного копирования

Основное предназначение блока — запись и хранение резервных копий базы данных, а также их восстановление. Блок собирается из Модулей резервного копирования. Каждый такой Модуль состоит из одного узла резервного копирования (Рис. 20).



Рис. 20 Модуль резервного копирования

Каждый отдельный узел резервного копирования:

- Содержит выделенные накопители SSD для загрузки ОС — обеспечение отказоустойчивости
- Применяется аппаратный RAID
- Оснащен 14 HDD для хранения данных
- Оснащен двухпортовыми сетевыми картами 100 Gigabit Ethernet для интерконнекта в рамках **Машины**
- Имеет дублированные интерфейсы данных (стандарт IEEE 802.3ad LACP) — повышение производительности, отказоустойчивость (в случае отказа одного из интерфейсов возможно снижение производительности)
- Содержит два блока питания в режиме резервирования по схеме 1+1
- Имеет два процессора Xeon не ниже 3-го поколения
- Использует от 384 Гбайт ОЗУ

Применяемое программное обеспечение:

- Для создания резервных копий используются встроенные средства резервного копирования ADH.
- Резервное копирование рассчитывается для объема от 4-х копий максимальной базы.

Система управления жизненным циклом Скала^р Генom

В **Машине больших данных Скала^р МБД.Х** применяется специализированное программное обеспечение **Скала^р Генom**.

Данный программный продукт обеспечивает:

- Автоматизацию обновлений и восстановления компонентов **Машины**

- Ведение электронного паспорта **Машины**
- Отслеживание состояния узлов
- Отслеживание конфигурации программно-аппаратного состава **Машины**
- Снижение влияния человеческого фактора — сокращение рисков, связанных с ошибками эксплуатирующего персонала

ПО **Скала^р Геном** устанавливается на узлы Модуля мониторинга и регистрации.

Система мониторинга и оповещения Скала^р Визион

В **Машине больших данных Скала^р МБД.Х** применяется специализированное программное обеспечение **Скала^р Визион**.

Данный программный продукт обеспечивает:

- Сбор данных о конфигурации элементов ПАК (**Машин**)
- Конфигурирование метрик мониторинга, настройку предупреждений
- Хранение метрик с возможностью настройки глубины хранения и управления жизненным циклом хранимых данных
- Мониторинг изменений объектов контроля в режиме, близком к реальному времени
- Установку информационных панелей для каждого типа ПАК

8. СПЕЦИФИЧНЫЕ ЧЕРТЫ

Проектирование и реализация **Машины Скала^Ар МБД.Х** осуществлялись с учетом ряда выбранных приоритетов, оказывающих непосредственное влияние на функциональные и эксплуатационные показатели. Наиболее значимые из них следующие —

Возможная интеграция с любыми источниками информации

- Доступна интеграция с такими источниками информации, как:
- OLTP, ERP, CRM
- Документы и почтовые сообщения
- Журналы веб-серверов, потоки посещений
- Данные социальных сетей
- Журналы промышленных систем
- Данные сенсоров и датчиков

Гарантия скорости обработки данных

В **Машине** используются параллельные и распределенные вычисления. Высокая скорость обработки данных обеспечивается применением высокопроизводительного оборудования: быстрых дисков SSD, высокоскоростных сетей внутреннего взаимодействия (100 Гбит/с), 48+ CPU на узел, достаточного объема памяти на узел.

Выполнение бизнес-задач на неограниченном объеме данных

Машина позволяет проводить анализ постоянно пребывающих больших данных, чей объем измеряется в ТБ, преобразовывать неструктурированных и частично структурированные данные в структурированную форму и работать с данными типа «ключ-значение».

Горизонтальное масштабирование

Горизонтальное масштабирование позволяет обеспечить рост системы в зависимости от потребностей бизнеса. Упор делается на простое наращивание вычислительной мощности и полезной емкости системы. Подобный подход позволяет расширять возможности существующих аналитических систем и обеспечить предельную загрузку оборудования.

9. ГАРАНТИРОВАННОЕ КАЧЕСТВО

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

Производство (комплектование и развёртывание ПО)

- При производстве используются высококачественные комплектующие
- Сборка продукции осуществляется строго в соответствии с утверждённым планом размещения компонентов
- Первичное развёртывание ПО осуществляется в автоматическом режиме
- Дополнительные настройки ПО осуществляются в соответствии с утверждённой методикой и пошаговой инструкцией
- Осуществляется функциональное тестирование сформированной **Машины**
- Отклонения от типового решения **Машины больших данных Скала^р МБД.Х** исключены

Передача в эксплуатацию

- **Машина больших данных Скала^р МБД.Х** полностью сформирована, протестирована, готова к размещению в сети Заказчика и подключению прикладного ПО
- В комплекте с **Машиной больших данных Скала^р МБД.Х** передаются паспорт решения, сертификат на поддержку
- Проводится обучение специалистов Заказчика работе с **Машиной больших данных Скала^р МБД.Х** (по запросу)

10. РЕАКЦИЯ МАШИНЫ НА ВОЗМОЖНЫЕ ОТКАЗЫ

Отказы, связанные со стандартными элементами Скала^р МБД.Х

В рамках **Машины больших данных Скала^р МБД.Х** обеспечена отказоустойчивость её основных элементов и процессов, в том числе:

- узлов (дублирование процессоров, источников питания и др.)
- дисковой подсистемы (RAID)
- Сети интерконнекта (дублирование сетевых интерфейсов)
- 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 **Скала^р МБД.Х** также восстанавливается.

Отказы, связанные с узлами кластера баз данных

Аппаратные сбои

Архитектура **Машин больших данных Скала^р МБД.Х** позволяет построить отказоустойчивый многоузловой кластер, малочувствительный к выходу из строя отдельных узлов в различных Модулях.

11. ВАРИАТИВНОСТЬ РЕШЕНИЯ

Приоритет производительности

Область применения:

- Data Warehouse как единое корпоративное хранилище архивных данных из разных источников
- информация оперативного доступа
- множественные аналитические запросы бизнеса

Варианты решения:

- увеличенный объём оперативной памяти
- повышение базовой частоты работы процессоров
- высокопроизводительные твердотельные накопители с подключением SAS (16×1,92 Тбайт)
- RAID 10

Приоритет объёма хранения

Область применения:

- база знаний предприятия
- данные исторического анализа

Вариант решения:

- стандартные параметры Вычислительного модуля
- SSD повышенного объёма (16×7,64 Тбайт)

Специальный тюнинг для повышения производительности

Вариант решения:

- может использоваться в комплексе с любым из вариантов
- требуется участие разработчиков прикладных систем
- достигается адаптацией настроек и конфигурации оборудования под структуру данных Заказчика, типы и периодичность запросов и др.

12. ТРЕБОВАНИЯ К РАЗМЕЩЕНИЮ РЕШЕНИЯ

Решение представляет собой серверный монтажный шкаф 19", высота 42U, с дальнейшей возможностью модульной расширяемости до 14 стоек.

Наполнение шкафа оборудованием и совокупный вес зависят от выбранного варианта решения и могут составлять от 400 до 800 кг.

Монтажный шкаф (стойка) может быть поставлена 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 развёртывании решения на нём будут выполнены настройки сетевых адресов в соответствии со структурой сети Заказчика. Заказчик должен предоставить необходимые данные в соответствии с номенклатурой компонентов решения.

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

Дальнейшие мероприятия по вводу в эксплуатацию осуществляются Заказчиком путём настройки прикладных программных систем.

13. ТЕХНИЧЕСКАЯ ПОДДЕРЖКА

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

- Пакет услуг по технической поддержке на первый год включен в поставку
- Заказчик может выбирать пакет в базовом режиме 9×5, или в расширенном режиме 24×7 (опция для 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 представлен в таблице ниже (Таблица 1).

Таблица 1. Пакеты услуг по технической поддержке Машин больших данных Скала^р МБД.Х

Услуги	Пакет «9×5»	Пакет «24×7»
«Обслуживание комплекса Скала^р в режиме 9×5» (в рабочее время по рабочим дням)	+	—
«Обслуживание комплекса Скала^р в режиме 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	+	+

Услуги	Пакет «9×5»	Пакет «24×7»
Расширенные опции обслуживания	—	+
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^{Ар} 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.

14. 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 мажорных и минорных релизов: мажорные релизы выпускаются раз в квартал, минорные релизы выпускаются при необходимости более быстрого введения в эксплуатацию небольших улучшений в системе.

Программное обеспечение СУБД Arenadata Hadoop (ADH) лицензируется согласно объему ресурсов (ядер) в Модулях вычисления и хранения, при этом на каждый Модуль выдается единая лицензия.

Программное обеспечение **Скала^р Визион** и **Скала^р Геном** поставляется исключительно в составе **Машин Скала^р МБД.Х**, и лицензируется по количеству вычислительных узлов.

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 мажорных и минорных релизов: мажорные релизы выпускаются раз в квартал, минорные релизы выпускаются при необходимости более быстрого введения в эксплуатацию небольших улучшений в системе.

О КОМПАНИИ

Скала^р (продукт **Группы Rubytech**) с 2015 года выпускает программно-аппаратные комплексы (**Машины**) и сегодня предлагает 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. Соответствуют критериям доверенности и использованию для объектов критической информационной инфраструктуры (КИИ).

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