



# Машина больших данных Скала^р МБД.С

Масштабируемая и отказоустойчивая Машина потоковой обработки и преобразования данных в реальном времени средствами Arenadata Streaming (AD.S) на основе Kafka и NiFi

## Технический обзор

версия 1.0 от 28.11.2023



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».

## ОГЛАВЛЕНИЕ

1. Введение .....	4
2. Кратко о Поточковой обработке и преобразовании данных .....	5
3. Сценарии применения Машины Скала^р МБД.С .....	7
4. Отличительные черты.....	8
5. Технологические преимущества.....	9
6. Состав решения .....	15
7. Специфичные черты.....	28
8. Гарантированное качество .....	29
9. Реакция на возможные отказы .....	30
10. Требования к размещению решения.....	31
11. Техническая поддержка .....	32
12. Лицензирование ПО Машины больших данных .....	34
О компании .....	35

### ВВЕДЕНИЕ

**Машина больших данных Скала^р МБД.С** — это программно-аппаратный комплекс (ПАК), предназначенный для потоковой обработки и преобразования данных средствами продукта Arenadata Streaming (далее — AD.S), созданного на основе Apache Kafka и NiFi, и расширенный служебным ПО Скала^р для управления и мониторинга комплекса.

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

**Машина больших данных Скала^р МБД.С** — комплексное решение, включающее в себя масштабируемые узлы для проведения вычислений и обработки данных, систему резервного копирования, сверхскоростную сетевую среду, систему интеллектуального управления.

Высокая производительность решения достигается в том числе применением оптимальных по производительности комплектующих, современных накопителей и сетевых протоколов.

Отказоустойчивость обеспечивается применением надежных комплектующих с резервированием критических компонентов, использованием устойчивых сетевых протоколов и специализированной версии ПО AD.S.

**Машина больших данных Скала^р МБД.С** содержит все необходимые элементы для функционирования высоконагруженной системы потоковой обработки и преобразования данных. Подключение к внешним сетям осуществляется с помощью стандартного интерфейса Ethernet.

**Скала^р МБД.С** предполагает использование кластеризации для повышения отказоустойчивости и повышения производительности обработки данных.

В **Машине Скала^р МБД.С** реализованы функции мониторинга состояния как аппаратных, так и программных компонентов решения, а также расширенные функции управления.

Продукция Скала^р включена в Единый реестр российской радиоэлектронной продукции.

Технические средства **Машин и модулей Скала^р** включены в Единый реестр российской радиоэлектронной продукции.

Программные компоненты **Машин Скала^р** включены в Единый реестр российских программ для ЭВМ и баз данных.

## КРАТКО О ПОТОКОВОЙ ОБРАБОТКЕ И ПРЕОБРАЗОВАНИИ ДАННЫХ

### ПОТОКОВАЯ ОБРАБОТКА

Брокер сообщений — центральный компонент для обеспечения взаимодействия в современной ИТ-архитектуре — является связующим интеграционным звеном между множеством слабо связанных систем и компонентов и основным средством реализации масштабной обработки потоковых данных. Применение брокеров сообщений позволяет построить инфраструктуру обработки потоковых данных с применением шаблонов «издатель — подписчик» и «очередь сообщений», обеспечить надежность и бесперебойность работы ИТ-системы в целом с учетом возможных сбоев и изменений во взаимодействующих элементах.

Apache Kafka — основной компонент AD.S — горизонтально масштабируемый программный брокер сообщений, являющийся фактическим стандартом в экосистеме Hadoop для организации высоконадежного асинхронного обмена сообщениями между различными системами и службами. Сообщения, обрабатываемые в Kafka, распределяются по темам (topics), которые, в свою очередь, разбиты на разделы (partitions), и хранятся в брокере упорядоченными в рамках одного раздела. Системы-производители (producers) отправляют сообщения в определенную тему, а брокер сообщений распределяет их по разделам и обеспечивает отказоустойчивое хранение. Процессы систем-потребители (consumers) читают сообщения из заданной темы в порядке их поступления, брокер для каждого потребителя и каждого раздела запоминает смещение (offset) — последнее сообщение, полученное потребителем, что позволяет приостанавливать и возобновлять работу без потери данных. Потребители могут быть объединены в группы (consumer groups), для каждой из которых устанавливается определенный набор разделов темы таким образом, чтобы масштабировать получение сообщений — несколько потребителей получают сообщения только из своих разделов, и несколькими параллельными процессами может вычитываться вся очередь без блокировок и дублирования.

Каждый узел Kafka в кластере называется брокером (broker), при этом система функционирует как единое целое. Один из брокеров автоматически выбирается в качестве управляющего (cluster controller), отвечая за административные операции (такие, как распределение разделов по брокерам и мониторинг отказов). Для каждого раздела система автоматически определяет ведущий брокер (leader), при этом возможно указать коэффициент репликации — на какое количество узлов распространять каждое сообщение, и таким образом регулировать уровень отказоустойчивости в зависимости от политики для каждого раздела.

Одно из типовых применений Kafka — в качестве компонента передачи сообщений в ETL-системах, обеспечивающего высокую пропускную способность, равномерную нагрузку и отказоустойчивость на этапах извлечения и загрузки.

### ПРЕОБРАЗОВАНИЕ ДАННЫХ

Для этапов преобразования и организации ETL-процессов в целом в экосистеме Hadoop наибольшее распространение получил инструмент Apache NiFi, также входящий в пакет AD.S.

NiFi — горизонтально масштабируемая система, интегрирующаяся практически со всеми распространенными системами обработки данных, наилучшим образом дополняемая системой Kafka как средством передачи сообщений. Масштабируемость NiFi позволяет задействовать всю мощь современных СУБД и брокера сообщений Kafka и сокращать окна загрузки до уровней, недостижимых для одноузловых ETL-систем. В то же время работа в кластере дает отказоустойчивость, которая в дополнение к отказоустойчивости на стороне Kafka, позволяет наладить бесперебойную работу в условиях интенсивного преобразования и масштабной загрузки данных.

NiFi оснащен веб-интерфейсом, обеспечивающим настройку, визуализацию и мониторинг процессов извлечения, загрузки и преобразования данных, что его выгодно отличает от других программных систем для оркестровки операций по обработке и передаче данных, приближая его по удобству применения к лучшим коммерческим тиражируемым корпоративным ETL-системам. При этом, будучи свободным программным обеспечением и частью экосистемы Hadoop, изначально создаваемых для сверхмасштабных применений, NiFi намного предпочтительнее большинства коммерческих ETL-систем для применения в условиях многотерабайтных объемов обработки, десятков и сотен узлов и многообразия задействованных разнообразных современных СУБД.

Главный исполнительный компонент NiFi — контроллер потоков (flow controller), управляющий расписаниями и запуском заданий по преобразованию и загрузке. Потоки-расширения (flow extensions) — плагины, обеспечивающие соединяемость с различными системами и исполняющие непосредственные операции по извлечению и загрузке, каждый из потоков-расширений — это отдельный экземпляр виртуальной Java-машины. Три метаинформационных компонента — репозиторий потоков (flowfile repository), репозиторий содержимого (content repository) и репозиторий провенанса (provenance repository) — служат для хранения в кластере информации об информационных пакетах (FlowFiles), их физическом содержимом (content) и сведений о происхождении данных (provenance). Графический компонент — веб-сервер — предоставляет пользовательский интерфейс для администраторов и инженеров данных, в котором разрабатываются и отслеживаются потоки. При этом все функции, доступные из веб-интерфейса, доступны через REST API, таким образом подготовка и мониторинг потоков могут быть реализованы программным способом.

## СЦЕНАРИИ ПРИМЕНЕНИЯ МАШИНЫ СКАЛА^Р МБД.С

**Машина Скала^р МБД.С** используется в процессах взаимодействия между различными системами в режиме, близком к реальному времени, для задач, где важна автоматизация обмена сообщениями и трансформация данных. Машина позволяет обрабатывать потоки запросов от различных систем и хранить данные от 4 ТБ прямо в очереди. Для **Машины Скала^р МБД.С** характерны следующие сценарии применения:

**Потоковая обработка данных** — позволяет решать задачи сбора и обработки сообщений с данными от различных систем-издателей и передавать их для обработки системам-потребителям в оперативном режиме, при этом обеспечивая асинхронность и отказоустойчивость, не нагружая взаимодействующие системы, но при этом используя максимум их производительности за счет высокого уровня параллелизма.

**Решение задач проверки, модификации и транспортировки данных между разнородными системами** — позволяет получать данные из источников в различных форматах, осуществлять приведение к заданному стандарту и передавать в целевую систему.

**Обработка событий для служб мониторинга** — позволяет создавать темы для сбора системной информации, на которые можно подписать функции обслуживания или управления. Это позволяет автоматически обрабатывать события. Например, если сервер сообщает об ошибке, система запускает бизнес-процесс для ремонта или замены оборудования.

**Потоковая сверхоперативная аналитика** — позволяет обмениваться сообщениями с оборудованием и анализировать потоки данных, например, с кассовых терминалов на точках продаж или с торговых автоматов, что позволяет получать оперативную информацию о необходимости пополнить определенные товарные позиции.

**Обработка потоков данных с устройств интернета вещей** — позволяет распространять обновления на все подписанные устройства, а также собирать показания датчиков и рассылать уведомления.

**Агрегация данных из различных источников** — позволяет реализовать функции агрегации и преобразования данных в распределенных приложениях для корпоративных витрин данных.

**Реализация концепции сетки данных (Data Mesh)** — распределенной архитектуры управления данными, подразумевающей децентрализованную организацию без создания единого хранилища данных и озера данных, при этом обеспечивающую оперативный доступ ко всем корпоративным данным за счет интеграции средствами Kafka и NiFi.

**Перемещение резервных копий данных в другие системы** — позволяет использовать сервис «издатель — подписчик» для создания конвейера переноса данных. После окончания резервного копирования в тему отправляется сообщение-триггер для подписанной функции, которая запускает процесс миграции или копирования данных.

## ОТЛИЧИТЕЛЬНЫЕ ЧЕРТЫ

### Преимущества Машины Скала^р МБД.С

#### Скорость ввода-вывода и хранение

Машина больших данных Скала^р МБД.С позволяет обрабатывать в параллельном асинхронном режиме миллионы сообщений в секунду, и обеспечивает надежное промежуточное хранение. Машина обеспечивает хранение существенных объемов сообщений, что необходимо для интеграции высокоинтенсивных ERP и MES систем, систем автоматизации продаж, CRM- и биллинговых систем, автоматизированных банковских систем, процессинговых систем, и других задач широкого круга отраслей.

#### Линейная масштабируемость

Компоненты Машины подобраны и сбалансированы для раскрытия всего потенциала масштабируемости, заложенные в составляющие продукта AD.S, таким образом, чтобы достичь практически любых пропускных способностей и неограниченной глубины хранения при наращивании количества узлов, сохраняя при этом экономическую эффективность и надлежащий уровень эксплуатационного качества.

#### Высокая доступность и отказоустойчивость

Кластер гарантирует доступность системы с учетом требуемого коэффициента репликации. Ни одно сообщение не будет утрачено и все сообщения будут обработаны, так как система сохраняет их при отсутствии активных подписчиков, а при аппаратных сбоях нагрузка переводится на реплики сообщений.



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

### Технические решения

- Блочная-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^р МБД.С** базируется на следующих принципах:

### Программный RAID1

- Производительнее аппаратного RAID-контроллера
- Не подвержен аппаратным сбоям
- Минимальное использование ОЗУ (требуется менее 4 ГБ ОЗУ)
- Минимальная снижение производительности в режиме восстановления RAID

### Спрогнозированная нагрузка

- Распараллеливание нагрузки достигается с помощью сегментирования
- Производительность можно выбирать встраиванием согласованного с задачей движка

### Выделенная сеть внутреннего взаимодействия

- Высокоскоростная сеть внутреннего взаимодействия, ускоряющая обмен данными между узлами, показана на рисунке (Рис. 1)
- Параллельная обработка запросов на узлах приводит к суммированию мощностей всех узлов

---

<sup>1</sup> Опция по умолчанию. Также могут применяться высокопроизводительные аппаратные контроллеры с резервированием.

- Создание параллельной синхронной копии не влияет на выполнение задания
- Все узлы взаимодействуют между собой с одинаковой скоростью

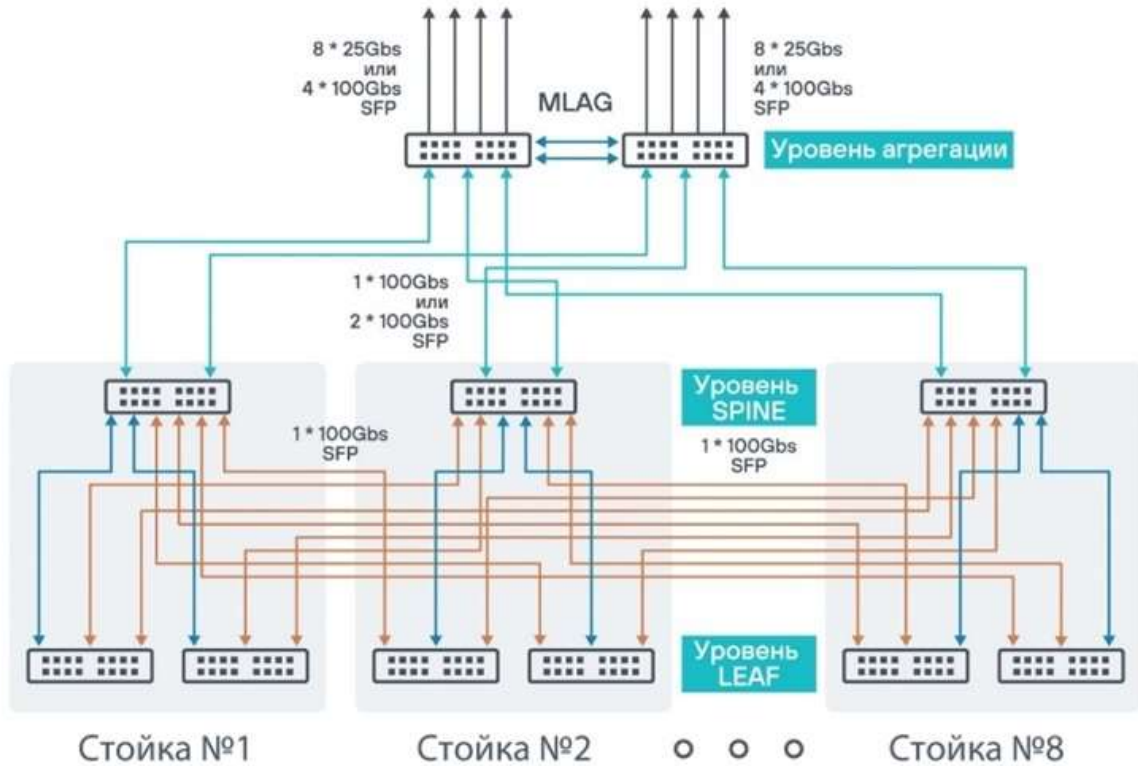


Рис. 1 Схема внутренних соединений Leaf-Spine с увеличением скорости при горизонтальном масштабировании

## Архитектура сервиса очереди сообщений на базе AD.S на основе Kafka

### Отказоустойчивость

В модуле вычисления и хранения отказоустойчивость реализована на уровне кластера Kafka-брокеров. Для хранения и обработки информации используется минимум два сервера в кластере. Использование дополнительных узлов повышает производительность и надежность системы. Координацию выполняет кластер Zookeeper, состоящий из трех узлов в составе модуля управления и распределения, который контролирует и координирует состояние брокеров, квоты, узлы, реплики, смещения.

### Особенности репликации данных

Данные сгруппированы по темам, которые разбиваются на разделы, у каждого из которых может быть несколько реплик — копий.

Реплики хранятся на брокерах, каждый из которых может хранить несколько тысяч реплик (Рис. 2).

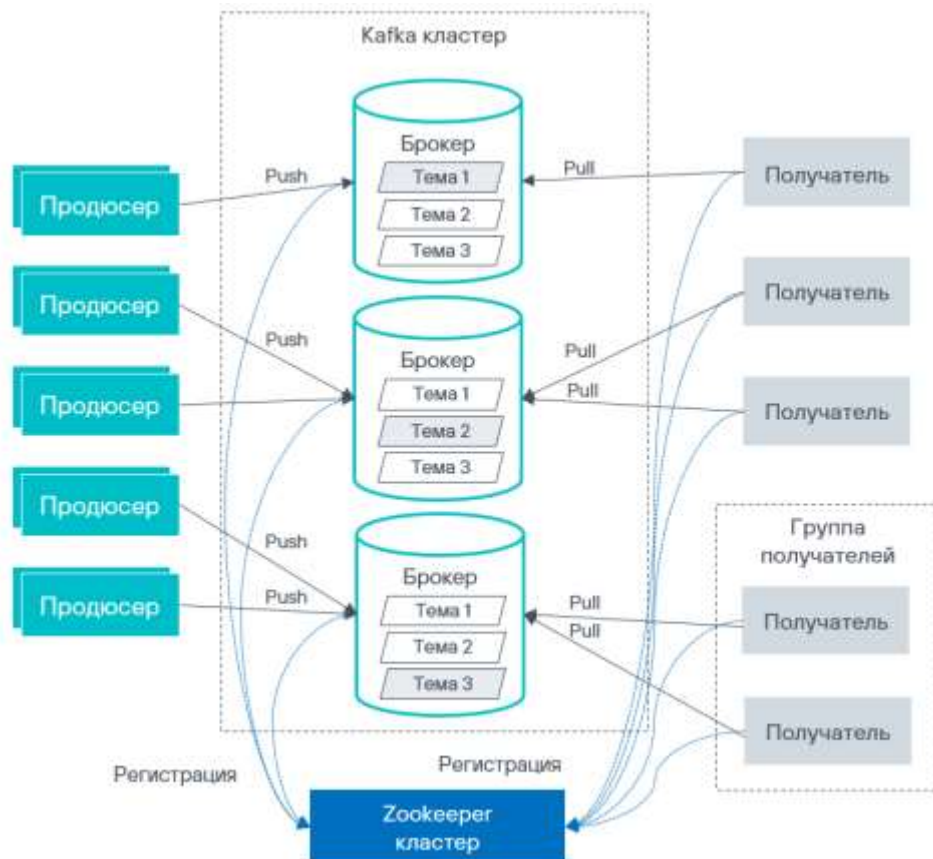


Рис. 2 Применение кластеров в Машине больших данных Скала^р МБД.С

### Применение кластеров

Кластеры решают задачи разделения данных, распределения нагрузки и масштабирования тем. Отказоустойчивость обеспечивается с помощью репликации.

Реплики делятся на два типа:

- Ведущие — реплики, через которые выполняются клиентские запросы
- Ведомые — реплики, которые копируют сообщения из ведущей реплики, тем самым поддерживая актуальное состояние по сравнению с ней

Применение кластеров позволяет расширить объем хранения сообщений и снять ограничения ресурсов.

Также кластеры при необходимости позволяют разделять доступ к запросам.

## Архитектура сервиса преобразования данных NiFi

Машина больших данных Скала^р МБД.С предоставляет гибкость в выборе способов работы с данными и в зависимости от конфигурации, включает сервис преобразования данных на базе ПО NiFi.

Система поддерживает множество процессоров для работы с разными источниками данных:

- Процессоры для NiFi (Рис. 3)
- Процессоры извлечения атрибутов
- Процессоры доступа к базе данных
- Процессоры приема данных
- Процессоры преобразования данных
- HTTP-процессоры
- Процессоры маршрутизации и посредничества
- Процессоры отправки данных
- Процессоры разделения и агрегации
- Процессоры системного взаимодействия
- Процессоры S3

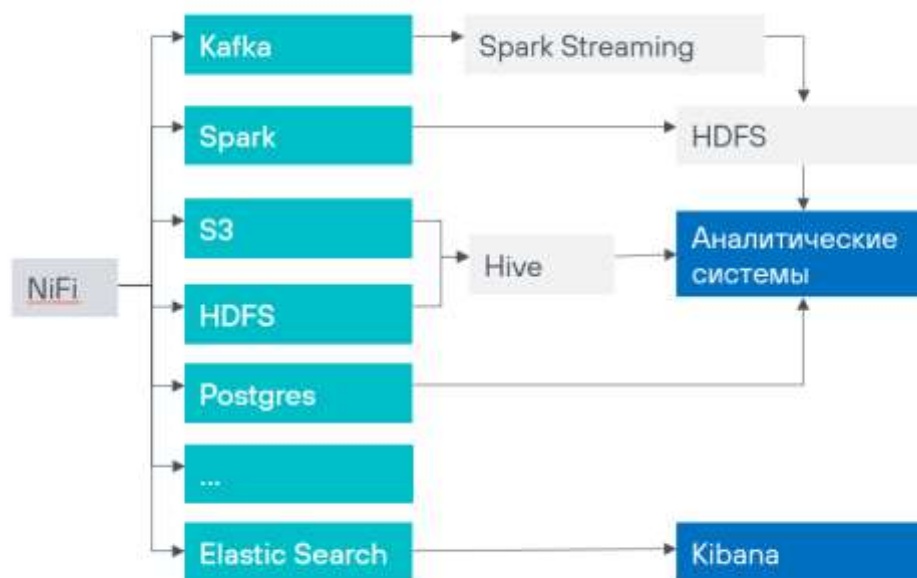


Рис. 3 Процессоры NiFi

### Удобство работы с интерфейсом для интеграции

Система имеет удобный веб интерфейс (Рис. 4), позволяющий легко настраивать компоненты NiFi.

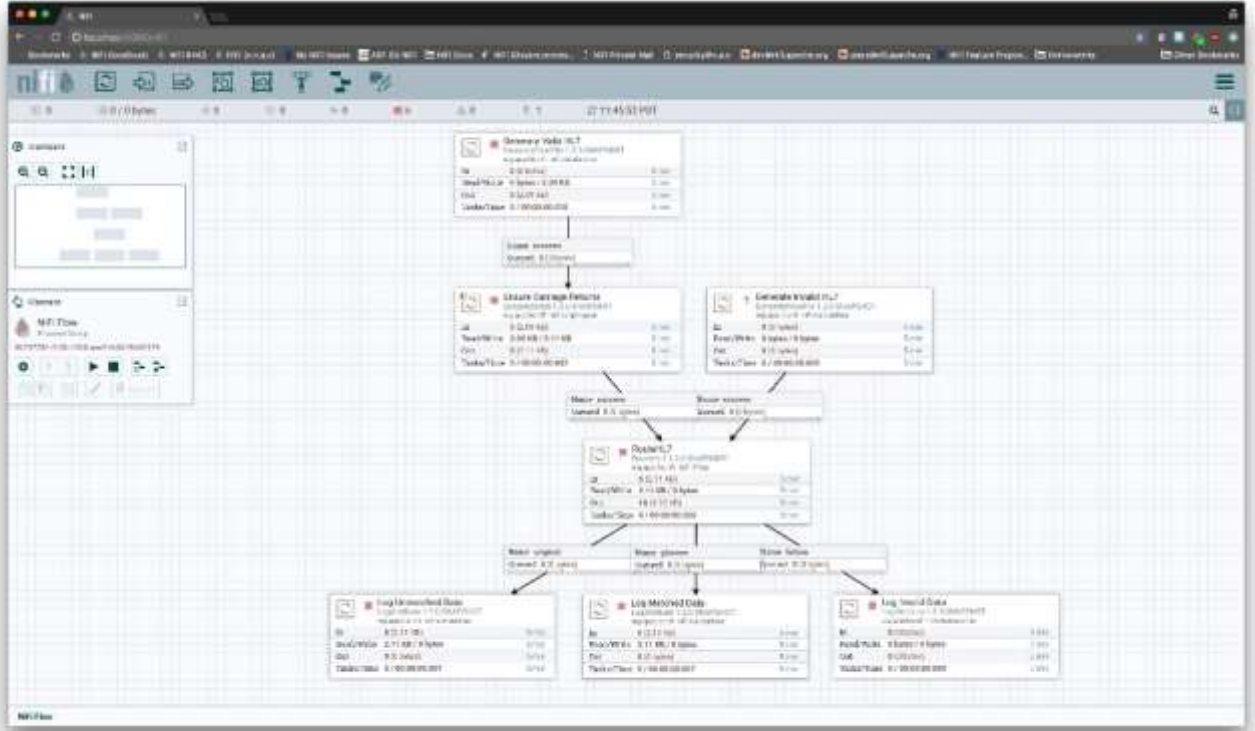


Рис. 4 Пример интерфейса NiFi

### СОСТАВ РЕШЕНИЯ

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

**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.

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

**Модуль** — это единица поставки Машин Скала^р, выполняющая определенные функции в соответствии с ее назначением. Модуль является единым и неделимым элементом спецификации и 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.С** состоит из следующих блоков (Рис. 5):

Блок коммутации и агрегации

Блок мониторинга и регистрации

Блок управления и распределения

Блок вычисления и хранения

Блок резервного копирования

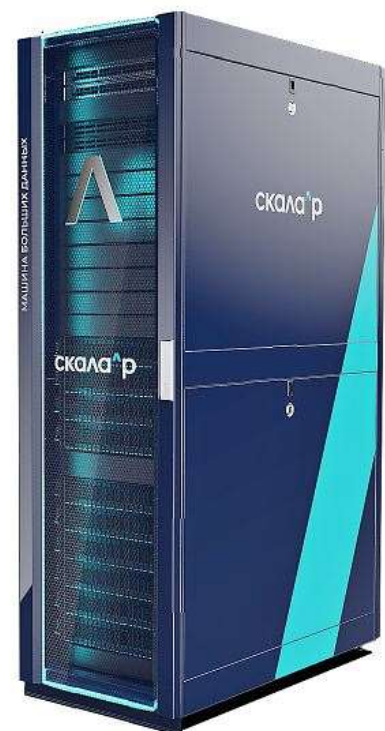


Рис. 5 Блоки в составе Машины больших данных Скала^р МБД.С

## Общая структура Машины

Структура Машины представлена на Рис. 6.

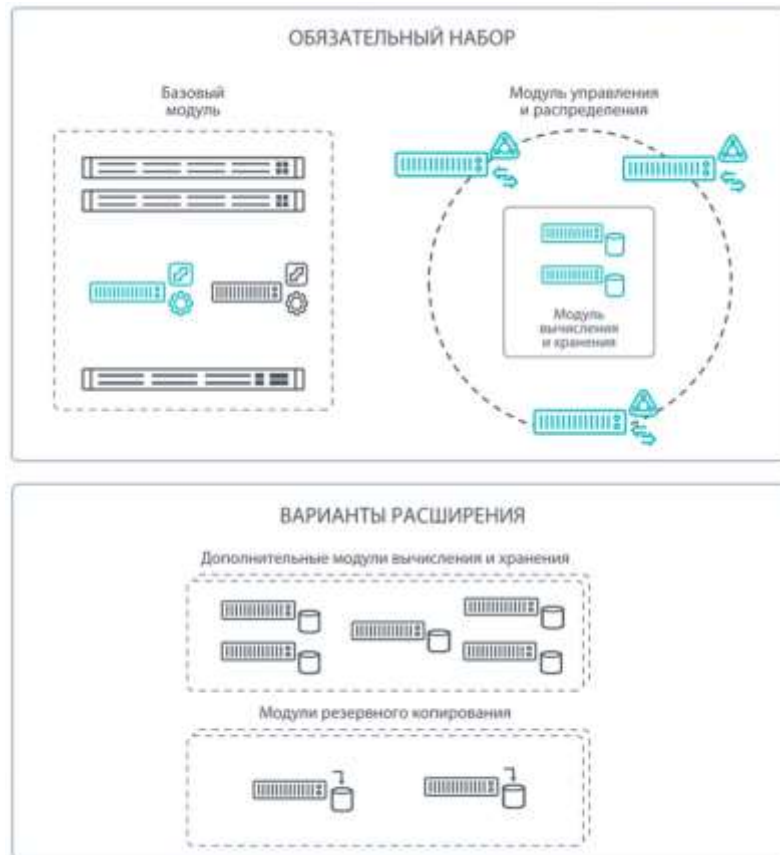


Рис. 6 Принципиальная схема Машины больших данных Скала^р МБД.С

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 и модулями резервного копирования.

### Базовый модуль

Базовый модуль — основа Машины. Модуль включает в себя узлы сетевого взаимодействия, мониторинга и регистрации, узел сетевого управления (Рис. 7). Он объединяет все модули и узлы Машины и способен доукомплектовываться отдельно стоящими секциями (стойками) расширения.



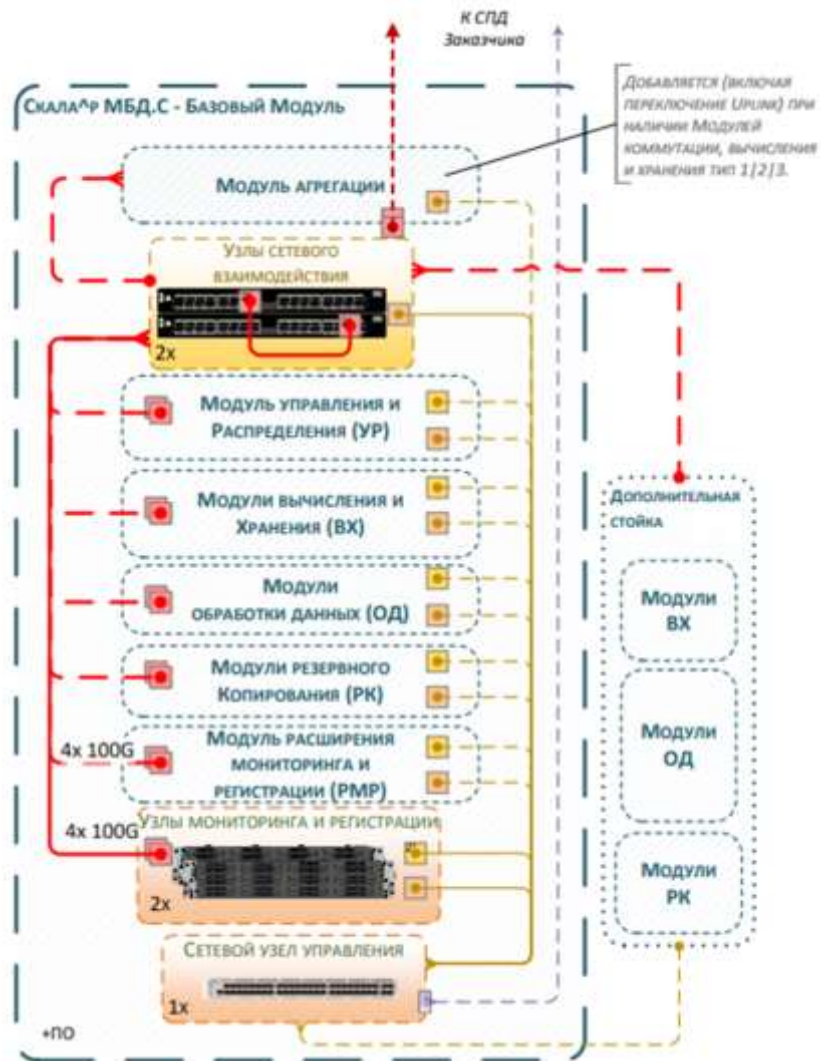


Рис. 7 Базовый модуль

## Блок коммутации и агрегации

Блок обеспечивает передачу данных между элементами **Машины МБД.С** (сеть внутреннего взаимодействия), информационный обмен с внешними сетями (сеть доступа), а также низкоуровневое управление узлами (сеть мониторинга и управления). Эффективное сетевое взаимодействие является важным фактором для быстрого и надежного функционирования Машины.

Схема сетевого взаимодействия представлена на Рис. 8.

### Реализованные подсети

- Internal VLAN — выделенная сеть внутреннего взаимодействия для обмена данными между узлами Машины, сеть резервного копирования, сеть кластерного взаимодействия. На схеме ниже обозначена стрелками оранжевого цвета
- External VLAN — сеть для подключения к сервисам Машины 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.

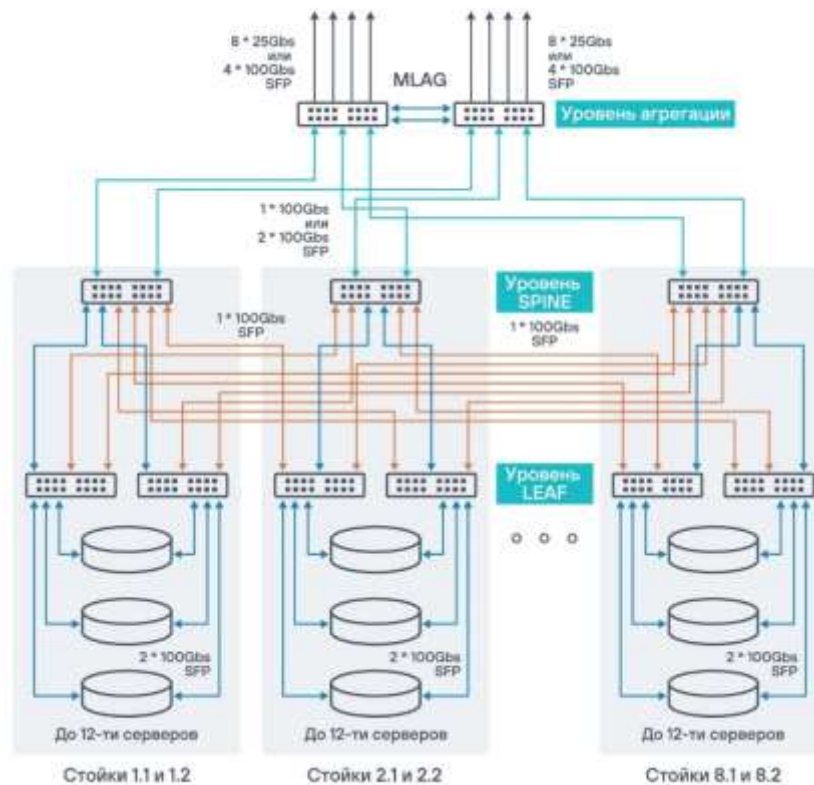


Рис. 8 Общая схема сетей внутреннего взаимодействия

### Комплект узлов сетевого взаимодействия

Узлы сетевого взаимодействия представлены внутри Машины в виде виртуальных коммутаторов, собранных по технологии MLAG из двух высокоскоростных аппаратных коммутаторов (Рис. 9).

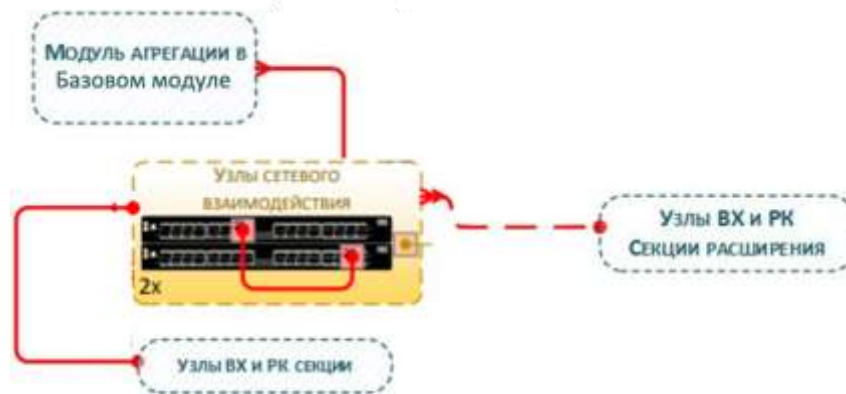


Рис. 9 Узлы сетевого взаимодействия

### Сети мониторинга и управления Машины

На рисунке ниже (Рис. 10) представлена архитектура сети управления:

- PXE (OS) VLAN — сеть для развертывания операционной системы по PXE, платформы МБД, мониторинга (стрелки оранжевого цвета)
- Агрегация (Ring VLAN) — резервная сеть кластерного взаимодействия, доступ к IPMI (стрелки зеленого цвета)
- IPMI VLAN — сеть управления оборудованием через интерфейсы удаленного управления (стрелки синего цвета)

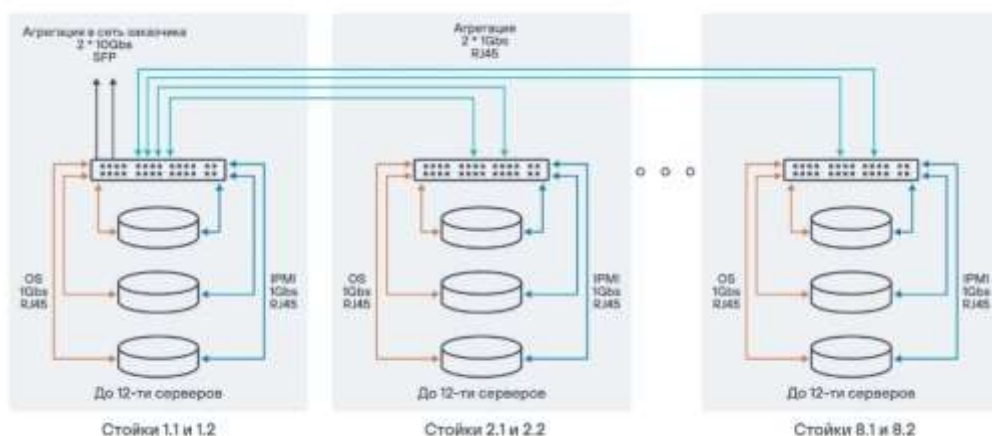


Рис. 10 Общая схема сетей мониторинга и управления

### Сетевой узел управления

Состоит из одного коммутатора сетевого управления с портами 1 и 10 Gigabit Ethernet для организации подсетей мониторинга, управления и служебного обмена (Рис. 11).

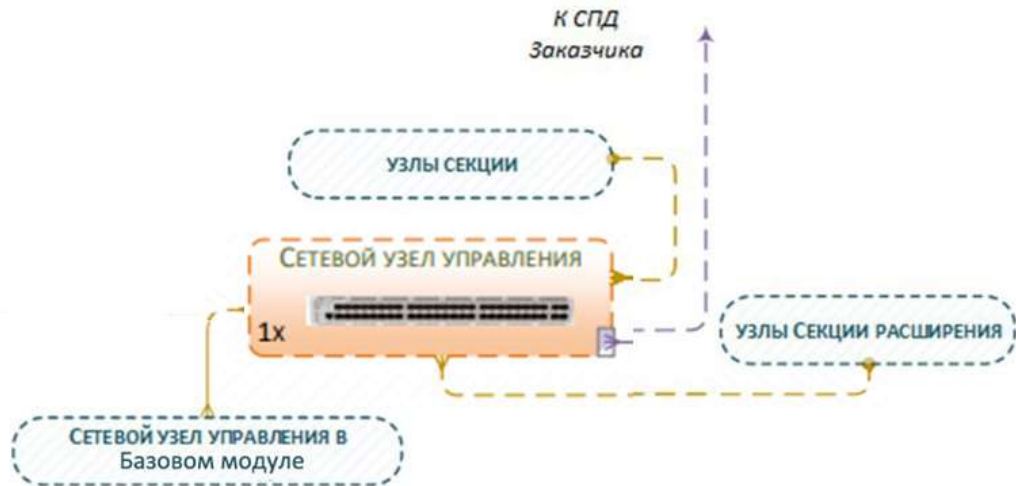


Рис. 11 Сетевой узел управления

### Модуль агрегации

Модуль агрегации — пара высокопроизводительных коммутаторов, собранных по технологии MLAG для агрегации коммутаторов сети внутреннего взаимодействия, применяется, когда Машина представлена множеством узлов, установленных в отдельно стоящих стойках, применяется при масштабировании Машины.

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

Модуль состоит из узлов сетевой агрегации (Рис. 12).



Рис. 12 Модуль агрегации

### Блок мониторинга и регистрации

Блок мониторинга и регистрации состоит из двух узлов мониторинга и регистрации, входящих в состав базового модуля Машины. Он обеспечивает управление Машиной на всех этапах жизненного цикла программных и аппаратных компонентов.

#### Основные функции управления реализуются с помощью ПО Скала^р

- 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. 13) состоит из двух высокопроизводительных узлов, объединенных в кластер отказоустойчивой конфигурации. Может быть расширен установкой дополнительного модуля расширения мониторинга и регистрации.



Рис. 13 Узлы мониторинга и регистрации

#### Каждый узел мониторинга и регистрации оптимизирован для своих специфических задач

- Использует SSD для обеспечения высокой производительности при хранении служебных данных
- Содержит выделенные SSD для загрузки ОС
- Оснащен двухпортовой сетевой картой 100 Gigabit Ethernet для организации сети внутреннего взаимодействия 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
- Использует от 256 ГБ ОЗУ

### Применяемое программное обеспечение

- ОС: Альт Линукс Сервер с виртуализацией или Базис.vCore
- Мониторинг и управление: ПО Скала^р Визион
- Управление эксплуатацией: ПО Скала^р Геном
- ПО Arenadata Cluster Manager (ADCM) и утилиты

### Модуль расширения мониторинга и регистрации

- Добавляется в Машину, если есть требование на увеличение отказоустойчивости системы мониторинга и регистрации, при использовании многофункциональной Машины, кратно количеству добавляемых функций (Рис. 14).



Рис. 14 Модуль расширения мониторинга и регистрации

### Блок управления и распределения

Основное предназначение блока-модуля — управление синхронизацией реплик БД и поддержание отказоустойчивого кластера Zookeeper (Рис. 15).



Рис. 15 Блок — 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 двухпортовой сетевой картой 100 Gigabit Ethernet для организации сети внутреннего взаимодействия 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
- Использует различные объемы памяти ОЗУ в зависимости от типа узла и проектного решения

#### Основная стойка коммутационного модуля

Модуль устанавливается в отдельную стойку, применяется для горизонтального масштабирования Машины сверх портовой емкости Базового модуля и может содержать следующие виды модулей Машины (Рис. 16):

- Узлы сетевого взаимодействия, вычисления и хранения, сетевые узлы управления
- Модули вычисления и хранения
- Модули резервного копирования

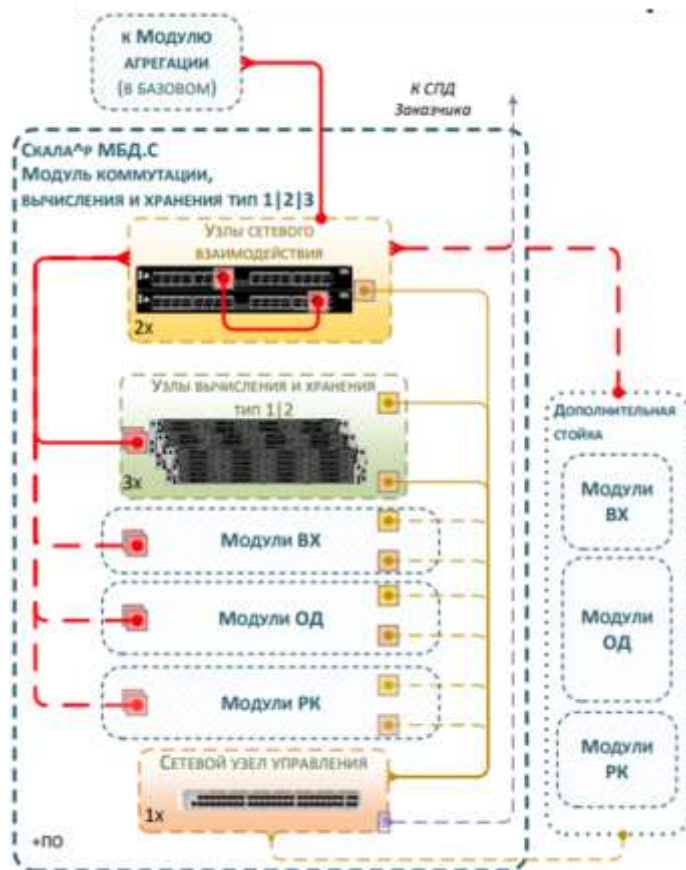


Рис. 16 Модуль коммутации, вычисления и хранения

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 стойки.

Горизонтальное масштабирование проводится в несколько этапов:

- Первый этап — это установка модулей дополнительно к Базовому. Как правило, это возможно, пока позволяет предельная мощность энергопотребления первой стойки и портовая емкость коммутаторов
- Далее устанавливается дополнительная отдельно стоящая стойка, в которой размещаются модули горизонтального масштабирования, вычисления и хранения, и резервного копирования. Такое расширение возможно до достижения ограничения количества портов в узлах внутреннего взаимодействия первой стойки
- Установка дополнительной полноценной второй стойки выполняется при достижении ограничения количества портов внутреннего взаимодействия в первой и дополнительной отдельно стоящей стойки. Стойка включает в себя набор узлов сетевого взаимодействия и требует установки дополнительного модуля агрегации в первую стойку с базовым модулем



## Блок вычисления и хранения

Блок вычисления и хранения состоит из типовых модулей вычисления и хранения в исполнении из 3-х узлов каждый (Рис. 17). Узлы соединяются с модулем сетевого взаимодействия и сетевым узлом управления.



Рис. 17 Модуль вычисления и хранения тип Т (3 узла)

### Каждый отдельный узел вычисления и хранения

- Содержит выделенные накопители SSD для загрузки ОС
- Использует локальные SAS SSD накопители для размещения образов данных и журналов (RAID 10), что обеспечивает повышение производительности (нет необходимости дополнительного внешнего обмена с системой хранения)
- Имеет дублированные интерфейсы данных (стандарт IEEE 802.3ad LACP) — повышение производительности, отказоустойчивость (в случае отказа одного из интерфейсов возможно снижение производительности)
- Оснащен двухпортовыми сетевыми картами 100 Gigabit Ethernet для сети внутреннего взаимодействия в рамках Машины
- Содержит два блока питания в режиме резервирования по схеме (1 + 1)
- Имеет два процессора
- Использует 512 ГБ ОЗУ в узле 1-го типа и 256 ГБ ОЗУ в узле 2-го типа

### Применяемое программное обеспечение

- ОС: Linux CentOS (ожидается переход на российские ОС)
- Специальная версия платформы AD.S
- Управление кластером средствами пакета AD.S

## Блок резервного копирования

Основное предназначение блока — запись и хранение резервных копий, а также их восстановление. Блок собирается из модулей резервного копирования. Каждый такой модуль состоит из одного узла резервного копирования (Рис. 18). Является опциональным.



Рис. 18 Модуль резервного копирования

### Каждый отдельный узел резервного копирования

- Содержит выделенные накопители SSD для загрузки ОС — обеспечение отказоустойчивости
- Применяется аппаратный или программный RAID в зависимости от конфигурации
- Оснащен от 14 HDD для хранения данных
- Имеет дублированные интерфейсы данных (стандарт IEEE 802.3ad LACP) — повышение производительности, отказоустойчивость (в случае отказа одного из интерфейсов возможно снижение производительности)
- Оснащен двухпортовой сетевой картой 100 Gigabit Ethernet для сети внутреннего взаимодействия в рамках Машины
- Содержит два блока питания в режиме резервирования по схеме 1+1
- Имеет два процессора
- Использует от 384 ГБ ОЗУ

### Применяемое программное обеспечение

Для создания резервных копии используются встроенные в ПО Arenadata средства резервного копирования. Резервное копирование рассчитывается для минимум 4-х копий максимальной базы.

## Система управления эксплуатацией Скала^р Геном

В **Маchine больших данных Скала^р МБД.С** применяется специализированное программное обеспечение Скала^р Геном.

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 Скала^р Визион

В **Маchine больших данных Скала^р МБД.С** 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

## СПЕЦИФИЧНЫЕ ЧЕРТЫ

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

*Приоритет обеспечения сохранности данных перед повышенной доступностью*

Такой подход обеспечивает гарантию сохранности данных при любых единичных отказах оборудования и быстрое восстановление из резервных копий в случае фатальных сбоев. В ряде сценариев возможны одновременные отказы разных компонентов без потери данных.

*Реализация вычислительного блока на аппаратном решении вместо использования виртуальной среды*

Это позволяет достичь максимума производительности на данном оборудовании, что исключает потери на среду виртуализации и повышает надежность системы в целом.

*Применение стандартного высоконадежного и производительного оборудования в качестве платформы для размещения компонентов решения взамен уникальных аппаратных разработок*

Обеспечение стабильного уровня производительности достигается за счет использования компонентов, проверенных лабораторией Скала^р. Отсутствие уникальных элементов также повышает надежность решения.

*Использование программных RAID отечественных производителей вместо отдельных аппаратных RAID*

Это позволяет достичь более высокой производительности, высокой гибкости в настройках (в зависимости от требований), снижение зависимости от производителей оборудования и гарантирует оптимальность алгоритмов RAID.

*Возможность применения типовых и сторонних решений для мониторинга и управления в дополнение к предустановленным*

Это позволяет сохранить ранее сделанные инвестиции в системы управления ИТ-инфраструктурой и дает возможность построения сквозных систем управления, в которые интегрируются Машины.

## ГАРАНТИРОВАННОЕ КАЧЕСТВО

**Машины больших данных Скала^р МБД.С** создаются на основе подтвержденной архитектуры с применением проверенных компонентов в тесном сотрудничестве с производителями и технологическими партнерами.

### Производство (сборка Машины и предустановка ПО)

При производстве используются только высококачественные комплектующие, а сборка продукции осуществляется строго в соответствии с технологическими картами. Первичное развертывание ПО осуществляется в автоматическом режиме, дополнительные настройки ПО осуществляются в соответствии с утвержденной пошаговой инструкцией. Сформированная Машина тщательно тестируется, таким образом отклонения от типового решения **Скала^р МБД.С** исключены.

### Передача в эксплуатацию

**Машина больших данных Скала^р МБД.С** полностью сформирована, протестирована, готова к размещению в сети Заказчика и установке прикладного ПО Заказчика. В комплекте с Машиной передается Паспорт изделия, эксплуатационная документация, Сертификат на техническую поддержку.

## РЕАКЦИЯ НА ВОЗМОЖНЫЕ ОТКАЗЫ

### Отказы, связанные со стандартными элементами Скала^р МБД.С

В рамках **Машины больших данных Скала^р МБД.С** обеспечена отказоустойчивость ее основных элементов, в том числе:

- Узлов (дублирование процессоров, источников питания и др.)
- Подсистемы ввода-вывода (RAID)
- Сети внутреннего взаимодействия (дублирование сетевых интерфейсов)
- Системы резервного копирования

Отказы перечисленных элементов обрабатываются 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.С также восстанавливается.

### Отказы, связанные с узлами кластера

#### Аппаратные сбои

Архитектура программного обеспечения, лежащего в основе Скала^р МБД.С, позволяет построить отказоустойчивый многоузловой кластер.

Отказоустойчивость обеспечивается за счет настройки уровня репликации тем, а также избыточности экземпляров сервиса Zookeeper, который отвечает за координацию работы Kafka и NiFi.

Даже одновременный отказ любых узлов с разными ролями, например, узла-брокера, узла-процессора NiFi и узла Zookeeper, не влияет на работоспособность системы и позволяет продолжать работу по обмену сообщениями. При этом общая производительность снижается.

#### Программные сбои и человеческий фактор

Кроме обеспечения отказоустойчивости, по требованию Заказчика в Машину может быть интегрирована система резервного копирования.

Одним из вариантов решения данной задачи является дублирование данных в Hadoop или S3, например, путем подписки дополнительных потребителей на темы Kafka.

Детальный порядок обеспечения отказоустойчивости кластера и рекомендации по действиям при его администрировании могут быть предоставлены по запросу.

## ТРЕБОВАНИЯ К РАЗМЕЩЕНИЮ РЕШЕНИЯ

Машина может быть смонтирована в 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 4x100 Gigabit Ethernet или до 8x10/25 Gigabit Ethernet. Требуемые трансиверы определяются на этапе формирования спецификации Машины.

При развертывании Машины на ней будут выполнены настройки сетевых адресов в соответствии со структурой сети Заказчика. Заказчик должен предоставить необходимые данные.

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

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

## ТЕХНИЧЕСКАЯ ПОДДЕРЖКА

Поставка **Машины больших данных Скала^р МБД.С** осуществляется с предварительными сборкой, тестированием и настройкой оборудования согласно требованиям Заказчика. Качественная поддержка Машины 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 9x5, или в расширенном режиме 24x7 (опция для критической функциональности)
- Срок начально приобретаемой технической поддержки может быть увеличен до 3 и 5 лет, также доступна пролонгация поддержки
- Возможно включение в состав стандартных пакетов дополнительных опций и услуг

Состав типовых пакетов услуг по технической поддержке представлен в таблице ниже.

*Пакеты услуг по технической поддержке Машины больших данных Скала^р МБД.С*

Услуга	Пакет «9x5»	Пакет «24x7»
Режим «Обслуживание комплекса Скала^р в режиме 9x5» (в рабочее время по рабочим дням)	+	+
Режим «Обслуживание комплекса Скала^р в режиме 24x7» (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: <ul style="list-style-type: none"> <li>■ устранение аппаратных неисправностей;</li> <li>■ техническое сопровождение ПО.</li> </ul>	+	+
Консультации по работе комплекса Скала^р	+	+



Услуга	Пакет «9x5»	Пакет «24x7»
«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 опции обслуживания	–	+
Времена реагирования и отклика, не более:		
Время регистрации обращений	30 минут, рабочие часы (9x5)	30 минут, круглосуточно (24x7)
Подключение специалиста к решению инцидентов критичного и высокого уровней	В течение 1 рабочего часа (9x5)	В течение 1 часа (24x7)

Примечание к срокам ремонта оборудования: комплекс **Скала^р МБД.С** 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.

## ЛИЦЕНЗИРОВАНИЕ ПО, ВХОДЯЩЕГО В СОСТАВ МАШИНЫ

Все наименования ПО лицензируются по модулям Машины и их количеству в Машине.

### Лицензирование ПО Скала^р МБД.С

Программное обеспечение СУБД Arenadata Streaming (AD.S) лицензируется согласно объему ресурсов в модулях вычисления и хранения и модулях обработки данных, при этом на каждый модуль выдается единая лицензия.

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

### Варианты лицензирования

Лицензирование ПО комплекса Скала^р МБД.С имеет три редакции:

- Фиксированная — приобретается бессрочная (постоянная) лицензия
- Временная — приобретается лицензия на период времени
- Подписка — ежемесячная подписка на лицензию и поддержку

### Политика обновления ПО

Команда Скала^р активно занимается развитием собственных программных продуктов Скала^р. 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

**Компания Скала^р** — разработчик и производитель модульной платформы для построения корпоративной ИТ-инфраструктуры. Компания выполняет роль технологического контрибьютора и объединяет в экосистему 25 партнеров – лидеров в своем классе решений.

**Скала^р** с 2015 года выпускает программно-аппаратные комплексы (Машины) и сегодня предлагает широкий технологический стек для решения задач виртуализации, создания транзакционных баз данных, хранения информации и аналитической обработки больших объемов данных. Продукты Скала^р включены в Реестр промышленной продукции, произведенной на территории Российской Федерации, и в Единый реестр российских программ для ЭВМ и БД.

**Машины Скала^р** являются серийно выпускаемыми преднастроенными комплексами и позволяют осуществлять быстрое развёртывание и ввод в эксплуатацию.

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](http://www.skala-r.ru).