Модульная платформа для высоконагруженных корпоративных и государственных информационных систем

Машина больших данных Скала^р МБД.КХ

скала^р

Скала^р сегодня

разработка и производство модульной платформы для высоконагруженных государственных и корпоративных информационных систем

8 лет

серийного выпуска

400+

комплексов в промышленной эксплуатации

6500+

вычислительных узлов

Линейка продуктов Скала^р

решения для высоконагруженных корпоративных и государственных систем по четырем направлениям

Динамическая инфраструктура

Машины виртуализации Скала^р МВ

на основе решений **Basis** для создания динамической конвергентной и гиперконвергентной инфраструктуры ЦОД и виртуальных рабочих мест пользователей

Управление большими данными

Машины больших данных Скала^р МБД.8

на основе решений **Arenadata и Picodata** для создания инфраструктуры хранения, преобразования, аналитической, статистической обработки данных с применением ИИ, а также распределенных вычислений

Высокопроизводительные базы данных

Машины баз данных Скала^р МБД

на основе решений **Postgres Pro** для замены Oracle Exadata в высоконагруженных системах с обеспечением высокой доступности и сохранности критически важных данных

Интеллектуальное хранение данных

Машины хранения данных Скала^р МХД

на основе технологии объектного хранения S3 для геораспределенных катастрофоустойчивых систем с сотнями миллионов объектов различного типа и обеспечения быстрого доступа к ним

Использование опыта технологических лидеров (гиперскейлеров)
Использование самых зрелых и перспективных технологий в кооперации с технологическими лидерами российского рынка в каждом из сегментов

ПАК Скала^р в Реестрах РФ

Машины
Модули
Компоненты

Реестр промышленной продукции, произведенной на территории Российской Федерации

Машины
Модули
Программное обеспечение

Машины
Модули
Программное обеспечение

Соответствуют критериям доверенного ПАК

Машины больших данных Скала^р МБД.8

высокопроизводительные хранилища и витрины данных на базе продуктов Arenadata и Picodata

Скала^р МБД.Г + Arenadata DB (ADB)

СУБД массово-параллельной обработки (на основе Greenplum)

Скала^р МБД.Т + Picodata

Распределенные вычисления в оперативной памяти (аналог Tarantool)

Скала^р МБД.С + Arenadata Streaming (ADS)

Потоковая обработка данных в реальном времени (на основе Kafka и NiFi)

Скала^р МБД.Х + Arenadata Hadoop (ADH)

Машина для обработки больших данных средствами экосистемы Hadoop

Скала^р МБД.КХ + Arenadata QuickMarts (ADQM)

Машина для быстрых аналитических витрин с реляционным доступом. Децентрализация, репликация, масштабируемость (на основе ClickHouse)

Машина больших данных Скала^р МБД.КХ для быстрых аналитических витрин

с применением продукта Arenadata QuickMarts (ADQM) на основе ClickHouse

Сценарии применения

- Сверхбыстрые витрины данных с задаваемой глубиной выборки
- Выполнения аналитических запросов в режиме реального времени (OLAP)
- Аналитические исследования, требующие быстрого перебора гипотез
- Анализ данных с IoT устройств

Преимущества

- Линейная масштабируемость
- Высокая доступность и катастрофоустойчивость
- Сжатие данных в десятки раз

Замещаемые технологии

- Vertica, Oracle, Teradata, Paraccel (Amazon RedShift)

Рекомендовано
от 2 ТБ

Скорость обработки аналитических запросов
до 100 раз
быстрее транзакционных систем

Обязательно для баз объемом
более 5 ТБ

Функциональная специфика Машины Скала^р МБД.КХ

Требования к сценарию работы

- Подавляющее большинство запросов — на чтение
- Данные обновляются достаточно большими пачками (> 1000 строк), а не по одной строке, или не обновляются вообще
- Высокая пропускная способность при обработке запросов (до миллиардов строк в секунду на один сервер)

Особенности ввода-вывода

- **Минимальный объем считываемых данных -** для запроса требуется небольшое количество столбцов, чтение только нужных данных, 20-кратное уменьшение операций ввода-вывода
- **Экономия дискового пространства** за счет блочного хранения
- **Эффективный кэш** за счет уменьшения операций ввода-вывода

Особенности вычислений

- **Векторный движок – снижение издержек на диспетчеризацию,** оптимизированный код операции
- **Кодогенерация – код запроса и** косвенные вызовы

Ключевые особенности и характеристики

- **Линейная масштабируемость** на петабайты данных, геокластеризация
- **Высокая доступность** – за счет применения репликации с фактором x3 как минимум в трех датацентрах
- **Сжатие данных** в десятки и сотни раз
- **Запросы** – встроенный диалект SQL, для работы с различными типами данных
- **Скорость обработки запросов** - до 100 раз быстрее классических СУБД, до 20 раз быстрее колоночных конкурентов

Сценарий: Ускорение систем визуализации данных для тысяч пользователей

- Высокоскоростная система для аналитических приложений
- Дополнительное ускорение за счет специализированной дисковой подсистемы
- Повышение скорости запросов за счет распределенной кластерной архитектуры
- Ускорение внутреннего взаимодействия между узлами за счет сети 100 Гбит/с
- Построение катастрофоустойчивого решения
- Модульная масштабируемость
- Разгрузка системы визуализации данных

Отвечая потребностям бизнеса

Производительность

Способы достижения высочайшей производительности, не требующие применения суперкомпьютеров

Доступность данных

Схема распределения потоков данных не препятствует выполнению вычислительных задач

Управляемость

Дополнительные программные сервисы, позволяющие управлять и чувствовать каждый такт работы всей системы

Производительность

Спрогнозированная нагрузка

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

Программный RAID

- Производительнее аппаратного RAID-контроллера
- Минимальное использование RAM (требуется менее 4GB RAM)
- Управление процессорными потоками
- Минимальная просадка производительности в режиме восстановления

Выделенный интерконнект

- Высокоскоростная сеть интерконнекта ускоряет распределение заданий, ETL и ELT
- Параллельная обработка запросов на узлах приводит к суммированию мощностей всех узлов
- Создание параллельной синхронной копии не влияет на выполнение задания
- Все серверы взаимодействуют между собой с одинаковой скоростью

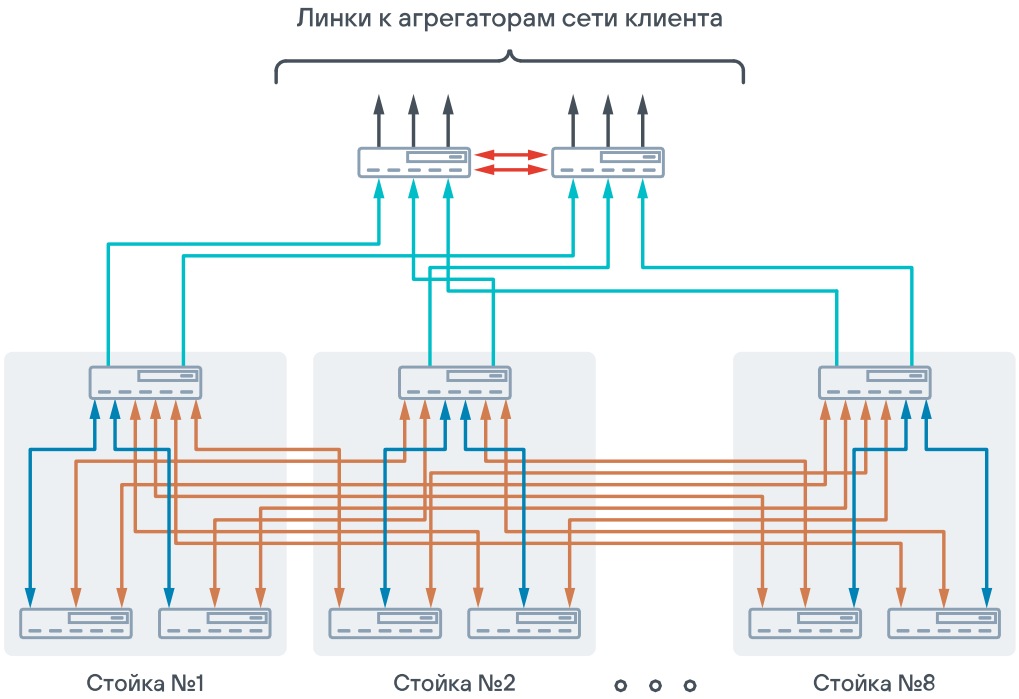


Схема внутренних соединений Leaf-Spine
с увеличением скорости при горизонтальном масштабировании

Доступность данных — синхронная копия БД

Синхронная репликация доступна для следующих таблиц семейства MergeTree

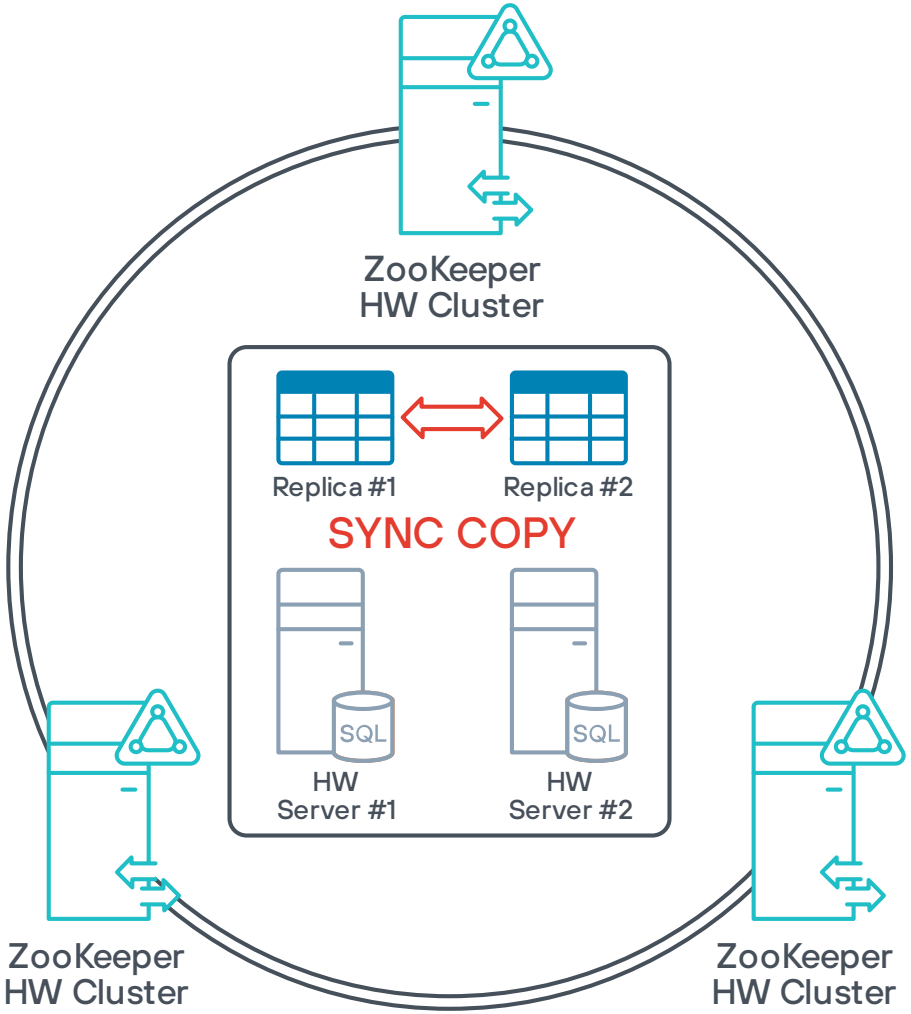
- ReplicatedMergeTree
- ReplicatedSummingMergeTree
- ReplicatedReplacingMergeTree
- ReplicatedAggregatingMergeTree
- ReplicatedCollapsingMergeTree
- ReplicatedVersionedCollapsingMergeTree
- ReplicatedGraphiteMergeTree

Метаинформация реплик

- Хранится в трехточечном кластере Модуля управления

Особенности репликации

- Работает на уровне отдельных таблиц, а не всего сервера. На узле могут быть расположены одновременно реплицируемые и нереплицируемые таблицы
- Не зависит от шардирования
- Основана на запросах INSERT и ALTER
- Не привязана к именам таблиц



Доступность данных — асинхронная копия БД

Особенности асинхронной репликации СУБД

Распределенный кластер

- Надежность за счет децентрализации и отсутствия единой точки отказа

Асинхронная multi-master репликация

- Обеспечение реплицирования данных в фоновом режиме. СУБД поддерживает полную идентичность данных на разных репликах, автоматически восстанавливая их после сбоев

Кворумный режим записи данных

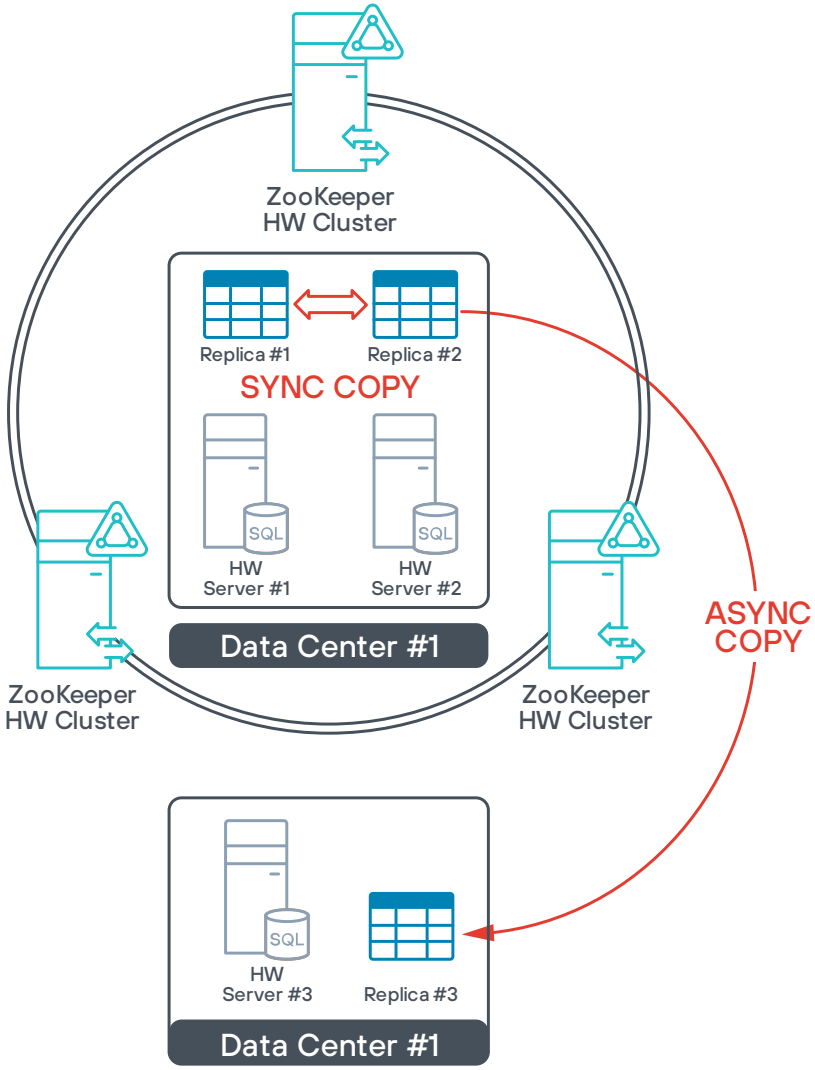
- Запись считается успешной только после того, как информация записана на несколько серверов — кворум. Так обеспечивается линеаризуемость и имитация синхронных реплик

Отставание асинхронной реплики

- Определяется шириной канала связи и задержками

Дополнительное применение

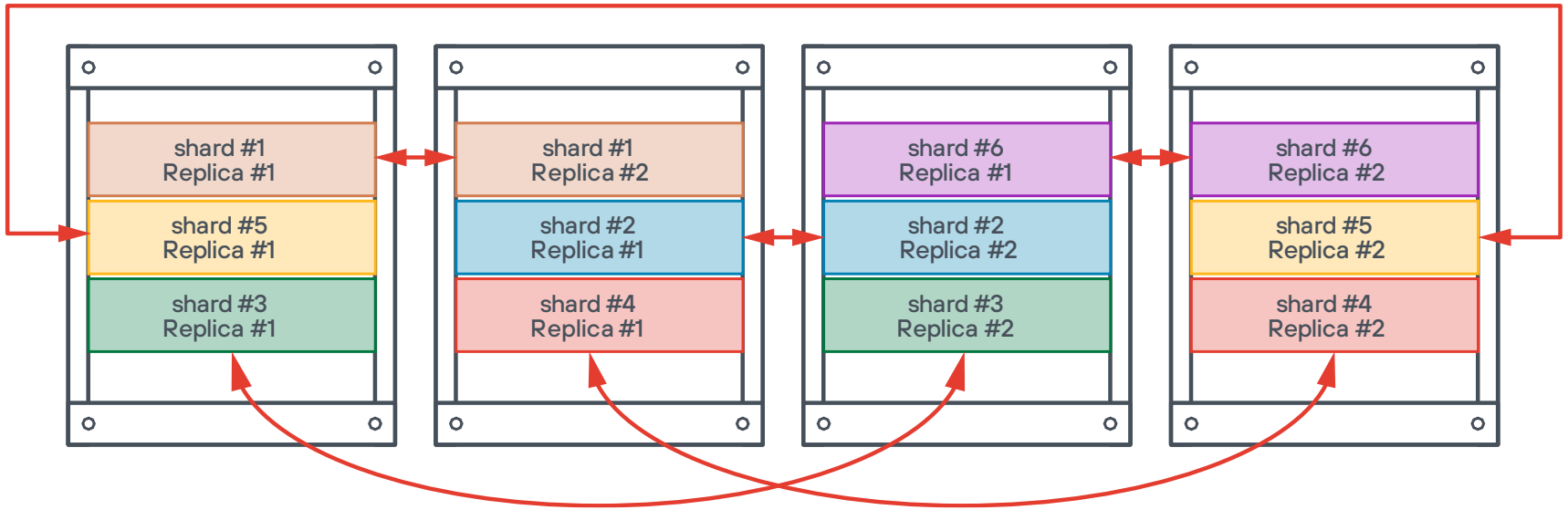
- Асинхронная реплика может быть использована для создания резервных копий, чтобы не останавливать основной синхронный контур



Расширение объема данных — шардирование

Главное преимущество

- Снимает ограничение ресурсов одного узла, увеличивая объем базы на десятки и сотни узлов
- Позволяет распараллеливать выполнение запросов, увеличивая скорость в десятки и сотни раз
- Позволяет с помощью реплик строить защиту таблицы / узла / кластера / стойки / зала / датацентра и т.д.



На схеме приведен вариант шардирования и кольцевой репликации шардов,
при котором можно допустить выход из строя целой серверной стойки без потери данных

Гибкость в выборе способы работы с данными

Движок — код, встроенный в таблицу, определяет:

- Как и где хранятся данные, куда их писать и откуда читать
- Какие запросы поддерживаются и каким образом
- Конкурентный доступ к данным
- Использование индексов, если есть
- Возможно ли многопоточное выполнение запроса
- Параметры репликации данных

MERGETREE

- MergeTree
- ReplacingMergeTree
- SummingMergeTree
- AggregatingMergeTree
- CollapsingMergeTree
- VersionedCollapsingMergeTree
- GraphiteMergeTree

LOG

- TinyLog
- StripeLog
- Log

Движки для интеграции

- Kafka
- MySQL
- ODBC
- JDBC
- S3
- EmbeddedRocksDB
- RabbitMQ
- PostgreSQL

Специальные движки

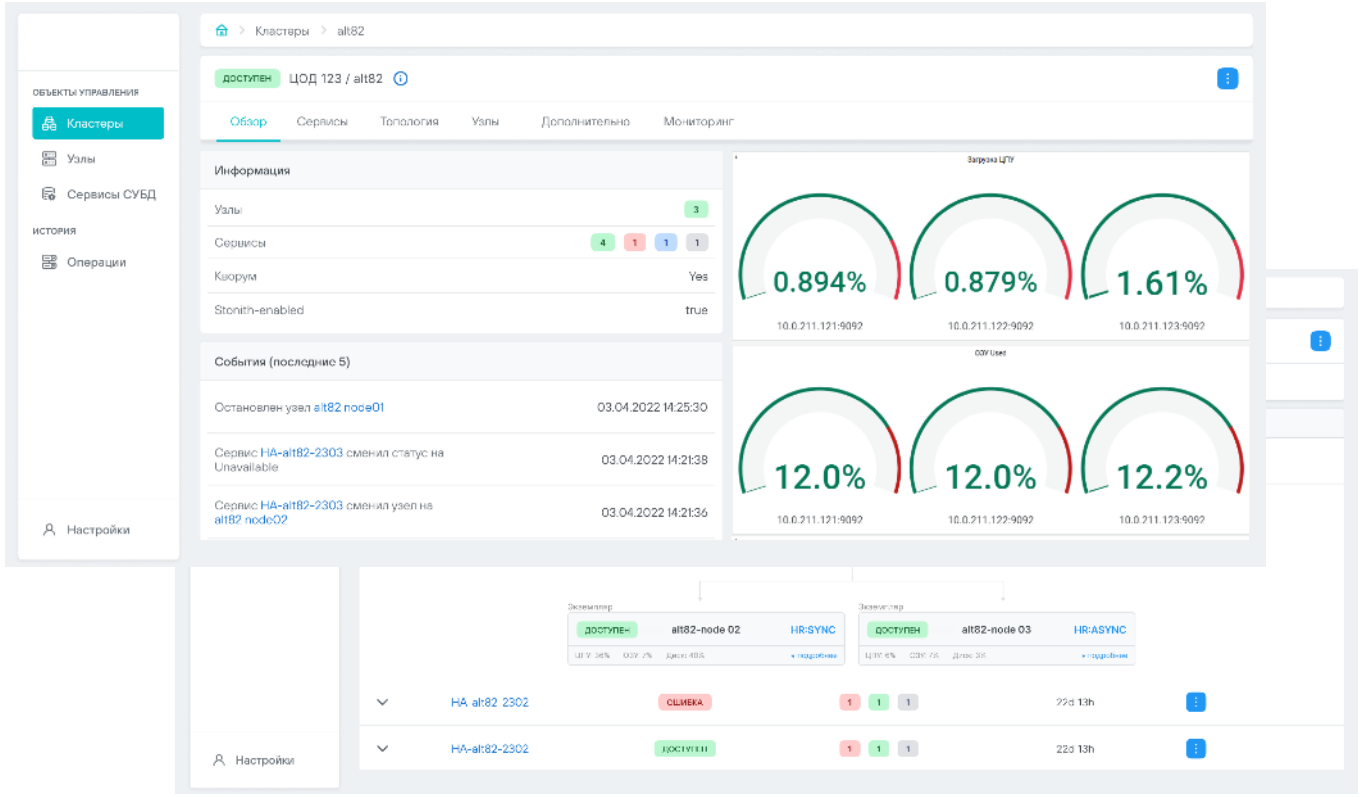
- Distributed
- MaterializedView
- Dictionary
- Merge
- File
- Null
- Set
- Join
- URL
- View
- Memory
- Buffer

Управляемость

Система управления жизненным циклом Скала^р Геном

Данный программный продукт обеспечивает:

- Контроль развертывания 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



Общий состав семейства Машин Скала^р МБД.8

Блок вычисления и хранения	▪ Высокопроизводительные кластеры ▪ Параллельные вычисления ▪ Отказоустойчивая архитектура	от 3х узлов
Блок коммутации и агрегации	▪ Объединение всех компонентов ▪ Высокоскоростное взаимодействие ▪ Отказоустойчивая схема сети	до 100 Гбит/с
Блок управления и распределения	▪ Интерфейс для запросов ▪ Расширяемость ▪ Сервисные функции	интеллектуальное управление
Блок мониторинга и регистрации	▪ Управление эксплуатацией ▪ Автоматизация процедур ▪ Мониторинг компонент Машины	50% экономия на эксплуатации

Блок резервного копирования*	▪ Хранение резервных копий:
	  ▪ Данные
	  ▪ Настройки и метаданные
		сохранность данных

* опция

Блок вычисления и хранения

Назначение:

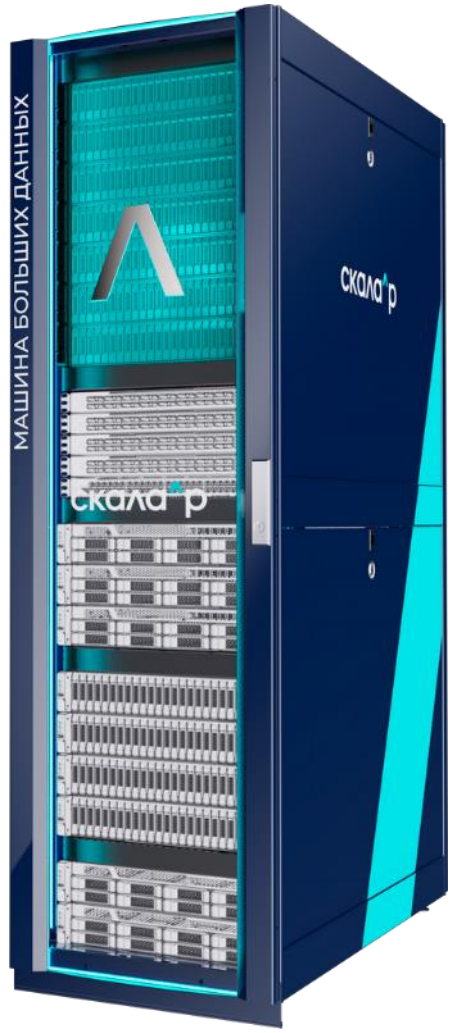
- Хранение таблиц БД и их синхронных и асинхронных реплик
- Быстрое вычисление запросов

Модификации составляющих модулей:

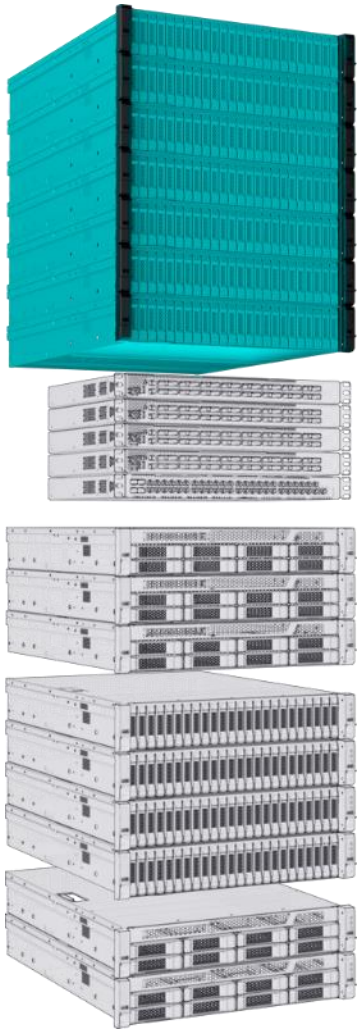
- По объему хранения и вычислений
- По производительности
- По предназначению: для продуктивной среды или для разработки

Расположение:

- В базовом блоке
- В стойках расширения
- В модулях расширения коммутации



Блок вычисления и хранения



Применимость:

- По параметрам модулей данного блока определяется производительность и объемы хранения МБД.КХ
- Расширение производительного объема и повышение производительности всей системы в 50% случаев происходит за счет дополнения модулей вычисления и хранения

Особенности:

- Самый высоконагруженный блок в Машине МБД.КХ
- Хранение строится на дисках SAS SSD 12G или NVMe SSD
- Количество процессорных ядер — от 80 на модуль
- Оперативная память от 384 ГБ до 1536 ГБ на модуль в зависимости от исполнения

Блок коммутации и агрегации

Назначение:

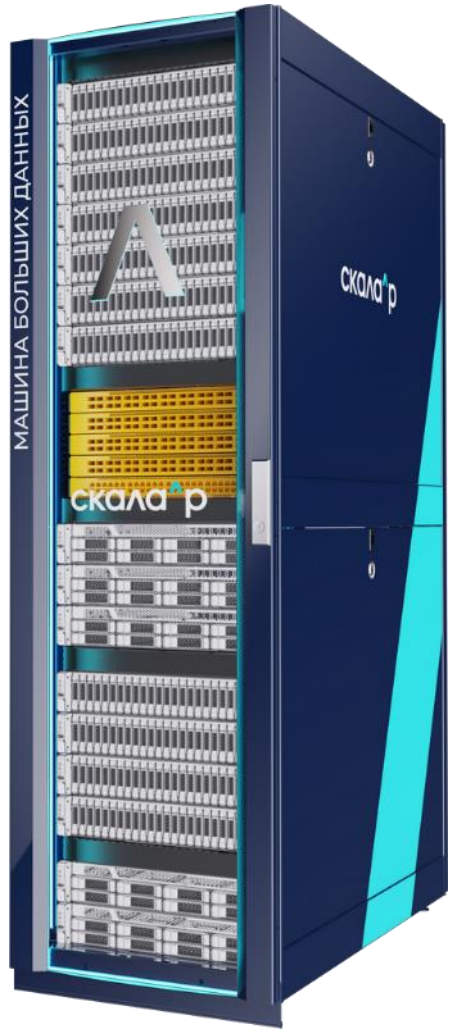
- Внутренний интерконнект на высокой скорости
- Агрегация по схеме Leaf-Spine или «звезда»
- Выделенная сеть для управления и мониторинга

Модификации составляющих модулей:

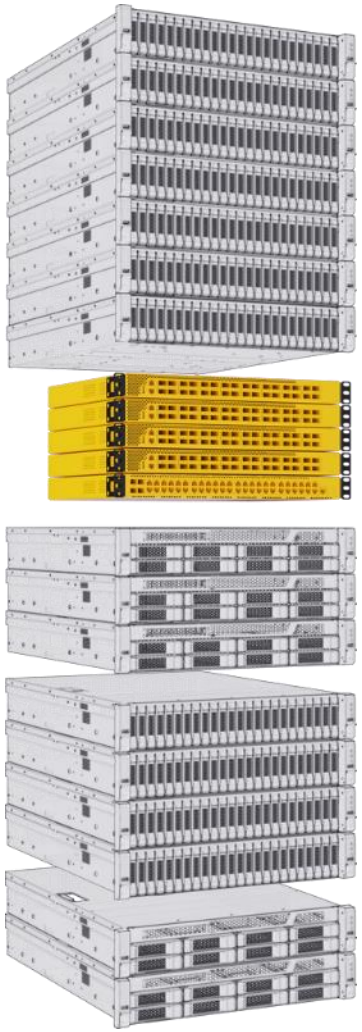
- Модуль агрегации в базовом блоке служит для соединения в одну сеть модулей коммутации и связи с инфраструктурой
- Модуль коммутации в каждой активной стойке

Расположение:

- В базовом блоке
- В модулях расширения коммутации



Блок коммутации и агрегации



Применимость:

- Соединение с инфраструктурой клиента
- Обеспечение скоростной внутренней коммутации
- Обеспечение отдельной сети для резервного копирования
- Обеспечение сетей для мониторинга и управления

Особенности:

- От трех до семи коммутаторов на стойку
- До трех параллельно действующих сетей для обеспечения отказоустойчивости

Блок управления и распределения

Назначение:

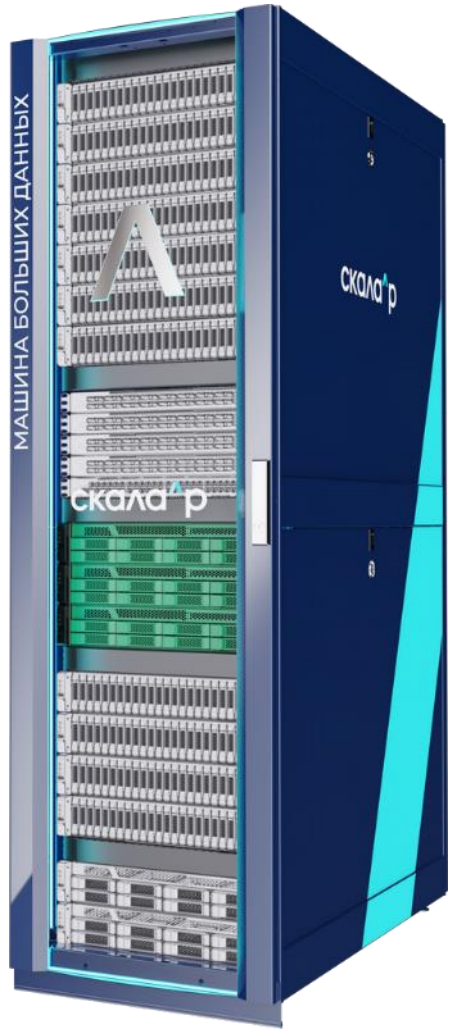
- Управление синхронизацией реплик БД
- Поддержание отказоустойчивого кластера

Модификации модулей:

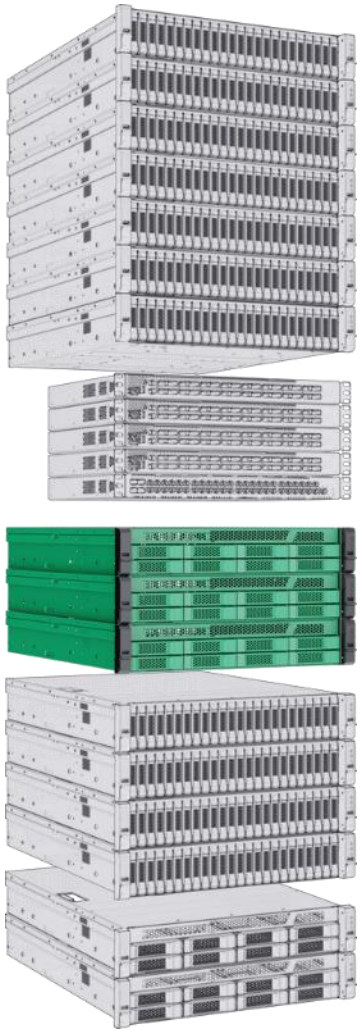
- Трехузловой кластер — стандартное решение
- Семиузловой кластер — решение для поддержания копий в резервном ЦОД с удаленным арбитром

Расположение:

- В 99% случаев в базовом модуле / модулях



Блок управления и распределения



Применимость:

- Является основой для поддержания репликации данных
- Может быть расширен резервными узлами

Особенности:

- Зафиксированы оптимальные конфигурации
- В отдельных случаях может использовать внешние относительно модуля базы данных для хранения метаинформации

Блок мониторинга и регистрации

Назначение:

- Управление Машиной от бизнес-модели до конкретных аппаратных компонентов
- Управление развертыванием, обновлением, жизненным циклом Машины

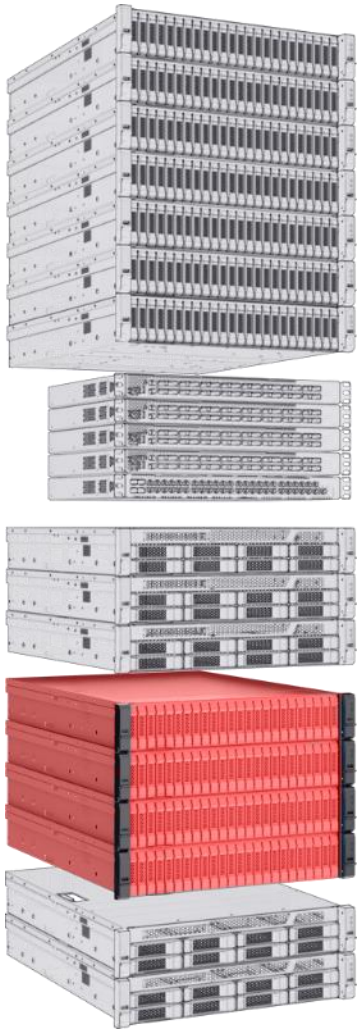
Модификации составляющих модулей:

- Один узел — абсолютный минимум без резервирования
- Два узла — взаимное резервирование с ручным переключением
- Четыре узла — стандартная отказоустойчивость высокой доступности с распределенным хранилищем

Расположение:

- В 99% случаев в базовом модуле

Блок мониторинга и регистрации



Применимость:

- Присутствует в любой Машине Больших Данных Скала^р
- Обязательно содержит ПО:
  - Скала^р Геном
  - Скала^р Визион
  - Аренадата кластер-менеджер
  - Аренадата инструменты
  - Аренадата Мониторинг

Особенности:

- Всегда в виртуальной среде
- Система управления виртуализацией входит в комплект

Блок резервного копирования

Назначение:

- Хранение резервных копий БД
- Хранение настроек и метаданных
- Пространство для ETL

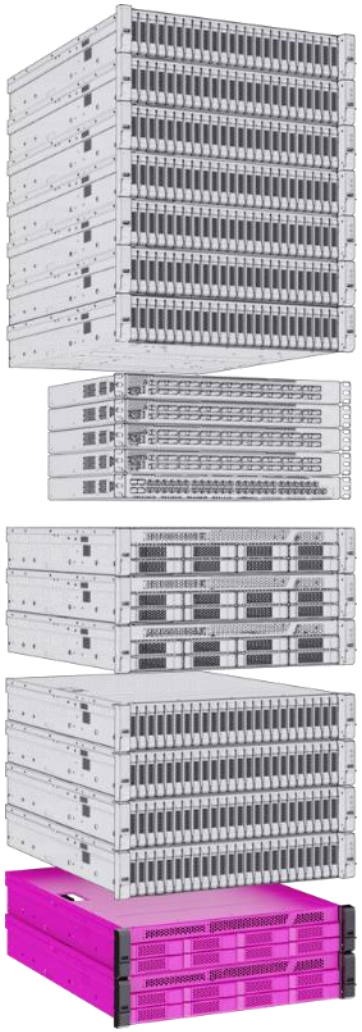
Модификации составляющих модулей:

- 2 размера базы
  - Неделя + инкременты
- 3 размера базы
  - Неделя + инкременты + текущий
- 4 размера базы
  - 2 недели + неделя + инкременты + текущий

Расположение:

- В стойках Машины равномерно

Блок резервного копирования



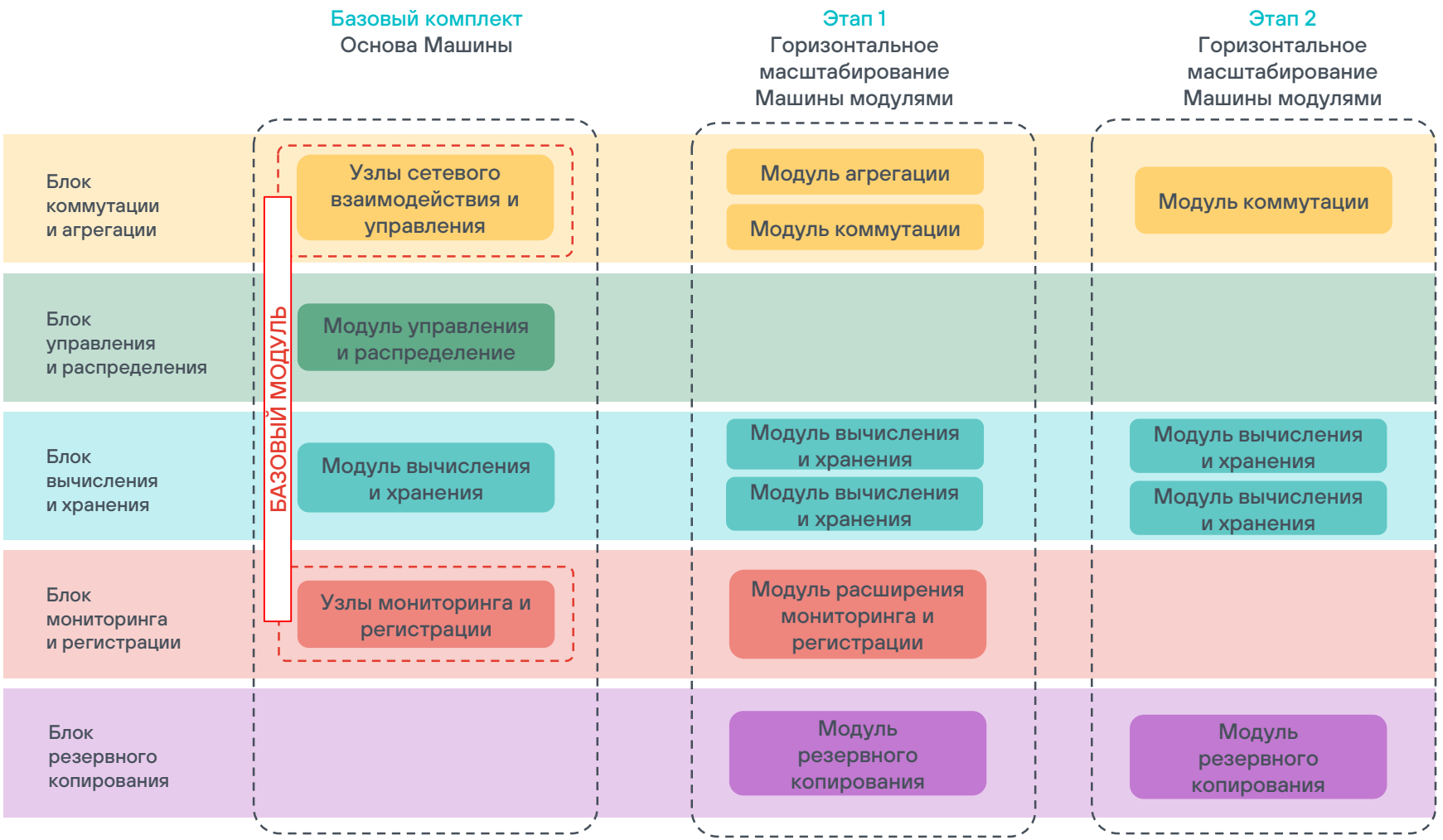
Применимость:

- Элемент, от которого можно отказаться с понижением надежности
- Возможно совмещение платформ для формирования теплого резерва (асинхронное копирование)
- Возможно использование для очень холодных данных

Особенности:

- Дисковое хранение
- RAID50
- Возможно подключение в выделенной параллельной сети
- Возможно иерархическое хранение (в разработке)

Принцип формирования состава Машин больших данных Скала^р по этапам поставки



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

Модуль — это единица поставки Машин Скала^р в составе спецификации

Техническая поддержка и услуги

Машины Скала^р поставляются с пакетами услуг технической поддержки:

техническая поддержка из «одного окна»

24x7

с поддержкой служб эксплуатации в круглосуточном режиме

возможность авансовой замены и ремонта оборудования по месту установки; опция невозврата накопителей с данными

1-5 лет

с возможностью продления

Круглосуточно

- 8-800-234-23-25
- tac@skala-r.ru
- личный кабинет Service Desk
- https://tac.scala-r.ru

В программу поддержки входит:

- решение инцидентов
- консультации по эксплуатации Машин
- предоставление обновлений ПО

Дополнительные профессиональные услуги

Программы дополнительных консультаций администрирования и эксплуатации Машин

Почему заказчики выбирают Скала^р

Глубокая интеграция и встречная оптимизация компонентов от платформенного ПО до микроконтроллеров:

- Высочайшая устойчивость
- Экстремальная производительность
- Стабильные показатели на предельных нагрузках

- Серийный выпуск, поддержка и сервисное обслуживание 24*7
- Быстрое развертывание и ввод в эксплуатацию
- Соответствие требованиям к критичным, высоконагруженным информационным системам
- Снижение совокупной стоимости владения (TCO)

Модульная платформа для высоконагруженных корпоративных и государственных информационных систем