

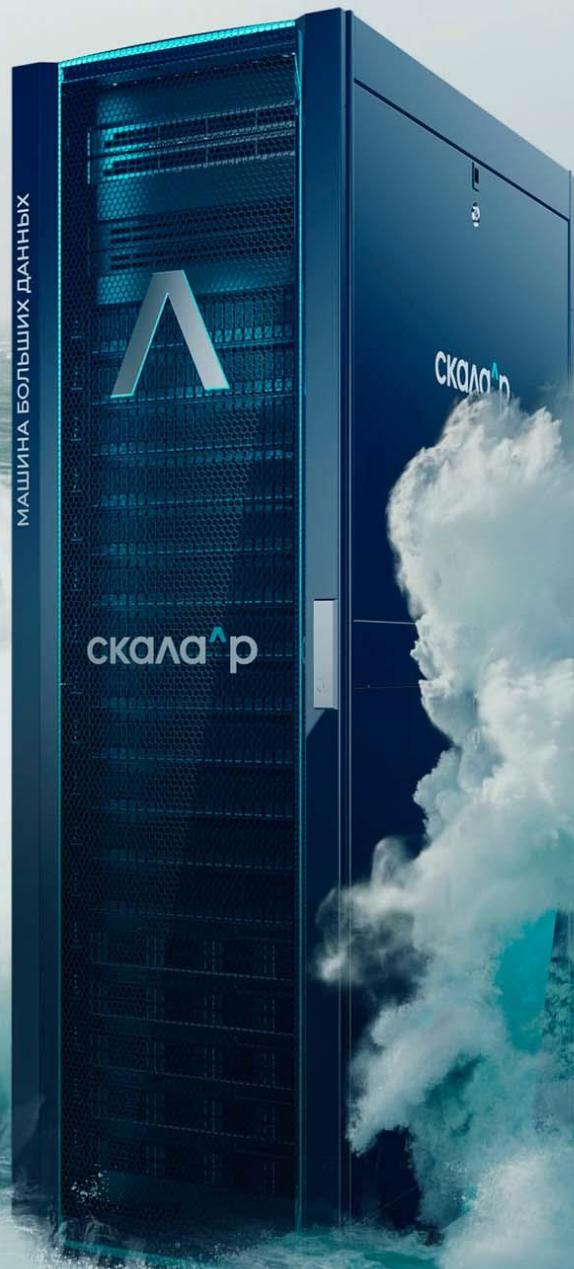


## Машина больших данных Скала^р МБД.Х

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

### Технический обзор

версия 3.0 от 15.12.2025



## ОГЛАВЛЕНИЕ

Перечень терминов и сокращений.....	6
1. Предисловие .....	10
1.1 Описание документа .....	10
1.2 Аудитория .....	10
1.3 Обратная связь .....	10
2. Введение .....	11
3. Отличительные черты .....	13
4. Подтвержденная безопасность.....	14
4.1 Операционные системы.....	14
4.2 Платформа .....	16
4.3 Средства защиты .....	16
5. Принципы создания Машины МБД.Х .....	18
5.1 Основные принципы.....	18
5.2 Хранение данных .....	19
5.2.1 HDFS .....	19
5.3 Компоненты ADH для управления ресурсами и координации.....	21
5.3.1 YARN.....	21
5.3.2 Zookeeper.....	21
5.4 Метаданные и каталогизация.....	22
5.4.1 Hive .....	22
5.5 Компоненты ADH для обработки данных.....	23
5.5.1 Spark .....	23
5.5.2 MapReduce .....	23
5.5.3 Tez.....	24
5.5.4 Flink .....	24
5.6 Базы данных и поиск.....	25
5.6.1 HBase .....	25
5.6.2 Solr .....	25
5.6.3 Zeppelin .....	26
5.7 Движки распределённых SQL-запросов .....	26

5.7.1 Impala .....	26
5.8 SQL-шлюзы и сервисы доступа.....	26
5.8.1 Phoenix.....	26
5.9 Компоненты ADH, обеспечивающие безопасность.....	27
5.9.1 Ranger .....	27
5.9.2 Knox .....	27
5.9.3 Kerberos .....	27
5.10 Управление, мониторинг и интерфейсы.....	27
5.10.1 Arenadata Cluster Manager (ADCM).....	27
5.10.2 Smart Storage Manager .....	28
5.10.3 Скала <sup>®</sup> Геном.....	29
5.10.4 Скала <sup>®</sup> Визион .....	29
5.11 Оркестрация рабочих процессов и загрузка данных .....	29
5.11.1 Airflow .....	29
5.11.2 Sqoop .....	30
6. Состав Машины .....	32
6.1 Подсистемы.....	32
6.1.1 Подсистема обеспечения базовых сервисов и Сетевая подсистема .....	34
6.1.2 Подсистема неструктурированной обработки больших данных .....	34
6.1.3 Подсистема управления .....	35
6.1.4 Подсистема резервного копирования .....	35
6.1.5 Подсистема управления данными .....	35
6.1.6 Подсистема преобразования данных .....	35
6.1.7 Подсистема управления доступом.....	36
6.1.8 Подсистема граничных функций .....	36
6.2 Модули.....	36
6.2.1 Базовый модуль .....	36
6.2.2 Модуль неструктурированной обработки .....	37
6.2.3 Модуль координации.....	38
6.2.4 Модуль резервного копирования .....	39
6.2.5 Специализированный модуль (для управления данными).....	40
6.2.6 Специализированный модуль (для задач преобразования данных) .....	41

---

6.2.7 Базовый модуль безопасности.....	41
6.2.8 Модуль граничных функций.....	42
7. Специфичные черты .....	44
8. Гарантируемое качество и полная готовность к промышленной эксплуатации .....	46
9. Реакция Машины на возможные отказы .....	48
10. Вариативность Машин .....	50
11. Требования к размещению Машины.....	52
12. Техническая поддержка .....	53
13. Лицензирование ПО в составе Модулей .....	55
13.1 Политика обновления ПО.....	55
О Компании .....	56

Документ носит исключительно информационный характер и является актуальным 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».

## ПЕРЕЧЕНЬ ТЕРМИНОВ И СОКРАЩЕНИЙ

<b>Термин, сокращение</b>	<b>Определение</b>
ACID	(англ. atomicity, consistency, isolation, durability) — набор требований к транзакционной системе, обеспечивающий наиболее надёжную и предсказуемую её работу — атомарность, согласованность, изоляцию, устойчивость
API	(англ. Application Programming Interface) — набор инструментов и протоколов для взаимодействия различных программ, позволяющий приложениям обмениваться данными и функциями без обозначения их внутренней структуры
Data Lake	«Озеро данных» — централизованное хранилище, в котором можно сохранять структурированные, полуструктурные и неструктурированные данные в исходном виде
Data Warehouse	«Хранилище данных» — это предметно-ориентированная, интегрированная, неизменяемая и поддерживающая хронологию совокупность данных, предназначенная для поддержки принятия управленческих решений, выполнения сложных аналитических запросов, агрегирующих большие объёмы исторических данных
Erasure coding	Механизм обеспечения отказоустойчивости и сохранности данных, при котором данные разбиваются на фрагменты, кодируются с добавлением избыточной информации и распределяются по различным узлам хранения, что позволяет восстановить исходные данные при потере части фрагментов, обеспечивая более высокую надёжность при меньших затратах на хранение по сравнению с репликацией
ETL	(англ. Extract, Transform, Load) — процесс, обеспечивающий перенос и подготовку данных для дальнейшего использования
JDBC	(англ. Java DataBase Connectivity) — платформенно независимый промышленный стандарт взаимодействия Java-приложений с различными СУБД
IoT	(англ. Internet of Things) — «интернет вещей», глобальная сеть взаимосвязанных устройств, которые способны обмениваться данными через интернет
MLAG	(англ. Multi-Switch Link Aggregation) — технология агрегации каналов, позволяющая одному или нескольким линкам с двух разных сетевых узлов быть объединенными вместе

Термин, сокращение	Определение
	таким образом, что для конечного устройства это выглядит как одиночное соединение
MPP	(англ. Massively Parallel Processing) — архитектура параллельных вычислительных систем, состоящих из множества независимых узлов, каждый из которых имеет собственную оперативную память, процессоры и диски. Обработка данных в такой системе (например, в аналитических базах данных) происходит путем одновременного выполнения операций на всех узлах, что позволяет очень быстро обрабатывать огромные объемы информации
NFS	(англ. Network File System) — протокол сетевого доступа к файловым системам
NL-SAS	Низкооборотный SAS — тип жёстких дисков, в котором сочетаются энергоэффективные и более дешёвые магнитные пластины от Nearline-дисков (используемых в больших массивах хранения) с высокопроизводительной и надежной SAS-электроникой. Оптимальное решение для создания экономичных систем хранения больших объёмов «холодных» и редко используемых данных, где важен баланс между стоимостью, ёмкостью и надёжностью
ODS	(англ. Operational Data Store) — предметно-ориентированная база данных, предназначенная для интеграции и хранения актуальных операционных данных из различных источников в почти реальном времени. Служит для выполнения оперативных запросов и отчётыности, занимая промежуточное положение между транзакционными системами и централизованным хранилищем данных (Data Warehouse)
RAID	(англ. Redundant Array of Independent Disks) — избыточный массив независимых накопителей, технология виртуализации данных для объединения нескольких физических дисковых устройств в логический модуль для повышения отказоустойчивости и производительности
SSD	(англ. Solid-State Drive) — тип накопителя данных, твердотельный, использующий флеш-память для хранения информации. Отличается компактностью и высокой скоростью чтения и записи данных
Ad-hoc-анализ	Гибкий, разовый, исследовательский метод анализа данных, направленный на решение конкретных, непредвиденных задач или поиск ответов на специфические вопросы, которые не предусмотрены в стандартных отчётах

Термин, сокращение	Определение
Бакетирование	Метод организации данных внутри партиции (или всей таблицы) для оптимизации запросов, особенно слияний и выборок. Данные распределяются по фиксированному количеству "бакетов" на основе хэш-функции от значений одного или нескольких столбцов. В отличие от партиционирования, которое создаёт отдельные папки (директории) для разных диапазонов и значений данных, бакетирование группирует данные в файлы внутри одной партиции, что позволяет эффективно отсекать ненужные данные при сканировании и ускорять операции соединения таблиц
БД	«База данных» — организованная совокупность структурированных данных, хранящаяся в электронном виде и управляемая системой управления базами данных (СУБД). Обеспечивает эффективный способ хранения, обработки, изменения и извлечения информации по запросу пользователей или приложений
ГИС	Государственные информационные системы — системы, которые создаются для реализации полномочий государственных органов и обеспечения обмена информацией между ними, а также в иных установленных федеральными законами целях
ЗОКИИ	Значимый объект критической информационной инфраструктуры
ИСПДн	Информационные системы персональных данных. Совокупность информации, содержащейся в базах данных, и обеспечивающих её обработку с использованием информационных технологий и технических средств
Кластер	Отказоустойчивая архитектура функционала Машины
Машина	Набор аппаратного и программного обеспечения в виде Модулей Скала <sup>®</sup> , соединённых вместе для обеспечения определённого метода обработки данных или предоставления ИТ-сервиса с заданными характеристиками. Зарегистрирована в ЕПРРП
Модуль	Функционально завершенный комплект сконфигурированных для выполнения заданных функций аппаратных и/или программных компонентов, аппаратных узлов и программного обеспечения (ПО), оформленный как самостоятельная единица продаж со своим кодом (part number) и стоимостью. Является единым и неделимым элементом спецификации. Зарегистрирован в ЕПРРП
Мультитенантность	Предоставление изолированного доступа к общим ресурсам разным арендаторам, то есть тенантам. Это

Термин, сокращение	Определение
	свойство программного обеспечения, которое позволяет нескольким пользователям или организациям использовать одну и ту же программную систему или приложение. Это означает, что каждый пользователь имеет свой собственный набор данных и настроек, но все они работают на одной и той же программной платформе. Это позволяет экономить ресурсы и упрощает управление системой, так как администратор может управлять всеми пользователями и их данными в единой системе
ОС	Операционная система
ПАК	Программно-аппаратный комплекс
ПО	Программное обеспечение
Подсистема	Логическое объединение компонентов по функциональному признаку, с целью пояснения состава и принципов действия ПАК
Партиционирование	Разбиение больших таблиц на более мелкие части для ускорения запросов и оптимизации работы базы данных
СУБД	Система управления базами данных
Узел	Вычислительный узел или сетевой узел (коммутатор) в составе Модуля, в зависимости от контекста
Шардирование	Разделение баз данных на независимые сегменты, каждый из которых управляется экземпляром СУБД на отдельном узле, обеспечивающее распределённые вычисления. Позволяет улучшить масштабирование и повысить отзывчивость сервиса. Большая база разделяется на несколько частей, называемых шардами, распределяемых по другим узлам и связываемых в одну систему
Экземпляр службы	Отдельная, запущенная копия программного обеспечения, выполняющая конкретную служебную функцию в рамках программных компонентов ПАК

## 1. ПРЕДИСЛОВИЕ

### 1.1 Описание документа

Этот технический обзор дает концептуальный и архитектурный обзоры **Машины больших данных Скала<sup>®</sup> МБД.Х.**

Документ описывает то, как оптимизированные программно-аппаратные комплексы отвечают современным вызовам, и фокусируется на **Машине больших данных Скала<sup>®</sup> МБД.Х.**

### 1.2 Аудитория

Эта брошюра предназначена для сотрудников компании Скала<sup>®</sup>, партнёров и Заказчиков, перед которыми ставятся задачи разработки, закупки, управления или эксплуатации **Машины больших данных Скала<sup>®</sup> МБД.Х.**

### 1.3 Обратная связь

Скала<sup>®</sup> и авторы этого документа будут рады обратной связи по нему.

Свяжитесь с командой Скала<sup>®</sup> по электронной почте [MBD8@skala-r.ru](mailto:MBD8@skala-r.ru).

## 2. ВВЕДЕНИЕ

**Машина больших данных Скала<sup>®</sup>р МБД.Х** — это программно-аппаратный комплекс, предназначенный для создания централизованного и безопасного хранилища данных (Data Lake), консолидирующего информацию любой структуры (структурную, слабоструктурированную, неструктурированную) в объемах от терабайт до петабайт. Комплекс — сертифицированный и готовый к промышленной эксплуатации продукт, построенный на базе отечественной платформы Arenadata Hadoop (ADH).

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

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

### Ключевые экономические преимущества

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

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

### Ключевые технологические преимущества

- **Единая платформа для всех данных:** консолидация сырых данных из различных источников (ERP, CRM, логи, датчики IoT) в распределённой файловой системе HDFS с возможностью их последующего преобразования, анализа и предоставления через единые интерфейсы (Hive, Impala)
- **Готовая аналитическая среда:** возможность выполнения сложных запросов к петабайтам данных за секунды (Impala), исследовательского анализа и

визуализации (Zeppelin), а также построения воспроизводимых рабочих процессов обработки данных (Airflow) и их интеграции с внешними системами (Sqoop)

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

Программно-аппаратные комплексы **Скала<sup>®</sup>** позволяют организациям быстро развернуть полнофункциональную, безопасную и масштабируемую платформу для работы с большими данными, соответствующую требованиям регуляторов и включённую в соответствующие реестры отечественной продукции. Программно-аппаратные комплексы **Скала<sup>®</sup>** включены в Единый реестр российской радиоэлектронной продукции и работают на ПО, включённом в реестр Минцифры РФ.

### 3. ОТЛИЧИТЕЛЬНЫЕ ЧЕРТЫ

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

- Продукты Скала<sup>▲</sup>р являются серийно выпускаемыми преднастроенными **Машинами** и позволяют осуществлять быстрое развёртывание 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<sup>▲</sup>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<sup>▲</sup>р** позволяет обеспечить снижение совокупной стоимости владения (TCO)
- **Машины** обладают широкими возможностями масштабируемости, способны хранить и обрабатывать терабайты данных
- Механизмы распределённой 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<sup>▲</sup>р для управления эксплуатацией и мониторинга работы компонентов **Машины**, а также использование средств управления от компании Arenadata, позволяет управлять параметрами СУБД, сервисами и приложениями, входящими в состав комплекса

## 4. ПОДТВЕРЖДЕННАЯ БЕЗОПАСНОСТЬ

### 4.1 Операционные системы

Машина больших данных Скала<sup>▲</sup>р МБД.Х поставляется с одной из сертифицированных операционных систем:

- **ОС Альт 8 СП** (сертификат ФСТЭК 3866 от 10.08.2018, действует до 10.08.2028)
- **РЕД ОС** (сертификат ФСТЭК №4060 от 12.01.2019, действует до 12.01.2029)
- **Astra Linux Special Edition** (сертификат ФСТЭК №2557 от 27.01.2012, действует до 27.01.2026)

#### ОС Альт 8 СП

Может применяться для защиты информации в:

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

ОС соответствует требованиям следующих нормативных документов:

- «Требования безопасности информации к операционным системам» (ФСТЭК России, 2016) и «Профиль защиты операционных систем типа А четвертого класса защиты. ИТ.ОС.А4.П3» (ФСТЭК России, 2017) по 4 классу защиты
- «Требования по безопасности информации к средствам контейнеризации» (ФСТЭК России, 2022) по 4 классу защиты
- «Требования по безопасности информации к средствам виртуализации» (ФСТЭК России, 2022) по 4 классу защиты
- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) по 4 уровню доверия

#### РЕД ОС

Может применяться для защиты информации в:

- значимых объектах критической информационной инфраструктуры 1 категории
- государственных информационных системах 1 класса защищённости

- в автоматизированных системах управления производственными и технологическими процессами 1 класса защищённости
- в информационных системах персональных данных при необходимости обеспечения 1 уровня защищённости персональных данных
- в информационных системах общего пользования 2 класса

ОС соответствует требованиям следующих нормативных документов:

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

### Astra Linux Special Edition

Может применяться для защиты информации в:

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

ОС соответствует требованиям следующих нормативных документов:

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

- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) по 1 уровню доверия

## 4.2 Платформа

Машина больших данных Скала<sup>▲</sup>р МБД.Х использует сертифицированную платформу **Arenadata Hadoop** (Сертификат ФСТЭК 4821 от 13.06.2024, действует до 13.06.2029).

Платформа может применяться для защиты информации в:

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

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

## 4.3 Средства защиты

**Обеспечивается полная совместимость с наложенными  
средствами защиты**

Сертифицированная система единой аутентификации **Avanpost FAM** (сертификат ФСТЭК 4492 от 13.12.2021, действует до 13.12.2026) соответствует требованиям нормативного документа «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020, приказ № 76) по 1 уровню доверия.

Сертифицированное антивирусное средство защиты **Kaspersky Endpoint Security для Linux** (сертификат ФСТЭК 2534 от 27.12.2011, действует до 27.12.2025) соответствует требованиям следующих нормативных документов:

- «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» (ФСТЭК России, 2020) — по 2 уровню доверия
- «Требования к средствам антивирусной защиты» (ФСТЭК России, 2012)
- «Профиль защиты средств антивирусной защиты типа Б 2 класса защиты. ИТ.САВ3.Б2.13» (ФСТЭК России, 2012)
- «Профиль защиты средств антивирусной защиты типа В второго класса защиты. ИТ.САВ3.В2.П3» (ФСТЭК России, 2012)
- «Профиль защиты средств антивирусной защиты типа Г второго класса защиты»

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 класса защищенности включительно и в ЗОКИИ до 1 категории включительно.

## 5. ПРИНЦИПЫ СОЗДАНИЯ МАШИНЫ МБД.Х

### 5.1 Основные принципы

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

#### Технологические принципы

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

#### Технические решения

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

#### Спрогнозированная нагрузка

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

## Выделенная сеть внутреннего взаимодействия

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

## 5.2 Хранение данных

### 5.2.1 HDFS

HDFS является основной распределённой файловой системой для хранения файлов большого размера, которые разбиваются на блоки и распределяются между вычислительными узлами. Каждый узел, хранящий данные, работает исключительно со своими собственными накопителями. Гарантия целостности информации в HDFS обеспечивается встроенными механизмами самой файловой системы. В процессе записи для каждого блока данных вычисляется специальная контрольная сумма, которая сохраняется вместе с ним. При дальнейшем чтении данных выполняется проверка этих контрольных сумм. Если обнаруживается повреждение какого-либо блока, система автоматически запускает процедуру его восстановления из соответствующей копии. Все эти операции выполняются автоматически, без необходимости вмешательства со стороны прикладных программ.

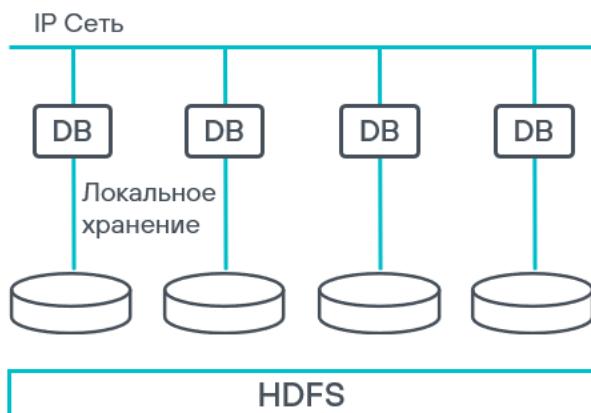


Рисунок 1. Архитектура без общих ресурсов (*shared nothing*)

Рисунки 1 и 2 отображают архитектуру HDFS. В контексте HDFS стойка — это отображение физического размещения вычислительных узлов и сетевой топологии. Правильная настройка позволяет кластеру идентифицировать, какие узлы находятся в одной физической стойке. Это знание используется политикой размещения реплик данных для их распределения между разными стойками, что обеспечивает устойчивость к потере целой стойки оборудования.

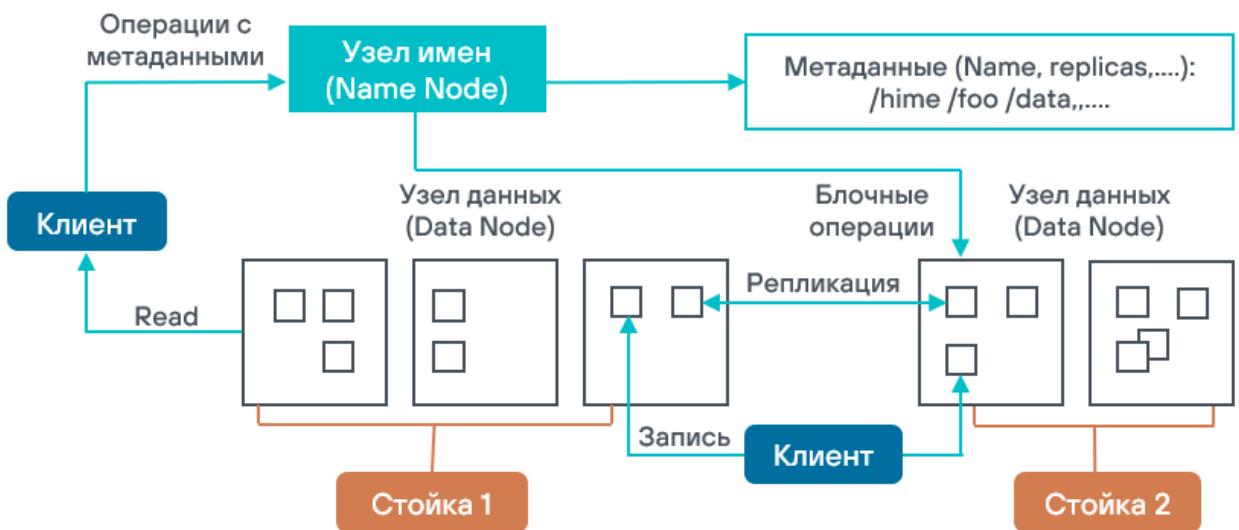


Рисунок 2. Архитектура HDFS

## Репликация данных

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

- узел имён (NameNode) выбирает новые узлы данных для размещения реплик
- узел имён выполняет балансировку размещения данных по узлам и составляет список узлов для репликации. Стратегия размещения копий данных в распределённой файловой системе спроектирована для гарантии сохранности информации при сбоях оборудования, включая отказ всей стойки. Блоки данных распределяются по вычислительным узлам, физически установленным в разных стойках
- ключевой принцип — минимизация риска одновременной потери нескольких копий одного блока. Кластер анализирует сетевую топологию, «понимая», какие узлы находятся в одной стойке, а какие — в разных. Это позволяет автоматически применять строгие правила, которые обеспечивают высочайший уровень отказоустойчивости и оптимальное использование пропускной способности внутренней сети кластера

Стандартный алгоритм размещения трёх копий блока работает следующим образом:

- первая копия записывается на тот узел, который первоначально принял данные от клиента
- вторая копия отправляется на любой другой узел, расположенный в стойке, отличной от стойки первого узла. Это защищает данные от полного выхода из строя одного шкафа с оборудованием
- третья копия создаётся на дополнительном узле, расположенном в той же стойке, что и узел со второй копией. Таким образом, данные защищены как от отказа отдельного узла, так и от потери всей стойки, поскольку копии распределены как минимум между двумя различными стойками

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

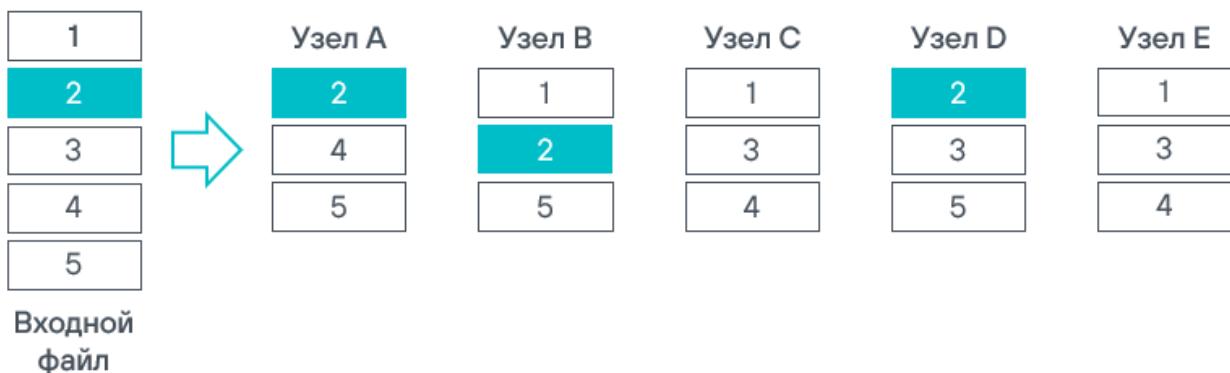


Рисунок 3. Схема распределения данных на HDFS

## 5.3 Компоненты ADH для управления ресурсами и координации

### 5.3.1 YARN

Фреймворк для централизованного управления вычислительными ресурсами (процессорное время, оперативная память) ПАК. YARN позволяет различным движкам обработки данных, таким как MapReduce, Spark, Flink и другим, работать совместно и эффективно, выделяя каждому приложению необходимые для его работы ресурсы.

#### Ключевые свойства

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

### 5.3.2 Zookeeper

Zookeeper — это надёжный и высокопроизводительный сервис координации для распределённых систем, который предоставляет простой программный интерфейс и гарантирует согласованность состояния в кластере.

## Ключевые свойства

- **Обеспечение отказоустойчивости:** сервис разворачивается в виде группы из нечётного числа рабочих экземпляров. Такая группа сохраняет работоспособность и способна принимать обновления даже при выходе из строя её части
- **Поддержание общего состояния:** предоставляет иерархическое пространство имён для хранения небольших порций конфигурационной информации и состояния системы. Этот реестр хранится в памяти всех участников группы, что обеспечивает высокую скорость доступа
- **Механизм уведомлений об изменениях:** клиенты могут подписываться на изменения в определённых узлах реестра, что позволяет эффективно реализовывать такие шаблоны распределённых систем, как выбор ведущего узла, управление блокировками и отслеживание доступности сервисов
- **Гарантии надёжности и порядка:** используется собственный алгоритм достижения консенсуса, который обеспечивает атомарность выполнения операций, их строгий порядок и надёжную запись на диск большинством участников группы

## 5.4 Метаданные и каталогизация

### 5.4.1 Hive

В составе **Машины** Hive предоставляет критически важный сервис работы с метаданными — Hive Metastore. Сервис Hive Metastore выполняет роль централизованного системного каталога, в котором хранятся определения таблиц, их схемы, а также информация о физическом расположении данных в хранилище и формате файлов, например колоночных ORC или Parquet. Каталог создает уровень абстракции, представляющий файлы в виде структурированных таблиц с определенной схемой. Эта абстракция используется многими компонентами экосистемы, включая Spark и Hive, для выполнения SQL-запросов и доступа к данным. Современные движки обработки, такие как Impala, также могут работать с данными, зарегистрированными в Hive Metastore, считывая метаданные о схеме и формате для их корректной интерпретации.

## Ключевые свойства

- Управление метаданными: централизованное хранилище схем и информации о таблицах для всей экосистемы
- Пакетная обработка: выполнение сложных ETL-задач с использованием декларативного SQL-подобного языка
- Поддержка структурированных и слабоструктурированных данных
- Интеграция с HDFS для хранения данных и с Hive Metastore для управления их структурой
- Возможность создания таблиц,パーティционирования и бакетирования
- Расширяемость через пользовательские функции (UDF)
- Оптимизация выполнения через интеграцию с фреймворками Tez и Spark для повышения скорости обработки

## 5.5 Компоненты ADH для обработки данных

### 5.5.1 Spark

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

#### Ключевые свойства

- Поддержка разнородных типов задач (пакетная обработка, стриминг, машинное обучение, графовые вычисления) в единой платформе
- Богатые API на Java, Scala, Python и R
- Встроенные библиотеки (Spark SQL, MLlib, Structured Streaming, GraphX)
- Оптимизация запросов через Catalyst Optimizer и Tungsten Engine
- Возможность работы с различными источниками данных (HDFS, S3, JDBC)
- Отказоустойчивость и масштабируемость на кластерах из тысяч узлов
- Решение бизнес-задач: высокая скорость обработки информации непосредственно в оперативной памяти позволяет сокращать время выполнения сложных процессов извлечения, преобразования и загрузки данных с часов до минут

### 5.5.2 MapReduce

Фундаментальная модель программирования и фреймворк для параллельной обработки больших объёмов данных, сыгравшая ключевую роль в становлении экосистемы Hadoop. Несмотря на появление более быстрых и удобных фреймворков, таких как Spark, которые стали стандартом для большинства задач пакетной обработки, MapReduce сохраняется в экосистеме. Его основное современное назначение — обеспечение обратной совместимости с устаревшими приложениями и инструментами, которые были разработаны напрямую для этой модели, а также выполнение узкого круга специфических фоновых операций, где его поведение было тщательно изучено и настроено. Для новых проектов и задач рекомендуется использовать современные фреймворки, такие как Spark.

#### Ключевые свойства

- Доказанная надежность и устойчивость к сбоям при работе с петабайтами данных. Архитектура фреймворка изначально создана для работы в условиях постоянных отказов оборудования. Он автоматически перезапускает упавшие задачи и гарантирует завершение обработки даже при выходе из строя отдельного узла
- Максимальная эффективность для задач, где главное — работа с накопителями. MapReduce оптимизирован для сценариев, когда скорость обработки ограничена не процессором или памятью, а пропускной способностью чтения и записи на диски. Это делает его часто более предпочтительным выбором для определённых

фоновых операций с данными, где важна общая надёжность, а не минимальная задержка

- Простая и мощная модель программирования через два ключевых этапа. Модель Сопоставление-Сокращение позволяет разработчикам описать сколь угодно сложный алгоритм пакетной обработки, разбив его на эти две хорошо понятные фазы, которые система затем параллелизует на всем кластере
- Фундаментальная основа экосистемы. MapReduce сыграл ключевую роль в становлении экосистемы Hadoop. Такие инструменты, как Hive и Sqoop, изначально создавались с опорой на MapReduce в качестве движка выполнения, что демонстрирует его историческую значимость. В современных конфигурациях эти системы поддерживают альтернативные, более производительные движки, но MapReduce остаётся доступным и проверенным вариантом для специфических задач пакетной обработки

### 5.5.3 Tez

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

#### Ключевые свойства

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

### 5.5.4 Flink

Система для параллельной обработки информации, созданная специально для работы с непрерывными потоками данных с минимально возможной задержкой. Её ключевой особенностью является единый подход к работе как с потоками, так и с пакетами данных, где обработка потока рассматривается как базовая операция, а пакетная обработка — как её частный случай.

#### Ключевые свойства

- Построения систем, работающих в реальном времени: непрерывного наблюдения за показателями, выявления аномалий и мошеннических схем, анализа сложных цепочек событий в потоке данных и обработки транзакционных операций
- Обработка ровно один раз — система гарантирует, что каждое событие в потоке данных будет учтено и обработано точно один раз, даже в случае сбоев
- Встроенная поддержка временных меток событий — позволяет корректно анализировать данные, поступившие с задержкой, и определять правильный

порядок событий на основе времени их возникновения, а не времени поступления в систему

- Обеспечение высокой пропускной способности и устойчивости к сбоям за счёт механизма контрольных точек — периодического автоматического сохранения текущего состояния вычислений в надёжном хранилище, что позволяет быстро восстановить работу после отказа
- Богатые возможности для сложных преобразований и работы с временными окнами — предоставляет развитые инструменты для агрегации и анализа данных за определённые периоды времени

## 5.6 Базы данных и поиск

### 5.6.1 HBase

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

#### Ключевые свойства

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

### 5.6.2 Solr

Масштабируемая платформа поиска и аналитики в реальном времени,строенная на базе Lucene. Помимо классического полнотекстового поиска, Solr предоставляет мощные возможности для индексации, поиска и аналитики по структурированным и слабоструктурированным данным из Hadoop. Ключевые функции включают сложную фильтрацию, фасетный поиск для агрегации и анализа данных по категориям, статистику, выделение результатов и автодополнение. Solr интегрируется с экосистемой Hadoop, поддерживая загрузку данных напрямую из HDFS, что делает его эффективным инструментом для построения поисковых интерфейсов поверх больших данных, аналитических панелей и систем мониторинга.

#### Ключевые свойства

- Мощный полнотекстовый поиск с поддержкой сложных запросов и ранжирования
- Горизонтальная масштабируемость за счет разделения индексов на независимые сегменты и их последующего копирования на разные узлы

- Функции фасетного поиска — это механизм для категоризации и фильтрации результатов поиска по различным свойствам данных, который позволяет пользователям интерактивно уточнять результаты. Помимо этого, платформа предоставляет возможности выделения найденных фрагментов в тексте и автоматического дополнения вводимых запросов.
- Интеграция с Hadoop (загрузка данных из HDFS)
- REST API и расширяемая архитектура за счёт плагинов

### 5.6.3 Zeppelin

Веб-интерфейс для интерактивной аналитики и совместной работы с данными. Позволяет создавать интерактивные запросы на SQL, Python, Scala и R, визуализировать результаты в виде графиков и диаграмм, а также делиться интерактивными документами (блокнотами) с коллегами. Интегрируется с Spark, Hive, Impala и другими движками обработки данных, что делает его удобным инструментом для исследовательского анализа и прототипирования.

## 5.7 Движки распределённых SQL-запросов

### 5.7.1 Impala

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

#### Ключевые свойства

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

## 5.8 SQL-шлюзы и сервисы доступа

### 5.8.1 Phoenix

SQL-слой поверх HBase, предоставляющий реляционный интерфейс для работы с данными через стандартный JDBC/ODBC-интерфейс. Позволяет использовать знакомый SQL-синтаксис для оперативной аналитики (OLTP) и взаимодействия с данными в HBase.

#### Ключевые свойства

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

- Полная совместимость с метаданными HBase

## 5.9 Компоненты ADH, обеспечивающие безопасность

### 5.9.1 Ranger

Фреймворк для централизованного управления безопасностью в экосистеме Hadoop. Обеспечивает соответствие требованиям безопасности и регулирующих стандартов в корпоративных средах, позволяя гибко управлять правами доступа.

#### Ключевые свойства

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

### 5.9.2 Knox

Шлюз безопасности (API Gateway) для экосистемы Hadoop, обеспечивающий единую точку входа для внешних приложений. Позволяет безопасно предоставлять доступ к сервисам Hadoop извне периметра кластера, не нарушая его внутреннюю безопасность.

#### Ключевые свойства

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

### 5.9.3 Kerberos

Протокол сетевой аутентификации, используемый для обеспечения безопасного доступа к сервисам экосистемы Hadoop. Обеспечивает взаимную аутентификацию клиентов и сервисов, предотвращая несанкционированный доступ и подмену узлов. Интегрируется с Ranger, Knox и внешними LDAP-каталогами для построения многоуровневой системы безопасности.

## 5.10 Управление, мониторинг и интерфейсы

### 5.10.1 Arenadata Cluster Manager (ADCM)

Инструмент для комплексного управления жизненным циклом кластера на базе Arenadata Hadoop. ADCM предоставляет единый веб-интерфейс и API для развёртывания, настройки, мониторинга, обновления и обслуживания всех компонентов распределённой системы.

## Ключевые свойства

- **Централизованное развертывание и конфигурация:** позволяет установить и настроить весь кластер, включая все сервисы Hadoop и их зависимости, из единой точки управления, что сокращает время развертывания и минимизирует человеческие ошибки
- **Управление конфигурациями:** обеспечивает централизованное хранение и версионирование конфигурационных файлов для всех сервисов кластера. Позволяет применять изменения конфигурации ко множеству узлов одновременно и выполнять откат при необходимости
- **Мониторинг работоспособности:** предоставляет сводную панель состояния кластера, отображающую статус всех сервисов и узлов, что позволяет оперативно выявлять проблемы
- **Управление обновлениями:** автоматизирует процесс обновления версий ПО кластера, управляя зависимостями и корректной последовательностью действий для каждого сервиса
- **Управление пользователями и безопасностью:** интегрируется с системами безопасности (такими как Kerberos и Ranger), упрощая настройку аутентификации и авторизации в кластере
- **Управление процедурами обслуживания:** предоставляет готовые рабочие процессы для выполнения стандартных операций, таких как добавление новых узлов в систему или временный вывод узлов на техническое обслуживание без остановки работы сервисов
- **Ролевое управление доступом:** поддерживает разграничение прав доступа для администраторов кластера и операторов в соответствии с их обязанностями

### 5.10.2 Smart Storage Manager

Предоставляет веб-интерфейс, с помощью которого можно создавать правила, запускать действия, проверять состояние их выполнения и следить за показателями кластера.

## Ключевые свойства

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

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

- сжатие данных в HDFS без ограничения доступа к ним для внешних приложений

### 5.10.3 Скала<sup>®</sup> Геном

Обеспечивает поддержку жизненного цикла ПАК.

#### Ключевые свойства

- Предоставление доступа к IPMI всех узлов ПАК
- Вывод узла в режим обслуживания
- Замена вышедшего из строя узла
- Формирование паспорта ПАК

### 5.10.4 Скала<sup>®</sup> Визион

Позволяет контролировать жизненный цикл ПАК, мониторить программные компоненты, собирать информацию о количестве и составе управляемых объектов.

#### Ключевые свойства

- Сбор данных о конфигурации элементов ПАК
- Сбор данных, отображение, мониторинг элементов программного обеспечения, активных компонентов модулей ПАК, служебных сервисов и сервисов баз данных
- Конфигурирование метрик мониторинга, настройка уведомлений
- Конфигурирование графического отображения на информационных панелях в виде графиков, отдельных значений, диаграмм, таблиц
- Хранение метрик с возможностью настройки глубины хранения и управления жизненным циклом хранимых данных
- Отображение в пользовательском графическом интерфейсе данных о состоянии объектов мониторинга
- Мониторинг сервисов, специфичных для различных типов ПАК

## 5.11 Оркестрация рабочих процессов и загрузка данных

### 5.11.1 Airflow

Платформа для оркестрации рабочих процессов (workflow), позволяющая описывать, планировать и мониторить сложные пайплайны обработки данных. Интегрируется с компонентами Hadoop (Spark, Hive, HDFS) и поддерживает возможность повторного выполнения задач, что критично для надёжности ETL-процессов.

#### Ключевые свойства

- **Декларативное описание пайплайнов:** рабочие процессы определяются на языке Python, что обеспечивает гибкость, управляемость версиями и возможность динамической генерации

- **Масштабируемая архитектура:** Машина использует модульную архитектуру, что позволяет гибко наращивать вычислительные мощности для выполнения задач и гарантировать высокий уровень доступности для критически важных компонентов
- **Богатый набор операторов:** включает предустановленные операторы для интеграции с ключевыми компонентами экосистемы Hadoop (HDFS, Spark, Hive) и внешними системами (базы данных, облачные сервисы, HTTP-запросы), а также возможность создания пользовательских операторов
- **Глубокий мониторинг и управление:** предоставляет веб-интерфейс для визуализации зависимостей задач, мониторинга их выполнения в реальном времени, ручного запуска/остановки, просмотра логов и анализа производительности
- **Политика повторных попыток и оповещения:** гибкая настройка правил повторного выполнения задач при сбоях, интеграция с системами оповещения для уведомления о статусе рабочих процессов
- **Поддержка временных зон и расписаний:** возможность задания сложных расписаний запуска заданий с учётом временных зон, что критично для кросс-региональных и кросс-платформенных процессов
- **Безопасность и управление доступом:** встроенная ролевая модель, поддержка аутентификации через LDAP/AD, OAuth, возможность изоляции выполнения задач для мультитенантных сред

## 5.11.2 Sqoop

Инструмент для переноса объёмных структурированных данных между реляционными базами данных и платформой Hadoop. Для выполнения операций импорта и экспорта Sqoop использует распределённые вычислительные фреймворки экосистемы, такие как MapReduce, что обеспечивает высокую пропускную способность при работе с большими объёмами информации.

### Ключевые свойства

- **Высокопроизводительное копирование:** разбивает задачу на множество параллельных процессов, что позволяет эффективно использовать вычислительные ресурсы кластера Hadoop для ускорения передачи терабайтов данных
- **Автоматическое отображение схемы данных:** автоматически генерирует схемы данных в Hive и HBase на основе исходных данных, упрощая процесс интеграции
- **Инкрементальный импорт:** поддержка импорта только новых или изменённых данных с момента последней операции, что снижает нагрузку на источники данных и ускоряет обновление хранилища
- **Сжатие данных при передаче:** возможность сжатия данных на лету с использованием эффективных кодеков снижает нагрузку на сеть и объём занимаемого хранилища в HDFS

- **Безопасное подключение к источникам:** поддержка безопасных соединений с СУБД, возможность хранения паролей в защищённом хранилище и интеграция с Kerberos для аутентификации
- **Оптимизация для различных СУБД:** использует нативные JDBC-драйверы и может применять специфичные для СУБД оптимизации
- **Экспорт данных из Hadoop:** возможность обратного переноса результатов обработки из HDFS или Hive в реляционные СУБД для интеграции с бизнес-приложениями и отчётностью

## 6. СОСТАВ МАШИНЫ

Ниже приведены термины, используемые для комплектации **Машины больших данных Скала<sup>®</sup> МБД.Х.**

**Машина** — это набор аппаратного и программного обеспечения в виде **Модулей Скала<sup>®</sup>**, соединенных вместе для обеспечения определенного метода обработки данных или предоставления ИТ-сервисов с заданными характеристиками.

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

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

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

### Комплекты поставки

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

Базовый комплект — это набор **Модулей Скала<sup>®</sup>**, минимально необходимый для функционирования всех подсистем, обеспечивающих выполнение основного функционала **Машины**.

Комплект модулей расширения — это набор **Модулей Скала<sup>®</sup>**, позволяющий, увеличить производительность, объём хранения и портовую ёмкость. Кроме того, можно добавить специальные **Модули Скала<sup>®</sup>**, позволяющие расширить функциональность ПАК.

### 6.1 Подсистемы

Функции **Машины больших данных Скала<sup>®</sup> МБД.Х** логически объединены в подсистемы. Часть подсистем обеспечивают основной функционал и всегда включены в **Машину**, а часть — дополнительный функционал и могут быть добавлены по требованию Заказчика.

Основной функционал — это минимальный набор подсистем, необходимых **Машине больших данных Скала<sup>®</sup> МБД.Х** для выполнения задач прямого назначения.

Дополнительный функционал — набор подсистем из Модулей, обеспечивающих расширение функций **Машины больших данных Скала<sup>®</sup> МБД.Х**.

## Машина больших данных Скала<sup>®</sup> МБД.Х. Технический обзор

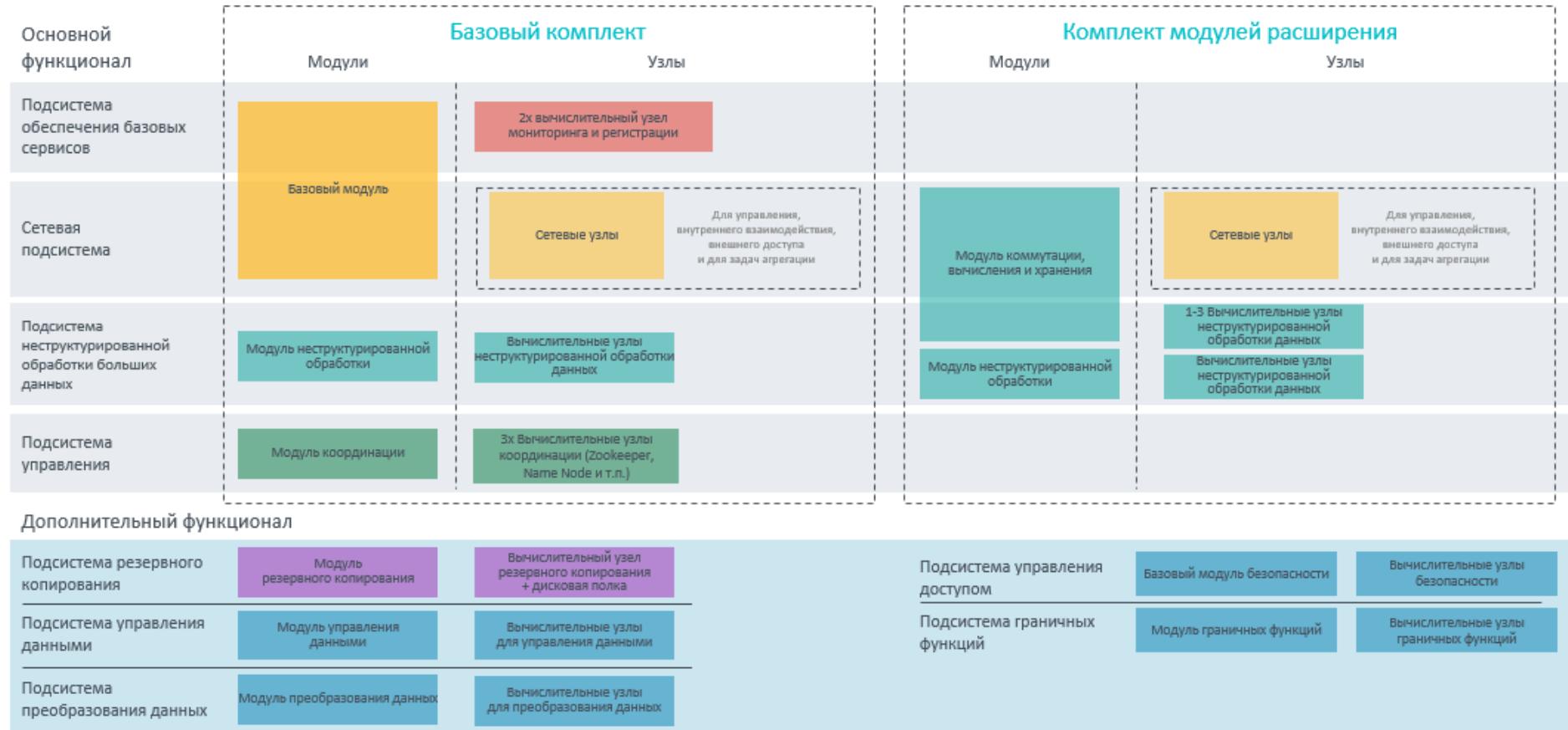


Рисунок 4. Комплектация Машины больших данных Скала<sup>®</sup> МБД.Х

### 6.1.1 Подсистема обеспечения базовых сервисов и Сетевая подсистема

Подсистема обеспечения базовых сервисов отвечает за мониторинг и управление аппаратными и программными компонентами **Машины больших данных Скала<sup>▲</sup>р МБД.Х**. В неё включены вычислительные узлы **Базового модуля** (см. п. 6.2.1), на которых предустановлено сервисное ПО **Скала<sup>▲</sup>р Геном** и **Скала<sup>▲</sup>р Визион**, выполняющее следующие функции:

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

Архитектура подсистемы обеспечения базовых сервисов обеспечивает отказоустойчивый режим работы.

Сетевая подсистема выполняет функций организации сетевой связности между всеми вычислительными узлами, входящими в состав **Машины больших данных Скала<sup>▲</sup>р МБД.Х**, и представляет собой набор сетевых узлов, которые организуют изолированные высокоскоростные сети:

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

Стартовый комплект сетевых узлов всегда размещается в **Базовом модуле** (см. п. 6.2.1).

### 6.1.2 Подсистема неструктурированной обработки больших данных

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

- масштабируемая и отказоустойчивая файловая система HDFS
- обработка огромных объёмов данных из разных источников с использованием распределённых федеративных запросов
- эффективное хранение данных разной температуры (горячие, тёплые, холодные) за счёт механизмов репликации и кодирования с стиранием (Erasure coding) в HDFS
- предоставление технических метаданных (определения таблиц, схемы, форматы файлов) через сервис Hive Metastore, что является основой для работы с данными

в высокопроизводительных колоночных форматах, таких как ORC и Parquet, с помощью инструментов обработки, например, Spark и Impala

- поиск данных с помощью Solr

Подсистема реализуется **Модулем неструктурированной обработки** (см. п. 6.2.2).

## 6.1.3 Подсистема управления

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

Подсистема реализуется **Модулем координации** (см. п. 6.2.3).

## 6.1.4 Подсистема резервного копирования

Подсистема резервного копирования предназначена для создания согласованных копий критических данных платформы и их конфигураций с целью аварийного восстановления. Она включает один или несколько **Модулей резервного копирования** (см. п. 6.2.4), которые предоставляют выделенное отказоустойчивое сетевое хранилище (например, по протоколу NFS) в качестве целевого для архивов.

Резервное копирование выполняется с помощью специализированных инструментов платформы Arenadata Hadoop и может включать многоуровневую стратегию:

- создание моментальных снимков (snapshots) распределённой файловой системы HDFS для быстрого восстановления данных от логических ошибок
- копирование данных между кластерами или в выделенное хранилище с помощью утилиты DistCp (Distributed Copy) для защиты от катастрофических сбоев и создания архивных копий за пределами основного кластера
- экспорт метаданных из Hive Metastore и других систем каталогов
- резервное копирование конфигураций и ключей безопасности с помощью Arenadata Cluster Manager (ADCM)

Политики резервного копирования (периодичность, глубина хранения, набор компонентов) настраиваются в соответствии с требованиями Заказчика к точке восстановления (RPO) и времени восстановления (RTO) и могут комбинировать указанные методы.

## 6.1.5 Подсистема управления данными

Отвечает за организацию совместной работы с данными на бизнес-уровне. Предоставляет бизнес-пользователям и аналитикам единый каталог для поиска, документирования и контроля качества корпоративных данных, а также для ведения бизнес-глоссария. В основе подсистемы лежит ПО Arenadata Catalog, которое интегрируется с техническим сервисом метаданных Hive Metastore для автоматического сбора информации об активах данных.

Подсистема реализуется **Модулем управления данными** (см. п. 6.2.5).

## 6.1.6 Подсистема преобразования данных

Состоит из **Модулей преобразования данных** (см. п. 6.2.6), которые необходимы для создания аналитических платформ, а также для интеграции, выгрузки и обработки данных из любых источников. Является основой для выстраивания и оркестрации ETL/ELT-процессов.

### 6.1.7 Подсистема управления доступом

Включает в себя набор инструментов и механизмов для защиты данных, управления доступом и обеспечения соответствия требованиям безопасности. В основе подсистемы лежит ПО Arenadata Platform Security.

Подсистема реализуется **Базовым модулем безопасности** (см. п. 6.2.7).

### 6.1.8 Подсистема граничных функций

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

Подсистема реализуется **Модулем граничных функций** (см. п. 6.2.8)

## 6.2 Модули

### 6.2.1 Базовый модуль

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Базовый модуль. Обеспечивает функционирование подсистемы обеспечения базовых сервисов и сетевой подсистемы.

#### Назначение

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

#### Узлы

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

#### Отказоустойчивость обеспечена

- Резервированием вычислительных узлов, отвечающих за мониторинг и управление компонентами **Машины**

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

### Применяемое программное обеспечение

- Скала<sup>▲</sup>р Визион
- Скала<sup>▲</sup>р Геном
- ОС, входящая в состав ПАК
- Сервисное ПО, входящее в состав Arenadata Cluster Manager (ADCM)
- ПО для управления пользователями и аутентификацией (опционально)

#### 6.2.2 Модуль неструктурированной обработки

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Модуль неструктурированной обработки. Обеспечивает функционирование подсистемы неструктурированной обработки больших данных, формируя ключевой компонент распределённого слоя хранения данных и предоставляя вычислительные ресурсы для пакетной обработки.

### Назначение

- Хранение данных любого типа обеспечивает надёжное, отказоустойчивое и масштабируемое хранение структурированных, слабоструктурированных (JSON, XML, лог-файлы) и неструктурных (изображения, видео, документы) данных
- Формирует основу для создания централизованного хранилища данных по модели Data Lake, предназначенного для консолидации сырых данных любого типа и формата. При использовании в связке с системой управления метаданными Hive и колоночными форматами хранения, такими как ORC, Машина может обеспечивать гарантии целостности данных, включая атомарность и согласованность операций на уровне таблиц, что позволяет реализовывать сценарии, характерные для аналитических хранилищ данных, такие как работа с обновляемыми витринами данных
- Высокопроизводительная пакетная обработка предоставляет вычислительные ресурсы для выполнения распределенных ETL-задач, сложной аналитики и работ по подготовке данных с использованием таких фреймворков, как Spark и MapReduce
- Обеспечение надёжности и эффективности реализует механизмы репликации и сжатия данных для гарантии их сохранности и снижения затрат на хранение

### Узлы

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

- тип 1 – оптимизирован для рабочих нагрузок, требующих высокой пропускной способности, минимальной задержки доступа к часто используемым («горячим»)

данным, и для ресурсоёмких вычислительных процессов. Оснащается процессорами с большим количеством ядер и высокой тактовой частотой, большим объёмом оперативной памяти и высокоскоростными накопителями NVMe SSD

- тип 2 – сконфигурирован для экономичного хранения больших объёмов данных при сохранении производительности. Комплектуется процессорами с оптимальным соотношением цены и производительности, объёмной оперативной памятью и высокояёмкими накопителями SAS/NL-SAS. Применяется для хранения «холодных» (редко используемых) и «тёплых» (среднего спроса) данных, а также для выполнения фоновых заданий, не требующих минимальной задержки отклика

### Отказоустойчивость обеспечена

- *Репликацией данных в HDFS*: данные автоматически реплицируются на разные узлы и стойки. Выход из строя одного или нескольких узлов не приводит к потере данных или неработоспособности кластера. HDFS автоматически восстанавливает фактор репликации на исправных узлах
- *Отказоустойчивостью сервисов*: критичные сервисы, такие как HDFS DataNode и Zookeeper, развернуты на множестве узлов. Отказ одного узла не влияет на доступность сервиса в целом
- Технологией RAID для дисков вычислительных узлов
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)

### Применяемое программное обеспечение

- Hadoop Distributed File System (HDFS)
- Spark для высокопроизводительной распределённой обработки данных в оперативной памяти и MapReduce для надёжной пакетной обработки
- Компоненты для управления жизненным циклом данных, включая перемещение между типами хранилищ (SSD/HDD), применение Erasure coding и компрессию файлов
- ОС, входящая в состав ПАК

#### 6.2.3 Модуль координации

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Модуль координации. Обеспечивает функционирование подсистемы управления.

### Назначение

- Обеспечение консенсуса и координации распределённых сервисов кластера
- Хранение и управление критическими метаданными распределённой файловой системы (HDFS)

- Предоставление сервиса распределённой конфигурации и синхронизации для компонентной экосистемы (HBase, Hive)
- Гарантия согласованности состояния кластера и отказоустойчивости ключевых управляющих сервисов

### Узлы

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

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

### Отказоустойчивость обеспечена

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

### Применяемое программное обеспечение

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

#### 6.2.4 Модуль резервного копирования

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Модуль резервного копирования. Обеспечивает функционирование подсистемы резервного копирования.

### Назначение

- Резервирование и восстановление данных: обеспечивает создание резервных копий данных, хранящихся в основных подсистемах **Машины** (например, в подсистеме неструктурированной обработки больших данных), и их последующее восстановление в случае необходимости
- Долгосрочное хранение резервных копий: предоставляет выделенное, экономичное хранилище для архивных данных, моментальных снимков состояния системы и других материалов резервного копирования

- Интеграция с экосистемой: функционирует как сетевое хранилище (NFS), обеспечивая совместимость со стандартными механизмами резервного копирования, входящими в состав платформы Arenadata Hadoop и других компонентов

## Узлы

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

## Отказоустойчивость обеспечена

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

## Применяемое программное обеспечение

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

### 6.2.5 Специализированный модуль (для управления данными)

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Специализированный модуль. Обеспечивает функционирование подсистемы управления данными (см. п. 6.1.5). Интерфейс основного функционального ПО Модуля предоставляет доступ к каталогу метаданных, бизнес-глоссарию, поиску, профилированию и проверке качества корпоративных данных.

## Назначение

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

## Узлы

В состав Модуля входит не менее двух вычислительных узлов

## Отказоустойчивость обеспечена

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

## Применяемое программное обеспечение

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

### 6.2.6 Специализированный модуль (для задач преобразования данных)

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Специализированный модуль. Обеспечивает функционирование подсистемы преобразования данных (см. п. 6.1.6).

## Назначение

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

## Узлы

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

## Отказоустойчивость обеспечена

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

## Применяемое программное обеспечение

- ПО для управления ETL-процессами

### 6.2.7 Базовый модуль безопасности

Базовый модуль безопасности обеспечивает функционирование подсистемы управления доступом (см. п.6.1.7).

## Назначение

Предоставляет комплексную систему, включающую управление доступом на основе политик, авторизацию и безопасный доступ к платформе и её сервисам, что помогает защитить конфиденциальные данные и обеспечить соответствие нормативным требованиям.

В Модуль входят компоненты Arenadata Platform Security (ADPS):

- Ranger Admin & UserSync — ПО управления политиками безопасности и синхронизации пользователей (развертываются на 2-3 узлах для отказоустойчивости)
- Ranger Key Management Service (KMS) — сервис управления ключами шифрования (развертывается на 2 узлах)
- Knox Gateway — шлюз безопасности (развертывается на 2 узлах для балансировки нагрузки и отказоустойчивости)

## Узлы

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

## Отказоустойчивость обеспечена

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

## Применяющееся программное обеспечение

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

### 6.2.8 Модуль граничных функций

Название в Едином реестре российской радиоэлектронной продукции — СКАЛА-Р Модуль граничных функций. Обеспечивает функционирование подсистемы граничных функций (см. п. 6.1.8).

## Назначение

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

## Узлы

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

## Отказоустойчивость обеспечена

- Интеграцией с YARN для управления ресурсами и автоматического перезапуска контейнеров с задачами при сбоях

- Использованием отказоустойчивых режимов исполнения для Spark и Flink
- Развертыванием критичных сервисов доступа в кластерном режиме с балансировкой нагрузки
- Резервированием сетевых коммутаторов (объединение сетевых узлов в MLAG-пару)
- Технологией RAID для системных накопителей вычислительных узлов

### Применяемое программное обеспечение

- YARN
- Spark (включая Spark SQL, MLlib)
- Flink
- Hive Server2
- ОС, входящая в состав ПАК

## 7. СПЕЦИФИЧНЫЕ ЧЕРТЫ

Проектирование **Машины больших данных Скала<sup>®</sup>р МБД.Х** осуществлялось с учётом ряда выбранных приоритетов, оказывающих непосредственное влияние на функциональные и эксплуатационные показатели. Каждый приоритет направлен на решение ключевых бизнес-задач.

### Возможная интеграция с любыми источниками информации

Комплекс поддерживает загрузку и обработку данных из всех основных источников, включая:

- OLTP и OLAP СУБД
- ERP, CRM-системы
- документы и почтовые сообщения
- журналы веб-серверов, потоки посещений
- данные социальных сетей
- журналы промышленных систем
- данные сенсоров и датчиков

Это позволяет создать единую точку истины для всей корпоративной информации, преодолевая разрозненность данных и обеспечивая целостность аналитической картины.

### Гарантия скорости обработки данных

В **Машине** используются параллельные и распределённые вычисления. Высокая скорость обработки достигается за счёт:

- применения высокопроизводительного оборудования: быстрых дисков SSD, высокоскоростных сетей внутреннего взаимодействия (до 100 Гбит/с), процессоров с большим числом ядер, значительного объёма оперативной памяти на узел
- архитектуры массово-параллельной обработки, где запрос выполняется одновременно на всех узлах кластера

### Выполнение бизнес-задач на неограниченном объёме данных

**Машина** предоставляет инструментарий для решения широкого круга аналитических задач на масштабируемой платформе:

- анализ постоянно поступающих больших данных, чей объём измеряется в терабайтах и петабайтах
- преобразование неструктурированных и частично структурированных данных (логи, JSON, XML) в структурированную форму, пригодную для глубокого анализа
- работа с данными типа «ключ-значение» для сценариев оперативного доступа

- гибкая интеграция с внешними системами: комплекс поддерживает загрузку данных из реляционных СУБД, CRM/ERP-систем, логов и датчиков через инструмент Sqoop, а также оркестрацию сквозных ETL/ELT-процессов с помощью Airflow. Это позволяет автоматизировать построение витрин данных и аналитических пайплайнсов от источника до результата

### Горизонтальное масштабирование

Горизонтальное масштабирование обеспечивает рост системы в зависимости от потребностей бизнеса. Такой подход позволяет:

- наращивать мощности строго в соответствии с потребностями бизнеса, без необходимости полной замены системы
- обеспечивать предельную загрузку оборудования, повышая его эффективность
- гарантировать предсказуемость инвестиций в ИТ-инфраструктуру — затраты на расширение мощности и объёма прямо пропорциональны растущим требованиям

Подобный подход позволяет расширять возможности существующих аналитических систем и обеспечить предельную загрузку оборудования.

## 8. ГАРАНТИРОВАННОЕ КАЧЕСТВО И ПОЛНАЯ ГОТОВНОСТЬ К ПРОМЫШЛЕННОЙ ЭКСПЛУАТАЦИИ

Качественные показатели **Машины больших данных Скала<sup>®</sup> 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 и сборка: без компромиссов

- Используются только **проверенные комплектующие**, отобранные под реальные нагрузки
- Все компоненты собираются строго по регламенту, в соответствии с утверждённой схемой размещения
- Развёртывание программного обеспечения и первичная настройка выполняются **автоматизированно**, чтобы исключить «ручные» ошибки
- Каждый комплекс перед отгрузкой проходит функциональное тестирование и проверку на наличие известных уязвимостей
- Отклонения от типового решения **Машины больших данных Скала<sup>®</sup> МБД.Х** исключены

### Передача в эксплуатацию: все готово к работе

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

### Техническая поддержка: от производителя, без посредников

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

## Сопровождение и развитие под задачи бизнеса

По запросу возможно:

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

Все доработки выполняются с участием тех, кто проектировал и создавал **Машину больших данных Скала<sup>®</sup> МБД.Х**, что снижает риски и обеспечивает стабильность в эксплуатации.

## 9. РЕАКЦИЯ МАШИНЫ НА ВОЗМОЖНЫЕ ОТКАЗЫ

### Отказы, связанные со стандартными элементами Скала<sup>®</sup> МБД.Х

В **Машине больших данных Скала<sup>®</sup> МБД.Х** обеспечена отказоустойчивость её основных элементов и процессов, в том числе:

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

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

### Отказы, связанные с узлами кластера

#### Аппаратные сбои

Архитектура программного обеспечения, лежащего в основе **Машины больших данных Скала<sup>®</sup> МБД.Х**, позволяет построить отказоустойчивый многоузловой кластер.

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

#### HDFS

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

#### Управляющие сервисы

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

#### Координация

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

Потеря кворума сервиса координации Zookeeper является критическим событием. Это нарушает работу механизмов автоматического переключения для высокодоступных сервисов, зависящих от него. Например, может быть заблокирована процедура автоматической смены активного узла NameNode в HDFS, что создает риск недоступности файловой системы для записи при отказе текущего ведущего узла. Для минимизации

такого риска Zookeeper развертывается в виде отказоустойчивого кластера минимум из трёх узлов.

## Программные сбои и человеческий фактор

### Мониторинг

Встроенная система мониторинга **Скала<sup>®</sup> Визион** отслеживает состояние всех сервисов и узлов, генерируя оповещения при деградации или сбое.

### Восстановление сервисов

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

### Резервное копирование и восстановление

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

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

## 10. ВАРИАТИВНОСТЬ МАШИН

Машины больших данных Скала<sup>®</sup>р МБД.Х проектируются под конкретные задачи Заказчика, что позволяет оптимизировать совокупную стоимость владения (TCO). Базовые конфигурации адаптируются под преобладающий тип рабочей нагрузки: высокопроизводительные вычисления, экономичное хранение больших объёмов или специализированные сценарии.

### Приоритет производительности

Область применения:

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

Варианты решения:

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

### Приоритет объёма и эффективности хранения

Область применения:

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

Вариант решения:

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

## Специализированный тюнинг под прикладные задачи

### Область применения

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

Подход к реализации проводится совместно архитекторами Скала<sup>®</sup> и техническими специалистами Заказчика и может включать:

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

## 11. ТРЕБОВАНИЯ К РАЗМЕЩЕНИЮ МАШИНЫ

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

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 будут выполнены настройки сетевых адресов в соответствии со структурой сети Заказчика. Заказчик должен предоставить необходимые данные в соответствии с номенклатурой компонентов **Машины больших данных Скала<sup>®</sup>р МБД.Х**.

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

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

## 12. ТЕХНИЧЕСКАЯ ПОДДЕРЖКА

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

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

Состав типовых пакетов услуг по технической поддержке **Машин больших данных Скала<sup>®</sup> МБД.Х** представлен в таблице 1.

Таблица 1. Пакеты услуг по технической поддержке

Услуги	Пакет «9×5»	Пакет «24×7»
«Режим предоставления услуг 9×5» (в рабочее время по рабочим дням)	+	—
«Режим предоставления услуг 24×7» (круглосуточно)	—	+
Предоставление доступа к системе регистрации запросов/инцидентов Service Desk	+	+
Предоставление доступа к базе знаний по продуктам Скала <sup>®</sup>	+	+
Предоставление обновлений лицензионного ПО Скала <sup>®</sup>	+	+
Диагностика, анализ и устранение проблем в работе комплекса Скала <sup>®</sup> , включая: <ul style="list-style-type: none"> <li>▪ устранение аппаратных неисправностей</li> <li>▪ техническое сопровождение ПО</li> </ul>	+	+
Консультации по работе комплекса Скала <sup>®</sup>	+	+
«Защита конфиденциальной информации» (неисправные носители информации не возвращаются Заказчиком)	Опция	Опция
Замена и ремонт оборудования по месту установки	+	+

Услуги	Пакет «9×5»	Пакет «24×7»
Доставка оборудования на замену за счет производителя	+	+
Расширенные параметры 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<sup>Ар</sup> 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 узлов.

## 13. ЛИЦЕНЗИРОВАНИЕ ПО В СОСТАВЕ МОДУЛЕЙ

Команда Скала<sup>®</sup> активно занимается развитием программных продуктов семейства **Машин больших данных Скала<sup>®</sup> МБД.8**, в которое входит комплекс **Машина больших данных Скала<sup>®</sup> МБД.Х**. Направления развития формируются на основе анализа мирового опыта использования систем подобного класса и пожеланий Заказчиков и партнеров. Новые функции реализуются в форме релизов, которые могут выходить несколько раз в год.

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

Программное обеспечение **Скала<sup>®</sup> Геном**, **Скала<sup>®</sup> Визион** поставляется исключительно в составе **Машин Скала<sup>®</sup>** и лицензируется по метрикам комплекса в соответствии с количеством узлов.

### 13.1 Политика обновления ПО

Команда Скала<sup>®</sup> активно занимается развитием собственных программных продуктов. Направления развития формируются на основе анализа мирового опыта использования систем подобного класса и пожеланий Заказчиков и партнеров. Новые функции реализуются в форме релизов. Обновления для **Машин**, находящихся в эксплуатации, производятся по согласованию с Заказчиком.

## О КОМПАНИИ

Скала<sup>®</sup> — модульная платформа для построения высоконагруженной ИТ-инфраструктуры, продукт Группы Rubytech.

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

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

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

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

Дополнительная информация — на сайте [www.skala-r.ru](http://www.skala-r.ru).