

**ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ**  
**ПЛАТФОРМА ДАННЫХ СЕЛЕНА**  
**РФ.DIS.00011–01 15**

### Аннотация

Документ РФ.DIS.00011-01 15 «Платформа Данных Селена» подготовлен на основе рекомендаций государственного стандарта ГОСТ 19.402—78 Единая система программной документации. Описание программы.

Электронная версия документа хранится в составе пакета программной и эксплуатационной документации на изделие РФ.DIS.00011 «Платформа Данных Селена» (далее по тексту Функциональные требования).

Ознакомление с документом «Платформа Данных Селена» персонала подразделения, принимающего участие в работе программного комплекса, производится под роспись с внесением соответствующей записи в журнал первичного инструктажа.

## СОДЕРЖАНИЕ

Обозначения и сокращения .....	4
1 Общие сведения .....	5
2 Описание логической структуры.....	7
3 Функциональные требования .....	10

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В документе РФ.DIS.00011-01 15 «Платформа Данных Селена» используются следующие обозначения и сокращения, имеющие соответствующие значения:

Обозначение или сокращение	Значение
ГОСТ	Государственный стандарт
РФ	Российская Федерация
СУБД	Система управления базами данных
УЗ	Учетная запись
AI	Искусственный интеллект
BI	Business intelligence. Система сбора, обработки и анализа данных
ETL	Extract, Transform, Load
JDBC	Java DataBase Connectivity
ML	Machine learning
ODBC	Open Database Connectivity
OLAP	Online analytical processing
MPP	Massively parallel processing
SIMD	Single instruction, multiple data
S3	Simple Storage Service
SQL	<b><i>Structured Query Language</i></b> — «язык структурированных запросов»

## 1 ОБЩИЕ СВЕДЕНИЯ

### 1.1 Наименование программы

Наименование программы на русском языке: Платформа Данных Селена.

Наименование программы в латинской транслитерации: Selena Data Platform.

Краткое наименование программы на русском языке: Селена.

Краткое наименование программы в латинской транслитерации: Selena.

### 1.2 Описание программы

Платформа Данных Селена предназначена для построения сверхбыстрой платформы хранения и обработки данных с массивно-параллельной обработкой (МРР), разработанная для упрощения и ускорения доступа к данным, быстрой аналитики в реальном времени.

Платформа Данных Селена обеспечивает требуемым функционалом все сферы бизнеса:

- финансы и страхование;
- телекоммуникации;
- производственный сектор;
- государственный сектор;
- промышленность;
- торговля;
- и т.д.

### 1.3 Цели внедрения

Целями внедрения Платформы Данных Селена являются:

- построение централизованного хранилища данных;
- снижение затрат на инфраструктуру;
- Реализация аналитики в реальном времени;
- Ускорение аналитических запросов;
- Упрощение архитектуры хранилища данных.

#### 1.4 Техническое описание

Платформа Данных Селена разработана на языках программирования:

- Java;
- C++;
- Kotlin;
- TypeScript.

Для функционирования Платформы Данных Селена требуется предварительно установленное программное обеспечение:

- пакет прикладных программ;
- базы данных источников информации;
- файловые ресурсы для размещения шаблонов документов;
- обозреватели для работы с веб интерфейсами Платформы Данных Селена;
- клиенты и консольные приложения для разработки запросов к данным на Платформе Данных Селена
- консольные приложения для настройки и обслуживания программы Платформа Данных Селена.

Компетенции обслуживающего персонала и пользователей Платформы Данных Селена должны обеспечивать бесперебойную работу всех модулей программного обеспечения в режиме реального времени с технологическими перерывами на техническое обслуживание оборудования и обновление программно-аппаратных средств.

#### 1.5 Возможности приложения

Основные возможности:

- централизованное хранение данных
- обработка любых типов данных на любых языках;
- загрузка данных из различных источников;
- обработка и доставка потоковых данных в хранилище;
- возможность последовательного и параллельного выполнения процессов обработки;
- консолидация и нормализация данных;
- управление потоками данных.

## 2 ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ

### 2.1 Логическая архитектура решения

Платформа Данных Селена основана на архитектуре открытых данных. Архитектура представлена на рисунке 1.

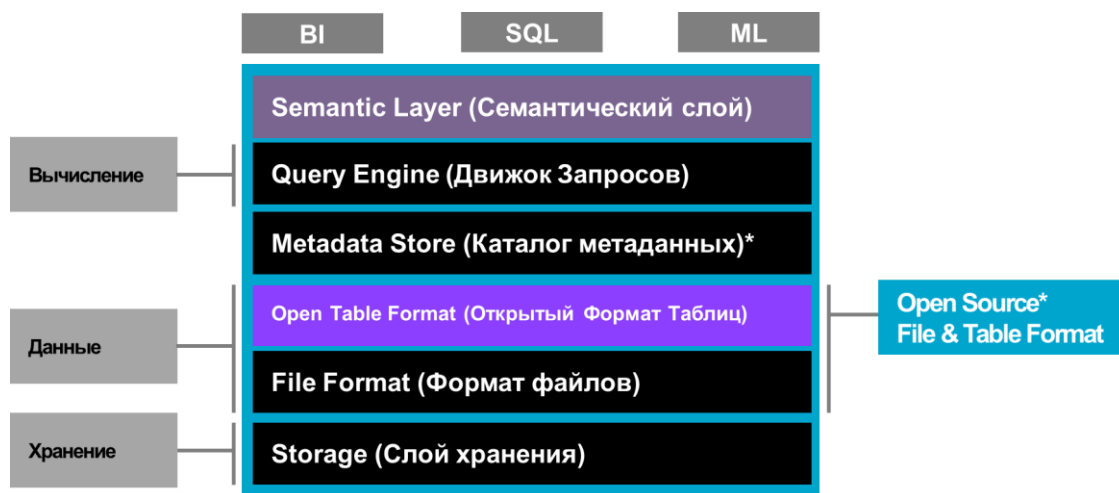


Рисунок 1. Архитектура открытых данных

Архитектура открытых данных используется при построении хранилищ данных Data Lakehouse. Это архитектурный подход, объединяющий лучшие практики, взятые из архитектур Data Warehouse и Data Lake, предоставляющий лучшие возможности хранения и обработки данных.

Платформа данных Селена строится на принципах и подходах, полностью соответствующих архитектуре Data Lakehouse, таких как:

- использование открытых форматов таблиц, обеспечивающий независимость единого слоя данных;
- использование каталога данных, который хранит в себе информацию о местонахождении данных, и являющийся обязательным компонентом платформы;
- использование независимого решения для хранения данных. В Селене используется S3, который является технологией для удобного и надежного хранения большого объема данных;

- использование движка запросов, для выполнения задач, связанных с обработкой данных и вычислениями;

Соблюдение этих подходов при построении хранилища данных обеспечивают следующие положительные преимущества:

- универсальность. По сравнению с блочными и файловыми хранилищами решение S3 позволяет хранить абсолютно любые данные;
- независимость и масштабируемость всех слоев архитектуры;
- доступность данных для сторонних систем. Все объекты хранилища располагаются в плоском адресном пространстве, без иерархии, упрощая доступ и работу с файлами;
- высокую надежность, данные хранятся одновременно в разных узлах;
- экономичность, версионность и многое другое.

## 2.2 Компоненты Платформы Данных Селена

Платформа данных Селена реализована с использованием следующих компонентов:

- движок обработки данных – реализует обработку запросов через кластер мастер-серверов (управление метаданными, соединениями, планированием) и кластер вычислительных серверов (исполнение SQL и вычисления);
- панель управления кластером (Cluster Manager) – предоставляет инструменты для администрирования кластера: создание, настройка, управление нодами, развертывание приложений и контроль доступа к данным и интерфейсам;
- компонент загрузки и управления потоками данных – обеспечивает импорт данных из внешних источников, визуальное конфигурирование ETL-процессов и настройку расписаний загрузки;
- компонент метакаталога данных – централизованное хранилище метаинформации, ускоряющее поиск и доступ к данным в объектном хранилище;



- компонент мониторинга – собирает метрики работоспособности кластера, статус компонентов и утилизацию ресурсов;
- компонент единой среды разработки – предоставляет пользователям интуитивный интерфейс для взаимодействия с данными;
- компонент хранения данных – отвечает за физическое сохранение информации в распределенном хранилище.

### 3 ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

Платформа Данных Селена представляет собой новый подход к хранению и работе с данными. Селена предлагает широкий набор функций для быстрой аналитики данных в реальном времени в больших масштабах.

#### 3.1 Функции, реализуемые вычислительным движком

В части компонента движка обработки данных Платформы Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает массово-параллельную обработку запросов благодаря реализации MPP-архитектуры;
- обеспечивает высокую скорость аналитических операций за счёт оптимизированного векторного движка выполнения запросов;
- обеспечивает поддержку SIMD-инструкций;
- обеспечивает выполнение сложных аналитических запросов с субсекундной задержкой;
- обеспечивает возможность определить и удалить ненужные части запроса, которые не влияют на конечный результат по средствам встроенного оптимизатора запросов;
- обеспечивает нахождение самого оптимального плана при выполнении запроса на основе собранной статистики данных по средствам встроенного оптимизатора запросов;
- обеспечивает автоматическое преобразование исходного запроса пользователей в альтернативную формулировку, с целью достижения того же результата за меньшее время и/или количества операций по средствам встроенного оптимизатора запросов;
- обеспечивает обработку петабайтных объёмов данных с линейной масштабируемостью;
- обеспечивает поддержку звёздообразных, снежинкообразных и Data Vault моделей данных;

- обеспечивает эффективное хранение и сжатие данных благодаря колоночному формату;
- обеспечивает автоматическое партиционирование и бакетирование данных для оптимизации запросов;
- обеспечивает ускорение аналитики через автоматически обновляемые материализованные представления;
- обеспечивает обработку полуструктурированных и неструктурированных данных;
- обеспечивает кэширование запросов, данных, результатов, планов запросов;
- обеспечивает интеграцию с системами потоковой и пакетной обработки данных, включая Apache Kafka, Flink и Spark;
- обеспечивает совместимость со стандартными интерфейсами JDBC/ODBC;
- обеспечивает поддержку различных открытых форматов хранения данных Iceberg, Hudi, Data lakes;
- обеспечивает поддержку различных форматов файлов, таких как Parquet, ORC, CSV;
- обеспечивает совместимость с протоколом MySQL;
- обеспечивает совместимость с BI-инструментами, такими как Форсайт, Tableau, Superset, Power BI, Metabase, FineBI, Pix BI, LuxMS BI;
- обеспечивает выполнение федеративных запросов к внешним источникам данных;
- обеспечивает взаимодействие с различными сторонними S3-совместимыми объектными хранилищами (в том числе с MinIO, Ceph, Zakroma, VK S3, MWS S3, Seaweed);
- обеспечивает поддержку ресурсных групп;
- обеспечивает программное управление и мониторинг через API;
- обеспечивает горизонтальное масштабирование без простоев;
- обеспечивает отказоустойчивость через автоматическую репликацию данных между узлами;
- обеспечивает поддержку multi-кластерных конфигураций;

- обеспечивает интеллектуальное распределение запросов и балансировку нагрузки;
- обеспечивает поддержку установки в гибридных и мультиоблачных средах;
- обеспечивает как пакетную, так и потоковую загрузку данных;
- обеспечивает встроенные возможности ETL для трансформации данных;
- обеспечивает инкрементальное обновление материализованных представлений;
- обеспечивает управление жизненным циклом данных;
- обеспечивает анализ временных рядов с поддержкой оконных функций;
- обеспечивает управление доступом на основе ролей (RBAC);
- обеспечивает интеграцию с системами аутентификации LDAP/Kerberos;
- обеспечивает шифрование данных при передаче и хранении;
- обеспечивает детальный аудит всех операций с данными;
- обеспечивает мониторинг производительности через встроенные метрики;
- обеспечивает анализ и профилирование запросов;
- обеспечивает автоматическую индексацию часто используемых колонок;
- обеспечивает ручную настройку запросов через подсказки оптимизатора;
- обеспечивает мониторинг через интеграцию с Prometheus и Grafana;
- обеспечивает обработку геопространственных данных и запросов;
- обеспечивает специализированные функции для анализа временных рядов;
- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes;
- обеспечивает расширение функциональности через пользовательские функции (UDF) Java, Python.

Компонент обработки данных предоставляет современное решение для аналитики больших данных, сочетающее высокую производительность запросов с гибкостью масштабирования. Компонент особенно эффективен для сценариев, требующих сложной аналитики в реальном времени на больших объемах данных,

включая интерактивные дашборды, оперативную отчетность и системы поддержки принятия решений.

### 3.2 Функции, реализуемые компонентом панели управления кластером (Cluster Manager)

В части компонента панели управления кластером Платформа Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает автоматизированный деплой кластеров в private cloud и on-premises средах;
- обеспечивает управление инфраструктурой как кодом через интеграцию Kubernetes;
- обеспечивает автоматическую проверку готовности нод и сервисов;
- обеспечивает динамическое масштабирование с добавлением/удалением нод без downtime;
- обеспечивает безопасный вывод нод из кластера;
- обеспечивает автоматическое восстановление упавших сервисов;
- обеспечивает локализацию на разные языки (русский\английский);
- обеспечивает централизованное управление параметрами через единый интерфейс;
- обеспечивает изменение настроек "на лету" без перезагрузки;
- обеспечивает отображение исходных значений параметров конфигурации;
- обеспечивает централизованное управление лицензиями и контроль сроков;
- обеспечивает установку квот ресурсов по лицензионным уровням;
- обеспечивает интеграцию с системами IAM (AD/LDAP/OIDC/SAML);
- обеспечивает синхронизацию прав доступа с метастором LakeKeeper;
- обеспечивает управление шифрованием данных через TLS/SSL интеграции;
- обеспечивает безопасное хранение и ротацию учетных данных;
- обеспечивает единый дашборд для визуализации состояния кластера;

- обеспечивает сбор и анализ логов компонентов;
- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes.

В части управления идентификацией и доступом обеспечиваются следующие функциональные возможности:

- обеспечивает многофакторную аутентификацию;
- обеспечивает интеграцию с внешними провайдерами аутентификации (OAuth 2.0, OIDC, SAML 2.0, LDAP/AD);
- обеспечивает единый вход (SSO) для кросс-сервисной авторизации;
- обеспечивает защиту от bruteforce-атак с автоматической блокировкой при множественных неудачных попытках;
- обеспечивает гибкие механизмы восстановления доступа через email/SMS;
- обеспечивает полный цикл управления учетными записями (создание, редактирование, деактивация);
- обеспечивает гибкую систему ролей и разрешений (RBAC) с возможностью кастомизации;
- обеспечивает функционал групповой работы с наследуемыми правами;
- обеспечивает мультитенантную архитектуру с разделением организаций;
- обеспечивает детализированный аудит всех действий пользователей и системных событий;
- обеспечивает интеграцию с SIEM-системами для централизованного сбора логов;
- обеспечивает настройку webhook-уведомлений о критических событиях;
- обеспечивает REST API для программного управления функциями IAM;
- обеспечивает возможность кастомизации интерфейсов аутентификации под корпоративный стиль;
- обеспечивает контроль активных сессий и принудительное завершение подключений.

Компонент панели управления кластером обеспечивает комплексное решение по управлению и конфигурированию всего кластера, а также поддерживает все необходимые функции по обеспечению и управлению доступом в распределенных системах с требованиями высокой безопасности и гибкости управления правами.

### 3.3 Функции, реализуемые компонентом загрузки и управления потоками данными

В части компонента загрузки и управления потоками данными Платформа Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает визуальное проектирование пайплайнов через drag-and-drop интерфейс;
- обеспечивает объединение задач в режиме потоковой передачи ориентированного ациклического графа;
- обеспечивает поддержку различных типов задач;
- обеспечивает создание зависимостей между задачами в workflow;
- обеспечивает распределенное выполнение задач с балансировкой нагрузки;
- обеспечивает планирование через cron-выражения и event-based триггеры;
- обеспечивает распределенное выполнение SQL-запросов;
- обеспечивает выполнение федеративных запросов к разнородным источникам;
- обеспечивает локализацию на разные языки (русский\английский);
- обеспечивает оптимизацию сложных аналитических запросов;
- обеспечивает встроенные коннекторы к реляционным СУБД, таким как MySQL, PostgreSQL, Oracle, SQL Server, MariaDB, Snowflake, Vertica;
- обеспечивает встроенные коннекторы к Data Lake-решениям, включая Hive, Iceberg, Delta Lake и Hudi;
- обеспечивает встроенные коннекторы к NoSQL-системами, такими как Cassandra, MongoDB, Redis, Elasticsearch и OpenSearch;
- обеспечивает встроенные коннекторы к платформам потоковой передачи данных, таким как Kafka;

- обеспечивает поддержку специализированных систем хранения и анализа данных, включая Greenplum, ClickHouse, Druid, DuckDB и тестовые решения TPC-DS, TPC-H;
- обеспечивает визуальный мониторинг состояния ETL-процессов;
- обеспечивает оповещение по электронной почте, Telegram, Http, Slack в зависимости от сработавшего триггера;
- обеспечивает детальную статистику выполнения задач;
- обеспечивает систему алертинга при сбоях или нарушениях SLA;
- обеспечивает возможность ручного перезапуска и отладки задач;
- обеспечивает историю выполнения с анализом временных метрик;
- обеспечивает работу с различными форматами данных (Parquet, ORC, и т.д.);
- обеспечивает REST API для интеграции с внешними системами;
- обеспечивает webhook-уведомления о событиях выполнения;
- обеспечивает поддержку плагинов для расширения функциональности;
- обеспечивает совместимость с Hadoop и облачными сервисами;
- обеспечивает ролевой доступ к ETL-процессам;
- обеспечивает интеграцию с корпоративными системами аутентификации;
- обеспечивает шифрование данных при передаче между компонентами;
- обеспечивает аудит всех действий пользователей системы;
- обеспечивает ограничение доступа к определенным источникам данных;
- обеспечивает горизонтальное масштабирование рабочих узлов;
- обеспечивает автоматическое восстановление после сбоев;
- обеспечивает поддержку multi-кластерных конфигураций;
- обеспечивает балансировку нагрузки между worker-узлами;
- обеспечивает географическое распределение обработки;
- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes;
- обеспечивает встроенные шаблоны типовых ETL-сценариев;



- обеспечивает инструменты профилирования производительности запросов;
- обеспечивает экспорт/импорт конфигураций ETL-процессов.

Компонент загрузки и управления потоками данных обеспечивает визуальное проектирование и выполнение ETL-процессов с поддержкой распределённой обработки, мониторинга и алертинга. Решение предлагает гибкое планирование задач, интеграцию с различными источниками данных и обеспечивает безопасность за счёт разграничения доступа и аудита действий.

### 3.4 Функции, реализуемые компонентом метакаталога данных

В части метакаталога данных Платформа Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает работу исполняемого файла без зависимостей от jvm или python;
- обеспечивает защиту данных через временные учётные данные и удалённую подпись s3-запросов;
- обеспечивает поддержку всех основных облачных платформ (aws, azure, gcp) и локальных s3-решений;
- обеспечивает интеграцию с корпоративными openid-провайдерами для аутентификации;
- обеспечивает развертывание в kubernetes через helm-чарты;
- обеспечивает одновременную аутентификацию через kubernetes service accounts и openid;
- обеспечивает генерацию событий изменений в формате cloudevents;
- обеспечивает контроль изменений через внешние системы верификации;
- обеспечивает мультитенантную архитектуру с единой точкой входа;
- обеспечивает кастомизацию каталога, хранилища секретов и системы авторизации;
- обеспечивает совместимость с spark, pyiceberg, trino и starrocks;
- обеспечивает горизонтальное масштабирование без локального состояния;

- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes;
- обеспечивает детализированный контроль доступа на базе openfga;

Компонент метакаталога данных обеспечивает детализированный контроль доступа, мультитенантность, верификацию изменений и совместимость с популярными аналитическими инструментами, обеспечивая безопасность и масштабируемость data-lake.

### 3.5 Функции, реализуемые компонентом мониторинга

В части компонента мониторинга Платформа Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает визуализацию метрик, логов и трейсов из различных источников данных;
- обеспечивает создание интерактивных и настраиваемых дашбордов;
- обеспечивает поддержку ad-hoc-запросов для исследования данных;
- позволяет детализировать данные по временным диапазонам;
- обеспечивает сравнение различных наборов данных;
- обеспечивает создание уведомлений на основе метрик, логов и трейсов;
- обеспечивает работу с множеством источников данных одновременно;
- обеспечивает комбинирование данных из разных источников в одном графике;
- обеспечивает возможность анализа данных мониторинга;
- обеспечивает фильтрацию и поиск в режиме реального времени;
- обеспечивает создание кастомных визуализаций;
- обеспечивает агрегацию метрик;
- обеспечивает анализ логов с live-стримингом;
- обеспечивает возможность расширять функциональность через плагины;
- обеспечивает подключение новых источников данных;

- обеспечивает гибкие настройки безопасности и доступа;
- обеспечивает масштабируемость для работы с большими объемами данных;
- обеспечивает работу в распределенных средах;
- предоставляет API для интеграции с внешними системами;
- обеспечивает экспорт и импорт дашбордов;
- обеспечивает совместную работу над дашбордами;
- обеспечивает историю изменений;
- обеспечивает наличие готовых шаблонов для быстрого создания дашбордов;
- обеспечивает возможность настраивать внешний вид дашбордов;
- обеспечивает локализацию на разные языки (русский\английский);
- обеспечивает сбор и хранение метрик в формате временных рядов;
- обеспечивает многомерную модель данных с использованием меток;
- обеспечивает анализ метрик мониторинга;
- обеспечивает возможность создавать гибкие правила оповещения на основе метрик;
- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes;
- обеспечивает различные методы обнаружения сервисов;
- обеспечивает простое развертывание и минимальную конфигурацию;
- обеспечивает возможность настраивать частоту сбора метрик;
- обеспечивает экспорт метрик в различные форматы;
- обеспечивает аутентификацию и авторизацию через SSO;

Компонента мониторинга обеспечивает комплексное решение для мониторинга состояния всех элементов платформы, сочетая мощные возможности сбора и анализа метрик с удобными инструментами визуализации и оповещения.

### 3.6 Функции, реализуемые компонентом единой среды разработки

В части компонента единой среды разработки Платформы Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает мультязыковую разработку (sql, python) в единой среде;
- обеспечивает интеллектуальный редактор с автодополнением и подсветкой синтаксиса;
- обеспечивает параллельное выполнение и сравнение различных версий запросов;
- обеспечивает интерактивную визуализацию результатов без переключения контекста;
- обеспечивает создание jupyter-подобных notebook'ов для анализа данных;
- обеспечивает интерактивный проводник структуры данных;
- обеспечивает систему аннотаций и тегирования для таблиц и столбцов;
- обеспечивает визуализацию lineage данных и зависимостей между объектами;
- обеспечивает историю выполненных запросов с возможностью повторного использования;
- обеспечивает шаринг запросов и дашбордов с контролем доступа;
- обеспечивает создание шаблонов для часто используемых запросов;
- обеспечивает совместную работу в реальном времени;
- обеспечивает экспорт результатов в форматы csv, tsv;
- обеспечивает экспорт блокнотов в формате pdf;
- обеспечивает поддержку плагинов для расширения функциональности;
- обеспечивает тонкую настройку прав доступа к запросам и данным;
- обеспечивает интеграцию с корпоративными iam (ldap, oauth, saml);
- обеспечивает аудит действий пользователей с возможностью воспроизведения;
- обеспечивает кэширование часто используемых запросов и результатов;
- обеспечивает оптимизацию больших запросов с индикацией прогресса;
- обеспечивает горячие клавиши для частых операций;

- обеспечивает систему квот для ресурсоемких операций;
- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes.

В части возможностей работы с генеративным ИИ обеспечиваются следующие функциональные возможности:

- обеспечивает преобразование естественного языка в sql-запросы с учетом структуры хранилища (text-to-sql);
- обеспечивает генерацию пояснений к sql-коду на естественном языке (sql-to-text);
- обеспечивает конвертацию sql-запросов в исполняемые блокноты;
- обеспечивает ai-анализ и оптимизацию sql-запросов;
- обеспечивает контекстно-зависимые подсказки для автодополнения sql;
- обеспечивает интеллектуальные подсказки на основе текущего контекста запроса;
- обеспечивает автоматическую генерацию шаблонного кода и анализа данных;
- обеспечивает объяснение ошибок sql и предлагает варианты исправлений;
- обеспечивает адаптацию к стилю работы пользователя на основе истории запросов;
- обеспечивает интеграцию с внешними llm (gpt, claude, yandexgpt, gigachat, qwen, qwq, deepseek).

Компонент единой среды разработки предоставляет возможности для анализа данных с поддержкой мультязыковой разработки, интеллектуальным редактором и инструментами визуализации. Интегрирует ИИ-помощника для автоматизации SQL-запросов, оптимизации кода и анализа данных.

### 3.7 Функции, реализуемые компонентом хранения данных

В части компонента хранения данных Платформа Данных Селена обеспечивает следующие функциональные возможности:

- обеспечивает поддержку s3-совместимого api для работы с объектным хранилищем;
- обеспечивает неограниченное масштабирование объема хранимых данных;
- обеспечивает гибкую систему организации данных через бакеты, префиксы и теги;
- обеспечивает версионирование объектов с возможностью восстановления предыдущих версий;
- обеспечивает автоматическую репликацию данных между узлами кластера;
- обеспечивает распределенную архитектуру для отказоустойчивости;
- обеспечивает кэширование часто запрашиваемых объектов;
- обеспечивает балансировку нагрузки между узлами кластера;
- обеспечивает географическое распределение данных (multi-site);
- обеспечивает оптимизацию для работы с большими файлами через мультипарт-загрузки;
- обеспечивает тонкую настройку прав доступа через iam;
- обеспечивает шифрование данных при передаче (tls) и хранении;
- обеспечивает поддержку временных токенов доступа (sts);
- обеспечивает интеграцию с внешними системами аутентификации (ldap, oidc);
- обеспечивает аудит операций с объектами;
- обеспечивает политики автоматического удаления устаревших данных;
- обеспечивает настройку правил архивирования объектов;
- обеспечивает поддержку immutable-объектов;
- обеспечивает управление репликацией между кластерами;
- обеспечивает полную совместимость с amazon s3 api;
- обеспечивает поддержку различных клиентских библиотек и sdk;
- обеспечивает использование как бэкенда для других сервисов;
- обеспечивает restful api для управления хранилищем;

- обеспечивает webhook-уведомления о событиях в хранилище;
- обеспечивает встроенные инструменты для анализа использования хранилища;
- обеспечивает поддержку лямбда-функций для обработки объектов;
- обеспечивает организацию "горячего" и "холодного" хранения;
- обеспечивает интеграцию с системами поиска по метаданным;
- обеспечивает кросс-региональную репликацию данных;
- обеспечивает встроенные метрики производительности и использования;
- обеспечивает интеграцию с системами мониторинга (prometheus, grafana);
- обеспечивает инструменты для диагностики и устранения проблем;
- обеспечивает плавное масштабирование кластера;
- обеспечивает поддержку квот и ограничений на использование ресурсов;
- обеспечивает возможность установки на различных операционных системах, в том числе Ubuntu, Centos, RHEL, Astra Linux, RedOS, Sber Linux;
- обеспечивает возможность установки с применением контейнерных технологий, таких как Podman, Kubernetes.

Компонент хранения данных предоставляет универсальное решение для хранения неструктурированных данных, сочетая высокую производительность и надежность.



Общество с ограниченной ответственностью «Платформа Данных Селена»  
123317, г.Москва, вн.тер.г. Муниципальный Округ Пресненский, ул. Литвина-  
Седого, д.2/13 К. В стр. 1, помещ. 1Ц  
ИНН 9703195303, КПП 770301001, ОГРН 1247700725491

[illegible]