

Описание функциональных характеристик Arenadata Enterprise SQL Umformer (ADESU)

Содержание:

Термины и определения	3
Сокращения и обозначения	3
1 Требования к программному обеспечению	5
2 Функциональные характеристики	5
2.1 Трансформации	5
2.2 Дополнительные функции ПО	6
2.3 Интерфейсы взаимодействия	7
2.4 Требования к исходным данным	8
2.5 Поддержка стороннего ПО	8
3 Требования к системам-источникам и системам-приемникам	9
3.1 Требования к системам-источникам	9
3.2 Требования к системам-приемникам	10

Термины и определения

Термин	Значение
ANSI SQL	Стандартный язык запросов к базам данных, разработанный Американским национальным стандартом (ANSI)
LLM-модели	Языковые модели, обученные на больших объемах текстовых данных

Сокращения и обозначения

Сокращение	Наименование
ADB	Arenadata DB
ADESU, ПО Umformer	Arenadata Enterprise SQL Umformer
API	(англ. – Application Programming Interface) – набор классов, процедур, функций, структур или констант, которыми одна компьютерная программа может взаимодействовать с другой программой
CPU	(англ. – Central Processing Unit) – центральный процессор
DDL	Data Definition Language
DML	Data Manipulation Language
IP	(англ. – Internet Protocol) – маршрутизируемый протокол сетевого уровня стека TCP/IP
LDAP	(англ. – Lightweight Directory Access Protocol) – протокол прикладного уровня для доступа к службе каталогов
LLMs	Large language models
MSSQL	Microsoft SQL Server
SQL	(англ. – Structured Query Language) – декларативный язык программирования, применяемый для создания, модификации и управления данными в реляционных базах данных
SSL	Secure Sockets Layer
UI	User interface
БД	База данных
ГБ	Гигабайт

Сокращение	Наименование
ИИ	Искусственный интеллект
ОС	Операционная система
ПО	Программное обеспечение
РФ	Российская Федерация
СУБД	Система управления базами данных
ТБ	Терабайт

1 Требования к программному обеспечению

Системные требования для узла развертывания ПО ADESU:

- OS: CentOS 7.9 / Ubuntu / Astra Linux;
- Системное ПО: Docker / Docker CE / Podman.

Требования к сетевой инфраструктуре:

- Внутренние подключения: сетевая связанность между узлами решения, отсутствие блокировок портов (TCP / UDP), производительность не ниже 1 Gb / сек.
- Подключение к внешним LLM: сетевой доступ между узлом SQL Umformer и внешним LLM-сервером.

Удаленный доступ по протоколу SSH с правами суперпользователя (sudo) к следующим узлам:

- узлу SQL Umformer;
- к узлам СУБД (при необходимости их развертывания).

2 Функциональные характеристики

2.1 Трансформации

ПО ADESU осуществляет трансформацию кода ANSI SQL системы-источника в запросы системы-приемника.

ПО ADESU выполняет следующие ключевые функции:

1. Трансформация sql-запросов с помощью LLM. Источником и приемником могут быть следующие СУБД: Oracle, ADB, MSSQL, Impala, Hive.

2. Трансформация функций и процедур Oracle в функции ADB с помощью LLM.
3. Трансформация функций и процедур Oracle в функции ADB с помощью ora2pg.

ПО ADESU работает в двух режимах:

1. Выполняются трансформации SQL-запросов, представленных в текстовом файле. При этом в одном файле может записано несколько запросов, перечисленных разделителем. Значение разделителя должно задаваться в качестве настройки.
2. Выполняются трансформации функций и процедур для системы-источника Oracle. При этом в качестве настройки есть возможность указания конкретной схемы и уровня каталога в БД.

2.2 Дополнительные функции ПО

После получения запроса системой-приемником осуществляется проверка синтаксиса полученного запроса. ПО ADESU проводит проверку синтаксиса, чтобы убедиться, что запрос корректно сформирован и готов к выполнению. Это включает проверку корректности синтаксиса, выявление ошибок и предупреждений.

ПО ADESU обеспечивает проверку на наличие ошибок и предупреждений, а также выдает отчеты о результатах проверки.

В режиме 1 ПО ADESU использует не менее трех проверок качества: *count*, состав полей и *pandas.DataFrame.compare*. Эти проверки позволяют убедиться, что преобразованные запросы возвращают корректные результаты.

ПО ADESU поддерживает возможность добавления дополнительных проверок по запросу Заказчика. Это позволяет гибко настраивать процесс валидации данных под конкретные требования.

В режиме 2 ПО ADESU осуществляет проверку создания преобразованной функции или процедуры в системе-приемнике. В случае неудачи выдается ошибка.

ПО ADESU осуществляет автоматическую корректировку сгенерированных sql-запросов и функций на основе ошибок их запуска/создания в ADB. Для корректировки функций используется UI-интерфейс, в котором можно редактировать сгенерированные функции.

ПО ADESU осуществляет фиксацию (логирование) функций в БД и файлах, в том числе учет времени генерации sql-запросов и их выполнения в БД.

ПО ADESU поддерживает подключение внешних LLM-сервисов и моделей от вендоров (в т.ч. совместимых с OpenAI API).

ПО ADESU поддерживает конфигурирование параметров моделей и используемых БД.

2.3 Интерфейсы взаимодействия

ПО ADESU предоставляет пользовательский интерфейс (UI) для взаимодействия с пользователями. UI является интуитивно понятным и удобным для использования.

UI может использоваться для создания веб-приложений, предназначенных для анализа данных, машинного обучения и визуализации информации. Примером UI является Streamlit.io. Он позволяет разработчикам быстро создавать интерактивные приложения без необходимости глубокого погружения в фронтенд-разработку. С его помощью существует возможность добавлять элементы интерфейса, такие как кнопки, ползунки, текстовые поля, графики и таблицы. Также можно использовать различные библиотеки для работы с данными, включая Pandas, Matplotlib, Plotly и другие.

UI обладает следующими свойствами:

- **Интерактивность:** у пользователей есть возможность взаимодействовать с приложением через интерфейс, меняя параметры и наблюдая за изменениями в реальном времени;
- **Визуализация:** поддержка различных библиотек для построения графиков и диаграмм для анализа данных;
- **Совместимость:** возможность интеграции с другими инструментами и библиотеками Python.

2.4 Требования к исходным данным

В части трансформации sql-запросов к исходным данным предъявляются следующие требования:

1. Источником для трансформации sql-запросов должен являться файл в кодировке *utf-8*, с разделителем (например, «;»).
2. Необходимо наличие целевых таблиц в ABD, т.к. операции DDL не поддерживаются.
3. Поддерживаются только запросы выборки данных (*select*).
Следующие DML-запросы не поддерживаются: *INSERT ... SELECT*, *MERGE*, *UPDATE / DELETE / INSERT*.

В части трансформации функций и процедур Oracle в ADB должны быть выполнены следующие требования к пакетам Oracle:

1. Каждая процедура и функция должны заканчиваться символом *END*, за которым следует название процедуры/функции.
2. Все глобальные переменные и типы пакетов должны быть объявлены до объявления процедур/функций.

2.5 Поддержка стороннего ПО

ПО ADESU использует только open-source LLM-модели и frameworks. Это обеспечивает прозрачность и возможность модификации ПО.

ПО ADESU совместимо с различными СУБД. Это включает поддержку таких СУБД, как MySQL, PostgreSQL, Oracle, Microsoft SQL Server и другие.

ПО ADESU поддерживает Docker для установки и настройки, с учетом ограничений работы Docker в Российской Федерации. Docker обеспечит легкость установки и развертывания ПО.

ПО ADESU поддерживает установку на различных платформах, в том числе в условиях отсутствия сети Интернет. Это включает поддержку локальных установок и автономного режима работы.

ПО ADESU поддерживает установку в условиях ограничений на установку стороннего ПО. Это включает поддержку установки без необходимости внесения изменений в систему безопасности или настройки сети.

3 Требования к системам-источникам и системам-приемникам

3.1 Требования к системам-источникам

ПО ADESU поддерживает следующие версии СУБД-источников:

- Oracle Database:
 - Версия ПО: 11g – 21c;
 - Аутентификация: внутренняя ролевая модель БД, без SSL / LDAP / Kerberos;
 - Необходимые ресурсы: CPU vCores: 4+, RAM: 8+ GB, Storage: 50+ GB;
- Microsoft SQL Server:
 - Версия ПО: 2005 – 2022;
 - Аутентификация: внутренняя ролевая модель БД, без SSL / LDAP / Kerberos;

- Ресурсы: CPU vCores: 4+, RAM: 4+ GB, Storage: 50+ GB.

3.2 Требования к системам-приемникам

ПО ADESU поддерживает следующие версии СУБД-приемника:

- Arenadata DB:
 - Версия ПО: 6.x (6.12 – 6.27);
 - Аутентификация: внутренняя ролевая модель БД, без SSL / LDAP / Kerberos;
 - Ресурсы: CPU vCores: 8+, RAM: 32+ GB, Storage: 100+ GB.