

Описание эксплуатации
программного обеспечения для
электронно-вычислительных машин
Arenadata Enterprise SQL Umformer (ADESU)

Содержание:

1	Термины и определения	3
2	Сокращения и обозначения	3
3	Введение.....	4
4	Запуск конвертера.....	5
5	Остановка конвертера	5
6	Запуск UI.....	6
7	Остановка UI	6

1 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

2 Сокращения и обозначения

Сокращение	Наименование
ADB	Arenadata DB
ADESU, ПО Umformer	Arenadata Enterprise SQL Umformer
API	(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	Маршрутизируемый протокол сетевого уровня стека TCP/IP.
MSSQL	Microsoft SQL Server
LDAP	(англ. – Lightweight Directory Access Protocol) – протокол прикладного уровня для доступа к службе каталогов
LLMs	Large language models
SSL	Secure Sockets Layer
SQL	(англ. – Structured Query Language) – декларативный язык программирования, применяемый для создания, модификации и управления данными в реляционных базах данных.
UI	User interface
БД	База данных
ГБ	Гигабайт
ОС	Операционная система
ПО	Программное обеспечение
СУБД	Система управления базами данных
ТБ	Терабайт

3 Введение

Arenadata Enterprise SQL Umformer (далее ПО Umformer) — это программный продукт, предназначенный для автоматического преобразования кода из одной системы в другую. Продукт будет использовать LLM-модели для трансформации запросов и обеспечивать проверку синтаксиса полученных запросов. Основные функции продукта:

- Трансформация sql-запросов с помощью LLM (источником и приемником могут быть следующие СУБД: Oracle, ADB, MSSQL, Impala, Hive);
- Трансформация функций и процедур Oracle в функции ADB с помощью LLM;

ПО Umformer является инструментом поддержки миграции с распространенных СУБД зарубежного производства на СУБД, использующие диалект PostgreSQL, например, ADB и ADPG.

Функционально Umformer состоит из двух блоков:

- Блок пакетной трансформации SQL-кода;
- Блок интерактивной доработки SQL-кода с поддержкой ML, реализованный в форме WEB UI.

ПО Umformer предназначено для работы совместно с внешней LLM, соответствующей открытой спецификации OpenAI. Элементы LLM в состав дистрибутива ПО Umformer не включаются и должны быть развернуты отдельно. ПО Umformer поставляется в виде архива, содержащего исполняемый код приложения, файлы конфигурации и docker image, обеспечивающий простую переносимость приложения.

ПО Umformer поддерживает два режима работы: конвертация пакетов Oracle и конвертация SQL скриптов из текстового файла. Текущий режим определяется значением параметра конфигурации `transformer_type: package` и `sql` соответственно.

4 Запуск конвертера

Запуск конвертера осуществляется из директории `umformer` командой:

```
docker run -it --rm -e INPUT_DB_PASSWORD='input_dbuser_passwd' -e  
OUTPUT_DB_PASSWORD='output_dbuser_passwd' -e OWN_TOKEN='token' -e  
INPUT_IP='10.0.0.2:1521/XEPDB1' -e OUTPUT_IP='10.0.0.3' -e  
LLM_RESULTS_DIR='llm_results' -v $(pwd)/llm_files:/llm/llm_files --name llm  
llm
```

Где:

- `INPUT_DB_PASSWORD` – пароль пользователя БД источника;
- `OUTPUT_DB_PASSWORD` – пароль пользователя БД приемника;
- `OWN_TOKEN` – токен доступа к LLM;
- `INPUT_IP` – адрес подключения к БД источнику, включая имя сервиса для OracleDB;
- `OUTPUT_IP` – адрес подключения к БД приемнику;
- `LLM_RESULTS_DIR` – директория для вывода результатов конвертации (по умолчанию `llm_results`).

5 Остановка конвертера

Остановка конвертера осуществляется путем остановки контейнера:

```
docker stop llm.
```

6 Запуск UI

Запуск UI осуществляется из директории *umformer* командой:

```
sudo docker run -p 8501:8501 --entrypoint "/bin/bash" -e PYTHONPATH=/llm -e  
INPUT_DB_PASSWORD='input_dbuser_passwd' -e  
OUTPUT_DB_PASSWORD='output_dbuser_passwd' -e OWN_TOKEN='token' -e  
INPUT_IP='10.0.0.2:1521/XEPDB1' -e OUTPUT_IP='10.0.0.3' -e  
LLM_RESULTS_DIR='llm_results' -it --rm -v $(pwd)/llm_files:/llm/llm_files --  
name llm llm -c "streamlit run ./src/web/init_app.py --  
server.fileWatcherType=none"
```

Где:

- INPUT_DB_PASSWORD – пароль пользователя БД источника;
- OUTPUT_DB_PASSWORD – пароль пользователя БД приемника;
- OWN_TOKEN – токен доступа к LLM;
- INPUT_IP – адрес подключения к БД источнику, включая имя сервиса для OracleDB;
- OUTPUT_IP – адрес подключения к БД приемнику;
- LLM_RESULTS_DIR – директория для вывода результатов конвертации (по умолчанию llm_results).

Для доступа к UI открыть в браузере страницу, указав актуальный адрес хоста, на котором развернуто ПО Umformer: *http://<адрес linux хост>:8501/*

7 Остановка UI

Остановка UI осуществляется путем остановки контейнера: `docker stop llm`.