

Выпускная квалификационная работа

# Определение внешних словарей в СУБД ClickHouse

Сабянин Максим Анатольевич, БПМИ155

Руководитель: руководитель команды ClickHouse, Миловидов Алексей Николаевич

Куратор: кандидат технических наук, Сухорослов Олег Викторович

# Предмет работы (1)

- Что такое внешние словари
- Настройки словарей
  - Источники данных: Postgres, Mongo, Redis, HTTP-сервер и другие
  - Типы укладки в памяти: Хеш-таблица, Cache-таблица
  - Время обновления
  - Структура словаря
- Изначальный способ определения словарей

# Предмет работы (2)

Пример XML-конфигурации словаря:

```
<dictionary>
  <name>currency</name>
  <source> ... </source>
  <lifetime> ... </lifetime>
  <layout><flat/></layout>
  <structure>
    <id>currency_id</id>
    <attribute>
      <name>code</name>
      <type>string</type>
    </attribute>
    <attribute>
      <name>name</name>
      <type>string</type>
    </attribute>
  </structure>
</dictionary>
```

```
<source>
  <mysql>
    <port>3306</port>
    <user>dw</user>
    <password>topsecret</password>
    <replica>
      <host>dwds1</host>
      <priority>1</priority>
    </replica>
    <db>ch</db>
    <table>dim_country</table>
    <where>country_key=0</where>
  </mysql>
</source>
```

# Цели и задачи работы (1)

Основная цель: добавить возможность определение внешнего словаря в клиенте ClickHouse. Например:

```
1 CREATE DICTIONARY database_name.dictionary_name
2 (
3   column1 UInt64 DEFAULT 0 expression 'rand64()' injective True,
4   column2 String DEFAULT 'null' injective True,
5   ...
6 )
7 PRIMARY KEY id_column
8 SOURCE(MYSQL(PORT 3306, USER 'dw',
9   |      |      |PASSWORD 'topsecret', REPLICA(HOST 'dwds1', PRIORITY 1),
10  |      |      |DB 'ch', TABLE 'dim_country',
11  |      |      |WHERE 'country_key>=0'))
12 LAYOUT(FLAT())
13 LIFETIME(MIN 10, MAX 100)
```

# Цели и задачи работы (2)

Основные задачи:

- Разобраться с устройством словарей: формат определения, устройство загрузки.
- Придумать новый формат с обратной совместимостью.
- Реализовать разбор SQL формата и поддержку работы в СУБД.
- Реализовать юнит-тестирование, интеграционное тестирование.

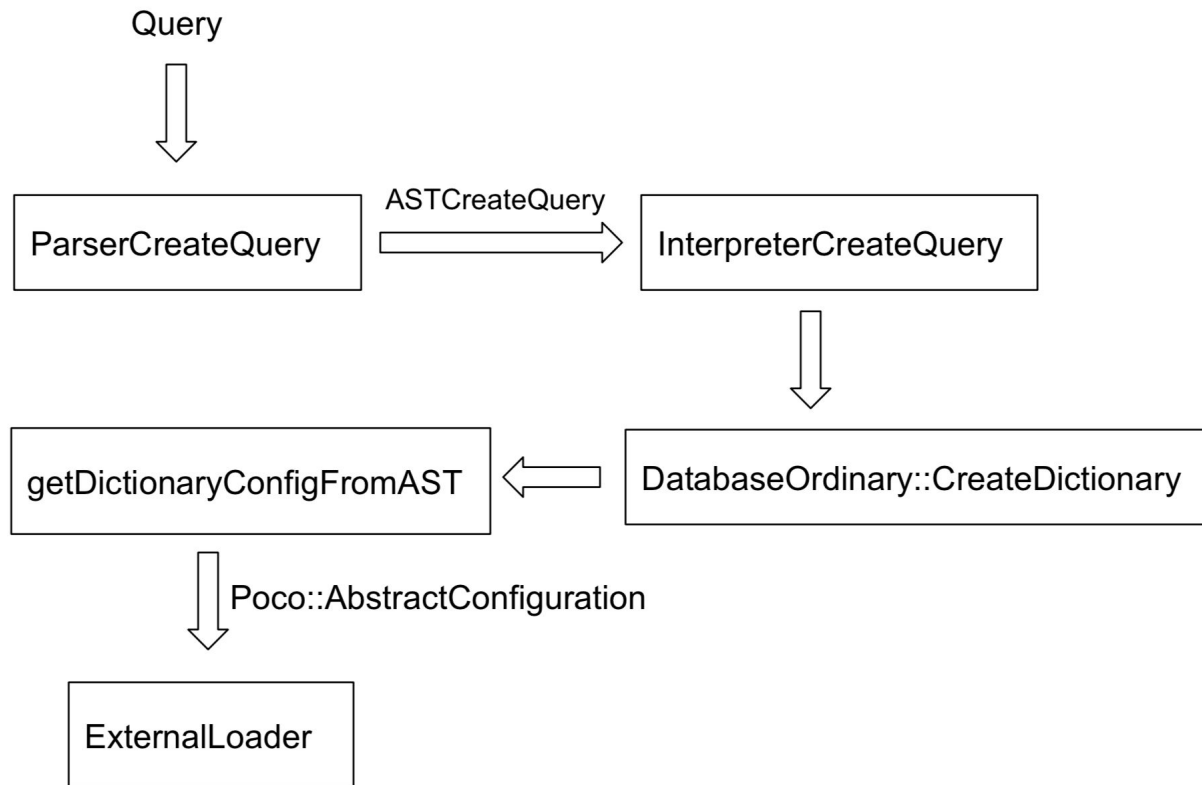
# Формальная постановка

Требуется придумать и реализовать новый формат определения словарей, в котором у любого словаря будет существовать соответствие с изначальным форматом.

# Дополнительные требования

Также хочется иметь возможность расширять возможности внешних словарей в старом и новом формате.

# Обработка запроса





# Тестирование

Алгоритм юнит-тестирования:

- 1) Подготавливается вход на языке запросов SQL с запросами на создание и работы со словарями
- 2) Вход с запросами подается в сервер кликхауса
- 3) Весь выходные сообщения записываются в файл
- 4) Выходной файл сравнивается с эталонным.

# Результаты работы

- Произведен анализ кодовой базы по работе со словарями.
- Придуман и реализован разбор и поддержка работы словарей, определяемых DDL запросом.
- Реализовано юнит-тестирование

# Вопросы