

Задача:

Необходимо написать 2 модуля и файлы sql для создания необходимых БД.

1. Модуль “Загрузки” (test_download.py):
 - a. Загружает файлы из сервиса
 - b. Записывает построчно в БД1 с необходимой доп информацией для дальнейшей обработки.
2. Модуль “Обработки” (test_preprocessing.py) :
 - a. Загружает данные из БД1 и помещает в БД2
 - b. Валидирует и преобразует согласно типу данных в таблице
 - c. Добавляет в необходимые таблицы согласно наименованию метода и колонок
 - d. Ошибочные данные помещаются в таблицу **data_error** с описанием ошибки
 - e. Вставка в базу в пакетном режиме.

Дано:

1. Внешний сервис с данными.
Данные отдаются файлами за определенную дату (пример:
reward-2017-02-01.csv.gz или input-2017-02-01.json.gz)
2. Сервер БД1 с таблицей
 - a. **raw_data** (
 1. api_source (varchar)
 2. api_method (varchar)
 3. api_date (timestampz)
 4. result (text)
 5. api_param (varchar)
 6. insert_ts (timestampz default 'now()'))
3. Сервер БД2 с таблицами
 - a. **data_method_input** (
 1. user (int)
 2. ts (timestampz)
 3. context (jsonb)
 4. ip (inet))
 - b. **data_method_reward_1** (
 1. user (int)
 2. ts (timestampz)
 3. reward_id (int)
 4. reward_money (int))
 - c. **data_method_reward_2** (

1. user (int)
2. ts (timestampz)
3. reward_id (int)
- PRIMARY KEY (user,reward_id)
-)

- d. **data_error** (
 1. api_source (varchar)
 2. api_method (varchar)
 3. api_date (timestampz)
 4. result (text)
 5. error_text (varchar))

4. Платформа

- a. СУБД: PostgreSQL >= 9.5
- b. OS: Ubuntu server 16.04
- c. Language: Python >= 3.x

5. Ссылки на файлы

- a. <https://s3.eu-central-1.amazonaws.com/gi-datastream/input-2017-02-01.json.gz>
- b. <https://s3.eu-central-1.amazonaws.com/gi-datastream/reward-2017-02-01.csv.gz>
- c.

