# Первый пункт задания

Для выполнения первого пункта использовал стек Oracle Autonomous Database + Python + CSV файл.

Workflow следующий:

База данных источник

База данных приемник

(хранение и реконсиляция)

Python обработчик

(прием, подготовка и загрузка данных)

CSV файл

В моем решении я использую одну и ту же базу данных в качестве источника и приемника. В реальном решении, конечно же, эти источники должны быть разными.

Вместо данного подхода можно было бы использовать Oracle Autonomous Database + Python + CSV файл + Apache Spark. Перенести все вычисления в Spark, вместо сервера базы данных.

# Результаты выполнения пункта 1

Изображение выглядит как текст

Автоматически созданное описание

57 секунд для реконсиляции 100.000 записей. Да, можно и быстрее. В коде расписываю возможные варианты ускорения.

# Второй пункт задания

Для выполнения второго пункта использовал Oracle Autonomous Database + Oracle REST Data Services (ORDS) + Python.

На стороне базы данных сделал:

* View admin.v\_recon\_tran\_aggregates
* REST модуль recon\_tran\_aggregates

Изображение выглядит как текст

Автоматически созданное описание

* Handler get\_all\_recods, который вытягивает все записи из v\_recon\_tran\_aggregates

Изображение выглядит как текст

Автоматически созданное описание

Далее, в программе на Python, с помощью библиотеки "requests", идет обращение к этому сервису.

# Результаты выполнения пункта 2

Изображение выглядит как текст

Автоматически созданное описание

Почти 10.000 записей за 2.6 секунды. В коде расписываю возможные варианты ускорения.

# Файлы с решением

* <https://github.com/mlysikov/homework/blob/master/task-2-solution-1.py>
* <https://github.com/mlysikov/homework/blob/master/task-2-solution-2.py>
* [https://github.com/mlysikov/homework/blob/master/transactions.csv](https://github.com/mlysikov/homework/blob/master/task-2-solution-2.py)

# Числа (факт)

* Трудоемкость (человеко-часы): 10,5
  + Анализ задания – 0,5
  + Выполнение пункта 1 задания (1-ый запрос) – 8
  + Выполнение пункта 2 задания (2-ый запрос) – 1
  + Генерация дополнительных тестовых данных и тестирование - 1
* Длительность: 4 дня
* Начало работ: 30.01.2021
* Окончание работ: 02.02.2021