Описание задачи для построения ETL-процесса (SCD1).

Разработать ETL процесс, получающий ежедневную выгрузку данных (предоставляется за 3 дня), загружающий ее в хранилище данных и ежедневно строящий отчет.

Выгрузка данных.

Ежедневно некие информационные системы выгружают три следующих файла:

- Список транзакций за текущий день. Формат – CSV.
- Список терминалов полным срезом. Формат – XLSX.
- Список паспортов, включенных в «черный список» - с накоплением с начала месяца. Формат –XLSX.

Предоставляется выгрузка за последние три дня

Сведения о картах, счетах и клиентах хранятся в СУБД PostgreSQL.

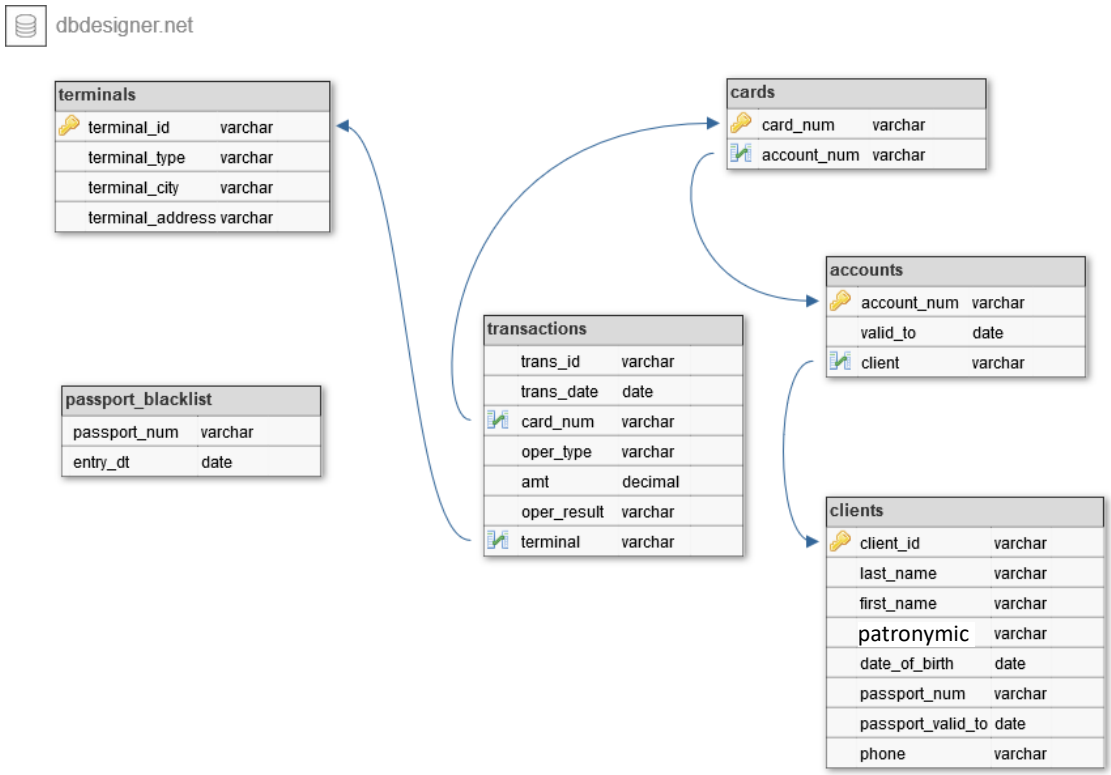
Реквизиты для подключения:

- Host: de-edu-db
- Port: 5432
- Database: bank
- User: bank_etl
- Password: bank_etl

Структура хранилища.

В качестве хранилища выступает база (edu).

Данные должны быть загружены в хранилище со следующей структурой:



Типы данных в полях можно изменять на однородные если для этого есть необходимость.

Имена полей менять нельзя.

Ко всем таблицам должны быть добавлены технические поля create_dt, update_dt;

Построение отчета.

По результатам загрузки ежедневно необходимо строить витрину отчетности по мошенническим операциям. Витрина строится накоплением, каждый новый отчет укладывается в эту же таблицу с новым report_dt. В витрине должны содержаться следующие поля:

event_dt	Время наступления события. Если событие наступило по результату нескольких действий – указывается время действия, по которому установлен факт мошенничества.
passport	Номер паспорта клиента, совершившего мошенническую операцию.
fio	ФИО клиента, совершившего мошенническую операцию.
phone	Номер телефона клиента, совершившего мошенническую операцию.
event_type	Описание типа мошенничества (номер).
report_dt	Дата, на которую построен отчет.

Признаки мошеннических операций.

Совершение операции при просроченном или заблокированном паспорте.

Совершение операции при недействующем договоре.

Совершение операций в разных городах в течение одного часа.

Правила именования таблиц.

DEAIAN.LAPP_STG_<TABLE_NAME>	Таблицы для стейджинговых таблиц.
DEAIAN.LAPP_DWH_FACT_<TABLE_NAME>	Таблицы фактов, загруженных в хранилище. В качестве фактов выступают сами транзакции и «черный список» паспортов. Имя таблиц – как в ER диаграмме.
DEAIAN.LAPP_DWH_DIM_<TABLE_NAME>	Таблицы измерений, в формате SCD1.
DEAIAN.LAPP_REP_FRAUD	Таблица с отчетом.
DEAIAN.LAPP_META_<TABLE_NAME>	Таблицы для хранения метаданных.

Обработка файлов

Выгружаемые файлы именуются согласно следующему шаблону:

- transactions_DDMMYYYY.txt
- passport_blacklist_DDMMYYYY.xlsx
- terminals_DDMMYYYY.xlsx

Предполагается что в один день приходит по одному такому файлу.

После загрузки соответствующего файла он должен быть переименован в файл с расширением .backup, чтобы при следующем запуске файл не искался и перемещен в каталог archive:

- transactions_DDMMYYYY.txt.backup
- passport_blacklist_DDMMYYYY.xlsx.backup
- terminals_DDMMYYYY.xlsx.backup

Данный проект должен содержать следующие файлы и каталоги:

main.py	Файл	Основной процесс обработки.
файлы с данными	Файл	Файлы, которые получили в качестве задания.
main.ddl	Файл	Файл с SQL кодом для создания всех необходимых объектов в базе edu.
main.cron	Файл	Файл для постановки вашего процесса на расписание, в формате crontab
archive	Каталог	Пустой, сюда должны перемещаться отработанные файлы