Добрый день, Гайк!
Небольшое описание разработанного ETL **процесса**

Программа состоит из 2 файлов.
start.py – оболочка запуска процесса.
main.py – основной модуль, загрузка данных, скрипты SQL, скрипты DDL

start.py

В этом модуле реализован поиск необходимых xlsx файлов для загрузки. В каталоге с файлом main.py ищутся файлы: transactions_DDMMYYYY.xlsx и passport_blacklist_DDMMYYYY.xlsx Предполагается, что в один день приходит по одному такому файлу, по этому реализованы следующие условия поиска.

1. Если найдено больше одного необходимого файла, или файлы вообще не найдены, то процедура пишет предубеждение и не далее не работает.
2. Если даты в названии файлов transactions_DDMMYYYY.xlsx и passport_blacklist_DDMMYYYY.xlsx не создают, то модуль выводит предупреждение, предлагает исправить фалы и прекращает дальнейшую работу.

После загрузки соответствующий файл переименуются в файл с расширением *.backup, чтобы при следующем запуске он не падал в поиск.

Загрузка данных осуществляется вызовом команды:
main.py passport_blacklist_DDMMYYYY.xlsx transactions_DDMMYYYY.xlsx

main.py

Основной модуль
Для работы процедуры необходимо установить следующие пакеты:
pip install virtualenv
pip install db-sqlite3
pip install pandas
pip install xlrd
И активировать виртуальное окружение.

Ко всем таблицам _FACT_добавлены технические поля create_dt (дата занесения записи в БД), update_dt (дата последнего изменения записи), deleted_flg (флаг удаленной записи).

Удаленные записи (а они могут быть даже среди транзакций! Я 2 года работал с основной бд АС ЦОД физиков в Сбербанке, и точно знаю, что в Сбербанке возможно всякое… **J**) помечаются флагом deleted_flg=1 (по умолчанию он =0). Если в последствие удаленная запись снова «появляется», то флаг снова становится = 0.

Описание таблиц

DE5_DWH_FACT_TERMINALS DE5_DWH_FACT_CARDS DE5_DWH_FACT_TRANSACTIONS DE5_DWH_FACT_ACCOUNTS DE5_DWH_FACT_CLIENTS DE5_DWH_FACT_PASSPORT_BLACKLIST	Таблицы фактов, загруженных в хранилище. В качестве фактов выступают сами транзакции и «черный список» паспортов.
DE5_STG_PASSP_BLACK_DT DE5_STG_TRANSACTIONS_DT DE5_STG_V_DWH_FACT_TERMINALS DE5_STG_V_DWH_FACT_CARDS DE5_STG_V_DWH_FACT_TRANSACTIONS DE5_STG_V_DWH_FACT_ACCOUNTS DE5_STG_V_DWH_FACT_CLIENTS DE5_STG_NEWROWS_DWH_FACT_TERMINALS DE5_STG_NEWROWS_DWH_FACT_CARDS DE5_STG_NEWROWS_DWH_FACT_TRANSACTIONS DE5_STG_NEWROWS_DWH_FACT_ACCOUNTS DE5_STG_NEWROWS_DWH_FACT_CLIENTS	Таблицы для размещения стейджинговых таблиц (первоначальная загрузка), промежуточное выделение инкремента. Временные таблицы и view.

DE5_STG_UPDATEROWS_DWH_FACT_TERMINALS DE5_STG_UPDATEROWS_DWH_FACT_CARDS DE5_STG_UPDATEROWS_DWH_FACT_TRANSACTIONS DE5_STG_UPDATEROWS_DWH_FACT_ACCOUNTS DE5_STG_UPDATEROWS_DWH_FACT_CLIENTS DE5_STG_DELETEROWS_DWH_FACT_TERMINALS DE5_STG_DELETEROWS_DWH_FACT_CARDS DE5_STG_DELETEROWS_DWH_FACT_TRANSACTIONS DE5_STG_DELETEROWS_DWH_FACT_ACCOUNTS DE5_STG_DELETEROWS_DWH_FACT_CLIENTS	
DE5_DWH_DIM_REP_FRAUD	Таблица справочник по признакам мошеннических операций
DE5_REP_FRAUD	Таблица витрина отчетности по мошенническим операциям DE5_REP_FRAUD