Что такое дата-пайплайны

Создание таблицы

In
```
CREATE TABLE _имя_таблицы_ (_название_поля_с_первичным_ключом_ serial PRIMARY KEY,
                            _имя_столбца1_ _тип_данных_,
                            _имя_столбца2_ _тип_данных_,
                            ...);
```

Удаление таблицы

In
```
DROP TABLE _имя_таблицы_;
```

Выдача прав доступа к таблице

In
```
GRANT ALL PRIVILEGES ON TABLE table_name TO username;
```

Выдача прав для работы с первичыми ключами

In
```
GRANT USAGE, SELECT ON SEQUENCE table_name_id_seq TO anthony;
```

Подключение к БД

In
```
from sqlalchemy import create_engine

# Задаем параметры подключения к БД, их можно узнать у администратора БД
db_config = {'user': 'my_user',         # имя пользователя
             'pwd': 'my_user_password', # пароль
             'host': 'localhost',       # адрес сервера
             'port': 5432,              # порт подключения
             'db': 'db_name'}           # название базы данных

# Формируем строку соединения с БД
connection_string = 'postgresql://{}:{}@{}:{}/{}'.format(db_config['user'],
                                                         db_config['pwd'],
                                                         db_config['host'],
                                                         db_config['port'],
                                                         db_config['db'])

# Подключаемся к БД
engine = create_engine(connection_string)
```

Выполняем запрос и сохраняем результат выполнения в DataFrame

In
```
# Sqlalchemy автоматически установит названия колонок такими же, как у таблицы в БД.
# Нам останется только указать индексную колонку с помощью index_col.

query = ''' SELECT column1, column2, column3
            FROM table_name
        '''

data_raw = pd.io.sql.read_sql(query, con = engine, index_col = 'column1')
```

Добавление строк в таблицу

In
```
df.to_sql(name = 'table_name', con = engine, if_exists = 'append', index = False))

# if_exists = 'replace' – совпадающие строки переписываются
# if_exists = 'append'  – совпадающие дублируются новыми копиями
```

Удаление строк из таблицы по условию

```
In  DELETE FROM _имя_таблицы_ WHERE _условия_для_поиска_записей_которые_нужно_стереть_;
```

Словарь

Дата-пайплайн
специальная программа, которая вызывается по расписанию, собирает, объединяет, трансформирует и сохраняет данные автоматически

Агрегирование или агрегация
процесс группировки и уменьшения размера данных