

Airflow Practice

Перед выполнением заданий рекомендуется ознакомиться с [туториалом по Airflow](#).

В качестве практической задачи, возьмем открытые API крипто-валютных бирж и будем регулярно получать данные, таким образом построим свое авто-пополняемое хранилище данных, с различными преобразованиями и автоматическим формированием статистики по Bitcoin в конце каждого дня по каждой бирже, которое в будущем, возможно, будет использоваться для тренировки трейдинговых ботов и тд.

В качестве API бирж предлагаем использовать следующие сервисы:

```
services = {  
    'bitfinex': 'https://api.bitfinex.com/v1/trades/btcusd?limit_trades=500',  
    'bitmex': 'https://www.bitmex.com/api/v1/trade?symbol=XBTCUSD&count=500&reverse=true',  
    'poloniex': 'https://poloniex.com/public?command=returnTradeHistory&currencyPair=USDT_BTC',  
}
```

Hint: для браузинга по API можно использовать [Postman](#) или [Insomnia](#)

Acceptance criteria:

- S3(minio) используется для хранения сырых данных
- Отдельная БД в PostgreSQL создана и используется для хранения статистики по дням
- Каждый час данные преобразуются в единый вид и складываются в отдельный бакет на S3 в parquet формате, а потом прочитывается статистика за прошлый час. Единый вид:
 - Название биржи
 - Время транзакции
 - Вид транзакции (buy/sell)
 - Количество купленных биткоинов
 - По какой цене
- Все DAGs запускаются по расписанию
 - Сбор данных с бирж каждую минуту
 - Агрегация данных и подсчет статистики каждый час
 - Имейл уведомление в конце каждого дня с почасовой статистикой
- Джобы для преобразования написаны на Spark