

Семинар №3

Инструменты Bid Data

1. Инструментарий:

[Лекция](#)

[Презентация](#)

Задание:

- 1) В докер контейнере собраны основные инструменты, с которыми мы будем работать
- 2) Выполняем команду docker compose up
- 3) Загрузить в hdfs датасет по ценам на жилье Airbnb, доступный на kaggle.com: <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

Notebooks

Задание:

- 1) Создать таблицу в hive с помощью spark kernel из датасета airbnb
- 2) Вывести график распределение цен
- 3) Сделать для в разных notebooks IDE

HUE

Задание:

- 1) Создать таблицу в hive с помощью spark kernel из датасета airbnb
- 2) Вывести график распределение цен
- 3) Сохранить результаты в HDFS
- 4) Выгрузить результаты из HDFS на локальный компьютер

2. Домашнее задание

1. Соберите данные о погоде в разных городах мира за последний месяц. Используйте открытые источники данных, такие как API погодных сервисов или веб-скрейпинг.

2. Выведете график изменения температуры в разных городах, график распределения температуры.

3. Сохранить результаты в HDFS

4. Выгрузить результаты из HDFS на локальный компьютер