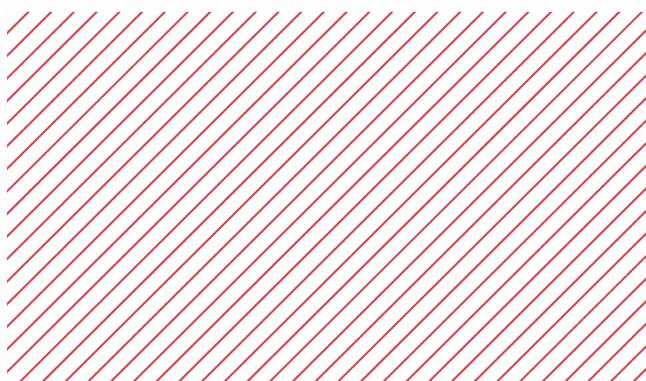


академия
больших
данных



HW02: Hive





Описание работы и критериев оценивания

В задании два блока:

- 1) Развёртывание локального Hive - 50 баллов
- 2) HiveQL - 50 баллов

Результаты ДЗ загрузить в репозиторий на Github и прислать ссылку на него в интерфейсе сдачи

Бонусы и штрафы:

- **100%** за плагиат
- **30%** за посылку решения в течение недели после deadline



Блок 1. Развёртывание локального Hive

- 1) Развернуть локальный Hive в любой конфигурации - 20 баллов
- 2) Подключиться к развернутому Hive с помощью любого инструмента: Hue, Python Driver, Zeppelin, любая IDE итд (15 баллов за любой инструмент, максимум 30 баллов)
- 3) Сделать скриншоты поднятого Hive и подключений в выбранными вами инструментах, добавить в репозиторий

Блок 2. Работа с Hive

- 1) Сделать таблицу artists в Hive и вставить туда значения, используя датасет <https://www.kaggle.com/pieca111/music-artists-popularity> - 15 баллов
- 2) Используя Hive найти (команды и результаты записать в файл и добавить в репозиторий):
 - a) Исполнителя с максимальным числом скретчей - 5 баллов
 - b) Самый популярный тэг на ластфм - 10 баллов
 - c) Самые популярные исполнители 10 самых популярных тегов ластфм - 10 баллов
 - d) Любой другой инсайт на ваше усмотрение - 10 баллов