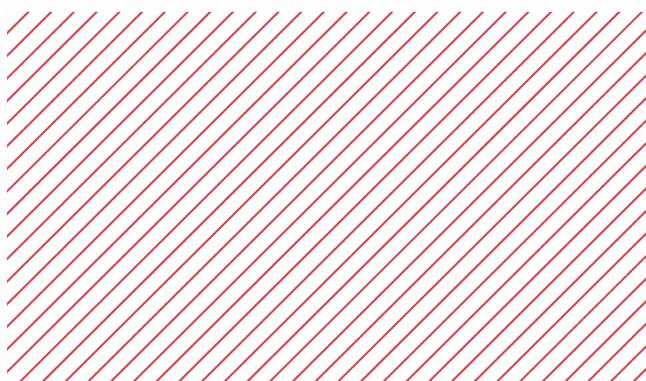


академия  
больших  
данных



## HW03: Hive QL



# Описание работы и критерии оценивания

---

Для данной задачи мы попрактикуемся в интерактивной среде Hue, где будем работать с Hive QL, создавая различные запросы.

Веб интерфейс Hue доступен по ссылке <https://demo.gethue.com/hue/accounts/login?next=/> (username:demo, password:demo)

Залогинившись, выберите Hive, как основную среду исполнения, после этого используйте базу default.

В работе, мы будем использовать две таблицы: sample\_07 и sample\_08, которые представляют информацию о внутреннем устройстве отделов двух компаний с указанием количества сотрудников и их зарплат.

В качестве решения, предлагается выслать код для запросов в формате .txt либо .sql и ответ на него на адрес [muzalevsky.ds@gmail.com](mailto:muzalevsky.ds@gmail.com)

Бонусы и штрафы:

- **100%** за плагиат в решениях (всем участникам процесса)
- **30%** за посылку решения в течение недели после deadline
- **10%** за несоответствие правилам оформления задачи



# Этапы задания (max 100 баллов)

---

1. Вывести название компании с максимальным количеством сотрудников и зарплатой по отделам. Названия компаний - названия файлов датасетов (20 баллов)
2. Вывести список ID отделов обоих компаний, количество сотрудников в которых превышает 50000 человек (20 баллов)
3. Для прошлого задания посчитать долю сотрудников отдела от всех сотрудников конкретной компании (20 баллов)
4. Проверить, существуют ли отделы, работающие в одной компании, но отсутствующие в другой (20 баллов)
5. Отсортировать отделы компаний по суммарной зарплате по убыванию. Использовать информацию по отделам обеих компаний суммарно (20 баллов)