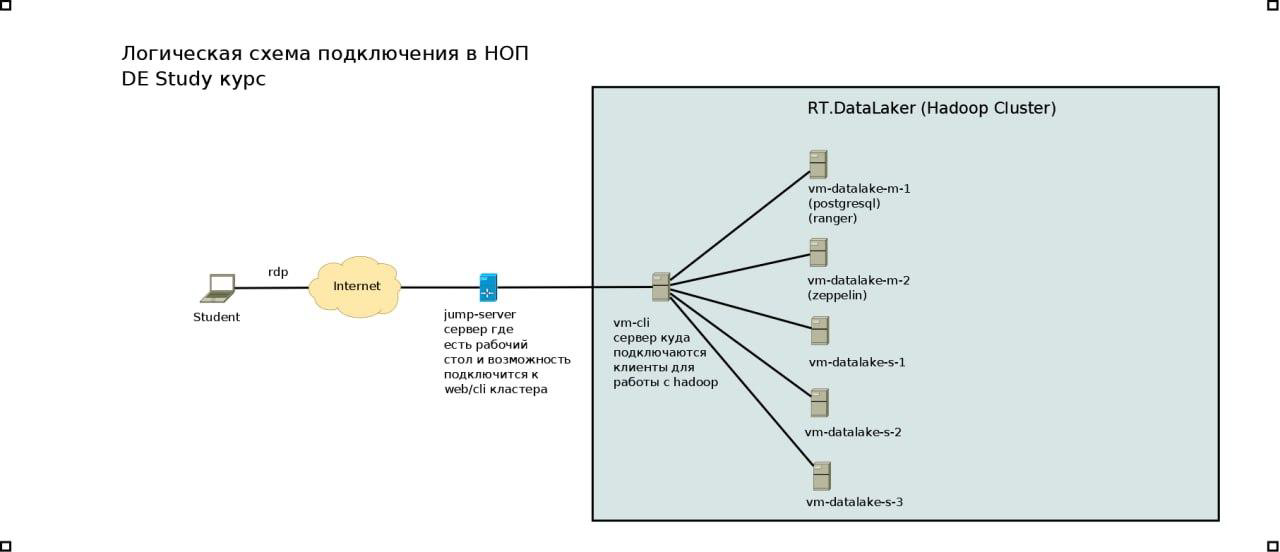
**Лабораторная работа №4**

**Олейников Михаил Николаевич (olejnikov)**

В текущей работе вам потребуется использовать продукт RT.DataLake, который развернут в рамках нашего курса. Схема подключения к учебному стенду отображена на рисунке и подключение было рассмотрено на наших вебинарах. (если появятся вопросы, пожалуйста напишите в общий учебный чат или личным сообщением преподавателю)

****

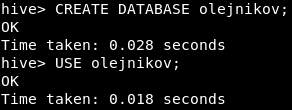
Предварительно забираем датасет:



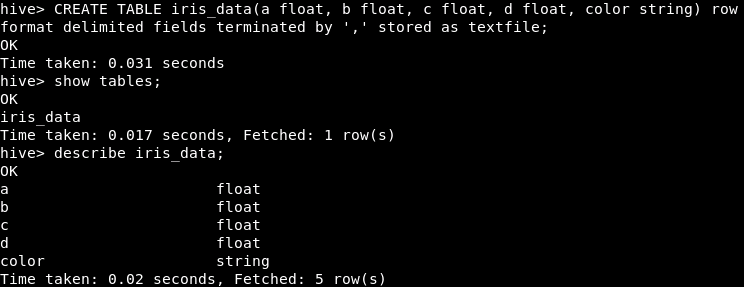
1. Соединитесь с HIVE, используя командную строку (команда hive)
2. Из консоли HIVE создайте свою собственную базу данных с именем вашего пользователя (например yakupov).

CREATE DATABASE имя\_вашего\_пользователя;

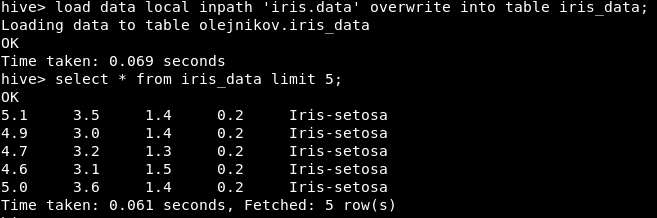
1. Переключитесь на вашу созданную базу данных из консоли HIVE командой USE имя\_вашего\_пользователя;



1. Создайте таблицу **iris\_data**



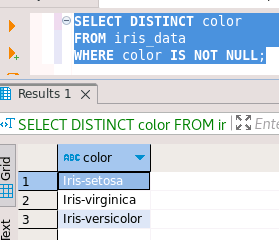
1. Загрузите данные из файла iris.data (ссылка на данные) в созданную таблицу **iris\_data**



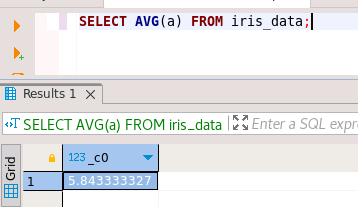


1. Выполните следующие SQL выражения

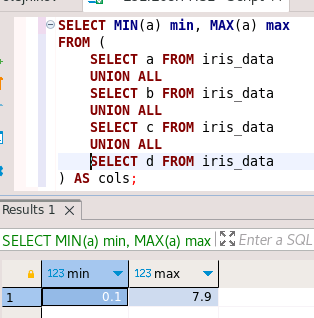
○ Вывести уникальные типы цветов ирисов на экран используя SQL запрос



○ Найти среднее по первому атрибуту данных файла, используя SQL запрос



○ Найти min, max элемент по всем числовым столбцам в рамках группы цветов ирисов, используя SQL запрос



(Предоставить преподавателю) приложите пожалуйста к вашей работе скриншоты результата выполнения ваших SQL запросов из шага f).





PS. Две пустые строки в конце датасета я не удалял, команды не было. Не знаю, правильно или нет, прошу обратную связь.