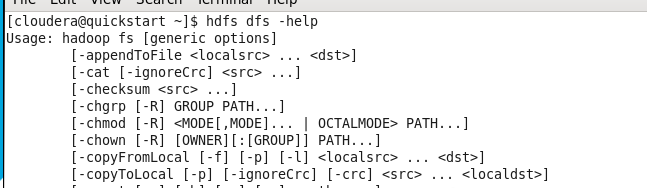
Практическая работа 3-1. Корпоративные платформы. Cloudera (Гадильшина В.Е., БД-231м)

**Задание к теме: Работа в HDFS в экосистеме cloudera**

**3.1.1. Развернуть виртуальное окружение.**

**+**

**3.1.2. Вывести с помощью команды help описание основных команды shell-клиента.**

****

**3.1.3. Просмотреть корневую директорию HDFS.**

**Изображение выглядит как текст, Шрифт, снимок экрана, белый

Автоматически созданное описание**

**3.1.4. Создать в HDFS в директории /user/mgpu поддиректорию ваше\_фио.**

****

**3.1.5. Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем, образованным вашими инициалами base64 /dev/urandom | head -c 10000000 > file.txt**

**Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание**

**Изображение выглядит как текст, программное обеспечение, Значок на компьютере, веб-страница

Автоматически созданное описание**

**3.1.6. Заархивировать созданный текстовый файл gzip -c file.txt > file.gz .**

**Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание**

**3.1.7. Скопировать текстовый файл и архив в директорию /user/mgpu/fio HDFS виртуальной машины.**

**Изображение выглядит как текст, Шрифт, число, снимок экрана

Автоматически созданное описание**

**3.1.8. Просмотреть файл и архив с помощью утилит cat, text в комбинации с каналами и утилитами head, tail -- привести не менее 3 вариантов команд и просмотра файла.**

****

**Изображение выглядит как текст, черно-белый, газета, бумага

Автоматически созданное описание**

**Изображение выглядит как текст, Шрифт, белый

Автоматически созданное описание**

**Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание**

****

**Изображение выглядит как текст, снимок экрана, Шрифт, информация

Автоматически созданное описание**

****

**Изображение выглядит как текст, Шрифт, черно-белый, белый

Автоматически созданное описание**

****

VA2wx

**Изображение выглядит как текст, Шрифт, белый, черно-белый

Автоматически созданное описание**

**Изображение выглядит как текст, Шрифт, черно-белый, белый

Автоматически созданное описание**

**Изображение выглядит как текст, Шрифт, снимок экрана, белый

Автоматически созданное описание**

**Изображение выглядит как текст, Шрифт, белый, черно-белый

Автоматически созданное описание**

**3.1.9. Создать копию файла file.txt вида date\_file.txt, где в начале имени файла-копии указана текущая дата. Вывести листинг.**

**Изображение выглядит как текст, снимок экрана, Шрифт

Автоматически созданное описание**

**Изображение выглядит как текст, снимок экрана, Шрифт, алгебра

Автоматически созданное описание**

**3.1.10. Вывести статистику по директории /user/mgpu/fio виртуальной машины.**

**Изображение выглядит как текст, снимок экрана, Шрифт, алгебра

Автоматически созданное описание**

**Изображение выглядит как текст, Шрифт, число, чек

Автоматически созданное описание**

**Изображение выглядит как текст, снимок экрана, дисплей, программное обеспечение

Автоматически созданное описание**

**3.1.11. Удалить поддиректорию /fio со всем содержимым.**

****

**Изображение выглядит как текст, Шрифт, линия, снимок экрана

Автоматически созданное описание**

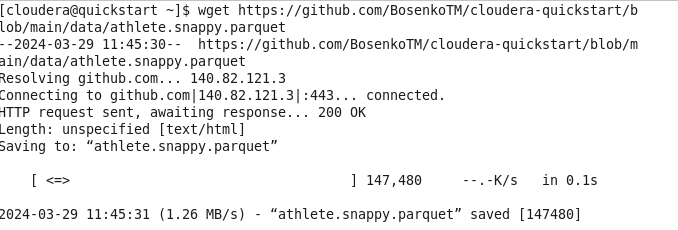
**3.1.12. Подсчитать количество слов в файле внутри HDFS с помощью методологии Map Reduce (размер файла не менее 128 Мб).**

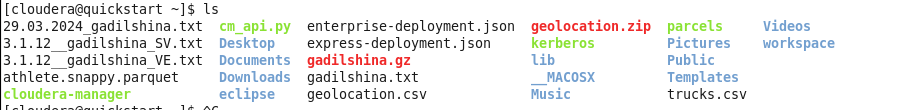
**Изображение выглядит как текст, снимок экрана, Шрифт, документ

Автоматически созданное описание**

**3.2. Создание таблицы в Hive**

1. **Скачать [датасет](https://github.com/BosenkoTM/cloudera-quickstart/blob/main/data/athlete.snappy.parquet) или**[**тут**](https://storage.googleapis.com/otus_sample_data/athlete.snappy.parquet)





1. **Через HUE загрузите файл в папку /user/cloudera/athlete.**

**+**

1. **В навигационном меню выберите Files.\**

**+**

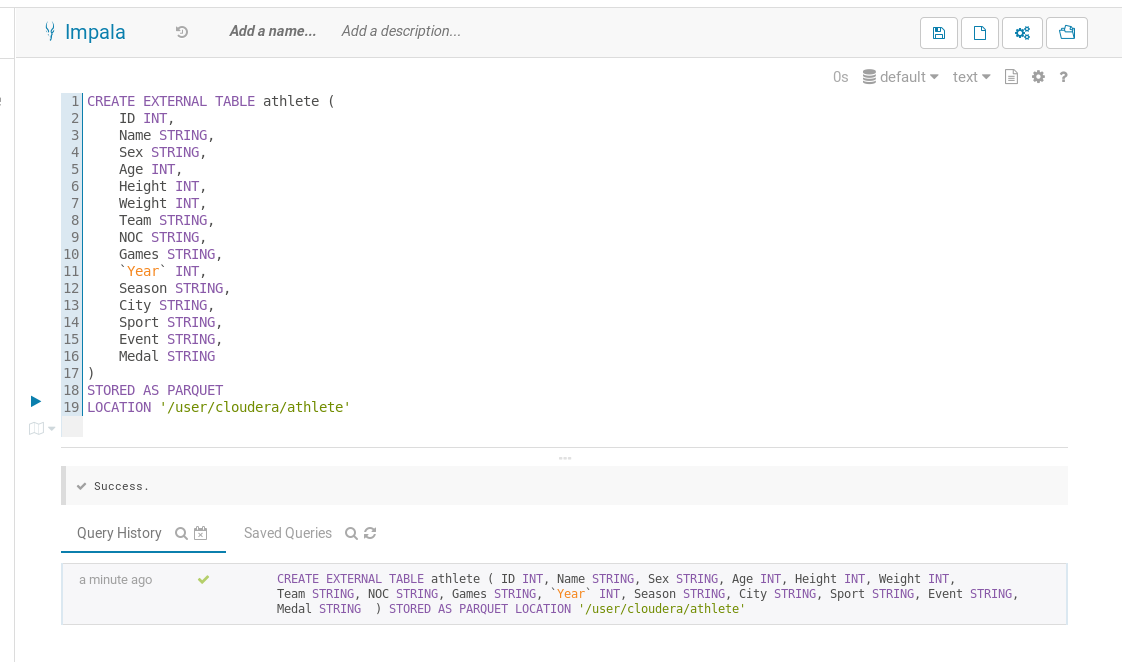
1. **Создайте**Изображение выглядит как текст, Шрифт, число, снимок экрана

   Автоматически созданное описание
2. **Загрузите файл в HDFS, нажав Upload.**

Изображение выглядит как текст, Шрифт, число, линия

Автоматически созданное описание

1. **Перейдите в “Editor > Hive” и выполните запрос:**



Изображение выглядит как текст, Шрифт, белый, снимок экрана

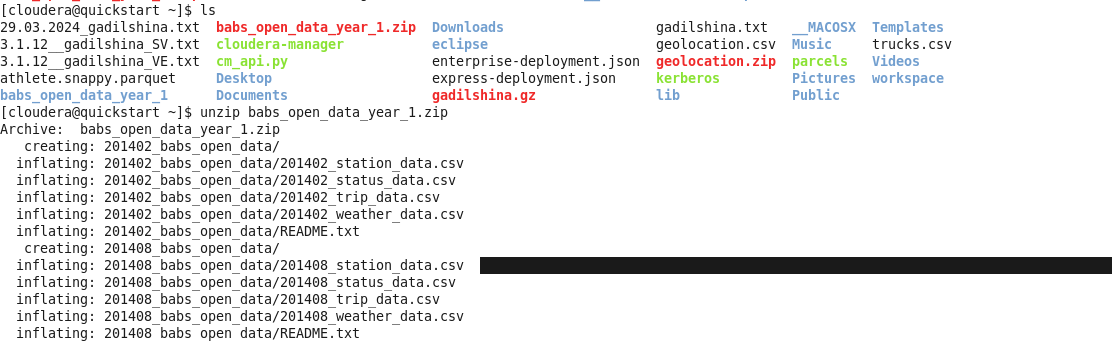
Автоматически созданное описание

**3.3 Проанализировать и визуализировать данные с помощью Impala (высокоскоростной механизм запросов SQL) или Hive.**

* **Загрузить и разархивировать**[**babs\_open\_data\_year\_1.zip**](https://disk.yandex.ru/d/JrboaizPXSh0Mg)**.**

Изображение выглядит как текст, снимок экрана, чек, Шрифт

Автоматически созданное описание



* **Перенести данные 201402\_trip\_data.csv в HDFS.**

Изображение выглядит как текст, чек, число, Шрифт

Автоматически созданное описание

* **Создать таблицу в Hive с привязкой к внешним данным 201402\_trip\_data.csv.**

Изображение выглядит как текст, программное обеспечение, Шрифт, снимок экрана

Автоматически созданное описание

Изображение выглядит как текст, Шрифт, белый, снимок экрана

Автоматически созданное описание

* **выполнить запрос**

select `startstation`, `endstation`, count(\*) as trips

from `default`.`201402\_trip\_data`

group by `startstation`, `endstation`

order by trips desc;

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

* **Создать гистограмму, щелкнув значок «Hue Bar»:**

Изображение выглядит как снимок экрана, График, текст, линия

Автоматически созданное описание

* **Установить ось X в качестве начальной станции, а ось Y — в качестве маршрута. Установить лимит 10.**

Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описание

* **Выгрузить результаты, выбрав CSV или Excel.**

