**Литература:**

<https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/FileSystemShell.html>

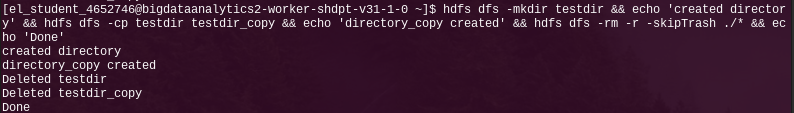
<https://hadoop.apache.org/docs/r2.7.1/hadoop-project-dist/hadoop-hdfs/HDFSCommands.html>

<https://www.oreilly.com/library/view/hadoop-with-python/9781492048435/ch01.html>

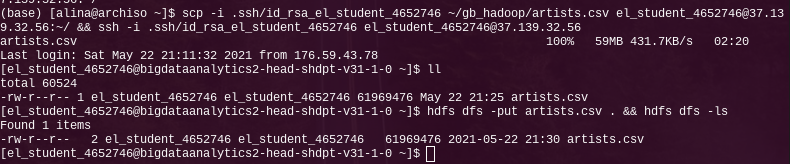
**Задания:**

1. Опробовать консольные утилиты для работы с кластером

* *Создать/скопировать/удалить папку*



* *Положить в HDFS любой файл*



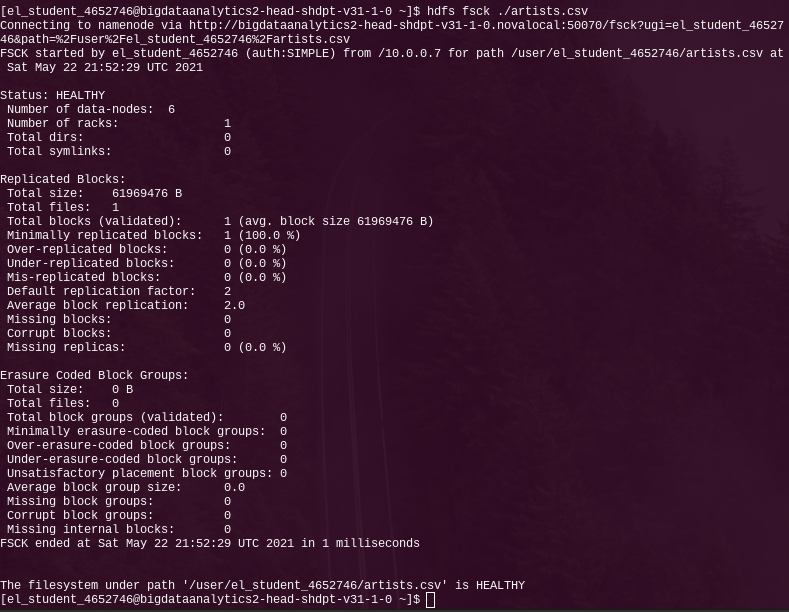
* *Скопировать/удалить этот файл*

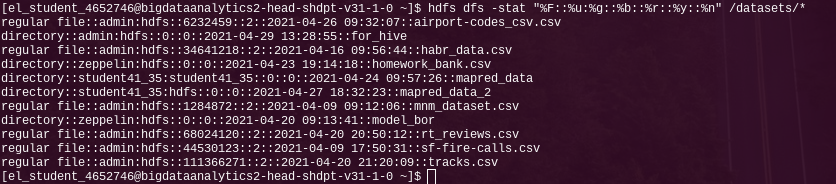


* *Просмотреть размер любой папки*

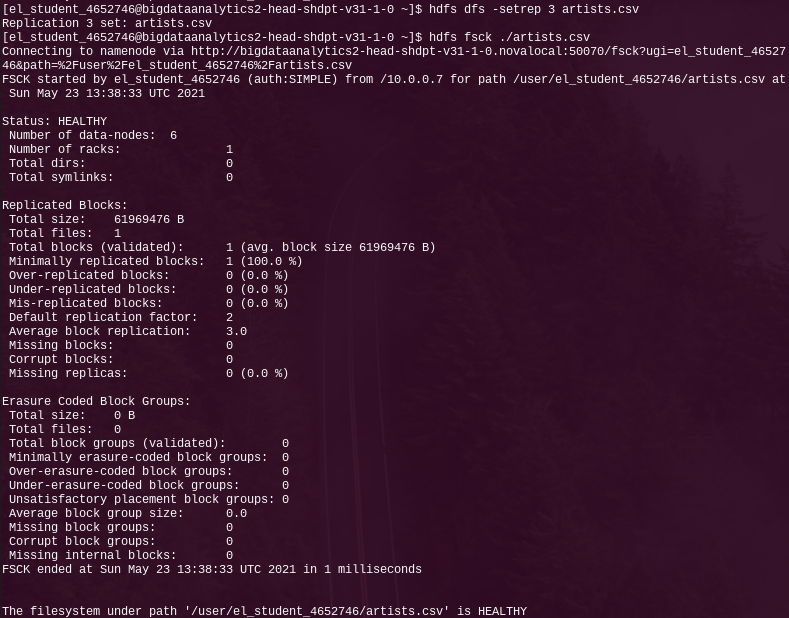


* *Посмотреть как файл хранится на файловой системе (см. команду fsck)*





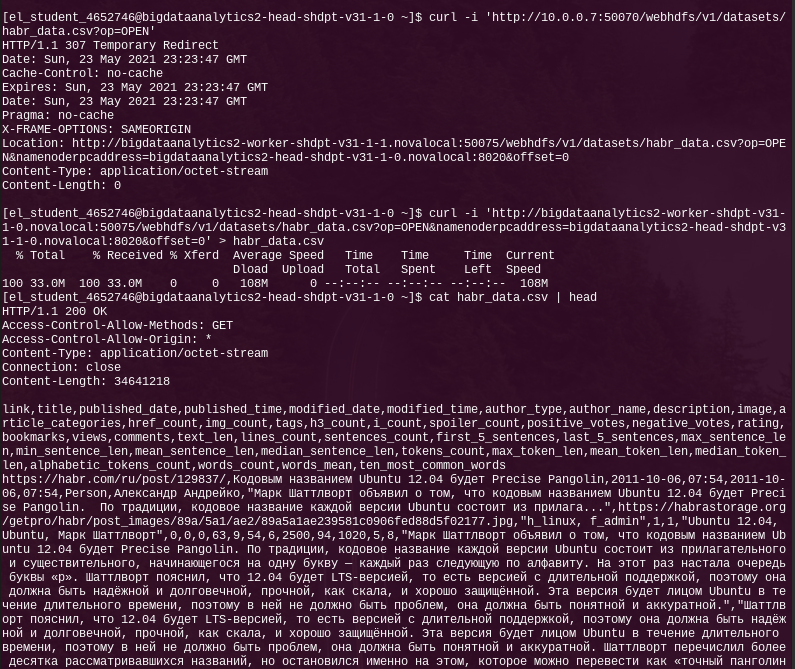
* \*Установить нестандартный фактор репликации (см. команду setrep)



\*2. Опробовать доступ для работы с кластером

* Используя утилиту CURL (нужна только консоль)

Попробовала прочитать файл из hdfs



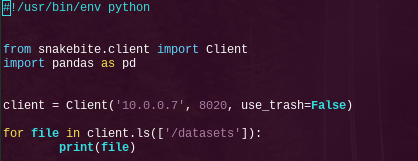
* Используя python3 + snakebite (python3 + virtual env + пакет snakebite)

10.0.0.7 – внутренний адрес NameNode

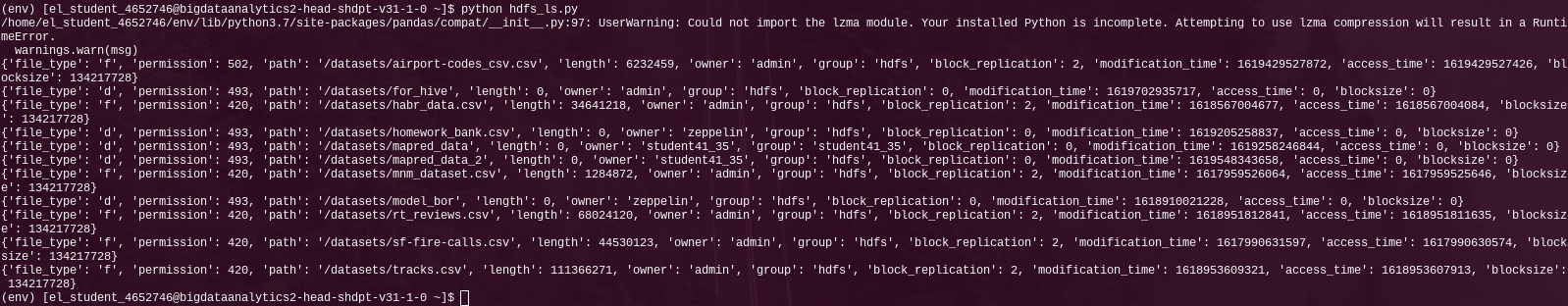
8020 - RPC port of the NameNode

http://10.0.0.7:50070/webhdfs/v1/?op=LISTSTATUS – ссылка на webhdfs

Написала простой скрипт.py:



Запуск скрипта:



Как создать и использовать энв на питоне

python3.7 -m venv ./env

source ./env/bin/activate

pip3 install snakebite