

Блок 2. Работа с HDFS

Флаги “-mkdir” и “-touchz”

1. `hdfs dfs -mkdir /new_folder`
2. `hdfs dfs -mkdir /new_folder/subfolder`
3. Trash это папка куда отправляются файлы после удаления. Реальное удаление файла из пространства имен HDFS и освобождение соответствующих блоков памяти происходит по истечении определенного количества времени.
Для того чтобы удалить файл минуя Trash, нужно при выполнении команды удаления (`hdfs dfs -rm`) добавить флаг `-skipTrash`.
4. `hdfs dfs -touchz /new_folder/subfolder/newFile`
5. `hdfs dfs -rm /new_folder/subfolder/newFile`
6. `hdfs dfs -rm -r /new_folder`

Флаги “-put”, “-cat”, “-tail”, “-cp”

1. `hdfs dfs -mkdir /folder`
`hdfs dfs -put text_file.txt /folder/`
2. `hdfs dfs -cat /folder/text_file.txt`
3. `hdfs dfs -cat /folder/text_file.txt | tail`
4. `hdfs dfs -cat /folder/text_file.txt | head`
5. `hdfs dfs -mkdir /folder1`
`hdfs dfs -cp /folder/text_file.txt /folder1/`

Флаги “hdfs dfs- setrep -w”, “hdfs fsck /path -files - blocks -locations”

1. `hdfs dfs -setrep -w 2 /folder/text_file.txt`
Работала с очень легким текстовым файлом (1 Кб), увеличение числа реплик заняло на пару секунд больше времени, чем удаление. Предположу, что при работе с большими данными процесс увеличения количества реплик значительно больше времени занимает, чем удаление.
2. `hdfs fsck /folder/text_file.txt -files -blocks -locations`
3. `hdfs fsck -blockId blk_1073741840_1016`