Hadoop Commands

Stefan

November 29, 2013

1 List of HDFS commands

- cat - copy files to stdout, similar to UNIX cat command. (prints file in terminal). Example:

  ```
  hadoop dfs -cat /user/hadoop/file4
  ```

- copyFromLocal copy single src, or multiple srcs from local file system to the destination filesystem. Source has to be a local file reference. Example:

  ```
  hadoop dfs -copyFromLocal localfile /user/hadoop/file1
  ```

- copyToLocal copy files to the local file system. Files that fail the CRC check may be copied with the -ignorecrc option. Files and CRCs may be copied using the -crc option. Destination must be a local file reference. Example:

  ```
  hadoop dfs -copyToLocal /user/hadoop/file localfile
  ```

- cp copy files from source to destination. This command allows multiple sources as well in which case the destination must be a directory. Similar to UNIX cp command. Example:

  ```
  hadoop dfs -cp /user/hadoop/file1 /user/hadoop/file2
  ```

- getmerge take a source directory and a destination file as input and concatenate files in src into the destination local file. Optionally addnl can be set to enable adding a newline character at the end of each file. Example:

```
hadoop dfs -getmerge /user/hadoop/mydir/ ~/result_file
```

- ls for a file returns stat on the file with the format:
  filename <number of replicas> size modificationDate modificationTime permissions userid groupid
  For a directory it returns list of its direct children as in UNIX, with the format:
  dirname <dir> modificationDime modificationTime permissions userid groupid
  Example:

  ```
  hadoop dfs -ls /user/hadoop/file1
  ```

- lsr recursive version of ls. Similar to UNIX ls -R command. Example:

  ```
  hadoop dfs -lsr /user/hadoop/
  ```

- mkdir create a directory. Behaves similar to UNIX mkdir -p command creating parent directories along the path (for bragging rights, what is the difference?) Example:

  ```
  hadoop dfs -mkdir /user/hadoop/dir1 /user/hadoop/dir2
  ```

- mv move files from source to destination similar to UNIX mv command. This command allows multiple sources as well in which case the destination needs to be a directory. Moving files across filesystems is not permitted. Example:

  ```
  hadoop dfs -mv /user/hadoop/file1 /user/hadoop/file2
  ```

- rm delete files, similar to UNIX rm command. Only deletes empty directories and files. Example:

  ```
  hadoop dfs -rm /user/hadoop/file1
  ```

- rmr recursive version of rm. Same as rm -r on UNIX. Example:

  ```
  hadoop dfs -rmr /user/hadoop/dir1/
  ```

- tail Displays last kilobyte of the file to stdout. Similar to UNIX tail command. Options:

- f output appended data as the file grows (follow)

```
hadoop dfs -tail /user/hadoop/file1
```

- test perform various test. Options:
  - e check to see if the file exists. Return 0 if true.
  - z check to see if the file is zero length. Return 0 if true.
  - d check return 1 if the path is directory else return 0.

  Example:

  ```
  hadoop dfs -test -e /user/hadoop/file1
  ```

- touchz create a file of zero length. Similar to UNIX touch command. Example:

  ```
  hadoop dfs -touchz /user/hadoop/file1
  ```