



Hadoop

HDFS

Objectives

- What is HDFS?
- HDFS commands?
- Removing files and folders
- Creating scripts and aliases for HDFS commands
- Labs

What is HDFS?

- HDFS (Hadoop Distributed File System) is a filesystem designed for storing very large files across a cluster of hosts. The file system is highly scalable and manages itself.
- Thru Hadoop tools, accessing the location of a file or set of files is transparent to the user. From the user perspective, a file on a Hadoop cluster acts as a single file residing on a traditional file system.
- The user need not be aware of the physical location of the file or files residing in the cluster.
- Files may be split across many drives on many hosts. Again, the user / developer is usually not concerned with physical placement of the files or slices of the files.
- HDFS is used to provide parallel processing of very large files. Files may be split across multiple hosts giving processors the responsibility of asynchronously processing their own piece of the file.

HDFS Commands

- **cat** – types the contents of argument(s) to stdout
 - Example:
 - `hadoop fs -cat /user/rico/fruit`
- **chmod** – change the permissions of file argument(s)
 - Example:
 - `hadoop fs –chmod 644 /user/rico/fruit`
- **copyToLocal** – Copy a file from hdfs to the local filesystem
 - Example:
 - `hadoop fs –copyToLocal /user/rico/fruit local.fruit`

HDFS Commands (continued)

- **copyFromLocal** – Copy a file from the local filesystem to hdfs
 - Example:
 - `hadoop fs -copyFromLocal fruit /user/rico/hfruit`
- **count** – Count files and directories
 - Example:
 - `hadoop fs -count -v /user/rico/`
- **cp** – Copy files within the hdfs
 - Example:
 - `hadoop fs -cp /user/rico/fruit /user/kip/fruit`

HDFS Commands (continued)

- **expunge** – Empty your recycle bin
 - Example:
 - `hadoop fs -expunge`
- **find** – find files by name with a starting directory
 - `hadoop fs -find / -name "fr*"`
- **help** – Command help
 - `hadoop fs -help`
 - `hadoop fs -help ls`
- **ls** – Show list of files
 - Example:
 - `hadoop fs -ls /user/rico/*`

HDFS Commands (continued)

- **mkdir** – Create a directory
 - Example:
 - `hadoop fs -mkdir /user/rico/foods`
- **moveFromLocal** – Like `copyFromLocal` but local source gets deleted
 - Example:
 - `hadoop fs -moveFromLocal apples /user/rico/foods`
- **moveToLocal** – Displays a “not Implemented yet” message
 - Example:
 - `hadoop fs -copyToLocal /user/rico/foods/apples`
 - `hadoop fs -rm /user/rico/foods/apples`

HDFS Commands (continued)

- **mv** – Moves files with the hdfs
 - Example:
 - `hadoop fs -mv /user/rico/food/dairy /user/rico/fruits`
- **rm** – Deletes files
 - Example:
 - `hadoop fs -rm /user/rico/foods/apples`
- **rmdir** – Removes directories
 - Example:
 - `hadoop fs -rmdir /user/rico/foods/unuseddir`
 - “Add” `-ignore-fail-on-non-empty` to remove non empty directories

HDFS Commands (continued)

- **stat** – Give me stats on file(s)
 - Example:
 - `hadoop fs -stat "%F %u" /user/rico/food/apples`
- **tail** – Displays last K of file
 - Example:
 - `hadoop fs -tail /user/rico/foods/bigfood`
- **test** – checks traits of file. Sets the UNIX return code.
 - Example:
 - `hadoop fs -test -f /user/rico/foods/apples`
 - `hadoop fs -test -d /user/rico/foods`
 - `hadoop fs -test -e /user/rico/foods/apples`

HDFS Commands (continued)

- **touchz** – Create an empty file
 - Example:
 - `hadoop fs -touchz /user/rico/foods/dairy`
- **usage** – Displays quick syntax help for a command.
 - Example:
 - `hadoop fs -usage ls`
- But wait, there's more! But the above are the most needed.

Labs

- Scripting, Hadoop HDFS command, and more (Lab 2)
 - Write a shell script
 - Create some files
 - Move files to and from the HDFS
 - Set up some aliases
 - Automate some activities