

HDFS Shell



Agenda

- **HDFS Shell**
- **Basic Shell Commands**
- **Moving Data**
- **Deleting Data**
- **File stats**

HDFS Shell Commands

- **Used to Interact with HDFS File System**
 - `$hdfs dfs -<command> -<option> <URI>`
- **Example**
 - `$hdfs dfs -ls /`
- **URI usage:**
 - HDFS: `$hdfs dfs -ls hdfs://localhost/to/path/dir`
 - Local: `$hdfs dfs -ls file:///to/path/file3`
- **Schema and namenode host is optional, default is used from the configuration In core-site.xml**
 - `fs.default.name` property

HDFS Shell Commands

- **cat - stream source to stdout**
 - entire file: `$hdfs dfs -cat /dir/file.txt`
- **Almost always a good idea to pipe to head or tail**
- **Get the first 25 lines of file.txt**
 - `$hdfs dfs -cat /dir/file.txt | head -n 25`

HDFS Shell Commands

- **cp - copy files from source to destination**
 - `$hdfs dfs -cp /dir/file1 /otherDir/file2`
- **ls - for a file displays stats, for a directory displays immediate children**
 - `$hdfs dfs -ls /dir/`
- **mkdir - create a directory**
 - `$hdfs dfs -mkdir /brandNewDir`

HDFS Shell - Moving Data

- **mv - move from source to destination**
 - `$hdfs dfs -mv /dir/file1 /dir2/file2`
- **put - copy file from local filesystem to hdfs**
 - `$hdfs dfs -put localfile /dir/file1`
 - Can also use `copyFromLocal`
- **get - copy file to the local filesystem**
 - `$hdfs dfs -get /dir/file localfile`
 - Can also use `copyToLocal`

HDFS Shell - Deleting Data

- **rm - delete files**
 - `$hdfs dfs -rm /dir/fileToDelete`
- **rm -r - delete directories recursively**
 - `$hdfs dfs -rm -r /dirWithStuff`

HDFS Shell - Filesystem Stats

- **du - displays length for each file/dir (in bytes)**
 - `$hdfs dfs -du /someDir/`
- **Add -h option to display in human readable format instead of bytes**
 - `$hdfs dfs -du -h /someDir`

Resources

- **Hadoop: The Definitive Guide**
 - Tom White (Author)
 - O'Reilly Media; 4th Edition.

