QA Consulting.

Hadoop Command Reference Guide

BIG DATA COURSE

Prepared by
Thomas Knowles

+447858368323

Thomas.Knowles@qa.com

Contents

Overview

This guide describes the most common commands needing during the Hadoop Admin, Hive and Impala Exercises. Basic SQL is assumed for interacting with Hive and Impala, but intricate tasks like creating tables from datasets are explained.

Anything not provided here will be found in the lecture slides or exercise guide.

Unix Commands

All commands must be conducted from the Terminal Prompt (not the Hive prompt). Prefacing with 'sudo' will be necessary to access certain filepaths. Square brackets indicate optional arguments – they are NOT part of the syntax.

Most of these commands can be used for HDFS administration too, but must be prefaced with 'hdfs dfs' OR 'hadoop fs' and a '-' before the command (turning it into a flag). All filepaths for the HDFS must be absolute.

Access the Manual Page for a Command

man [command]

Listing Files in Directory. No filepath = current directory

ls [filepath]

Printing the Contents of a File

cat [filepath]

Moving a File

mv [source filepath] [destination filepath]

Removing a File. r = with sub-directories; f = ignore warnings

rm [-r] [-f]

Copying a File

cp [source filepath] [destination filepath]

Moving a File

mv [source filepath] [destination filepath]

Creating a new directory

mkdir [filepath]

Adding Permissions (r – read, w – write, x – execute) to a File

chmod +u[permissions] [filename]

Changing the Ownership of a File

chown [user][:usergroup]

Create a new Directory. Add '-p' to create parent directories too

mkdir [-p] [filepath]

Delete a Directory

rmdir [filepath]

Download and Install Package. [pm] will be 'apt-get' or 'yum'

sudo [pm] update

sudo [pm] install [package name]

Create a new Text File. [editor] can be 'vi', or any other installed editor

sudo [editor] [filepath].txt

Change Ownership of a File

chown [filename]

Run the previous Command with Sudo permissions

sudo !!

Vi Commands

To access commands, you will need to 'esc' from INSERT mode.

Enter INSERT (text entry) mode

i

Save and Exit

:wq

Exit without Saving

:q!

Copy Selection

yy

Paste Selection

p

Find String after cursor

/string

Find String before cursor

?string

Next/Previous Match

n/N

Write to a New File

:w [filename]

Write to a Pre-existing File

:w! [filename]

Hadoop Commands

All commands must be conducted from the Terminal prompt (not the Hive prompt). Square brackets indicate optional arguments – they are NOT part of the syntax. All filepaths for the HDFS must be absolute.

Most of the Unix Commands can also be used here. To convert a Unix command to a Hadoop command, preface it with 'hadoop fs' or 'hdfs dfs' and convert the command to a flag with '-'.

Using Unix commands in Hadoop - Example

hadoop fs -mv [source_filepath] [destination_filepath]

Moving a File from Host to Hadoop

hadoop fs -put [host_filepath] [hdfs_filepath]

Alternative for Host to Hadoop File Transfer (if 'put' doesn't work)

hadoop fs -copyFromLocal [host_filepath] [hdfs_filepath]

Moving a File from Hadoop to Host

hadoop fs -get [hdfs_filepath] [host_filepath]

Alternative for Hadoop to Host File Transfer (if 'get' doesn't work)

hadoop fs -copyToLocal [hdfs_filepath] [host_filepath]

Set Replication Number

hadoop fs -setrep [replication_number] [filepath]

Hive Commands

Hive is mostly just a subset of standard SQL (HiveQL). Basics are covered in the PowerPoint slides, which Hive shares with SQL. The more verbose commands are repeated below for ease of reference.

Creating a Table with Primitives, Array and Map

```
CREATE TABLE jobs
        (id INT,
        title STRING,
        salary INT,
        posted TIMESTAMP,
        headquarters_address ARRAY <STRING>,
        reviews STRUCT <review1 : STRING, review2 : STRING, review3 : STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ‘,’;
```

Optional Lines for Complex Types

```
COLLECTION ITEMS TERMINATED BY ','
MAP KEYS TERMINATED BY ':'
```

Loading Data into Table

```
LOAD DATA [LOCAL] INPATH ‘filepath’ [OVERWRITE] INTO TABLE ‘table_name’;
```

Accessing Complex Data Types

```
array[index]
map[key]
struct.element
```