

How to use Hadoop VM

VM details

The VM runs a 32-bit version of Xubuntu 14.04 LTS. I tested it with 512 MB of RAM and it goes pretty smoothly.

Hadoop is installed in /home/hadoop/hadoop

I added \$HADOOP_HOME/sbin and \$HADOOP_HOME/bin to PATH, in order to make it easier to execute Hadoop commands.

Username & Password

Username: hadoop

Password: hadoop

How to start Hadoop

I created two scripts, **start-hadoop.sh** and **stop-hadoop.sh**, that automate the process of starting and stopping Hadoop, respectively.

start-hadoop.sh formats the files system, creates the directories necessary to run MR jobs, and starts namenode and datanode.

stop-hadoop.sh stops all the daemons.

Start Hadoop:

> start-hadoop.sh

Stop hadoop:

> stop-hadoop.sh

How to run a simple wordcount

I created a folder **books** in \$HOME with some books to use with wordcount.

Transfer books to HDFS:

```
> hdfs -put $HOME/books input
```

Run wordcount:

```
> hadoop jar  
$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-  
examples-2.6.4.jar wordcount input output
```

View output files:

```
> hdfs dfs -cat output/*
```

or

```
> hdfs dfs -get output <local_folder>
```

```
> less <local_folder>/*
```