Hadoop Installation Document

Abhay Dandekar

Local Hadoop Installation

Version 1.0

Hadoop Installation Document

Abhay Dandekar

Scope

This document is designed as a guide to local installation of various hadoop components on your local.

At actual production installations, a distribution like Cloudera Distribution / Hontonworks Distribution are preferred. Local installations are preferred only for local development work.

Installations Steps

HDFS Installation

Download the tarball from the download location mentioned below.
Link : Apache_link_for_v3.3.1

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export HADOOP_HOME=<YOUR_UNTARRED_LOCATION>
3. Edit the configuration files for namenode dir and datanode dir
    a. Add the following entries to core-site.xml

```
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
</property>
```

    b. Add the following entries to hdfs-site.xml:

```
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
<property>
  <name>dfs.namenode.name.dir</name>
  <value>file:///<SOME_PATH_ON_YOUR_MACHINE>/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///<SOME_PATH_ON_YOUR_MACHINE>/datanode</value>
</property>
```

4. If this is your first installation, format HDFS using the following command
    a. $ hdfs namenode -format
5. Start HDFS using the startup script
    a. $ bin/start-dfs.sh
6. Your HDFS should startup
7. Check the installation using the following command
    a. $ hdfs dfs -ls /

YARN Installation

Download the tarball from the download location mentioned below.
Link : Apache_link_for_v3.3.1

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export HADOOP_HOME=<YOUR_UNTARRED_LOCATION>
3. Edit the configuration files for YARN and Map-Reduce.
    a. Edit etc/hadoop/yarn-site.xml

```
<property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
</property>
```

    b. Copy your mapred-site.xml.template to mapred-site.xml
    c. Edit etc/hadoop/mapred-site.xml

```
<property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
</property>
```

4. Start YARN using the following script
    a. $ sbin/start-yarn.sh

Pig Installation

Download the tarball from the download location mentioned below.
Link : Inet-Link

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export HADOOP_HOME=<YOUR_HADOOP_LOCATION>
3. Export PATH=$PATH:<YOUR_UNTAR_LOCATION>/bin

Sqoop Installation

Download the tarball from the download location mentioned below.
Link : Inet-Link

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export HADOOP_HOME=<YOUR_HADOOP_LOCATION>
3. Export SQOOP_HOME=<YOUR_UNTAR_LOCATION>
4. Export PATH=$PATH:<YOUR_UNTAR_LOCATION>/bin
5. Store all your JDBC jars (e.g mysql-connector-<version>.jar) under $SQOOP_HOME/lib

Hive Installation

Download the tarball from the download location mentioned below.
Link : Inet-Link

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export HADOOP_HOME=<YOUR_HADOOP_LOCATION>
3. Export HIVE_HOME=<YOUR_HIVE_UNTARRED_LOCATION>
4. Export PATH=$PATH:<YOUR_UNTAR_LOCATION>/bin
5. Create HDFS dir using the following commands
    a. $HADOOP_HOME/bin/hadoop fs -mkdir -p /user/hive/warehouse
    b. $HADOOP_HOME/bin/hadoop fs -mkdir -p /tmp
6. Provide updated permissions using the following commands
    a. $HADOOP_HOME/bin/hadoop fs -chmod g+w /user/hive/warehouse
    b. $HADOOP_HOME/bin/hadoop fs -chmod g+w /tmp
7. Initialise the metastore DB using the schema_tool
    a. cd $HIVE_HOME # ENSURE YOU ARE ALWAYS IN HIVE_HOME
    b. $HIVE_HOME/bin/schematool -dbType derby -initSchema
8. Start you HIVE prompt
    a. cd $HIVE_HOME # ENSURE YOU ARE ALWAYS IN HIVE_HOME
    b. ./bin/hive
9. Test your hive prompt. Execute the following on hive prompt
    a. show databases # THIS COMMAND SHOULD NOT ERROR OUT

Note : Recommended tutorial :
https://cwiki.apache.org/confluence/display/Hive/GettingStarted#GettingStarted-InstallationandConfiguration

Spark Installation

Download the tarball from the download location mentioned below.
Link : Inet-Link

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export HADOOP_HOME=<YOUR_HADOOP_LOCATION>
3. Export SPARK_HOME=<YOUR_UNTAR_LOCATION>
4. Start master using the following command
    a. $ sbin/start-master.sh
5. Start slave using the following command
    a. $ sbin/start-slave.sh spark://<your_master_host>:7077
    b. Note : 7077 is the default port, it may vary across versions

Flume Installation

Download the tarball from the download location mentioned below.
Link : Inet-Link

Follow the steps below for installation.

1. Untar the downloaded jar
    a. $ tar -xvzf <YOUR_DOWNLOADED_TAR_BALL_LOCATION>
2. Export FLUME_HOME=<YOUR_UNTAR_LOCATION>