Install java sudo apt-get install openjdk-7-jdk

Ctrl c 是终止

bin/hdfs dfs -command

-mkdir

-rm -r

-put localdir hdfsdir

-get hdfsdir localdir

Bin/hdfs

/home/ubuntu/hadoop-2.7.3/etc/hadoop/mapred-site.xml //changing to local is running on local . Changing to yarn is running on yarn

Local模式下运行程序是在hdfs上，需要用hdfs dfs - get 获取output到本地

0. Install java sudo apt-get install openjdk-7-jdk

1.Download Hadoop

Wget ..........

Tar -xvf hadoop......

Prepare to Start the Hadoop Cluster

Unpack the downloaded Hadoop distribution. In the distribution, edit the file etc/hadoop/hadoop-env.sh to define some parameters as follows:

export JAVA\_HOME=/usr

Pseudo-Distributed Operation

1.

Pseudo-Distributed Operation

Hadoop can also be run on a single-node in a pseudo-distributed mode where each Hadoop daemon runs in a separate Java process.

Configuration

Use the following:

etc/hadoop/core-site.xml:

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

etc/hadoop/hdfs-site.xml:

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

2.修改权限

ubuntu@ip-172-31-37-58:~$ chmod 777 ~/.ssh

3.

1. ssh-keygen -t rsa

Press enter for each line

2. cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

3. chmod og-wx ~/.ssh/authorized\_keys

4.Now check that you can ssh to the localhost without a passphrase:

$ ssh localhost

5.

The following instructions are to run a MapReduce job locally. If you want to execute a job on YARN, see [YARN on Single Node](https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html" \l "YARN_on_Single_Node).

Format the filesystem:

$ bin/hdfs namenode -format

Start NameNode daemon and DataNode daemon:

$ sbin/start-dfs.sh

The hadoop daemon log output is written to the $HADOOP\_LOG\_DIR directory (defaults to $HADOOP\_HOME/logs).

6.Run some of the examples provided:

$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar grep input output 'dfs[a-z.]+'

7.Examine the output files: Copy the output files from the distributed filesystem to the local filesystem and examine them:

$ bin/hdfs dfs -get output output

$ cat output/\*

or

View the output files on the distributed filesystem:

$ bin/hdfs dfs -cat output/\*

8. When you’re done, stop the daemons with:

$ sbin/stop-dfs.sh

1. YARN

Configure parameters as follows:etc/hadoop/mapred-site.xml:

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

etc/hadoop/yarn-site.xml:

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

</configuration>

10. Start ResourceManager daemon and NodeManager daemon:

$ sbin/start-yarn.sh

11. When you’re done, stop the daemons with:

$ sbin/stop-yarn.sh