Installation Errors and Exceptions

Hadoop

Error: PriviledgedActionException as:hduser cause:java.io.IOException: File could only be replicated to 0 nodes, instead of 1

The datanode and namenode are in inconsistent state as the namespace ids don't match. Thus datanode is not able to recogonize the namenode it was started by.

To rectify this /tmp/hadoop-hduser/dfs is to be deleted from both master node and slave node so that all the previously associated state is no longer available. After then, we start the dfs again, all the ids get updated and the problem is resolved.

Error: Shutting down DataNode at java.net.UnknownHostException:

This problem was caused when the master node was not able to resolve the address of a host in the cluster. The host names are to be verified in /etc/hosts file where the details about all the related hosts are given. Mostly this problem occurs because of error in specifying the host names on slaves.
eg.
127.0.0.1 A B #A and B are the host names

Master node is not able to identify which is the host among A and B. So it keeps on trying. This exception is finally given when it is unable to resolve the host name.

To solve this problem, we must properly and uniquely assign host names for each hosts and there should be no ambiguity in their names.

ERROR: org.apache.hadoop.hdfs.server.namenode.NameNode: java.net.BindException: Problem binding to master..

This error arises when we start DFS. The reason is a previous instance is already running I.e The Java process for NameNode is already running in the system.

To solve this, firstly identify which process is running by executing “sudo jps” and kill the process explicitly. The cluster is then restarted again. Now the exception gets resolved

Error:SafeModeException:org.apache.hadoop.hdfs.server.namenode.SafeModeException tion:

On startup, the NameNode enters a special state called Safemode. Replication of data blocks does not occur when the NameNode is in the Safemode state.

If this error occurs, we need to delete /app/hadoop/tmp/* reformat the namenode and restart the cluster.
We may leave the safe mode forcibly by typing the command:
hadoop dfsadmin -safemode leave

Mahout:

Error : While installing mahout you get **Build Failure** in between.

Solution:
In /usr/local/apache-maven-3.0.4/conf/settings.xml change the proxy setting and port number accordingly.
When you downloaded the mahout check whether pom.xml is present or not.

Rhadoop:

The rmr package offers Hadoop MapReduce functionalities in R.

Make sure you have install rjava, rcpp. and all the packages ("RJSONIO"); ("itertools");' ("digest"); Then Install rmr.

When it gives following error while installing rmr,

Error : ** R ** inst ** preparing package for lazy loading ** help * instialling help indices ** building package indices ** testing if installed package can be loaded .onLoad failed in loadNamespace()**

Check whether rJava is present in /usr/local/lib/R/site-library/
If rJava is present then check rJava.so is present or not in /usr/local/lib/R/site-library/rJava/ libs/rJava.so
If present, then in R execute command library(rmr). If this gives an error, it's clear that rmr is not installed properly. We have to install rmr1.3.1 instead of rmr2. The version currently present on github is rmr2_2.0.0. Older one can be found in github repository. After installing rmr1.3.1, set the hadoop environment variables in the system's bashrc by:

export HADOOP_HOME=/usr/lib/hadoop
export HADOOP_CONF=/etc/hadoop/conf

Now install rmr, it will be set up properly.

Similar is the case with rhdfs. It gives the error while installing

```
installing source package ‘rhdfs’ ... ** R ** inst ** preparing package for lazy loading ** help *** instialling help indices ** building package indices ** testing if installed package can be loaded Error : .onLoad failed in loadNamespace() for 'rhdfs', details: call: fun(libname, pkgname) error: rhdfs Error: loading failed Execution halted ERROR: loading failed
removing ‘/usr/local/lib/R/site-library/rhdfs’
```

If we check the installation by calling library(rhdfs) it also gives the error :

```
** R ** inst ** preparing package for lazy loading ** help *** instialling help indices ** building package indices ** testing if installed package can be loaded Error : .onLoad failed in loadNamespace() for 'rhdfs'
```

The solution is you must set the Hadoop HOME , CONF , CMD and STREAMING path in the bashrc file.

export HADOOP_HOME=/usr/lib/hadoop

export HADOOP_CONF=/etc/hadoop/conf
export HADOOP_CMD=/usr/bin/hadoop
export HADOOP_STREAMING=/usr/lib/hadoop/contrib/streaming/hadoop-streaming-1.0.3.jar

While installing the package of rmr1.3.1 you get the older version of rhdfs. Instead of installing the older version, install the new version of rhdfs. Your error would get resolved.

Giraph

java.lang.IllegalStateException: run: Caught an unrecoverable exception onlineZooKeeperServers: Failed to connect in 10 tries!

After downloading and extracting zookeper, in conf/zoo.cfg set clientPort=2181 then start the zookeeper by executing bin/zkServer.sh

Starting zookeeper is necessary for giraph to run on hadoop cluster, as zookeeper allows distributed process to coordinate with each other. If zookeper is not present then giraph wouldn't run on hadoop cluster.

While executing the job we get the following error:

```
org.apache.giraph.graph.BspServiceWorker.failureCleanup(BspServiceWorker.java:735)
    at org.apache.giraph.graph.GraphMapper.run(GraphMapper.java:648)
     at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:764)
    at org.apache.hadoop.mapred.MapTask.run(MapTask.java:370)
    at org.apache.hadoop.mapred.Child$4.run(Child.java:255)
    at java.security.AccessController.doPrivileged(Native Method)
    at javax.security.auth.Subject.doAs(Subject.java:396)
```

Check for this file giraph-0.2-SNAPSHOT-for-hadoop-0.20.203.0-jar-with-dependencies.jar in /usr/local/giraph/target and ensure that the number of workers argument should not be greater than the number of map slots.