Hadoop Configuration Simplified : Master Slave Architecture

Posted by **datumengineering** on February 11, 2012

Master-Slave: Hadoop configuration has simple components which can be divided as Master and Slave components. Master: NameNode, Secondary namenode and Job tracker. Slave: data node and task tracker.

Following are the key points of configuration:

1. Namenode always be localize during hadoop configuration in cluster environment.

2. Entries for master and slave are in master and slaves files repectively. Master translate to namenode host and job tracker.

3. All configuration can be categorize in 4 aspects:

- Environment variable configuration
- Hadoop-HDFS configuration.
- Master-Slave configuration.
- Map Reduce configuration.

I) **Environment configuration**: Set up all environment variables to run the scripts in HDFS environment.

{Files: hdfs-env.sh}

II) **Hadoop-HDFS configuration**: This includes entries for hosts for namenode, secondary namenode, job trackers and task trackers.

{Files: core-site.xml, hdfs-site.xml, mapred-site.xml}

III) **Text file configuration**: There are 2 text entries in hadoop configuration for master and slaves. One for master node i.e. Host entry for master node and other for slaves i.e. Host entry for all slaves machin where data node and task tracker will run.

{Files: masters, slaves}

IV) **Hadoop metrics and job log configuration**: in map reduce configuration setting it captures all the metrics and logs related to map reduce program.

{Files: hadoop-metrics.properties, log4j.properties}

This is how brief function of different components together:

Namenode and job tracker starts at local machine, starts secondary node on each machine listed in the masters file. Eventually start tasktracker and datanode on each machine listed in the masters file. In a cluster environment namenode, secondary namenode and job tracker run on single machine as a master node. However in a large cluster it can be sparated. When namenode and jobtracker are on separate node their slaves files should be in synch.

Source:http://datumengineering.wordpress.com/2012/02/11/hadoop-configuration-simplified-master-slave-architecture/