Cloudera Standard Installation Steps
Chen Sun 1000983

Note: I installed Cloudera on 2 machines (A and B) and used later to run the WordCount.java, but most of the steps which indicated how to install and configure Cloudera and services are screenshots from A machine, WordCount runnable screenshots are token from B.

I used VirtualBox and installed Ubuntu 64bit 12.04 precise system on both my laptop (ram 2GB, i3 dual cores) and SCU library's computer (ram 8GB, i7, 8 cores) to do the Cloudera Standard Installation. After my laptop completed the installation the virtual machine crashed once I started the Cloudera services. So I just do the below screen shot on my library's computer's installation.

1. First install an Ubuntu 64bit 12.04 precise system on VirtualBox

2. Download the Cloudera-manager-installer.bin from its website and use sudo 's privilege to run it.
3. It takes long time to installing the Cloudera Manager Server.

4. When the previous step is done, use browser and input 'localhost:7180' to go the Cloudera Manager's host web page.

5. Before proceed on, I strongly suggest to do these two steps' configurations for ubuntu system. The first one is for configuration of sudo's ssh command; The second one is for IP address for host.
6. Just input your hosts'/host's name to which you want to install hadoop to find it and continue.
7. Here this step is close related to the step 5 in which I configured the `sudo ssh`.

8. In the below step, if you haven't configured well as my picture in step 5, you would get stuck at authentication.
9. Complete it.

10. The following step took a lot of time, even in the library high performed system.
Cluster Installation

Installing Selected Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

- CDH 4.3.0-1.cdh4.3.0.p0.22: 0%
- SOLR 0.9.1-1.cdh4.3.0.p0.275: 1%
- IMPALA 1.0.1-1.p0.431: 2%

Back 1 2 3 4 5 Continue
Cluster Installation

Installing Selected Parcels

The selected parcels are being downloaded and installed on all the hosts in the cluster.

- CDH 4.3.0-1.cdh4.3.0.p0.22
- SOLR 0.9.1-1.cdh4.3.0.p0.275
- IMPALA 1.0.1-1.p0.431

Back | Continue
11. The next step is to inspect hosts for correctness.

12. There would be 1~2 warning. Doesn't matter.

13. Hadoop database step up
Starting Service is approaching the end. But this is also where my own laptop system crashed.
15. After all. It's done.

16. I don't think the bad health matters. All due to the 18G virtual disk is too small to fulfill the good
Some time you restart the bad health entries they would become good but at once they turn to bad again. So leave it.

17. Now I started to compile WordCount.java as Tutorial do and I found it doesn't work. It seems hadoop frameworks still missing (In fact Cloudera didn't offer any hadoop frame work, my opinion).
18. So I downloaded hadoop-1.1.2 from Apache Hadoop’s website. Untar it, and started configuration. I googled how to configure and found answer from this website:

19. After that we can compile WordCount.java (I missed taking the screen shot for the compile step.)

20. The first time to configure the hadoop should set environment variables as below:

   export HADOOP_HOME=/usr/lib/hadoop
   export JAVA_HOME=/usr/lib/jvm/j2sdk1.6-oracle
   export PATH=$PATH:$HADOOP_HOME/bin

21. Next is to make input directory under hadoop's distributed file system. DO not make output directory at first.

   As below picture. There is permission problems. After google, I found solution is to uncheck
the box as indicated in the zookeeper1. (If you don't want to start this service, just ignore these 2 steps)

21. Then restart this service

22. There is still safe mode protected problem. So we need exit safe mode using the command as shown in the second picture below.
23. Started to make input file and run
(Stopped some services for computer's performance)
24. Info and Results: