Ubuntu 16.04 hadoop2.7.6和hive2.3.0安装

hadoop 和hive安装包自行下载

<http://mirror.bit.edu.cn/apache/hadoop/common/>

<https://archive.apache.org/dist/hive/hive-2.3.0/>

解压到usr/local下并更名为hadoop 和hive

**tar –zxvf 安装包目录 /usr/local**

**cd /sr/local**

**mv 解压后文件夹名 hadoop或者hive**

安装jdk1.8 配置环境变量。

**sudo apt install openjdk-8-jre-headless**

**sudo apt install openjdk-8-jdk-headless**

**sudo vim /etc/profile**

#hadoop hive

export HADOOP\_HOME=/usr/local/hadoop

export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64

export HIVE\_HOME=/usr/local/hive

export PATH=.:${JAVA\_HOME}/bin:${HADOOP\_HOME}/bin:${HADOOP\_HOME}/sbin:${HIVE\_HOME}/bin:$PATH

**source /etc/profile**

设置Hadoop的JAVA\_HOME

**cd hadoop**

**vi etc/hadoop/hadoop-env.sh**

export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64

ssh免密登录，可能出现的问题

<https://blog.csdn.net/wenyun_kang/article/details/77413714>

<https://blog.csdn.net/someby/article/details/85329063>

**1、hadoop伪分布式安装**

这种模式是，1个节点上运行，HDFS daemon的 NameNode 和 DataNode、YARN daemon的 ResourceManger 和 NodeManager，分别启动单独的java进程**。**

# **vi etc/hadoop/core-site.xml**

<configuration>

<property>

             <name>hadoop.tmp.dir</name>

             <value>file:/usr/local/hadoop/tmp</value>

             <description>Abase for other temporary directories.</description>

        </property>

    <property>

        <name>fs.defaultFS</name>

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

    </property>

</configuration>

# **vi etc/hadoop/hdfs-site.xml**

<configuration>

     <property>

         <name>dfs.replication</name>

         <value>1</value>

</property>

        <property>

             <name>dfs.namenode.name.dir</name>

             <value>file:/usr/local/hadoop/tmp/dfs/name</value>

        </property>

        <property>

             <name>dfs.datanode.data.dir</name>

             <value>file:/usr/local/hadoop/tmp/dfs/data</value>

        </property>

</configuration>

建立配置文件中存储数据的文件夹

**mkdir tmp**

**mkdir tmp/dfs**

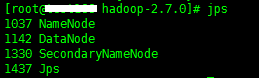
**mkdir tmp/dfs/name**

**mkdir tmp/dfs/data**

格式化文件系统 **hdfs namenode -format**

启动namenode和datanode： **sbin/start-dfs.sh**

**jps**



#### 访问NameNode的web页面

http://localhost:50070/

创建HDFS

**hdfs dfs -mkdir /user**

**hdfs dfs -mkdir /user/test**

拷贝input文件到HDFS目录下

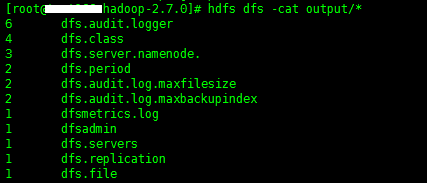
**hdfs dfs -put etc/hadoop /user/test/input**

查看 **hadoop fs -ls /user/test/input**

执行mapreduce实例

|  |
| --- |
| **hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.6.jar grep /user/test/input output 'dfs[a-z.]+'** |

查看结果 **hdfs dfs -cat output/\***



停止服务 sbin/stop-dfs.sh

修改mapreduce配置文件

**cp etc/hadoop/mapred-site.xml.template etc/hadoop/mapred-site.xml**

**vi etc/hadoop/mapred-site.xml**

<configuration>

    <property>

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

        <value>yarn</value>

    </property>

</configuration>

**vi etc/hadoop/yarn-site.xml**

<configuration>

    <property>

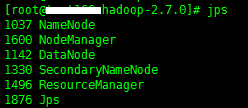
        <name>yarn.nodemanager.aux-services</name>

        <value>mapreduce\_shuffle</value>

    </property>

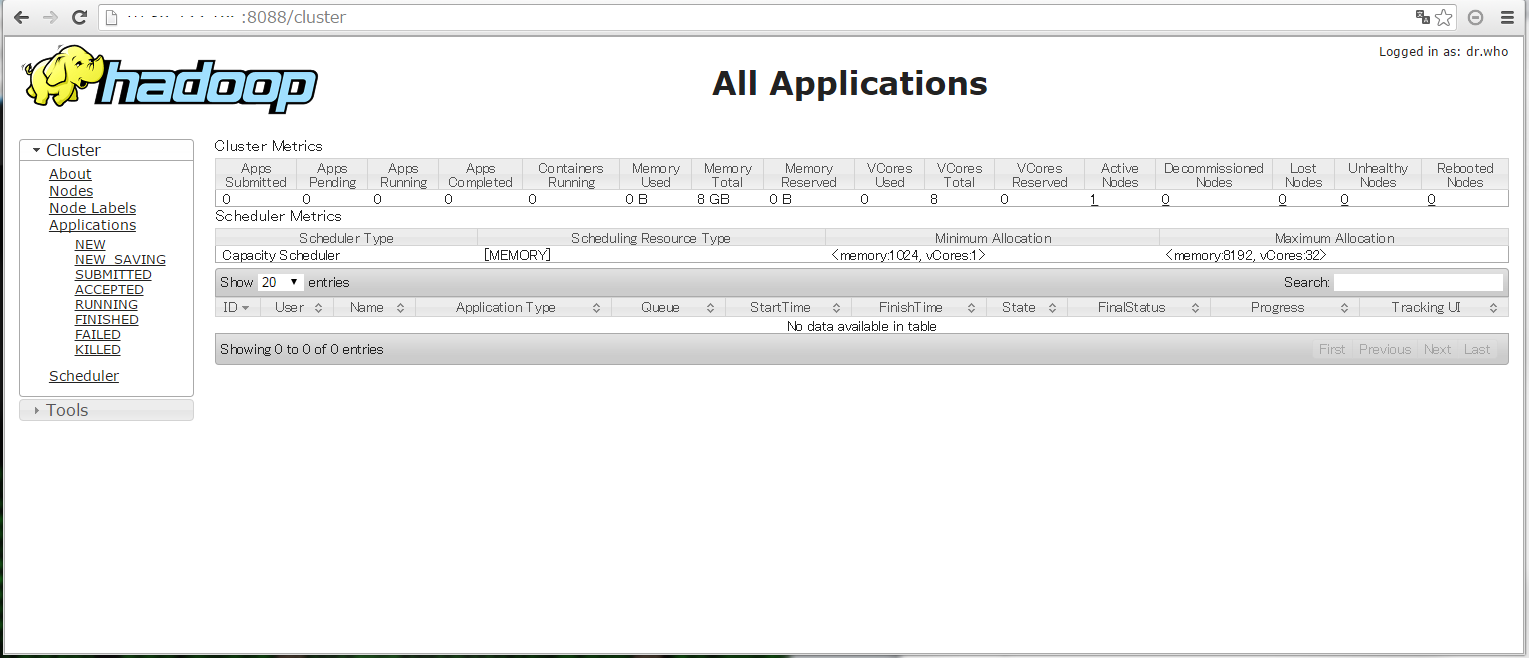
</configuration>

**sbin/start-all.sh**



访问ResourceManger的web页面

http://localhost:8088/



2. mysql安装

**sudo apt-get install mysql-server** 会自动安装client等其他组件

**mysql -uroot -p 密码** 登录验证

3、 hive安装

**cd /usr/local/hive/conf**

**cp hive-env.sh.template hive-env.sh**

**cp hive-default.xml.template hive-site.xml**

**vim** **hive-env.sh**

export HADOOP\_HOME=/usr/local/hadoop

export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-amd64

export HIVE\_HOME=/usr/local/hive

在hdfs中创建目录，并授权，用于存储文件

**hdfs dfs -mkdir -p /user/hive/warehouse**

**hdfs dfs -mkdir -p /user/hive/tmp**

**hdfs dfs -mkdir -p /user/hive/log**

**hdfs dfs -chmod -R 777 /user**

添加mysql jdbc包 <https://dev.mysql.com/downloads/connector/j/> 选择5.1.47版本

解压后将文件夹里的jar 包复制到 hive/lib下

修改hive-site.xml (hive/conf下若没有则新建 **vi hive-site.xml**)

<configuration>

<property>

<name>hive.metastore.warehouse.dir</name>

<value>/user/hive/warehouse</value>

</property>

<property>

<name>hive.metastore.local</name>

<value>true</value>

</property>

<property>

<name>javax.jdo.option.ConnectionURL</name>

<value>jdbc:mysql://localhost:3306/hive?createDatabaseIfNotExist=true&amp;useSSL=false</value>

</property>

<property>

<name>javax.jdo.option.ConnectionDriverName</name>

<value>com.mysql.jdbc.Driver</value>

</property>

<property>

<name>javax.jdo.option.ConnectionUserName</name>

<value>root</value>

</property>

<property>

<name>javax.jdo.option.ConnectionPassword</name>

<value>123</value>

</property>

<!-- 显示表的列名 -->

<property>

<name>hive.cli.print.header</name>

<value>true</value>

</property>

<!-- 显示数据库名称 -->

<property>

<name>hive.cli.print.current.db</name>

<value>true</value>

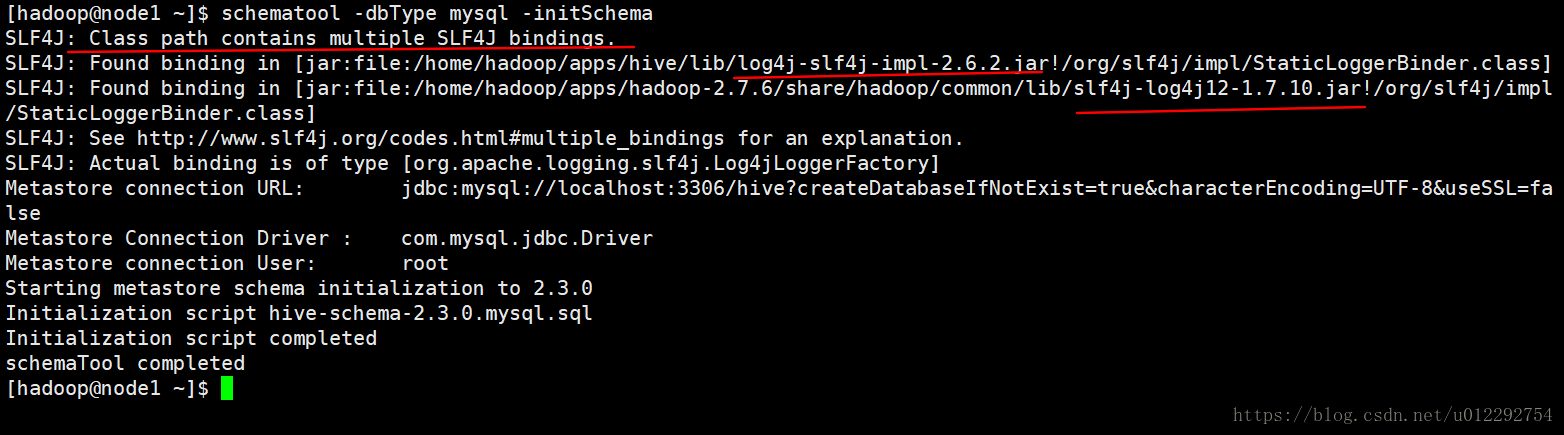
</property>

</configuration>

初始化hive：

**schematool -dbType mysql -initSchema**

如果包冲突，删除/hive/lib下的划线jar包 删除hive数据库，重新执行



**hive** 启动

show databases;