1.Hive安装 www.linuxidc.com

1. 需要环境：

   Jdk 1.6

   Hadoop 0.20.X

2. 安装步骤：

tar zxvf hive-0.8.1-bin.tar.gz
sudo mkdir /home/hive cd /home/hive
sudo mv /mnt/hgfs/sharedir/hive-0.8.1-bin ./

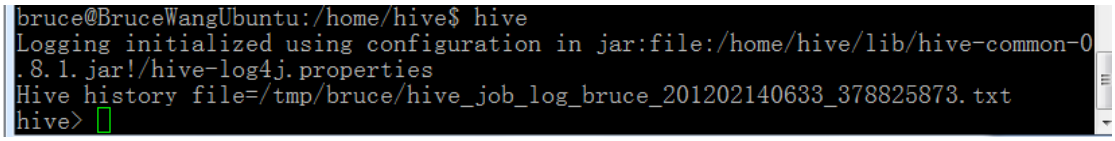
3. 配置环境变量

   执行：vi ~/.bashrc，在结尾处添加：

export HIVE_HOME=hive 的安装目录 export PATH=${HIVE_HOME}/bin:${PATH}

   应用设置执行：source ~/.bashrc

   执行 hive 进入 hive shell 界面，如下图所示：



2.测试

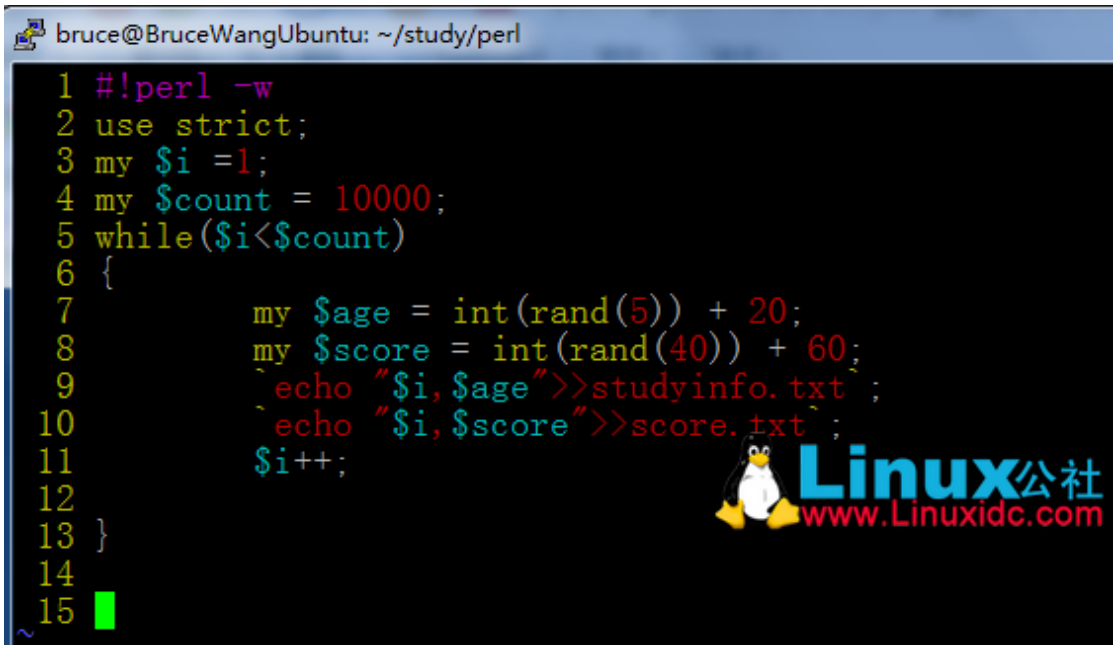
Linux公社（LinuxIDC.com）于 2006 年 9 月 25 日注册并开通网站，Linux现在已经成为一种广受关注和支持的一种操作系统，IDC是互联网数据中心，LinuxIDC就是关于Linux的数据中心。

LinuxIDC.com提供包括Ubuntu，Fedora，SUSE技术，以及最新IT资讯等Linux专业类网站。

1. 创建表 studyinfo 和 score，如下所示：

```
hive> create table studyinfo(id int,age int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 19.029 seconds
hive>
    > create table scores(id int, score int)
    > row format delimited
    > fields terminated by ','
    > stored as textfile;
OK
Time taken: 0.108 seconds
```

2. 用 perl 创建测试数据，如下所示：



3. 将数据导入到 hive 中

```
hive> load  data local inpath '/home/bruce/study/perl/score.txt' overwrite into
table scores;
Copying data from file:/home/bruce/study/perl/score.txt
Copying file: file:/home/bruce/study/perl/score.txt
Loading data to table default.scores
Deleted hdfs://BruceWangUbuntu:9000/user/hive/warehouse/scores
OK
Time taken: 2.447 seconds
hive> load  data local inpath '/home/bruce/study/perl/studyinfo.txt' overwrite i
nto table studyinfo;
Copying data from file:/home/bruce/study/perl/studyinfo.txt
Copying file: file:/home/bruce/study/perl/studyinfo.txt
Loading data to table default.studyinfo
Deleted hdfs://BruceWangUbuntu:9000/user/hive/warehouse/studyinfo
OK
Time taken: 0.472 seconds
```

4. 查看文件存储

   在hdfs空间中查看 www.linuxidc.com

```
bruce@BruceWangUbuntu:/home/hadoop$ bin/hadoop dfs  -lsr   /user/
drwxr-xr-x   - bruce supergroup          0 2012-02-11 09:17 /user/bruce
drwxr-xr-x   - bruce supergroup          0 2012-02-11 09:17 /user/bruce/input
-rw-r--r--   3 bruce supergroup         12 2012-02-11 09:17 /user/bruce/input/Name.txt
drwxr-xr-x   - bruce supergroup          0 2012-02-11 09:08 /user/bruce/input22
drwxr-xr-x   - bruce supergroup          0 2012-02-14 07:00 /user/hive
drwxr-xr-x   - bruce supergroup          0 2012-02-14 07:00 /user/hive/warehouse
drwxr-xr-x   - bruce supergroup          0 2012-02-14 07:00 /user/hive/warehouse/scores
-rw-r--r--   1 bruce supergroup      78885 2012-02-14 07:00 /user/hive/warehouse/scores/score.txt
drwxr-xr-x   - bruce supergroup          0 2012-02-14 07:00 /user/hive/warehouse/studyinfo
-rw-r--r--   1 bruce supergroup      78885 2012-02-14 07:00 /user/hive/warehouse/studyinfo/studyinfo.txt
```

5. 查询

出错“FAILED: Hive Internal Error: java.lang.RuntimeException(Error while making MR scratch directory - check filesystem config (null))”

解决方法：

把 hadoop 中的 core-site 中的 ip 换成 master 的 hostname 即可，保证每台机器的/etc/hosts 一样，其他配置文件中可以用 ip，重启 hadoop。

`hadoop dfsadmin -safemode leave`

```
hive> select count(*) from  studyinfo;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201202140733_0001, Tracking URL = http://BruceWangUbuntu:5003
0/jobdetails.jsp?jobid=job_201202140733_0001
Kill Command = /home/hadoop/bin/../bin/hadoop job  -Dmapred.job.tracker=BruceWan
gUbuntu:9001 -kill job_201202140733_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2012-02-14 07:43:26,652 Stage-1 map = 0%,  reduce = 0%
2012-02-14 07:43:38,599 Stage-1 map = 100%,  reduce = 0%
2012-02-14 07:43:51,727 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_201202140733_0001
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   HDFS Read: 78885 HDFS Write: 5 SUCESS
Total MapReduce CPU Time Spent: 0 msec
OK
9999
Time taken: 40.508 seconds
```



```
hive> select *from  studyinfo where id<4;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201202140733_0002, Tracking URL = http://BruceWangUbuntu:5003
0/jobdetails.jsp?jobid=job_201202140733_0002
Kill Command = /home/hadoop/bin/../bin/hadoop job  -Dmapred.job.tracker=BruceWan
gUbuntu:9001 -kill job_201202140733_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2012-02-14 07:44:45,820 Stage-1 map = 0%,  reduce = 0%
2012-02-14 07:44:48,851 Stage-1 map = 100%,  reduce = 0%
2012-02-14 07:44:51,876 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_201202140733_0002
MapReduce Jobs Launched:
Job 0: Map: 1   HDFS Read: 78885 HDFS Write: 15 SUCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1       21
2       21
3       22
Time taken: 14.309 seconds
hive>
```