about云分析discuz论坛apache日志hadoop大数据项目：hive与hbase是如何整合使用的

此篇需要具有hive、hbase整合基础，可以参考：
hive与hbase整合原理介绍
hbase0.96与hive0.12整合高可靠文档及问题总结

这里说一下项目实现的思路，此篇属于项目需求文档：数据的清洗以及导入hbase、hive准备工作

一、准备工作

首先我们启动hadoop、hbase、hive
如下图所示：

1.启动hadoop

```
1. start-dfs.sh
```



```
1. start-yarn.sh
```

```
aboutyun@master:~$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/hadoop/logs/yarn-aboutyun-resour
ger-master.out
slave1: starting nodemanager, logging to /usr/hadoop/logs/yarn-aboutyun-no
ger-slave1.out
slave2: starting nodemanager, logging to /usr/hadoop/logs/yarn-aboutyun-no
ger-slave2.out
```

2.启动hbase

```
1.  start-hbase.sh
```

```
aboutyun@master:~$ start-hbase.sh
slave2: starting zookeeper, logging to /usr/hbase/bin/../logs/hbase-aboutyun-
keeper-slave2.out
master: starting zookeeper, logging to /usr/hbase/bin/../logs/hbase-aboutyun-
keeper-master.out
slave1: starting zookeeper, logging to /usr/hbase/bin/../logs/hbase-aboutyun-
keeper-slave1.out
starting master, logging to /usr/hbase/bin/../logs/hbase-aboutyun-master-mast
out
slave2: starting regionserver, logging to /usr/hbase/bin/../logs/hbase-abouty
regionserver-slave2.out
slave1: starting regionserver, logging to /usr/hbase/bin/../logs/hbase-abouty
regionserver-slave1.out
```

```
1.  hbase shell
```

```
aboutyun@master:~$ hbase shell
14/08/01 23:07:59 INFO Configuration.deprecation: hadoop.native.lib is dep
d. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.96.0-hadoop2, r1531434, Fri Oct 11 15:28:08 PDT 2013

hbase(main):001:0>
```

3.启动hive

启动服务端

```
1.  hive --service metastore
```


启动客户端

```
1.  hive
```


🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒

上面准备工作完毕，下面我们开始创建表，其实也比较简单。

我们这个项目主要是对日志进行清洗，并且导入hbase、与hive中，为了避免重复导入，因此我们采用了hbase与hive整合，这样在导入hbase的同时，hive同时也导入了，避免了重复工作。那么这里的建表工作就显的比较重要。

CREATE TABLE
LogTable(Key string,IpAddress string,Url string,UserBrowser string,OSstring,AccressTime string) STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES ("hbase.columns.mapping" =
":key,Info:IpAddress,Info:Url,Info:UserBrowser,Info:OS,Info:AccressTime")
TBLPROPERTIES ("hbase.table.name" = "LogTable");

上面分别创建了hive以及hbase对应的表LogTable
1.查看hive表

```
1.  show tables;
```

我们看到创建hive表

```
hive> CREATE TABLE LogTable(Key string,IpAddress string,Url string,UserBrow
tring,OS string,AccessTime string) STORED BY 'org.apache.hadoop.hive.hbase
eStorageHandler' WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,Info
dress,Info:Url,Info:UserBrowser,Info:OS,Info:AccessTime") TBLPROPERTIES ("
.table.name" = "LogTable");                                    创建表
OK
Time taken: 6.654 seconds
hive> show tables;
OK                                                             创建成功
logtable
Time taken: 0.189 seconds, Fetched: 1 row(s)
hive>
```

2.查看hbase表

```
1.  list
```

```
hbase(main):008:0> list
TABLE
LogTable
1 row(s) in 0.1240 seconds

=> ["LogTable"]
```

🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒

上图我们已经我们就实现了在插入hbase的同时hive同样能够查询得到。

这里我们小小的展示一下：

```
1.  hbase(main):010:0> put 'LogTable','192.168.1.10

    http://www.aboutyun.com/thread-8472-1-1.html

    123','Info:IpAddress','192.168.1.10'

2.  0 row(s) in 0.6850 seconds

3.
```

🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒

验证结果：

hbase验证

ROW COLUMN+CELL
192.168.1.10 http:// column=Info:IpAddress, timestamp=1406908690106, value=192.
www.aboutyun.com/thr 168.1.10
ead-8472-1-1.html 12
3
1 row(s) in 0.1050 seconds

上图红字部分为RowKey

hive验证：

hive> select * from logtable;
OK

192.168.1.10 http://www.aboutyun.com/thread-8472-1-1.html 123 192.168.1.10
NULL   NULL   NULL   NULL
Time taken: 3.024 seconds, Fetched: 1 row(s)

这个成功之后，我们剩下的工作就是清洗插入，然后hive只写简单的sql就能统计了

本文链接：
http://www.aboutyun.com/thread-8597-1-1.html

🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒 🍒

对于云技术、大数据爱好者、可以关注：about云腾讯认证空间，每天更新最新内容，经典文章
http://user.qzone.qq.com/278595437/main

about云官方群**39327136、371358502、322273151**

云计算爱好者群

惯用手机可以关注微信：

云资源、云技术、疑问解答邮件订阅地址：
http://list.qq.com/cgi-bin/qf_invite?id=5e272bf95516e801681caf0fda5304dcadda1f67b4bbe4d8

淘宝云技术入门、hadoop、openstack及其它视频：

http://aboutyun.taobao.com/?v=1