



趣游

Gamewave Interactive
Technology Co., Ltd.

Hive与HBase整合经验谈

史东杰@趣游

QQ: 85573879

Hive与HBase整合经验谈

1

Hive与hbase整合的原理

2

Hive与hbase整合后的使用场景

3

Hive与hbase整合后的使用方法

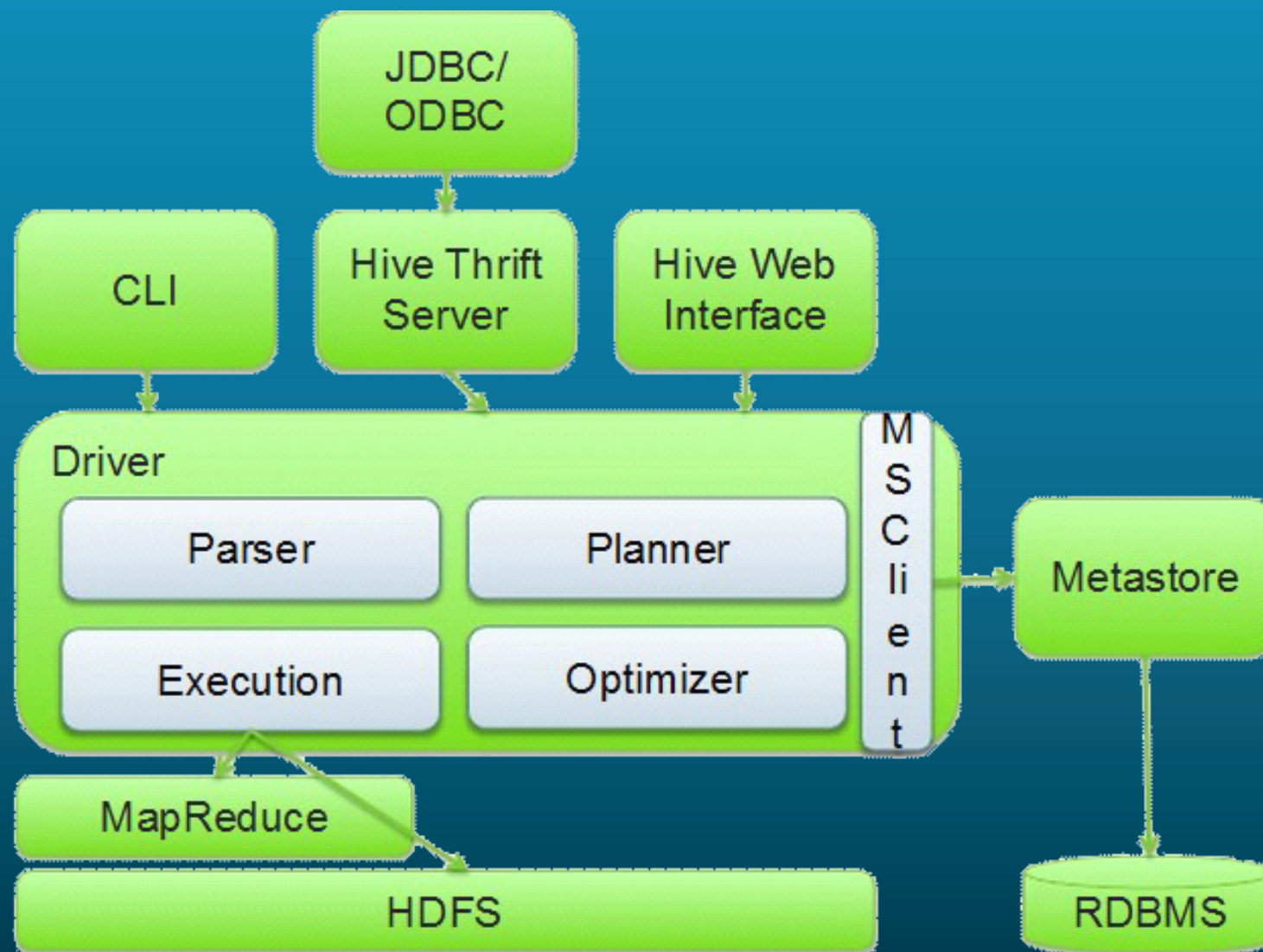
4

应用中的一些经验分享

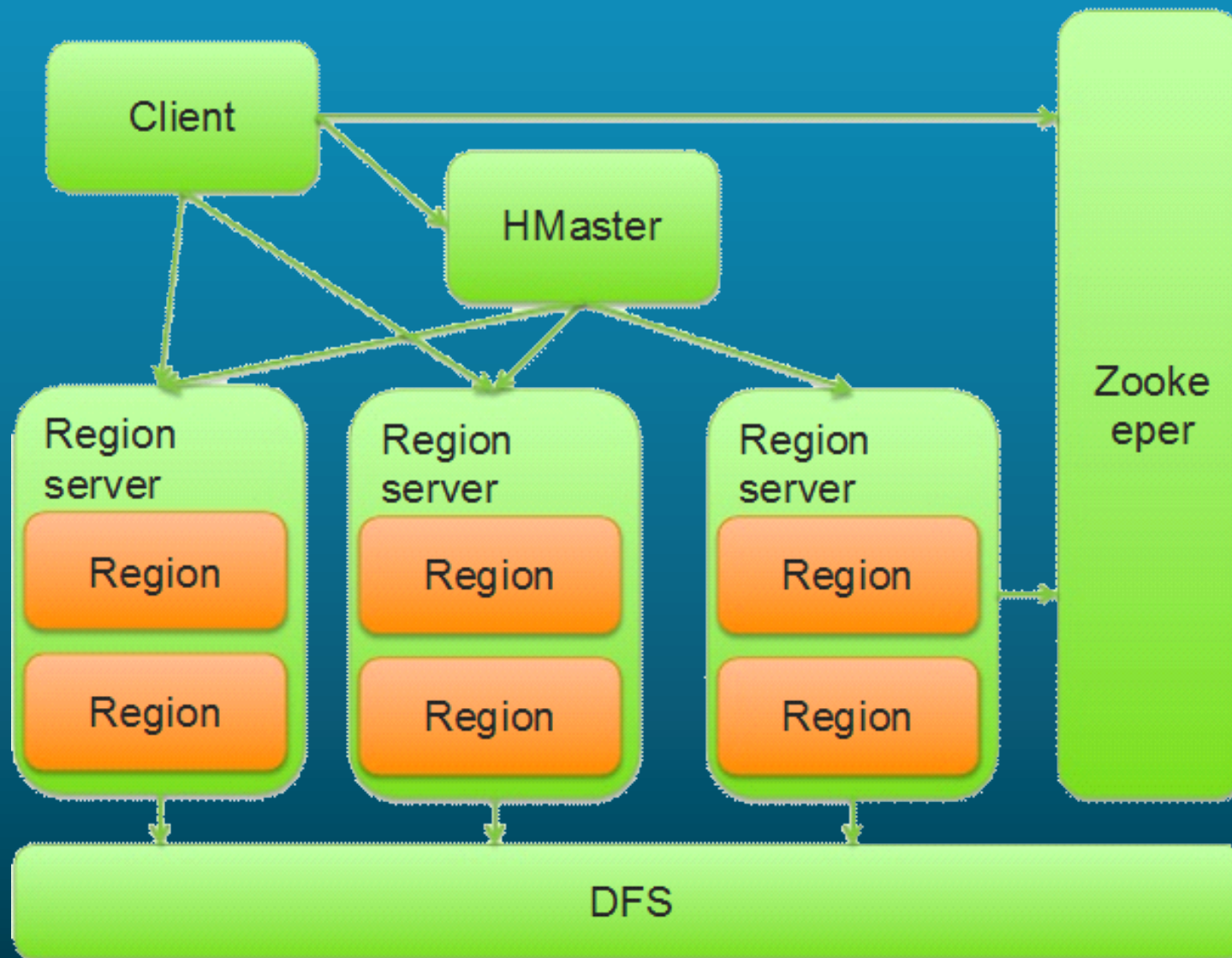
Hive与hbase整合的原理

- ❖ Hive与hbase的结构
- ❖ Hive与hbase整合后的结构图
- ❖ Hive与hbase整合的原理

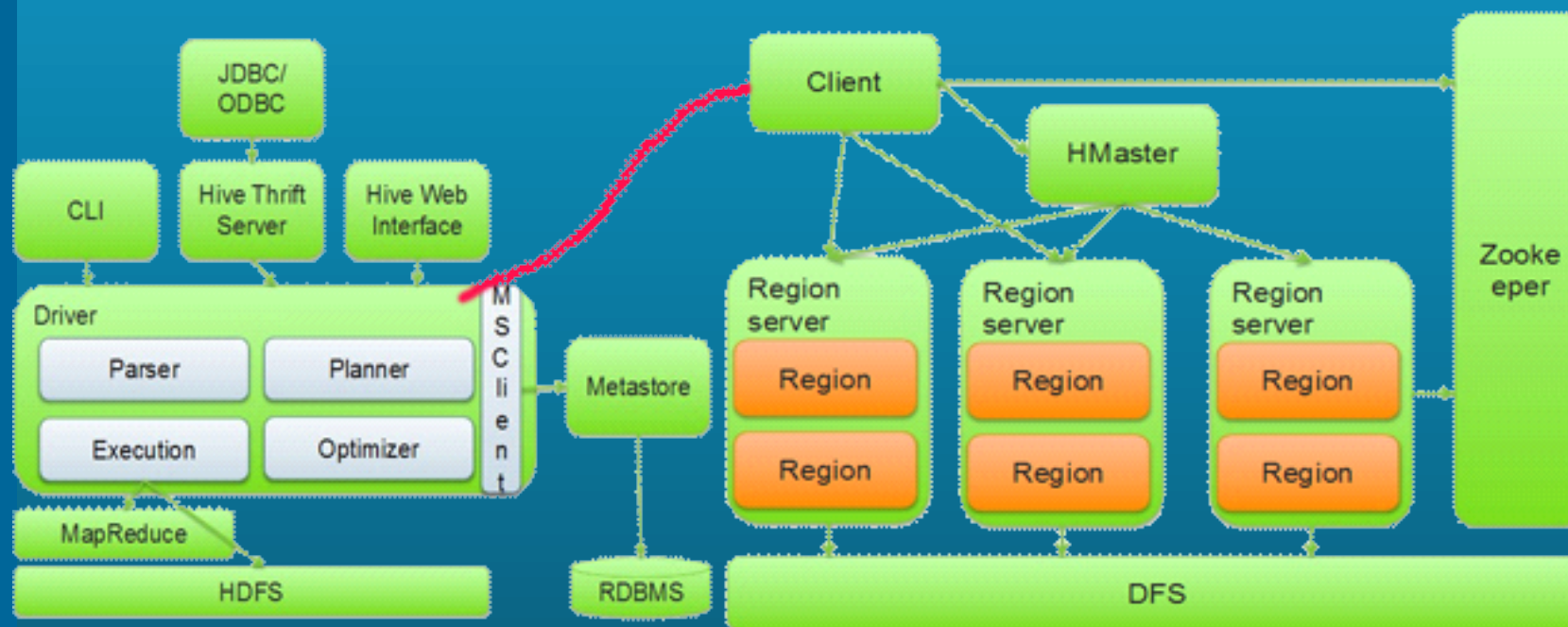
Hive的结构



hbase的结构



Hive与hbase整合后的结构图



Hive与hbase整合的原理

Hive与HBase的整合功能的实现是利用两者本身对外的API接口互相进行通信，相互通信主要是依靠hive-hbase-handler.jar工具类 (Hive Storage Handlers)

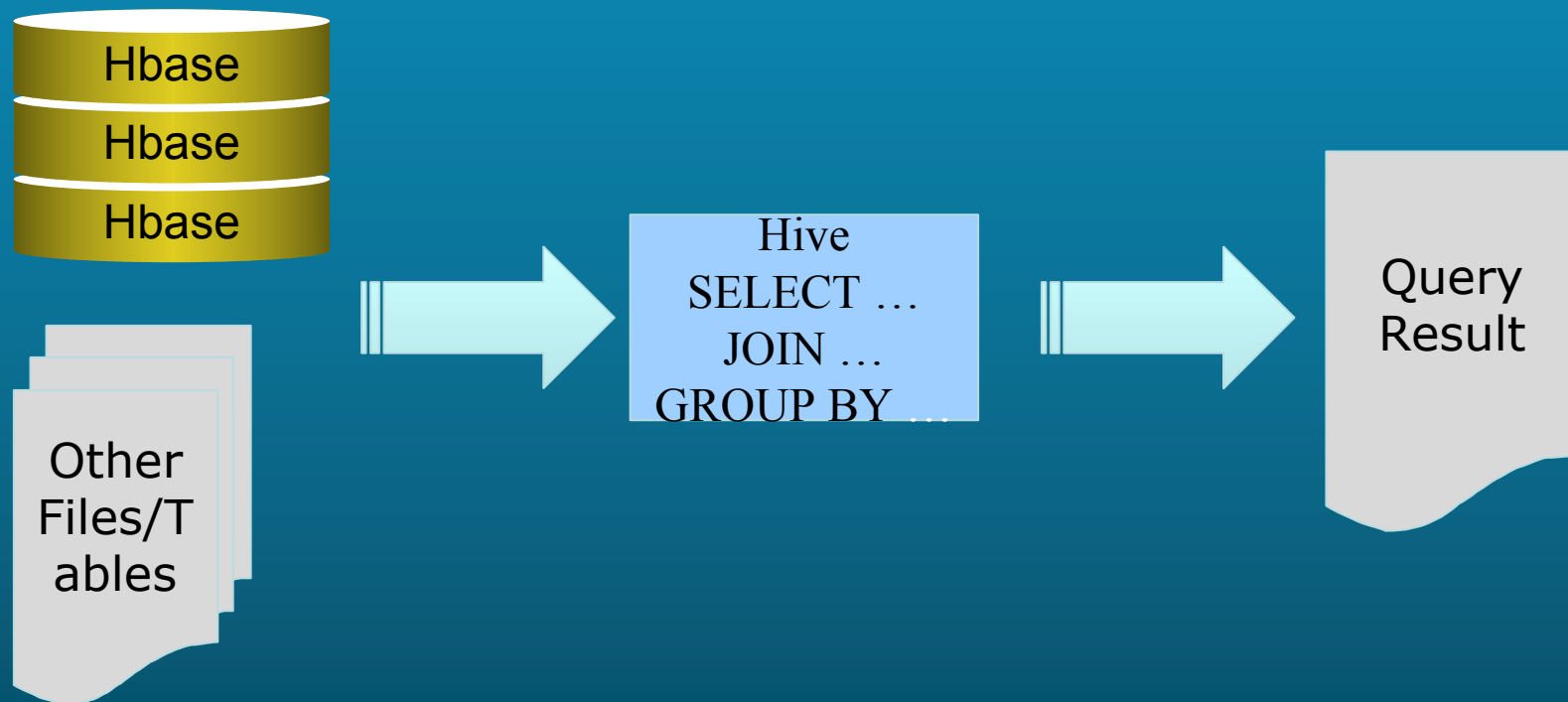
Hive与hbase整合后的使用场景

场景一：



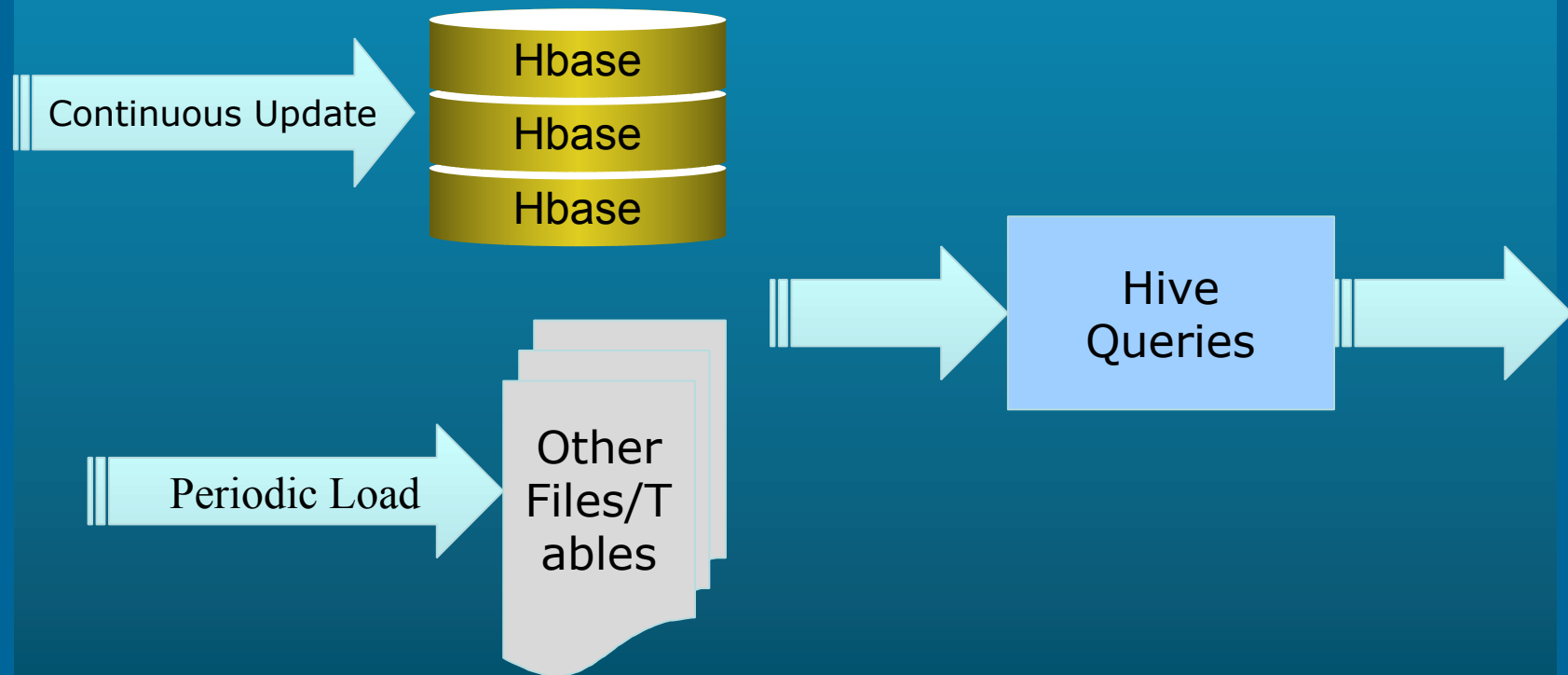
Hive与hbase整合后的使用场景

场景二:



Hive与hbase整合后的使用场景

场景三：



Hive与hbase整合后的使用方法

整合后Hive的启动方法

单节点启动 `hive -hiveconf`

`hbase.master=master:60000`

集群启动

`hive -hiveconf`

`hbase.zookeeper.quorum=node1,node2,node3`

Hive与hbase整合后的使用方法

在Hive中创建Hbase识别的表方法

```
CREATE TABLE hbase_table (key int, value string) STORED BY  
'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH  
SERDEPROPERTIES ("hbase.columns.mapping" = ":key,cf1:val")  
TBLPROPERTIES ("hbase.table.name" = "ceshi");
```

hbase.table.name 定义在hbase的table名称

hbase.columns.mapping 定义在hbase的列族，里面的:key 是固定值而且是唯一值

创建有分区的表

```
CREATE TABLE hbase_table(key int, value string) partitioned by (day  
string) STORED BY  
'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH  
SERDEPROPERTIES ("hbase.columns.mapping" = ":key,cf1:val")  
TBLPROPERTIES ("hbase.table.name" = "ceshi");
```

Hive与hbase整合后的使用方法

因为在hive中创建的和Hbase整合的表不支持load data导入数据需要在Hive中创建新表

```
create table pokes (foo int,bar string)row format delimited fields terminated by ',';
```

导入数据

```
load data local inpath '/home/1.txt' overwrite into table pokes;
```

将数据导入到hbase_table表中

```
insert overwrite table hbase_table select * from pokes;
```

然后可以在hive中查看到导入的数据

进入Hbase 查看表结构

```
# hbase shell
```

```
hbase(main):005:0>describe `ceshi`
```

查看Hbase表中数据

```
hbase(main):005:0>scan `ceshi`
```

Hive与hbase整合后的使用方法

对于Hbase中已经存在的表使用方法

```
CREATE EXTERNAL TABLE hbase_table(key int, info  
map<string,string>) STORED BY  
'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH  
SERDEPROPERTIES ("hbase.columns.mapping" = "cf1:val")  
TBLPROPERTIES("hbase.table.name" = "ceshi");
```

创建好表后，可以在hive中查看到hbase表中的数据

```
Hive>select * from hbase_table;  
1 {"val":"qqq"} {"val":"aaa"} {"val":"bbb"}  
2 {"val":"qqq"} {} {"val":"bbb"}
```

获取val中的数值方法

```
Select info[val] from hbase_table;
```

应用中的一些经验分享

在应用中需要使用hive thrift server

没与hbase整合前 启动方法为hive --service hiveserver

整合后如果使用启动命令

```
hive --service hiveserver -hiveconf
```

```
hbase.zookeeper.quorum=node0,node1,node2
```

会返回错误信息

```
java.lang.NumberFormatException: For input string: "-hiveconf"
```

应用中的一些经验分享

解决方法一:

Hive --service hiveserver启动后

```
Hive>set hbase.zookeeper.quorum=node0,node1,node2;
```

解决方法二:

修改hive-site.xml文件配置

```
<property>
```

```
  <name>hbase.zookeeper.quorum</name>
```

```
  <value>node1,node2,node3</value>
```

```
</property>
```

然后启动Hive --service hiveserver

应用中的一些经验分享

在hive中创建有分区的表与hbase整合后，使用
`Select * from hbase_test;` 命令无法看到表中的记录，使用
`Select id,name from hbase_test;` 可以看到记录。

分析原因：

`select *` 在普通表中是直接读取HDFS文件。在这里是直接从HBase进行查询。应该是查询成功了，但是hive并不会展示。
其它情况需要运行mapreduce，就没有问题了。

整合的表不支持修改

FAILED: Error in metadata: Cannot use ALTER TABLE on a non-native table

应用中的一些经验分享

在使用Hive 进行多个列进行group by 的查询分享

Hive中的字段为id,,name,time,num

time格式为yyMMddHHmmss格式，需求为按照每小时进行统计记录求num总数,并且返回结果中显示time的完整格式。

```
Select id,name,collect_set(time)[0],sum(num) from ceshi group by  
id,name,substr(time,9,2) ;
```



趣游

Gamewave Interactive
Technology Co., Ltd.

Thank You!