四种方法把mongDB迁移数据到hive或Hbase

方法一：用mongoexport 导出csv,然后hive关联csv

```
/bin/mongoexport -h IP --port 27017 -d mobp2p -c ncontacts -f
id,name,addtime,querytime -q '{addtime:{$gt: '$cts'}}' --csv -o
/data/ncontacts.csv

然后hive关联csv文件
create table if not exists test1.ncontacts (
id string,
name string,
addtime int
) row format delimited fields terminated by ',';
load data local inpath '/mobankerdata1/output.csv' overwrite into table
test1.ncontacts ;
```

方法二：用jar包直接关联mongdb。

```
add jar /home/yz/mongo-java-driver-3.4.1.jar;
add jar /home/yz/mongo-hadoop-core-2.0.1.jar;
add jar /home/yz/mongo-hadoop-hive-2.0.1.jar;

CREATE TEMPORARY TABLE IF NOT EXISTS mcookie_tmp
(
   id STRING,
   userId STRING,
   addtime STRING,
   userHead STRING,
   userCookie STRING,
```

```
   ip STRING,
   type STRING,
   addProduct STRING,
   addChannel STRING
)
STORED BY 'com.mongodb.hadoop.hive.MongoStorageHandler'
WITH SERDEPROPERTIES('mongo.columns.mapping'='{
"id":"_id",
"userId":"userId",
"addtime":"addtime",
"userHead":"userHead",
"userCookie":"userCookie",
"ip":"ip",
"type":"type",
"addProduct":"addProduct",
"addChannel":"addChannel"
}')
TBLPROPERTIES('mongo.uri'='mongodb://10.139.54.69:27017/mobp2p.mcookie');
```

方法三：用bson文件关联 先dump mongdb的数据

```
--hdfs dfs -mkdir /user/hive/warehouse/mobp2p/mobile
--hdfs dfs -put /data/mongodata/mobile.bson
/user/hive/warehouse/mobp2p/mobile
--hive
--添加jar包
add jar /home/yz/mongo-java-driver-3.4.1.jar;
add jar /home/yz/mongo-hadoop-core-2.0.1.jar;
add jar /home/yz/mongo-hadoop-hive-2.0.1.jar;
use mobp2p;
--创建临时表，关联到BSON文件：（1）一定要用临时表关联，否则删除表会一起删除
MongoDB中的数据；（2）注意表字段关联
create external table IF NOT EXISTS mobile_bson
(
```

```
   id STRING,
   addtime BIGINT,
   user_id STRING,
   type STRING,
   device_id STRING,
   phone_os STRING,
   phone_model STRING,
   phone_name STRING,
   phone_user_info STRING,
   phone_number STRING,
   photos STRING,
   ip STRING,
   apps STRING,
   appNames STRING,
   add_product STRING,
   idfa STRING,
   mac STRING,
   imei STRING,
   imsi STRING,
   networktype STRING,
   version STRING,
   qqNumbers STRING
)
ROW FORMAT SERDE "com.mongodb.hadoop.hive.BSONSerDe"
WITH SERDEPROPERTIES('mongo.columns.mapping'='{
"id":"_id",
"addtime":"addtime",
"user_id":"user_id",
"type":"type",
"device_id":"device_id",
"phone_os":"phone_os",
"phone_model":"phone_model",
"phone_name":"phone_name",
"phone_user_info":"phone_user_info",
"phone_number":"phone_number",
```

```
"photos":"photos",
"ip":"ip",
"apps":"apps",
"appNames":"appNames",
"add_product":"add_product",
"mac":"mac",
"imei":"imei",
"imsi":"imsi",
"networktype":"networktype",
"version":"version",
"qqNumbers":"qqNumbers"
}')
STORED AS INPUTFORMAT "com.mongodb.hadoop.mapred.BSONFileInputFormat"
OUTPUTFORMAT "com.mongodb.hadoop.hive.output.HiveBSONFileOutputFormat"
location '/user/hive/warehouse/mobp2p/mobile';
```

方法四：自定义分割符导出

由于mongDB mongoexport 不能指定分割符号 ，这种方法可以自定义分割符

```
vim myjsfile.js
cursor = db.apps.find({'id':{$gt:"0"}},{'_id':0, 'id':1,
'name':1});
while ( cursor.hasNext() ) {
var row = cursor.next();
print( row['id'] + "\t" + row['name'] );
}
mongo localhost:27017/test myjsfile.js >> output.csv

然后csv文件
create table if not exists test1.lbs_tmp (
id string,
name string
```

```
) row format delimited fields terminated by ',';
load data local inpath '/mobankerdata1/output.csv' overwrite into
table test1.lbs_tmp;
```