**HIVE查询原始日志要用到的命令和SQL**

注意：现网上有数据运行的时候轻易不要执行，执行缓慢会影响现网的日志运算。

1.登陆HTCDN-ZJHZ-NOC-CLOUD-01

2.输入命令hive -t -h localhost

[localhost:10000] hive> create external table if not exists huatong\_test  (sessionid STRING,  channelname STRING, reqtime DOUBLE,  method STRING,  url STRING,  statuscode SMALLINT,  responsetime Double,  headlen INT,  bodylen BIGINT,  ishit STRING,  iscache TINYINT,  userip STRING,  cdnid STRING,  devicetype STRING,  referer STRING,  backtime STRING, backtype STRING, backlen STRING)  partitioned by(loglocation STRING)  row format delimited fields terminated by '|'  stored as SEQUENCEFILE ;

3.往临时表中添加数据 注意loglocation不能重复

[localhost:10000] hive> alter table huatong\_test add partition(loglocation='1')     location 'hdfs://hdfs1/flume/test/huatong/picture/20140302/1/001';

[localhost:10000] hive> alter table huatong\_test add partition(loglocation='2')     location 'hdfs://hdfs1/flume/test/huatong/picture/20140302/2/001';

……

统计SQL：

1.统计ISP为其他的IP

select distinct userip ,ipsearch(userip)  from huatong\_aime  where  cast(reqtime as bigint)>=unix\_timestamp('2014-03-17 15:00:00')  and  cast(reqtime as bigint)<unix\_timestamp('2014-03-17 16:00:00')  and channelname ='img04.taobaocdn.com'  and  ipsearch(userip) like '%|0';

2.统计004日志中状态码为200，有重复url的

select  \*  from huatong\_wushan where  backtype ='04' and backlen > 0 and statuscode ='200' group by url having count(\*)>1 limit 10;

3.统计命中率是否合理

CREATE EXTERNAL TABLE huatong\_tmp(url STRING) ;

INSERT OVERWRITE TABLE huatong\_tmp  select url from huatong\_wushan  where  backtype ='04' and backlen > 0 and statuscode ='200' group by url having count(\*)>1;

insert overwrite local directory '/tmp/expired'  SELECT huatong\_wushan.sessionid ,  huatong\_wushan.channelname , huatong\_wushan.reqtime ,  huatong\_wushan.method ,  huatong\_wushan.url ,  huatong\_wushan.statuscode ,  huatong\_wushan.responsetime ,  huatong\_wushan.headlen ,  huatong\_wushan.bodylen ,  huatong\_wushan.ishit ,  huatong\_wushan.iscache ,  huatong\_wushan.userip ,  huatong\_wushan.cdnid ,  huatong\_wushan.devicetype ,  huatong\_wushan.referer ,  huatong\_wushan.backtime , huatong\_wushan.backtype , huatong\_wushan.backlen   FROM huatong\_wushan LEFT SEMI JOIN huatong\_tmp on (huatong\_wushan.url = huatong\_tmp.url);

3.查询状态为200，EXPIRED的日志

 insert overwrite local directory '/tmp/expired'  select \* from huatong\_wushan  where ishit='EXPIRED' and statuscode='200' ;

-- ４.统计探针

[localhost:10000] hive> select sessionid,channelname,reqtime,method,url,statuscode,responsetime,headlen,bodylen,ishit,iscache,userip,cdnid,devicetype,referer,backtime,backtype,backlen from huatong\_test where url ='/bao/uploaded/i4/T1dINAFBNXXXXXXXXX\_!!0-item\_pic.jpg\_250x250.jpg';