### 分区表

数据分区的概念以及存在很久了，通常使用分区来水平分散压力，将数据从物理上移到和使用最频繁的用户更近的地方，以及实现其目的。

hive中有分区表的概念，我们可以看到分区具重要性能优势，而且分区表还可以将数据以一种符合逻辑的方式进行组织，比如分层存储

分区表分别有静态分区和动态分区

### 静态分区

（1）创建分区表

hive> create table emp(name string,sal float,address string)

partitioned by (country string,mdate string)

row format delimited fields terminated by ","

stored as textfile;

（2）数据源

[hadoop@h201 hhh]$ cat emp.txt

zs,10000,bj

ls,20000,bj

ww,30000,bj

zl,40000,sh

qq,50000,bj

ee,60000,sh

（3）导入数据

hive> load data local inpath '/home/hadoop/hhh/emp.txt' into table emp partition (country ="china",mdate="20200210");

（4）查看数据

hive> select \* from emp;

（查看hdfs上的情况）

（5）查看分区

hive> show partitions emp;

1. 添加分区

hive> alter table emp add partition(country ="usa",mdate="20200209");

1. 删除分区

hive> alter table emp drop partition(country ="usa",mdate="20200209");

### 外部表静态分区

hive> create external table emp1(name string,sal float,address string)

partitioned by (country string,mdate string)

row format delimited fields terminated by ","

stored as textfile;

导入数据

hive> load data local inpath '/home/hadoop/hhh/emp.txt' into table emp1 partition (country ="china",mdate="20200211");

[hadoop@h201 hhh]$ hadoop fs -rmr /hive1/emp1

hive> select \* from emp1;

（查询时没有数据了）

hive> show partitions emp1;

（分区表信息还存在）

country=china/mdate=20200211

[hadoop@h201 hhh]$ hadoop fs -mkdir /h1

[hadoop@h201 hhh]$ cp emp.txt empp.txt

[hadoop@h201 hhh]$ hadoop fs -put empp.txt /h1

hive> alter table emp1 partition(country="china",mdate="20200211") set location 'hdfs://h201:9000/h1';

hive> select \* from emp1;

(数据又回来了)

### 动态分区

往hive分区表中插入数据时，如果需要创建的分区很多，比如以表中某个字段进行分区存储，则需要复制粘贴修改很多sql去执行，效率低。因为hive是批处理系统，所以hive提供了一个动态分区功能，其可以基于查询参数的位置去推断分区的名称，从而建立分区。

hive> create table t5(name string,sal bigint,city string)

partitioned by (ct string)

row format delimited fields terminated by ","

stored as textfile;

1. 配置参数

set hive.exec.dynamic.partition=true; //使用动态分区

set hive.exec.dynamic.partition.mode=nonstrick;//非严格模式，如果模式是strict，则必须有一个静态分区，且放在最前面

set hive.exec.max.dynamic.partitions.pernode=10000;//每个节点生成动态分区的最大个数

set hive.exec.max.dynamic.partitions=100000;//生成动态分区的最大个数

set hive.exec.max.created.files=150000;//一个任务最多可以创建的文件数目

set dfs.datanode.max.xcievers=8192;//限定一次最多打开的文件数

set hive.merge.mapfiles=true; //map端的结果进行合并

set mapred.reduce.tasks =20000; //设置reduce task个数

1. 创建临时表

hive> create table t55(name string,sal bigint,city string) row format delimited fields terminated by "," ;

hive> load data local inpath '/home/hadoop/hhh/emp.txt' into table t55;

1. 动态分区载入数据

hive> insert into table t5 partition(ct) select name,sal,city,city as ct from t55;

（city as ct 为分区条件，别名必须跟分区名相同）

1. 查看分区

hive> show partitions t5;

OK

ct=bj

ct=sh

也可以观察hdfs

[hadoop@h201 hhh]$ hadoop fs -lsr /hive1/t5