1. Hive数据库操作

创建数据库

create database if not exists itcast;

显示数据库

show databases;

查看数据库详情

desc database itcast;

切换数据库

use XXX

删除数据库等等等等

1. Hive内部表操作

创建表和修改表

针对基本类型建表：

在Hadoop001机器/export/data目录下创建hivedata目录，该文件夹下创建user.txt文件，并导入数据。

创建表

create table t\_user(

id int,

name string,

age int)

row format delimited fields terminated by ‘,’;

使用Linux终端将Linux文件上传到HDFS中的Hive表目录，实现hive表映射

hdfs dfs -put /xxx/04\_students.txt /user/t\_user

或者load

load data local inpath '/root/04\_students.txt' into table t\_user

创建表后，可以通过Web UI打开Hive内部表所在HDFS路径：

/user/hive/warehouse/itcast.db/t\_user

直接访问内部信息

select \* from t\_user;

创建文件student.txt

创建表

create table t\_student(

id int,

name string,

hobby map<string, string>

)

row format delimited fields terminated by ‘,’

collection items terminated by ‘-’

map keys terminated by ‘:’;

查询数据(什么都没有)

select \* from t\_students;

导入数据

load data local inpath '/path/file\_name' into table table\_name;

such as

load data local inpath '/hviedata/student.txt' into table t\_user;

再次查询数据

select \* from t\_students;

1. hive外部表操作

结构化文件students.txt

将文件上传到HDFS上的/stu路径下，模拟生产环境下的数据文件

hadoop fs -mkdir /stu

hadoop fs -put students.txt /stu

创建一张外部表

create external table students\_ext(

Sno int,

Sname string,

Sex string,

Sage int,

Sdept string

)

row format delimited fields terminated by ‘,’

location /stu;

查看itcast数据库中的数据表

show tables;

查看数据表内容

select \* from students\_ext;

注意：hive创建内部表会将数据移动到数据库指向的路径，而外部表则不会移动数据

1. hive分区表操作

Hive分区表一共有两种，分别是普通分区和动态分区

4.1普通分区

结构化数据文件user\_p.txt

创建分区表

create table t\_user\_p(

id int,

name string

)

partitioned by (country string)

row format delimited fields terminated by ‘,’;

加载数据是将数据文件移动到与Hive表对应的位置，从本地复制或移动到HDFS中的操作。由于分区表不能使用Hadoop命令移动文件，所以需要使用Load命令

load data local inpath ‘/hivedata/user\_p.txt’ into table t\_user\_p partition(Country=’USA’)

查看数据

select \* from t\_user\_p;

再次，新增分区

alter table t\_user\_p add partition (country=’China’) location ‘/user/hive/warehouse/itcast.db/t\_user\_p/country=China’;

通过HDFS的UI可以看到新增的分区country=China

接着，修改分区

alter table t\_user\_p partition (country=’China’) rename to partition (country=’Japan’)

最后，删除分区

alter table t\_user\_p drop if exists partition (country=’Japan’);

此时HDFS UI中已经看不到这个分区了

4.2动态分区

开启动态分区功能

set hive.exec.dynamic.partition=true;

set hive.exec.dynamic.partition.mode=nonstrict ;

现有原始化数据文件dynamic\_partition\_table.txt

现演示动态分区的数据插入操作——将dynamic\_partition\_table中的数据按照时间插入到目标表d\_p\_t相应分区中

首先创建原始数据表

create table dynamic\_partition\_table (

day string,

ip string

)

row format delimited fields terminated by ‘,’;

其次加载数据文件至原始表

load data local inpath ‘/export/data/hivedata/dynamic\_partition\_table.txt’ into table dynamic\_partition\_table;

再次，创建目标表

create table d\_p\_t(ip string)

partitioned by (month string, day string);

接着，动态插入

insert overwrite table d\_p\_t partition (month, day)

select ip, substr(day, 1, 7) as month, day

from dynamic\_partition\_table;

最后，查看目标表中的分区数据

show partitions d\_p\_t ;

1. hive桶表操作

开启分桶功能

set hive.enforce.bucketing=true;

set mapreduce.job.reduces=4;

创建桶表

以学生编号Sno分为4个桶

create table stu\_buck(

Sno int,

Sname string,

Sex string,

Sage int,

Sdept string

)

clustered by(Sno) into 4 buckets

row format delimited fields terminated by ‘,’;

将students.txt文件复制到/hivedata目录下

。。

加载数据到桶表中，但是不能用load了，桶表直接收表类，不能直接接受文件，所以在创建一个临时的students表

create table students\_tmp(

Sno int,

Sname string,

Sex string,

Sage int,

Sdept string

)

row format delimited fields terminated by ‘,’;

接着，加载数据至students\_tmp表

load data local inpath ‘/hivedata/students.txt’ into table students\_tmp;

最后，将数据导入stu\_back表

insert overwrite table stu\_buck

select \* from students\_tmp cluster by(Sno);

针对每个桶的文件，可以使用hadoop命令查看内容

Hadoop fs -cat /user/hive/warehouse/itcast.db/stu\_back/000000\_0