

Hive练习6 索引

编程要求

student.txt 中的数据格式如下：

```
1 001,Xiaohong,18,female,96-88-90.5
2 002,Xiaoliang,17,male,95-88-93.5
3 003,Xiaoming,19,male,86.5-98-91
4 004,Xiaoguang,18,male,88-80-94
5 005,Xiaohua,16,female,97-58.5-88
```

- 创建 test2 数据库；
- 根据以上数据创建 student 表；
- 将 /home/student.txt 中的数据导入到表 student 中；
- 根据学号 Sno 创建索引 student_index ；
- 删除索引 student_index。

代码

```
#***** Begin *****#
echo "
create database test2;
create table test2.student(
    Sno INT,
    name STRING,
    age INT,
    sex STRING,
    score STRUCT<Chinese:FLOAT,Math:FLOAT,English:FLOAT>
)
row format delimited fields terminated by ','
collection items terminated by '-';

load data local inpath '/home/student.txt'
overwrite into table student;

create index student_index on table student(Sno)
as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
with deferred rebuild
IN TABLE student_index_table;

DROP INDEX IF EXISTS student_index on student;
"
#***** End *****#
```

解析

导入本地数据到 hive 表中

- /home/shoppings.txt 目录下数据格式如下:

```
1 1,Nike Air Max 97 ND,1299,Nick,shoes,55,Guangzhou-Chinese-510000
2 2,Nike Air VaporMax Flyknit 3,1599,Nick,shoes,42,Fuzhou-Chinese-350000
3 3,W MS CLS TRJK jacket,599,Adidas,clothes,Hangzhou-Chinese-310000
4 4,W MS CLS TRPT pants,549,Adidas,clothes,Hangzhou-Chinese-310000
5 5,ADIDAS NEO PARKHOOD BP package,369,Adidas,accessories,Guangzhou-Chinese-510000
```

- 在数据库 shopping 中根据数据分隔方式创建表 items_info:

```
CREATE TABLE IF NOT EXISTS shopping.items_info(
  id INT COMMENT 'item id',
  name STRING COMMENT 'item name',
  price FLOAT COMMENT 'item price',
  category STRING COMMENT 'item category',
  brand STRING COMMENT 'item brand',
  stock INT COMMENT 'item stock',
  address STRUCT<city:STRING, country:STRING, zip:INT> COMMENT 'item sales address')
COMMENT 'goods information table'
row format delimited fields terminated by ',' //字段以','分隔
collection items terminated by '-' //集合以'-'分隔
TBLPROPERTIES ('creator'='Xiaoming','date'='2019-01-01');
```

- 进入到数据库 shopping 中: use shopping;
- 导入数据到表 items_info 中:

```
load data local inpath '/home/shoppings.txt' overwrite into table items_info;
```

- 查看导入的数据: select * from items_info;

Create 创建索引

创建索引的语法为:

```
CREATE INDEX index_name ON TABLE base_table_name (col_name,...)
AS index_type
[with DEFERRED REBUILD]
[INDEXPROPERTIES (property_name=property_value,...)]
[IN TABLE index_table_name]
[[ROW FORMAT ...] STORED AS ... | STORED BY]
[LOCATION hdfs_path] [TBLPROPERTIES (...)] [COMMENT "index comment"];
```

属性参数说明:

- with DEFERRED REBUILD: 用于构建一个空索引。

例子:

- 创建索引 items_index:

```
create index items_index on table items_info(id)
as 'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler'
with deferred rebuild
IN TABLE items_index_table;
```

Alter 修改索引

修改索引的语法为：

```
ALTER INDEX index_name ON table_name [PARTITION partition_spec] REBUILD;
```

`ALTER INDEX REBUILD` 用于重建之前创建索引时使用关键字 `WITH DEFERREDREBUILD` 建立的所有或之前建立的索引，若指定关键字 `PARTITION`，则只针对相应分区建立索引。

DROP 删除索引

删除索引的语法为：

```
DROP INDEX [IF EXISTS] index_name ON table_name;
```

- 删除索引 `items_index`：

```
drop index if exists items_index on items_info;
```