

Hive练习4 表分区

编程要求

请根据右侧命令行内的提示，在 `Begin - End` 区域内进行 `sql` 语句代码补充，具体任务如下：

`student` 表结构：

INFO	TYPE	COMMENT
Sno	INT	student sno
name	STRING	student name
age	INT	student age
sex	STRING	student sex
score	STRUCT	student score

- 创建数据库 `test4`
- 在数据库 `tets4` 中，创建分区表 `student`，表结构如上所示，和第二、三关相同，设置分区列为：`stu_year` 类型 `STRING`、`subject` 类型 `STRING`
- 添加两个分区：`stu_year='2018',subject='Chinese'` 和 `stu_year='2018',subject='Math'`
- 重命名表分区：将 `2018/Math` 分区重命名为 `2018/English`
- 删除表分区：将 `2018/Chinese` 分区删除

按照以上要求，在代码栏填写命令。每个要求对应一条命令，共 `5` 条命令，以 `;` 隔开。

由于 `hive` 启动时间较长，测评时请耐心等待，大概需要时间：`1-3` 分钟。

答案

```
#***** Begin *****#
echo "

CREATE DATABASE test4;
CREATE TABLE IF NOT EXISTS test4.student(
    Sno INT COMMENT 'student sno',
    name STRING COMMENT 'student name',
    age INT COMMENT 'student age',
    sex STRING COMMENT 'student sex',
    score STRUCT<Chinese:FLOAT, Math:FLOAT, English:FLOAT> COMMENT 'student
score'
)
PARTITIONED BY (stu_year STRING, subject STRING);
ALTER TABLE student ADD PARTITION (stu_year='2018', subject='Chinese')
LOCATION '/hive/test4/student/2018/Chinese'
PARTITION (stu_year='2018', subject='Math')
LOCATION '/hive/test4/student/2018/Math';
```

```
ALTER TABLE student PARTITION (stu_year='2018', subject='Math') RENAME TO
PARTITION (stu_year='2018', subject='English');
ALTER TABLE student DROP IF EXISTS PARTITION (stu_year='2018',
subject='Chinese');
"
#***** End *****#
```

解析

创建分区表

使用 `shopping` 数据库创建一张商品信息分区表 `items_info2`，按商品品牌 `p_brand` 和商品分类 `p_category` 进行分区：

```
CREATE TABLE IF NOT EXISTS shopping.items_info2(
name STRING COMMENT 'item name',
price FLOAT COMMENT 'item price',
category STRING COMMENT 'item category',
brand STRING COMMENT 'item brand',
type STRING COMMENT 'item type',
stock INT COMMENT 'item stock',
address STRUCT<street:STRING, city:STRING, state:STRING, zip:INT> COMMENT 'item
sales address')
COMMENT 'goods information table'
PARTITIONED BY (p_category STRING,p_brand STRING)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY '\t'
COLLECTION ITEMS TERMINATED BY ','
TBLPROPERTIES ('creator'='Xiaoming','date'='2019-01-01');
```

添加分区

向表 `items_info2` 添加两个分区：

```
ALTER TABLE items_info2 ADD PARTITION (p_category='clothes',p_brand='playboy')
LOCATION '/hive/shopping/items_info2/playboy/clothes'
PARTITION (p_category='shoes',p_brand='playboy') LOCATION
'/hive/shopping/items_info2/playboy/shoes';
```

重命名表分区的语法为：

```
ALTER TABLE table_name PARTITION partition_spec RENAME TO PARTITION
partition_spec;
```

1. 重命名 `items_info2` 表的表分区 `playboy/clothes` 为 `playboy/T-shirt`：

```
ALTER TABLE items_info2 PARTITION (p_category='clothes',p_brand='playboy')
RENAME TO PARTITION (p_category='T-shirt',p_brand='playboy');
```

交换表分区

交换表分区的语法为：

```
ALTER TABLE table_name_1 EXCHANGE PARTITION (partition_spec) WITH TABLE table_name_2;
```

该操作移动表 `table_name_1` 中特定分区下的数据到具有相同表模式且不存储在相应分区的 `table_name_2` 中。

表分区信息持久化

表分区信息持久化的语法为：

```
MSCK REPAIR TABLE table_name;
```

该操作作用于同步表 `table_name` 在 HDFS 上的分区信息到 Hive 位于 RDBMS 的 metastore 中。

删除表分区

删除表分区的语法为：

```
ALTER TABLE table_name DROP [IF EXISTS] PARTITION partition_spec, PARTITION partition_spec,...;
```

该操作会删除与特定分区相关的数据以及 metadata。

1. 删除表 `items_info2` 的 `playboy/shoes` 表分区：

```
ALTER TABLE items_info2 DROP IF EXISTS PARTITION (p_category='shoes',p_brand='playboy');
```