任务 4-3 HiveQL操作分区表

幻灯片43：任务4-3 HiveQL操作分区表（标题页）

好，现在我们来学习一个非常重要的内容——分区表。

分区表是Hive中用来优化查询性能（xìng néng）的一种技术。

幻灯片44-45：4.3.1 任务描述 —— 分区表概念

随着系统运行时间越来越长，表里面的数据量会越来越大。

Hive查询数据通常是全表扫描（quán biǎo sǎo miáo），就是把整个表从头到尾看一遍。

如果表很大，这样查询会非常慢。

为了解决这个问题，Hive引进了分区技术。

分区就是把一个大表的数据，按照某个字段（zì duàn）分成多个小部分存储。

每个小部分就是一个分区，对应一个子目录（zǐ mù lù）。

> 💡 *举个例子：比如你有一整年的销售数据，你可以按月份分成12个分区。查询3月份的数据时，Hive只扫描3月份的分区，不需要扫描全年数据，速度就快了很多！*

幻灯片46-47：分区表语法树和创建

创建分区表的关键语法是 PARTITIONED BY。

在CREATE TABLE的时候，用 PARTITIONED BY 指定分区字段。

注意：分区字段不能是表里已有的字段，不能重名（chóng míng）。

分区字段是一个虚拟字段（xū nǐ zì duàn），数据并不存在底层文件中。

幻灯片48：4.3.2 任务准备知识 —— 1. 查询分区表

查看分区信息的语法：

```
SHOW PARTITIONS [db_name.]table_name
[PARTITION(partition_column = partition_col_value, ...)];
```

这个命令可以查看分区表有哪些分区。

幻灯片49：4.3.2 —— 2. 添加分区

添加分区的语法：

```
ALTER TABLE table_name ADD [IF NOT EXISTS]
PARTITION (partition_column = partition_col_value, ...)
[LOCATION 'location'];
```

可以一次添加多个分区。

幻灯片50：4.3.2 —— 3. 重命名分区

重命名分区的语法：

```
ALTER TABLE table_name PARTITION (...)
RENAME TO PARTITION (...);
```

幻灯片51：4.3.2 —— 4. 删除分区 / 5. 修改分区

删除分区的语法：

```
ALTER TABLE table_name DROP [IF EXISTS]
PARTITION (...) [PURGE];
```

还可以修改分区的存储位置和文件格式：

```
ALTER TABLE ... PARTITION (...) SET LOCATION 'new_location';
ALTER TABLE ... PARTITION (...) SET FILEFORMAT file_format;
```

幻灯片52-58：4.3.3 任务实施 —— 分区表实战

好，我们来实际操作一下分区表。

（1）创建分区表：在 hive_database 中创建一个按省份和城市分区的表。

注意：分区字段名不能跟表中已有的列名重复（chóng fù）！

（2）查看分区信息：

【实习命令 / 실습 명령】

```
SHOW PARTITIONS hive_database.partitioned_table;
```

（3）添加分区——同时添加两个分区：

【实习命令 / 실습 명령】

```
ALTER TABLE hive_database.partitioned_table
ADD PARTITION (province='HeBei', city='HanDan')
    location '/user/hive/warehouse/hive_database.db/HeBei'
PARTITION (province='ShanDong', city='JiNan')
    location '/user/hive/warehouse/hive_database.db/ShanDong';
```

（4）重命名分区——把HeBei/HanDan改为HuBei/WuHan：

【实习命令 / 실습 명령】

```
ALTER TABLE hive_database.partitioned_table
PARTITION (province='HeBei', city='HanDan')
RENAME TO PARTITION (province='HuBei', city='WuHan');
```

（5）删除分区：

【实习命令 / 실습 명령】

```
ALTER TABLE hive_database.partitioned_table1
DROP PARTITION (province='HuBei', city='WuHan');
```

幻灯片59：分区表注意事项

最后，关于分区表有几个重要的注意事项（zhù yì shì xiàng）：

第一，分区不是建表必须的，它是一种优化手段（shǒu duàn），可选的。

第二，分区字段不能是表中已有的字段，不能重名。

第三，分区字段是虚拟字段，数据不存储在底层文件中。

第四，分区字段的值可以手动指定（静态分区）或者自动推断（tuī duàn）（动态分区）。

第五，Hive支持多重（duō chóng）分区，就是在分区的基础上再分区，分得更细。

项目小结

幻灯片101：项目小结

好，同学们，今天的课程到这里就全部结束了！
我们来回顾（huí gù）一下今天学了什么。

任务4-1：学习了HiveQL操作数据库——创建、查询、查看、切换、修改和删除数据库。
任务4-2：学习了HiveQL操作数据表——创建内部表和外部表、查看、修改和删除表。
任务4-3：学习了HiveQL操作分区表——分区的概念、创建、添加、重命名和删除分区。
任务4-4：学习了HiveQL操作分桶表——分桶的概念、创建分桶表、分桶的好处。
任务4-5：学习了HiveQL操作视图——视图的概念、创建、查看、修改和删除视图。
任务4-6：学习了HiveQL相关操作——索引的概念、创建、查看、重建和删除索引。

今天的内容比较多，但是大部分操作的模式（mó shì）是一样的：
CREATE 创建、SHOW 查看列表、DESCRIBE 查看详情、ALTER 修改、DROP 删除。
掌握（zhǎng wò）了这个规律（guī lǜ），就能举一反三（jǔ yī fǎn sān）了。

课后大家一定要在虚拟机（xū nǐ jī）上动手练习这些命令，光看不练是学不会的！
下次课我们会继续学习Hive的数据操作。
同学们，今天辛苦（xīn kǔ）了，下课！谢谢大家！