Hive 数据仓库应用教程

项目四 Hive数据定义

—— 第1回 强义原稿（jiǎng yì yuán gǎo）——

讲课对象：专科大学学生

讲课语言：中文（口语体）

目录 / CONTENTS

1. 任务4-1 HiveQL操作数据库
2. 任务4-2 HiveQL操作数据表
3. 任务4-3 HiveQL操作分区表
4. 任务4-4 HiveQL操作分桶表
5. 任务4-5 HiveQL操作视图
6. 任务4-6 HiveQL相关操作

课程开始

幻灯片1：项目四 Hive数据定义（封面）

同学们好!

欢迎（huān yíng）来到我们Hive数据仓库（shù jù cāng kù）课程。

今天我们要学习的内容是项目四——Hive数据定义。

什么是数据定义呢?

简单来说，就是我们怎么在Hive里面创建数据库、创建表、管理（guǎn lǐ）表。

就像你要整理（zhěng lǐ）房间，首先得有柜子（guì zi），然后柜子里有抽屉（chōu ti），数据库就是柜子，表就是抽屉。

这一章的内容非常重要，是后面学习数据查询的基础（jī chǔ）。

我们一步一步来学，不要着急。

幻灯片2：目录 / CONTENTS

我们先来看一下今天的学习内容目录（mù lù）。

总共（zǒng gòng）有六个任务：

第一个：任务4-1，HiveQL操作数据库——教大家怎么创建和管理数据库。

第二个：任务4-2，HiveQL操作数据表——教大家怎么创建和管理数据表。

第三个：任务4-3，HiveQL操作分区表——这是一种优化（yōu huà）查询速度的特殊表。

第四个：任务4-4，HiveQL操作分桶表——这也是一种优化查询的方法。

第五个：任务4-5，HiveQL操作视图——视图（shì tú）是一个虚拟的表。

第六个：任务4-6，HiveQL相关操作——包括索引（suǒ yǐn）等其他操作。

好，我们现在正式开始!

任务 4-1 HiveQL操作数据库

幻灯片3：任务4-1 HiveQL操作数据库（标题页）

好，我们现在开始第一个任务——HiveQL操作数据库。

HiveQL就是Hive的查询语言（chá xún yǔ yán），跟SQL很像。

如果你学过SQL，那这些内容对你来说会很容易（róng yì）理解。

即使没有学过也没关系，我会一个一个解释（jiě shì）清楚。

幻灯片4：4.1.1 任务描述

这个任务主要讲什么呢?

为了方便数据的存储（cún chǔ）和管理，Hive数据仓库会把数据分类（fēn lèi）放到不同的数据库里。

可以把数据库理解为一个大文件夹（wén jiàn jiā），里面放着很多表。

在这个任务里，我们会学习数据库的六个基本操作：

创建（chuàng jiàn）、查询（chá xún）、查看、切换（qiē huàn）、修改（xiū gǎi）、删除（shān chú）。

这些操作是最基础的，一定要掌握（zhǎng wò）。

幻灯片5：4.1.2 任务准备知识 —— 1. 创建数据库

我们先来看怎么创建数据库。

创建数据库的语法（yǔ fǎ）是这样的：

```
CREATE (DATABASE|SCHEMA) [IF NOT EXISTS] database_name
[COMMENT database_comment]
[LOCATION hdfs_path]
[WITH DBPROPERTIES (property_name=property_value, ...)];
```

我来一个一个解释：

CREATE DATABASE 或者 CREATE SCHEMA——这两个是一样的意思，都是创建数据库。

IF NOT EXISTS——这个是可选（kě xuǎn）的，意思是「如果数据库不存在就创建」。

加上这个的好处是：如果数据库已经存在了，它不会报错（bào cuò）。

database_name——就是你想给数据库起的名字。

COMMENT——可选的，就是给数据库加一个说明，告诉别人这个数据库是干什么的。

LOCATION——可选的，指定数据库的文件在HDFS上存在哪个位置（wèi zhì）。

WITH DBPROPERTIES——可选的，用来设置（shè zhì）数据库的一些属性（shǔ xìng）。

> *举个例子：就像你在电脑上新建一个文件夹，你可以选择放在哪里，也可以给它写个备注。创建数据库就是类似的操作。*

幻灯片6：4.1.2 任务准备知识 —— 注意事项

大家注意几个重要的事情：

第一，蓝色的字是语法的关键字（guān jiàn zì），就是固定（gù dìng）不变的部分。

第二，方括号 [ ] 里面的内容是可选的，你可以写也可以不写。

第三，竖线（shù xiàn）| 表示二选一，左边或者右边选一个用就行。

第四，写SQL语句的时候，顺序（shùn xù）很重要，要按照语法树的顺序来写。

不能乱（luàn）写，不然会报错。

幻灯片7：4.1.2 任务准备知识 —— 2. 查询数据库

学完了创建，我们来看怎么查询已经有的数据库。

语法很简单：

```
SHOW (DATABASES|SCHEMAS) [LIKE 'identifier_with_wildcards'];
```

SHOW DATABASES 就是列出（liè chū）所有的数据库。

后面可以加 LIKE 做模糊查询（mó hú chá xún）。

什么是模糊查询呢？

就是你不知道完整的名字，只知道一部分，比如名字以 'h' 开头的数据库。

你可以写 LIKE 'h*'，这个星号（xīng hào）* 代表任意字符（zì fú）。

> 💡 *比如说：你有100个数据库，但你只想找名字以'h'开头的，用LIKE 'h*'就能找到了。*

幻灯片8：4.1.2 任务准备知识 —— 3. 查看数据库

查看数据库是看一个数据库的详细（xiáng xì）信息。

语法如下：

```
DESCRIBE|DESC (DATABASES|SCHEMAS) [EXTENDED] db_name;
```

DESCRIBE 可以简写成 DESC，都是查看的意思。

EXTENDED 是可选的，加了它会显示更多的信息，包括数据库的属性。

db_name 就是你要查看的数据库名字。

> 💡 *SHOW是看有哪些数据库（列出名字），DESCRIBE是看某一个数据库的详细信息，两个不一样哦！*

幻灯片9：4.1.2 任务准备知识 —— 4. 切换数据库

切换数据库非常简单，就一行代码：

```
USE db_name;
```

USE 后面跟你要切换到的数据库名字就行了。

切换之后，你后面执行（zhí xíng）的所有操作都是在这个数据库里面的。

就像你打开了一个文件夹，后面的操作都是在这个文件夹里做的。

幻灯片10：4.1.2 任务准备知识 —— 5. 修改数据库

修改数据库有两种操作：

第一种是修改数据库的属性：

```
ALTER (DATABASE|SCHEMA) database_name SET DBPROPERTIES
(property_name=property_value, ...);
```

第二种是修改数据库的所有者（suǒ yǒu zhě）：

```
ALTER (DATABASE|SCHEMA) database_name SET OWNER [USER|ROLE] user_or_role;
```

ALTER 就是修改的意思。

SET DBPROPERTIES 是设置属性。

SET OWNER 是设置所有者，就是改变谁拥有（yōng yǒu）这个数据库。

幻灯片11：4.1.2 任务准备知识 —— 6. 删除数据库

删除数据库的语法：

```
DROP (DATABASE|SCHEMA) [IF EXISTS] database_name [RESTRICT|CASCADE];
```

DROP 就是删除的意思。

IF EXISTS 的意思是如果数据库存在就删除，不存在就不管了，也不会报错。

最后有一个重要的选择：RESTRICT 还是 CASCADE。

RESTRICT 是默认（mò rèn）的，意思是数据库里面如果有表，就不让你删除。

CASCADE 的意思是：即使数据库里有表，也把所有的表一起（yì qǐ）全部删掉。

> *小心使用CASCADE！它会把数据库里面的所有表都删掉，很危险（wēi xiǎn）！就像你把一个文件夹连同里面所有文件一起删了一样。*

幻灯片12：4.1.3 任务实施 —— 1. 创建数据库

好，现在我们来做实际操作。

我们要在Hive客户端（kè hù duān）工具Beeline里面创建一个数据库，名字叫 h_sample。

同时，我们指定数据存放在HDFS的 /hive_db/create_db/ 目录下。

【实习命令 / 실습 명령】

```
CREATE DATABASE IF NOT EXISTS h_sample
LOCATION '/hive_db/create_db/';
```

大家看到了吗？我们用了 IF NOT EXISTS，这样如果h_sample已经存在了，也不会报错。

LOCATION 指定了存储位置。

幻灯片13：4.1.3 任务实施 —— 2. 查询已有的数据库

创建好之后，我们来查询一下。

首先，查看Hive中所有的数据库：

【实习命令 / 실습 명령】

```
SHOW DATABASES;
```

这样就能看到所有数据库的列表。

然后，如果我们只想查找名字以 'h' 开头的数据库：

【实习命令 / 실습 명령】

```
SHOW DATABASES LIKE 'h*';
```

这样就只会显示以 h 开头的数据库了。

幻灯片14-15：4.1.3 任务实施 —— 3~6. 查看/切换/修改/删除

接下来我们把查看、切换、修改和删除一起做一下。

查看数据库 h_sample 的详细信息：

【实习命令 / 실습 명령】

```
DESC DATABASE EXTENDED h_sample;
```

切换到数据库 h_sample：

【实习命令 / 실습 명령】

```
USE h_sample;
```

修改数据库 h_sample 的属性，给它加一个日期属性：

【实习命令 / 실습 명령】

```
ALTER DATABASE h_sample SET DBPROPERTIES ("date"="2022-05-18");
```

最后，删除数据库 h_sample：

【实习命令 / 실습 명령】

```
DROP DATABASE IF EXISTS h_sample;
```

好，到这里，数据库的基本操作我们就学完了！

大家有没有觉得（jué de）其实不难？创建用CREATE，查看用SHOW和DESC，切换用USE，修改用ALTER，删除用DROP。

记住这几个关键字就可以了！