大 数 据 库 系 统

6.4 Hive的基本操作

6.4 Hive的基本操作

◆ **本节内容**

6.4.1 Hive中数据库的操作

◆ **数据库的操作**

创建数据库

删除数据库

查看数据库信息

6.4.1 Hive中数据库的操作

◆ **官方文档：数据库创建语句**

Create Database

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

The uses of SCHEMA and DATABASE are interchangeable – they mean the same thing. CREATE DATABASE was added in Hive 0.6 (HIVE-675). The WITH DBPROPERTIES clause was added in Hive 0.7 (HIVE-1836).

例如：创建名为"tmp1"的数据库：

create database if not exists tmp1;

```
hive (default)> create database if not exists tmp1;
OK
Time taken: 0.097 seconds
hive (default)> █
```

6.4.1 Hive中数据库的操作

◆ **数据准备**

在HDFS上"/"下建立一个student.txt文件，输入以下内容（学号string，姓名string两个字段）作为我们表的数据

```
1001    zhangsan
1002    lisi
1003    wangwu
1004    zhaoliu

中间用TAB键分隔

-- INSERT --
```

6.4.1 Hive中数据库的操作

◆ 创建表tmp1

create database if not exists tmp1;　←——————　创建tmp1数据库

use tmp1;　←————　指定当前数据库为tmp1

create table if not exists tmp1_table(

number string,

name string

) row format delimited fields terminated by '\t';

创建名为"tmp3_table"的内部表，指定数据分隔符为TAB

load data inpath ' /student.txt' into table tmp1_table;　←————　将student.txt中的数据导入表中

1、创建表前先使用数据库"tmp1"，注意hive命令行后的括号内有当前数据库名称

```
hive (default)> use tmp1;
OK
Time taken: 0.015 seconds
hive (tmp1)> █
```

2、创建表"tmp1_table"

```
hive (tmp1)> create table if not exists tmp1_table(
            > number string,
            > name string
            > ) row format delimited fields terminated by '\t';
OK
Time taken: 0.086 seconds
hive (tmp1)> █
```

3、使用show tables命令显示数据库中的所有表名

```
hive (tmp1)> show tables;
OK
tmp1_table
Time taken: 0.062 seconds, Fetched: 1 row(s)
hive (tmp1)>
```

4、将"/"目录下的student.txt文件里的内容导入到表中

```
hive (tmp1)> load data inpath '/student.txt' into table tmp1_table;
Loading data to table tmp1.tmp1_table
Table tmp1.tmp1_table stats: [numFiles=1, numRows=0, totalSize=49, rawDataSize=0]
OK
Time taken: 0.256 seconds
hive (tmp1)> █
```

5、显示表"tmp1_table"的内容

```
hive (tmp1)> select * from tmp1_table;
OK
1001    zhangsan
1002    lisi
1003    wangwu
1004    zhaoliu
Time taken: 0.052 seconds, Fetched: 4 row(s)
```

6.4.1 Hive中数据库的操作

◆ 上述操作后，会在hdfs的/user/hive/warehouse下创建tmp1.db目录（数据库目录），并在tmp1.db目录下会创建"tmp1_table"目录（表目录）

Browse Directory

/user/hive/warehouse/tmp1.db							Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	hpsk	supergroup	0 B	0	0 B	tmp1_table

6.4.1 Hive中数据库的操作

Create Database

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

The uses of SCHEMA and DATABASE are interchangeable – they mean the same thing. CREATE DATABASE was added in Hive 0.6 (HIVE-675). The WITH DBPROPERTIES clause was added in Hive 0.7 (HIVE-1836).

◆ 官方描述中还有个"LOCATION"关键字

6.4.1 Hive中数据库的操作

例：创建数据库tmp2，存放于hdfs下的"/hive/tmp2"目录：

create database if not exists tmp2 location '/hive/tmp2';

不加"LOCATION"，则默认将数据库存放在hdfs上的
/user/hive/warehouse目录下

加"LOCATION"，则由用户指定数据库在hdfs上的目录（通常外部表用）

1、在hdfs上创建目录"/hive/tmp2"，并创建数据库"tmp2"；

```
hive (tmp1)> dfs -mkdir -p /hive/tmp2;
hive (tmp1)> create database if not exists tmp2 location '/hive/tmp2';
OK
Time taken: 0.038 seconds
hive (tmp1)>
```

2、使用tmp2数据库并创建表;

```
hive (tmp1)> use tmp2;
OK
Time taken: 0.01 seconds
hive (tmp2)> create table if not exists tmp2_table(
           > number string,
           > name string
           > ) row format delimited fields terminated by '\t';
OK
Time taken: 0.044 seconds
hive (tmp2)> load data local inpath '/opt/modules/hive-0.13.1-bin/student.txt' into table tmp2_table;
Copying data from file:/opt/modules/hive-0.13.1-bin/student.txt
Copying file: file:/opt/modules/hive-0.13.1-bin/student.txt
Loading data to table tmp2.tmp2_table
Table tmp2.tmp2_table stats: [numFiles=1, numRows=0, totalSize=49, rawDataSize=0]
OK
Time taken: 0.232 seconds
hive (tmp2)> show tables;
OK
```

3、发现"/hive"目录下多了个"tmp2"目录，在"tmp2"下有"tmp2_table"目录

Browse Directory

/hive/tmp2						Go!
Permission	Owner	Group	Size	Replication	Block Size	Name
drwxr-xr-x	hpsk	supergroup	0 B	0	0 B	tmp2_table

6.4.1 Hive中数据库的操作

◆ 数据库删除语句

Drop Database

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

The uses of SCHEMA and DATABASE are interchangeable – they mean the same thing. DROP DATABASE was added in Hive 0.6 (HIVE-675).

6.4.1 Hive中数据库的操作

例：尝试使用drop命令删除刚刚创建的tmp1数据库

输入命令：

drop database tmp1;

```
hive (tmp2)> drop database tmp1;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. InvalidOperationEx
ception(message:Database tmp1 is not empty. One or more tables exist.)
hive (tmp2)>
```

发现出错，为什么？

数据库tmp1不是个空数据库，里面有至少一张表

6.4.1 Hive中数据库的操作

Drop Database

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

The uses of SCHEMA and DATABASE are interchangeable – they mean the same thing. DROP DATABASE was added in Hive 0.6 (HIVE-675).

◆ **如果要删除非空数据库：**

drop database tmp1 cascade;

1、删除非空数据库；

```
hive (tmp2)> drop database tmp1 cascade;
OK
Time taken: 0.765 seconds
```

2、使用show databases，显示所有数据库；

```
hive (tmp2)> show databases;
OK
default
student
tmp2
Time taken: 0.022 seconds, Fetched: 3 row(s)
hive (tmp2)> █
```

找不到tmp1数据库，成功删除！

create database if not exists tmp1;

use tmp1;

create table if not exists tmp1_table(

number string,

name string

) row format delimited fields terminated by '\t';

load data inpath /student.txt' into table tmp1_table;

6.4.1 Hive中数据库的操作

到默认的warehouse目录下也找不到"tmp1"的目录;

Browse Directory

/user/hive/warehouse							Go!
Permission	Owner	Group	Size	Replication	Block Size	Name	
drwxr-xr-x	hpsk	supergroup	0 B	0	0 B	stu_tmp	
drwxr-xr-x	hpsk	supergroup	0 B	0	0 B	student.db	

注意：删除时会删除元数据及HDFS的目录！

6.4.1 Hive中数据库的操作

◆ **显示数据库**：show databases

◆ **模糊查找数据库**：show databases like

例：查找名称中带有"db_hive"的数据库

```
hive> show databases like 'db_hive*';
OK
db_hive
db_hive_1
```

6.4.1 Hive中数据库的操作

◆ **官方文档：查看数据库信息**

Describe Database

> (i) **Version information**
> As of Hive 0.7.

```
DESCRIBE DATABASE [EXTENDED] db_name;
DESCRIBE SCHEMA [EXTENDED] db_name;     -- (Note: Hive 0.15.0 and later)
```

DESCRIBE DATABASE shows the name of the database, its comment (if one has been set), and its root location on the filesystem. The uses of SCHEMA and DATABASE are interchangeable – they mean the same thing. DESCRIBE SCHEMA is added in Hive 0.15 (HIVE-8803).

EXTENDED also shows the database properties.

6.4.1 Hive中数据库的操作

◆ **查看tmp2数据库的详细信息：**

desc database EXTENDED tmp2;

加上"EXTENDED"参数表示显示详细的信息

```
hive (tmp2)> desc database tmp2;
OK
tmp2            hdfs://bigdata-training01.hpsk.com:8020/hive/tmp2        hpsk
Time taken: 0.013 seconds, Fetched: 1 row(s)
```

总结

◆ **数据库命令：**

- ✓ create database if not exists dbname [location]

- ✓ use dbname

- ✓ drop database dbname [cascade]

- ✓ desc database EXTENDED

- ✓ show databases