

第 5 章 DML 数据操作

5.1 数据导入

5.1.1 向表中装载数据 (Load)

1) 语法

```
hive> load data [local] inpath '/opt/module/datas/student.txt' [overwrite] into table student  
[partition (partcol1=val1,...)];
```

- (1) load data:表示加载数据
- (2) local:表示从本地加载数据到 hive 表; 否则从 HDFS 加载数据到 hive 表
- (3) inpath:表示加载数据的路径
- (4) overwrite:表示覆盖表中已有数据, 否则表示追加
- (5) into table:表示加载到哪张表
- (6) student:表示具体的表
- (7) partition:表示上传到指定分区

2) 实操案例

(0) 创建一张表

```
hive (default)> create table student(id string, name string) row format delimited fields  
terminated by '\t';
```

(1) 加载本地文件到 hive

```
hive (default)> load data local inpath '/opt/module/datas/student.txt' into table  
default.student;
```

(2) 加载 HDFS 文件到 hive 中

上传文件到 HDFS

```
hive (default)> dfs -put /opt/module/datas/student.txt /user/atguigu/hive;
```

加载 HDFS 上数据

```
hive (default)> load data inpath '/user/atguigu/hive/student.txt' into table default.student;
```

(3) 加载数据覆盖表中已有的数据

上传文件到 HDFS

```
hive (default)> dfs -put /opt/module/datas/student.txt /user/atguigu/hive;
```

加载数据覆盖表中已有的数据

```
hive (default)> load data inpath '/user/atguigu/hive/student.txt' overwrite into table
default.student;
```

5.1.2 通过查询语句向表中插入数据（Insert）

1) 创建一张分区表

```
hive (default)> create table student(id int, name string) partitioned by (month string) row
format delimited fields terminated by '\t';
```

2) 基本插入数据

```
hive (default)> insert into table student partition(month='201709') values(1,'wangwu');
```

3) 基本模式插入（根据单张表查询结果）

```
hive (default)> insert overwrite table student partition(month='201708')
select id, name from student where month='201709';
```

4) 多插入模式（根据多张表查询结果）

```
hive (default)> from student
insert overwrite table student partition(month='201707')
select id, name where month='201709'
insert overwrite table student partition(month='201706')
select id, name where month='201709';
```

5.1.3 查询语句中创建表并加载数据（As Select）

详见 4.5.1 章创建表。

根据查询结果创建表（查询的结果会添加到新创建的表中）

```
create table if not exists student3
as select id, name from student;
```

5.1.4 创建表时通过 Location 指定加载数据路径

1) 创建表，并指定在 hdfs 上的位置

```
hive (default)> create table if not exists student5(
id int, name string
)
row format delimited fields terminated by '\t'
location '/user/hive/warehouse/student5';
```

2) 上传数据到 hdfs 上

```
hive (default)> dfs -put /opt/module/datas/student.txt /user/hive/warehouse/student5;
```

3) 查询数据

```
hive (default)> select * from student5;
```

5.1.5 Import 数据到指定 Hive 表中

注意：先用 export 导出后，再将数据导入。

```
hive (default)> import table student2 partition(month='201709') from  
'/user/hive/warehouse/export/student';
```

5.2 数据导出

5.2.1 Insert 导出

1) 将查询的结果导出到本地

```
hive (default)> insert overwrite local directory '/opt/module/datas/export/student'  
  
select * from student;
```

2) 将查询的结果格式化导出到本地

```
hive (default)> insert overwrite local directory '/opt/module/datas/export/student1'  
  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'  
  
select * from student;
```

3) 将查询的结果导出到 HDFS 上(没有 local)

```
hive (default)> insert overwrite directory '/user/atguigu/student2'  
  
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t'  
  
select * from student;
```

5.2.2 Hadoop 命令导出到本地

```
hive (default)> dfs -get /user/hive/warehouse/student/month=201709/000000_0  
  
/opt/module/datas/export/student3.txt;
```

5.2.3 Hive Shell 命令导出

基本语法：(hive -f/-e 执行语句或者脚本 > file)

```
[atguigu@hadoop102 hive]$ bin/hive -e 'select * from default.student;' >  
  
/opt/module/datas/export/student4.txt;
```

5.2.4 Export 导出到 HDFS 上

```
hive (default)> export table default.student to '/user/hive/warehouse/export/student';
```

5.2.5 Sqoop 导出

后续课程专门讲。

5.3 清除表中数据 (Truncate)

注意：Truncate 只能删除管理表，不能删除外部表中数据

```
hive (default)> truncate table student;
```