

Hive insert into语句用法

在Hive0.8开始支持Insert into语句，它的作用是在一个表格里面追加数据。
标准语法语法如下：

用法一：

```
INSERT OVERWRITE TABLE tablename1 [PARTITION  
(partcol1=val1, partcol2=val2 ...)] [IF NOT EXISTS]]  
select_statement1 FROM from_statement;
```

用法二：

```
INSERT INTO TABLE tablename1 [PARTITION  
(partcol1=val1, partcol2=val2 ...)]  
select_statement1 FROM from_statement;
```

举例：

```
hive> insert into table cite  
> select * from tt;
```

这样就会将tt表格里面的数据追加到cite表格里面。并且在cite数据存放目录生成了一个新的数据文件，这个新文件是经过处理的，列之间的分割是cite表格的列分割符，而不是tt表格列的分隔符。

◦ (1)、如果两个表格的维度不一样，将会插入错误：

```
hive> insert into table cite  
> select * from cite_standby;
```

```
FAILED: SemanticException [Error 10044]: Line 1:18 Cannot insert into  
target table because column number/types are different 'cite':  
Table insclause-0 has 2 columns, but query has 1 columns.
```

从上面错误提示看出，查询的表格cite_standby只有一列，而目标表格（也就是需要插入数据的表格）有2列，由于列的数目不一样，导致了上面的语句不能成功运行，我们需要保证查询结果

列的数目和需要插入数据表格的列数目一致，这样才行。

(2)、在用extended关键字创建的表格上插入数据将会影响到其它的表格的数据，因为他们共享一份数据文件。

(3)、如果查询出来的数据类型和插入表格对应的列数据类型不一致，将会进行转换，但是不能保证转换一定成功，比如如果查询出来的数据类型为int，插入表格对应的列类型为string，可以通过转换将int类型转换为string类型；但是如果查询出来的数据类型为string，插入表格对应的列类型为int，转换过程可能出现错误，因为字母就不可以转换为int，转换失败的数据将会为NULL。

(4)、可以将一个表查询出来的数据插入到原表中：

```
hive> insert into table cite  
> select * from cite;
```

结果就是相当于复制了一份cite表格中的数据。

(5)、和insert overwrite的区别：

```
hive> insert overwrite table cite  
> select * from tt;
```

上面的语句将会用tt表格查询到的数据覆盖cite表格已经存在的数据。

更多关于Hive insert into的用法，可以参见Hive

官方wiki：<https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DML>

本博客文章除特别声明，全部都是原创！

原创文章版权归过往记忆大数据（[过往记忆](#)）所有，未经许可不得转载。

本文链接: [【】](#)（）