

## 创建数据库表操作

### 创建数据库表语法

```
CREATE [EXTERNAL] TABLE [IF NOT EXISTS] table_name

    [(col_name data_type [COMMENT col_comment], ...)]

    [COMMENT table_comment]

    [PARTITIONED BY (col_name data_type [COMMENT col_comment], ...)]

    [CLUSTERED BY (col_name, col_name, ...)]

    [SORTED BY (col_name [ASC|DESC], ...)] INTO num_buckets BUCKETS

    [ROW FORMAT row_format]

    [STORED AS file_format]

    [LOCATION hdfs_path]
```

说明：

- 1、CREATE TABLE 创建一个指定名字的表。如果相同名字的表已经存在，则抛出异常；用户可以用 IF NOT EXISTS 选项来忽略这个异常。
- 2、EXTERNAL 关键字可以让用户创建一个外部表，在建表的同时指定一个指向实际数据的路径（LOCATION），Hive 创建内部表时，会将数据移动到数据仓库指向的路径；若创建外部表，仅记录数据所在的路径，不对数据的位置做任何改变。在删除表的时候，内部表的元数据和数据会被一起删除，而外部表只删除元数据，不删除数据。
- 3、LIKE 允许用户复制现有的表结构，但是不复制数据。
- 4、ROW FORMAT **DELIMITED** [FIELDS TERMINATED BY char] [COLLECTION

```
ITEMS TERMINATED BY char] [MAP KEYS TERMINATED BY char] [LINES  
TERMINATED BY char] | SERDE serde_name [WITH SERDEPROPERTIES  
(property_name=property_value,  
property_name=property_value, ...)]
```

用户在建表的时候可以自定义 SerDe 或者使用自带的 SerDe。如果没有指定 ROW FORMAT 或者 ROW FORMAT DELIMITED, 将会使用自带的 SerDe。在建表的时候, 用户还需要为表指定列, 用户在指定表的列的同时也会指定自定义的 SerDe, Hive 通过 SerDe 确定表的具体的列的数据。

## 5、STORED AS

SEQUENCEFILE|TEXTFILE|RCFILE

如果文件数据是纯文本, 可以使用 STORED AS TEXTFILE。

## hive 建表初体验

```
use mydb01;  
create table stu(id int,name string);  
insert into stu values (1,"Tom");  
select * from stu;
```

```
0: jdbc:hive2://node03:10000> use mydb01;  
INFO : Compiling command(queryId=root_20190712093232_bcba2cca-2531-45b7-b654-da403afd8ead): use mydb01  
INFO : Semantic Analysis Completed  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=root_20190712093232_bcba2cca-2531-45b7-b654-da403afd8ead); Time taken: 0.007 seconds  
INFO : Concurrency mode is disabled, not creating a lock manager  
INFO : Executing command(queryId=root_20190712093232_bcba2cca-2531-45b7-b654-da403afd8ead): use mydb01  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=root_20190712093232_bcba2cca-2531-45b7-b654-da403afd8ead); Time taken: 0.014 seconds  
INFO : OK  
No rows affected (0.043 seconds)  
0: jdbc:hive2://node03:10000>
```

```
0: jdbc:hive2://node03:10000> create table stu(id int,name string);  
INFO : Compiling command(queryId=root_20190712093333_0f3d7f7b-bba4-4200-8438-d6790840e9ab): create table stu(id int  
,name string)  
INFO : Semantic Analysis Completed  
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)  
INFO : Completed compiling command(queryId=root_20190712093333_0f3d7f7b-bba4-4200-8438-d6790840e9ab); Time taken: 0.078 seconds  
INFO : Concurrency mode is disabled, not creating a lock manager  
INFO : Executing command(queryId=root_20190712093333_0f3d7f7b-bba4-4200-8438-d6790840e9ab): create table stu(id int  
,name string)  
INFO : Starting task [Stage-0:DDL] in serial mode  
INFO : Completed executing command(queryId=root_20190712093333_0f3d7f7b-bba4-4200-8438-d6790840e9ab); Time taken: 0.121 seconds  
INFO : OK  
No rows affected (0.218 seconds)  
0: jdbc:hive2://node03:10000>
```

```

0: jdbc:hive2://node03:10000> insert into stu values (1,"minglan");
INFO : Compiling command(queryId=root_20190712094545_892a7f12-7923-413f-8f8b-c4ff3dbdb364): insert into stu values
(1,"minglan")
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:_col0, type:int, comment:null), FieldSchema(name:
_col1, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=root_20190712094545_892a7f12-7923-413f-8f8b-c4ff3dbdb364); Time taken: 0
.201 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712094545_892a7f12-7923-413f-8f8b-c4ff3dbdb364): insert into stu values
(1,"minglan")
INFO : Query ID = root_20190712094545_892a7f12-7923-413f-8f8b-c4ff3dbdb364
INFO : Total jobs = 3
INFO : Launching Job 1 out of 3
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Number of reduce tasks is set to 0 since there's no reduce operator
INFO : Starting Job = job_1562889576719_0002, Tracking URL = http://node01:8088/proxy/application_1562889576719_000
2/
INFO : Kill Command = /export/servers/hadoop-2.6.0-cdh5.14.0/bin/hadoop job -kill job_1562889576719_0002
INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
INFO : 2019-07-12 09:45:38,471 Stage-1 map = 0%, reduce = 0%
INFO : 2019-07-12 09:45:40,667 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.98 sec
INFO : MapReduce Total cumulative CPU time: 1 seconds 980 msec
INFO : Ended Job = job_1562889576719_0002
INFO : Starting task [Stage-7:CONDITIONAL] in serial mode
INFO : Stage-4 is selected by condition resolver.
INFO : Stage-3 is filtered out by condition resolver.

```

```

INFO : Stage-5 is filtered out by condition resolver.
INFO : Starting task [Stage-4:MOVE] in serial mode
INFO : Moving data to: hdfs://node01:8020/user/hive/warehouse/mydb01.db/stu/.hive-staging_hive_2019-07-12_09-45-28_
184_8779695298220063920-1/-ext-10000 from hdfs://node01:8020/user/hive/warehouse/mydb01.db/stu/.hive-staging_hive_20
19-07-12_09-45-28_184_8779695298220063920-1/-ext-10000
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table mydb01.stu from hdfs://node01:8020/user/hive/warehouse/mydb01.db/stu/.hive-staging_hiv
e_2019-07-12_09-45-28_184_8779695298220063920-1/-ext-10000
INFO : Starting task [Stage-2:STATS] in serial mode
INFO : Table mydb01.stu stats: [numFiles=2, numRows=2, totalSize=20, rawDataSize=16]
INFO : MapReduce Jobs Launched:
INFO : Stage-Stage-1: Map: 1 Cumulative CPU: 1.98 sec HDFS Read: 3866 HDFS Write: 287929 SUCCESS
INFO : Total MapReduce CPU Time Spent: 1 seconds 980 msec
INFO : Completed executing command(queryId=root_20190712094545_892a7f12-7923-413f-8f8b-c4ff3dbdb364); Time taken: 1
3.648 seconds
INFO : OK
No rows affected (13.871 seconds)

```

创建表并指定字段之间的分隔符

```
create table if not exists stu2(id int ,name string) row format delimited fields
terminated by '\t' stored as textfile location '/user/stu2';
```

```

0: jdbc:hive2://node03:10000> create table if not exists stu2(id int ,name string) row format delimited fields term
inated by '\t' stored as textfile location '/user/stu2';
INFO : Compiling command(queryId=root_20190712133434_076b11f5-37c9-43ff-a17b-097b36ddb218): create table if not ex
ists stu2(id int ,name string) row format delimited fields terminated by '\t' stored as textfile location '/user/stu
2'
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=root_20190712133434_076b11f5-37c9-43ff-a17b-097b36ddb218); Time taken: 0
.162 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712133434_076b11f5-37c9-43ff-a17b-097b36ddb218): create table if not ex
ists stu2(id int ,name string) row format delimited fields terminated by '\t' stored as textfile location '/user/stu
2'
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=root_20190712133434_076b11f5-37c9-43ff-a17b-097b36ddb218); Time taken: 0
.286 seconds
INFO : OK
No rows affected (0.473 seconds)
0: jdbc:hive2://node03:10000>

```

根据查询结果创建表

```
create table stu3 as select * from stu;
```

```

0: jdbc:hive2://node03:10000> create table stu3 as select * from stu2;
INFO : Compiling command(queryId=root_20190712133535_6734fc72-2088-42bc-9418-22d49dc089fe): create table stu3 as se
lect * from stu2
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:stu2.id, type:int, comment:null), FieldSchema(n
ame:stu2.name, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=root_20190712133535_6734fc72-2088-42bc-9418-22d49dc089fe); Time taken: 0
.48 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712133535_6734fc72-2088-42bc-9418-22d49dc089fe): create table stu3 as se
lect * from stu2
INFO : Query ID = root_20190712133535_6734fc72-2088-42bc-9418-22d49dc089fe
INFO : Total jobs = 3
INFO : Launching Job 1 out of 3
INFO : Starting task [Stage-1:MAPRED] in serial mode
INFO : Number of reduce tasks is set to 0 since there's no reduce operator
INFO : Starting Job = job_1562908564764_0001, Tracking URL = http://node01:8088/proxy/application_1562908564764_000
1/
INFO : Kill Command = /export/servers/hadoop-2.6.0-cdh5.14.0/bin/hadoop job -kill job_1562908564764_0001
INFO : Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
INFO : 2019-07-12 13:35:20,616 Stage-1 map = 0%, reduce = 0%
INFO : 2019-07-12 13:35:22,735 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.28 sec
INFO : MapReduce Total cumulative CPU time: 1 seconds 280 msec
INFO : Ended Job = job_1562908564764_0001
INFO : Starting task [Stage-7:CONDITIONAL] in serial mode
INFO : Stage-4 is selected by condition resolver.
INFO : Stage-3 is filtered out by condition resolver.
INFO : Stage-5 is filtered out by condition resolver.

```

根据已经存在的表结构创建表

```

create table stu4 like stu2;
0: jdbc:hive2://node03:10000> create table stu4 like stu2;
INFO : Compiling command(queryId=root_20190712133636_fd0e7858-d260-4792-a11d-502132c79c64): create table stu4 like
stu2
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=root_20190712133636_fd0e7858-d260-4792-a11d-502132c79c64); Time taken: 0
.017 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712133636_fd0e7858-d260-4792-a11d-502132c79c64): create table stu4 like
stu2
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=root_20190712133636_fd0e7858-d260-4792-a11d-502132c79c64); Time taken: 0
.076 seconds
INFO : OK
No rows affected (0.114 seconds)
0: jdbc:hive2://node03:10000>

```

查询表的类型

```

desc formatted stu2;
0: jdbc:hive2://node03:10000> desc formatted stu2;
INFO : Compiling command(queryId=root_20190712133737_d0bb2cbe-f4ff-4e1b-a57b-38ad224518f6): desc formatted stu2
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:col_name, type:string, comment:from deserializ
er), FieldSchema(name:data_type, type:string, comment:from deserializer), FieldSchema(name:comment, type:string, comm
ent:from deserializer)], properties:null)
INFO : Completed compiling command(queryId=root_20190712133737_d0bb2cbe-f4ff-4e1b-a57b-38ad224518f6); Time taken: 0
.077 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712133737_d0bb2cbe-f4ff-4e1b-a57b-38ad224518f6): desc formatted stu2
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=root_20190712133737_d0bb2cbe-f4ff-4e1b-a57b-38ad224518f6); Time taken: 0
.04 seconds
INFO : OK

```

col_name	data_type	comment
# col_name	data_type	comment
id	int	NULL
name	string	NULL
# Detailed Table Information	NULL	NULL
Database:	default	NULL
Owner:	root	NULL
CreateTime:	Fri Jul 12 13:34:11 CST 2019	NULL
LastAccessTime:	UNKNOWN	NULL
Protect Mode:	None	NULL
Retention:	0	NULL
Location:	hdfs://node01:8020/user/stu2	NULL

## 操作案例

分别创建老师与学生表，并向表中加载数据

创建老师表：

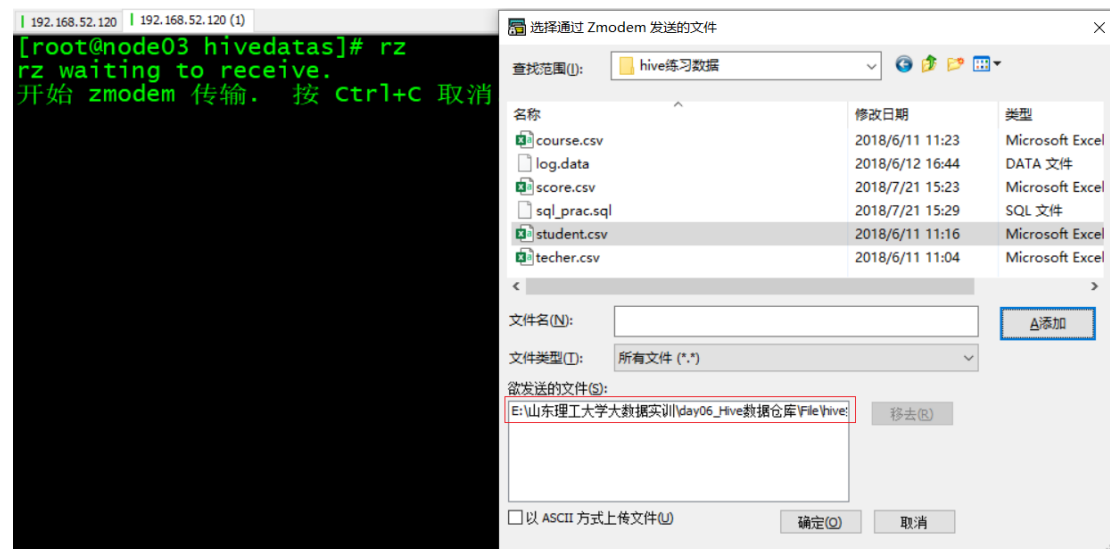
```
create table teacher (t_id string,t_name string) row format delimited
fields terminated by '\t';
```

创建学生表：

```
create table student (s_id string,s_name string,s_birth string , s_sex
string ) row format delimited fields terminated by '\t';
```

从本地文件系统向表中加载数据

```
[root@node03 servers]# mkdir hivedatas
[root@node03 servers]# ll
total 24
drwxr-xr-x.  8 1106 4001 4096 Jan  7  2018 apache-flume-1.6.0-cdh5.14.0-bin
drwxr-xr-x. 11 root root 4096 Jul  7 16:52 hadoop-2.6.0-cdh5.14.0
drwxr-xr-x. 11 1106 4001 4096 Jul 10 09:19 hive-1.1.0-cdh5.14.0
drwxr-xr-x.  2 root root 4096 Jul 12 13:40 hivedatas
drwxr-xr-x.  8 root root 4096 Jul  5 06:56 jdk1.8.0_141
drwxr-xr-x. 15 root root 4096 Jul  7 07:19 zookeeper-3.4.5-cdh5.14.0
[root@node03 servers]#
```



```
[root@node03 hivedatas]# ll
total 4
-rw-r--r--. 1 root root 200 Jun 11  2018 student.csv
[root@node03 hivedatas]#
```

```
load data local inpath '/export/servers/hivedatas/student.csv' into
```

```
table student;
```

```
0: jdbc:hive2://node03:10000> load data local inpath '/export/servers/hivedatas/student.csv' into table student;
INFO : Compiling command(queryId=root_20190712134646_437822d2-0887-4e22-bf11-bc2ebe6ae69c): load data local inpath
'/export/servers/hivedatas/student.csv' into table student
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=root_20190712134646_437822d2-0887-4e22-bf11-bc2ebe6ae69c); Time taken: 0
.048 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712134646_437822d2-0887-4e22-bf11-bc2ebe6ae69c): load data local inpath
'/export/servers/hivedatas/student.csv' into table student
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.student from file:/export/servers/hivedatas/student.csv
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Table default.student stats: [numFiles=1, totalSize=200]
INFO : Completed executing command(queryId=root_20190712134646_437822d2-0887-4e22-bf11-bc2ebe6ae69c); Time taken: 0
.349 seconds
INFO : OK
No rows affected (0.422 seconds)
0: jdbc:hive2://node03:10000>
```

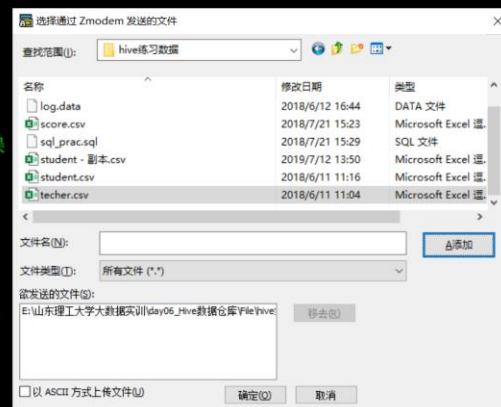
```
0: jdbc:hive2://node03:10000> load data local inpath '/export/servers/hivedatas/student.csv' overwrite into table student;
INFO : Compiling command(queryId=root_20190712134747_01b64081-5fe8-4bae-8329-8494cba38411): load data local inpath
'/export/servers/hivedatas/student.csv' overwrite into table student
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:null, properties:null)
INFO : Completed compiling command(queryId=root_20190712134747_01b64081-5fe8-4bae-8329-8494cba38411); Time taken: 0
.023 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=root_20190712134747_01b64081-5fe8-4bae-8329-8494cba38411): load data local inpath
'/export/servers/hivedatas/student.csv' overwrite into table student
INFO : Starting task [Stage-0:MOVE] in serial mode
INFO : Loading data to table default.student from file:/export/servers/hivedatas/student.csv
INFO : Starting task [Stage-1:STATS] in serial mode
INFO : Table default.student stats: [numFiles=1, totalSize=200]
INFO : Completed executing command(queryId=root_20190712134747_01b64081-5fe8-4bae-8329-8494cba38411); Time taken: 0
.284 seconds
INFO : OK
No rows affected (0.325 seconds)
0: jdbc:hive2://node03:10000>
```

从 hdfs 文件系统向表中加载数据（需要提前将数据上传到 hdfs 文件系统，其实就是一个移动文件的操作）

```
cd /export/servers/hivedatas
hdfs dfs -mkdir -p /hivedatas
hdfs dfs -put /export/servers/hivedatas/teacher.csv /hivedatas/
load data inpath '/hivedatas/teacher.csv' into table teacher;
```

```
[root@node03 hivedatas]# cd /export/servers/hivedatas
You have new mail in /var/spool/mail/root
[root@node03 hivedatas]# hdfs dfs -mkdir -p /hivedatas
```

```
192.168.52.120 | 192.168.52.120 (1)
[root@node03 hivedatas]# rz
rz waiting to receive.
开始 zmodem 传输. 按 Ctrl+C 取消.
You have new mail in /var/spool/mail/root
[root@node03 hivedatas]# rz
rz waiting to receive.
开始 zmodem 传输. 按 Ctrl+C 取消.
Transferring student.csv...
100% 200 bytes 200 bytes/s 00:00:01 0 错误
[root@node03 hivedatas]# ll
total 4
-rw-r--r--. 1 root root 200 Jun 11 2018 student.csv
[root@node03 hivedatas]# ll
total 4
-rw-r--r--. 1 root root 200 Jun 11 2018 student.csv
[root@node03 hivedatas]# cd /export/servers/hivedatas
You have new mail in /var/spool/mail/root
[root@node03 hivedatas]# hdfs dfs -mkdir -p /hivedatas
[root@node03 hivedatas]# rz
rz waiting to receive.
开始 zmodem 传输. 按 Ctrl+C 取消.
```



创建普通表,并向表中加载数据

```
create table course (c_id string,c_name string,t_id string) row format  
delimited fields terminated by '\t';
```

```
load data local inpath '/export/servers/hivedatas/course.csv' into  
table course;
```

```
create table score (s_id string,c_id string,s_score int) row format  
delimited fields terminated by '\t';
```

```
load data local inpath '/export/servers/hivedatas/score.csv' into  
table score;
```