题目：Hive 实验

姓名：南希诺

日期：2022/10/19

实验环境：Ubuntu 21.04

实验内容和完成情况：

创建，修改和删除数据库

create database if not exists hive; #创建数据库

show databases; #查看 Hive 中包含数据库

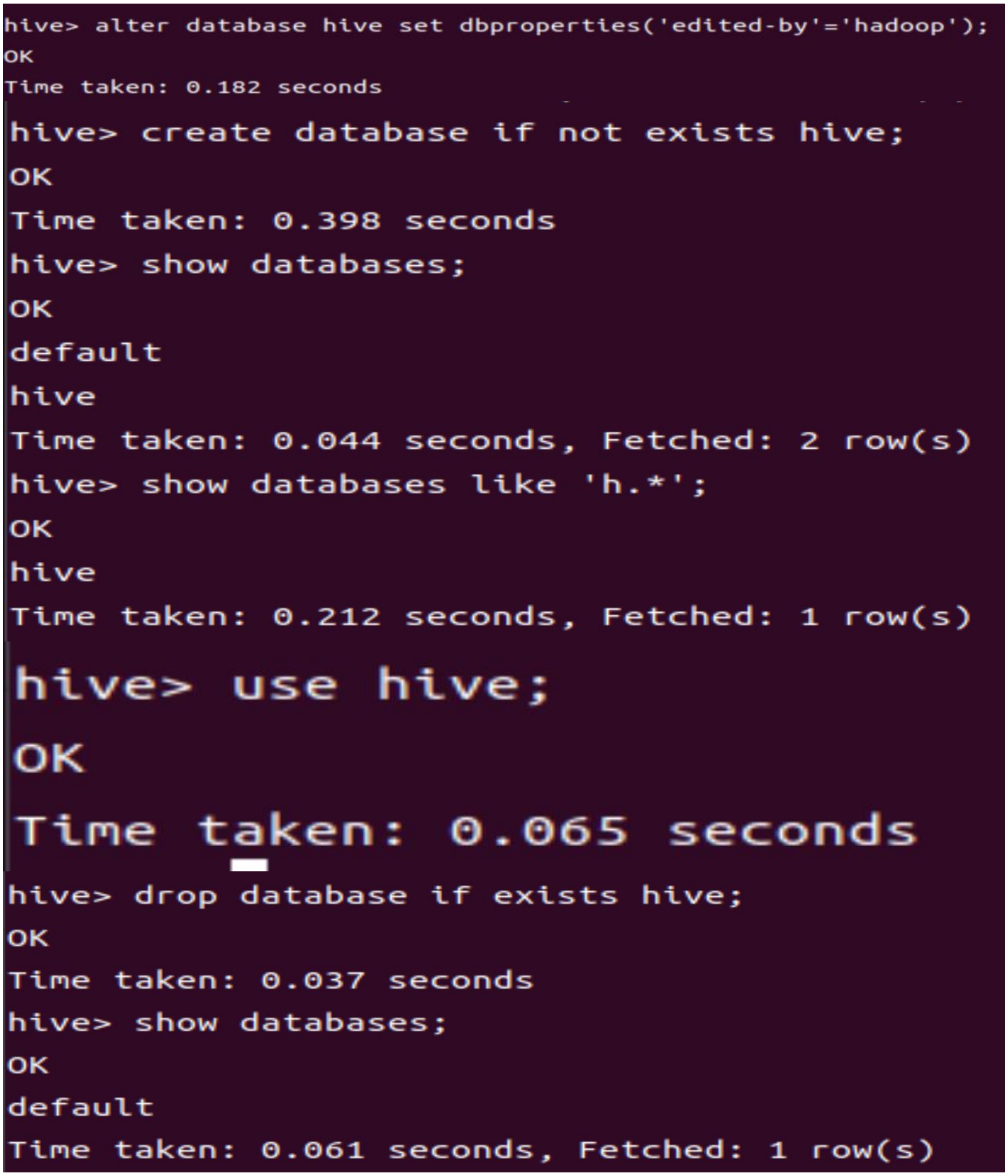
show databases like 'h.*'; #查看 Hive 中以 h 开头数据库

use hive; #切换到 hive 数据库下

drop database if exists hive; #删除不含表的数据库

为 hive 设置键值对属性

alter database hive set dbproperties('edited-by'='hadoop');



```
hive> alter database hive set dbproperties('edited-by'='hadoop');
OK
Time taken: 0.182 seconds
hive> create database if not exists hive;
OK
Time taken: 0.398 seconds
hive> show databases;
OK
default
hive
Time taken: 0.044 seconds, Fetched: 2 row(s)
hive> show databases like 'h.*';
OK
hive
Time taken: 0.212 seconds, Fetched: 1 row(s)
hive> use hive;
OK
Time taken: 0.065 seconds
hive> drop database if exists hive;
OK
Time taken: 0.037 seconds
hive> show databases;
OK
default
Time taken: 0.061 seconds, Fetched: 1 row(s)
```

创建、修改、删除数据表

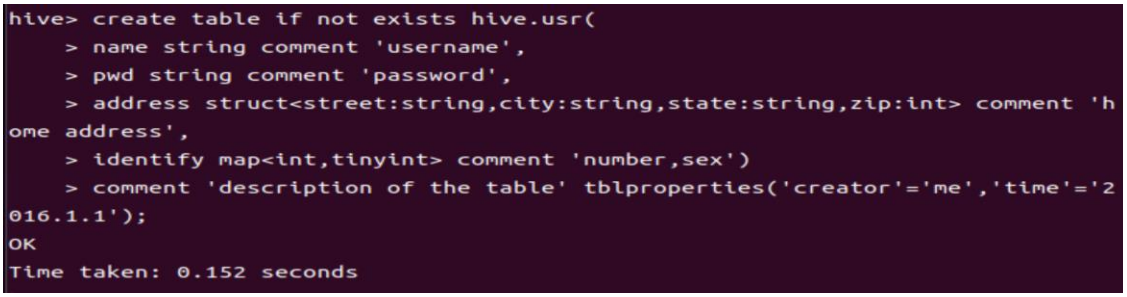
创建内部表

```
create table if not exists hive.usr(
name string comment 'username',
pwd string comment 'password',
```

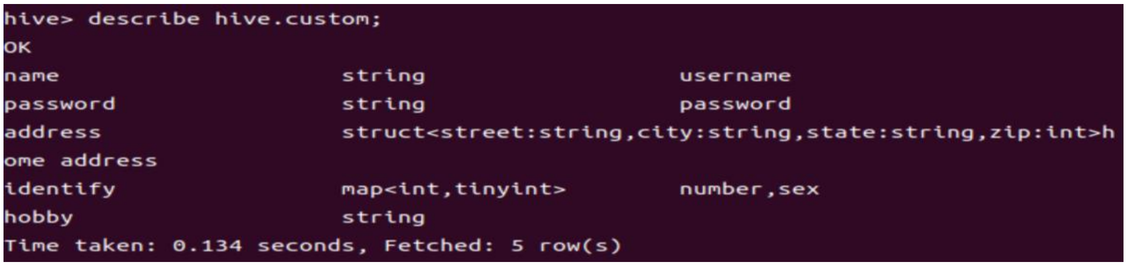
address struct<street:string,city:string,state:string,zip:int> comment 'home address',

identify map<int,tinyint> comment 'number,sex')

comment 'description of the table' tblproperties('creator'='me','time'='2016.1.1');



```
hive> create table if not exists hive.usr(
    > name string comment 'username',
    > pwd string comment 'password',
    > address struct<street:string,city:string,state:string,zip:int> comment 'h
ome address',
    > identify map<int,tinyint> comment 'number,sex')
    > comment 'description of the table' tblproperties('creator'='me','time'='2
016.1.1');
OK
Time taken: 0.152 seconds
```



```
hive> describe hive.custom;
OK
name                    string                  username
password                string                  password
address                 struct<street:string,city:string,state:string,zip:int>h
ome address
identify                map<int,tinyint>        number,sex
hobby                   string
Time taken: 0.134 seconds, Fetched: 5 row(s)
```

创建外部表

create external table if not exists usr2(

name string,

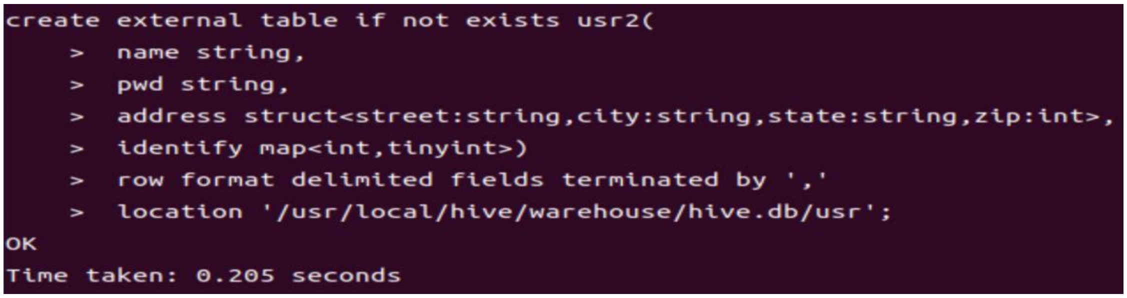
pwd string,

address struct<street:string,city:string,state:string,zip:int>,

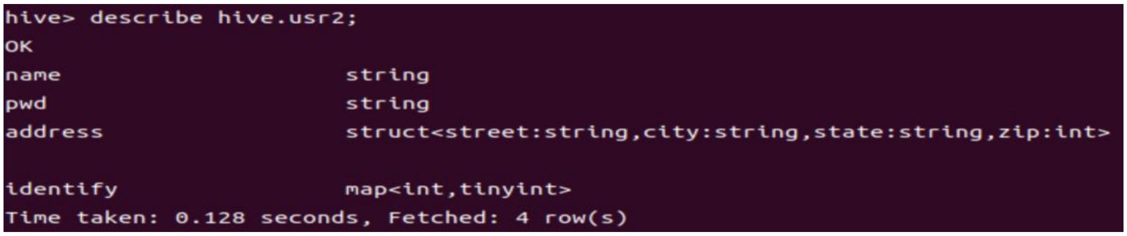
identify map<int,tinyint>)

row format delimited fields terminated by ','

location '/usr/local/hive/warehouse/hive.db/usr';



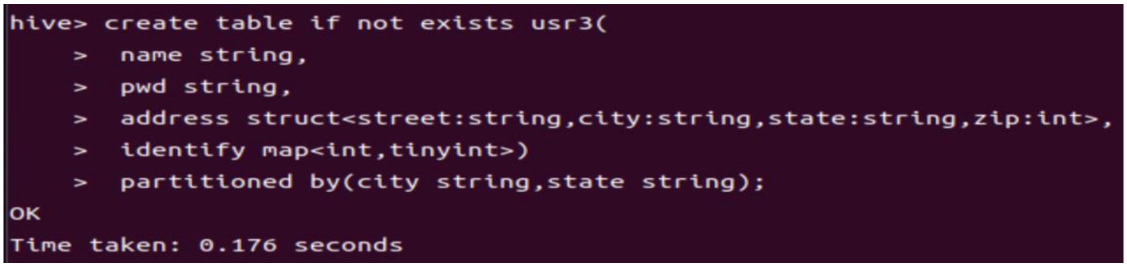
```
create external table if not exists usr2(
    >  name string,
    >  pwd string,
    >  address struct<street:string,city:string,state:string,zip:int>,
    >  identify map<int,tinyint>)
    >  row format delimited fields terminated by ','
    >  location '/usr/local/hive/warehouse/hive.db/usr';
OK
Time taken: 0.205 seconds
```



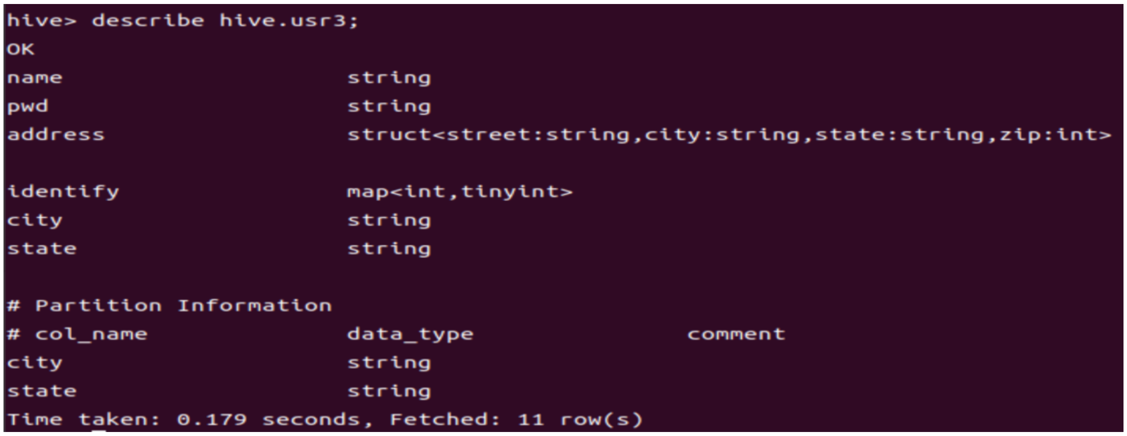
```
hive> describe hive.usr2;
OK
name                    string
pwd                     string
address                 struct<street:string,city:string,state:string,zip:int>

identify                map<int,tinyint>
Time taken: 0.128 seconds, Fetched: 4 row(s)
```

创建分区表

```
create table if not exists usr3(
 name string,
 pwd string,
 address struct<street:string,city:string,state:string,zip:int>,
 identify map<int,tinyint>)
 partitioned by(city string,state string);
```



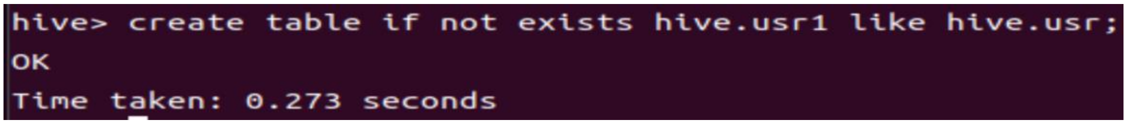
```
hive> create table if not exists usr3(
    >  name string,
    >  pwd string,
    >  address struct<street:string,city:string,state:string,zip:int>,
    >  identify map<int,tinyint>)
    >  partitioned by(city string,state string);
OK
Time taken: 0.176 seconds
```



```
hive> describe hive.usr3;
OK
name                    string
pwd                     string
address                 struct<street:string,city:string,state:string,zip:int>

identify                map<int,tinyint>
city                    string
state                   string

# Partition Information
# col_name              data_type               comment
city                    string
state                   string
Time taken: 0.179 seconds, Fetched: 11 row(s)
```

复制 usr 表的表模式

create table if not exists hive.usr1 like hive.usr;



```
hive> create table if not exists hive.usr1 like hive.usr;
OK
Time taken: 0.273 seconds
```

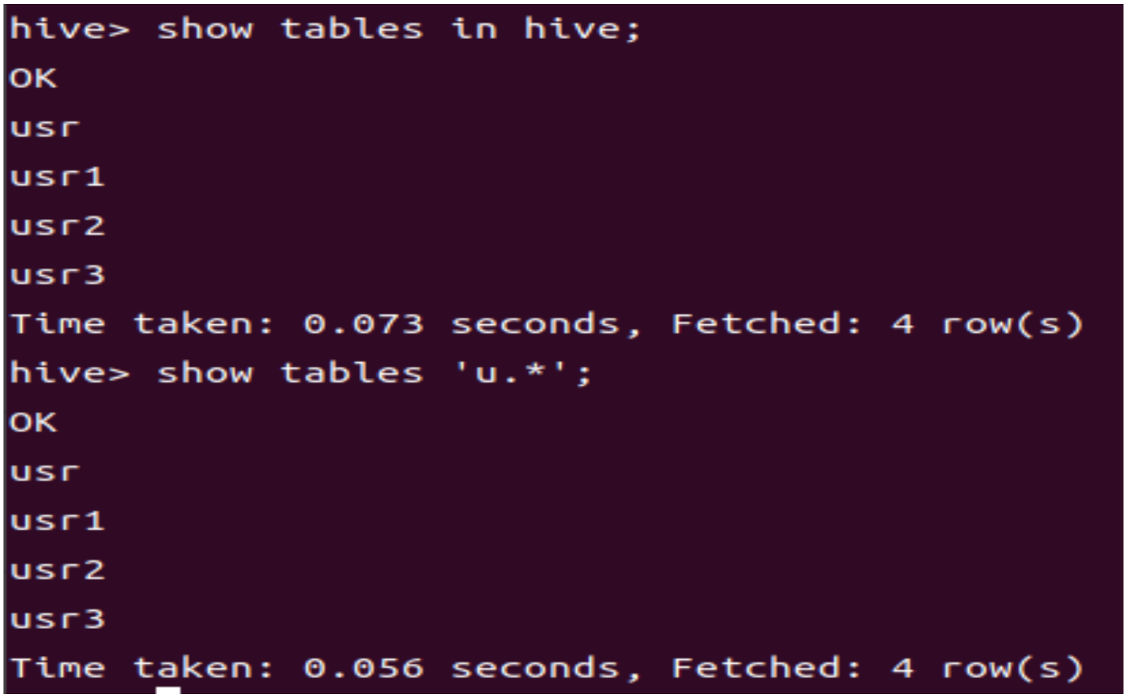
查看表信息

show tables in hive;

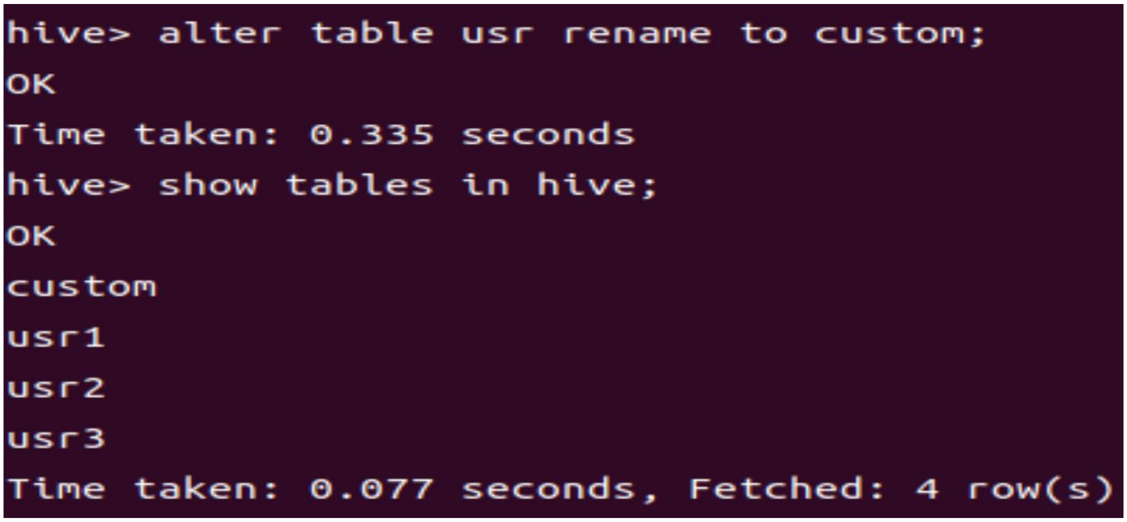
show tables 'u.*'; #查看 hive 中以 u 开头的表

describe hive.usr; #查看 usr 表相关信息

alter table usr rename to custom; #重命名表



```
hive> show tables in hive;
OK
usr
usr1
usr2
usr3
Time taken: 0.073 seconds, Fetched: 4 row(s)
hive> show tables 'u.*';
OK
usr
usr1
usr2
usr3
Time taken: 0.056 seconds, Fetched: 4 row(s)
```



```
hive> alter table usr rename to custom;
OK
Time taken: 0.335 seconds
hive> show tables in hive;
OK
custom
usr1
usr2
usr3
Time taken: 0.077 seconds, Fetched: 4 row(s)
```

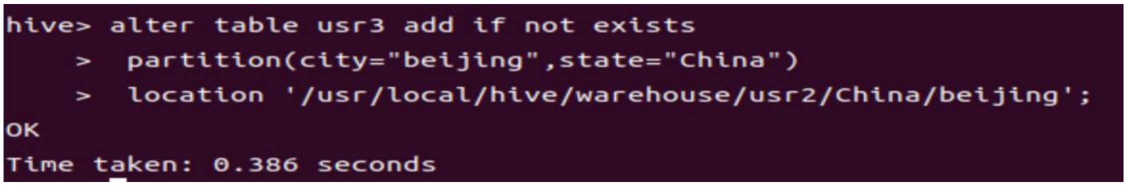
为表增加一个分区

注意刚才新建立的是分区表是 usr3，不是 usr2！！！

alter table usr3 add if not exists

partition(city="beijing",state="China")

location '/usr/local/hive/warehouse/usr2/China/beijing';

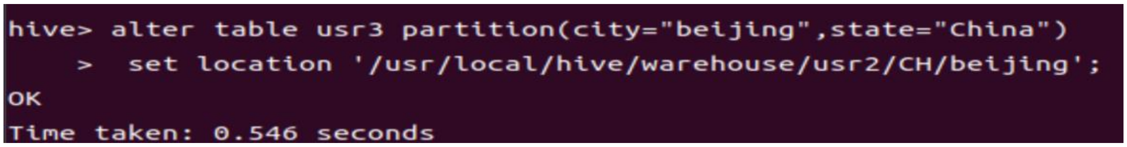


```
hive> alter table usr3 add if not exists
    >  partition(city="beijing",state="China")
    >  location '/usr/local/hive/warehouse/usr2/China/beijing';
OK
Time taken: 0.386 seconds
```

修改分区路径

alter table usr3 partition(city="beijing",state="China")

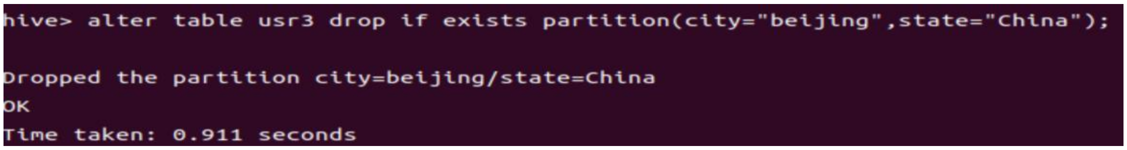
set location '/usr/local/hive/warehouse/usr2/CH/beijing';



```
hive> alter table usr3 partition(city="beijing",state="China")
    >  set location '/usr/local/hive/warehouse/usr2/CH/beijing';
OK
Time taken: 0.546 seconds
```

删除分区

alter table usr3 drop if exists partition(city="beijing",state="China");



```
hive> alter table usr3 drop if exists partition(city="beijing",state="China");

Dropped the partition city=beijing/state=China
OK
Time taken: 0.911 seconds
```

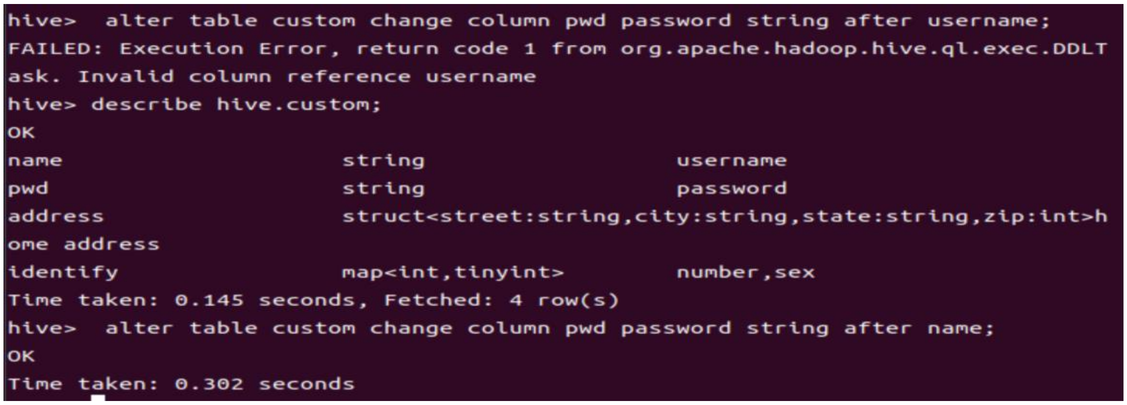
修改列信息

附报错及解决办法：

alter table custom change column pwd password string after name;

- 刚才不是改了表名了吗，记得把表名也改了！

- Failed 的话是因为你用的是 comment 的名字，得用一开始的列名，不能用“nickname”!

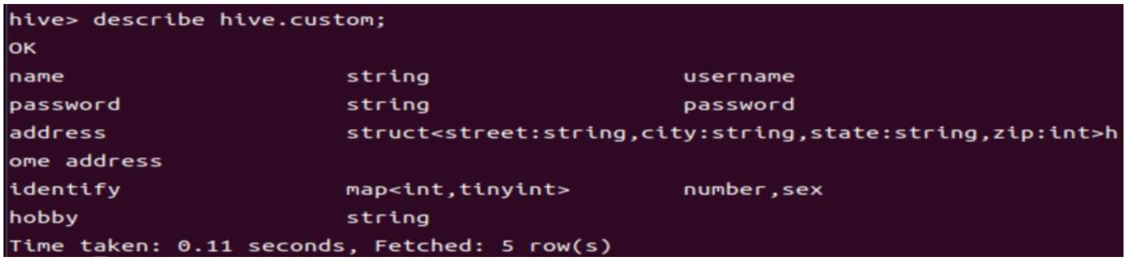


```
hive>  alter table custom change column pwd password string after username;
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLT
ask. Invalid column reference username
hive> describe hive.custom;
OK
name                    string                  username
pwd                     string                  password
address                 struct<street:string,city:string,state:string,zip:int>h
ome address
identify                map<int,tinyint>        number,sex
Time taken: 0.145 seconds, Fetched: 4 row(s)
hive>  alter table custom change column pwd password string after name;
OK
Time taken: 0.302 seconds
```

增加列

alter table custom add columns(hobby string); #增加列

use hive; #切换到 hive 数据库下

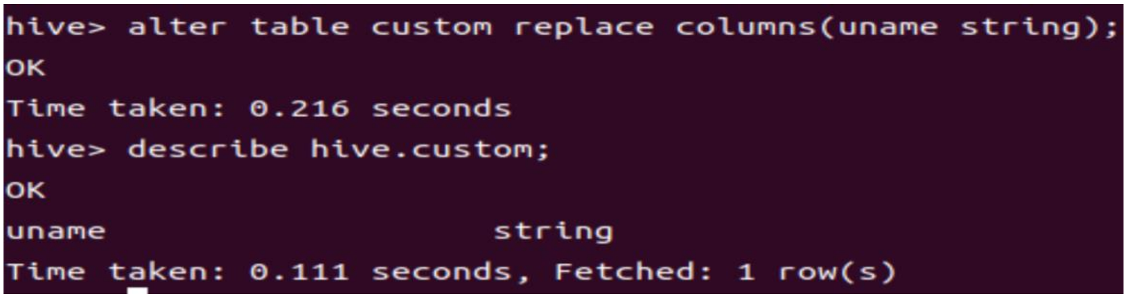


```
hive> describe hive.custom;
OK
name                    string                  username
password                string                  password
address                 struct<street:string,city:string,state:string,zip:int>h
ome address
identify                map<int,tinyint>        number,sex
hobby                   string
Time taken: 0.11 seconds, Fetched: 5 row(s)
```

删除替换列

就是说把所有的列删除，并替换成这个列。

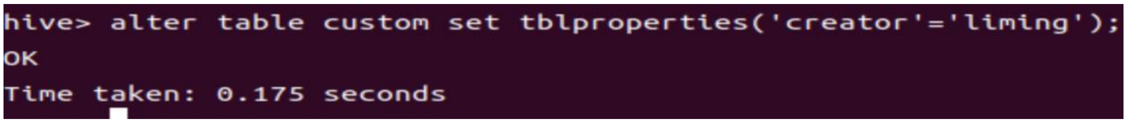
alter table custom replace columns(uname string); #删除替换列



```
hive> alter table custom replace columns(uname string);
OK
Time taken: 0.216 seconds
hive> describe hive.custom;
OK
uname                           string
Time taken: 0.111 seconds, Fetched: 1 row(s)
```

修改表属性

alter table custom set tblproperties('creator'='liming'); #修改表属性



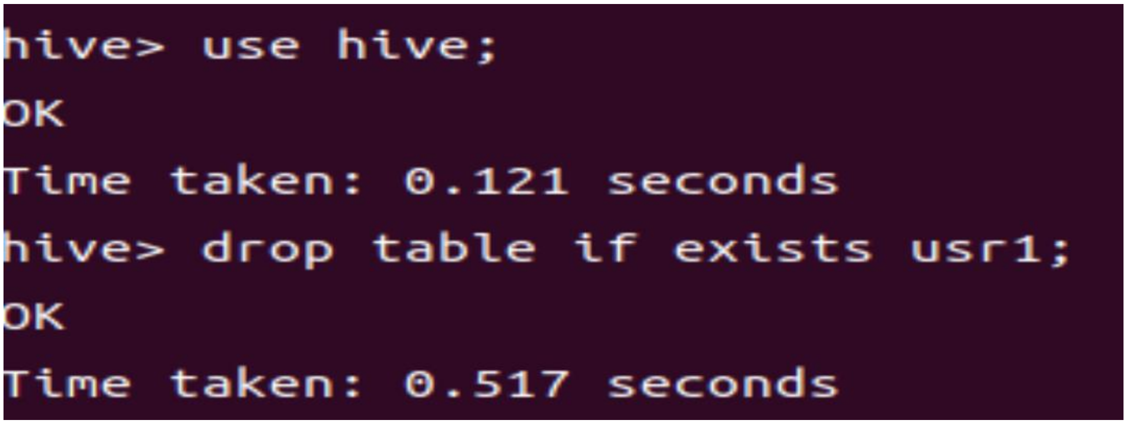
```
hive> alter table custom set tblproperties('creator'='liming');
OK
Time taken: 0.175 seconds
```

修改存储属性

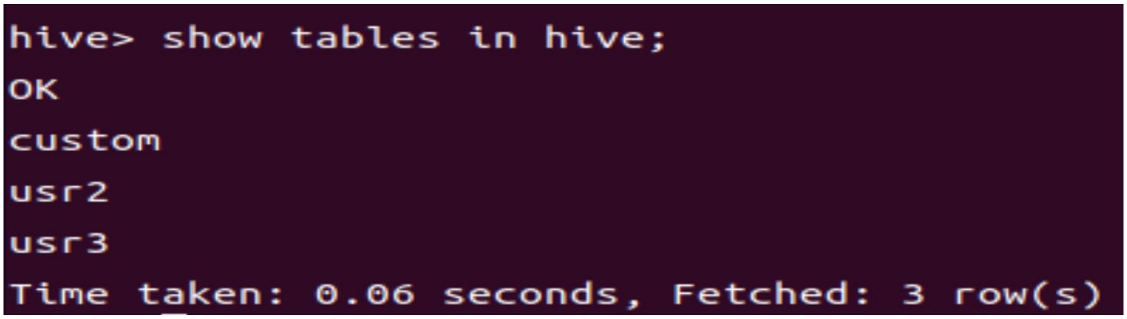
alter table usr2 partition(city="beijing",state="China"); #修改存储属性

删除表

drop table if exists usr1; #删除表



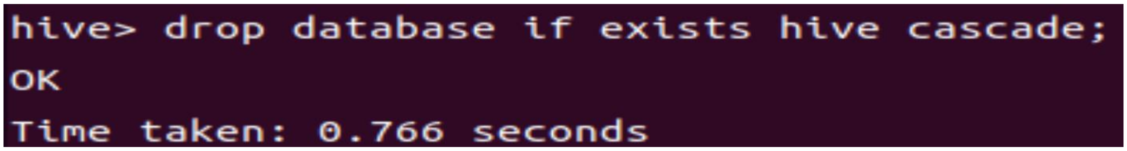
```
hive> use hive;
OK
Time taken: 0.121 seconds
hive> drop table if exists usr1;
OK
Time taken: 0.517 seconds
```



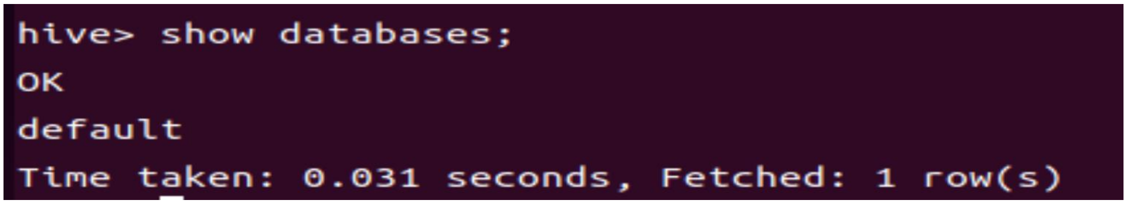
```
hive> show tables in hive;
OK
custom
usr2
usr3
Time taken: 0.06 seconds, Fetched: 3 row(s)
```

删除数据库及其表

drop database if exists hive cascade; #删除数据库和它中的表



```
hive> drop database if exists hive cascade;
OK
Time taken: 0.766 seconds
```



```
hive> show databases;
OK
default
Time taken: 0.031 seconds, Fetched: 1 row(s)
```

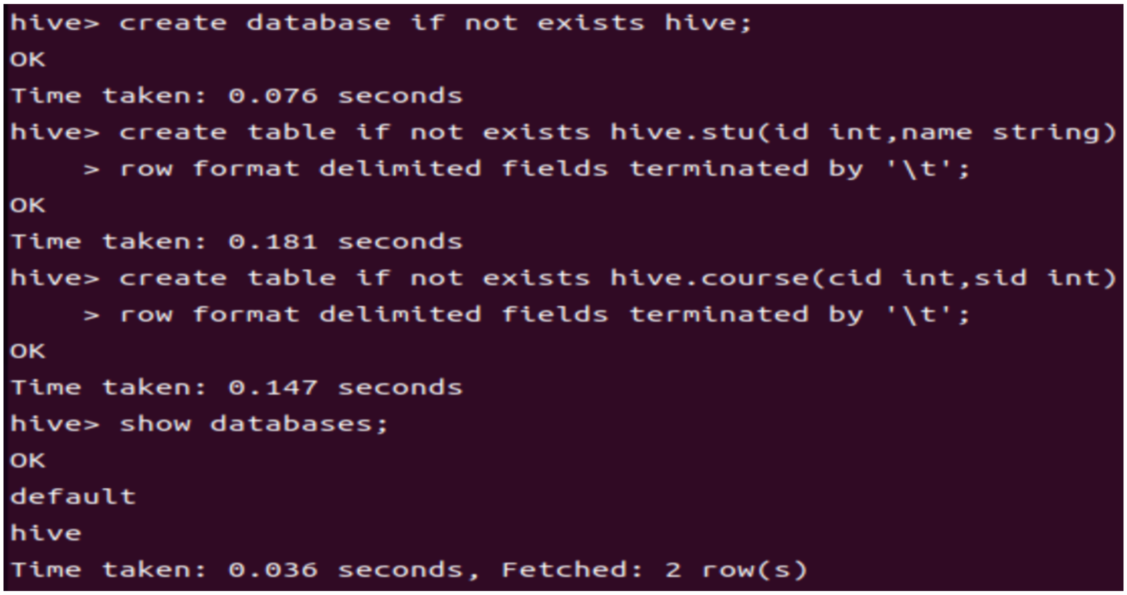
添加数据

create table if not exists hive.stu(id int,name string)

row format delimited fields terminated by '\t';

create table if not exists hive.course(cid int,sid int)

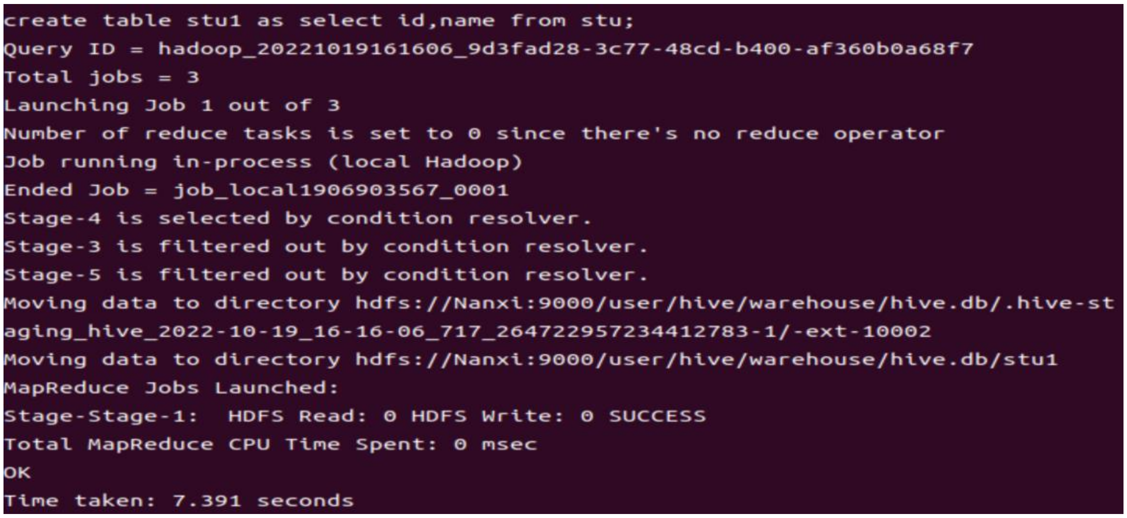
row format delimited fields terminated by '\t';



```
hive> create database if not exists hive;
OK
Time taken: 0.076 seconds
hive> create table if not exists hive.stu(id int,name string)
    > row format delimited fields terminated by '\t';
OK
Time taken: 0.181 seconds
hive> create table if not exists hive.course(cid int,sid int)
    > row format delimited fields terminated by '\t';
OK
Time taken: 0.147 seconds
hive> show databases;
OK
default
hive
Time taken: 0.036 seconds, Fetched: 2 row(s)
```

用 stu 表创建 stu1 表

create table stu1 as select id,name from stu;



```
create table stu1 as select id,name from stu;
Query ID = hadoop_20221019161606_9d3fad28-3c77-48cd-b400-af360b0a68f7
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
Ended Job = job_local1906903567_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://Nanxi:9000/user/hive/warehouse/hive.db/.hive-st
aging_hive_2022-10-19_16-16-06_717_264722957234412783-1/-ext-10002
Moving data to directory hdfs://Nanxi:9000/user/hive/warehouse/hive.db/stu1
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 0 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 7.391 seconds
```

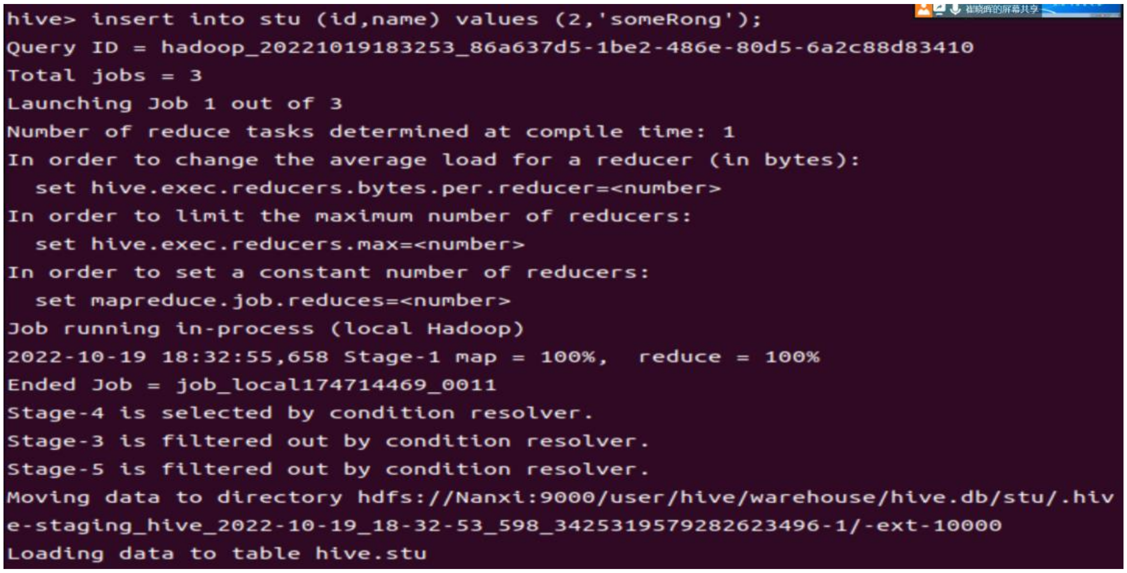
添加数据

insert into stu (id,name) values (1,'Nanxi');

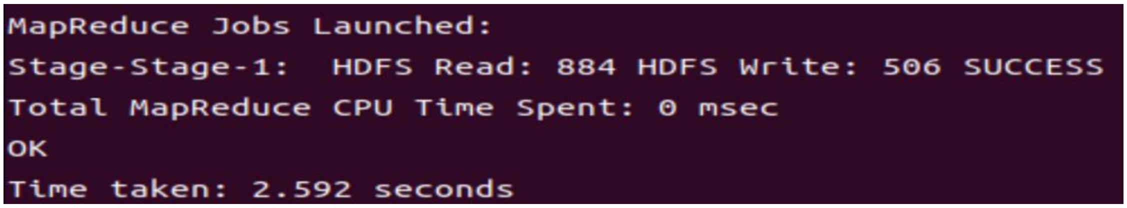
insert into stu (id,name) values (2,'someRong');

insert into stu (id,name) values (3,'Light');

```
insert into stu (id,name) values (4,'Xueying');
insert into stu (id,name) values (5,'BigGrandpa');
insert into stu (id,name) values (6,'WhiteBird');
insert into stu (id,name) values (7,'Eating');
insert into stu (id,name) values (8,'Seasea');

select * from stu;

insert into course (cid,sid) values (37,1);
insert into course (cid,sid) values (36,2);
insert into course (cid,sid) values (37,3);
insert into course (cid,sid) values (27,4);
insert into course (cid,sid) values (30,5);
insert into course (cid,sid) values (32,6);
insert into course (cid,sid) values (21,7);
insert into course (cid,sid) values (30,8);

select * from course;
```

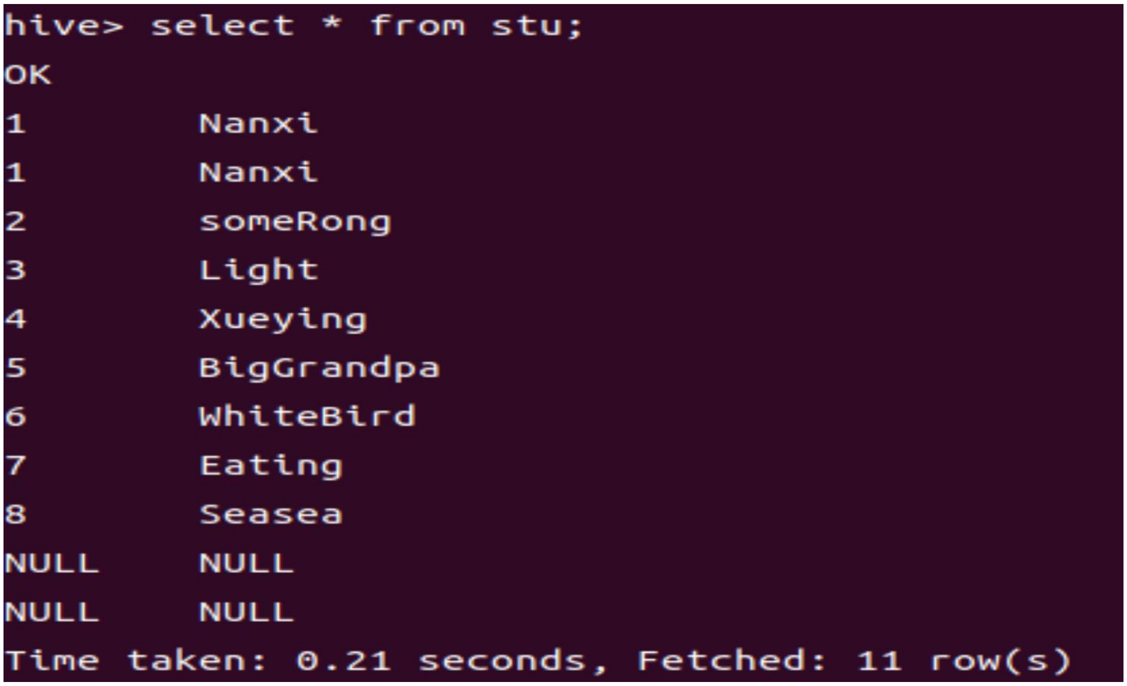


```
hive> insert into stu (id,name) values (2,'someRong');
Query ID = hadoop_20221019183253_86a637d5-1be2-486e-80d5-6a2c88d83410
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2022-10-19 18:32:55,658 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local174714469_0011
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://Nanxi:9000/user/hive/warehouse/hive.db/stu/.hiv
e-staging_hive_2022-10-19_18-32-53_598_3425319579282623496-1/-ext-10000
Loading data to table hive.stu
```



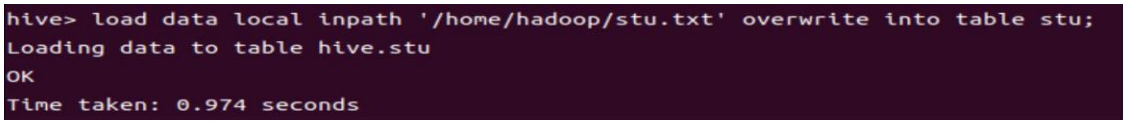
```
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 884 HDFS Write: 506 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
Time taken: 2.592 seconds
```

插入如下：



```
hive> select * from stu;
OK
1       Nanxi
1       Nanxi
2       someRong
3       Light
4       Xueying
5       BigGrandpa
6       WhiteBird
7       Eating
8       Seasea
NULL    NULL
NULL    NULL
Time taken: 0.21 seconds, Fetched: 11 row(s)
```

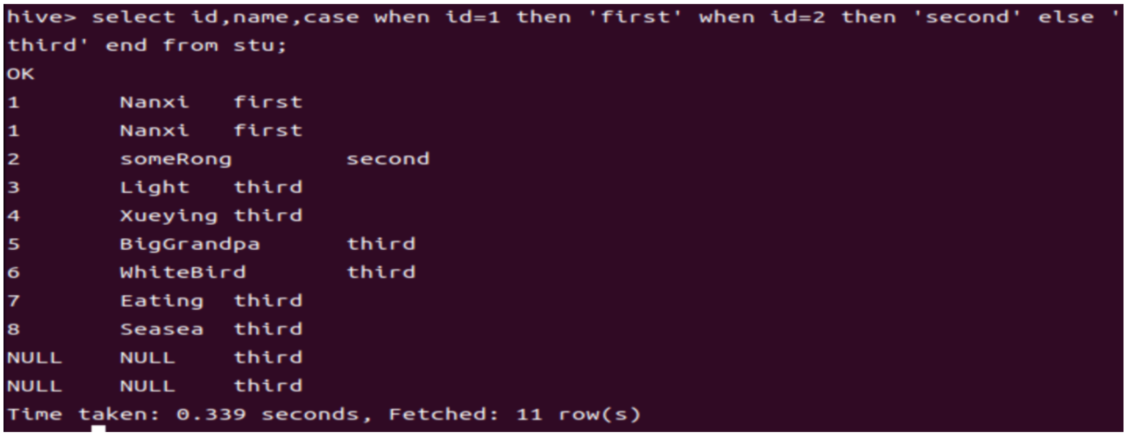
load data local inpath '/home/hadoop/stu.txt' overwrite into table stu;



```
hive> load data local inpath '/home/hadoop/stu.txt' overwrite into table stu;
Loading data to table hive.stu
OK
Time taken: 0.974 seconds
```

查看数据

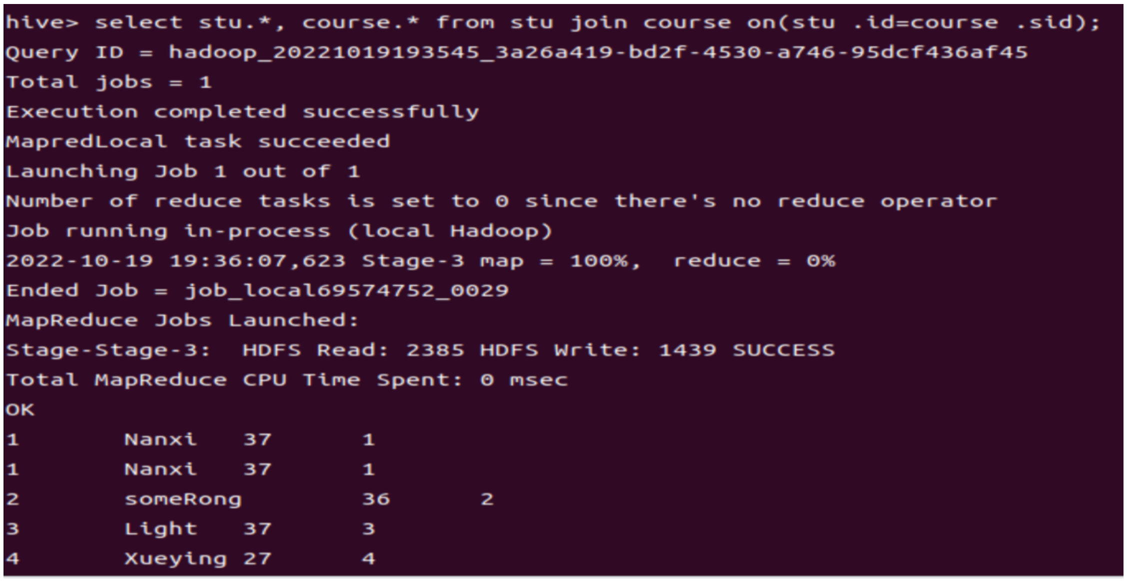
select id,name,case when id=1 then 'first' when id=2 then 'second' else 'third' end from stu;



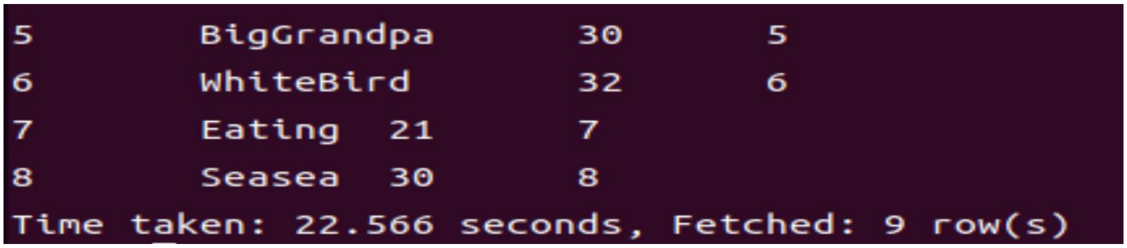
```
hive> select id,name,case when id=1 then 'first' when id=2 then 'second' else '
third' end from stu;
OK
1       Nanxi   first
1       Nanxi   first
2       someRong        second
3       Light   third
4       Xueying third
5       BigGrandpa      third
6       WhiteBird       third
7       Eating  third
8       Seasea  third
NULL    NULL    third
NULL    NULL    third
Time taken: 0.339 seconds, Fetched: 11 row(s)
```

连接

select stu.*, course.* from stu join course on(stu .id=course .sid);



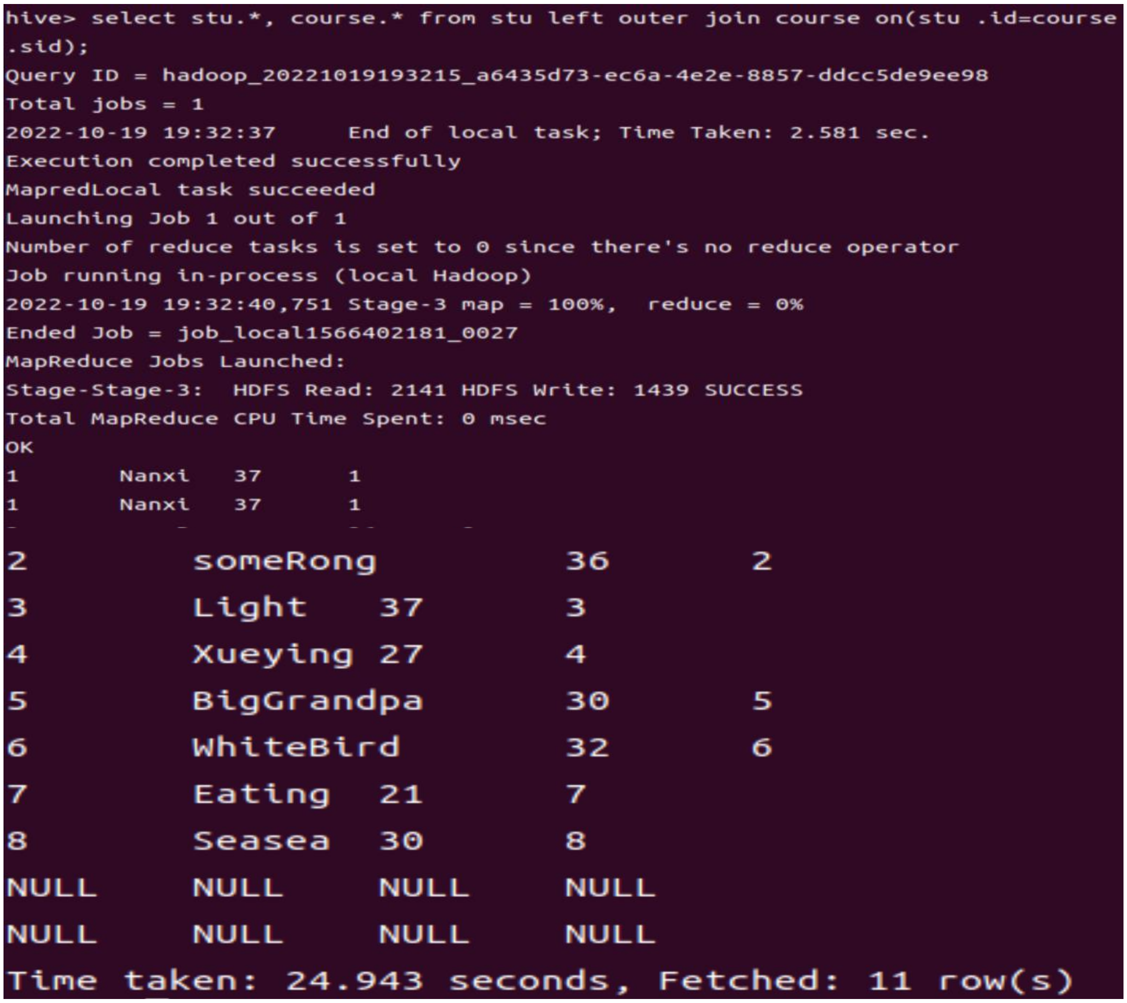
```
hive> select stu.*, course.* from stu join course on(stu .id=course .sid);
Query ID = hadoop_20221019193545_3a26a419-bd2f-4530-a746-95dcf436af45
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2022-10-19 19:36:07,623 Stage-3 map = 100%,  reduce = 0%
Ended Job = job_local69574752_0029
MapReduce Jobs Launched:
Stage-Stage-3:  HDFS Read: 2385 HDFS Write: 1439 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1       Nanxi   37      1
1       Nanxi   37      1
2       someRong        36      2
3       Light   37      3
4       Xueying 27      4
```



```
5       BigGrandpa      30      5
6       WhiteBird       32      6
7       Eating  21      7
8       Seasea  30      8
Time taken: 22.566 seconds, Fetched: 9 row(s)
```

左外连接

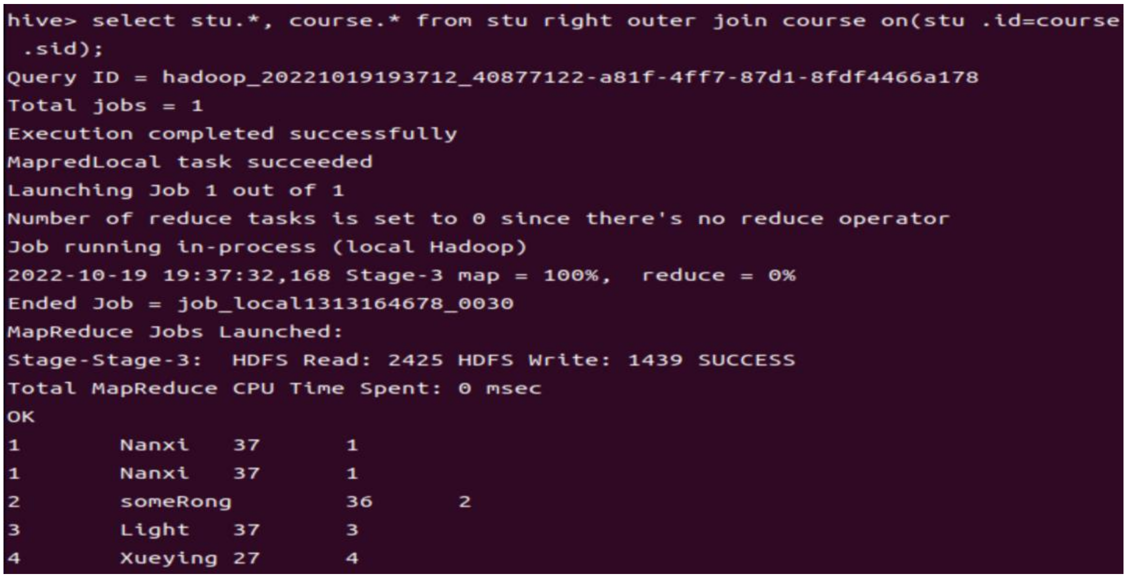
select stu.*, course.* from stu left outer join course on(stu .id=course .sid);



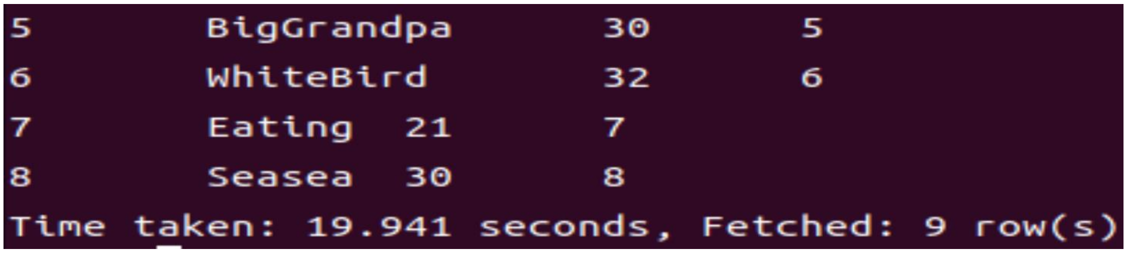
```
hive> select stu.*, course.* from stu left outer join course on(stu .id=course
.sid);
Query ID = hadoop_20221019193215_a6435d73-ec6a-4e2e-8857-ddcc5de9ee98
Total jobs = 1
2022-10-19 19:32:37     End of local task; Time Taken: 2.581 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2022-10-19 19:32:40,751 Stage-3 map = 100%,  reduce = 0%
Ended Job = job_local1566402181_0027
MapReduce Jobs Launched:
Stage-Stage-3:  HDFS Read: 2141 HDFS Write: 1439 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1       Nanxi   37      1
1       Nanxi   37      1
2       someRong        36      2
3       Light   37      3
4       Xueying 27      4
5       BigGrandpa      30      5
6       WhiteBird       32      6
7       Eating  21      7
8       Seasea  30      8
NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL
Time taken: 24.943 seconds, Fetched: 11 row(s)
```

右外连接

select stu.*, course.* from stu right outer join course on(stu .id=course .sid);



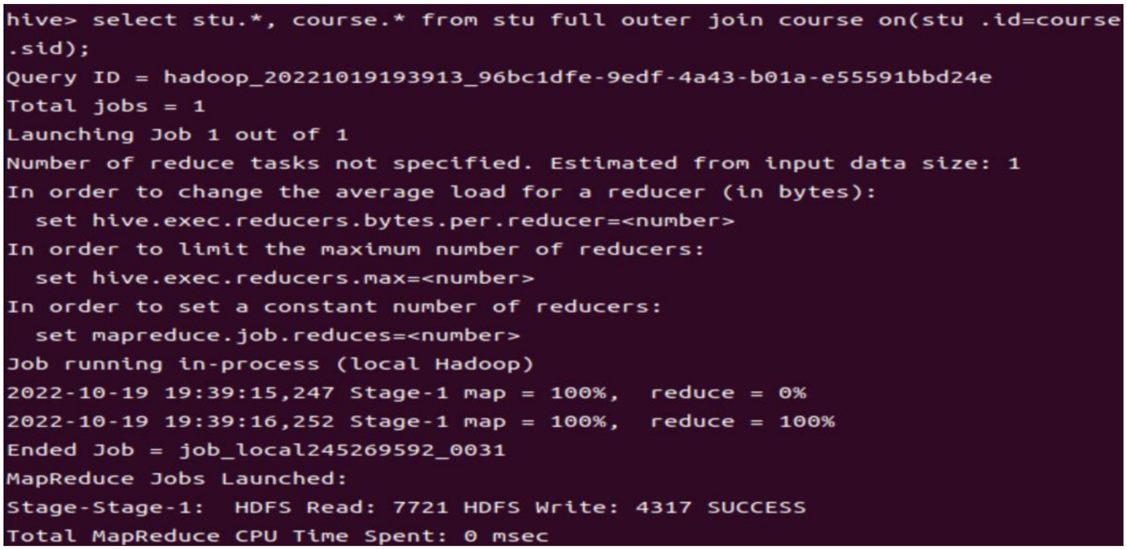
```
hive> select stu.*, course.* from stu right outer join course on(stu .id=course
 .sid);
Query ID = hadoop_20221019193712_40877122-a81f-4ff7-87d1-8fdf4466a178
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Job running in-process (local Hadoop)
2022-10-19 19:37:32,168 Stage-3 map = 100%,  reduce = 0%
Ended Job = job_local1313164678_0030
MapReduce Jobs Launched:
Stage-Stage-3:  HDFS Read: 2425 HDFS Write: 1439 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
OK
1       Nanxi   37      1
1       Nanxi   37      1
2       someRong        36      2
3       Light   37      3
4       Xueying 27      4
```



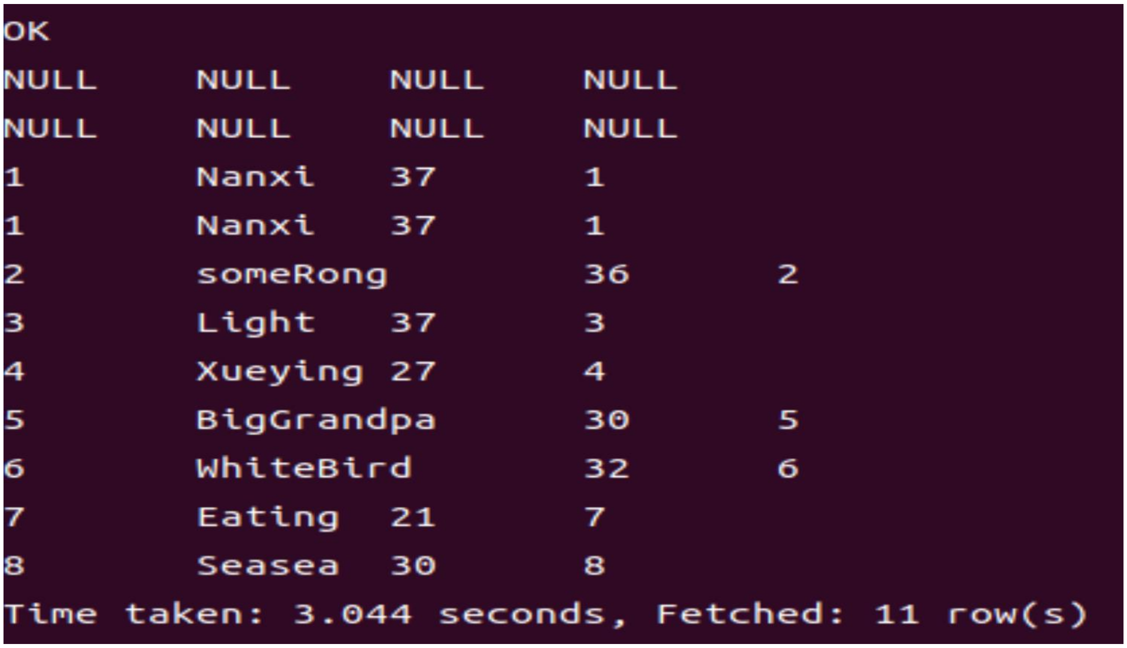
```
5       BigGrandpa      30      5
6       WhiteBird       32      6
7       Eating  21      7
8       Seasea  30      8
Time taken: 19.941 seconds, Fetched: 9 row(s)
```

full outer join

select stu.*, course.* from stu full outer join course on(stu .id=course .sid);



```
hive> select stu.*, course.* from stu full outer join course on(stu .id=course
.sid);
Query ID = hadoop_20221019193913_96bc1dfe-9edf-4a43-b01a-e55591bbd24e
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Job running in-process (local Hadoop)
2022-10-19 19:39:15,247 Stage-1 map = 100%,  reduce = 0%
2022-10-19 19:39:16,252 Stage-1 map = 100%,  reduce = 100%
Ended Job = job_local245269592_0031
MapReduce Jobs Launched:
Stage-Stage-1:  HDFS Read: 7721 HDFS Write: 4317 SUCCESS
Total MapReduce CPU Time Spent: 0 msec
```



```
OK
NULL    NULL    NULL    NULL
NULL    NULL    NULL    NULL
1       Nanxi   37      1
1       Nanxi   37      1
2       someRong        36      2
3       Light   37      3
4       Xueying 27      4
5       BigGrandpa      30      5
6       WhiteBird       32      6
7       Eating  21      7
8       Seasea  30      8
Time taken: 3.044 seconds, Fetched: 11 row(s)
```

出现的问题及总结：

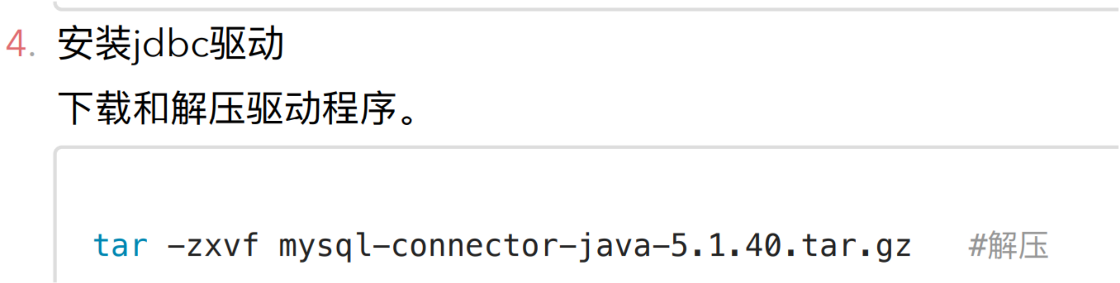
教程的错误或可改正地方如下：

1.hive-site.xml 文件中，应该加上&useSSL=false，否则会出现 warning 提示，虽然不影响使用但是比较麻烦：

```
<property>

    <name>javax.jdo.option.ConnectionURL</name>
```

```
<value>jdbc:mysql://localhost:3306/hive?createDatabaseIfNotExist=true&amp;useSSL=false</value>
   <description>JDBC connect string for a JDBC metastore</description>
 </property>
```

2.提供的包和安装的包不匹配，不能使用提供的 tar 包，必须自己下载 tar.gz 包并解压

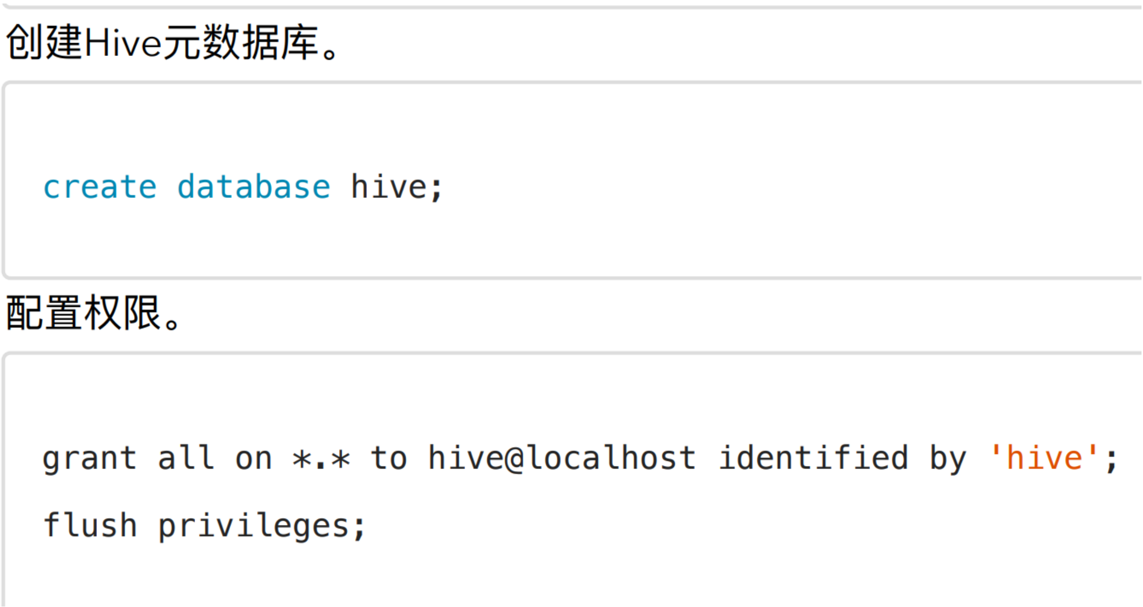


4. 安装jdbc驱动

下载和解压驱动程序。

```
tar -zxvf mysql-connector-java-5.1.40.tar.gz   #解压
```

3.这两步之前加上"先启动 mysql，登录后在 mysql 中执行下列操作"——mysql 安装后可能会有初始密码，不过这个自己想想搜一下能解决，不用写到教程里



创建Hive元数据库。

```
create database hive;
```

配置权限。

```
grant all on *.* to hive@localhost identified by 'hive';
flush privileges;
```

4.在 mysql 中配置权限步骤出错，应该这么写：

```
CREATE USER 'hive'@'localhost' IDENTIFIED BY 'hive';

GRANT ALL ON *.* TO 'hive'@'localhost';
```

启动 hive 报错：……Exception in thread "main" java.lang.NoSuchMethodError: com.google.common.base.Preconditions.checkArgument(ZLjava/lang/String;Ljava/lang/Object;)V……

：删除 hive 中低版本的 guava-14.0.1.jar 包，将 hadoop 中的 **guava-27.0-jre.jar** 复制到 hive 的 lib 目录下即可。

Exception in thread "main" java.lang.RuntimeException: com.ctc.wstx.exc.WstxUnexpectedCharException: Unexpected character (CTRL-CHAR, code 10); missing element name?
at [row,col,system-id]: [6,77,"file:/usr/local/hive/conf/hive-site.xml"]

按报错打开一看，发现是复制粘贴老师给的 xml 文件时，格式发生了变化导致错误，</value>标签中的/和 value 不能有空格，到对应的行列删了之后会发现标签的</value>变为蓝色，那样就改对了，再 ./bin/hive 即可。

5.sql 语句错误：

创建内部表正确语法：

```
create table if not exists hive.usr(

name string comment 'username',

pwd string comment 'password',

address struct<street:string,city:string,state:string,zip:int> comment 'home address',

identify map<int,tinyint> comment 'number,sex')

comment 'description of the table' tblproperties('creator'='me','time'='2016.1.1');
```

```
alter table usr3 add if not exists

    partition(city="beijing",state="China")

    location '/usr/local/hive/warehouse/usr3/China/beijing'; #为表3增加一个分区，表2没有分区属性

alter table usr3 drop if exists  partition(city="beijing",state="China");#注意换成英文标点
```