

1. 在 hxl 数据库下创建表

```
hive> create table tb_emp_info
> (id int,
> name string,
> age int,
> tel string)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY '|'
> STORED AS TEXTFILE;
```

OK

Time taken: 0.296 seconds

```
hive> show tables in hxl;
```

OK

tb_emp_info

Time taken: 0.073 seconds

2. 准备加载数据

```
[hadoop1@node1 hive]$ more tb_emp_info.txt
```

```
1|name1|25|13188888888888
2|name2|30|13888888888888
3|name3|3|147896221
4|name4|56|899314121
5|name5|12|899314121
6|name6|9|899314121
7|name7|32|899314121
8|name8|42|158964
9|name9|86|899314121
10|name10|45|789541
```

3. 本地系统加载文件数据

进入到 tb_emp_info.txt 文件所在的目录, 然后执行 hive 进入到 hive 模式

```
[hadoop1@node1 hive]$ hive
```

```
hive> use hxl;
```

OK

Time taken: 0.103 seconds

```
hive> load data local inpath 'tb_emp_info.txt' into table tb_emp_info;
```

Copying data from file:/home/hadoop1/file/hive/tb_emp_info.txt

Copying file: file:/home/hadoop1/file/hive/tb_emp_info.txt

Loading data to table hxl.tb_emp_info

OK

Time taken: 0.694 seconds

若是分区表的话，需要指点导入的分区，如：

```
hive> load data local inpath 'login.txt' into table tb_sso_ver_login_day
partition(statedate=20141201);
```

4. 查看加载进去的数据

```
hive> select * from tb_emp_info;
```

OK

1	name1	25	131888888888888
2	name2	30	138888888888888
3	name3	3	147896221
4	name4	56	899314121
5	name5	12	899314121
6	name6	9	899314121
7	name7	32	899314121
8	name8	42	158964
9	name9	86	899314121
10	name10	45	789541

5. 可以进入到 hdfs 目录下查看该表对应的文件

```
hive> dfs -ls /user/hive/warehouse/hxl.db/tb_emp_info;
```

Found 1 items

```
-rw-r--r--      3 hadoop1 supergroup          214 2014-10-28 17:31
/user/hive/warehouse/hxl.db/tb_emp_info/tb_emp_info.txt
```

HDFS 文件导入到 Hive 表

1. 查看 hdfs 系统上的文件

```
$hadoop fs -cat /user/hadoop1/myfile/tb_class.txt
```

输出部分

```
0|班级 0|2014-10-29 14:10:17|2014-10-29 14:10:17
1|班级 1|2014-10-29 14:10:17|2014-10-29 14:10:17
2|班级 2|2014-10-29 14:10:17|2014-10-29 14:10:17
3|班级 3|2014-10-29 14:10:17|2014-10-29 14:10:17
4|班级 4|2014-10-29 14:10:17|2014-10-29 14:10:17
5|班级 5|2014-10-29 14:10:17|2014-10-29 14:10:17
6|班级 6|2014-10-29 14:10:17|2014-10-29 14:10:17
7|班级 7|2014-10-29 14:10:17|2014-10-29 14:10:17
8|班级 8|2014-10-29 14:10:17|2014-10-29 14:10:17
```

2. 创建表

```
create table tb_class_info
```

```
(id int,  
class_name string,  
createtime timestamp ,  
modifytime timestamp)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY '|'   
STORED AS TEXTFILE;
```

3. 导入表

```
load data inpath '/user/hadoop1/myfile/tb_class.txt' into table tb_class_info;
```

导出到本地系统

将 hive 表数据导出到 os 目录/home/hadoop1/file/from_hive, 导出后会在 tb_emp_info 目录下生成文件名为 000000_0 的文件, 好像不能指定具体的文件名导出

```
hive> insert overwrite local directory '/home/hadoop1/file/tb_emp_info' select  
* from tb_emp_info;
```

导出到 hdfs 系统

```
hive> insert overwrite directory '/home/hadoop1/hive_output/tb_emp_info' select  
* from tb_emp_info;
```

Total MapReduce jobs = 2

Launching Job 1 out of 2

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job_201410300915_0003, Tracking URL =

http://node1:50030/jobdetails.jsp?jobid=job_201410300915_0003

Kill Command = /usr1/hadoop/libexec/./bin/hadoop

job -Dmapred.job.tracker=http://192.168.56.101:9001 -kill

job_201410300915_0003

Hadoop job information for Stage-1: number of mappers: 1; number of reducers:
0

2014-10-30 10:35:42,580 Stage-1 map = 0%, reduce = 0%

2014-10-30 10:35:44,595 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 0.38
sec

2014-10-30 10:35:45,603 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 0.38
sec

MapReduce Total cumulative CPU time: 380 msec

Ended Job = job_201410300915_0003

Ended Job = -926438869, job is filtered out (removed at runtime).

Moving data to:

hdfs://192.168.56.101:9000/tmp/hive-hadoop1/hive_2014-10-30_10-35-38_612_637
7994640941387871/-ext-10000

```
Moving data to: /home/hadoop1/hive_output/tb_emp_info
Failed with exception Unable to rename:
hdfs://192.168.56.101:9000/tmp/hive-hadoop1/hive_2014-10-30_10-35-38_612_637
7994640941387871/-ext-10000 to: /home/hadoop1/hive_output/tb_emp_info
FAILED: Execution Error, return code 1 from
org.apache.hadoop.hive.ql.exec.MoveTask
MapReduce Jobs Launched:
Job 0: Map: 1      Cumulative CPU: 0.38 sec      HDFS Read: 768 HDFS Write: 428
SUCCESS
Total MapReduce CPU Time Spent: 380 msec
```

解决办法:

将 hive.insert.into.multilevel.dirs 参数的值由默认 false 修改为 true

```
hive.insert.into.multilevel.dirs
true
```