

解决Hive中show create table乱码问题

我目前使用的Hive版本是apache-hive-1.2.0-bin，每次在使用 show create table 语句的时候如果你字段中有中文注释，那么Hive得出来的结果如下：

```
hive> show create table iteblog;
OK
CREATE TABLE `iteblog` (
  `id` bigint COMMENT 'id',
  `uid` bigint COMMENT '(7id',
  `name` string COMMENT '(7
)
ROW FORMAT SERDE
  'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
  'hdfs://user/iteblog/hive/warehouse/dw_source_log.db/iteblog'
TBLPROPERTIES (
  'transient_lastDdlTime'='1465380093')
Time taken: 0.52 seconds, Fetched: 15 row(s)
```

COMMENT 后面的东西变成了乱码了。。。是我Mysql中元数据字符集不对吗？其实不是，罪魁祸首

首是

Hive内部

的代码没有处理好

中文。从网上搜索解决方法看到了这

个Issue给了一个方法：<https://issues.apache.org/jira/browse/HIVE-11837>

，其中的HIVE-11837.1.patch就是具体方法，我们可以修改ql/src/java/org/apache/hadoop/hive/ql/exec/DDLTask.java类中的两句代码：

```
diff --git a/ql/src/java/org/apache/hadoop/hive/ql/exec/DDLTask.java b/ql/src/java/org/apache/hadoop/hive/ql/exec/DDLTask.java
index 6fca9f7ec86574a6053af3672c551c6a63aa4870..661367f27b69f9796140808eda53a3bbcdcbdb11 100644
--- a/ql/src/java/org/apache/hadoop/hive/ql/exec/DDLTask.java
+++ b/ql/src/java/org/apache/hadoop/hive/ql/exec/DDLTask.java
```

```

@@ -2048,7 +2048,7 @@ private int showCreateTable(Hive db, DataOutputStream outputStream,
String tableName

    if (tbl.isView()) {
        String createTab_stmt = "CREATE VIEW `" + tableName + "` AS " + tbl.getViewExpandedTe
xt();
-        outputStream.writeBytes(createTab_stmt.toString());
+        outputStream.write(createTab_stmt.toString().getBytes("UTF-8"));
        return 0;
    }

@@ -2196,7 +2196,7 @@ else if (sortCol.getOrder() == BaseSemanticAnalyzer.HIVE_COLUMN_
ORDER_DESC) {
    }
    createTab_stmt.add(TBL_PROPERTIES, tbl_properties);

-    outputStream.writeBytes(createTab_stmt.render());
+    outputStream.write(createTab_stmt.render().getBytes("UTF-8"));
} catch (IOException e) {
    LOG.info("show create table: " + stringfyException(e));
    return 1;
}

```

也就是使用+号标识的代码替换-号标识的代码（因为这个patch并不是基于apache-hive-1.2.0-src打的，所以你可以看到apache-hive-1.2.0-src中的代码和上面略有不同），然后我们使用[《Hive 1.2.1源码编译》](#)文中介绍的方法去编译修改好的代码。编译成功之后，使用编译的hive-exec-1.2.0.jar替换你之前的hive-exec-1.2.0.jar文件即可，然后重启Hive客户端，现在来show create table xx：

```

hive> show create table iteblog;
OK
CREATE TABLE `iteblog` (
  `id` bigint COMMENT '自增id',
  `uid` bigint COMMENT '用户id',
  `name` string COMMENT '用户姓名')
ROW FORMAT SERDE
  'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'
STORED AS INPUTFORMAT
  'org.apache.hadoop.mapred.TextInputFormat'
OUTPUTFORMAT
  'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
LOCATION
  'hdfs://user/iteblog/hive/warehouse/dw_source_log.db/iteblog'
TBLPROPERTIES (

```

```
'transient_lastDdlTime'='1465380093')  
Time taken: 0.075 seconds, Fetched: 14 row(s)
```

可以看到，已经不乱码了,终于可以不被乱码影响工作了。

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: [【】（）](#)