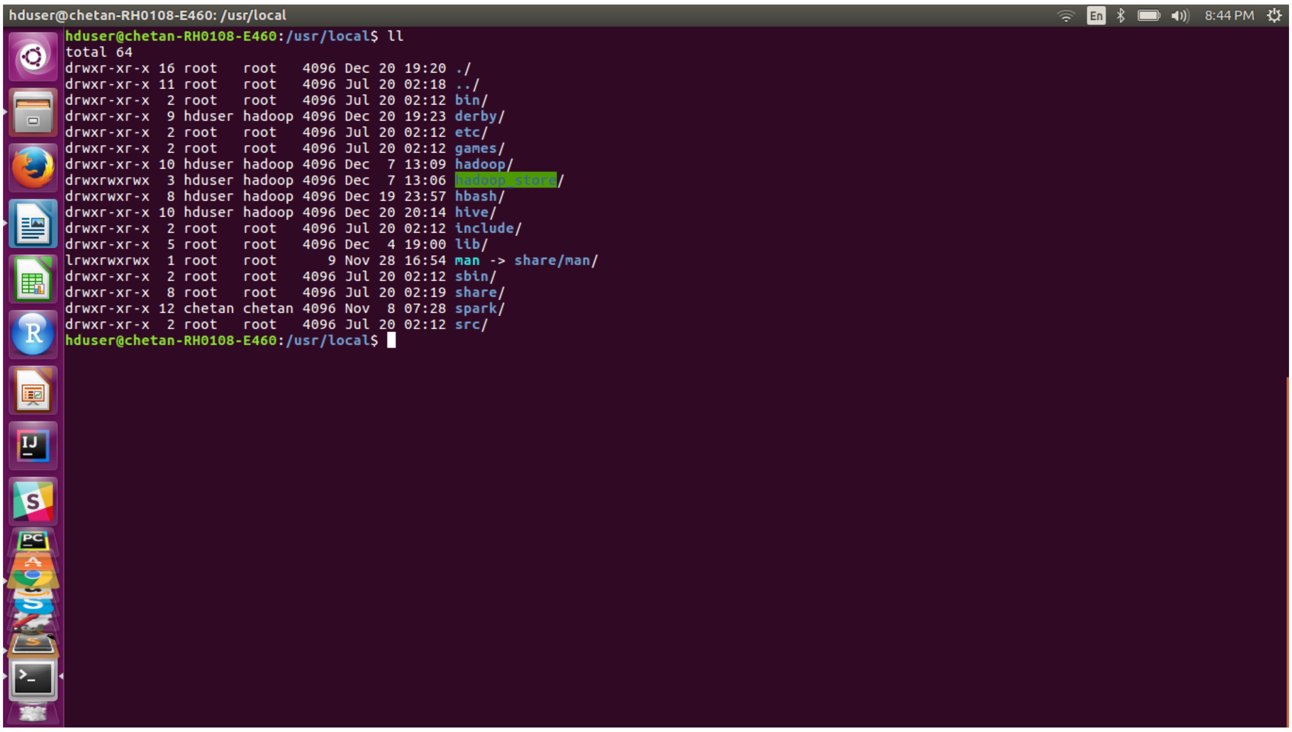
Apache Hive 2.0.1 Configuration with Ubuntu 14.x / 16.x Documentation

Prerequisites:
Apache Hadoop 2.6.x / 2.7.x must be configured and running.

1. Download Apache Hive 2.0.1

sudo wget http://redrockdigimark.com/apachemirror/hive/hive-2.0.1/apache-hive-2.0.1-bin.tar.gz
sudo tar zxvf apache-hive-2.0.1-bin.tar.gz
sudo mv apache-hive-2.0.1-bin /usr/local/hive
sudo chown hadoop:hduser hive



```
hduser@chetan-RH0108-E460:/usr/local$ ll
total 64
drwxr-xr-x 16 root   root   4096 Dec 20 19:20 ./
drwxr-xr-x 11 root   root   4096 Jul 20 02:18 ../
drwxr-xr-x  2 root   root   4096 Jul 20 02:12 bin/
drwxr-xr-x  9 hduser hadoop 4096 Dec 20 19:23 derby/
drwxr-xr-x  2 root   root   4096 Jul 20 02:12 etc/
drwxr-xr-x  2 root   root   4096 Jul 20 02:12 games/
drwxr-xr-x 10 hduser hadoop 4096 Dec  7 13:09 hadoop/
drwxrwxrwx  3 hduser hadoop 4096 Dec  7 13:06 hadoop_store/
drwxrwxr-x  8 hduser hadoop 4096 Dec 19 23:57 hbash/
drwxr-xr-x 10 hduser hadoop 4096 Dec 20 20:14 hive/
drwxr-xr-x  2 root   root   4096 Jul 20 02:12 include/
drwxr-xr-x  5 root   root   4096 Dec  4 19:00 lib/
lrwxrwxrwx  1 root   root      9 Nov 28 16:54 man -> share/man/
drwxr-xr-x  2 root   root   4096 Jul 20 02:12 sbin/
drwxr-xr-x  8 root   root   4096 Jul 20 02:19 share/
drwxr-xr-x 12 chetan chetan 4096 Nov  8 07:28 spark/
drwxr-xr-x  2 root   root   4096 Jul 20 02:12 src/
hduser@chetan-RH0108-E460:/usr/local$ 
```

Figure 1: *Hadoop – Hive OS User and Group Level Permissions*

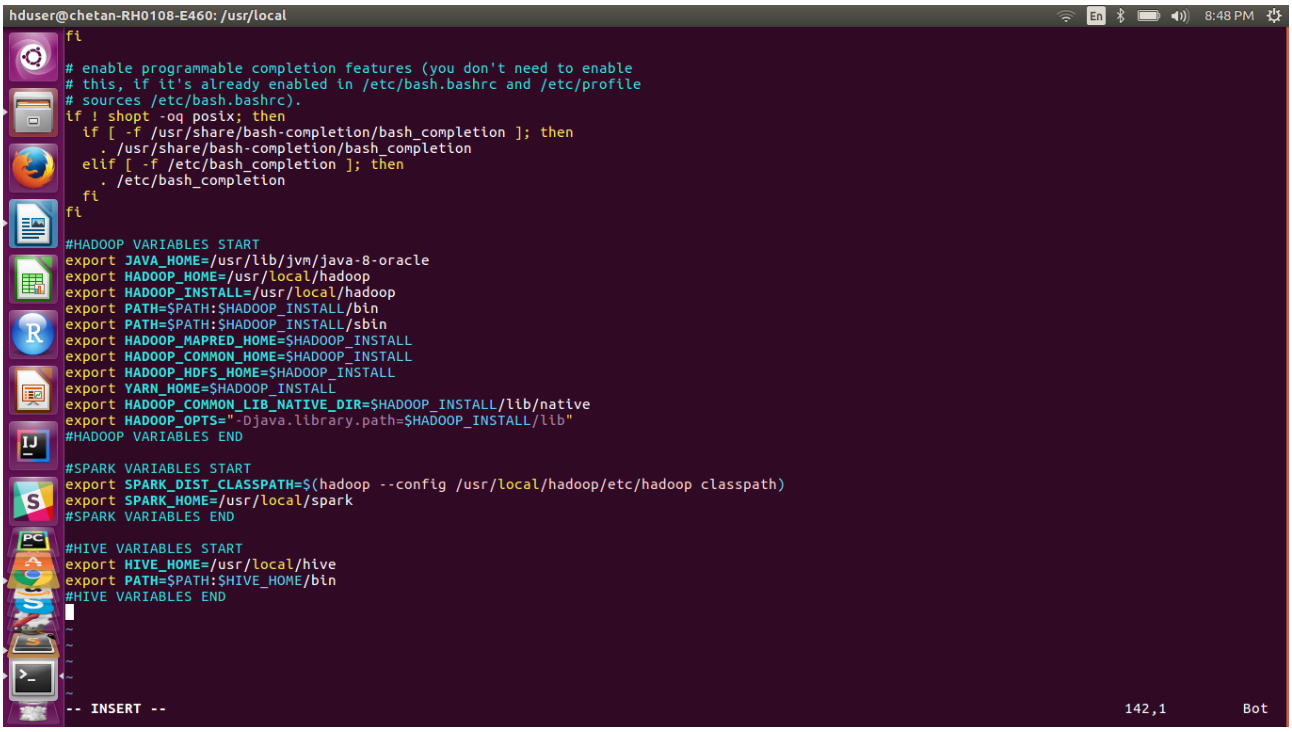
2. Create Hive Directories within HDFS and grant appropriate permissions

```
hadoop fs -mkdir /tmp
hadoop fs -mkdir -p /usr/local/hive/warehouse
hadoop fs -chmod g+w /tmp
hadoop fs -chmod g+w /usr/local/hive/warehouse
```

3. Configure Environment Variables

```
sudo  vim ~/.bashrc
```

```
#HIVE VARIABLES START
export HIVE_HOME=/usr/local/hive
export PATH=$PATH:$HIVE_HOME/bin
#HIVE VARIABLES END
```



```
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

#HADOOP VARIABLES START
export JAVA_HOME=/usr/lib/jvm/java-8-oracle
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
#HADOOP VARIABLES END

#SPARK VARIABLES START
export SPARK_DIST_CLASSPATH=$(hadoop --config /usr/local/hadoop/etc/hadoop classpath)
export SPARK_HOME=/usr/local/spark
#SPARK VARIABLES END

#HIVE VARIABLES START
export HIVE_HOME=/usr/local/hive
export PATH=$PATH:$HIVE_HOME/bin
#HIVE VARIABLES END

~
~
~
~
~
-- INSERT --                                                          142,1         Bot
```

Figure 2: *Environment variables – Apache Hadoop, Apache Hive at ~/.bashrc file*

4. Install MySQL for Hive Metastore

```
sudo apt-get install mysql-server
sudo apt-get install libmysql-java
```

4.1 Provide softlink to mysql connector at Hive Library directory

```
sudo ln -s /usr/share/java/mysql-connector-java.jar $HIVE_HOME/lib/mysql-connector-java.jar
```

4.2 Create Hive MetaStore Schema and execute hive-schema script there.

```
mysql -u root -p
```

```
mysql> CREATE DATABASE metastore;
mysql> USE metastore;
mysql> SOURCE usr/local/hive/scripts/metastore/upgrade/mysql/hive-schema-0.14.0.mysql.sql;

mysql> CREATE USER 'hiveuser'@'%' IDENTIFIED BY 'hivepassword';
mysql> GRANT all on *.* to 'hiveuser'@localhost identified by 'hivepassword';
mysql> flush privileges;
```

5. hive-site.xml configuration setup

Add below at conf/hive-site.xml , if not there then create it.

```
<property>
    <name>javax.jdo.option.ConnectionURL</name>
    <value>jdbc:mysql://localhost/metastore?createDatabaseIfNotExist=true</value>
    <description>metadata is stored in a MySQL server</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionDriverName</name>
    <value>com.mysql.jdbc.Driver</value>
    <description>MySQL JDBC driver class</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionUserName</name>
    <value>hiveuser</value>
    <description>user name for connecting to mysql server</description>
  </property>
  <property>
    <name>javax.jdo.option.ConnectionPassword</name>
    <value>hivepassword</value>
    <description>password for connecting to mysql server</description>
  </property>
```

Replace below 3 properties tag with whatever already exist by default. otherwise it will throw an error

"java.net.URISyntaxException: Relative path in absolute URI: ${system:java.io.tmpdir%7D/$%7Bsystem:user.name%7D"

```
<property>
   <name>hive.querylog.location</name>
   <value>$HIVE_HOME/iotmp</value>
   <description>Location of Hive run time structured log file</description>
</property>
```

```
<property>
   <name>hive.exec.local.scratchdir</name>
   <value>$HIVE_HOME/iotmp</value>
   <description>Local scratch space for Hive jobs</description>
</property>

<property>
   <name>hive.downloaded.resources.dir</name>
   <value>$HIVE_HOME/iotmp</value>
   <description>Temporary local directory for added resources in the remote file
system.</description>
</property>
```

6. Testing Hive with MySQL Metastore

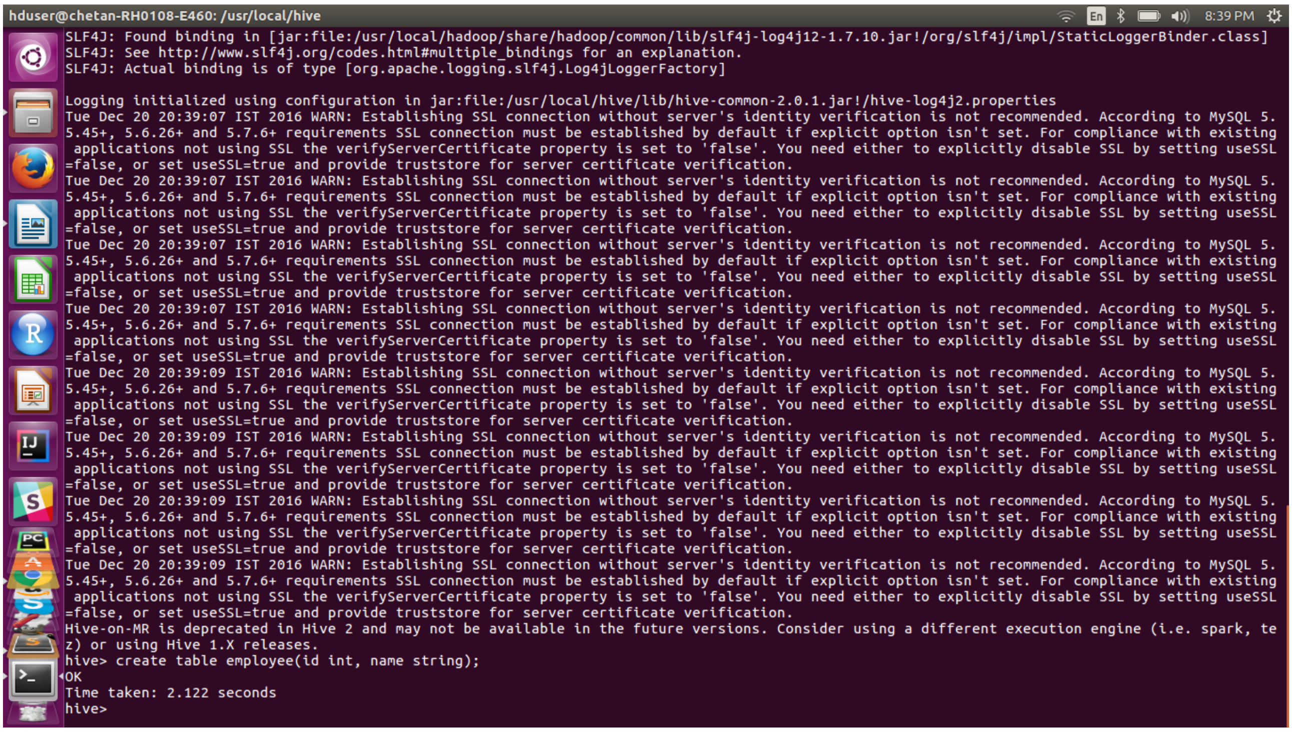
Go to **/usr/local/hive** and start hive by **bin/hive** command

Now Hive is running for you, You can test it out with MySQL Metastore.

```
hive> create table employee(id int, name string);
```

```
mysql -u root -p
mysql> use metastore;
mysql> show tables;
mysql> select * from TBLS;
```

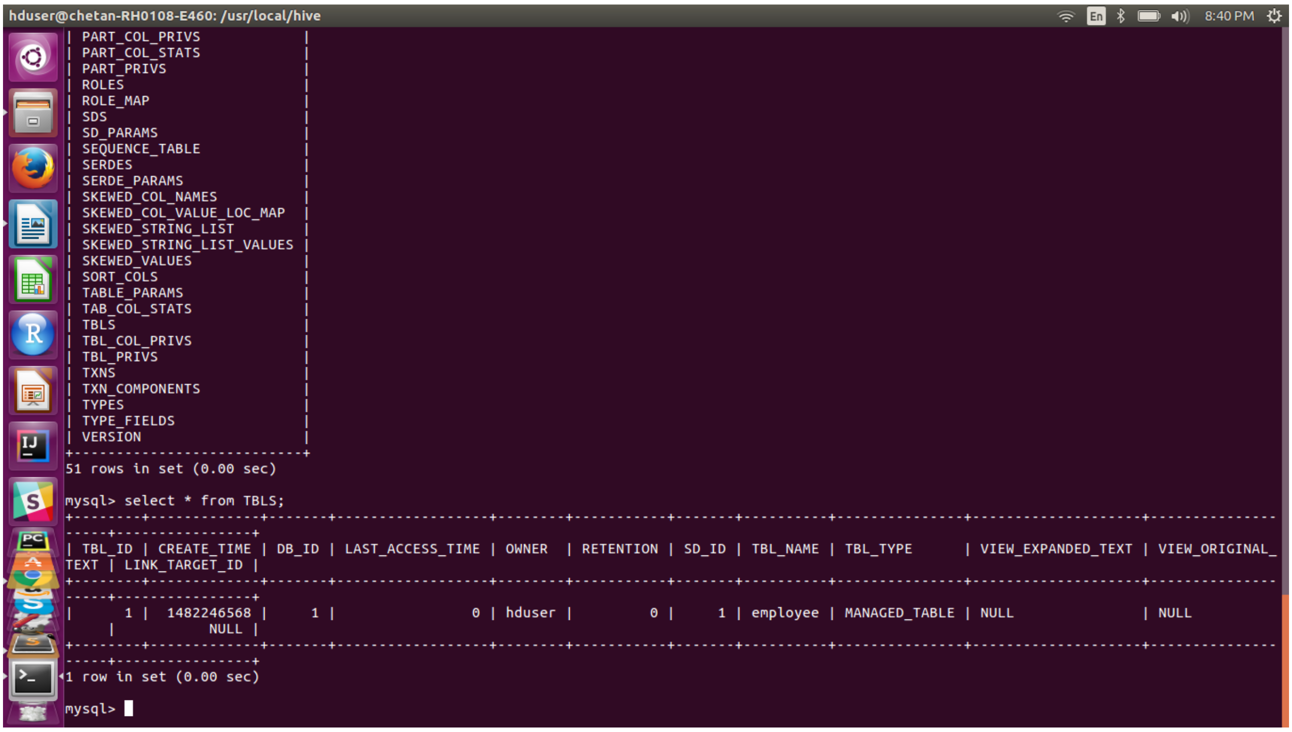
You can see your table over there.



```
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-2.0.1.jar!/hive-log4j2.properties
Tue Dec 20 20:39:07 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:07 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:07 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:07 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:09 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:09 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:09 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Tue Dec 20 20:39:09 IST 2016 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.
5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For compliance with existing
 applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly disable SSL by setting useSSL
=false, or set useSSL=true and provide truststore for server certificate verification.
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, te
z) or using Hive 1.X releases.
hive> create table employee(id int, name string);
OK
Time taken: 2.122 seconds
hive>
```

Figure 3: *Running Hive Query at Hive Shell*



```
| PART_COL_PRIVS            |
| PART_COL_STATS            |
| PART_PRIVS                |
| ROLES                     |
| ROLE_MAP                  |
| SDS                       |
| SD_PARAMS                 |
| SEQUENCE_TABLE            |
| SERDES                    |
| SERDE_PARAMS              |
| SKEWED_COL_NAMES          |
| SKEWED_COL_VALUE_LOC_MAP  |
| SKEWED_STRING_LIST        |
| SKEWED_STRING_LIST_VALUES |
| SKEWED_VALUES             |
| SORT_COLS                 |
| TABLE_PARAMS              |
| TAB_COL_STATS             |
| TBLS                      |
| TBL_COL_PRIVS             |
| TBL_PRIVS                 |
| TXNS                      |
| TXN_COMPONENTS            |
| TYPES                     |
| TYPE_FIELDS               |
| VERSION                   |
+---------------------------+
51 rows in set (0.00 sec)

mysql> select * from TBLS;
+--------+-------------+-------+------------------+--------+-----------+-------+----------+---------------+--------------------+--------------
-----+----------------+
| TBL_ID | CREATE_TIME | DB_ID | LAST_ACCESS_TIME | OWNER  | RETENTION | SD_ID | TBL_NAME | TBL_TYPE      | VIEW_EXPANDED_TEXT | VIEW_ORIGINAL_
TEXT | LINK_TARGET_ID |
+--------+-------------+-------+------------------+--------+-----------+-------+----------+---------------+--------------------+--------------
-----+----------------+
|      1 |  1482246568 |     1 |                0 | hduser |         0 |     1 | employee | MANAGED_TABLE | NULL               | NULL
     |           NULL |
+--------+-------------+-------+------------------+--------+-----------+-------+----------+---------------+--------------------+--------------
-----+----------------+
1 row in set (0.00 sec)

mysql> 
```

Figure 4: *Already created table metastore data at MySQL*