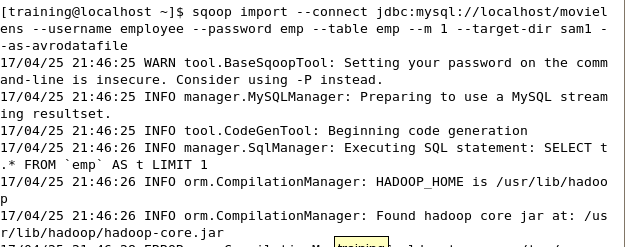
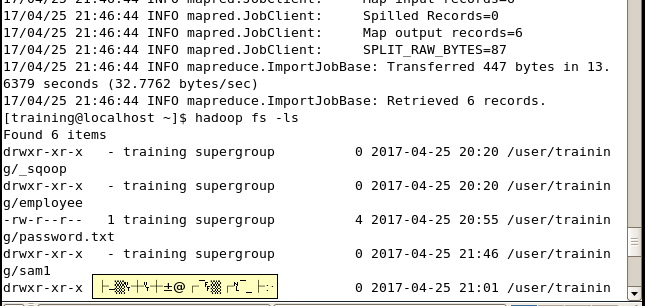
**ASSIGNMENT 15.6**

**1)IMPORT MYSQL TABLE AS AVRO FILE IN HDFS**

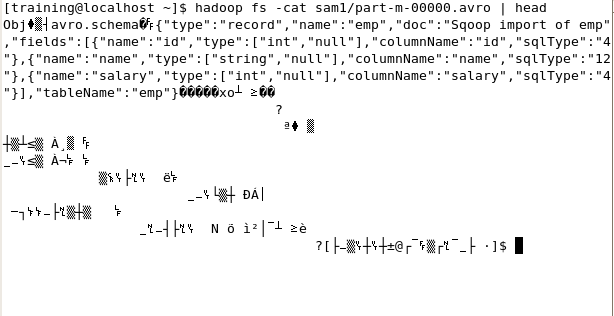
**SQOOP COMMAND FOR AVRODATAFILE**

****

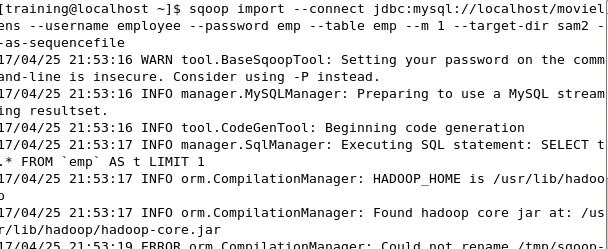
**DISPLAY THE DIRECTORY STRUCTURE**

****

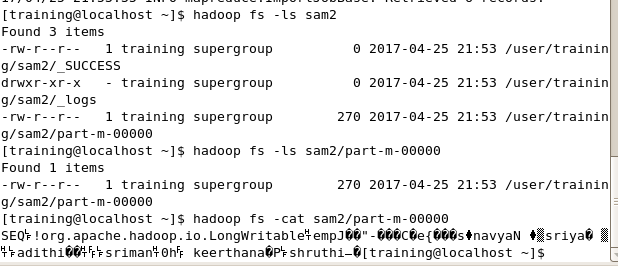
**DISPLAY THE CONTENTS OF AVRO FILE**

****

**2) IMPORT MYSQL DATA INTO HDFS AS SEQUENCE FILE**

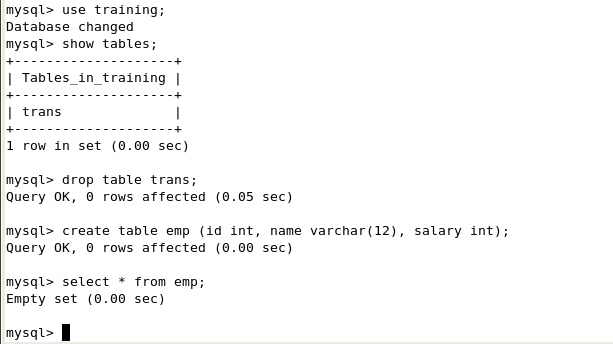
****

**DISPLAY THE DIRECTORY AND SEQUENCE FILE CONTENTS**

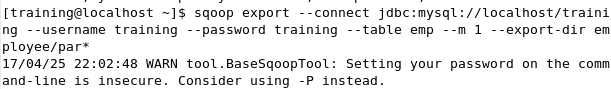
****

**3)SQOOP EXPORT**

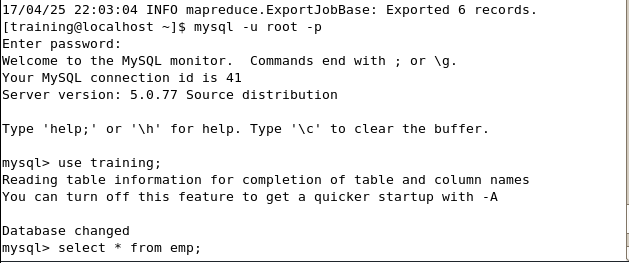
**CREATE AN EMPTY TABLE IN MYSQL**

****

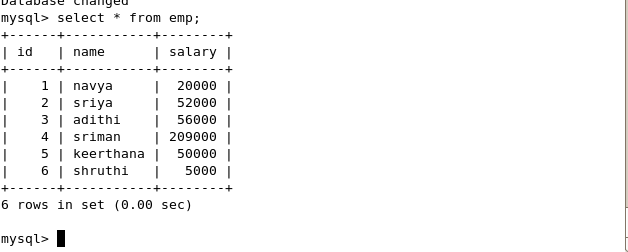
**EXECUTE SQOOP EXPORT COMMAND**

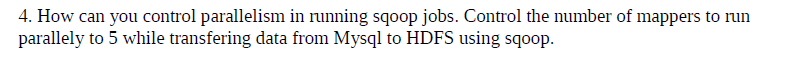
****

**OPEN MYSQL AND USE THE DATABASE**

****

**DISPLAY THE TABLE FIELDS AND CONTENTS**

****



**.Controlling Parallelism**

Specifies number of map tasks that can run in parallel. Default is 4. To optimize performance, set the number of map tasks to a value lower than the maximum number of connections that the database supports.

Use the parameter --num-mappers if you want Sqoop to use a different number of mappers. For example, to suggest 10 concurrent tasks, use the following Sqoop command:

**sqoop import --connect**

**jdbc:mysql://localhost/sqoop --username**

**sqoop --password**

**sqoop --table**

**cities --num-mappers 5**

Controlling the amount of parallelism that Sqoop will use to transfer data is the main way to control the load on your database. Using more mappers will lead to a higher number of concurrent data transfer tasks, which can result in faster job completion. However, it will also increase the load on the database as Sqoop will execute more concurrent queries.