ASSIGNMENT 15.6

1)IMPORT MYSQL TABLE AS AVRO FILE IN HDFS

```
[training@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/moviel
ens --username employee --password emp --table emp --m 1 --target-dir sam1 -
-as-avrodatafile
17/04/25 21:46:25 WARN tool.BaseSqoopTool: Setting your password on the comm
and-line is insecure. Consider using -P instead.
17/04/25 21:46:25 INFO manager.MySQLManager: Preparing to use a MySQL stream
ing resultset.
```

```
Found 6 items
drwxr-xr-x   - training supergroup          0 2017-04-25 20:20 /user/trainin
g/_sqoop
drwxr-xr-x   - training supergroup          0 2017-04-25 20:20 /user/trainin
g/employee
-rw-r--r--   1 training supergroup          4 2017-04-25 20:55 /user/trainin
g/password.txt
drwxr-xr-x   - training supergroup          0 2017-04-25 21:46 /user/trainin
g/sam1
```

OUTPUT

```
[training@localhost ~]$ hadoop fs -cat sam1/part-m-00000.avro | head
Obj♦▒┤avro.schema▒F{"type":"record","name":"emp","doc":"Sqoop import of emp"
,"fields":[{"name":"id","type":["int","null"],"columnName":"id","sqlType":"4
"},{"name":"name","type":["string","null"],"columnName":"name","sqlType":"12
"},{"name":"salary","type":["int","null"],"columnName":"salary","sqlType":"4
"}],"tableName":"emp"}▒▒▒▒▒xo┴ ≥▒▒
                                    ?
                                     ª♦ ▒
┼▒┴≤▒ Ȧ▒ F
_ᵕ≤▒ Ȧ¬F ┣
          ▒ᵏ┣ᵏ  ëF
                    _ᵕ└▒┼ ÐÁ|
 ¬┐┣┣_┣ᵏ▒┼▒   ┣
               _ᵏ_┤┣ᵏ  N ö ì² |¯┴ ≥è
                                  ?[┣_▒ᵏ┼ᵏ┼±@┌¯F▒┌ᵏ¯_┣ ·]$ ▉
```

2) IMPORT MYSQL DATA INTO HDFS AS SEQUENCE FILE

```
training@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/moviel
ns --username employee --password emp --table emp --m 1 --target-dir sam2 -
as-sequencefile
7/04/25 21:53:16 WARN tool.BaseSqoopTool: Setting your password on the comm
nd-line is insecure. Consider using -P instead.
7/04/25 21:53:16 INFO manager.MySQLManager: Preparing to use a MySQL stream
ng resultset.
7/04/25 21:53:16 INFO tool.CodeGenTool: Beginning code generation
```

```
[training@localhost ~]$ hadoop fs -cat sam2/part-m-00000
SEQ!org.apache.hadoop.io.LongWritableempJ""-CeseshnavyaN sriya
adithisrimanh0h keerthanaPshruthi-[training@localhost ~]$
```

3)SQOOP EXPORT

CREATE A TABLE IN MYSQL

```
mysql> create table emp (id int, name varchar(12), salary int);
Query OK, 0 rows affected (0.00 sec)
```

SQOOP EXPORT

```
[training@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/traini
ng --username training --password training --table emp --m 1 --export-dir em
ployee/par*
17/04/25 22:02:48 WARN tool.BaseSqoopTool: Setting your password on the comm
and-line is insecure. Consider using -P instead.
```

OUTPUT

```
mysql> select * from emp;
+------+-----------+--------+
| id   | name      | salary |
+------+-----------+--------+
|    1 | navya     |  20000 |
|    2 | sriya     |  52000 |
|    3 | adithi    |  56000 |
|    4 | sriman    | 209000 |
|    5 | keerthana |  50000 |
|    6 | shruthi   |   5000 |
+------+-----------+--------+
6 rows in set (0.00 sec)

mysql>
```

4. How can you control parallelism in running sqoop jobs. Control the number of mappers to run parallely to 5 while transfering data from Mysql to HDFS using sqoop.

4)CONTROLLING PARALLELISM

- It 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.