Experiment - 02

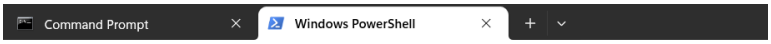
Aim: To install Sqoop tool and execute Hadoop Commands to transfer data between Hadoop and relational database servers

Theory: Sqoop (SQL-to-Hadoop) is an open-source tool designed to transfer data between relational databases and Hadoop ecosystems like HDFS (Hadoop Distributed File System), Hive, and HBase. Sqoop facilitates efficient data import and export, enabling seamless integration between large-scale data stored in Hadoop and traditional databases.

Key Features of Sqoop:

1. **Data Import and Export:**
   - Sqoop allows importing data from relational databases like **MySQL**, **PostgreSQL**, **Oracle**, and others into Hadoop HDFS, Hive, or HBase.
   - It also enables the export of data from Hadoop back into relational databases.
2. **Parallel Data Transfer:**
   - Sqoop uses **MapReduce** to divide the import/export job into multiple parallel tasks, allowing high-speed data transfer, which is both reliable and fault-tolerant.
3. **Support for Incremental Loads:**
   - Sqoop can import data incrementally, fetching only the newly added records by checking timestamp columns or primary keys. This minimizes redundant data imports.
4. **Support for Multiple File Formats:**
   - Data imported using Sqoop can be stored in various formats like **text**, **Avro**, **Sequence**, or **Parquet**, making it versatile and suitable for different use cases.
5. **Data Compression:**
   - Sqoop supports data compression techniques, such as **Gzip** and **Snappy**, allowing efficient storage of large datasets in Hadoop.

MySQL Database Table to be used



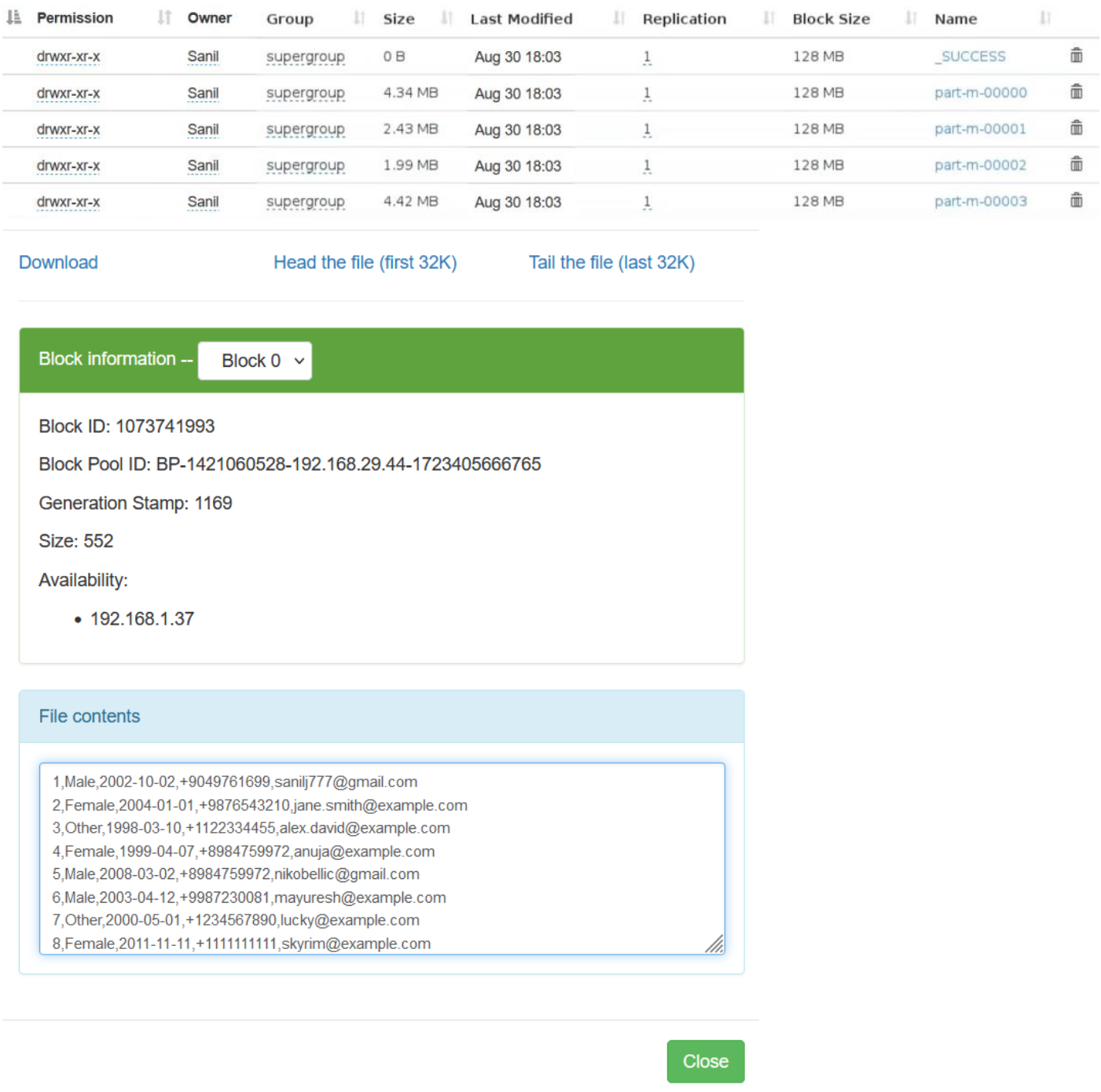

```
+---------------------------+
| Tables_in_telecomservice |
+---------------------------+
| customers                 |
| device                    |
| location                  |
| service                   |
| telecom_service           |
| time                      |
+---------------------------+
6 rows in set (0.01 sec)

mysql> select * from customers;
+------------+--------+------------+-------------+------------------------+
| customerID | gender | birthDate  | contactNo   | email_ID               |
+------------+--------+------------+-------------+------------------------+
|          1 | Male   | 2002-10-02 | +9049761699 | sanilj777@gmail.com    |
|          2 | Female | 2004-01-01 | +9876543210 | jane.smith@example.com |
|          3 | Other  | 1998-03-10 | +1122334455 | alex.david@example.com |
|          4 | Female | 1999-04-07 | +8984759972 | anuja@example.com      |
|          5 | Male   | 2008-03-02 | +8984759972 | nikobellic@gmail.com   |
|          6 | Male   | 2003-04-12 | +9987230081 | mayuresh@example.com   |
|          7 | Other  | 2000-05-01 | +1234567890 | lucky@example.com      |
|          8 | Female | 2011-11-11 | +1111111111 | skyrim@example.com     |
|          9 | Male   | 2004-04-04 | +1234567890 | saivee@example.com     |
|         10 | Female | 2004-10-02 | +8850227919 | pihu@example.com       |
+------------+--------+------------+-------------+------------------------+
10 rows in set (0.00 sec)
```

Sqoop Commands

1. Sqoop **Import** Table from Database

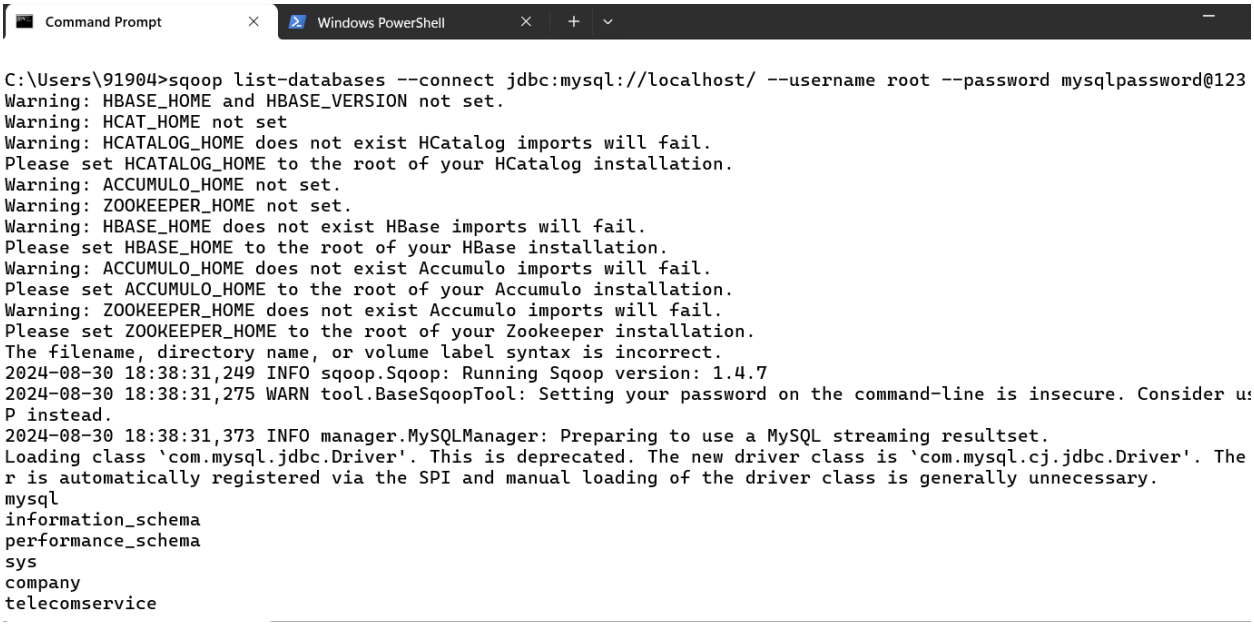
>sqoop import --connect jdbc:mysql://localhost/telecomservice --username root --password mysqlpassword@123 --table customers



Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	Sanil	supergroup	0 B	Aug 30 18:03	1	128 MB	_SUCCESS
drwxr-xr-x	Sanil	supergroup	4.34 MB	Aug 30 18:03	1	128 MB	part-m-00000
drwxr-xr-x	Sanil	supergroup	2.43 MB	Aug 30 18:03	1	128 MB	part-m-00001
drwxr-xr-x	Sanil	supergroup	1.99 MB	Aug 30 18:03	1	128 MB	part-m-00002
drwxr-xr-x	Sanil	supergroup	4.42 MB	Aug 30 18:03	1	128 MB	part-m-00003

2. List Databases using Sqoop

>sqoop list-databases --connect jdbc:mysql://localhost/ --username root --password mysqlpassword@123




```
C:\Users\91904>sqoop list-tables --connect jdbc:mysql://localhost/telecomservice --username root --password mysqlpasswor
d@123
Warning: HBASE_HOME and HBASE_VERSION not set.
Warning: HCAT_HOME not set
Warning: HCATALOG_HOME does not exist HCatalog imports will fail.
Please set HCATALOG_HOME to the root of your HCatalog installation.
Warning: ACCUMULO_HOME not set.
Warning: ZOOKEEPER_HOME not set.
Warning: HBASE_HOME does not exist HBase imports will fail.
Please set HBASE_HOME to the root of your HBase installation.
Warning: ACCUMULO_HOME does not exist Accumulo imports will fail.
Please set ACCUMULO_HOME to the root of your Accumulo installation.
Warning: ZOOKEEPER_HOME does not exist Accumulo imports will fail.
Please set ZOOKEEPER_HOME to the root of your Zookeeper installation.
The filename, directory name, or volume label syntax is incorrect.
2024-08-30 18:45:48,863 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2024-08-30 18:45:48,887 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -
P instead.
2024-08-30 18:45:49,038 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
Loading class `com.mysql.jdbc.Driver'. This is deprecated. The new driver class is `com.mysql.cj.jdbc.Driver'. The drive
r is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
customers
device
location
service
telecom_service
time
```

4. Using **hadoop fs -cat** to view the contents of the imported database

>hadoop fs -cat /user/part-m-00000

```
C:\Users\91904>hadoop fs -cat /user/part-m-00000
The filename, directory name, or volume label syntax is
1,Male,2002-10-02,+9049761699,sanilj777@gmail.com
2,Female,2004-01-01,+9876543210,jane.smith@example.com
3,Other,1998-03-10,+1122334455,alex.david@example.com
4,Female,1999-04-07,+8984759972,anuja@example.com
5,Male,2008-03-02,+8984759972,nikobellic@gmail.com
6,Male,2003-04-12,+9987230081,mayuresh@example.com
7,Other,2000-05-01,+1234567890,lucky@example.com
8,Female,2011-11-11,+1111111111,skyrim@example.com
9,Female,2004-04-04,+1234567890,saivee@example.com
10,Female,2004-10-02,+8850227919,pihu@example.com
```

5. Using **grep** along with the **cat** command to display a specific row

>hadoop fs -cat /user/part-m-00000 | findstr = "__"

```
C:\Users\91904>hadoop fs -cat /user/part-m-00000 | findstr "sanil"
The filename, directory name, or volume label syntax is incorrect.
1,Male,2002-10-02,+9049761699,sanilj777@gmail.com

C:\Users\91904>hadoop fs -cat /user/part-m-00000.txt | findstr "Male"
The filename, directory name, or volume label syntax is incorrect.
1,Male,2002-10-02,+9049761699,sanilj777@gmail.com
5,Male,2008-03-02,+8984759972,nikobellic@gmail.com
6,Male,2003-04-12,+9987230081,mayuresh@example.com

C:\Users\91904>hadoop fs -cat /user/part-m-00000.txt | findstr "Female"
The filename, directory name, or volume label syntax is incorrect.
2,Female,2004-01-01,+9876543210,jane.smith@example.com
4,Female,1999-04-07,+8984759972,anuja@example.com
8,Female,2011-11-11,+1111111111,skyrim@example.com
9,Female,2004-04-04,+1234567890,saivee@example.com
10,Female,2004-10-02,+8850227919,pihu@example.com
```

Conclusion: Hence, we successfully installed the Sqoop tool and executed Hadoop commands to transfer data between Hadoop and RDBMS servers