Assignment_2

2023-10-18

Architecture of Big Data System

Assignment 2

##Submitted By : Preetham Gouda

Sqoop Commands

Questions

1. Grant privileges to database(s) in MySQL.
2. List all databases in MySQL using Sqoop.
3. List tables in selected database using Sqoop.
4. Import table from local DB to HDFS using Sqoop by target dir method.
5. Import table from local DB to HDFS using Sqoop by warehouse dir method.
6. Import selected data from specified table in local DB to HDFS.
7. Import only selected fields (columns) from local table to HDFS.
8. Import all tables present in local DB to HDFS
9. Import all table excluding specified table(s) into HDFS.
10. Export table present in HDFS to local DB.

1. Grant privileges to database(s) in MySQL

```
GRANT ALL PRIVILEGES ON *.* to 'user'@'localhost' IDENTIFIED BY 'password';
```

```
Query OK, 0 rows affected, 1 warning (0.02 sec)
```

---

2. List all databases in MySQL using Sqoop.

```
sqoop list-databases --connect jdbc:mysql://localhost/?useSSL=false --username user --password password;
```

```
Warning: /opt/sqoop/../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /opt/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /opt/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /opt/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
/opt/hadoop/libexec/hadoop-functions.sh: line 2366: HADOOP_ORG.APACHE.SQOOP.SQOOP_USER: bad substitution
/opt/hadoop/libexec/hadoop-functions.sh: line 2461: HADOOP_ORG.APACHE.SQOOP.SQOOP_OPTS: bad substitution
2023-10-06 09:29:20,113 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2023-10-06 09:29:20,369 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider
using -P instead.
2023-10-06 09:29:20,706 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'. Th
e driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
information_schema
mysql
performance_schema
retail_db
sys
testDatabase
```

---

3. List tables in selected database using Sqoop

```
sqoop list-tables --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password password;
```

```
Warning: /opt/sqoop/../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /opt/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /opt/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /opt/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
/opt/hadoop/libexec/hadoop-functions.sh: line 2366: HADOOP_ORG.APACHE.SQOOP.SQOOP_USER: bad substitution
/opt/hadoop/libexec/hadoop-functions.sh: line 2461: HADOOP_ORG.APACHE.SQOOP.SQOOP_OPTS: bad substitution
2023-10-06 09:31:29,140 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7
2023-10-06 09:31:29,436 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider
using -P instead.
2023-10-06 09:31:29,783 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'. Th
e driver is automatically registered via the SPI and manual loading of the driver class is generally unnecessary.
categories
customers
departments
order_items
orders
products
```

---

4. Import table from local DB to HDFS using Sqoop by target dir method.

```

sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password password --table
customers --target-dir '/customer_tb';
```

```
2023-10-06 09:38:34,259 INFO mapreduce.ImportJobBase: Transferred 931.1768 KB in 66.1862 seconds (14.069 KB/sec)
2023-10-06 09:38:34,269 INFO mapreduce.ImportJobBase: Retrieved 12435 records.
```

---

5. Import table from local DB to HDFS using Sqoop by warehouse dir method.

```

sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password password --table
orders --warehouse-dir '/orders_tb';
```

```
2023-10-06 09:46:57,316 INFO mapreduce.ImportJobBase: Transferred 2.861 MB in 58.9446 seconds (49.7014 KB/sec)
2023-10-06 09:46:57,326 INFO mapreduce.ImportJobBase: Retrieved 68883 records.
```

---

6. Import selected data from specified table in local DB to HDFS.

```
sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password password --table
orders --warehouse-dir '/orders_id_tb' --where "order_id<200";
```

```
2023-10-06 09:53:04,869 INFO mapreduce.ImportJobBase: Transferred 8.1035 KB in 48.4842 seconds (171.1484 bytes/se
c)
2023-10-06 09:53:04,896 INFO mapreduce.ImportJobBase: Retrieved 199 records.
```

---

7. Import only selected fields (columns) from local table to HDFS

```
sqoop import --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password password --column
s order_id,order_status --table orders --warehouse-dir '/orders_id_status_tb' --where "order_id<200";
```

```
2023-10-06 09:56:45,690 INFO mapreduce.ImportJobBase: Transferred 2.8398 KB in 48.7203 seconds (59.6876 bytes/se
c)
2023-10-06 09:56:45,702 INFO mapreduce.ImportJobBase: Retrieved 199 records.
```

---

8. Import all tables present in local DB to HDFS

```
sqoop import-all-tables --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password passwo
rd --warehouse-dir '/retail_db';
```

```
2023-10-06 10:04:30,912 INFO mapreduce.ImportJobBase: Transferred 169.915 KB in 50.1136 seconds (3.3906 KB/sec)
2023-10-06 10:04:30,927 INFO mapreduce.ImportJobBase: Retrieved 1345 records.
```

---

9. Import all table excluding specified table(s) into HDFS.

```

sqoop import-all-tables --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password passwo
rd --warehouse-dir '/retail_db_partial' --exclude-tables orders,customers,departments,products;
```

```
2023-10-06 10:08:24,379 INFO mapreduce.ImportJobBase: Transferred 5.1583 MB in 49.4886 seconds (106.7338 KB/sec)
2023-10-06 10:08:24,391 INFO mapreduce.ImportJobBase: Retrieved 172198 records.
Skipping table: orders
Skipping table: products
```

---

10. Export table present in HDFS to local DB.

```
--Create a skeleton copy of the table to be exported in local machine
CREATE TABLE test_categories LIKE categories;
```

```
#Export table from hdfs to local machine
sqoop export --connect jdbc:mysql://localhost/retail_db?useSSL=false --username user --password password --table
test_categories --export-dir /retail_db/categories
```

```
023-10-08 09:29:02,426 INFO mapreduce.ExportJobBase: Transferred 2.1592 KB in 54.3225 seconds (40.7014 bytes/sec)
2023-10-08 09:29:02,434 INFO mapreduce.ExportJobBase: Exported 58 records.
```