CN7031 - Big Data Analytics

Student ID: 3059676

Name: Mann Shrestha

LAB 3

Step 1: Accessing MySQL shell

Step 2: Show Database and Tables

```
mysql> SHOW DATABASES;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| cm                 |
| firehose           |
| hue                |
| metastore          |
| mysql              |
| nav                |
| navms              |
| oozie              |
| retail_db          |
| rman               |
| sentry             |
+--------------------+
12 rows in set (0.00 sec)
```

```
mysql> SHOW TABLES;
+---------------------+
| Tables_in_retail_db |
+---------------------+
| categories          |
| customers           |
| departments         |
| order_items         |
| orders              |
| products            |
+---------------------+
6 rows in set (0.00 sec)
```

Step 3: Exit from MySQL

Exercise 3: Importing data from MySQL database to Hadoop HDFS

```
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=20
        Job Counters
                Launched map tasks=2
                Other local map tasks=2
                Total time spent by all maps in occupied slots (ms)=13881
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=13881
                Total vcore-milliseconds taken by all map tasks=13881
                Total megabyte-milliseconds taken by all map tasks=14214144
        Map-Reduce Framework
                Map input records=1345
                Map output records=1345
                Input split bytes=236
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=524
                CPU time spent (ms)=7670
                Physical memory (bytes) snapshot=545239040
                Virtual memory (bytes) snapshot=5623541760
                Total committed heap usage (bytes)=542638080
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
25/10/23 13:38:12 INFO mapreduce.ImportJobBase: Transferred 55.6396 KB in 29.5615 seconds (1.8822 KB/sec)
25/10/23 13:38:12 INFO mapreduce.ImportJobBase: Retrieved 1345 records.
[cloudera@quickstart ~]$
```

Verification

```
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/
Found 7 items
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:35 /user/hive/warehouse/categories
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:36 /user/hive/warehouse/customers
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:36 /user/hive/warehouse/departments
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:37 /user/hive/warehouse/order_items
drwxrwxrwx   - cloudera supergroup          0 2022-06-25 00:58 /user/hive/warehouse/order_items_dfd
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:37 /user/hive/warehouse/orders
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:38 /user/hive/warehouse/products
[cloudera@quickstart ~]$ hdfs dfs -ls /user/hive/warehouse/categories/
Found 4 items
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:34 /user/hive/warehouse/categories/.metadata
drwxr-xr-x   - cloudera supergroup          0 2025-10-23 13:35 /user/hive/warehouse/categories/.signals
-rw-r--r--   1 cloudera supergroup       1491 2025-10-23 13:35 /user/hive/warehouse/categories/4f85b6b6-dda7-4e63-b099-99c0ea
b0e8f8.parquet
-rw-r--r--   1 cloudera supergroup       1520 2025-10-23 13:35 /user/hive/warehouse/categories/f1ef04ce-daa1-4998-a21d-e8c2b3
78727b.parquet
[cloudera@quickstart ~]$
```

/user/hive/warehouse								Go!

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	cloudera	supergroup	0 B	Thu Oct 23 21:35:30 +0100 2025	0	0 B	categories
drwxr-xr-x	cloudera	supergroup	0 B	Thu Oct 23 21:36:01 +0100 2025	0	0 B	customers
drwxr-xr-x	cloudera	supergroup	0 B	Thu Oct 23 21:36:32 +0100 2025	0	0 B	departments
drwxr-xr-x	cloudera	supergroup	0 B	Thu Oct 23 21:37:09 +0100 2025	0	0 B	order_items
drwxrwxrwx	cloudera	supergroup	0 B	Sat Jun 25 08:58:30 +0100 2022	0	0 B	order_items_dfd
drwxr-xr-x	cloudera	supergroup	0 B	Thu Oct 23 21:37:39 +0100 2025	0	0 B	orders
drwxr-xr-x	cloudera	supergroup	0 B	Thu Oct 23 21:38:10 +0100 2025	0	0 B	products

Exporting data from Hadoop to MySQL

```
                    File System Counters
                            FILE: Number of bytes read=0
                            FILE: Number of bytes written=606896
                            FILE: Number of read operations=0
                            FILE: Number of large read operations=0
                            FILE: Number of write operations=0
                            HDFS: Number of bytes read=501
                            HDFS: Number of bytes written=953839
                            HDFS: Number of read operations=16
                            HDFS: Number of large read operations=0
                            HDFS: Number of write operations=8
                    Job Counters
                            Killed map tasks=1
                            Launched map tasks=4
                            Other local map tasks=4
                            Total time spent by all maps in occupied slots (ms)=103179
                            Total time spent by all reduces in occupied slots (ms)=0
                            Total time spent by all map tasks (ms)=103179
                            Total vcore-milliseconds taken by all map tasks=103179
                            Total megabyte-milliseconds taken by all map tasks=105655296
                    Map-Reduce Framework
                            Map input records=12439
                            Map output records=12439
                            Input split bytes=501
                            Spilled Records=0
                            Failed Shuffles=0
                            Merged Map outputs=0
                            GC time elapsed (ms)=2062
                            CPU time spent (ms)=7220
                            Physical memory (bytes) snapshot=800034816
                            Virtual memory (bytes) snapshot=11136077824
                            Total committed heap usage (bytes)=799539200
                    File Input Format Counters
                            Bytes Read=0
                    File Output Format Counters
                            Bytes Written=953839
25/10/23 15:09:03 INFO mapreduce.ImportJobBase: Transferred 931.4834 KB in 55.5216 seconds (16.7769 KB/sec)
25/10/23 15:09:03 INFO mapreduce.ImportJobBase: Retrieved 12439 records.
[cloudera@quickstart ~]$
```

```
|       6226 | John         | Jones       | XXXXXXXXX    | XXXXXXXXX    | 1773 Hazy Bay                 | We
slaco       | TX           |      78596  |              |
|       6227 | David        | Weber       | XXXXXXXXX    | XXXXXXXXX    | 3803 Dusty Edge               | Ne
w Brunswick  | NJ          |      08901  |              |
|       6228 | Roger        | Davila      | XXXXXXXXX    | XXXXXXXXX    | 8688 Amber Creek Round        | Ca
guas        | PR           |      00725  |              |
|       6229 | Mary         | Smith       | XXXXXXXXX    | XXXXXXXXX    | 8341 Stony Avenue             | Ri
chmond      | IN           |      47374  |              |
|       6230 | Kenneth      | Cameron     | XXXXXXXXX    | XXXXXXXXX    | 4694 Burning Front            | Ca
guas        | PR           |      00725  |              |
|       6231 | Mary         | Smith       | XXXXXXXXX    | XXXXXXXXX    | 2250 Cinder Knoll             | Ha
rlingen     | TX           |      78550  |              |
|       6232 | Janet        | Webb        | XXXXXXXXX    | XXXXXXXXX    | 9767 Tawny Nook               | Lo
s Angeles   | CA           |      90047  |              |
|       6233 | Mary         | Elliott     | XXXXXXXXX    | XXXXXXXXX    | 9003 Easy Gate Drive          | Ca
guas        | PR           |      00725  |              |
|       6234 | Mary         | Smith       | XXXXXXXXX    | XXXXXXXXX    | 4775 Wishing Nook             | Br
onx         | NY           |      10458  |              |
|       6235 | Stephen      | Brown       | XXXXXXXXX    | XXXXXXXXX    | 8598 Jagged Treasure Circuit  | Ca
guas        | PR           |      00725  |              |
|       6236 | Mary         | Martinez    | XXXXXXXXX    | XXXXXXXXX    | 7253 Sunny Acres              | Ch
icago       | IL           |      60629  |              |
|       6237 | Bobby        | Anderson    | XXXXXXXXX    | XXXXXXXXX    | 5509 Dewy Hickory Centre      | El
 Cajon      | CA           |      92020  |              |
|       6238 | Mary         | Young       | XXXXXXXXX    | XXXXXXXXX    | 7894 Heather Gardens          | Ca
guas        | PR           |      00725  |              |
|       6239 | Karen        | Mcdonald    | XXXXXXXXX    | XXXXXXXXX    | 7231 Silent Square            | Ju
ana Diaz    | PR           |      00795  |              |
|       6240 | Virginia     | Elliott     | XXXXXXXXX    | XXXXXXXXX    | 5735 Quaking Cider Highway    | Ca
guas        | PR           |      00725  |              |
|       6241 | Mary         | Clay        | XXXXXXXXX    | XXXXXXXXX    | 7994 Easy Estates             | Ar
lington Heights | IL       |      60004  |              |
|       6242 | David        | Smith       | XXXXXXXXX    | XXXXXXXXX    | 9174 Umber Zephyr Run         | Ca
guas        | PR           |      00725  |              |
+------------+--------------+-------------+--------------+--------------+-------------------------------+---
------------------+--------------+-------------+
12439 rows in set (0.02 sec)

mysql>
```