



**SQOOP**

**Sqoop** – jest to aplikacja interfejsu wiersza poleceń do przesyłania danych między relacyjnymi bazami danych i Hadoop'em.

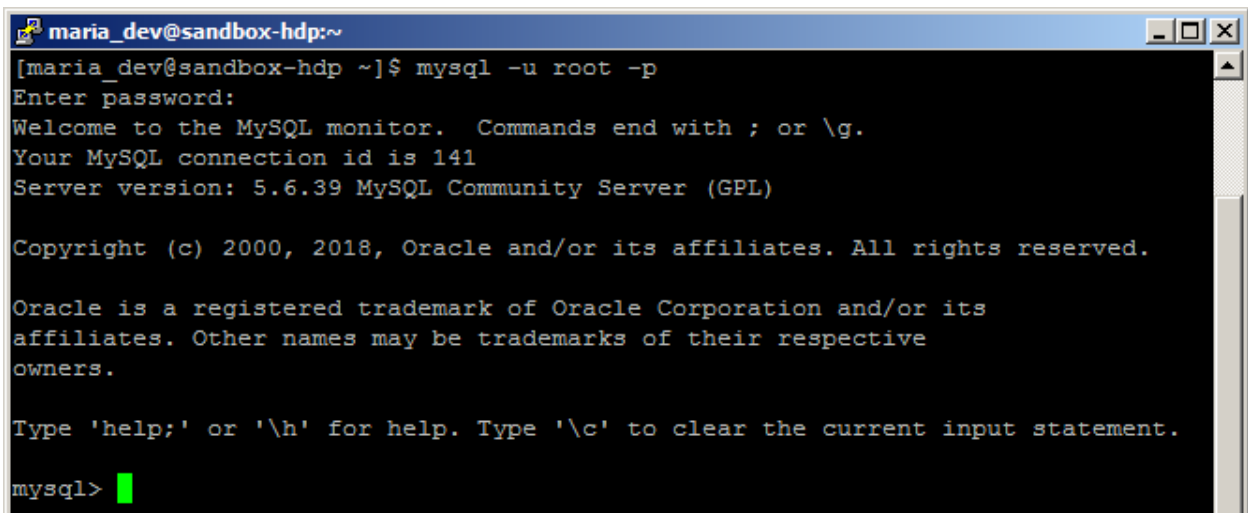
Zanim zaczniemy kopiować dane z MySQL do Hadoop musimy stworzyć baze danych i tabele w MySQL.

Po zalogowaniu się (login/hasło: *maria\_dev*) do wirtualnej maszyny korzystając z Putty, pobierz skrypt który stworzy nam tabele i wypełni ją danymi:

```
wget https://danilewicz.blob.core.windows.net/public/data.sql
```

Zaloguj się do MySQL wpisując: `mysql -u root -p`

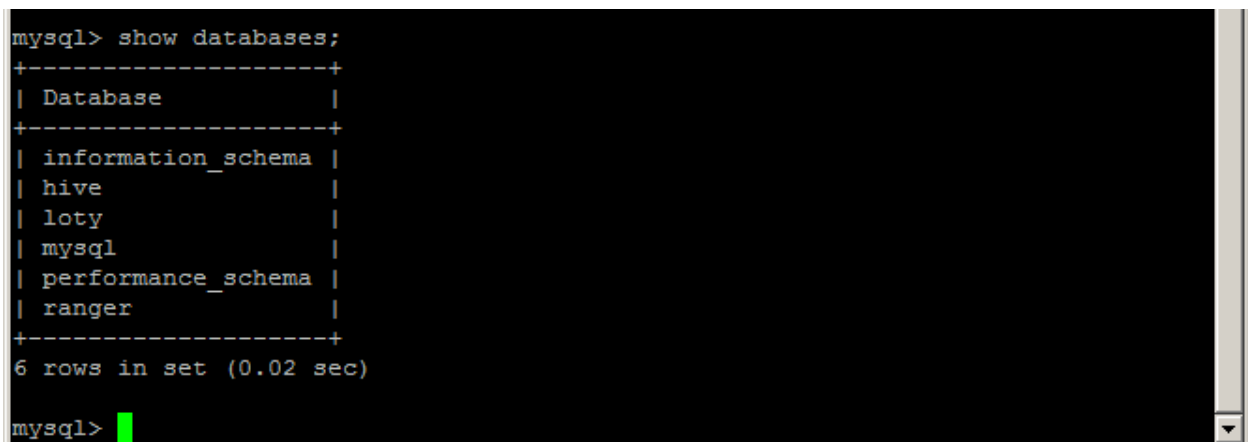
Zostaniesz poproszony o hasło. Hasłem dla użytkownika root jest: `hadoop`



```
maria_dev@sandbox-hdp:~  
[maria_dev@sandbox-hdp ~]$ mysql -u root -p  
Enter password:  
Welcome to the MySQL monitor.  Commands end with ; or \g.  
Your MySQL connection id is 141  
Server version: 5.6.39 MySQL Community Server (GPL)  
  
Copyright (c) 2000, 2018, Oracle and/or its affiliates. All rights reserved.  
  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
mysql>
```

Stwórz baze danych o nazwie '*loty*' wpisując: `create database loty;`

Sprawdź czy baza została utworzona wpisując: `show databases;`



```
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| hive |  
| loty |  
| mysql |  
| performance_schema |  
| ranger |  
+-----+  
6 rows in set (0.02 sec)  
  
mysql>
```

Wybierz tabelę na której będziesz pracował wpisując: `use loty;`

Aby uruchomić skrypt, który ściągnąłeś na początku wpisz: `source data.sql;`

Sprawdź czy tabela 'rok1987' została stworzona wpisując: `show tables;`

```
mysql> show tables;
+-----+
| Tables_in_loty |
+-----+
| rok1987        |
+-----+
1 row in set (0.00 sec)

mysql>
```

Pozostaje sprawdzić czy tabela nie jest pusta. Wpisz: `SELECT DepTime, ArrTime FROM rok1987 LIMIT 10;`

```
mysql> SELECT DepTime, ArrTime FROM rok1987 LIMIT 10;
+-----+-----+
| DepTime | ArrTime |
+-----+-----+
| 741     | 912     |
| 729     | 903     |
| 741     | 918     |
| 729     | 847     |
| 749     | 922     |
| 728     | 848     |
| 728     | 852     |
| 731     | 902     |
| 744     | 908     |
| 729     | 851     |
+-----+-----+
10 rows in set (0.00 sec)

mysql>
```

Pozostało dać dostęp Sqoop do MySQL. Wpisz: `GRANT ALL PRIVILEGES ON loty.* to '@'localhost';`

Wyloguj się z MySQL wpisując: `exit`

## Import danych z MySQL do HDFS:

```
scoop import --connect jdbc:mysql://localhost/loty --driver com.mysql.jdbc.Driver --table rok1987 -m 1
```

gdzie m oznacza liczbę zadań mapper'a które będą importować dane.

```
maria_dev@sandbox-hdp:~  
Map-Reduce Framework  
  Map input records=10000  
  Map output records=10000  
  Input split bytes=87  
  Spilled Records=0  
  Failed Shuffles=0  
  Merged Map outputs=0  
  GC time elapsed (ms)=295  
  CPU time spent (ms)=2750  
  Physical memory (bytes) snapshot=133419008  
  Virtual memory (bytes) snapshot=2133569536  
  Total committed heap usage (bytes)=41943040  
File Input Format Counters  
  Bytes Read=0  
File Output Format Counters  
  Bytes Written=971059  
19/03/01 11:43:49 INFO mapreduce.ImportJobBase: Transferred 948.2998 KB in 33.84  
94 seconds (28.0153 KB/sec)  
19/03/01 11:43:49 INFO mapreduce.ImportJobBase: Retrieved 10000 records.  
[maria_dev@sandbox-hdp ~]$
```

Wejdź do Ambari i sprawdź czy dane zostały skopiowane do HDFS 'a:

Name	Size	Last Modified	Owner	Group	Permission
↶					
📁 _SUCCESS	0.1 kB	2019-03-01 12:43	maria_dev	hdfs	-rw-r--r--
📁 part-m-00000	948.3 kB	2019-03-01 12:43	maria_dev	hdfs	-rw-r--r--

Na koniec usuń folder 'rok1987' z HDFS.



## Import danych z MySQL do Hive:

```
scoop import --connect jdbc:mysql://localhost/loty --driver com.mysql.jdbc.Driver --table rok1987 -m 1 -hive-import
```

Wejdź do Ambari i sprawdź czy tabela została utworzona w Hive:

The screenshot shows the Ambari Hive interface. At the top, there's a navigation bar with 'Ambari', 'Sandbox', '0 ops', '0 alerts', and a user profile 'maria\_dev'. Below this is the 'HIVE' section with tabs for 'QUERY', 'JOBS', 'TABLES', 'SAVED QUERIES', 'UDFs', and 'SETTINGS'. A '+ NEW JOB' and '+ NEW TABLE' button are on the right. The 'QUERY' tab is active, showing a 'Worksheet1' editor with a SQL query: `1 SELECT COUNT(Year) AS Liczba_Wierszy FROM rok1987;`. Below the editor are buttons for 'Execute', 'Save As', 'Insert UDF', and 'Visual Explain'. To the right of the editor is a 'Tables(4)' sidebar listing 'loty', 'rok1987' (highlighted), 'sample\_07', and 'sample\_08'. At the bottom, the 'RESULTS' tab is active, showing a table with one column 'liczba\_wierszy' and one row with the value '10000'.

## Export danych z Hive do MySQL:

Dane z których korzysta Hive znajdują się w folderze: [/apps/hive/warehouse/](#)

Name >	Size >	Last Modified >	Owner >	Group >	Permission
foodmart.db	--	2018-02-01 11:46	hive	hadoop	drwxrwxrwx
loty	--	2019-02-28 11:31	hive	hadoop	drwxrwxrwx
rok1987	--	2019-03-01 13:10	maria_dev	hadoop	drwxrwxrwx
sample_07	--	2018-02-01 11:47	hive	hadoop	drwxrwxrwx
sample_08	--	2018-02-01 11:47	hive	hadoop	drwxrwxrwx
xademo.db	--	2018-02-01 11:40	hive	hadoop	drwxrwxrwx

Zanim skopiujemy dane z Hive do MySQL musimy utworzyć tabelę do której będziemy kopiować dane.

Ściągnij skrypt który utworzy pustą tabelę 'hive1987':

`wget https://danilewicz.blob.core.windows.net/public/emptydata.sql`

Zaloguj się do MySQL: `mysql -u root -p`

Podaj hasło: `hadoop`

Wybierz bazę danych 'loty': `use loty;`

Uruchom skrypt: `source emptydata.sql;`

Sprawdź czy tabela została utworzona: `show tables;`

Wydź z MySQL: `exit`

Teraz możemy wyeksportować dane z Hive do MySQL:

```
sqoop export --connect jdbc:mysql://localhost/loty -m 1 --driver com.mysql.jdbc.Driver --table hive1987 --export-dir /apps/hive/warehouse/rok1987 --input-fields-terminated-by '\0001'
```

```
maria_dev@sandbox-hdp:~  
Map-Reduce Framework  
    Map input records=10000  
    Map output records=10000  
    Input split bytes=161  
    Spilled Records=0  
    Failed Shuffles=0  
    Merged Map outputs=0  
    GC time elapsed (ms)=119  
    CPU time spent (ms)=3450  
    Physical memory (bytes) snapshot=170061824  
    Virtual memory (bytes) snapshot=2124206080  
    Total committed heap usage (bytes)=42467328  
File Input Format Counters  
    Bytes Read=0  
File Output Format Counters  
    Bytes Written=0  
19/03/01 12:51:28 INFO mapreduce.ExportJobBase: Transferred 948.46 KB in 19.5631 seconds (48.4822 KB/sec)  
19/03/01 12:51:28 INFO mapreduce.ExportJobBase: Exported 10000 records.  
[maria_dev@sandbox-hdp ~]$
```

Pozostaje sprawdzić czy dane zostały skopiowane:

```
maria_dev@sandbox-hdp:~  
+-----+-----+  
| 1855 | 1756 |  
| 1838 | 1732 |  
| 1202 | 1039 |  
| 1156 | 1040 |  
| 1150 | 1039 |  
| 1217 | 1056 |  
| 1155 | 1042 |  
| 1153 | 1039 |  
| 1155 | 1041 |  
| 1211 | 1052 |  
| 1155 | 1040 |  
| 1209 | 1054 |  
| 1217 | 1056 |  
| 1218 | 1057 |  
| 1157 | 1041 |  
| 1208 | 1042 |  
+-----+-----+  
10000 rows in set (0.02 sec)  
  
mysql>
```