

SQOOP INSTALLATION

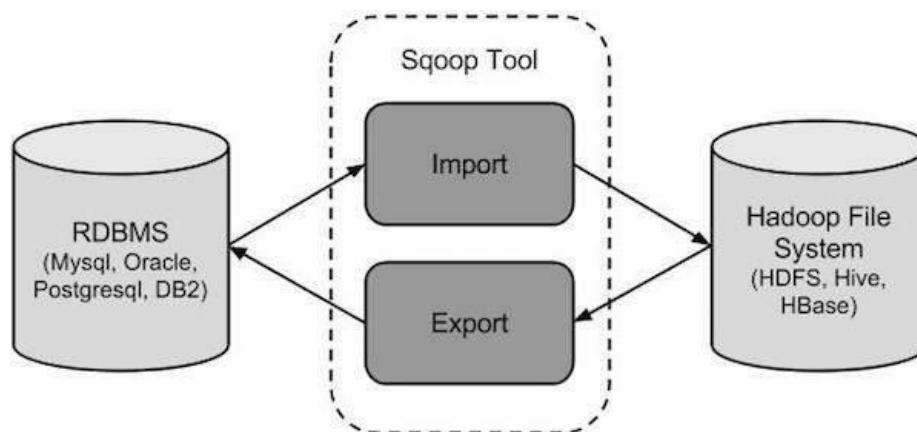
Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. The traditional application management system, that is, the interaction of applications with relational database using RDBMS, is one of the sources that generate Big Data. Such Big Data, generated by RDBMS, is stored in **Relational Database Servers** in the relational database structure.

When Big Data storages and analyzers such as MapReduce, Hive, HBase, Cassandra, Pig, etc. of the Hadoop ecosystem came into picture, they required a tool to interact with the relational database servers for importing and exporting the Big Data residing in them. Here, Sqoop occupies a place in the Hadoop ecosystem to provide feasible interaction between relational database server and Hadoop's HDFS.

Sqoop: "SQL to Hadoop and Hadoop to SQL"

Sqoop is a tool designed to transfer data between Hadoop and relational database servers. It is used to import data from relational databases such as MySQL, Oracle to Hadoop HDFS, and export from Hadoop file system to relational databases. It is provided by the Apache Software Foundation.

The following image describes the workflow of Sqoop.



Sqoop Import

The import tool imports individual tables from RDBMS to HDFS. Each row in a table is treated as a record in HDFS. All records are stored as text data in text files or as binary data in Avro and Sequence files.

Sqoop Export

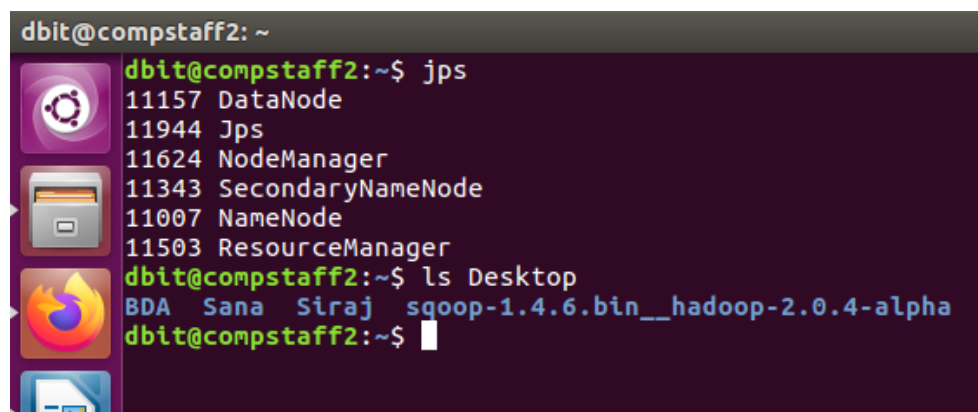
The export tool exports a set of files from HDFS back to an RDBMS. The files given as input to Sqoop contain records, which are called as rows in table.

Those are read and parsed into a set of records and delimited with user-specified delimiter.

STEPS to install Sqoop :

1. Extract the Sqoop Package from the tar file pasted on the Desktop. The extracted package can be seen, listed in the list of files and folders of the Desktop using the **ls** command.

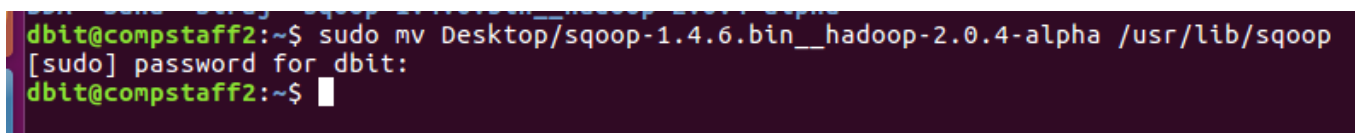
Is Desktop



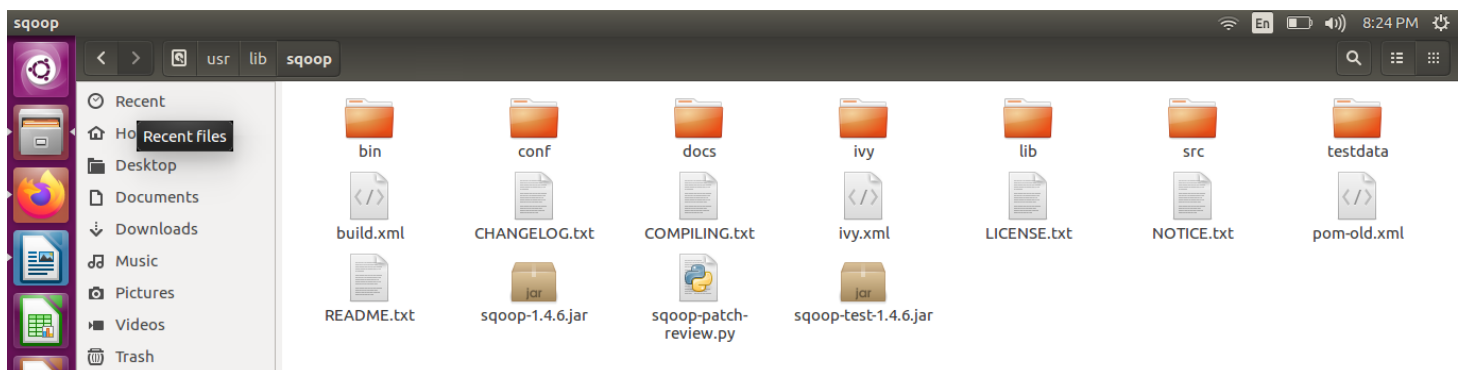
```
dbit@compstaff2: ~  
dbit@compstaff2:~$ jps  
11157 DataNode  
11944 Jps  
11624 NodeManager  
11343 SecondaryNameNode  
11007 NameNode  
11503 ResourceManager  
dbit@compstaff2:~$ ls Desktop  
BDA Sana Siraj sqoop-1.4.6.bin__hadoop-2.0.4-alpha  
dbit@compstaff2:~$
```

2. Move this extracted folder (sqoop-1.4.6.bin__hadoop-2.0.4-alpha) from Desktop to the directory /usr/lib/sqoop using the sudo mv command.

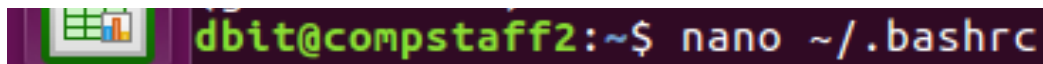
sudo mv Desktop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha /usr/lib/sqoop



```
dbit@compstaff2:~$ sudo mv Desktop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha /usr/lib/sqoop  
[sudo] password for dbit:  
dbit@compstaff2:~$
```



3. Sqoop environment can be set up only by appending the following lines by executing **nano ~/.bashrc** command.

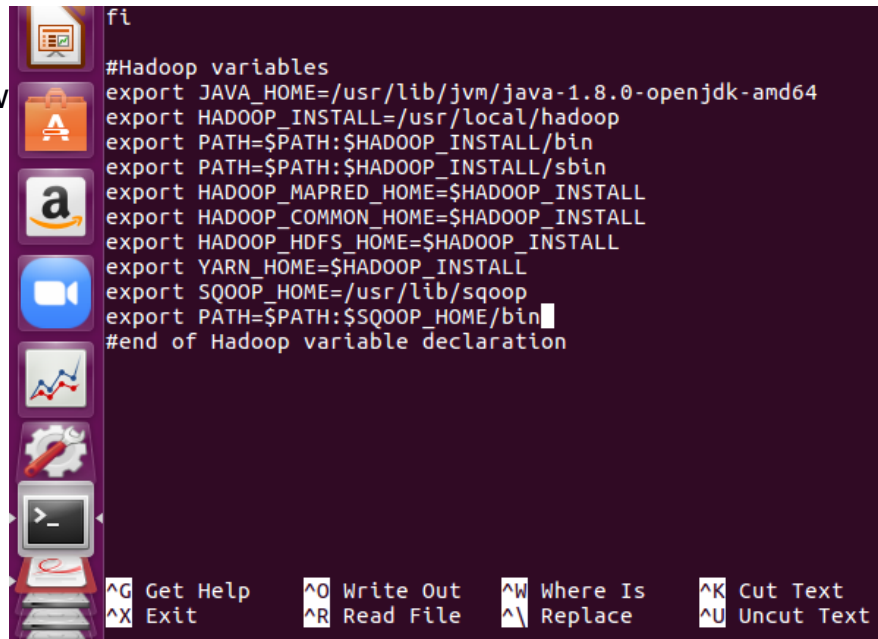


Append the following lines in this file.

```
export SQOOP_HOME=/usr/lib/sqoop
export PATH=$PATH:$SQOOP_HOME/bin
```

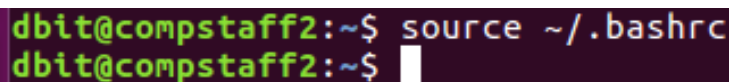
Ctrl+XY....Enter

4. Now
bashrc

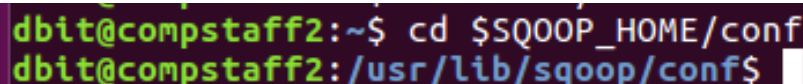


save this
file

permanently by the command **source ~/.bashrc**

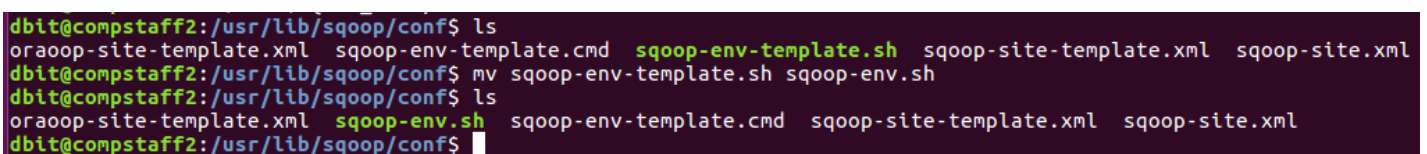


5. To configure Sqoop with Hadoop we need to edit a file **sqoop-env.sh** which is present in the directory path **\$SQOOP_HOME/conf**.



Now move the contents of the template file **sqoop-env-template.sh** to **sqoop-env.sh** using the **mv** command.

mv sqoop-env-template.sh sqoop-env.sh



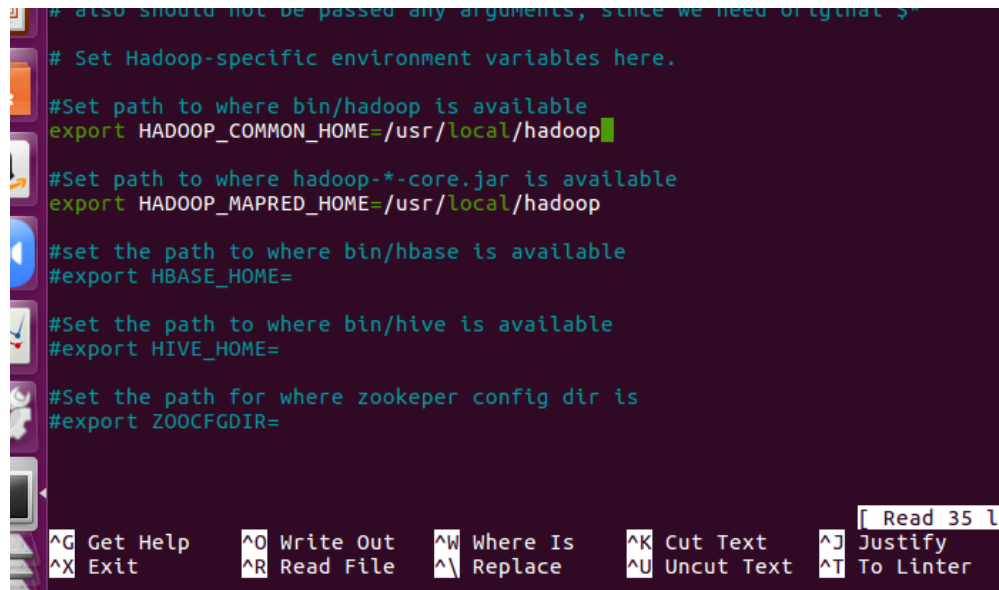
To add contents in the **sqoop-env.sh** file use the command:

nano sqoop-env.sh

```
dbit@compstaff2:/usr/lib/sqoop/conf$ nano sqoop-env.sh
```

Add these lines in the file code:

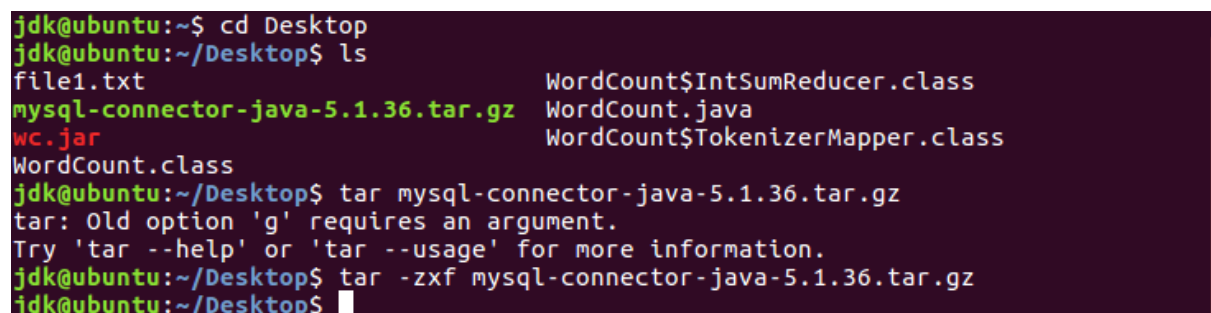
```
export HADOOP_COMMON_HOME=/usr/local/hadoop
export HADOOP_MAPRED_HOME=/usr/local/hadoop
```



```
# also should not be passed any arguments, since we need original $
# Set Hadoop-specific environment variables here.
#Set path to where bin/hadoop is available
export HADOOP_COMMON_HOME=/usr/local/hadoop
#Set path to where hadoop-*-core.jar is available
export HADOOP_MAPRED_HOME=/usr/local/hadoop
#set the path to where bin/hbase is available
#export HBASE_HOME=
#Set the path to where bin/hive is available
#export HIVE_HOME=
#Set the path for where zookeeper config dir is
#export ZOO_CFG_DIR=
```

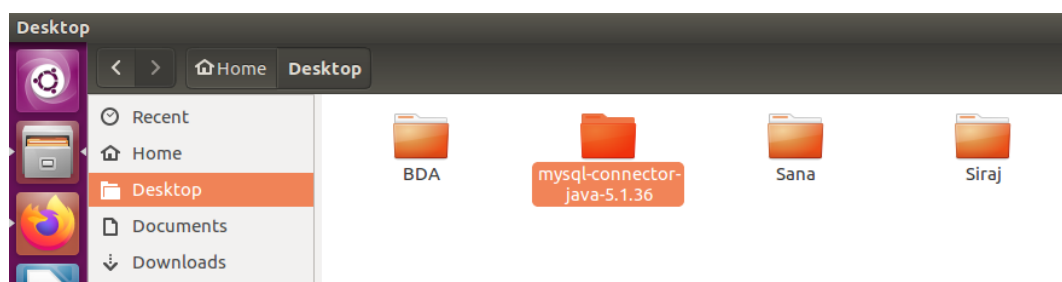
Ctrl+X...Y....Enter

6. Now copy, add or download the **mysql-connector-java-5.1.36.tar.gz** file onto the Desktop. Extract this file in the same file location.



```
jdk@ubuntu:~$ cd Desktop
jdk@ubuntu:~/Desktop$ ls
file1.txt                               WordCount$IntSumReducer.class
mysql-connector-java-5.1.36.tar.gz       WordCount.java
wc.jar                                  WordCount$TokenizerMapper.class
WordCount.class
jdk@ubuntu:~/Desktop$ tar mysql-connector-java-5.1.36.tar.gz
tar: Old option 'g' requires an argument.
Try 'tar --help' or 'tar --usage' for more information.
jdk@ubuntu:~/Desktop$ tar -zxf mysql-connector-java-5.1.36.tar.gz
jdk@ubuntu:~/Desktop$
```

OR can be done through GUI.



7. Move this extracted file to the location /usr/lib/sqoop/lib using the mv command.

```
cd Desktop
ls
cd mysql-connector-java-5.1.36
ls
mv mysql-connector-java-5.1.36-bin.jar /usr/lib/sqoop/lib
ls /usr/lib/sqoop/lib
```

```
ompstaff2: ~/Desktop/mysql-connector-java-5.1.36
dbit@compstaff2:~$ cd Desktop
dbit@compstaff2:~/Desktop$ ls
BDA mysql-connector-java-5.1.36 Sana Siraj
dbit@compstaff2:~/Desktop$ cd mysql-connector-java-5.1.36
dbit@compstaff2:~/Desktop/mysql-connector-java-5.1.36$ ls
build.xml CHANGES COPYING docs mysql-connector-java-5.1.36-bin.jar README README.txt src
dbit@compstaff2:~/Desktop/mysql-connector-java-5.1.36$ mv mysql-connector-java-5.1.36-bin.jar /usr/lib/sqoop/lib
dbit@compstaff2:~/Desktop/mysql-connector-java-5.1.36$ ls /usr/lib/sqoop/lib
ant-contrib-1.0b3.jar          hsqldb-1.8.0.10.jar          kite-hadoop-compatibility-1.0.0.jar  parquet-generator-1.4.1.jar
ant-eclipse-1.0-jvm1.2.jar    jackson-annotations-2.3.0.jar mysql-connector-java-5.1.36-bin.jar  parquet-hadoop-1.4.1.jar
avro-1.7.5.jar               jackson-core-2.3.1.jar      opencsv-2.3.jar                    parquet-jackson-1.4.1.jar
avro-mapred-1.7.5-hadoop2.jar jackson-core-asl-1.9.13.jar paranamer-2.3.jar                  slf4j-api-1.6.1.jar
commons-codec-1.4.jar         jackson-databind-2.3.1.jar  parquet-avro-1.4.1.jar              snappy-java-1.0.5.jar
commons-compress-1.4.1.jar    jackson-mapper-asl-1.9.13.jar parquet-column-1.4.1.jar            xz-1.0.jar
commons-io-1.4.jar           kite-data-core-1.0.0.jar    parquet-common-1.4.1.jar
commons-jexl-2.1.1.jar        kite-data-hive-1.0.0.jar    parquet-encoding-1.4.1.jar
commons-logging-1.1.1.jar     kite-data-mapreduce-1.0.0.jar parquet-format-2.0.0.jar
dbit@compstaff2:~/Desktop/mysql-connector-java-5.1.36$
```

8. To check if Sqoop has been installed correctly we move to the directory **\$SQOOP_HOME/bin** and use the command **sqoop version** to check for sqoop installation success.

```
cd /usr/lib/sqoop/bin
cd $SQOOP_HOME/bin
sqoop version
```

```
compstaff2: /usr/lib/sqoop/bin
dbit@compstaff2:~$ cd /usr/lib/sqoop/bin
dbit@compstaff2:/usr/lib/sqoop/bin$ cd $SQOOP_HOME/bin
dbit@compstaff2:/usr/lib/sqoop/bin$ sqoop version
Warning: /usr/lib/sqoop/./hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /usr/lib/sqoop/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/lib/sqoop/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
20/09/14 20:56:43 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Sqoop 1.4.6
git commit id c0c5a81723759fa575844a0a1eae8f510fa32c25
Compiled by root on Mon Apr 27 14:38:36 CST 2015
dbit@compstaff2:/usr/lib/sqoop/bin$
```

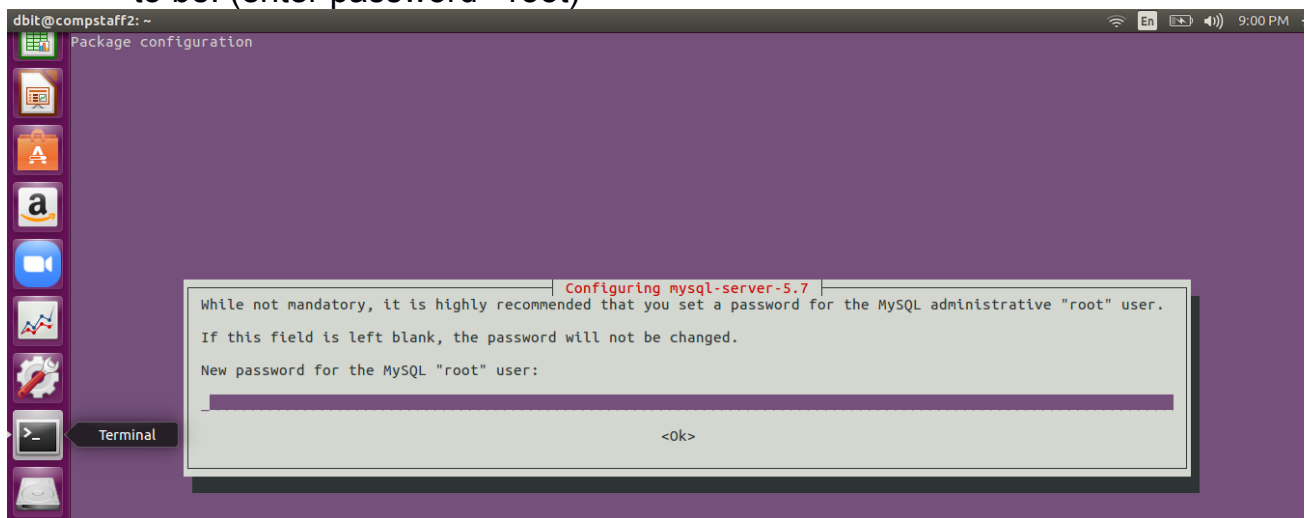
STEPS for MYSQL INSTALLATION

1. After Sqoop installation MySQL has to be installed as well. Firstly, install all the required libraries using the command **sudo apt-get install mysql-server**.

```
compstaff2: ~
dbit@compstaff2:~$ sudo apt-get install mysql-server
[sudo] password for dbit:
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-4.4.0-21 linux-headers-4.4.0-21-generic linux-image-4.4.0-21-generic linux-image-extra-4.4.0-21-generic
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  libaio1 libhtml-template-perl mysql-client-5.7 mysql-client-core-5.7 mysql-common mysql-server-5.7
Suggested packages:
  libipc-sharedcache-perl mailx tinyca
The following NEW packages will be installed:
  libaio1 libhtml-template-perl mysql-client-5.7 mysql-client-core-5.7 mysql-common mysql-server mysql-server-5.7
0 upgraded, 8 newly installed, 0 to remove and 302 not upgraded.
Need to get 17.8 MB of archives.
After this operation, 156 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
```

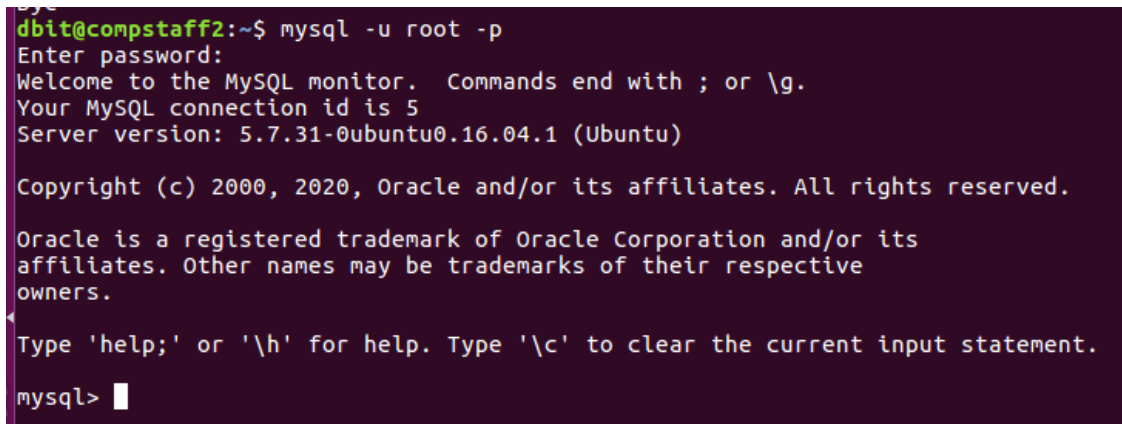
After getting this message press **Y** to continue. The libraries get downloaded successfully.

2. Now a screen like the one below appears. It prompts the user to set a password for the user "root". So, we can set the password as we wish it to be. (enter password - root)



3. To login to the MySQL user, use the following command:

mysql -u root -p

A terminal window with a dark purple background. The prompt is 'dbit@compstaff2:~\$'. The command 'mysql -u root -p' has been entered. The output shows 'Enter password:', 'Welcome to the MySQL monitor. Commands end with ; or \g.', 'Your MySQL connection id is 5', and 'Server version: 5.7.31-0ubuntu0.16.04.1 (Ubuntu)'. It also displays copyright and trademark information for Oracle. The prompt 'mysql>' is shown at the bottom with a cursor.

```
dbit@compstaff2:~$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 5
Server version: 5.7.31-0ubuntu0.16.04.1 (Ubuntu)

Copyright (c) 2000, 2020, 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> █
```

It will be asked to enter the password for the corresponding user. Enter the password. Now the MySQL script will run and the user will be logged in. This verifies the successful completion of the MySQL installation onto the system.