Installing MySQL on Mac

1. **Download MySQL from Homebrew and Install mysql.**

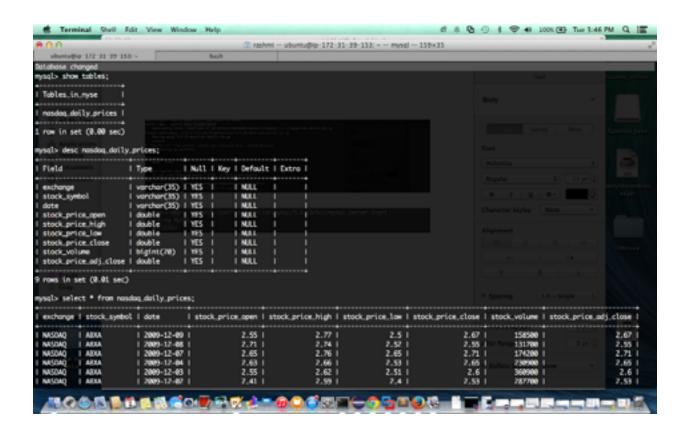2. **Star MySQL server.**

Creating Database, Table and loading the data

1. Open Mysql command line client, give username as "root" & password. MySQL command line will open. Now we can create tables, databases etc.

2. Create a new database and use new database using command given below.
create database nyse;
use nyse;

3. Create table and insert records

/usr/local/Cellar/mysql/5.6.19/bin/mysql -uroot nyse < tables.sql
/usr/local/Cellar/mysql/5.6.19/bin/mysql -uroot nyse < data.sql

Installing Sqoop on Mac

1. Download & untar sqoop-1.4.5.bin_hadoop-0.23.tar.gz

Rashmis-MacBook-Pro:Downloads rashmi$ tar -zxvf sqoop-1.4.5.bin__hadoop-0.23.tar.gz

2. Setup HADOO_HOME

```
Rashmis-MacBook-Pro:sqoop-1.4.5.bin__hadoop-0.23 rashmi$ export HADOOP_HOME=/usr/local/Cellar/hadoop/2.6.0/
```

3. Download MySQL connector.

```
Rashmis-MacBook-Pro:lib rashmi$ cp ~/Downloads/mysql-connector-java-5.1.17-bin.jar .
```

Importing data from MySQL to HDFS using Sqoop

1. Execute Scoop command to get the data from MySQL into HDFS.

```
Rashmis-MacBook-Pro:sqoop-1.4.5.bin__hadoop-2.0.4-alpha rashmi$ ./bin/sqoop import --connect jdbc:mysql://localhost/nyse  --table nasdaq_daily_prices --target-dir /user/rashmi/nasdaq_daily_prices --username root --split-by exchange
Warning: /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/../../hbase does not exist! HBase imports will fail.
Please set $HBASE_HOME to the root of your HBase installation.
Warning: /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/../../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/../../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/../../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
14/12/05 21:15:23 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
14/12/05 21:15:23 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
14/12/05 21:15:23 INFO tool.CodeGenTool: Beginning code generation
14/12/05 21:15:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nasdaq_daily_prices` AS t LIMIT 1
```

14/12/05 21:15:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nasdaq_daily_prices` AS t LIMIT 1

14/12/05 21:15:23 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/Cellar/hadoop/2.6.0

Note: /tmp/sqoop-rashmi/compile/ecd5bff9cfba766fd893f9ff88a4af15/nasdaq_daily_prices.java uses or overrides a deprecated API.

Note: Recompile with -Xlint:deprecation for details.

14/12/05 21:15:24 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-rashmi/compile/ecd5bff9cfba766fd893f9ff88a4af15/nasdaq_daily_prices.jar

14/12/05 21:15:24 WARN manager.MySQLManager: It looks like you are importing from mysql.

14/12/05 21:15:24 WARN manager.MySQLManager: This transfer can be faster! Use the --direct

14/12/05 21:15:24 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.

14/12/05 21:15:24 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)

14/12/05 21:15:24 INFO mapreduce.ImportJobBase: Beginning import of nasdaq_daily_prices

14/12/05 21:15:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

14/12/05 21:15:25 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar

14/12/05 21:15:25 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps

14/12/05 21:15:25 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032

14/12/05 21:15:27 INFO db.DBInputFormat: Using read commited transaction isolation

14/12/05 21:15:27 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`exchange`), MAX(`exchange`) FROM `nasdaq_daily_prices`

14/12/05 21:15:27 WARN db.TextSplitter: Generating splits for a textual index column.

14/12/05 21:15:27 WARN db.TextSplitter: If your database sorts in a case-insensitive order, this may result in a partial import or duplicate records.

14/12/05 21:15:27 WARN db.TextSplitter: You are strongly encouraged to choose an integral split column.

14/12/05 21:15:27 WARN db.BigDecimalSplitter: Set BigDecimal splitSize to MIN_INCREMENT

14/12/05 21:15:27 INFO mapreduce.JobSubmitter: number of splits:1

14/12/05 21:15:27 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1417842624965_0001

14/12/05 21:15:28 INFO impl.YarnClientImpl: Submitted application application_1417842624965_0001

14/12/05 21:15:28 INFO mapreduce.Job: The url to track the job: http://Rashmis-MacBook-Pro.local:8088/proxy/application_1417842624965_0001/

```
14/12/05 21:15:28 INFO mapreduce.Job: Running job: job_1417842624965_0001
14/12/05 21:15:37 INFO mapreduce.Job: Job job_1417842624965_0001 running in uber
mode : false
14/12/05 21:15:37 INFO mapreduce.Job:  map 0% reduce 0%
14/12/05 21:15:43 INFO mapreduce.Job:  map 100% reduce 0%
14/12/05 21:15:44 INFO mapreduce.Job: Job job_1417842624965_0001 completed
successfully
14/12/05 21:15:44 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=114886
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=125
                HDFS: Number of bytes written=538464
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Other local map tasks=1
                Total time spent by all maps in occupied slots (ms)=4634
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=4634
                Total vcore-seconds taken by all map tasks=4634
                Total megabyte-seconds taken by all map tasks=4745216
        Map-Reduce Framework
                Map input records=9999
                Map output records=9999
                Input split bytes=125
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=51
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=149946368
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=538464
```

14/12/05 21:15:44 INFO mapreduce.ImportJobBase: Transferred 525.8438 KB in 18.9459 seconds (27.755 KB/sec)
14/12/05 21:15:44 INFO mapreduce.ImportJobBase: Retrieved 9999 records.
Rashmis-MacBook-Pro:sqoop-1.4.5.bin__hadoop-2

2. Verify the data in HDFS.

Rashmis-MacBook-Pro:sqoop-1.4.5.bin__hadoop-2.0.4-alpha rashmi$ hadoop fs -ls /user/rashmi
14/12/05 21:16:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - rashmi supergroup   0 2014-12-05 21:15 /user/rashmi/nasdaq_daily_prices
Rashmis-MacBook-Pro:sqoop-1.4.5.bin__hadoop-2.0.4-alpha rashmi$ hadoop fs -ls /user/rashmi/nasdaq_daily_prices
14/12/05 21:16:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 rashmi supergroup   0 2014-12-05 21:15 /user/rashmi/nasdaq_daily_prices/_SUCCESS
-rw-r--r--   3 rashmi supergroup   538464 2014-12-05 21:15 /user/rashmi/nasdaq_daily_prices/part-m-00000

Run MapReduce program on the dataset

1. Run NYSE.jar MapReduce program on the input data. input is at location "/user/rashmi/nasdaq_daily_prices/part-m-00000 /" and the output is generated at the location "/user/rashmi/nasdaq_daily_prices/output"

Rashmis-MacBook-Pro:Desktop rashmi$ hadoop jar NYSE.jar com.rashmi.mapreduce.NYSE.NYSEJob /user/rashmi/nasdaq_daily_prices/part-m-00000 /user/rashmi/nasdaq_daily_prices/output
14/12/08 16:00:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

```
14/12/08 16:00:07 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
14/12/08 16:00:07 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing
not performed. Implement the Tool interface and execute your application with
ToolRunner to remedy this.
14/12/08 16:00:08 INFO input.FileInputFormat: Total input paths to process : 1
14/12/08 16:00:08 INFO mapreduce.JobSubmitter: number of splits:1
14/12/08 16:00:08 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1417842624965_0007
14/12/08 16:00:08 INFO impl.YarnClientImpl: Submitted application
application_1417842624965_0007
14/12/08 16:00:08 INFO mapreduce.Job: The url to track the job: http://Rashmis-
MacBook-Pro.local:8088/proxy/application_1417842624965_0007/
14/12/08 16:00:08 INFO mapreduce.Job: Running job: job_1417842624965_0007
14/12/08 16:00:14 INFO mapreduce.Job: Job job_1417842624965_0007 running in uber
mode : false
14/12/08 16:00:14 INFO mapreduce.Job:  map 0% reduce 0%
14/12/08 16:00:20 INFO mapreduce.Job:  map 100% reduce 0%
14/12/08 16:00:25 INFO mapreduce.Job:  map 100% reduce 100%
14/12/08 16:00:26 INFO mapreduce.Job: Job job_1417842624965_0007 completed
successfully
14/12/08 16:00:26 INFO mapreduce.Job: Counters: 49
        File System Counters
                FILE: Number of bytes read=257659
                FILE: Number of bytes written=727285
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=538595
                HDFS: Number of bytes written=44
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=3430
                Total time spent by all reduces in occupied slots (ms)=3251
                Total time spent by all map tasks (ms)=3430
                Total time spent by all reduce tasks (ms)=3251
                Total vcore-seconds taken by all map tasks=3430
                Total vcore-seconds taken by all reduce tasks=3251
                Total megabyte-seconds taken by all map tasks=3512320
```

```
                Total megabyte-seconds taken by all reduce tasks=3329024
        Map-Reduce Framework
                Map input records=9999
                Map output records=9999
                Map output bytes=237655
                Map output materialized bytes=257659
                Input split bytes=131
                Combine input records=0
                Combine output records=0
                Reduce input groups=3
                Reduce shuffle bytes=257659
                Reduce input records=9999
                Reduce output records=3
                Spilled Records=19998
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=62
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=358612992
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=538464
        File Output Format Counters
                Bytes Written=44
Rashmis-MacBook-Pro:Desktop rashmi$
```

```
Rashmis-MacBook-Pro:zookeeper rashmi$ hadoop fs -ls /user/rashmi/nasdaq_daily_prices/
output
14/12/08 16:00:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--   3 rashmi supergroup          0 2014-12-08 16:00 /user/rashmi/
nasdaq_daily_prices/output/_SUCCESS
```

```
-rw-r--r--   3 rashmi supergroup        44 2014-12-08 16:00 /user/rashmi/
nasdaq_daily_prices/output/part-r-00000
Rashmis-MacBook-Pro:zookeeper rashmi$ hadoop fs -text /user/rashmi/
nasdaq_daily_prices/output/part-r-00000
14/12/08 16:00:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
AAME      4.9223E7
ABXA      7.62147968E8
AIPC      6.2642592E8
Rashmis-MacBook-Pro:zookeeper rashmi$
```

Exporting MapReduce output to MySQL using Sqoop

1.Create table stock_volume in nyse MySQL database.

```
create table stock_volume ( stock_symbol VARCHAR(35) ,
total_stock_volume double) engine innodb;
```
2. Grant privileges on nyse MySQL database.

mysql> grant all privileges on nyse.* to ''@localhost ;

3. Execute sqoop export with all the required parameters to export MapReduce output into stock_volume table in MySQL

Rashmis-MacBook-Pro:zookeeper rashmi$ /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/sqoop export --connect jdbc:mysql://localhost/nyse --table stock_volume --export-dir /user/rashmi/nasdaq_daily_prices/output --username root --input-fields-terminated-by '\t' --input-lines-terminated-by '\n'
Warning: /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/../../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /Users/rashmi/sqoop-1.4.5.bin__hadoop-2.0.4-alpha/bin/../../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
14/12/08 16:03:51 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5
14/12/08 16:03:51 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
14/12/08 16:03:51 INFO tool.CodeGenTool: Beginning code generation

14/12/08 16:03:51 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nasdaq_daily_prices_total_volume` AS t LIMIT 1
14/12/08 16:03:51 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `nasdaq_daily_prices_total_volume` AS t LIMIT 1
14/12/08 16:03:51 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/Cellar/hadoop/2.6.0/libexec
Note: /tmp/sqoop-rashmi/compile/fa2c9bdd67e402b57c339a5a6b9435f2/nasdaq_daily_prices_total_volume.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
14/12/08 16:03:53 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-rashmi/compile/fa2c9bdd67e402b57c339a5a6b9435f2/nasdaq_daily_prices_total_volume.jar
14/12/08 16:03:53 INFO mapreduce.ExportJobBase: Beginning export of nasdaq_daily_prices_total_volume
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hadoop/2.6.0/libexec/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hbase/0.98.8/libexec/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/Cellar/hive/0.14.0/libexec/lib/hive-jdbc-0.14.0-standalone.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
14/12/08 16:03:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
14/12/08 16:03:53 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
14/12/08 16:03:54 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
14/12/08 16:03:54 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
14/12/08 16:03:54 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
14/12/08 16:03:54 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
14/12/08 16:03:55 INFO input.FileInputFormat: Total input paths to process : 1
14/12/08 16:03:55 INFO input.FileInputFormat: Total input paths to process : 1
14/12/08 16:03:55 INFO mapreduce.JobSubmitter: number of splits:4
14/12/08 16:03:55 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
14/12/08 16:03:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1417842624965_0008

14/12/08 16:03:55 INFO impl.YarnClientImpl: Submitted application application_1417842624965_0008

14/12/08 16:03:55 INFO mapreduce.Job: The url to track the job: http://Rashmis-MacBook-Pro.local:8088/proxy/application_1417842624965_0008/

14/12/08 16:03:55 INFO mapreduce.Job: Running job: job_1417842624965_0008

14/12/08 16:04:02 INFO mapreduce.Job: Job job_1417842624965_0008 running in uber mode : false

14/12/08 16:04:02 INFO mapreduce.Job: map 0% reduce 0%

14/12/08 16:04:12 INFO mapreduce.Job: map 25% reduce 0%

14/12/08 16:04:13 INFO mapreduce.Job: map 100% reduce 0%

14/12/08 16:04:13 INFO mapreduce.Job: Job job_1417842624965_0008 completed successfully

14/12/08 16:04:13 INFO mapreduce.Job: Counters: 30

    File System Counters

        FILE: Number of bytes read=0

        FILE: Number of bytes written=458004

        FILE: Number of read operations=0

        FILE: Number of large read operations=0

        FILE: Number of write operations=0

        HDFS: Number of bytes read=738

        HDFS: Number of bytes written=0

        HDFS: Number of read operations=16

        HDFS: Number of large read operations=0

        HDFS: Number of write operations=0

    Job Counters

        Launched map tasks=4

        Data-local map tasks=4

        Total time spent by all maps in occupied slots (ms)=33292

        Total time spent by all reduces in occupied slots (ms)=0

        Total time spent by all map tasks (ms)=33292

        Total vcore-seconds taken by all map tasks=33292

        Total megabyte-seconds taken by all map tasks=34091008

    Map-Reduce Framework

        Map input records=3

        Map output records=3

        Input split bytes=616

        Spilled Records=0

        Failed Shuffles=0

        Merged Map outputs=0

        GC time elapsed (ms)=174

        CPU time spent (ms)=0

        Physical memory (bytes) snapshot=0

        Virtual memory (bytes) snapshot=0

```
        Total committed heap usage (bytes)=541065216
    File Input Format Counters
        Bytes Read=0
    File Output Format Counters
        Bytes Written=0
14/12/08 16:04:13 INFO mapreduce.ExportJobBase: Transferred 738 bytes in 18.7436
seconds (39.3734 bytes/sec)
14/12/08 16:04:13 INFO mapreduce.ExportJobBase: Exported 3 records.
```

```
mysql> select * from stock_volume;
+--------------+--------------------+
| stock_symbol | total_stock_volume |
+--------------+--------------------+
| AIPC         |          626426000 |
| ABXA         |          762148000 |
| AAME         |           49223000 |
+--------------+--------------------+
3 rows in set (0.00 sec)
```