

Experiment No 2 : Using Sqoop

Aim: To install sqoop and execute basic commands .

Procedure:

1. After Installing sqoop we will use two basic commands like import and export which will help us import data from rdbms to hdfs and vice versa

2. For an example we are going to export a table products that is stored in database retail_db in rdbms to hdfs file system


3. To import data into hdfs we will use the command
>sqoop import --connect jdbc:mysql://localhost/retail_db
--username root --password cloudera --table products

4. To export data from hdfs to rdbms we will first have to create a table with the same description that we had in the hdfs with the help of create table

5. Then we will use the command
>sqoop export --connect jdbc:mysql://localhost//retail_db
--username root --password cloudera --table yo --export-dir
/user/cloudera/products

Result: You were able to import and export from hdfs to rdbms and vice versa .

Screenshots:



The screenshot shows a terminal window titled "cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox". The terminal output shows the execution of a sqoop import command and its output.

```
21/03/07 11:34:10 ERROR tool.BaseSqoopTool: Unrecognized argument: --table
21/03/07 11:34:10 ERROR tool.BaseSqoopTool: Unrecognized argument: products

Try --help for usage instructions.
[cloudera@quickstart ~]$ sqoop import --connect jdbc:mysql://localhost/retail_db --username root --password cloudera
--table products
Warning: /usr/lib/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
21/03/07 11:34:23 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.2
21/03/07 11:34:23 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P i
nstead.
21/03/07 11:34:24 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
21/03/07 11:34:24 INFO tool.CodeGenTool: Beginning code generation
21/03/07 11:34:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `products` AS t LIMIT 1
21/03/07 11:34:25 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `products` AS t LIMIT 1
21/03/07 11:34:25 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-cloudera/compile/3131d1741f6e69bca83c4ae970b569ab/products.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
21/03/07 11:34:29 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-cloudera/compile/3131d1741f6e69bca83c4ae9
70b569ab/products.jar
21/03/07 11:34:30 WARN manager.MySQLManager: It looks like you are importing from mysql.
21/03/07 11:34:30 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
21/03/07 11:34:30 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
21/03/07 11:34:30 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
21/03/07 11:34:30 INFO mapreduce.ImportJobBase: Beginning import of products
21/03/07 11:34:30 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker
.address
21/03/07 11:34:30 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
21/03/07 11:34:32 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
21/03/07 11:34:33 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
21/03/07 11:34:37 INFO db.DBInputFormat: Using read committed transaction isolation
21/03/07 11:34:37 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`product_id`), MAX(`product_id`) FRO
M `products`
21/03/07 11:34:37 INFO mapreduce.JobSubmitter: number of splits:4
21/03/07 11:34:38 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1614751261682_0001
21/03/07 11:34:39 INFO impl.YarnClientImpl: Submitted application application_1614751261682_0001
21/03/07 11:34:39 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application 161
4751261682_0001/
21/03/07 11:34:39 INFO mapreduce.Job: Running job: job_1614751261682_0001
21/03/07 11:34:58 INFO mapreduce.Job: Job job_1614751261682_0001 running in uber mode : false
21/03/07 11:34:58 INFO mapreduce.Job: map 0% reduce 0%
21/03/07 11:35:38 INFO mapreduce.Job: map 100% reduce 0%
21/03/07 11:35:40 INFO mapreduce.Job: Job job_1614751261682_0001 completed successfully
21/03/07 11:35:40 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=540104
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=474
  HDFS: Number of bytes written=173993
  HDFS: Number of read operations=16
```

The terminal window also shows the Cloudera desktop environment with a taskbar at the bottom containing icons for the terminal, a file browser, and a web browser.

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System

Sun Mar 7, 11:40 AM cloudera

Browsing HDFS - Mozilla Firefox

Browsing HDFS

localhost:50070/explorer.html#/user/cloudera/products

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

Browse Directory

/user/cloudera/products Go!

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	0 B	1	128 MB	_SUCCESS
-rw-r--r--	cloudera	cloudera	40.45 KB	1	128 MB	part-m-00000
-rw-r--r--	cloudera	cloudera	42.64 KB	1	128 MB	part-m-00001
-rw-r--r--	cloudera	cloudera	41.21 KB	1	128 MB	part-m-00002
-rw-r--r--	cloudera	cloudera	45.62 KB	1	128 MB	part-m-00003

Hadoop, 2014.

```
es.acmesports.sports/Under+Armour+Men%27s+Highlight+MC+Football+Cleat
| 990 | 45 | Under Armour WWP Freedom Flag T-Shirt | 27.99 | http://imag
es.acmesports.sports/Under+Armour+WWP+Freedom+Flag+T-Shirt
| 991 | 45 | Bushnell 2014 Tour V3 Patriot Pack Laser Rang | 299.99 | http://imag
es.acmesports.sports/Bushnell+2014+Tour+V3+Patriot+Pack+Laser+Rangefinder
| 992 | 45 | Under Armour Men's Highlight MC Football Clea | 129.99 | http://imag
es.acmesports.sports/Under+Armour+Men%27s+Highlight+MC+Football+Cleat
| 993 | 45 | Under Armour Men's Highlight MC Football Clea | 129.99 | http://imag
es.acmesports.sports/Under+Armour+Men%27s+Highlight+MC+Football+Cleat
| 994 | 45 | Primos Truth Cam Supercharged Blackout 7 MP G | 99.99 | http://imag
es.acmesports.sports/Primos+Truth+Cam+Supercharged+Blackout+7+MP+Game+Camera
| 995 | 45 | Bushnell 2014 Tour V3 Slope Patriot Pack Lase | 399.99 | http://imag
es.acmesports.sports/Bushnell+2014+Tour+V3+Slope+Patriot+Pack+Laser+Rangefinder
| 996 | 45 | Bushnell Golf Tour v3 Patriot Pack Laser Rang | 299.99 | http://imag
es.acmesports.sports/Bushnell+Golf+Tour+v3+Patriot+Pack+Laser+Rangefinder
| 997 | 45 | Under Armour Men's Highlight MC Football Clea | 129.99 | http://imag
es.acmesports.sports/Under+Armour+Men%27s+Highlight+MC+Football+Cleat
| 998 | 45 | Bushnell NEO-X Golf GPS Watch | 179.99 | http://imag
es.acmesports.sports/Bushnell+NEO-X+Golf+GPS+Watch
| 999 | 45 | Timberland PRO Men's Pit Boss 6" Steel Toe Wo | 99.99 | http://imag
es.acmesports.sports/Timberland+PRO+Men%27s+Pit+Boss+6"+Steel+Toe+Work+Boot
| 1000 | 45 | Nikon Monarch 5 10x42 Binoculars | 329.99 | http://imag
es.acmesports.sports/Nikon+Monarch+5+10x42+Binoculars
| 1001 | 45 | GoPro HERO3+ Silver Edition Camera | 299.99 | http://imag
es.acmesports.sports/GoPro+HERO3+Silver+Edition+Camera
| 1002 | 45 | Under Armour Women's Ignite VI Camo Slide | 34.99 | http://imag
es.acmesports.sports/Under+Armour+Women%27s+Ignite+VI+Camo+Slide
| 1003 | 45 | Nike Men's Free TR 5.0 (LSA Pack) Training Sh | 109.99 | http://imag
es.acmesports.sports/Nike+Men%27s+Free+TR+5.0+%28LSA+Pack%29+Training+Shoe
| 1004 | 45 | Field & Stream Sportsman 16 Gun Fire Safe | 399.98 | http://imag
es.acmesports.sports/Field+%26+Stream+Sportsman+16+Gun+Fire+Safe
| 1005 | 45 | Diamond Infinite Edge RTH Compound Bow Packag | 349.99 | http://imag
es.acmesports.sports/Diamond+Infinite+Edge+RTH+Compound+Bow+Package+-+Mossy+Oak...
| 1006 | 45 | O'Brien Neoprene Pet Life Vest | 39.99 | http://imag
es.acmesports.sports/O%27Brien+Neoprene+Pet+Life+Vest
| 1007 | 45 | Realtree Outfitters Men's Cobra Realtree Xtra | 49.99 | http://imag
es.acmesports.sports/Realtree+Outfitters+Men%27s+Cobra+Realtree+Xtra+Hiking+Shoe
| 1008 | 45 | Cuddeback Ambush IR Scouting Camera | 119.99 | http://imag
es.acmesports.sports/Cuddeback+Ambush+IR+Scouting+Camera
+-----+-----+-----+-----+-----+-----+
1345 rows in set (0.00 sec)

mysql> describe yo
-> ;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| product_id | int(11) | NO | PRI | NULL | |
| product_category_id | int(11) | YES | | NULL | |
| product_name | varchar(45) | YES | | NULL | |
| product_description | varchar(255) | YES | | NULL | |
| product_price | float | YES | | NULL | |
| product_image | varchar(255) | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
6 rows in set (0.00 sec)

mysql> S
```

cloudera@quickstart:~ cloudera@quickstart:~ Browsing HDFS - Mozil...

Right Ctrl

cloudera-quickstart-vm-5.4.2-0-virtualbox [Running] - Oracle VM VirtualBox

File Machine View Input Devices Help

Applications Places System Sun Mar 7, 12:05 PM cloudera

cloudera@quickstart:~

File Edit View Search Terminal Help

```
21/03/07 12:00:45 INFO input.FileInputFormat: Total input paths to process : 4
21/03/07 12:00:45 INFO mapreduce.JobSubmitter: number of splits:3
21/03/07 12:00:45 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use
mapreduce.map.speculative
21/03/07 12:00:45 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1614751261682_0002
21/03/07 12:00:46 INFO impl.YarnClientImpl: Submitted application application_1614751261682_0002
21/03/07 12:00:46 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_161
4751261682_0002/
21/03/07 12:00:46 INFO mapreduce.Job: Running job: job_1614751261682_0002
21/03/07 12:01:00 INFO mapreduce.Job: Job job_1614751261682_0002 running in uber mode : false
21/03/07 12:01:00 INFO mapreduce.Job: map 0% reduce 0%
21/03/07 12:01:32 INFO mapreduce.Job: map 100% reduce 0%
21/03/07 12:01:33 INFO mapreduce.Job: Job job_1614751261682_0002 completed successfully
21/03/07 12:01:33 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=404394
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=178670
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=21
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
  Job Counters
    Launched map tasks=3
    Data-local map tasks=3
    Total time spent by all maps in occupied slots (ms)=89790
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=89790
    Total vcore-seconds taken by all map tasks=89790
    Total megabyte-seconds taken by all map tasks=91944960
  Map-Reduce Framework
    Map input records=1345
    Map output records=1345
    Input split bytes=696
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=1479
    CPU time spent (ms)=4110
    Physical memory (bytes) snapshot=313942016
    Virtual memory (bytes) snapshot=4512583680
    Total committed heap usage (bytes)=182255616
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=0
21/03/07 12:01:34 INFO mapreduce.ExportJobBase: Transferred 174.4824 KB in 52.0674 seconds (3.3511 KB/sec)
21/03/07 12:01:34 INFO mapreduce.ExportJobBase: Exported 1345 records.
[cloudera@quickstart ~]$ sqoop export --connect jdbc:mysql://localhost/retail_db --username root --password cloudera
--table yo --export-dir /user/cloudera/products
```

cloudera@quickstart:~ cloudera@quickstart:~ Browsing HDFS - Mozil...

Right Ctrl