

HADOOP RDBMS TO HDFS

www.HadoopExam.com in Association with QuickTechie.com

Importing data from RDBMS to HDFS (Single table import)

Note: These instructions should be used with the HadoopExam CCP:Data Engineer Training. Where it is executed and you can do hands on with trainer.

1. Connecting to existing MySQL Database

```
mysql --user=retail_dba --password=cloudera retail_db
```

2. Show all the available tables

```
show tables;
```

3. View/Count data from a table in MySQL

```
select count(1) from categories;
```

```
select * from categories;
```

4. Check the currently available data in HDFS directory

```
hdfs dfs -ls
```

5. Import Single table

```
sqoop import \  
  --connect jdbc:mysql://quickstart:3306/retail_db \  
  --username=retail_dba \  
  --password=cloudera \  
  --table=categories
```

Important Note: Please check you don't have space between before or after '=' sign.
Sqoop use the MapReduce framework to copy data from RDBMS to hdfs

6. Check the result

```
hdfs dfs -ls
```

7. Check the content inside the directory

```
hdfs dfs -ls categories
```

8. Read the data from one of the partition

```
hdfs dfs -cat categories/part-m-00000
```

Importing data from RDBMS to HDFS (Single table import) with Specifying a Target Directory

1. Specifying target directory in import command

```
sqoop import \  
  --connect jdbc:mysql://quickstart:3306/retail_db \  
  --username=retail_dba \  
  --password=cloudera \  
  --table=categories \  
  --target-dir=categories_target
```

2. Check that data has been copied or not

```
hdfs dfs -ls
```

3. Check the content in output directory

```
hdfs dfs -ls categories_target
```

4. Check the content in one of the partition file.

```
hdfs dfs -cat categories_target/part-m-00000
```

5. **Specifying parent directory so that you can copy more than one table in a specified target directory.**

6. Command to specify warehouse directory

```
sqoop import \  
  --connect jdbc:mysql://quickstart:3306/retail_db \  
  --username=retail_dba \  
  --password=cloudera \  
  --table=categories \  
  --warehouse-dir=categories_warehouse
```

7. Check the output

```
hdfs dfs -ls
```

8. See the contest in the directory

hdfs dfs -ls categories_warehouse

9. See the content specific to table

hdfs dfs -ls categories_warehouse/categories

10. See the content in one of the file (partition)

hdfs dfs -cat categories_warehouse/categories/part-m-00000

Please check other Material Provided by www.HadoopExam.com

 430 Q & A Click Here Cloudera Hadoop Developer Certification CCD-410	 200 + Q & A Click Here Cloudera Hadoop Administrator Certification CCA-500	 262 Q & A Click Here Cloudera Hadoop HBase Certification CCB-400	 266 Q & A Click Here AWS Developer Certification Associate Level
 235 Q & A Click Here Cloudera Data Science Certification DS-200	 Click Here Hadoop Training With HandsOn	 Click Here Package Deal	 144 Q & A Click Here AWS Certified Solutions Architect Professional Level
 300 + Q & A Click Here AWS Certified SysOps Administrator Associate Level	 474 + Q & A Click Here AWS Certified Solutions Architect Associate Level	 Click Here Click Here for AWS Package Deal	 490 Q & A Click Here SAS Base Certification A00-211

 365 Q & A Click Here SAS Advance Certification A00-212	 86+ Q & A Click Here SAS Certified Statistical Business Analyst A00-240	 85 Q & A Click Here SAS Certified Platform Administrator 9 A00-250	 Click Here SAS Packaged Deal
---	--	--	---



234 Q & A
[Click Here](#)

EMC Data Scientist Associate Certification
E20-007 (EMCDSA)

Data Science certification really needs a good and in depth knowledge of statistics cum BigData Hadoop knowledge. It also require you to have good knowledge in like the main phases of the Data Analytics Lifecycle, analyzing and exploring data with R, statistics for model building and evaluation, the theory and methods of advanced analytics and statistical modeling, the technology and tools that can be used for advanced analytics, operationalizing an analytics project, and data visualization techniques. Successful candidates will achieve the EMC Proven Professional – Data Science Associate credential. Hence to clear the real exam it really needs very well preparation. So HadoopExam Learning Resources brings Data Science Certification Simulator with 234 Practice Questions, which can help you to prepare for this exam in lesser time. **Practice - practice - practice!** The EMC:DS E20-007 Exam Simulator offers you the opportunity to take 4 sample Exams before heading out for the real thing. Be ready to succeed on exam day!

Upcoming Releases

1. [Apache Spark Training](#)
2. [Apache Spark Certification material](#)
3. [MongoDB Certification Material](#)
4. [Android Certification](#)
5. [Java Certification](#)
6. [AWS Trainings](#)
7. [Data Science Training](#)