



Vineeta Singh

Jun 1, 2023, 5:51 PM (5 days ago) ☆ ↶ ⋮

to me ▾

Database credentials Username: [bigdatamind4385](#) Password: hadooplab234 Database: bigdatamind4385 Command to login MySQL: Command : mysql -u bigdatamind4385 -phadooplab234 Note: user has no privileges to create Database, we already created one database for user. User have to use that created database only.

Thanks a lot.

Thanks for the mail.

Got it, thanks!

```
result1
result_RohanB
result_S
result_a
result_aakash
result_abhinav
result_aman
result_aman1
result_amol
result_anjana
result_ankita
result_apurva
result_aviral
result_balwant
result_chandra
result_chetan
result_divya
result_divyani
result_gaurav
result_gauravb
result_mandar
result_mandar_1
result_mv
result_piyush
result_prajakta
result_sailesh
result_sandeep
result_shubham
result_spurthi
result_sumit
student_master
-----
35 rows in set (0.00 sec)

mysql> create database my_db;
ERROR 1044 (42000): Access denied for user 'bigdatamind4385'@'%' to database 'my_db'
mysql> create database bigdatalab46644;
ERROR 1044 (42000): Access denied for user 'bigdatamind4385'@'%' to database 'bigdatalab46644'
mysql> select * from student_master;
ERROR 1146 (42S02): Table 'bigdatamind4385.student_master' doesn't exist
mysql>
```

For local linux based

User: root

Password: password


```

ns_err_pid12234.log ns_err_pid13207.log ns_err_pid16000.log ns_err_pid22730.log ns_err_pid20311.log ns_err_pid30041.log ns_err_pid40643.log ns_err_pid303.log
hs_err_pid12296.log hs_err_pid15211.log hs_err_pid18863.log hs_err_pid22758.log hs_err_pid26369.log hs_err_pid30854.log hs_err_pid4703.log hs_err_pid9291.log
[bigdatalab456422@ip-10-1-1-204 opt]$ ls
anaconda3
apache-mahout-distribution-0.12.2.tar.gz cloudera conda mahout-0.12.2 sbt scripts
[bigdatalab456422@ip-10-1-1-204 opt]$ cd cloudera/
[bigdatalab456422@ip-10-1-1-204 cloudera]$ ls
cm cm-agent cm_cdp_cdh_log4j_jndi_removal.sh csd installer log4shell-backup parcel-cache parcel-repo parcels
[bigdatalab456422@ip-10-1-1-204 cloudera]$ cd parcels
[bigdatalab456422@ip-10-1-1-204 parcels]$ ls -l CDH
-rwxrwxrwx 1 root root 31 Jan 13 2021 CDH -> CDH-6.2.1-1.cdh6.2.1.p0.1425774
[bigdatalab456422@ip-10-1-1-204 parcels]$ cd CDH
[bigdatalab456422@ip-10-1-1-204 CDH]$ ls
bin etc include jars lib lib64 libexec meta share
[bigdatalab456422@ip-10-1-1-204 CDH]$ cd etc
[bigdatalab456422@ip-10-1-1-204 etc]$ ls
bash_completion.d flume-ng hadoop-httpfs hbase hive hive-webhcat impala kudu pig security solr sqoop
default hadoop-kms hadoop-kms hbase-solr hive-hcatalog hue kafka oozie rc.d sentry spark zookeeper
[bigdatalab456422@ip-10-1-1-204 etc]$ pwd
/opt/cloudera/parcels/CDH/etc
[bigdatalab456422@ip-10-1-1-204 etc]$ ls hive
conf.dist
[bigdatalab456422@ip-10-1-1-204 etc]$ cd hive
[bigdatalab456422@ip-10-1-1-204 hive]$ cd conf.dist/
[bigdatalab456422@ip-10-1-1-204 conf.dist]$ ls
beeline-log4j2.properties.template hive-exec-log4j2.properties hive-site.xml llap-cli-log4j2.properties.template parquet-logging.properties
hive-env.sh.template hive-log4j2.properties ivysettings.xml llap-daemon-log4j2.properties.template
[bigdatalab456422@ip-10-1-1-204 conf.dist]$

```

→
While joining an important data with some other lookup table, always use left outer join so that you don't lose any data

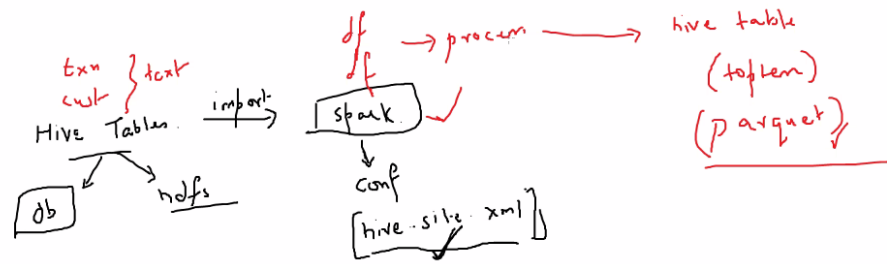
```

2023-06-06 11:25:28,659 Stage-2 map = 100%, reduce = 0%, Cumulative CPU 2.87 sec
2023-06-06 11:25:40,137 Stage-2 map = 100%, reduce = 100%, Cumulative CPU 2.87 sec
MapReduce Total cumulative CPU time: 5 seconds 290 msec
Ended Job = job_1685754149182_0790
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.9 sec HDFS Read: 4427025 HDFS Write: 288019 HDFS EC Read: 0 SUCCESS
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 5.29 sec HDFS Read: 293511 HDFS Write: 366 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 16 seconds 190 msec
OK
4009485 1973.3
4006425 1732.09
4000221 1671.47
4003228 1640.63
4006606 1628.94
4006467 1605.95
4004927 1576.71
4008321 1560.79
4000815 1557.82
4001051 1488.67
Time taken: 77.772 seconds, Fetched: 10 row(s)
hive (sandeep_training)> select * from customer where custno = 4009485;
Query ID = bigdatalab45644_20230606112618_b84318a3-2a58-41e1-b58f-19323e66a3d8
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
23/06/06 11:26:18 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
23/06/06 11:26:18 INFO client.RMProxy: Connecting to ResourceManager at ip-10-1-1-204.ap-south-1.compute.internal/10.1.1.204:8032
Starting Job = job_1685754149182_0794, Tracking URL = http://ip-10-1-1-204.ap-south-1.compute.internal:6066/proxy/application_1685754149182_0794/
Kill Command = /opt/cloudera/parcels/CDH-6.2.1-1.cdh6.2.1.p0.1425774/lib/hadoop/bin/hadoop job -kill job_1685754149182_0794
Hadoop job information for Stage=1: number of mappers: 1; number of reducers: 0
2023-06-06 11:26:28,270 Stage-1 map = 0%, reduce = 0%
2023-06-06 11:26:37,600 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.4 sec
MapReduce Total cumulative CPU time: 4 seconds 400 msec
Ended Job = job_1685754149182_0794
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Cumulative CPU: 4.4 sec HDFS Read: 397076 HDFS Write: 131 HDFS EC Read: 0 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 400 msec
OK
4009485,Stuart House 58 Teacher
Time taken: 20.436 seconds, Fetched: 1 row(s)
hive (sandeep_training)>

```

Handwritten red annotations:

- A bracket on the left side of the ID list (4009485 to 4001051) is labeled "Topken list".
- To the right of the list, "50,000" and "10,000" are written.
- Below these, a flow diagram shows: `imp -> Join -> lookup`.
- Below the flow diagram, it says "left outer join".



spark can read data fm any src
write to any dest