

Creating Database in ITV Hive

So in order to use hive we have to create our own hive warehouse directory after launching hive cli.

```
set hive.metastore.warehouse.dir=/user/Your_Username/warehouse;
```

Currently it is set to user/hive/warehouse. See below:

```
hive> set hive.metastore.warehouse.dir;  
hive.metastore.warehouse.dir=/user/hive/warehouse
```

Change to your user dedicated path:

```
set hive.metastore.warehouse.dir=/user/<your-itv-username>/warehouse; hive>
```

```
set hive.metastore.warehouse.dir=/user/itv000197/warehouse; After that you can
```

create your own database & start using that db.

Also put this property in the .hiverc file. So that you don't have to set in every session.

```
[itv000197@g02 ~]$ cat .hiverc  
set hive.metastore.warehouse.dir=/user/itv000197/warehouse;  
[itv000197@g02 ~]$ █
```

Refer:

<https://discuss.itversity.com/t/unable-to-create-or-use-hive-database-getting-permission-denied-error/23169>

Connecting Beeline

beeline

!connect jdbc:hive2://m02.itversity.com:10000/;auth=noSasl

```
[itv000197@g02 ~]$ beeline ✓
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoop-3.3.0/share/hadoop/common/lib/slf4j-log4j12-1.7.25.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.apache.logging.slf4j.Log4jLoggerFactory]
Beeline version 3.1.2 by Apache Hive
beeline> !connect jdbc:hive2://m02.itversity.com:10000/;auth=noSasl ✓
Connecting to jdbc:hive2://m02.itversity.com:10000/;auth=noSasl
Enter username for jdbc:hive2://m02.itversity.com:10000/: itv000197
Enter password for jdbc:hive2://m02.itversity.com:10000/: *****
Connected to: Apache Hive (version 3.1.2)
Driver: Hive JDBC (version 3.1.2)
Transaction isolation: TRANSACTION_REPEATABLE_READ
0: jdbc:hive2://m02.itversity.com:10000/> show databases;
```

Provide your lab username & password.

Problem 1 : TableSample does not work

```
itv622793@g02:~ x itv622793@g02:~ x
hive> show partitions products_partitioned_buckets;
OK
category=fashion
category=furniture
category=mobile
category=mobiles
category=toys
Time taken: 0.184 seconds, Fetched: 5 row(s)
hive> select * from products_partitioned_buckets TABLESAMPLE(bucket 1 out of 4);
OK
7      makeup  100.0   fashion
10     table   269.0   furniture
6      truck   7.99    toys
2      doll    8.99    toys
Time taken: 0.233 seconds, Fetched: 4 row(s)
hive> █
```

Cloudera Image

```
cloudera@quickstart:~
hive> select * from products_partitioned_buckets
> TABLESAMPLE(bucket 1 out of 4);
OK
8      earrings 69.0    fashion
Time taken: 0.111 seconds, Fetched: 1 row(s)
hive> █
```

Problem 2 : map side join hint does not work

```

hive> set hive.auto.convert.join;
hive.auto.convert.join=false
hive> set hive.ignore.mapjoin.hint;
hive.ignore.mapjoin.hint=false
hive> select /*+ mapjoin(o) */ o.order_customer_id,o.order_id,o.order_date,o.order_status,c.customer_id,c.customer_fname,c.customer_l
name,c.customer_state from orders o join customers c on (o.order_customer_id = c.customer_id) limit 5;
Query ID = itv000307_20210516054039_08139383-deb4-4164-a77c-9429e6bbc632
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reducers=<number>
Starting Job = job_1615503744169_16974, Tracking URL = http://m02.itversity.com:19088/proxy/application_1615503744169_16974/
Kill Command = /opt/hadoop/bin/mapred job -kill job_1615503744169_16974
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-05-16 05:40:48,707 Stage-1 map = 0%, reduce = 0%
2021-05-16 05:40:54,025 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.34 sec
2021-05-16 05:40:58,293 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.12 sec
MapReduce Total cumulative CPU time: 8 seconds 120 msec
Ended Job = job_1615503744169_16974
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 8.12 sec HDFS Read: 3975726 HDFS Write: 443 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 120 msec
OK
1 22945 2013-12-13 00:00:00.0 COMPLETE 1 Richard Hernandez TX
2 57963 2013-08-02 00:00:00.0 ON_HOLD 2 Mary Barrett CO
2 33865 2014-02-18 00:00:00.0 COMPLETE 2 Mary Barrett CO
2 15192 2013-10-29 00:00:00.0 PENDING_PAYMENT 2 Mary Barrett CO
2 67863 2013-11-30 00:00:00.0 COMPLETE 2 Mary Barrett CO
Time taken: 19.943 seconds, Fetched: 5 row(s)

```

Cloudera image

```

hive> set hive.ignore.mapjoin.hint=false;
hive> set hive.auto.convert.join=false;
hive> SELECT /*+ MAPJOIN(o) */ c.customer_id, c.customer_fname, c.customer_lname
, o.order_id, o.order_date FROM orders o JOIN customers c ON (o.order_customer_i
d = c.customer_id) limit 5;
Query ID = cloudera_20210415110808_a44067a5-b98c-41c7-bb9d-642cf98bbfb4
Total jobs = 1
Execution log at: /tmp/cloudera/cloudera_20210415110808_a44067a5-b98c-41c7-bb9d-
642cf98bbfb4.log
2021-04-15 11:08:15 Starting to launch local task to process map join; m
aximum memory = 1013645312
2021-04-15 11:08:19 Dump the side-table for tag: 0 with group count: 12405 i
nto file: file:/tmp/cloudera/965b3043-7786-4801-bda5-f7d5e00e4cb2/hive 2021-04-1
5 11-08-03 688 8471395495282544247-1/-local-10003/HashTable-Stage-1/MapJoin-o-10
--.hashtable
2021-04-15 11:08:20 Uploaded 1 File to: file:/tmp/cloudera/965b3043-7786-480
1-bda5-f7d5e00e4cb2/hive 2021-04-15 11-08-03 688 8471395495282544247-1/-local-10
003/HashTable-Stage-1/MapJoin-o-10--.hashtable (2278888 bytes)
2021-04-15 11:08:20 End of local task; Time Taken: 4.846 sec.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1618506163761_0002, Tracking URL = http://quickstart.cloudera
:8088/proxy/application_1618506163761_0002/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1618506163761_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2021-04-15 11:08:40,144 Stage-1 map = 0%, reduce = 0%
2021-04-15 11:08:56,994 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.58 se
c
MapReduce Total cumulative CPU time: 3 seconds 580 msec
Ended Job = job_1618506163761_0002
MapReduce Jobs Launched:

```

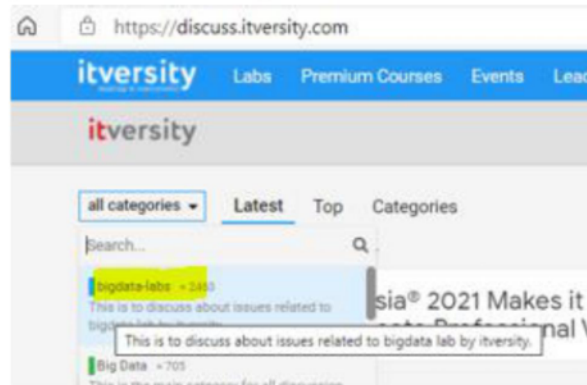


WEEK4 SUPPORT FOR ITVERSITY BY TRENDYTECH

Other Information:

You can search as well as post your queries as mentioned below.

1. Web → [discuss.ITVersity.com](https://discuss.itversity.com)([ITVersity](https://discuss.itversity.com)) → make sure to post your queries inside bigdata-labs.



2. ITV Slack - Join "labs-support channel".

https://join.slack.com/t/itv-data-engineers/shared_invite/zt-cmyt24hd-oxY8_YfCtLew~~Q_kg3U2~w

You can also post a query in "labs-support channel" ITV Slack

Note: In case if you can't see the labs-support channel then you can search for that channel and join.

3. You can also email on support@itversity.com