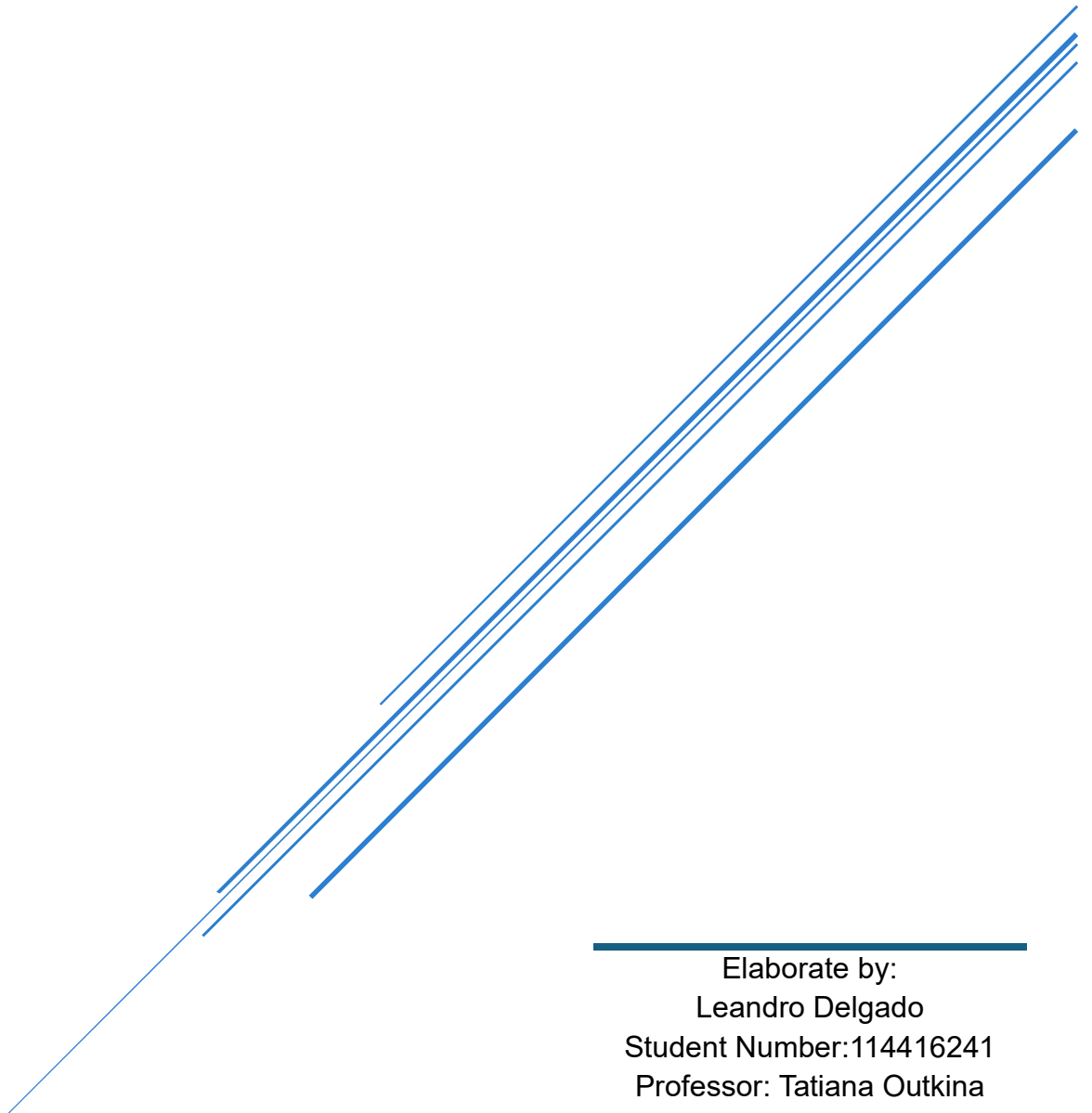
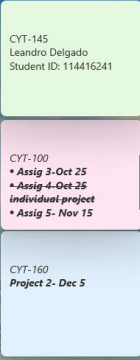


PROJECT 3

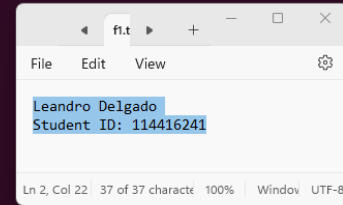
Install and configure Hive, run simple HQL queries



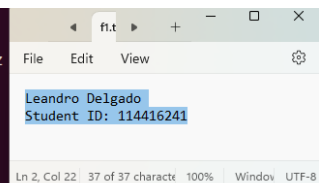
Elaborate by:
Leandro Delgado
Student Number:114416241
Professor: Tatiana Outkina
Course: CYT-180



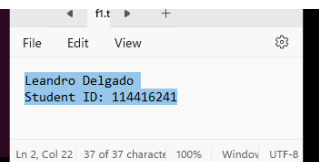
```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~  
leandro-dash-cyt-180@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ su hadoop  
Password:  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/home/leandro-dash-cyt-180$ start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.  
WARNING: This is not a recommended production deployment configuration.  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [localhost]  
Starting datanodes  
Starting secondary namenodes [leandro-dash-cyt-180-VMware-Virtual-Platform]  
Starting resourcemanager  
Starting nodemanagers  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/home/leandro-dash-cyt-180$ jps  
3504 SecondaryNameNode  
3156 NameNode  
3895 NodeManager  
4247 Jps  
3771 ResourceManager  
3278 DataNode  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/home/leandro-dash-cyt-180$ cd hadoop/share/hadoop/mapreduce  
bash: cd: hadoop/share/hadoop/mapreduce: Permission denied  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/home/leandro-dash-cyt-180$ ls  
ls: cannot open directory '.': Permission denied  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/home/leandro-dash-cyt-180$ wget https://dldcn.apache.org/hive/hive-4.0.1/apache-hive-4.0.1-bin.tar.gz  
pathconf: Permission denied  
--2024-10-23 22:19:31-- https://dldcn.apache.org/hive/hive-4.0.1/apache-hive-4.0.1-bin.tar.gz  
Resolving dldcn.apache.org (dldcn.apache.org)... 151.101.2.132, 2a04:4e42::644  
Connecting to dldcn.apache.org (dldcn.apache.org)|151.101.2.132|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 408225272 (389M) [application/x-gzip]  
apache-hive-4.0.1-bin.tar.gz: Permission denied
```



```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/home/leandro-dash-cyt-180$ cd ~  
wget https://dldcn.apache.org/hive/hive-4.0.1/apache-hive-4.0.1-bin.tar.gz  
--2024-10-23 22:27:59-- https://dldcn.apache.org/hive/hive-4.0.1/apache-hive-4.0.1-bin.tar.gz  
Resolving dldcn.apache.org (dldcn.apache.org)... 151.101.2.132, 2a04:4e42::644  
Connecting to dldcn.apache.org (dldcn.apache.org)|151.101.2.132|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 408225272 (389M) [application/x-gzip]  
Saving to: 'apache-hive-4.0.1-bin.tar.gz'  
  
apache-hive-4.0.1-bin.tar.gz 100%[=====] 389.31M 7.55MB/s in 42s  
  
2024-10-23 22:28:41 (9.20 MB/s) - 'apache-hive-4.0.1-bin.tar.gz' saved [408225272/408225272]  
  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$
```



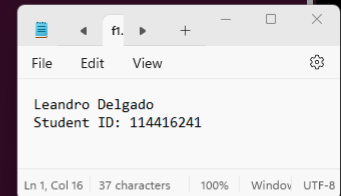
```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ tar -xzf apache-hive-4.0.1-bin.tar.gz  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ ls  
apache-hive-4.0.1-bin Desktop Downloads hadoop-3.3.6.tar.gz Music Public Templates  
apache-hive-4.0.1-bin.tar.gz Documents hadoop hadoopdata Pictures snap Videos  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ mv apache-hive-4.0.1-bin hive  
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$
```



```

hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive$ ls
bin  conf  contrib  examples  hcatalog  jdbc  lib  LICENSE  licenses  licenses.xml  NOTICE  RELEASE_NOTES.txt  scripts
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive$ cd conf
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ ls
beeline-log4j2.properties.template  hive-exec-log4j2.properties.template  llap-cli-log4j2.properties.template
hive-default.xml.template            hive-log4j2.properties.template        llap-daemon-log4j2.properties.template
hive-env.sh.template                 ivysettings.xml                        parquet-logging.properties
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ cp hive-default.xml.template hive-site.xml
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ nano hive-site.xml
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ nano ~/.bashrc
nano: invalid option -- '/'
Type 'nano -h' for a list of available options.
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ nano ~/.bashrc:
nano: invalid option -- '/'
Type 'nano -h' for a list of available options.
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ nano ~/.bashrc
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ source ~/.bashrc
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -ls
ls: '.': No such file or directory
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x - hadoop supergroup          0 2024-10-04 16:58 /logs
drwxr-xr-x - hadoop supergroup          0 2024-10-04 16:57 /test1
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -rm -r /test1
Deleted /test1
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x - hadoop supergroup          0 2024-10-04 16:58 /logs
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -mkdir -p /user/hive/warehouse
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -ls /

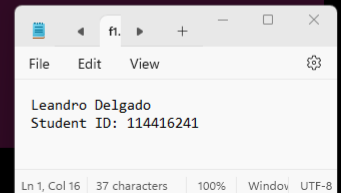
```



```

hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/hive/conf
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x - hadoop supergroup          0 2024-10-04 16:58 /logs
drwxr-xr-x - hadoop supergroup          0 2024-10-24 00:21 /user
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -chomd g+w /user
-chomd: Unknown command
Usage: hadoop fs [generic options]
    [-appendToFile [-n] <localsrc> ... <dst>]
    [-cat [-ignoreCrc] <src> ...]
    [-checksum [-v] <src> ...]
    [-chgrp [-R] GROUP PATH...]
    [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
    [-chown [-R] [OWNER][:[GROUP]] PATH...]
    [-concat <target path> <src path> <src path> ...]
    [-copyFromLocal [-f] [-p] [-l] [-d] [-t <thread count>] [-q <thread pool queue size>] <localsrc> ... <dst>]
    [-copyToLocal [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
    [-count [-q] [-h] [-v] [-t <storage type>] [-u] [-x] [-e] [-s] <path> ...]
    [-cp [-f] [-p | -p[topax]] [-d] [-t <thread count>] [-q <thread pool queue size>] <src> ... <dst>]
    [-createSnapshot <snapshotDir> [<snapshotName>]]
    [-deleteSnapshot <snapshotDir> <snapshotName>]
    [-df [-h] [<path> ...]]
    [-du [-s] [-h] [-v] [-x] <path> ...]
    [-expunge [-immediate] [-fs <path>]]
    [-find <path> ... <expression> ...]
    [-get [-f] [-p] [-crc] [-ignoreCrc] [-t <thread count>] [-q <thread pool queue size>] <src> ... <localdst>]
    [-getfacl [-R] <path>]
    [-getfattr [-R] {-n name | -d} [-e en] <path>]
    [-getmerge [-nl] [-skip-empty-file] <src> <localdst>]
    [-head <file>]
    [-help [cmd ...]]
    [-ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [-e] [<path> ...]]
    [-mkdir [-p] <path> ...]

```



```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -chmod g+w /user
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -chmod g+wx /user
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$ hdfs dfs -ls /
Found 2 items
drwxr-xr-x - hadoop supergroup 0 2024-10-04 16:58 /logs
drwxrwxr-x - hadoop supergroup 0 2024-10-24 00:21 /user
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/hive/conf$
```

File Edit View
Leandro Delgado
Student ID: 114416241
Ln 1, Col 16 | 37 characters | 100% | Window UTF-8

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
GNU nano 7.2 /home/hadoop/.bashrc *
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

# Hadoop Path
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$HIVE_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

export HIVE_HOME=/home/hadoop/apache-hive-3.1.2-bin
export PATH=$PATH:$HIVE_HOME/bin
[ line 134/135 (99%), col 33/33 (100%), char 4407/4408 (99%) ]
^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute ^C Location M-U Undo M-A Set Mark
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^_ Go To Line M-E Redo M-6 Copy
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/hadoop
GNU nano 7.2 hive-config.sh *
shift
confdir=$1
shift
HIVE_CONF_DIR=$confdir
;;
--auxpath)
shift
HIVE_AUX_JARS_PATH=$1
shift
;;
*)
break;
;;
esac
done

# Allow alternate conf dir location.
HIVE_CONF_DIR="${HIVE_CONF_DIR:-$HIVE_HOME/conf}"

export HIVE_CONF_DIR=$HIVE_CONF_DIR
export HIVE_AUX_JARS_PATH=$HIVE_AUX_JARS_PATH

# Default to use 256MB
export HADOOP_HEAPSIZE=${HADOOP_HEAPSIZE:-256}
export HADOOP_HOME=/home/hadoop/hadoop
```

^G Help ^O Write Out ^W Where Is ^K Cut ^T Execute ^C Location M-U Undo M-A Set Mark
^X Exit ^R Read File ^\ Replace ^U Paste ^J Justify ^_ Go To Line M-E Redo M-6 Copy


```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin$ bin/schematool -initSchema -dbType derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/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:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Metastore connection URL:      jdbc:derby;;databaseName=metastore_db;create=true
Metastore Connection Driver :  org.apache.derby.jdbc.EmbeddedDriver
Metastore connection User:     APP
Starting metastore schema initialization to 3.1.0
Initialization script hive-schema-3.1.0.derby.sql
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin$ 
Initialization script completed
schemaTool completed
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin$
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin$ bin/hive
Initialization script completed
schemaTool completed
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/apache-hive-3.1.2-bin$ bin/hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/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:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Hive Session ID = 2d8e2da7-e0e2-48a8-bb3d-fb7e4361c63f

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 1e25ff6f-ae83-4daf-889b-838dc4bc0fde
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive>
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin$ bin/hive
Initialization script completed
schemaTool completed
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/apache-hive-3.1.2-bin$ bin/hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/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:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Hive Session ID = 2d8e2da7-e0e2-48a8-bb3d-fb7e4361c63f

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 1e25ff6f-ae83-4daf-889b-838dc4bc0fde
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show databases;
OK
default
Time taken: 0.278 seconds, Fetched: 1 row(s)
hive>
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hive> create database 30midtermdb;
OK
Time taken: 0.049 seconds
hive>
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hive> use 30midtermdb;
OK
Time taken: 0.019 seconds
hive> 
```



```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.... x hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: /

hive> use 30midtermdb;
OK
Time taken: 0.019 seconds
hive> CREATE TABLE insure_data_managed (age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, charges DOUBLE, region STRING)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.241 seconds
hive> CREATE external Table insure_data_ext ( age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, charges DOUBLE, region STRING )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.03 seconds
hive>
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.... x hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: /

hive> use 30midtermdb;
OK
Time taken: 0.019 seconds
hive> CREATE TABLE insure_data_managed (age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, charges DOUBLE, region STRING)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.241 seconds
hive> 
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hive> use 30midtermdb;
OK
Time taken: 0.019 seconds
hive> CREATE TABLE insure_data_managed (age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, charges DOUBLE, region STRING)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.241 seconds
hive> CREATE external Table insure_data_ext ( age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, charges DOUBLE, region STRING )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.03 seconds
hive> cp /home/leandro-dash-cyt-180/Downloads/insurance.csv /tmp/insurance.csv
[1]+  Stopped                  bin/hive
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin$ bin/hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/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:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Hive Session ID = fba3b753-3d84-4ea3-8de7-eba463174863

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution
```

23	female	23.18	2	no	northwest	14426.07385
20	female	30.59	0	no	northeast	2459.7201
32	female	41.1	0	no	southwest	3989.841
43	female	34.58	1	no	northwest	7727.2532
34	male	42.13	2	no	southeast	5124.1887
30	male	38.83	1	no	southeast	18963.17192
18	female	28.215	0	no	northeast	2200.83085
41	female	28.31	1	no	northwest	7153.5539
35	female	26.125	0	no	northeast	5227.98875
57	male	40.37	0	no	southeast	10982.5013
29	female	24.6	2	no	southwest	4529.477
32	male	35.2	2	no	southwest	4670.64
37	female	34.105	1	no	northwest	6112.35295
18	male	27.36	1	yes	northeast	17178.6824
43	female	26.7	2	yes	southwest	22478.6
56	female	41.91	0	no	southeast	11093.6229
38	male	29.26	2	no	northwest	6457.8434
29	male	32.11	2	no	northwest	4433.9159
22	female	27.1	0	no	southwest	2154.361
52	female	24.13	1	yes	northwest	23887.6627
40	female	27.4	1	no	southwest	6496.886
23	female	34.865	0	no	northeast	2899.48935
31	male	29.81	0	yes	southeast	19350.3689
42	female	41.325	1	no	northeast	7650.77375
24	female	29.925	0	no	northwest	2850.68375
25	female	30.3	0	no	southwest	2632.992
48	female	27.36	1	no	northeast	9447.3824
23	female	28.49	1	yes	southeast	18328.2381
45	male	23.56	2	no	northeast	8603.8234
20	male	35.625	3	yes	northwest	37465.34375
62	female	32.68	0	no	northwest	13844.7972

```

hive> hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/apache-hive-3.1.2-bin$
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~/apache-hive-3.1.2-bin$ bin/hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hadoop/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:/home/hadoop/hadoop/share/hadoop/common/lib/slf4j-reload4j-1.7.36.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]
Hive Session ID = d5d9a03c-993b-45e1-9a81-5937ef467c5c

Logging initialized using configuration in jar:file:/home/hadoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 3036fe35-5852-45db-bfb9-e2291d335194
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> set hive.exec.dynamic.partitioned=True;
Query returned non-zero code: 1, cause: hive configuration hive.exec.dynamic.partitioned does not exists.
hive> set hive.exec.dynamic.partition.mode=nonstrict;
hive> CREATE TABLE insurance_partitioned (age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, region STRING, charges DOUBLE)
> ;
OK
Time taken: 0.482 seconds
hive> CREATE TABLE insurance_partitioned (age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, region STRING, charges DOUBLE);
FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. AlreadyExistsException(message:Table hive.default.insurance_partitioned already exists)
hive> drop table insurance partitioned

```

```

hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/leandro-dash-cyt-180$ cd /
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/$ cd home/hadoop
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:/$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [leandro-dash-cyt-180-VMware-Virtual-Platform]
Starting resourcemanager
Starting nodemanagers
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ hdfs dfs -ls
ls: '.': No such file or directory
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ hdfs dfs -ls /user/hive
Found 1 items
drwxrwxr-x - hadoop supergroup 0 2024-10-30 21:00 /user/hive/warehouse
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ hdfs dfs -ls /user/hive/warehouse
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2024-10-30 21:57 /user/hive/warehouse/30midtermdb.db
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ hdfs dfs -ls /user/hive/warehouse/30midtermdb.db
Found 3 items
drwxr-xr-x - hadoop supergroup 0 2024-10-30 21:39 /user/hive/warehouse/30midtermdb.db/insure_data_ext
drwxr-xr-x - hadoop supergroup 0 2024-10-30 21:58 /user/hive/warehouse/30midtermdb.db/insure_data_managed
drwxr-xr-x - hadoop supergroup 0 2024-10-30 22:00 /user/hive/warehouse/30midtermdb.db/insure_ext
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$ hdfs dfs -ls /user/hive/warehouse/30midtermdb.db/insure_data_managed
Found 1 items
-rw-r--r-- 1 hadoop supergroup 55628 2024-10-30 21:58 /user/hive/warehouse/30midtermdb.db/insure_data_managed/insurance.csv
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform:~$

```



```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtua... x hadoop@leandro-dash-cyt-180-VMware-Virtua... x leandro-dash-cyt-180@leandro-dash-cyt-180-V... x
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.275 seconds
hive> SET hive.exec.dynamic.partition = true;
hive> SET hive.exec.dynamic.partition.mode = nonstrict;
hive> CREATE TABLE insure_partitioned (age INT, sex STRING, bmi DOUBLE, children INT, smoker STRING, charges DOUBLE)
> partitioned by (region STRING)
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY ',';
OK
Time taken: 0.03 seconds
hive> show tables;
OK
insure_data_managed
insure_partitioned
Time taken: 0.015 seconds, Fetched: 2 row(s)
hive> LOAD DATA LOCAL INPATH 'file:///tmp/insurance.csv' into table insure_data_managed;
Loading data to table 30midtermdb.insure_data_managed
OK
Time taken: 0.289 seconds
hive> INSERT into insure_partitioned PARTITION (region)
> SELECT age, sex, bmi, children, smoker, region , charges FROM insure_data_managed;
Query ID = hadoop_20241030224034_ae2070f2-3551-4978-b661-d4b411276b3a
Total jobs = 3
Launching Job 1 out of 3
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>
```

```
hadoop@leandro-dash-cyt-180-VMware-Virtual-Platform: ~/apache-hive-3.1.2-bin
hadoop@leandro-dash-cyt-180-VMware-Virtua... x hadoop@leandro-dash-cyt-180-VMware-Virtua... x leandro-dash-cyt-180@leandro-dash-cyt-180-V... x
> ;
OK
Time taken: 0.174 seconds
hive> INSERT into insure_partitioned PARTITION (region)
> SELECT age, sex, bmi, children, smoker, region , charges FROM insure_data_managed;
FAILED: SemanticException [Error 10096]: Dynamic partition strict mode requires at least one static partition column. To
turn this off set hive.exec.dynamic.partition.mode=nonstrict
hive> SET hive.exec.dynamic.partition = true;
hive> SET hive.exec.dynamic.partition.mode = nonstrict;
hive> INSERT into insure_partitioned PARTITION (region)
> SELECT age, sex, bmi, children, smoker, region , charges FROM insure_data_managed;
Query ID = hadoop_20241030232332_e88b9bb0-2573-4c62-bac2-c6c0af77cc51
Total jobs = 3
Launching Job 1 out of 3
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.reduces=<number>
Job running in-process (local Hadoop)
2024-10-30 23:23:34,491 Stage-1 map = 0%, reduce = 0%
Ended Job = job_local1377132762_0001 with errors
Error during job, obtaining debugging information...
FAILED: Execution Error, return code 2 from org.apache.hadoop.hive.ql.exec.mr.MapRedTask
MapReduce Jobs Launched:
Stage-Stage-1: HDFS Read: 8192 HDFS Write: 2345 FAIL
Total MapReduce CPU Time Spent: 0 msec
hive>
```