SUSHMITHA P.A
1NT19IS170
C2

EXERCISE 5

Opening the hadoop environment

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP_HOME
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin/
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 4168.  Stop it first.
Starting datanodes
localhost: datanode is running as process 4319.  Stop it first.
Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]
admin1-HP-280-G4-MT-Business-PC: secondarynamenode is running as process 4570.  Stop it first.
Starting resourcemanager
resourcemanager is running as process 4758.  Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 4910.  Stop it first.
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps
10917 Jps
4758 ResourceManager
4168 NameNode
4570 SecondaryNameNode
4910 NodeManager
4319 DataNode
```

Dropping the database:

rm -rf metastore_db

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ cd $HIVE_HOME
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin$ ./conf
bash: ./conf: Is a directory
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin$ cd conf
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin/conf$ ls
beeline-log4j2.properties.template  hive-env.sh.template                ivysettings.xml                          metastore_db
derby.log                           hive-exec-log4j2.properties.template  llap-cli-log4j2.properties.template      parquet-logging.properties
hive-default.xml.template           hive-log4j2.properties.template       llap-daemon-log4j2.properties.template
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin/conf$ rm -rf metastore_db
```

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin/conf$ ls
beeline-log4j2.properties.template  hive-env.sh.template                ivysettings.xml                          parquet-logging.properties
derby.log                           hive-exec-log4j2.properties.template  llap-cli-log4j2.properties.template
hive-default.xml.template           hive-log4j2.properties.template       llap-daemon-log4j2.properties.template
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin/conf$ schematool -initSchema -db Type derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdoop/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/hdoop/hadoop-3.2.1/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]
HiveSchemaTool:Parsing failed.  Reason: Unrecognized option: -db
usage: schemaTool
 -alterCatalog <arg>               Alter a catalog, requires
                                   --catalogLocation and/or
                                   --catalogDescription parameter as well
 -catalogDescription <arg>         Description of new catalog
 -catalogLocation <arg>            Location of new catalog, required when
                                   adding a catalog
 -createCatalog <arg>              Create a catalog, requires
                                   --catalogLocation parameter as well
 -dbOpts <databaseOpts>            Backend DB specific options
 -dbType <databaseType>            Metastore database type
 -driver <driver>                  driver name for connection
 -dryRun                           list SQL scripts (no execute)
 -fromCatalog <arg>                Catalog a moving database or table is
                                   coming from.  This is required if you
                                   are moving a database or table.
 -fromDatabase <arg>               Database a moving table is coming
                                   from.  This is required if you are
                                   moving a table.
 -help                             print this message
 -ifNotExists                      If passed then it is not an error to
```

schematool -initSchema -dbType derby

```
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin/conf$ schematool -initSchema -dbType derby
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdoop/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/hdoop/hadoop-3.2.1/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]
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
```

```
Initialization script completed
schemaTool completed
hdoop@admin1-HP-280-G4-MT-Business-PC:~/apache-hive-3.1.2-bin/conf$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdoop/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/hdoop/hadoop-3.2.1/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]
Hive Session ID = f19fea24-a3e4-4214-8a18-f3045f179122

Logging initialized using configuration in jar:file:/home/hdoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = c33cf6cc-4008-476d-8756-ca99e19587e3
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 u
sing Hive 1.X releases.
```

Creating the database:

show databases;

create database if not exists salesdb;

use salesdb;

```
hive> show databases;
OK
default
Time taken: 0.451 seconds, Fetched: 1 row(s)
hive> create database if not exist salesdb;
FAILED: ParseException line 1:23 missing KW_EXISTS at 'exist' near '<EOF>'
line 1:29 extraneous input 'salesdb' expecting EOF near '<EOF>'
hive> create database if not exists salesdb;
OK
Time taken: 0.114 seconds
hive> use salesdb;
OK
Time taken: 0.037 seconds
```

Creating table:

create table sales(product string,price int,payment_type string,name string,city string,state string,country string);

desc sales;

```
hive> create table sales(product string,price int,payment_type string,name string,city string,state string,country string);
OK
Time taken: 0.418 seconds
hive> desc sales;
OK
product                 string
price                   int
payment_type            string
name                    string
city                    string
state                   string
country                 string
Time taken: 0.112 seconds, Fetched: 7 row(s)
```

Inserting values into the table:

```
hive> insert into sales values('P1',5000,'visa','sushmitha','bangalore','karnataka','india');
hive> insert into sales values('P2',6000,'mastercard','gowtham','mangalore','karnataka','india');
```

```
hive> select * from sales;
OK
P1      5000    visa    sushmitha       bangalore       karnataka       india
P2      6000    mastercard      gowtham mangalore       karnataka       india
P3      7000    Rupay   soni    velore  Tamil Nadu      india
P4      8000    Rupay   tejaswini       mumbai  Maharastra      india
P5      3000    Rupay   alia    kolkata west bengal     india
P6      2000    visa    anuj    kolkata west bengal     india
Time taken: 0.162 seconds, Fetched: 6 row(s)
```

Dropping the table:

create table temp(Product string,Country string)

drop table temp;

```
hive> create table temp(product string,country string);
OK
Time taken: 0.063 seconds
hive> drop table temp;
OK
Time taken: 0.708 seconds
```

Alter table:

alter table sales rename to sales;

```
hive> alter table sales rename to 1NT19IS170_SALES;
OK
Time taken: 0.212 seconds
hive> show tables;
OK
1nt19is170_sales
Time taken: 0.018 seconds, Fetched: 1 row(s)
hive> alter table 1NT19IS170_SALES rename to sales;
OK
Time taken: 0.111 seconds
```

3.Built in operators:

Arithmetic Operator in Hive

select product,name,price+1500 from sales;

```
hive> select product,name,price+1500 from sales;
OK
P1      sushmitha       6500
P2      gowtham 7500
P3      soni    8500
P4      tejaswini       9500
P5      alia    4500
P6      anuj    3500
Time taken: 0.093 seconds, Fetched: 6 row(s)
```

select product,name,(price*10)/2 from sales;

```
hive> select product,name,(price*10)/2 from sales;
OK
P1      sushmitha       25000.0
P2      gowtham 30000.0
P3      soni    35000.0
P4      tejaswini       40000.0
P5      alia    15000.0
P6      anuj    10000.0
Time taken: 0.115 seconds, Fetched: 6 row(s)
```

Relational Operators in Hive

select * from sales where price>4500;

```
hive> select * from sales where price>4500;
OK
P1      5000    visa    sushmitha       bangalore       karnataka       india
P2      6000    mastercard      gowtham mangalore       karnataka       india
P3      7000    Rupay   soni    velore  Tamil Nadu      india
P4      8000    Rupay   tejaswini       mumbai  Maharastra      india
Time taken: 0.194 seconds, Fetched: 4 row(s)
```

4.Built in functions:

Mathematical Functions in Hive

select product,name,sqrt(price) from sales;

```
hive> select product,name,sqrt(price) from sales;
OK
P1      sushmitha       70.71067811865476
P2      gowtham 77.45966692414834
P3      soni    83.66600265340756
P4      tejaswini       89.44271909999159
P5      alia    54.772255750516614
P6      anuj    44.721359549995796
Time taken: 0.07 seconds, Fetched: 6 row(s)
```

select product,name,floor(price) from sales;

```
hive> select product,name,floor(price) from sales;
OK
P1      sushmitha       5000
P2      gowtham 6000
P3      soni    7000
P4      tejaswini       8000
P5      alia    3000
P6      anuj    2000
Time taken: 0.064 seconds, Fetched: 6 row(s)
```

select product,name,cos(price) from sales;

```
hive> select product,name,cos(price) from sales;
OK
P1      sushmitha       0.15466840618074712
P2      gowtham 0.9039115103477952
P3      soni    0.8620134342951074
P4      tejaswini       0.06564512751032302
P5      alia    -0.9756821998857504
P6      anuj    -0.36745954910083134
Time taken: 0.07 seconds, Fetched: 6 row(s)
```

Aggregate Functions in Hive

select max(price) from sales where state='Karnataka';

```
hive> select max(price) from sales where state='karnataka';
Query ID = hdoop_20220628134629_2e32f4c5-4fac-4a0a-99de-c47d775efda6
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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>
Starting Job = job_1656398423323_0011, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1656398423323_0011/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1656398423323_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-28 13:46:35,007 Stage-1 map = 0%,  reduce = 0%
2022-06-28 13:46:39,081 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.14 sec
2022-06-28 13:46:44,201 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.67 sec
MapReduce Total cumulative CPU time: 3 seconds 670 msec
Ended Job = job_1656398423323_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.67 sec   HDFS Read: 15510 HDFS Write: 104 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 670 msec
OK
6000
Time taken: 15.82 seconds, Fetched: 1 row(s)
```

Other built-in functions

select product, upper(name) from sales;

```
hive> select product,upper(name) from sales;
OK
P1      SUSHMITHA
P2      GOWTHAM
P3      SONI
P4      TEJASWINI
P5      ALIA
P6      ANUJ
Time taken: 0.102 seconds, Fetched: 6 row(s)
```

select product, concat(name,state) from sales;

```
hive> select product,concat(name,state)from sales;
OK
P1      sushmithakarnataka
P2      gowthamkarnataka
P3      soniTamil Nadu
P4      tejaswiniMaharastra
P5      aliawest bengal
P6      anujwest bengal
Time taken: 0.094 seconds, Fetched: 6 row(s)
```

select product, trim(name) from sales;
(Removes white spaces from both the ends)

```
hive> select product,trim(name) from sales;
OK
P1      sushmitha
P2      gowtham
P3      soni
P4      tejaswini
P5      alia
P6      anuj
Time taken: 0.069 seconds, Fetched: 6 row(s)
```

5.VIEWS:

create view visa as select * from sales where payment_type="Visa";
select * from visa;

```
hive> create view visa as select * from sales where payment_type='visa';
OK
Time taken: 0.099 seconds
hive> select * from visa;
OK
P1      5000    visa    sushmitha       bangalore       karnataka       india
P6      2000    visa    anuj    kolkata west bengal     india
Time taken: 0.111 seconds, Fetched: 2 row(s)
```

create view master as select * from sales where payment_type='Mastercard';
select * from master;

```
hive> create view master as select * from sales where payment_type='mastercard';
OK
Time taken: 0.104 seconds
hive> select * from mastercard;
FAILED: SemanticException [Error 10001]: Line 1:14 Table not found 'mastercard'
hive> select * from master;
OK
P2      6000    mastercard      gowtham mangalore       karnataka       india
Time taken: 0.092 seconds, Fetched: 1 row(s)
```

6.HIVEQL

GROUPBY:

select count(*)as count,country from sales group by country;

```
hive> select count(*)as count ,country from sales group by country;
Query ID = hdoop_20220628135531_605a20a4-28d4-46d8-8a8e-5dcdafc45f7c
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.reduces=<number>
Starting Job = job_1656398423323_0012, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1656398423323_0012/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1656398423323_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-28 13:55:36,079 Stage-1 map = 0%,  reduce = 0%
2022-06-28 13:55:40,162 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.28 sec
2022-06-28 13:55:44,232 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.48 sec
MapReduce Total cumulative CPU time: 2 seconds 480 msec
Ended Job = job_1656398423323_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.48 sec   HDFS Read: 14750 HDFS Write: 107 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 480 msec
OK
6       india
```

select count(*) as count,state from sales group by state;

```
hive> select count(*) as count,state from sales group by state;
Query ID = hdoop_20220628135616_f7c242fa-5762-4ab7-b5f7-90d501fbe2b0
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.reduces=<number>
Starting Job = job_1656398423323_0013, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1656398423323_0013/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1656398423323_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-28 13:56:23,229 Stage-1 map = 0%,  reduce = 0%
2022-06-28 13:56:27,342 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.29 sec
2022-06-28 13:56:31,415 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.51 sec
MapReduce Total cumulative CPU time: 2 seconds 510 msec
Ended Job = job_1656398423323_0013
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.51 sec   HDFS Read: 14885 HDFS Write: 187 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 510 msec
OK
1       Maharastra
1       Tamil Nadu
2       karnataka
2       west bengal
Time taken: 15.739 seconds, Fetched: 4 row(s)
```

select product,name from sales group by product,name;

```
hive> select product,name from sales group by product,name;
Query ID = hdoop_20220628135658_1f008a7d-6089-49b3-80dc-0e61f6ef3a54
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.reduces=<number>
Starting Job = job_1656398423323_0014, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1656398423323_0014/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1656398423323_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-28 13:57:03,420 Stage-1 map = 0%,  reduce = 0%
2022-06-28 13:57:07,531 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.22 sec
2022-06-28 13:57:10,582 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.54 sec
MapReduce Total cumulative CPU time: 2 seconds 540 msec
Ended Job = job_1656398423323_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.54 sec   HDFS Read: 13764 HDFS Write: 220 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 540 msec
OK
P1      sushmitha
P2      gowtham
P3      soni
P4      tejaswini
P5      alia
P6      anuj
```

WHERE

select * from sales where state="Karnataka";

```
hive> select * from sales where state='karnataka';
OK
P1      5000    visa    sushmitha       bangalore       karnataka       india
P2      6000    mastercard      gowtham mangalore       karnataka       india
Time taken: 0.116 seconds, Fetched: 2 row(s)
```

ORDERBY:

select * from sales order by price;

```
hive> select * from sales order by price;
Query ID = hdoop_20220628135902_a1af60d6-23e8-4c39-96a9-0b344aed0345
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 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>
Starting Job = job_1656398423323_0015, Tracking URL = http://admin1-HP-280-G4-MT-Business-PC:8088/proxy/application_1656398423323_0015/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill job_1656398423323_0015
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-06-28 13:59:07,550 Stage-1 map = 0%,  reduce = 0%
2022-06-28 13:59:11,646 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.6 sec
2022-06-28 13:59:15,737 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 2.87 sec
MapReduce Total cumulative CPU time: 2 seconds 870 msec
Ended Job = job_1656398423323_0015
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 2.87 sec   HDFS Read: 13363 HDFS Write: 441 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 870 msec
OK
P6      2000    visa    anuj    kolkata west bengal     india
P5      3000    Rupay   alia    kolkata west bengal     india
P1      5000    visa    sushmitha       bangalore       karnataka       india
P2      6000    mastercard      gowtham mangalore       karnataka       india
P3      7000    Rupay   soni    velore  Tamil Nadu      india
P4      8000    Rupay   tejaswini       mumbai  Maharastra      india
Time taken: 15.572 seconds, Fetched: 6 row(s)
```