

BDMH: Assignment I

Group 45

Harsh Bandhey
Ridam Pal

2017234
PhD19201

Question 1

Docker Run

Commands:

```
docker pull raptor4/harsh17234_ridamp_a1  
docker run -it raptor4/harsh17234_ridamp_a1
```

For a non-persistent non-ported container image. Use -v and -p flags accordingly for mounting volumes and ports.

Volume mount points for

MongoDB /data/db

MySQL /var/lib/mysql

Usage:

Run “mongo”, “hive” and “mysql -u root -proot” for accessing MongoDB, hive and MySQL shells respectively.

MySQL Question 1.

```
root@016ed015911bc:/
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 5
Server version: 5.7.29-0ubuntu0.18.04.1 (Ubuntu)

Copyright (c) 2000, 2020, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show database;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'database' at line 1
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| mydb      |
| mysql     |
| performance_schema |
| sys       |
+-----+
5 rows in set (0.01 sec)

mysql> use mydb;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables
+-----+
| Tables_in_mydb |
+-----+
| Group45         |
+-----+
1 row in set (0.00 sec)

mysql> select * from Group45;
+----+-----+-----+-----+-----+-----+
| id | technique | d_id | n_samples | t_samples | pubmed_id |
+----+-----+-----+-----+-----+-----+
| ID1 | Affymetrix Array | GSE62232 | 91 | 81 HCC and 10 non-tumor | 25822088 |
| ID2 | Affymetrix Array | GSE63067 | 18 | steatosis, healthy and Non-alcoholic steatohepatitis | 25993042 |
| ID3 | Affymetrix Array | GSE64041 | 125 | 60 pairs HCC and non-tumor; 5 normal | 27499918 |
| ID4 | Affymetrix Array | GSE67115 | 103 | HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor | 29538454 |
| ID5 | Affymetrix Array | GSE72981 | 30 | Subcutaneous and orthotopic HCC tissue | 26520397 |
+----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)

mysql>
```

MySQL commands and demos:

(With pre-existing data from Group45.csv)

Show databases command is used for showing all the databases stored within it.

```
mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| mydb      |
| mysql     |
| performance_schema |
| sys       |
+-----+
5 rows in set (0.01 sec)
```

Use command inherits the database we need to access from the list of databases.

Show tables command used to show all the tables present in the database.

```
mysql> use mydb;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables
-> ;
+-----+
| Tables_in_mydb |
+-----+
| Group45         |
+-----+
1 row in set (0.00 sec)
```

1) Select

Select command query states about selecting all the rows from the table Group45

```
mysql> select * from Group45;
+----+-----+-----+-----+-----+-----+
| id | technique | d_id | n_samples | t_samples | pubmed_id |
+----+-----+-----+-----+-----+-----+
| ID1 | Affymetrix Array | GSE62232 | 91 | 81 HCC and 10 non-tumor | 25822088 |
| ID2 | Affymetrix Array | GSE63067 | 18 | Steatosis, healthy and Non-alcoholic steatohepatitis | 25993042 |
| ID3 | Affymetrix Array | GSE64041 | 125 | 60 pairs HCC and non-tumor, 5 normal | 27499918 |
| ID4 | Affymetrix Array | GSE69715 | 103 | HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor | 29538454 |
| ID5 | Affymetrix Array | GSE72981 | 30 | Subcutaneous and orthotopic HCC tissue | 26520397 |
+----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

This query is selecting all the rows from table Group45 **where** 'id=ID1'.

```
mysql> select * from Group45 where id="ID1";
+----+-----+-----+-----+-----+-----+
| id | technique | d_id | n_samples | t_samples | pubmed_id |
+----+-----+-----+-----+-----+-----+
| ID1 | Affymetrix Array | GSE62232 | 91 | 81 HCC and 10 non-tumor | 25822088 |
+----+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)
```

2) Insert

This query states about inserting rows into the table Group45

```
mysql> INSERT INTO `Group45` VALUES ('ID1','Affymetrix Array','GSE62232','91','81 HCC and 10 non-tumor','25822088');
Query OK, 1 row affected (0.01 sec)

mysql> mysql> INSERT INTO `Group45` VALUES ('ID1','Affymetrix Array','GSE62232','91','81 HCC and 10 non-tumor','25822088'); select * from Group45;
+----+-----+-----+-----+-----+-----+
| id | technique | d_id | n_samples | t_samples | pubmed_id |
+----+-----+-----+-----+-----+-----+
| ID1 | Affymetrix Array | GSE62232 | 91 | 81 HCC and 10 non-tumor | 25822088 |
| ID2 | Affymetrix Array | GSE63067 | 18 | Steatosis, healthy and Non-alcoholic steatohepatitis | 25993042 |
| ID3 | Affymetrix Array | GSE64041 | 125 | 60 pairs HCC and non-tumor, 5 normal | 27499918 |
| ID4 | Affymetrix Array | GSE69715 | 103 | HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor | 29538454 |
| ID5 | Affymetrix Array | GSE72981 | 30 | Subcutaneous and orthotopic HCC tissue | 26520397 |
| ID1 | Affymetrix Array | GSE62232 | 91 | 81 HCC and 10 non-tumor | 25822088 |
+----+-----+-----+-----+-----+-----+
6 rows in set (0.08 sec)

mysql>
```

3) Delete and Drop

Delete command is used for deleting the rows from the table.

Drop command is used for removing the table from the database.

```
mysql> DELETE FROM Group45;
Query OK, 6 rows affected (0.01 sec)

mysql> DROP TABLE Group45;
Query OK, 0 rows affected (0.03 sec)

mysql>
```

4) Create Table

Create table command is used for creating the Table45 hence defining the data types of each entry along with each column name.

```
mysql> CREATE TABLE `Group45` (
->   `id` varchar(25) COLLATE utf8_unicode_ci DEFAULT NULL,
->   `technique` varchar(255) COLLATE utf8_unicode_ci DEFAULT NULL,
->   `d_id` varchar(255) COLLATE utf8_unicode_ci DEFAULT NULL,
->   `n_samples` varchar(25) COLLATE utf8_unicode_ci DEFAULT NULL,
->   `t_samples` varchar(255) COLLATE utf8_unicode_ci DEFAULT NULL,
->   `pubmed_id` varchar(25) COLLATE utf8_unicode_ci DEFAULT NULL
-> );
Query OK, 0 rows affected (0.03 sec)

mysql>
```

5) Update

Update command helps in updating the specific row **where** 'id=ID1' with the updated value.

```
mysql> update Group45 set t_samples="Edit" where id="ID1";
Query OK, 0 rows affected (0.01 sec)
Rows matched: 1  Changed: 0  Warnings: 0

mysql> select * from Group45 where id="ID1";
+----+-----+-----+-----+-----+-----+
| id | technique | d_id | n_samples | t_samples | pubmed_id |
+----+-----+-----+-----+-----+-----+
| ID1 | Affymetrix Array | GSE62232 | 91 | Edit | 25822088 |
+----+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)
```

Question 2.

MongoDB commands and demos:

(With pre-existing data from Group45.csv)

Show dbs command is used for showing all the databases stored within it.

MongoDB Question 2.

```
root@eeb43294eb3: /
Welcome to the MongoDB shell.
For interactive help, type "help".
For more comprehensive documentation, see
  http://docs.mongodb.org/
Questions? Try the support group
  http://groups.google.com/group/mongodb-user
Server has startup warnings:
2020-02-14T15:09:17.784+0000 I STORAGE [initandlisten] ** WARNING: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine
2020-02-14T15:09:17.784+0000 I STORAGE [initandlisten] See http://dochub.mongodb.org/core/prodnotes-filesystem
2020-02-14T15:09:17.784+0000 I STORAGE [initandlisten]
2020-02-14T15:09:18.433+0000 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2020-02-14T15:09:18.433+0000 I CONTROL [initandlisten] Read and write access to data and configuration is unrestricted.
2020-02-14T15:09:18.433+0000 I CONTROL [initandlisten] ** WARNING: You are running this process as the root user, which is not recommended.
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten]
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten] ** WARNING: This server is bound to localhost.
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten] Remote systems will be unable to connect to this server.
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten] Start the server with --bind_ip <address> to specify which IP
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten] addresses it should serve responses from, or with --bind_ip_all to
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten] bind to all interfaces. If this behavior is desired, start the
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten] server with --bind_ip 127.0.0.1 to disable this warning.
2020-02-14T15:09:18.434+0000 I CONTROL [initandlisten]
---
Enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc).

The monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.

To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
---
> show dbs
admin 0.000GB
config 0.000GB
local 0.000GB
mgdb 0.000GB
> use mgdb
switched to db mgdb
> show collections
Group45
> db.Group45.find()
{ "_id" : ObjectId("5e46b820b1509b82dc910edd"), " ID" : "ID1", "Profiling Technique" : "Affymetrix Array", "Dataset ID" : "GSE62232", "No of Samples" : 91, "Type of Samples" : "91 HCC and 10 non-tumor", "Pubmed ID" : 25822088 }
{ "_id" : ObjectId("5e46b820b1509b82dc910ede"), " ID" : "ID2", "Profiling Technique" : "Affymetrix Array", "Dataset ID" : "GSE63067", "No of Samples" : 18, "Type of Samples" : "Steatosis, healthy and Non-alcoholic steatohepatitis", "Pubmed ID" : 25993042 }
{ "_id" : ObjectId("5e46b820b1509b82dc910edf"), " ID" : "ID3", "Profiling Technique" : "Affymetrix Array", "Dataset ID" : "GSE64041", "No of Samples" : 125, "Type of Samples" : "60 pairs HCC and non-tumor, 5 normal", "Pubmed ID" : 27499918 }
{ "_id" : ObjectId("5e46b820b1509b82dc910ee0"), " ID" : "ID4", "Profiling Technique" : "Affymetrix Array", "Dataset ID" : "GSE69715", "No of Samples" : 103, "Type of Samples" : "HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor", "Pubmed ID" : 29538454 }
{ "_id" : ObjectId("5e46b820b1509b82dc910ee1"), " ID" : "ID5", "Profiling Technique" : "Affymetrix Array", "Dataset ID" : "GSE72981", "No of Samples" : 30, "Type of Samples" : "Subcutaneous and orthotopic HCC tissue", "Pubmed ID" : 26520397 }
> db.Group45.find().pretty()
```

1. **find()** method is used in query for displaying the documents in a random and unstructured way.
2. **Insert ()** method is used in query for inserting rows in the database.
3. **Update()** method is used in query for updating the document in the database.
4. **Remove()** method is used in query for removing rows in the database.
5. **Drop()** method common is used to drop the whole collection.

The query for each command has been shown in the images below where all the queries related to each method have been shown.

```
root@eeb43294aeb3: /
@ (shell):2:0
> db.Group45.insert({ "ID": "ID6", "Everythng Else":"Nothing"})
writeResult({ "nInserted" : 1 })
> db.Group45.find().pretty()
{
  "_id" : ObjectId("5e46b820b1509b82dc910edd"),
  "ID" : "ID1",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE62232",
  "No of Samples" : 91,
  "Type of Samples" : "81 HCC and 10 non-tumor",
  "Pubmed ID" : 25822088
}
{
  "_id" : ObjectId("5e46b820b1509b82dc910ede"),
  "ID" : "ID2",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE63067",
  "No of Samples" : 18,
  "Type of Samples" : "Steatosis, healthy and Non-alcoholic steatohepatitis",
  "Pubmed ID" : 25993042
}
{
  "_id" : ObjectId("5e46b820b1509b82dc910edf"),
  "ID" : "ID3",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE64041",
  "No of Samples" : 125,
  "Type of Samples" : "60 pairs HCC and non-tumor, 5 normal",
  "Pubmed ID" : 27499918
}
{
  "_id" : ObjectId("5e46b820b1509b82dc910ee0"),
  "ID" : "ID4",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE69715",
  "No of Samples" : 103,
  "Type of Samples" : "HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor",
  "Pubmed ID" : 29538454
}
{
  "_id" : ObjectId("5e46b820b1509b82dc910ee1"),
  "ID" : "ID5",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE72981",
  "No of Samples" : 30,
  "Type of Samples" : "Subcutaneous and orthotopic HCC tissue",
  "Pubmed ID" : 26520397
}
{
  "_id" : ObjectId("5e46b9bba037f449b7dc3820"),
  "ID" : "ID6",
  "Everythng Else" : "Nothing"
}
```

```

> db.Group45.update({ "ID": "ID6"},{$set:{'title':'Extra Tittle'}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
> db.Group45.find({"ID":"ID6"}).pretty()
{
  "_id" : ObjectId("5e46b9bba037f449b7dc3820"),
  "ID" : "ID6",
  "Everyting Else" : "Nothing",
  "title" : "Extra Tittle"
}

```

```

C:\ Select root@eeb43294aeb3: /
@ (shell):1:1
> db.Group45.remove({"ID":"ID6"})
WriteResult({ "nRemoved" : 1 })
> db.Group45.find({"ID":"ID6"}).pretty()
> db.Group45.find().pretty()
{
  "_id" : ObjectId("5e46b820b1509b82dc910edd"),
  "ID" : "ID1",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE62232",
  "No of Samples" : 91,
  "Type of Samples" : "81 HCC and 10 non-tumor",
  "Pubmed ID" : 25822088
}

{
  "_id" : ObjectId("5e46b820b1509b82dc910ede"),
  "ID" : "ID2",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE63067",
  "No of Samples" : 18,
  "Type of Samples" : "Steatosis, healthy and Non-alcoholic steatohepatitis",
  "Pubmed ID" : 25993042
}

{
  "_id" : ObjectId("5e46b820b1509b82dc910edf"),
  "ID" : "ID3",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE64041",
  "No of Samples" : 125,
  "Type of Samples" : "60 pairs HCC and non-tumor, 5 normal",
  "Pubmed ID" : 27499918
}

{
  "_id" : ObjectId("5e46b820b1509b82dc910ee0"),
  "ID" : "ID4",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE69715",
  "No of Samples" : 103,
  "Type of Samples" : "HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor",
  "Pubmed ID" : 29538454
}

{
  "_id" : ObjectId("5e46b820b1509b82dc910ee1"),
  "ID" : "ID5",
  "Profiling Technique" : "Affymetrix Array",
  "Dataset ID" : "GSE72981",
  "No of Samples" : 30,
  "Type of Samples" : "Subcutaneous and orthotopic HCC tissue",
  "Pubmed ID" : 26520397
}

```


Hive Question 3.

Hive commands and demos:

(With pre-existing data from Group45.csv)

Show databases command is used for showing all the databases stored within it.

1. **Create databases** command is used for creating the database.
Create table command is used for creating the table within the databases.
2. **Select** command is used to select a set of rows from the table.
3. **Alter** command is used in the query to edit/update columns of the table.
4. **Drop** command is used to drop the table from the database although the database still persists.

The query for all the commands has been shown in the images attached below.

```
root@677d90630142:/usr/local/hive/lib
hbase-common-2.0.0-alpha4.jar
hbase-hadoop-compat-2.0.0-alpha4.jar
hbase-hadoop2-compat-2.0.0-alpha4-tests.jar
hbase-hadoop2-compat-2.0.0-alpha4.jar
hbase-http-2.0.0-alpha4.jar
hbase-mapreduce-2.0.0-alpha4.jar
hbase-metrics-2.0.0-alpha4.jar
hbase-metrics-api-2.0.0-alpha4.jar
hbase-prefix-tree-2.0.0-alpha4.jar
hbase-procedure-2.0.0-alpha4.jar
javax.servlet.jsp-api-2.3.1.jar
javax.ws.rs-api-2.0.1.jar
javolution-5.5.1.jar
jaxb-api-2.2.11.jar
jcodings-1.0.18.jar
jcommander-1.32.jar
jdo-api-3.0.1.jar
jersey-client-2.25.1.jar
jersey-common-2.25.1.jar
jersey-container-servlet-core-2.25.1.jar
transaction-api-1.1.jar
validation-api-1.1.0.Final.jar
velocity-1.5.jar
websocket-api-9.3.20.v20170531.jar
websocket-client-9.3.20.v20170531.jar
websocket-common-9.3.20.v20170531.jar
websocket-server-9.3.20.v20170531.jar
websocket-servlet-9.3.20.v20170531.jar
zookeeper-3.4.6.jar
root@677d90630142:/usr/local/hive/lib# hive
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/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 = 80448107-5b04-4e28-a8cc-7f1e2b61d7eb

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = 14b98b57-345c-484b-8b29-4f8976684006
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> create database hdb;
OK
Time taken: 0.774 seconds
hive> show databases;
OK
default
hdb
Time taken: 0.18 seconds, Fetched: 2 row(s)
hive> use hdb;
OK
Time taken: 0.037 seconds
hive> exit;
root@677d90630142:/usr/local/hive/lib# hive
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
WARNING: HADOOP_PREFIX has been replaced by HADOOP_HOME. Using value of HADOOP_PREFIX.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/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 = b6cb5be7-7aed-4a5b-987f-2ad963b21956

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive Session ID = f6c3e83d-228e-4065-b99d-7ea9c3c519b0
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>
```



```

hive> create database hdb;
OK
Time taken: 0.774 seconds
hive> show databases;
OK
default
hdb
Time taken: 0.18 seconds, Fetched: 2 row(s)
hive> use hdb;
OK
Time taken: 0.037 seconds

```

```

hive> CREATE EXTERNAL TABLE IF NOT EXISTS Group45(
>   id String,
>   technique String,
>   d_id String,
>   n_samples String,
>   t_samples String,
>   pubmed_id String
> )
> ROW FORMAT DELIMITED
> FIELDS TERMINATED BY '\t'
> STORED AS TEXTFILE
> LOCATION '/user/root/hdb';
OK
Time taken: 1.34 seconds
hive>

```

```

hive> Select * from Group45
> ;
OK

```

ID	Profiling Technique	Dataset ID	No of Samples	Type of Samples	Pubmed ID
ID1	Affymetrix Array	GSE62232	91	81 HCC and 10 non-tumor	25822088
ID2	Affymetrix Array	GSE63067	18	Steatosis, healthy and Non-alcoholic steatohepatitis	25993042
ID3	Affymetrix Array	GSE64041	125	60 pairs HCC and non-tumor, 5 normal	27499918
ID4	Affymetrix Array	GSE69715	103	HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor	29538454
ID5	Affymetrix Array	GSE72981	30	Subcutaneous and orthotopic HCC tissue	26520397

```

Time taken: 2.048 seconds, Fetched: 6 row(s)
hive>

```

```

> ;
hive> ALTER TABLE Group45 REPLACE COLUMNS(id String, technique String, d_id String, n_samples String, t_samples String);
OK
Time taken: 0.123 seconds
hive> Select * from Group45
> ;
OK

```

ID	Profiling Technique	Dataset ID	No of Samples	Type of Samples
ID1	Affymetrix Array	GSE62232	91	81 HCC and 10 non-tumor
ID2	Affymetrix Array	GSE63067	18	Steatosis, healthy and Non-alcoholic steatohepatitis
ID3	Affymetrix Array	GSE64041	125	60 pairs HCC and non-tumor, 5 normal
ID4	Affymetrix Array	GSE69715	103	HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor
ID5	Affymetrix Array	GSE72981	30	Subcutaneous and orthotopic HCC tissue

```

Time taken: 0.121 seconds, Fetched: 6 row(s)
hive>

```

```

hive> DROP TABLE IF EXISTS Group45;
OK
Time taken: 0.404 seconds
hive>

```

JSON Question 4.

Using pandas and json library we convert the Group45.csv to json format.

```
root@677d90630142:/usr/local/hive/lib# cd /
root@677d90630142:/# python3 /code/answer3.py
python3: can't open file '/code/answer3.py': [Errno 2] No such file or directory
root@677d90630142:/# cd /code/
root@677d90630142:/code# ls
Group45.csv  answer1.py  answer2.py  answer4.py
root@677d90630142:/code# python3 answer4.py
[{"ID": "ID1", "Profiling Technique": "Affymetrix Array", "Dataset ID": "GSE62232", "No of Samples": 91, "Type of Samples": "81 HCC and 10 non-tumor", "Pubmed ID": 25822088}, {"ID": "ID2", "Profiling Technique": "Affymetrix Array", "Dataset ID": "GSE63067", "No of Samples": 18, "Type of Samples": "Steatosis, healthy and Non-alcoholic steatohepatitis", "Pubmed ID": 25993042}, {"ID": "ID3", "Profiling Technique": "Affymetrix Array", "Dataset ID": "GSE64041", "No of Samples": 125, "Type of Samples": "60 pairs HCC and non-tumor, 5 normal", "Pubmed ID": 27499918}, {"ID": "ID4", "Profiling Technique": "Affymetrix Array", "Dataset ID": "GSE69715", "No of Samples": 103, "Type of Samples": "HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor", "Pubmed ID": 29538454}, {"ID": "ID5", "Profiling Technique": "Affymetrix Array", "Dataset ID": "GSE72981", "No of Samples": 30, "Type of Samples": "Subcutaneous and orthotopic HCC tissue", "Pubmed ID": 26520397}]
```

```
{
  "ID": "ID1",
  "Dataset ID": "GSE62232",
  "No of Samples": 91,
  "Profiling Technique": "Affymetrix Array",
  "Pubmed ID": 25822088,
  "Type of Samples": "81 HCC and 10 non-tumor"
},
{
  "ID": "ID2",
  "Dataset ID": "GSE63067",
  "No of Samples": 18,
  "Profiling Technique": "Affymetrix Array",
  "Pubmed ID": 25993042,
  "Type of Samples": "Steatosis, healthy and Non-alcoholic steatohepatitis"
},
{
  "ID": "ID3",
  "Dataset ID": "GSE64041",
  "No of Samples": 125,
  "Profiling Technique": "Affymetrix Array",
  "Pubmed ID": 27499918,
  "Type of Samples": "60 pairs HCC and non-tumor, 5 normal"
},
{
  "ID": "ID4",
  "Dataset ID": "GSE69715",
  "No of Samples": 103,
  "Profiling Technique": "Affymetrix Array",
  "Pubmed ID": 29538454,
  "Type of Samples": "HCV-associated hepatocellular carcinoma (HCC) and adjacent non tumor"
},
{
  "ID": "ID5",
  "Dataset ID": "GSE72981",
  "No of Samples": 30,
  "Profiling Technique": "Affymetrix Array",
  "Pubmed ID": 26520397,
  "Type of Samples": "Subcutaneous and orthotopic HCC tissue"
}
```