Big Data in Healthcare
Assignment 1

Swati Verma(MT19073)
Pragya Dara(MT19126)

Question 1: Mysql

Step by step procedure to run mysql on docker:

1. **Listing all docker images**
   $ Docker images

```
Pragyas-MacBook-Pro:~ pragyadara$ docker images
REPOSITORY           TAG                 IMAGE ID            CREATED             SIZE
<none>               <none>              77fa6244e5a2        6 minutes ago       465MB
mysql                8.0.19              791b6e40940c        10 days ago         465MB
mysql                latest              791b6e40940c        10 days ago         465MB
ubuntu               latest              ccc6e87d482b        3 weeks ago         64.2MB
mysql/mysql-server   5.7                 2a6c84ecfcb2        4 weeks ago         334MB
mysql/mysql-server   8.0                 a7a39f15d42d        4 weeks ago         381MB
mysql/mysql-server   latest              a7a39f15d42d        4 weeks ago         381MB
```

2. **Creating and running container**
   $ docker run --name final_container -e MYSQL_ROOT_PASSWORD=pragyasql -d mysql:latest
   $ docker exec -it final_container mysql -u root -p

3. **Shows the status of all the running processes with their Ids.**
   $ docker ps -a

```
Pragyas-MacBook-Pro:~ pragyadara$ docker ps -a
CONTAINER ID        IMAGE               COMMAND                  CREATED             STATUS                     PORTS                 NAMES
7fd3a35c87f5        mysql:latest        "docker-entrypoint.s…"   47 seconds ago      Up 46 seconds              3306/tcp, 33060/tcp   final_container
03aab13fc4af        mysql:latest        "docker-entrypoint.s…"   4 minutes ago       Up 4 minutes               3306/tcp, 33060/tcp   new_container
7ace4f0b9ee0        mysql               "docker-entrypoint.s…"   12 minutes ago      Up 12 minutes              3306/tcp, 33060/tcp   mysql1
ff59fba81c9f        mysql:latest        "docker-entrypoint.s…"   45 hours ago        Exited (1) 44 hours ago                          pragya_ques1_final
20f8e60cd2a3        mysql:8.0.19        "docker-entrypoint.s…"   6 days ago          Exited (255) 12 hours ago  3306/tcp, 33060/tcp   pragya_ques1
Pragyas-MacBook-Pro:~ pragyadara$ docker exec -it final_container mysql:latest -u root  -p
OCI runtime exec failed: exec failed: container_linux.go:346: starting container process caused "exec: \"mysql:latest\": executable file not found in $PATH": unknown
Pragyas-MacBook-Pro:~ pragyadara$ docker exec -it final_container mysql -u root  -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 8
Server version: 8.0.19 MySQL Community Server - GPL

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.
```

4. **Creating database mydb, creating table mytable and inserting values.**

> **create** database mydb;

> **use** mydb;

> **create** table mytable(id char(3) PRIMARY KEY, profiling_technique varchar(30), dataset_id char(8), no_of_samples int, type_of_samples varchar (500), pumbed_id int);

> **insert** into mytable values('ID1', 'Affymetrix Array', 'GSE45050', 16, 'HCC, Cirrhosis and adjacent non-tumor', 24497316);

> **insert** into mytable values('ID2', 'Affymetrix Array', 'GSE45267', 87, '48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients' , 30411085);

> **insert** into mytable values('ID3', 'Affymetrix Array', GSE45434, 16 , ' HCC ' , NA);

> **insert** into mytable values('ID4', 'Affymetrix Array', GSE45435, 31, ' HCC ' , NA);

> **insert** into mytable values('ID5', 'Affymetrix Array', GSE51401, 64, ' Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC ' , NA);

```
mysql> create database mydb;
Query OK, 1 row affected (0.01 sec)

mysql> use mydb;
Database changed
mysql> create table mytable(id char(3) PRIMARY KEY, profiling_technique varchar(10), dataset_id char(8), no_of_samples int, type_of_samples varchar(500),pumbed_id int);
Query OK, 0 rows affected (0.02 sec)

mysql> insert into mytable values('ID1', 'Affymetrix Array','GSE45050',16,'HCC, Cirrhosis and adjacent non-tumor',24497316);
ERROR 1406 (22001): Data too long for column 'profiling_technique' at row 1
mysql> drop table mytable;
Query OK, 0 rows affected (0.02 sec)

mysql> create table mytable(id char(3) PRIMARY KEY, profiling_technique varchar(30), dataset_id char(8), no_of_samples int, type_of_samples varchar(500),pumbed_id int);
Query OK, 0 rows affected (0.02 sec)

mysql> insert into mytable values('ID1', 'Affymetrix Array','GSE45050',16,'HCC, Cirrhosis and adjacent non-tumor',24497316);
Query OK, 1 row affected (0.01 sec)

mysql> insert into mytable values('ID2', 'Affymetrix Array','GSE45267',87,'48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients',30411085);
Query OK, 1 row affected (0.01 sec)

mysql> insert into mytable values('ID3', 'Affymetrix Array','GSE45434',16,'HCC',NULL);
Query OK, 1 row affected (0.01 sec)

mysql> insert into mytable values('ID4', 'Affymetrix Array','GSE45435',31,'HCC',NULL);
Query OK, 1 row affected (0.01 sec)

mysql> insert into mytable values('ID5', 'Affymetrix Array','GSE51401',64,'Primary uncultured CD31+ and CD105+ tumor endothelial (TEC), non-tumor endothelial cells(NEC), remnant cells from tumor(TC) and n
on-tumor liver tissue(NTC) of HCC',NULL);
Query OK, 1 row affected (0.00 sec)

mysql> select * from mytable;
+-----+--------------------+------------+--------------+-----------------------------------------------------------------------------------------------------------------------------------------------------------------+
------------------+------------+
| id  | profiling_technique | dataset_id | no_of_samples | type_of_samples
                  | pumbed_id |
+-----+--------------------+------------+--------------+-----------------------------------------------------------------------------------------------------------------------------------------------------------------+
------------------+------------+
| ID1 | Affymetrix Array    | GSE45050   |           16 | HCC, Cirrhosis and adjacent non-tumor
                  |  24497316 |
| ID2 | Affymetrix Array    | GSE45267   |           87 | 48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients
                  |  30411085 |
| ID3 | Affymetrix Array    | GSE45434   |           16 | HCC
                  |      NULL |
| ID4 | Affymetrix Array    | GSE45435   |           31 | HCC
                  |      NULL |
| ID5 | Affymetrix Array    | GSE51401   |           64 | Primary uncultured CD31+ and CD105+ tumor endothelial (TEC), non-tumor endothelial cells(NEC), remnant cells from tumor(TC) and non-tumor liver t
issue(NTC) of HCC |      NULL |
+-----+--------------------+------------+--------------+-----------------------------------------------------------------------------------------------------------------------------------------------------------------+
------------------+------------+
5 rows in set (0.00 sec)
```

Executing various mysql commands:

I. update mytable set no_of_samples=29 where id='ID3';

The update command updates and sets the value of number_of_samples to 29 where id is equal to ID3.

```
mysql> update mytable set no_of_samples=29 where id='ID3';
Query OK, 1 row affected (0.00 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> select * from mytable;
+-----+---------------------+------------+---------------+------------------------------------------------------------------------------------------------------------------------------------------------------
------------------+-----------+
| id  | profiling_technique | dataset_id | no_of_samples | type_of_samples
                  | pumbed_id |
+-----+---------------------+------------+---------------+------------------------------------------------------------------------------------------------------------------------------------------------------
------------------+-----------+
| ID1 | Affymetrix Array    | GSE45050   |            16 | HCC, Cirrhosis and adjacent non-tumor
                  |  24497316 |
| ID2 | Affymetrix Array    | GSE45267   |            87 | 48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients
                  |  30411085 |
| ID3 | Affymetrix Array    | GSE45434   |            29 | HCC
                  |      NULL |
| ID4 | Affymetrix Array    | GSE45435   |            31 | HCC
                  |      NULL |
| ID5 | Affymetrix Array    | GSE51401   |            64 | Primary uncultured CD31+ and CD105+ tumor endothelial (TEC), non-tumor endothelial cells(NEC), remnant cells from tumor(TC) and non-tumor liver t
issue(NTC) of HCC |      NULL |
+-----+---------------------+------------+---------------+------------------------------------------------------------------------------------------------------------------------------------------------------
------------------+-----------+
5 rows in set (0.00 sec)
```

II. Delete from mytable where no_of_samples> 80;

The above delete command deletes all the number_of_samples whose values are greater than 80.

```
mysql> delete from mytable where no_of_samples>80;
Query OK, 1 row affected (0.00 sec)

mysql> select * from mytable;
+-----+---------------------+------------+---------------+------------------------------------------------------------------------------------------------------------------------------------------------------
------------------+-----------+
| id  | profiling_technique | dataset_id | no_of_samples | type_of_samples
                  | pumbed_id |
+-----+---------------------+------------+---------------+------------------------------------------------------------------------------------------------------------------------------------------------------
------------------+-----------+
| ID1 | Affymetrix Array    | GSE45050   |            16 | HCC, Cirrhosis and adjacent non-tumor
                  |  24497316 |
| ID3 | Affymetrix Array    | GSE45434   |            29 | HCC
                  |      NULL |
| ID4 | Affymetrix Array    | GSE45435   |            31 | HCC
                  |      NULL |
| ID5 | Affymetrix Array    | GSE51401   |            64 | Primary uncultured CD31+ and CD105+ tumor endothelial (TEC), non-tumor endothelial cells(NEC), remnant cells from tumor(TC) and non-tumor liver t
issue(NTC) of HCC |      NULL |
+-----+---------------------+------------+---------------+------------------------------------------------------------------------------------------------------------------------------------------------------
------------------+-----------+
4 rows in set (0.00 sec)
```

III. Insert into mytable values('ID6','Affymetrix Array', 'GSE73126', 53, 'HCC', NULL);
The above insert command inserts a new tuple with the value of id=ID6, profiling_technique= " Affymetrix Array " , dataset_id= "'GSE73126" , no_of_samples = 53, types_of_samples= " HCC" and pumbed_id= NULL.

```
mysql> insert into mytable values('ID6','Affymetrix Array','GSE73126',53,'HCC',NULL);
Query OK, 1 row affected (0.00 sec)

mysql> select * from mytable;
+-----+---------------------+------------+---------------+------------------------------------------
------------------+-----------+
| id  | profiling_technique | dataset_id | no_of_samples | type_of_samples
                  | pumbed_id |
+-----+---------------------+------------+---------------+------------------------------------------
------------------+-----------+
| ID1 | Affymetrix Array    | GSE45050   |            16 | HCC, Cirrhosis and adjacent non-tumor
                  |  24497316 |
| ID3 | Affymetrix Array    | GSE45434   |            29 | HCC
                  |      NULL |
| ID4 | Affymetrix Array    | GSE45435   |            31 | HCC
                  |      NULL |
| ID5 | Affymetrix Array    | GSE51401   |            64 | Primary uncultured CD31+ and CD105+ tumor endothelial (TEC), non-tumor endothelial cells(NEC), remnant cells from tumor(TC) and non-tumor liver t
issue(NTC) of HCC |      NULL |
| ID6 | Affymetrix Array    | GSE73126   |            53 | HCC
                  |      NULL |
+-----+---------------------+------------+---------------+------------------------------------------
------------------+-----------+
5 rows in set (0.00 sec)
```

IV. Update mytable set pumbed_id=35678123 where dataset_id='GCE73126' ;
The above update command set the value of pumbed_id to 3578123 where dataset_id is 'GCE73126'.

```
mysql> update mytable set pumbed_id=35678123 where dataset_id='GSE73126';
Query OK, 1 row affected (0.01 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> select * from mytable;
+-----+---------------------+------------+---------------+------------------------------------------
------------------+-----------+
| id  | profiling_technique | dataset_id | no_of_samples | type_of_samples
                  | pumbed_id |
+-----+---------------------+------------+---------------+------------------------------------------
------------------+-----------+
| ID1 | Affymetrix Array    | GSE45050   |            16 | HCC, Cirrhosis and adjacent non-tumor
                  |  24497316 |
| ID3 | Affymetrix Array    | GSE45434   |            29 | HCC
                  |      NULL |
| ID4 | Affymetrix Array    | GSE45435   |            31 | HCC
                  |      NULL |
| ID5 | Affymetrix Array    | GSE51401   |            64 | Primary uncultured CD31+ and CD105+ tumor endothelial (TEC), non-tumor endothelial cells(NEC), remnant cells from tumor(TC) and non-tumor liver t
issue(NTC) of HCC |      NULL |
| ID6 | Affymetrix Array    | GSE73126   |            53 | HCC
                  |  35678123 |
+-----+---------------------+------------+---------------+------------------------------------------
------------------+-----------+
5 rows in set (0.00 sec)
```

4. **Creating new image from the container's change.**

   $ **docker commit** 7fd3a35c87f5;

```
Pragyas-MacBook-Pro:~ pragyadara$ docker commit 7fd3a35c87f5
sha256:7d87f3eafe9c8067c80311f5ea8d05664e224557695f34c226687a49cd8e7c71
Pragyas-MacBook-Pro:~ pragyadara$
```

5. **Pushing image inside the repository.**

   Create tag:

   $ docker tag mysql:latest paudara/14_mt19126_mt19073_a1:ques1

   $ **docker push** paudara/14_mt19126_mt19073_a1: ques1

```
Pragyas-MacBook-Pro:~ pragyadara$ docker tag mysql:latest paudara/14_mt19126_mt19073_a1:ques1
Pragyas-MacBook-Pro:~ pragyadara$ docker push paudara/14_mt19126_mt19073_a1:ques1
The push refers to repository [docker.io/paudara/14_mt19126_mt19073_a1]
cdae83e68539: Pushed
c1f58dc402c7: Pushed
b7a46f8264c1: Pushed
15669f7521b3: Pushed
15669f7521b3: Pushing  96.64MB/350.1MB
6b5c7baa4da8: Pushed
76db703007bc: Pushed
cee57cdf5101: Pushed
1a527f11e03e: Pushed
4dac9b6b28ce: Pushed
605f8f2fe1e5: Pushed
e0db3ba0aaea: Pushing [==============================>                ]  36.e0db3ba0aaea: Pushed
ques1: digest: sha256:283caa87ee6a3997b32beaff33e75974fd689edfd2df0084a45e300d8a001c91 size: 2828
Pragyas-MacBook-Pro:~ pragyadara$ docker pull paudara/14_mt19126_mt19073_a1:ques1
ques1: Pulling from paudara/14_mt19126_mt19073_a1
Digest: sha256:283caa87ee6a3997b32beaff33e75974fd689edfd2df0084a45e300d8a001c91
Status: Image is up to date for paudara/14_mt19126_mt19073_a1:ques1
docker.io/paudara/14_mt19126_mt19073_a1:ques1
```

Question 2: MongoDB

Step by step procedure to run mysql on docker:

1. **Pulling mongo image from docker.**

   $ docker pull mongo

```
Pragyas-MacBook-Pro:~ pragyadara$ docker pull mongo
Using default tag: latest
latest: Pulling from library/mongo
5c939e3a4d10: Already exists
c63719cdbe7a: Already exists
19a861ea6baf: Already exists
651c9d2d6c4f: Already exists
85155c6d5fac: Pull complete
85fb0780fd97: Pull complete
85b3b1a901f5: Pull complete
6a882e007bb6: Pull complete
f7806503a70f: Pull complete
5732cde4308d: Pull complete
8f892a804391: Pull complete
afc61ce39de5: Pull complete
479082b17a4a: Pull complete
Digest: sha256:14b612325925ca60d9ccbc710aa4c2dbfb74106229f60f4fee9d42fab0281f6f
Status: Downloaded newer image for mongo:latest
```

2. **Creating a container.**
   $docker run --name mongo_container -d mongo:latest

```
Pragyas-MacBook-Pro:~ pragyadara$ docker run --name mongo_container -d mongo:latest
fa769db76aa2fb7705dee413e5fead8f5815f11f59dcc16f1851ba23e5824bfener -d mongo:tag
```

3. **Executing a docker container.**
   $ docker exec -it mongo_container bash

```
Pragyas-MacBook-Pro:~ pragyadara$ docker exec -it mongo_container bash
root@fa769db76aa2:/# mongo
MongoDB shell version v4.2.3
connecting to: mongodb://127.0.0.1:27017/?compressors=disabled&gssapiServiceName=mongodb
Implicit session: session { "id" : UUID("ffb99bf8-e473-46fa-98e5-a1b0569471af") }
MongoDB server version: 4.2.3
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:
2020-02-12T17:26:32.428+0000 I  STORAGE  [initandlisten]
2020-02-12T17:26:32.428+0000 I  STORAGE  [initandlisten] ** WARNING: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine
2020-02-12T17:26:32.429+0000 I  STORAGE  [initandlisten] **          See http://dochub.mongodb.org/core/prodnotes-filesystem
2020-02-12T17:26:32.985+0000 I  CONTROL  [initandlisten]
2020-02-12T17:26:32.986+0000 I  CONTROL  [initandlisten] ** WARNING: Access control is not enabled for the database.
2020-02-12T17:26:32.986+0000 I  CONTROL  [initandlisten] **          Read and write access to data and configuration is unrestricted.
2020-02-12T17:26:32.986+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()
---
```

4. **Command to show all the databases.**
   > show dbs

```
> show dbs
admin   0.000GB
config  0.000GB
local   0.000GB
> use mydb
```

5. **Inserting values in MongoDB collection.**

> **db.mytable.save**({ id: “ID1”, profiling_technique : “ Affymetrix Array”, dataset_id : “ GSE45050”, no_of_samples :16, type_of_samples : “ HCC , Cirrhosis and adjacent non-tumor”, pumbed_id : 24497316 })

> **db.mytable.save**({ id: “ID2”, profiling_technique : “ Affymetrix Array”, dataset_id : “ GSE45267”, no_of_samples :87, type_of_samples : “ 48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients”, pumbed_id : 30411085})

> **db.mytable.save**({ id: “ID3”, profiling_technique : “ Affymetrix Array”, dataset_id : “ GSE45434”, no_of_samples : 16, type_of_samples : “ HCC ”, pumbed_id : null})

>**db.mytable.save**({ id: “ID4”, profiling_technique : “ Affymetrix Array”, dataset_id : “ GSE45435”, no_of_samples : 31, type_of_samples : “ HCC ”, pumbed_id : null})

>**db.mytable.save**({ id: “ID5”, profiling_technique : “ Affymetrix Array”, dataset_id : “ GSE51401”, no_of_samples : 64, type_of_samples : “ Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC”, pumbed_id : null})

```
> db.mytable.save({id:"ID1",profiling_technique:"Affymetrix Array",dataset_id:"GSE45050",no_of_samples: 16,type_of_samples:"HCC, Cirrhosis and adjacent non-tumor",pumbed_id:24497316})
WriteResult({ "nInserted" : 1 })
> db.mytable.save({id:"ID2",profiling_technique:"Affymetrix Array",dataset_id:"GSE45267",no_of_samples:87,type_of_samples:"48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients",pumbed_id:30411085})
WriteResult({ "nInserted" : 1 })
> db.mytable.save({id:"ID3",profiling_technique:"Affymetrix Array",dataset_id:"GSE45434",no_of_samples:16,type_of_samples:"HCC",pumbed_id:null})
WriteResult({ "nInserted" : 1 })
> db.mytable.save({id:"ID4",profiling_technique:"Affymetrix Array",dataset_id:"GSE45435",no_of_samples:31,type_of_samples:"HCC",pumbed_id:null})
WriteResult({ "nInserted" : 1 })
> db.mytable.save({id:"ID5",profiling_technique:"Affymetrix Array",dataset_id:"GSE51401",no_of_samples:64,type_of_samples:"Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC",pumbed_id:null})
WriteResult({ "nInserted" : 1 })
```

Executing four commands of mongoDB:

I. **db.updateOne**({id:’ID3’},{$set : { ‘no_of_samples’:29} })

The updateOne command updates and sets the value of number_of_samples to 29 where id is equal to ID3

```
> db.mytable.updateOne({id:'ID3'}, { $set: { "no_of_samples":29} } )
{ "acknowledged" : true, "matchedCount" : 1, "modifiedCount" : 1 }
> db.inventory.find( {} )
> db.mytable.find( {} )
{ "_id" : ObjectId("5e4440ff413d469de491ed47"), "id" : "ID1", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45050", "no_of_samples" : 16, "type_of_samples" : "HCC, Cirrhosis and adjacent non-tumor", "pumbed_id" : 24497316 }
{ "_id" : ObjectId("5e444195413d469de491ed48"), "id" : "ID2", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45267", "no_of_samples" : 87, "type_of_samples" : "48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients", "pumbed_id" : 30411085 }
{ "_id" : ObjectId("5e444212413d469de491ed49"), "id" : "ID3", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45434", "no_of_samples" : 29, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e44422c413d469de491ed4a"), "id" : "ID4", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45435", "no_of_samples" : 31, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e444274413d469de491ed4b"), "id" : "ID5", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE51401", "no_of_samples" : 64, "type_of_samples" : "Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC", "pumbed_id" : null }
```

II. **db.mytable.remove**({ ‘no_of_samples’ : {$gt :80 } })

The above remove command deletes all the number_of_samples whose values are greater than 80.

```
> db.mytable.remove({"no_of_samples": {$gt : 80}})
WriteResult({ "nRemoved" : 1 })
> db.mytable.find( {} )
{ "_id" : ObjectId("5e4440ff413d469de491ed47"), "id" : "ID1", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45050", "no_of_samples" : 16, "type_of_samples" : "HCC, Cirrhosis and adjacent non-tumor", "pumbed_id" : 24497316 }
{ "_id" : ObjectId("5e444212413d469de491ed49"), "id" : "ID3", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45434", "no_of_samples" : 29, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e44422c413d469de491ed4a"), "id" : "ID4", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45435", "no_of_samples" : 31, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e444274413d469de491ed4b"), "id" : "ID5", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE51401", "no_of_samples" : 64, "type_of_samples" : "Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC", "pumbed_id" : null }
```

III. **db.mytable.save({**id:’ID6’, profiling_technique:’Affymetrix Array’, dataset_id : ’GSE73126’, no_of_samples:53, type_of_samples:’HCC’, pumbed_id:null})

The above save command inserts a new tuple with the value of id=ID6, profiling_technique= “ Affymetrix Array ” , dataset_id= “‘GSE73126” , no_of_samples = 53, types_of_samples= “ HCC” and pumbed_id= NULL.

```
> db.mytable.save({no:"ID6",profiling_technique:"Affymetrix Array",dataset_id:"GSE73126",no_of_samples:53 ,type_of_samples:"HCC",pumbed_id:null})
WriteResult({ "nInserted" : 1 })
> db.mytable.find( {} )
{ "_id" : ObjectId("5e4440ff413d469de491ed47"), "id" : "ID1", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45050", "no_of_samples" : 16, "type_of_samples" : "HCC, Cirrhosis and adjacent non-tumor", "pumbed_id" : 24497316 }
{ "_id" : ObjectId("5e444212413d469de491ed49"), "id" : "ID3", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45434", "no_of_samples" : 29, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e44422c413d469de491ed4a"), "id" : "ID4", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45435", "no_of_samples" : 31, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e444274413d469de491ed4b"), "id" : "ID5", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE51401", "no_of_samples" : 64, "type_of_samples" : "Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e444812413d469de491ed4c"), "no" : "ID6", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE73126", "no_of_samples" : 53, "type_of_samples" : "HCC", "pumbed_id" : null }
```

IV. **db.mytable.updateOne**({ dataset_id: ‘GSE73126}, { $set :{ ‘pumbed_id’ :35678123} })

The above updateOne command set the value of pumbed_id to 3578123 where dataset_id is ’GCE73126’.

```
> db.mytable.updateOne({dataset_id:'GSE73126'}, { $set: { "pumbed_id":35678123} } )
{ "acknowledged" : true, "matchedCount" : 1, "modifiedCount" : 1 }
> db.mytable.find( {} )
{ "_id" : ObjectId("5e4440ff413d469de491ed47"), "id" : "ID1", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45050", "no_of_samples" : 16, "type_of_samples" : "HCC, Cirrhosis and adjacent
non-tumor", "pumbed_id" : 24497316 }
{ "_id" : ObjectId("5e444212413d469de491ed49"), "id" : "ID3", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45434", "no_of_samples" : 29, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e44422c413d469de491ed4a"), "id" : "ID4", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE45435", "no_of_samples" : 31, "type_of_samples" : "HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e444274413d469de491ed4b"), "id" : "ID5", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE51401", "no_of_samples" : 64, "type_of_samples" : "Primary uncultured CD31+ and
 CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC", "pumbed_id" : null }
{ "_id" : ObjectId("5e444812413d469de491ed4c"), "no" : "ID6", "profiling_technique" : "Affymetrix Array", "dataset_id" : "GSE73126", "no_of_samples" : 53, "type_of_samples" : "HCC", "pumbed_id" : 35678123
 }
```

6. **Committing changes and pushing the newly created image into a docker repository.**
   $docker commit fa769db76aa2

```
Pragyas-MacBook-Pro:~ pragyadara$ docker commit fa769db76aa2
sha256:c27764ae82d7d5ae0e2fad3c0b9d65133f13f9f7fd4bcde3f8be9ff2bd3faabd
Pragyas-MacBook-Pro:~ pragyadara$ docker tag mongo:latest paudara/14_mt19126_mt19073_a1:ques2
Pragyas-MacBook-Pro:~ pragyadara$ docker push paudara/14_mt19126_mt19073_a1:ques2
The push refers to repository [docker.io/paudara/14_mt19126_mt19073_a1]
d7283debbc5d: Mounted from library/mongo
c1627b609564: Mounted from library/mongo
545e0cb89d63: Mounted from library/mongo
1771747534c7: Mounted from library/mongo
af5953d78f77: Mounted from library/mongo
30d62c5c51c9: Mounted from library/mongo
309081bde8f2: Mounted from library/mongo
f191801be505: Mounted from library/mongo
69b7a17e61f7: Mounted from library/mongo
f55aa0bd26b8: Mounted from library/mongo
1d0dfb259f6a: Mounted from library/mongo
21ec61b65b20: Mounted from library/mongo
43c67172d1d1: Mounted from library/mongo
ques2: digest: sha256:ee221d2fbe26df2c765088e9d14cf1a7d361ef30f3bc09f52595400745ffcff7 size: 303
```

Question 3. Hadoop

1. Pull Image

```
Pragyas-MacBook-Pro:~ pragyadara$ docker images
REPOSITORY                      TAG                 IMAGE ID            CREATED             SIZE
<none>                          <none>              c27764ae82d7        36 hours ago        386MB
<none>                          <none>              d2dcdb251acc        39 hours ago        465MB
<none>                          <none>              77fa6244e5a2        40 hours ago        465MB
mongo                           latest              e43a2492d00f        3 days ago          386MB
paudara/14_mt19126_mt19073_a1   ques2               e43a2492d00f        3 days ago          386MB
mysql                           8.0.19              791b6e40940c        12 days ago         465MB
mysql                           latest              791b6e40940c        12 days ago         465MB
ques1                           latest              791b6e40940c        12 days ago         465MB
paudara/14_mt19126_mt19073_a1   ques1               791b6e40940c        12 days ago         465MB
ubuntu                          latest              ccc6e87d482b        4 weeks ago         64.2MB
mysql/mysql-server              5.7                 2a6c84ecfcb2        4 weeks ago         334MB
mysql/mysql-server              8.0                 a7a39f15d42d        4 weeks ago         381MB
mysql/mysql-server              latest              a7a39f15d42d        4 weeks ago         381MB
harisekhon/hbase                latest              856957168a1c        6 months ago        416MB
```

2. **Run hbase**
   $ docker run -ti harisekhon/hbase

```
Pragyas-MacBook-Pro:~ pragyadara$ docker run -ti harisekhon/hbase
+ set -euo pipefail
+ '[' -n '' ']'
+++ dirname /entrypoint.sh
++ cd /
++ pwd
+ srcdir=/
+ export HBASE_HOME=/hbase
+ HBASE_HOME=/hbase
+ export JAVA_HOME=/usr
+ JAVA_HOME=/usr
+ echo ================================================================================
================================================================================
+ echo '                              HBase Docker Container'
                             HBase Docker Container
+ echo ================================================================================
================================================================================
+ echo
```

3. **Hbase shell gets started**

```
Now starting HBase Shell...

+ /hbase/bin/hbase shell
2020-02-14 06:39:56,648 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library fo
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.1.3, rda5ec9e4c06c537213883cca8f3cc9a7c19daf67, Mon Feb 11 15:45:33 CST 2019
Took 0.0049 seconds
```

4. **Creating database**
   > create_namespace 'mydb'

```
hbase(main):001:0> create_namespace 'mydb'
Took 0.6900 seconds
```

5. **Creating table**
   >create 'mydb:mytable','tag'

```
hbase(main):002:0> create 'mydb:mytable','tag'
Created table mydb:mytable
Took 0.8604 seconds
```

6. **Insert data into table:**

put 'mydb:mytable','ID1','tag:Profiling_Technique','Affymetrix Array'

put 'mydb:mytable','ID2','tag:Profiling_Technique','Affymetrix Array'

put 'mydb:mytable','ID3','tag:Profiling_Technique','Affymetrix Array'

put 'mydb:mytable','ID4','tag:Profiling_Technique','Affymetrix Array'

put 'mydb:mytable','ID5','tag:Profiling_Technique','Affymetrix Array'

put 'mydb:mytable','ID1','tag:Dataset_id','GSE45050'

put 'mydb:mytable','ID2','tag:Dataset_id','GSE45267'

put 'mydb:mytable','ID3','tag:Dataset_id','GSE45434'
put 'mydb:mytable','ID4','tag:Dataset_id','GSE45435'

put 'mydb:mytable','ID5','tag:Dataset_id','GSE51401'

put 'mydb:mytable','ID1','tag:No_of_samples','16'

put 'mydb:mytable','ID2','tag:No_of_samples','87'

put 'mydb:mytable','ID3','tag:No_of_samples','16'

put 'mydb:mytable','ID4','tag:No_of_samples','31'

put 'mydb:mytable','ID5','tag:No_of_samples','64'

put 'mydb:mytable','ID1','tag:Type_of_samples','HCC, Cirrhosis and adjacent non-tumor'

put 'mydb:mytable','ID2','tag:Type_of_samples','48 primary HCC samples, as well as those of 39 non-cancerous tissues, from 61 patients'

put 'mydb:mytable','ID3','tag:Type_of_samples','HCC'

put 'mydb:mytable','ID4','tag:Type_of_samples','HCC'

put 'mydb:mytable','ID5','tag:Type_of_samples','Primary uncultured CD31+ and CD105+ tumor endothelial cells (TEC), non-tumor endothelial cells (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC'

put 'mydb:mytable','ID1','tag:Pubmed_ID','24497316'

put 'mydb:mytable','ID2','tag:Pubmed_ID','30411085'

put 'mydb:mytable','ID3','tag:Pubmed_ID',''

put 'mydb:mytable','ID4','tag:Pubmed_ID',''

put 'mydb:mytable','ID5','tag:Pubmed_ID',''

```
hbase(main):032:0> scan 'mydb:mytable'
ROW                                                   COLUMN+CELL
 ID1                                                  column=tag:Dataset_id, timestamp=1581662891066, value=GSE45050
 ID1                                                  column=tag:No_of_samples, timestamp=1581663054986, value=16
 ID1                                                  column=tag:Profiling_Technique, timestamp=1581662717558, value=Affymetrix Array
 ID1                                                  column=tag:Pubmed_ID, timestamp=1581663445779, value=24497316
 ID1                                                  column=tag:Type_of_samples, timestamp=1581663234110, value=HCC, Cirrhosis and adjacent non-tumor
 ID2                                                  column=tag:Dataset_id, timestamp=1581662909734, value=GSE45267
 ID2                                                  column=tag:No_of_samples, timestamp=1581663065617, value=87
 ID2                                                  column=tag:Profiling_Technique, timestamp=1581662727107, value=Affymetrix Array
 ID2                                                  column=tag:Pubmed_ID, timestamp=1581663464075, value=30411085
 ID2                                                  column=tag:Type_of_samples, timestamp=1581663324025, value=48 primary HCC samples, as well as those o
 ID3                                                  column=tag:Dataset_id, timestamp=1581662936785, value=GSE45434
 ID3                                                  column=tag:No_of_samples, timestamp=1581663074866, value=16
 ID3                                                  column=tag:Profiling_Technique, timestamp=1581662734898, value=Affymetrix Array
 ID3                                                  column=tag:Type_of_samples, timestamp=1581663362976, value=HCC
 ID4                                                  column=tag:Dataset_id, timestamp=1581662950134, value=GSE45435
 ID4                                                  column=tag:No_of_samples, timestamp=1581663085962, value=31
 ID4                                                  column=tag:Profiling_Technique, timestamp=1581662744054, value=Affymetrix Array
 ID4                                                  column=tag:Type_of_samples, timestamp=1581663372421, value=HCC
 ID5                                                  column=tag:Dataset_id, timestamp=1581662965113, value=GSE51401
 ID5                                                  column=tag:No_of_samples, timestamp=1581663098116, value=64
 ID5                                                  column=tag:Profiling_Technique, timestamp=1581662752293, value=Affymetrix Array
 ID5                                                  column=tag:Type_of_samples, timestamp=1581663396120, value=Primary uncultured CD31+ and CD105+ tumor
                                                      ls (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC
```

Executing four commands of Hadoop:

I. **put** 'mydb:mytable','ID3','tag:No_of_samples','29'

The above command sets the value of No_of_samples =29 where id is equal to ID3.

```
hbase(main):036:0> put 'mydb:mytable','ID3','tag:No_of_samples','29'
Took 0.0040 seconds
hbase(main):037:0> scan 'mydb:mytable'
ROW                                    COLUMN+CELL
 ID1                                   column=tag:Dataset_id, timestamp=1581662891066, value=GSE45050
 ID1                                   column=tag:No_of_samples, timestamp=1581663054986, value=16
 ID1                                   column=tag:Profiling_Technique, timestamp=1581662717558, value=Affymetrix Array
 ID1                                   column=tag:Pubmed_ID, timestamp=1581663445779, value=24497316
 ID1                                   column=tag:Type_of_samples, timestamp=1581663234110, value=HCC, Cirrhosis and adjacent non-tumor
 ID2                                   column=tag:Dataset_id, timestamp=1581662909734, value=GSE45267
 ID2                                   column=tag:No_of_samples, timestamp=1581663065617, value=87
 ID2                                   column=tag:Profiling_Technique, timestamp=1581662727107, value=Affymetrix Array
 ID2                                   column=tag:Pubmed_ID, timestamp=1581663464075, value=30411085
 ID2                                   column=tag:Type_of_samples, timestamp=1581663324025, value=48 primary HCC samples, as well as those o
 ID3                                   column=tag:Dataset_id, timestamp=1581662936785, value=GSE45434
 ID3                                   column=tag:No_of_samples, timestamp=1581663933630, value=29
 ID3                                   column=tag:Profiling_Technique, timestamp=1581662734898, value=Affymetrix Array
 ID3                                   column=tag:Pubmed_ID, timestamp=1581663861013, value=
 ID3                                   column=tag:Type_of_samples, timestamp=1581663362976, value=HCC
 ID4                                   column=tag:Dataset_id, timestamp=1581662950134, value=GSE45435
 ID4                                   column=tag:No_of_samples, timestamp=1581663085962, value=31
 ID4                                   column=tag:Profiling_Technique, timestamp=1581662744054, value=Affymetrix Array
 ID4                                   column=tag:Pubmed_ID, timestamp=1581663868066, value=
 ID4                                   column=tag:Type_of_samples, timestamp=1581663372421, value=HCC
 ID5                                   column=tag:Dataset_id, timestamp=1581662965113, value=GSE51401
 ID5                                   column=tag:No_of_samples, timestamp=1581663098116, value=64
 ID5                                   column=tag:Profiling_Technique, timestamp=1581662752293, value=Affymetrix Array
 ID5                                   column=tag:Pubmed_ID, timestamp=1581663874491, value=
 ID5                                   column=tag:Type_of_samples, timestamp=1581663396120, value=Primary uncultured CD31+ and CD105+ tumor
                                       ls (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC
```

II. **delete** 'mydb:mytable','ID2','tag:No_of_samples'

The above command removes the row corresponding to ID2.

```
ROW                                    COLUMN+CELL
 ID1                                   column=tag:Dataset_id, timestamp=1581662891066, value=GSE45050
 ID1                                   column=tag:No_of_samples, timestamp=1581663054986, value=16
 ID1                                   column=tag:Profiling_Technique, timestamp=1581662717558, value=Affymetrix Array
 ID1                                   column=tag:Pubmed_ID, timestamp=1581663445779, value=24497316
 ID1                                   column=tag:Type_of_samples, timestamp=1581663234110, value=HCC, Cirrhosis and adjacent non-tumor
 ID2                                   column=tag:Dataset_id, timestamp=1581662909734, value=GSE45267
 ID2                                   column=tag:Profiling_Technique, timestamp=1581662727107, value=Affymetrix Array
 ID2                                   column=tag:Pubmed_ID, timestamp=1581663464075, value=30411085
 ID2                                   column=tag:Type_of_samples, timestamp=1581663324025, value=48 primary HCC samples, as well as those o
 ID3                                   column=tag:Dataset_id, timestamp=1581662936785, value=GSE45434
 ID3                                   column=tag:No_of_samples, timestamp=1581663933630, value=29
 ID3                                   column=tag:Profiling_Technique, timestamp=1581662734898, value=Affymetrix Array
 ID3                                   column=tag:Pubmed_ID, timestamp=1581663861013, value=
 ID3                                   column=tag:Type_of_samples, timestamp=1581663362976, value=HCC
 ID4                                   column=tag:Dataset_id, timestamp=1581662950134, value=GSE45435
 ID4                                   column=tag:No_of_samples, timestamp=1581663085962, value=31
 ID4                                   column=tag:Profiling_Technique, timestamp=1581662744054, value=Affymetrix Array
 ID4                                   column=tag:Pubmed_ID, timestamp=1581663868066, value=
 ID4                                   column=tag:Type_of_samples, timestamp=1581663372421, value=HCC
 ID5                                   column=tag:Dataset_id, timestamp=1581662965113, value=GSE51401
 ID5                                   column=tag:No_of_samples, timestamp=1581663098116, value=64
 ID5                                   column=tag:Profiling_Technique, timestamp=1581662752293, value=Affymetrix Array
 ID5                                   column=tag:Pubmed_ID, timestamp=1581663874491, value=
 ID5                                   column=tag:Type_of_samples, timestamp=1581663396120, value=Primary uncultured CD31+ and CD105+ tumor
                                       ls (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC
```

III. **put** 'mydb:mytable','ID6','tag:Profiling_Technique','Affymetrix Array'

put 'mydb:mytable','ID6','tag:Dataset_id','GSE73126'

put 'mydb:mytable','ID6','tag:Type_of_samples','HCC'

put 'mydb:mytable','ID6','tag:Pubmed_ID',''

put 'mydb:mytable','ID6','tag:No_of_samples','53'

The above put command inserts a new tuple

```
hbase(main):046:0> scan 'mydb:mytable'
ROW                                     COLUMN+CELL
 ID1                                    column=tag:Dataset_id, timestamp=1581662891066, value=GSE45050
 ID1                                    column=tag:No_of_samples, timestamp=1581663054986, value=16
 ID1                                    column=tag:Profiling_Technique, timestamp=1581662717558, value=Affymetrix Array
 ID1                                    column=tag:Pubmed_ID, timestamp=1581663445779, value=24497316
 ID1                                    column=tag:Type_of_samples, timestamp=1581663234110, value=HCC, Cirrhosis and adjacent non-tumor
 ID2                                    column=tag:Dataset_id, timestamp=1581662909734, value=GSE45267
 ID2                                    column=tag:Profiling_Technique, timestamp=1581662727107, value=Affymetrix Array
 ID2                                    column=tag:Pubmed_ID, timestamp=1581663464075, value=30411085
 ID2                                    column=tag:Type_of_samples, timestamp=1581663324025, value=48 primary HCC samples, as well as those o
 ID3                                    column=tag:Dataset_id, timestamp=1581662936785, value=GSE45434
 ID3                                    column=tag:No_of_samples, timestamp=1581663933630, value=29
 ID3                                    column=tag:Profiling_Technique, timestamp=1581662734898, value=Affymetrix Array
 ID3                                    column=tag:Pubmed_ID, timestamp=1581663861013, value=
 ID3                                    column=tag:Type_of_samples, timestamp=1581663362976, value=HCC
 ID4                                    column=tag:Dataset_id, timestamp=1581662950134, value=GSE45435
 ID4                                    column=tag:No_of_samples, timestamp=1581663085962, value=31
 ID4                                    column=tag:Profiling_Technique, timestamp=1581662744054, value=Affymetrix Array
 ID4                                    column=tag:Pubmed_ID, timestamp=1581663868066, value=
 ID4                                    column=tag:Type_of_samples, timestamp=1581663372421, value=HCC
 ID5                                    column=tag:Dataset_id, timestamp=1581662965113, value=GSE51401
 ID5                                    column=tag:No_of_samples, timestamp=1581663098116, value=64
 ID5                                    column=tag:Profiling_Technique, timestamp=1581662752293, value=Affymetrix Array
 ID5                                    column=tag:Pubmed_ID, timestamp=1581663874491, value=
 ID5                                    column=tag:Type_of_samples, timestamp=1581663396120, value=Primary uncultured CD31+ and CD105+ tumor
                                        ls (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC
 ID6                                    column=tag:Dataset_id, timestamp=1581664116962, value=GSE73126
 ID6                                    column=tag:Profiling_Technique, timestamp=1581664087423, value=Affymetrix Array
 ID6                                    column=tag:Pubmed_ID, timestamp=1581664189897, value=
 ID6                                    column=tag:Type_of_samples, timestamp=1581664162934, value=HCC
```

IV. **put** 'mydb:mytable','ID6','tag:Pubmed_ID','35678123'

The above command updates and sets the value of pubmed_id to '35678123' where id= ID6.

```
hbase(main):048:0> scan 'mydb:mytable'
ROW                                     COLUMN+CELL
 ID1                                    column=tag:Dataset_id, timestamp=1581662891066, value=GSE45050
 ID1                                    column=tag:No_of_samples, timestamp=1581663054986, value=16
 ID1                                    column=tag:Profiling_Technique, timestamp=1581662717558, value=Affymetrix Array
 ID1                                    column=tag:Pubmed_ID, timestamp=1581663445779, value=24497316
 ID1                                    column=tag:Type_of_samples, timestamp=1581663234110, value=HCC, Cirrhosis and adjacent non-tumor
 ID2                                    column=tag:Dataset_id, timestamp=1581662909734, value=GSE45267
 ID2                                    column=tag:Profiling_Technique, timestamp=1581662727107, value=Affymetrix Array
 ID2                                    column=tag:Pubmed_ID, timestamp=1581663464075, value=30411085
 ID2                                    column=tag:Type_of_samples, timestamp=1581663324025, value=48 primary HCC samples, as well as those o
 ID3                                    column=tag:Dataset_id, timestamp=1581662936785, value=GSE45434
 ID3                                    column=tag:No_of_samples, timestamp=1581663933630, value=29
 ID3                                    column=tag:Profiling_Technique, timestamp=1581662734898, value=Affymetrix Array
 ID3                                    column=tag:Pubmed_ID, timestamp=1581663861013, value=
 ID3                                    column=tag:Type_of_samples, timestamp=1581663362976, value=HCC
 ID4                                    column=tag:Dataset_id, timestamp=1581662950134, value=GSE45435
 ID4                                    column=tag:No_of_samples, timestamp=1581663085962, value=31
 ID4                                    column=tag:Profiling_Technique, timestamp=1581662744054, value=Affymetrix Array
 ID4                                    column=tag:Pubmed_ID, timestamp=1581663868066, value=
 ID4                                    column=tag:Type_of_samples, timestamp=1581663372421, value=HCC
 ID5                                    column=tag:Dataset_id, timestamp=1581662965113, value=GSE51401
 ID5                                    column=tag:No_of_samples, timestamp=1581663098116, value=64
 ID5                                    column=tag:Profiling_Technique, timestamp=1581662752293, value=Affymetrix Array
 ID5                                    column=tag:Pubmed_ID, timestamp=1581663874491, value=
 ID5                                    column=tag:Type_of_samples, timestamp=1581663396120, value=Primary uncultured CD31+ and CD105+ tumor
                                        ls (NEC) ,remnant cells from tumor (TC) and non-tumor liver tissue (NTC)of HCC
 ID6                                    column=tag:Dataset_id, timestamp=1581664116962, value=GSE73126
 ID6                                    column=tag:Profiling_Technique, timestamp=1581664087423, value=Affymetrix Array
 ID6                                    column=tag:Pubmed_ID, timestamp=1581664261113, value=35678123
 ID6                                    column=tag:Type_of_samples, timestamp=1581664162934, value=HCC
```

Link for docker hub :

https://hub.docker.com/repository/docker/paudara/14_mt19126_mt19073_a1