

Name: Rahul Kumar
Batch: B1

USN:1BM17CS070
Section: 7-B

BDA LAB EXPERIMENTS REPORT

Lab1 - 08-10-20

Question: Perform the following DB operations using MongoDB.

1. Create a database "Student" with the following attributes Rollno, Age, ContactNo, Email-Id.
2. Insert appropriate values
3. Write a query to update Email-Id of a student with rollno 10.
4. Replace the student name from "ABC" to "FEM" of rollno 11.
5. Export the created table into local file system
6. Drop the table
7. Import a given csv dataset from the local file system into mongodb collection.

Queries:

```
> use Student
switched to db Student
> db.createCollection("Students")
{ "ok" : 1 }
>
db.Students.insert({_id:1,rollno:1,name:"Aakash",age:21,email:"aakash@gmail.com",contact:"9778901232"});
WriteResult({ "nInserted" : 1 })
>
db.Students.insert({_id:2,rollno:5,name:"Aditya",age:22,email:"aditya@gmail.com",contact:"9878310255"});
WriteResult({ "nInserted" : 1 })
>
db.Students.insert({_id:3,rollno:10,name:"Hritik",age:21,email:"hritik@yahoo.com",contact:"8612993321"});
WriteResult({ "nInserted" : 1 })
>
db.Students.insert({_id:4,rollno:11,name:"Rahul",age:22,email:"rahul@gmail.com",contact:"7088910567"});
WriteResult({ "nInserted" : 1 })
```

```

>
db.Students.insert({_id:5,rollno:15,name:"Shivangi",age:22,email:"shivangi@gmail.com",contact
:"8901467672"});
WriteResult({ "nInserted" : 1 })

> db.Students.find().pretty()

> db.Students.update({rollno:10},{ $set:{email:"hritik.gmail.com"}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })

> db.Students.find({rollno :10})
{ "_id" : 3, "rollno" : 10, "name" : "Hritik", "age" : 21, "email" : "hritik.gmail.com", "contact" :
"8612993321" }

> db.Students.update({rollno:11},{ $set:{name:"Rahul Kumar"}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })

> db.Students.find({rollno :11})
{ "_id" : 4, "rollno" : 11, "name" : "Rahul Kumar", "age" : 22, "email" : "rahul@gmail.com",
"contact" : "7088910567" }

```

TO EXPORT THE STUDENTS DATABASE INTO LOCAL CSV FILE :

```

C:\Program Files\MongoDB\Server\4.4\bin> mongoexport -d Student -c Students -f
rollno,name,age,email,contact --csv -o c:\bda\output.csv
2020-10-08T15:09:17.246+0530   csv flag is deprecated; please use --type=csv instead
2020-10-08T15:09:17.864+0530   connected to: mongod://localhost/
2020-10-08T15:09:17.876+0530   exported 5 records

```

TO IMPORT INTO STUDENTS DATABASE FROM THE LOCAL CSV FILE:

```

C:\Program Files\MongoDB\Server\4.4\bin>mongoimport -d Student -c Students --type csv --
headerline --file C:\bda\input.csv
2020-10-08T15:25:01.211+0530   connected to: mongod://localhost/
2020-10-08T15:25:01.216+0530   1 document(s) imported successfully. 0 document(s) failed to
import.

```

Output Screenshots:

```

> use Student
switched to db Student
> db.createCollection("Students")
{ "ok" : 1 }
> db.Students.insert({_id:1,rollno:1,name:"Aakash",age:21,email:"aakash@gmail.com",contact:"9778901232"});
WriteResult({ "nInserted" : 1 })
> db.Students.insert({_id:2,rollno:5,name:"Aditya",age:22,email:"aditya@gmail.com",contact:"9878310255"});
WriteResult({ "nInserted" : 1 })
> db.Students.insert({_id:3,rollno:10,name:"Hritik",age:21,email:"hritik@yahoo.com",contact:"8612993321"});
WriteResult({ "nInserted" : 1 })
> db.Students.insert({_id:4,rollno:11,name:"Rahul",age:22,email:"rahul@gmail.com",contact:"7088910567"});
WriteResult({ "nInserted" : 1 })
> db.Students.insert({_id:5,rollno:15,name:"Shivangi",age:22,email:"shivangi@gmail.com",contact:"8901467672"});
WriteResult({ "nInserted" : 1 })
> db.Students.find().pretty()

```

```

> db.Students.find().pretty()
{
  "_id" : 1,
  "rollno" : 1,
  "name" : "Aakash",
  "age" : 21,
  "email" : "aakash@gmail.com",
  "contact" : "9778901232"
}
{
  "_id" : 2,
  "rollno" : 5,
  "name" : "Aditya",
  "age" : 22,
  "email" : "aditya@gmail.com",
  "contact" : "9878310255"
}
{
  "_id" : 3,
  "rollno" : 10,
  "name" : "Hritik",
  "age" : 21,
  "email" : "hritik@yahoo.com",
  "contact" : "8612993321"
}
{
  "_id" : 4,
  "rollno" : 11,
  "name" : "Rahul",
  "age" : 22,
  "email" : "rahul@gmail.com",
  "contact" : "7088910567"
}
{
  "_id" : 5,
  "rollno" : 15,
  "name" : "Shivangi",
  "age" : 22,
  "email" : "shivangi@gmail.com",
  "contact" : "8901467672"
}

```

```

> db.Students.update({rollno:10},{set:{email:"hritik.gmail.com"}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
> db.Students.find().pretty()
{
  "_id" : 1,
  "rollno" : 1,
  "name" : "Aakash",
  "age" : 21,
  "email" : "aakash@gmail.com",
  "contact" : "9778901232"
}
{
  "_id" : 2,
  "rollno" : 5,
  "name" : "Aditya",
  "age" : 22,
  "email" : "aditya@gmail.com",
  "contact" : "9878310255"
}
{
  "_id" : 3,
  "rollno" : 10,
  "name" : "Hritik",
  "age" : 21,
  "email" : "hritik.gmail.com",
  "contact" : "8612993321"
}
{
  "_id" : 4,
  "rollno" : 11,
  "name" : "Rahul",
  "age" : 22,
  "email" : "rahul@gmail.com",
  "contact" : "7088910567"
}
{
  "_id" : 5,
  "rollno" : 15,
  "name" : "Shivangi",
  "age" : 22,
  "email" : "shivangi@gmail.com",
  "contact" : "8901467672"
}

```

```

> db.Students.find({rollno :10})
{ "_id" : 3, "rollno" : 10, "name" : "Hritik", "age" : 21, "email" : "hritik.gmail.com", "contact" : "8612993321" }
> db.Students.update({rollno:11},{set:{name:"Rahul Kumar"}})
WriteResult({ "nMatched" : 1, "nUpserted" : 0, "nModified" : 1 })
> db.Students.find({rollno :11})
{ "_id" : 4, "rollno" : 11, "name" : "Rahul Kumar", "age" : 22, "email" : "rahul@gmail.com", "contact" : "7088910567" }
>

```

```
C:\Program Files\MongoDB\Server\4.4\bin> mongoexport -d Student -c Students -f rollno,name,age,email,contact --csv -o c:\bda\output.csv
2020-10-08T15:09:17.246+0530 csv flag is deprecated; please use --type=csv instead
2020-10-08T15:09:17.864+0530 connected to: mongodb://localhost/
2020-10-08T15:09:17.876+0530 exported 5 records
```

```
C:\Program Files\MongoDB\Server\4.4\bin>mongoimport -d Student -c Students --type csv --headerline --file C:\bda\input.csv
2020-10-08T15:34:17.594+0530 connected to: mongodb://localhost/
2020-10-08T15:34:17.598+0530 2 document(s) imported successfully. 0 document(s) failed to import.
```

```
{
  "_id" : 3,
  "rollno" : 10,
  "name" : "Hritik",
  "age" : 21,
  "email" : "hritik.gmail.com",
  "contact" : "8612993321"
}
{
  "_id" : 4,
  "rollno" : 11,
  "name" : "Rahul Kumar",
  "age" : 22,
  "email" : "rahul@gmail.com",
  "contact" : "7088910567"
}
{
  "_id" : 5,
  "rollno" : 15,
  "name" : "Shivangi",
  "age" : 22,
  "email" : "shivangi@gmail.com",
  "contact" : "8901467672"
}
{
  "_id" : ObjectId("5f7ee421a8643fdc2f751bdf"),
  "rollno" : 20,
  "name" : "Tanmay",
  "age" : 22,
  "email" : "tanmay@gmail.com",
  "contact" : NumberLong("9778145670")
}
{
  "_id" : ObjectId("5f7ee421a8643fdc2f751be0"),
  "rollno" : 21,
  "name" : "Zaheer",
  "age" : 22,
  "email" : "zaheer@gmail.com",
  "contact" : NumberLong("8780345671")
}
>
```

	A	B	C	D	E	F	G
1	rollno	name	age	email	contact		
2	1	Aakash	21	aakash@g	9.78E+09		
3	5	Aditya	22	aditya@gr	9.88E+09		
4	10	Hritik	21	hritik.gmai	8.61E+09		
5	11	Rahul Kum	22	rahul@gm	7.09E+09		
6	15	Shivangi	22	shivangi@	8.9E+09		
7							

Lab2 - 08-10-20

Question: Perform the following DB operations using MongoDB.

1. Create a collection by name Customers with the following attributes.
Cust_id, Acc_Bal, Acc_Type
2. Insert at least 5 values into the table
3. Write a query to display those records whose total account balance is greater than 1200 of account type 'Z' for each customer_id.
4. Determine Minimum and Maximum account balance for each customer_id.
5. Export the created collection into local file system
6. Drop the table
7. Import a given csv dataset from the local file system into mongodb collection.

Queries:

```

use Customer
switched to db Customer
> db.createCollection("Customers");
{ "ok" : 1 }
> db.Customers.insert({_id:1,cust_id:1001,acc_bal:25000.0,acc_type:"Savings"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:2,cust_id:1001,acc_bal:10000.0,acc_type:"Current"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:3,cust_id:1002,acc_bal:1000.0,acc_type:"Current"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:4,cust_id:1002,acc_bal:2000.0,acc_type:"Savings"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:5,cust_id:1003,acc_bal:200.0,acc_type:"Current"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:6,cust_id:1003,acc_bal:700.0,acc_type:"Savings"});
WriteResult({ "nInserted" : 1 })

```

CUSTOMERS WHOSE TOTAL BALANCE IS GREATER THAN 1200 IN CURRENT ACCOUNT:

```
> db.Customers.aggregate({$match : {acc_type:"Current"}}, {$group : {_id:"$cust_id", totalBaln:
{$sum :"$acc_bal"}}, {$match : {totalBaln : {$gt : 1200}}})
{ "_id" : 1001, "totalBaln" : 10000 }
```

CUSTOMERS WHOSE TOTAL BALANCE IS GREATER THAN 1200 IN SAVINGS ACCOUNT:

```
> db.Customers.aggregate({$match : {acc_type:"Savings"}}, {$group : {_id:"$cust_id", totalBaln:
{$sum :"$acc_bal"}}, {$match : {totalBaln : {$gt : 1200}}})
{ "_id" : 1002, "totalBaln" : 2000 }
{ "_id" : 1001, "totalBaln" : 25000 }
```

MAX BALANCE FOR ALL CUSTOMERS ACROSS ANY ACCOUNT TYPE:

```
> db.Customers.aggregate({$group : {_id:"$cust_id", maxBaln: {$max : "$acc_bal"}}});
{ "_id" : 1002, "maxBaln" : 2000 }
{ "_id" : 1001, "maxBaln" : 25000 }
{ "_id" : 1003, "maxBaln" : 700 }
```

MIN BALANCE FOR ALL CUSTOMERS ACROSS ANY ACCOUNT TYPE:

```
> db.Customers.aggregate({$group : {_id:"$cust_id", minBaln: {$min : "$acc_bal"}}});
{ "_id" : 1002, "minBaln" : 1000 }
{ "_id" : 1003, "minBaln" : 200 }
{ "_id" : 1001, "minBaln" : 10000 }
```

COMMAND TO EXPORT CUSTOMERS DATABASE TO LOCAL CSV FILE:

```
C:\Program Files\MongoDB\Server\4.4\bin> mongoexport -d Customer -c Customers -f
cust_id,acc_bal,acc_type --csv -o c:\bda\C_output.csv
2020-10-08T16:09:58.955+0530 csv flag is deprecated; please use --type=csv instead
2020-10-08T16:09:59.586+0530 connected to: mongodb://localhost/
2020-10-08T16:09:59.598+0530 exported 6 records
```

```
> db.Customers.drop() //COMMAND TO DROP A TABLE
true //TABLE DROPPED
> db.Customers.find() //TABLE EMPTY
```

COMMAND TO IMPORT LOCAL CSV FILE DATA INTO CUSTOMERS DATABASE:

```
C:\Program Files\MongoDB\Server\4.4\bin>mongoimport -d Customer -c Customers --type csv -
-headerline --file C:\bda\C_output.csv
2020-10-08T16:11:48.847+0530   connected to: mongodb://localhost/
2020-10-08T16:11:48.878+0530   6 document(s) imported successfully. 0 document(s) failed to
import.
```

AFTER IMPORT, FINAL OUTPUT:

```
> db.Customers.find()
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2a"), "cust_id" : 1001, "acc_bal" : 10000, "acc_type"
: "Current" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2b"), "cust_id" : 1002, "acc_bal" : 1000, "acc_type" :
"Current" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2c"), "cust_id" : 1002, "acc_bal" : 2000, "acc_type" :
"Savings" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2d"), "cust_id" : 1003, "acc_bal" : 200, "acc_type" :
"Current" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2e"), "cust_id" : 1003, "acc_bal" : 700, "acc_type" :
"Savings" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2f"), "cust_id" : 1001, "acc_bal" : 25000, "acc_type" :
"Savings" }
>
```

Output Screenshots:

```
> use Customer
switched to db Customer
> db.createCollection("Customers");
{ "ok" : 1 }
> db.Customers.insert({_id:1,cust_id:1001,acc_bal=25000.0,acc_type="Savings"});
uncaught exception: SyntaxError: missing : after property id :
@(shell):1:47
> db.Customers.insert({_id:1,cust_id:1001,acc_bal:25000.0,acc_type:"Savings"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:2,cust_id:1001,acc_bal:10000.0,acc_type:"Current"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:3,cust_id:1002,acc_bal:1000.0,acc_type:"Current"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:4,cust_id:1002,acc_bal:2000.0,acc_type:"Savings"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:5,cust_id:1003,acc_bal:200.0,acc_type:"Current"});
WriteResult({ "nInserted" : 1 })
> db.Customers.insert({_id:6,cust_id:1003,acc_bal:700.0,acc_type:"Savings"});
WriteResult({ "nInserted" : 1 })
```



```

> db.Customers.aggregate({$match : {acc_type:"Current"}}, {$group : {_id:"$cust_id", totalBaln: {$sum : "$acc_bal"}}, {$match : {totalBaln : {$gt : 1200}}})
{ "_id" : 1001, "totalBaln" : 10000 }
> db.Customers.aggregate({$match : {acc_type:"Savings"}}, {$group : {_id:"$cust_id", totalBaln: {$sum : "$acc_bal"}}, {$match : {totalBaln : {$gt : 1200}}})
{ "_id" : 1002, "totalBaln" : 2000 }
{ "_id" : 1001, "totalBaln" : 25000 }
> db.Customers.aggregate({$group : {_id:"$cust_id", maxBaln: {$max : "$acc_bal"}}});
{ "_id" : 1002, "maxBaln" : 2000 }
{ "_id" : 1001, "maxBaln" : 25000 }
{ "_id" : 1003, "maxBaln" : 700 }
> db.Customers.aggregate({$group : {_id:"$cust_id", minBaln: {$min : "$acc_bal"}}});
{ "_id" : 1002, "minBaln" : 1000 }
{ "_id" : 1003, "minBaln" : 200 }
{ "_id" : 1001, "minBaln" : 10000 }

```

```

C:\Program Files\MongoDB\Server\4.4\bin> mongoexport -d Customer -c Customers -f cust_id,acc_bal,acc_type --csv -o c:\bda\C_output.csv
2020-10-08T16:09:58.955+0530   csv flag is deprecated; please use --type=csv instead
2020-10-08T16:09:59.586+0530   connected to: mongodb://localhost/
2020-10-08T16:09:59.598+0530   exported 6 records

```

```

C:\Program Files\MongoDB\Server\4.4\bin> mongoimport -d Customer -c Customers --type csv --headerline --file C:\bda\C_output.csv
2020-10-08T16:11:48.847+0530   connected to: mongodb://localhost/
2020-10-08T16:11:48.878+0530   6 document(s) imported successfully. 0 document(s) failed to import.

```

```

> db.Customers.drop()
true
> db.Customers.find()
> db.Customers.find()
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2a"), "cust_id" : 1001, "acc_bal" : 10000, "acc_type" : "Current" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2b"), "cust_id" : 1002, "acc_bal" : 1000, "acc_type" : "Current" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2c"), "cust_id" : 1002, "acc_bal" : 2000, "acc_type" : "Savings" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2d"), "cust_id" : 1003, "acc_bal" : 200, "acc_type" : "Current" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2e"), "cust_id" : 1003, "acc_bal" : 700, "acc_type" : "Savings" }
{ "_id" : ObjectId("5f7eececf91c9ca3e5bda2f"), "cust_id" : 1001, "acc_bal" : 25000, "acc_type" : "Savings" }
>

```

	A	B	C	D	
1	cust_id	acc_bal	acc_type		
2	1001	25000	Savings		
3	1001	10000	Current		
4	1002	1000	Current		
5	1002	2000	Savings		
6	1003	200	Current		
7	1003	700	Savings		
8					
9					
10					
11					

Lab3 - 05-11-20

Question: Perform the following DB operations using Cassandra:

1. Create a keyspace by name Employee
2. Create a column family by name Employee-Info with attributes
Emp_Id Primary Key, Emp_Name, Designation, Date_of_Joining, Salary, Dept_Name
3. Insert the values into the table in batch
3. Update Employee name and Department of Emp-Id 121
4. Sort the details of Employee records based on salary
5. Alter the schema of the table Employee_Info to add a column Projects which stores a set of Projects done by the corresponding Employee.
6. Update the altered table to add project names.
7. Create a TTL of 15 seconds to display the values of Employees.

Queries:

```
> CREATE KEYSPACE Employee with replication = {'class': 'SimpleStrategy', 'replication_factor': 1};
```

```
> use employee ;
```

```
> CREATE COLUMNFAMILY Employee_Info(
... Emp_Id int PRIMARY KEY ,
... Emp_name text,
... Designation text,
... DOJ text,
... Salary float,
... Dept_Name text);
```

> describe employee_info

```
> BEGIN BATCH INSERT INTO employee_info(emp_id,emp_name, designation , doj, salary ,
dept_name ) values (100,'albert','clerk','08-11-1998',75000,'marketing'); INSERT INTO
employee_info(emp_id, emp_name , designation , doj, salary , dept_name ) values
(121,'einstein','clerk','18-12-1998',7500,'sales'); INSERT INTO employee_info(emp_id,
emp_name , designation , doj, salary , dept_name ) values (234,'maxwell','manager','08-11-
1945',65000,'tech'); INSERT INTO employee_info(emp_id, emp_name , designation , doj, salary
, dept_name ) values (170,'albert','manager','12-1-1988',10000,'sales'); APPLY BATCH ;
cqlsh:employee> SELECT * from employee_info;
```

```
cqlsh:employee> update employee_info set emp_name = 'shivangi', dept_name = 'deployment'
where emp_id = 121;
```

```
cqlsh:employee> SELECT * from employee_info;
```

```
cqlsh:employee> alter table employee_info add projects set<text>;
```

```
cqlsh:employee> SELECT * from employee_info;
```

```
cqlsh:employee> update employee_info set projects = {'ML','Algo'} WHERE emp_id = 234;
```

```
cqlsh:employee> SELECT * from employee_info;
```

```
> INSERT INTO employee_info(emp_id, emp_name , designation , doj, salary , dept_name )
values (371,'Bharath','manager','12-10-2015',18000,'tech') USING TTL 15;
```

```
cqlsh:employee> select ttl(emp_name) from employee_info where emp_id = 371;
```

Output Screenshots:

```
Select Command Prompt - cqlsh
cqlsh> create keyspace Employee with replication={'class':'SimpleStrategy','replication_factor':1};
cqlsh> use Employee
...;
cqlsh:employee> create columnfamily Employee_Info(Emp_Id int PRIMARY KEY,Emp_name text,Designation text,DOJ text,Salary float,Dept_Name text);
cqlsh:employee> describe employee_info;

CREATE TABLE employee.employee_info (
  emp_id int PRIMARY KEY,
  dept_name text,
  designation text,
  doj text,
  emp_name text,
  salary float
) WITH bloom_filter_fp_chance = 0.01
AND caching = {'keys': 'ALL', 'rows_per_partition': 'NONE'}
AND comment = ''
AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max_threshold': '32', 'min_threshold': '4'}
AND compression = {'chunk_length_in_kb': '64', 'class': 'org.apache.cassandra.io.compress.LZ4Compressor'}
AND crc_check_chance = 1.0
AND dclocal_read_repair_chance = 0.1
AND default_time_to_live = 0
AND gc_grace_seconds = 864000
AND max_index_interval = 2048
AND memtable_flush_period_in_ms = 0
AND min_index_interval = 128
AND read_repair_chance = 0.0
AND speculative_retry = '99PERCENTILE';

cqlsh:employee> BEGIN BATCH INSERT INTO employee_info(emp_id,emp_name, designation , doj, salary , dept_name ) values (100,'hritik','clerk','08-11-1998',75000,'marketing')
; INSERT INTO employee_info(emp_id, emp_name , designation , doj, salary , dept_name ) values (121,'prateek','clerk','18-12-1997',7500,'sales'); INSERT INTO employee_info(e
mp_id, emp_name , designation , doj, salary , dept_name ) values (234,'rahul','manager','08-11-1998',65000,'tech'); INSERT INTO employee_info(emp_id, emp_name , designation
, doj, salary , dept_name ) values (170,'shivangi','manager','12-1-1988',70000,'sales'); APPLY BATCH ;
cqlsh:employee> select * from employee_info;

emp_id | dept_name | designation | doj          | emp_name | salary
-----+-----+-----+-----+-----+-----
234 | tech | manager | 08-11-1998 | rahul | 65000
121 | sales | clerk | 18-12-1997 | prateek | 7500
100 | marketing | clerk | 08-11-1998 | hritik | 75000
170 | sales | manager | 12-1-1988 | shivangi | 70000

(4 rows)
```

```
cqlsh:employee> update employee_info set emp_name = 'shivangi', dept_name = 'deployment' where emp_id = 121;
cqlsh:employee> select * from employee_info;

emp_id | dept_name | designation | doj          | emp_name | salary
-----+-----+-----+-----+-----+-----
234 | tech | manager | 08-11-1998 | rahul | 65000
121 | deployment | clerk | 18-12-1997 | shivangi | 7500
100 | marketing | clerk | 08-11-1998 | hritik | 75000
170 | sales | manager | 12-1-1988 | shivangi | 70000

(4 rows)
cqlsh:employee> alter table employee_info add projects set<text>;
cqlsh:employee> select * from employee_info;

emp_id | dept_name | designation | doj          | emp_name | projects | salary
-----+-----+-----+-----+-----+-----+-----
234 | tech | manager | 08-11-1998 | rahul | null | 65000
121 | deployment | clerk | 18-12-1997 | shivangi | null | 7500
100 | marketing | clerk | 08-11-1998 | hritik | null | 75000
170 | sales | manager | 12-1-1988 | shivangi | null | 70000

(4 rows)
cqlsh:employee> update employee_info set projects = {'ML','Algo'} WHERE emp_id = 234;
cqlsh:employee> select * from employee_info;

emp_id | dept_name | designation | doj          | emp_name | projects | salary
-----+-----+-----+-----+-----+-----+-----
234 | tech | manager | 08-11-1998 | rahul | {'Algo', 'ML'} | 65000
121 | deployment | clerk | 18-12-1997 | shivangi | null | 7500
100 | marketing | clerk | 08-11-1998 | hritik | null | 75000
170 | sales | manager | 12-1-1988 | shivangi | null | 70000

(4 rows)
```

```

cqlsh:employee> INSERT INTO employee_info(emp_id, emp_name , designation , doj, salary , dept_name ) values (371,'Bharath','manager','12-10-2015',18000,'tech') USING TTL 15
;
cqlsh:employee> select * from employee_info;

emp_id | dept_name | designation | doj          | emp_name | projects | salary
-----+-----+-----+-----+-----+-----+-----
234 | tech | manager | 08-11-1998 | rahul | {'Algo', 'ML'} | 65000
121 | deployment | clerk | 18-12-1997 | shivangi | null | 7500
100 | marketing | clerk | 08-11-1998 | hritik | null | 75000
170 | sales | manager | 12-1-1988 | shivangi | null | 70000
371 | tech | manager | 12-10-2015 | Bharath | null | 18000

(5 rows)
cqlsh:employee> select ttl(emp_name) from employee_info where emp_id = 371;

ttl(emp_name)
-----
3

(1 rows)
cqlsh:employee> select ttl(emp_name) from employee_info where emp_id = 371;

ttl(emp_name)
-----

(0 rows)
cqlsh:employee>

```

Lab4 - 05-11-20

Question: Perform the following DB operations using Cassandra.

1. Create a keyspace by name Library
2. Create a column family by name Library-Info with attributes
Stud_Id Primary Key,
Counter_value of type Counter,
Stud_Name, Book-Name, Book-Id, Date_of_issue
3. Insert the values into the table in batch
3. Display the details of the table created and increase the value of the counter
4. Write a query to show that a student with id 112 has taken a book "BDA" 2 times.
5. Export the created column to a csv file
6. Import a given csv dataset from local file system into Cassandra column family

Queries:

```
cqlsh> create keyspace Library with replication = { 'class':'SimpleStrategy', 'replication_factor':1 };
```

```
cqlsh> use Library;
```

```
cqlsh:library> create columnfamily Library_info (stud_id int, counter_value counter, stud_name
varchar, book_name varchar, book_id int, DOI varchar, PRIMARY KEY(stud_id, stud_name,
book_name, book_id, DOI));
```

```
cqlsh:library> update Library_info set counter_value = counter_value + 1 where stud_id = 110
AND stud_name = 'Ram' AND book_name = 'BDA' AND book_id = 1000 AND DOI = '2017-10-
15';
```

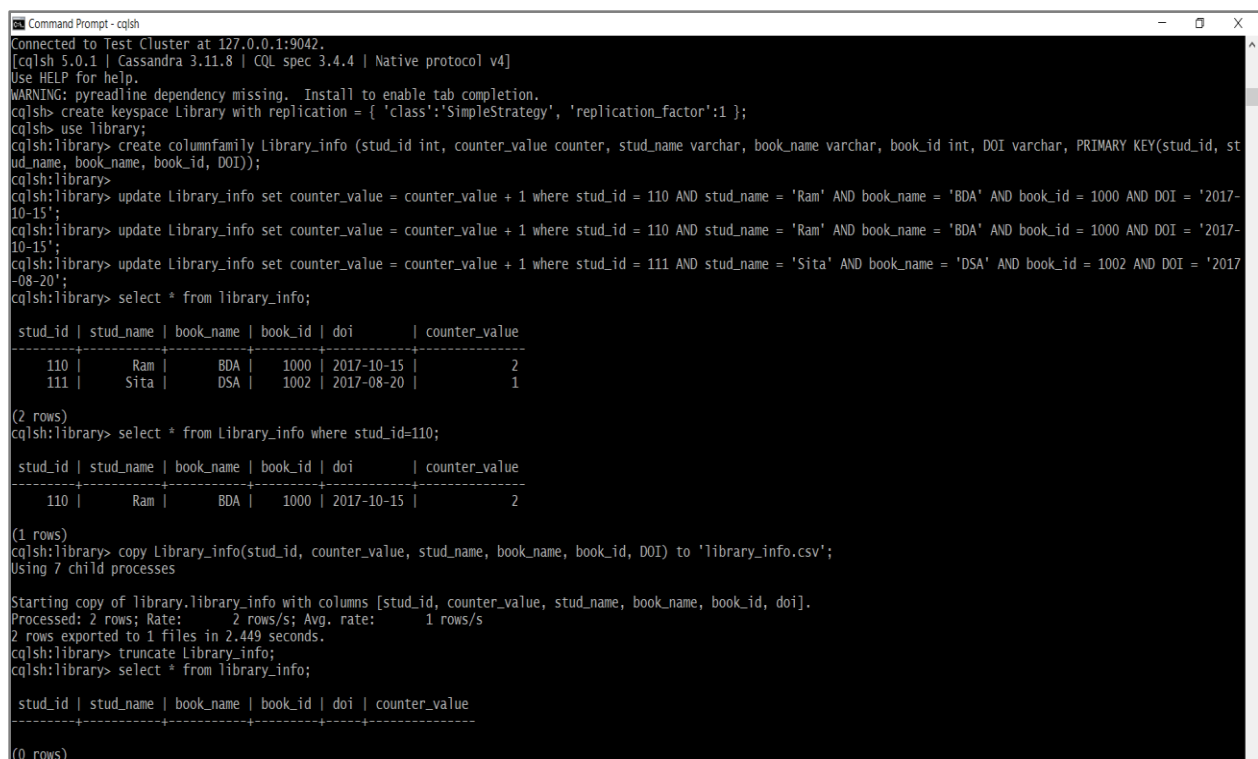
```
cqlsh:library> update Library_info set counter_value = counter_value + 1 where stud_id = 110
AND stud_name = 'Ram' AND book_name = 'BDA' AND book_id = 1000 AND DOI = '2017-10-
15';
```

```
cqlsh:library> update Library_info set counter_value = counter_value + 1 where stud_id = 111
AND stud_name = 'Sita' AND book_name = 'DSA' AND book_id = 1002 AND DOI = '2017-08-20';
```

```
cqlsh:library> select * from Library_info;
cqlsh:library> select * from Library_info where stud_id=110;
cqlsh:library> copy Library_info(stud_id, counter_value, stud_name, book_name, book_id, DOI)
to 'library_info.csv';
```

```
cqlsh:library> truncate Library_info;
cqlsh:library> select * from Library_info;
cqlsh:library> copy Library_info(stud_id, counter_value, stud_name, book_name, book_id, DOI)
from 'library_info.csv';
cqlsh:library> select * from Library_info;
```

Output Screenshots:



```
Command Prompt - cqlsh
Connected to Test Cluster at 127.0.0.1:9042.
[cqlsh 5.0.1 | Cassandra 3.11.8 | CQL spec 3.4.4 | Native protocol v4]
Use HELP for help.
WARNING: pyreadline dependency missing. Install to enable tab completion.
cqlsh> create keyspace Library with replication = { 'class':'SimpleStrategy', 'replication_factor':1 };
cqlsh> use library;
cqlsh:library> create columnfamily Library_info (stud_id int, counter_value counter, stud_name varchar, book_name varchar, book_id int, DOI varchar, PRIMARY KEY(stud_id, stud_name, book_name, book_id, DOI));
cqlsh:library>
cqlsh:library> update Library_info set counter_value = counter_value + 1 where stud_id = 110 AND stud_name = 'Ram' AND book_name = 'BDA' AND book_id = 1000 AND DOI = '2017-10-15';
cqlsh:library> update Library_info set counter_value = counter_value + 1 where stud_id = 110 AND stud_name = 'Ram' AND book_name = 'BDA' AND book_id = 1000 AND DOI = '2017-10-15';
cqlsh:library> update Library_info set counter_value = counter_value + 1 where stud_id = 111 AND stud_name = 'Sita' AND book_name = 'DSA' AND book_id = 1002 AND DOI = '2017-08-20';
cqlsh:library> select * from library_info;

stud_id | stud_name | book_name | book_id | doi | counter_value
-----+-----+-----+-----+-----+-----
110 | Ram | BDA | 1000 | 2017-10-15 | 2
111 | Sita | DSA | 1002 | 2017-08-20 | 1
(2 rows)
cqlsh:library> select * from Library_info where stud_id=110;

stud_id | stud_name | book_name | book_id | doi | counter_value
-----+-----+-----+-----+-----+-----
110 | Ram | BDA | 1000 | 2017-10-15 | 2
(1 rows)
cqlsh:library> copy Library_info(stud_id, counter_value, stud_name, book_name, book_id, DOI) to 'library_info.csv';
Using 7 child processes

Starting copy of library.Library_info with columns [stud_id, counter_value, stud_name, book_name, book_id, doi].
Processed: 2 rows; Rate: 2 rows/s; Avg. rate: 1 rows/s
2 rows exported to 1 files in 2.449 seconds.
cqlsh:library> truncate Library_info;
cqlsh:library> select * from library_info;

stud_id | stud_name | book_name | book_id | doi | counter_value
-----+-----+-----+-----+-----+-----
(0 rows)
```

```
cqlsh:library> copy Library_info(stud_id, counter_value, stud_name, book_name, book_id, DOI) from 'library_info.csv';
Using 7 child processes
```

```
Starting copy of library.Library_info with columns [stud_id, counter_value, stud_name, book_name, book_id, doi].
```

```
Process ImportProcess-10: 1 rows/s; Avg. rate: 1 rows/s
```

```
Process ImportProcess-8:
```

```
Process ImportProcess-9:
```

```
Process ImportProcess-11:
```

```
ITPPraceback (most recent call last):
```

```
ITraceback (most recent call last):
```

```
Process ImportProcess-12:
```

```
Process ImportProcess-14:
```

```
Process ImportProcess-13:
```

```
raceback (most recent call last):
```

```
raceback (most recent call last):
```

```
T TT raceback (most recent call last):
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
raceback (most recent call last):
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
self.run()
```

```
raceback (most recent call last):
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
self.run()
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
self.run()
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
File "D:\python2.7\lib\multiprocessing\process.py", line 267, in _bootstrap
```

```
self.run()
```

```
self.run()
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
self.run()
```

```
self.close()
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2328, in run
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2332, in close
```

```
self.close()
```

```
self.close()
```

```
self.close()
```

```
File "C:\Program Files\apache-cassandra-3.11.8\bin\..\pylib\cqlshlib\copyutil.py", line 2332, in close
```

```
self.close()
```

```
self._session.cluster.shutdown()
```

```
self.close()
```

```
self.close()
```

```
Processed: 2 rows; Rate: 0 rows/s; Avg. rate: 1 rows/s
```

```
2 rows imported from 1 files in 2.757 seconds (0 skipped).
```

```
cqlsh:library> select * from library_info;
```

stud_id	stud_name	book_name	book_id	doi	counter_value
110	Ram	BDA	1000	2017-10-15	2
111	Sita	DSA	1002	2017-08-20	1

```
(2 rows)
```

```
cqlsh:library>
```

Lab5 - 26-11-20

Question: Develop a MapReduce program to count the number of occurrences of words in a given file.

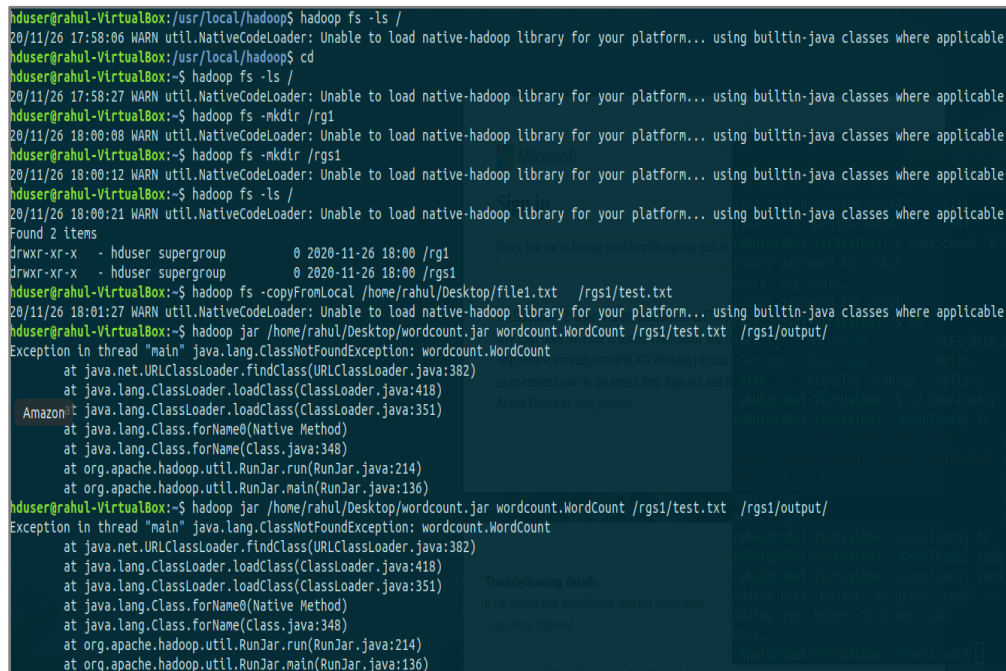
Queries:

BDA-lab 7: Hadoop Word count program

Commands

```
su hduser
cd /usr/local/hadoop/sbin/
./start-dfs.sh
./start-yarn.sh
jps
cd ~
hadoop fs -ls /
hadoop fs -mkdir /rgs1
hadoop fs -copyFromLocal /home/rahul/Documents/lab7/text.txt /rgs1/text.txt
hadoop fs -ls /rgs1
## executing the program ##
hadoop jar /home/rahul/Documents/lab7/wordcount.jar WordCount /rgs1/text.txt /rgs1/output/
hadoop fs -ls /rgs1/output
hadoop fs -cat /rgs1/output/part-r-00000
hadoop fs -copyToLocal /rgs1/output/part-r-00000 $HOME/wordcount-output.txt
./usr/local/hadoop/sbin/stop-all.sh
```

Output Screenshots:



```
hduser@rahul-VirtualBox: /usr/local/hadoop$ hadoop fs -ls /
20/11/26 17:58:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@rahul-VirtualBox: /usr/local/hadoop$ cd
hduser@rahul-VirtualBox:~$ hadoop fs -ls /
20/11/26 17:58:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@rahul-VirtualBox:~$ hadoop fs -mkdir /rgs1
20/11/26 18:00:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@rahul-VirtualBox:~$ hadoop fs -mkdir /rgs1
20/11/26 18:00:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@rahul-VirtualBox:~$ hadoop fs -ls /
20/11/26 18:00:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x  - hduser supergroup          0 2020-11-26 18:00 /rg1
drwxr-xr-x  - hduser supergroup          0 2020-11-26 18:00 /rgs1
hduser@rahul-VirtualBox:~$ hadoop fs -copyFromLocal /home/rahul/Desktop/file1.txt /rgs1/test.txt
20/11/26 18:01:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@rahul-VirtualBox:~$ hadoop jar /home/rahul/Desktop/wordcount.jar wordcount.WordCount /rgs1/test.txt /rgs1/output/
Exception in thread "main" java.lang.ClassNotFoundException: wordcount.WordCount
    at java.net.URLClassLoader.findClass(URLClassLoader.java:382)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:418)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:351)
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:348)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
hduser@rahul-VirtualBox:~$ hadoop jar /home/rahul/Desktop/wordcount.jar wordcount.WordCount /rgs1/test.txt /rgs1/output/
Exception in thread "main" java.lang.ClassNotFoundException: wordcount.WordCount
    at java.net.URLClassLoader.findClass(URLClassLoader.java:382)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:418)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:351)
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:348)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
```



```

hduuser@rahul-VirtualBox:~$ hadoop jar /home/rahul/Desktop/wordcount.jar WordCount /rgs1/test.txt /rgs1/output/
20/11/26 18:14:28 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/11/26 18:14:28 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
20/11/26 18:14:28 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
20/11/26 18:14:28 INFO InputFileInputFormat: Total input paths to process : 1
20/11/26 18:14:28 INFO mapreduce.JobSubmitter: number of splits:1
20/11/26 18:14:29 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local1271569334_0001
20/11/26 18:14:29 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
20/11/26 18:14:29 INFO mapreduce.Job: Running job: job_local1271569334_0001
20/11/26 18:14:29 INFO mapred.LocalJobRunner: OutputCommitter set in config null
20/11/26 18:14:29 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
20/11/26 18:14:29 INFO mapred.LocalJobRunner: Waiting for map tasks
20/11/26 18:14:29 INFO mapred.LocalJobRunner: Starting task: attempt_local1271569334_0001_m_000000_0
20/11/26 18:14:29 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
20/11/26 18:14:29 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/rgs1/test.txt:0+89
20/11/26 18:14:29 INFO mapred.MapTask: (EQUATOR) 0 kvt 26214396(104857584)
20/11/26 18:14:29 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
20/11/26 18:14:29 INFO mapred.MapTask: soft limit at 83886080
20/11/26 18:14:29 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
20/11/26 18:14:29 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
20/11/26 18:14:29 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer
20/11/26 18:14:29 INFO mapred.LocalJobRunner:
20/11/26 18:14:29 INFO mapred.MapTask: Starting flush of map output
20/11/26 18:14:29 INFO mapred.MapTask: Spilling map output
20/11/26 18:14:29 INFO mapred.MapTask: bufstart = 0; bufend = 169; bufvoid = 104857600
20/11/26 18:14:29 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26214320(104857280); length = 77/6553600
20/11/26 18:14:29 INFO mapred.MapTask: Finished spill 0
20/11/26 18:14:29 INFO mapred.Task: Task:attempt_local1271569334_0001_m_000000_0 is done. And is in the process of committing
20/11/26 18:14:29 INFO mapred.LocalJobRunner: map
20/11/26 18:14:29 INFO mapred.Task: Task 'attempt_local1271569334_0001_m_000000_0' done.
20/11/26 18:14:29 INFO mapred.LocalJobRunner: Finishing task: attempt_local1271569334_0001_m_000000_0
20/11/26 18:14:29 INFO mapred.LocalJobRunner: map task executor complete.
20/11/26 18:14:29 INFO mapred.LocalJobRunner: Waiting for reduce tasks
20/11/26 18:14:29 INFO mapred.LocalJobRunner: Starting task: attempt_local1271569334_0001_r_000000_0
20/11/26 18:14:29 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
20/11/26 18:14:29 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@68b63d9e
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=334338464, maxSingleShuffleLimit=83584616, mergeThreshold=228663392, ioSortFactor=10, memToMemMergeOutputsThreshold=10
20/11/26 18:14:29 INFO reduce.EventFetcher: attempt_local1271569334_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
20/11/26 18:14:29 INFO reduce.LocalFetcher: LocalFetcher#1 about to shuffle output of map attempt_local1271569334_0001_m_000000_0 decomp: 111 len: 115 to MEMORY
20/11/26 18:14:29 INFO reduce.InMemoryMapOutput: Read 111 bytes from map-output for attempt_local1271569334_0001_m_000000_0
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 111, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->111
20/11/26 18:14:29 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
20/11/26 18:14:29 INFO mapred.LocalJobRunner: 1 / 1 copied.
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: Merging 1 sorted segments
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: Down to the last merge-pass, with 1 segments left of total size: 105 bytes
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: Merged 1 segments, 111 bytes to disk to satisfy reduce memory limit
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: Merging 1 files, 115 bytes from disk
20/11/26 18:14:29 INFO reduce.MergeManagerImpl: Merged 0 segments, 0 bytes from on-disk map-outputs

```

```

20/11/26 18:14:29 INFO mapred.Task: Task:attempt_local1271569334_0001_r_000000_0 is done. And is in the process of committing
20/11/26 18:14:29 INFO mapred.LocalJobRunner: 1 / 1 copied.
20/11/26 18:14:29 INFO mapred.Task: Task attempt_local1271569334_0001_r_000000_0 is allowed to commit now
20/11/26 18:14:29 INFO output.FileOutputCommitter: Saved output of task 'attempt_local1271569334_0001_r_000000_0' to hdfs://localhost:54310/rgs1/output/_temporary/0/task_local1271569334_0001_r_000000
20/11/26 18:14:29 INFO mapred.LocalJobRunner: reduce > reduce
20/11/26 18:14:29 INFO mapred.Task: Task attempt_local1271569334_0001_r_000000_0 done.
20/11/26 18:14:29 INFO mapred.LocalJobRunner: Finishing task: attempt_local1271569334_0001_r_000000_0
20/11/26 18:14:29 INFO mapred.LocalJobRunner: reduce task executor complete.
20/11/26 18:14:30 INFO mapreduce.Job: Job job_local1271569334_0001 running in uber mode : false
20/11/26 18:14:30 INFO mapreduce.Job: map 100% reduce 100%
20/11/26 18:14:30 INFO mapreduce.Job: Job job_local1271569334_0001 completed successfully
20/11/26 18:14:30 INFO mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=6982
  FILE: Number of bytes written=508711
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=178
  HDFS: Number of bytes written=69
  HDFS: Number of read operations=13
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
Map-Reduce Framework
  Map input records=5
  Map output records=20
  Map output bytes=169
  Map output materialized bytes=115
  Input split bytes=101
  Combine input records=20
  Combine output records=10
  Reduce input groups=10
  Reduce shuffle bytes=115
  Reduce input records=10
  Reduce output records=10
  Spilled Records=20
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=0
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
  Total committed heap usage (bytes)=525336576
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  ID_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=89
File Output Format Counters
  Bytes Written=69

```

```
Combine output records=10
Reduce input groups=10
Reduce shuffle bytes=115
Reduce input records=10
Reduce output records=10
Spilled Records=20
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ns)=0
CPU time spent (ns)=0
Physical memory (bytes) snapshot=0
Virtual memory (bytes) snapshot=0
Total committed heap usage (bytes)=525336576

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=89
File Output Format Counters
Bytes Written=69
hduser@rahul-VirtualBox:~$ hadoop fs -ls /rgs1
20/11/26 18:14:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x  - hduser supergroup          0 2020-11-26 18:14 /rgs1/output
-rw-r--r--  1 hduser supergroup        89 2020-11-26 18:01 /rgs1/test.txt
hduser@rahul-VirtualBox:~$ hadoop fs -cat /output/part-00000
cat: /output/part-00000: No such file or directory
hduser@rahul-VirtualBox:~$ hadoop fs -cat /output/
20/11/26 18:16:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: /output/: No such file or directory
hduser@rahul-VirtualBox:~$ hadoop fs -cat /output/part-r-00000
20/11/26 18:16:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: /output/part-r-00000: No such file or directory
hduser@rahul-VirtualBox:~$ hadoop fs -cat /rgs1/output/part-r-00000
20/11/26 18:17:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
are 1
brother 1
family 1
hl 1
how 5
ts 4
job 1
sister 1
you 1
your 4
```

Lab6 - 10-12-20

Question: For the given file, create a Map Reduce program to find the average temperature for each year from NCDC data set

Queries: ### Commands to execute lab8 - Average Map Reduce program ###

```
su hduser
sudo cp -r /home/rahul/Documents/Average /home/hduser/lab8
sudo cp /home/rahul/Documents/1901-input.txt /home/hduser/lab8
cd /home/hduser/lab8
# create MANIFEST.MF file with contents : Main-Class : AverageDriver
sudo nano MANIFEST.MF
sudo javac -classpath $(hadoop classpath) -d . *.java
sudo jar cvf average.jar *.class
jar -tf average.jar
cd /usr/local/hadoop/sbin/
./start-dfs.sh
./start-yarn.sh
jps
cd ~
hadoop fs -ls /
hadoop fs -mkdir /rgs2
hadoop fs -copyFromLocal /home/hduser/lab8/1901-input.txt /rgs2/1901-input.txt
hadoop fs -ls /rgs2
```

executing the program

hadoop jar /home/hduser/lab8/average.jar AverageDriver /rgs2/1901-input.txt /rgs2/output/

hadoop fs -ls /rgs2/output

hadoop fs -cat /rgs2/output/part-r-00000

hadoop fs -copyToLocal /rgs2/output/part-r-00000 \$HOME/average-output.txt

/usr/local/hadoop/sbin/stop-all.sh

Output Screenshots:

```
hduser@rahul-VirtualBox: ~  
File Edit View Search Terminal Help  
rahul@rahul-VirtualBox:~$ su hduser  
Password:  
hduser@rahul-VirtualBox:/home/rahul$ sudo cp -r /home/rahul/Documents/Average /home/hduser/lab8  
[sudo] password for hduser:  
cp: cannot stat '/home/rahul/Documents/Average': No such file or directory  
hduser@rahul-VirtualBox:/home/rahul$ sudo cp -r /home/rahul/Documents/Average /home/hduser/lab8  
hduser@rahul-VirtualBox:/home/rahul$ sudo cp /home/rahul/Documents/1901-input.txt /home/hduser/lab8
```

```
hduser@rahul-VirtualBox: ~  
File Edit View Search Terminal Help  
hduser@rahul-VirtualBox:~$ sudo su  
root@rahul-VirtualBox:/home/hduser# chmod 777 lab8  
root@rahul-VirtualBox:/home/hduser# pwd  
/home/hduser  
root@rahul-VirtualBox:/home/hduser# ls -la  
total 190752  
drwxrwxrwx 8 hduser hadoop 4096 Dec 20 14:16 .  
drwxr-xr-x 4 root root 4096 Nov 26 15:51 ..  
-rw-r----- 1 hduser hadoop 2247 Nov 26 18:49 .bash_history  
-rwxrwxrwx 1 hduser hadoop 220 Nov 26 15:51 .bash_logout  
-rwxrwxrwx 1 hduser hadoop 4266 Nov 26 17:07 .bashrc  
drwxrwxrwx 2 hduser hadoop 4096 Nov 26 16:41 .cache  
-rwxrwxrwx 1 hduser hadoop 8980 Nov 26 15:51 examples.desktop  
drwxrwxrwx 3 hduser hadoop 4096 Nov 26 16:34 .font  
drwxr-xr-x 2 hduser hadoop 4096 Nov 26 17:02 hadoop-2.6.0  
-rw-r--r-- 1 root root 195257684 Nov 26 16:52 hadoop-2.6.0.tar.gz  
drwxrwxrwx 2 root root 4096 Dec 20 14:25 .local  
drwxr-xr-x 3 hduser hadoop 4096 Nov 26 17:06 .local  
-rwxrwxrwx 1 hduser hadoop 807 Nov 26 15:51 .profile  
drwxrwxrwx 2 hduser hadoop 4096 Nov 26 16:41 .ssh  
-rw-r--r-- 1 hduser hadoop 0 Nov 26 17:00 .sudo_as_admin_successful  
root@rahul-VirtualBox:/home/hduser# exit  
exit  
hduser@rahul-VirtualBox:~$ pwd  
/home/hduser  
hduser@rahul-VirtualBox:~$ cd lab8  
hduser@rahul-VirtualBox:~/lab8$ ls  
1901-input.txt AverageMapper.java MANIFEST.MF  
AverageDriver.java AverageReducer.java  
hduser@rahul-VirtualBox:~/lab8$ sudo javac -classpath $(hadoop classpath) -d . *.java  
Note: AverageDriver.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.  
hduser@rahul-VirtualBox:~/lab8$ sudo jar cvf average.jar *.class  
added manifest  
adding: AverageDriver.class(in = 1393) (out= 808)(deflated 41%)  
adding: AverageMapper.class(in = 1862) (out= 803)(deflated 56%)  
adding: AverageReducer.class(in = 1616) (out= 681)(deflated 57%)  
hduser@rahul-VirtualBox:~/lab8$ jar -tf average.jar  
META-INF/  
META-INF/MANIFEST.MF  
AverageDriver.class  
AverageMapper.class  
AverageReducer.class  
hduser@rahul-VirtualBox:~/lab8$ cd /usr/local/hadoop/sbin/
```

```
hduser@rahul-VirtualBox: ~
File Edit View Search Terminal Help
hduser@rahul-VirtualBox: /usr/local/hadoop/sbin$ ./start-dfs.sh
2012/12/20 14:35:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-rahul-VirtualBox.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-rahul-VirtualBox.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
hduser@0.0.0.0's password: 0.0.0.0: Permission denied, please try again.

0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-rahul-VirtualBox.out
2012/12/20 14:36:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
hduser@rahul-VirtualBox: ~
File Edit View Search Terminal Help
hduser@rahul-VirtualBox: /usr/local/hadoop/sbin$ ./start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-rahul-VirtualBox.out
hduser@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-rahul-VirtualBox.out
hduser@rahul-VirtualBox: /usr/local/hadoop/sbin$ jps
20418 ResourceManager
20770 NodeManager
19363 NameNode
19769 SecondaryNameNode
19545 DataNode
20891 Jps
hduser@rahul-VirtualBox: /usr/local/hadoop/sbin$ cd ~ && echo windows or unix at any time using these
hduser@rahul-VirtualBox: ~$ hadoop fs -ls /
2012/12/20 14:38:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x  - hduser supergroup          0 2020-11-26 18:00 /rg1
drwxr-xr-x  - hduser supergroup          0 2020-11-26 18:14 /rgs1
hduser@rahul-VirtualBox: ~$ hadoop fs -mkdir /rgs2
2012/12/20 14:38:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
Activities Terminal Sun 14:48
hduser@rahul-VirtualBox: ~
File Edit View Search Terminal Help
hduser@rahul-VirtualBox: ~$ hadoop fs -copyFromLocal /home/hduser/lab8/1901-input.txt /rgs2/1901-input.txt
2012/12/20 14:40:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduser@rahul-VirtualBox: ~$ hadoop fs -ls /rgs2
2012/12/20 14:40:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 hduser supergroup 888190 2020-12-20 14:40 /rgs2/1901-input.txt
hduser@rahul-VirtualBox: ~$ hadoop fs -ls /rgs2/1901-input.txt /rgs2/output/
2012/12/20 14:41:04 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2012/12/20 14:41:05 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id
2012/12/20 14:41:05 INFO Jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
2012/12/20 14:41:05 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2012/12/20 14:41:05 INFO Input.FileInputFormat: Total input paths to process : 1
2012/12/20 14:41:05 INFO mapreduce.JobSubmitter: number of splits:1
2012/12/20 14:41:05 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local387256883_0001
2012/12/20 14:41:05 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
2012/12/20 14:41:05 INFO mapreduce.Job: Running job: job_local387256883_0001
2012/12/20 14:41:05 INFO mapred.LocalJobRunner: OutputCommitter set in config null
2012/12/20 14:41:05 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
2012/12/20 14:41:05 INFO mapred.LocalJobRunner: Waiting for map tasks
2012/12/20 14:41:05 INFO mapred.LocalJobRunner: Starting task: attempt_local387256883_0001_m_000000_0
2012/12/20 14:41:05 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2012/12/20 14:41:05 INFO mapred.MapTask: Processing split: hdfs://localhost:54310/rgs2/1901-input.txt:0+888190
2012/12/20 14:41:05 INFO mapred.MapTask: (EQUATOR) 0 kvt 26214396(104857584)
2012/12/20 14:41:05 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
2012/12/20 14:41:05 INFO mapred.MapTask: soft limit at 83886080
2012/12/20 14:41:05 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600
2012/12/20 14:41:05 INFO mapred.MapTask: kvstart = 26214396; length = 6553600
2012/12/20 14:41:05 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTaskMapOutputBuffer
2012/12/20 14:41:06 INFO mapred.LocalJobRunner:
2012/12/20 14:41:06 INFO mapred.MapTask: Starting flush of map output
2012/12/20 14:41:06 INFO mapred.MapTask: Spilling map output
2012/12/20 14:41:06 INFO mapred.MapTask: bufstart = 0; bufend = 59076; bufvoid = 104857600
2012/12/20 14:41:06 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26188144(104752576); length = 26253/6553600
2012/12/20 14:41:06 INFO mapred.MapTask: Finished spill 0
2012/12/20 14:41:06 INFO mapred.Task: Task:attempt_local387256883_0001_m_000000_0 is done. And is in the process of committing
2012/12/20 14:41:06 INFO mapred.LocalJobRunner: map
2012/12/20 14:41:06 INFO mapred.Task: Task 'attempt_local387256883_0001_m_000000_0' done.
2012/12/20 14:41:06 INFO mapred.LocalJobRunner: Finishing task: attempt_local387256883_0001_m_000000_0
2012/12/20 14:41:06 INFO mapred.LocalJobRunner: map task executor complete.
2012/12/20 14:41:06 INFO mapred.LocalJobRunner: Waiting for reduce tasks
2012/12/20 14:41:06 INFO mapred.LocalJobRunner: Starting task: attempt_local387256883_0001_r_000000_0
2012/12/20 14:41:06 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
2012/12/20 14:41:06 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@7ff04a54
2012/12/20 14:41:06 INFO reduce.MergeManagerImpl: MergeManager: memoryLimit=334338464, maxSingleShuffleLimit=83584616, mergeThreshold=226663392, isSortFactor=10, memToMemMergeOutputsThreshold=10
2012/12/20 14:41:06 INFO reduce.EventFetcher: attempt_local387256883_0001_r_000000_0 Thread started: EventFetcher for fetching Map Completion Events
2012/12/20 14:41:06 INFO reduce.LocalFetcher: LocalFetcher#1 about to shuffle output of map attempt_local387256883_0001_m_000000_0 decomp: 72206 len: 72210 to MEMORY
2012/12/20 14:41:06 INFO reduce.InMemoryMapOutput: Read 72206 bytes from map-output for attempt_local387256883_0001_m_000000_0
2012/12/20 14:41:06 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 72206, inMemoryMapOutputs.size() -> 1, commitMemory -> 72206
2012/12/20 14:41:06 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
2012/12/20 14:41:06 INFO mapred.LocalJobRunner: 1 / 1 copied.
2012/12/20 14:41:06 INFO reduce.MergeManagerImpl: FinalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs
2012/12/20 14:41:06 INFO reduce.Merger: Merging 1 sorted segments
2012/12/20 14:41:06 INFO reduce.Merger: Down to the last merge-pass, with 1 segments left of total size: 72199 bytes
2012/12/20 14:41:06 INFO reduce.MergeManagerImpl: Merged 1 segments, 72206 bytes to disk to satisfy reduce memory limit
2012/12/20 14:41:06 INFO reduce.MergeManagerImpl: Merging 1 files, 72210 bytes from disk
```



```
Activities Terminal Sun 14:49 hduiser@rahul-VirtualBox -
File Edit View Search Terminal Help
20/12/20 14:41:06 INFO mapred.LocalJobRunner: Finishing task: attempt_local387256883_0001_r_000000_0
20/12/20 14:41:06 INFO mapred.LocalJobRunner: reduce task executor complete.
20/12/20 14:41:06 INFO mapreduce.Job: Job local387256883_0001 running in uber mode : false
20/12/20 14:41:06 INFO mapreduce.Job: map 100% reduce 100%
20/12/20 14:41:06 INFO mapreduce.Job: Job local387256883_0001 completed successfully
20/12/20 14:41:06 INFO mapreduce.Job: Counters: 38
File System Counters
  FILE: Number of bytes read=1508934
  FILE: Number of bytes written=720510
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1776380
  HDFS: Number of bytes written=8
  HDFS: Number of read operations=13
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=4
Map-Reduce Framework
  Map input records=6565
  Map output records=6564
  Map output bytes=59076
  Map output materialized bytes=72210
  Input split bytes=107
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=72210
  Reduce input records=6564
  Reduce output records=1
  Spilled Records=13128
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=35
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
  Total committed heap usage (bytes)=520093696
Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=888190
File Output Format Counters
  Bytes Written=8
hduiser@rahul-VirtualBox:~$ hadoop fs -ls /rgs2/output
20/12/20 14:41:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r-- 1 hduiser supergroup 0 2020-12-20 14:41 /rgs2/output/_SUCCESS
-rw-r--r-- 1 hduiser supergroup 8 2020-12-20 14:41 /rgs2/output/part-r-000000
hduiser@rahul-VirtualBox:~$ hadoop fs -cat /rgs2/output/part-r-000000
20/12/20 14:41:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1901 46
hduiser@rahul-VirtualBox:~$ hadoop fs -copyToLocal /rgs2/output/part-r-000000 $HOME/average-output.txt
20/12/20 14:41:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hduiser@rahul-VirtualBox:~$ /usr/local/hadoop/sbin/stop-all.sh
This script is deprecated. Instead use stop-dfs.sh and stop-yarn.sh
20/12/20 14:42:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Stopping namenodes on [localhost]
hduiser@localhost's password:
localhost: stopping namenode
hduiser@localhost's password:
localhost: stopping datanode
Stopping secondary namenodes [0.0.0.0]
hduiser@0.0.0.0's password:
0.0.0.0: stopping secondarynamenode
20/12/20 14:42:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
stopping yarn daemons
stopping resourcemanager
hduiser@localhost's password:
localhost: stopping nodemanager
no proxyserver to stop
hduiser@rahul-VirtualBox:~$
```

Lab7 - 17-12-20

Question: Write Queries in Hive to do the following

1. Create an external table named with the following attributes
-> Empl_ID ->Emp_Name -> Designation -> Salary
2. Load data into table from a given file

3. Create a view to Generate a query to retrieve the employee details who earn a salary of more than Rs 30000.
4. Alter the table to add a column Dept_Id and Generate a query to retrieve the employee details in order by using Dept_Id
5. Generate a query to retrieve the number of employees in each department whose salary is greater than 30000
6. Create another table Department with attributes
-> Dept_Id ->Dept_name ->Emp_Id
- 7.Display the cumulative details of each employee along with department details

Queries:

```
hive>create database if not exists Employee;
```

```
hive> use Employee;
```

```
hive> create external table if not exists Employee (Empl_ID int, Emp_Name String, Designation String, Salary int) row format delimited fields terminated by ',' lines terminated by '\n';
```

```
hive> load data local inpath '/home/rahul/Downloads/employee.txt' overwrite into table Employee;
```

```
hive> select * from Employee;
```

```
hive> create view Employee_view as select * from Employee where Salary>30000;
```

```
hive> select * from Employee_view;
```

```
hive> alter table Employee add columns (Dept_ID int);
```

```
hive> load data local inpath '/home/rahul/Downloads/employee_dept.txt' overwrite into table Employee;
```

```
hive> select * from Employee order by Dept_ID;
```

```
hive> select count(*),Dept_ID from Employee where Salary > 30000 group by Dept_ID;
```

```
hive> create table if not exists Department (Dept_ID int , Dept_name String, Emp_ID int) row format delimited fields terminated by ',' lines terminated by '\n';
```

```
hive> load data local inpath '/home/rahul/Downloads/dept.txt' overwrite into table Department;
```

```
hive> select * from Department;
```

```
hive> select e.Empl_ID, e.Emp_Name, e.Designation, e.Salary, e.Dept_ID, d.Dept_Name from  
Employee e join Department d ON (d.Dept_ID = e.Dept_ID);
```
