

ACADEMIC YEAR 2020-2021



KNOWLEDGE • CHARACTER • UNITY

BIGDATA LABORATORY

Report on,

Learning Activity II-Programming Assignment

Submitted by,

Anjani Saurav (1NT18IS030)

Submitted to,

Disha DN,

Assistant Professor,
Department of Information Science and Engineering
Nitte Meenakshi Institute of Technology
Bangalore-064

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

(An autonomous institution with A+ Grade by NAAC /UGC, Affiliated to Visvesvaraya Technological University, Belgaum, Approved by UGC/AICTE/Govt. of Karnataka)

Yelahanka, Bengaluru-560064

Big data lab- Learning activity II

Table of Contents

1. List of Figures.	3
2. Brief about Hadoop and Map Reduce.	4
3. Hadoop Map-Reduce problem statement	5
3.1. Dataset description.	6
3.2. Source Code	6
3.3. Results and snapshots.	7
4. Brief about Hive.	9
5. Hive problems / use cases.	10
5.1. Dataset description.	11
5.2. Use-cases results and snapshots.	12
6. References.	17

Big data lab- Learning activity II

List of Figures

- | | |
|--------------------------|---|
| 1. Map Reduce algorithm | 4 |
| 2. Architecture of Hive. | 9 |

Big data lab- Learning activity II

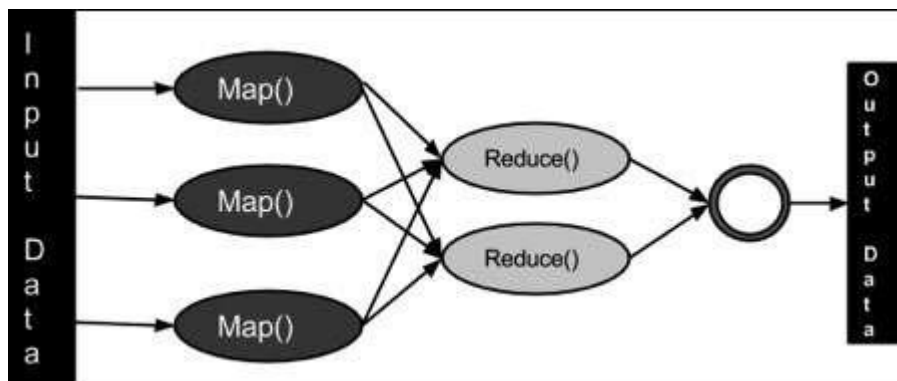
Hadoop and Map Reduce

Hadoop is an open-source framework that allow to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

MapReduce is a processing technique and a program model for distributed computing based on java.

The MapReduce algorithm contains two important:

- 1 Map : It takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs).
- 2 Reduce : It takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.



MapReduce job –

(Input) $\langle k1, v1 \rangle \rightarrow \text{map} \rightarrow \langle k2, v2 \rangle \rightarrow \text{reduce} \rightarrow \langle k3, v3 \rangle$ **(Output).**

The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes.

Big data lab- Learning activity II

Problem Statement

Using the Hadoop Map Reduce framework compute the following :-

1. Total number of employees who are eligible for the pay raise.
2. Total number of cumulative awards the company had this year.
3. How many total awards were obtained by the employee whose salary is 30000 ?
4. Count the number of employees who had paid the Tax.

Big data lab- Learning activity II

Dataset description

Employee Dataset:

name	ssn	salary	awards	tax_paid	eligible_for_pay_raise
ajay	61301	30000	2	yes	yes
aman	61302	40000	1	yes	no
alok	61303	45000	2	yes	yes
bhavya	61304	55000	4	no	no
divya	61305	30000	3	yes	yes
abhishek	61306	43000	2	no	no
subham	61307	25000	1	yes	no
saini	61308	69000	3	yes	yes
sanket	61309	30000	2	no	yes
supriya	61310	35000	1	yes	no
mishra	61311	42000	3	yes	yes
kaushal	61312	37000	2	no	yes
mallya	61313	28000	2	yes	no
sujan	61314	30000	2	yes	yes
mukund	61315	29000	1	no	no
kshitija	61316	54000	0	yes	no
pandey	61317	30000	0	no	no
rajan	61318	55000	1	yes	no
skrikant	61319	26000	1	yes	no
talpade	61320	30000	1	no	no
raji	61321	30000	2	no	yes
saumya	61322	40000	2	no	yes
pallavi	61323	48000	3	yes	yes
shreya	61324	39000	4	no	yes
vaibhav	61325	60000	4	yes	yes

Big data lab- Learning activity II

Results and Snapshots

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~ 190x48

Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=8147
Total time spent by all reduces in occupied slots (ms)=4436
Total time spent by all map tasks (ms)=8147
Total time spent by all reduce tasks (ms)=4436
Total vcore-milliseconds taken by all map tasks=8147
Total vcore-milliseconds taken by all reduce tasks=4436
Total megabyte-milliseconds taken by all map tasks=8342528
Total megabyte-milliseconds taken by all reduce tasks=4542464

Map-Reduce Framework
  Map input records=25
  Map output records=13
  Map output bytes=182
  Map output materialized bytes=44
  Input split bytes=198
  Combine input records=13
  Combine output records=2
  Reduce input groups=1
  Reduce shuffle bytes=44
  Reduce input records=2
  Reduce output records=1
  Spilled Records=4
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=146
  CPU time spent (ms)=1489
  Physical memory (bytes) snapshot=810929056
  Virtual memory (bytes) snapshot=7678844928
  Total committed heap usage (bytes)=724566016
  Peak Map Physical memory (bytes)=291999744
  Peak Map Virtual memory (bytes)=2557898752
  Peak Reduce Physical memory (bytes)=232288256
  Peak Reduce Virtual memory (bytes)=2564149248

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=1053
File Output Format Counters
  Bytes Written=13
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$ hdfs dfs -cat ~/emp1/part*/
13
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~ 189x48

Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=4840
Total time spent by all reduces in occupied slots (ms)=1897
Total time spent by all map tasks (ms)=4840
Total time spent by all reduce tasks (ms)=1897
Total vcore-milliseconds taken by all map tasks=4840
Total vcore-milliseconds taken by all reduce tasks=1897
Total megabyte-milliseconds taken by all map tasks=4956160
Total megabyte-milliseconds taken by all reduce tasks=1942528

Map-Reduce Framework
  Map input records=25
  Map output records=25
  Map output bytes=550
  Map output materialized bytes=60
  Input split bytes=198
  Combine input records=25
  Combine output records=2
  Reduce input groups=1
  Reduce shuffle bytes=60
  Reduce input records=2
  Reduce output records=1
  Spilled Records=4
  Shuffled Maps =2
  Failed Shuffles=0
  Merged Map outputs=2
  GC time elapsed (ms)=234
  CPU time spent (ms)=1640
  Physical memory (bytes) snapshot=765779968
  Virtual memory (bytes) snapshot=7669706752
  Total committed heap usage (bytes)=690351616
  Peak Map Physical memory (bytes)=290217884
  Peak Map Virtual memory (bytes)=2553573376
  Peak Reduce Physical memory (bytes)=189542408
  Peak Reduce Virtual memory (bytes)=2564186112

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=1053
File Output Format Counters
  Bytes Written=21
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$ hdfs dfs -cat ~/emp2/part*/
Cumulative Awards
49
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$
```

Big data lab- Learning activity II

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~ 190x48

Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=5085
Total time spent by all reduces in occupied slots (ms)=1948
Total time spent by all map tasks (ms)=5085
Total time spent by all reduce tasks (ms)=1948
Total vcore-milliseconds taken by all map tasks=5085
Total vcore-milliseconds taken by all reduce tasks=1948
Total megabyte-milliseconds taken by all map tasks=5125120
Total megabyte-milliseconds taken by all reduce tasks=1994752

Map-Reduce Framework
Map input records=25
Map output records=7
Map output bytes=273
Map output materialized bytes=94
Input split bytes=198
Combine input records=7
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=94
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=205
CPU time spent (ms)=1620
Physical memory (bytes) snapshot=812105728
Virtual memory (bytes) snapshot=7669370880
Total committed heap usage (bytes)=716177488
Peak Map Physical memory (bytes)=331747328
Peak Map Virtual memory (bytes)=255281408
Peak Reduce Physical memory (bytes)=187662336
Peak Reduce Virtual memory (bytes)=2563112960

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=1053
File Output Format Counters
Bytes Written=38

hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$ hdfs dfs -cat ~/emp3/part*
Total awards whose salary is 30000 12
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: ~ 190x48

Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=6455
Total time spent by all reduces in occupied slots (ms)=3302
Total time spent by all map tasks (ms)=6455
Total time spent by all reduce tasks (ms)=3302
Total vcore-milliseconds taken by all map tasks=6455
Total vcore-milliseconds taken by all reduce tasks=3302
Total megabyte-milliseconds taken by all map tasks=6609920
Total megabyte-milliseconds taken by all reduce tasks=3381248

Map-Reduce Framework
Map input records=25
Map output records=15
Map output bytes=105
Map output materialized bytes=42
Input split bytes=198
Combine input records=15
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=42
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=141
CPU time spent (ms)=3410
Physical memory (bytes) snapshot=769273856
Virtual memory (bytes) snapshot=7672782848
Total committed heap usage (bytes)=707788800
Peak Map Physical memory (bytes)=282503852
Peak Map Virtual memory (bytes)=2556184704
Peak Reduce Physical memory (bytes)=186814464
Peak Reduce Virtual memory (bytes)=2562281472

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=1053
File Output Format Counters
Bytes Written=12

hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$ hdfs dfs -cat ~/emp4/part*
Paid TAX
15
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:~$
```


Big data lab- Learning activity II

Hive

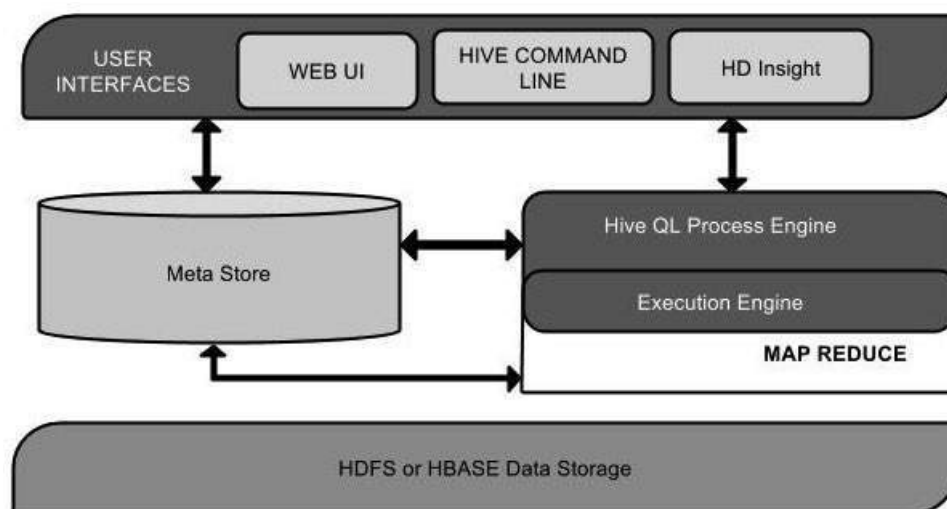
Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy.

Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

Features of Hive :-

- It stores schema in a database and processed data into HDFS.
- It is designed for OLAP.
- It provides SQL type language for querying called HiveQL or HQL.
- It is familiar, fast, scalable and extensible.

Architecture of Hive :-



Big data lab- Learning activity II

Dataset description

Employee Dataset :

name	ssn	salary	awards	tax_paid	eligible_for_pay_raise
ajay	61301	30000	2	yes	yes
aman	61302	40000	1	yes	no
alok	61303	45000	2	yes	yes
bhavya	61304	55000	4	no	no
divya	61305	30000	3	yes	yes
abhishek	61306	43000	2	no	no
subham	61307	25000	1	yes	no
saini	61308	69000	3	yes	yes
sanket	61309	30000	2	no	yes
supriya	61310	35000	1	yes	no
mishra	61311	42000	3	yes	yes
kaushal	61312	37000	2	no	yes
mallya	61313	28000	2	yes	no
sujan	61314	30000	2	yes	yes
mukund	61315	29000	1	no	no
kshitija	61316	54000	0	yes	no
pandey	61317	30000	0	no	no
rajan	61318	55000	1	yes	no
skrikant	61319	26000	1	yes	no
talpade	61320	30000	1	no	no
raji	61321	30000	2	no	yes
saumya	61322	40000	2	no	yes
pallavi	61323	48000	3	yes	yes
shreya	61324	39000	4	no	yes
vaibhav	61325	60000	4	yes	yes

Orders Dataset :

order_id	custssn	amount
61	61302	50000
62	61304	30000
63	61304	10000
64	61302	10000
65	61308	8000
66	61309	50000
67	61309	15000
68	61310	16000
69	61306	17000
70	61301	18000
71	61315	12000
72	61316	30000
73	61317	31000
74	61318	35000
75	61319	2100
76	61320	1000
77	61321	16000
78	61322	20000
79	61323	21000
80	61324	12000
81	61306	12000
82	61307	1100
83	61308	501
84	61309	190
85	61310	11111
86	61311	9999
87	61312	9991
88	61313	2390
89	61314	7891
90	61315	65000

Big data lab- Learning activity II

Problem Statement

Using HiveQL language perform following Queries :-

- 1 Insert 5 records using the INSERT command.
- 2 Demonstrate the Alter command for the following cases,
 - 2.a Rename the table name to “Emp”.
 - 2.b Rename the column name “Eligible for Pay raise” to “Eligibility”.
- 3 Count the number of Employees who are eligible for pay raise who had paid the tax.
- 4 Extract all the users ordered by the Name who had paid the tax but are not eligible for pay raise.
- 5 Create a separate view containing “SSN and Salary” and call the view as sal_ssn_view.
- 6 Display count (eligibility) fields grouped by the SSN.
- 7 Display the (Name, SSN) of employees whose salary is >40000 but < 48000.
- 8 Create Another table called orders with the following fields (custssn = SSN in the Employee) and perform the following joins over custssn .
 - 1)Outer
 - 2.Left outer
 - 3.Right outer.

Big data lab- Learning activity II

Results and Snapshots

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 271x69

hive> INSERT INTO TABLE employee VALUES
> ('anurag',61326,61000,2,'yes','yes'),
> ('prince',61327,60000,5,'no','yes'),
> ('pk',61329,48000,1,'no','yes'),
> ('rajat',61329,33000,0,'no','no'),
> ('chiku',61330,20000,0,'yes','no');
Query ID = hadoop_26210009114207_03ade000-4b0d-4453-851e-edeebd129fd7
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes-per-reducer=number
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=number
In order to set a constant number of reducers:
  set mapreduce.job.reduces=number
Starting Job = job_1623217562077_0001, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:8088/proxy/application_1623217562077_0001/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0001
Hadoop job information for Stage-1: number of mappers: 1, number of reducers: 1
2021-06-09 11:43:16,456 Stage-1 map = 0%, reduce = 0%
2021-06-09 11:43:21,682 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.4 sec
2021-06-09 11:43:26,857 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.1 sec
MapReduce Total cumulative CPU time: 5 seconds 100 msec
Ended Job = job_1623217562077_0001
Job 1 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://localhost:9000/user/hive/warehouse/employedb.db/employee/.hive-staging_hive_2021-06-09_11-43-07_020_6041612324536571935-1/-ext-10000
Loading data to table employedb.employee
MapReduce Job Launched
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.1 sec HDFS Read: 20779 HDFS Write: 811 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 100 msec
OK
hive> select * from employee;
OK
anurag 61326 61000 2 yes yes
prince 61327 60000 5 no yes
pk 61329 48000 1 no yes
rajat 61329 33000 0 no no
chiku 61330 20000 0 yes no
anurag 61301 30000 2 yes yes
prince 61302 40000 2 yes no
pk 61303 45000 2 yes yes
rajat 61304 55000 4 no no
chiku 61305 30000 3 yes yes
anurag 61306 43000 2 no no
prince 61307 25000 1 yes no
pk 61308 69000 3 yes yes
rajat 61309 30000 2 no yes
chiku 61310 35000 1 yes no
anurag 61311 42000 3 yes yes
prince 61312 37000 2 no yes
pk 61313 28000 2 yes no
rajat 61314 30000 2 yes yes
chiku 61315 29000 1 no no
anurag 61316 34000 0 yes no
prince 61317 38000 0 no no
pk 61318 50000 1 yes no
rajat 61319 26000 1 yes no
chiku 61320 30000 1 no no
anurag 61321 30000 2 no yes
prince 61322 40000 2 no yes
pk 61323 48000 3 yes yes
rajat 61324 39000 4 no yes
chiku 61325 60000 4 yes yes
Time taken: 0.119 seconds, Fetched: 30 row(s)
hive>
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 211x53

hive> ALTER TABLE employee RENAME TO emp;
OK
Time taken: 0.679 seconds
hive> show tables;
OK
emp
Time taken: 0.638 seconds, Fetched: 1 row(s)
hive> ALTER TABLE emp CHANGE eligible_for_pay_raise eligibility string;
OK
Time taken: 0.661 seconds
hive> desc emp;
OK
name string
ssn int
salary int
rewards int
tax paid string
eligibility string
Time taken: 0.633 seconds, Fetched: 6 row(s)
hive>
```

Big data lab- Learning activity II

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 211x53

hive> SELECT COUNT(*) FROM emp WHERE tax_paid='yes' AND eligibility='yes';
Query ID = hadoop_20210609115516_a284d3ae-45a9-45c8-9297-67881c9c1211
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623217562077_0002, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:8088/proxy/application_1623217562077_0002/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-09 11:55:22,439 Stage-1 map = 0%, reduce = 0%
2021-06-09 11:55:27,627 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 3.32 sec
MapReduce Total cumulative CPU time: 6 seconds 40 msec
Ended Job = job_1623217562077_0002
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.04 sec HDFS Read: 15782 HDFS Write: 101 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 40 msec
OK
9
Time taken: 10.183 seconds, Fetched: 1 row(s)
hive>
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 190x48

hive> SELECT name FROM emp
> WHERE tax_paid='yes' AND eligibility='no'
> ORDER BY name;
Query ID = hadoop_20210609154831_d336eeb2-7ca2-4814-9f06-74ce2b981400
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1623217562077_0014, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:8088/proxy/application_1623217562077_0014/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-09 15:48:41,981 Stage-1 map = 0%, reduce = 0%
2021-06-09 15:49:06,733 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.97 sec
MapReduce Total cumulative CPU time: 5 seconds 490 msec
Ended Job = job_1623217562077_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.49 sec HDFS Read: 12981 HDFS Write: 240 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 490 msec
OK
aman
chiku
kshitiija
maliya
rajan
akrikant
subham
supriya
Time taken: 42.583 seconds, Fetched: 8 row(s)
hive>
```

Big data lab- Learning activity II

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJ: /usr/local/hadoop/sbin 190x48
hive> CREATE VIEW sal_ssn_view AS
> SELECT ssn,salary FROM emp;
OK
Time taken: 0.094 seconds
hive> SELECT * FROM sal_ssn_view;
OK
61326 61000
61327 66000
61328 44000
61329 33000
61330 28000
61301 30000
61302 40000
61303 45000
61304 55000
61305 30000
61306 43000
61307 25000
61308 69000
61309 30000
61310 35000
61311 42000
61312 37000
61313 28000
61314 30000
61315 29000
61316 54000
61317 30000
61318 55000
61319 26000
61320 30000
61321 30000
61322 40000
61323 48000
61324 39000
61325 60000
Time taken: 0.086 seconds, Fetched: 30 row(s)
hive>
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJ: /usr/local/hadoop/sbin 190x48
hive> SELECT salary,COUNT(eligibility) FROM emp GROUP BY salary;
Query ID = hadoop_20210609141041_cb339c63-9c85-427e-a230-8506be546996
Total jobs = 1
Number of reduce tasks not specified. Estimated from input data size: 1
Launching Job 1 out of 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reducers=<number>
Starting Job = job_1623217562077_0013, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJ:8088/proxy/application_1623217562077_0013/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0013
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-09 14:10:49.317 Stage-1 map = 0%, reduce = 0%
2021-06-09 14:10:54.452 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.54 sec
2021-06-09 14:10:59.562 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.61 sec
MapReduce Total cumulative CPU time: 3 seconds 610 msec
Ended Job = job_1623217562077_0013
MapReduce Jobs Launched:
  Stage-Stage 1: Map: 1 Reduce: 1 Cumulative CPU: 3.61 sec HDFS Read: 14900 HDFS Write: 507 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 610 msec
OK
25000 1
26000 1
28000 2
29000 1
30000 7
33000 1
35000 1
37000 1
39000 1
40000 2
42000 1
43000 1
44000 1
45000 1
48000 1
54000 1
55000 2
60000 1
61000 1
66000 1
69000 1
Time taken: 20.532 seconds, Fetched: 21 row(s)
hive>
```

Big data lab- Learning activity II

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 190x48

hive> SELECT name,ssn FROM emp
> WHERE salary > 40000 AND salary < 48000;
OK
pk 61328
alok 61303
abhishek 61306
mishra 61311
Time taken: 0.092 seconds, Fetched: 4 row(s)
hive>
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 237x56

hive> SELECT e.name,e.ssn,o.orderId,o.amount FROM emp e JOIN
o ON (e.ssn=o.custidssn);
Query ID = hadoop_20210609135037_95f51fcf-c608-4c6e-af6b-76bd2446d2ec
Total jobs = 1
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
2021-06-09 13:50:44 Dump the side-table for tag: 1 with group count: 22 into file: file:/tmp/hadoop/103323fd-dc1c-4a3a-9cf1-120eb87f3311/hive 2021-06-09 13:50-37_183_1344190831503681371-1/-local-10004/HashTable-Stage-3/MapJoin-mapfil
e31--hashtable2021-06-09 13:50:44 Uploaded 1 File to: file:/tmp/hadoop/103323fd-dc1c-4a3a-9cf1-120eb87f3311/hive 2021-06-09 13:50-37_183_1344190831503681371-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile31--hashtable (884 bytes)
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1623217562077_0008, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:8088/proxy/application_1623217562077_0008/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-09 13:50:52.540 Stage-3 map = 0%, reduce = 0%
2021-06-09 13:50:57.653 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 3.28 sec
MapReduce Total cumulative CPU time: 3 seconds 280 msec
End of Job = job_1623217562077_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 3.28 sec HDFS Read: 10570 HDFS Write: 1096 SUCCESS
Total MapReduce CPU Time Spent: 3 seconds 280 msec
OK
ajay 61301 70 10000
aman 61302 61 50000
aman 61302 64 10000
bhavya 61304 62 30000
bhavya 61304 63 10000
abhishek 61306 69 17000
abhishek 61306 81 12000
adham 61307 82 1100
saini 61308 65 8000
saini 61308 83 501
sanket 61309 86 50000
sanket 61309 67 15000
sanket 61309 84 150
supriya 61310 68 16000
supriya 61310 85 11111
mishra 61311 86 9999
kaushal 61312 89 9991
nallya 61313 88 2390
sujan 61314 89 7891
mukund 61315 71 12000
mukund 61315 80 65000
ashitija 61316 72 30000
pandey 61317 73 31000
rajan 61318 74 35000
skrikant 61319 75 2100
talpade 61320 76 1000
raj 61321 77 16000
saanya 61322 78 20000
gallavi 61323 79 21000
bhavya 61324 80 12000
Time taken: 22.628 seconds, Fetched: 30 row(s)
hive>
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 211x53

Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1623217562077_0009, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:8088/proxy/application_1623217562077_0009/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0009
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-09 13:53:25,321 Stage-3 map = 0%, reduce = 0%
2021-06-09 13:53:29,519 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.71 sec
MapReduce Total cumulative CPU time: 2 seconds 710 msec
Ended Job = job_1623217562077_0009
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.71 sec HDFS Read: 9946 HDFS Write: 1108 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 710 msec
OK
anurag NULL NULL
prince NULL NULL
pk NULL NULL
rajat NULL NULL
chiku NULL NULL
ajay 70 18000
aman 61 50000
aman 64 10000
alok NULL NULL
bhavya 62 30000
bhavya 63 10000
divya NULL NULL
abhishek 69 17000
abhishek 81 12000
subham 82 1100
saini 65 8000
saini 83 501
sanket 66 50000
sanket 67 15000
sanket 84 190
supriya 68 10000
supriya 85 11111
mishra 86 9999
kaushal 87 9991
mallya 88 2390
mujan 89 7891
mukund 71 12000
mukund 90 65000
kshitiya 72 30000
pandey 73 31000
rajan 74 35000
skrikant 75 2100
talpade 76 1000
raji 77 16000
saumya 78 20000
pallavi 79 21000
shreya 80 12000
vaibhav NULL NULL
Time taken: 21.208 seconds, Fetched: 38 row(s)
hive>
```

```
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin
hadoop@asr-VivoBook-ASUSLaptop-X412FJC-X412FJ: /usr/local/hadoop/sbin 237x56

hive> SELECT e.name,o.orderid,o.amount FROM emp e RIGHT OUTER JOIN
> orders o ON (e.sname,custid);
Query ID = hadoop_20210609135505_7fe07f36-6216-492a-964a-0e4fbc3997c4
Total jobs = 1
SLF4J: Found binding in [jar:file:/usr/local/hadoop/hive/lib/log4j-slf4j-impl-2.10.0.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2021-06-09 13:55:11 Starting to launch local task to process emp join: maximum memory = 239075320
2021-06-09 13:55:12 Uploaded 1 file to: file:/tmp/hadoop/103323f6-dc1c-4a3a-9cfd-120eb07f3111/hive_2021-06-09_13-55-05_573_2106176038447780768-1/-local-10084/HashTable-Stage-3/MapJoin-mapfile50--hashtable (1096 bytes)
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1623217562077_0010, Tracking URL = http://asr-VivoBook-ASUSLaptop-X412FJC-X412FJ:8088/proxy/application_1623217562077_0010/
Kill Command = /usr/local/hadoop/bin/mapred job -kill job_1623217562077_0010
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-09 13:55:20,740 Stage-3 map = 0%, reduce = 0%
2021-06-09 13:55:25,064 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.8 sec
MapReduce Total cumulative CPU time: 2 seconds 800 msec
Ended Job = job_1623217562077_0010
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 2.8 sec HDFS Read: 9343 HDFS Write: 916 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 800 msec
OK
aman 61 50000
bhavya 62 30000
bhavya 63 10000
aman 64 10000
saini 65 8000
sanket 66 50000
sanket 67 15000
supriya 68 10000
abhishek 69 17000
ajay 70 18000
mukund 71 12000
kshitiya 72 30000
pandey 73 31000
rajan 74 35000
skrikant 75 2100
talpade 76 1000
raji 77 16000
saumya 78 20000
pallavi 79 21000
shreya 80 12000
abhishek 81 12000
subham 82 1100
saini 83 501
sanket 84 190
supriya 85 11111
mishra 86 9999
kaushal 87 9991
mallya 88 2390
mujan 89 7891
mukund 90 65000
Time taken: 21.397 seconds, Fetched: 30 row(s)
hive>
```


Big data lab- Learning activity II

References

1. Hadoop-[https://github.com/1NT18IS009/1NT18IS009_abhishek_B_bdLab/tree/master/Programming%20Exercise/Exercise1\(hadoop\)](https://github.com/1NT18IS009/1NT18IS009_abhishek_B_bdLab/tree/master/Programming%20Exercise/Exercise1(hadoop))
2. Hive-[https://github.com/1NT18IS009/1NT18IS009_abhishek_B_bdLab/tree/master/Programming%20Exercise/Exercise2\(hive\)](https://github.com/1NT18IS009/1NT18IS009_abhishek_B_bdLab/tree/master/Programming%20Exercise/Exercise2(hive))