

Lab 02

Map Reduce Programming

IT413 No SQL Databases, Winter'2021, DAIICT, Gandhinagar; pm_jat

In this lab also we continue doing some hands on Map Reduce Programming. Here you create your own programs. Even if you do not know much about java and Hadoop, you should be able to attempt it.

Do following-

1. Consider a given employee file (employee.csv) with following attributes
(eno, name, dob, gender, salary, sup_eno, dno)

Values are separated by comma.

Write Map Reduce program to compute following from the file:

```
SELECT eno, name, gender from employee where salary > 30000;
```

You should able to do this by creating a map only MR job. Some idea you should get from following blog:

<https://data-flair.training/blogs/map-only-job-in-hadoop-mapreduce/>

2. Considering same above, write map reduce program to compute following from the file:

```
SELECT dno, min(salary), avg(salary), max(salary) from employee group by dno;
```

Attempt Creating Following on “web access log” in Lab01:

3. Create another Map Reduce program that reads from same log file, and outputs following summaries on Monthly Basis –

(a) Total number of requests.

(b) Total downloaded size

It should output: <Year-Month, Number of Requests, Download Size> for every months like Dec-2016, Jan-2017, and so!

4. Create another Map Reduce program that lists Timestamp, URL for which http response status has been 404.

Submission Required:

1. Source Code in single compressed file.