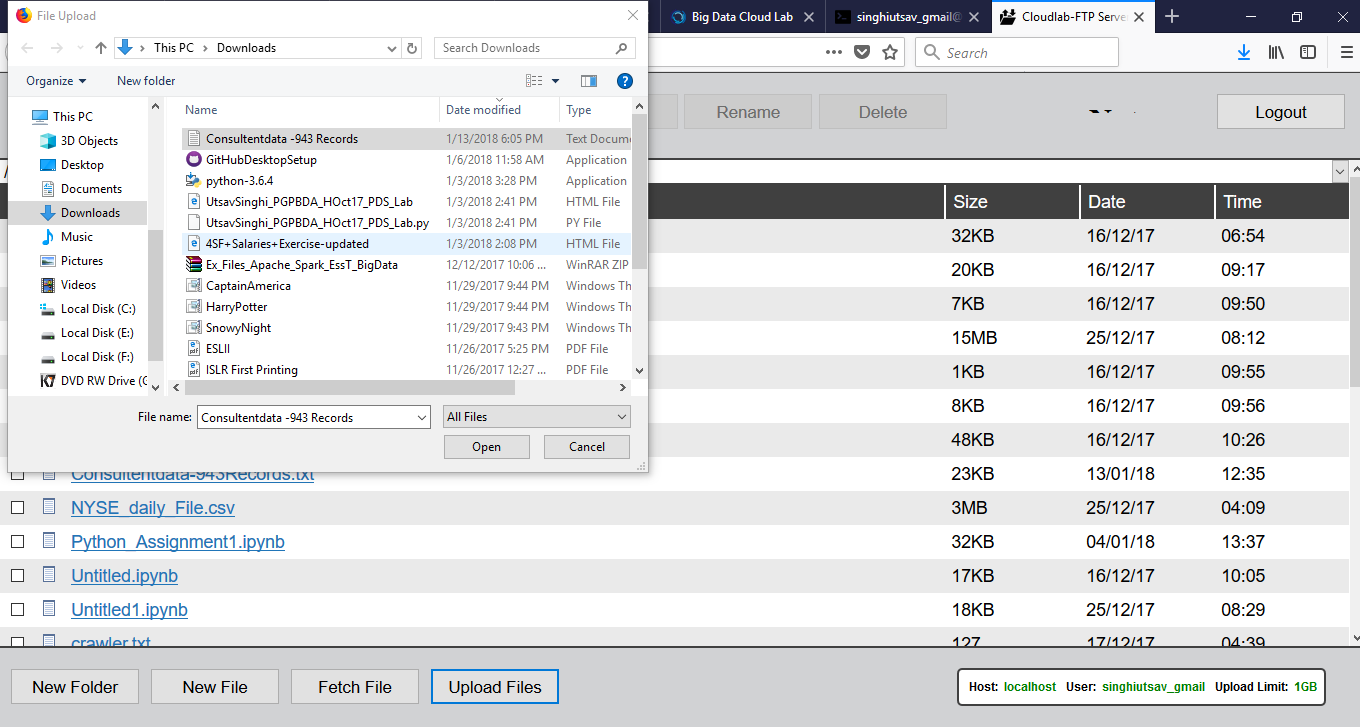
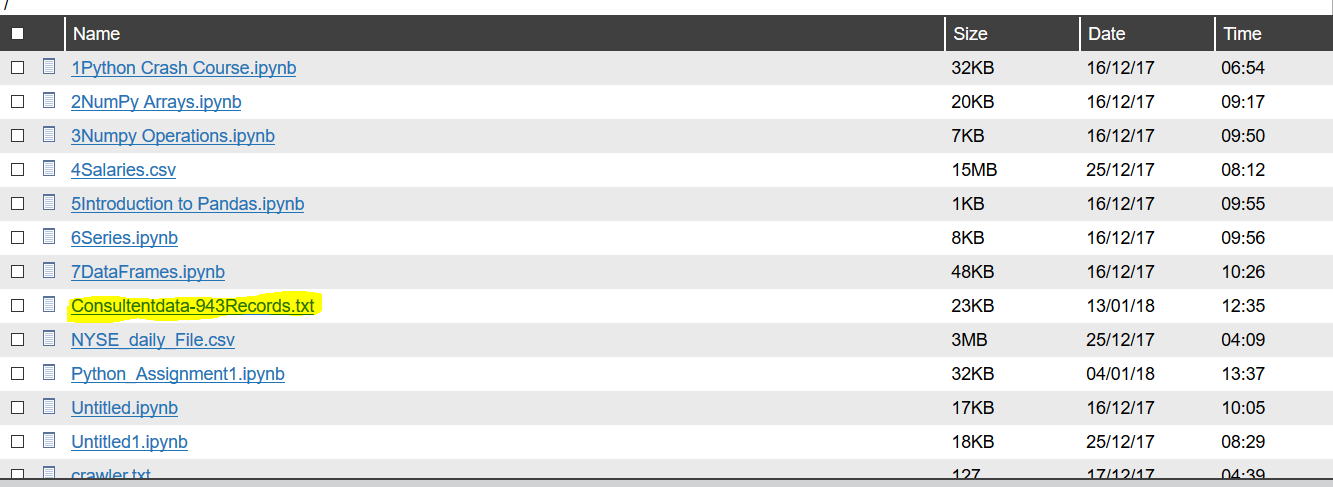
Using FTP, uploading file in local machine



File Uploaded in local machine



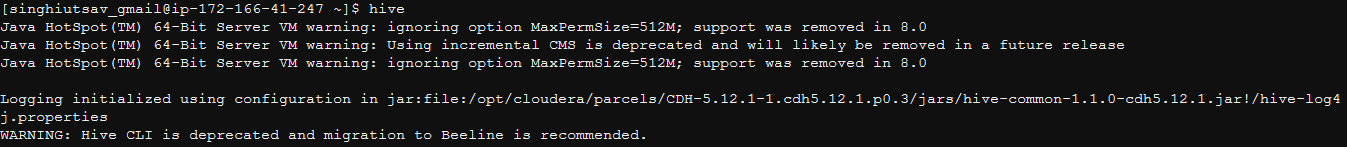
Login to Webconsole and creating new directory using mkdir



Now putting file in HDFS

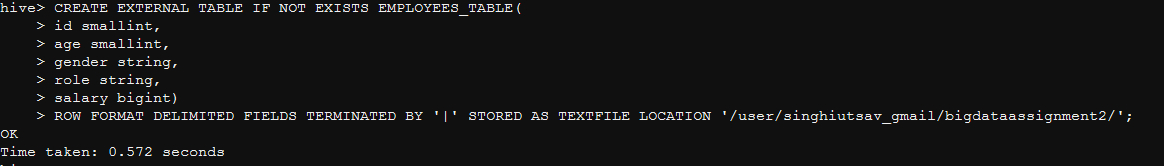


Entering Hive



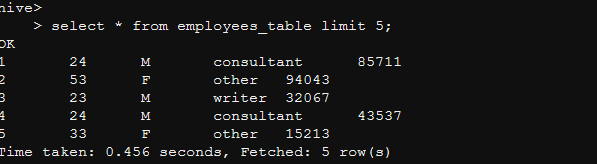
As per requirement, Creating External Table as “employees\_table” having 5 fields.

* CREATE EXTERNAL TABLE IF NOT EXISTS EMPLOYEES\_TABLE( Id smallint, age smallint, gender string, role string, salary bigint) ROW FORMAT DELIMITER FIELDS TERMINATED BY ‘|’ STORED AS TEXTFILE LOCATION ‘/user/singhiutsav\_gmail/bigdataassignment2/’;



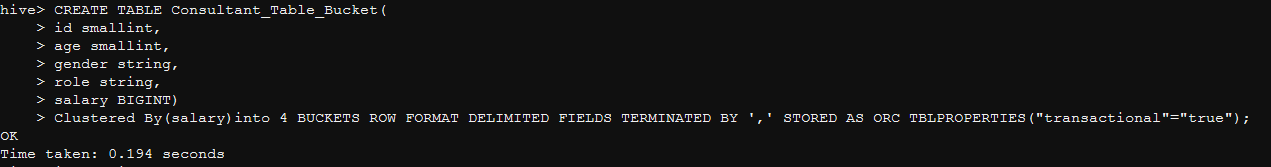
Checking Correct Data been Inserted

* Select \* from employees\_table limit 5;



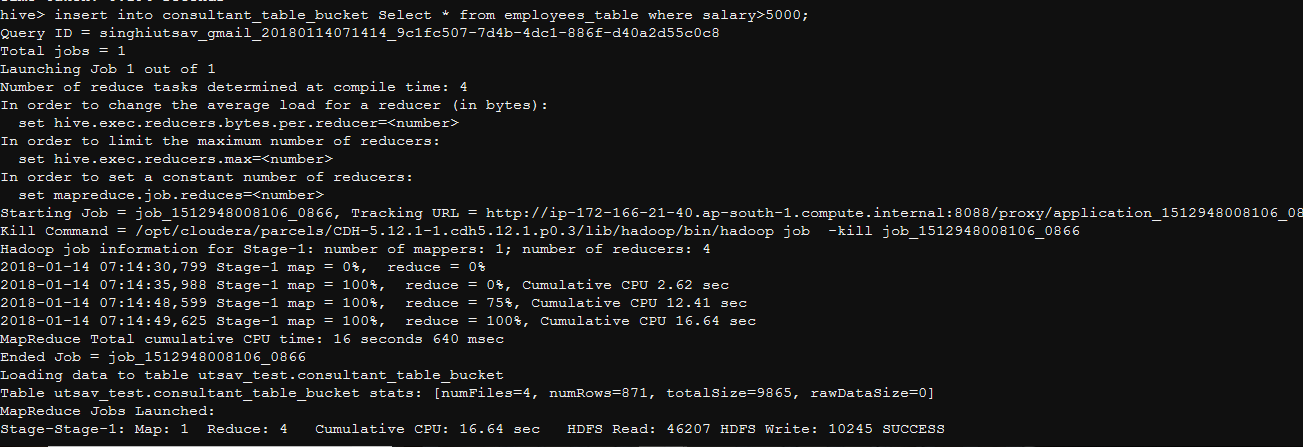
As per requirement, we will be Creating 4 buckets

* CREATE TABLE Consultant\_Table\_Bucket( id smallint, age smallint, gender string, role string, salary bigint) CLUSTERED BY(salary) into 4 buckets row format delimited fields terminated by ‘,’ stored as orc tblproperties(“transactional”=”true”);



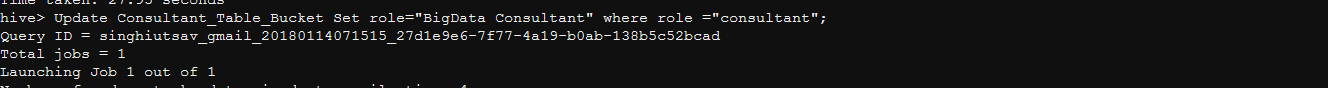
Inserting Data in Consultant\_Table\_Bucket

* Insert into consultant\_table\_bucket Select \* from employee\_table where salary > 5000



Updating table Consultant\_table\_bucket table role =’ consultant’ will be updated role=’BigData Consultant’

* Updating Consultant\_Table\_Bucket set role =’BigData Consultant’ where role= ‘Consultant’



Selecting maximum and minimum salary from table Consultant\_Table\_Bucket

