

HIVE BUCKETING:

1.CREATE DATABASE

```
~
hive history file=/tmp/training/hive_job_log_training_202303141324_204364239.txt
hive> create database bucketdb;
OK
Time taken: 7.105 seconds
hive> show databases;
OK
bucketdb
default
part
std
Time taken: 0.681 seconds
hive> use █
```

2.CREATE TABLE:

```
~
hive> create table sada(id int,firstname string,lastname string)
> row format delimited
> fields terminated by ','
> stored as textfile;
OK
Time taken: 7.78 seconds
hive> show tables;
OK
sada
student4
userratings
Time taken: 0.358 seconds
hive> █
```

3.LOAD DATA FROM LOCAL TO TABLE:

```
hive> load data local inpath '/home/training/aditya/bucketfile' into table sada;  
Copying data from file:/home/training/aditya/bucketfile  
Copying file: file:/home/training/aditya/bucketfile  
Loading data to table default.sada  
OK  
Time taken: 1.05 seconds  
hive> █
```

CREATE BUCKETING TABLE:

```
hive> create table bucketwala(id int,firstname string,lastname string)  
    > clustered by (id) into 5 buckets  
    > row format delimited  
    > fields terminated by ','  
    > stored as textfile;  
OK  
Time taken: 0.12 seconds  
hive> █
```

OVERWRITE DATA NORMAL TABLE TO BUCKETING TABLE:

```
hive> insert overwrite table bucketwala select *from sada;
Total MapReduce jobs = 2
Launching Job 1 out of 2
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_202303141318_0001, Tracking URL = http://0.0.0.0:50030/jobdetails.jsp?jobid=job_202303141318_0001
Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=0.0.0.0:8021 -kill job_202303141318_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2023-03-14 13:44:38,409 Stage-1 map = 0%, reduce = 0%
2023-03-14 13:44:43,518 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2023-03-14 13:44:44,546 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2023-03-14 13:44:45,572 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2023-03-14 13:44:46,591 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2023-03-14 13:44:47,633 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 1.4 sec
MapReduce Total cumulative CPU time: 1 seconds 400 msec
Ended Job = job_202303141318_0001
Ended Job = -162565767, job is filtered out (removed at runtime).
Moving data to: hdfs://0.0.0.0:8020/tmp/hive-training/hive_2023-03-14_13-44-21_292_8572309140725563283/-ext-10000
Loading data to table default.bucketwala
rmr: DEPRECATED: Please use 'rm -r' instead.
Deleted /user/hive/warehouse/bucketwala
14 Rows loaded to bucketwala
MapReduce Jobs Launched:
Job 0: Map: 1 Cumulative CPU: 1.4 sec HDFS Read: 0 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 1 seconds 400 msec
OK
Time taken: 26.922 seconds
hive> █
```

BUCKING IS DONE:

```
[training@localhost aditya]$ hadoop fs -cat /user/hive/warehouse/bucketwala/0000
00 0
1,aditya,anal
2,rahul,tiuwari
3,sandeep,sanini
4,ajay,singh
5,richa,sharma
6,shalu,sharma
7,pankag,singh
8,pandy,ji
9,abhishek,sharma
10,anjali,sharma
11,neha,sharma
12,pandul,sharma
13,khusi,sharma
\n,\n,\n
[training@localhost aditya]$ █
```