Page No.
Date

Assignment No. 4

* **Aim:** HBase & Hive.

* **Problem statement:**
Write a application using HBase & Hive QL for flight information system with including

1. Creating, dropping & altering database tables
2. Creating an external hive table to connect to HBase for customer Information table.
3. Load table with data, join table with Hive & insert row values & fields
4. Create index
5. Find the average temperature departure delay per day in year.

* **Objectives:**
- To understand various integration methods of NOSQL.
- To analyze the performance of distributed processing.

Page No.:
Date:

* Theory :-

→ HBase :- HBase is a distributed column oriented database built on the top of the Hadoop file system. It is open source project & horizontally scalable.
HBase is also model that similar to google Big table designed to provide quick random access to huge tolerance provided by HDFS.

- table is collection of rows
- row is collection of column families.
- column family is a collection of columns
- column is a collection of key value pairs

→ Starting of HBase
start_hbase.sh
hbase shell

→ Creating & droping tables
Syntax:

Create '<tablename>', '<column family>' [, '<column family>']

eg. create flight 'finfo' 'fsch':

Page No.
Date

② Listing tables →
Syntax → list

③ insert record

Syntax →
put <tablename> <row no> <column family>
<column name> <value>

eg put 'flight', 'info' 'source' 'pune'

eg
: get 'flight' for reading data

→ creating external Hive table to concept to the HBase for customer Information table

→ Load the ~~data~~ table with data, insert new values & field in the table with Hive

→ ADD these .jar files in HIVE on Hive prompt

① zookeeper.jar
② guava.jar
③ Hbase_client.jar
④ hbase_common.jar
⑤ Hbase_protocol.jar

⑥ hbase_shell.jar
⑦ hbase_shell.jar
⑧ hbase_thrift.jar
⑨ hive_hbasehandler.jar

→ set value of variables in Hive.

set hbase.zookeeper.quorum = localhost;
set hbase.regionserver.dns.nameserver = true
set hive.exec.dynamic.partition = true
set hive.exec.dynamic.partition.mode = nonstrict
set hive.exec.stagingdir = /temp/hivestage
set hive.exec.max.dynamic = 1000
set hive.exec.max.dynamic.periode = 1000

hive > create table hive_table emp(id int, name string, esal string) stored by apache.hadoop.hive.hbase.HBaseStorageHandler

hive > load data local inputpath '/home/hduser/desktop/empdbnew.txt' into table 'empdbnew'

hive > select * from hive_table emp

→ find the average departure delay per day in 2008.

Page No.
Date

Calculate avg. salary

hive> select sum(delay) from hbase.
flight_new;

→ Create index :-

create index hbase_index on table
hbase_flight_new(delay) AS 'org.apache.
hadoop.hive.ql.index.compact
Compact - IndexHandler' with deferred
rebuild

hive> show index ON hbase_flight_new.

* Conclusion :- Thus, In this assignment I have completed & studied the assignment of appln using HBase & HiveQL for flight information system.