Experiment 5: Hive and HiveQL

1. jps
2. open mysql

create table empsal (eid int, esal int);

insert into empsal values (1,1000), (2,2000), (3,3000), (4,4000), (5,5000);

select \* from empsal;

1. Import data into hdfs using sqoop:

after jps command, type:

sqoop import --connect jdbc:mysql://localhost:3306/emp --table empsal --username hadoop --password 123456 --target-dir /sqooptranser/emp -m 1

<press enter key>

1. Load recently imported hdfs data into hive and perform analysis:

type command in terminal

hive

in hive prompt, type

use test;

create external table emp (eid int, esal int) ROW FORMAT DELIMITED FIELDS TERMINATED BY ‘,’ LINES TERMINATED BY ‘\n’ LOCATION ‘/sqooptranser/emp1’;

select \* from emp;

select \* from emp where esal>3000;

**HIVE EXPERIMENT**

**hive**

**hive> create database test2;**

**use test2;**

**create table emp2**

**> (eid int,**

**> esal int)**

**> row format delimited**

**> fields terminated by','**

**> lines terminated by '\n'**

**> STORED AS TEXTFILE;**

**LOAD DATA LOCAL INPATH '/home/hadoop/hive/sample1.txt' OVERWRITE INTO TABLE emp2;**

**select \* from emp2;**

**create table employee (eid int, name String, salary String, destination String)**

**> comment 'Employee Details'**

**> row format delimited**

**> fields terminated by ','**

**> lines terminated by '\n'**

**> STORED AS TEXTFILE;**

**LOAD DATA LOCAL INPATH '/home/hadoop/hive/sample.txt' OVERWRITE INTO TABLE employee;**

**select \* from employee;**

**SELECT \* FROM employee WHERE salary>30000;**

**SELECT eid, name,salary FROM employee ORDER BY salary;**

**ALTER TABLE employee ADD COLUMNS (**

**> dept STRING COMMENT 'Department name');**

**SELECT salary,count(\*) FROM employee GROUP BY salary;**

**Inner join**

**SELECT c.eid, c.name, c.destination, o.esal**

**> FROM employee c JOIN emp2 o**

**> ON (c.salary = o.esal);**

### **Left Outer Join**

**hive> SELECT c.eid, c.name, c.destination, o.esal**

**> FROM employee c LEFT OUTER JOIN emp2 o**

**> ON (c.salary = o.esal);**

### **Right Outer Join**

**hive> SELECT c.eid, c.name, c.destination, o.esal**

**> FROM employee c RIGHT OUTER JOIN emp2 o**

**> ON (c.salary = o.esal);**

### **Full Outer Join**

**hive> SELECT c.eid, c.name, c.destination, o.esal**

**> FROM employee c FULL OUTER JOIN emp2 o**

**> ON (c.salary = o.esal);**