

HBase Homework

HBase, Pig, and Hive Integration

Teamwork – Due on 4/28/2015

Extract the files emp1.txt and emp2.txt in HBase.zip from the Support Files folder in BlackBoard to ~/hbase. These files represent how table emp(eid, fname, lname, age, dno, city, sal) has been vertically fragmented.

A: Hive Integration – use emp1.txt for this section. It represents eid, fname, dno, and city.

```
[training@localhost ~]$ cat hbase/emp1.txt
```

1	Saeed	2	MPLS
2	Joe	1	Edina
3	Sam	2	
5	Pat	1	Edina
6	Kurt	1	MPLS
7	Bob		MPLS

Write the necessary Hive scripts that load the contents of file emp1.txt as a table called 'emp' in HBase.

1. Show the contents of your Hive staging table
2. Use Hive and show the contents of the external table you create in Hive that stores its data in HBase
3. How many rows are in this table
4. Scan table emp in HBase
 - a. Using HBase, make sure that you add 'Edina' as default for all employees who do not have a city

B: Pig integration – use emp2.txt for this section. It represents eid, lname, age, and salary.

```
[training@localhost ~]$ cat hbase/emp2.txt
```

```
1,Rahimi,63,2000
2,Krummel,27,1000
4,Boyd,,3500
5,Zimmer,72
6,Abraham,66,6000
7,Jones,95,5000
```

Write the necessary Pig scripts that load the contents of file emp2.txt into the same HBase table that you used in Step A. How many rows are in table emp in HBase now? Why?

C:

1: Using Pig

1. Load the data in emp as a bag into Pig
2. Figure out the average salary for all employees
3. Print the count of employees who live in MPLS

2: Using Hive

1. Load the data in emp as a new table in Hive
2. Figure out the average age for all employees
3. Print the count of employees who live in Edina

Deliverables: Turn in a print out of your Hive, Pig, HBase scripts with the output for all three sections. Make sure that the output of each step and sub-step are right under the commands. Turn in this output in class when due. To save paper and to make sure your print out is readable, start Hive in silent mode using -S. Email me the source scripts.