

## **Hive Homework**

### **Individual Work**

**Due: On or before the first test**

**Data Set:** Start with the file(s) that your Pig homework as your dataset.

**To Do:** Write a set of Hive scripts that load these files into a staging external table called, **Teller**. Create a managed partitioned table, called **PtTeller** that partitions the rows in the table Teller dynamically based on the teller's name.

**A:** Working with table Teller, write a set of Hive scripts that print the answers the following questions:

1. The total amount of money and total number of bills collected by all tellers.
2. The teller name and the total amount of money collected and the number of bills and the average money for each teller (as done in the MR job).
3. Print the teller name and the number of bills for those tellers that have collected the same number of bills.

**B:** Working with the PtTeller table, write a set of scripts that print the answers to the following questions:

1. The number of partitions in this table.
2. The teller name and the total amount of money collected and the number of bills and the average money for each teller (as done in the MR job).
3. The teller or the tellers' names that have collected the maximum amount of money.

**Deliverables:** Turn in your Hive scripts and a print out of running them and the outputs in class.

**Note:** for the final run of your scripts run Hive in the silent mode and pipe the results of running your scripts into a text file called **tellers.txt**. Print this file, bring it to class and Email me the Hive scripts.