Big Data Systems - CS4545/CS6545
Hands-on 1
Due: January 20, 2020 at 5 pm

In this hands-on you will become familiar with how to build relational algebra expression with Apache Calcite.

INSTRUCTIONS FOR THE CODE

1. Launch the BigDataSystems Master VM

2. Get the handson1.zip file by the typing the command in a terminal:
$ wget http://www.cs.unb.ca/~sray/teaching/bds/handson1.zip

3. Unzip handson1.zip and there are 4 top level elements:
- org-apache-calcite-jdbc.properties
- src/
- lib/
- this file

4. Create an Eclipse Java project. You can launch the Eclipse IDE already included in the VM under:
/home/bigdata/eclipse/java-mars/eclipse

5. Overwrite your Eclipse project's workspace content with the top level elements in step #3 (except for this file)

6. Include all the jar files in /lib into your project.

INSTRUCTIONS FOR THE DATASET

1. The data files are under ./src/resources/company/. The tables are stored as .csv files. The schema of the tables are given below:

```
EMPLOYEE(EMPID:int,NAME:string,DEPTNO:int,GEN-
DER:string,CITY:string,AGE:int,SLACKER:boolean,MGRID:int,JOINEDAT:date,SALARY:float)
DEPT(DEPTNO:int,NAME:string)
CERTIFICATE(EMPID:int, COURSEID:int,COMPLETIONDATE:date)
COURSE(COURSEID:int,TITLE:string, CATEGORYID:int)
CCATEGORY(CATID:int, CATNAME:string)
```

The included java program RelAlgebraDemo.java uses Calcite CSV adapter to treat the above .csv files as data tables and thus enables you to run queries on them directly without any database engine.

EXAMPLES

Look at methods ***example0*** through ***example7*** in RelAlgebraDemo.java for various examples of how to write relational algebra expression using Apache Calcite. You can run them by uncommenting the following line in ***runAll():***

```
// runExample(builder, r); // handson1: uncomment this to run examples
```

Please refer to lecture notes L03_CS4545_CS6545_QueryProcessing.pptx posted in D2L for more on these examples.

For a tutorial on Apache Calcite see https://calcite.apache.org/docs/algebra.html
For Apache Calcite Java APIs see https://calcite.apache.org/apidocs/org/apache/calcite/tools/RelBuilder.html

INSTRUCTIONS FOR THE TASK

Look for the methods: ***exercise1***, ***exercise2*** and ***exercise3*** in RelAlgebraDemo.java and solve them. In particular, using the code template from the method ***exercise0,*** write the relational algebra expressions for following queries:

1. exercise1: Show the name, salary of the employees who earn more than 45k
2. exercise2: Show the names of employees who work in Sales department
3. exercise3: Show the name of the employees, the course titles and the dates of completion of the courses they took

Tip: it may be useful to write down the relational algebra expression in a piece of paper before you code it using Calcite. This piece of paper need not be submitted for the hands-on.

NOTE

The files modified in the virtual machines are not persisted. So, before you shutdown your virtual machine, you may want to save your modified file in your UNB account.

You can access your UNB user account's files from the Big Data Systems Master VM by accessing the folder /media/sf_FCS-HomeDir.
For instance, execute the commands below to copy a file from your user account to the VM:

```
cd /media/sf_FCS-HomeDir
cp <src file>  /home/bigdata/DataScience/workspace/.
```

SUBMISSION INSTRUCTIONS :

1) Submit the following:
    a. ***RelAlgebraDemo.java*** file with your **solution**
    b. ***Soln.txt file*** with the **relational algebra expression** generated by Calcite based on your solution
2) Submit through Desire To Learn (D2L) course drop box Hands-on1
3) Hands-ons not submitted electronically via Desire To Learn or submitted after the due date will NOT be marked.