

# COMP9318 Data Warehousing and Data Mining

*Session 1, 2009*

## Quiz 1

*Due Date: 23:59 6 May 2009 (Wed)*

### 1 MDX Queries

Consider the [Sales] data cube bundled with Mondrian. Write down an MDX query and its output for the following operation:

Slice the [Sales] cube on the condition of time is 1997 Q2 and show the total Store Sales ([Store Sales]) and total Store Costs ([Store Cost]) for all the USA stores grouped by states. Results for each state are shown in the same row.

You may use the Mondrian server at <http://snare09.cse.unsw.edu.au:8080/mondrian-embedded/index.html>. Notify me if the server is down.

### 2 ETL

#### 2.1 Kettle

- Download and install kettle v3.1.0.
- Choose “No repository” when starting kettle.<sup>1</sup>

#### 2.2 Data

- Download the data file `archdata.txt` from the course homepage <http://www.cse.unsw.edu.au/~cs9318/09s1/quiz/archdata.txt>.<sup>2</sup>
- Create a folder named `q1` under kettle’s main directory, and place the downloaded file there. This ensures that you can access the input file by the following relative path: `q1/archdata.txt`.

---

<sup>1</sup>Run `spoon.bat` in windows or `./spoon.sh` in Linux..

<sup>2</sup>The data file was downloaded from <http://www.tbae.state.tx.us/PublicInfo/Rosters.shtml> on 14 Apr 2009. It contains information about architects registered in Texas, USA. See more information.

## 2.3 Your Task

You need to construct a `kettle` transformation such that it can display state and the number of records in each state. The result should be sorted by the name of the state in ascending order. The output should be a text file in `q1/out.txt`, which look like the following:

```
State StateCount
  6
Ab 2
AE 2
AK 7
AL 47
...
```

You need to save the transformation as `q1.ktr`.

Note:

- Make sure you refer to the input file by the following relative path: `q1/archdata.txt`.
- Save your transformation periodically in case `kettle` hangs.

Comment: It is easy to see that there are records whose state information is not valid or in the USA. This gives you some idea of the “dirty” (e.g., `N/`) and “missing” (e.g., the null value) data in the real world.

## SUBMISSION DETAILS

1. This is an individual assignment.
2. You should submit your two files: `q1.txt` containing the MDX query and its output and `q1.ktr` for the transformation task. You can use the following command:

```
give cs9318 q1 q1.txt q1.ktr
```

3. To facilitate marking, please make sure you have included your name and student ID in the `q1.txt` file.
4. **Late submission policy: 0 if late!**
5. Copying assignments is **unacceptable**.

Hint to Q1: [DIM] .Members will show all the members defined in the dimension DIM.