USER ID:

User Study for CineCubes

http://www.cs.uoi.gr/~pvassil/projects/cinecubes/userStudies/2014-01/

Context

In this test, you are required to perform the task of a data analyst working with census data. We will give you (i) a generic querying tool, to work with sessions of queries that you construct on your own and (ii) our CineCubes system for automatically creating reports for data analysts. Your task will be to **create a report twice**, the first time **without CineCubes** and the second time **with CineCubes** available. Then, you will also be asked to complete a **questionnaire**, too.

Data set Description

We will work with the "Adult" dataset referring to data from 1994 USA census. The measure of interest is Working Hours per Week, assessing how much people work per week.

There are several dimensions (Age, Native Country, Education, Occupation, Marital status, Work class, Gender, and Race) in the Adult data set and a single measure, Hours per Week. The most fundamental aspect for these dimensions is the hierarchical structuring of the data. We uniformly name the levels of the hierarchy as L0, L1, … etc. The idea behind the hierarchies is that a value in a higher level in the hierarchy practically groups several values in the lower levels. So for example, the value `Post-Grad` in level L2 in the hierarchy of the dimension Education is mapped to 2 values in level L1 (`MSc` and `PhD`). This means that all the records in the data set that have a value `MSc` or a value `PhD` at the Education attribute can be grouped in the same group if we decide to see the values of the dataset grouped by Education.L2.

See the hierarchies for the dimensions in the following figures. Dimension Age (not shown) is organized in years, 5-year, 10-year and 20-year intervals as well as *.

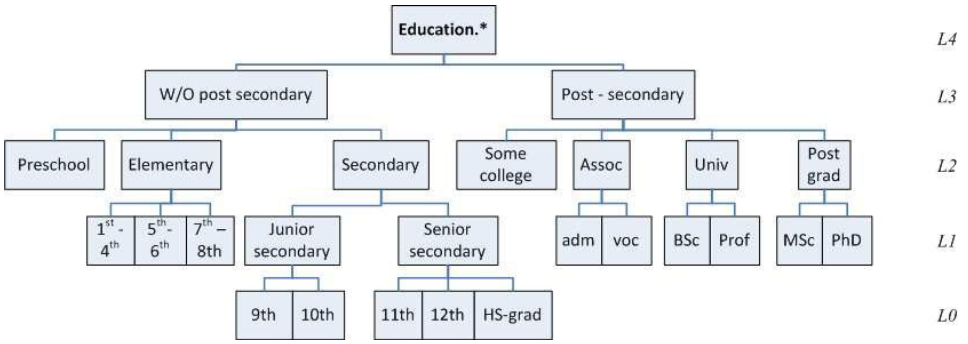

Figure 1 The hierarchy for the dimension *Education*

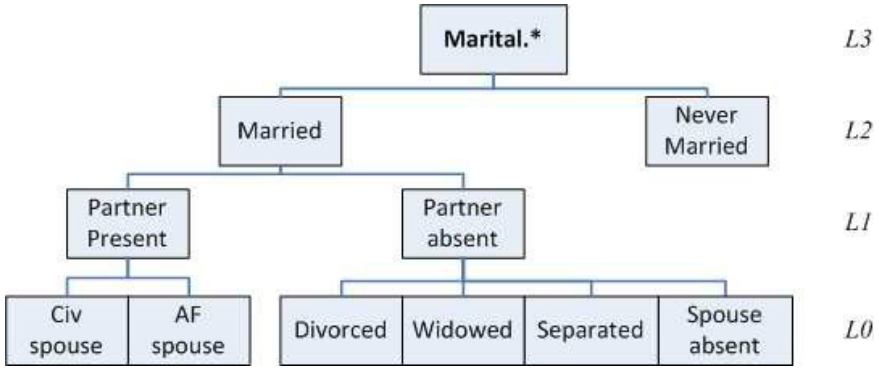

Figure 2 The hierarchy for the dimension *Marital Status*

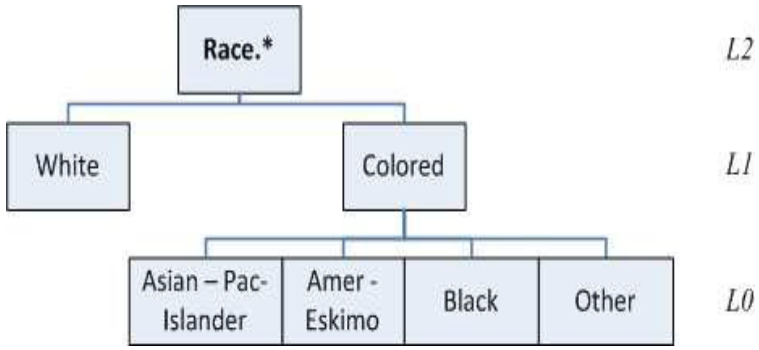

Figure 3 The hierarchy for the dimension Race

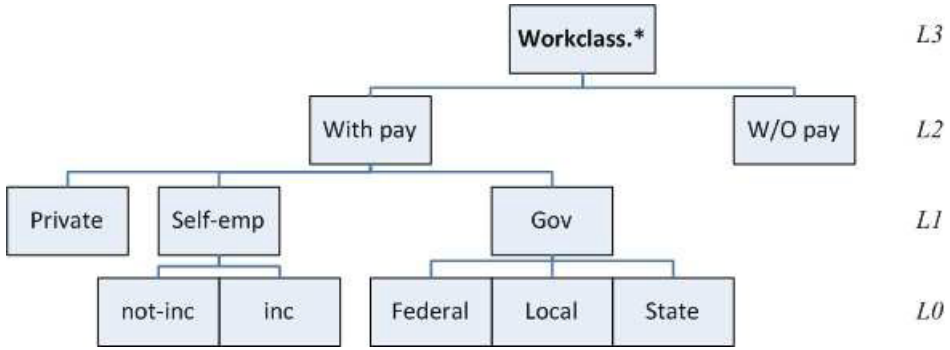

Figure.4 The hierarchy for the dimension Work class

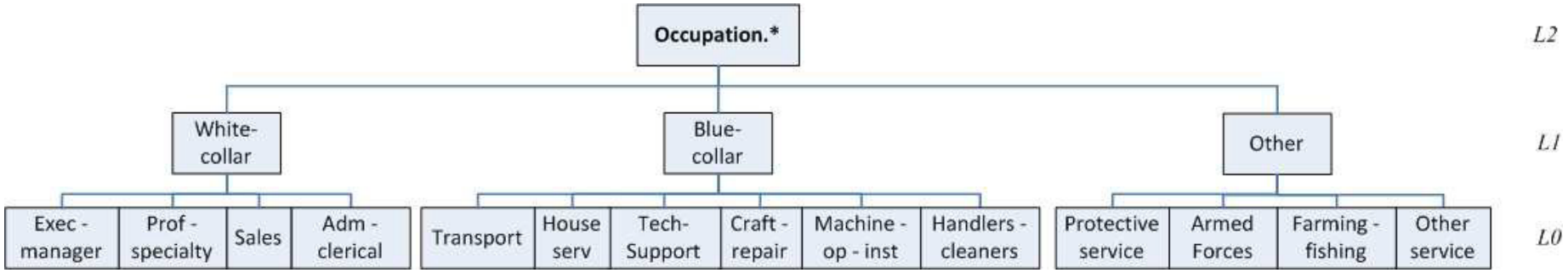

Figure 5 The hierarchy for the dimension *Occupations*

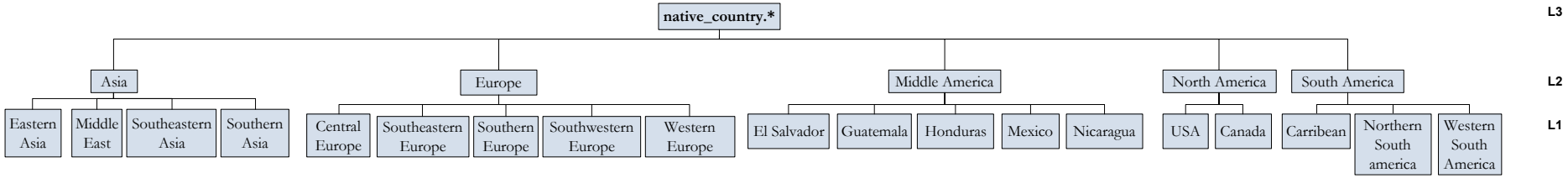

Figure 6 The hierarchy for the dimension Native Country (L0 ignored due to numerous values)

Tool description

The user interface requires you to work with data as they would typically appear in a spreadsheet analytic tool. So, the queries are assumed to abide by a 2D principle giving measures arranged by rows and columns. This is why all our queries are grouped by two groupers. Since we work with a measure of working hours the aggregate function is preset to Average. One can add as many selection conditions as he thinks fit.

In the above screen, we answer the following query: "***What is the effect of age (organized in 10 year intervals) and gender on the average number of working hours for people with country of origin in North America which are of age in 37-56?***"

Observe:

- the measure of interest as the average working hours in box: `Measure`
- the specification of the for the subset of the population via two `Selection Conditions`: native country in North America (level 2) and age in 37-56 (level 3).
- the role of the age group of level 2 (in 10year groups) and gender of level 0 (i.e., both sexes) in the `Grouper` boxes

The system is tailored for the particular data set, so one can construct the query via point-n-click. We need to give a name to the query too.

Objective

The goal is to construct a report where you analyze the situation for the following task

"***How is age in (10 year intervals) and education (post-secondary vs. without post-secondary) affecting the average number of working hours for people country of origin in North America (level 2) independently of their occupation?***"

We will need you to

- **count the time** you have spent for each report
- give a **list of bullets with highlight remarks** on the question
- construct a **textual description** (possibly with some charts if you like / have the time)

To help you, we list a few possible ways of searching for highlights and

(a) report on the a first answer along with important observations

(b) try to find out whether the observed status is typical or not

(c) possibly explain the details of the observed status

(d) … please think out of the box and do not stick only to what we suggest …

Questionnaire

Once done, you will have to complete a questionnaire: please go to

http://www.cs.uoi.gr/~pvassil/projects/cinecubes/userStudies/2014-01/

and follow the link to the on-line questionnaire.