Lab exercises for the Data Warehouse part of CS636 in November 2005

You can make a maximum of 41 points in these exercises. Each exercise counts as many points as shown (the first three exercises are warm-ups that do not count). Please note that your answers to all exercises should be handed in by December 10.

For all Star Tracker (ST) exercises,

- use the Grocery database that comes with ST,
- use ST's operators (instead of, e.g., computing the results by hand),
- briefly explain how you have used these operators,
- provide a snapshot of the result, and
- briefly comment on this result.

In the MSc labs, you will find demonstrators in case you have questions. Moreover, you might find the ST manual as Appendix D of the Kimball book helpful.

Exercise 1. Consider the two parts of ST's main window (report definition and report results), and make sure that you understand the following terms: fact table, calculated field, (non-)additive facts, dimensions, dimension tables and attributes, groups.

Exercise 2. Use ST's Browse feature to familiarise with the dimensions.

Exercise 3. Go to ST's GROUPS tab and define a new group of Products, e.g., those that are healthy. Define new comparison groups for the store and the date dimensions.

Exercise 4. (2/41) Use ST to generate a report showing the sales of Diet products in November 1995.

Exercise 5. (1/41) We would like to see the total sales made per product: summarise the sales data (in dollar) for November 1994 and for November 1995 for each product.

Exercise 6. (2/41) Compare the total dollar sales from November 1994 with those from November 1995 and compute the percentage change between them.

Exercise 7. (4/41) Consider the Sales Schema from Slide 41 "Star Schema Example" (second handout) and write SQL queries for the following "Roll-Ups":

1. Summarise the Sales per Year.
2. Summarise the Sales per Year, City, and Colour of Product.
3. For each combination of product description and size, find the colours of which the most product units were sold, and show these sales by size of the products of this colour .

Exercise 8. (6/41) Consider a data warehouse storing data about sales, where the total items sold are stored, organised by customer order and product. Each customer order includes the name of the customer and the date of the order; each product includes a description of the product and its price.

1. Devise the relational schema (specifying the relations, the attributes, the primary keys, and the foreign keys) of the above data warehouse using the star schema.
2. Write a SQL query to answer the following question: "Which customer(s) made an order containing at least five products with different descriptions?"
3. Write a SQL query for the following report: "Which customer(s) made the largerst order (i.e., those that would result in the largest bill)?"
4. Consider to add a new level *product categories* to the product dimension. Devise the new relational star schema, and write a SQL query for the following report: "Select the total number of products sold per product category".

Exercise 9. (3/41) Assume you are building a movie warehouse with an enterprise model containing three relations:

movie(title,director, year)
european(director)
critique(title,review)

and (so far) two source relations:

source1(title,year,director) for movies since 1960 with european directors and
source2(title,critique) for movies since 1990.

Describe the links between these relations for the global as view and for the local as view as SQL queries. I.e., for each table in the warehouse, formulate it as a SQL query over the source relations. For each source, formulate a SQL query S over the warehouse relations such that S *would* return the same tuples that you find in this source.

Exercise 10. (3/41) Consider the cube lattice on slide 48 "Cube Aggregation Lattice" and expand it with those views that can be built by rolling-up the time hierarchy on slide 51 "Time: Dimension with Interesting Hierarchy" , following the example on slide 50 "Dimension Hierachies". How many cubes are in this expanded lattice (explain your answer)? What implications does your count of cubes have?

The remaining exercises are all about the analysis of multi-dimensional data, and should familiarise you with the kind of analysis one wants to make in OLAP: use Star Tracker's operators on groups to answer the following questions and, for each question, briefly describe how you used Star Tracker to obtain the answer, show a snapshots of the intermediate results, and briefly interprete the final result. Please note that only the first exercises are explained in detail and step by step: I assume that you become more familiar with Star Tracker and learn how to use it to analyse data.

Exercise 11. (7/41) In this example, we want to check whether promotions on the products of a brand, the "National Bottle" brand, have worked:

1. which promotions have been made on "National Bottle" products?
2. use appropriate time windows (and ST groups for these!) to compare, for the latest promotion, the unit sales before, after, and during this promotion (taking into account the time after the promotion is crucial to see the effect of a promotion). Display, for the "National Bottle" product, the unit sales before, during, and after this promotion, and show their comparisons as percentages.

Exercise 12. (7/41) This exercise is similar to, but a bit more detailed than the previous one: we analyse effects of promotions (of the Western Vegetable brand) to the sales of the Big Can brand.

1. Identify promotions of the Western Vegetable brand; identify the stores that had these promotions and carry Big Can products and, for each promotion, specify a store group.
2. for each of the promotions, define time groups of equal length for before and during the promotion (i.e., the lengths of these groups depend on the length of the promotion).
3. List, for each promotion and each participating store, the total sales (in unit sales and dollar sales) of Big Can before and during the promotion, and the differences as percentages.
4. Now summarise and compare total sales (in unit sales and dollar sales) of Big Can of stores participating in promotions (which promotion worked better?), and compare them with those not participating in promotions.
5. What does this analysis tell you about Western Vegetable promotions?

Exercise 13. (6/41) We would like to see those stores that sold above average units of candy in November 1994. Use ST's break rows and exception conditions to list those stores.
Now we want to see, for each month of the last quarter of 1994, those stores that sold above average units of candy within this month. What does the result tell you?