**44-560 Advanced Topics in Database Systems**

**The Grocery Store Exercise KEY**

We begin by creating a data warehouse for the grocery store chain. The data warehouse will obtain its data from the operational database modeled in part 1 of this exercise. The business process to be modeled is grocery store sales at the point-of-sale (POS) terminals.

With this data warehouse

* We want to be able to analyze sales by store or by product, over any time period. In particular, we want information about the quantity sold of a particular product, the dollar amount from the sale, the dollar cost, and the dollar profit.
* We are interested in asking questions such as “How many of product X were sold at store Y on date Z?”.
* We may want to roll up sales data by department or brand.
* We may want to compare sales in stores based on the number of square feet in the store or the floor plan type of the floor.
* We may want to compare sales over cities, districts or regions.
* We want to maintain information about individual transactions in the event that we want to do market basket analysis in the future.

Note that we are *not* interested in tracking inventory or analyzing shelf space requirements. Phone and fax numbers for stores and departments are not needed. We are not interested in inventory or shipping information, such as quantity on hand, units per retail case, units per shipping cases. However, we do want to store weight information. We are not interested in personnel information, such as names of managers.

We will first complete the four-step process for data warehouse design as described in Chapter 2 of the text by Kimball and Ross.

Step 1: Select the business process. This has been done. The business process to be modeled is grocery store sales at the point-of-sale (POS) terminals.

Step 2: Declare the grain. The grain is an individual line item on a POS transaction. This fine grain is necessary if we are to be able to do market basket analysis as specified.

Step 3: Determine the dimensions. Date, Product, and Store. In addition, the POS transaction will be a degenerate dimension

Step 4: Determine the facts. SalesQuantity, SalesDollarAmount, CostDollarAmount, ProfitDollarAmount

**Your task:** Create the dimensional model for the data warehouse described above.

Your solution may be slightly different from that shown below. Here is a description of some of the allowed variations:

StoreDim: Your model *must* include the surrogate key StoreKey, in addition to StoreNumber, StoreName, City, State, floor plan type, total square feet, and the district and region description. District square miles is not addressed as being needed or not, so it could be omitted. Similarly for the district and region phones, and for the square feet allocated to grocery, frozen, and meat. It is explicitly stated that phone numbers for stores are *not* needed, so Phone and Fax should be omitted from Store.

ProductDim: It is explicitly stated that we are not interested in analyzing shelf requirements, or inventory, or shipping information, so height, width, depth, units per retail case, units per shipping case, and cases per pallet must be omitted. We do want weight information so weight and unit of measure must be included. In addition, brand and department information must be included. Unit cost and price will change frequently and should *not* be included in the dimension table.

Storing the keys for region, district, department, and brand (from the operational database) is not addressed in the requirements stated, but they should be stored, unless explicitly told not to store them.

