AM05 Data Mgmt – Lab 4

This lab has two parts. The first part asks you create a star schema suitable for a data warehouse or data mart. The second part ask you to parse text data with regular expressions.

Part 1. DataMart Design

Problem 1. You are to construct a *star schema* for Simplified Automobile Insurance Company. The relevant dimensions, dimension attributes, and dimension sizes are as follows

Dimension	Dimension Attributes	Dimension Size comment
InsuredParty	1. InsuredPartyID 2. Name	There is an average of two insured parties for each policy and covered item.
CoverageItem	1. CoverageKey 2. Description	There is an average of 10 covered items per policy.
Agent	1. AgentID 2. AgentName	There is one agent for each policy and covered item.
Policy	1. PolicyID 2. Type	The company has approximately 1 million policies at the present time.
Period	1. DateKey 2. FiscalPeriod	

Facts to be recorded for each combination of these dimensions are PolicyPremium, Deductible, and NumberOfTransactions.

a. Design a *star schema* for this problem. See example on next page for the format you should follow.

b. Estimate the number of rows in the fact table, using the assumptions stated previously.

c. Estimate the total size of the fact table (in bytes), assuming that each field has an average of 5 bytes.

Problem 2. Simplified Automobile Insurance Company would like to add a Claims dimension to its star schema. Attributes of Claim are ClaimID, ClaimDescription, and ClaimType. Attributes of the fact table are now PolicyPremium, Deductible and MonthlyClaimTotal

a. Extend the star schema from the previous problem to include these new data.

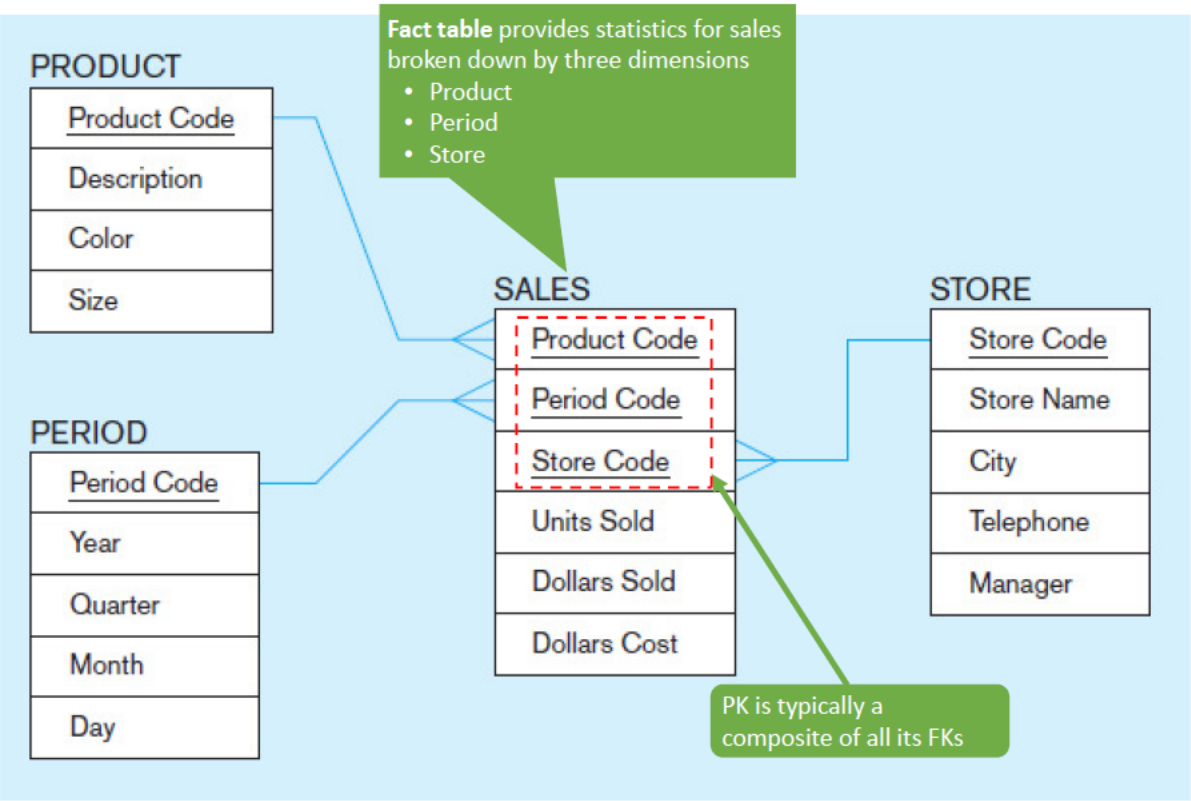
b. Calculate the estimated number of rows in the fact table, assuming that the company experiences an average of 2000 claims per month.

Part 2. Parsing text data with regular expressions

We starting working with text data from a feed of 911 calls from Monroe county New York. I demonstrated in class how to use regular expressions to extract some of the specific data items using regular expressions. Please,

1. Develop regular expressions to extract all the data fields from the raw text data.
2. Use UPDATE queries to place the extracted data in the appropriate columns of the database table specified in the supplied sql file that accompanied the class notes. This will require casting to the data types that match those in the specification of the table.
3. Run some select queries on the data to explore common event types and the proportion of calls responded to by different agencies.

Example of star schema



Note: while the operational database has records of each sale, here we get summary data (units sold, dollars sold, dollars cost)