IST722: Class Exercise 4

This is an individual assignment.

Before you begin, please make sure you've read and understand 1) our class honor code, 2) course policies on late work and 3) participation policies as posted on the syllabus. "I didn't know" is not an excuse.

You should cite your sources in a standard format like MPA or APA and include a list of works cited.

Your Name:	Bhavya Shah
Your Email:	bhshah@syr.edu

Instructions (Refer Units 4 & 5)

Answer each of the following questions as concisely as possible. More is not necessarily better. Please justify your answer by citing your sources from the assigned readings from our textbooks, our class lectures, or online if directed to do so. Be sure to cite in text and include a list of works cited. Place your answer below each question. When you're finished, print out this document and bring it to class as part of your participation grade.

Questions

[1] Summarize (SCD) Slowly Changing Dimensions and (RCD) Rapidly Changing Dimensions. Give examples.

Ans - Slowly Changing Dimensions (SCD) in data warehousing refer to dimensions that change gradually over time and require historical tracking or a strategy to address the change. They can be categorized into three types:
SCD Type 1: overwriting existing attributes
SCD Type 2: Add a new dimension row
SCD Type 3: Add a new dimension attribute
Example are employee has address change or customer has name change.

Rapidly Changing Dimensions (RCD), on the other hand, are dimensions that experience frequent and significant changes. It is loosely defined as not often and with no consistency.
Examples are age of a person changing annually or change of a customer's address that does not happen frequently.

[2] Explain a) conformed dimensions b) role-play dimensions c) degenerate dimensions d) factless facts. Give examples.

Ans - a) Conformed dimensions: These dimensions are essential in data warehousing as shared reference dimensions across fact tables, enabling drill-across operations, faster star schema development, and ensuring data consistency. Two types: Identical Dimensions (e.g., "Date" with consistent attributes) and Perfect Subset of an Existing Dimension (e.g., "Employee" with tailored attributes).

b) Role-play dimensions: The same physical dimension plays more than one logical dimensional **role**. This is common among the date dimensions. Stored in the same physical table, just aliased as a view. Implemented as multiple FK's in the fact table to the same dimension table. Examples: **Date:** Order Date, Shipping Date, Delivery Date, which is Same Date dimension, **Address:** Ship to, Bill to which is Same Address Dimension.

c) Degenerate dimensions: Degenerate dimensions are those that we keep in the fact table because they change too often to be kept in their own dimension, and there are a lot of them. Some of them are business keys since they give access to operational data in the ODS. They are typically the fact table's primary key and any attribute. A degenerate dimension is one that is not a dimension key or fact in the fact table. Order Number, Flight Number, etc.

d) Factless facts: A factless fact table is a form of data warehouse fact table that contains just dimension keys and no quantitative measurements. It records events or relationships between dimensions without requiring numerical data. Example: A Student Enrollment factless fact table with dimension keys Student ID, Course ID, Semester ID, and Enrollment Date, intended to manage student enrollment events without explicit quantitative metrics.

[3] What is the best choice for PK in a Dimension table? What is the best choice for PK in a Fact table?

Ans - Surrogate keys should be used in dimension tables for PK in the table. Dimension table properties should not have null values. If attributes do not have a value it should be assigned one, such as "Unknown", "Not Shipped", "Not Delivered" etc.

[4] What are Database Schemas? How are they useful? Give examples.

Ans - The purpose of a database schema is to logically arrange items such as tables, views, and procedures. Schema ensures the security of objects. The schema provides an overview of the data in the database and makes it easier to track down information in the database. It also has integrity limitations and is used to keep data consistent.

In a data warehouse, a schema organizes information into dimension tables (e.g., products, customers, time) and a fact table (e.g., sales data). Dimension tables provide details about the entities, while the fact table holds measurable data associated with those entities.

[5] What are Conceptual, Logical and Physical Models in the data warehouse context? Give examples of each.

Ans - The Conceptual Model is a high-level representation of the data warehouse, focusing on business requirements and entities, not technical details. For a retail data warehouse, it includes entities like "Products," "Customers," and "Sales."

The Logical Model is a detailed description that bridges the conceptual and physical models. It defines data organization, relationships, and business rules. For instance, it specifies attributes like "Product_ID," "Product_Name," and their relationships.

The Physical Model involves implementing the data warehouse, considering specific database technology, data types, and optimization. It includes details like the database platform (e.g., Oracle, SQL Server) and storage methods.

WORKS CITED:

Class slides and Professor Fudge’s videos.