IST722: Unit 05 Participation Questions

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:	Rashika Singh
Your Email:	rsingh37@syr.edu

Instructions

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] What is the best choice for PK in a Dimension table? When are NULL values acceptable in the attributes of dimension tables? Explain.

Dimension tables should use Surrogate keys for PK in the table. The attributes in dimension tables should not have null values. Attributes without a value should be assigned a value such as -1= "Unknown", -2= "Not Shipped", -3= "Not Delivered".

[2] What is the best choice for PK in a Fact table? When are NULL values acceptable in the fact values of a fact table? Explain.

Fact tables should use composite keys composed of dimension foreign keys and degenerate dimensions. Foreign keys in the fact table should not be null they should be assigned for unknown number such as -1. Nulls are allowed in fact tables, but they are not allowed in calculations.

[3] What are Database Schemas? How are they useful?

Database schema is to logically group objects such as tables, views, procedures. Schemas are namespaces which facilitate the separation, management and ownership of database objects. Objects are securable by schema. The schema gives overview of data in the database and makes it easier to track information present in the database. It also contains integrity constraints and is used to maintain data consistency.

[4] Discuss Conceptual, Logical and Physical Models in the data warehouse context.

Conceptual Model start with entity and entity relationships (High-Level)

Logical Model add attributes, Primary Keys, Foreign Keys (Detailed)

Physical Model add table names, column names, column data types (construction)

Internal Model: no user abstraction

External model: user abstraction: e.g.- views

[5] What are the three ways we can improve the performance of a star schema?

The three ways to improve performance of star schema are:

1) Summary tables:
Aggregate popular roll-up data and use ETL process to create
2) Materialized views or indexed views:
Copy of the query result
3) Partitioning:
Organize the fact table into partitions by date
4) Indexes:
Single clustered index for order of table and multiple indexes for improving search of a table

WORKS CITED:

Professor fudge videos and lecture slides