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:	Akshay Bhala
Your Email:	abhala@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] 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?

A database schema represents the logical configuration of all or part of a relational database. It can exist both as a visual representation and as a set of formulas known as integrity constraints that govern a database. These formulas are expressed in a data definition language, such as SQL. As part of a data dictionary, a database schema indicates how the entities that make up the

database relate to one another, including tables, views, stored procedures, and more. A database schema is a sketch of a planned database. It doesn’t actually have any data in it.

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

ERD feature	**Conceptual**	**Logical**	**Physical**
Entity (name)	Yes	Yes	Yes
Relationship	Yes	Yes	Yes
Column		Yes	Yes
Column’s Type		Optional	Yes
Primary Key			Yes
Foreign Key			Yes

In the table, it summarizes the characteristics of the three data model:

- The **Conceptual Model** Is To Establish The Entities, Their Attributes, And Their Relationships.
- The **Logical Data Model** Defines The Structure Of The Data Elements And Set The Relationships Between Them.
- The **Physical Data Model** Describes The Database-Specific Implementation Of The Data Model.

[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:

https://www.lucidchart.com/pages/database-diagram/database-schema

https://online.visual-paradigm.com/knowledge/visual-modeling/conceptual-vs-logical-vs-physical-data-model/

Lecture discussions