IST722: Unit 08 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 Data Quality? List three types of DQ rules with examples.

Data Quality are activities to ensure the data in the data warehouse is correct and complete so that we can trust the data in the data warehouse.

Three types of data quality rules are:

1. Incoming Data Validation- Rules are checked as staged data enters the data warehouse. Example- mobile number should be only 10 characters.
2. Cross-Reference validation- Rules to check incoming data against the data already in the Data Warehouse. Example- Monday's website visitors against the running average of last 3 Mondays.
3. Data warehouse Internal validation- Check data warehouse against itself, typically for aggregates. Example- Yearly sales in summary table matches actual number of sales by year.

[2] Explain Data Lineage and Dependency.

Data Lineage is the ability to look at a data element and see how it was populated. Audit Tables help here.

Dependency is in opposite direction of lineage where we look at the source table and identify the cubes and star schemas which use it. Custom Metadata tables are used for dependency.

[3] What is Master Data? How is this different from Dimensions in a DW? Give examples.

Master Data is the main entity of the organization which maintains central copy of the organization's data. Master Data stores union of the entity data which is available as compared to dimension data which stores data relevant to the

business process. Example- Customer is master data but particular process that deals with online customers is dimension data.

[4] How can a company benefit from Master Data Management? Does every company need an MDM strategy?

Master Data Management is creating a single “Reference Copy” of key business entities. It offers an organization a single version of what “Customer” means and helps reduce inconsistencies in data distributed throughout the enterprise so that it provides clean and reliable data. It also provides automated rules and utilities to maintain “golden records” for business entities. It is not mandatory to have master data management for each company if the people there know their roles and responsibilities.

[5] Explain the composition and function of a data governance committee or board.

Data Governance consists of employees such as Chief Information Officer (CIO) / Chief Data Officer, risk managers, data administrators and data custodians. The Data Governance board manages overall data of organization and address data availability, usability, integrity and security. It defines procedures and definitions of how to deal with data situations and establish plans for these procedures. Data administrators/ stewards control information in the organization. Data custodian control access, documentation and auditing whereas CIO is for setting goals, approving policies and managing risks.

WORKS CITED:

Professor Fudge slides