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

Data Quality is the activity performed to ensure that data in the data warehouse is correct and complete.

Three types of DQ rules are:

1. Data warehouse internal validation – In internal validation, data present in the data warehouse is checked, usually by aggregation. For example, Cumulative GPA should match the average of individual Grade points over all semesters
2. Incoming data validation – Check the data that is incoming, basically as staged data enters the data warehouse. For example, when we have a binary Value, we ensure that nothing other than True or false has been entered.
3. Cross-reference validation – Cross-reference validation is used to check the incoming data against data that is already present in the warehouse. For example, we know that website visitors on Mondays are usually about 15% of overall visitors; so we can check number of visitors on any given Monday to cross validate our data.

[2] Explain Data Lineage and Dependency.

Data Lineage is being able to look at a data element and see how it was populated in. It is basically tracking how an element got into the data warehouse.

Dependency allows us to track things in the opposite direction, i.e. to identify tables that have values in the data warehouse.

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

Master data is the core entity of the organization. It is the central copy where entire organization's entity data is stored. Master data, unlike dimension data, stores the complete union of entity data available. Whereas, dimension stores data only relevant to business processed. For example, customer data is master data. But if we're looking at a certain business process that deals with online customers, that's dimension data.

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

Master Data Management is highly beneficial to any company. It helps avoid data duplication and in turn improve data quality. It also helps with increasing data accuracy and allowing informed decision making. Although having an MDM strategy is very beneficial, it is not mandatory to have it, if in a firm each individual knows their role and what they're doing, they can manage without an MDM strategy.

[5] Explain the composition and function of a data governance committee or board.
A data governance committee is composed of all levels of employees, from Chief Information Officer (CIO) / CDO, to risk managers, to data administrators, to data custodians. The data governance board manages overall data of an organization and address data availability, usability, integrity and security. They define procedures & definitions of how to deal with data situations and establish a plan for these procedures. Data administrators or data stewards have administrative control over and are accountable for organization's information. Data custodian is responsible for maintaining documentation, controlling access over data and auditing DQ rules. CIOs are involved in setting goals for the organization, approval of policies and risk management.

WORKS CITED:

Lecture discussions

Lecture Videos