IST722: Class Exercise 1

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 Deepak Shah
Your Email:	bhshah@syr.edu

Instructions (Refer Unit 1)

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] Why is data in organizations the more important asset?

Ans – Data is a more important asset because it enables organizations to gain valuable insights, make informed decisions, and drive innovation, leading to improved efficiency, competitive advantage, and better customer experiences. Asset is about adding money and value to the organization and hence data is a more important asset[1].

[2] What are the 4 characteristics of a data warehouse? Provide your own novel examples of each.

Ans – The four characteristics of data warehouse are:
a) SUBJECT ORIENTED: Subject-oriented in a data warehouse means organizing entities around business processes for a comprehensive view. For instance, in a data warehouse focusing on orders, products, and shipping, the emphasis would be on understanding products within the context of orders and shipments rather than solely analyzing products independently.

b) TIME VARIANT: A data warehouse stores historical data and enables analysis over time. It captures and maintains a record of changes and trends, allowing users to analyze data at different points in time. Example, a change in customer details like address etc in a system can be analyzed with the change in time as when the change was executed.

c) INTEGRATED: This means all the data is centralized in one place where everything is visible and can be retrieved. Example, a website customer and a marketing OLTP customer will be centralized as data warehouse customer.

d) NON-VOLATILE: This refers to the fact that the data is never removed or deleted. Even if there is any change in the data it is saved and added over time. For example, if an order is created for a product and then if the customer returns the product. The order row will not be deleted and a new row for reverse/return order will be generated[4].

[3] You probably noticed we made a "copy" of the data from the source system to the data warehouse. Can you think of three reasons why the data must be a copy and you cannot just use the original data?

Ans - Three reasons why a copy of the data is made from the source system to the data warehouse are:

a) To draw insights, data is collected from many sources for integration. As a result, data may be translated, indexed, and formatted in a way that promotes efficient reporting and analysis without affecting the performance of the source system by making a copy.

b) Historical data may be collected/retrieved and standardized by producing a copy to maintain consistency and coherence across multiple systems.

c) The data warehouse copy enables enterprises to establish access rules, adopt data governance principles, and safeguard sensitive information[2][4].

[4] What is the difference between business intelligence and data warehouse?

Ans – Data warehousing is a foundation through which we base Business intelligence. Business intelligence (BI) is the process of using tools and techniques to analyze data and make informed decisions. A data warehouse is a central repository that stores integrated historical data from various sources, providing accurate information for analysis and reporting. A data warehouse is a key component of a business intelligence architecture[2][4].

[5] How do Inmon and Kimball approaches to Data Warehousing differ? Be brief.

Ans – Inmon's approach differs from Kimball's in terms of their perspective on data and business processes. Inmon's approach can be considered a bottom-up approach as it focuses on data first and then incorporates business processes. On the other hand, Kimball's approach considers the required business processes and then designs the data warehouse accordingly.
Kimball's Data Warehouse includes data marts within the data warehouse itself to provide business intelligence. In contrast, Inmon's Data Warehouse normalizes the data first and sources the data marts from outside the data warehouse to support business intelligence.
In summary, Inmon's approach aims to create a centralized, integrated data warehouse as the primary data source, while Kimball's approach emphasizes dimensional modeling for optimized reporting and analysis. [3][4].

WORKS CITED:

[1] King, D. (2021, July 27). *Data is an Asset*. Association Analytics. https://associationanalytics.com/blog/data-is-an-asset/

[2] Inmon, W. H., & Linstedt, D. (2014). Data Architecture: A Primer for the Data Scientist

[3] Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling

[4] Video lectures on Introduction to Data Warehousing by Prof. Michael Fudge and Class Discussions with Prof. Humayu Khan.