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: | Jaya Varshini Prabakar |
| Your Email: | jprabaka@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?

Data is a more important asset for business because it increases profits, reduces costs, helps business to know its customers and understand their needs.

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

The four characteristics of Data Warehouse are:

* **Subject Oriented**: The data stored is **domain specific**. E.g Students, Grades, Subjects, Professors, Departments, these entities come under **university information** on students and Professors.
* **Non** – **Volatile:** The data can’t be changed over time. E.g My example for this is ROM(Read only Memory) in computers, the data is still there even after the system is turned off. Likewise, the previous data will be present in the Data Warehouse when the new data is added.
* **Integrated**: Combines data from various sources. E.g Collecting data from **Walmart’s** online sales platform and Customer Relationship Management Systems and integrating into a centralized Data Warehouse repository.
* **Time-Variant:** Data is represented as it is at that time, E.g Due to non-volatile property we have historical data, insights like **periodic snapshots** like profit margins, customer count and sales revenue in each quarter or annually can be obtained.

[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?

The 3 reasons for storing data copy instead of using original data:

1. **Data Integration:** Data Warehouse integrates data from multiple sources. Thus, providing a common unified format is achieved.
2. **Historical Analysis:** By storing a copy we aren’t disturbing the original source which may undergo updates and modifications. And this also creates an independent copy of data that allows for historical analysis.
3. **Making it Domain Specific:** The structure of data in original is not suitable for analysis and querying based on data- warehouse based structure like star schema.

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

Business intelligence involves analytical decision making, presentation of actional information. Data Warehouse is a centralized infrastructure that stores historical and integrated data for BI analysis. We can say that Data warehouse is the foundation for Business intelligence.

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

The main difference in Inmon and Kimball approach can be seen in their architecture.

|  |  |  |
| --- | --- | --- |
|  | **Architecture** | **Modelling** |
| **Inmon** | Top-down approach | Relational Modeling (normalized data) |
| **Kimball** | Bottom-up Approach | Dimensional Modelling (Star Schema) |

WORKS CITED:

Professor Michael A Fudge. Introduction to Data Warehousing [Video file].