A - Introduction to Data Warehousing

1. Why is data in organizations the more important asset? Video 1.3.1
   1. Answer: Without data you can’t identify customer, their needs. Data helps to identify what products you should and should not put for sale. Data helps to identify expenses and profitability. Each organization level has different requirements for data. Starts with Operational Database (OLTP), online transaction processing system. Stored in rational databases or files, data is stored as efficiently as possible (highly normalized), optimized for speed and to get current required information, designed to capture data not report it. Designed to support the operational need of the organization. Transactional databases are complex. Issues with transactional databases: Difficult and time consuming with joins and sub joins, resource-intensive and impossible to query data you no longer have.
2. What are the 4 characteristics of a data warehouse? Video 1.3.3
   1. Answer:
      1. Subject Orientation - Data is domain specific. Built around business entities/processes. Used by all functional areas.
      2. Non-volatility – Data are never removed or changed, data always grows with time.
      3. Integrated – Centralized in one place, holds the entire organization data. Single version of the truth.
      4. Time-variant – Flow of data through time. Data reflected as a point in time.
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? Video 1.5.1
   1. Answer: In Case errors occurs, keeping a backup enables us to go back to our original data. Provides security and saves up cost and time, helps to maintain consistent copies of data after major updates to the data. Different versions of the data can be used by multiple users at any point in time.
4. What is the difference between business intelligence and data warehouse? Video 1.4.1
   1. Answer:
      1. Business intelligence – Analytical and decision-support capabilities of the data warehouse. Informed decision-making, the visual or front-end that the users see.
      2. Data warehousing – as a foundation, by which we base the business intelligence.
5. How do Inmon and Kimball approaches to Data Warehousing differ? Videos 1.7.1, 3.3.2
   1. Answer:

|  |  |  |
| --- | --- | --- |
|  | Inmon | Kimball |
| Father of | Data warehouse | Business intelligence |
| Invented | Data warehouse | Dimensional models |
| Data warehouse is | Normalized tables | Dimensional models |
| Purpose of data warehouse | Data integration | Query |
| Approach: | Data-first (iterative, bottom-up). Relational modelling. Entity-relational model. Data marts – Not believed to be part of the data warehouse.  Inmon advocates the hub-and-spoke architecture (e.g., the Corporate Information Factory), while | Process-first (waterfall, top-down). Dimensional modelling. Fact tables and dimensions. Data marts – believes they are part of the data warehouse).  Kimball promotes the data mart bus architecture with conformed dimensions. |

1. Types of analytics? Video 1.4.2
   1. Answer: Business Intelligence:
      1. Retrospective: Traditional business intelligence (What happened?) Hindsight
      2. Diagnostic: Analytic dashboard (Why did it happen?)
      3. Descriptive: Real-time dashboard (What is happening now?) Insight
   2. Business Analytics:
      1. Predictive: Machine learning/forecasting (What is likely to happen?) Foresight
      2. Prescriptive: Decide or act (What should I do about it?)
2. Tables – Start with either Dim or Fact.
   1. Dim tables represent – Dimension tables. Business entities/Categories.
   2. Fact tables – Business process.