Name: Polasa Shobharani

Student ID: S1365603

Reading Check 4

Chapter 5

15points

**Instructions:** Please **highlight the answer(s)** that you are choosing. Once completed, please submit your highlighted version back into the assignment task. Ensure you **replace the name and student ID at the top of this page**. Please reach out to me if you have trouble uploading your assignment.

1. What is the term that are the logical mappings between the physical data stored in the technology stack, along with the business terminology (1 points)

These were Managed by individual Vendors through unique **Semantic layers:** logicalmappings between the physical data stored in the technology stack, along with business terminology.

1. List and define the **4 Common Approaches to Data Modeling** (8 points)

Common Approaches to Data Modeling: The analytical dara landscape today predominantly uses four main patterns of data modeling.

**Normalized Modeling:** As discussed in previous chapters, source systems such ERPs might consist of thousands of meticulously defined transactional tables. ACorporate Information Factory (CIF), attributed to Bil Inmon and Claudia Imhoff, mirrors this complexity, necessitar- ing substantial data joins (i.e., combinations) to arrive at the final analytical result.'

To provide more accessible structures to end-users, Inmon suggested the creation of data marts derived from this centralized data warehouse. Data marts might exist separately for finance, marketing, sales, and so on, designed to cater to the specific needs of business units or departments within an organization.

The data within these marts undergoes denormalization or restructuring into a dimensional model (like a scar schema or a snowflake schema). This process optimizes the readability and performance of the data based on the particular business units needs. As a result, denormalization simplifies the risk for business users when querying the data or generating reports.

**Dimensional (Denormalized) Modeling:** Another widely employed approach to data modeling for analytics dimensional modeling, also known as the Kimball Method. The brainchild of Ralph Kimball, Margy Ross, and other data warehousing experts in the 1990s, this method was defined by its bottom-up approach, as opposed to Inmon's top-down CIF.3 The Kimball Method centers around a "star schema" in which dimension tables surround a central fact table. The fact table holds the measurable, quantitative data (facts), while the dimension tables contain descriptive attributes related to the facts. Owing to its simplicity, the star schema design is more intuitive and straightforward for end-users to comprehend and navigate. This clarity also simplifies writing queries and generating through indexing and partitioning.

its most granular, transactional level. However, aggregated star schemas are possible, although they come with the compromise of sacrificing detail for improved performance.

, which maintain consistency across different star schema

**Data Vault Modeling:** The data vault modeling approach has been gaining traction since the 2010s and is favored for its flexibility, scalability, and

adaptability.

The model consists of three main components that allow the data vault model to handle complex, interconnected datasets efficiently.

• Hubs: Representing business keys or identifiers.

• Links: Defining relationships between data elements.

• Satellites: Representing descriptive attributes of data.

The data vault model shines because of its agility and scalability and its design to accommodate swift changes in business requirements. This makes choice for dynamic, fast-pared scenarios that deal with large and diverse volumes of data. It also excels in data models that incorporate information from various systems by maintaining the representation of hte source systems' data and facilitating easier ongoing integration of new source.

**One Big Table (OBT) Modeling:** One big table (OBT) is a newer approach that involves the creation of expansive denormalized models. This method offers the advantage of speed and ease in initial setup, but it could lead to high computational costs and data redundancy.

nI the Modern Data Stack context-where storage costs are a fraction of compute costs-the main argument for OBT lies in its cost-effectiveness, performance, and shorter implementation time? Evidence suggests that wide tables are 25-50 percent faster than star schema queries, primarily because they eliminate the need for joins.

1. Within the Data Vault Modeling, what are the **3 components** that allow the data vault to handle complex datasets efficiently? (3 points)

The data vault modeling approach has been gaining traction since the 2010s and is favored for its flexibility, scalability, and adaptability.

The model consists of three main components that allow the data vault model to handle complex, interconnected datasets efficiently:

* **Hubs**: Representing business keys or identifiers.
* **Links**: Defining relationships between data elements.
* **Satellites**: Representing descriptive attributes of data.

1. List and define the **3 main components of dbt** (Data Build Tool) (3 points)

The tool consists of three main components:

* **Development**: Where users craft transformations in SQL and implement business logic.
* **Testing**: Where users conduct tests to ensure the transformed data is formatted correctly.
* **Documentation**: Where documentation for the transformations is produced.