Dimensional Modelling Contd.

2023-July-05

CSCI 32092 – Data Mining and Data Warehousing

Indrajith Ekanayake

Contents

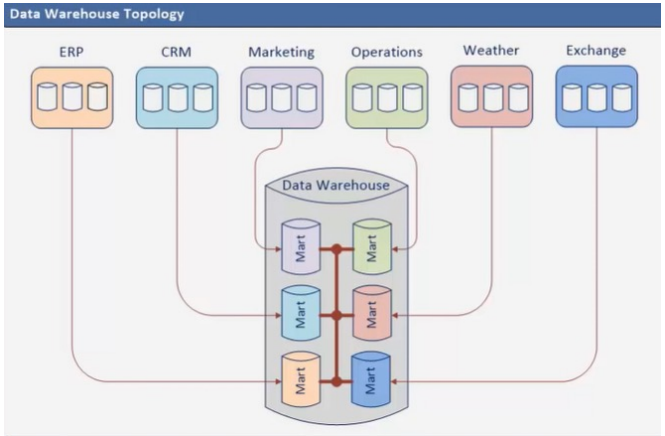
Bus Matrix

Introduced in Kimball Datawarehouse model and this section we will look how to create a bus matrix for given scenario.

Bus Architecture

- Way of taking individual data marts and combining them in a way that forms an integrated data warehouse.
- In an integrated data warehouse, we should be able to pull data from multiple data marts and utilize that data in integrated way to perform analytics.

Data warehouse topology:
There are individual subject areas on the top, each feeding data into individual Datamarts. These individual Datamart plugged into a bus which forms the Data Warehouse



Source: https://www.leapfrogbi.com/

Bus Matrix

- The rows of the bus matrix correspond to data marts. The columns of the bus matrix correspond to dimensions in each data mart.
- You should create separate matrix rows if the sources are different, the processes are different, or if the matrix row represents more than what can reasonably be tackled in a single implementation iteration.

This simple diagram helps in communicating immensely. When you have a Bus Matrix, you can easily recognize the reuse of certain dimensions. Ex: In the given example the Date and Product dimensions reused in all data marts

	COMMON DIMENSIONS							
BUSINESS PROCESSES	Date	Product	Store	Promotion	Warehouse	Vendor	Contract	Shipper
Retail Sales	X	X	X	X				
Retail Inventory	X	X	X					
Retail Deliveries	X	X	X					
Warehouse Inventory	X	X			X	X		
Warehouse Deliveries	X	X			X	X		
Purchase Orders	X	X			X	X	X	X

Figure 3.8 Sample data warehouse bus matrix.

Source: Kimball book

Task 1:

Create Bus Matrix for the below Case scenario?

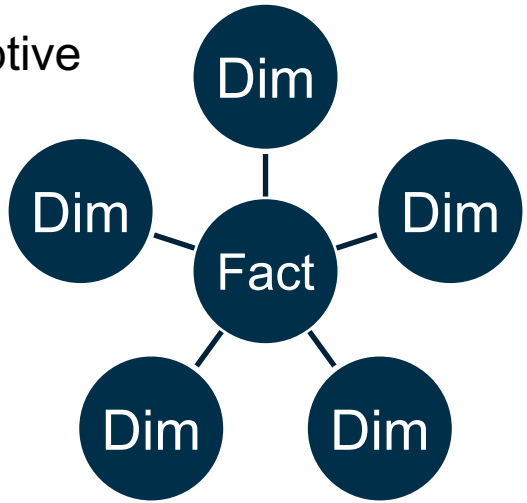
Case scenario 1:
Fashion store ABC has outlets all over the island. They have customers who are registered and earn points when they buy from the store. ABC also runs promotions during seasons where they offer a percentage discount to a particular product category.

Dimensional Modeling Techniques

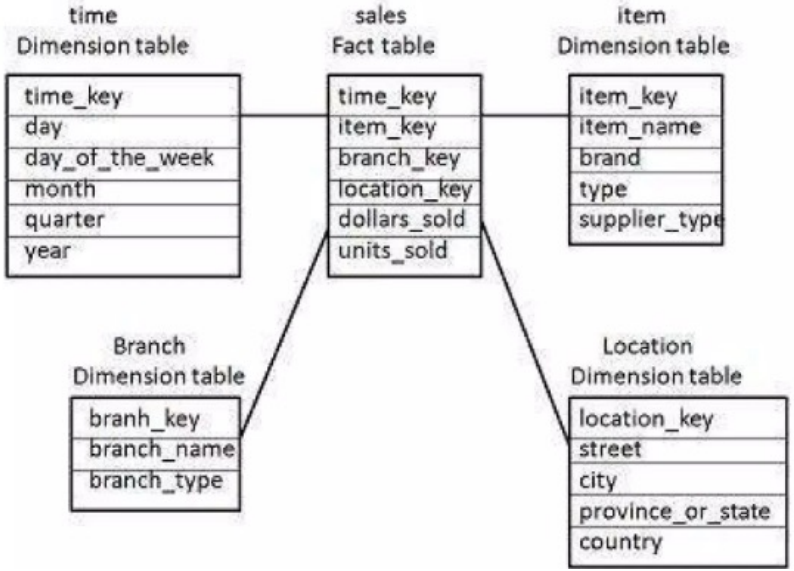
This section provides an overview to 3-dimensional modeling techniques

Star Schema

- This is the simplest data warehouse schema, and it is called such because the diagram resembles a star, with points radiating from a center.
- Singe data (fact) table surrounded by multiple descriptive (dimension) tables
- Facts represent the main Business Processes
- Dimensions represent the main entities involved.



Star Schema Cont.

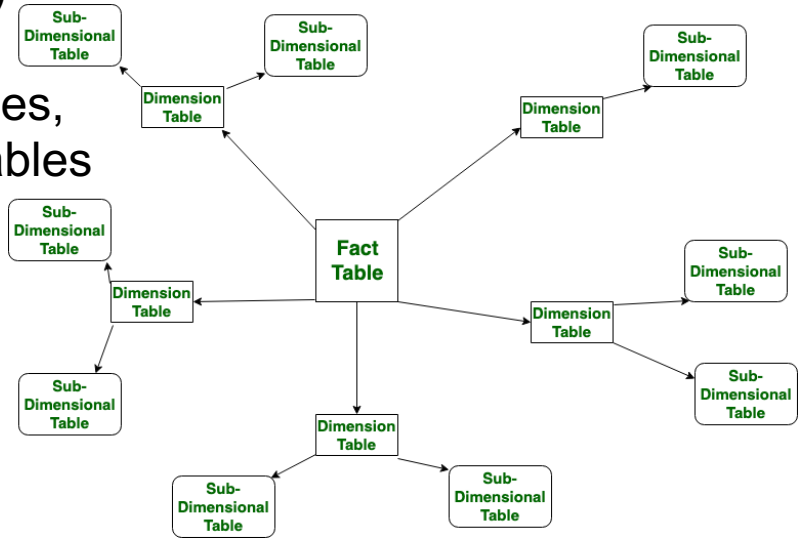


Source: https://www.mssqltips.com/sqlservertip/5690/create-a-star-schema-data-model-in-sql-server-using-the-microsoft-toolset/

Snowflake Schema

- The fact tables, dimension tables as well as sub dimension tables are contained. (dimension tables are normalized)

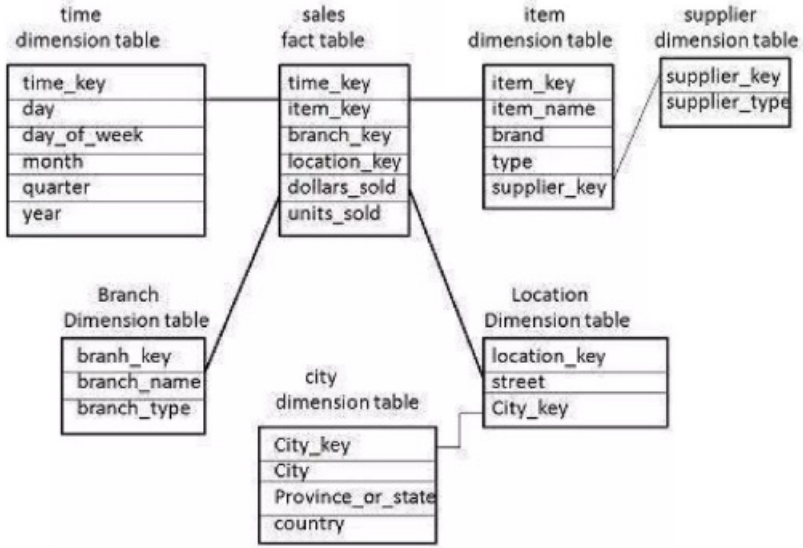
- This schema forms a snowflake with fact tables, dimension tables as well as sub-dimension tables



Source: Geeks for geeks

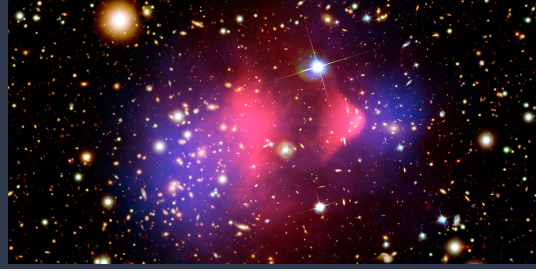
Snowflake Schema Cont.

Unlike star schema, the dimension table in snowflake schema are normalized. For example, item dimension table in star schema is normalized and split into two-dimension tables called item and supplier.

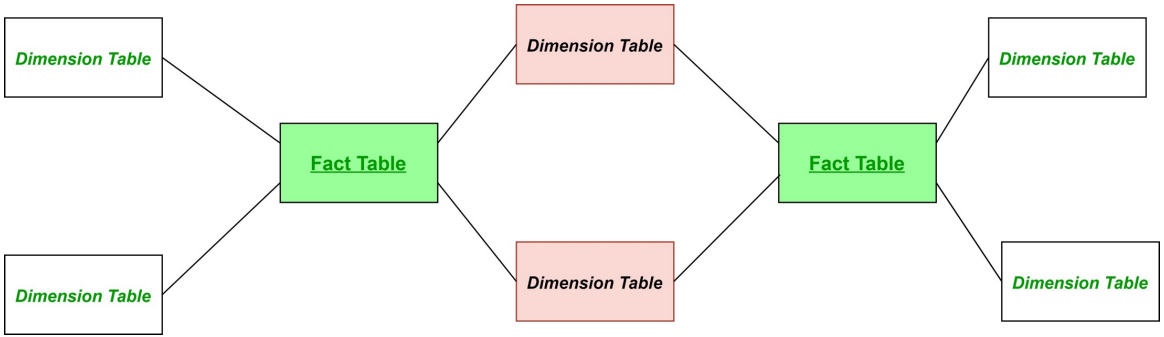


Source: https://www.mssqltips.com/sqlservertip/5690/create-a-star-schema-data-model-in-sql-server-using-the-microsoft-toolset/

Galaxy Schema/ Fact Constellation



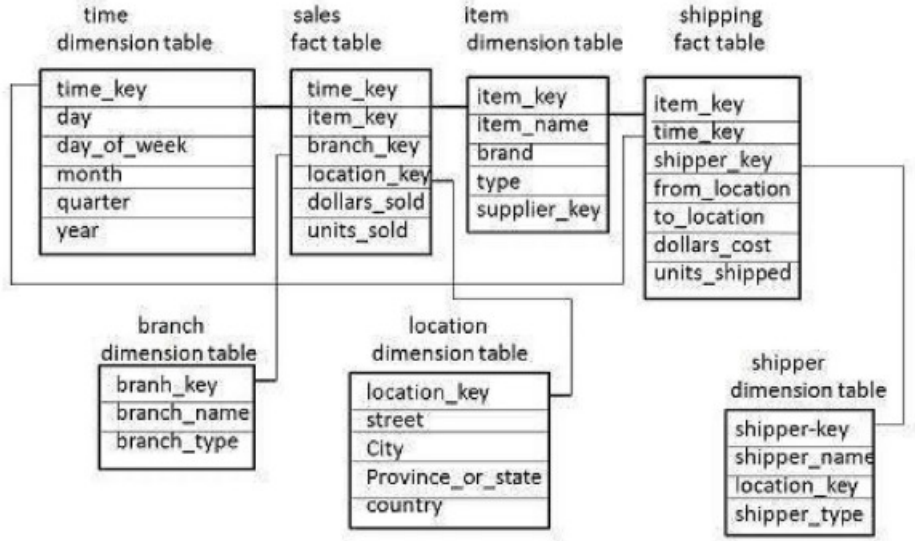
- It is a collection of multiple fact tables having some common dimension tables.
- It can be viewed as a collection of several star schemas and hence, also known as Galaxy schema.
- It is one of the widely used schema for Data warehouse designing and it is much more complex than star and snowflake schema.



Source: Geeks for geeks

Fact Constellation Cont.

Time, item and location dimensions are shared between sales and shipping fact tables.



Source: https://www.mssqltips.com/sqlservertip/5690/create-a-star-schema-data-model-in-sql-server-using-the-microsoft-toolset/

The 4-Step Design Process

This section explains the 4-step design process for dimensional modeling

The 4-Step Design Process

1. Choose the Business Process/ Choose the data mart
2. Identify the grain
3. Identify the dimension
4. Identify the fact

1.Choose the Business Process

- A business process is a low-level activity performed by an organization, such as taking orders, invoicing, receiving payments, handling service calls, registering students, performing a medical procedure, or processing claims.

- Listen carefully to the business to identify the organization's business processes because business users can't readily answer the question. The performance measurements users want to analyze in the DW/BI system result from business process events.

2. Identify the grain

- Preferably you should develop dimensional models for the most atomic information
- captured by a business process. Atomic data is the most detailed information collected;
- such data cannot be subdivided further.
- A data warehouse almost always demands data expressed at the lowest possible
- grain of each dimension not because queries want to see individual low-level rows,
- but because queries need to cut through the details in very precise ways.

3. Identify the dimension

- You need to decorate fact tables with a robust set of dimensions representing all possible descriptions that take on single values in the context of each measurement.
- If you are clear about the grain, the dimensions typically can easily be identify end as they represent the “who, what, where, when, why, and how” associated with the event.
- Examples of common dimensions include date, product, customer, employee, and facility.
- With the choice of each dimension, you then list all the discrete, text-like attributes that flesh out each dimension table.

4. Identify the fact

- Facts are determined by answering the question, “What is the process measuring? “
- Business users are keenly interested in analyzing these performance metrics.
- All candidate facts in a design must be true to the grain defined in step 2. Facts that clearly belong to a different grain must be in a separate fact table.
- Typical facts are numeric additive figures, such as quantity ordered or rand cost amount.

Task 2:

Case scenario 2:
Fashion store ABC has outlets all over the island. They have customers who are registered and earn points when they buy from the store. ABC also runs promotions during seasons where they offer a percentage discount to a particular product category.

ABC company has stores all over the island to help to supply goods to the outlets. They want to find out their stock levels for each product in each of their locations.

Create a star schema for this?

Some additional resources:

- Dimensional Modeling Techniques: https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/
- Create a Star Schema Data Model in SQL Server using the Microsoft Toolset: https://www.mssqltips.com/sqlservertip/5690/create-a-star-schema-data-model-in-sql-server-using-the-microsoft-toolset/
- Four-Step Dimensional Design Process: https://www.linkedin.com/pulse/four-step-dimensional-design-process-brian-moodley/
- Dimensional Modeling types: https://hevodata.com/learn/dimensional-data-modelling/

If you have any questions message me on LinkedIn