|  |
| --- |
|  |
| Star Schema Basics |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| REVISION HISTORY | | | | | |
| Ver. | Description of Change | Author | Date | Approved | |
| Name | Effective Date |
| 1.0 | Initial status | Valeryia\_Lupanava | 08-NOV-2017 |  |  |

Contents

[1. Star Schema 3](#_Toc497924179)

[1.1. Step 1: Select the Business Process 3](#_Toc497924180)

[1.2. Step 2: Declare the Grain 3](#_Toc497924181)

[1.3. Step 3: Identify the Dimensions 3](#_Toc497924182)

[1.4. Step 4: Identify the Facts 4](#_Toc497924183)

[1.5. Step 5: Star Schema 6](#_Toc497924184)

[2. Snowflake Schema 7](#_Toc497924185)

[2.1. Step 1: Select the Business Process 7](#_Toc497924186)

[2.2. Step 2: Declare the Grain 7](#_Toc497924187)

[2.3. Step 3: Identify the Dimensions 7](#_Toc497924188)

[2.4. Step 4: Identify the Facts 9](#_Toc497924189)

[2.5. Step 5: Snowflake Schema 10](#_Toc497924190)

[3. Conclusion 11](#_Toc497924191)

# Star Schema

## Step 1: Select the Business Process

Business Process for analysis is company’s sales per different metrics.

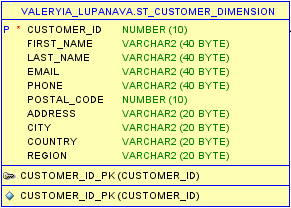
## Step 2: Declare the Grain

Fact granularity is Daily Sales Amount per certain Customer and Employee in a specific location.

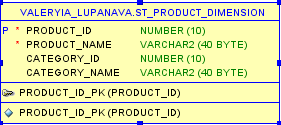
## Step 3: Identify the Dimensions

Schema should contains next dimensions:

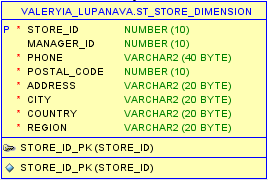
1. Customer dimension (Customer ID is Primary Key);



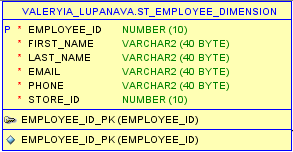
1. Product dimension (Product ID is Primary Key);



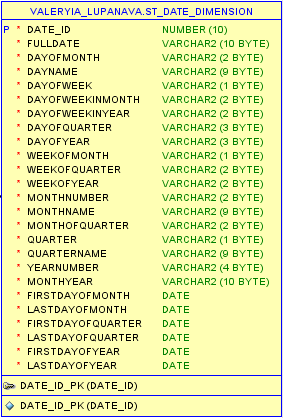
1. Store dimension (Store ID is Primary Key);



1. Employee dimension (Employee ID is Primary Key);



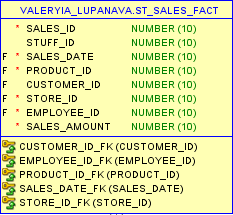
1. Date dimension (Date ID is Primary Key).



## Step 4: Identify the Facts

Schema should contains the fact table “Sales” with next data:

1. Sales ID;
2. Stuff ID;
3. Sales Date as Foreign Key from Date Dimension;
4. Product ID as Foreign Key from Product Dimension;
5. Customer ID as Foreign Key from Customer Dimension;
6. Store ID as Foreign Key from Store Dimension;
7. Employee ID as Foreign Key from Employee Dimension;
8. Sales Amount as measure of Sales.



## C:\Users\Valeryia\Desktop\Topic 06 - Star Schema Basics\Star_schema.jpg Step 5: Star Schema

# Snowflake Schema

## Step 1: Select the Business Process

Business Process for analysis is company’s sales per different metrics.

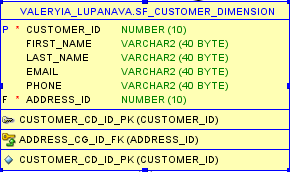
## Step 2: Declare the Grain

Fact granularity is Daily Sales Amount per certain Customer and Employee in a specific location.

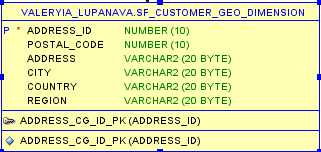
## Step 3: Identify the Dimensions

Schema should contains next dimensions:

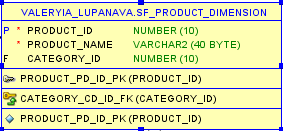
1. Customer dimension (Customer ID is Primary Key);



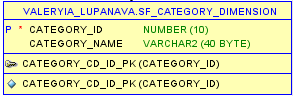
1. Customer geo dimension (Parent table is Customer Dimension; Address ID is Primary Key);



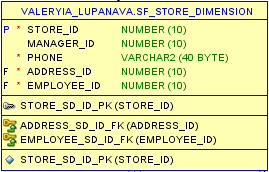
1. Product dimension (Product ID is Primary Key);



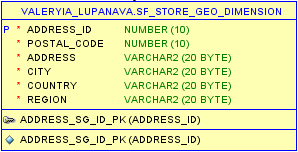
1. Category dimension (Parent table is Product Dimension; Category ID is Primary Key);



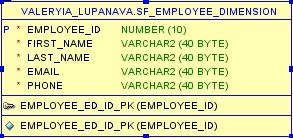
1. Store dimension (Store ID is Primary Key);



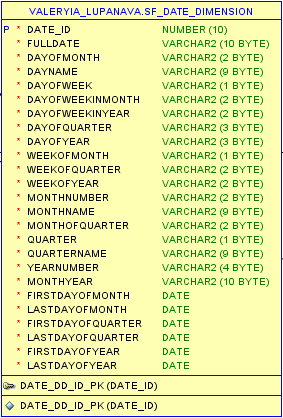
1. Store geo dimension (Parent table is Store Dimension; Address ID is Primary Key);



1. Employee dimension (Parent table is Store Dimension; Employee ID is Primary Key);



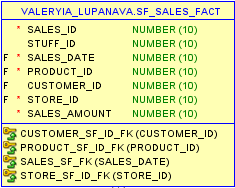
1. Date dimension.



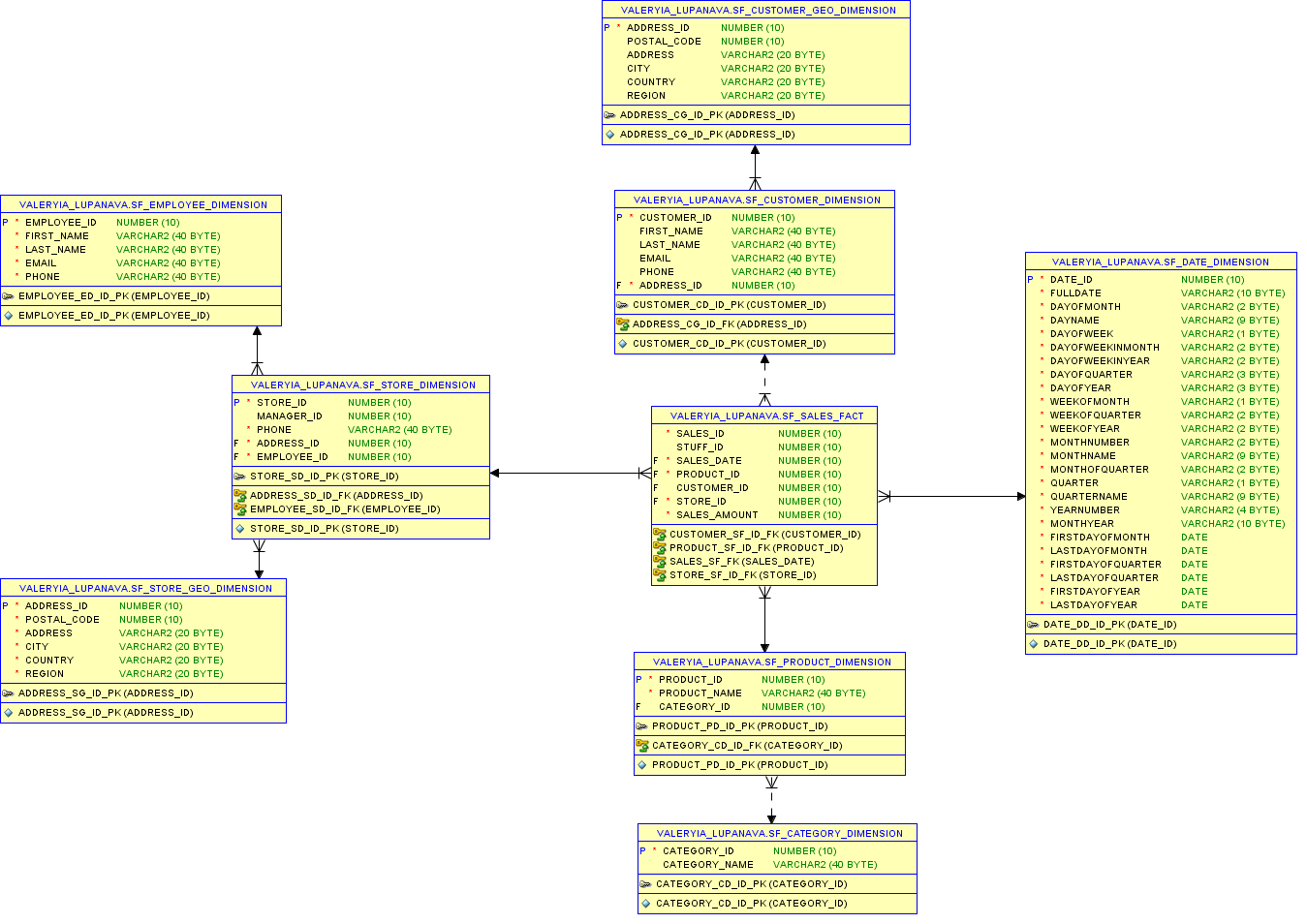
## Step 4: Identify the Facts

Schema should contains the fact table “Sales” with next data:

1. Sales ID;
2. Stuff ID;
3. Sales Date as Foreign Key from Date Dimension;
4. Product ID as Foreign Key from Product Dimension;
5. Customer ID as Foreign Key from Customer Dimension;
6. Store ID as Foreign Key from Store Dimension;
7. Employee ID as Foreign Key from Employee Dimension;
8. Sales Amount as measure of Sales.



## Step 5: Snowflake Schema



# Conclusion

To begin with the granularity of the data is determined by the lowest level of granularity of each dimension table. For example, a fact table may store weekly or monthly totals for specific products. The lower the granularity, the more records that will exist in the fact table. The granularity also determines how far users can drill down. At schemas, designed at this task, the granularity also could be lower or higher.

Star schemas are easy for end users and applications to understand and navigate. With a well-designed schema, users can quickly analyze large, multidimensional data sets. The main advantages of star schemas are:

1. Query performance: because a star schema database has a small number of tables and clear join paths, queries run faster;
2. Load performance and administration: structural simplicity also reduces the time required to load large batches of data into a star schema database;
3. Built-in referential integrity: a star schema has referential integrity built in when data is loaded. Referential integrity is enforced because each record in a dimension table has a unique primary key, and all keys in the fact tables are legitimate foreign keys drawn from the dimension tables;
4. Easily understood: a star schema is easy to understand and navigate, with dimensions joined only through the fact table.