Introduction to Data Warehouse

Introduction

Data Warehouse basics

Benefits over Transactional Database

Dimensional Modeling
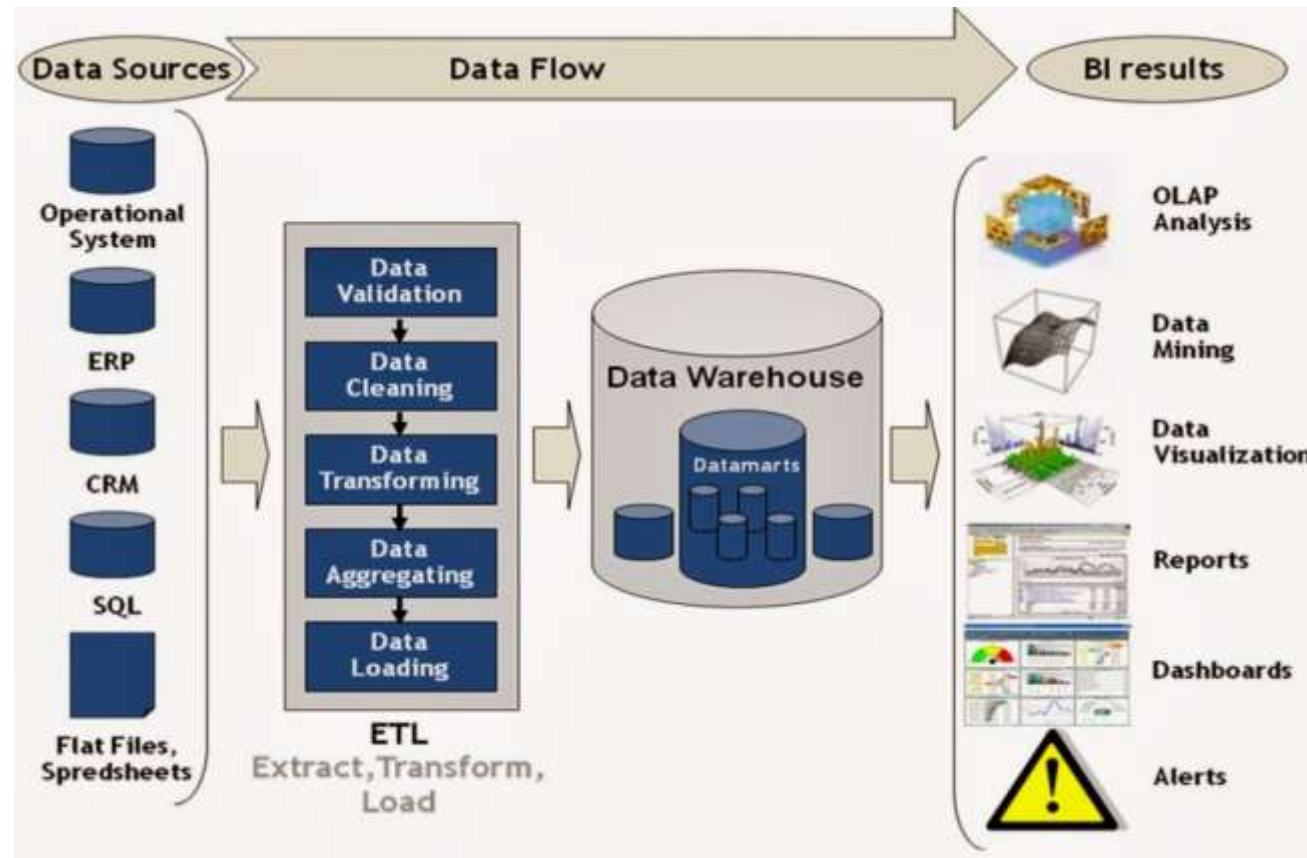
Facts and Fact Table

Dimensions and Dimension Table

Star vs Snowflake Schema

Why we need Data warehouse?

What is Data Warehouse

Concerns About a Data Warehouse

Extra piece of software that needs maintenance

The data is already present in the operational systems

However, extracting information from operational data is complicated

arif.works

Problems a Data Warehouse Can Solve

We have so much data, but we can't make anything of it
I only want to know what is important
We need to slice and dice the data
Business people need to access data easily
Numbers between departments don't match and we don't know who's right
We want people to make decisions based on facts

arif.works

Ideal Data Warehouse Solution
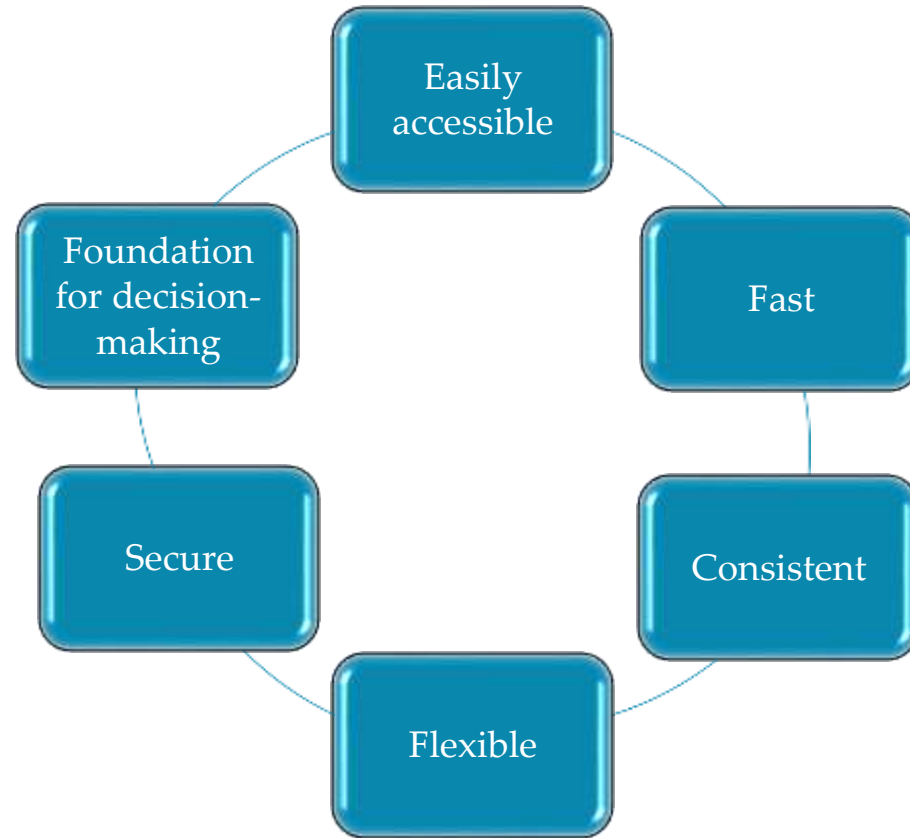
Responsibilities of a Data Warehouse Designer

| Understand the business users goal and objectives | Deliver accurate, trustworthy and relevant information | Sustain the DW environment |

OLTP vs OLAP

OLTP	OLAP
Online Transactional Processing	Online Analytical Processing
• Many small transaction	• Low volume but complex queries
• Current data	• Historic, non-volatile data
• Used to run the business	• Used to analyze the business
• Highly detailed	• Consolidated and summarized
• Typically in the GB scale	• TB and above scale
• Processing performance limit	• No limit, pause/resume compute

arif.works

Dimensional Modeling

Database design method optimized for data warehouse solutions

Popular technique because it addresses two important requirements

    1. Deliver data in an understandable format

    2. Deliver fast query performance

Key word is "simplicity"

Sustain the DW environment

Elements of Dimensional Model

Facts	• The measurements or metrics or facts from your business process
Dimensions	• For providing the context of a business process event
Attributes	• The various characteristics of a dimension
Star Schema	• And/or OLAP cubes

arif.works

"We sell Cake and other Products in various locations and measure our achievements over time."

Facts and Fact Tables

Fact = a business measure
- Sales
- Profit
- Volume
- Number of transactions

Fact table = table that stores the performance measurements resulting from an organization's business process events

Facts and Fact Tables

Facts answer questions like:

    What are we doing? (sell, buy, count)

    What do we want to achieve? (more sales, bigger profit)

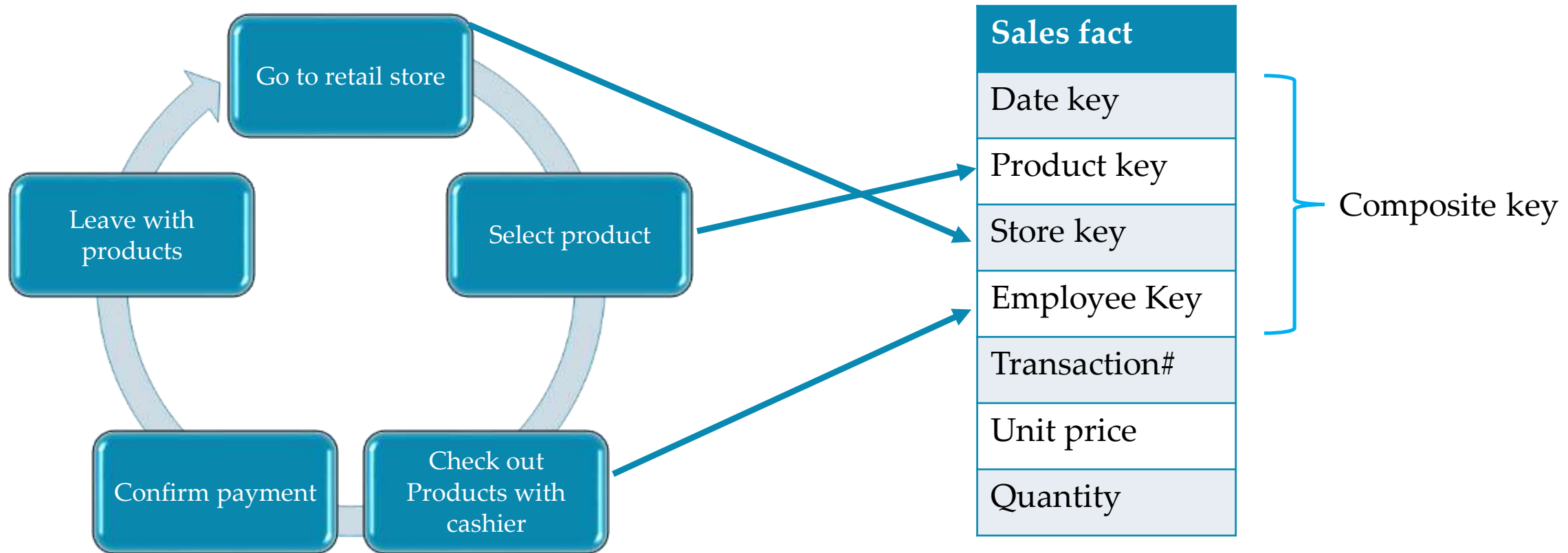1 row in the fact table is 1 measurement in real life

Fact columns in a fact table should be additive
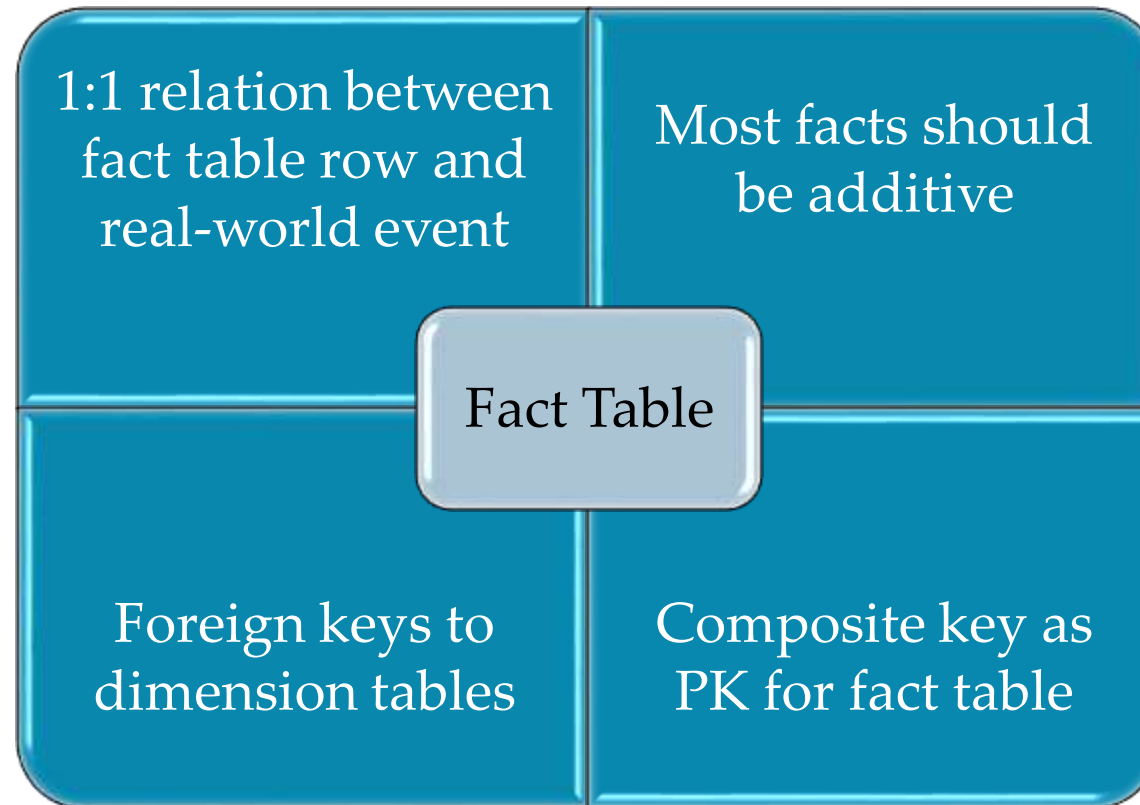
Facts make sense in combination with dimensions

    Linked with foreign keys

    Dat/Time dimension is present in most data warehouses

Example of a Fact Table
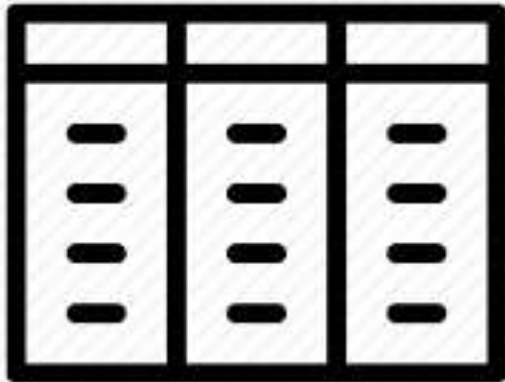
Characteristics of Fact Tables

1:1 relation between fact table row and real-world event

Most facts should be additive

Fact Table

Foreign keys to dimension tables

Composite key as PK for fact table

What are Dimensions?

Companions to a fact table

Textual context associated with a business

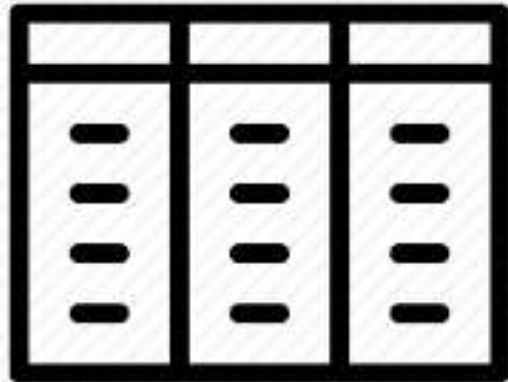process measurement event

Questions Answered by Dimension Tables

Example of a Dimension Table

Sales fact
Date key
Product key
Store key
Employee Key
Transaction#
Unit price
Quantity

Product Dimension
Product key
Product name
Brand name
Category name
Subcategory name
Package type
Package size
Weight
Weight unit of measure

Process cycle:
- Go to retail store
- Select product
- Check out Products with cashier
- Confirm payment
- Leave with products

arif.works

Characteristics of Dimension Tables

No limit for the number of attribute in a dimension table

Common to have tables with 50 to 100 attributes

Some dimension tables have only a handful of attributes

Have fewer rows than fact tables

But can be much wider

Defined by a single primary key

Basis for referential integrity with the fact table

Denormalized

Flattened many-to-one relationships within a single dimension table

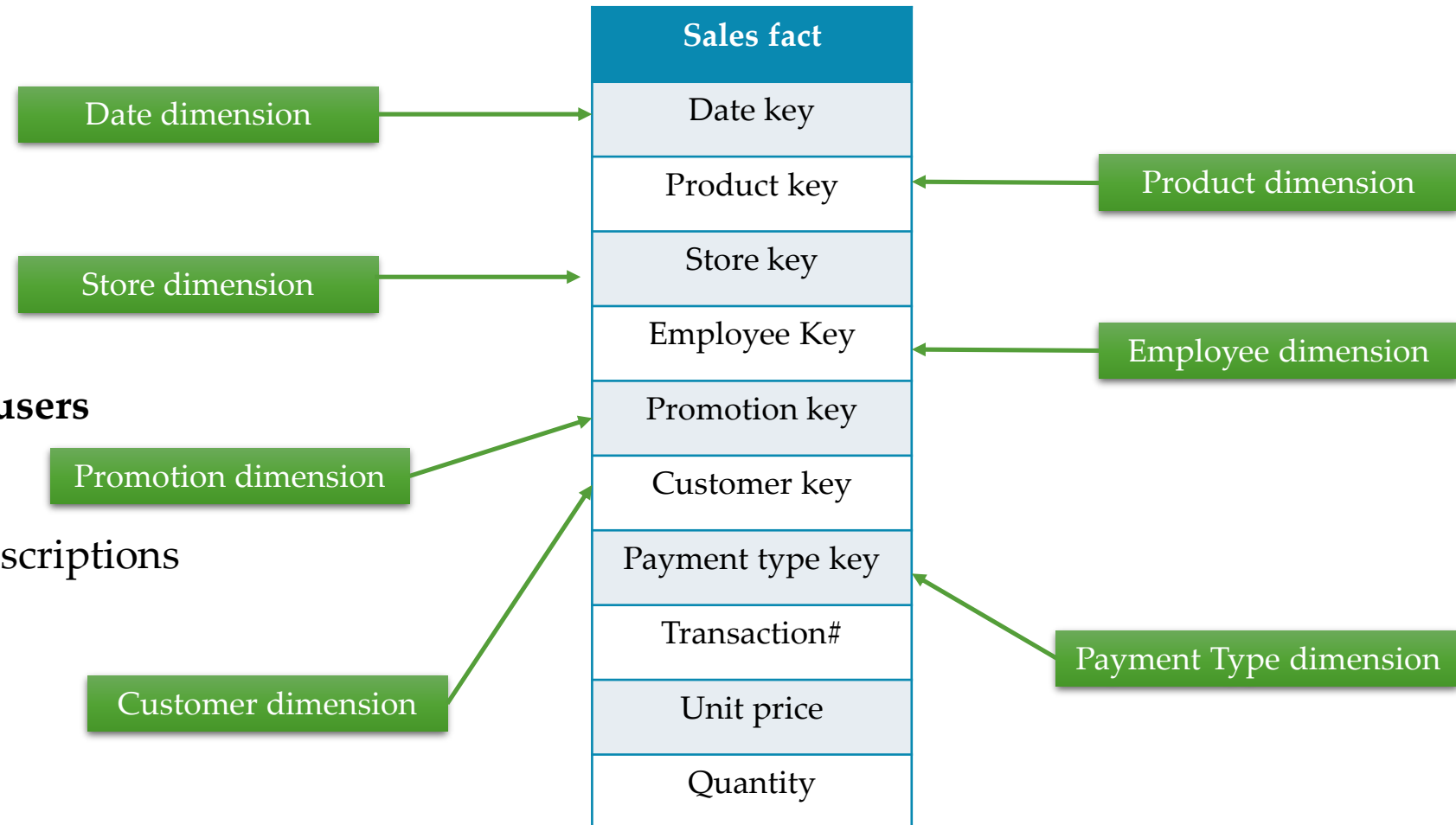arif.works

Example of a Dimension Table

Characteristics

   Simple

   Symmetric

Highly-recognizable by business users

   Reduced number of tables

   Use of meaningful business descriptions

Extensible

Sales fact
Date key
Product key
Store key
Employee Key
Promotion key
Customer key
Payment type key
Transaction#
Unit price
Quantity

Date dimension → Date key

Product dimension → Product key

Store dimension → Store key

Employee dimension → Employee Key

Promotion dimension → Promotion key

Customer dimension → Customer key

Payment Type dimension → Payment type key

arif.works
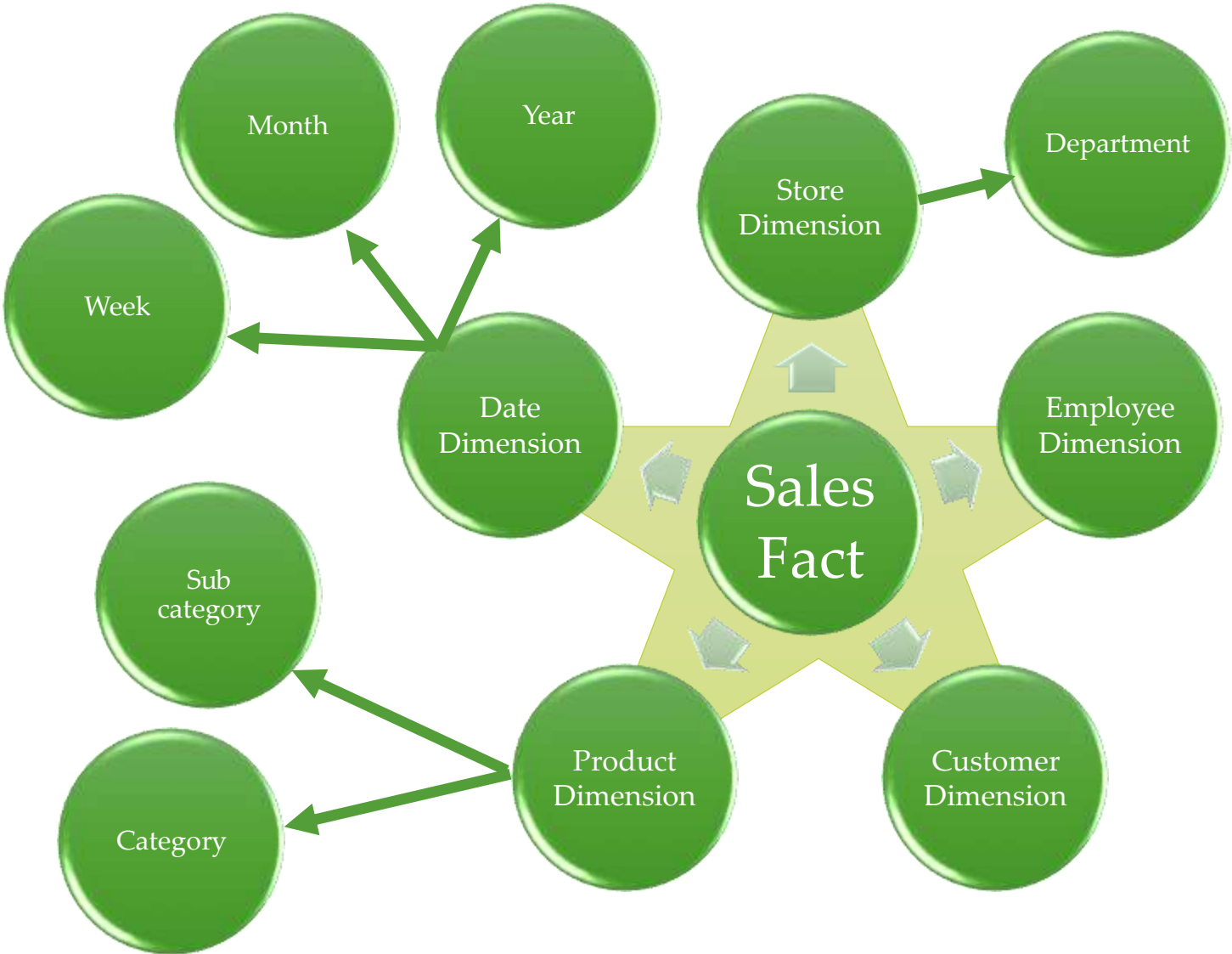
Dimensional Model as a Star Schema

Dimensional Model as a Snow flake

Summary

What is Data Warehouse

Why we need Data Warehouse

Data Warehouse (OLAP) vs Transactional Database (OLTP)

Dimensional Modeling

Facts and Fact Tables

Dimensions and Dimension tables

Star vs Snowflake Schema