

# Concepts: Data Warehouse



**Eric Tremblay**  
**Oracle Specialist**  
[www.data-warehouse.ca](http://www.data-warehouse.ca)  
[eric.tremblay@data-warehouse.ca](mailto:eric.tremblay@data-warehouse.ca)

# Objective

**Describes the main steps in the design of a data warehouse. Presents techniques for its use and challenges in its development.**

# Definition

# Definition

**« A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process »**

**--- Bill Inmon**

# The Goals of a Data Warehouse

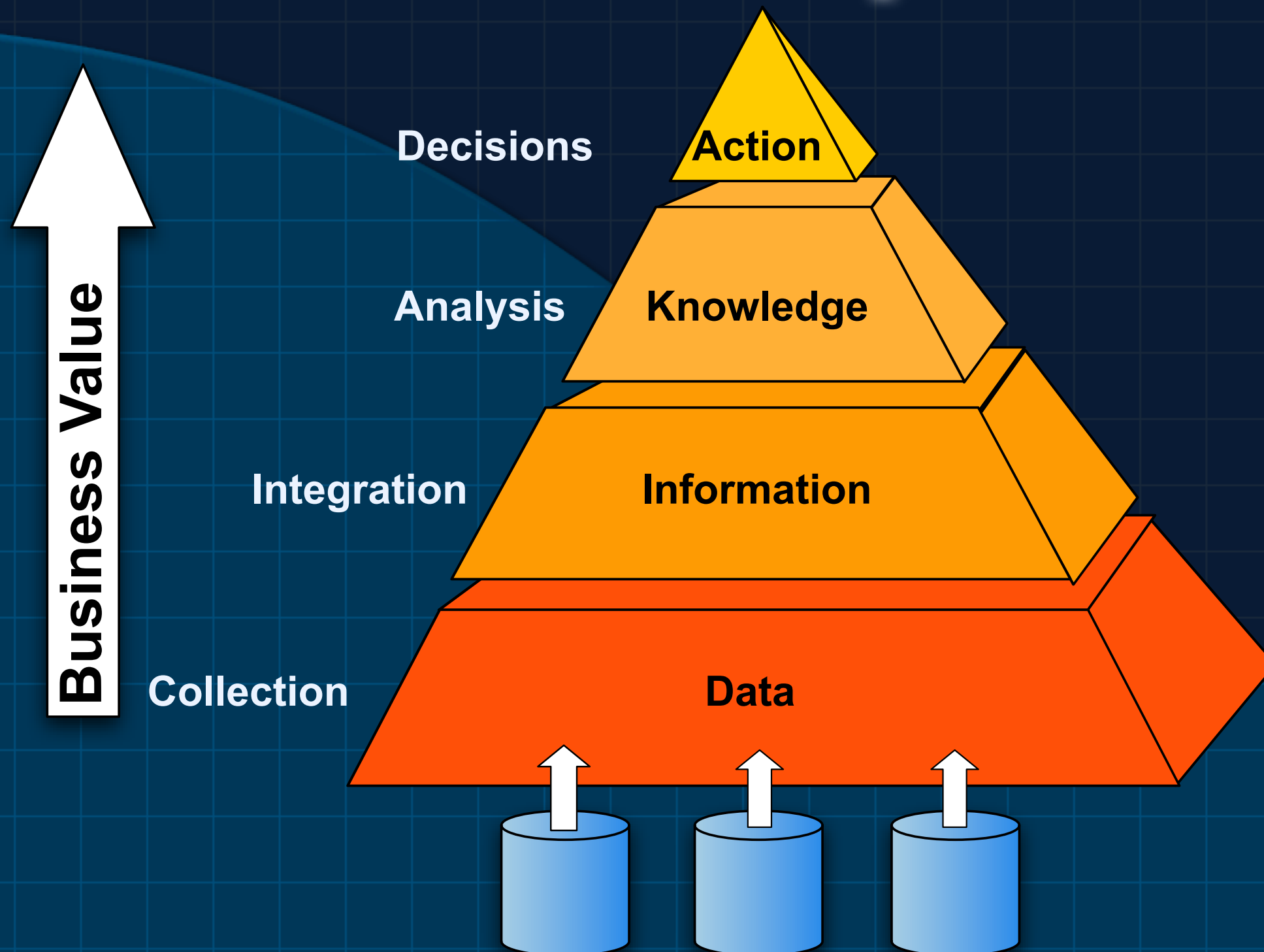
# The Goals of a Data Warehouse

- Access to company information
- Coherent company information
- A consistent, total and unified sight the company data
- The data published is stored for fast consultation

# The Goals of a Data Warehouse

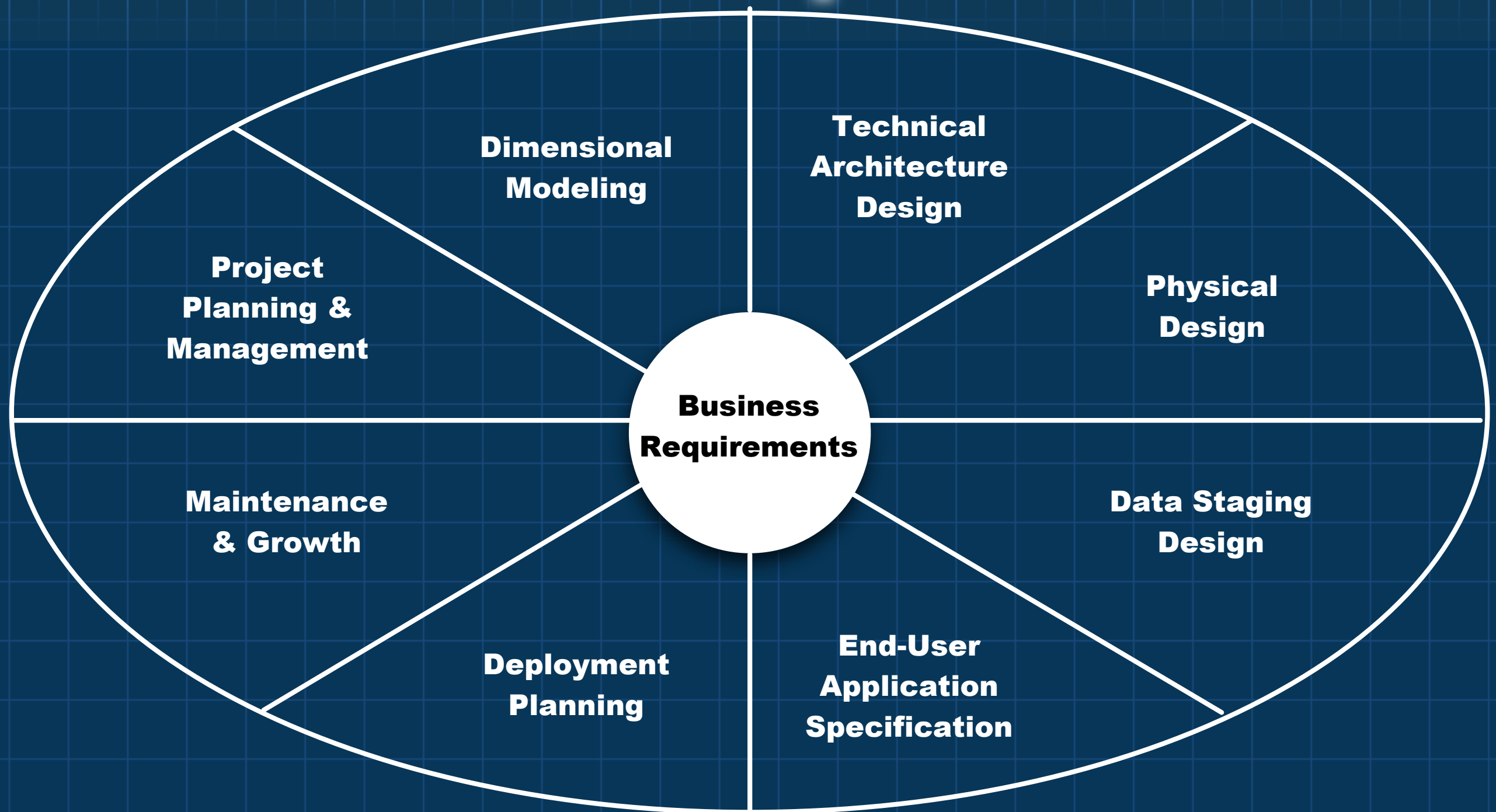
- **Quality of the information in the data warehouse**
- **The presentation tools to display the information is part of the data warehouse**
- **Support business intelligence applications**

# Business Intelligence Concept





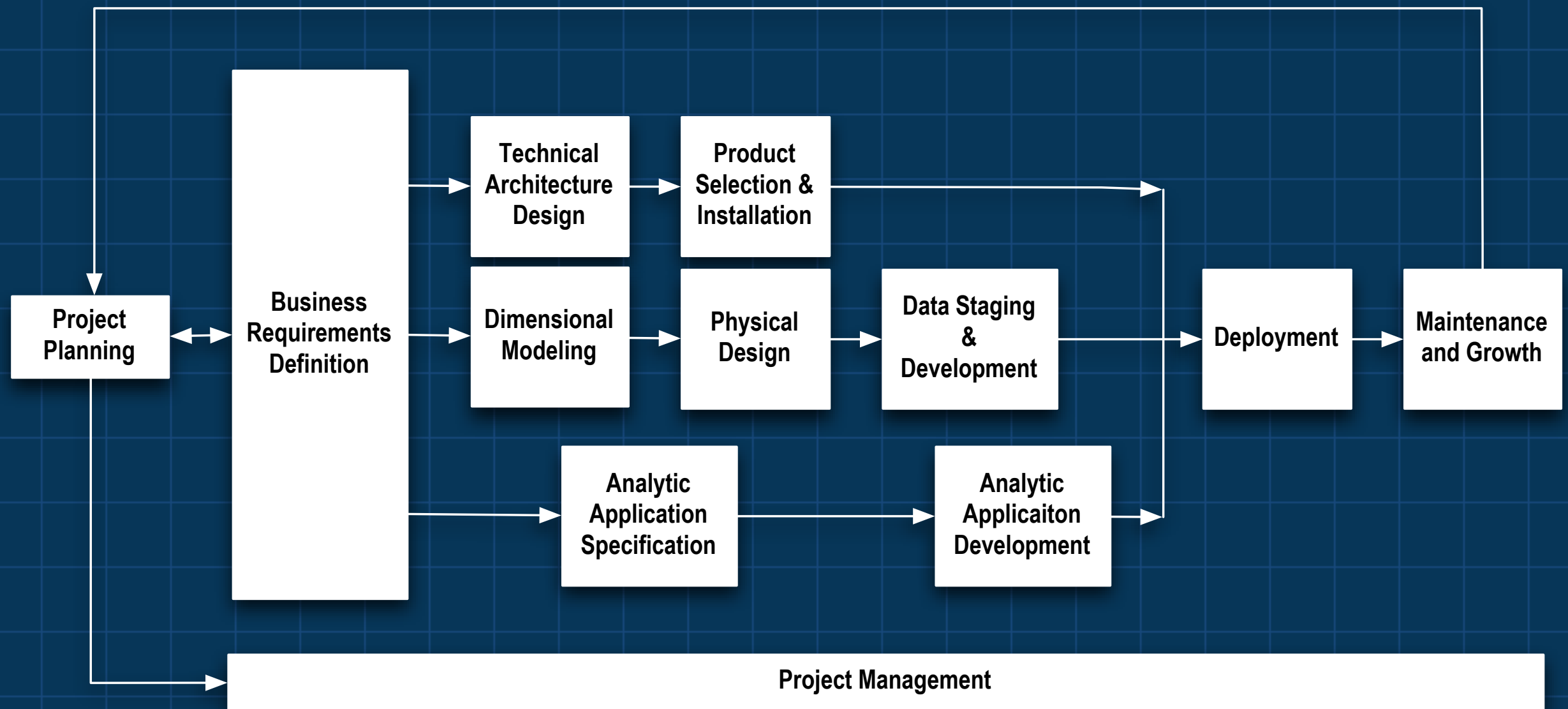
# Business Requirements



# How to build a Data Warehouse

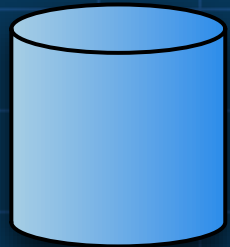
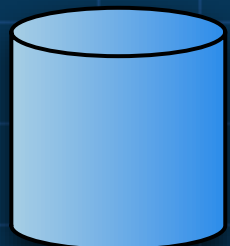
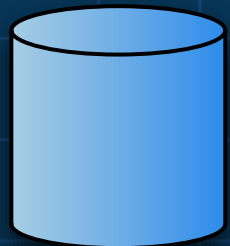
- Identify the problems and the business processes
- Identify the grain
- Identify the dimensions
- Identify the facts

# Data Warehouse Lifecycle diagram



# Elements of a Data Warehouse

**DATA  
SOURCES**



**DATA WAREHOUSE**

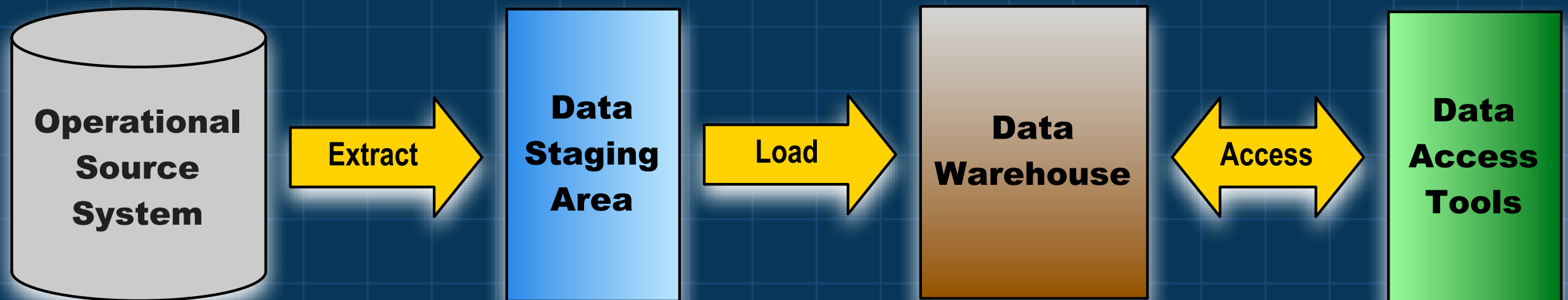


**DATA QUERY  
TOOLS**

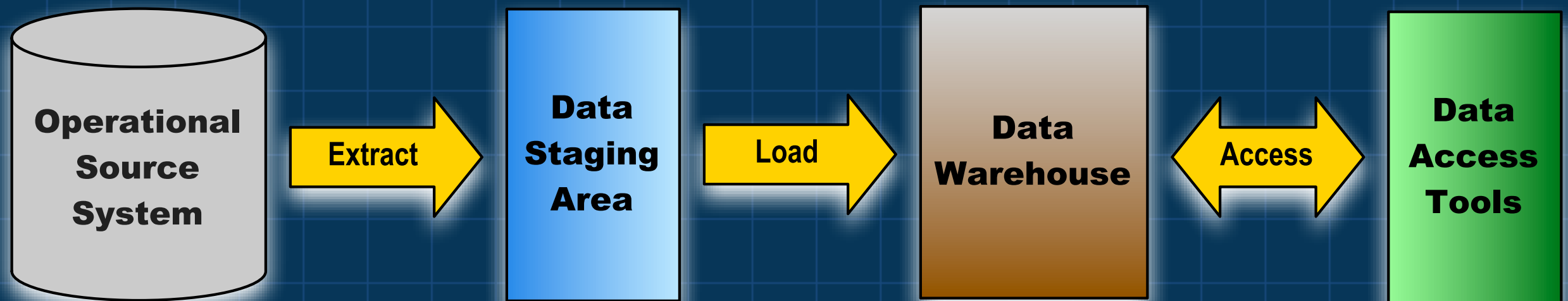


**Query Tools  
Reporting  
Analysis  
Data Mining**

# Elements of a Data Warehouse



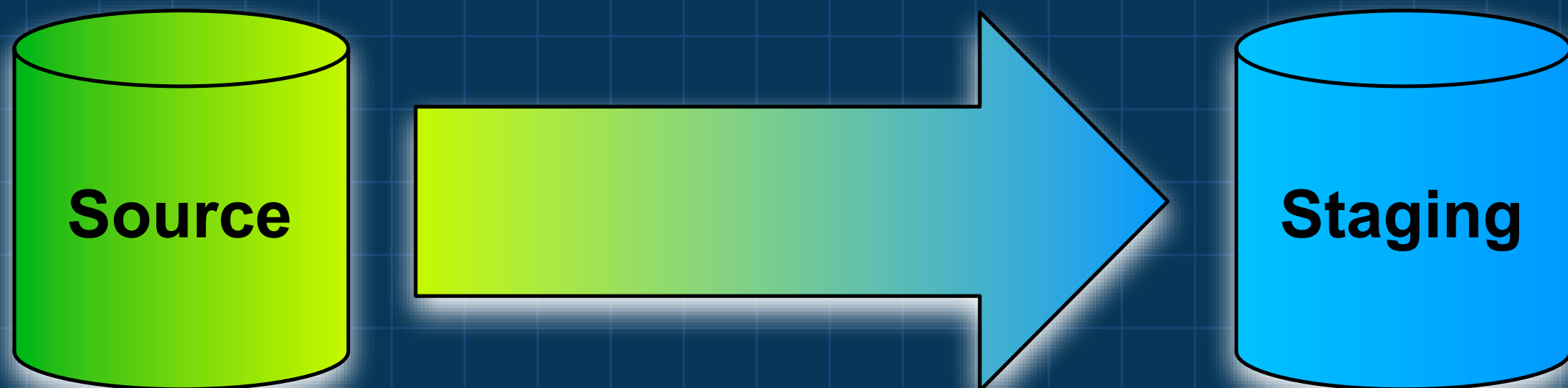
# Elements of a Data Warehouse



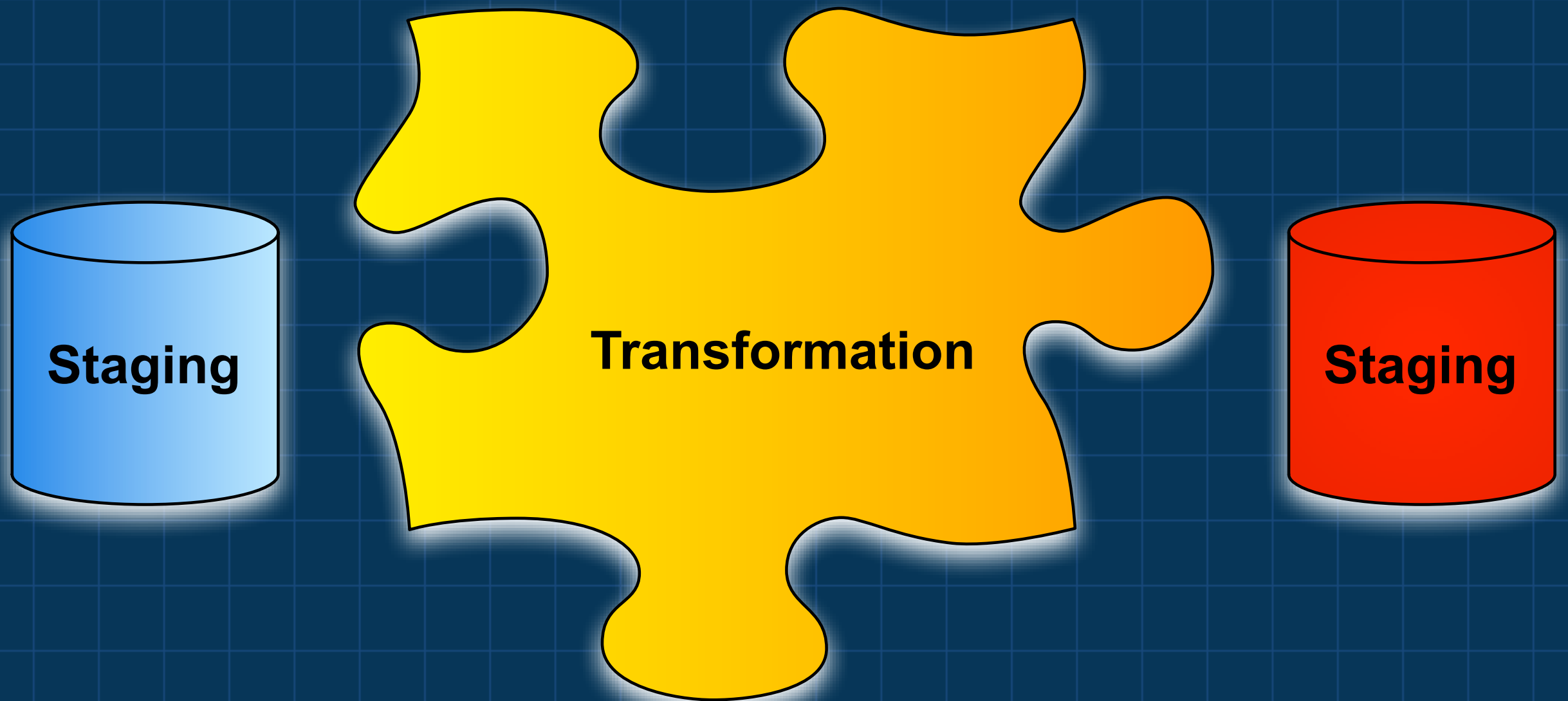
## Flow of information in a Data Warehouse



# Extraction

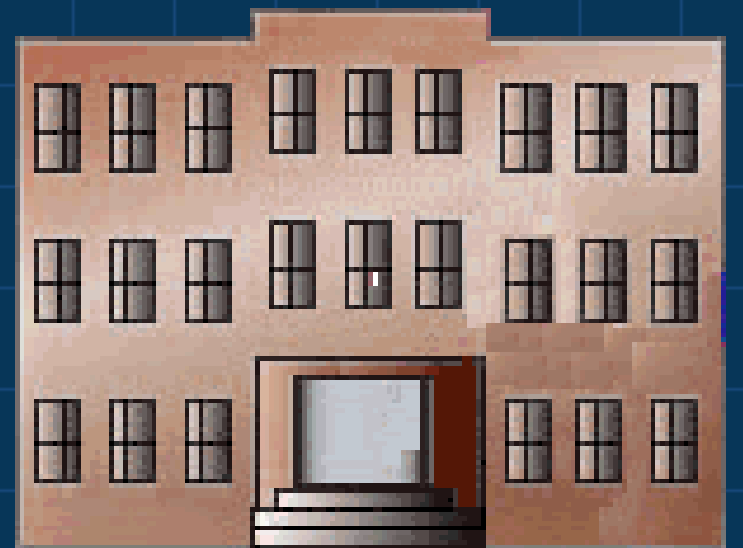
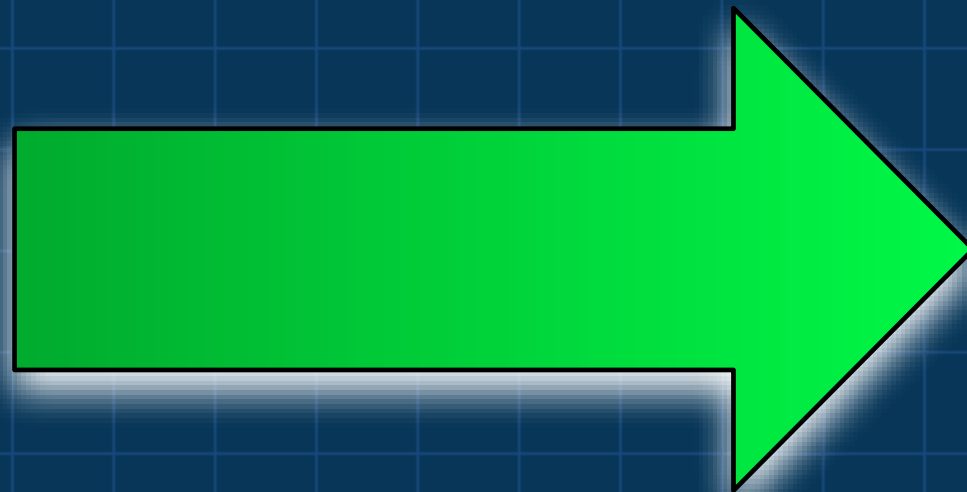


# Transformation

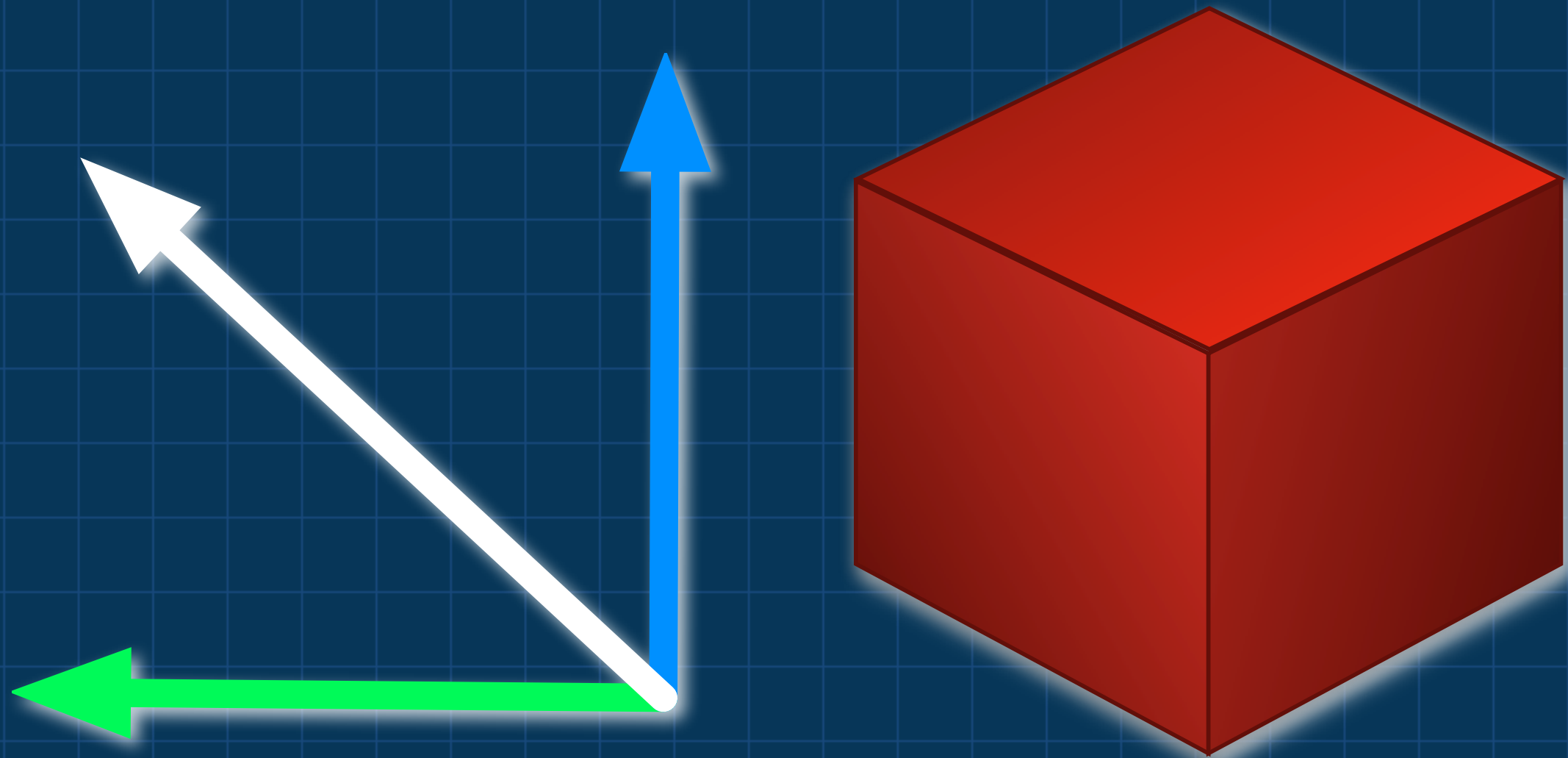




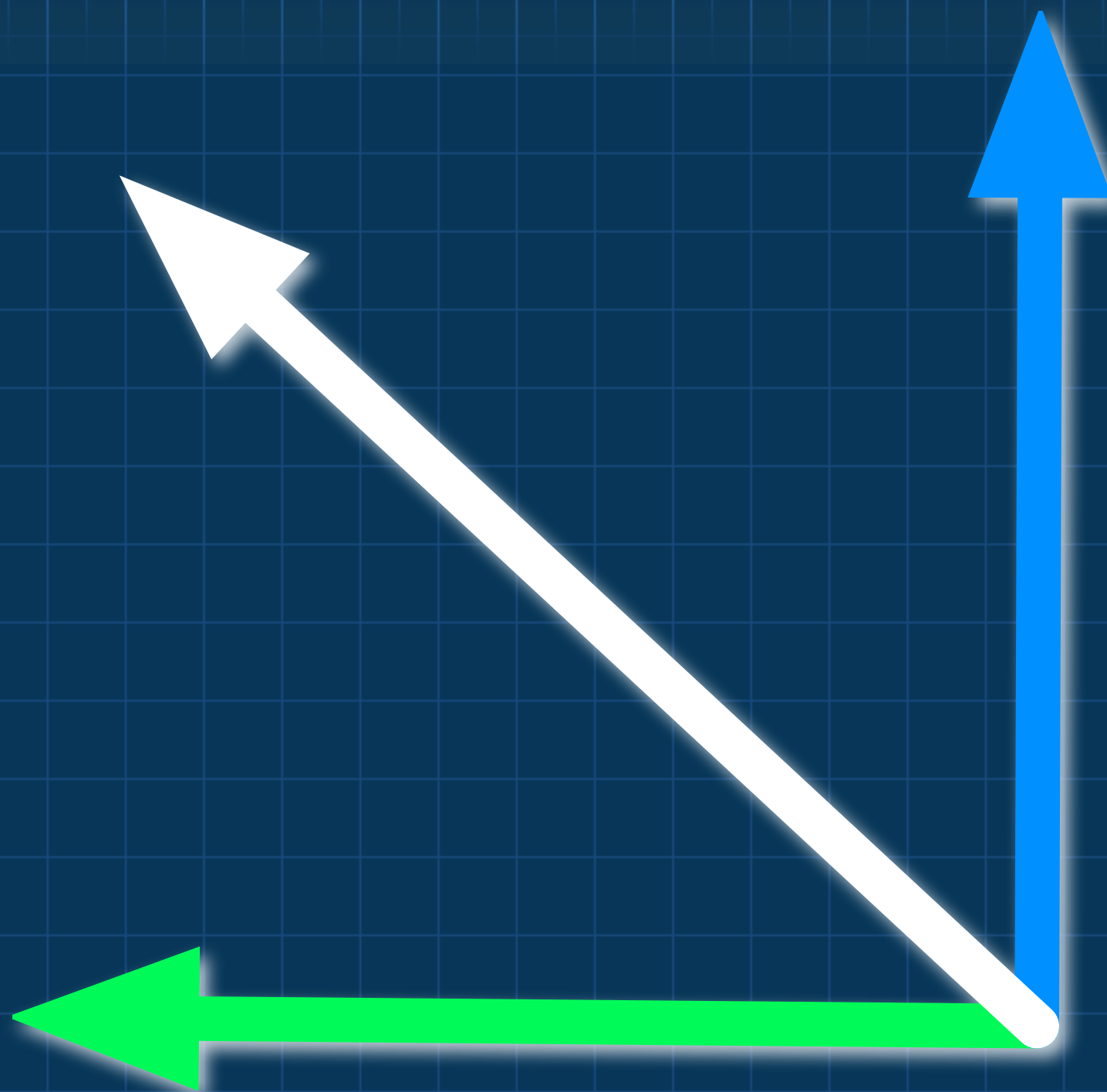
# Loading



# Dimension & Fact Table



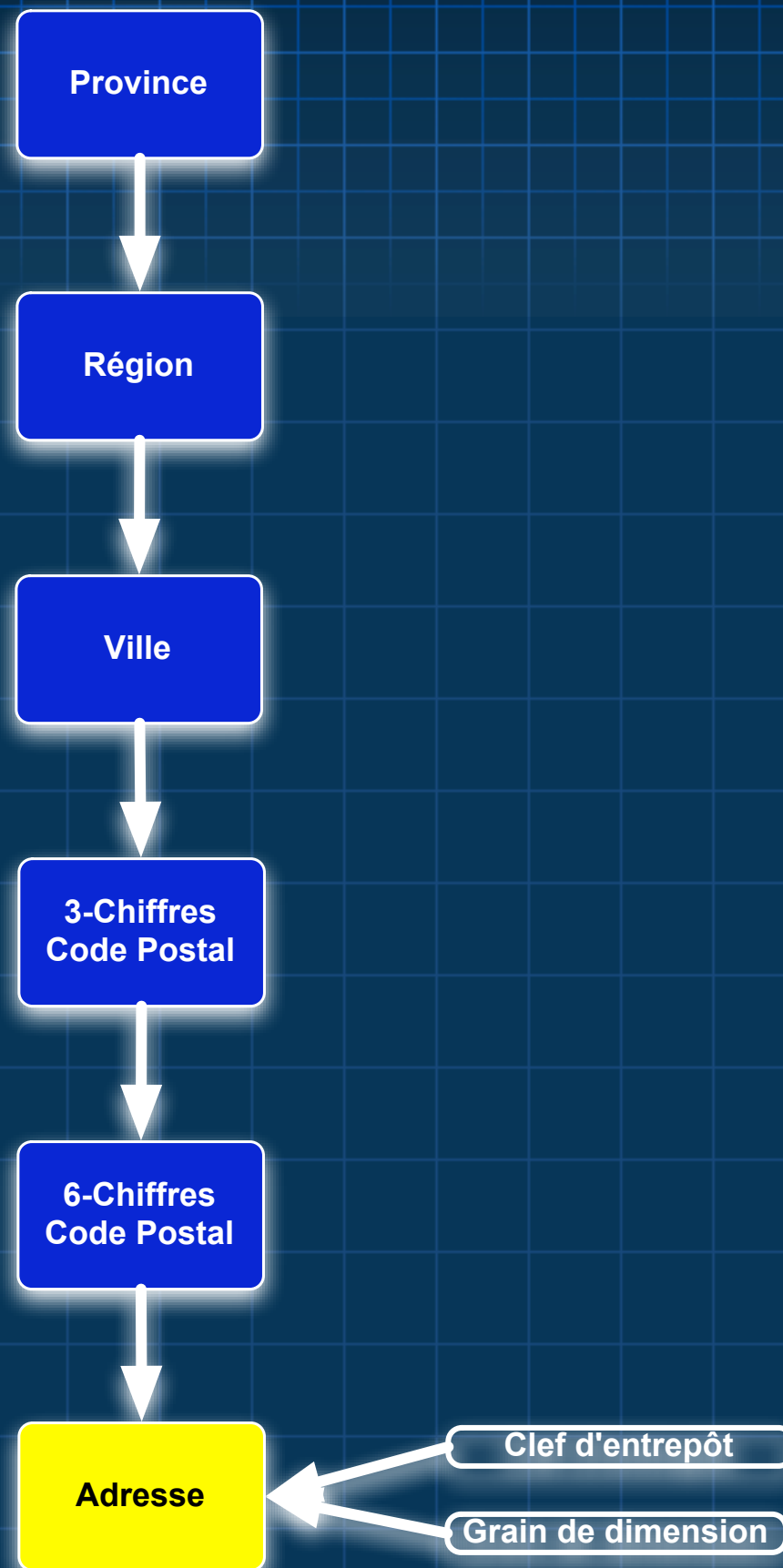
# Dimension Table



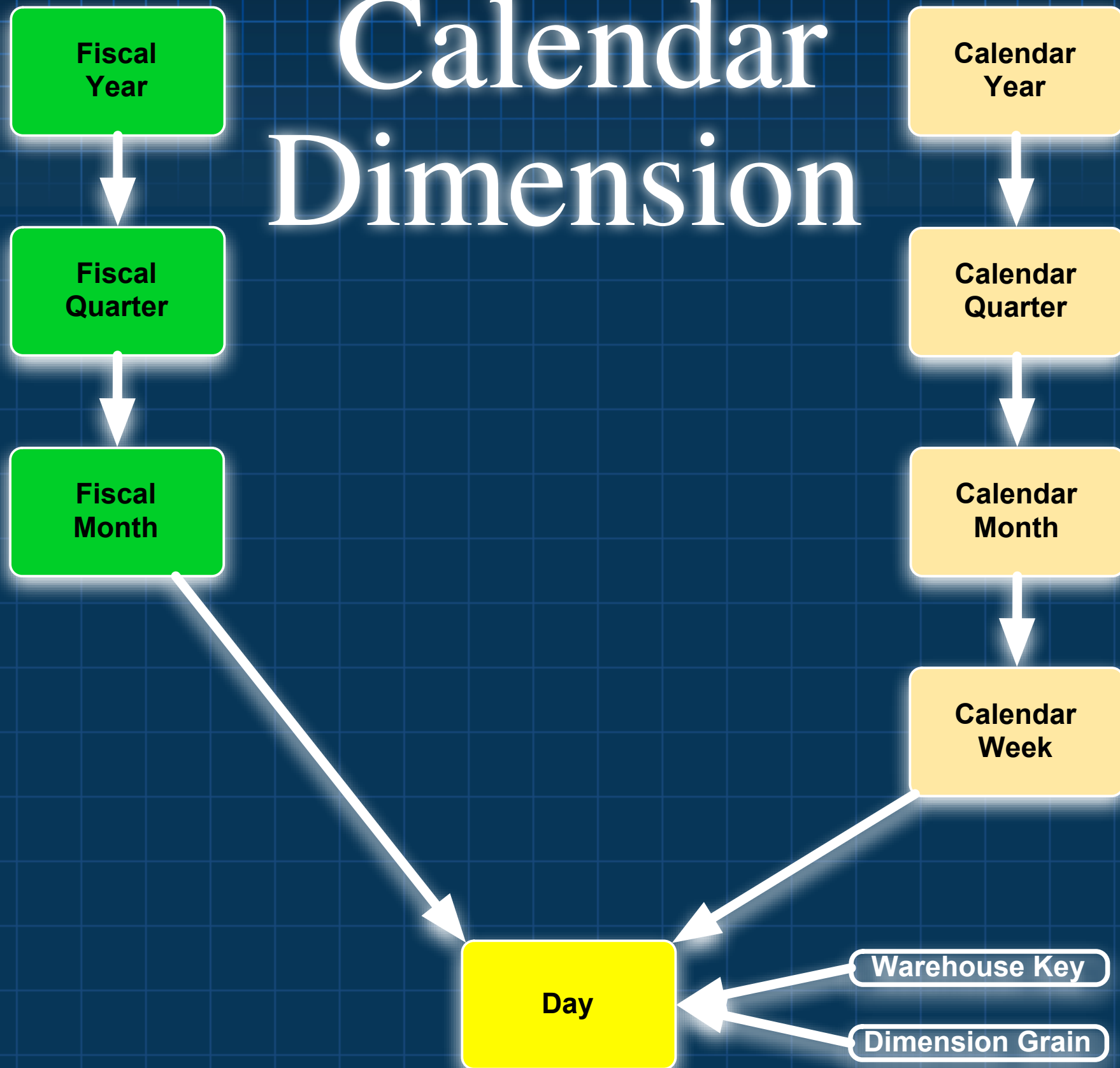
# Dimension Table

- Simple primary key
- Textual attributes rich and adapted to the user
- Hierarchical reports
- Few codes; Few codes; codes should be decoded according to their descriptions
- Relatively small

# Dimension Client Location

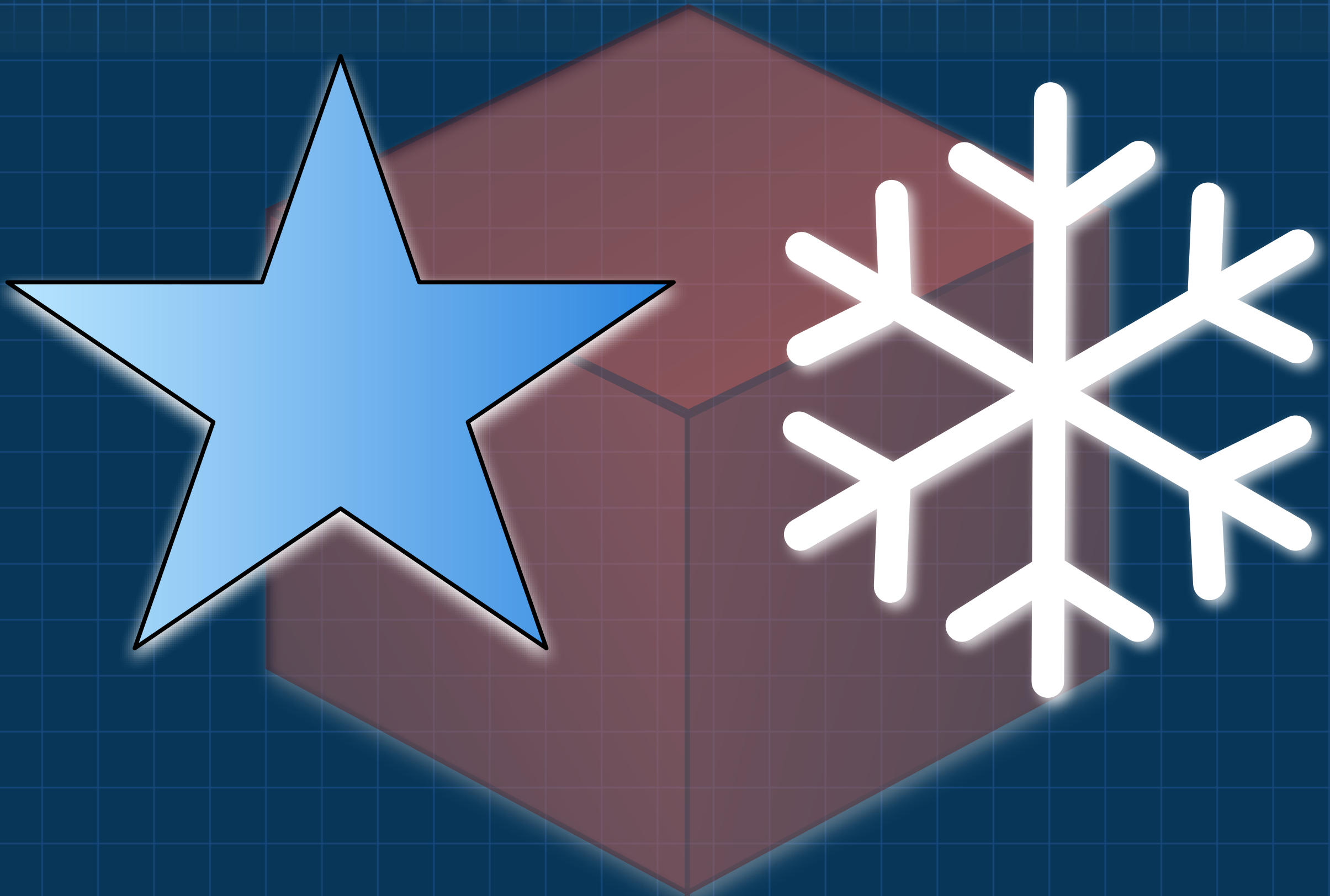


# Calendar Dimension



# Multidimensional Models

Star & Snowflake Schema





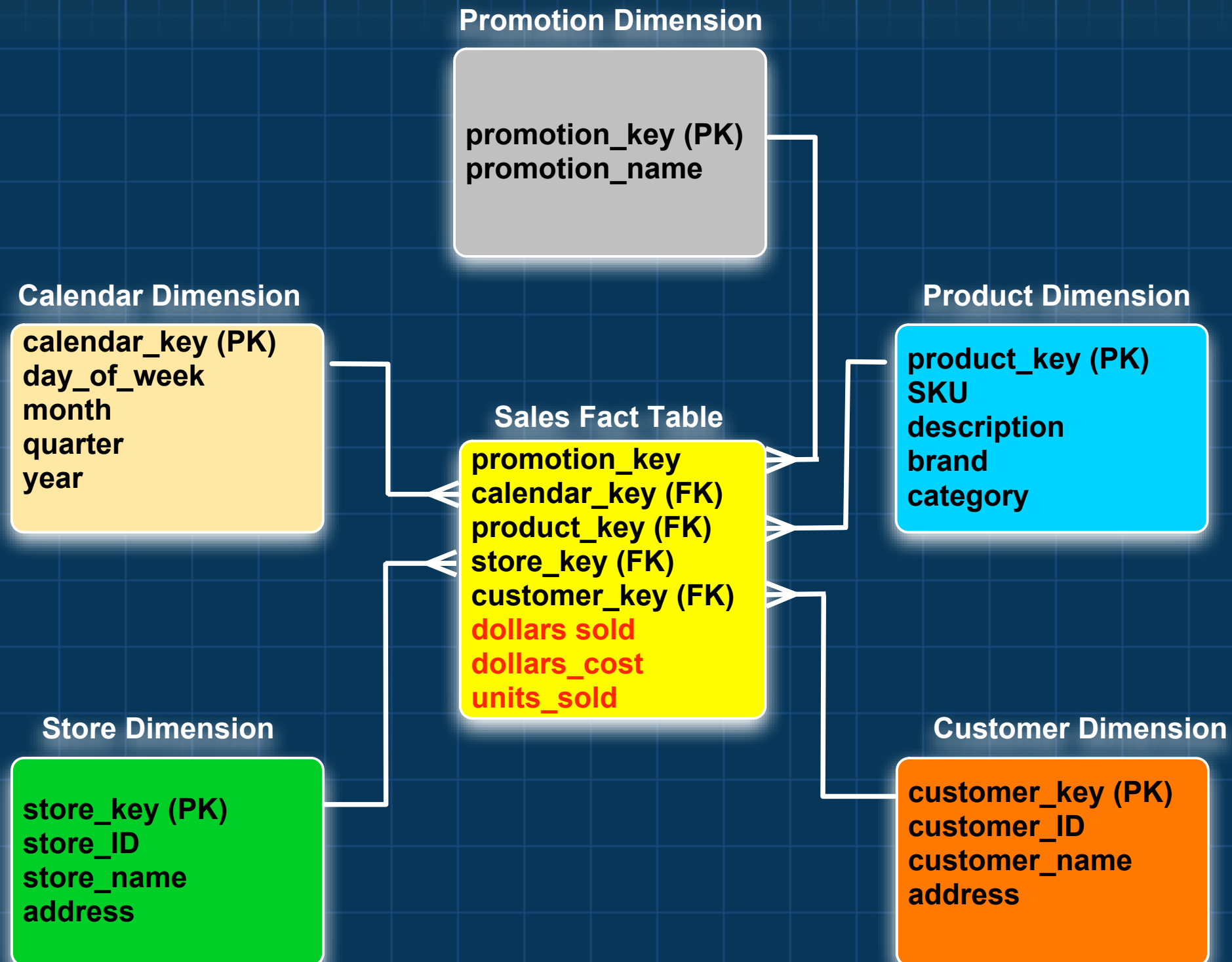
# Fact Table

- Composed Primary key
- The date/time is almost always a key
- The facts are usually numerical
- The facts are in general additive



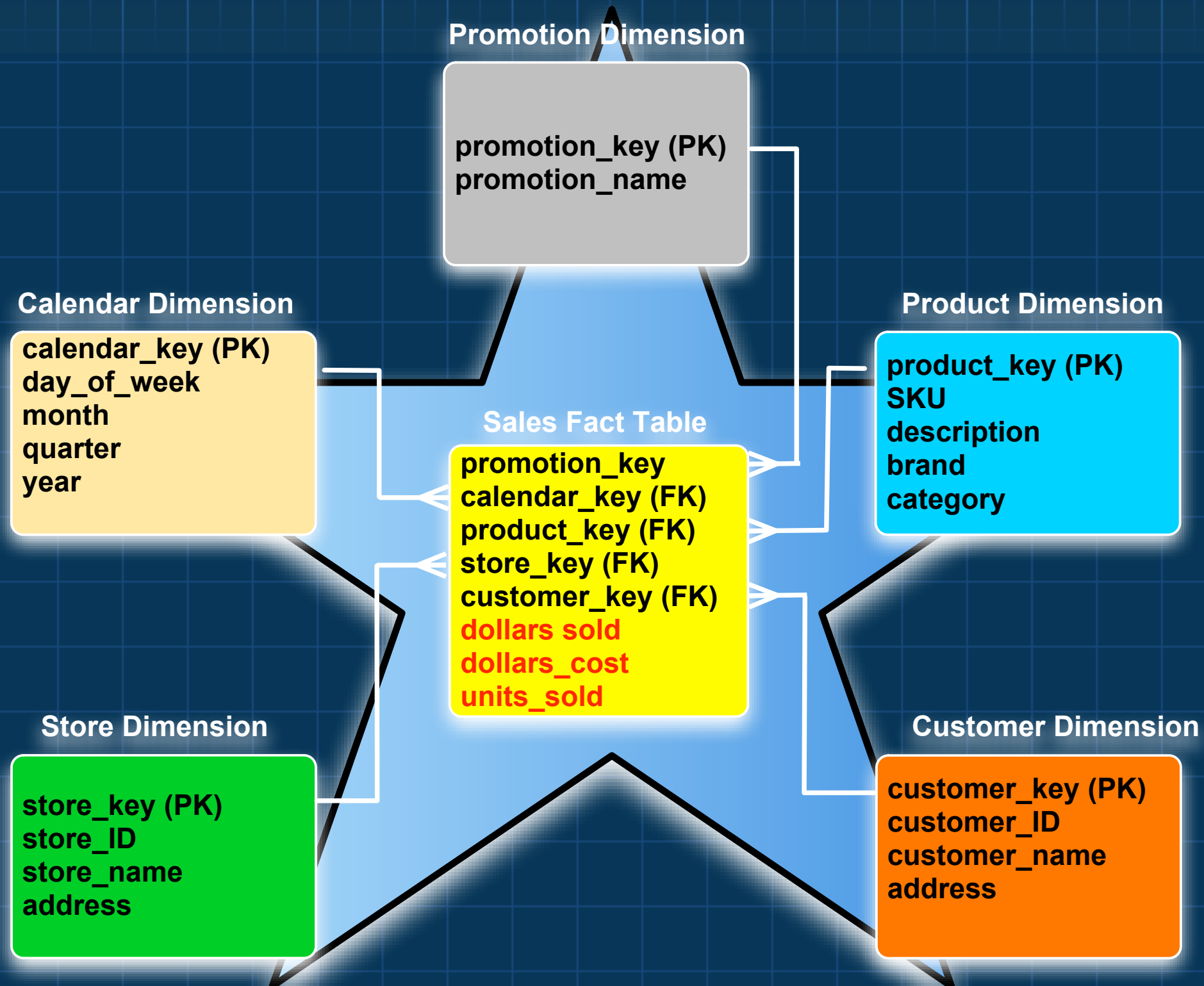
# Fact Table

## Star Schema



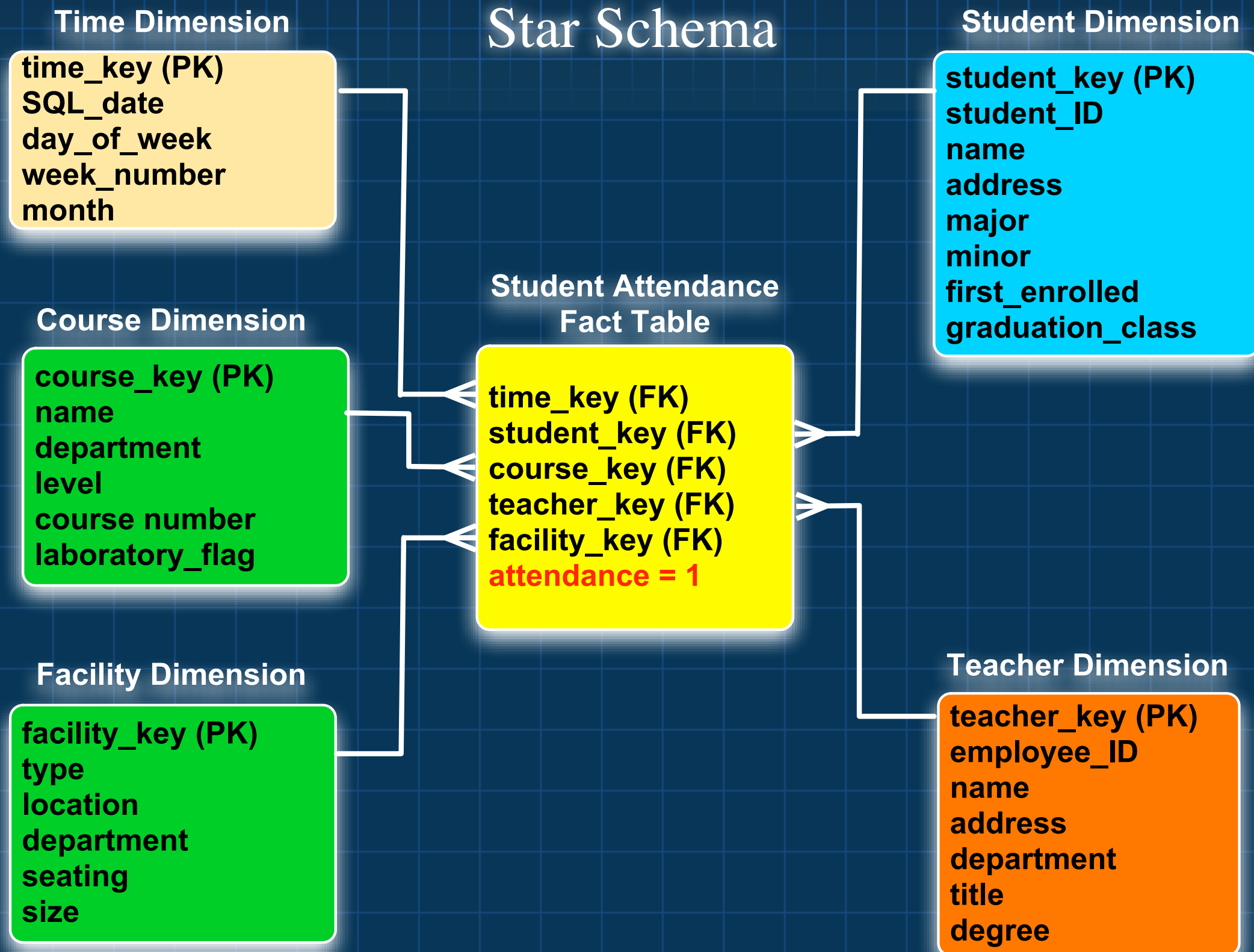
# Fact Table

## Star Schema



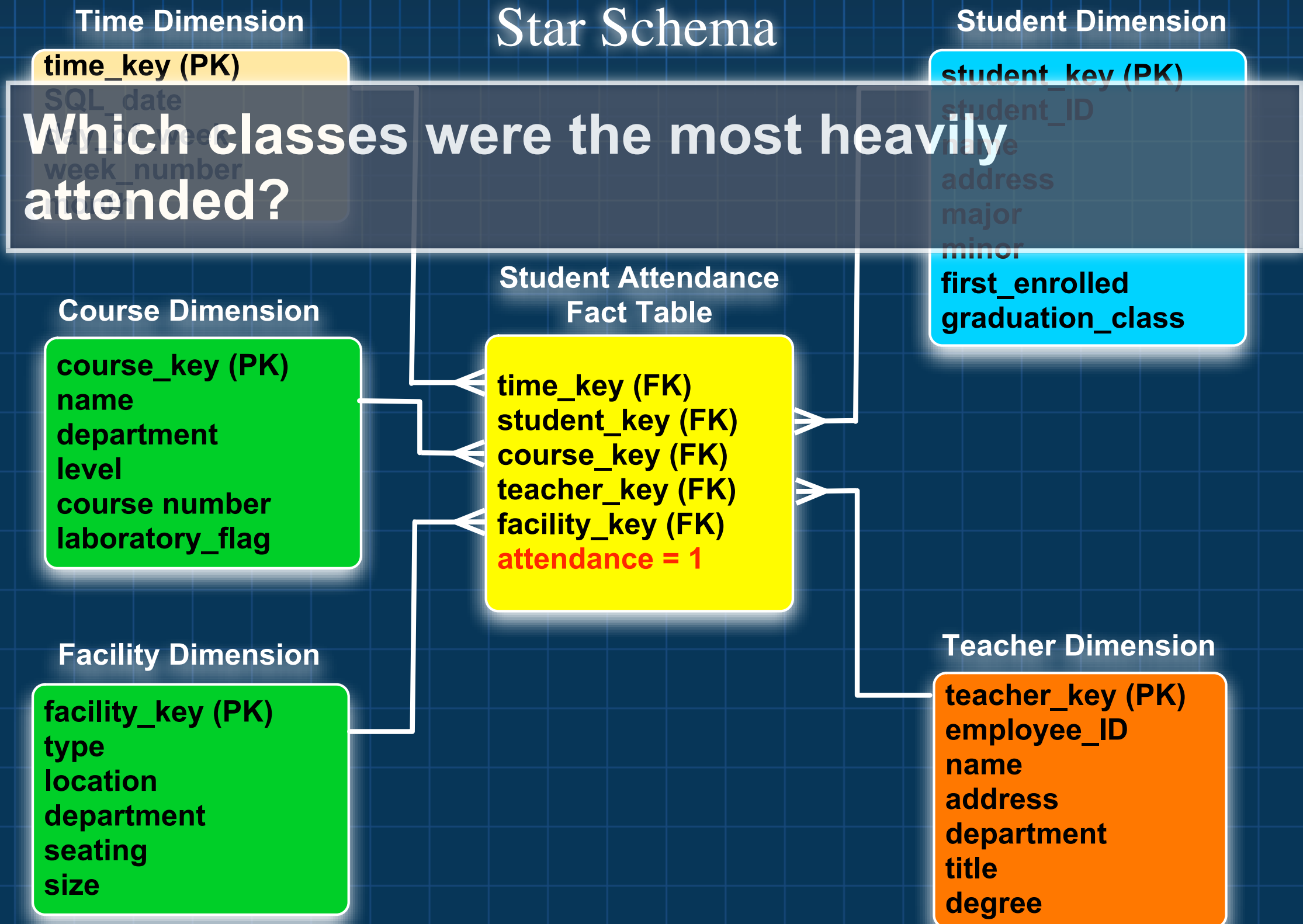
# Fact Table

## Star Schema



# Fact Table

## Star Schema



# Fact Table

## Star Schema

### Time Dimension

**time\_key (PK)**

SQL\_date

year

week\_number

**Which classes were the most heavily attended?**

### Student Dimension

**student\_key (PK)**

student\_ID

name

address

major

minor

first\_enrolled

graduation\_class

**Which classes were the most lightly used?**

### Course Dimension

**course\_key (PK)**

name

department

level

course number

laboratory\_flag

### Student Attendance

#### Fact Table

time\_key (FK)

student\_key (FK)

course\_key (FK)

teacher\_key (FK)

facility\_key (FK)

**attendance = 1**

### Facility Dimension

**facility\_key (PK)**

type

location

department

seating

size

### Teacher Dimension

**teacher\_key (PK)**

employee\_ID

name

address

department

title

degree

# Fact Table

## Star Schema

### Time Dimension

**time\_key (PK)**

SQL\_date

year

week\_number

**Which classes were the most heavily attended?**

### Student Dimension

**student\_key (PK)**

student\_ID

name

address

major

minor

first\_enrolled

graduation\_class

**Which classes were the most lightly used?**

### Course Dimension

**course\_key (PK)**

name

department

level

laboratory\_flag

**Which teachers taught the most students?**

### Student Attendance Fact Table

time\_key (FK)

student\_key (FK)

course\_key (FK)

teacher\_key (FK)

facility\_key (FK)

attendance = 1

### Facility Dimension

**facility\_key (PK)**

type

location

department

seating

size

### Teacher Dimension

**teacher\_key (PK)**

employee\_ID

name

address

department

title

degree



# Fact Table

## Star Schema

### Time Dimension

**time\_key (PK)**

SQL\_date

year

week\_number

Which classes were the most heavily attended?

### Student Dimension

**student\_key (PK)**

student\_ID

name

address

major

minor

first\_enrolled

graduation\_class

Which classes were the most lightly used?

### Course Dimension

**course\_key (PK)**

name

department

level

laboratory\_flag

Which teachers taught the most students?

### Student Attendance Fact Table

time\_key (FK)

student\_key (FK)

course\_key (FK)

teacher\_key (FK)

attendance = 1

Which teachers taught classes in facilities belonging to other departments?

### Facility Dimension

type

location

department

seating

size

### Teacher Dimension

**teacher\_key (PK)**

employee\_ID

name

address

department

title

degree

# Fact Table

## Snowflake Schema

### Calendar Dimension

**calendar\_key (PK)**  
**month\_key (FK)**  
**year**

**month\_key (PK)**  
**year**

**year\_key (PK)**

### Store Dimension

**store\_key (PK)**  
**store\_ID**  
**store\_nom**  
**address**

### Product Dimension

**product\_key (PK)**  
**SKU**  
**description**  
**brand\_key (FK)**  
**category**

**brand\_key (PK)**  
**brand\_description**

### Customer Dimension

**customer\_key (PK)**  
**customer\_ID**  
**customer\_name**  
**address**

### Sales Fact Table

**calendar\_key (FK)**  
**product\_key (FK)**  
**store\_key (FK)**  
**customer\_key (FK)**  
**dollars\_sold**  
**dollars\_cost**  
**units\_sold**



# Slowly Changing Dimensions

# Slowly Changing Dimensions

**Type 1: Overwrite changed attribute.**

A fact is associated with only the current value of a dimension column.

# Slowly Changing Dimensions

## **Type 1: Overwrite changed attribute.**

A fact is associated with only the current value of a dimension column.

## **Type 2: Add new dimension record.**

A fact is associated with only the original value of a dimension column.

# Slowly Changing Dimensions

## **Type 1: Overwrite changed attribute.**

A fact is associated with only the current value of a dimension column.

## **Type 2: Add new dimension record.**

A fact is associated with only the original value of a dimension column.

## **Type 3: Use field for 'old' value.**

A fact is associated with both the original value and with the current value of a dimension column.

# Slowly Changing Dimensions

**Scenario: Customer last name changes from Pharand to Smith: Update Cust.Lname**

Original dimension table			
Cust Key	Cust ID	Cust Lname	Cust Fname
1	TE123	Tremblay	Eric
2	PJ456	Pharand	Josée

Slowly changing dimension Type 1			
Cust Key	Cust ID	Cust Lname	Cust Fname
1	TE123	Tremblay	Eric
2	PJ456	Smith	Josée

# Slowly Changing Dimensions

Slowly changing dimension Type 2						
Cust Key	Cust ID	Cust Lname	Cust Fname	Row Current?	Row Start	Row Stop
1	TE123	Tremblay	Eric	yes	12/3/99	
2	PJ456	Pharand	Josée	no	2/20/00	9/15/00
3	PJ789	Smith	Josée	yes	9/16/00	

# Slowly Changing Dimensions

Slowly changing dimension Type 2						
Cust Key	Cust ID	Cust Lname	Cust Fname	Row Current?	Row Start	Row Stop
1	TE123	Tremblay	Eric	yes	12/3/99	
2	PJ456	Pharand	Josée	no	2/20/00	9/15/00
3	PJ789	Smith	Josée	yes	9/16/00	

Slowly changing dimension Type 3				
Cust Key	Cust ID	Cust Lname	Cust Fname	Old Value
1	TE123	Tremblay	Eric	
2	PJ456	Smith	Josée	Pharand

# Normalisation



# Normalisation

# 3NF

# Normalisation



# Normalisation

# Normalisation

Simplicity

# Normalisation

Simplicity

Query Performance

# Normalisation

Simplicity

Query Performance

Minor disk space saving

# Normalisation

Simplicity

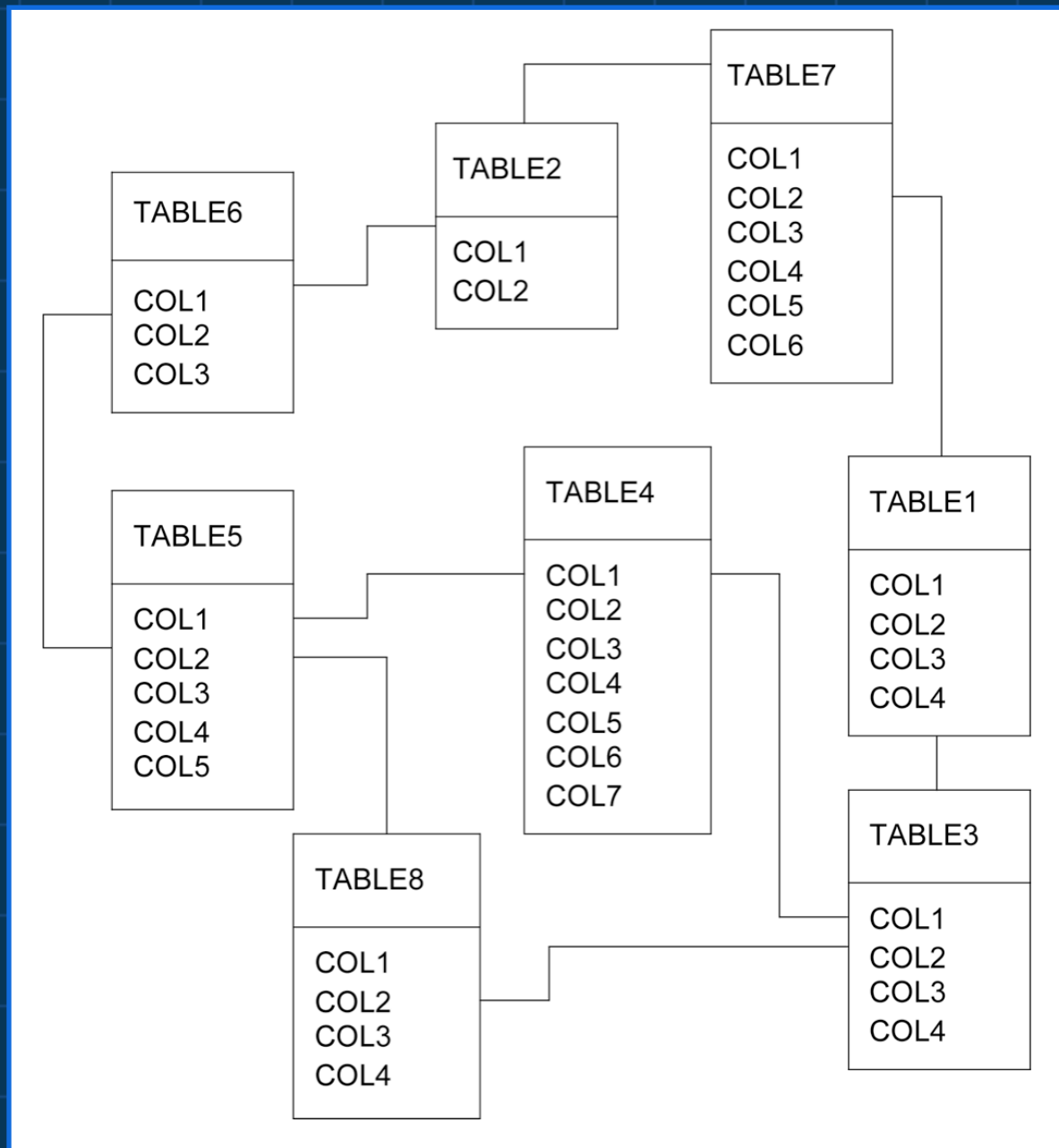
Query Performance

Minor disk space saving

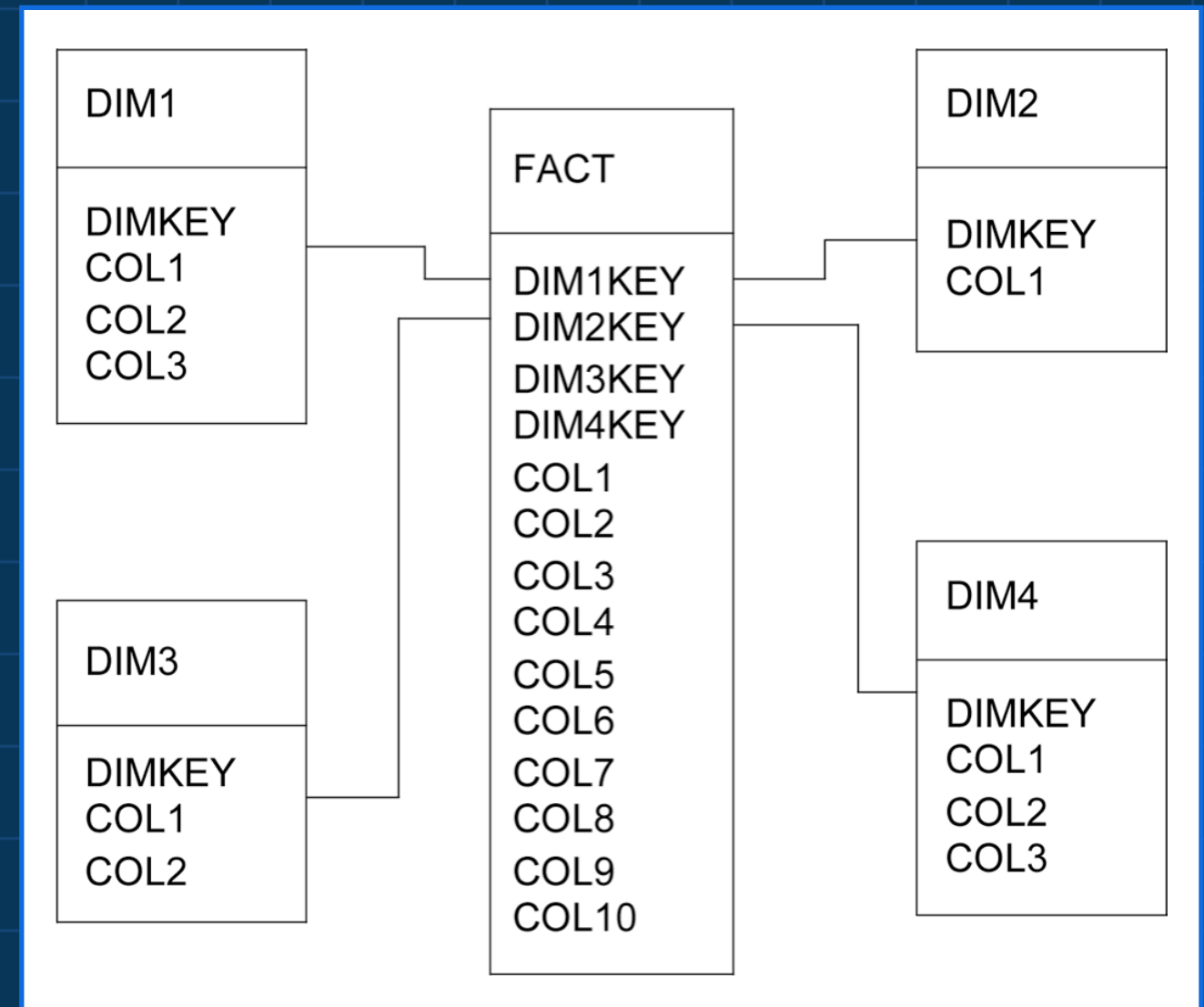
Slows down the users' ability to  
browse

# The Difference

## ERD



## Star Schema



(Star Schema)

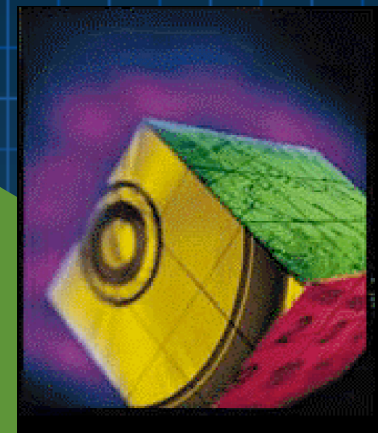


# The Difference

Data Base	Data Warehouse
Actual	Historical
Internal	Internal and External
Isolated	Integrated
Transactions	Analysis
Normalised	Dimensional
Dirty	Clean and Consistent
Detailed	Detailed and Summary



**Ad Hoc Query  
and Analysis**

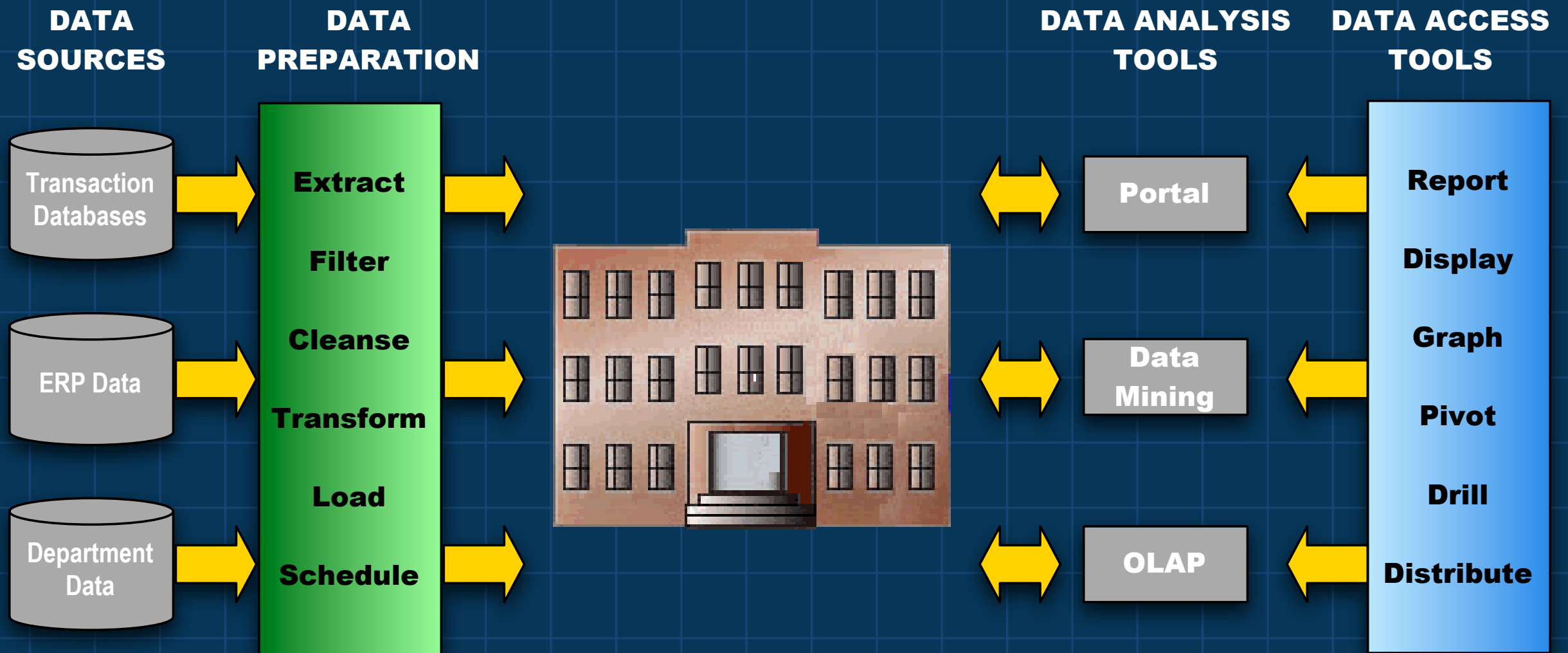


**Advanced  
analysis**

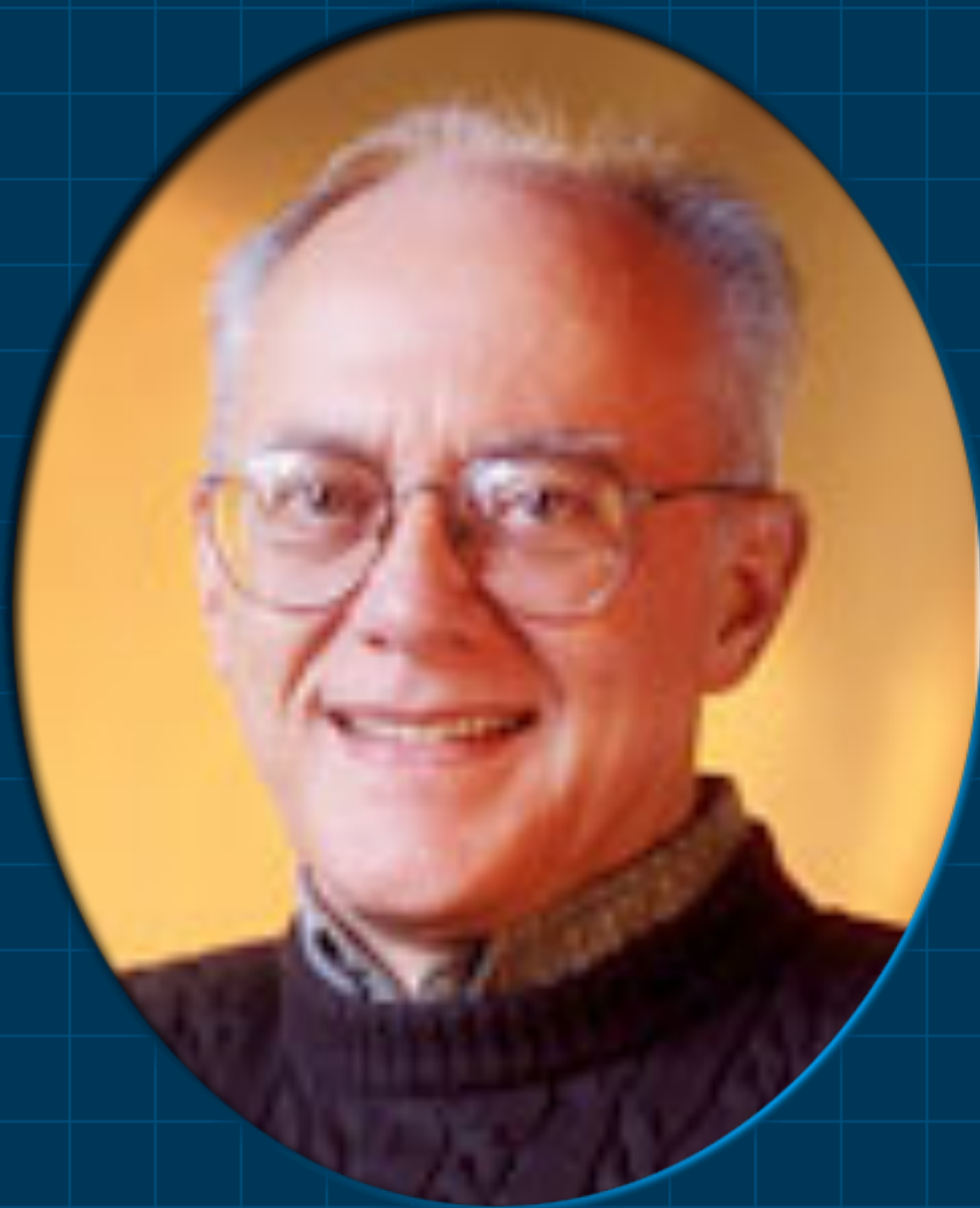


**Enterprise  
reporting**

# Elements of a Data Warehouse



# Ralph Kimball



○ Data Warehouse  
Guru

○ [www.rkimball.com](http://www.rkimball.com)



# The Data Warehouse Lifecycle Toolkit

2<sup>nd</sup>  
Edition

Practical Techniques  
for Building Data  
Warehouse and  
Business Intelligence  
Systems

**Ralph Kimball**  
Margy Ross  
Warren Thornthwaite  
Joy Mundy  
Bob Becker



# The Data Warehouse Toolkit

Second Edition

The Complete  
Guide to  
Dimensional  
Modeling

**Ralph Kimball**  
Margy Ross

