

**Data Warehouse  
Enablement Through  
Bi-Level Data Modeling**

**Charles W. Bachman**

# Overview:

---

- ✓ • **decision support premise**
- **decision support problem**
- **data warehouse solution**
- **data complexity problem**
- **“bi- level” data model solution**

# Decision Support Premise

---

**That there is valuable information locked up in the current operational systems and other sources which, if it were readily available, would contribute to improved business decision making and lead to a more efficient and profitable business.**

# Overview:

---

- decision support premise
- ✓ • decision support problem
- data warehouse solution
- data complexity problem
- “bi- level” data model solution

# Decision Support Problem

---

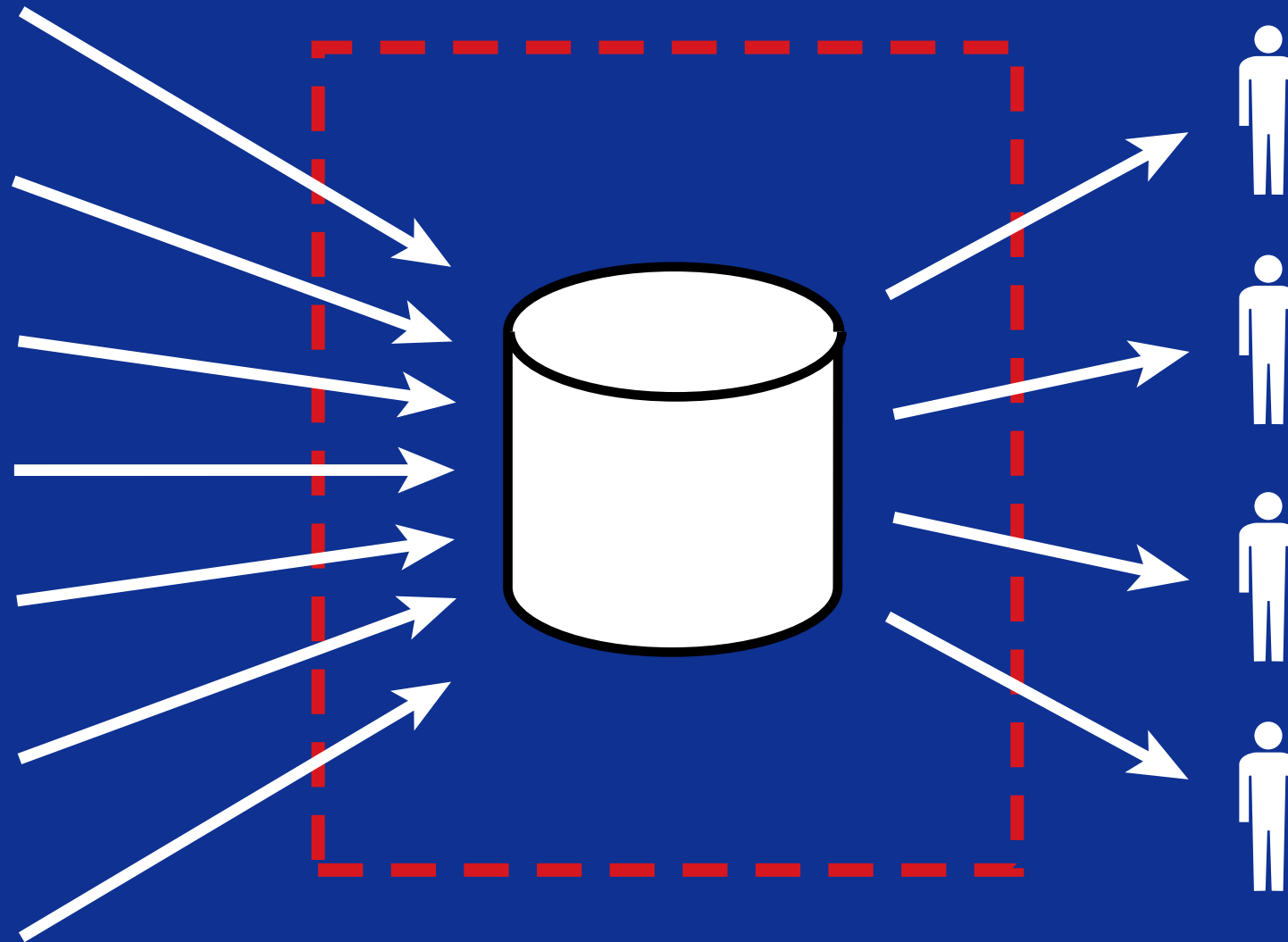
- **many, independently developed operational systems**
- **inadequate capacity and performance**
- **inappropriate data organization**
- **conflicting objectives**
- **insufficient data content**

# Integrated Database Solution

---

- **single, integrated database**
- **unchallenged capacity and performance**
- **multiple, concurrent data organizations**
- **unified objectives**
- **integrated, current, archival and external data**

# Integrated Database Solution



**single, integrated, all-purpose database**

# Integrated Database Solution

---

- **idealized**
- **impractical**
- **too risky**
- **not in our time**
- **maybe never**

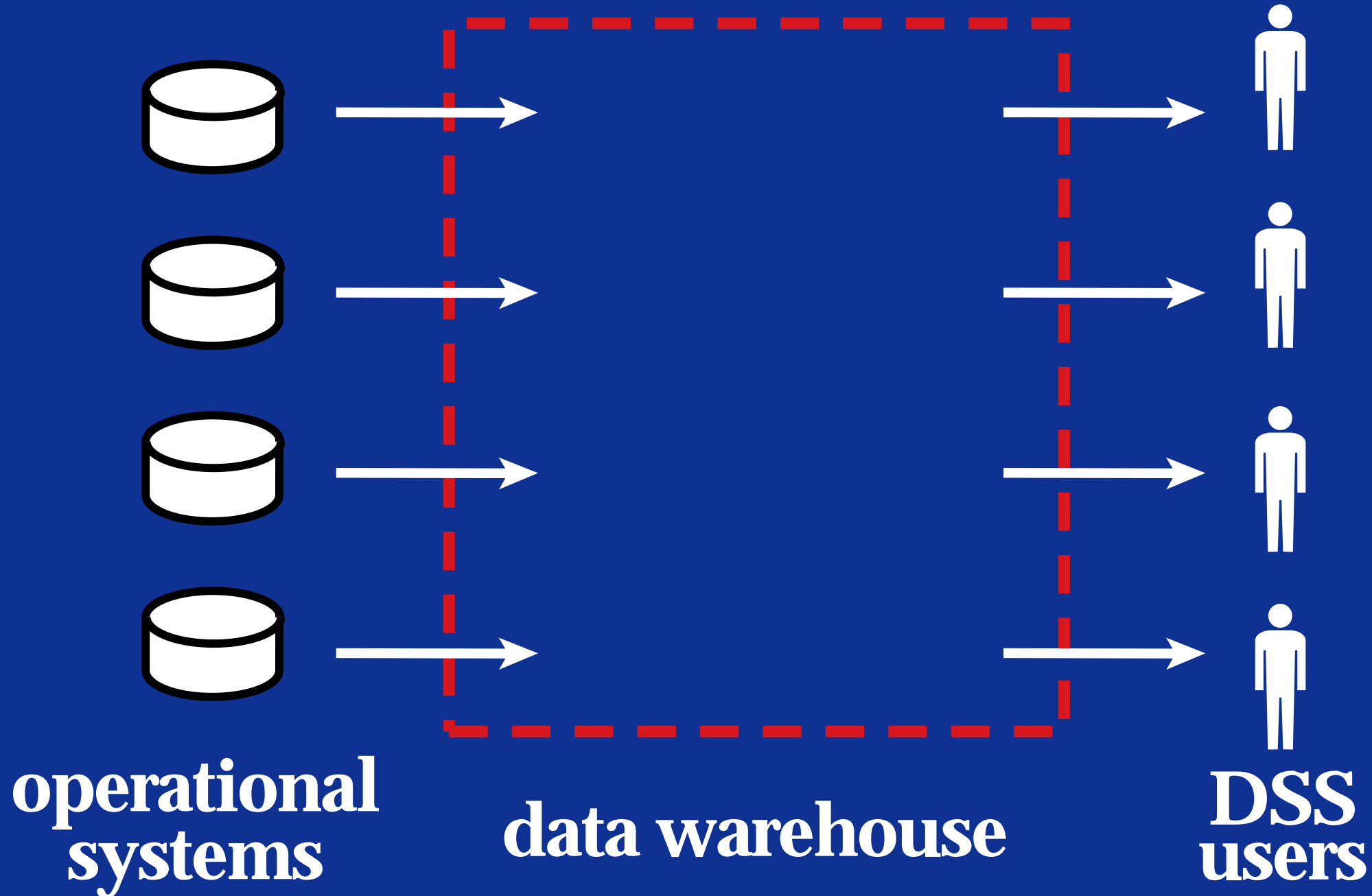


# Overview:

---

- decision support premise
- decision support problem
- ✓ • data warehouse solution
- data complexity problem
- “bi- level” data model solution

# Data Warehouse Solution

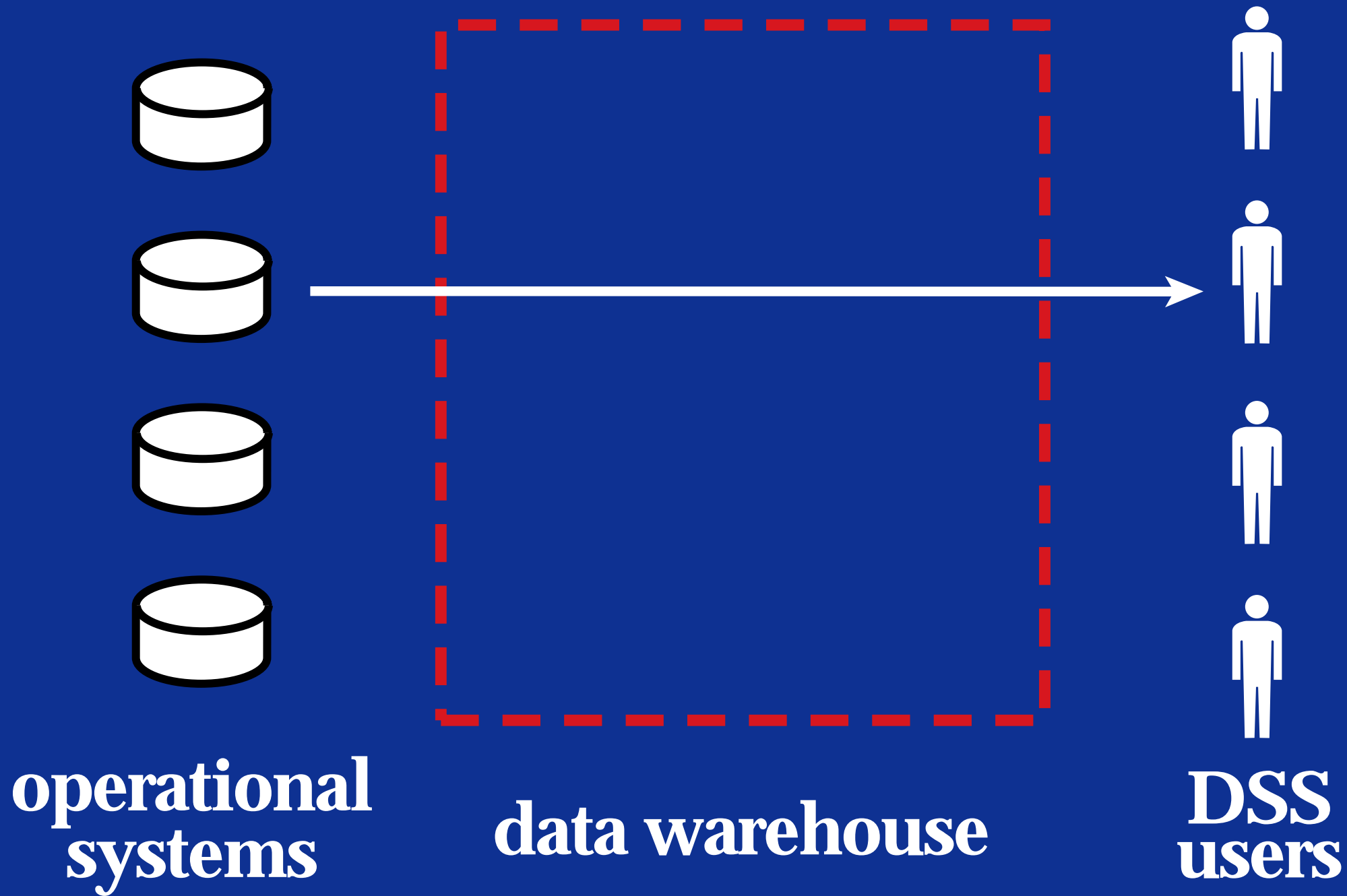


# Data Warehouse Solution

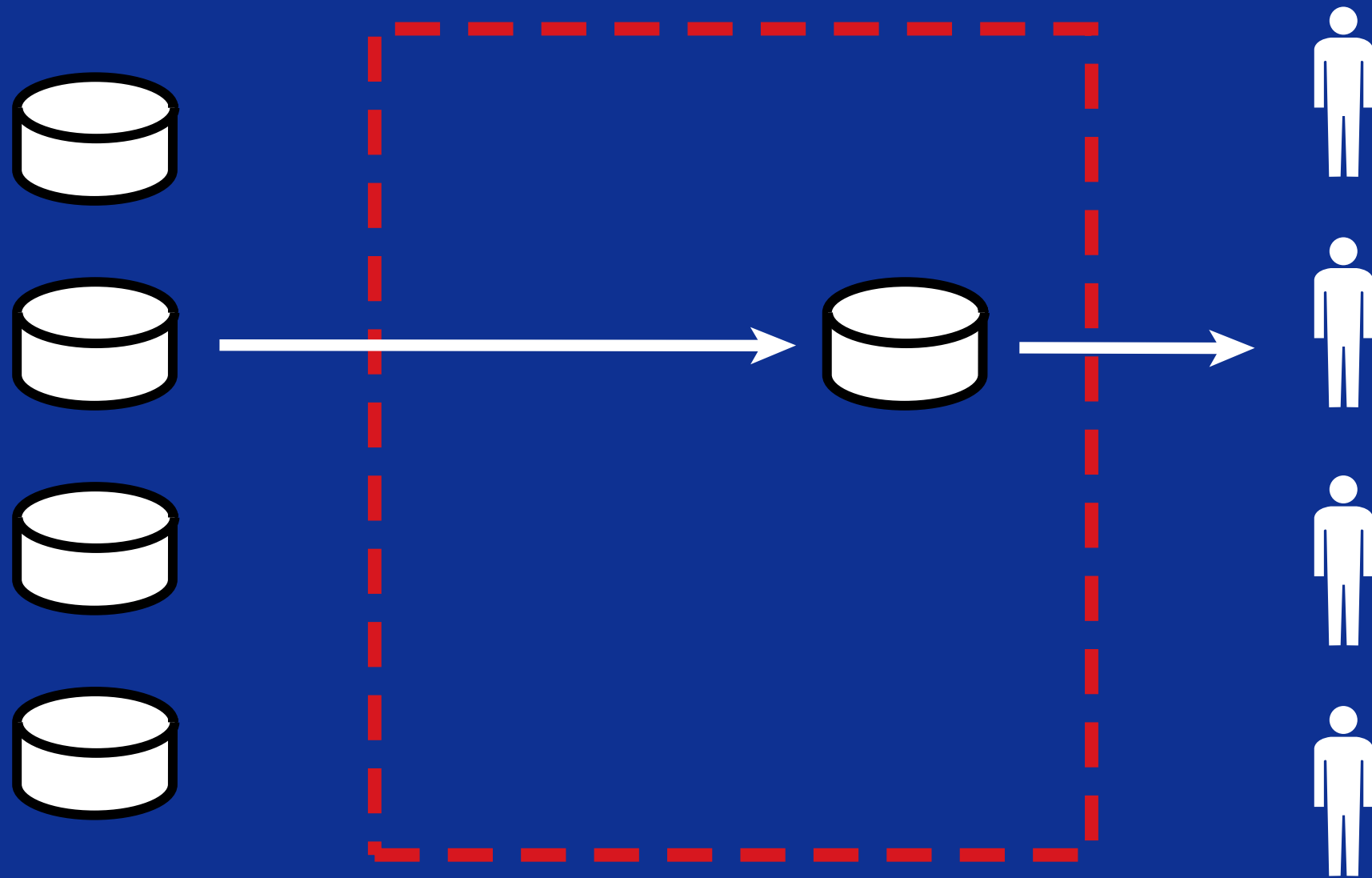
---

- pragmatic
- practical
- minimize risk
- compromising
- potentially, very expensive

# Data Warehouse Solution (stage 0)

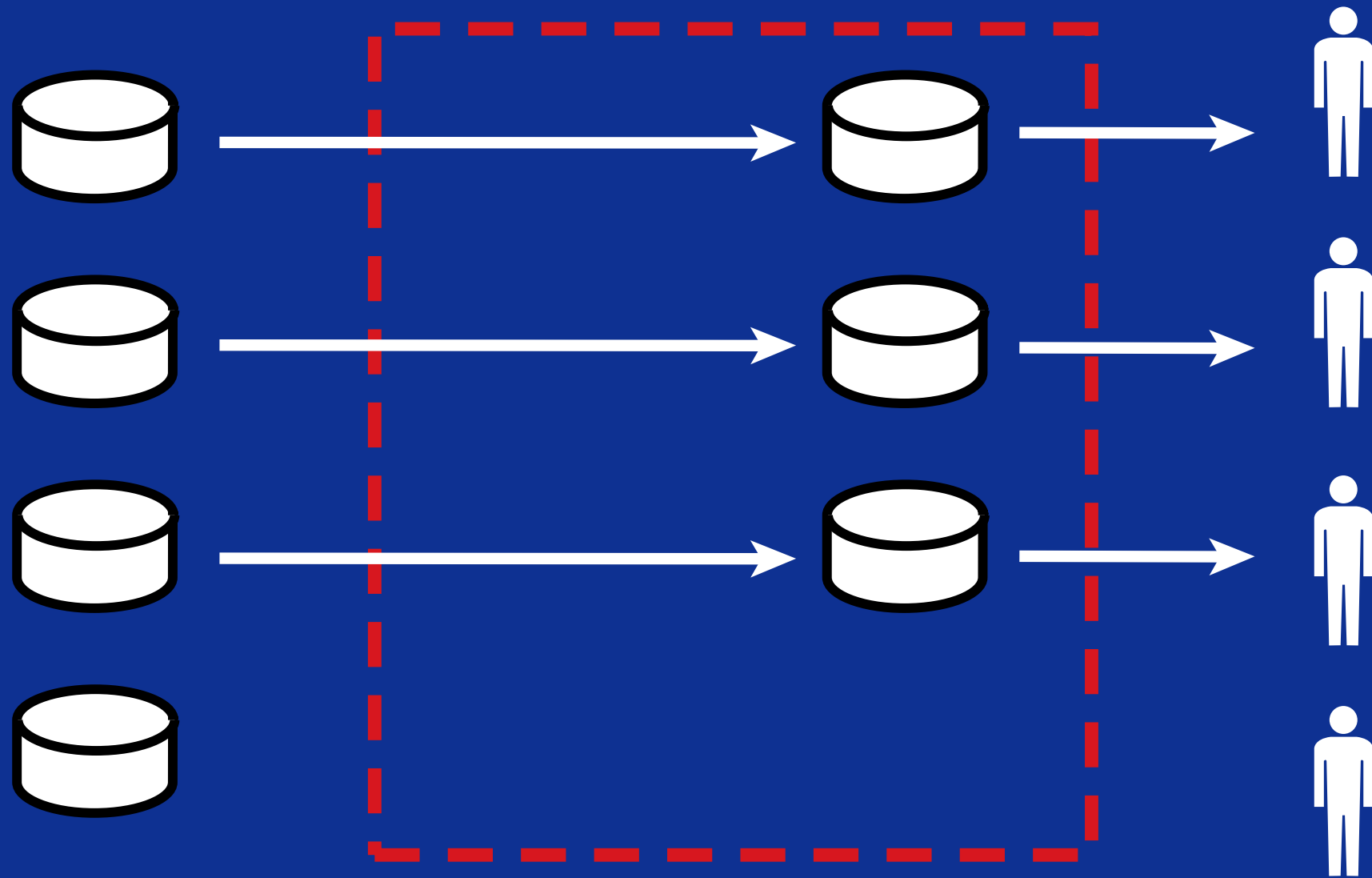


# Data Warehouse Solution (stage 1)



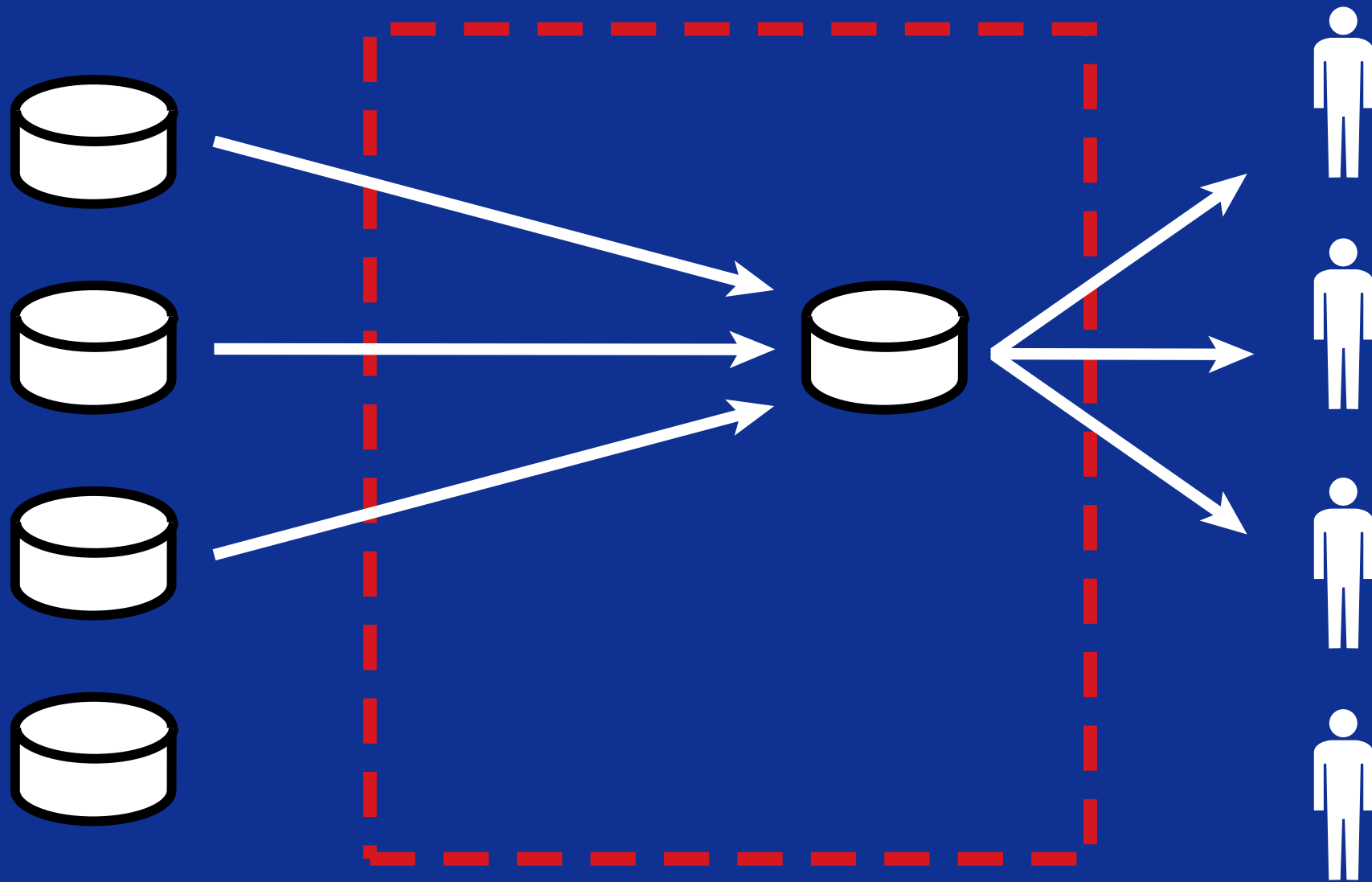
data warehouse

# Data Warehouse Solution (stage 1)



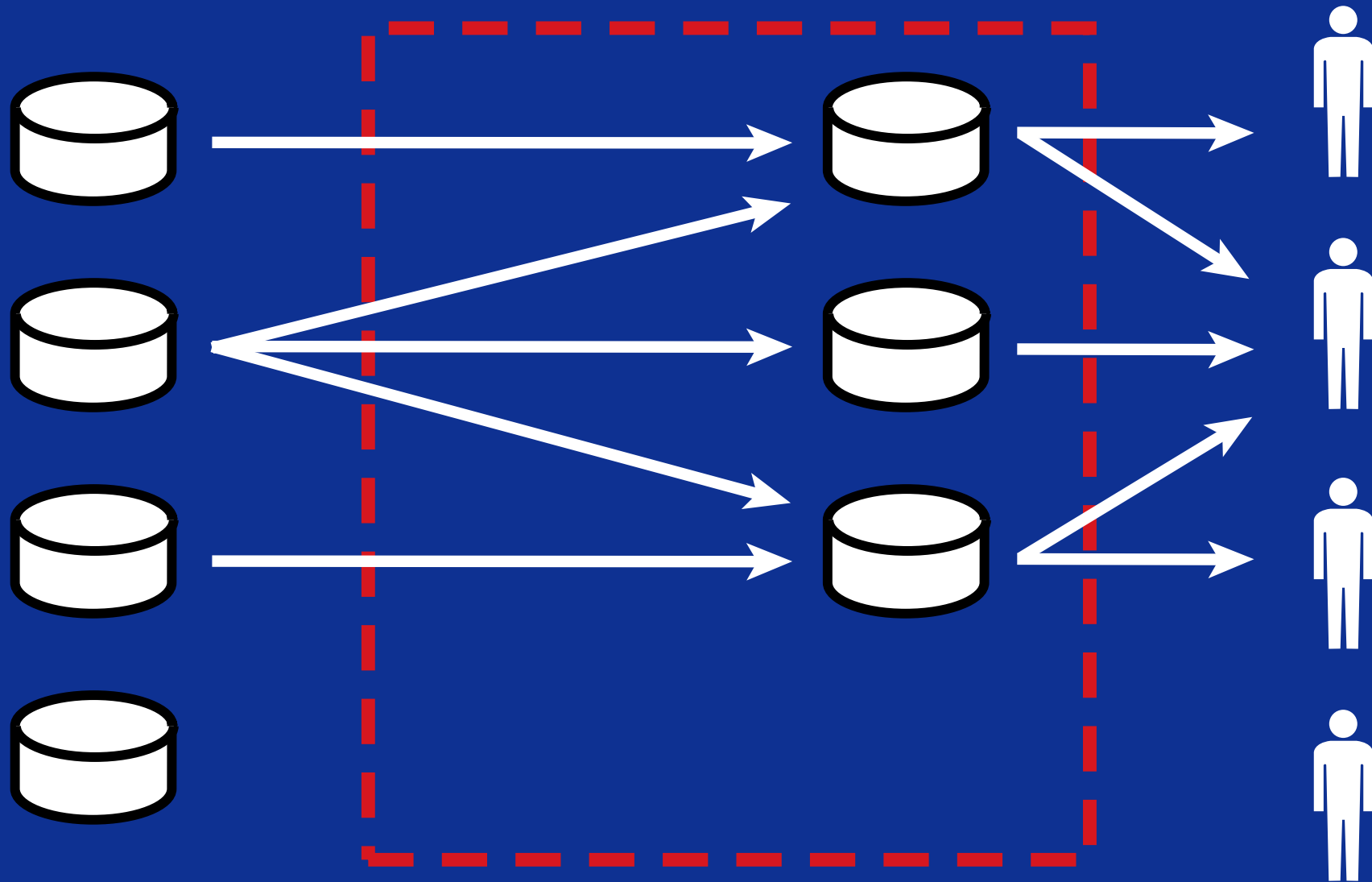
data warehouse

# Data Warehouse Solution (stage 2)



data warehouse

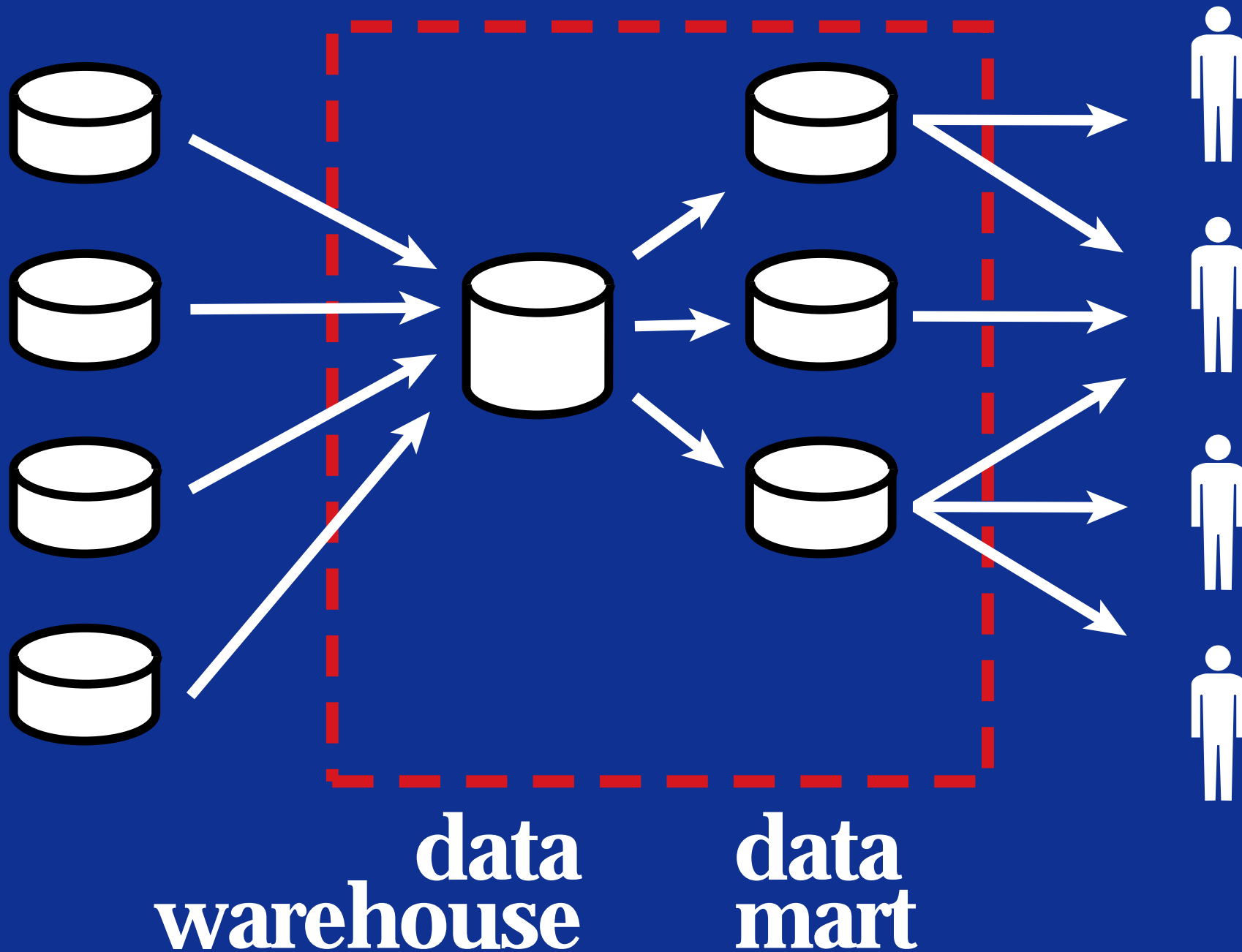
# Data Warehouse Solution (stage 3)



data warehouse



# Data Warehouse Solution (stage 4)



# **Data Warehouse Solution (stage 4)**

---

**The “data warehouse” in the stage 4 solution provides the one, consistent, integrated database that was sought by the “integrated database” solution.**

**All data marts are constructed from the information obtained from that single data warehouse and thus yield consistent answers to all users.**

# Operational Systems

# Decision Support Systems

repetitive decisions

limited data

response intensive

frequent update

operational worker

one of a kind decisions

data intensive

relaxed response

read only

knowledge worker

# OLTP versus OLAP

---

The term “Online Transaction Processing” (OLTP) is used to characterize operational systems.

The term “Online Analytic Processing” (OLAP) is used to characterize the newer decision support systems.

But look out for “data mining.”

# OLAP/ROLAP/MOLAP/DOLAP

---

- **OLAP**
- **Relational - OLAP**
- **Multidimensional - OLAP**
- **Distributed - OLAP**

# **Data Mining**

## **(off-line analytical processing)**

---

**Data mining is executed by batch programs, (demons) which, acting on their own, intelligently survey masses of data looking for significant patterns that are obscured by the sheer mass of the available data.**

# Overview:

---

- decision support premise
- decision support problem
- data warehouse solution
- ✓ • data complexity problem
- “bi- level” data model solution

# All Kinds of Data

---

- **base and summary data**
- **current and archival data**
- **internal and external data**
- **content and context data**
- **real and meta data**



# Base and Summary Data

---

- **last sales transaction**
- **individual sales transactions for the last 24 hours**
- **summary of sales by store, product and hour of day, for the last 24 hours**
- **summary of daily sales, for the last thirteen weeks**
- **summary of monthly sales, by district, for the last ten years**

# Current and Archival Data

---

- **current sales, current inventory, current bank balances, current salaries, etc.**
- **sales transactions for the past ten years**
- **personnel records as they have appeared, change by change, since employment**
- **claims history by insurance policy, since the policy was written**

# Internal and External Data

## Internal Data

- company's truck sales
- company's automobile sales

## External Data

- industry truck sales
- industry automobile sales

# Content and Contextual Data

---

- changed sales district boundaries
- changed product grouping
- changed accounting rules
- changed pricing structure
- changed product packaging
- changed fiscal year

Question:

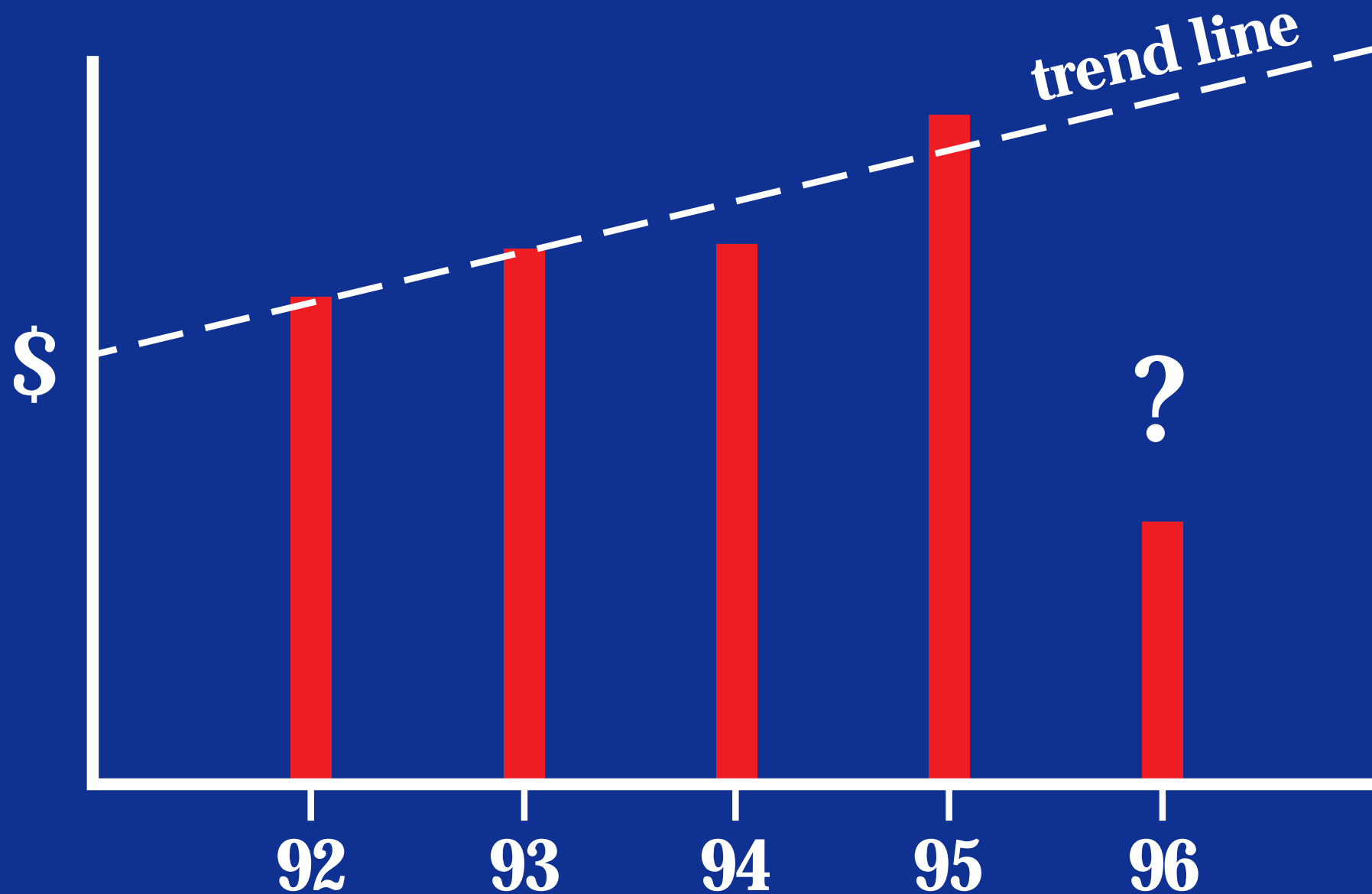
How are the numbers to be understood?

# Seventy Five (75) Units

---

- kilograms?
- gallons?
- feet?
- cubic feet?
- cartons?
- metric tons?
- cars?

# Why were the Chicago district sales so low in 1996?

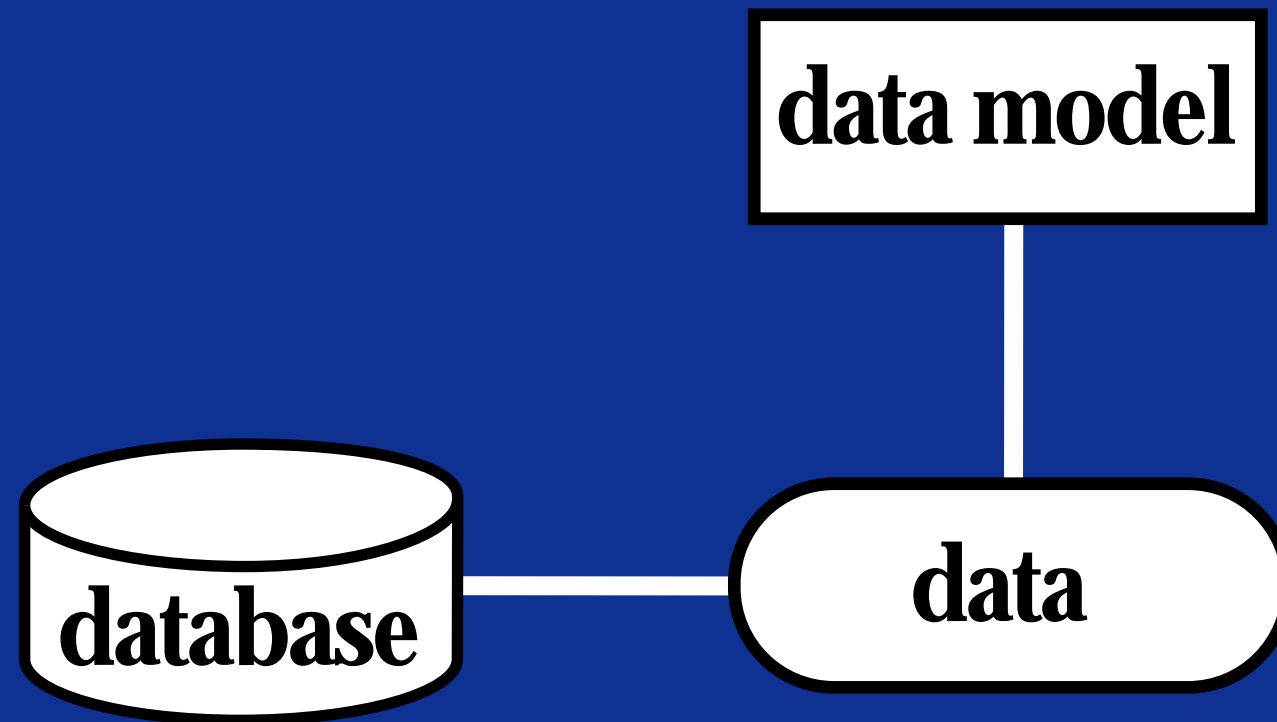


# Real and Meta Data

<b>“meta”</b>	<b>order no</b>	<b>date</b>	<b>quantity</b>
	<b>12498</b>	<b>May 1</b>	<b>10</b>
	<b>12944</b>	<b>May 7</b>	<b>21</b>
<b>“real”</b>	<b>13001</b>	<b>June 6</b>	<b>75</b>
	<b>13749</b>	<b>July 9</b>	<b>15</b>
	<b>14992</b>	<b>Aug 9</b>	<b>23</b>

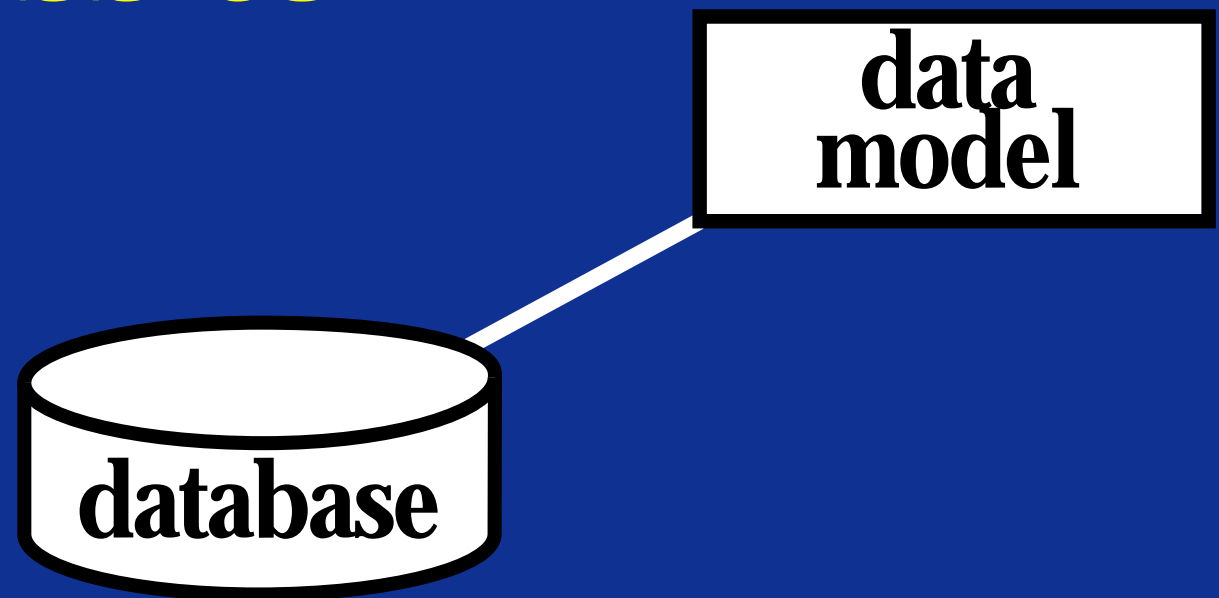
# Data and Data Model

A data model describes the data in a database.

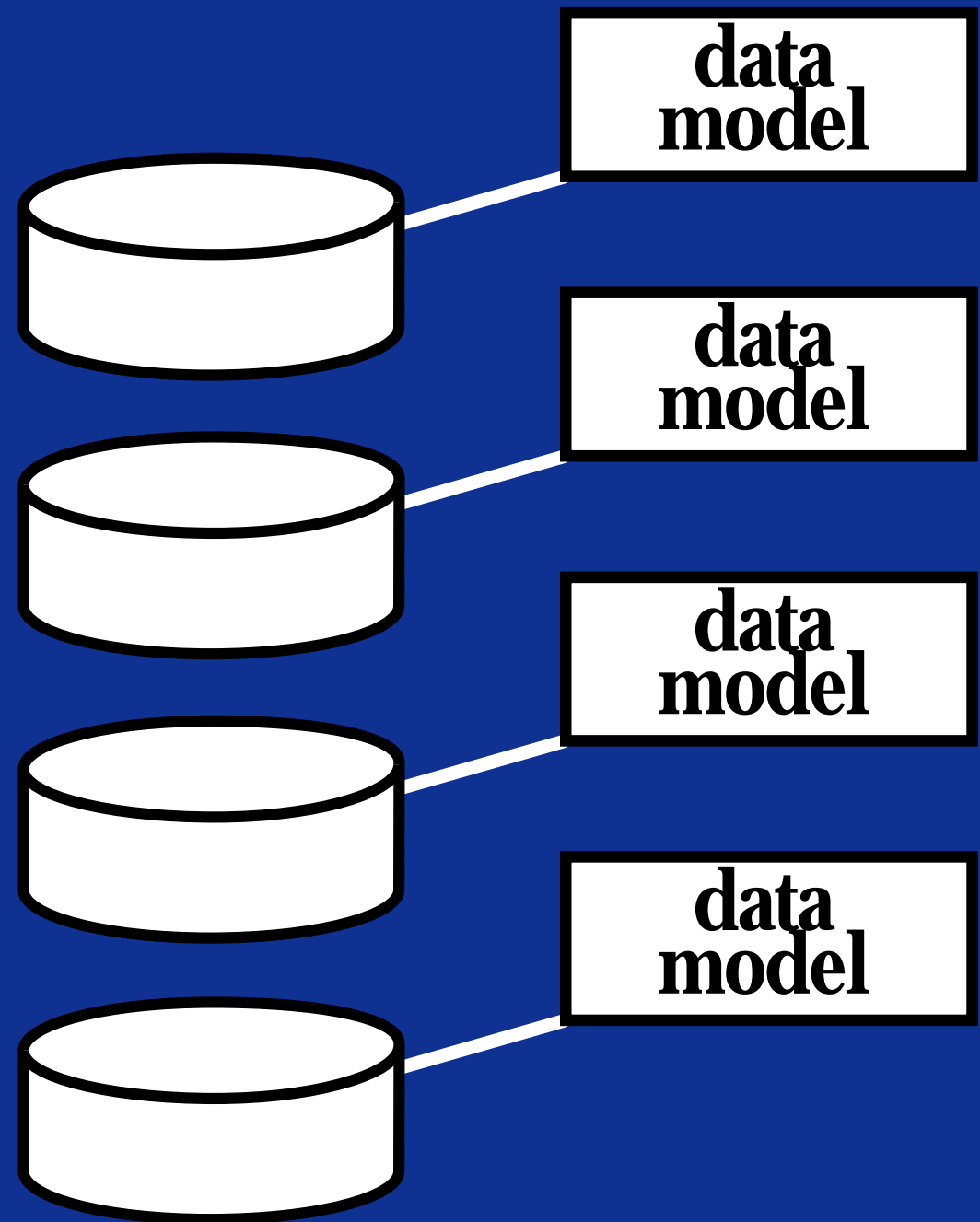




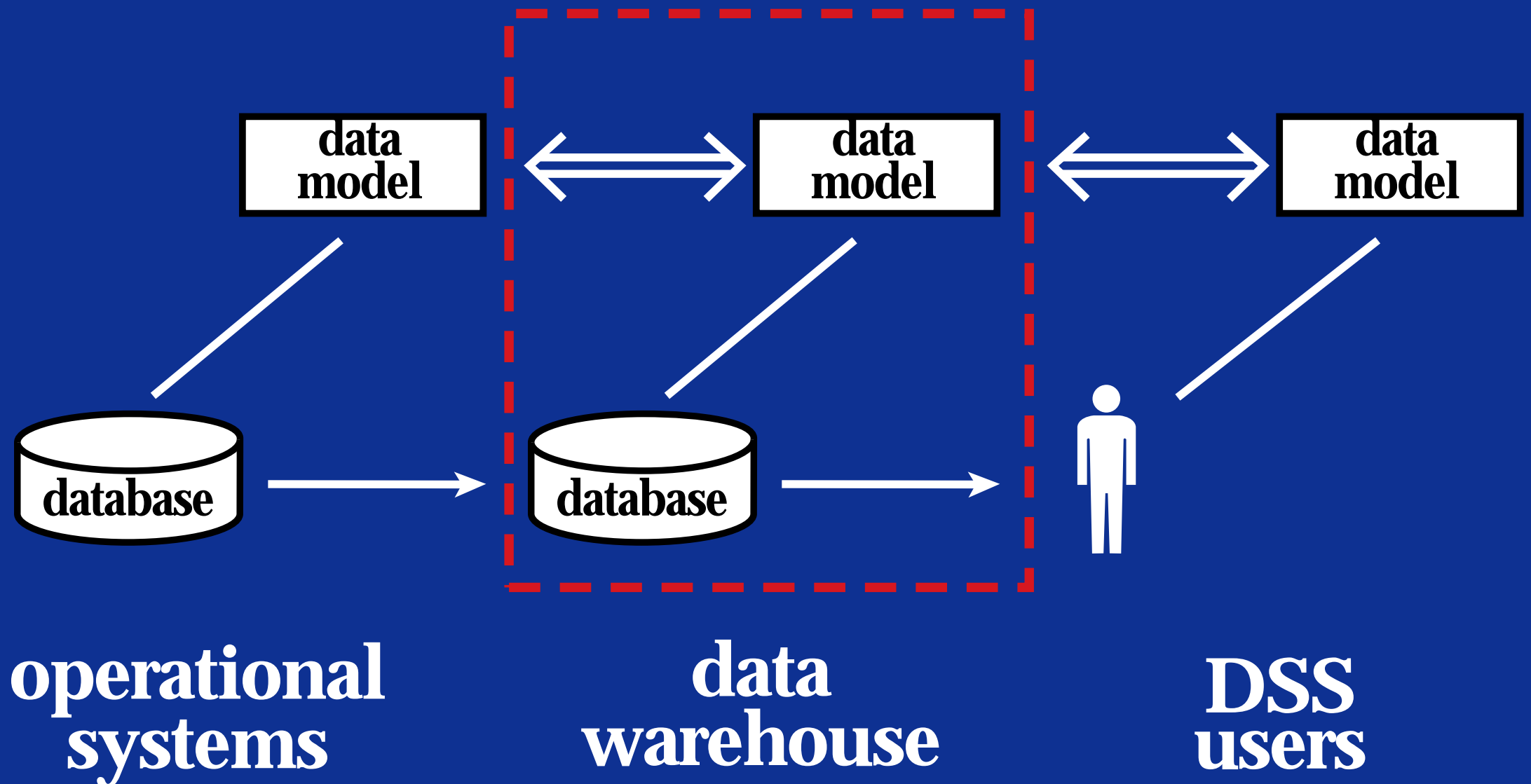
# Data Models, with Their Data Descriptions, Control Access to Databases.



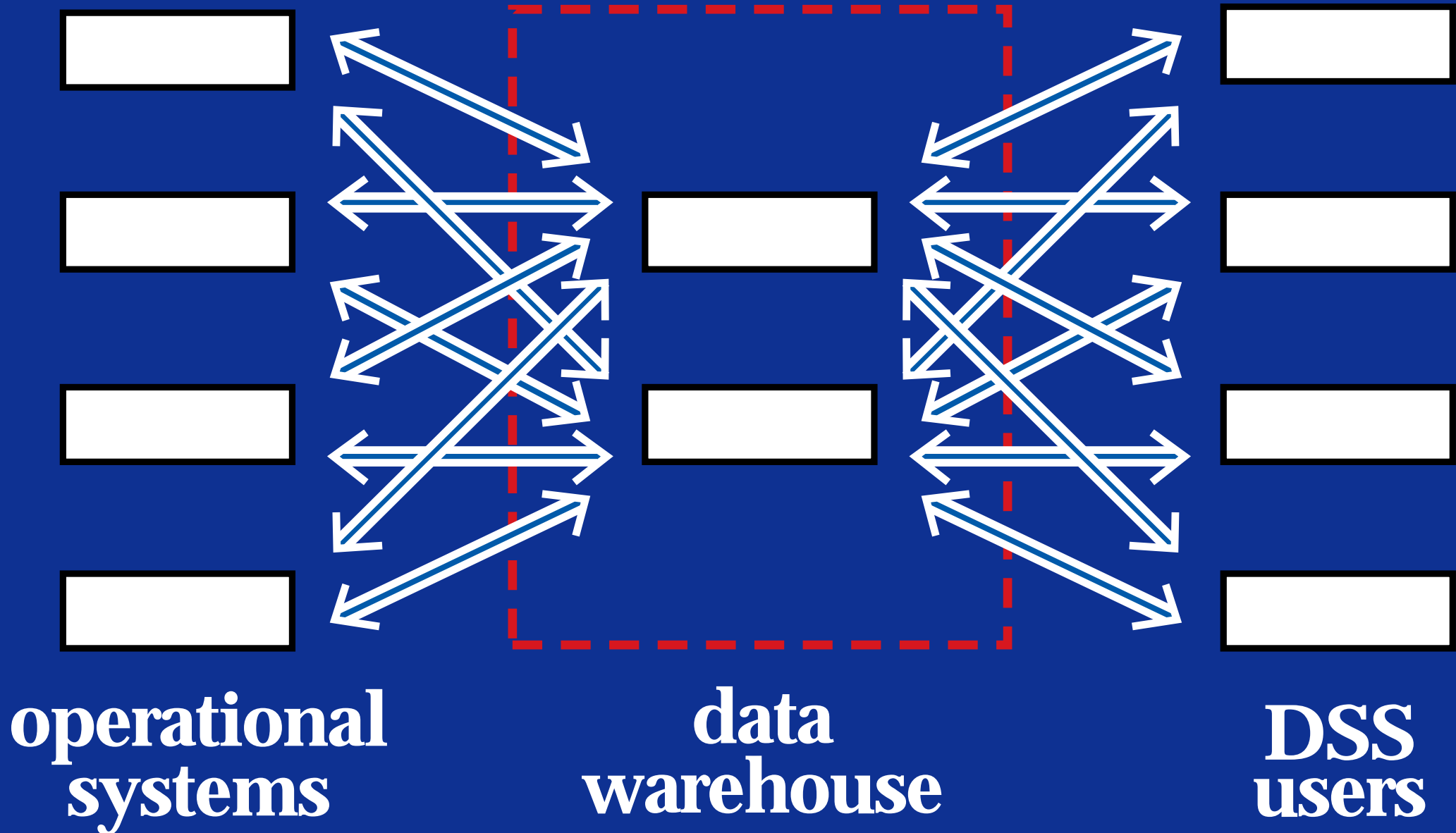
# Operational Databases, with Their Data Models



# Data Transformations and Data Model Maps



# Single Level Data Models and Their Data Model Maps



# ***m* x (*m* - 1) Mappings**

---

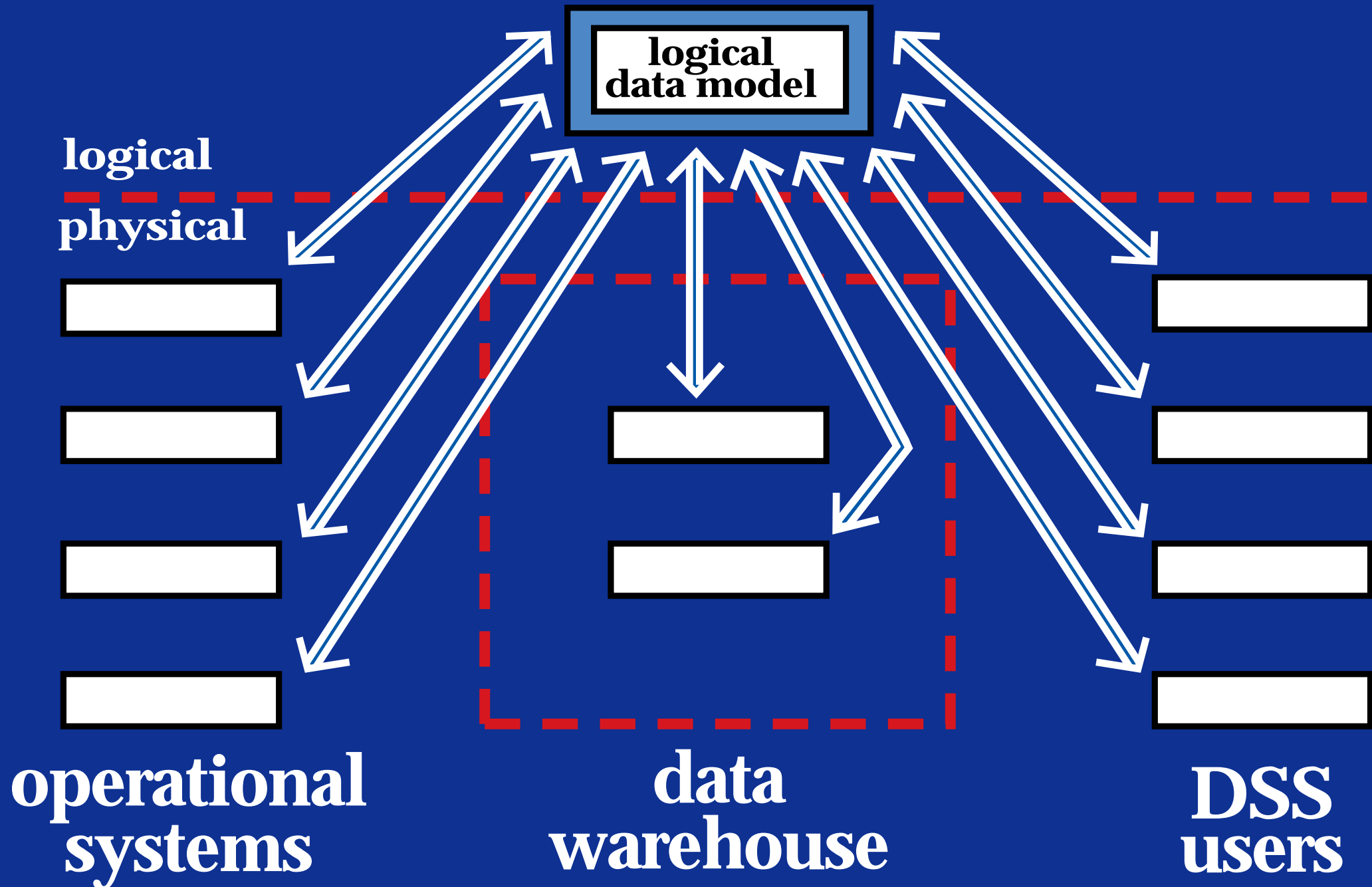
If there are “*m*” data models, then there are potentially *m* x (*m* - 1) mappings to create and maintain.

# Overview:

---

- **decision support premise**
- **decision support problem**
- **data warehouse solution**
- **data complexity problem**
- ✓ • **“bi- level” data model solution**

# Bi-Level Data Models



# “*m*” Mappings

---

With Bi-Level Data Modeling, if there are “*m*” physical data models, there are only the *m* mappings, one between each physical data model and the single logical data model.



# DBA's Challenge

---

**The complexity of the data warehouse solutions, with so many operational systems, external sources, data marts and a single integrating data warehouse, creates a real challenge for the database administrators.**

# Meta Data Focus

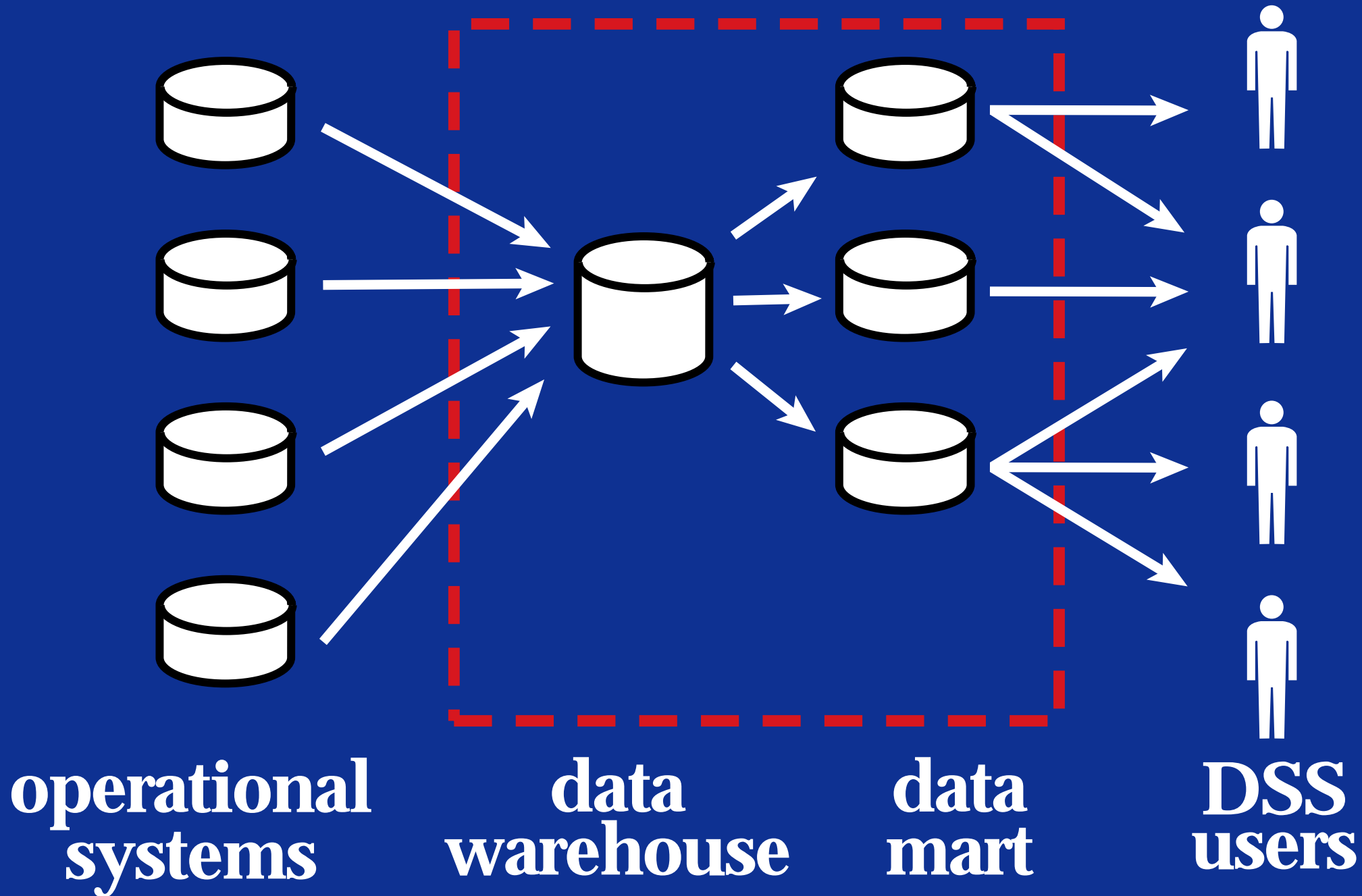
---

**Just keeping track of the various data descriptions used by the various files, databases, programs, and the mappings that join them, is a challenge.**

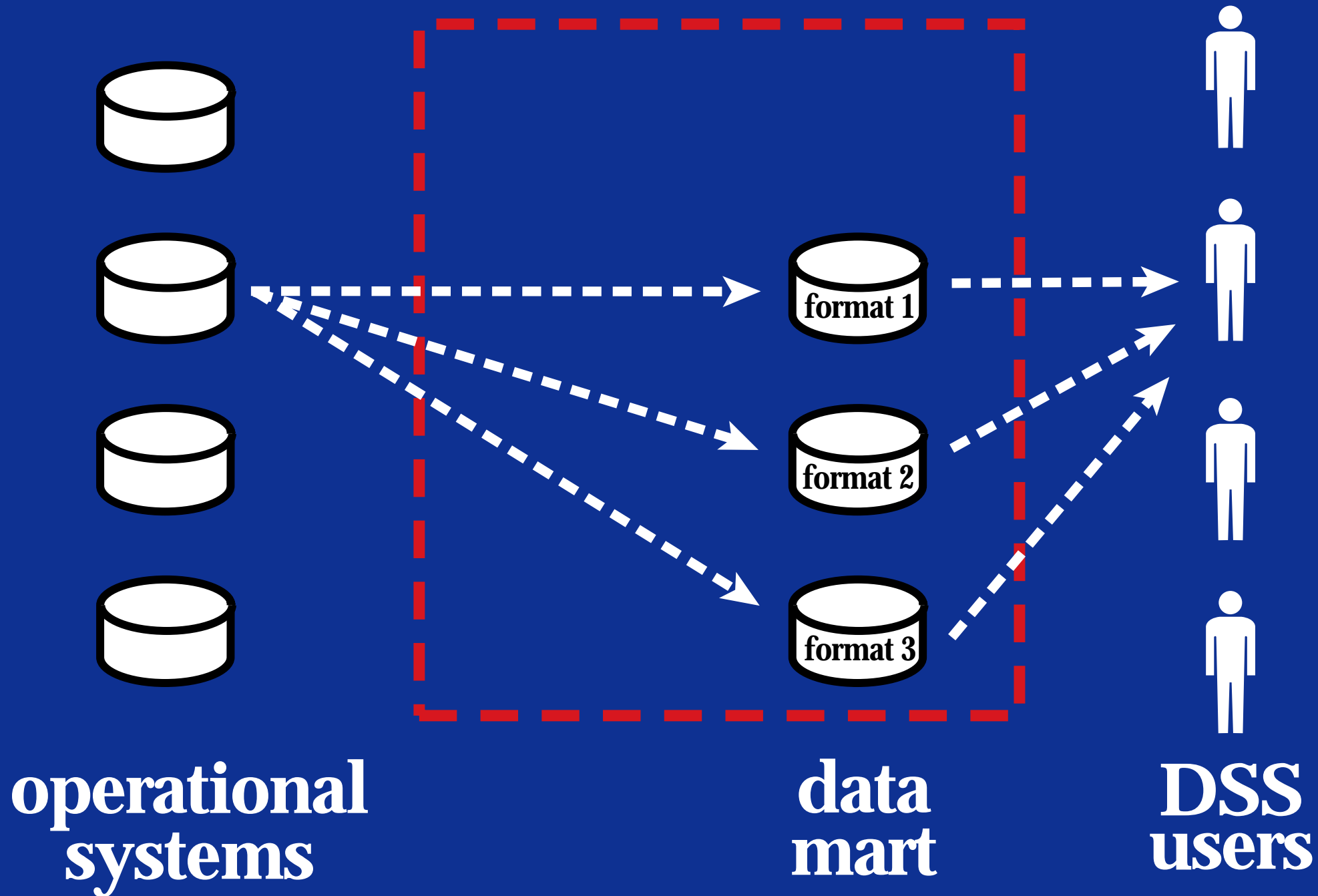
**Bi-level data models must handle a number of programming languages, query languages, and database management systems:**

- **COBOL**
- **PL/1**
- **IMS**
- **IDMS**
- **many varieties of SQL**
- **Object Oriented databases**
- **Multidimensional databases**
- **and whatever might come next**

# Data Warehouse (stage 4)



# Data Warehouse Solution (stage 1)



**The “bi-level data model” approach operates on the well known principle of “abstraction.”**

**It uses a single, higher level logical data model (conceptual schema) to understand the semantics of differing physical level data models.**

# Levels of Abstraction

**Logical Data Model**

**Physical Data Model**

**Machine Model**

**Micro Code Model**

**Physical Model**

# Meta Data Challenge

---

While the single, integrated database solution to operational and decision support systems seems continually to be beyond reach, the single, integrated meta data model to support the data warehouse solution is available today.



# Logical Data Model

---

- **key to bi-level data modeling**
- **provides platform independence**
- **target for reverse engineering existing databases and files**
- **source for forward engineering to new databases**
- **synchronizes logical and physical descriptions**

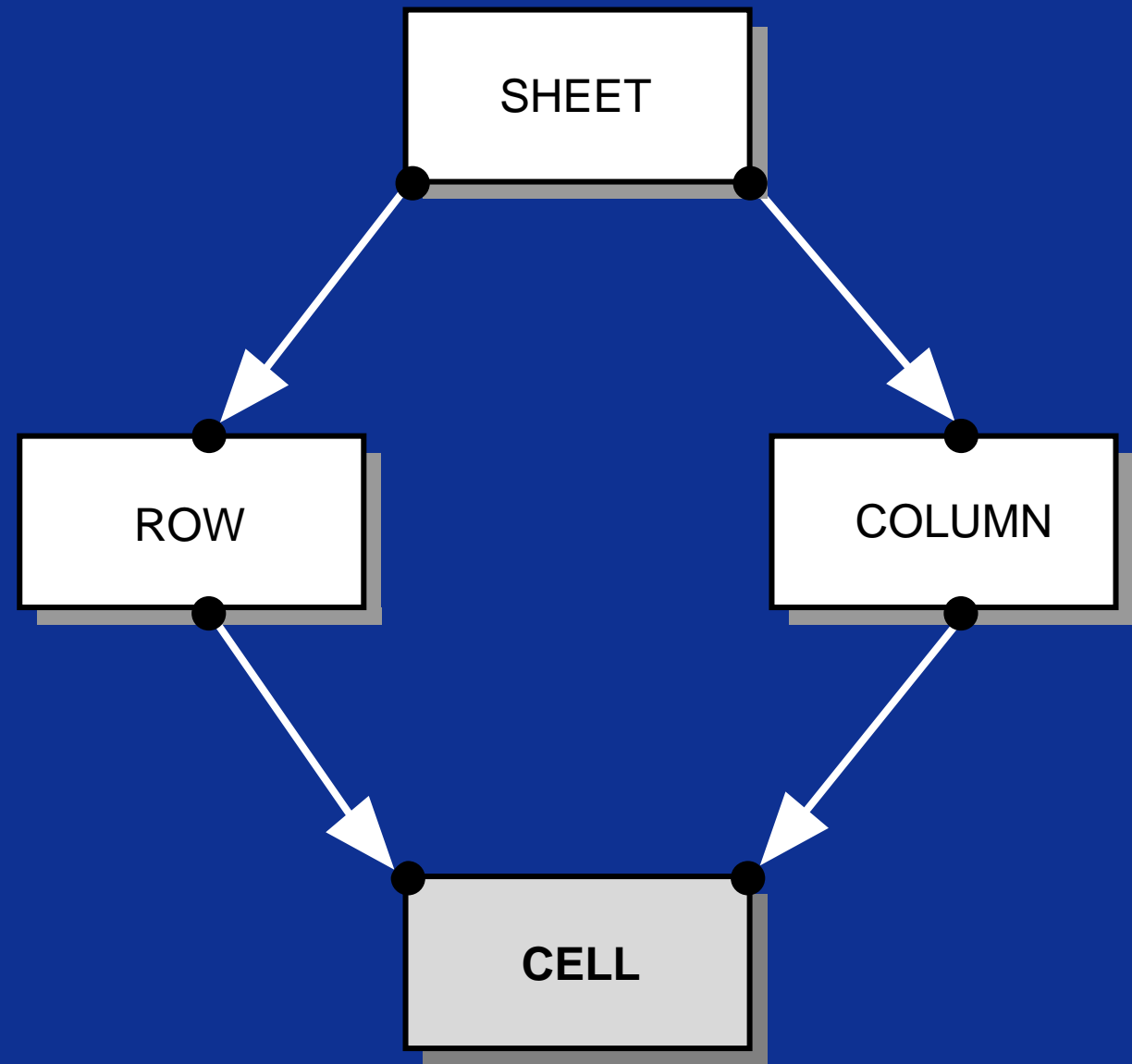
# Physical Models for Data Warehousing

---

- **spread sheets**
- **multidimensional schema**
- **star schema**
- **snowflake schema**

# Spread Sheet

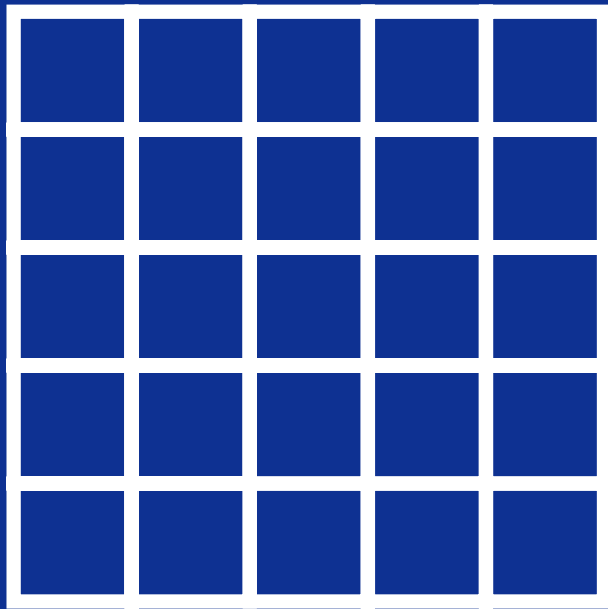
	a	b	c	d
1				
2				
3				
4				



**1 - fact**

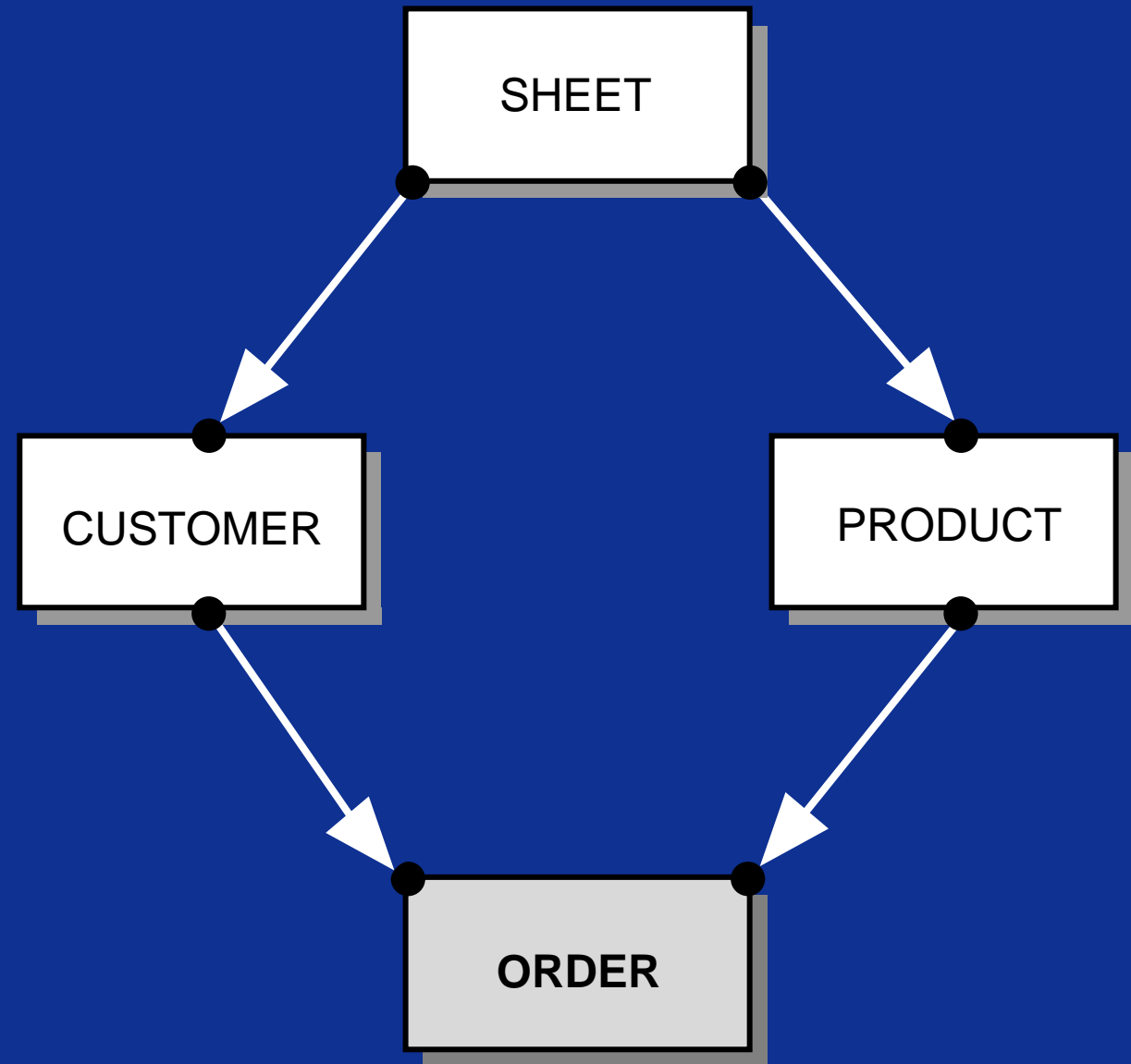
**2 - dimensions**

# Spread Sheet

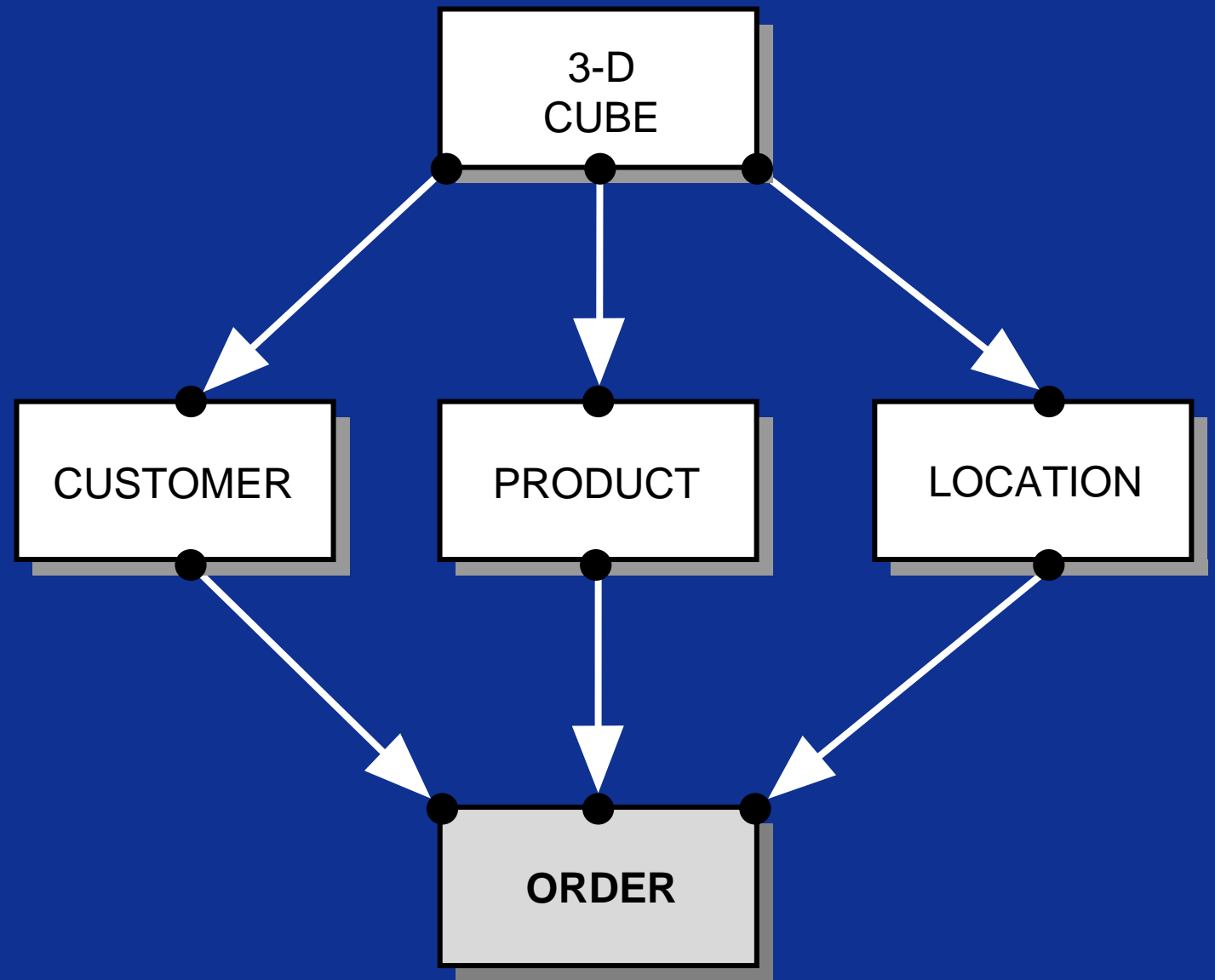
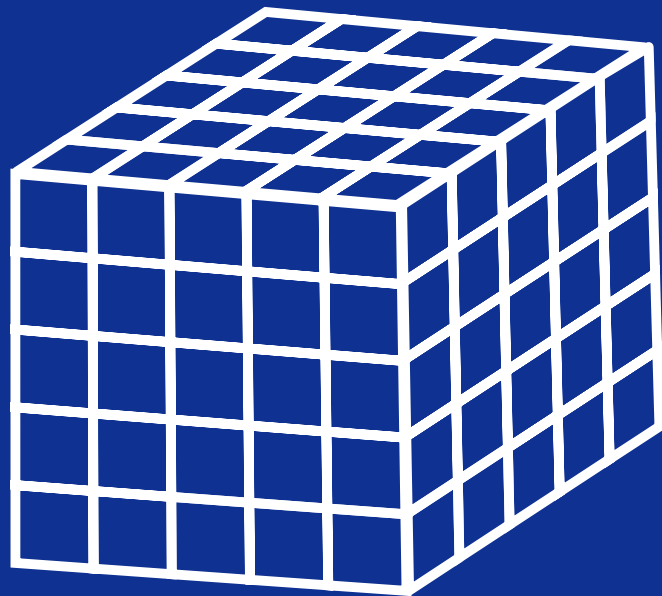



**1 - fact**

**2 - dimensions**



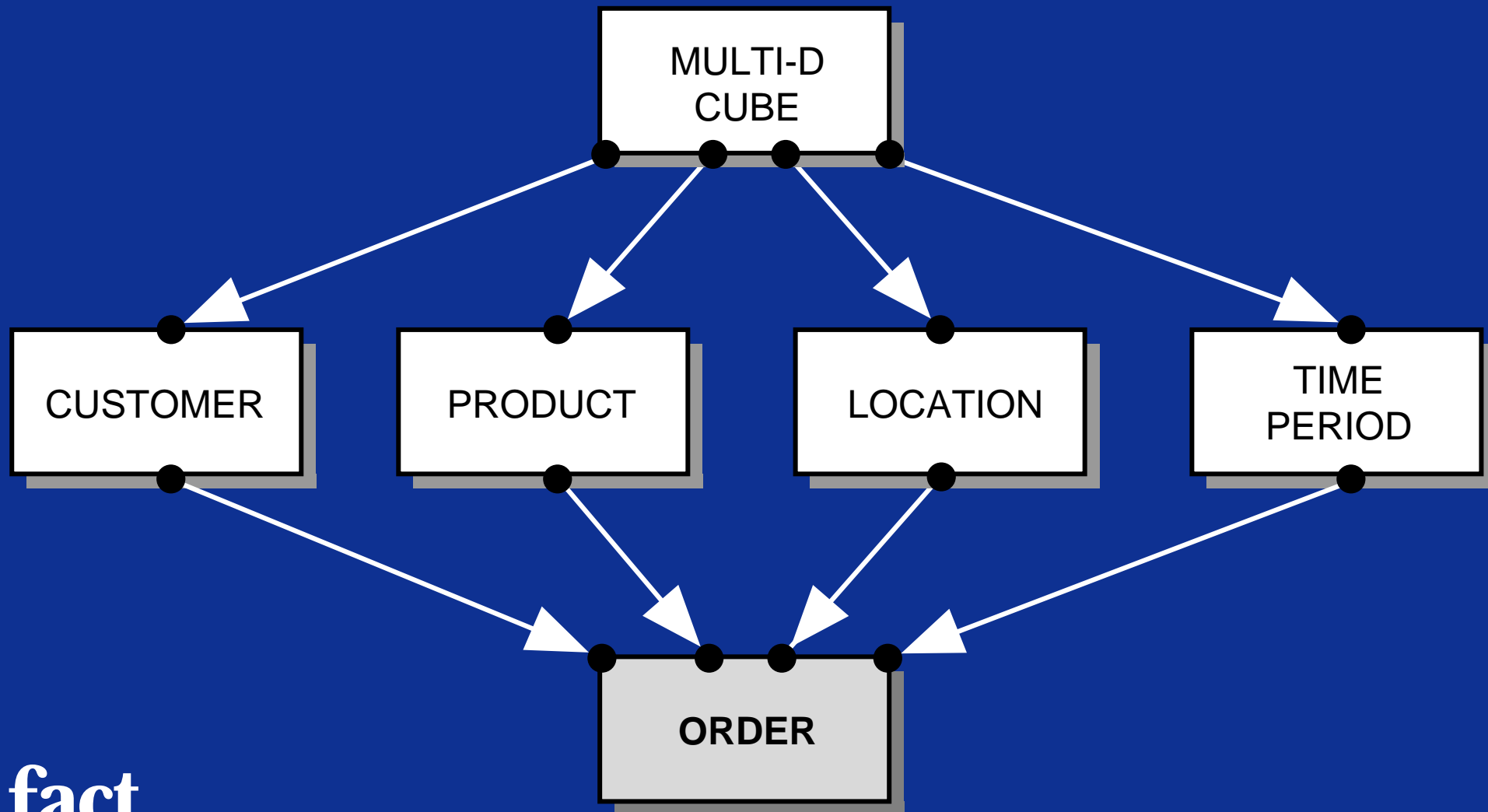
# Cube



**1 - fact**

**3 - dimensions**

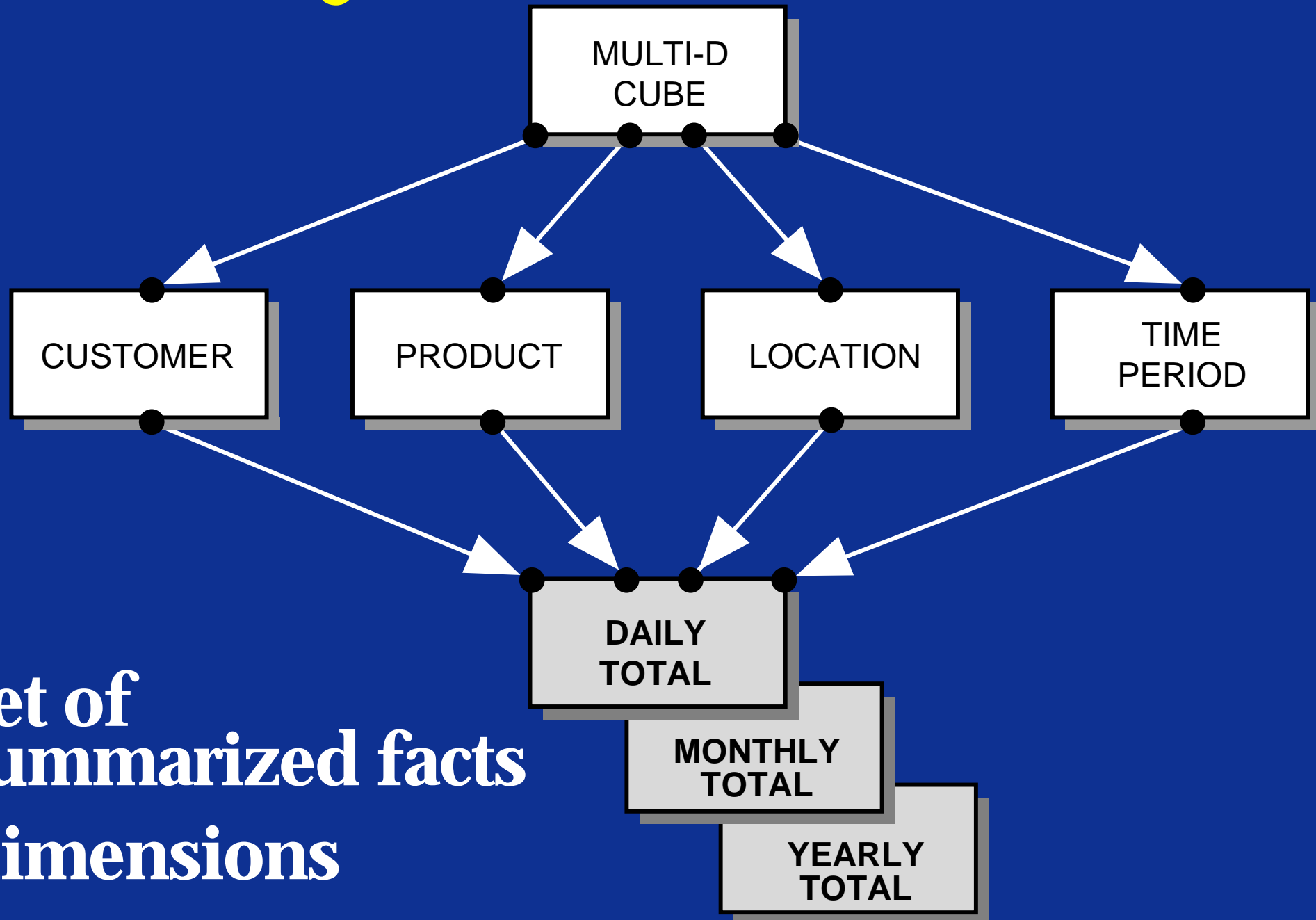
# Multidimensional Cube



**1 - fact**

**4 - dimensions, or more**

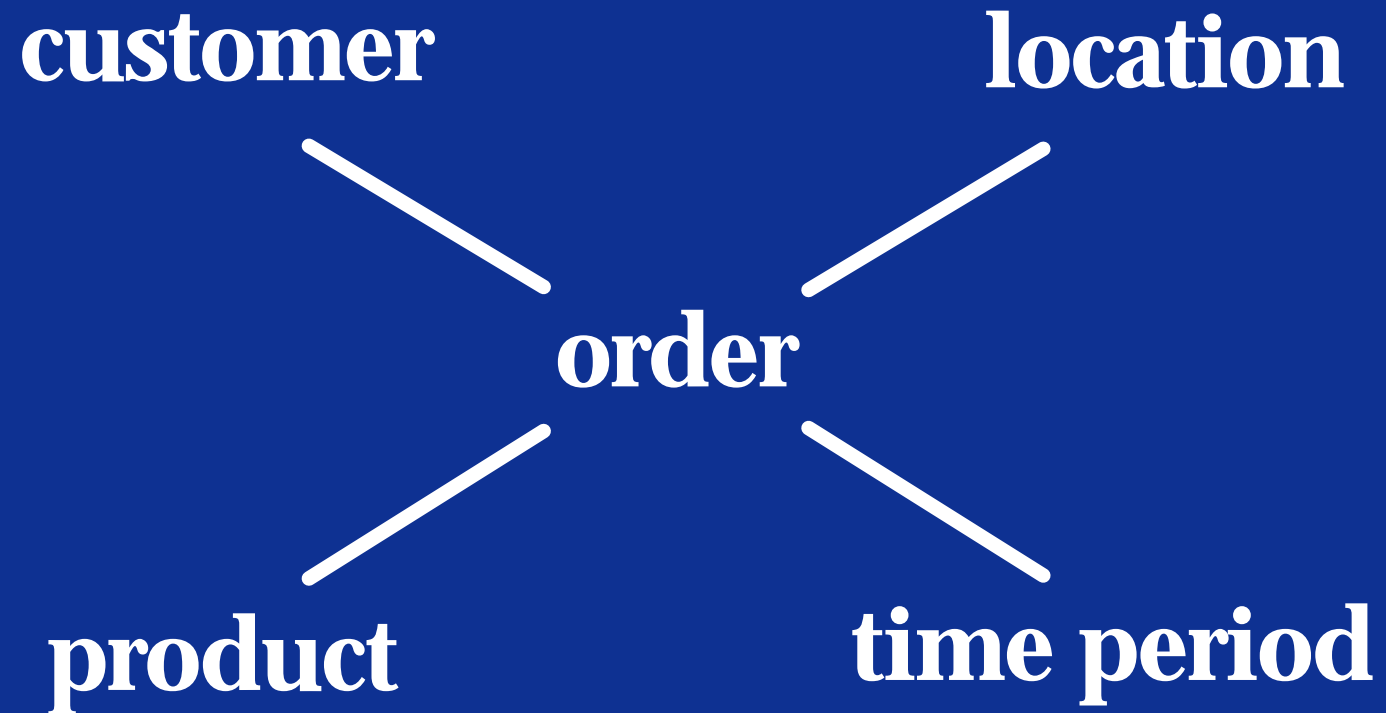
# Summary Data in the Cube



**1 - set of  
summarized facts**

**4 - dimensions**

# Star Schema

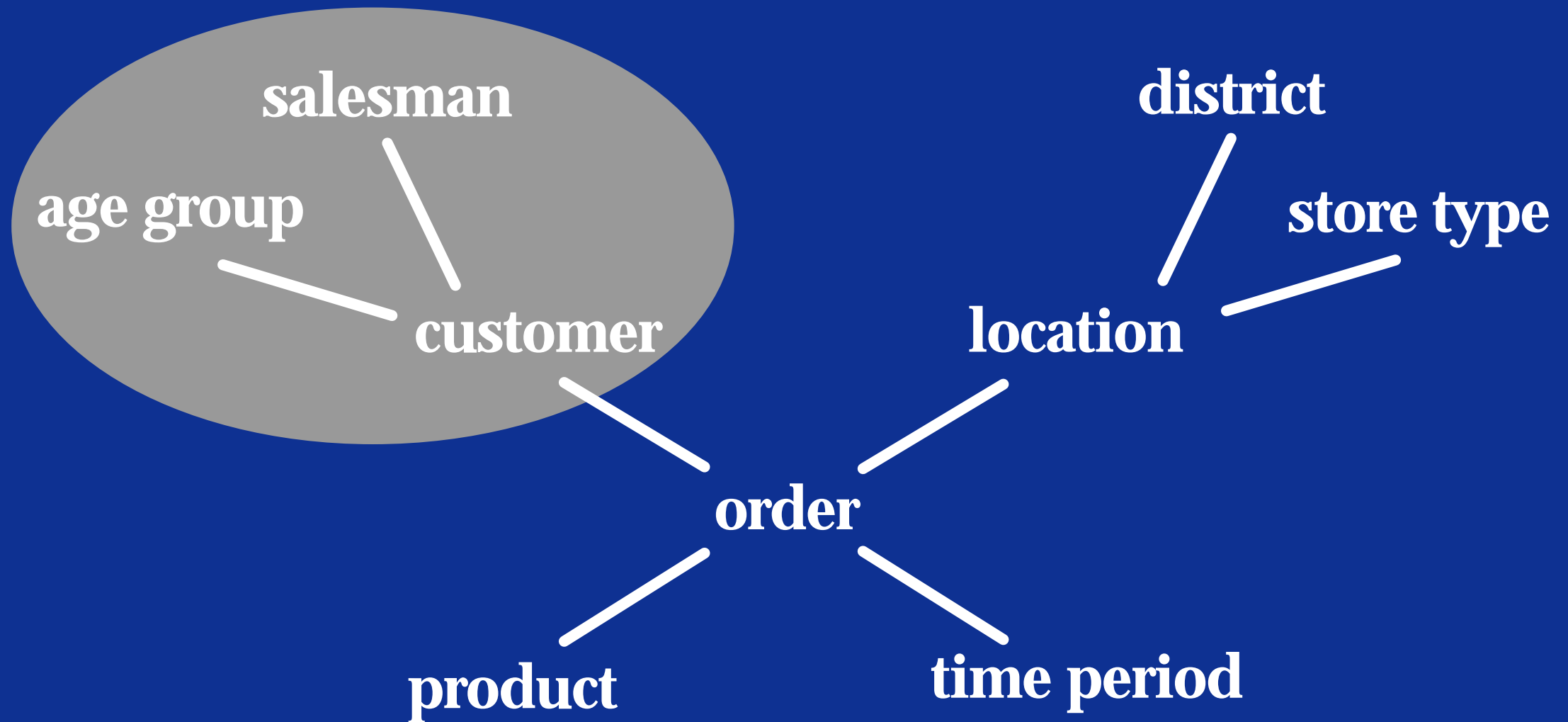


**1 - fact**

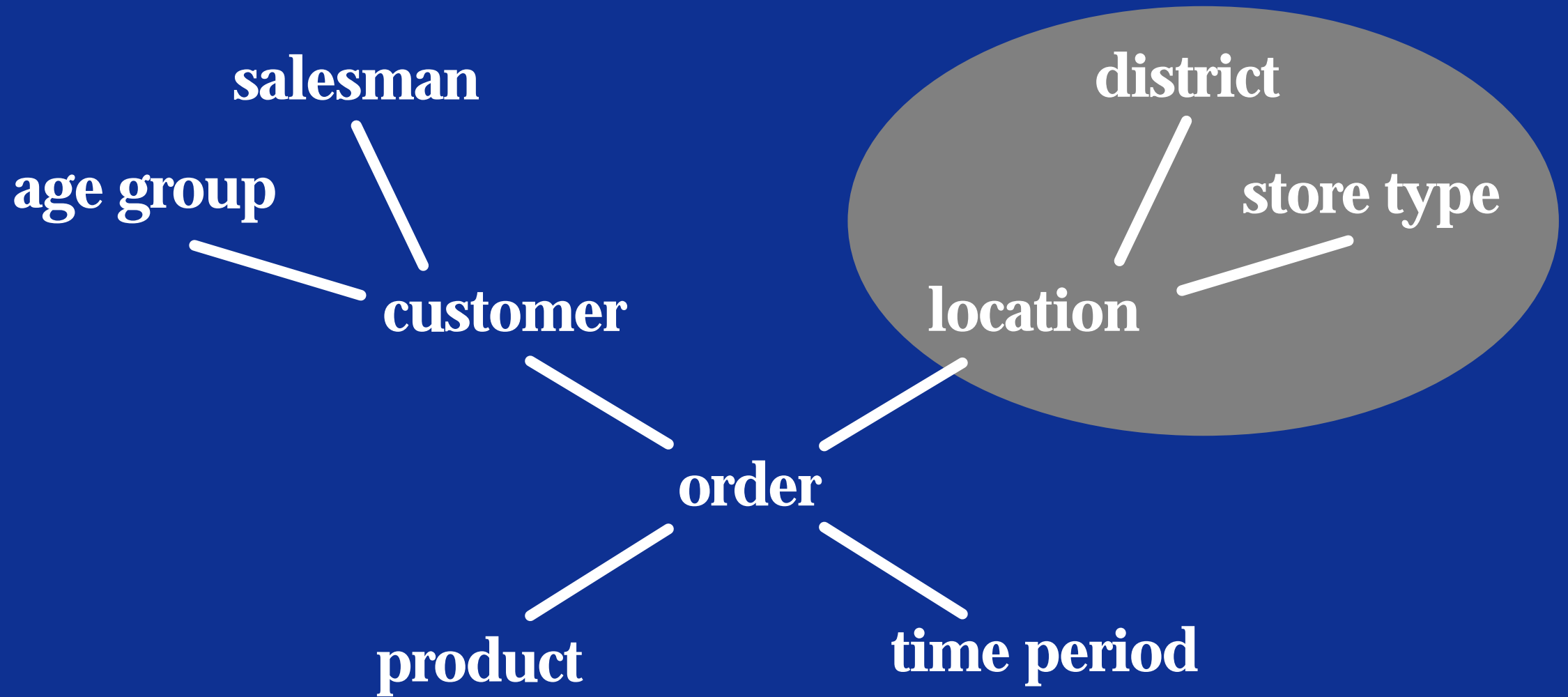
**4 - dimensions**



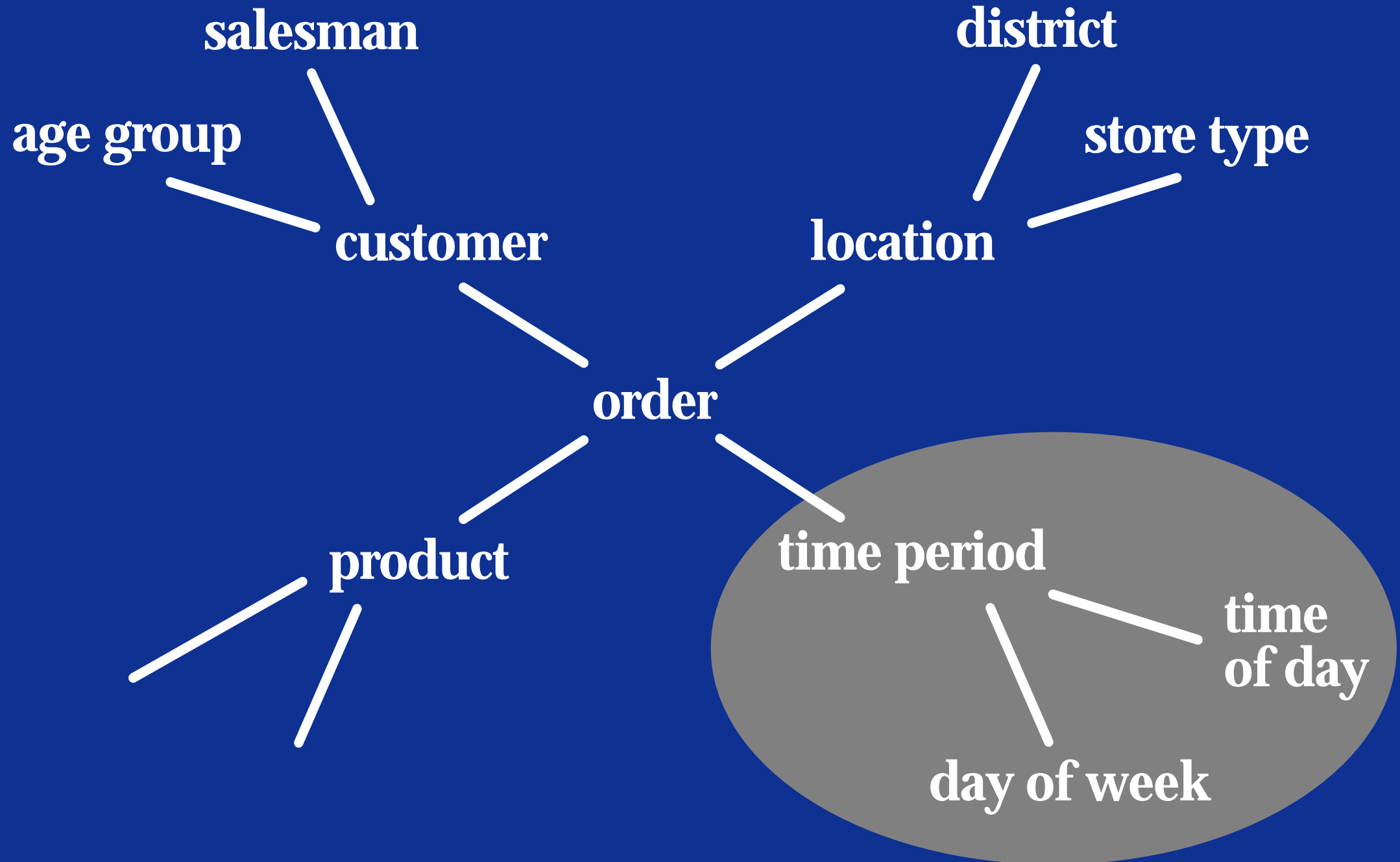
# Snow Flake Schema



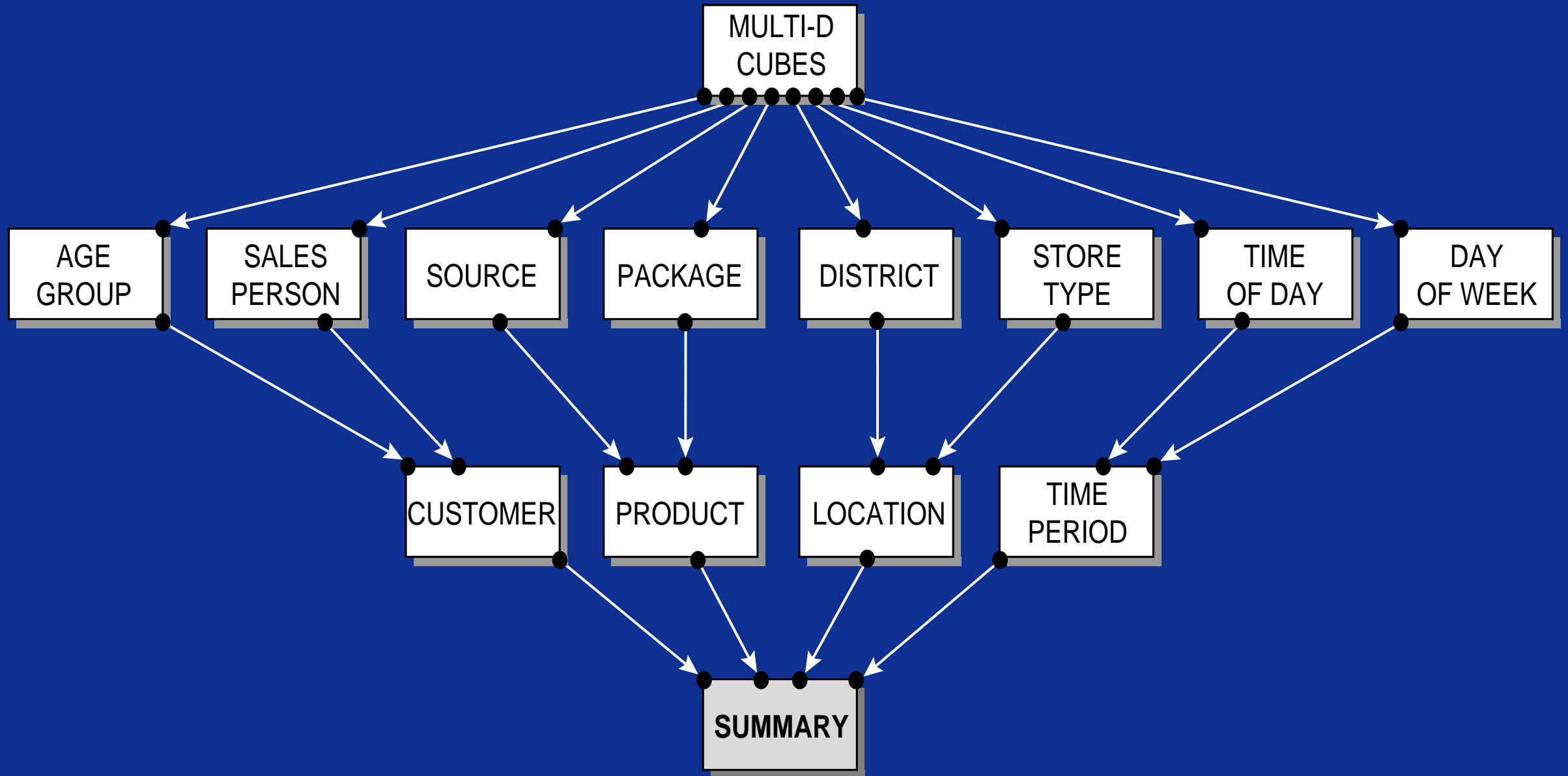
# Snow Flake Schema



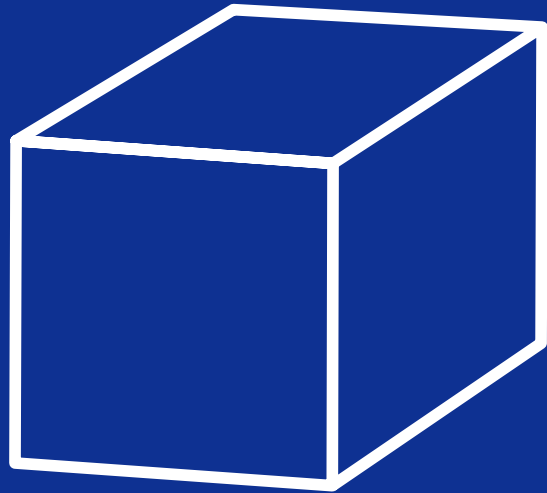
# Snow Flake Schema



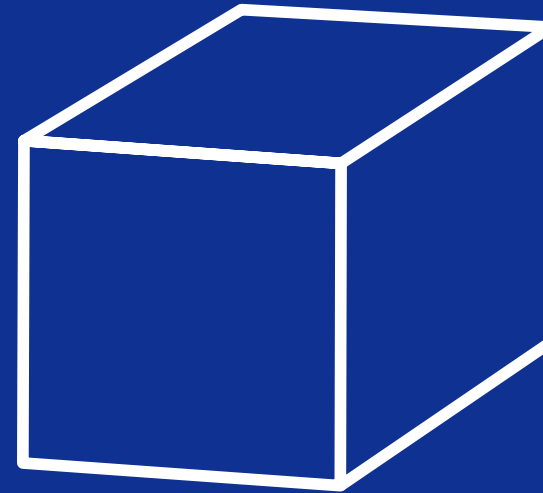
# Snow Flake Schema



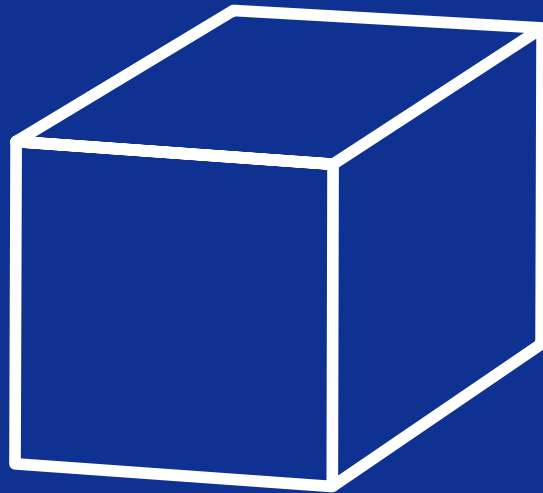
# Multiple Multi-Cube Databases



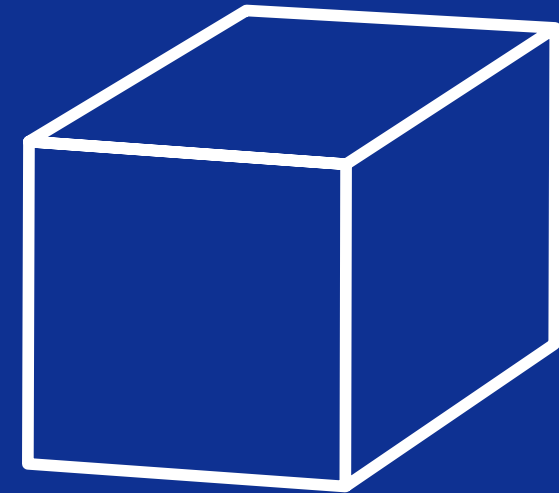
**orders**



**purchases**



**production**



**collections**

**The OLAP/MOLAP/ROLAP/DOLAP view of the business world is a narrow view, designed and organized to answer questions about a specified type of fact, quickly and easily. It offers a rifle shot into the designated collection of corporate data.**

**Multi-disciplinary questions crossing from one star schema or multidimensional cube to another generally require separate queries by the person raising the question.**

# Summary: Bi-Level Data Model Architecture

---

- records the flow of data,
- defines the necessary data transformations,
- assures data consistency, and
- catalogues the available data for decision support workers.