

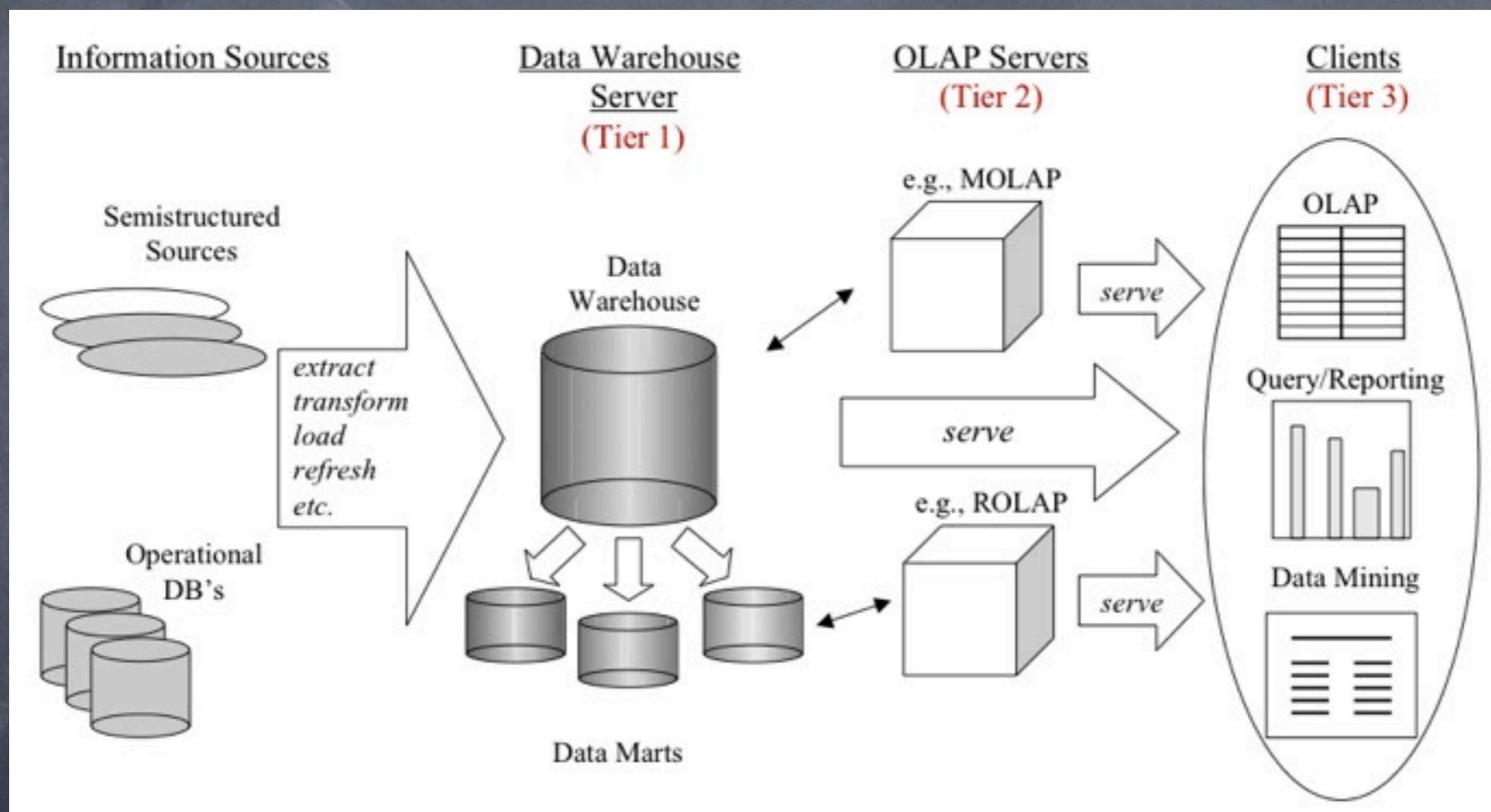
Online Analytical Processing

Vaibhav Bajpai

Decision Support Systems

- ⦿ Architecture
 - ⦿ Information Sources
 - ⦿ Data Warehouse
 - ⦿ OLAP Servers
 - ⦿ OLAP Clients

DSS Architecture



DSS are used to make business decisions based on data collected by OLTP.

Architecture Information Sources

- ⦿ Operational Databases
- ⦿ ERP system
- ⦿ Semi-Structured Sources
 - ⦿ does not conform with formal structure of relational data models but contains tags to separate semantic elements (XML).

Architecture

ETL

- ⦿ Extract-Transform-Load is a process involving
 - ⦿ Extracting data from information sources.
 - ⦿ Transforming to fit operational needs. (cleansing)
 - ⦿ Loading it into the Data Warehouse.

Architecture Data Warehouse

- ⦿ DW is a repository of data to support management decision making process.
 - ⦿ Characteristics
 - ⦿ Basic Elements
 - ⦿ Conceptual Models
 - ⦿ Data Marts

Data Warehouse Characteristics

- ⦿ Subject-Oriented
- ⦿ Integrated (security, single-version)
- ⦿ Time Variant (particular time-period)
- ⦿ Non-Volatile (never removed!)

Data Warehouse Basic Elements

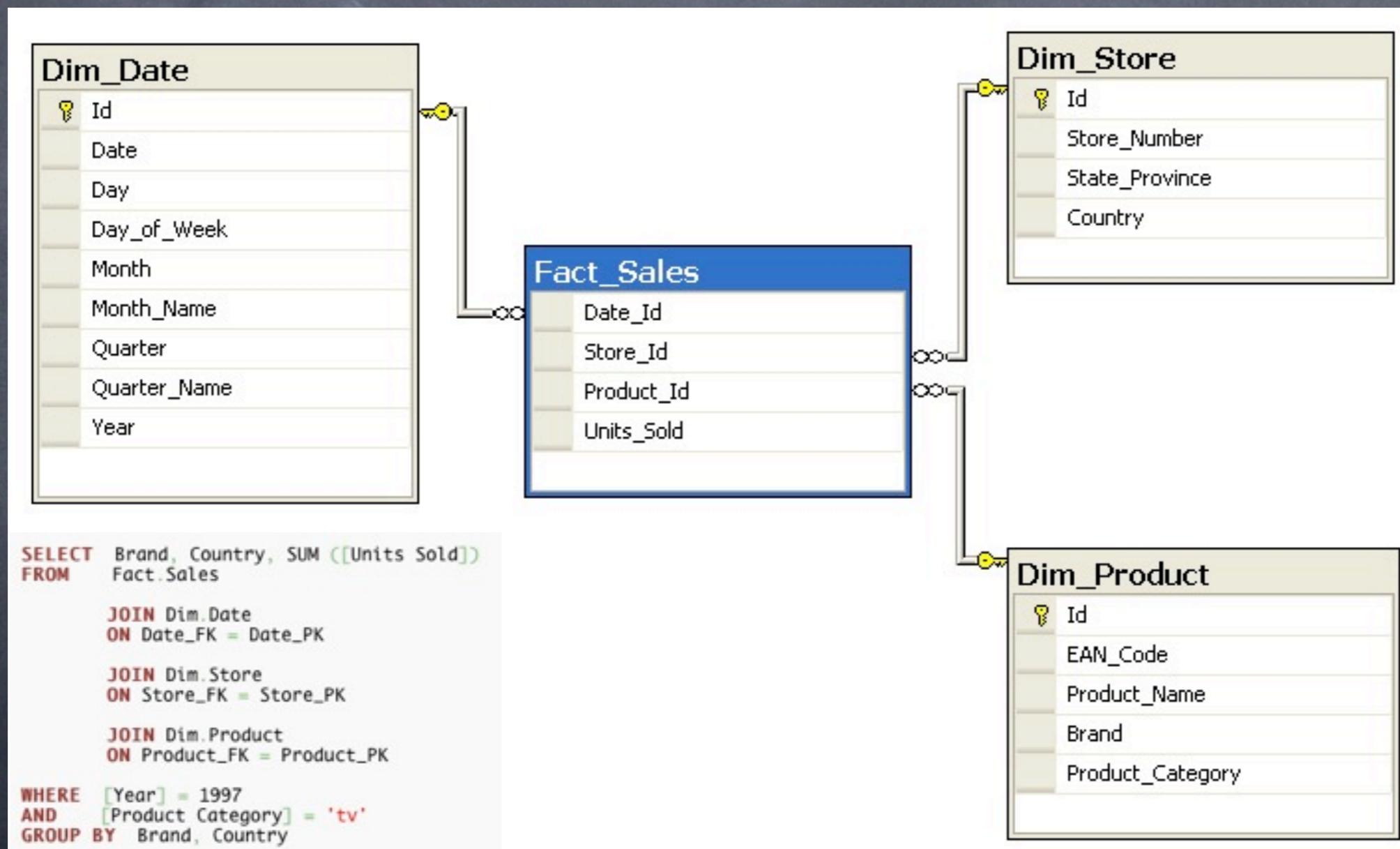
- ⦿ Facts (measures + fKeys to dimTables)
- ⦿ Measures (additive + non-additive + semi-additive)
- ⦿ Dimensions (measures from different perspective)
- ⦿ Hierarchies (classification of dimensions)

Data Warehouse Conceptual Models

- ⦿ Star Schema
- ⦿ Snowflake Schema (normalized)
- ⦿ Galaxy Schema (many fact tables)

Conceptual Models

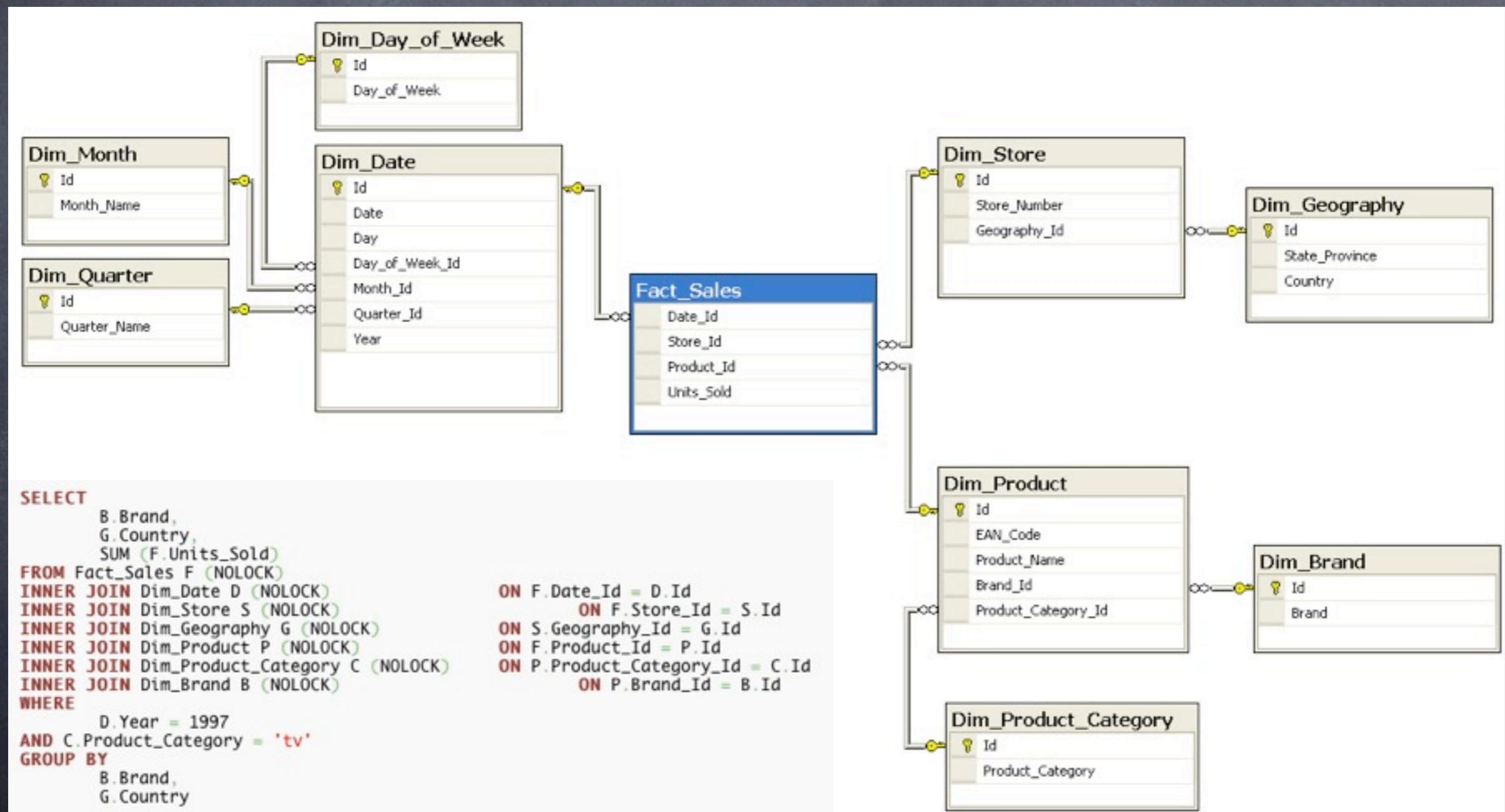
Star Schema



good for large data warehouses!

Conceptual Models

Snowflake Schema



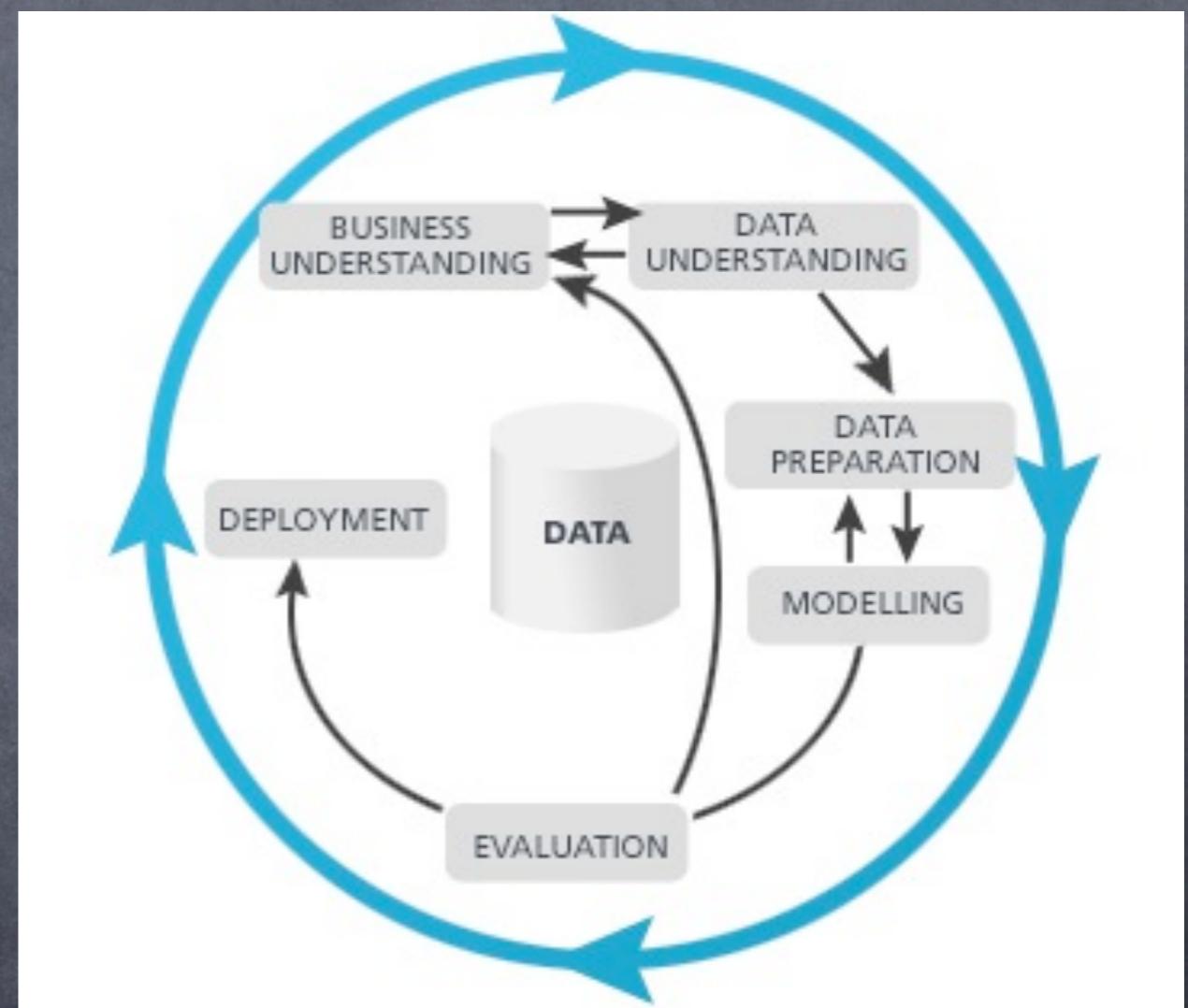
good for small data warehouses!

Data Warehouse Data Marts

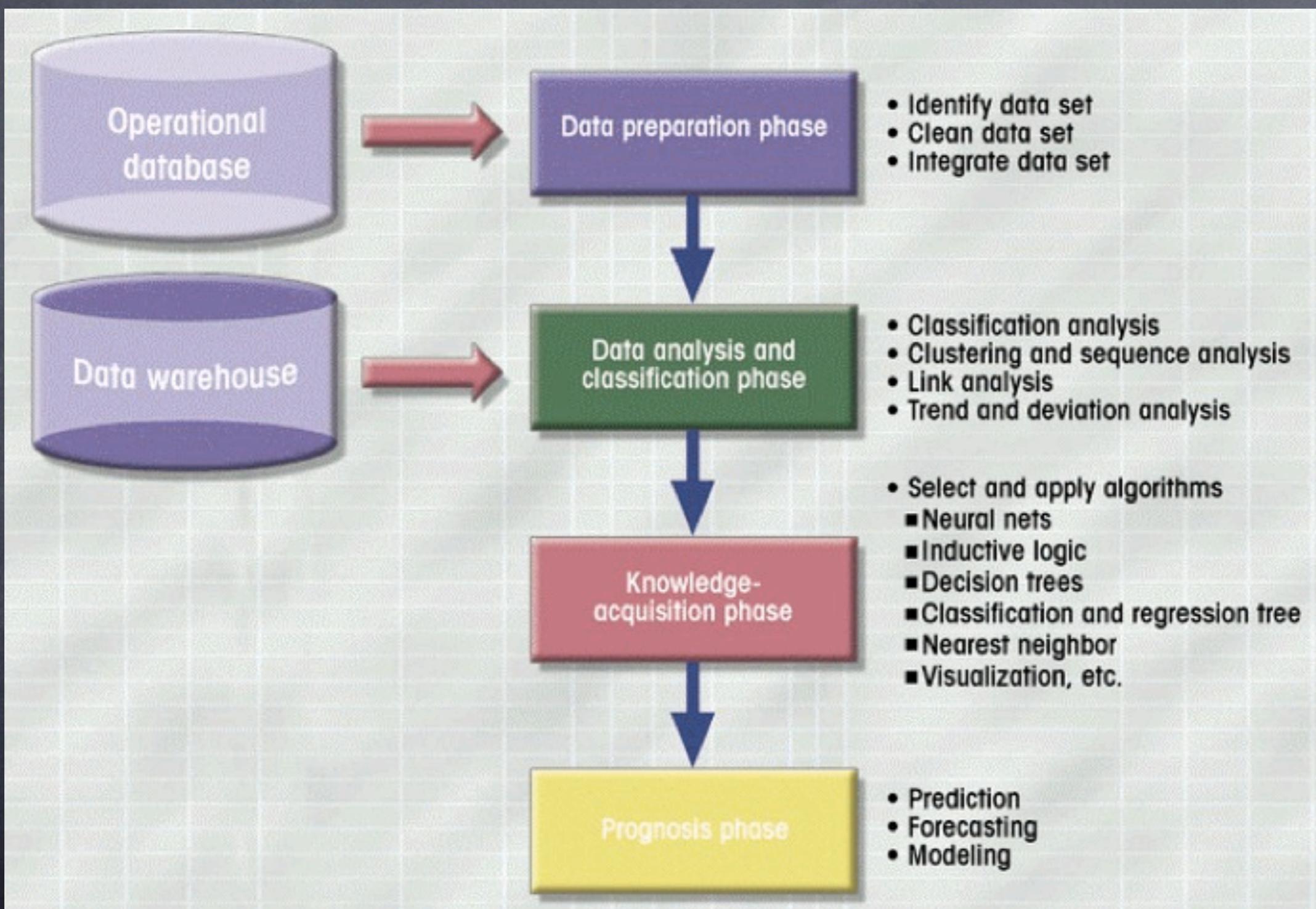
- ⦿ DM is a subset of DW oriented to specific business line or team.
- ⦿ It is an access layer of DW used to get data out to the users.

Data Mining

- process of extracting patterns from data.
- transforms data into business intelligence.



Data Mining

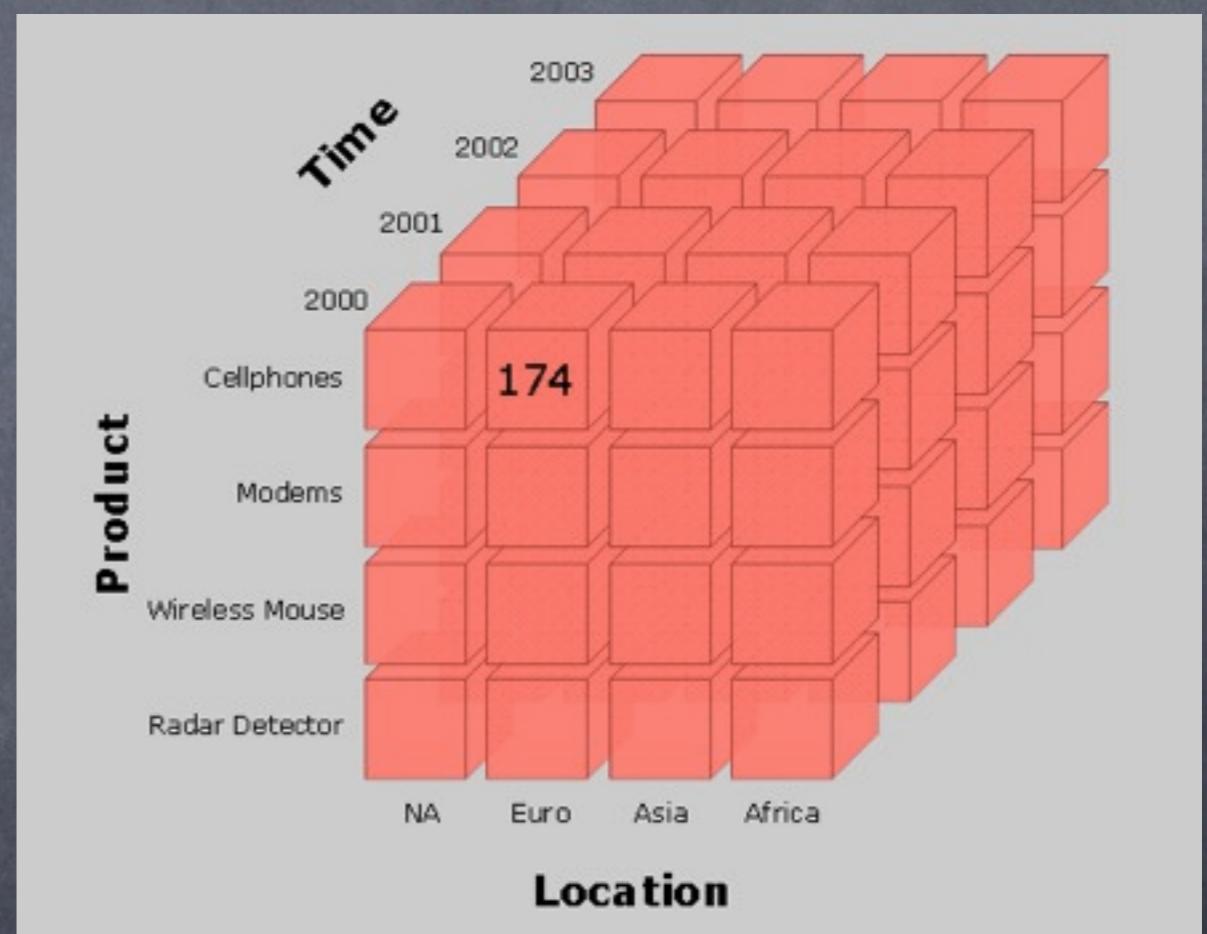


OLAP

- ⦿ Data Cube
- ⦿ Aggregation
- ⦿ Navigational Operations

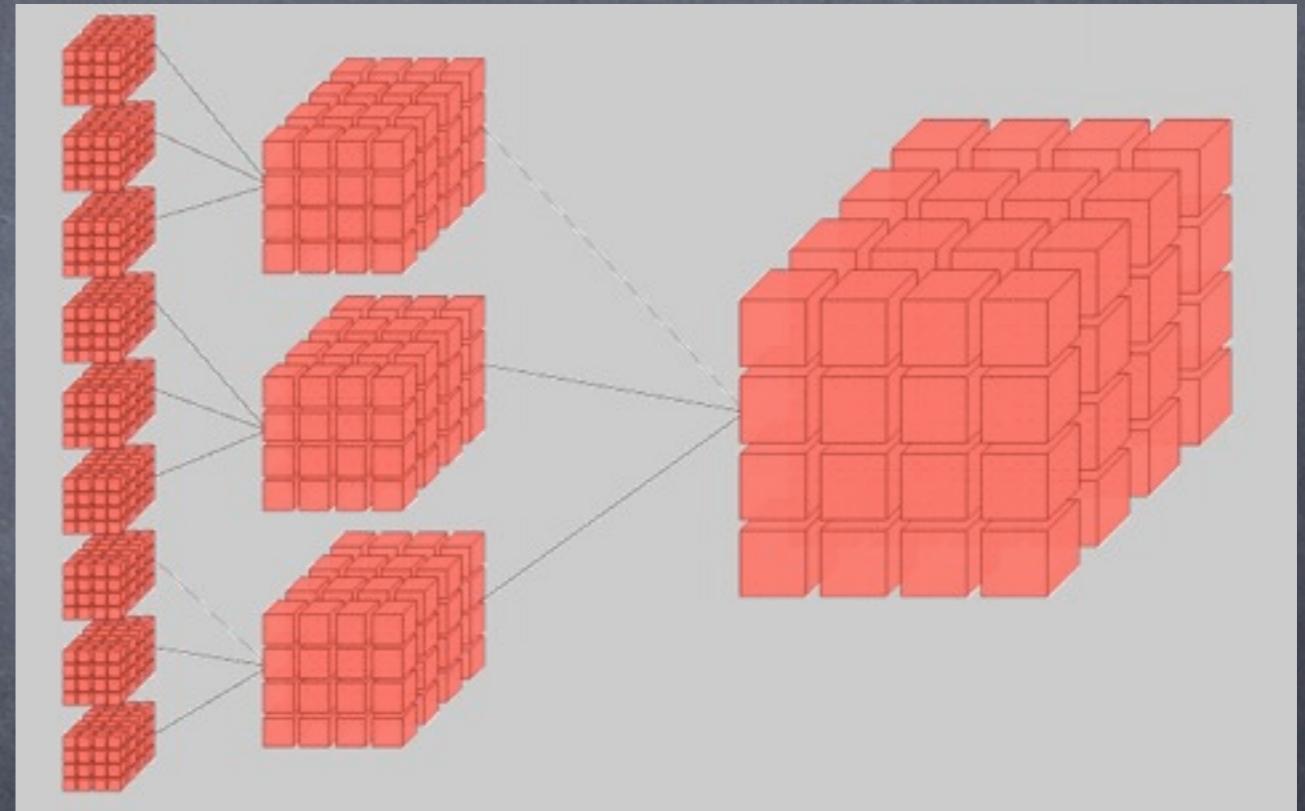
OLAP Data Cube

- core of an OLAP system.
- provides the m-Dim way to look at the summary of the data.
- m-Dim generalization of group by operator
- are sparse in nature



OLAP Aggregation

- pre-calculated summaries of data.
- answers are ready before the questions!
- improve query response time.
- aggregations are stored in the data cube



OLAP

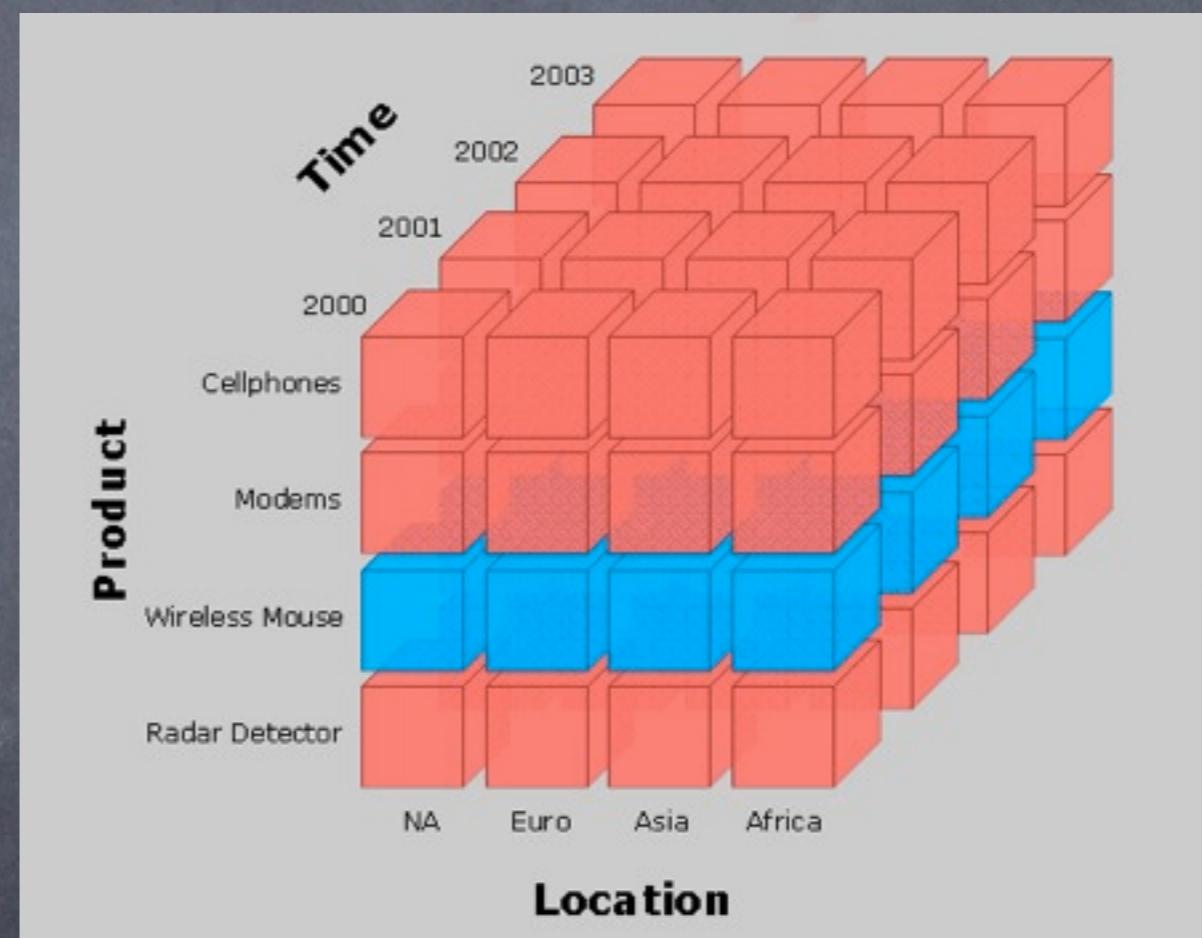
Navigational Operations

- ⦿ Roll-up (lower to higher aggregation)
- ⦿ Drill-down (higher to lower aggregation)
- ⦿ Slicing
- ⦿ Dicing
- ⦿ Pivot

Navigational Operations

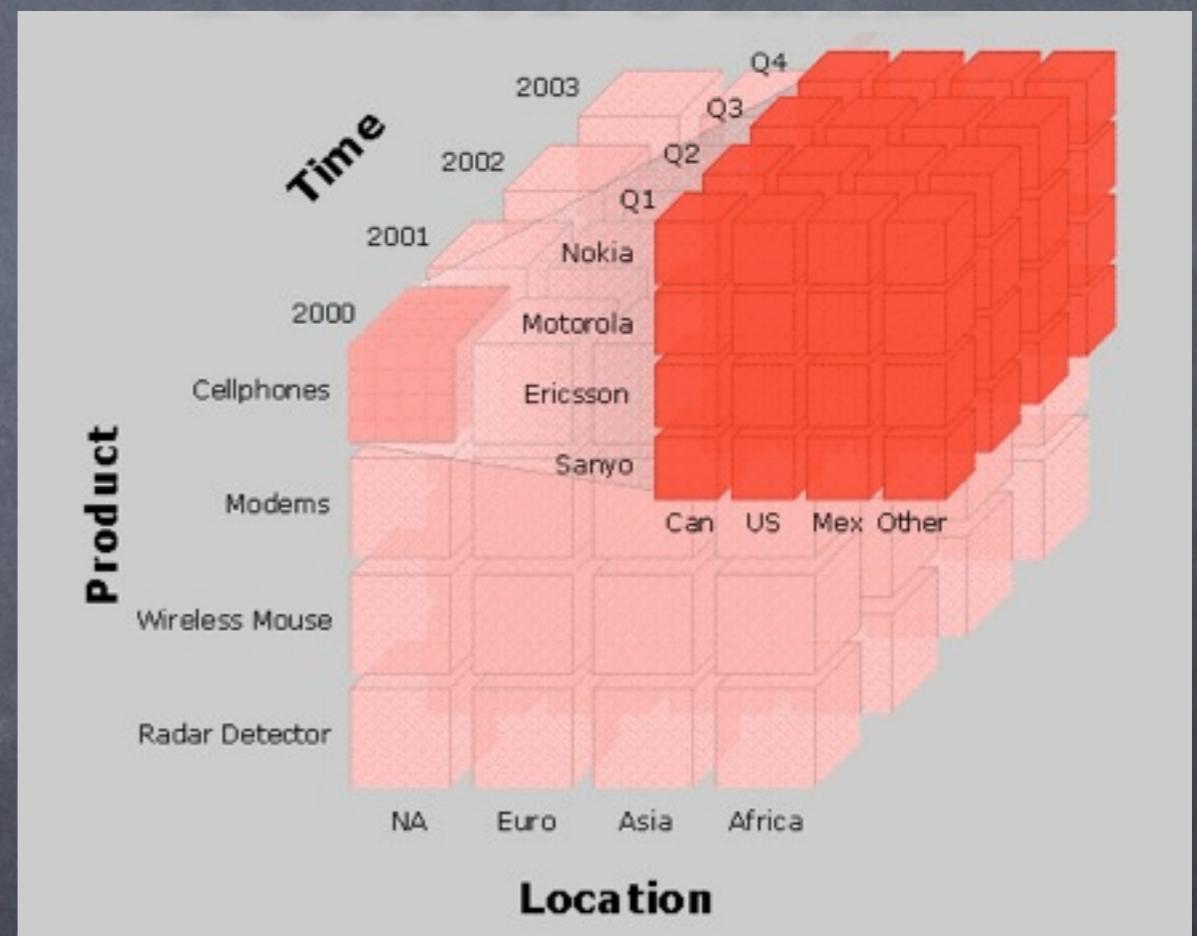
Slicing

- a subset of a mDim array corresponding to a single value for one or more members of the dimensions not in the subset.



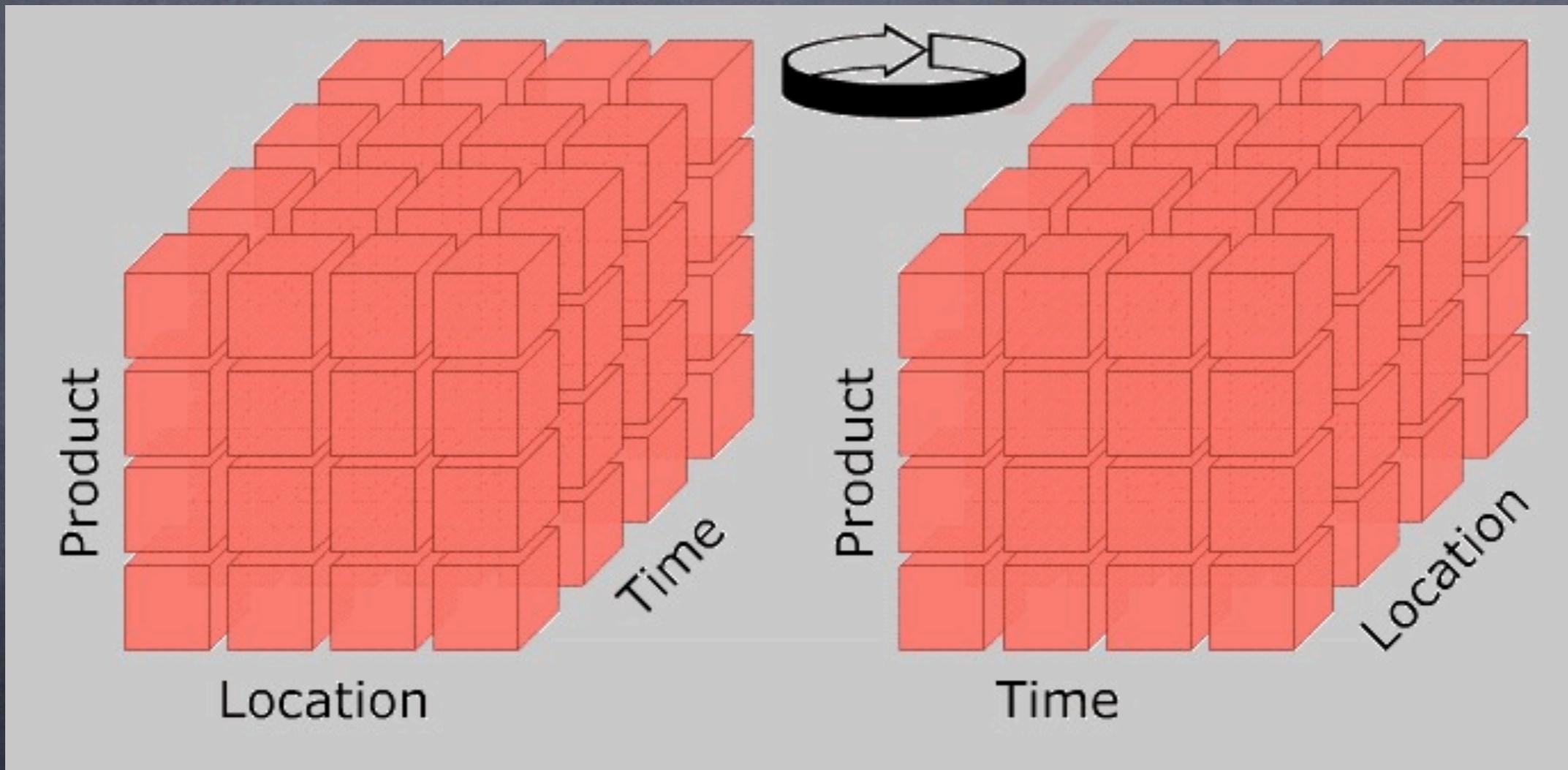
Navigational Operations Dicing

- is a slice on more than two dimensions of a cube.
- i.e. more than two consecutive slices.



Navigational Operations

Pivoting



OLAP Approaches

- ⦿ ROLAP
- ⦿ MOLAP
- ⦿ HOLAP

OLAP Approaches

ROLAP

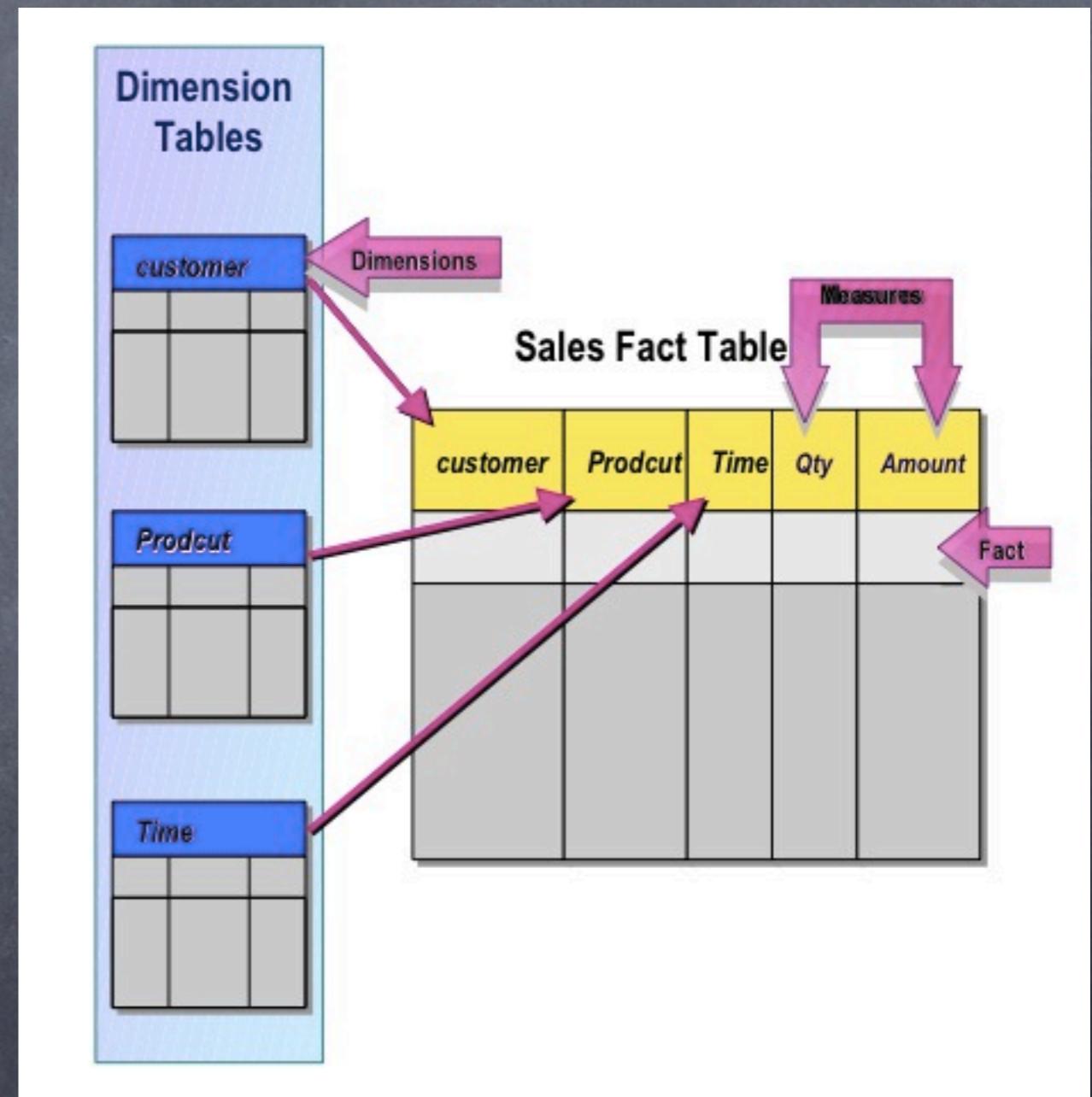
- ⦿ Overview
- ⦿ Architecture
- ⦿ Methods of Cubing
- ⦿ Performance Evaluation

ROLAP Overview

- uses a RDBMS as a source.
 - however, a DB designed for OLTP will not function well with ROLAP.
- does not require pre-aggregation
 - generate SQL queries at appropriate level at request time.

ROLAP Architecture

- scalable!
- needs additional attributes to define position in m-Dim space.



ROLAP

Methods of Cubing

- ⦿ Sort-based Methods (pipeSort)
- ⦿ Hash-based Methods (pipeHash)

ROLAP

Performance Evaluation

- ⦿ ROLAP is CPU-bound.
- ⦿ slow
 - ⦿ requires more disk
 - ⦿ requires more IOs
 - ⦿ requires more IO time

OLAP Approaches

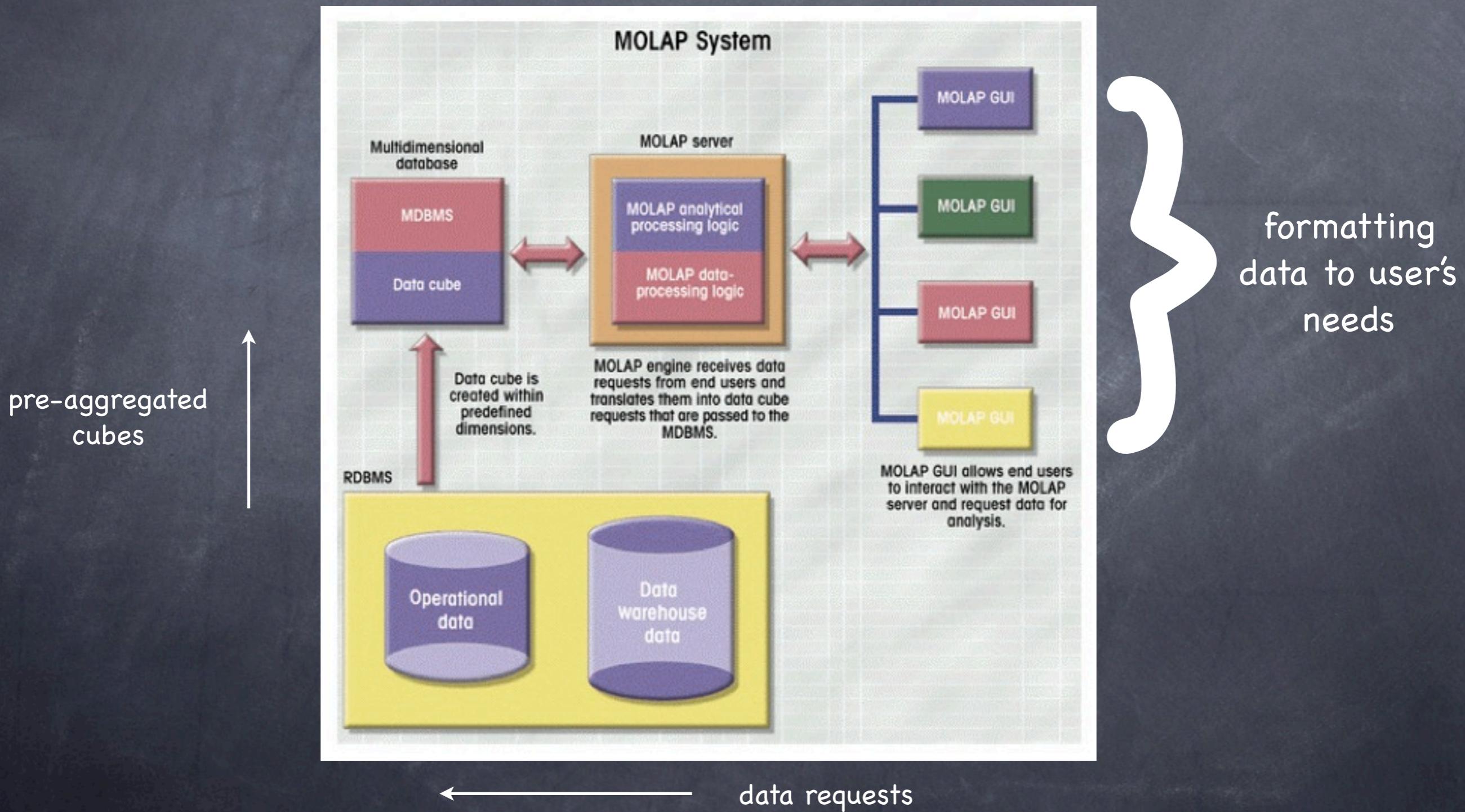
MOLAP

- ⦿ Overview
- ⦿ Architecture
- ⦿ Storage Issues

MOLAP Overview

- core is a m-Dim data cube.
- allows position based computation.
- the cube is **very** sparse (not-scalable).
- extremely fast!

MOLAP Architecture



MOLAP Storage Issues

- ⦿ Chunking
- ⦿ Chunk-offset Compression

MOLAP I Storage Issues

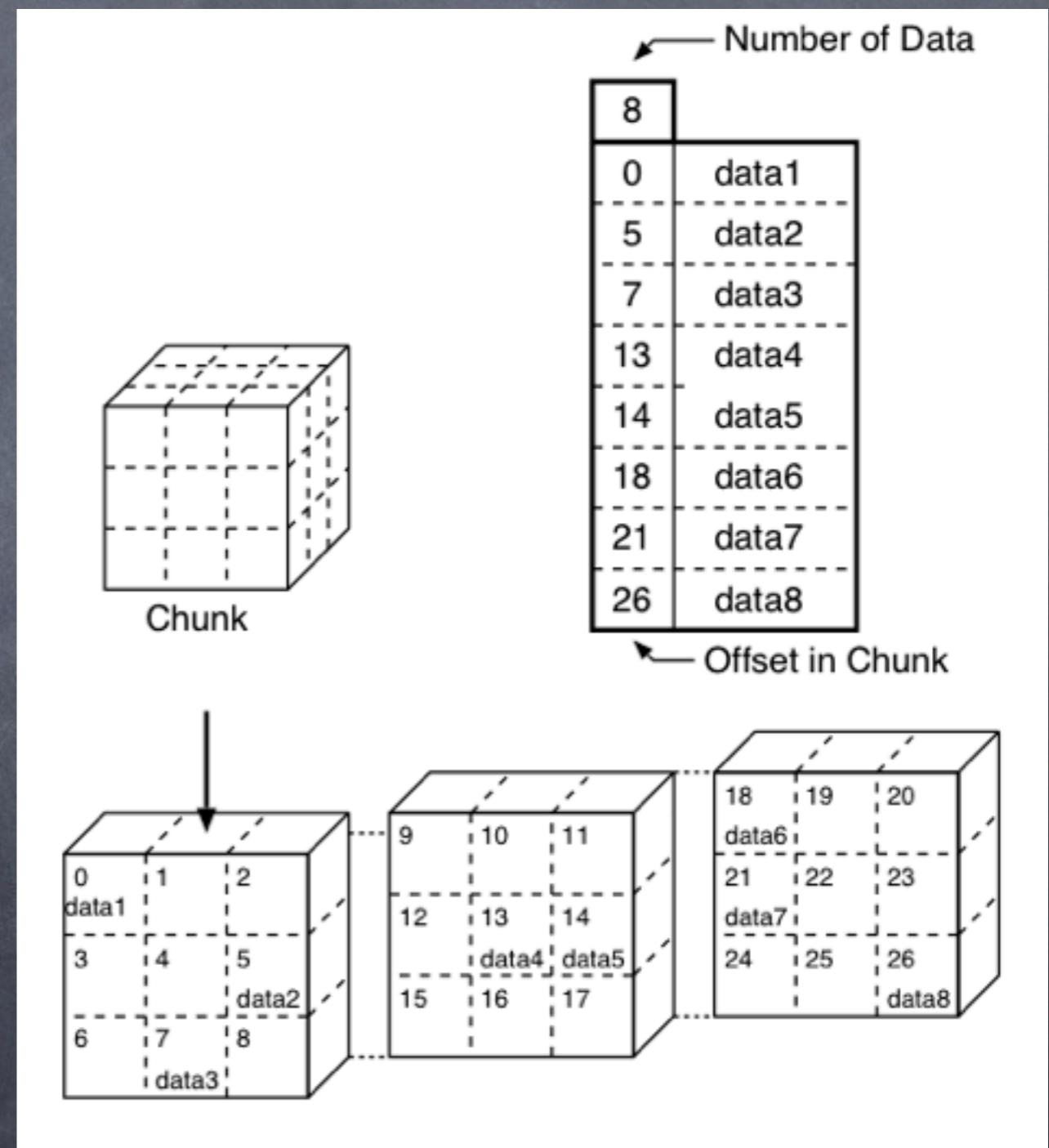
Chunking

- ➊ dividing m-Dim array into small chunks.
- ➋ allows chunks to fit into available memory for in-memory computations.

MOLAP | Storage Issues

Chunk-offset Compression

- store a pair for each valid entry.
- solves the sparse-array problem.



OLAP Approaches

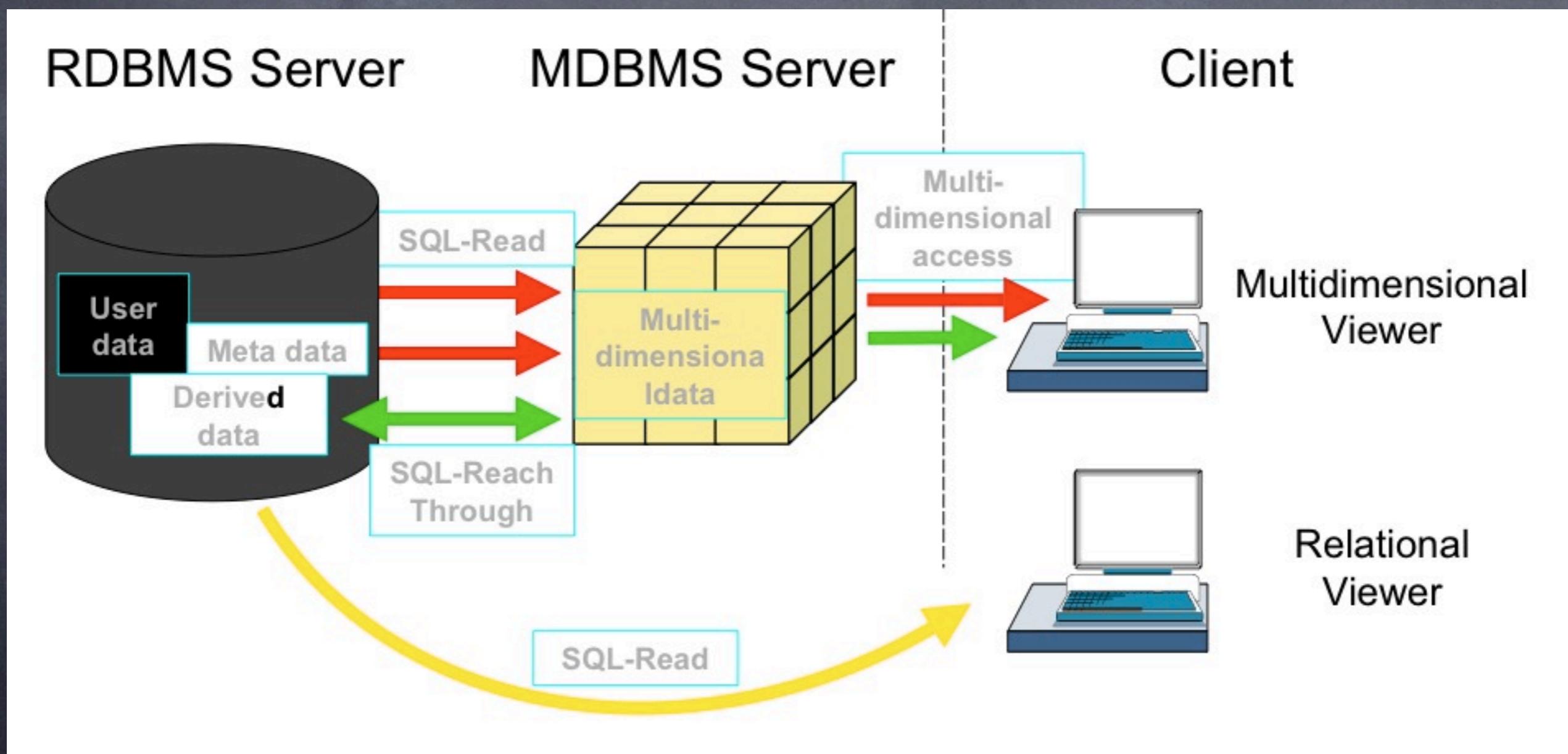
HOLAP

- ⦿ Overview
- ⦿ Architecture

HOLAP Overview

- store detailed data in RDBMS
- store aggregated data in MDBMS

HOLAP Architecture



When to choose What?

- ⦿ Performance is a concern?
 - ⦿ MOLAP!
- ⦿ Data Volume and Scalability is a concern?
 - ⦿ ROLAP!

Thank You!