OLAP

- Perceive the demand for online analytical processing (OLAP) .
- Review major functions and features of OLAP
- Understand Dimensional analysis, learn meaning of hypercubes, drill-down and roll-up, and slice and dice operation
- Examine OLAP models

OLAP

Data Warehouse provides the best opportunity for analysis and OLAP is the vehicle for carrying out involved analysis.

- Demand for On-Line Analytical Processing
- Need For Multi-Dimensional Analysis
- Fast Access and Powerful Calculations

OLAP

Need for Multidimensional Analysis:

- "How many units of Product A did we sell in the store A, in Mumbai?"
- "How much revenue did the new product X generate during the last three months, broken down by individual months, in the South Central Territory, by individual stores, broken down by promotions, compared to estimates, and compared to the previous version of the product?"
- For effective analysis, user must have easy methods of performing complex analysis across several business dimensions.

OLAP

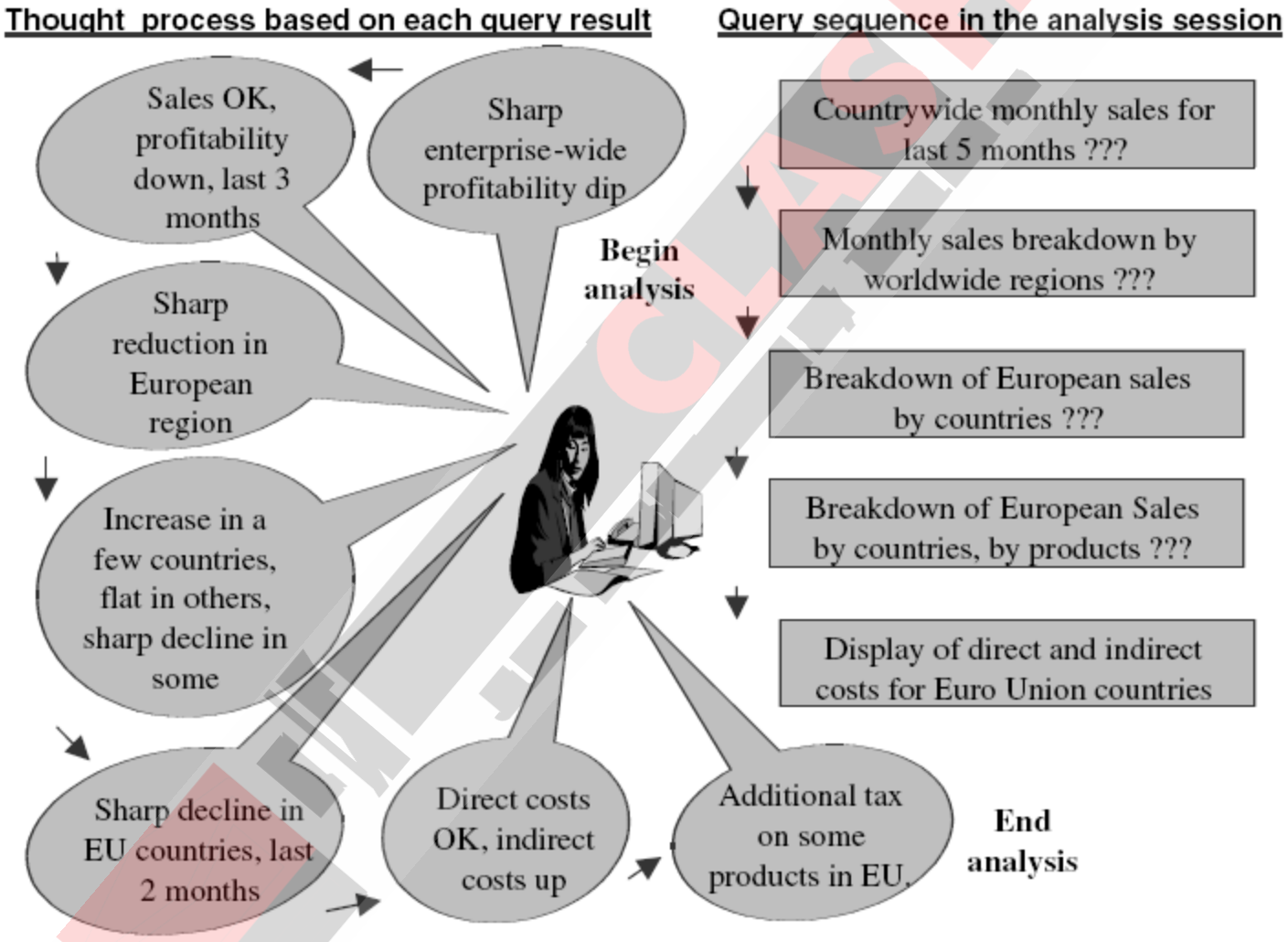


Figure 15-1 Query steps in an analysis session.

OLAP is the Answer

- Enables analysts, executives, and managers to gain useful insights from the presentation of data.
- Can reorganize metrics along several dimensions and allow data to be viewed from different perspectives.
- Supports multidimensional analysis.
- Is able to drill down or roll up within each dimension.
- Is capable of applying mathematical formulas and calculations to measures.
- Provides fast response, facilitating speed-of-thought analysis.
- Complements the use of other information delivery techniques such as data mining.
- Improves the comprehension of result sets through visual presentations using graphs and charts.
- Can be implemented on the Web.
- Designed for highly interactive analysis.

OLAP

LINE	TOTAL SALES
Clothing	$12,836,450
Electronics	$16,068,300
Video	$21,262,190
Kitchen	$17,704,400
Appliances	$19,600,800
Total	$87,472,140

1 High level summary by product line

2 Drill down by year

LINE	1998	1999	2000	TOTAL
Clothing	$3,457,000	$3,590,050	$5,789,400	$12,836,450
Electronics	$5,894,800	$4,078,900	$6,094,600	$16,068,300
Video	$7,198,700	$6,057,890	$8,005,600	$21,262,190
Kitchen	$4,875,400	$5,894,500	$6,934,500	$17,704,400
Appliances	$5,947,300	$6,104,500	$7,549,000	$19,600,800
Total	$27,373,200	$25,725,840	$34,373,100	$87,472,140

3 Rotate columns to rows

YEAR	Clothing	Electronics	Video	Kitchen	Appliances	TOTAL
1998	$3,457,000	$5,894,800	$7,198,700	$4,875,400	$5,947,300	$27,373,200
1999	$3,590,050	$4,078,900	$6,057,890	$5,894,500	$6,104,500	$25,725,840
2000	$5,789,400	$6,094,600	$8,005,600	$6,934,500	$7,549,000	$34,373,100
Total	$12,836,450	$16,068,300	$21,262,190	$17,704,400	$19,600,800	$87,472,140

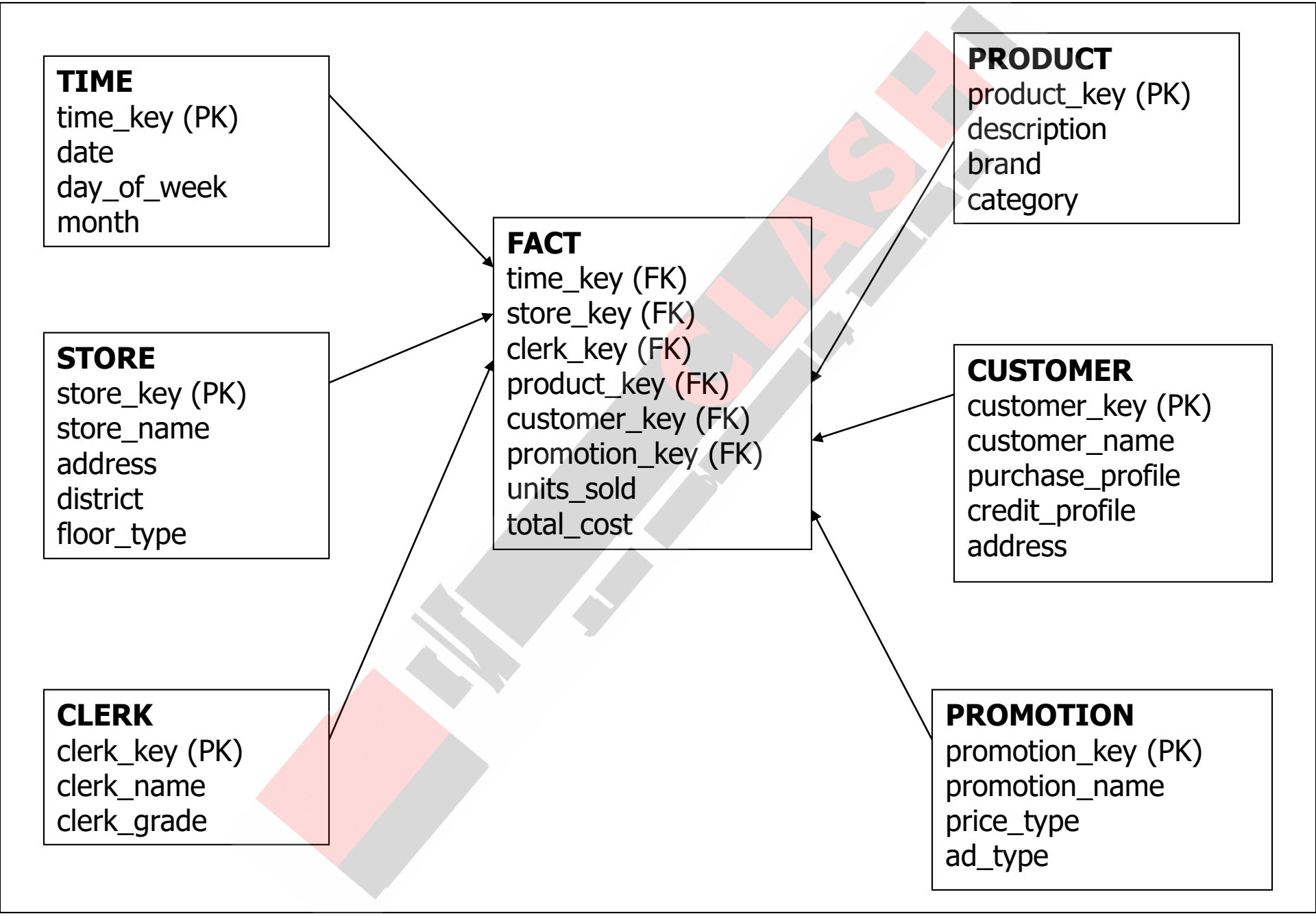
Figure 15-3 Simple OLAP session.

Dimensional Modeling

- Two important concepts
  - Fact
    - Numeric measurements, represent business activity/event
    - Are pre-computed
    - Example: Profit, quantity sold
  - Dimension
    - Qualifying characteristics, perspective to a fact
    - Example: date (Date, month, quarter, year)

Dimensional Modeling (cont.)

- Facts are stored in fact table
- Dimensions are represented by dimension tables
- Dimensions are degrees in which facts can be judged
- Each fact is surrounded by dimension tables
- Looks like a star so called Star Schema


TIME
time_key (PK)
date
day_of_week
month
STORE
store_key (PK)
store_name
address
district
floor_type
CLERK
clerk_key (PK)
clerk_name
clerk_grade
FACT
time_key (FK)
store_key (FK)
clerk_key (FK)
product_key (FK)
customer_key (FK)
promotion_key (FK)
units_sold
total_cost
PRODUCT
product_key (PK)
description
brand
category
CUSTOMER
customer_key (PK)
customer_name
purchase_profile
credit_profile
address
PROMOTION
promotion_key (PK)
promotion_name
price_type
ad_type

- Inside Dimension table
  - Key attribute of dimension table, for identification
  - Large no of columns, wide table
  - Non-calculated attributes
  - Un-normalized in Star schema
  - Ability to drill-down and roll-up are two ways of exploiting dimensions
  - Can have multiple hierarchies
  - Relatively small number of records

Inside Fact table

- Have two types of attributes
  - Key attributes, for connections
  - Facts
- Inside fact table
  - Concatenated key
  - Large number of records
  - Limited attributes

Warehouse Database Schema

- Design should reflect multidimensional view
  - Star Schema
  - Snowflake Schema
  - Fact Constellation Schema

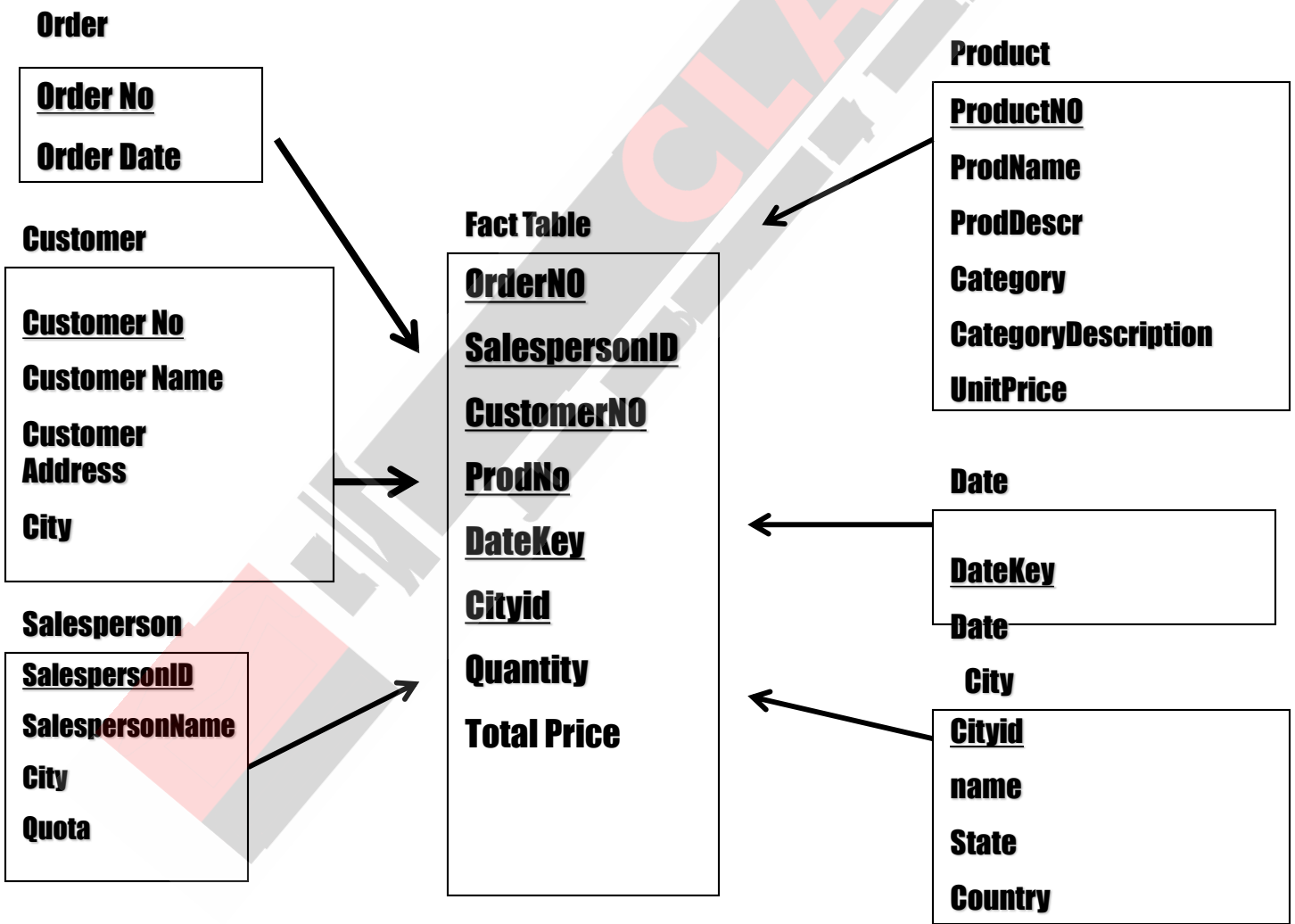
Conceptual Modeling of Data Warehouses

- Modeling data warehouses: dimensions & measures
  - Star schema: A fact table in the middle connected to a set of dimension tables
  - Snowflake schema: A refinement of star schema where some dimensional hierarchy is *normalized* into a set of smaller dimension tables, forming a shape similar to snowflake
  - Fact constellations: Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called *galaxy schema* or fact constellation

Star Schema

- Creating the STAR schema is the fundamental data design technique for the data warehouse.
- Star schema provides a multidimensional view.
- *Star schema :* A dimensional model with fact table in the middle & dimension tables arranged around fact table.
- The key component of the STAR schema is the set of dimension tables.
- The STAR schema structure is a structure that can be easily understood by the users and with which they can comfortably work.
- When a query is made against the data warehouse, the results of the query are produced by combining or joining one or more dimension tables with the fact table.

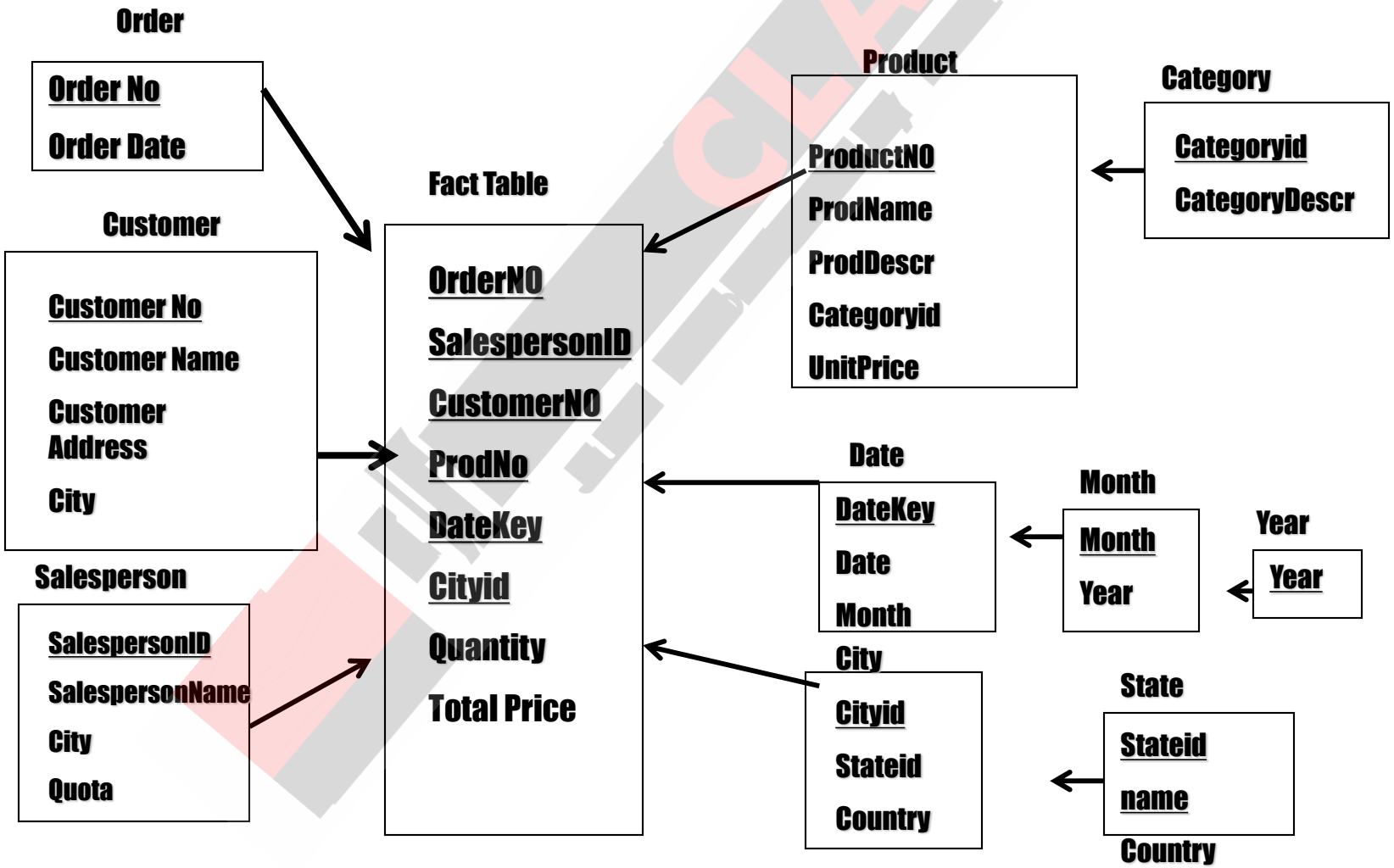
Example of a Star Schema



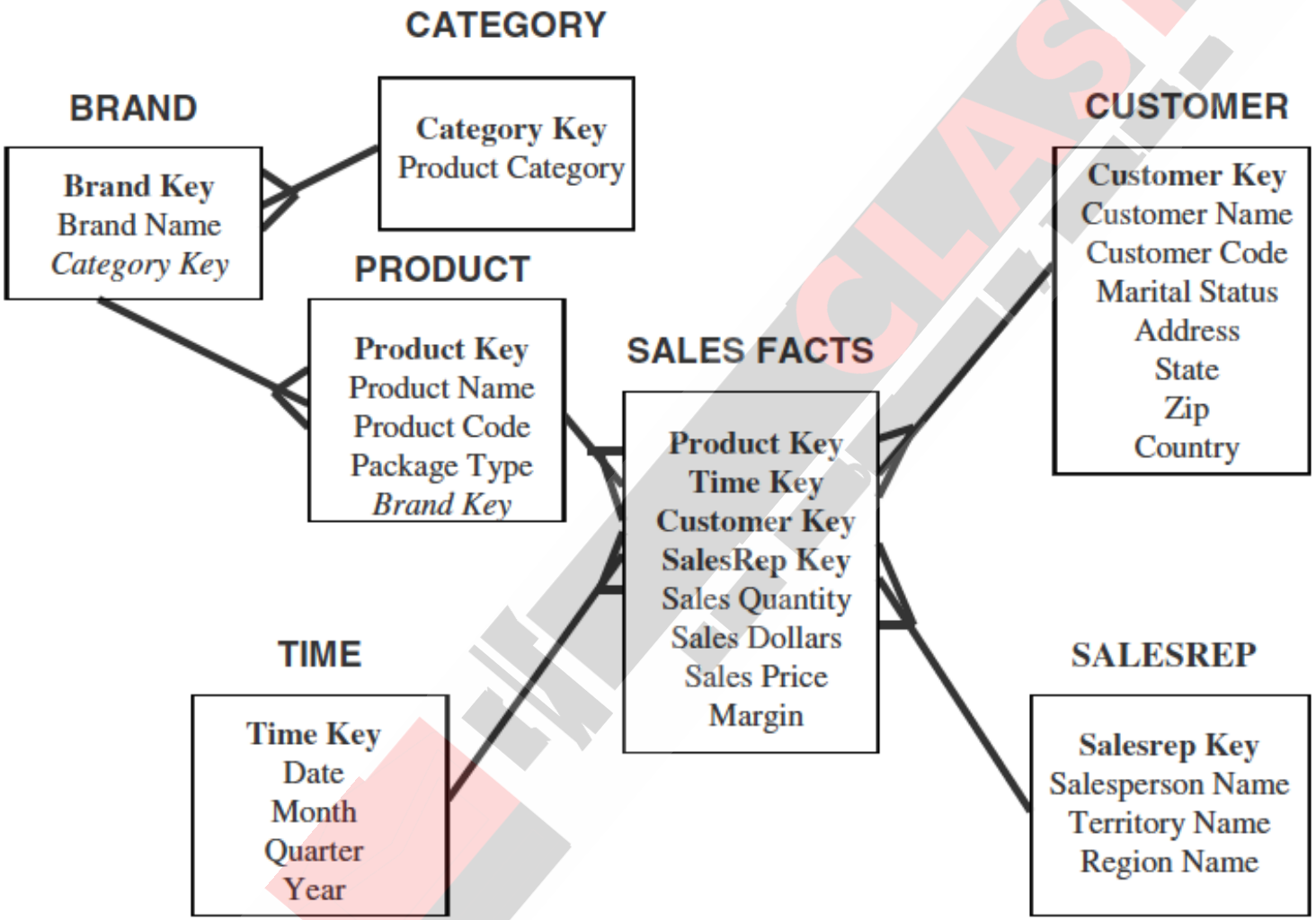
The Snowflake Schema

- Snowflaking involves normalization of dimensions in Star Schema
- Reasons:
- To save storage space
- To optimize some specific queries

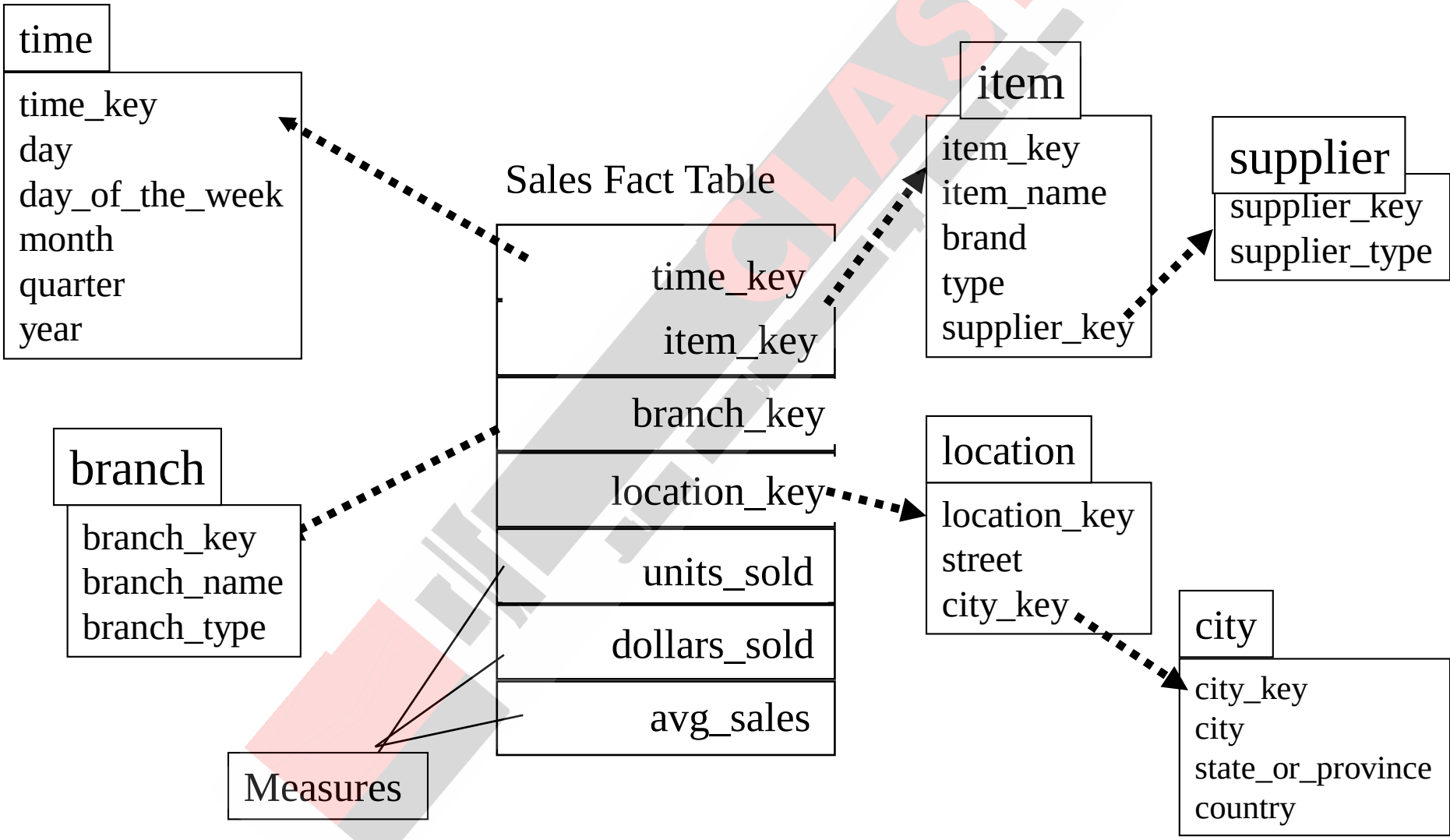
Example of a Snowflake Schema



Snowflake Schema Example



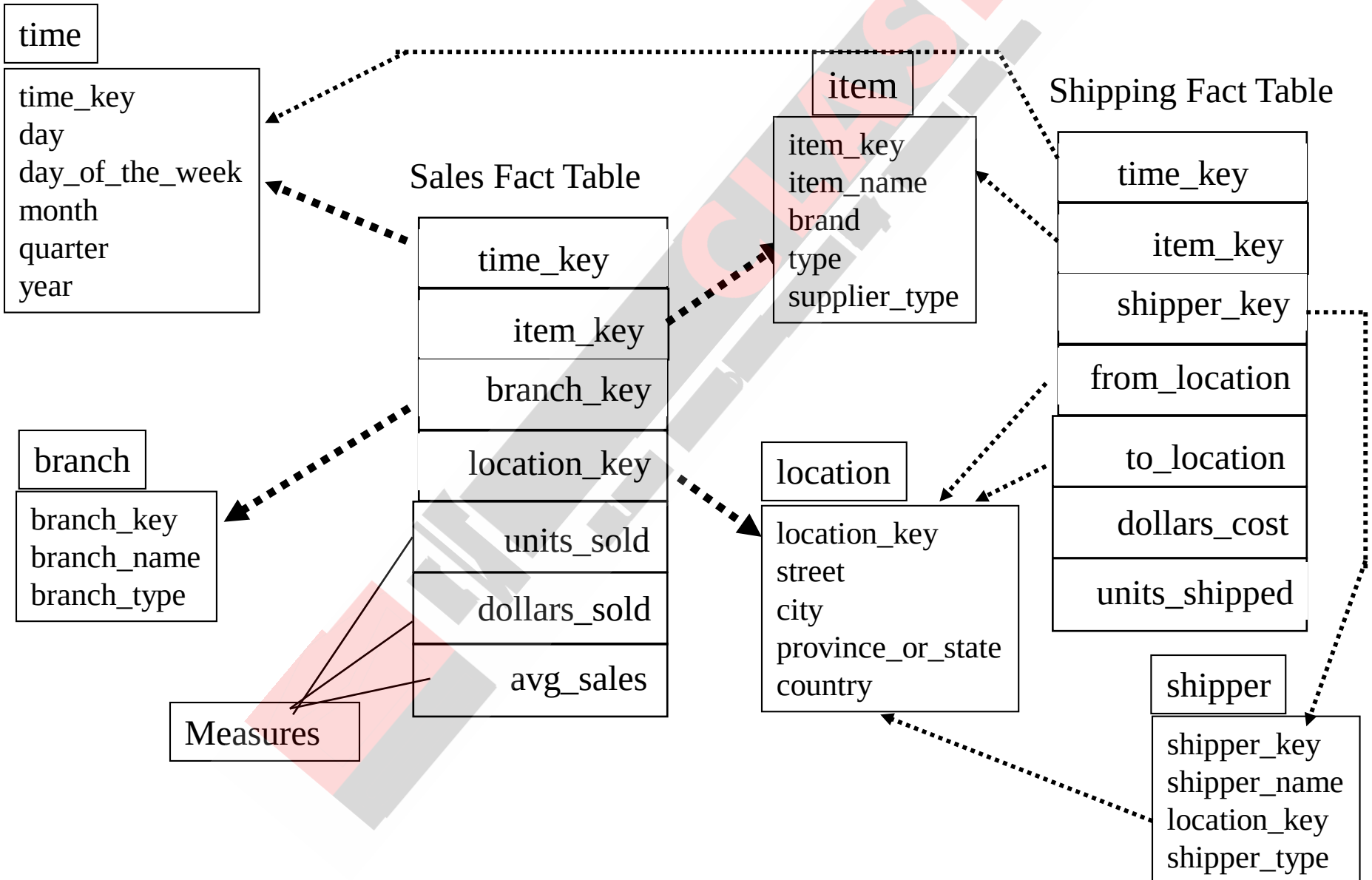
Example of Snowflake Schema



Fact constellation or Galaxy Schema:

Multiple Fact tables that share many dimension tables

Example of Fact Constellation



What is OLAP

- Basic idea: converting data into information that decision makers need
- A Concept to analyze data by multiple dimension in a structure called *data cube*

OLAP

- On-Line Analytical Processing (OLAP)

is a category of software technology
that enables analysts and executives
to gain insight into data
through fast, consistent, interactive access
in a wide variety of possible views of information
that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user.

OLAP Characteristics

- Let business users have a multidimensional and logical view of the data in the data warehouse,
- Facilitate interactive query and complex analysis for the users,
- Allow users to drill down for greater details or roll up for aggregations of metrics
- Along a single business dimension or across multiple dimensions, provide ability to perform intricate calculations and comparisons, and
- Present results in a number of meaningful ways, including charts and graphs.

OLAP

BASIC FEATURES	Multidimensional analysis	Consistent performance	Fast response times for interactive queries
	Drill-down and roll-up	Navigation in and out of details	Slice-and-dice or rotation
	Multiple view modes	Easy scalability	Time intelligence (year-to-date, fiscal period)
ADVANCED FEATURES	Powerful calculations	Cross-dimensional calculations	Pre-calculation or pre-consolidation
	Drill-through across dimensions or details	Sophisticated presentation & displays	Collaborative decision making
	Derived data values through formulas	Application of alert technology	Report generation with agent technology

Figure 15-4 General features of OLAP.

Hypercubes

- A representation that accommodates more than 3 dimensions.
- i.e. represents multidimensional data.
- An OLAP cube is a data structure suitable for fast analysis of multi-dimensional problems, converting raw data into compelling data visualizations.

OLAP

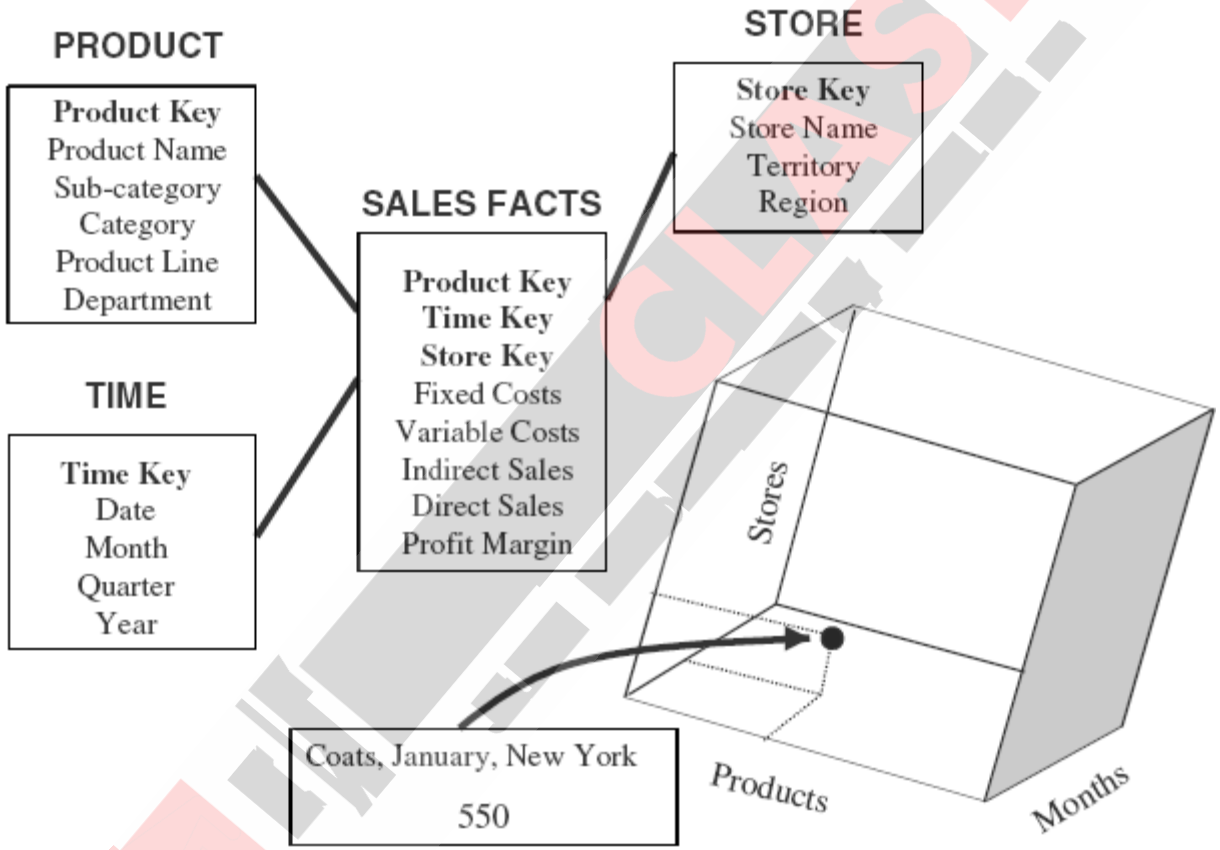


Figure 15-5 Simple STAR schema.

Example 2 - Hypercube

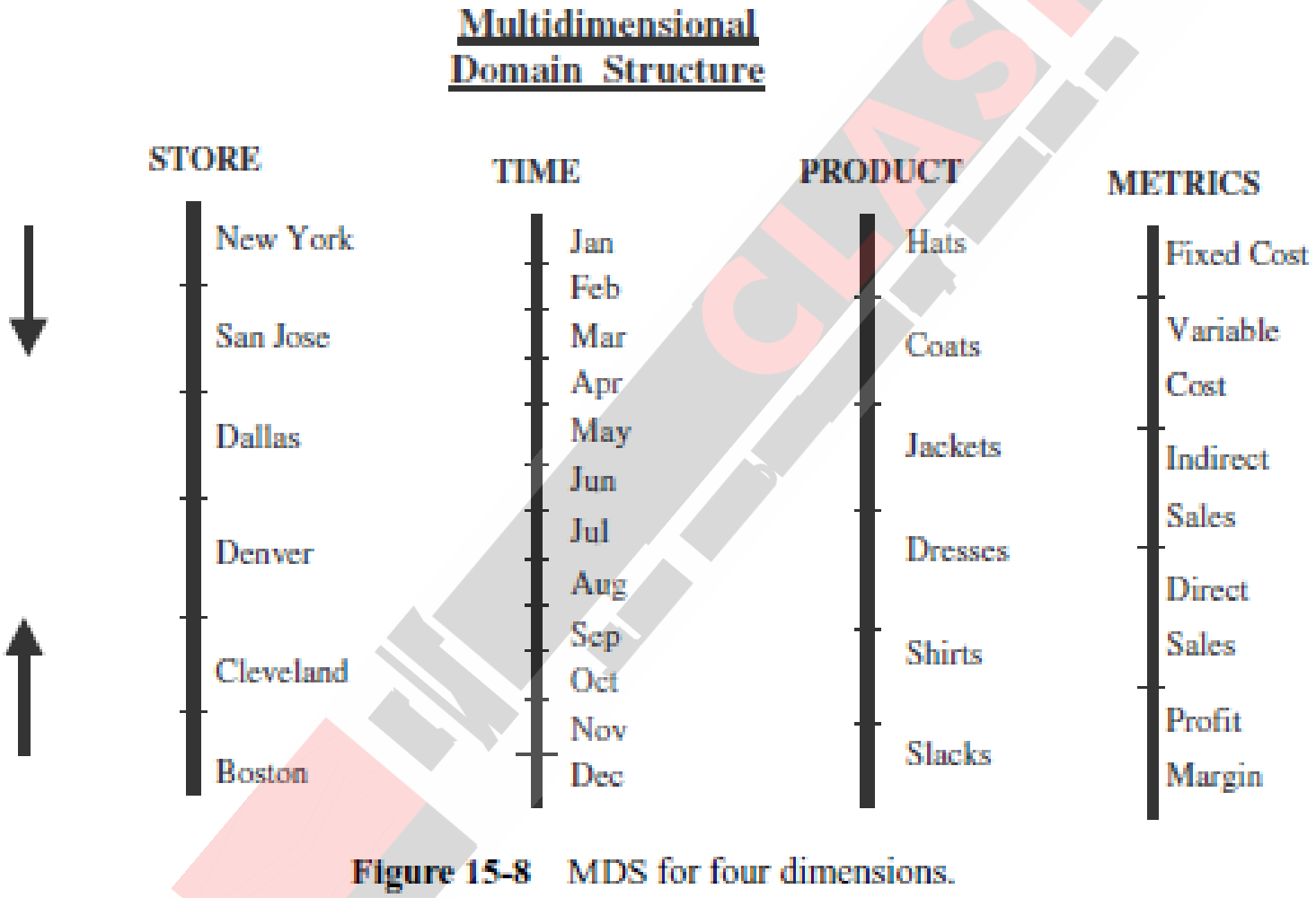
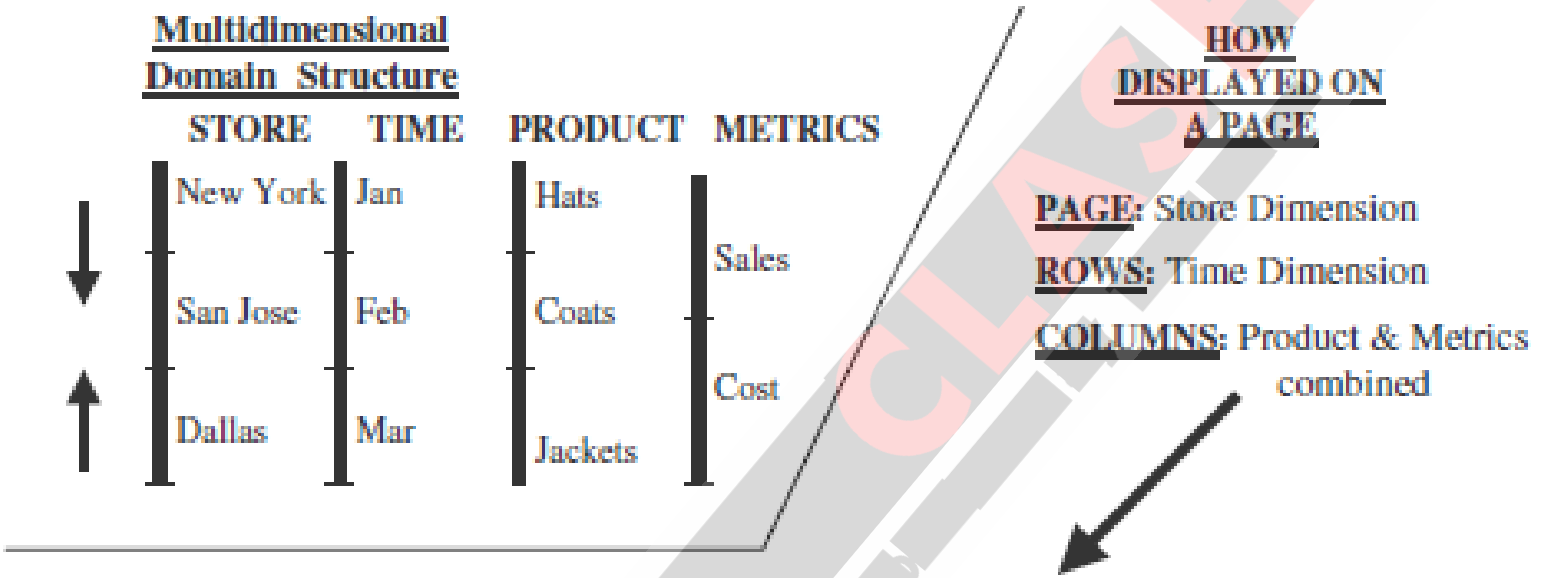


Figure 15-8 MDS for four dimensions.

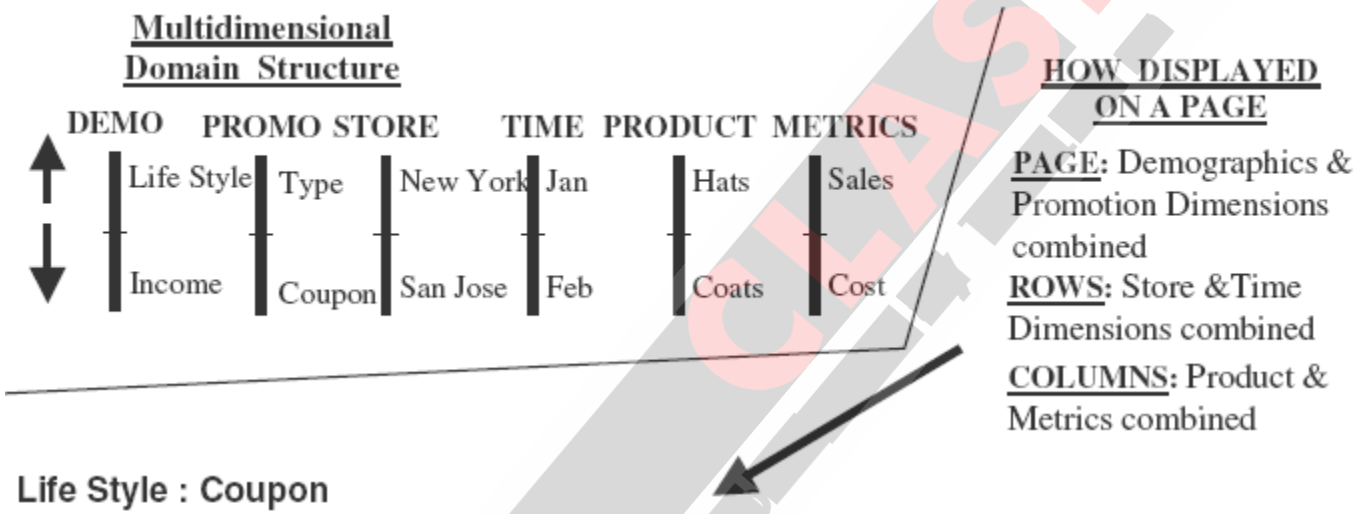


New York Store

	Hats:Sales	Hats:Cost	Coats:Sales	Coats:Cost	Jackets:Sales	Jackets:Cost
Jan	450	350	550	450	500	400
Feb	380	280	460	360	400	320
Mar	400	310	480	410	450	400

Figure 15-9 Page displays for four-dimensional data.

OLAP



Life Style : Coupon

		Hats	Hats	Coats	Coats
		Sales	Cost	Sales	Cost
New York	Jan	220	170	270	220
	Feb	190	140	230	180
Boston	Jan	200	160	240	200
	Feb	180	130	220	170

Figure 15-11 Page displays for six-dimensional data.

OLAP

Drill-Down and Roll-Up

Store: New York

Sub-categories

PAGES: STORE dimension

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

	Outer	Dress	Casual
Jan	1,100	1,020	490
Feb	1,080	1,040	500
Mar	1,050	980	470
Apr	970	1,000	480
May	1,010	1,080	520
Jun	910	1,100	330
Jul	880	1,120	250
Aug	960	1,320	230
Sep	870	1,280	210
Oct	910	1,240	250
Nov	980	1,380	260
Dec	1,080	1,520	310

Figure 15-13 Three-dimensional display with roll-up.

Major Features & Functions of OLAP (Typical OLAP Operations)

- Drill-down (roll down)- *from higher level summary to lower level summary or detailed data, or introducing new dimensions*
- Roll-Up- *by climbing up hierarchy or by dimension reduction*
- Slice and dice- *project and select*
- Pivot (rotate)- *reorient the cube, visualization.*
- Drill-across- *involving (across) more than one fact table*
- Drill-through- *through the bottom level of the cube to its back-end relational tables.*

Roll-Down(Drill-Down) & Roll-up (Drill up)

Example: Roll-up and drill-down features with reference to the product dimension hierarchies.

- It shows the rolling up to higher hierarchical levels of aggregation and the drilling down to lower levels of detail.
- *Roll-Up and Drill-Down*
  - Drilling Down to Child Members.
- *Drill-Across* -(Eg: Let me see other data) - Executes queries involving (i.e., across) more than one fact table.
- *Drill-Through-* Uses relational SQL facilities to drill through the bottom level of a data cube down to its back-end relational tables.

- Jumping from OLAP Back to the Source Data

- In OLAP you can drill up and drill down to view different levels in the data.

-Some OLAP systems also allow you to jump back to the source data.

-You can select data you find interesting in the cube and drill through to the source data to view extra detail.

OLAP

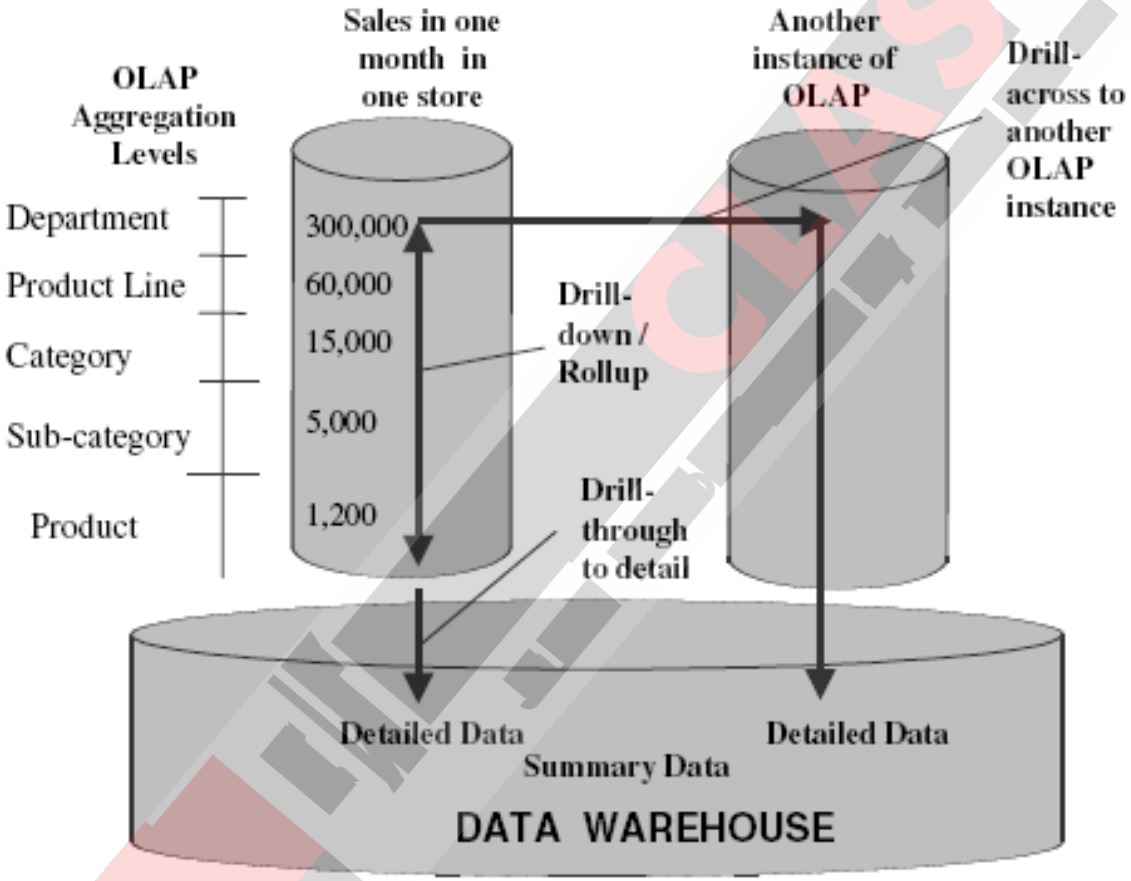


Figure 15-12 Roll-up and drill-down features of OLAP.

slicing, dicing

- Three important concepts associated with analyzing data using OLAP cubes and an OLAP reporting tool are slicing, dicing, and rotating.

- In fig 1, time, product, and location represent the dimensions of the cube, while 174 represents the measure. A dimension is a category of data and a measure is a fact or value.

.

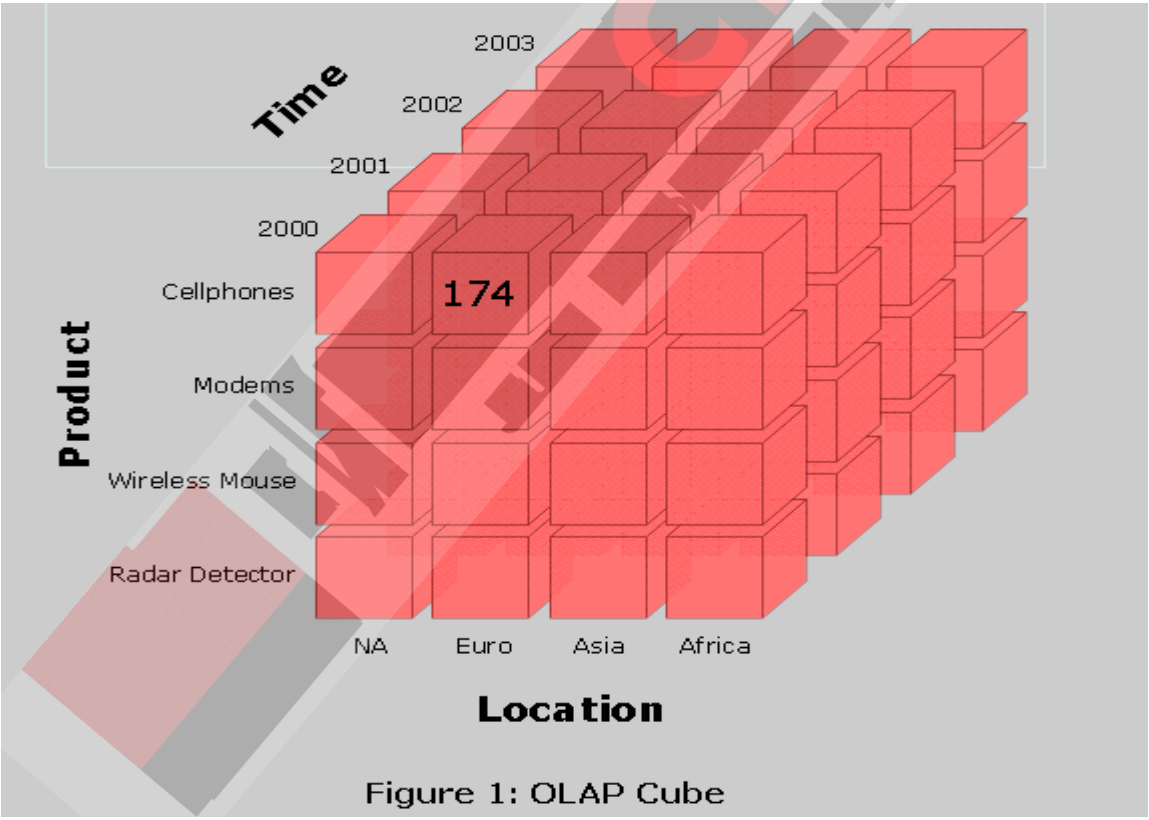


Figure 1: OLAP Cube

Slice

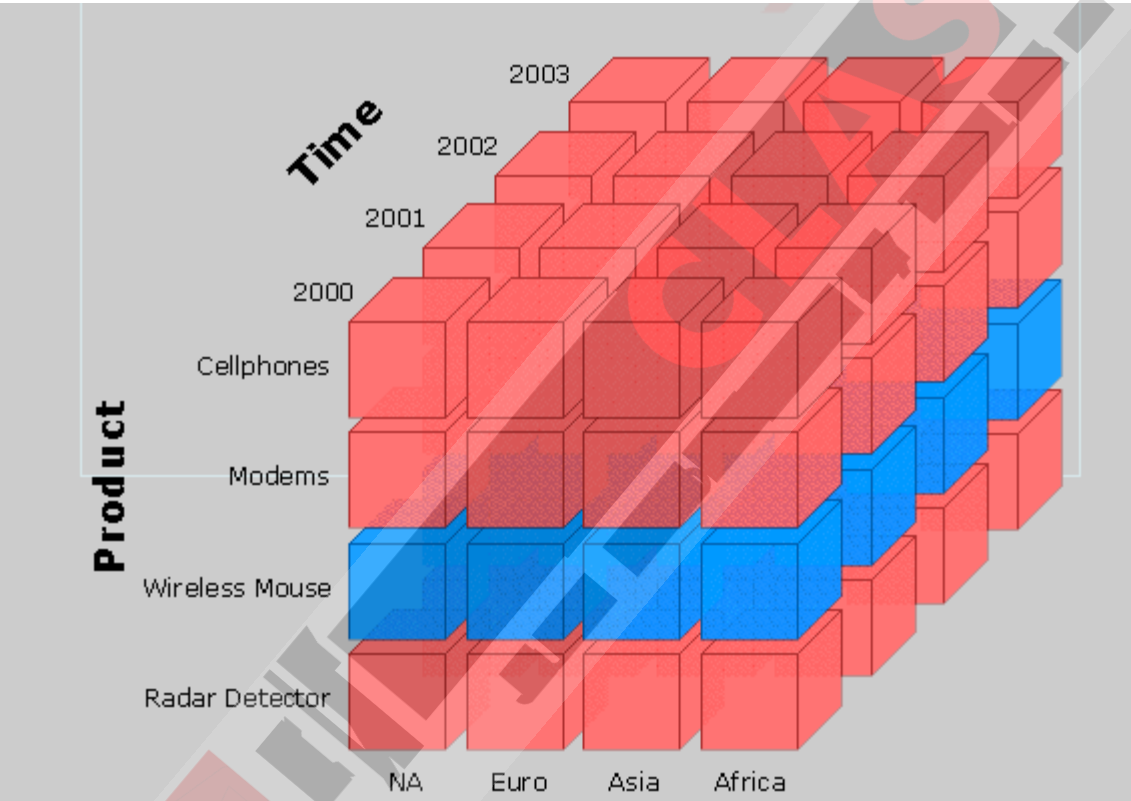


Figure 2: Slicing-Wireless Mouse

Slice

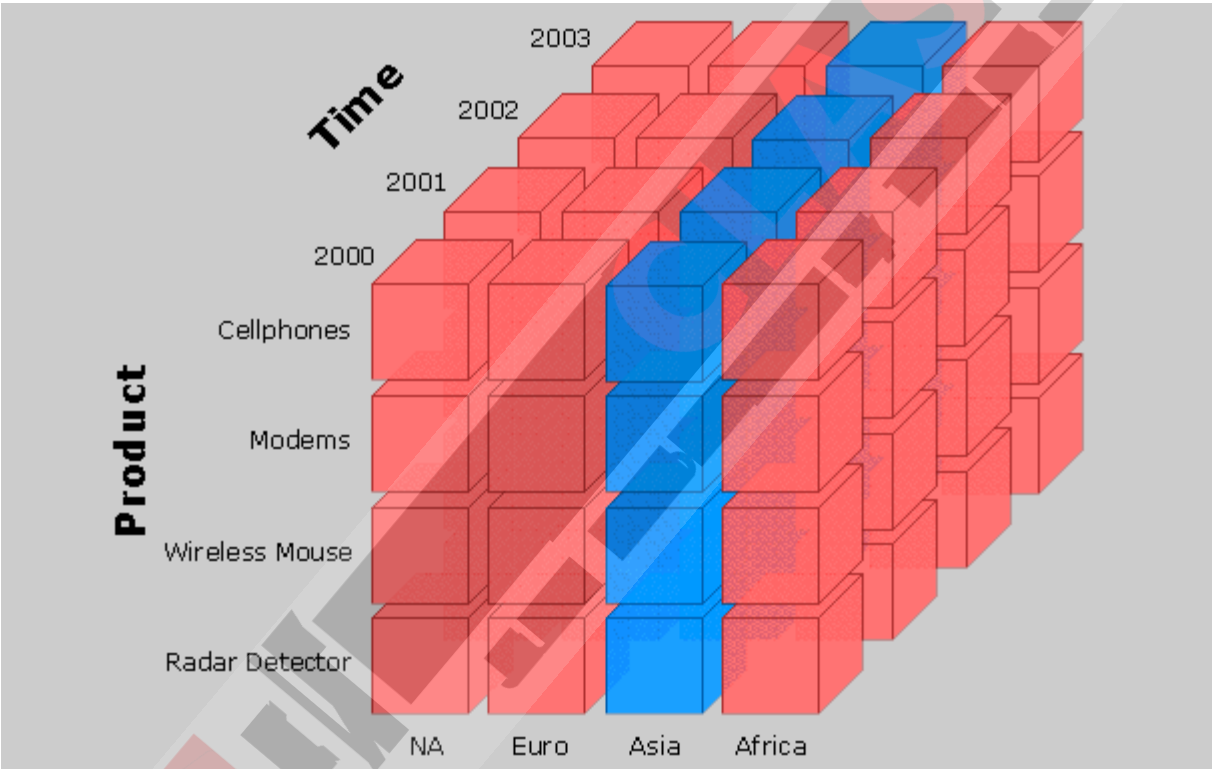


Figure 3: Slicing-Asia

Dicing

- In dicing, you define a subcube of the original space.
- The data you see is that of one cell from the cube. Dicing provides you the smallest available slice.

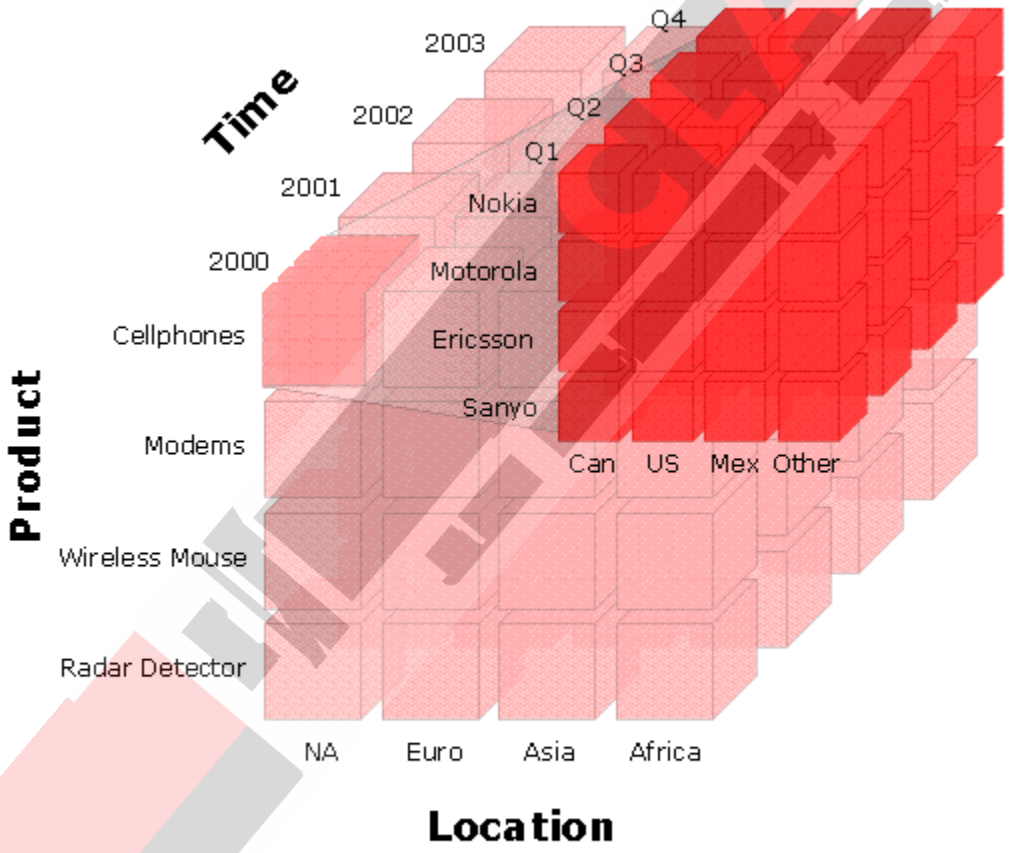


Figure 4: Dicing

Rotate (Pivote)

- Visualization operation that rotate the data axes in view in order to provide an alternative presentation of the data
- Rotating changes the dimensional orientation of the report from the cube data.
- Rotating and pivoting are the same thing.

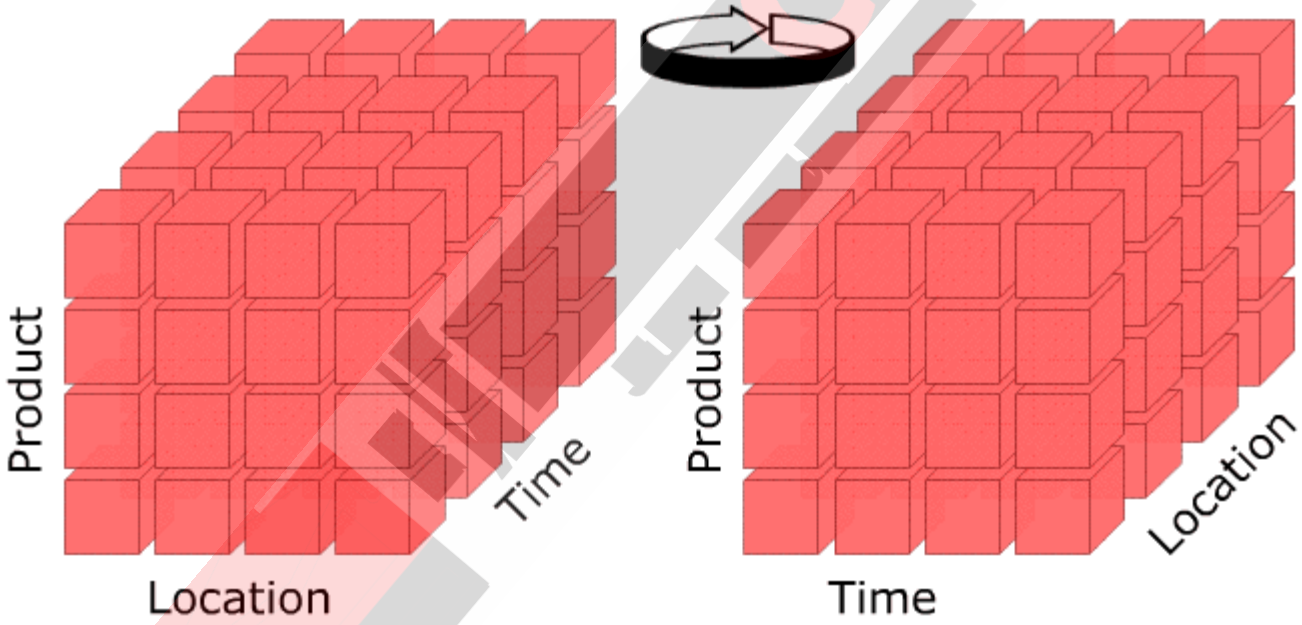


Figure 5: Rotating

Slice-and-Dice , Rotation

Advantages

- The users can view the data from many angles, understand the numbers better, and arrive at meaningful conclusions.

Uses and Benefits of OLAP

- Increased productivity of business managers, executives, and analysts
- Inherent flexibility of OLAP systems means that users may be self-sufficient in running their own analysis without IT assistance.
- Benefit for IT developers because using software specifically designed for the system development results in faster delivery of applications.
- Self-sufficiency of users, resulting in reduction in backlog.
- Faster delivery of applications following from the previous benefits.
- More efficient operations through reducing time on query executions and in network traffic.
- Ability to model real-world challenges with business metrics and dimensions.

OLAP Architecture

OLAP MODELS

Cubes in a data warehouse are stored in different modes.

- ROLAP (Relational online analytical processing)
- MOLAP (Multidimensional OLAP)

Relational OLAP (ROLAP)

- Data is stored as rows and columns in relational form.
- This model presents data to the users in the form of business dimensions.
- A semantic layer of metadata is created to hide the storage structure to the user and present data multidimensionality.
- The metadata layer supports mapping of dimensions to the relational tables.
- Enables multiple multidimensional views of 2D relational tables to be created, avoiding structuring data around the desired view.

Characteristics:

- Supports all the basic OLAP features and functions.
- Stores data in a relational form.
- Supports some form of aggregation.

Relational OLAP

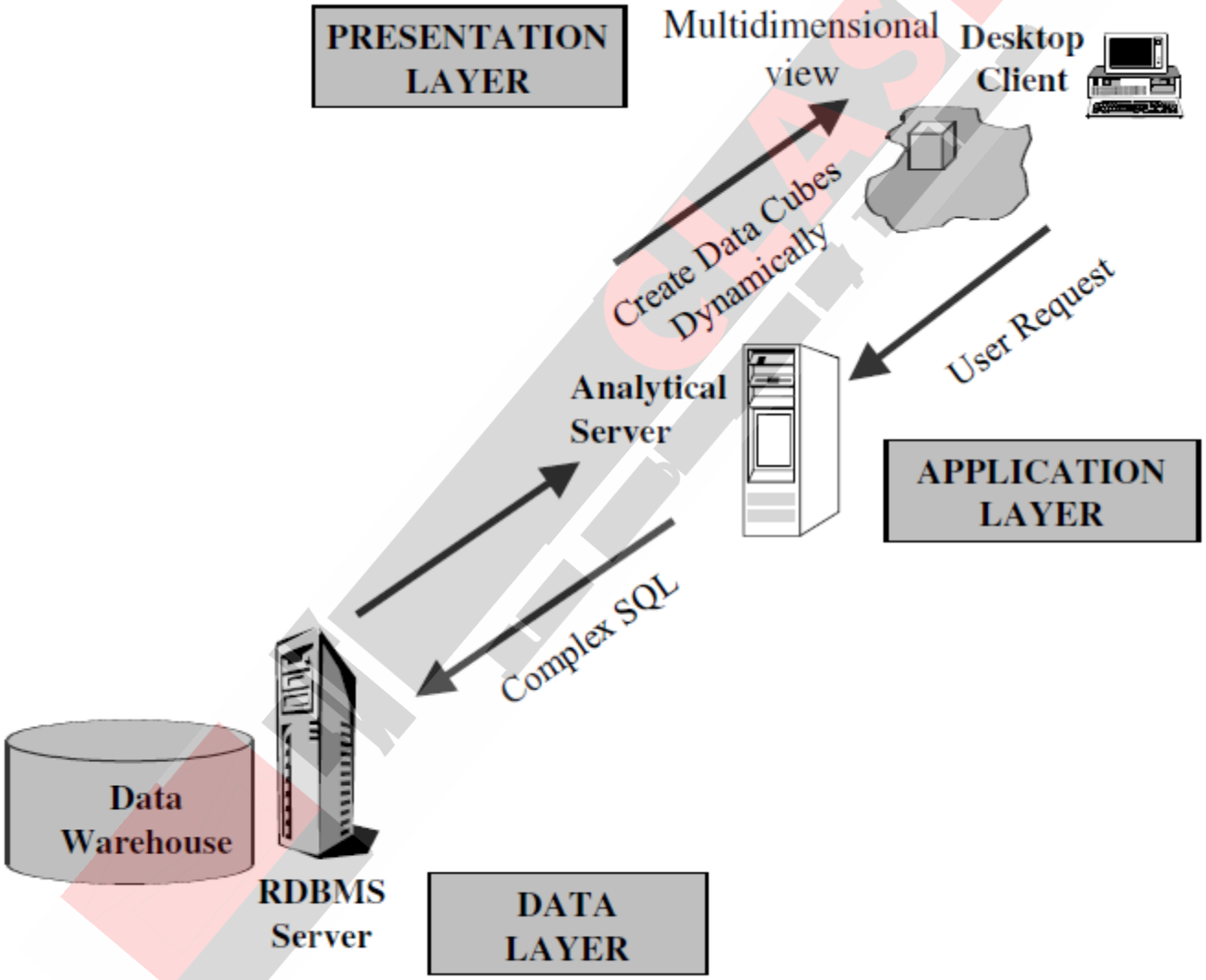


Figure 15-17 The ROLAP model.

MOLAP

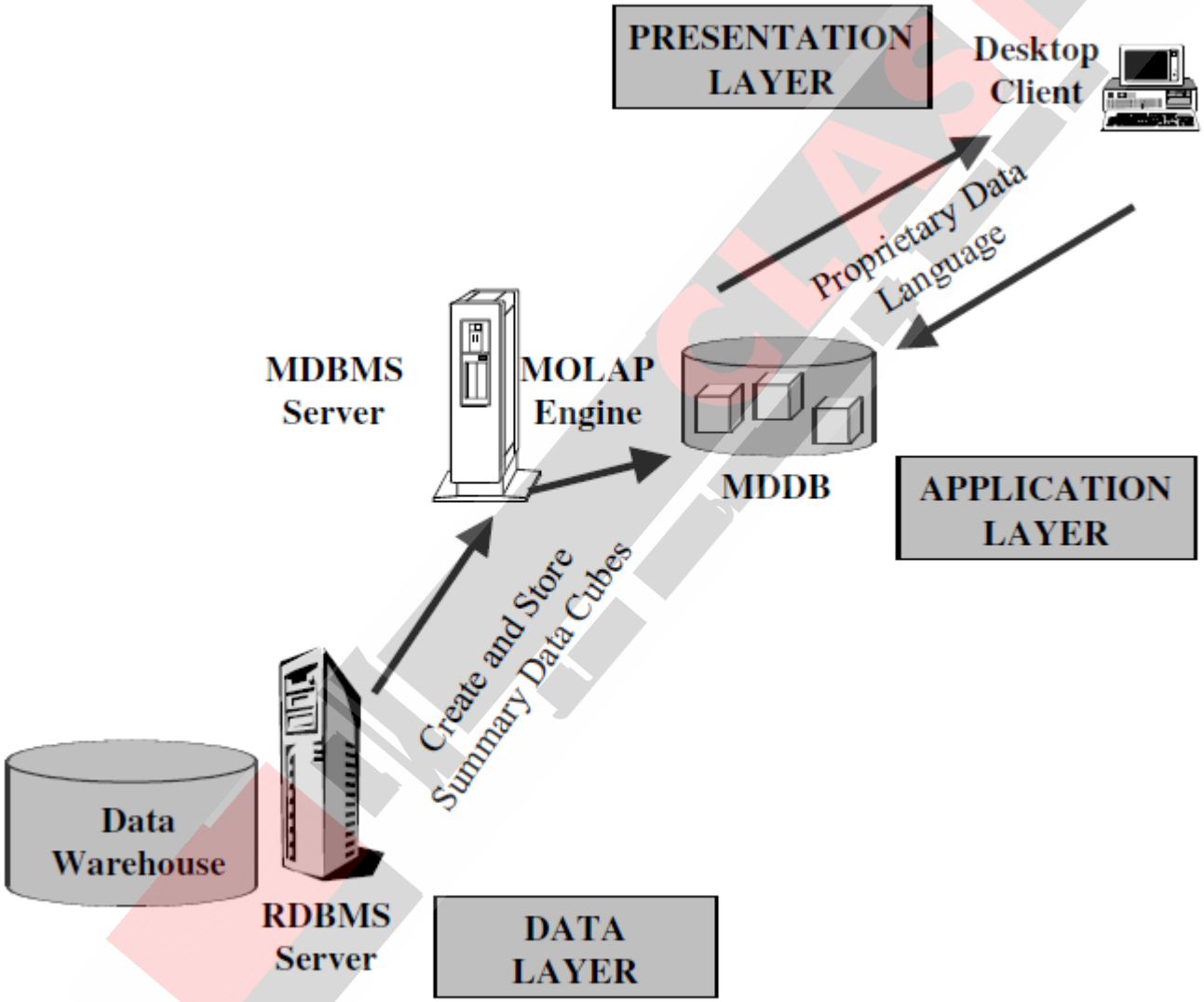


Figure 15-16 The MOLAP model.

MOLAP

- This is the traditional mode in OLAP analysis.
- Data for analysis is stored in specialized multidimensional databases in form of multidimensional cubes.
- 3 layers in the multitier architecture- presentation, application and data layer
- Pre-calculated multidimensional data cubes are stored in multidimensional databases.
- The MOLAP engine in the application layer pushes a multidimensional view of the data from the MDDBs (multidimensional DBMS) to the users.

Advantages : It provides excellent query performance and the cubes are built for fast data retrieval.

- All calculations are pre-generated when the cube is created and can be easily applied while querying data.

ROLAP versus MOLAP

The choice between ROLAP and MOLAP depends,

- on how important query performance is for users.
- on the complexity of the queries.

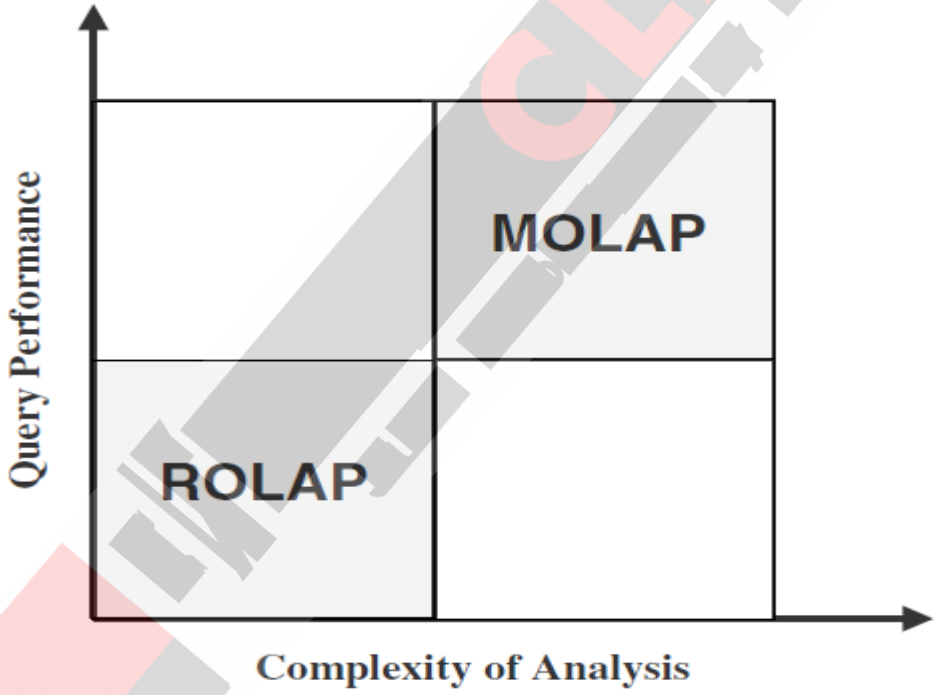


Figure 15-18 ROLAP or MOLAP?

- MOLAP is the choice for faster response and more intensive queries.

ROLAP versus MOLAP

- MOLAP (multidimensional OLAP) tools utilize a pre-calculated data set, commonly referred to as a data cube, that contains all the possible answers to a given range of questions.

- ROLAP does not require the pre-computation and storage of information.

- Instead, ROLAP tools access the data in a relational database and generate SQL queries to calculate information at the appropriate level when an end user requests it.

OLAP

Data Design and Considerations:

- An OLAP system stores and uses much less data compared to a data warehouse.
- Data in the OLAP system is summarized. You will find data at the lowest level of detail in the data warehouse.
- OLAP data is more flexible for processing and analysis partly because there is much less data to work with.