Online Analytical Processing

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
  - Limitations to other analysis methods

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?"
- User doesn't stops here with single multidimensional query. he continues to ask for further comparisons to similar products, comparisons among territories, view of results by rotating the presentation between columns & rows.

OLAP

Need for Multidimensional Analysis:

- For effective analysis, user must have easy methods of performing complex analysis across several business dimensions.
- Moreover those methods must have easy & flexible access to information.

OLAP

Need for Multidimensional Analysis:

Decision makers must be able to analyze data:

- **along any number of dimensions**
- **at any level of aggregation**
- **with the capability of viewing results in a variety of ways**
- **with ability to drill down & roll up along the hierarchies of every dimension.**

Without a solid system for true multidimensional analysis, a data warehouse is incomplete.

OLAP : Fast Access and Powerful Calculations

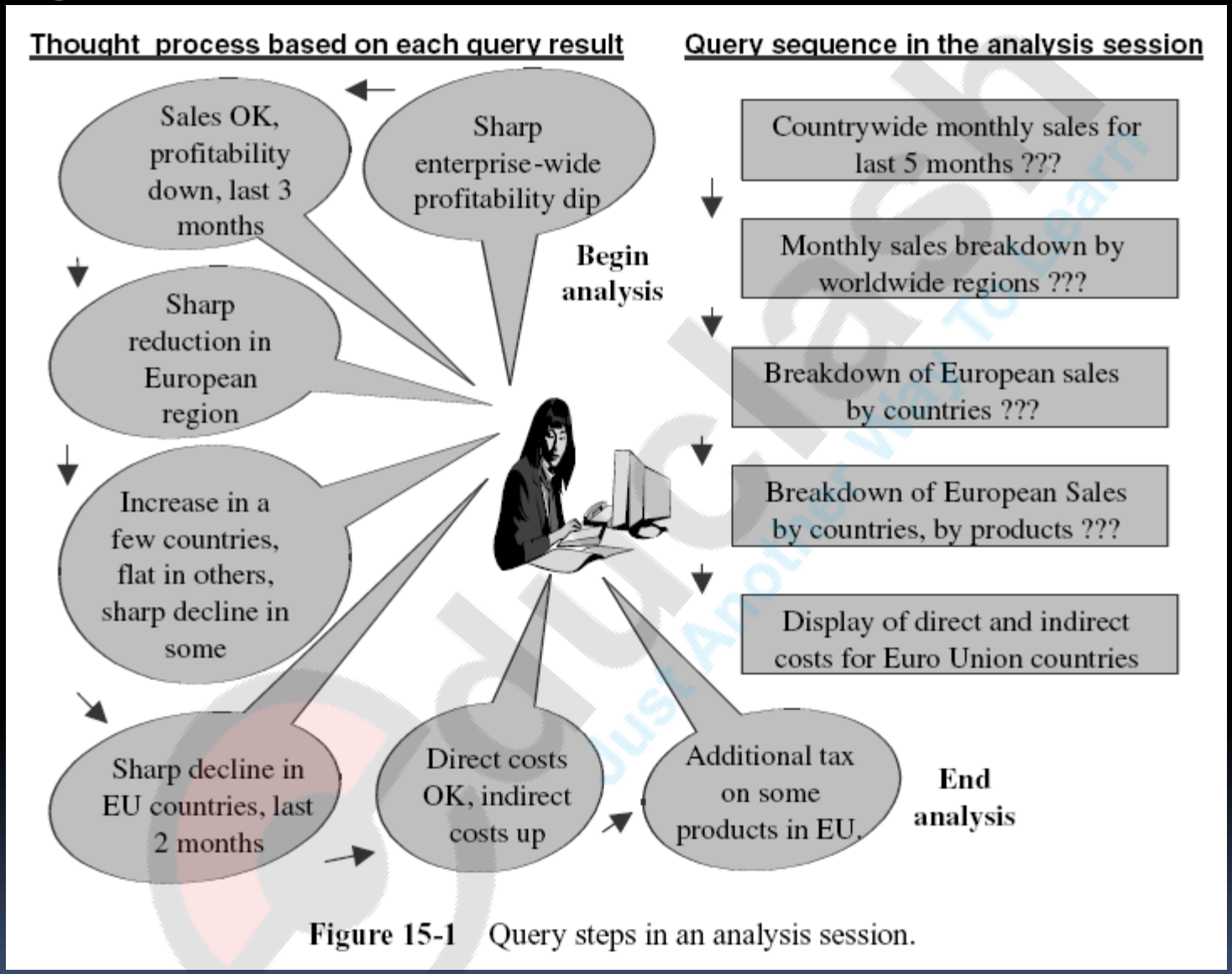


Figure 15-1 Query steps in an analysis session.

In real world analysis sessions queries require complex and powerful calculations: rollups, dril down, trend analysis etc,.

OLAP : Limitations of Other Analysis Methods

- Reports writers : formats output reports, point and click issuing SQL calls
  - → do not support multi dimensionality
  - → no drill down or rotates
  - → Once run cannot alter the presentation of the result data sets
- Spreadsheets :"what if analysis" ,pivot table, aggregations
  - → Multidimensional analysis get difficult with increasing no. of columns
  - → Difficult with increasing hierarchies
  - → No rollup , drill down possible
- Structured Query Language
  - → Vocabulary unsuited for analysis
  - → Multiple joins , unable to handle time series data

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.

OLAP

OLAP Definition :

- The term OLAP was introduced in a paper entitled "Providing On-Line Analytical Processing to User Analysis" by **Dr. E. F. Codd** in 1993.
- The paper defined **12 guidelines** for an OLAP system.
- OLAP Definition: "OLAP is a category of software technology that enables analysts, managers 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 Guidelines

- Multidimensional conceptual view
- Transparency — Underlying data repository should be transparent
- Accessibility — Access only to data that is actually needed
- Consistent reporting pe — No degradation in reporting as DIM size increases
- Client/Server architecture — Interopretability and performance
- Generic dimensionality — No bias towards any single data dimension
- Dynamic Sparse matrix handling — Adapt schema
- Multiuser support — End user to work concurrently with same or different models
- Unrestricted cross dimensional operation
- Institutive data manipulation — Avoid use of menu or multiple trips to UI
- Flexible report — Every dim and subsets must be able to be displayed with equal ease
- Unlimited dimension and aggregation levels

OLTP vs. OLAP

- On-Line Transaction Processing (OLTP):
  - technology used to perform updates on operational or transactional systems (e.g., point of sale systems)
- On-Line Analytical Processing (OLAP):
  - technology used to perform complex analysis of the data in a data warehouse

OLTP vs. OLAP

	OLTP	OLAP
User	Clerk, IT Professional	Knowledge worker
Function	Day to day operations	Decision support
DB Design	Application-oriented (E-R based)	Subject-oriented (Star, snowflake)
Data	Current, Isolated	Historical, Consolidated
View	Detailed, Flat relational	Summarized, Multidimensional
Usage	Structured, Repetitive	Ad hoc
Unit of work	Short, Simple transaction	Complex query
Access	Read/write	Read Mostly
# Records accessed	Tens	Millions
#Users	Thousands	Hundreds
Db size	100 MB-GB	100GB-TB

OLAP

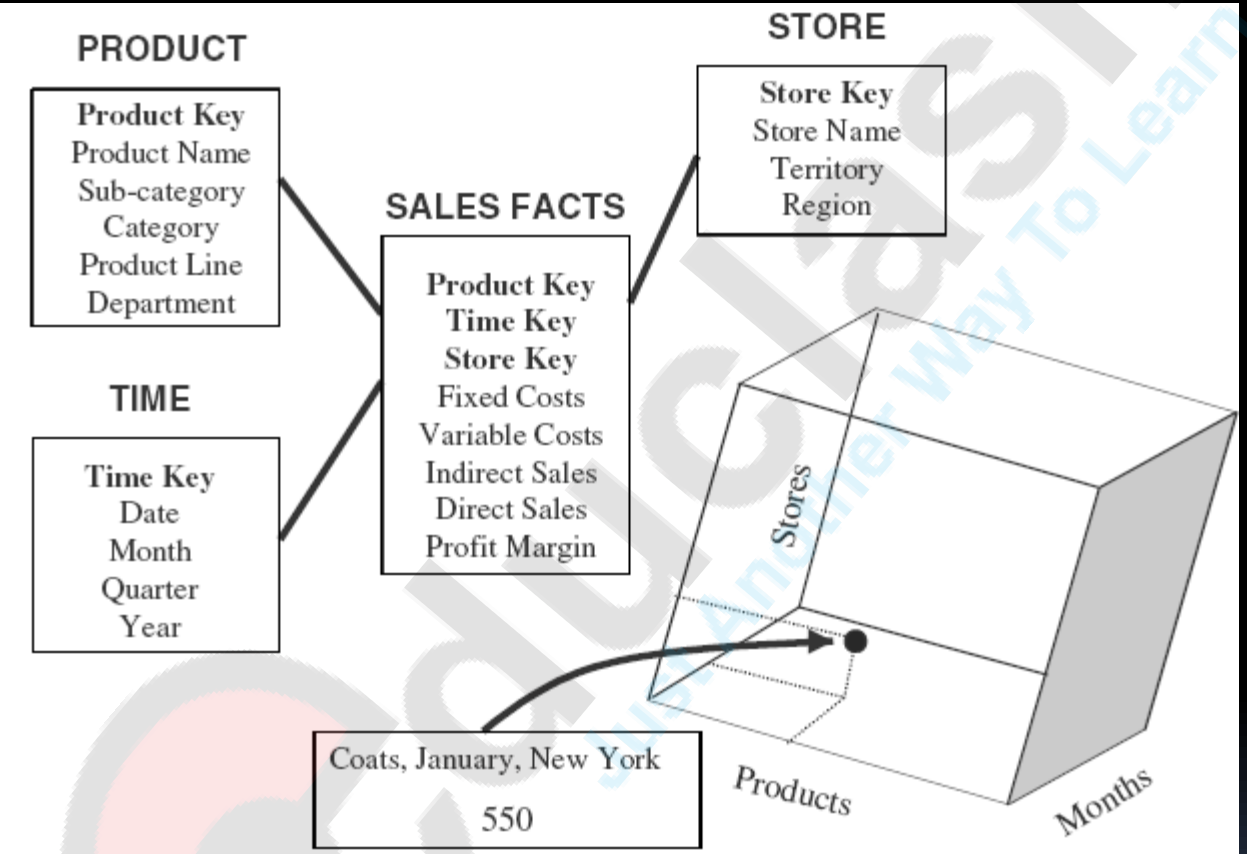


Figure 15-5 Simple STAR schema.

How do we represent this cube on a 2 D page

Store: New York

PAGES: STORE dimension

Products

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

	Hats	Coats	Jackets	Dresses	Shirts	Slacks
Jan	200	550	350	500	520	490
Feb	210	480	390	510	530	500
Mar	190	480	380	480	500	470
Apr	190	430	350	490	510	480
May	160	530	320	530	550	520
Jun	150	450	310	540	560	330
Jul	130	480	270	550	570	250
Aug	140	570	250	650	670	230
Sep	160	470	240	630	650	210
Oct	170	480	260	610	630	250
Nov	180	520	280	680	700	260
Dec	200	560	320	750	770	310

Figure 15-6 A Three-dimensional display.

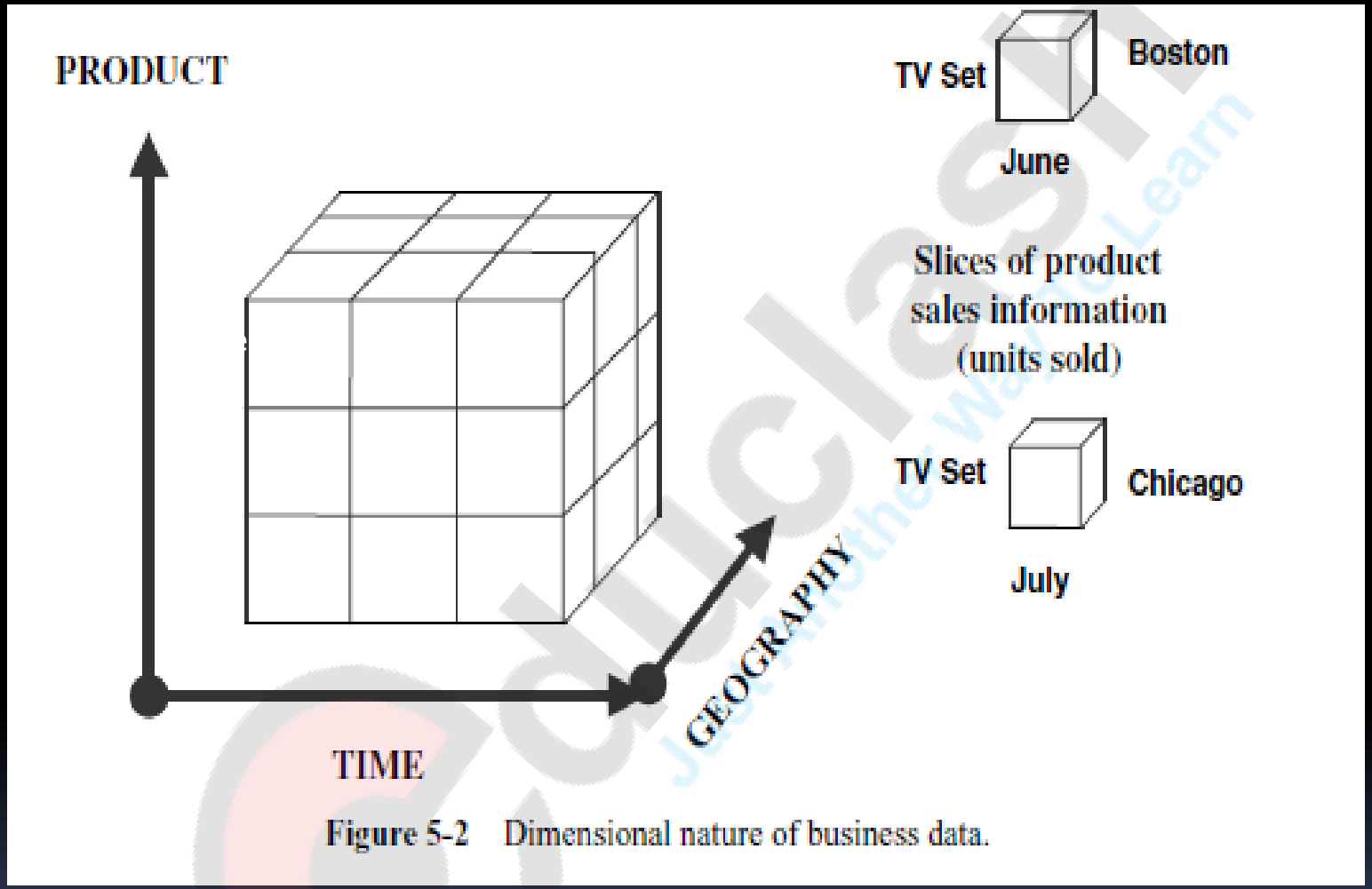
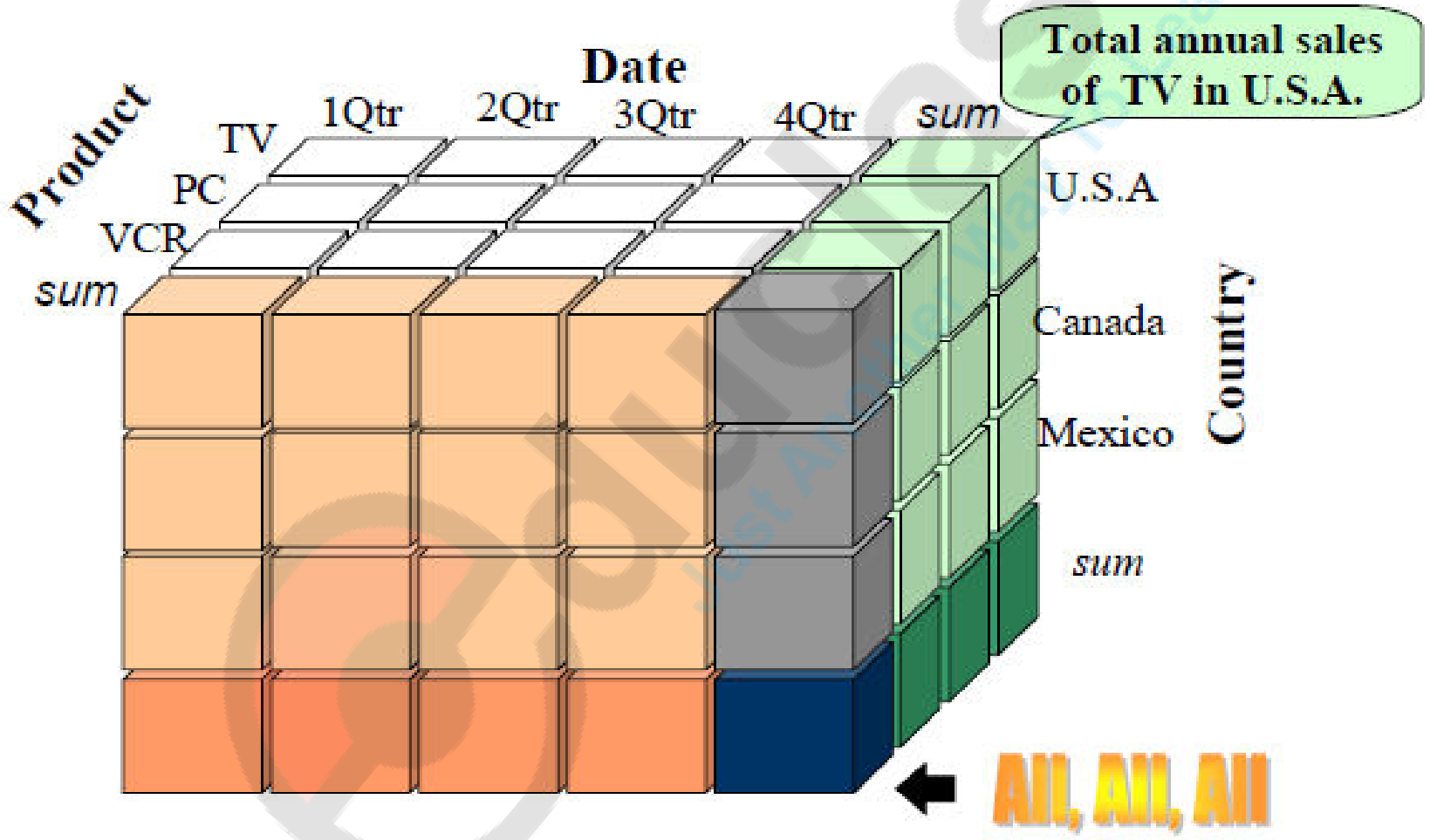


Figure 5-2 Dimensional nature of business data.

A Sample Data Cube



Sales Volume as a function of time, city and product

OLAP

Store: New York

PAGES: STORE dimension

Products

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

	Hats	Coats	Jackets	Dresses	Shirts	Slacks
Jan	200	550	350	500	520	490
Feb	210	480	390	510	530	500
Mar	190	480	380	480	500	470
Apr	190	430	350	490	510	480
May	160	530	320	530	550	520
Jun	150	450	310	540	560	330
Jul	130	480	270	550	570	250
Aug	140	570	250	650	670	230
Sep	160	470	240	630	650	210
Oct	170	480	260	610	630	250
Nov	180	520	280	680	700	260
Dec	200	560	320	750	770	310

Figure 15-6 A Three-dimensional display.

OLAP: simple queries and result sets during a multidimensional analysis sessiion

Query

Display the total sales of all products for past five years in all stores.

Display of Results

Rows: Year numbers 2000, 1999, 1998, 1997, 1996

Columns: Total Sales for all products

Page: One store per page

Query

Compare total sales for all stores, product by product, between years 2000 and 1999.

Display of Results

Rows: Year numbers 2000, 1999; difference; percentage increase or decrease

Columns: One column per product, showing all products

Page: All stores

Query → Show above query by comparisons only for reduced sales

Query → Show above query by switching row and columns

Query → Show above query by switching pages with rows

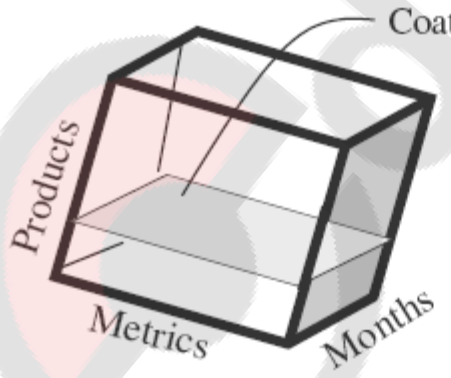
OLAP

PRODUCT: Coats

PAGES: PRODUCT dimension COLUMNS: Metrics

ROWS: TIME dimension

	Fixed Cost	Variable Cost	Indirect Sales	Direct Sales	Profit Margin
Jan	340	110	230	320	100
Feb	270	90	200	260	100
Mar	310	100	210	270	70
Apr	340	110	210	320	80
May	330	110	230	300	90
Jun	260	90	150	300	100
Jul	310	100	180	300	70
Aug	380	130	210	360	60
Sep	300	100	180	290	70
Oct	310	100	170	310	70
Nov	330	110	210	310	80
Dec	350	120	200	360	90



Multidimensional Domain Structure

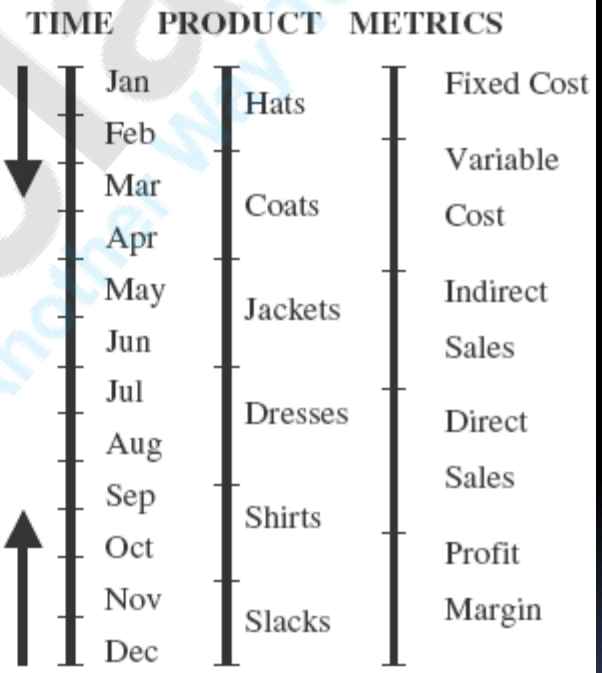
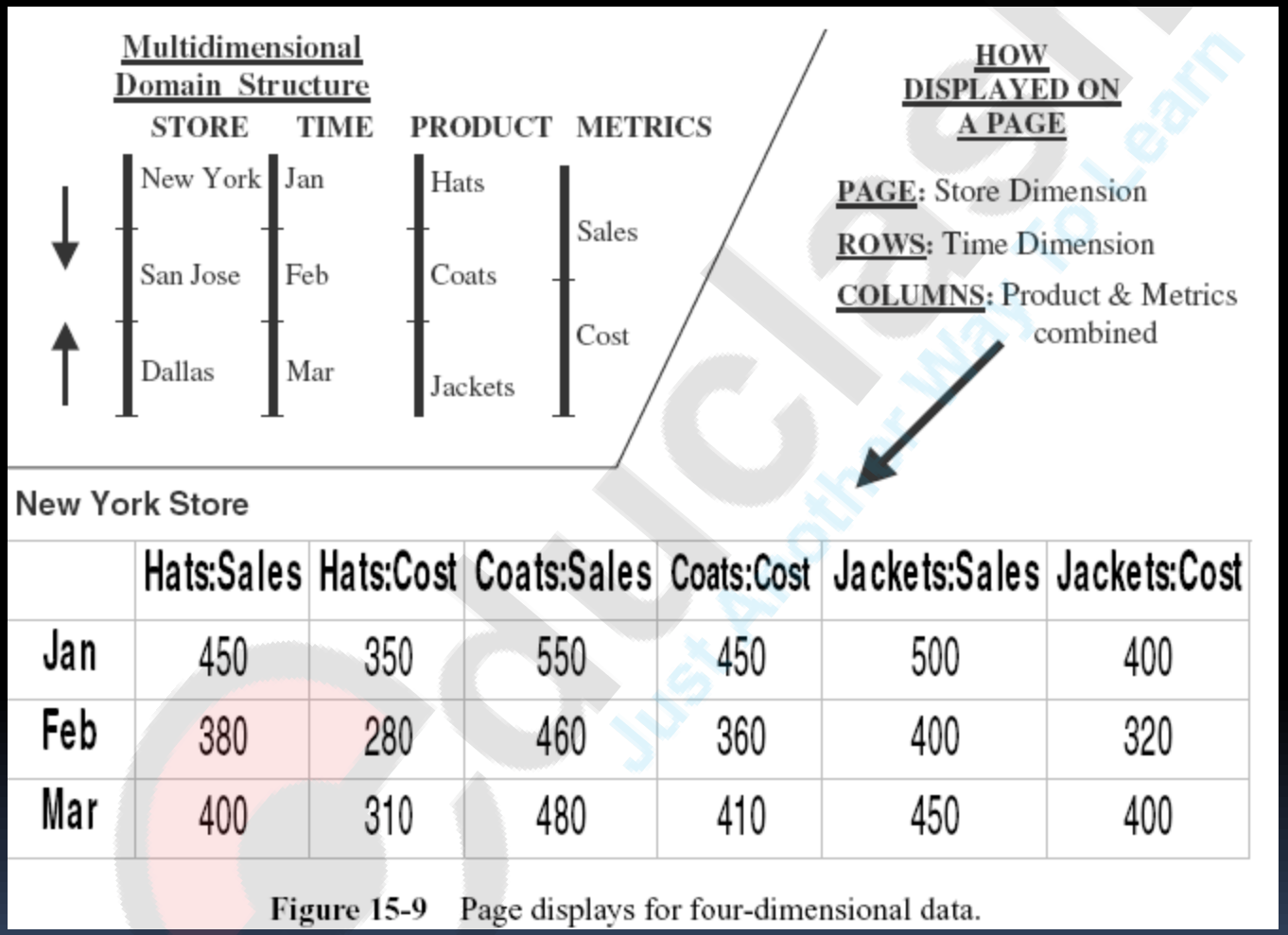


Figure 15-7 Display of columns, rows, and pages.

OLAP



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.

MDS → intuitive representation that accommodates more than 3 dimensions or Hypecube

OLAP

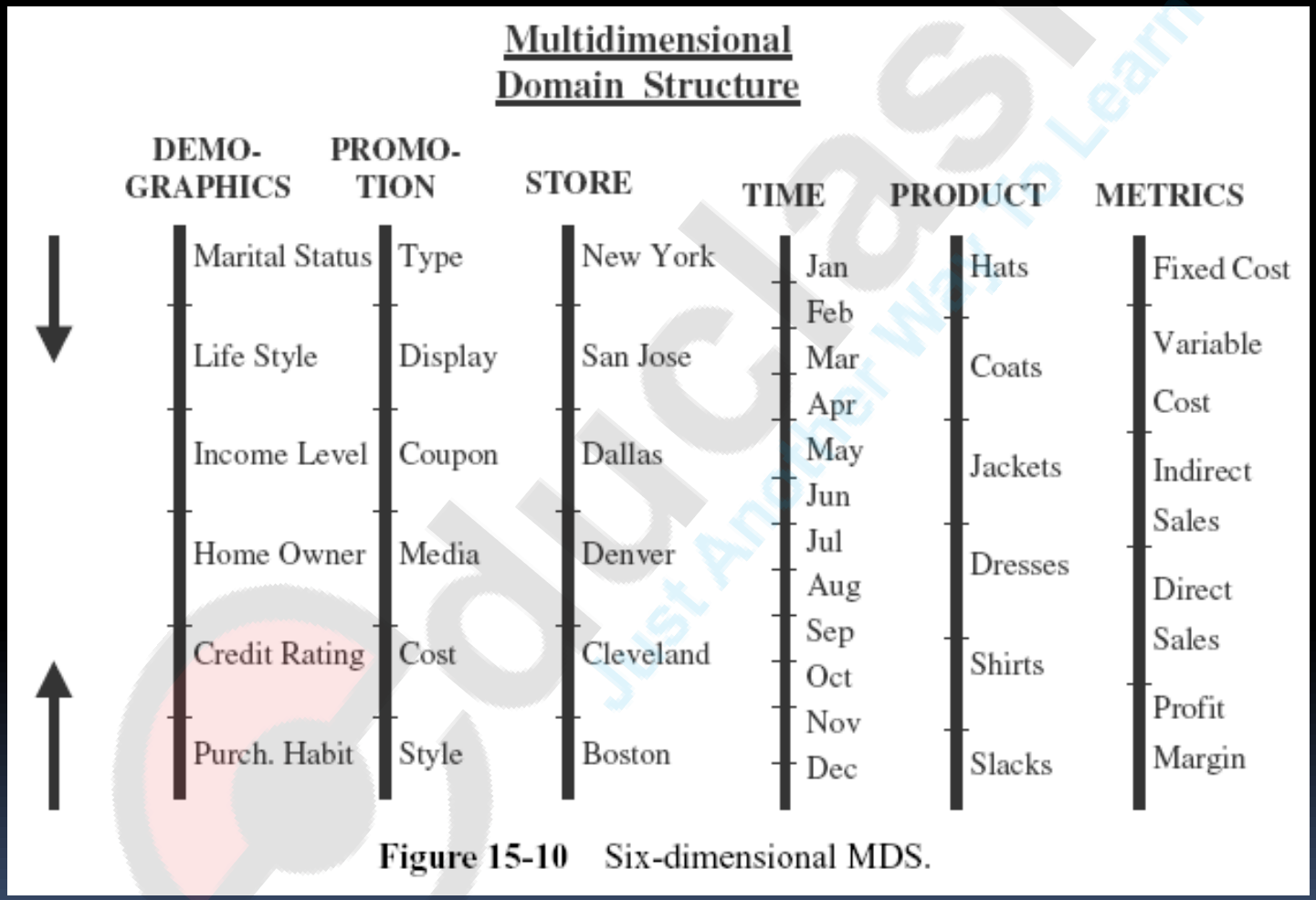
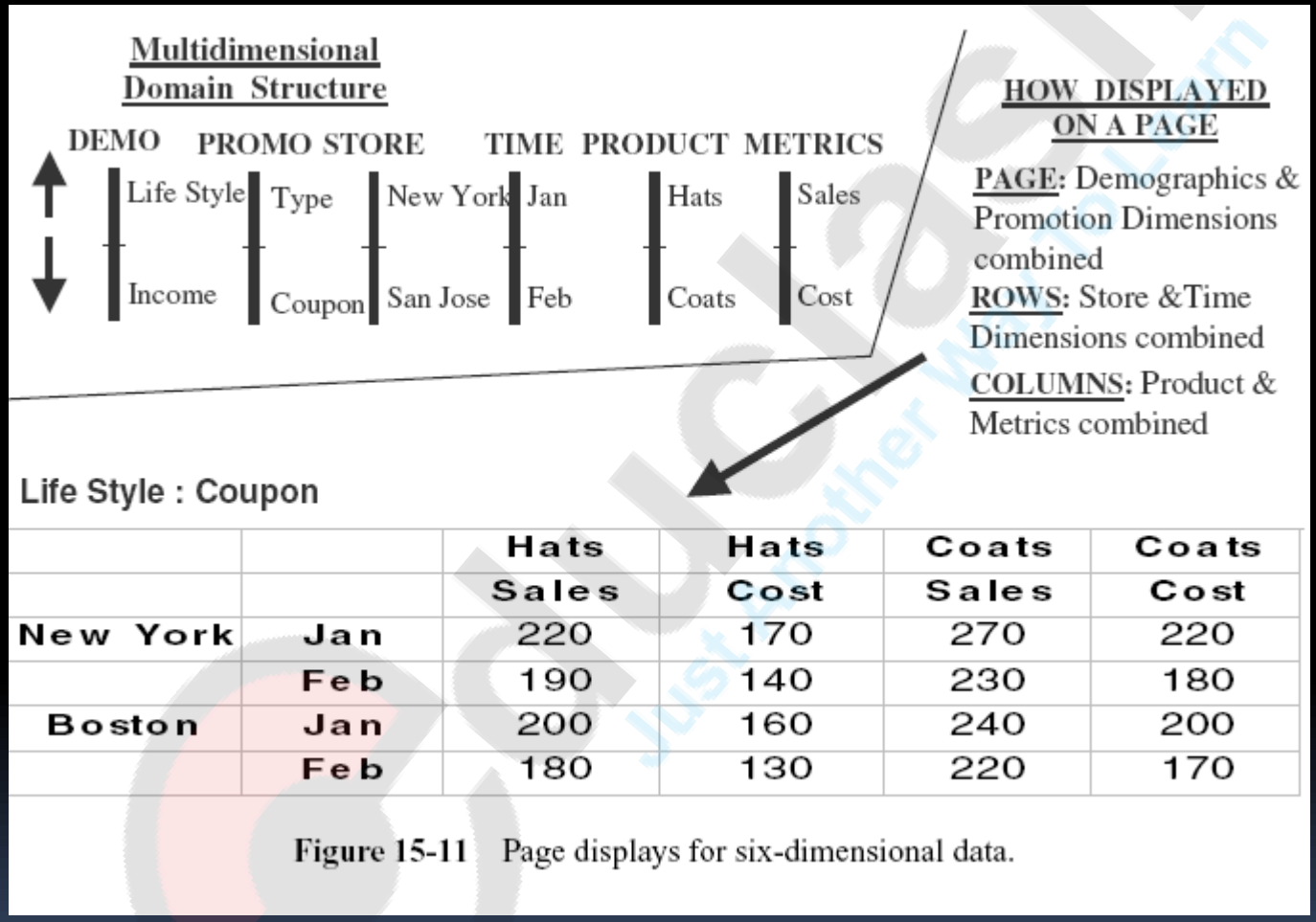


Figure 15-10 Six-dimensional MDS.

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.

Two significant aspects of OLAP are:

- Drill-down & Roll-up
- Slice-and-Dice or Rotation

Drill Down

Region	Sales variance
Africa	105%
Asia	57%
Europe	122%
North America	97%
Pacific	85%
South America	163%

Nation	Sales variance
China	123%
Japan	52%
India	87%
Singapore	95%

OLAP

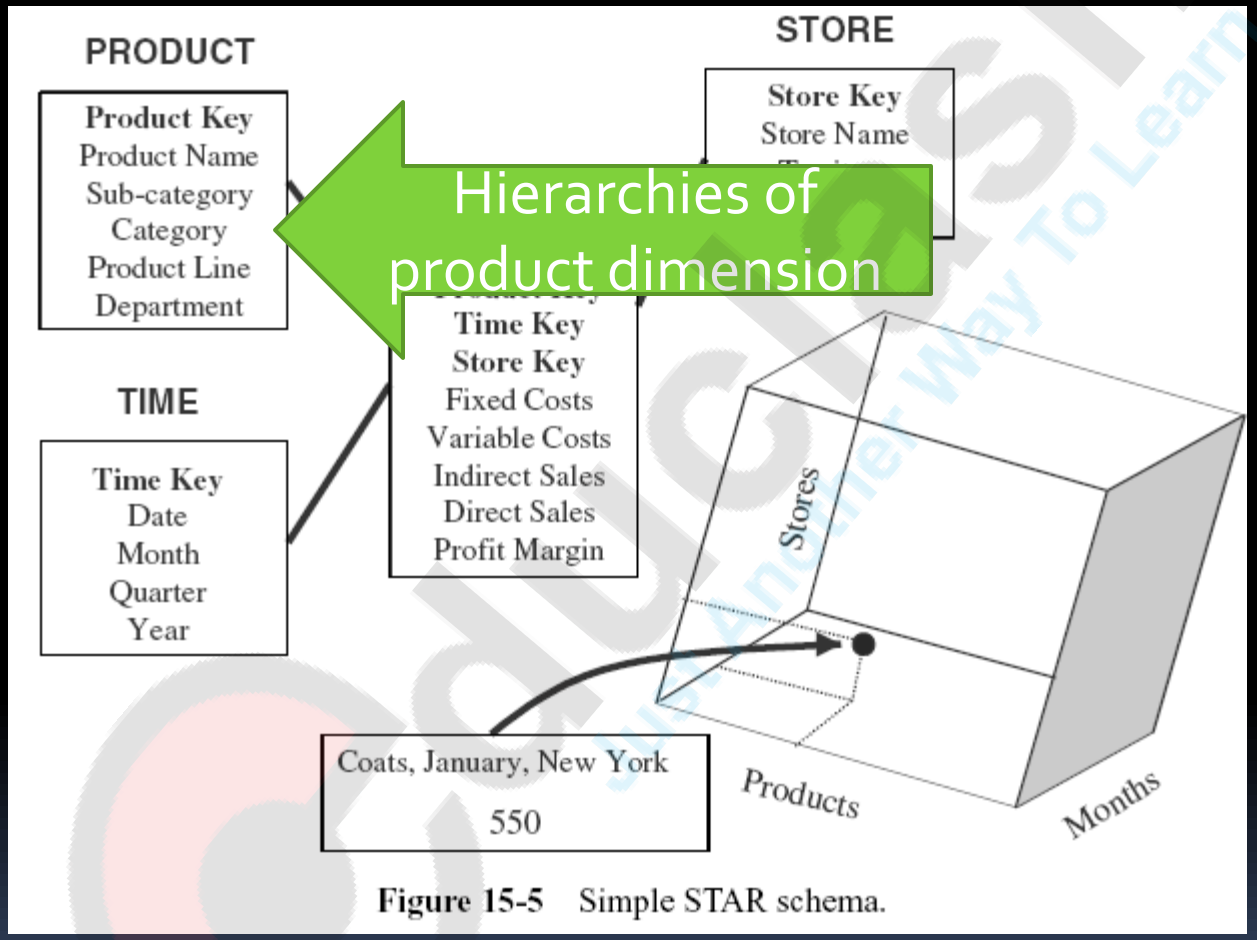


Figure 15-5 Simple STAR schema.

OLAP

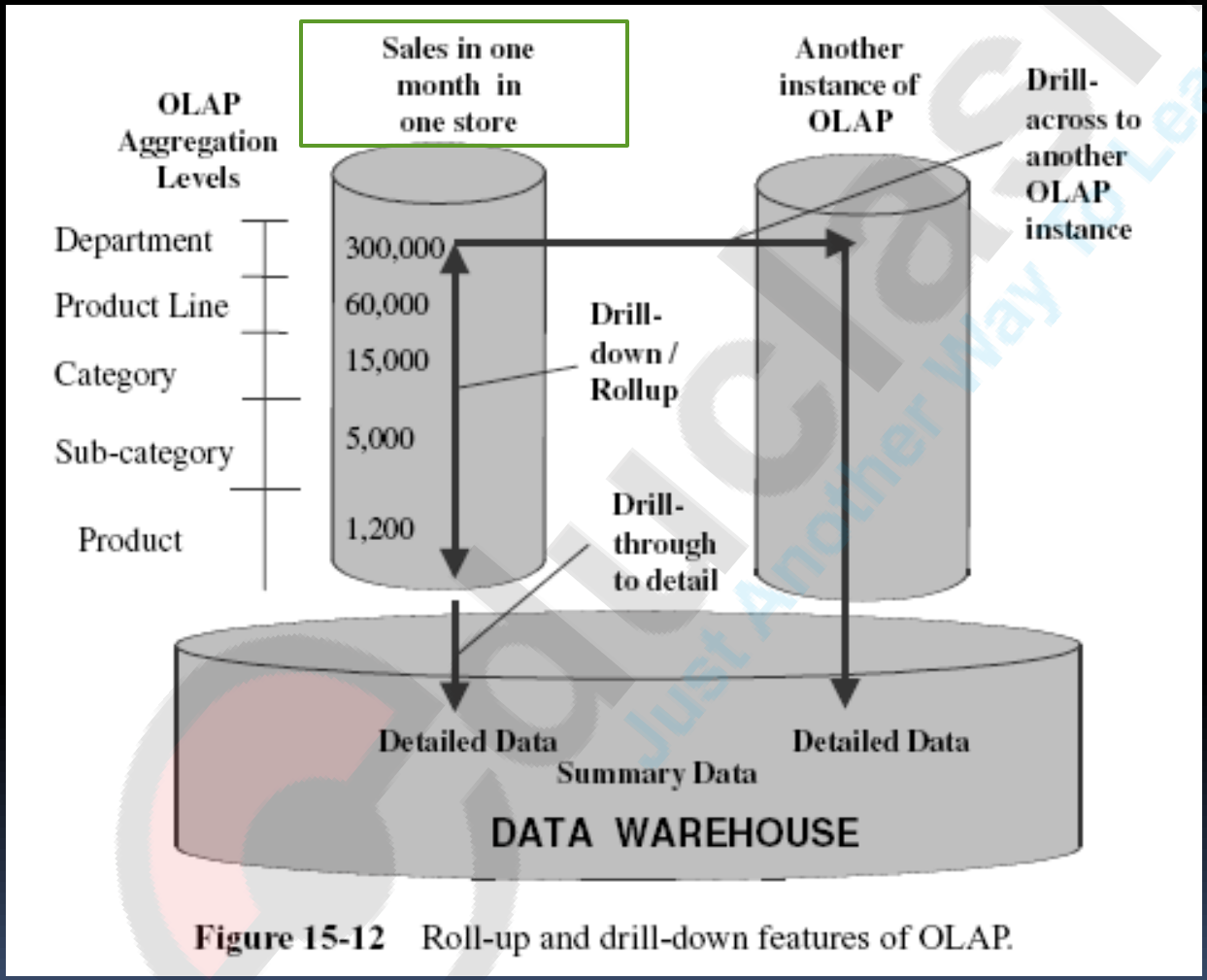


Figure 15-12 Roll-up and drill-down features of OLAP.

OLAP

While drilling down or rolling up how the page display changes on the spreadsheets?

Store: New York

PAGES: STORE dimension

Products

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

Months

	Hats	Coats	Jackets	Dresses	Shirts	Slacks
Jan	200	550	350	500	520	490
Feb	210	480	390	510	530	500
Mar	190	480	380	480	500	470
Apr	190	430	350	490	510	480
May	160	530	320	530	550	520
Jun	150	450	310	540	560	330
Jul	130	480	270	550	570	250
Aug	140	570	250	650	670	230
Sep	160	470	240	630	650	210
Oct	170	480	260	610	630	250
Nov	180	520	280	680	700	260
Dec	200	560	320	750	770	310

Figure 15-6 A Three-dimensional display.

OLAP

Drill-Down and Roll-Up

Store: New York

Sub-categories

PAGES: STORE dimension

COLUMNS: PRODUCT dimension

ROWS: TIME dimension

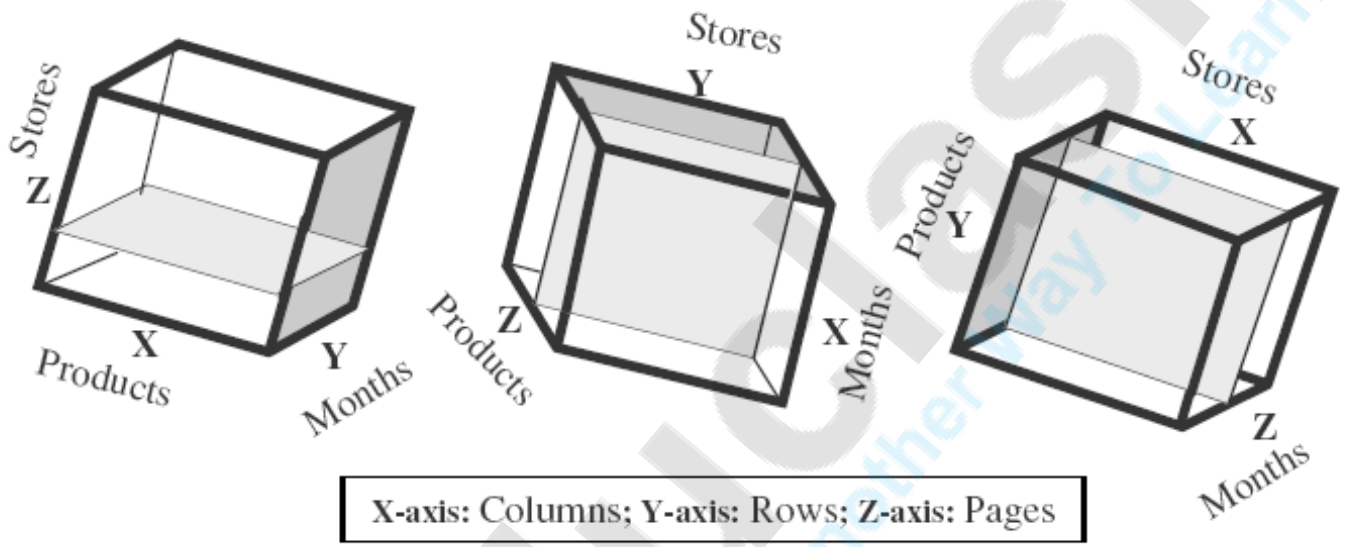
Months

	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.

OLAP

Slice-and-Dice or Rotation



Store: New York

	Hats	Coats	Jackets
Jan	200	550	350
Feb	210	480	390
Mar	190	480	380

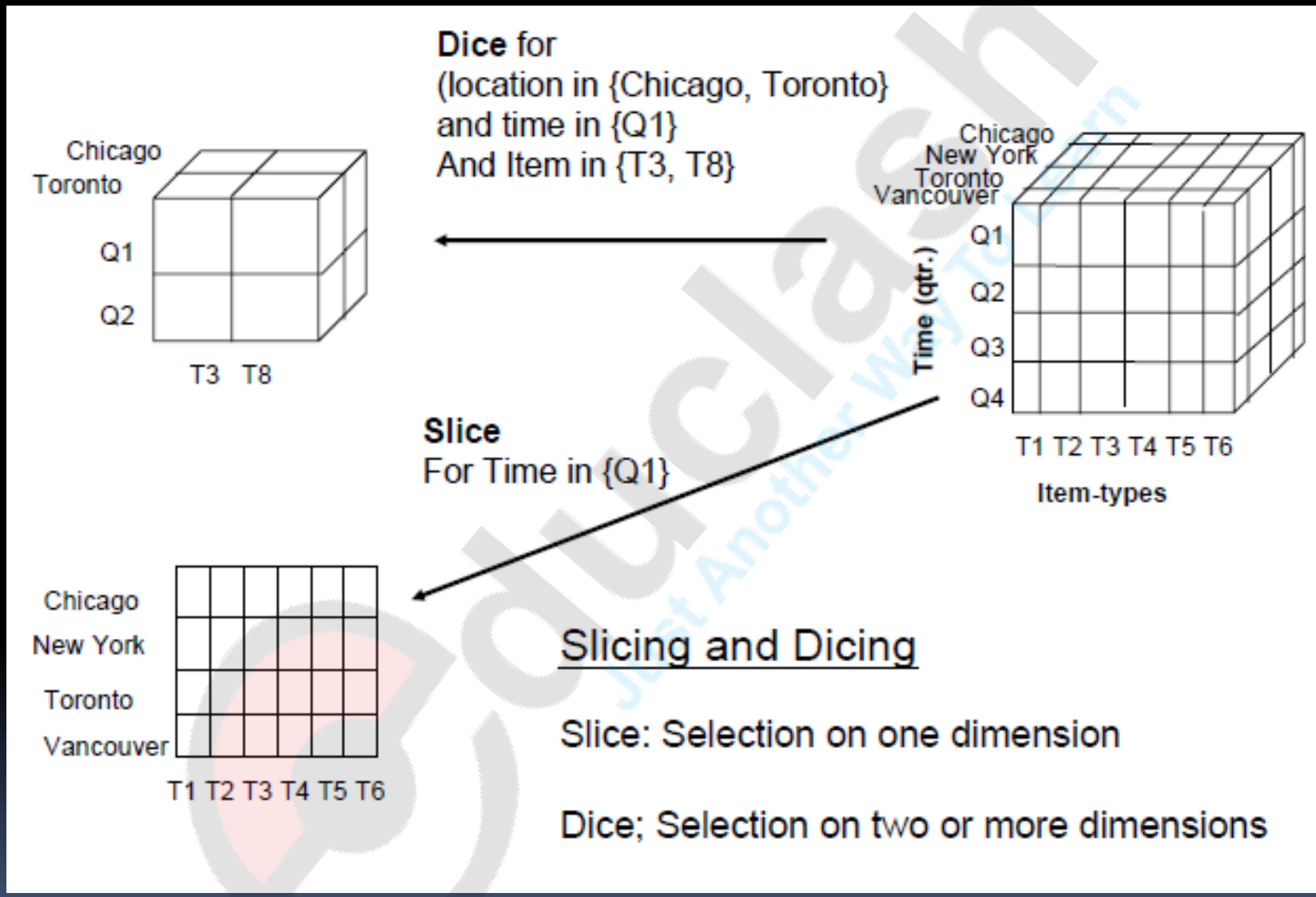
Product: Hats

	Jan	Feb	Mar
New York	200	210	190
Boston	210	250	240
San Jose	130	90	70

Month: January

	New York	Boston	San Jose
Hats	200	210	130
Coats	550	500	200
Jackets	350	400	100

Figure 15-14 Slicing and dicing.


Dice for
(location in {Chicago, Toronto}
and time in {Q1}
And Item in {T3, T8}
Chicago
Toronto
Q1
Q2
T3 T8
Chicago
New York
Toronto
Vancouver
Time (qtr.)
Q1
Q2
Q3
Q4
T1 T2 T3 T4 T5 T6
Item-types
Slice
For Time in {Q1}
Chicago
New York
Toronto
Vancouver
T1 T2 T3 T4 T5 T6
Slicing and Dicing
Slice: Selection on one dimension
Dice; Selection on two or more dimensions

Rotation (Pivot Table)

		Region			
Year	Data	Asia	Europe	North America	Grand Total
1995	Sum of Hardware	97	23	198	318
	Sum of Software	83	41	425	549
1996	Sum of Hardware	115	28	224	367
	Sum of Software	78	65	410	553
1997	Sum of Hardware	102	25	259	386
	Sum of Software	55	73	497	625
Total Sum of Hardware		314	76	681	1071
Total Sum of Software		216	179	1332	1727

		Year			
Region	Data	1995	1996	1997	Grand Total
Asia	Sum of Hardware	97	115	102	314
	Sum of Software	83	78	55	216
Europe	Sum of Hardware	23	28	25	76
	Sum of Software	41	65	73	179
North America	Sum of Hardware	198	224	259	681
	Sum of Software	425	410	497	1332
Total Sum of Hardware		318	367	386	1071
Total Sum of Software		549	553	625	1727

Uses & Benefits:

Now we have studied features of OLAP in sufficient detail, we can deduce the enormous benefits of OLAP.

Summary of benefits of OLAP:

- Given on P-393; Paulraj

FASMI characteristics of OLAP:

- Fast
- Analytic
- Shared
- Multidimensional
- Information

Types of OLAP Models: pg. no. 393 paulraj

- MOLAP (Multidimensional OLAP)
- ROLAP (Relational OLAP)

MOLAP (Multidimensional OLAP)

- Implemented by storing data multidimensionally, i.e. data is stored in the form of hypercube.
- OLAP engine resides on a special server(MDDBS Server).
- Proprietary MDDB stores data in the form of hypercubes.
- We need to run special extraction & aggregation jobs to create these data cubes in MDDB from the relational database of DW.

OLAP

OLAP MODELS

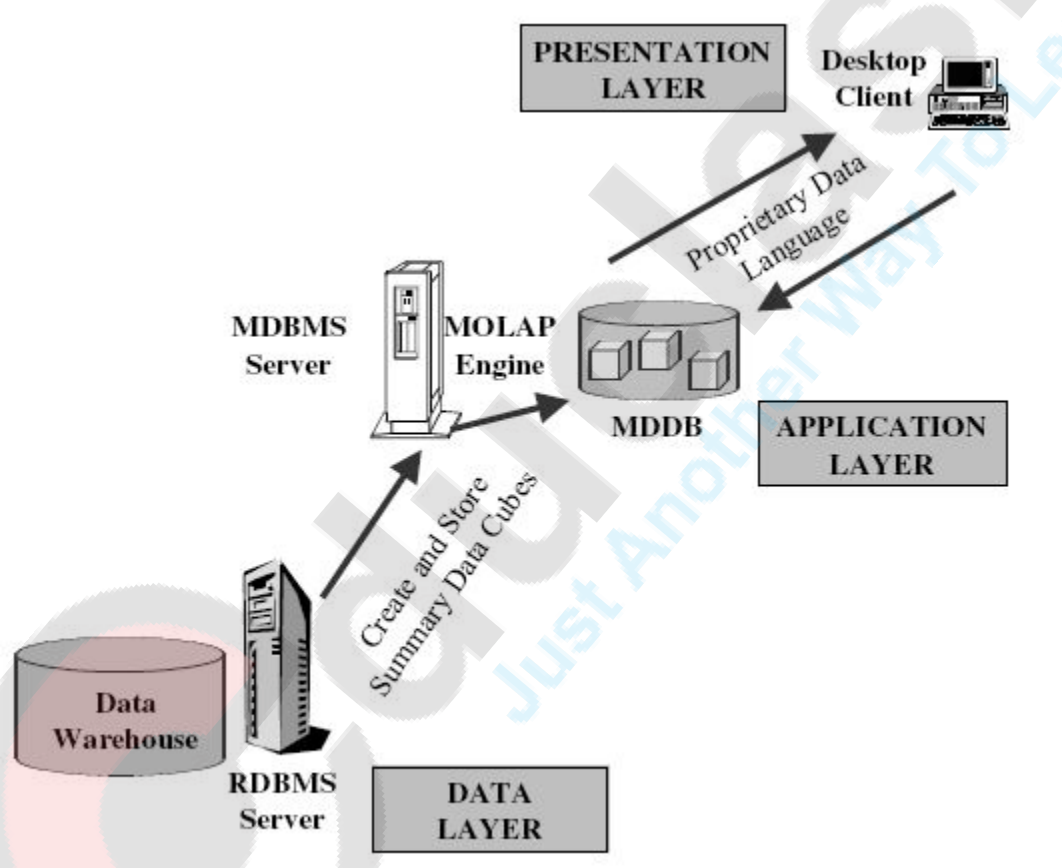


Figure 15-16 The MOLAP model.

MOLAP (Multidimensional OLAP)

- Data stored in array-based structures
- good for smaller warehouses
- optimized for canned queries.
  - Canned queries are predefined queries.
  - Mostly canned queries contain prompts that allow you to customize the query for your specific needs.
- Commercial offerings of MOLAP are available.
- Examples: Hyperion Essbase, Fusion (Information Builders)

ROLAP (Relational OLAP)

- ROLAP relies on existing relational DBMS of the DW. OLAP features are provided against the relational database.
- OLAP engine resides on the desktop.
- Prefabricated multidimensional cubes are not created beforehand & stored in special DB.
- The relational data is presented as virtual multidimensional data cubes.

OLAP

OLAP MODELS

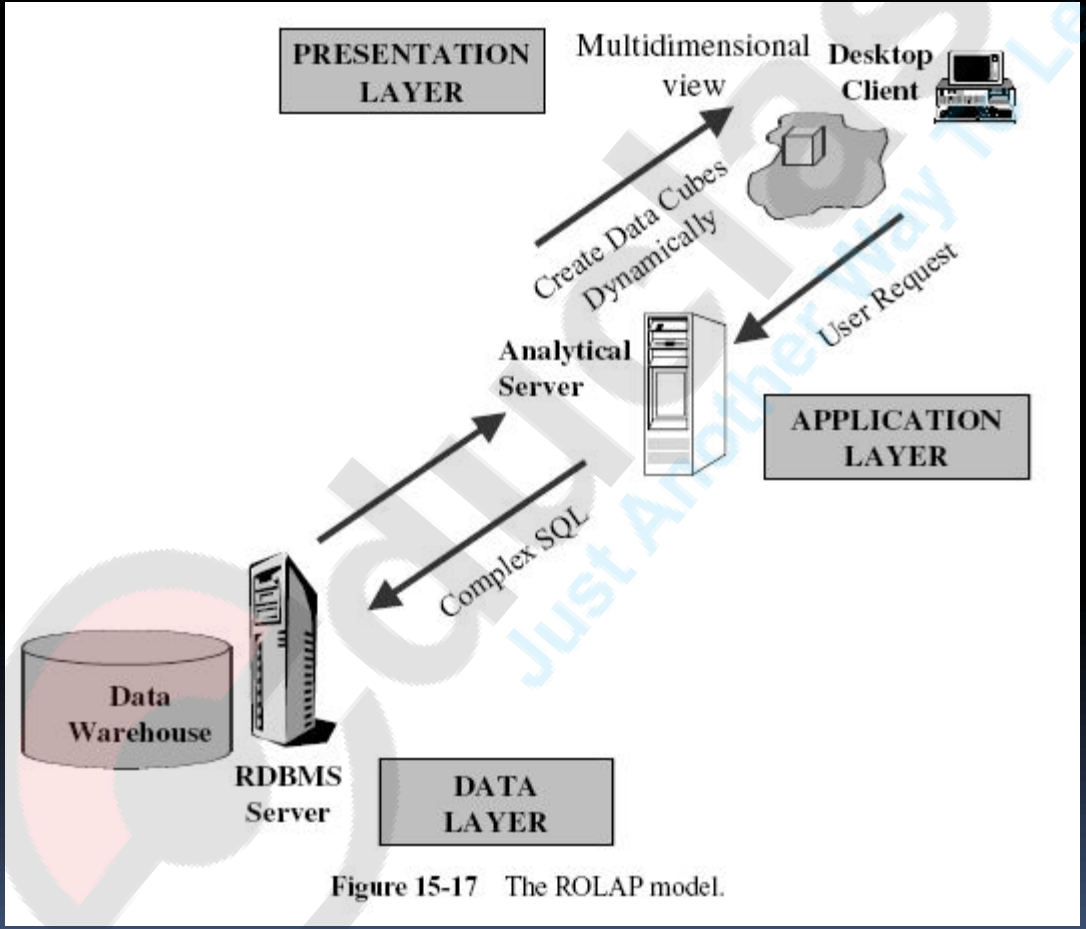


Figure 15-17 The ROLAP model.

ROLAP (Relational OLAP)

- Data stored in relational tables
- ROLAP is more flexible
- May pay a performance penalty to realize flexibility
- Commercial offerings of ROLAP are available
- Examples: Microstrategy Intelligence Server, MetaCube (Informix/IBM)

OLAP

OLAP MODELS

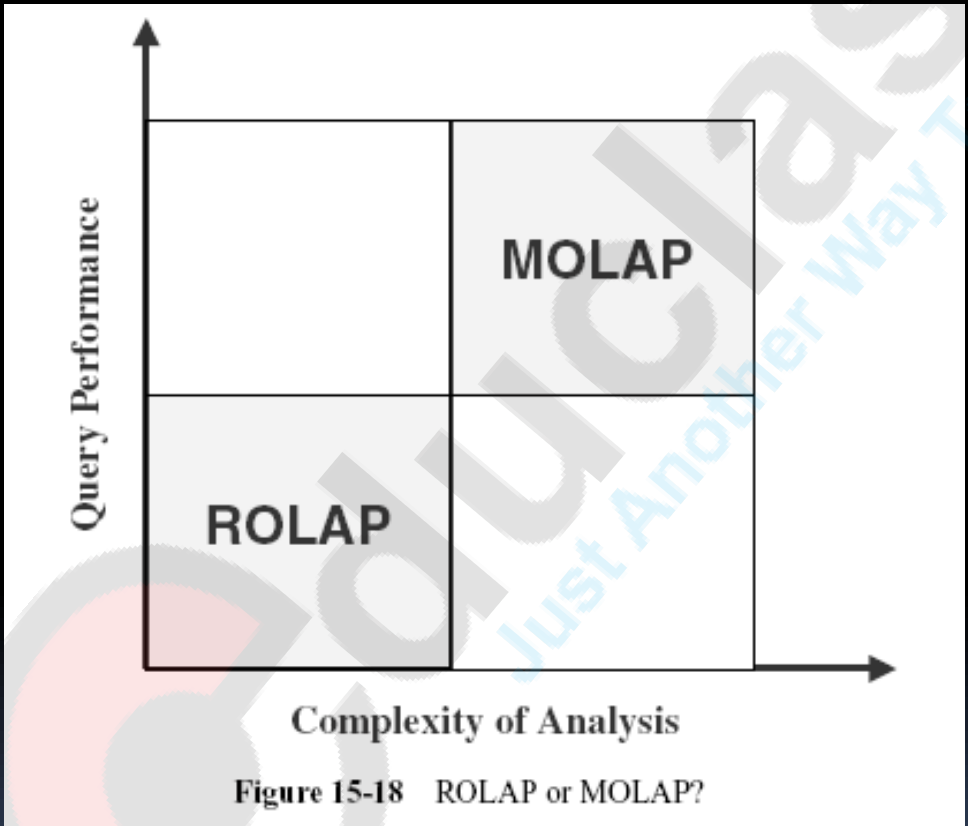


Figure 15-18 ROLAP or MOLAP?

OLAP

ROLAP versus MOLAP.

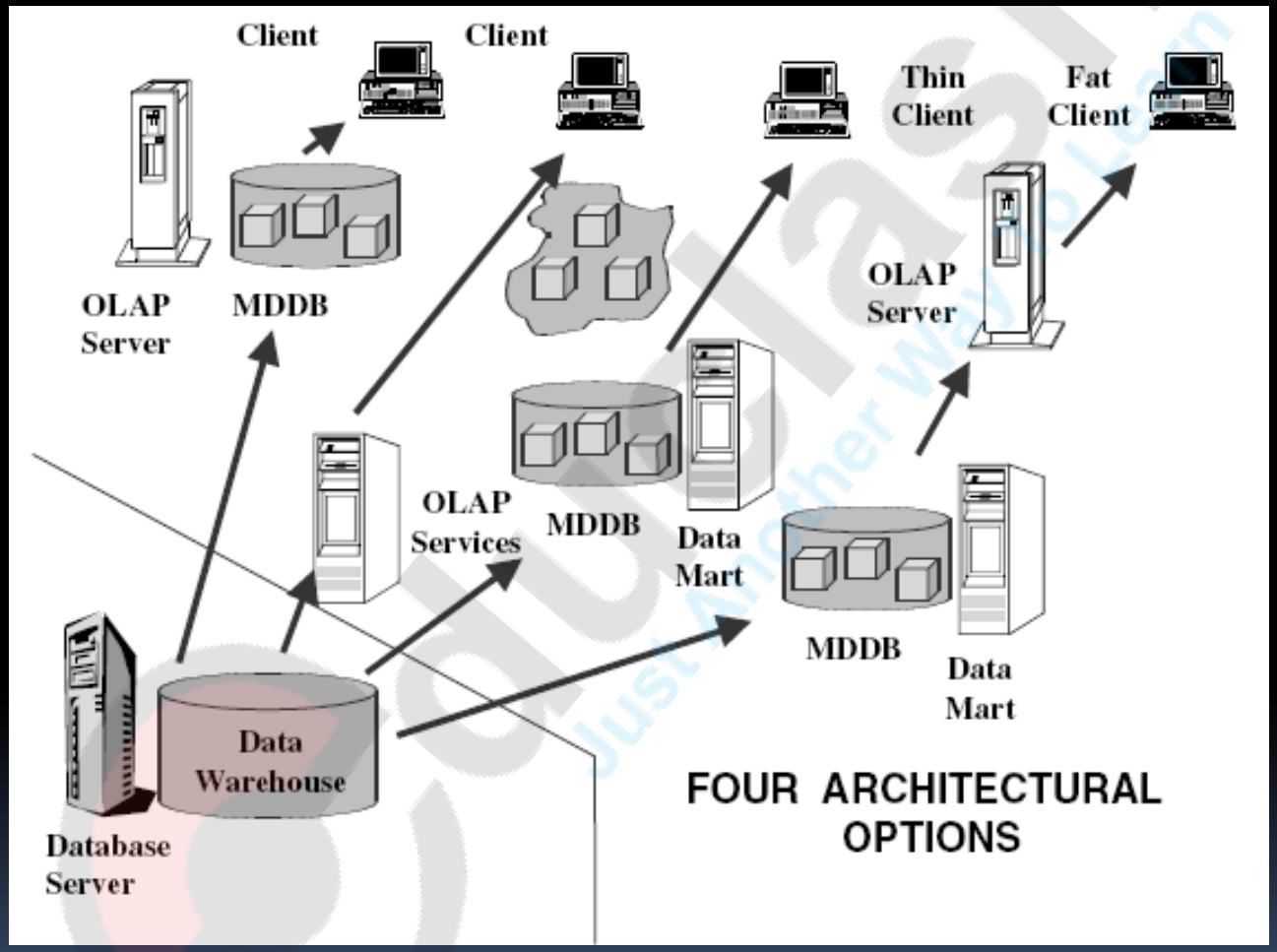
	Data Storage	Underlying Technologies	Functions and Features
ROLAP	Data stored as relational tables in the warehouse. Detailed and light summary data available. Very large data volumes. All data access from the warehouse storage.	Use of complex SQL to fetch data from warehouse. ROLAP engine in analytical server creates data cubes on the fly. Multidimensional views by presentation layer.	Known environment and availability of many tools. Limitations on complex analysis functions. Drill-through to lowest level easier. Drill-across not always easy.
MOLAP	Data stored as relational tables in the warehouse. Various summary data kept in proprietary databases (MDDBs) Moderate data volumes. Summary data access from MDDB, detailed data access from warehouse.	Creation of pre-fabricated data cubes by MOLAP engine. Proprietary technology to store multidimensional views in arrays, not tables. High speed matrix data retrieval. Sparse matrix technology to manage data sparsity in summaries.	Faster access. Large library of functions for complex calculations. Easy analysis irrespective of the number of dimensions. Extensive drill-down and slice-and-dice capabilities.

Thanks

Tools: Warehouse Servers

- The RDBMS dominates:
  - Oracle 8i/9i
  - IBM DB2
  - Microsoft SQL Server
  - Informix (IBM)
  - Red Brick Warehouse (Informix/IBM)
  - NCR Teradata
  - Sybase…

OLAP



OLAP architectural options.

OLAP

Types & levels of data in OLAP systems:

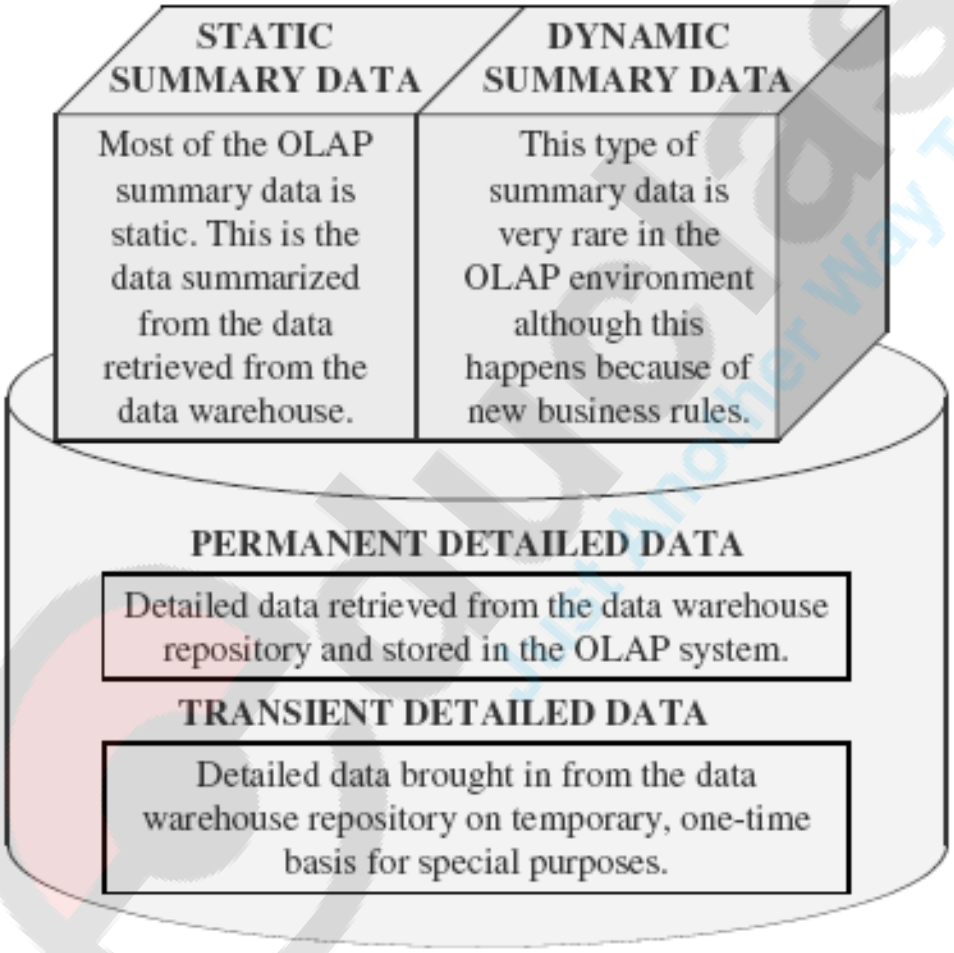


Figure 15-21 Data modeling considerations for OLAP.

OLAP

Implementation Steps:

- Dimensional modeling
- Design and building of the MDDB
- Selection of the data to be moved into the OLAP system
- Data acquisition or extraction for the OLAP system
- Data loading into the OLAP server
- Computation of data aggregation and derived data
- Implementation of application on the desktop
- Provision for user training

OLAP

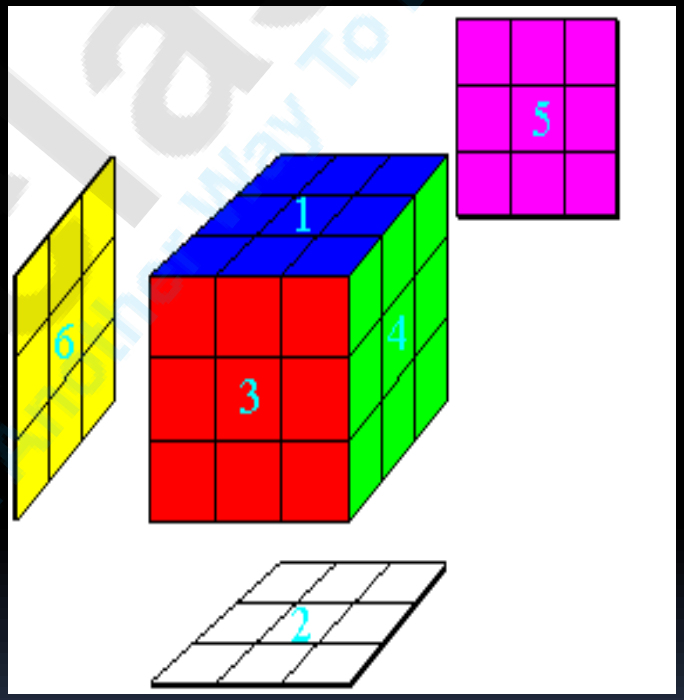
Summary:

- OLAP is critical because its multidimensional analysis, fast access, and powerful calculations exceed that of other analysis methods.
- OLAP is defined on the basis of Codd's initial twelve guidelines.
- OLAP characteristics include multidimensional view of the data, interactive and complex analysis facility, ability to perform intricate calculations, and fast response time.
- Dimensional analysis is not confined to three dimensions that can be represented by a physical cube. Hypercubes provide a method for representing views with more dimensions.
- ROLAP and MOLAP are the two major OLAP models. The difference between them lies in the way the basic data is stored. Ascertain which model is more suitable for your environment.
- OLAP tools have matured. Some RDBMSs include support for OLAP.

It has been claimed that for complex queries OLAP cubes can produce an answer in around 0.1% of the time required for the same query on OLTP relational data

Data Mining versus OLAP

- OLAP - Online Analytical Processing
  - Provides you with a very good view of what is happening, but can not predict what will happen in the future or why it is happening



Results of Data Mining Include:

- Forecasting what may happen in the future
- Classifying people or things into groups by recognizing patterns
- Clustering people or things into groups based on their attributes
- Associating what events are likely to occur together
- Sequencing what events are likely to lead to later events

Sample of OLAP products

BusinessObjects and WebIntelligence	Business Objects
Crystal Analysis and Reports	Business Objects
DB2 Cube Views	IBM
DB2 OLAP Server	IBM
EAP	OutlookSoft
e-Planning	Cognos (Adaytum)
Enterprise and Pillar	Hyperion Solutions
Essbase	Hyperion Solutions
Executive Suite	CIP-Global
Executive Viewer	Temtec
Express and the Oracle9*i* OLAP Option	Oracle
Khalix	Longview Solutions
Magnitude	Cartesis

Just a snippet from http://www.olapreport.com/ProductsIndex.htm