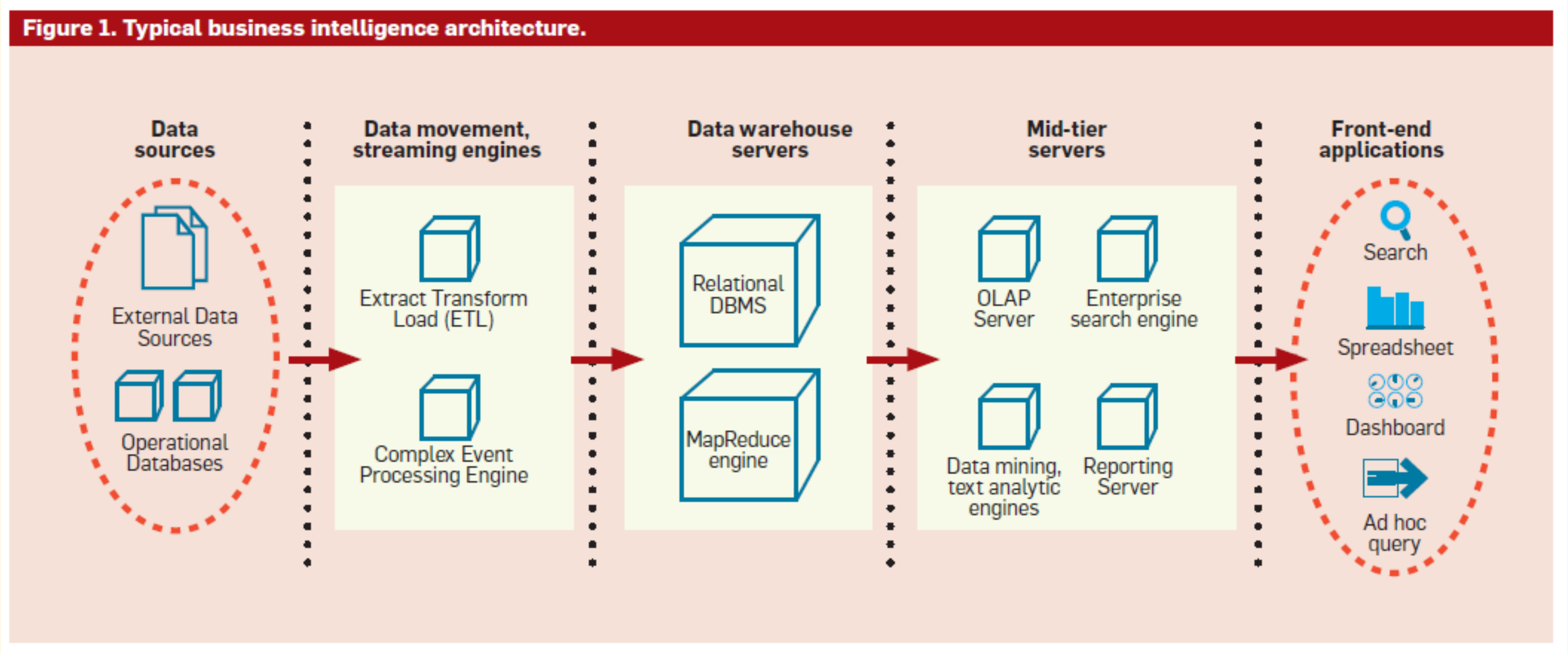
BUSINESS INTELLIGENCE

OLAP: On-Line Analytical Processing

BI Architecture



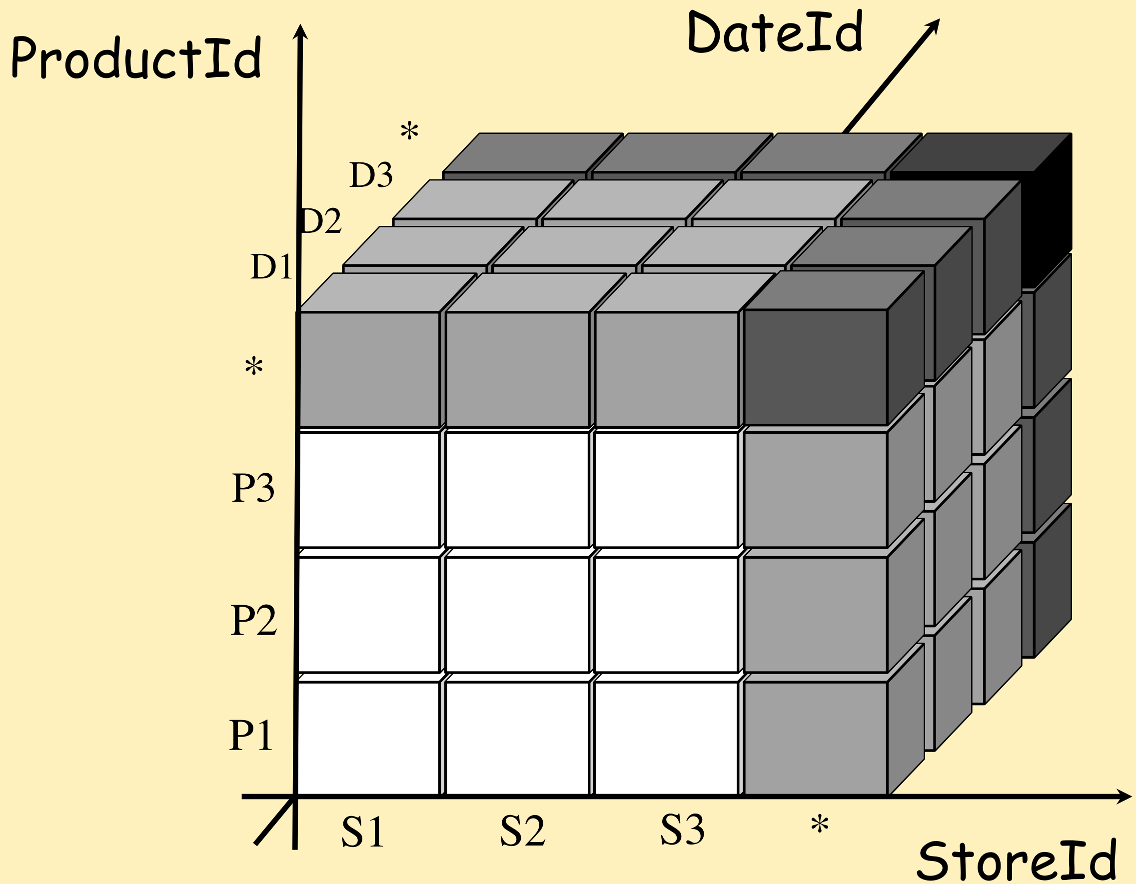
WHICH DBMS FOR DATAWAREHOUSING?

ON-LINE ANALYTICAL PROCESSING (OLAP)

- An OLAP server provides a *multidimensional view* starting from a datawarehouse

- The multidimensional view can be navigated through pivot tables, reports, 2-D or 3-D plots, or it can be queried using a query language (eg., MDX – MultiDimensional eXpressions)

The multidimensional model is useful to understand interactive data analysis, and how to improve the execution performance.



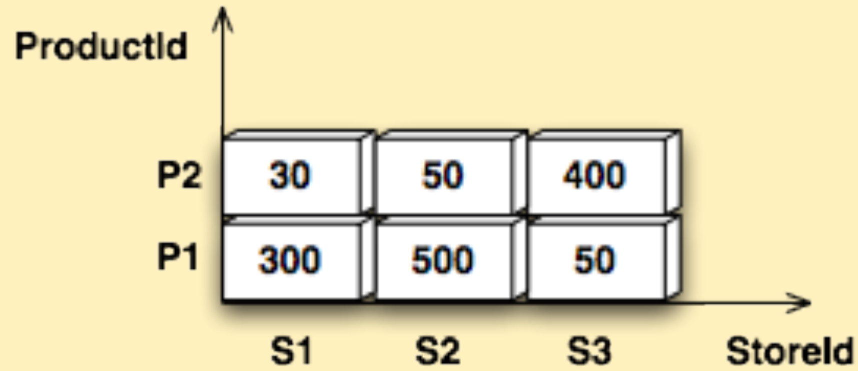
2-D CUBE

M

Sales

StoreId	ProductId	Qty
S1	P1	300
S2	P1	500
S3	P1	50
S1	P2	30
S2	P2	50
S3	P2	400

Fact Table



2-D Cube

CROSS TABULATION

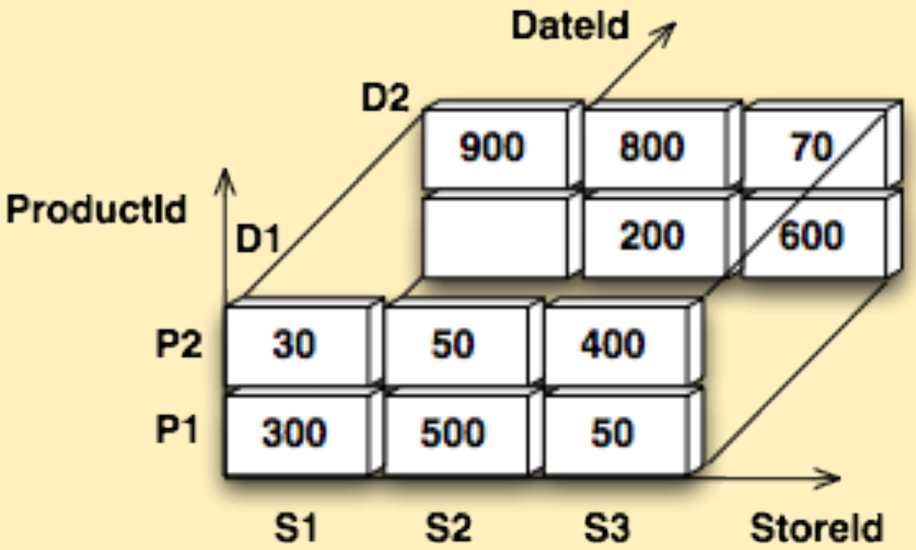
	StoreId		
ProductId	S1	S2	S3
P1	300	500	50
P2	30	50	400

3-D CUBE

Sales

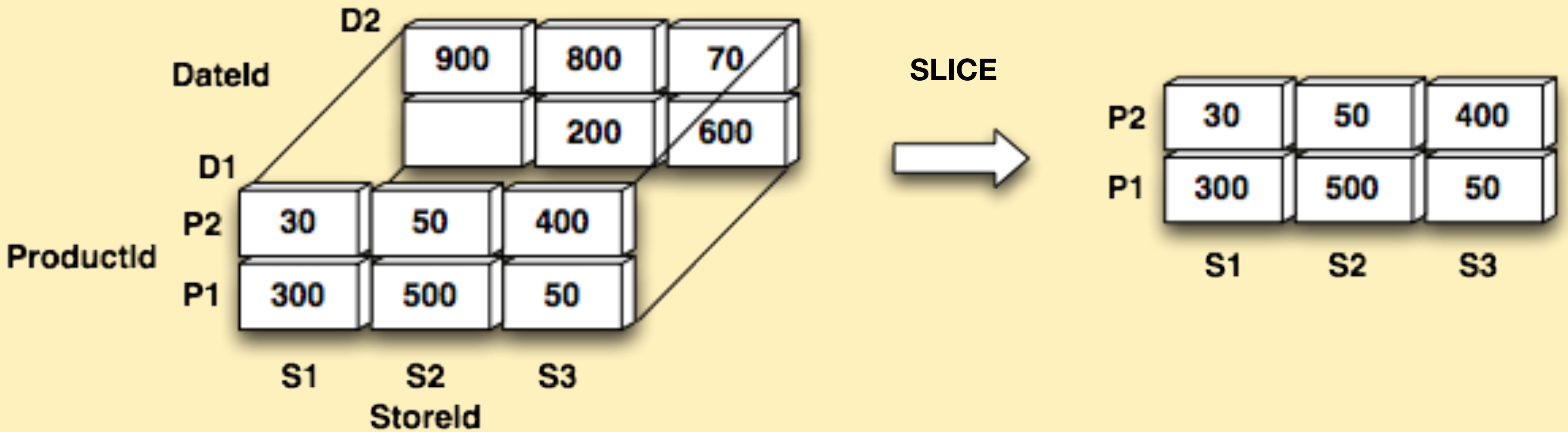
StoreId	ProductId	DateId	Qty
S1	P1	D1	300
S2	P1	D1	500
S3	P1	D1	50
S1	P2	D1	30
S2	P2	D1	50
S3	P2	D1	400
S2	P1	D2	200
S3	P1	D2	600
S1	P2	D2	900
S2	P2	D2	800
S3	P2	D2	70

Fact Table



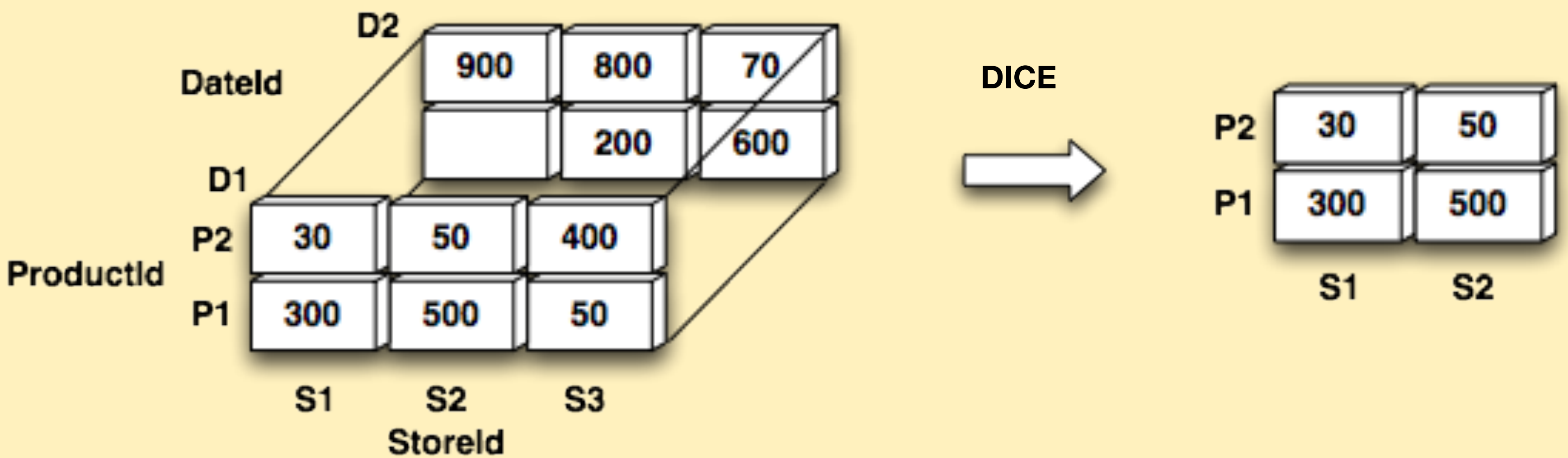
3-D Cube

Sales **SLICE FOR** DateId = ‘D1’;



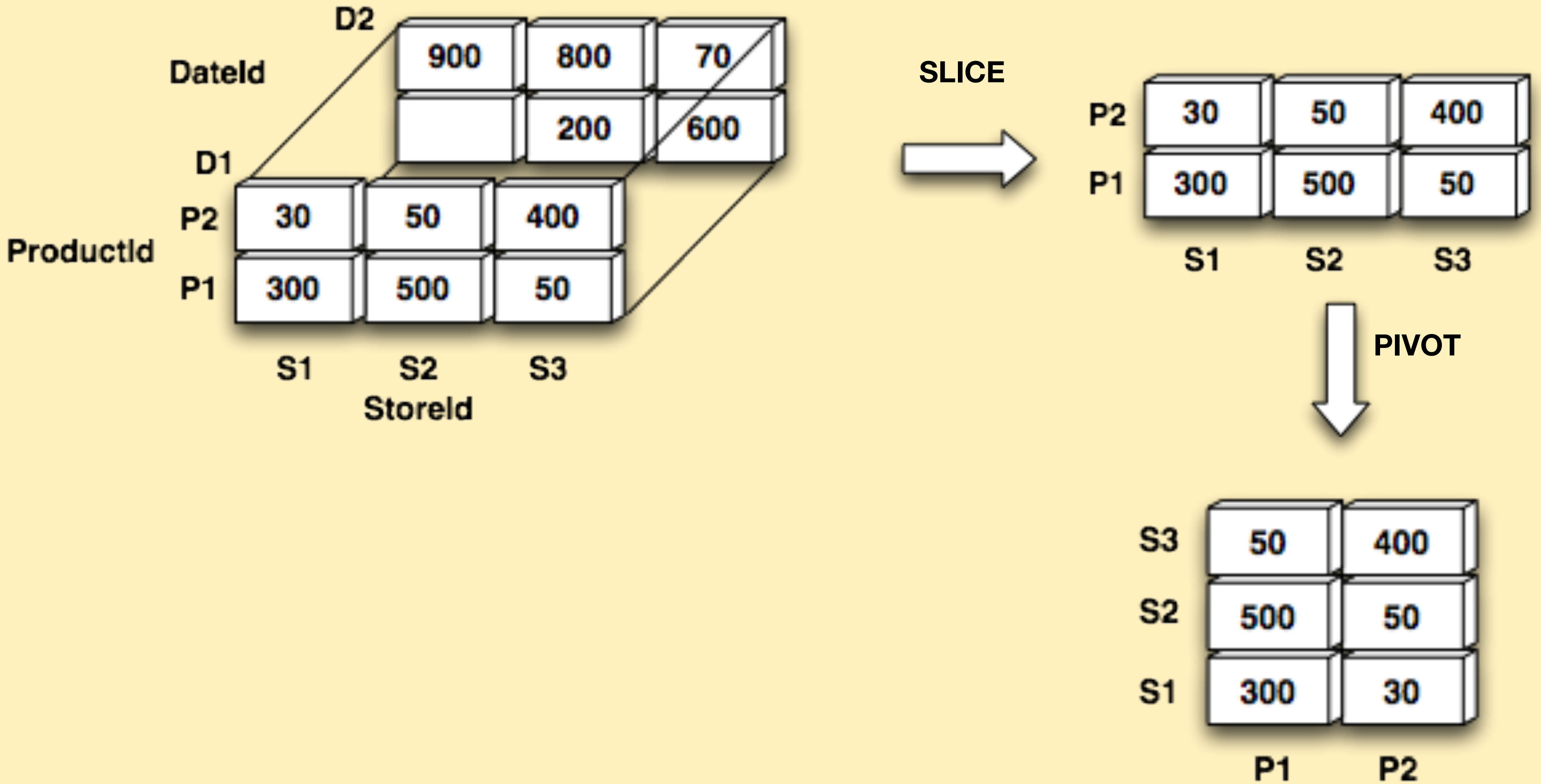
CUBE OPERATOR: DICE

Sales **DICE FOR** DateId = 'D1'
StoreId **IN** ('S1', 'S2');



CUBE OPERATOR: PIVOT

PIVOT (Sales **SLICE FOR** DateId = 'D1');



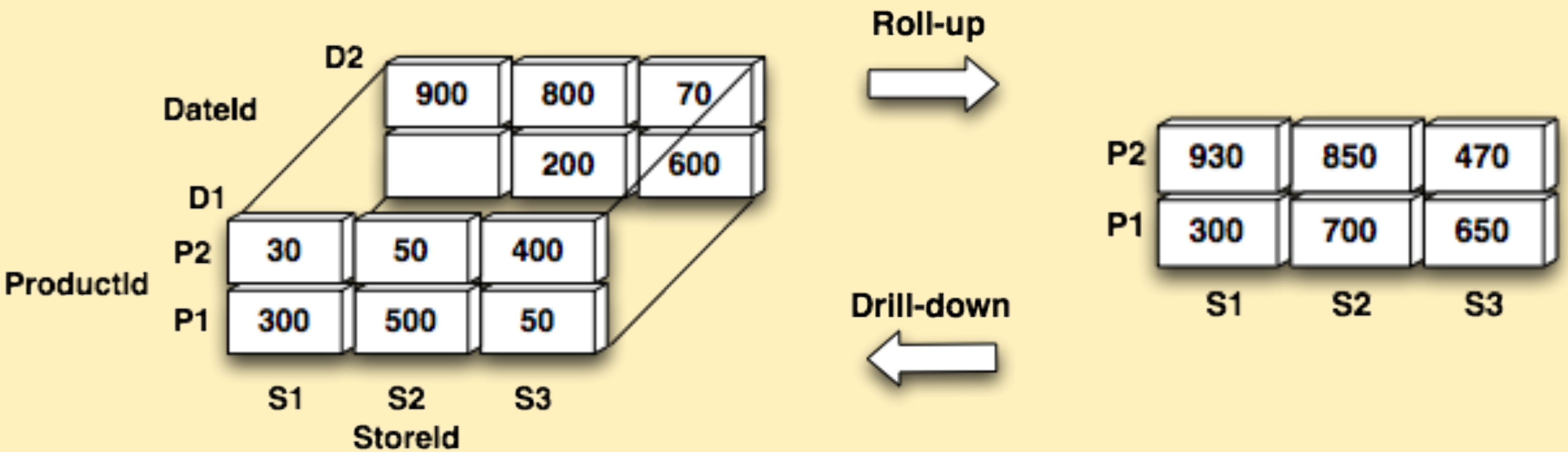
Rotate: reorient the cube, visualization, 3D to series of 2D planes

Roll-up aggregates data by **dimension reduction** or by navigating attribute hierarchy (**Drill-down** is the reverse of roll-up)

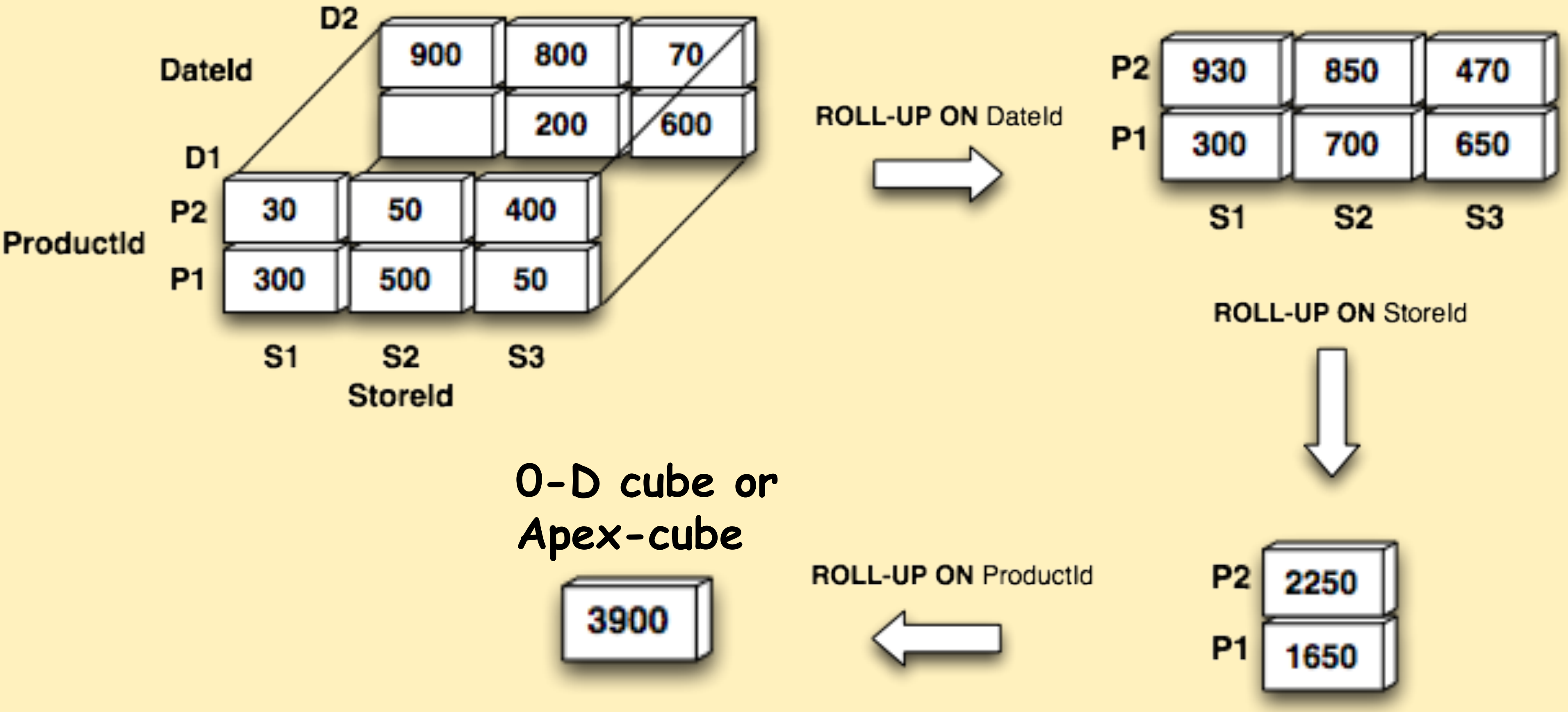
Hypothesis: one measure and aggregations by **sum**.

SALES **ROLL-UP ON** DateId

(total Qty by **ProductId** and by **StoreId**)



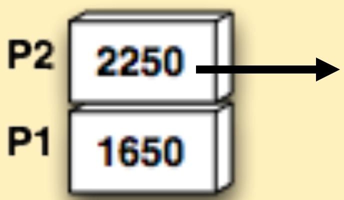
CUBE OPERATORS: ROLL-UP and DRILL-DOWN



Drill-through produces the facts that satisfy a cell coordinate

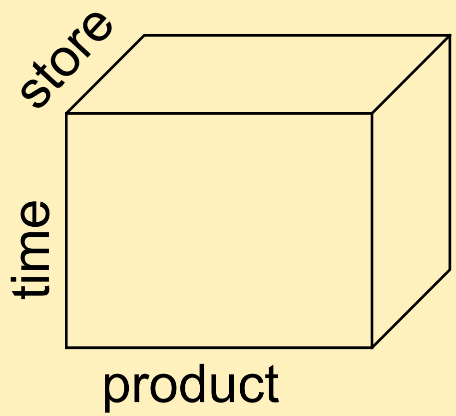
Sales

StoreId	ProductId	DateId	Qty
S1	P1	D1	300
S2	P1	D1	500
S3	P1	D1	50
S1	P2	D1	30
S2	P2	D1	50
S3	P2	D1	400
S2	P1	D2	200
S3	P1	D2	600
S1	P2	D2	900
S2	P2	D2	800
S3	P2	D2	70

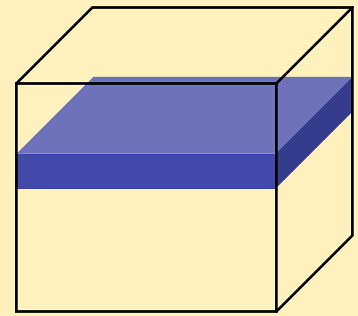


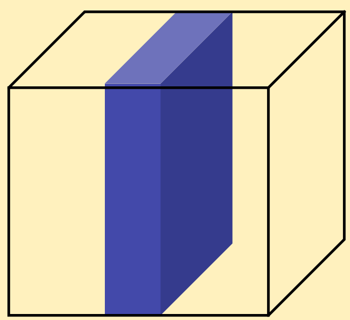
StoreId	ProductId	DateId	Qty
S1	P2	D1	30
S2	P2	D1	50
S3	P2	D1	400
S1	P2	D2	900
S2	P2	D2	800
S3	P2	D2	70

CUBE NAVIGATION BY DIFFERENT USERS



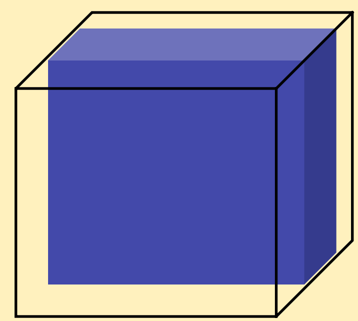
Finance manager look at sales
of a period compared to the previous period
for any product and any market





Product managers look at sales
of some products
in any period and in any market

Branch manager look at sales
of his/her stores
for any product and any period



TEXTUAL NOTATION FOR CUBE OPERATORS

Hypothesis: one measure and aggregations by **sum**.

Sales(StoreId, ProductId, DateId)

is the cube with dimensions **StoreId**, **ProdottoId**, **DataId**, and measure M

A cube operation is denoted by substituting a dimension with a value

TEXTUAL NOTATION FOR CUBE OPERATORS (cont)

Sales(StoreId, ProductId, ‘D1’) **slice**

Sales(‘S1’, ProductId, ‘D1’) **dice**

Sales(‘S1’, ‘P1’, ‘D1’) **dice**

TEXTUAL NOTATION FOR CUBE OPERATORS (cont.)

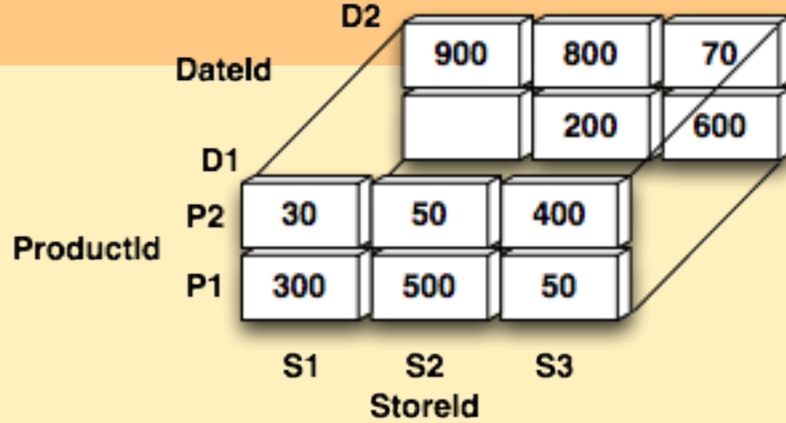
Each dimension domain is extended with the value "*", that means summarize data (**sum**) by all the dimension values.

Sales(StoreId, ProductId, *)

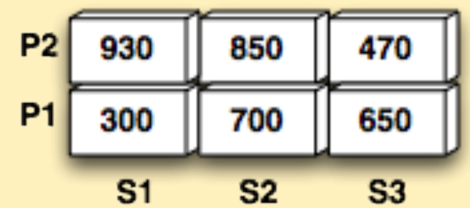
Sales by **roll-up on** DateId **with sum**(M)

CUBE OPERATORS: EXAMPLES

Sales(StoreId, ProductId, DateId) =

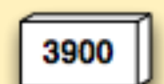


Sales(StoreId, ProductId, *) =



Sales(StoreId, *, *) =

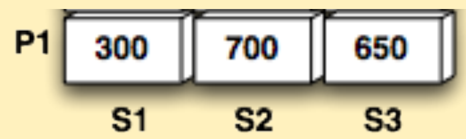
Sales(*, *, *) =



CUBE OPERATORS: EXAMPLES

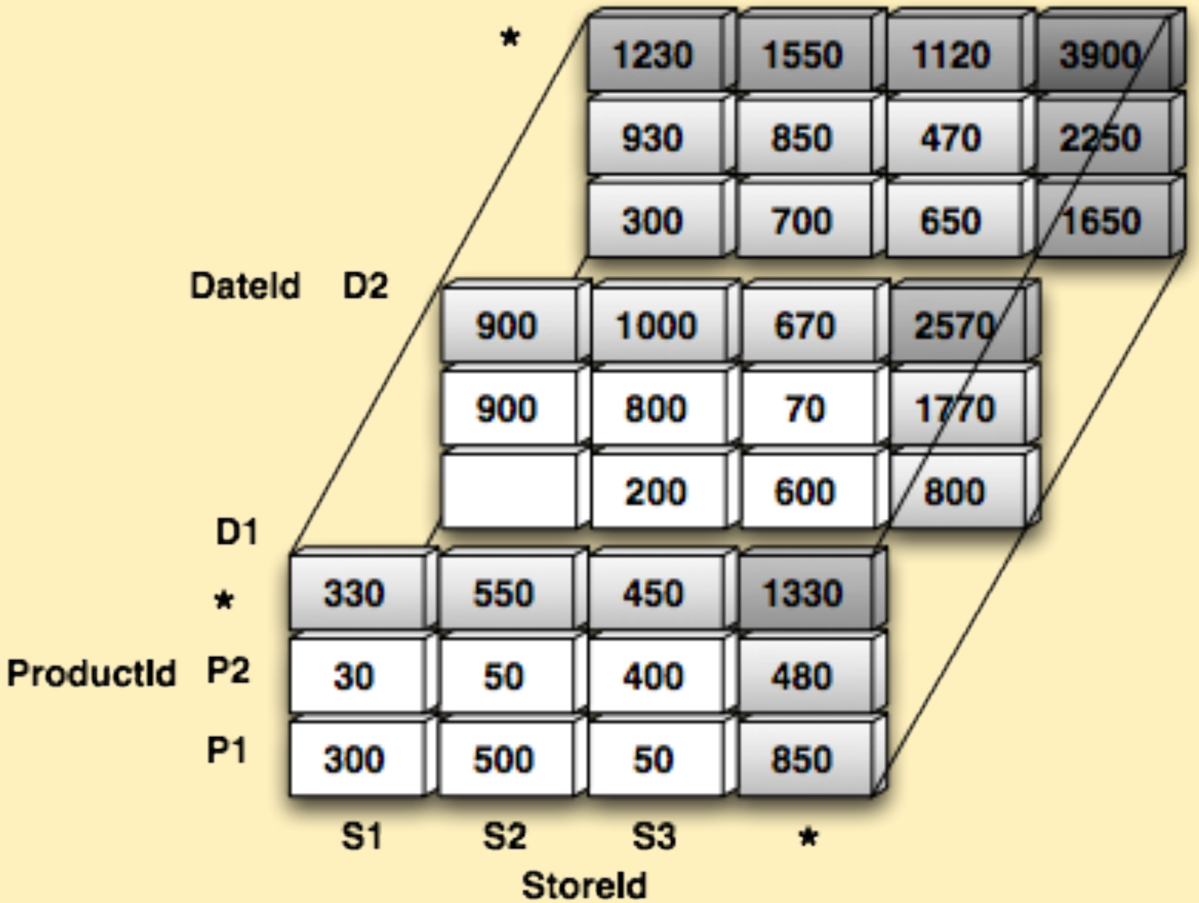
- What is

Sales(StoreId, 'P1', *) =



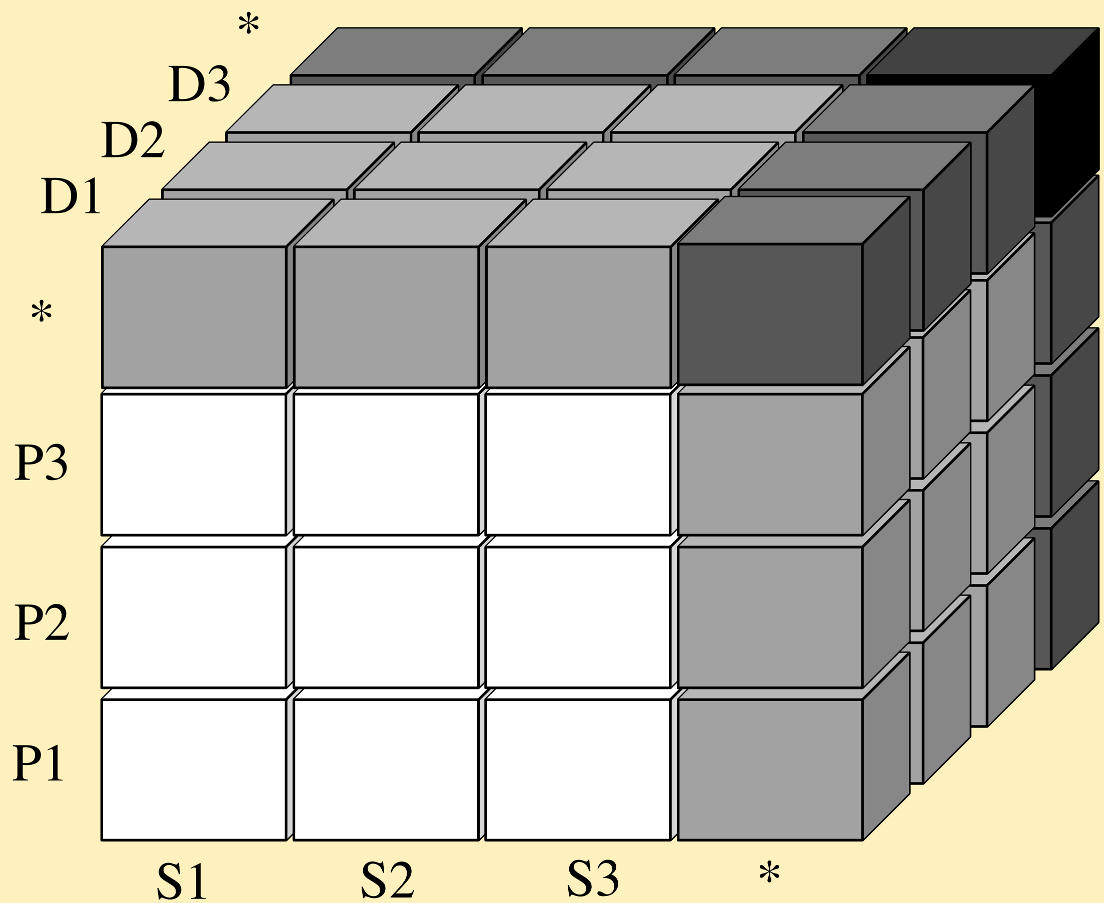
EXTENDED CUBE

A data cube is extended with the value '*' for each dimensions, and in the corresponding cells is stored the **sum** of the **measure**.



EXTENDED CUBE

With the '*' values, the cube becames a set of **cuboids**:



- **white** cells are the data cube
- **gray** cells are roll-up by a dimension,
- **dark gray** cells are roll-up by two dimensions
- **black** cells are roll-up by all dimensions.

EXTENDED CROSS TABULATION

Sales

StoreId	ProductId	Qty
S1	P1	300
S2	P1	500
S3	P1	50
S1	P2	30
S2	P2	50
S3	P2	400

CROSS TABULATION

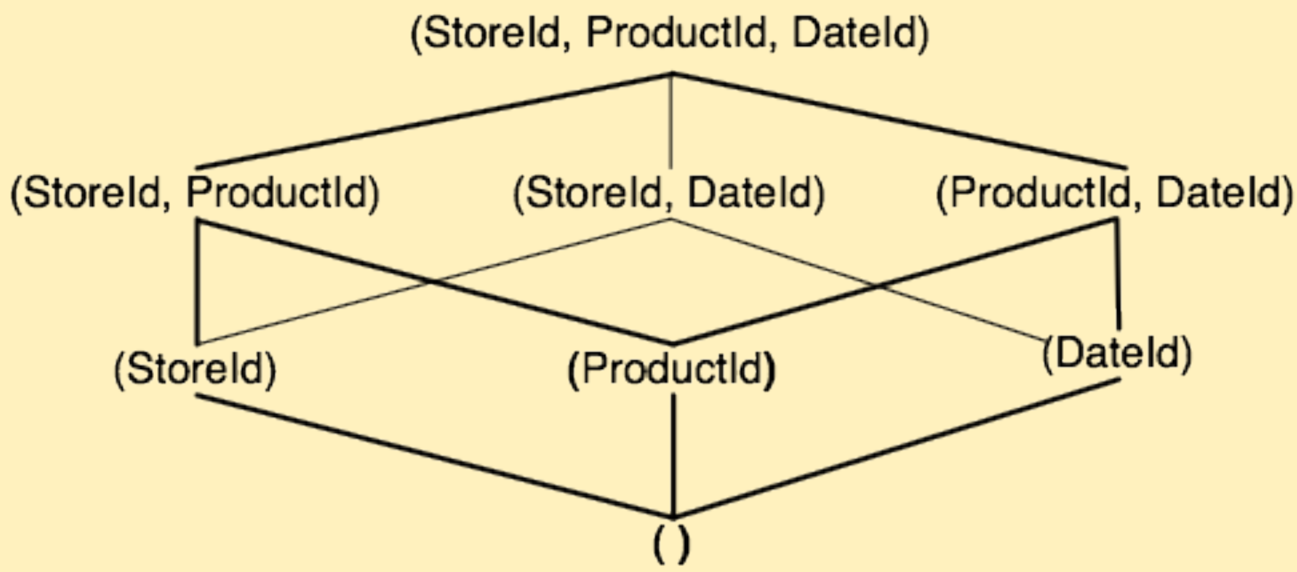
	StoreId		
ProductId	S1	S2	S3
P1	300	500	50
P2	30	50	400

EXTENDED CROSS TABULATION

	StoreId			
ProductId	S1	S2	S3	Total
P1	300	500	50	**850**
P2	30	50	400	**480**
Total	**330**	**550**	**450**	**1330**

On the set of cuboids is defined the following partial order relation:

C1 ≤ C2 if C1 dimensions are included in C2 dimensions.



HOW MANY CUBOIDS? HOW MANY CELLS?

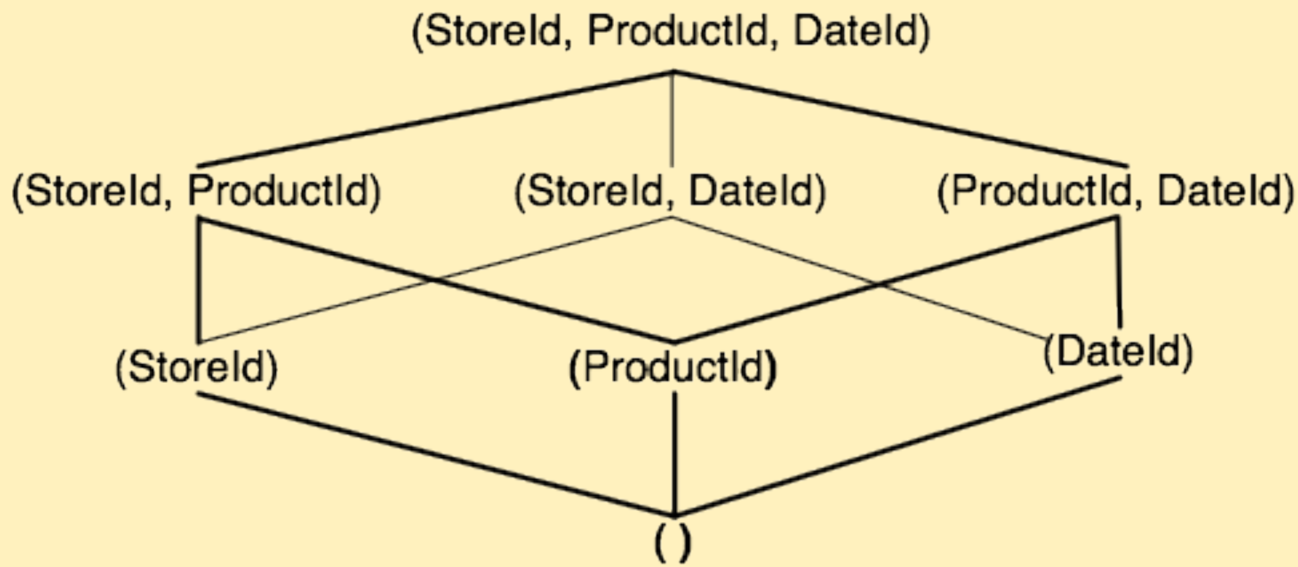
- $D = \{d_1, \ldots, d_N\}$ dimensions (degenerate or flat)
  - 2^N cuboids
- Let $\#d_i$ = number of values for dimension d_i
- How many cells in total?

$$\sum_{C \subseteq D} \prod_{d \in C} \#d = \prod_{i=1..N} (\#d_i + 1)$$

CUBOIDS MATERIALIZATION

Complete

Partial



AGGREGATION FUNCTIONS TYPES

$$V = V_1 \cup V_2 \qquad\qquad V_1 \cap V_2 = \varnothing$$

Distributive

E.g., sum(), min(), max(), count()

sum(V) = sum(V_1) + sum(V_2) sum({v}) = v

count(V) = count(V_1) + count(V_2) count({v}) = 1

Algebraic

E.g., avg(), standard_deviation()

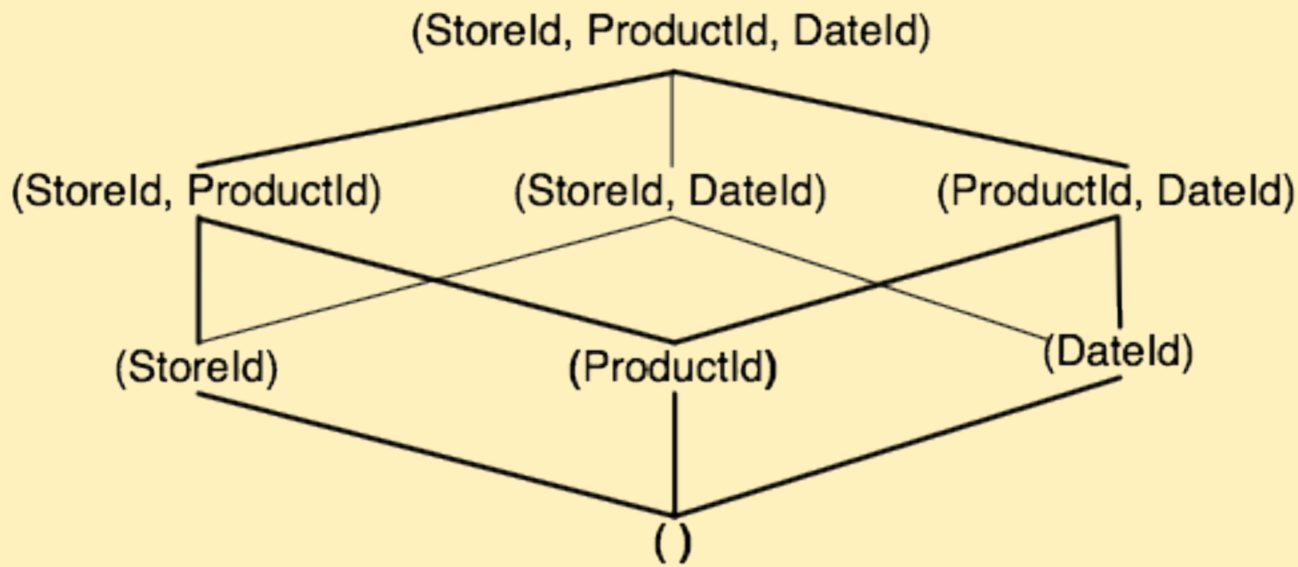
avg(V) = sum(V)/count(V)

$$\text{var}(V) = \frac{\text{sum}(V^2) - \text{sum}(V)^2 / \text{count}(V)}{\text{count}(V)-1} \qquad \text{sum}(\{v\}^2) = v*v$$

Holistic

E.g., median(), mode(), rank().

Complete



Partial

If the materialization is partial, which cuboids do we select?

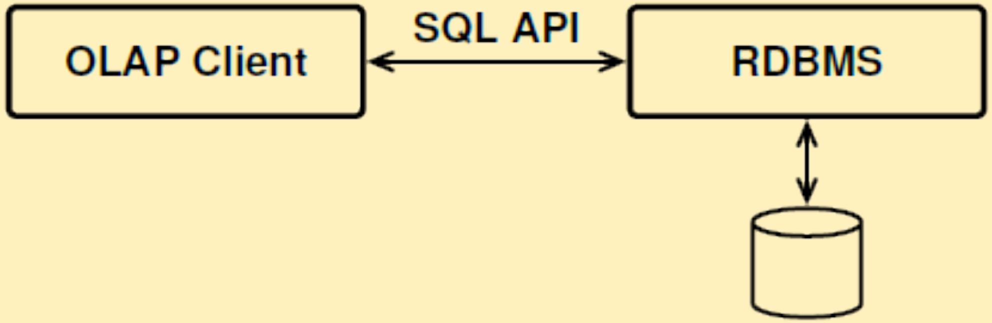
OLAP SYSTEMS SOLUTIONS

OLAP refers to the technique of performing complex business analysis over the information stored in a data warehouse.

We will see how report developers use SQL to write queries, but there are business intelligence tools that allows a user or a developer to make data analysis and to build beautiful reports without any knowledge of SQL...

which is generated automatically.

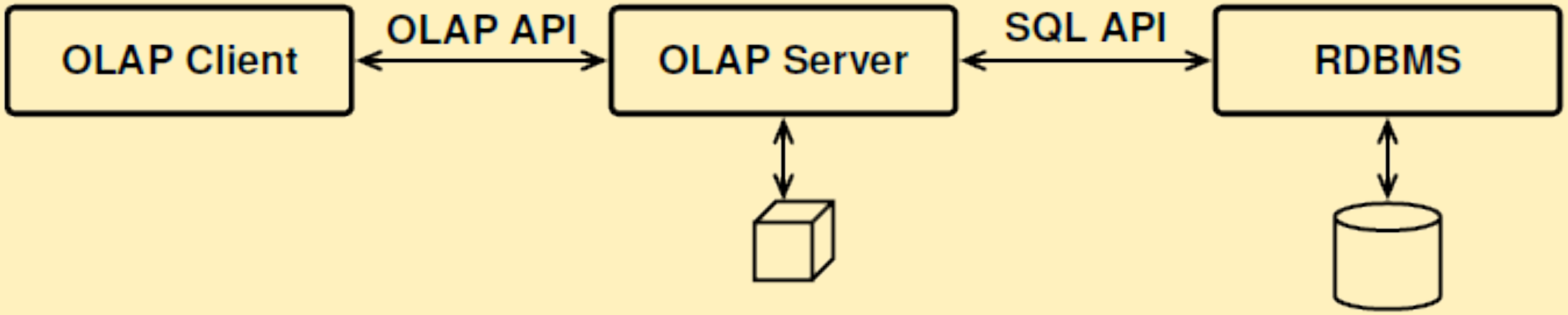
OLAP SYSTEMS: SOLUTION 1



The DW is managed by a specialized RDBMS (Relational **Data Server**)

The **OLAP Client** provides presentation and reporting tools to deal with data analysis and visualization, and interacts with the **Data Server**.

OLAP SYSTEMS: SOLUTION 2



The **OLAP Client** interacts with an **OLAP Server**, that supports multidimensional data and operations, and can be one of the following type:

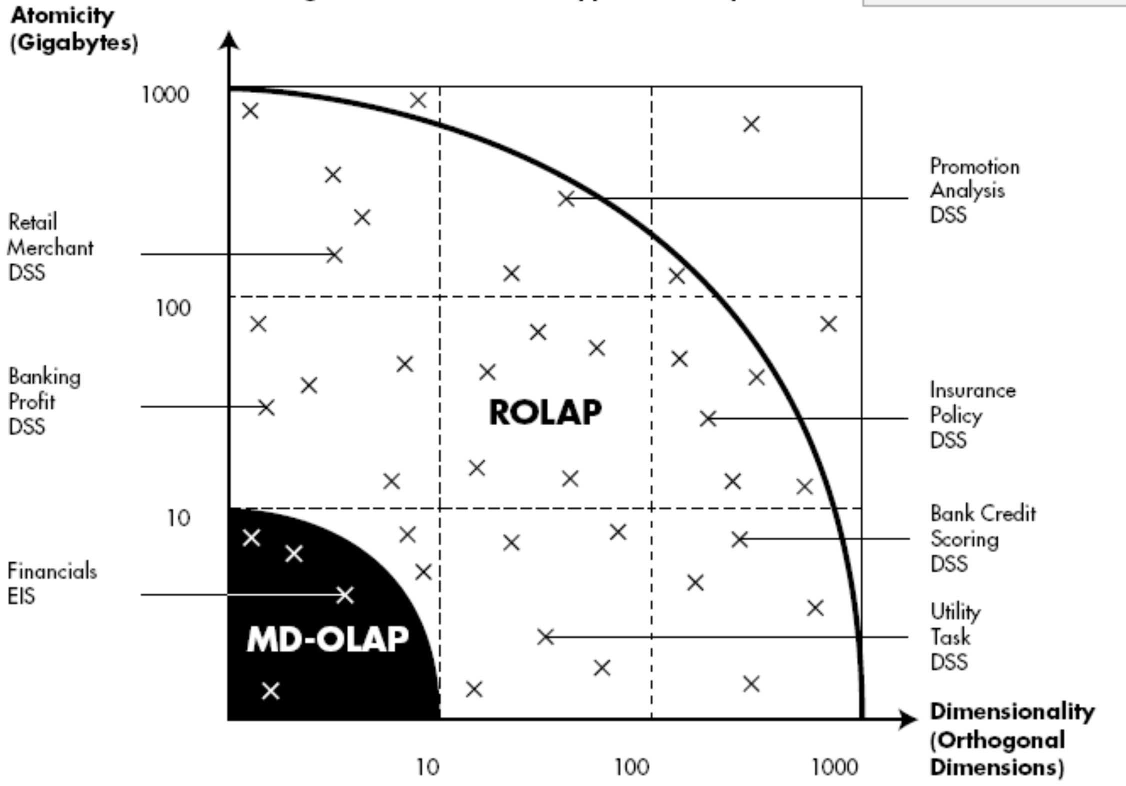
- **MOLAP**, which stores in the local memory both the data cube, taken from the Data Server, and the aggregates of the extended cube, using a specialized data structure. **A MOLAP server does not support SQL, but MDX.**
- **ROLAP** which stores both the data and the aggregates of the extended cube in the Data Server. ROLAP servers may also implement functionalities not supported in the SQL of the Data server.
- **HOLAP** which stores the data in the Data Server, and the aggregates of the extended cube in the local memory.

The ROLAP case

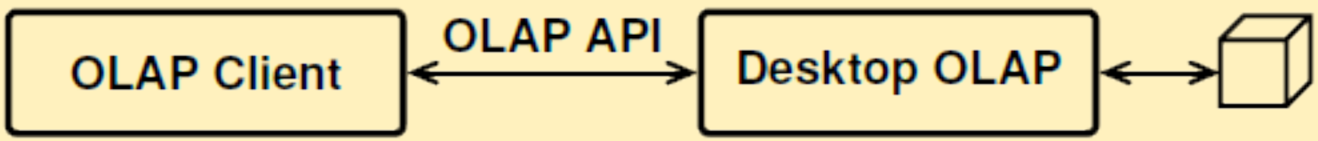
Data storage modes

OLAP Server	MOLAP	ROLAP	HOLAP	Offline
Essbase	Yes	Yes	Yes	
icCube	Yes	No	No	Offline Cubes
Microsoft Analysis Services	Yes	Yes	Yes	Local cubes, PowerPivot for Excel
MicroStrategy Intelligence Server	Yes	Yes	Yes	MicroStrategy Office, Dynamic Dashboards
Mondrian OLAP server	No	Yes	No	
Oracle Database OLAP Option	Yes	Yes	Yes	
Palo	Yes	No	No	
SAS OLAP Server	Yes	Yes	Yes	
TM1	Yes	No	No	
SAP NetWeaver BW	Yes	Yes	No	

Figure 20: ROLAP Meets Application Requirements



OLAP SYSTEMS: SOLUTION 3

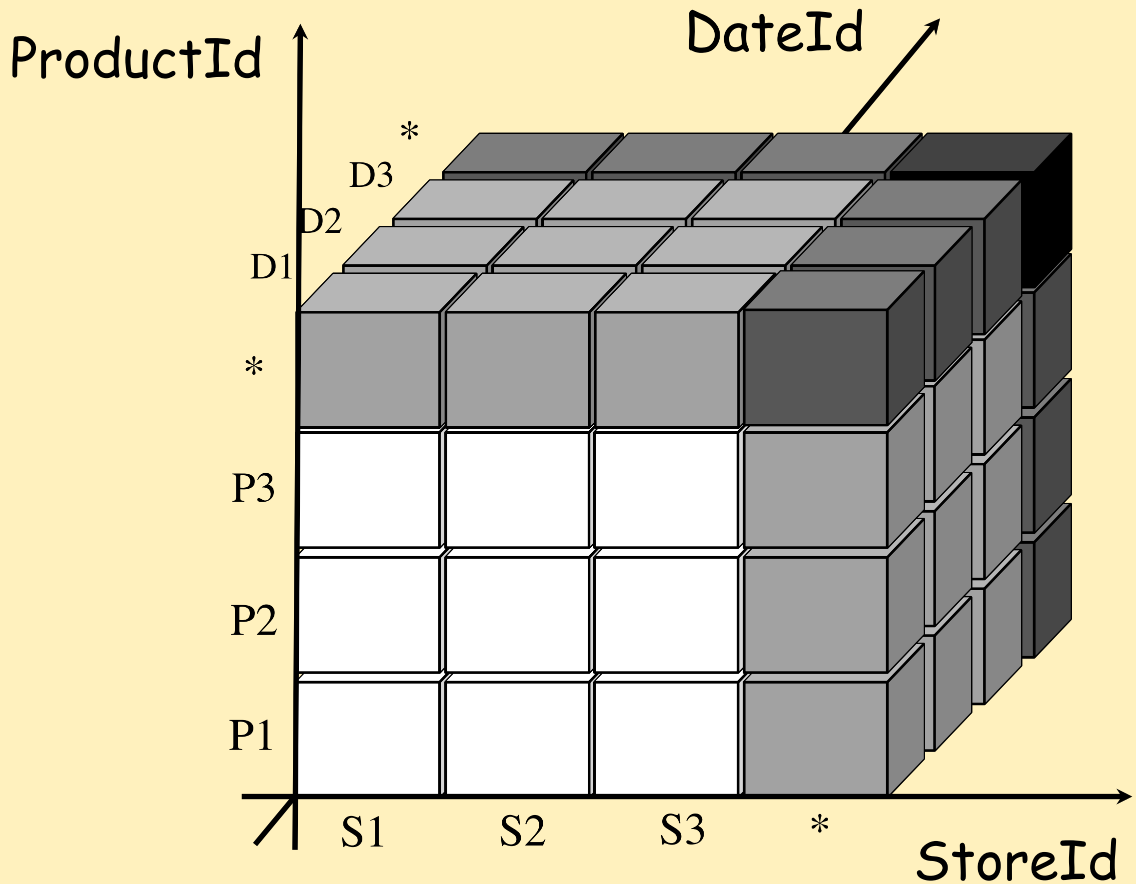


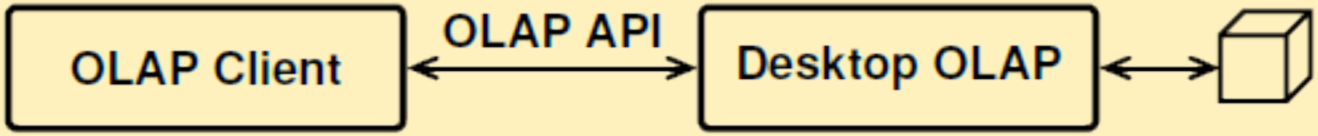
The **OLAP client** interacts with a local **DOLAP** system (**Desktop OLAP**) which manages small amount of data extracted from the **OLAP server**, the **Data server** or an **operational DBMS**. It a good choice for those who travel and move extensively, by using portable computers.

E.g., Microsoft Power Pivot (Add-in of Excel)

MULTIDIMENSIONAL MODEL (CUBE)

The multidimensional model is useful to understand interactive data analysis, and how to improve the execution performance.





The **OLAP client** interacts with a local **DOLAP** system (**Desktop OLAP**) which manages small amount of data extracted from the **OLAP server**, the **Data server** or an **operational DBMS**. It a good choice for those who travel and move extensively, by using portable computers.

DEMO WITH Microsoft Power Pivot (Add-in of Excel)

Power BI - Overview and Learning

Microsoft Power BI is a collection of online services and features that enables you to find and visualize data, share discoveries, and collaborate in intuitive new ways. There are two experiences now available for Power BI: the *new experience*, generally referred to as **Power BI**, and the *previous experience* which is referred to as **Power BI for Office 365**.

OLAP SYSTEMS