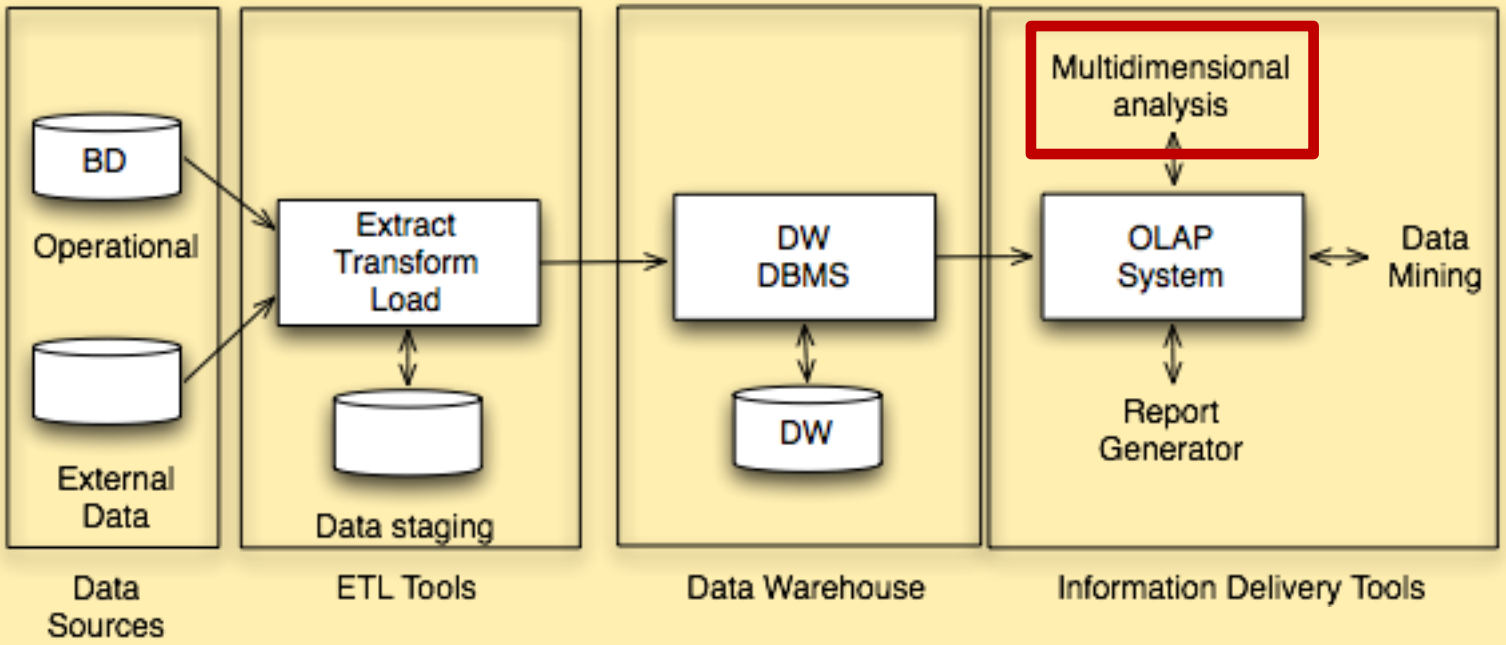
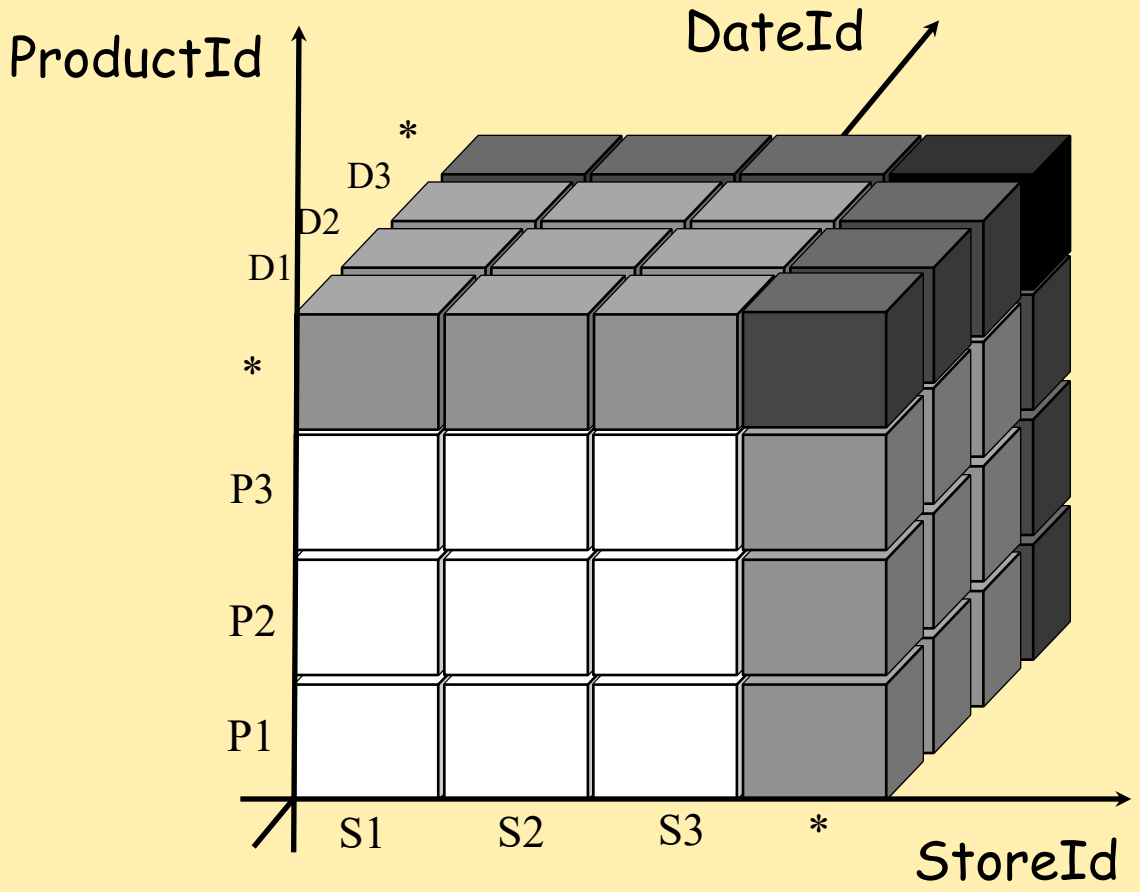
MULTIDIMENSIONAL MODEL

- Today, we introduce a logical model for the analysis of aggregated data:
  - **Multidimensional model**



MULTIDIMENSIONAL MODEL (CUBE)

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



2-D CUBE: A matrix

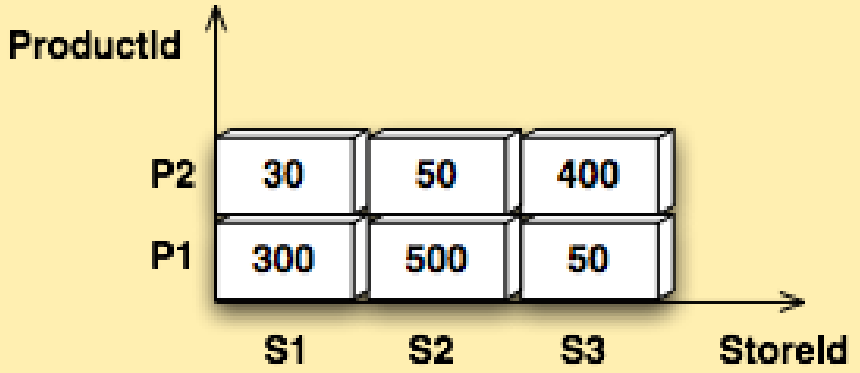
Hypothesis: one measure and aggregations by **sum**.

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

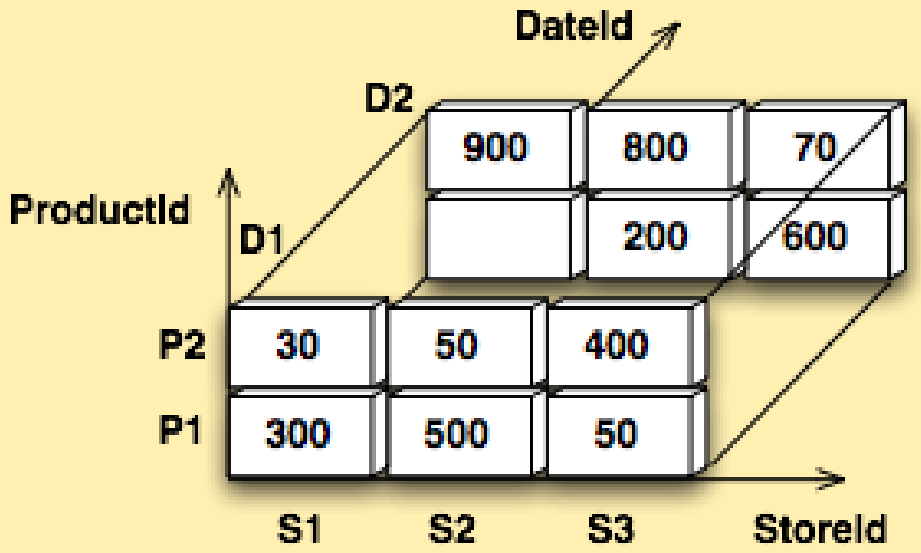
	StoreId		
ProductId	S1	S2	S3
P1	300	500	50
P2	30	50	400

3-D CUBE: A 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

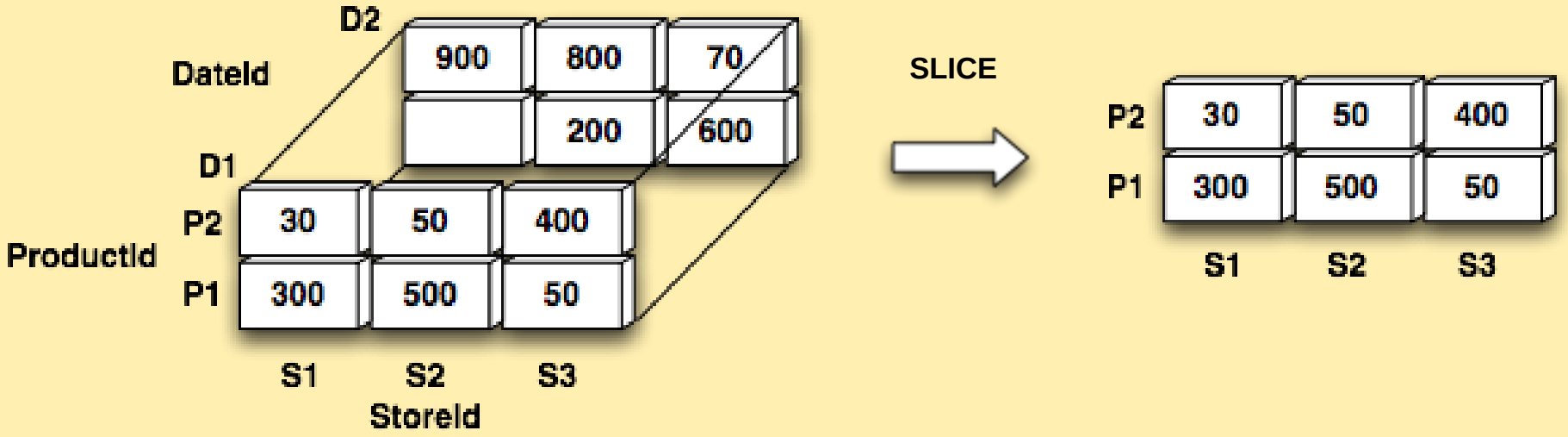
Fact Table



3-D Cube

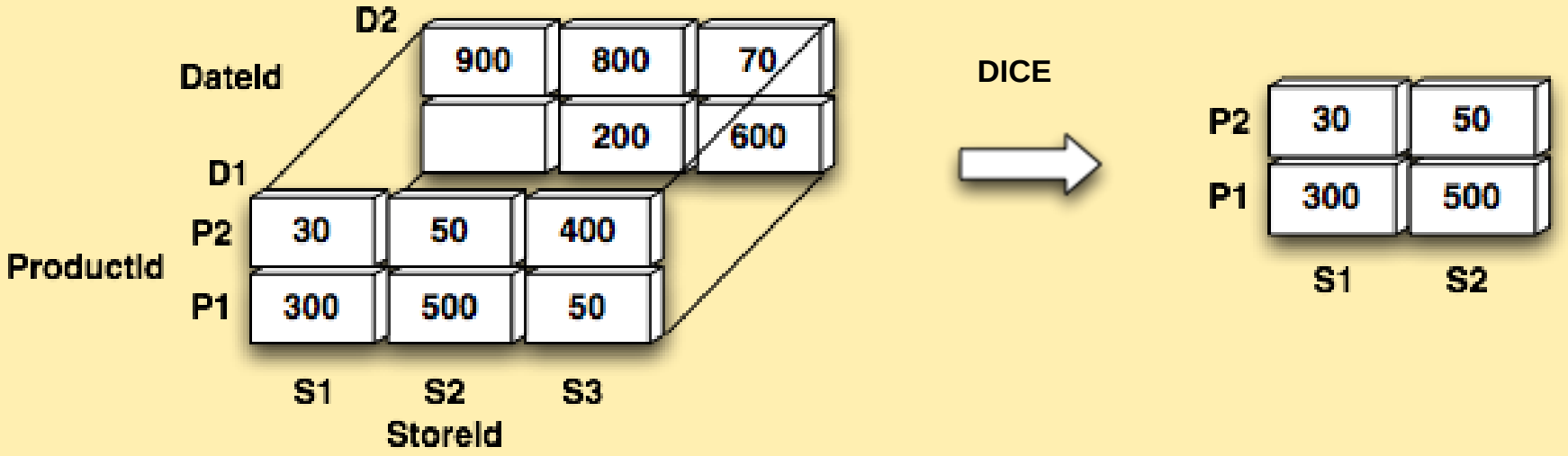
CUBE OPERATOR: SLICE

Sales **SLICE FOR** DateId = 'D1';



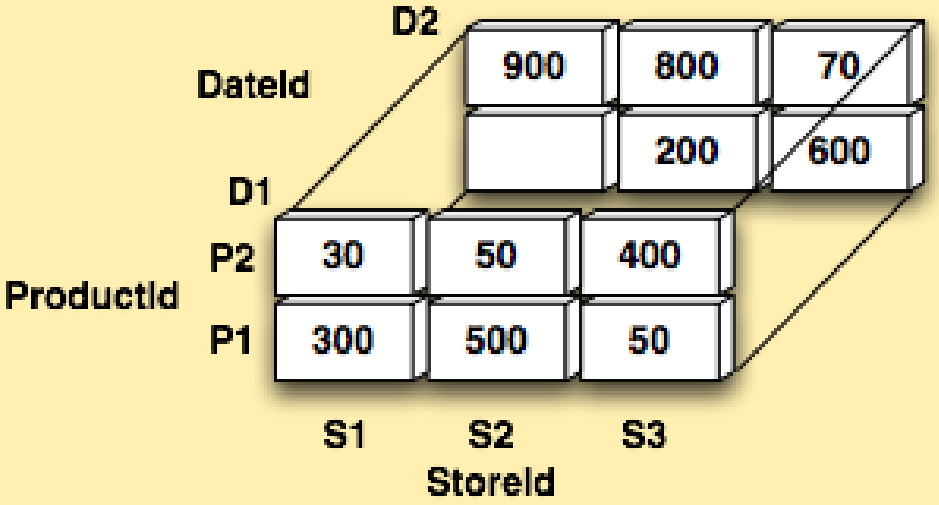
CUBE OPERATOR: DICE

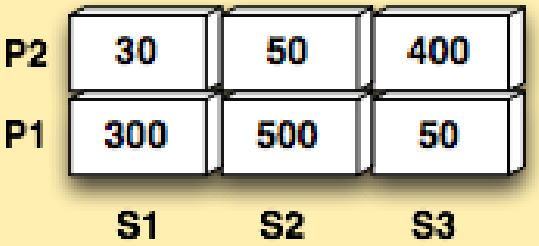
Sales **DICE FOR** DateId = 'D1'
StoreId **IN** ('S1', 'S2');

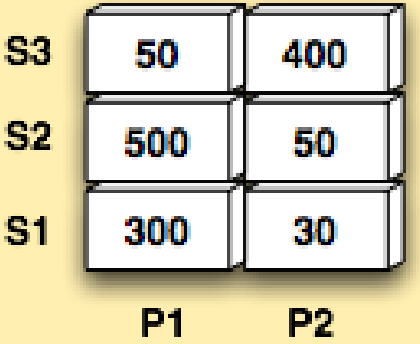


CUBE OPERATOR: PIVOT

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







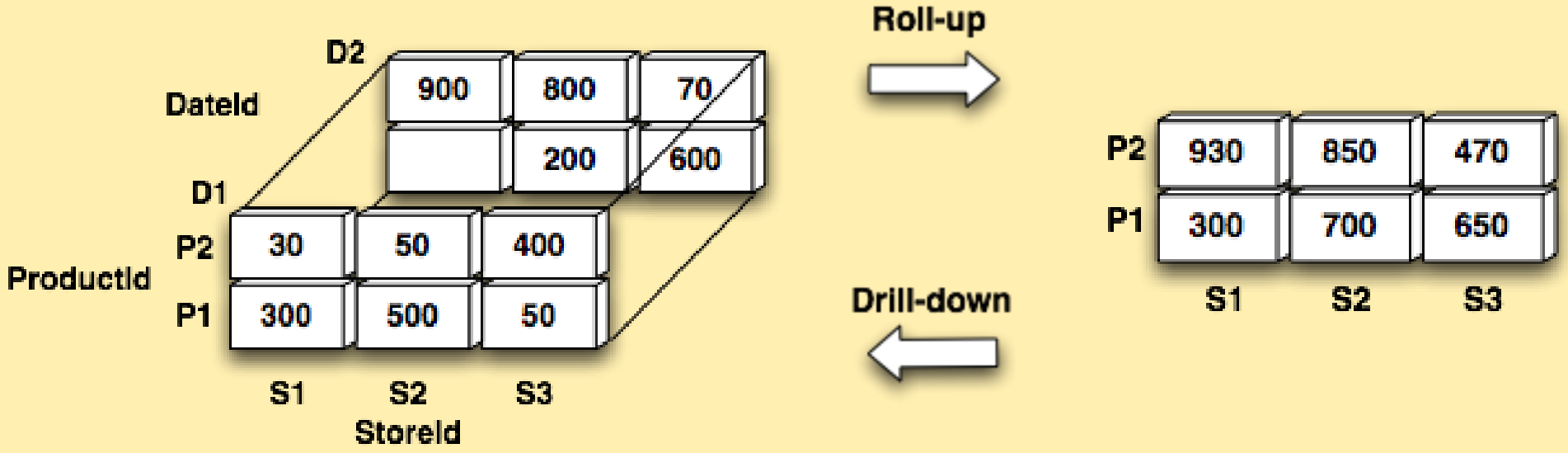
Rotate: reorient the cube, visualization, 3D to series of 2D planes

CUBE OPERATORS: ROLL-UP and DRILL-DOWN

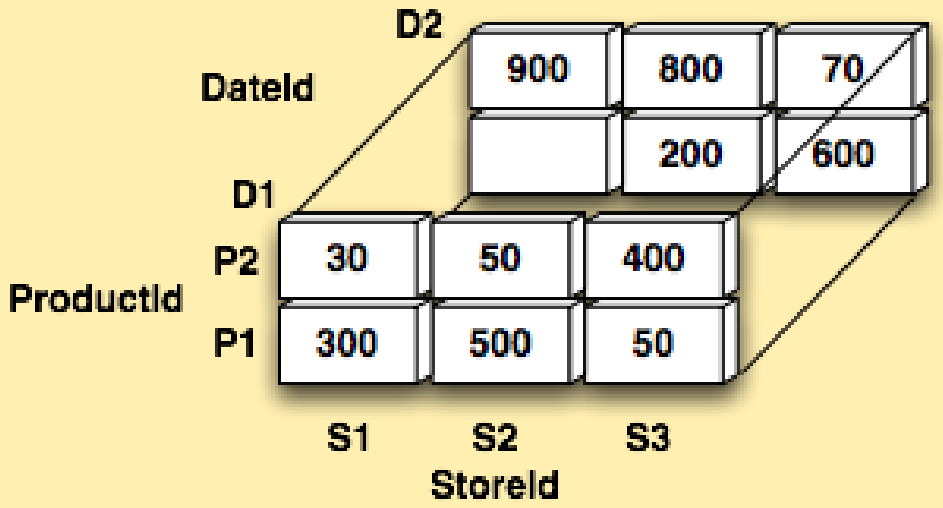
Roll-up aggregates data by **dimension reduction** or by going up in attribute hierarchy (**Drill-down** is the reverse of roll-up)

SALES **ROLL-UP ON** DateId

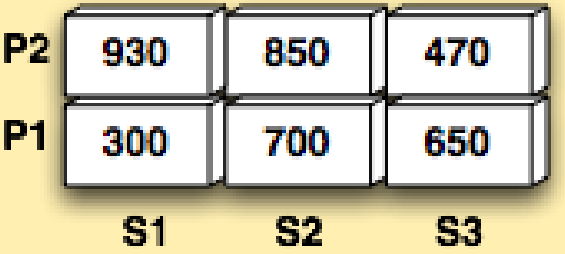
(total Qty by **ProductId** and by **StoreId**)



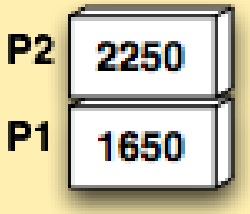
CUBE OPERATORS: ROLL-UP and DRILL-DOWN



ROLL-UP ON DateId



ROLL-UP ON StoreId



ROLL-UP ON ProductId

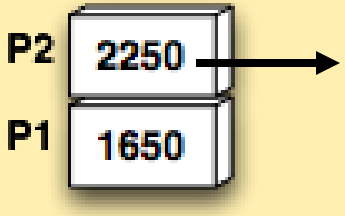
0-D cube or
Apex-cube

CUBE OPERATORS: DRILL THROUGH

Drill-through produces the facts that satisfy a cell coordinate

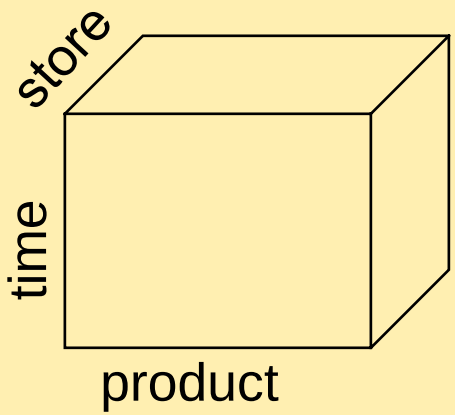
Sales

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

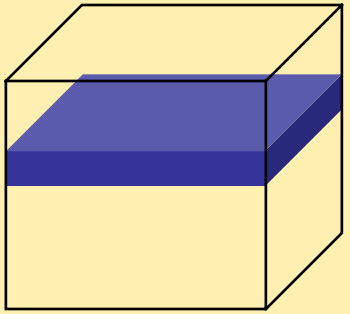


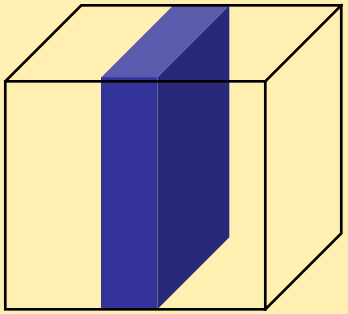
StoreId	ProductId	DateId	Qty
S1	P2	D1	30
S2	P2	D1	50
S3	P2	D1	400
S1	P2	D2	900
S2	P2	D2	800
S3	P2	D2	70

CUBE NAVIGATION BY DIFFERENT USERS

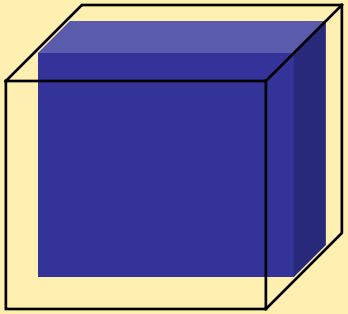


Finance manager look at sales
of a period compared to the previous period
for any product and any market



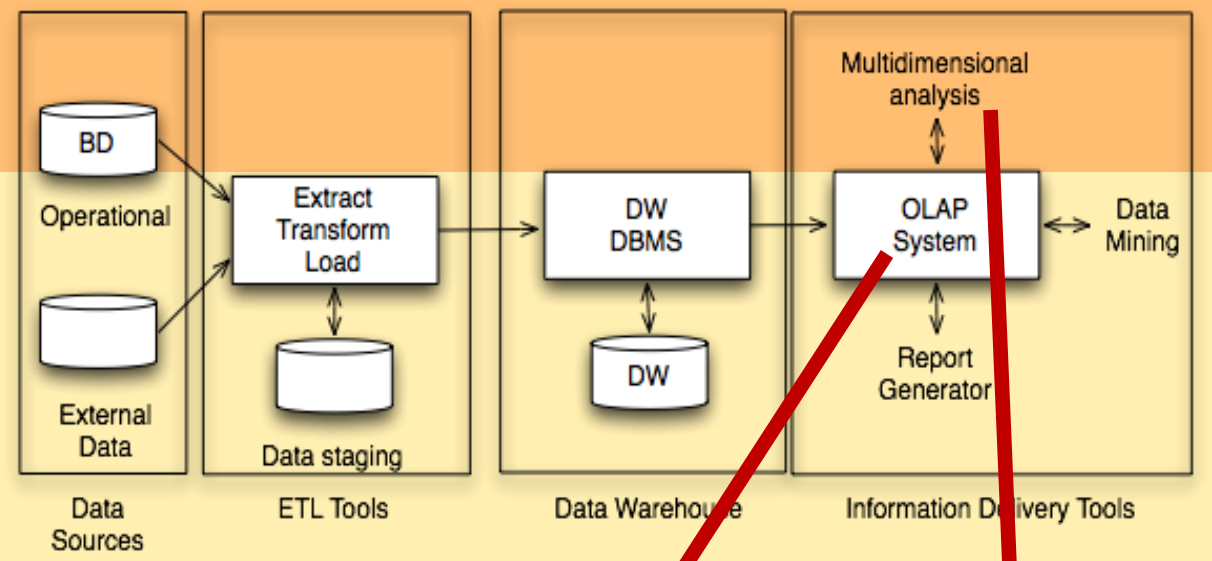


Branch manager look at sales
of his/her stores
for any product and any period



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

Demo



- Excel
  - HerbalTeas -> Pivot Table
  - DW -> Power Pivot or SSAS -> Pivot Table
- Pivot Tables & Charts
  - for multidimensional analysis

PowerPivot
SSAS
...

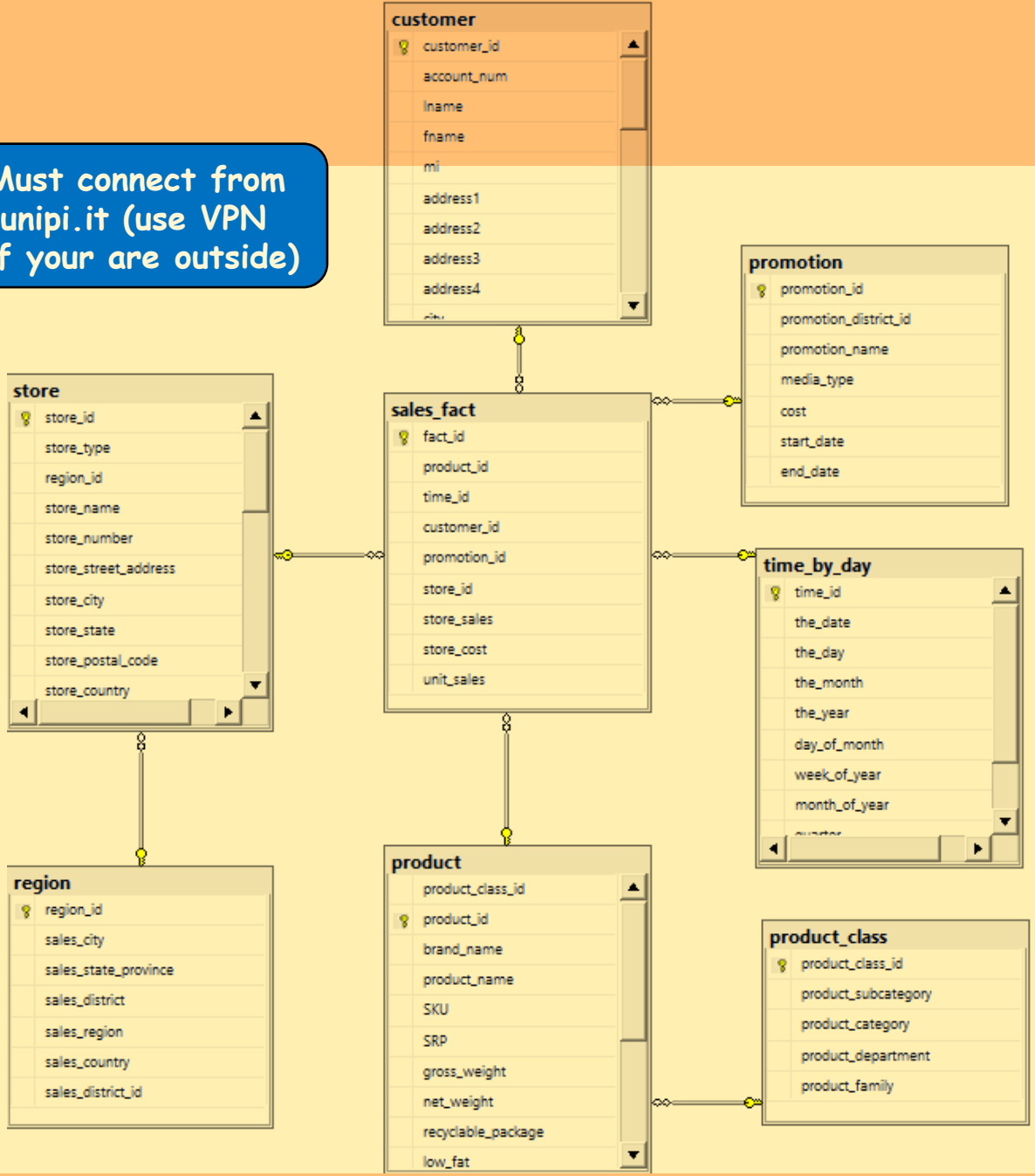
Excel
PowerBI
Microstrategy
...

More at the Lab of Data Science!

Foodmart datawarehouse

- RDBMS: Microsoft SQL Server
- **SQL Server: lds.di.unipi.it**
- **Login: dsd**

Must connect from .unipi.it (use VPN if your are outside)



TEXTUAL NOTATION FOR CUBE OPERATORS

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.)

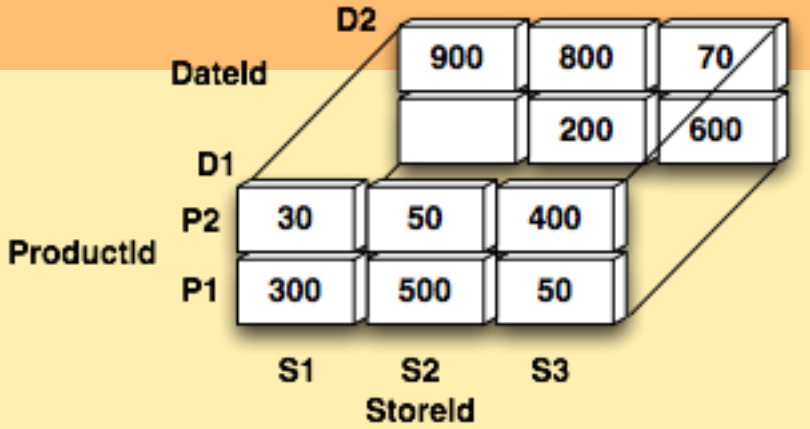
Each dimension domain is extended with the value "*", that means summarize data (**sum**) by all the dimension values.

Sales(StoreId, ProductId, *)

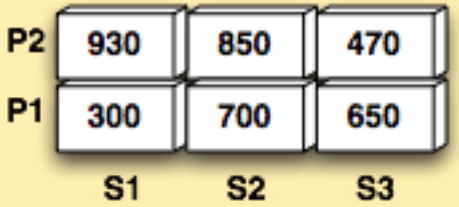
Sales by **roll-up on** DateId **with sum**(M)

CUBE OPERATORS: EXAMPLES

Sales(StoreId, ProductId, DateId) =



Sales(StoreId, ProductId, *) =



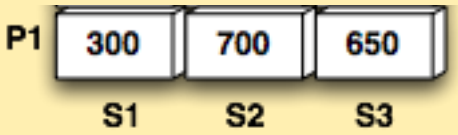
Sales(StoreId, *, *) =

Sales(*, *, *) =

CUBE OPERATORS: EXAMPLES

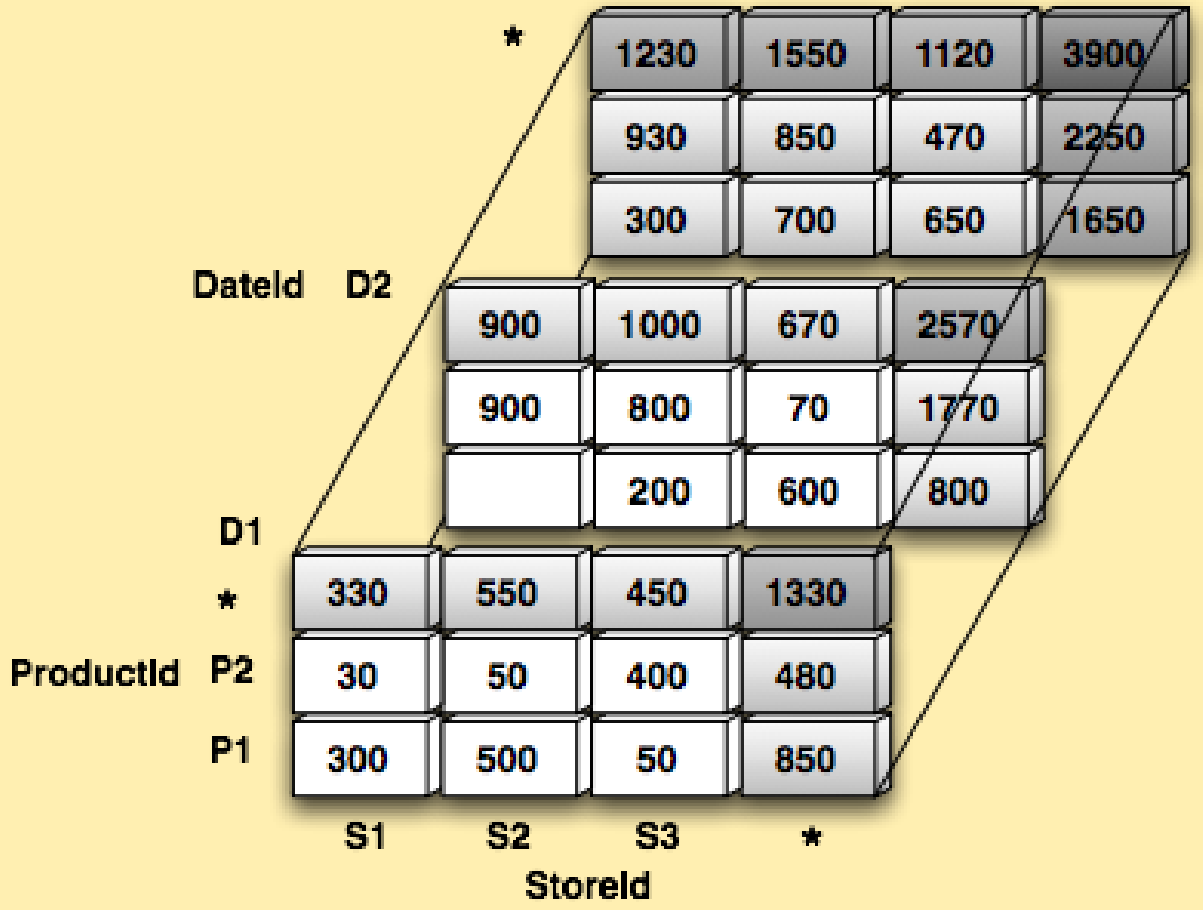
- What is

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 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

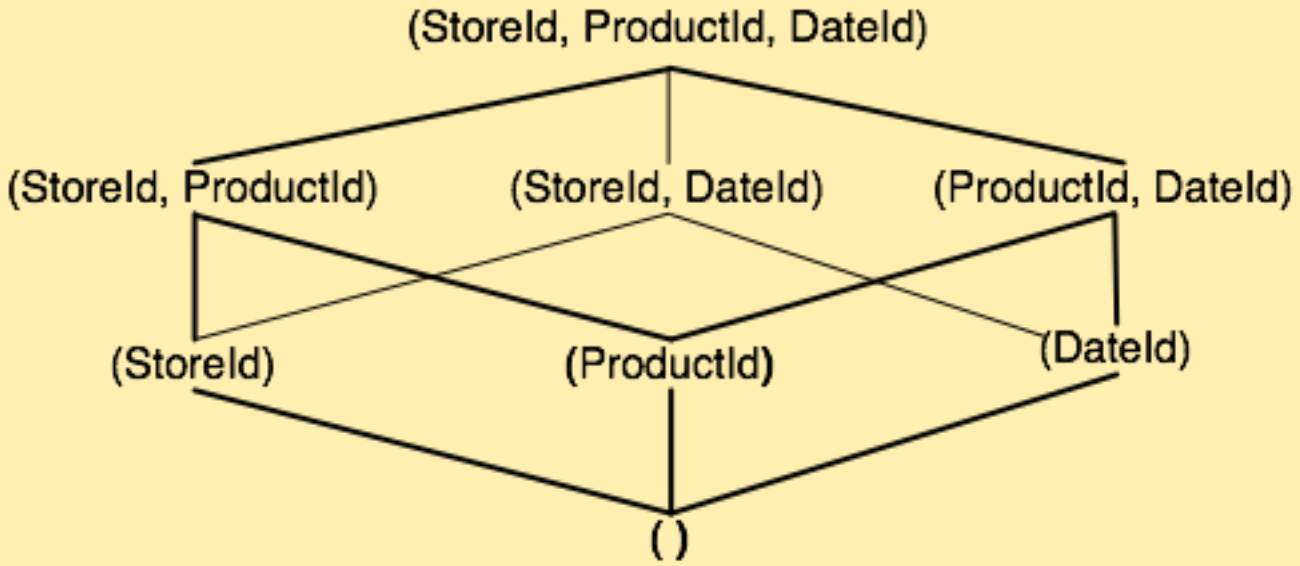
EXTENDED CROSS TABULATION

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

DW LATTICE: A LATTICE OF CUBOIDES

Cuboid = a cube on a subset of dimensions
On the set of cuboids is defined the following partial order relation (lattice):

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



HOW MANY CUBOIDS? HOW MANY CELLS?

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

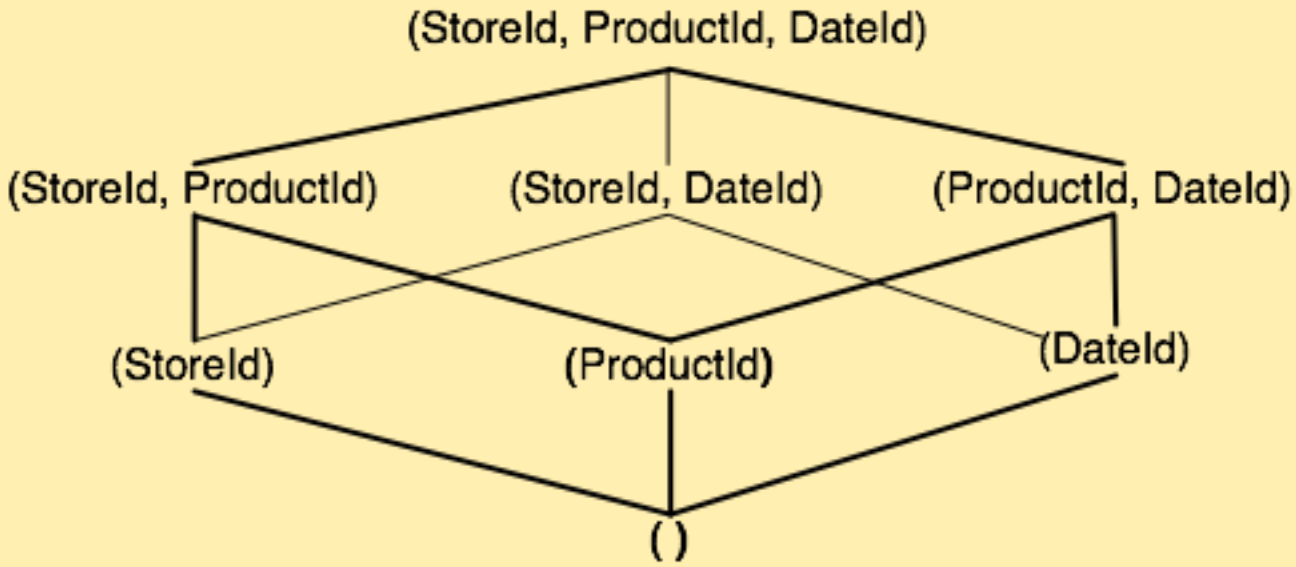
$$\sum_{C \subseteq D} \prod_{d \in C} \#\mathrm{d}$$

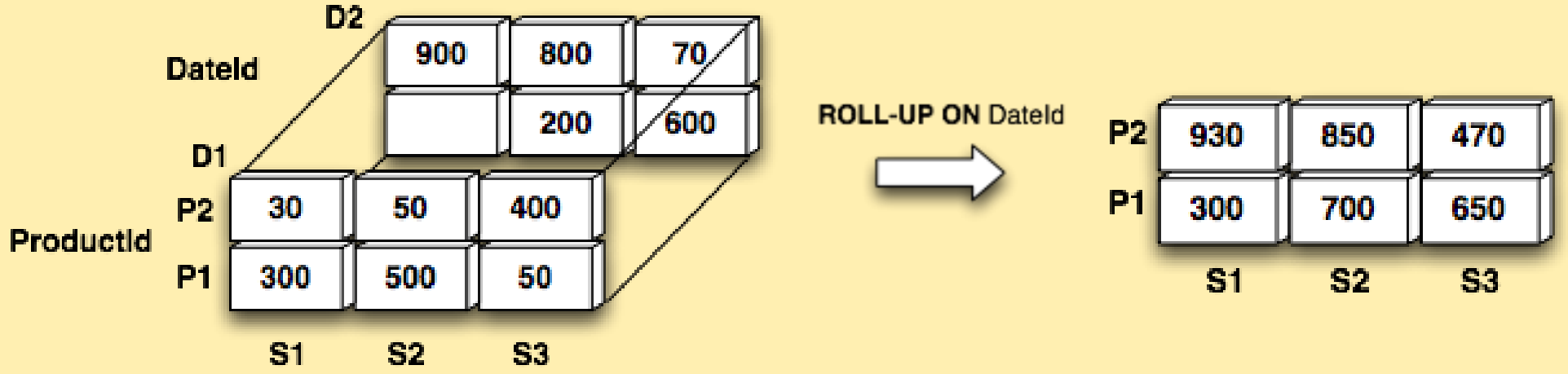
$$= \prod_{i=1..N} (\#d_i + 1)$$

CUBOIDS MATERIALIZATION

Complete

Partial





AGGREGATION FUNCTIONS TYPES

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

Distributive

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

$sum(V) = sum(V_1) + sum(V_2)$ $\qquad sum(\{v\}) = v$

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

Algebraic

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

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

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

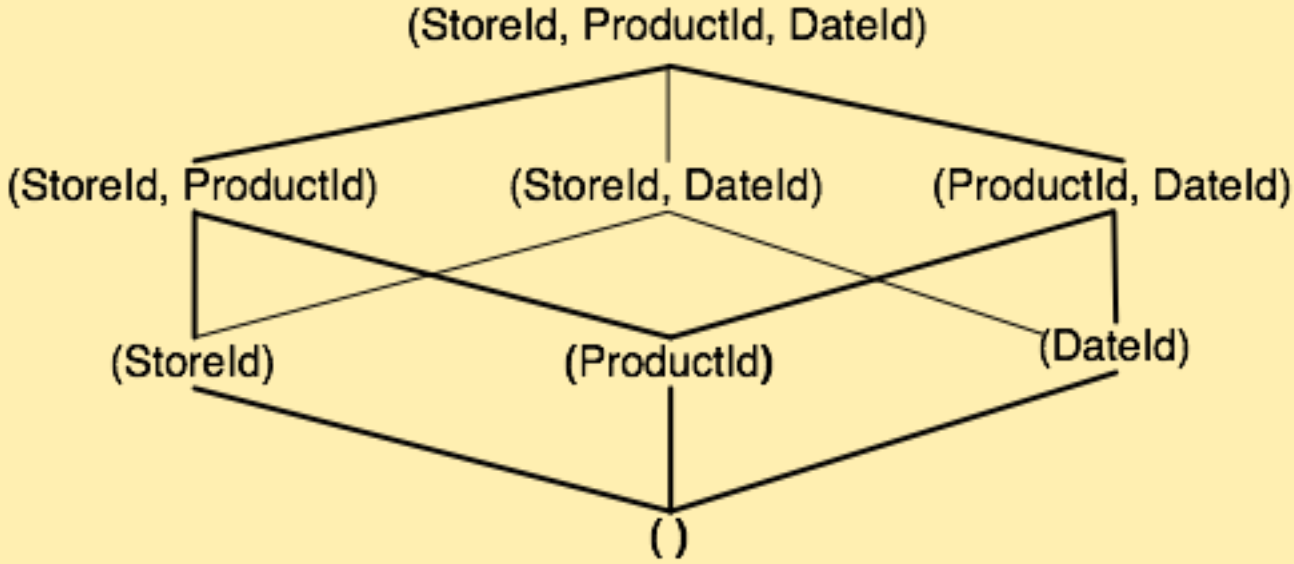
Holistic

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

CUBOIDS MATERIALIZATION

Complete

Partial



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