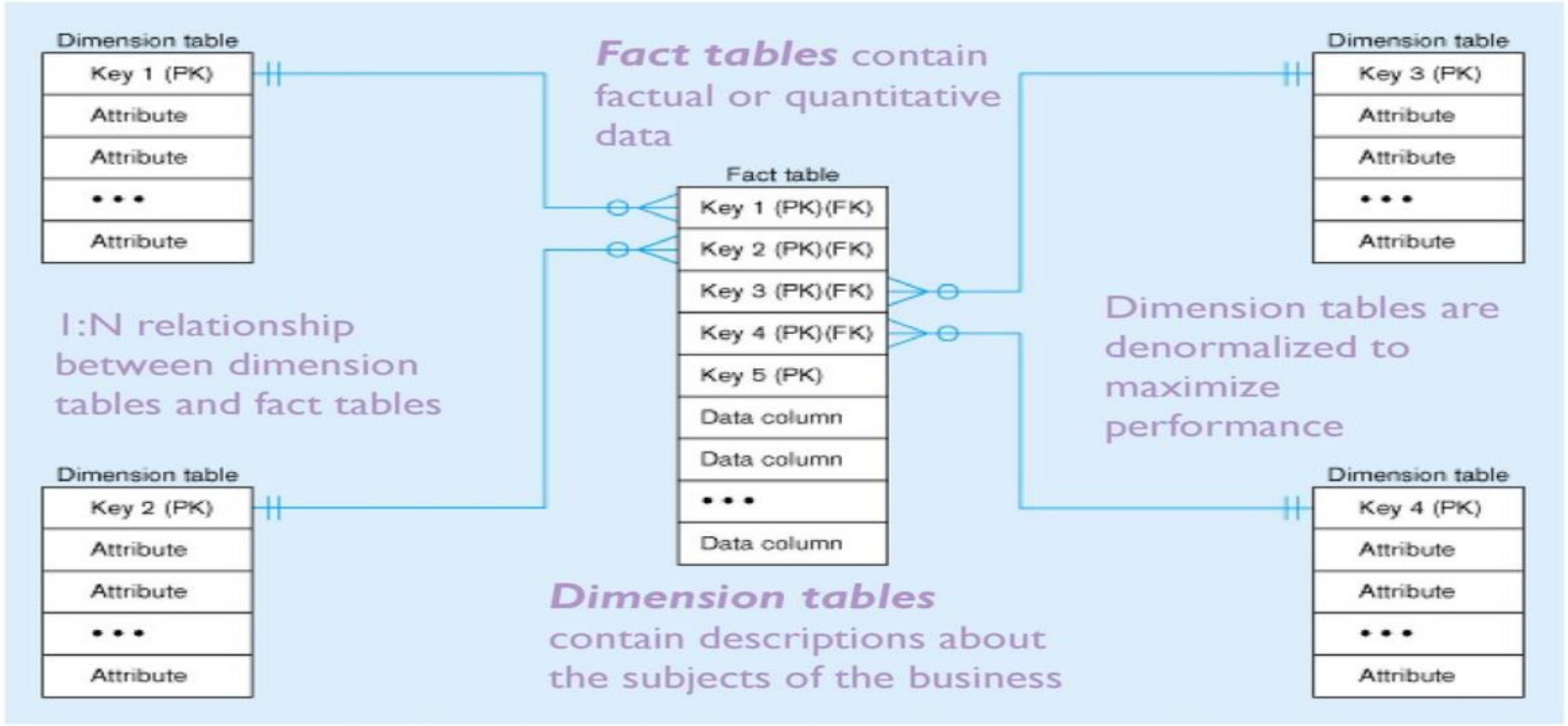
Data Warehouse

Sirojul Munir S.SI, M.KOM – Semester Genap TA 20182

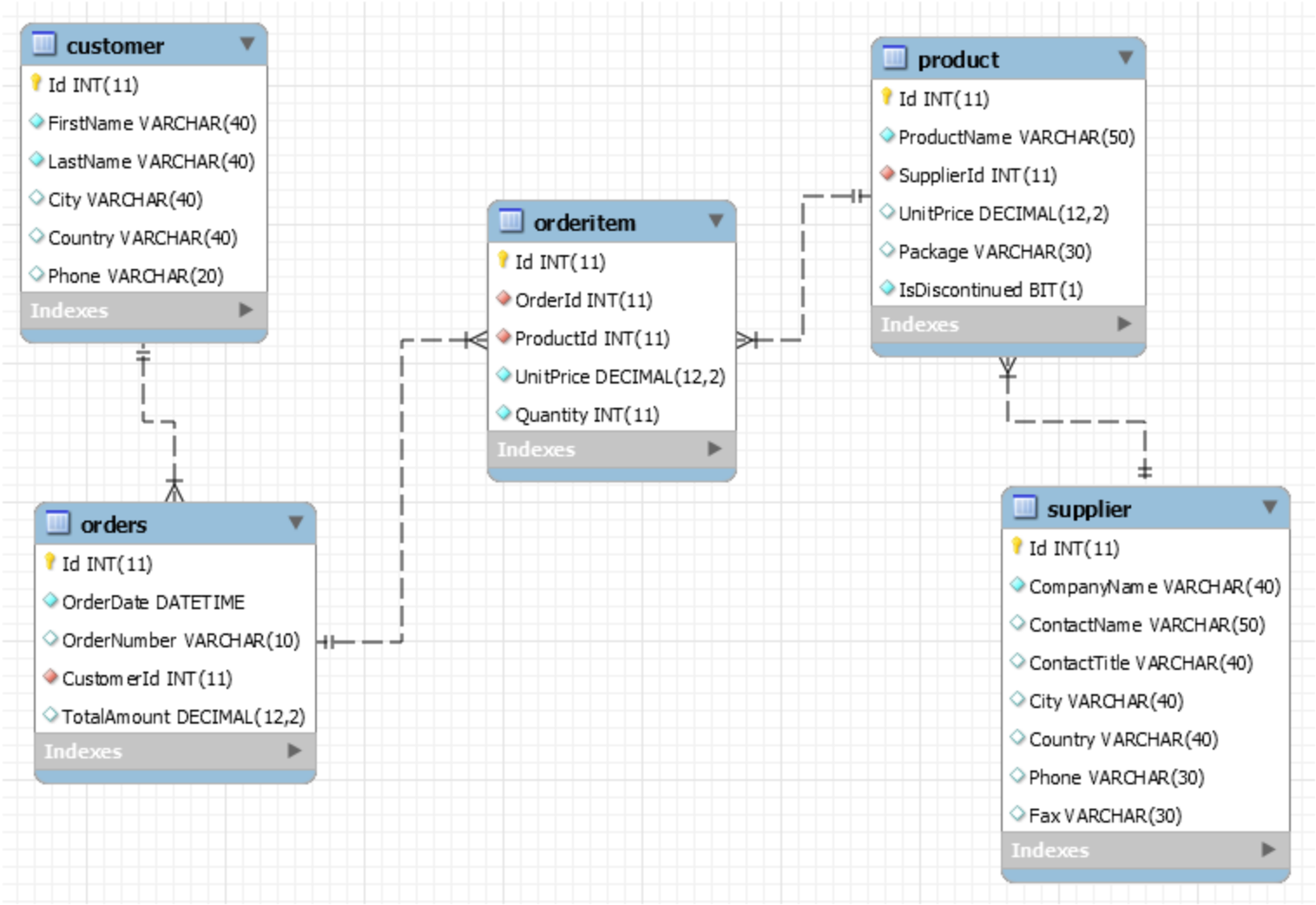
STAR SCHEMA - DATA CUBES

Star Schema

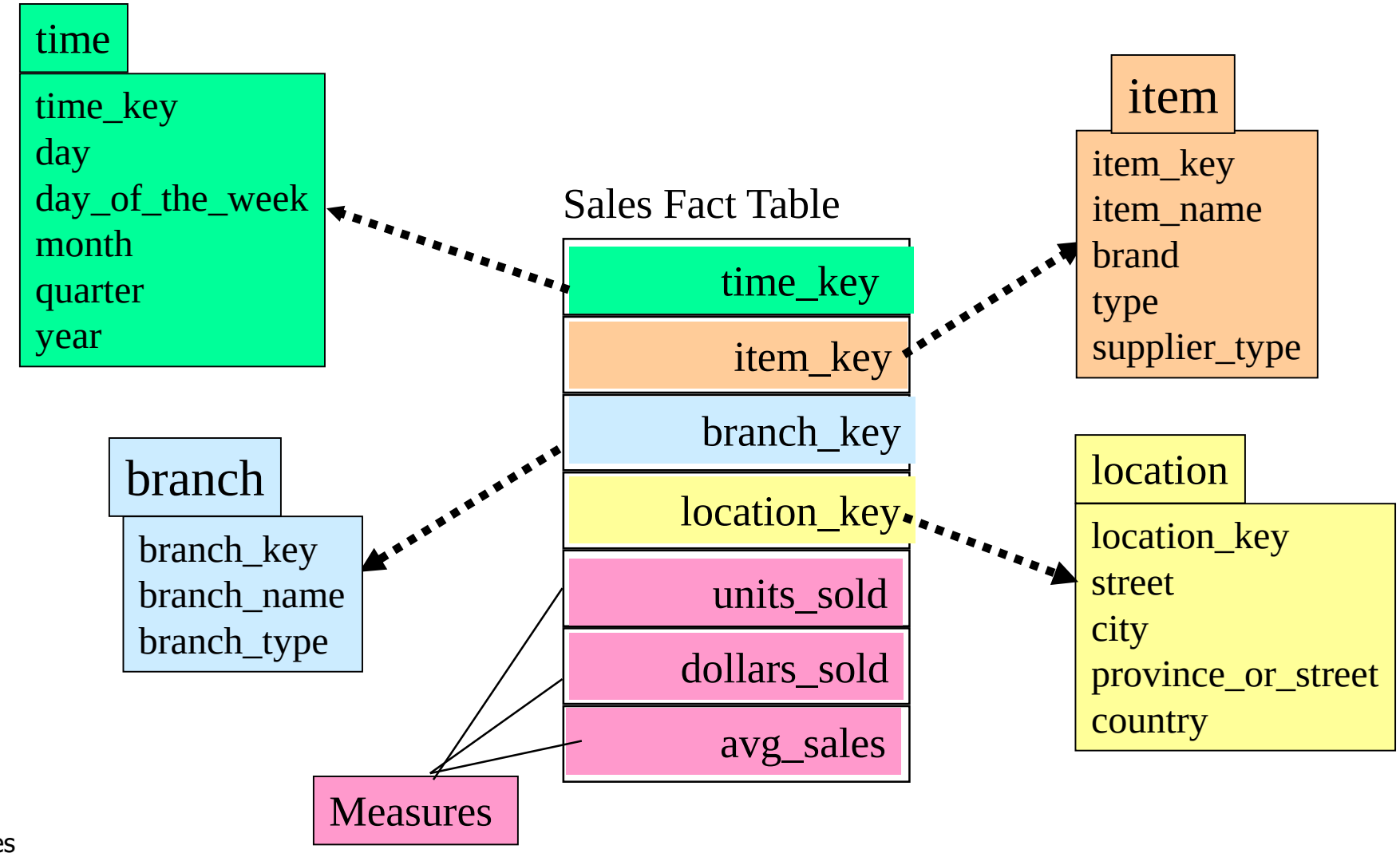


Excellent for ad-hoc queries,
but bad for online transaction processing

Studi Kasus : Product Orders

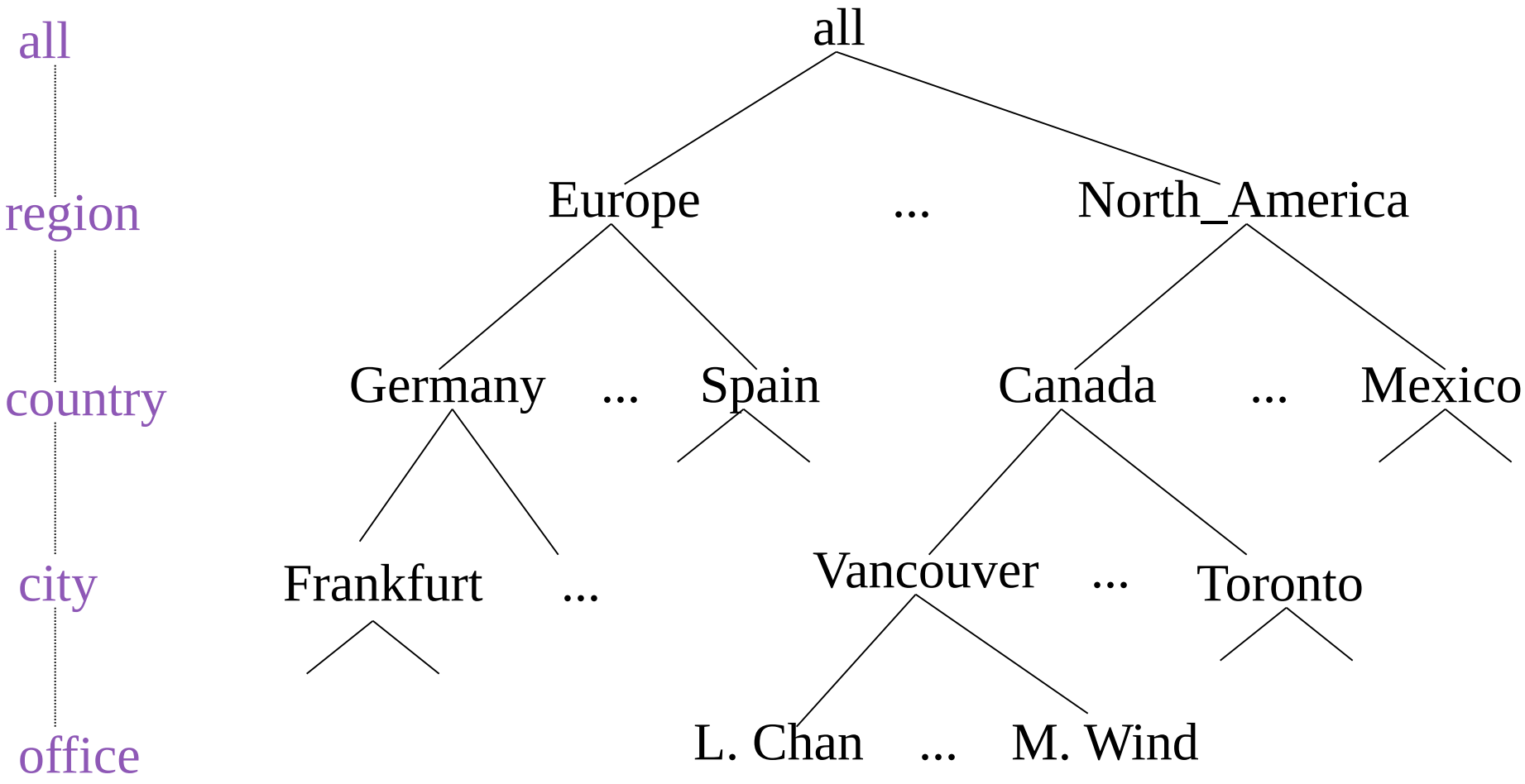


Contoh Star Schema



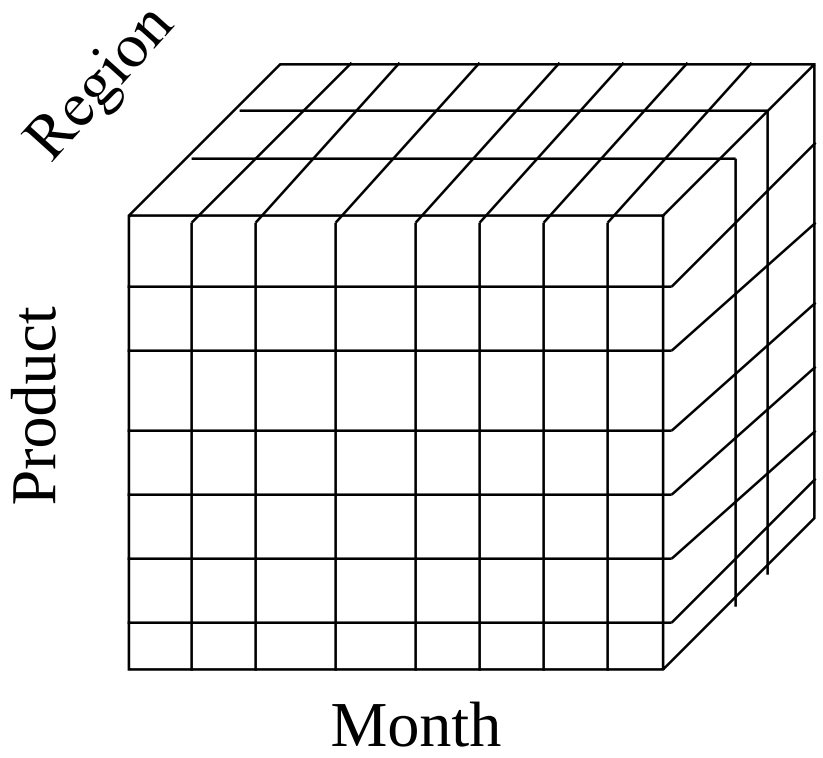
Han: Data Cubes

A Concept Hierarchy : Dimension Location

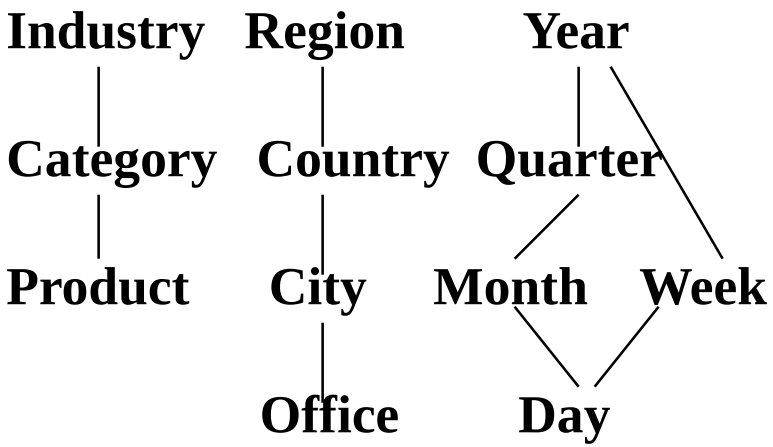


Multidimensi Data

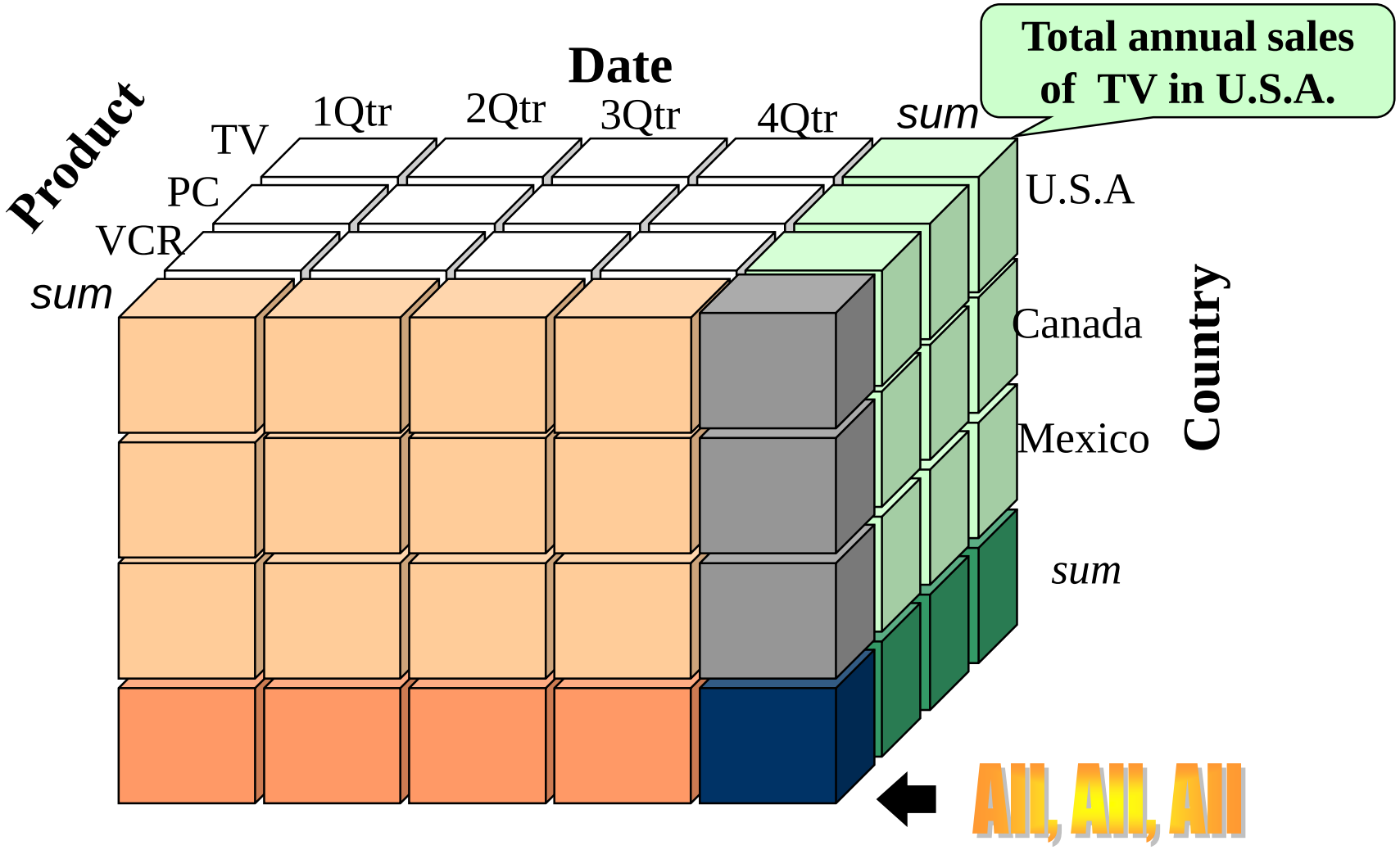
Sales volume as a function of product, month, and region



Dimensions: Product, Location, Time
Hierarchical summarization paths



Contoh: Data Cubes



Data Penjualan 2018

Sales of Jakarta	TV	DVD	Audio
Januari	2000	1000	500
Februari	1500	1500	1000
Maret	500	1000	2000
April	500	1000	2000

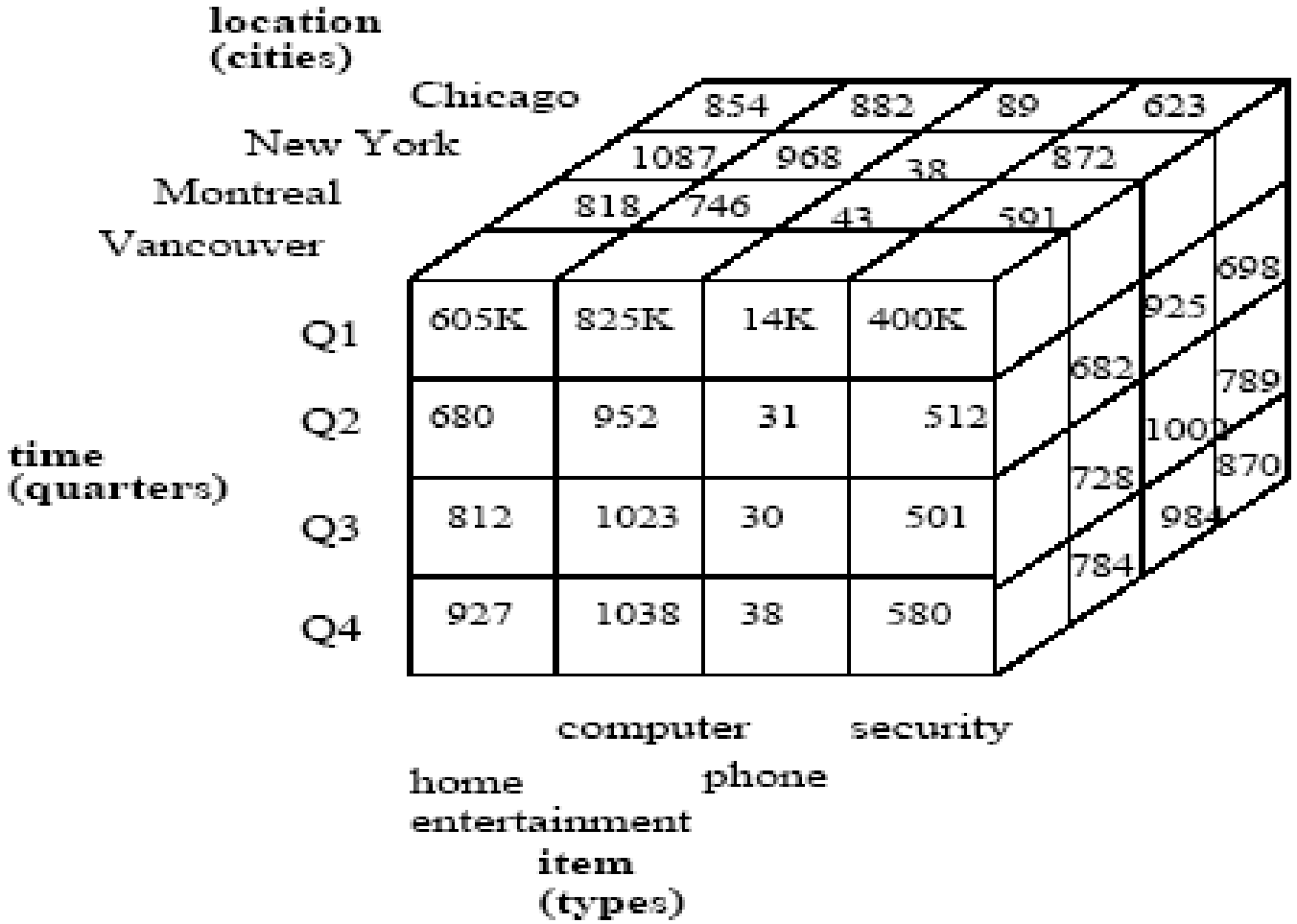
Data Penjualan 2018

Sales of Depok	TV	DVD	Audio
Januari	1000	100	500
Februari	500	200	100
Maret	800	300	200
April	1500	400	200

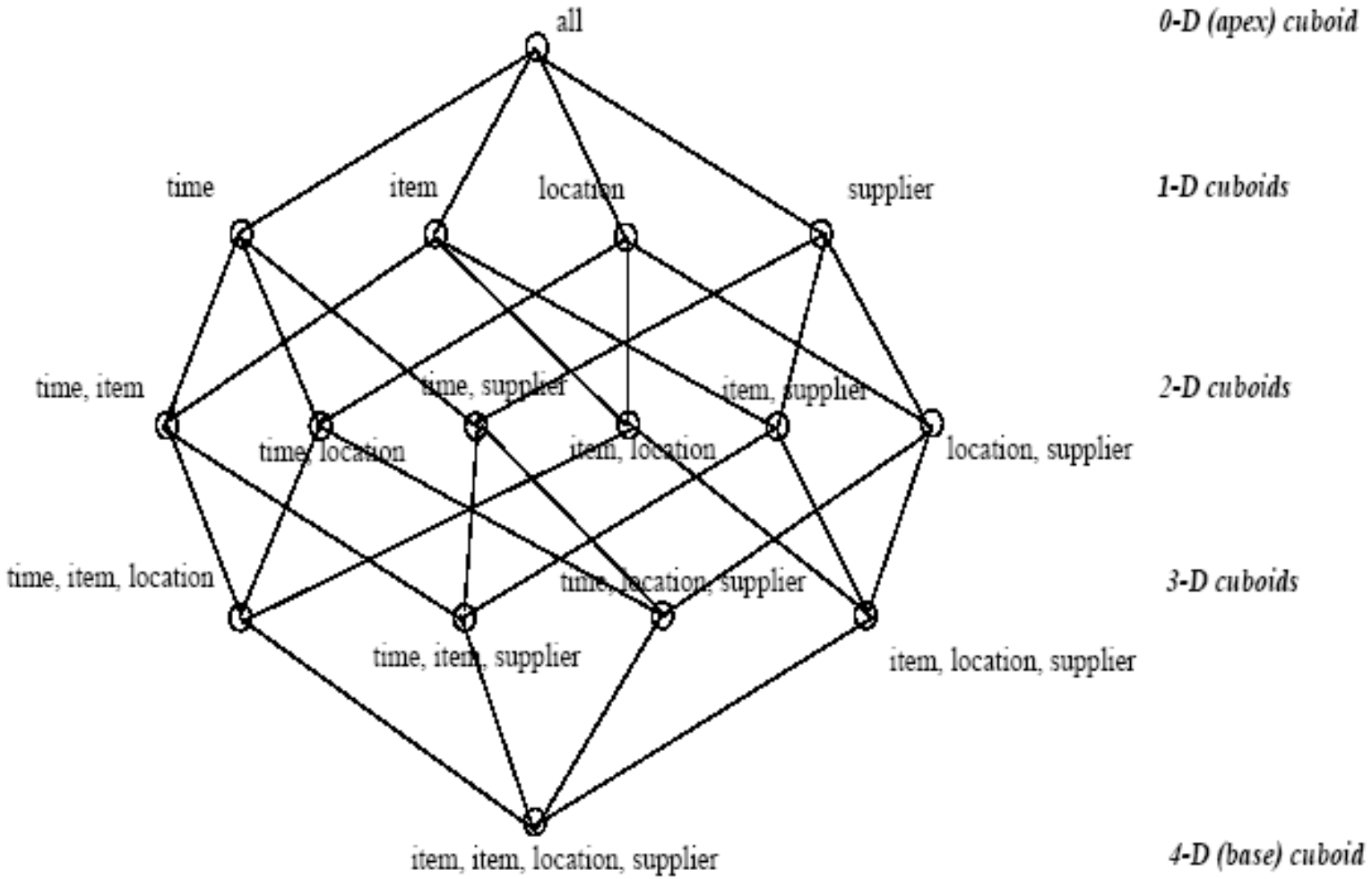
Multidimensional (3D) Table

Items	TV		DVD		Audio	
City	Jakarta	Depok	Jakarta	Depok	Jakarta	Depok
Januari	2000	1000	1000	100	500	500
Febaruari	1500	500	1500	200	1000	100
Maret	500	800	1000	300	2000	200
April	500	1500	1000	400	2000	200

Sample Data cube



Cuboids


all
time
item
location
supplier
time, item
time, supplier
item, supplier
time, location
item, location
location, supplier
time, item, location
time, location, supplier
time, item, supplier
item, location, supplier
item, item, location, supplier
0-D (apex) cuboid
1-D cuboids
2-D cuboids
3-D cuboids
4-D (base) cuboid

Data Cube

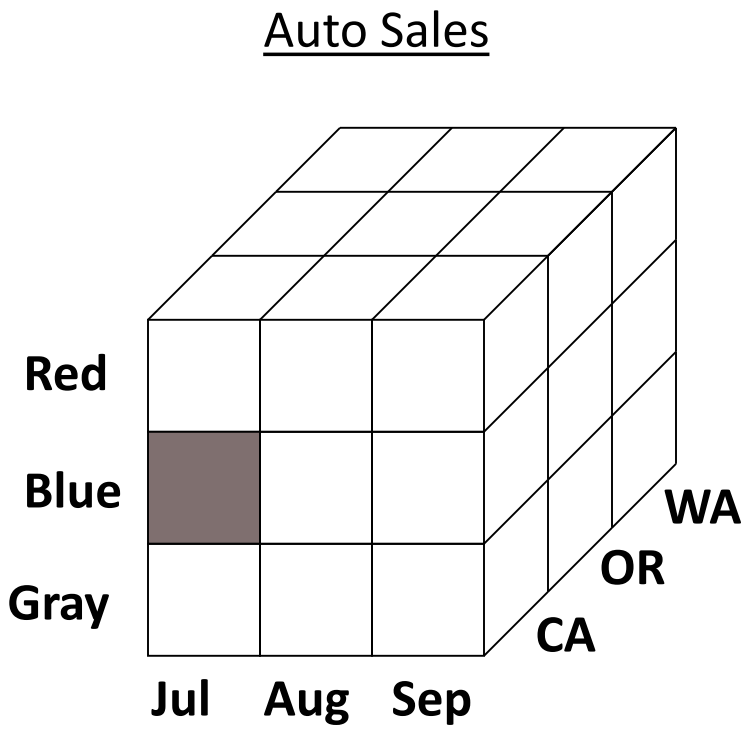
Axes of the cube represent attributes of the data records

- Generally discrete-valued / categorical
- e.g. color, month, state
- Called dimensions

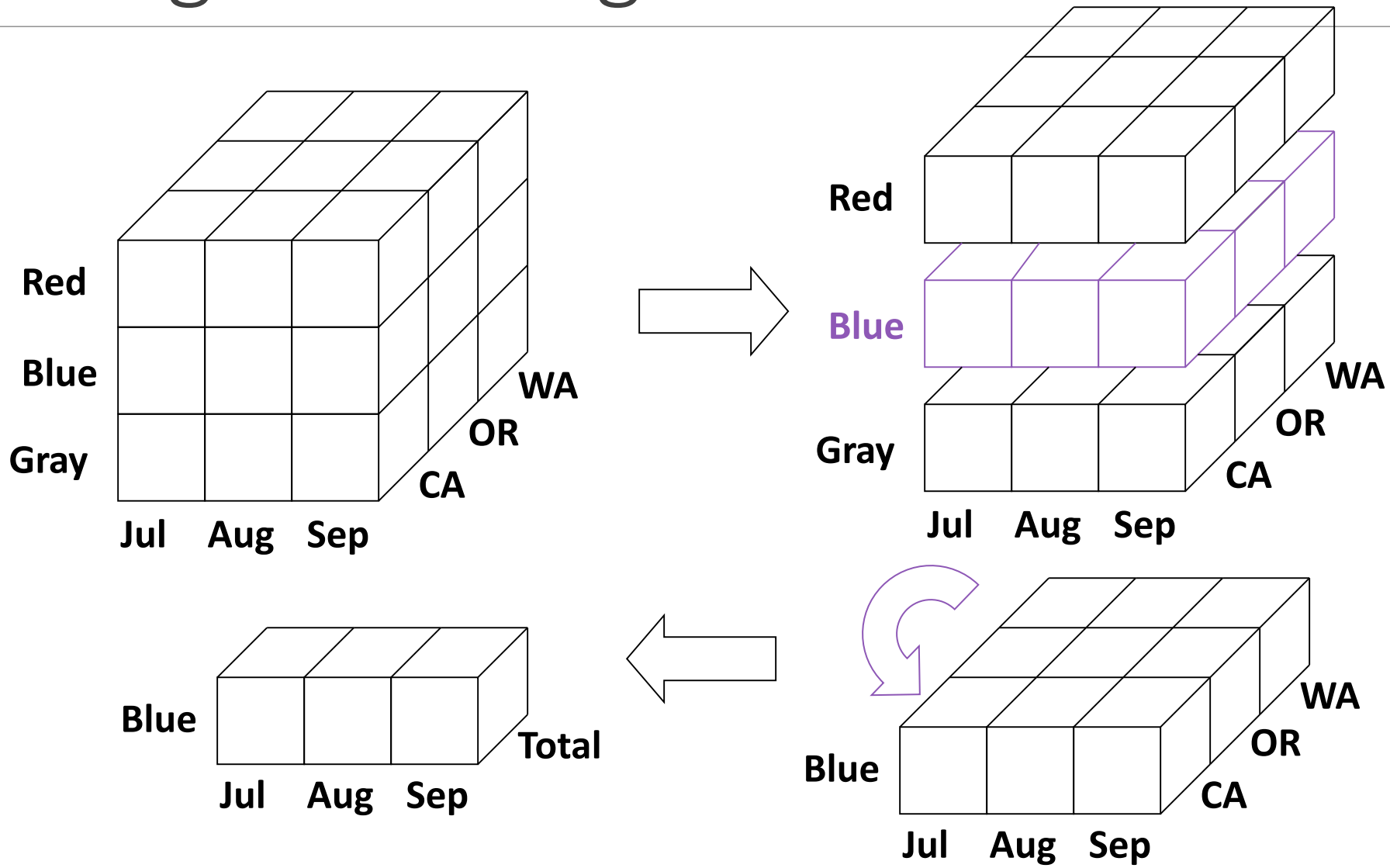
Cells hold aggregated measurements

- e.g. total $ sales, number of autos sold
- Called facts

Real data cubes have >> 3 dimensions



Slicing and Dicing



Querying the Data Cube

Cross-tabulation

- "Cross-tab" for short
- Report data grouped by 2 dimensions
- Aggregate across other dimensions
- Include subtotals

Operations on a cross-tab

- Roll up (further aggregation)
- Drill down (less aggregation)

Number of Autos Sold

	CA	OR	WA	Total
Jul	45	33	30	108
Aug	50	36	42	128
Sep	38	31	40	109
Total	133	100	112	345

Roll Up and Drill Down

Number of Autos Sold

	CA	OR	WA	Total
Jul	45	33	30	108
Aug	50	36	42	128
Sep	38	31	40	109
Total	133	100	112	345

Roll up
by Month

Number of Autos Sold

CA	OR	WA	Total
133	100	112	345

Drill down
by Color

Number of Autos Sold

	CA	OR	WA	Total
Red	40	29	40	109
Blue	45	31	37	113
Gray	48	40	35	123
Total	133	100	112	345

“Standard” Data Cube Query

Measurements

- Which fact(s) should be reported?

Filters

- What slice(s) of the cube should be used?

Grouping attributes

- How finely should the cube be diced?
- Each dimension is either:
  - (a) A grouping attribute
  - (b) Aggregated over (“Rolled up” into a single total)
- n dimensions → 2^n sets of grouping attributes
- Aggregation = projection to a lower-dimensional subspace

Latihan

1. Buat query yang menghasilkan skema bintang !
2. Buat query yang mengambil salah satu dimensi data (product, location , time) !