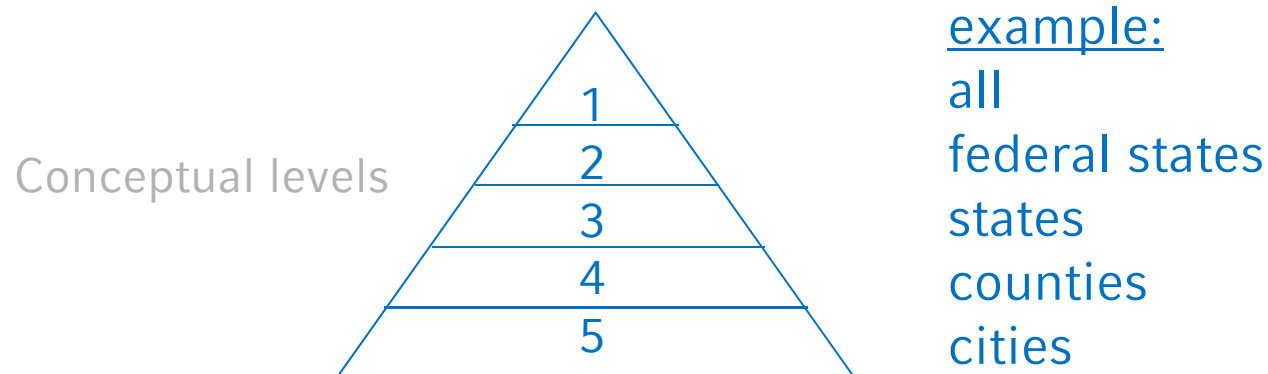


Daten Generalisierung:

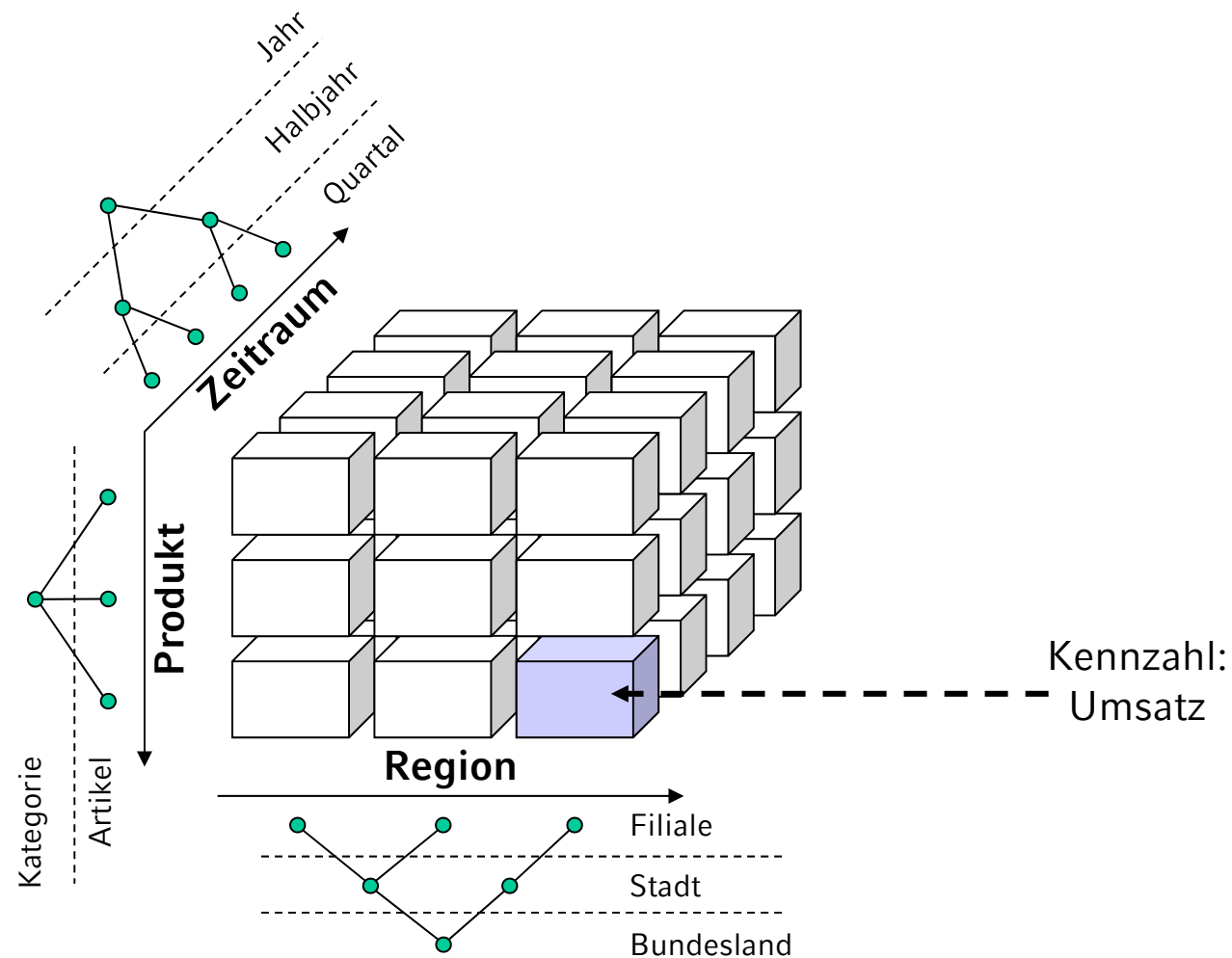
- Prozess, der Anwendungsdaten schrittweise von niedrigen auf höhere konzeptuelle Level aggregiert



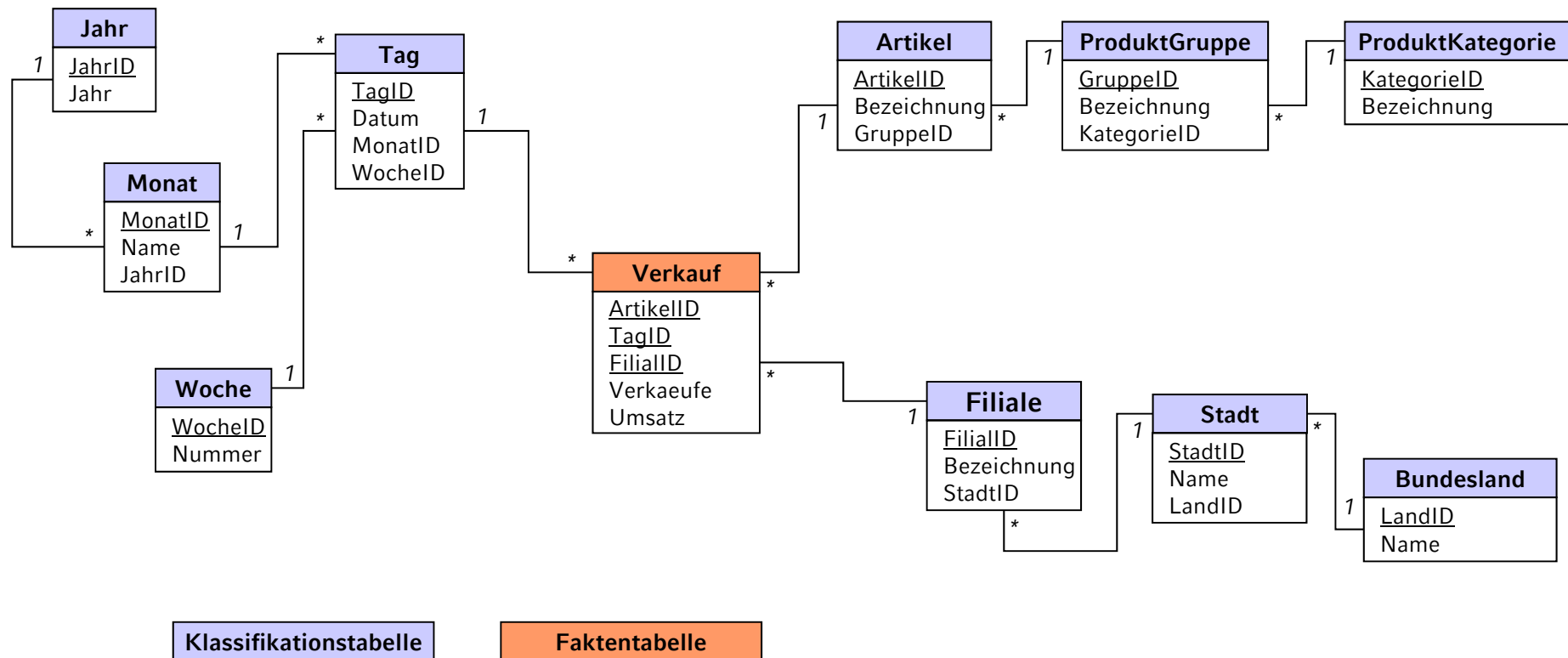
- Ansätze:
 - Data-cube approach (OLAP / Roll-up) → manuell
 - Attribute-oriented induction (AOI) → automatisiert

- OLAP basiert auf Data Warehouses und ist Basis aller BI-Systeme
- Datawarehouses:
 - Datenmodell sollte bzgl. Analyseprozess optimiert werden
 - Datenanalyse im Entscheidungsprozess
 - Betriebswirtschaftliche Kennzahlen stehen im Mittelpunkt (z.B. Erlöse, Gewinne, Verluste, Umsätze, ...)
=> Fakten
 - Betrachtung dieser Kennzahlen aus unterschiedlichen Perspektiven (z.B. zeitlich, regional, produktbezogen, ...)
=> Dimensionen
 - Unterteilung der Auswertungsdimensionen möglich (z.B. zeitlich: Jahr, Quartal, Monat; regional: Bundesländer, Bezirke, Städte/Gemeinden; ...)
=> Hierarchien, Konsolidierungsebenen
 - Data-Cubes
 - Grundlage der multidimensionalen Datenanalyse: Datenwürfel (Data-Cube)
 - Kanten des Cubes: Dimensionen
 - Zellen des Cubes: ein oder mehrere Kennzahlen (als Funktion der Dimension)
 - Anzahl der Dimensionen: Dimensionalität des Cubes

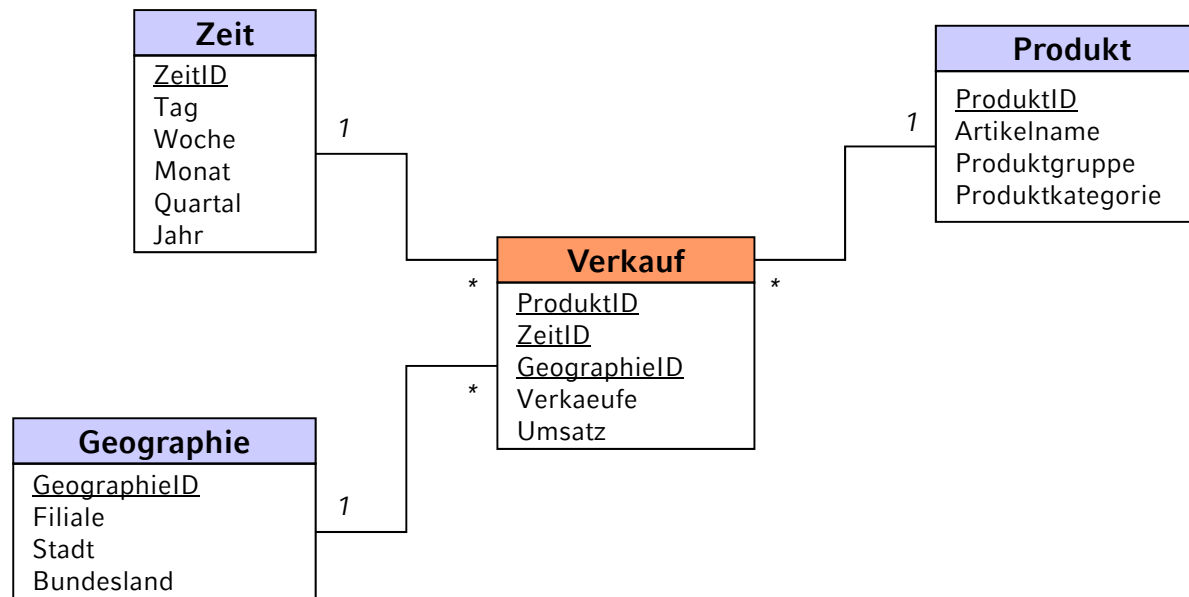
- Beispiel



- Umsetzung des Data Cube Modells meist relational:
 - Snowflake-Schema (ist in 3NF, also redundanzfrei)
 - Faktentabelle
 - Eigen Tabelle für jede Klassifikationsebene entlang der Dimensionen



- Star-Schema (redundant aber effizienter)
 - Faktentabelle
 - Denormalisierte Dimensionstabelle (eine pro Dimension, dadurch weniger Joins entlang der Dimensions-Hierarchien notwendig)



OLAP-Operationen zur Analyse:

- *Roll up*: summarize data
 - by climbing up hierarchy or by dimension reduction
- *Drill down*: reverse of roll-up
 - from higher level summary to lower level summary or detailed data, or introducing new dimensions
- *Slice and dice*:
 - selection on one (slice) or more (dice) dimensions
- *Pivot (rotate)*:
 - reorient the cube, visualization, 3D to series of 2D planes
- Other operations
 - *drill across*: involving (across) more than one fact table
 - *drill through*: through the bottom level of the cube to its back-end relational tables (using SQL)

Beispiel: Roll up / Drill down

Query:

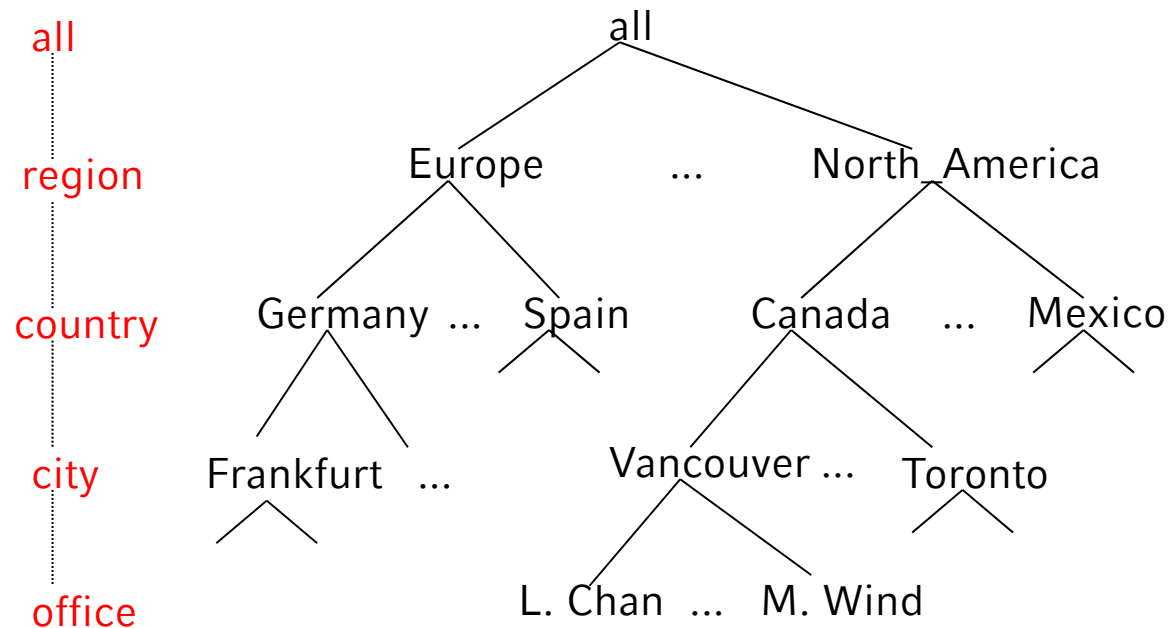
```
SELECT *
FROM facts
WHERE condition
(table: results)
```

*aggregating
values*



Roll-up:

```
SELECT ...
FROM results
GROUP BY country, year
```

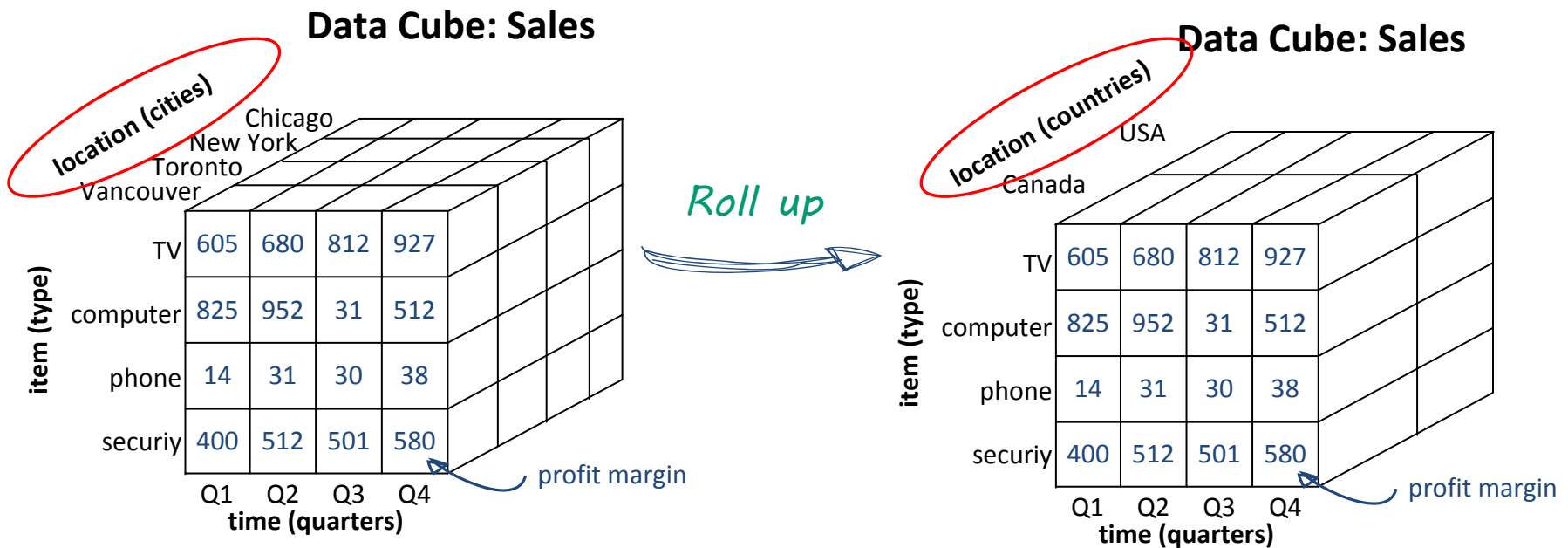


*expand
results*



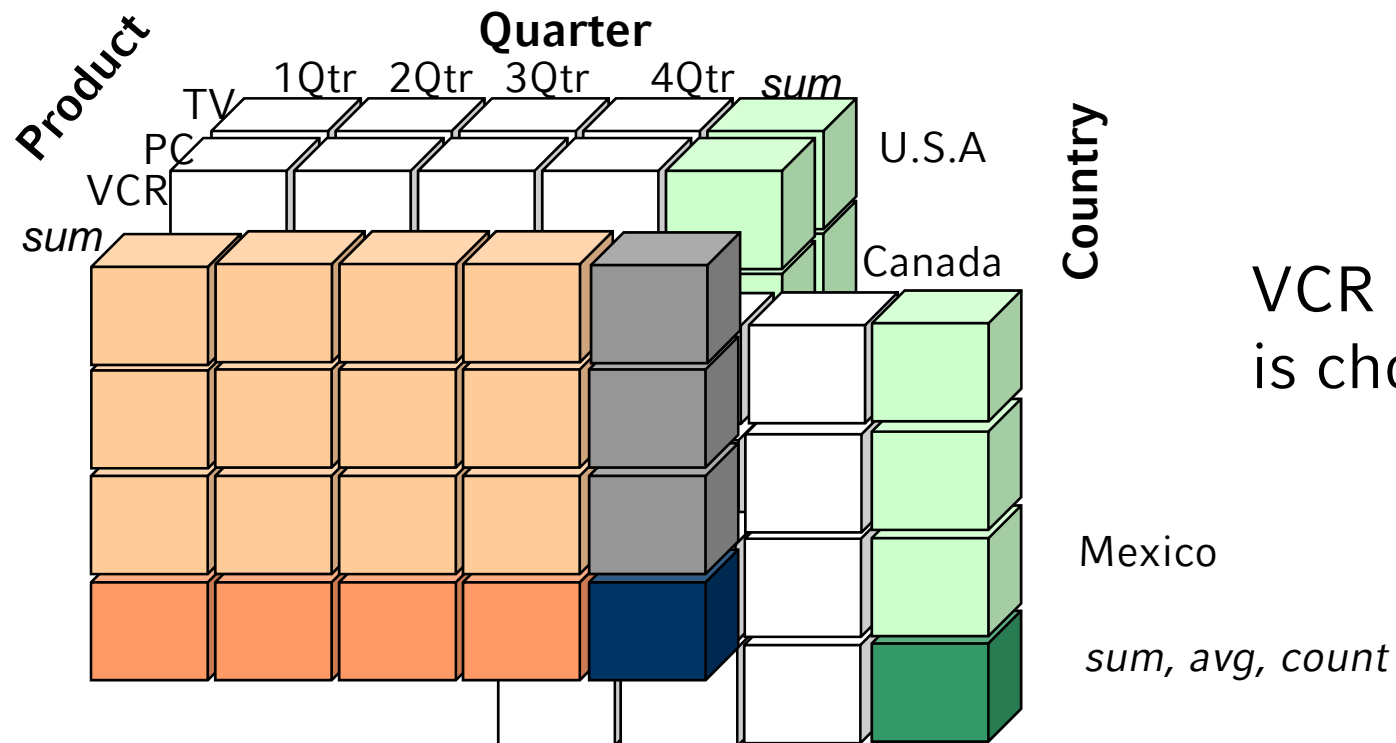
Drill down:

```
GROUP BY city, year
```



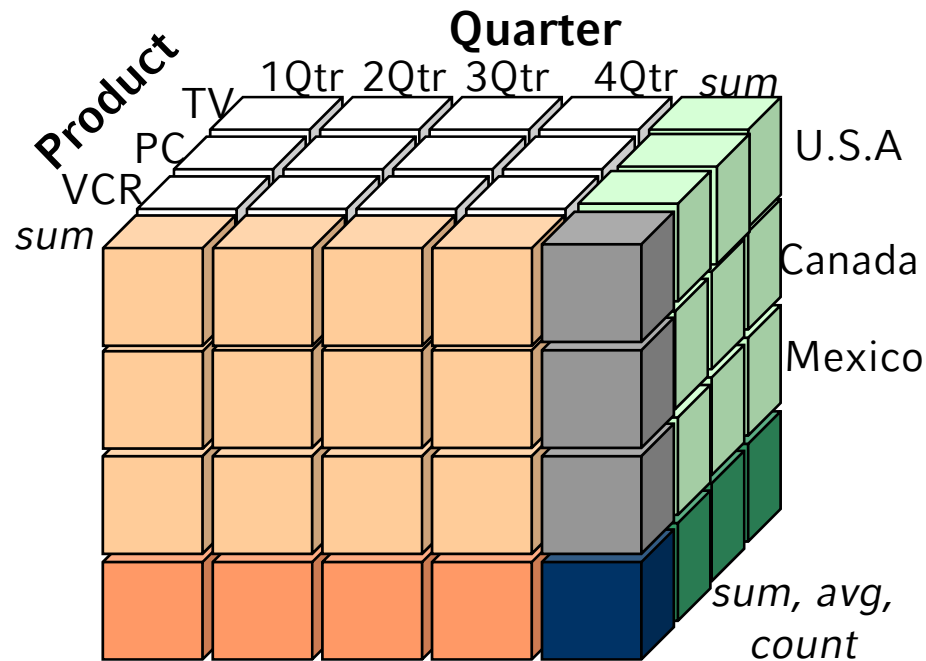
Beispiel: Slice

```
SELECT income
FROM   time t, product p, country c
WHERE  p.name = 'VCR'
```



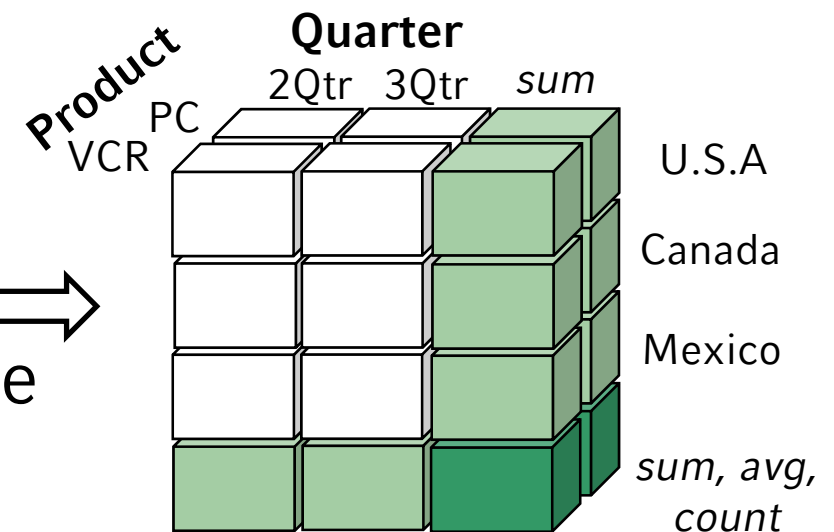
```

SELECT income
FROM   time t, product p, country c
WHERE  p.name = 'VCR' OR p.name = 'PC'
AND    t.quarter BETWEEN 2 AND 3
  
```



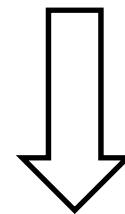
sub-data cube over PC, VCR and quarters 2 and 3 is extracted

→
dice



Beispiel: Pivot (rotate)

Quarter 1			Quarter 2			Quarter 3		
TV	PC	VCR	TV	PC	VCR	TV	PC	VCR
...
...



Pivot (rotate)

TV			PC			VCR		
Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
...
...

- Vorteile
 - Effiziente Implementatierung der Daten-Generalisierung
 - Verschiedene Aggregate als Kennzahlen möglich
 - Z.B. count, sum, average, max
 - Generalisierung (und Spezialisierung) durch *roll-up* (and *drill-down*) im Data Cube
 - Etablierte Technik um Reports (Visualisierungen von Cube-Ebenen) zu erstellen
- Nachteile
 - Nur *simple nonnumeric data* und *simple aggregated numeric values*.
 - Keine intelligente Analyse, insbesondere keine Hinweise, welche Dimensionen/Hierarchie-Level genutzt werden sollen