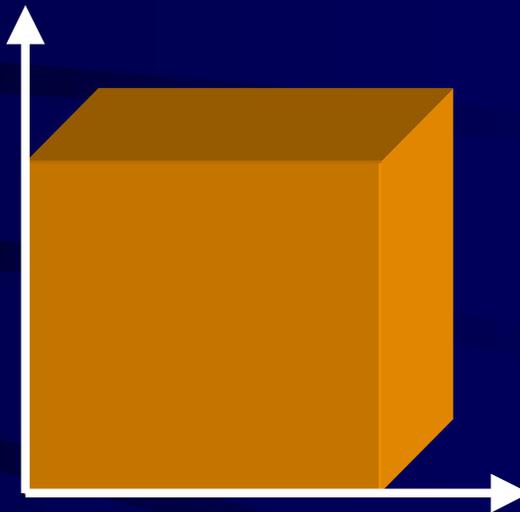


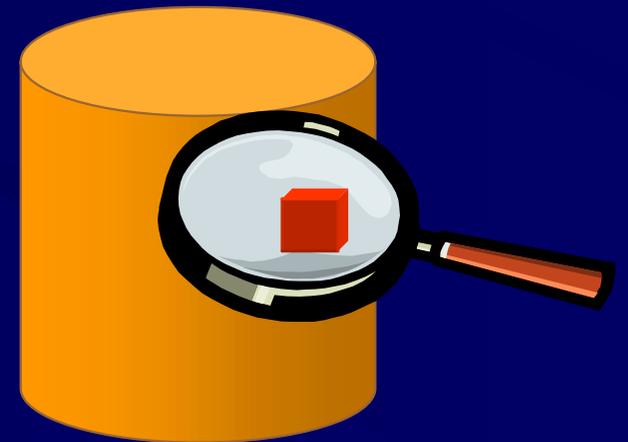
Logische Modelle für OLAP



Burkhard Schäfer

Übersicht

- Einführung in OLAP
- Multidimensionale Daten: Hypercubes
- Operationen
- Formale Grundlagen
- Zusammenfassung



Einführung in OLAP

- Verfahren zur Analyse großer Datenbestände
- Data Warehouse als Datenquelle
- Multidimensionales Datenmodell

- Anwendung:
Analyse historischer Geschäftsdaten,
Entscheidungsunterstützung

Datenquelle

- Data Warehouse: flache Struktur

Produkt	Volumen	Ort	Kunde	Zeit	Umsatz
Milch	1,00	Alzey	Müller	1.10.03	1,00
Rotwein	0,75	Mainz	Meyer	1.11.03	8,50
Vodka	0,70	KL	Chekov	9.01.03	12,00

- Auswahl zu analysierender Attribute

VK (Produkt, Volumen, Ort, Kunde, Zeit, Umsatz)

- Klassifikation in bestimmende und abhängige Attribute

Hypercube

- N-dimensionale Datenstruktur

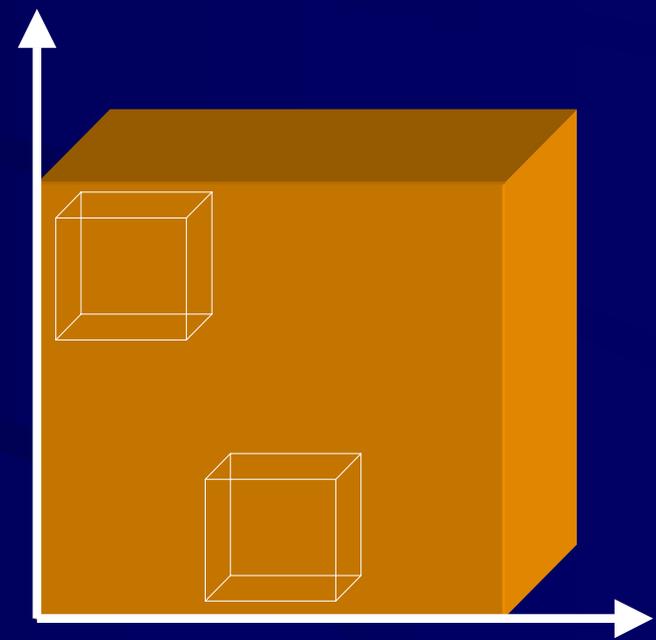
- Dimensionen:
bestimmende Attribute

→ Achsen

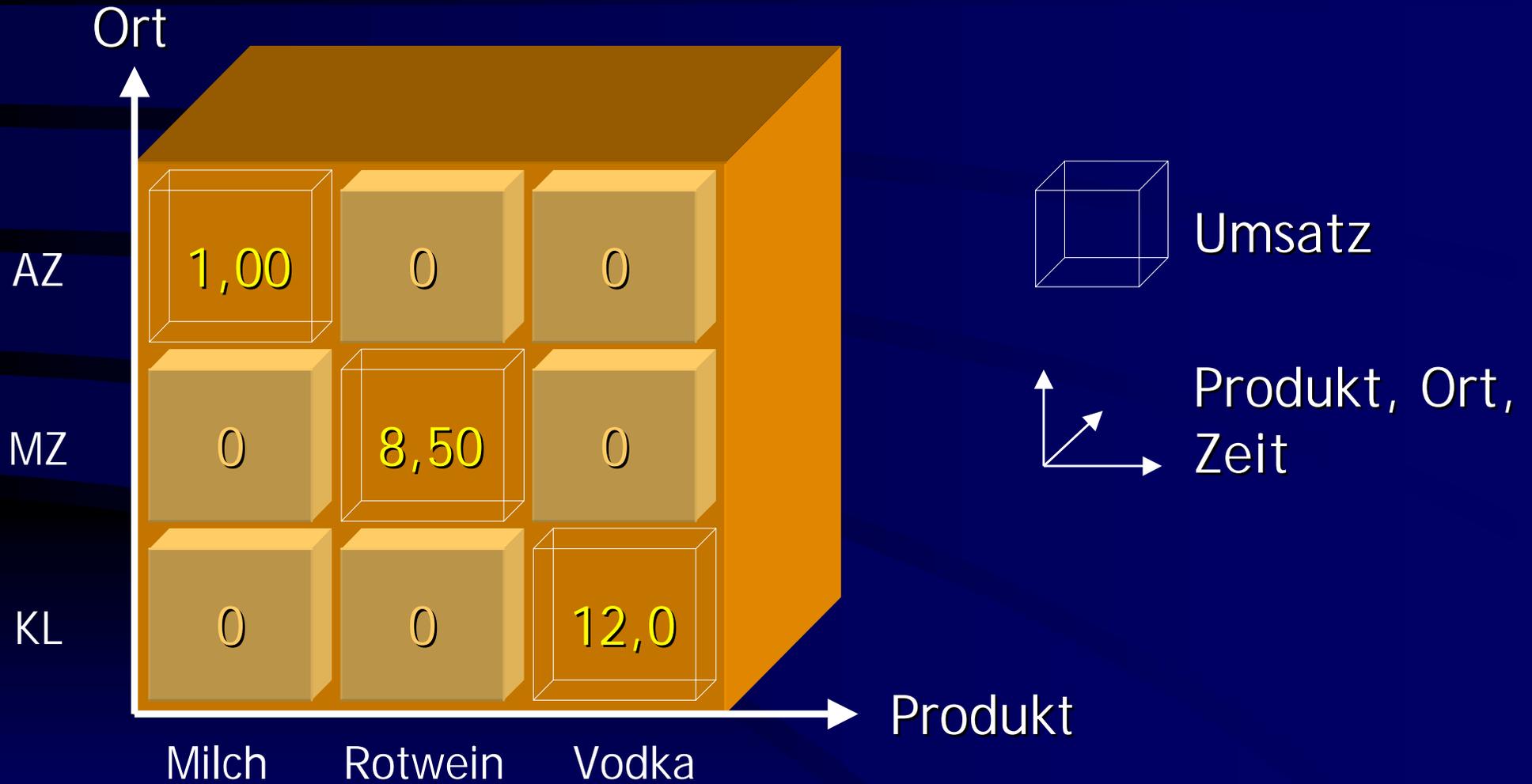
- Maße:
abhängige Attribute

→ Zellen

mit Dimensionswerten als Koordinaten



Hypercube Beispiel



Informationsgewinnung

- Problem:
Daten im Hypercube zu detailliert
- Idee:
 - Zusammenfassung “ähnlicher” Daten
 - Auswahl an Detaillierungsgraden
 - Hierarchien

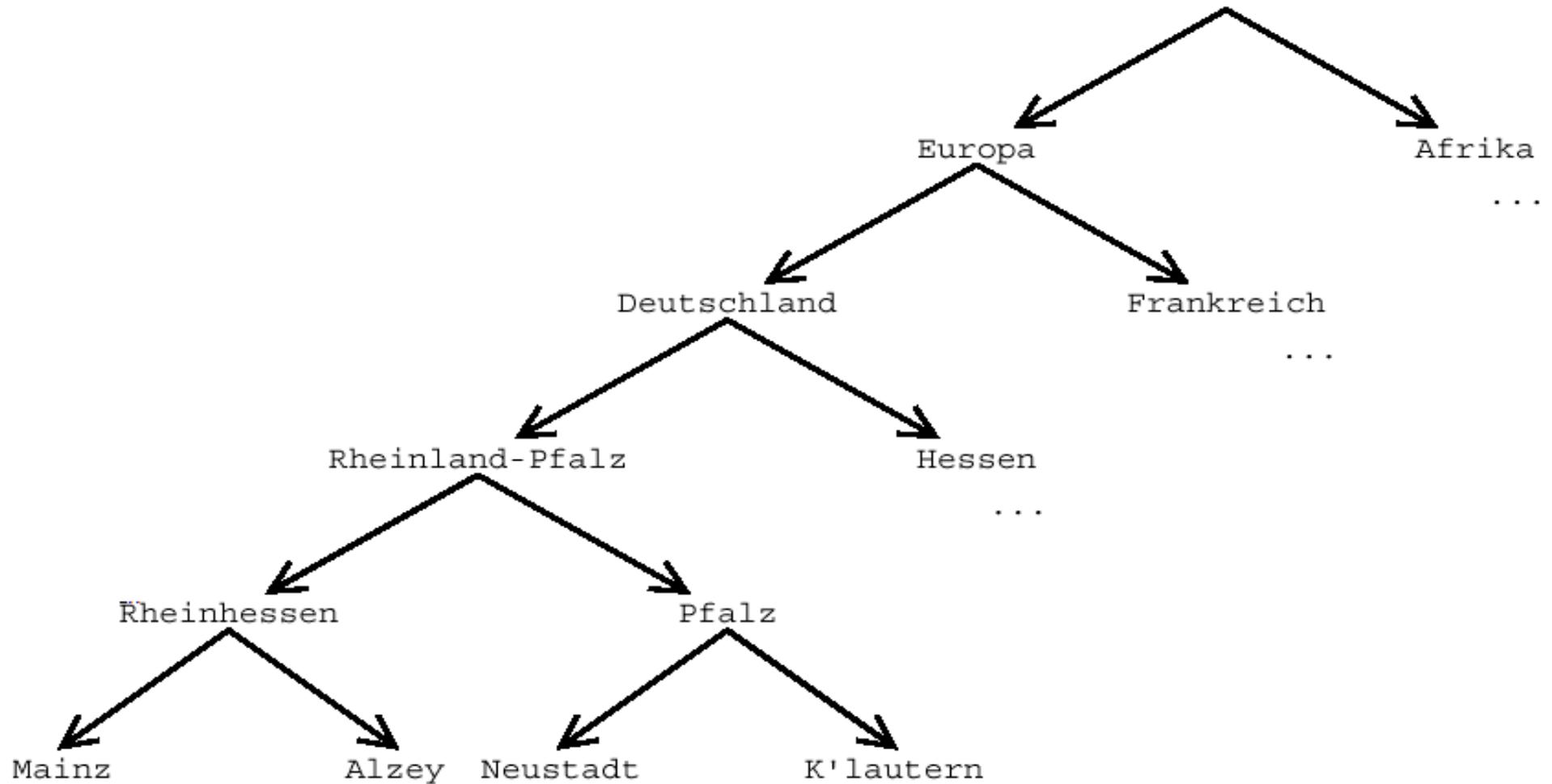
Hierarchien

- Hierarchisch organisierte Aggregationsebenen auf Dimensionen
- Part-Of-Relation zwischen Ebenen

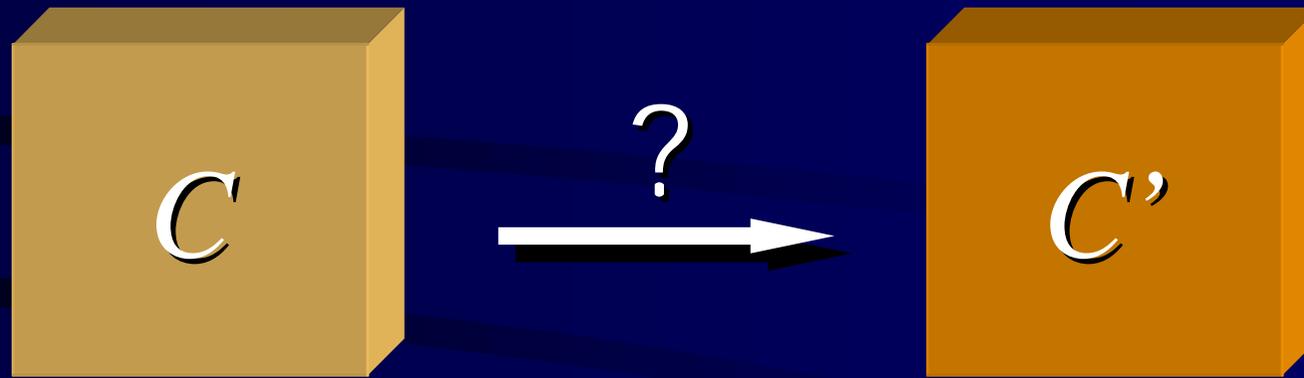


- Mit Prädikaten oder Abbildungen definiert
in(Pfalz, Kaiserslautern), in(Pfalz, Neustadt)
 $f_{1 \rightarrow 2}(\text{Alzey}) = \text{Rheinhessen}$

Hierarchie auf Ort



OLAP-Operationen

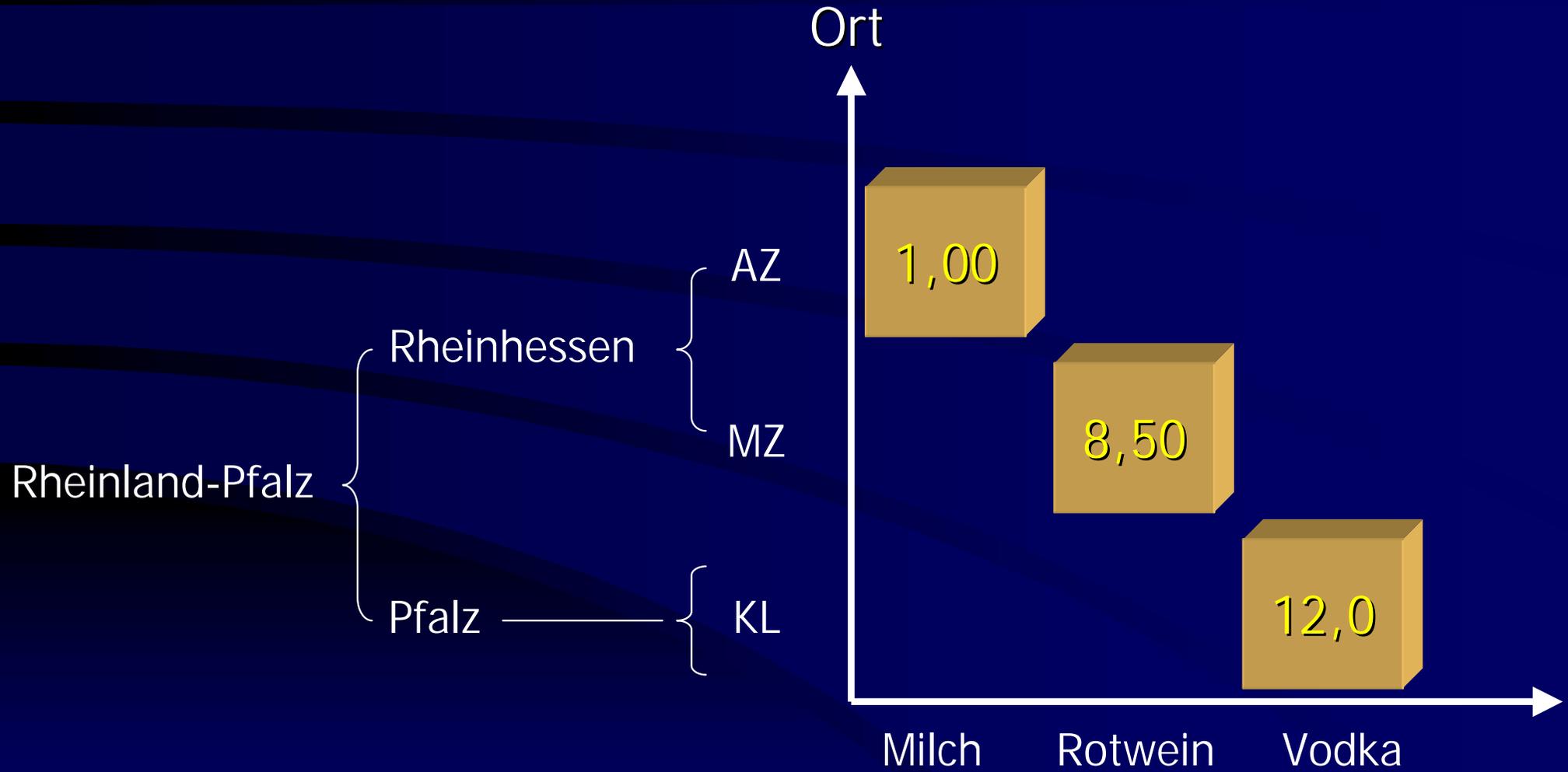


- Drill-Down und Roll-Up
- Slicing und Dicing
- Pivoting
- Push und Pull

Drill-Down und Roll-Up

- Navigation zwischen Hierarchieebenen
- **Drill-Down:**
 - detailliertere Ebene Land → Bundesland
- **Roll-Up:**
 - grobere Ebene Stadt → Region
 - Zusammenfassen von Werten (Maßen) der darunterliegenden Ebene mit Aggregationsfunktion

Roll-Up



Roll-Up

1. Roll-Up Stadt → Region



Roll-Up

1. Roll-Up Stadt → Region
2. Roll-Up Produkt → Typ

Rheinland-Pfalz {
Rheinhessen
Pfalz

Ort



Mehrere Einträge
→ Aggregation



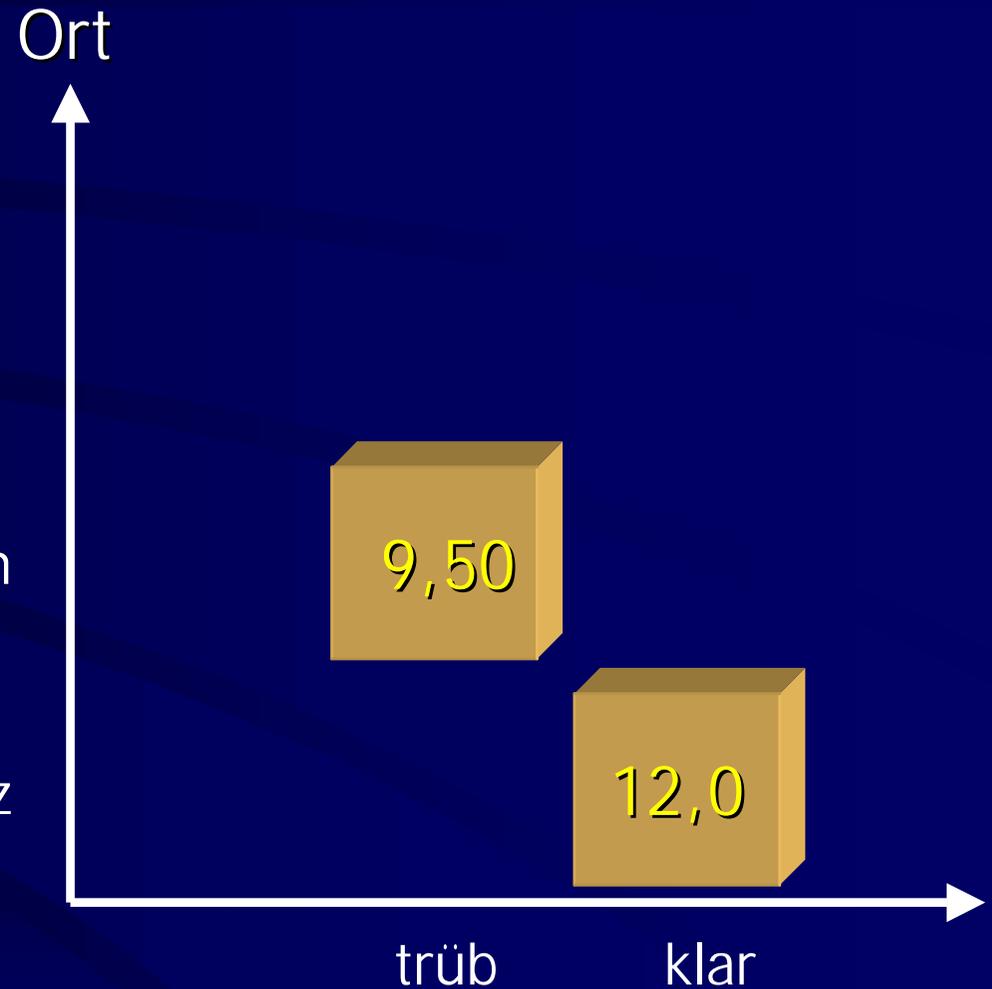
trüb

klar

Roll-Up

1. Roll-Up Stadt \rightarrow Region
2. Roll-Up Produkt \rightarrow Typ
3. Aggregation mit Σ

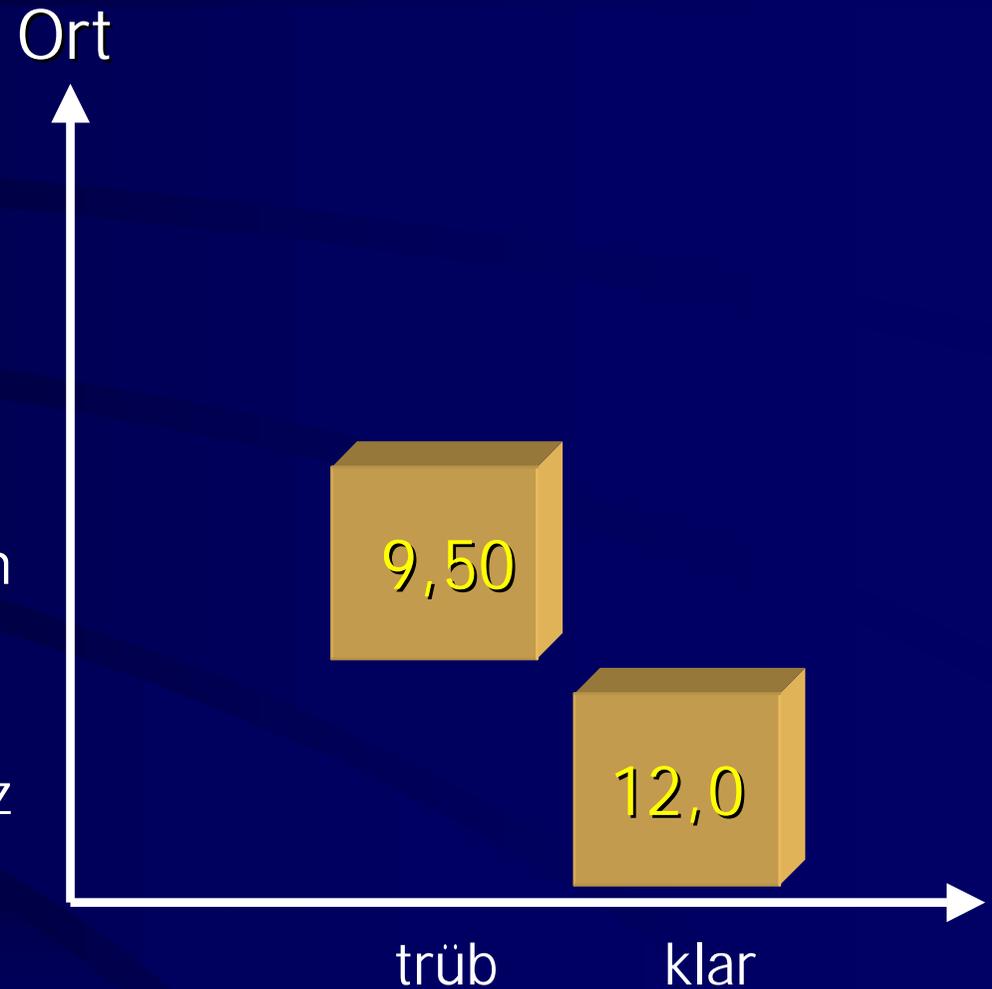
Rheinland-Pfalz {
Rheinhessen
Pfalz



Roll-Up

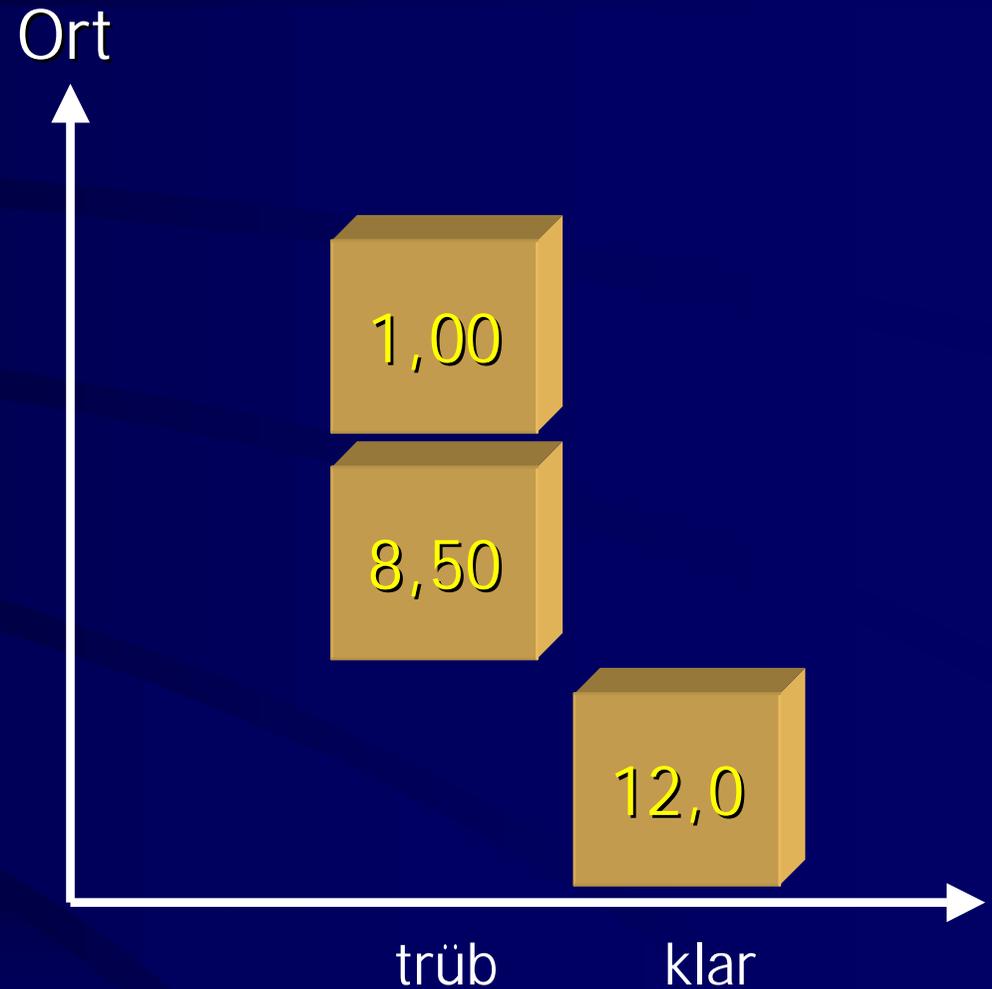
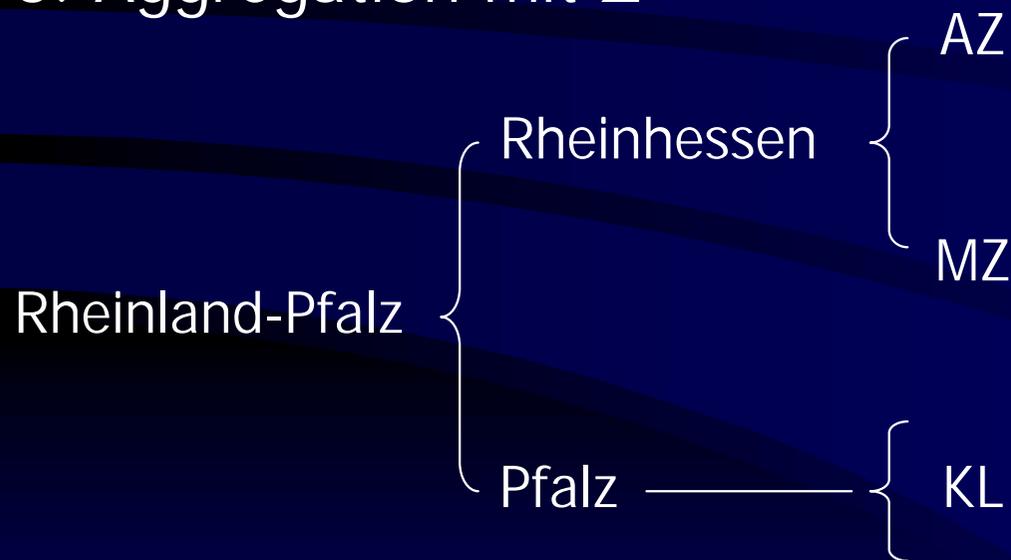
1. Roll-Up Stadt → Region
(Aggregation unnötig)
2. Roll-Up Produkt → Typ
3. Aggregation mit Σ

Rheinland-Pfalz {
Rheinhessen
Pfalz



Drill-Down

- ~~1. Roll-Up Stadt → Region
(Aggregation unnötig)~~
2. Roll-Up Produkt → Typ
3. Aggregation mit Σ



Slicing und Dicing

- Restriktionen auf Hypercubes
- Selektives Entfernen von Zellen

Slicing:

- Restriktion auf einer Dimension

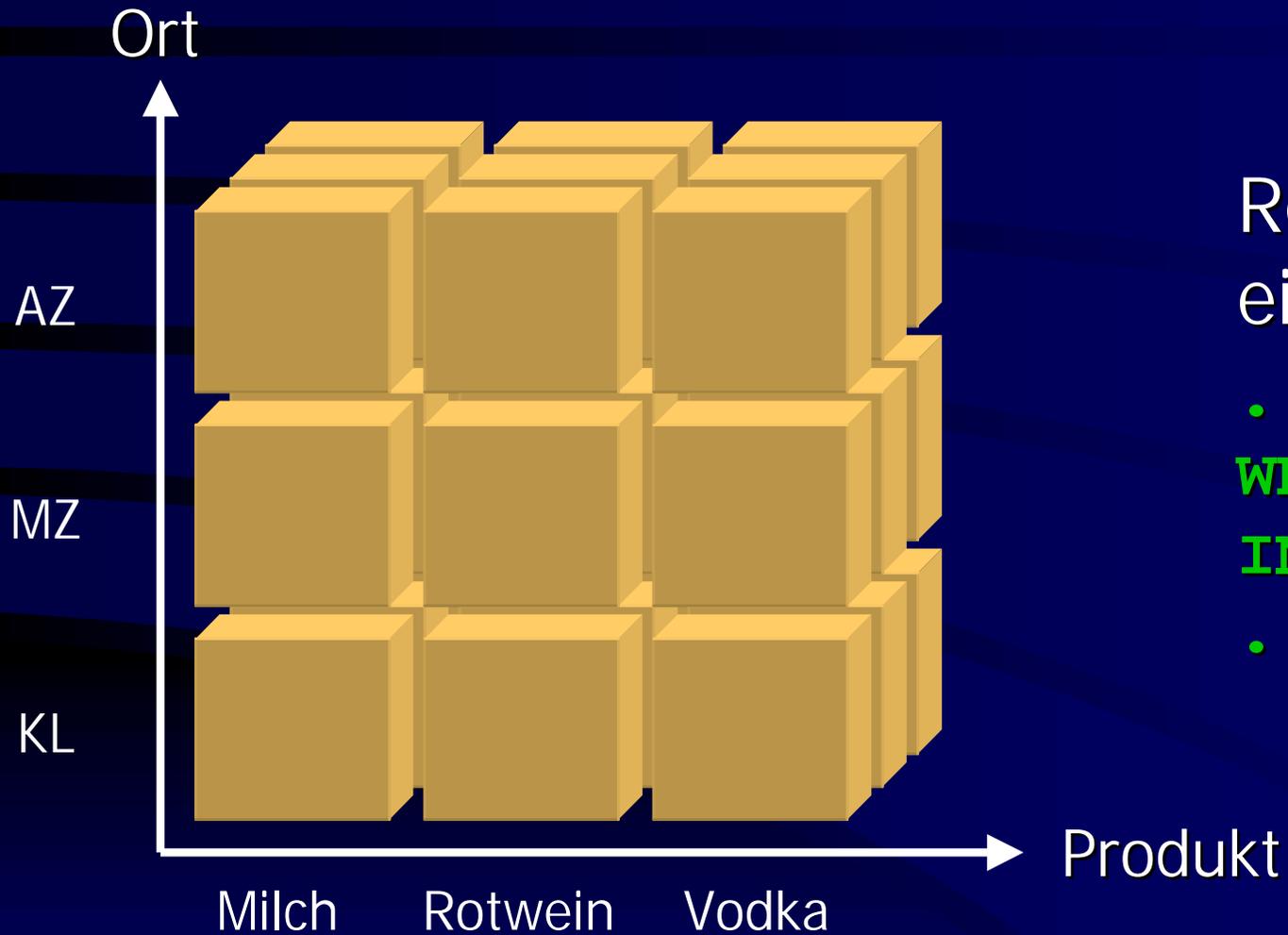
Dicing:

- Restriktion auf mehreren Dimensionen

Slicing und Dicing

- Auswahl von Zellen über
 - Intervalle $5 < \text{Umsatz} < 10$
 - Mengen Ort in (KL, MZ)
 - Prädikate $P(m) \text{ gdw } m.\text{Umsatz} \text{ MOD } 2 = 0$

Slicing



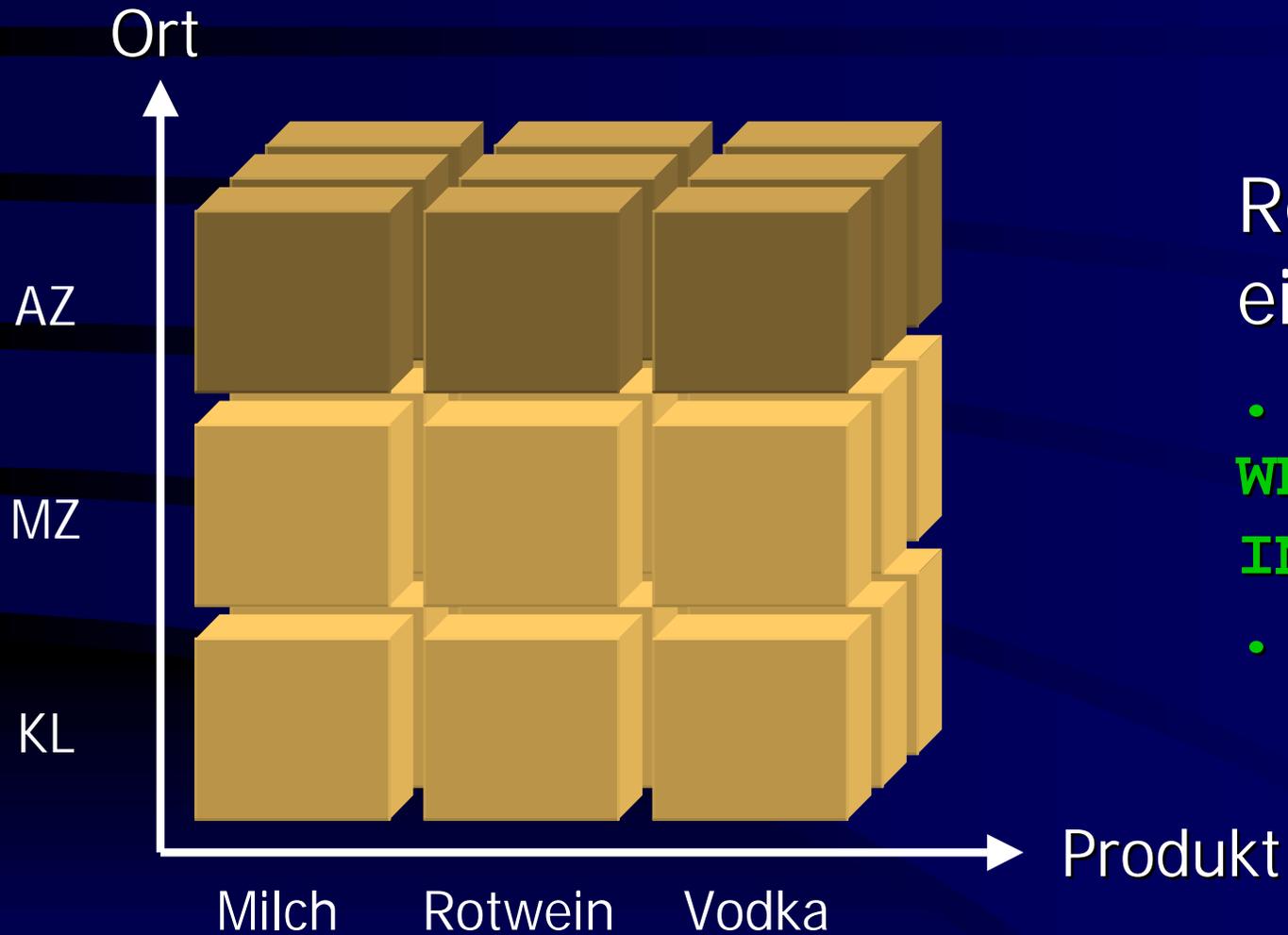
Restriktion auf
einer Dimension:

...

```
WHERE Ort  
IN (MZ, KL)
```

...

Slicing



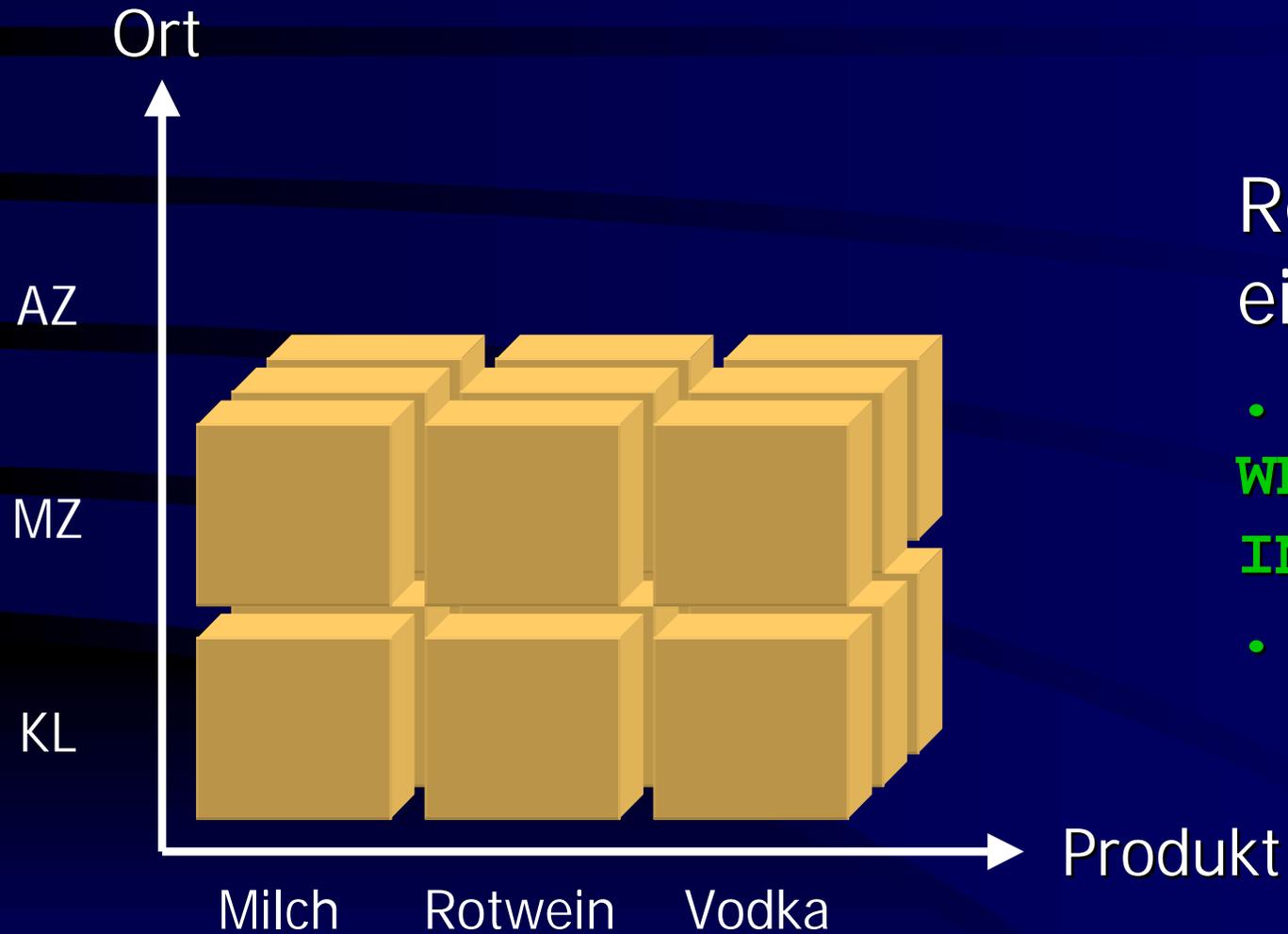
Restriktion auf
einer Dimension:

...

```
WHERE Ort  
IN (MZ, KL)
```

...

Slicing



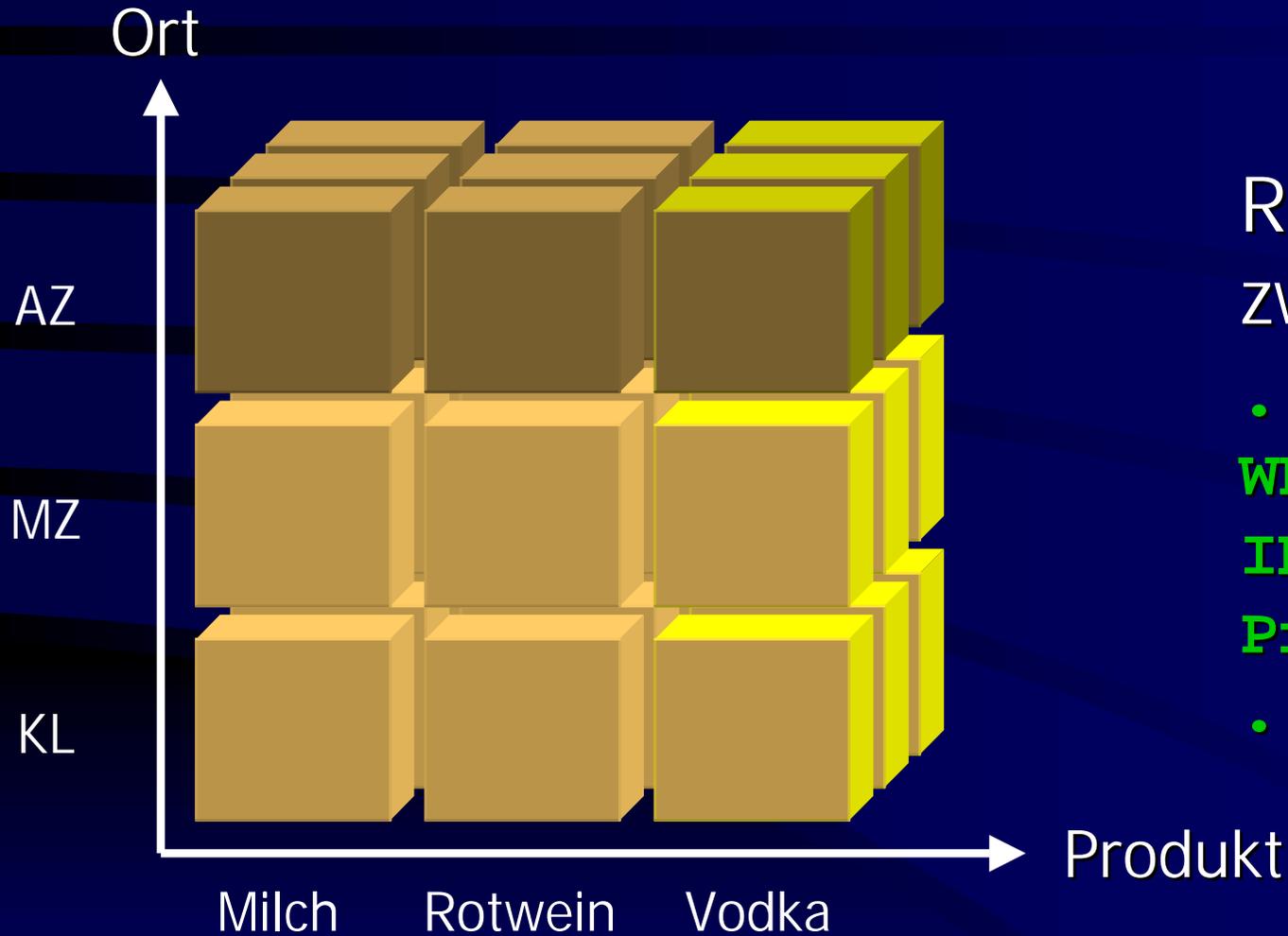
Restriktion auf
einer Dimension:

...

```
WHERE Ort  
IN (MZ, KL)
```

...

Dicing



Restriktion auf
zwei Dimensionen:

...

```
WHERE Ort  
IN (MZ, KL) AND  
Produkt <> Vodka
```

...

Dicing



Restriktion auf
zwei Dimensionen:

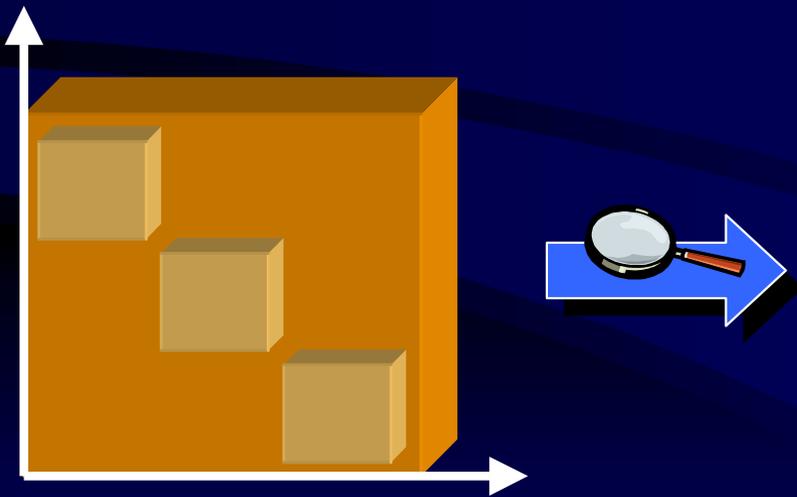
...

```
WHERE Ort  
IN (MZ, KL) AND  
Produkt <> Vodka
```

...

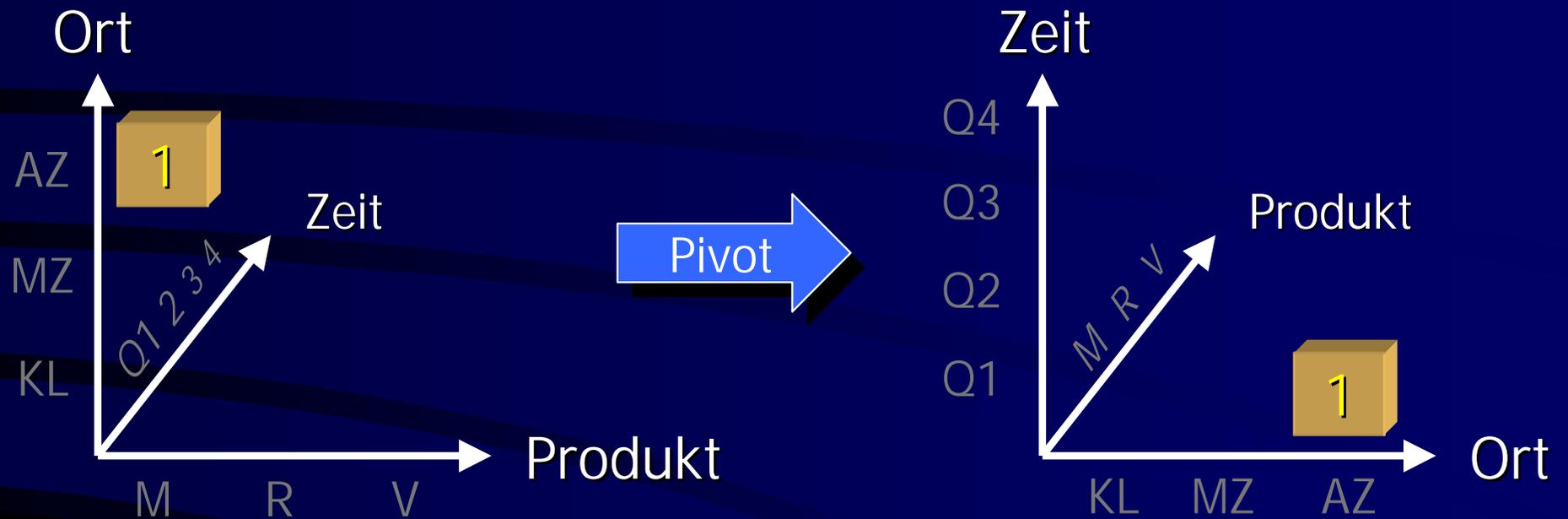
Pivoting

- Wunsch: Anpassung der Darstellung
- Weg: Vertauschen von Dimensionen



		Measures	
Products	Region	Measures[0]	Measures[1]
+Category[0]	-All Region[0]	1,017.41	787.35
	+Region[0]	1,019.25	1,123.19
	+Region[1]	877.00	1,148.39
	+Region[2]	752.89	1,069.97
	+Region[3]	829.57	916.46
	+Region[4]	920.09	1,017.18
+Category[1]	+All Region[0]	1,032.02	898.40
+Category[2]	+All Region[0]	963.97	1,016.30

Pivoting

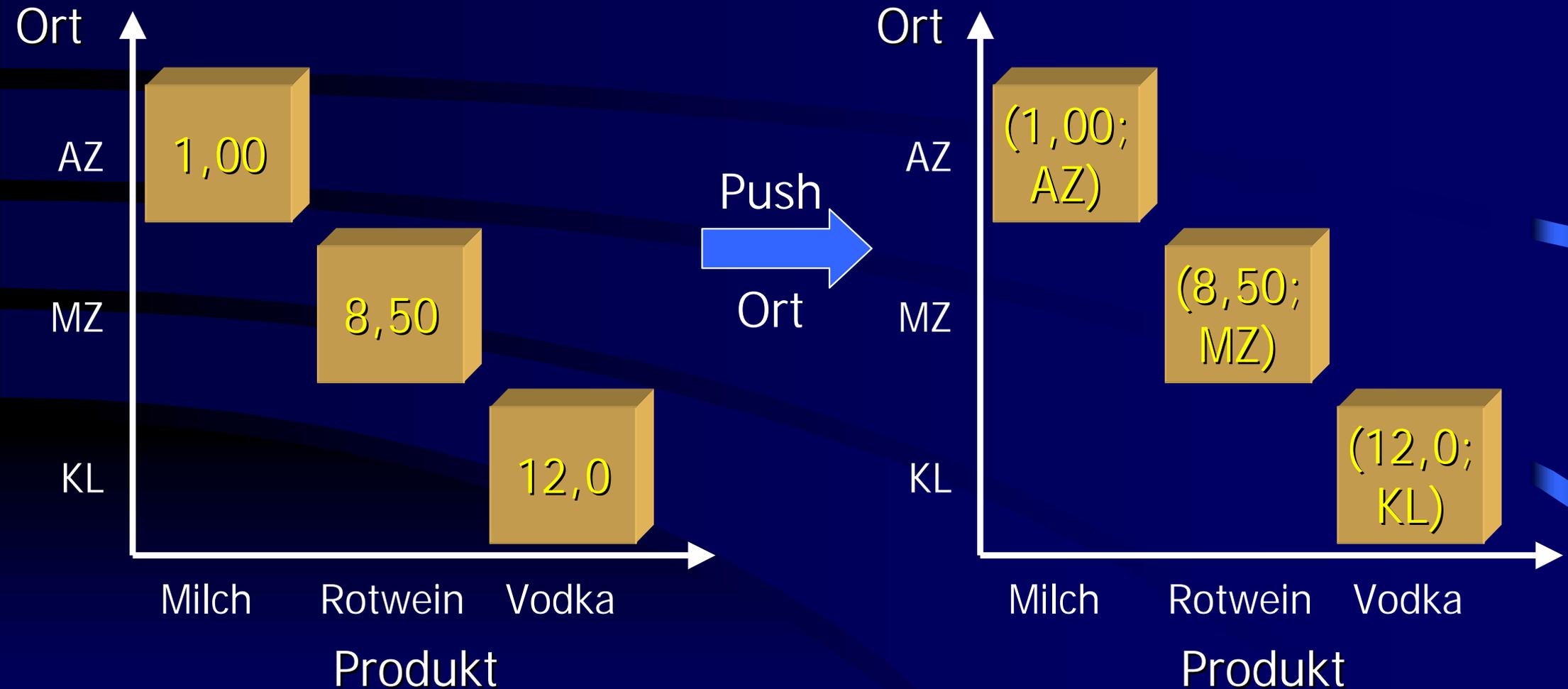


- Dimensionen und Maße bleiben gleich
- Anordnung ändert sich

Push und Pull

- Konvertierung Dimensionen \leftrightarrow Maße
- Symmetrie, Aggregate auf Dimensionen
- **Push:**
 - Dimensionswert in Maßtuplel einfügen
 - Dimension entfernen (optional)
- **Pull:**
 - neue Dimension erstellen
 - Dimensionswert aus Maßtuplel bilden

Push und Pull



Formale Grundlagen

- Datenmodell
 - formales Erfassen eines Hypercubes
 - kompakte Schreibweise
 - äquivalente Behandlung von Maßen und Dimensionen
- Operationen
 - übliche OLAP-Funktionalität
 - exakte Bestimmung möglicher Cube-Transformationen

Der Hypercube C

- Hypercube als 5-Tupel

$$C = (k, dn, dt, e, v)$$

- Menge aller Hypercubes

$$Cube = \mathbb{N}_+ \times DN \times DT \times E \times \mathbb{N}_0$$

$$DN = String^k$$

$$DT = Dom^k$$

$$E = dt_1 \times \dots \times dt_k$$

Der Hypercube C

$$C = (k, dn, dt, e, v)$$

Bedeutung der Komponenten:

- k* Anzahl der Attribute
- dn* Namen der Attribute
- dt* Datentypen der Attribute
- e* Werte der Attribute
- v* Attribute, die eine Dimension sind

Der Hypercube C

$$C = (k, dn, dt, e, v)$$

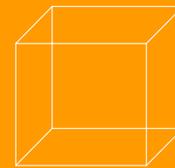
k Anzahl der Attribute

$$k \in \mathbb{N}_+$$

Gesamtanzahl aller Attribute im Hypercube. Sowohl Maße als auch Dimensionen enthalten.



Produkt, Ort, Zeit



Umsatz

Produkt, Ort, Zeit, Umsatz

$$k=4$$

Der Hypercube C

$$C = (k, dn, dt, e, v)$$

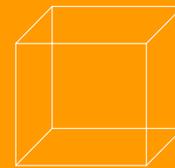
dn Namen der Attribute

$$dn \in DN = String^k$$

k-Tupel, das Namen aller Attribute enthält



Produkt, Ort,
Zeit



Umsatz

$$dn = (\text{Produkt}, \text{Ort}, \text{Zeit}, \text{Umsatz})$$

Der Hypercube C

$$C = (k, dn, dt, e, v)$$

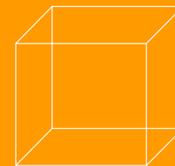
dt Typen der Attribute

$$dt \in DT = Dom^k$$

k-Tupel mit Datentypen
aller Attribute



Produkt, Ort,
Zeit



Umsatz

$$dt = (\textit{String}, \textit{String}, \textit{Date}, \textit{Float})$$

Der Hypercube C

$$C = (k, dn, dt, e, v)$$

e Rohdaten

$$e \in E = dt_1 \times \dots \times dt_k$$

Menge mit k -Tupeln, die
Attributwerte enthalten

$$e = \left\{ \begin{array}{l} ('Milch', 'Alzey', 1.10.03, 1.00), \\ ('Rotwein', 'Mainz', 1.11.03, 8.50), \\ ('Vodka', 'Kaiserslautern', 9.01.03, 12.00) \end{array} \right\}$$

Der Hypercube C

$$C = (k, dn, dt, e, v)$$

v Dimensionsspalten

$$v \in \mathbb{N}_0 \quad v \leq k$$

Attribute, die (von links gezählt) Dimension sind

Produkt, Ort, Zeit, Umsatz

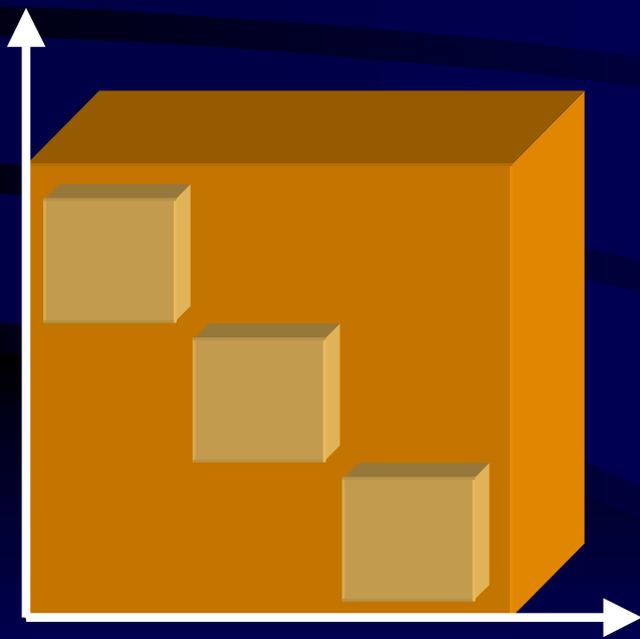
$v=3$

Dimensionen

Maß

Der Hypercube C

$$C = (k, dn, dt, e, v)$$



Sicht-
konfiguration

Daten

Metadaten

Operationen

- Menge von Inferenzregeln zur Manipulation von Hypercubes
- Ziel:
übliche OLAP-Funktionen darstellbar
- Operationen:
swap, restrict, transform, aggregate, copy,
push, pull, project

Swap-Operation

$$\text{swap} : \text{Cube} \times \mathbb{N}_+ \times \mathbb{N}_+ \rightarrow \text{Cube}$$

- $\text{swap}(C, i, j)$ vertauscht im Hypercube C die Dimensionen i und j .

$$\begin{array}{l} \frac{(k, (dn_1, \dots, dn_i, \dots, dn_j, \dots, dn_k), (dt_1, \dots, dt_i, \dots, dt_j, \dots, dt_k)) \dots}{(k, (dn_1, \dots, dn_j, \dots, dn_i, \dots, dn_k), (dt_1, \dots, dt_j, \dots, dt_i, \dots, dt_k)) \dots} \\ \dots \frac{\{(e_1, \dots, e_i, \dots, e_j, \dots, e_k), \dots\}, v)}{\{(e_1, \dots, e_j, \dots, e_i, \dots, e_k), \dots\}, v} \end{array}$$

- Entspricht der Pivot-Funktion

Swap-Operation

$dn = (\text{Ort}, \text{Produkt}, \text{Zeit})$

$dt = (\text{String}, \text{String}, \text{Date})$



$\text{swap}(\mathbf{C}, 1, 3)$



$dn = (\text{Zeit}, \text{Produkt}, \text{Ort})$

$dt = (\text{Date}, \text{String}, \text{String})$



$\text{swap}(\mathbf{C}, 1, 2)$



$dn = (\text{Produkt}, \text{Zeit}, \text{Ort})$

$dt = (\text{String}, \text{Date}, \text{String})$

(analog für e)

Restrict-Operation

$$\text{restrict} : \text{Cube} \times \mathbb{N}_+ \times \text{Praed} \rightarrow \text{Cube}$$

- $\text{restrict}(C, i, P)$ entfernt alle Elemente aus Hypercube C , für die Attribut i das Prädikat P nicht erfüllt.

$$\text{restrict} : \frac{(k, dn, dt, e, v)}{(k, dn, dt, \{\tilde{e} \mid \tilde{e} \in e \wedge P(\tilde{e}_i)\}, v)}$$

- Wird für Slicing und Dicing benötigt

Restrict-Operation

Sei $C \in \text{Cube}$ wie im Beispiel mit

$$e = \left\{ \begin{array}{l} ('Milch', 'Alzey', 1.10.03, 1.00), \\ ('Rotwein', 'Mainz', 1.11.03, 8.50), \\ ('Vodka', 'Kaiserslautern', 9.01.03, 12.00) \end{array} \right\}$$

Restriktion auf Umsatz $\rightarrow i=4$

Mit $P(u)$ gdw $u > 5$ ergibt $\text{restrict}(C, i, P)$

$$e' = \left\{ \begin{array}{l} ('Rotwein', 'Mainz', 1.11.03, 8.50), \\ ('Vodka', 'Kaiserslautern', 9.01.03, 12.00) \end{array} \right\}$$

Transform-Operation

$$\text{transform} : \text{Cube} \times \mathbb{N}_+ \times \mathcal{F}_{dn} \times \mathcal{F}_{dt} \times \mathcal{F}_e \rightarrow \text{Cube}$$

- $\text{transform}(C, i, f_{dn}, f_{dt}, f_e)$ wendet die Transformationsfunktionen f_{dn} , f_{dt} und f_e auf das Attribut i des Hypercube C an.

$$\frac{(k, (dn_1, \dots, dn_i, \dots, dn_k), (dt_1, \dots, dt_i, \dots, dt_k), \dots)}{(k, (dn_1, \dots, f_{dn}(dn_i), \dots, dn_k), (dt_1, \dots, f_{dt}(dt_i), \dots, dt_k), \dots)} \dots$$

$$\dots \frac{\{(e_{1,j}, \dots, e_{i,j}, \dots, e_{k,j}), \dots\}, v)}{\{(e_{1,j}, \dots, f_e(e_{i,j}), \dots, e_{k,j}), \dots\}, v)}$$

- Erster Teil eines Roll-Up-Vorgangs

Transform-Operation

- Anzugebende Funktionen:

$$\begin{array}{ll} f_{dn} : String \rightarrow String & f_{dn} : dn_i \mapsto \widetilde{dn}_i \\ f_{dt} : Dom \rightarrow Dom & f_{dt} : dt_i \mapsto \widetilde{dt}_i \\ f_e : dt_i \rightarrow \widetilde{dt}_i & f_e : e_i \mapsto \widetilde{e}_i \end{array}$$

f_{dn} transformiert Namen des Attributs
 f_{dt} transformiert Datentyp des Attributs
 f_e transformiert Werte

- Pro Hierarchieebene ein Satz Funktionen

Transform-Operation

- Transformationsfunktionen:

$f_{e_1 \rightarrow 2}$: *Mainz* \mapsto *Rheinhessen*,
Alzey \mapsto *Rheinhessen*,
Neustadt \mapsto *Pfalz*,
Kaiserslautern \mapsto *Pfalz*

$f_{dn_1 \rightarrow 2}$: *Stadt* \mapsto *Region*

$f_{dt_1 \rightarrow 2}$: *String* \mapsto *String*

- Transformation mit $i=2$ und f_* wie vor:
(‘Milch’, ‘Alzey’, 1.10.03, 1.00)
 \rightarrow *(‘Milch’, ‘Rheinhessen’, 1.10.03, 1.00)*,

Aggregate-Operation

$$\text{aggregate} : \text{Cube} \times \mathcal{F}_{\text{aggr}}^{k-v} \rightarrow \text{Cube}$$

- Aggregiert alle Tupel mit identischen Dimensionswerten
- Eine Aggregatsfunktion pro Maß

- Zweiter Teil eines Roll-Up-Vorgangs

Aggregate-Operation

$$e = \left\{ \begin{array}{l} ('Milch', 'Alzey', 1.10.03, 1.00), \\ ('Milch', 'Alzey', 1.11.03, 1.00), \\ ('Vodka', 'Kaiserslautern', 9.01.03, 12.00) \end{array} \right\}$$



transform(...)

$$e' = \left\{ \begin{array}{l} ('Milch', 'Alzey', Q4, 1.00), \\ ('Milch', 'Alzey', Q4, 1.00), \\ ('Vodka', 'Kaiserslautern', Q1, 12.00) \end{array} \right\}$$



aggregate(C', (Σ))

v=3

$$e'' = \left\{ \begin{array}{l} ('Milch', 'Alzey', Q4, 2.00), \\ ('Vodka', 'Kaiserslautern', Q1, 12.00) \end{array} \right\}$$

Push- und Pull-Operation

$push : Cube \rightarrow Cube$

$pull : Cube \rightarrow Cube$

$push(C)$ verwandelt
Dimension v in Maß

$pull(C)$ verwandelt
Maß v in Dimension

$push : \frac{(k, dn, dt, e, v)}{(k, dn, dt, e, v - 1)}$

$pull : \frac{(k, dn, dt, e, v)}{(k, dn, dt, e, v + 1)}$

- Entspricht nicht den gewohnten OLAP-Operationen für Push und Pull.

Push- und Pull-Operation

$dn = (\text{Produkt}, \text{Zeit}, \text{Ort}, \text{Umsatz}) \quad v = 3$



$dn = (\text{Produkt}, \text{Zeit}, \text{Ort}, \text{Umsatz}) \quad v = 2$



$dn = (\text{Produkt}, \text{Zeit}, \text{Ort}, \text{Umsatz}) \quad v = 1$



$dn = (\text{Produkt}, \text{Zeit}, \text{Ort}, \text{Umsatz}) \quad v = 2$

Ausdrucksmächtigkeit

- Pivoting

$swap \circ \dots \circ swap \circ C$

- Slicing

$restrict \circ C$ mit $P(u)$

- Dicing

$restrict \circ \dots \circ restrict \circ C$

- Push

$push \circ swap \circ copy \circ C$

Ausdrucksmächtigkeit

- Pull

$$pull \circ C$$

- Roll-Up

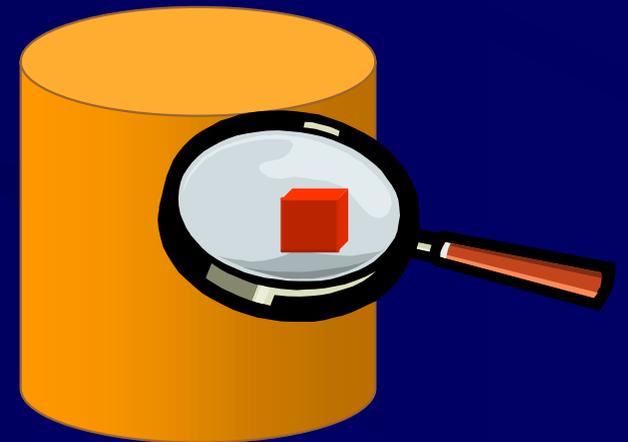
$$aggregate(f_{aggr}^*) \circ transform(f_{dn}, f_{dr}, f_e) \circ C$$

- Drill-Down

korrespondierenden Roll-Up aus der Transformationskette entfernen

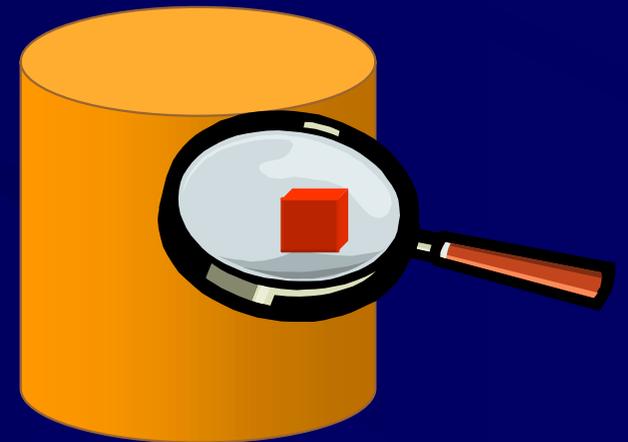
Zusammenfassung

- Einführung in OLAP
- Operationen anschaulich
- Formales Datenmodell mit gleicher Ausdrucksmächtigkeit

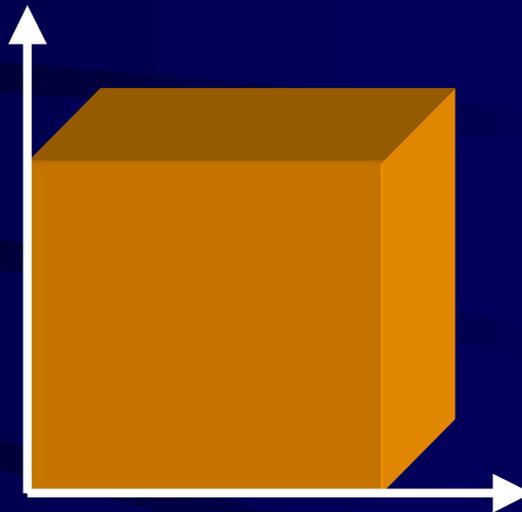


Ausblick

- Intelligent Roll-Ups
 - Generalisieren eines Problems
 - Untersuchen aller Roll-Up-Pfade, um Problem einzugrenzen
- Werkzeugunterstützung
- Anfragesprachen
- OLAP über XML



Logische Modelle für OLAP



Burkhard Schäfer

