

# Microsoft SQL Server 2005 Analysis Services

## 2. SQL Server Entwickler Konferenz

Thomas Trefz

- Vom Data Warehouse zur Analyse
- Das Gerüst: Dimensionen und Hierarchien
- Flexibel: Die Parent-Child Hierarchie
- Mehrdimensional: Der Cube
- Der Kern: Kennzahlen
- MDX: Abfragen und Funktionen
- Individuell: Berechnete Elemente
- Diskussion

- Vermittlung eines Grundwissens...
- ... denn Analysis Services in 3 Stunden ?
- Mehr Praxisanteile wie Folien...
- ... aber die Basiskenntnisse sind wichtig!
- Fragen sind nicht nur am Ende möglich
- Lachen ist Gesund und daher erlaubt
- „Viele Wege führen nach Rom“
- Skript und Beispiele über die SEK Web Site

- Ein kleines Demo: Adventure Works DW
- Ein noch kleineres Beispiel: STONE

- Analysis Services sind Bestandteil von
  - Standard und Enterprise Edition
- Data Warehousing hat viele Einsatzgebiete
  - Und ist nicht abhängig von
    - Unternehmensgröße
    - Datenbankgröße
    - ...
- Projekte bestehen im wesentlichen aus
  - Reden, Reden, Reden, ....

- eine dauerhafte, integrierte Sammlung von Daten aus unterschiedlichen Quellen zum Zweck der Analyse bzw. Entscheidungsunterstützung  
[Chaudhuri & Dayal 1997]
- eine themenorientierte, integrierte, zeit-variante und permanente Sammlung von Daten zur Entscheidungsunterstützung  
[Inmon 1996]

- Zweck des DW ist nicht Erfüllung einer dedizierten Aufgabe (z.B. Personaldatenverwaltung), sondern Unterstützung übergreifender Auswertungsmöglichkeiten aus verschiedenen Perspektiven
- Das DW enthält alle Daten - unternehmensweit - über ein Subjekt (Kunden, Produkte, Regionen ...) und nicht „versteckt“ in verschiedenen Anwendungen

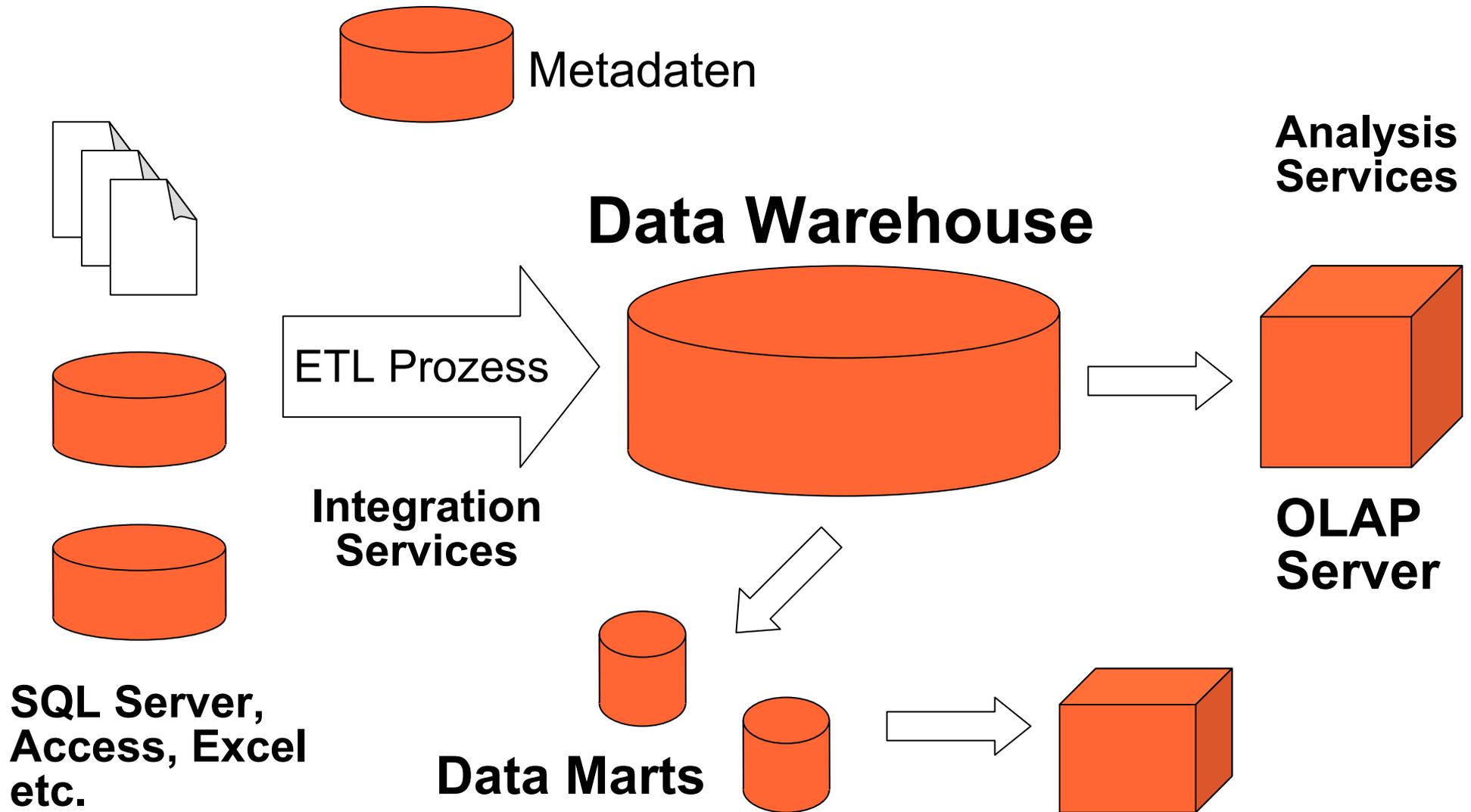
- Integrierte Datenbasis (**integrated**):  
Daten aus mehreren verschiedenen, meist heterogenen Datenquellen zusammengefasst und vereinheitlicht (z.B. Anschriften Strasse - Straße – Str.)
- Überwindung der Heterogenität auf verschiedenen Ebenen (System, Schema, Daten)

- Nicht-flüchtige Datenbasis (**non-volatile**):  
Vorhandene Daten im DW werden i.a. nicht mehr geändert, das Hinzufügen neuer Daten ist natürlich notwendig. Stabile, persistente Datenbasis.

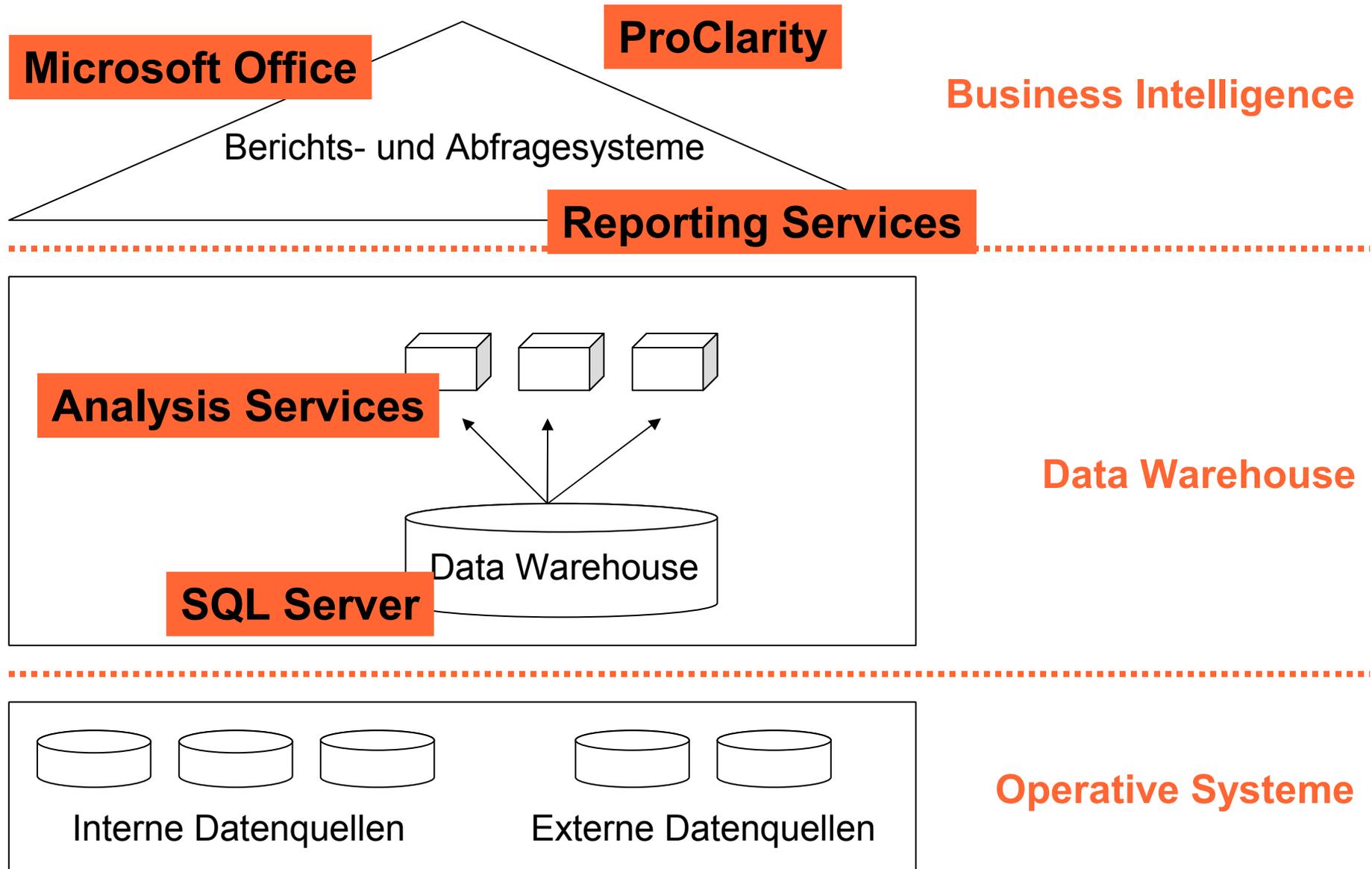
- Historische Daten (**time-variant**):  
ein Vergleich der Daten über die Zeit ist möglich (Zeitreihenanalyse). Die Speicherung erfolgt über längeren Zeitraum (2 – 10 Jahre). Alle Objekt im DW haben einen Zeitbezug.

- **Inmon**: Das DW ist ein Teil des Business Intelligence Systems. Es gibt ein DW welches die Informationen in der 3. Normalform speichert. Data Marts erhalten ihre Daten aus dem DW.
- **Kimball**: Ein DW ist eine Sammlung von Data Marts. Informationen werden ausschließlich im Dimensional Model gespeichert.

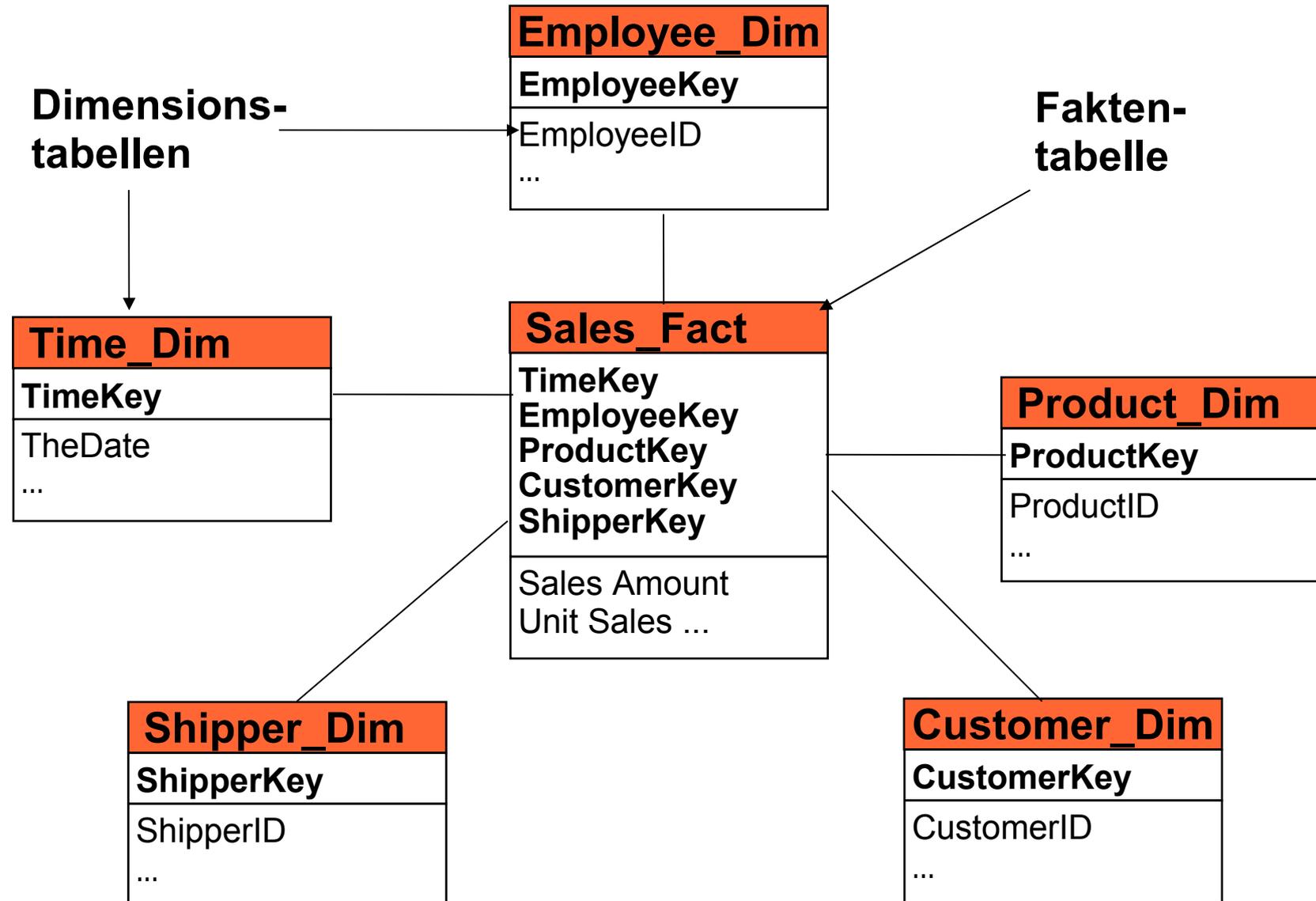
# Grundstruktur (1)



# Grundstruktur (2)



# Data Warehouse Inside



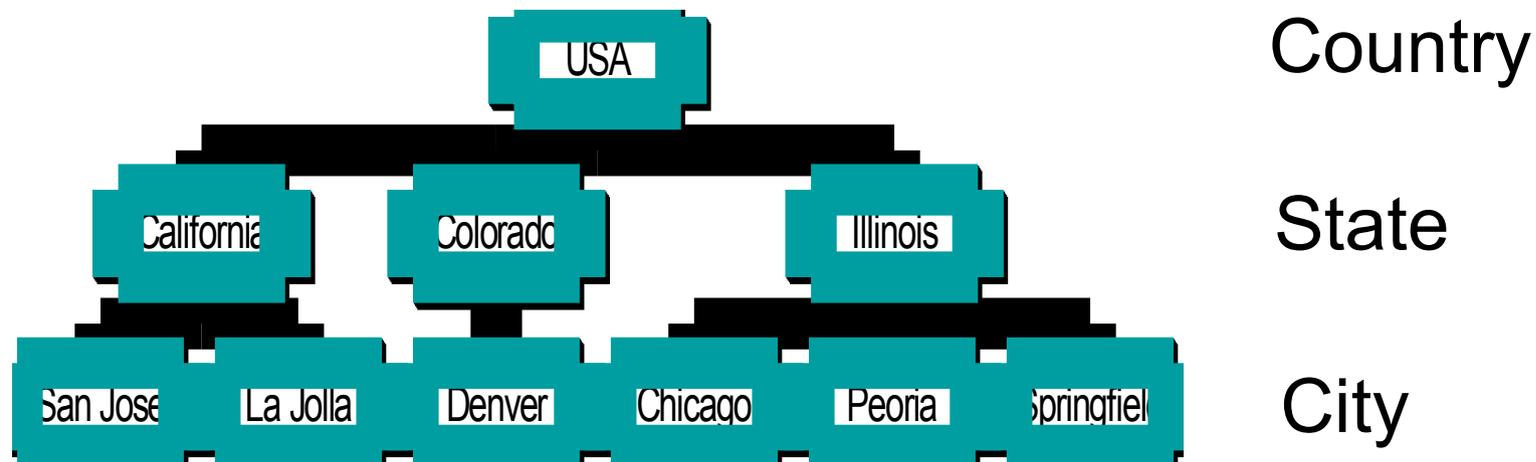
<i>customer_key</i>	<i>product_key</i>	<i>time_key</i>	<i>quantity_sales</i>	<i>amount_sales</i>
201	25	134	400	10,789

Die **Granularität** der **Sales\_Fact**-Tabelle wird durch die unterste Detailschicht definiert, die in allen (!) Dimensionen gespeichert ist

Alle Datensätze in der Tabelle haben die gleiche Granularität (!)

- Sind die „Achsen“ des Cubes
- In der Regel über 3 Dimensionen pro Cube
- Bilden die **Aggregationsebenen** des Cubes
- Werden aus den **Dimensionstabellen** im Data Warehouse gebildet.
- Dimensionen besitzen **Hierarchien**
- Hierarchien besitzen 1 oder mehrere **Ebenen**
- Ebenen beinhalten **Elemente**

- Die Inhalte der Dimensionen sind Daten die sich selten ändern (!)
- Kimball: **Slowly Changing Dimensions**
- Für jede Dimension die Änderungsstrategie festlegen (Typisierung nach Kimball):
  - Typ 1: Überschreiben der Daten in der Dimensionstabelle
  - Typ 2: Datensatz in der Dimensionstabelle duplizieren



- Jede Ebene einer Dimension entspricht einer Spalte in der Dimensionstabelle
- Jedes Element einer Dimensionsebene besitzt dieselbe Anzahl von Vorgängern

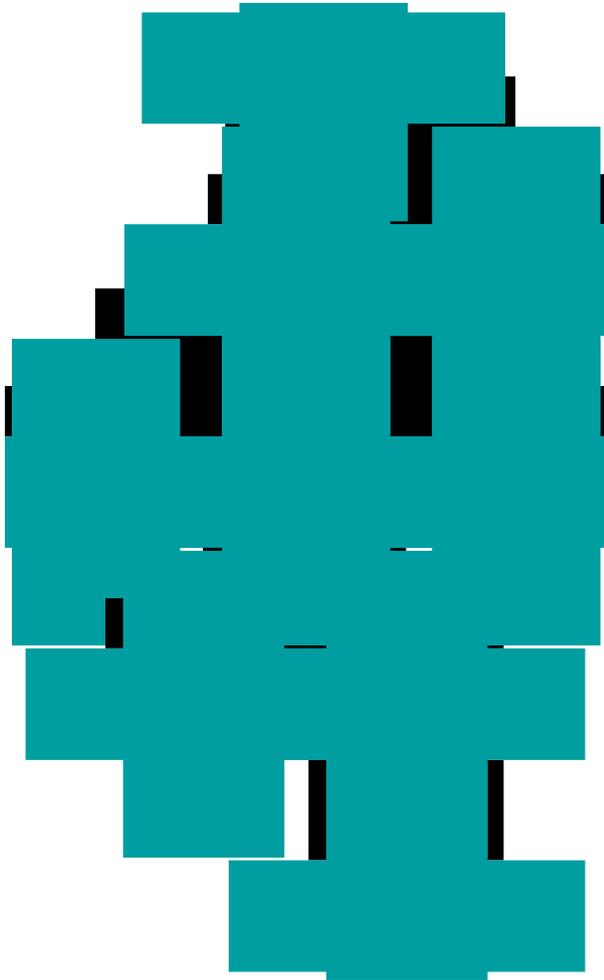
# Unregelmäßige Hierarchien



- logisch übergeordnetes Objekt befindet sich nicht auf der Ebene unmittelbar über dem betreffenden Element

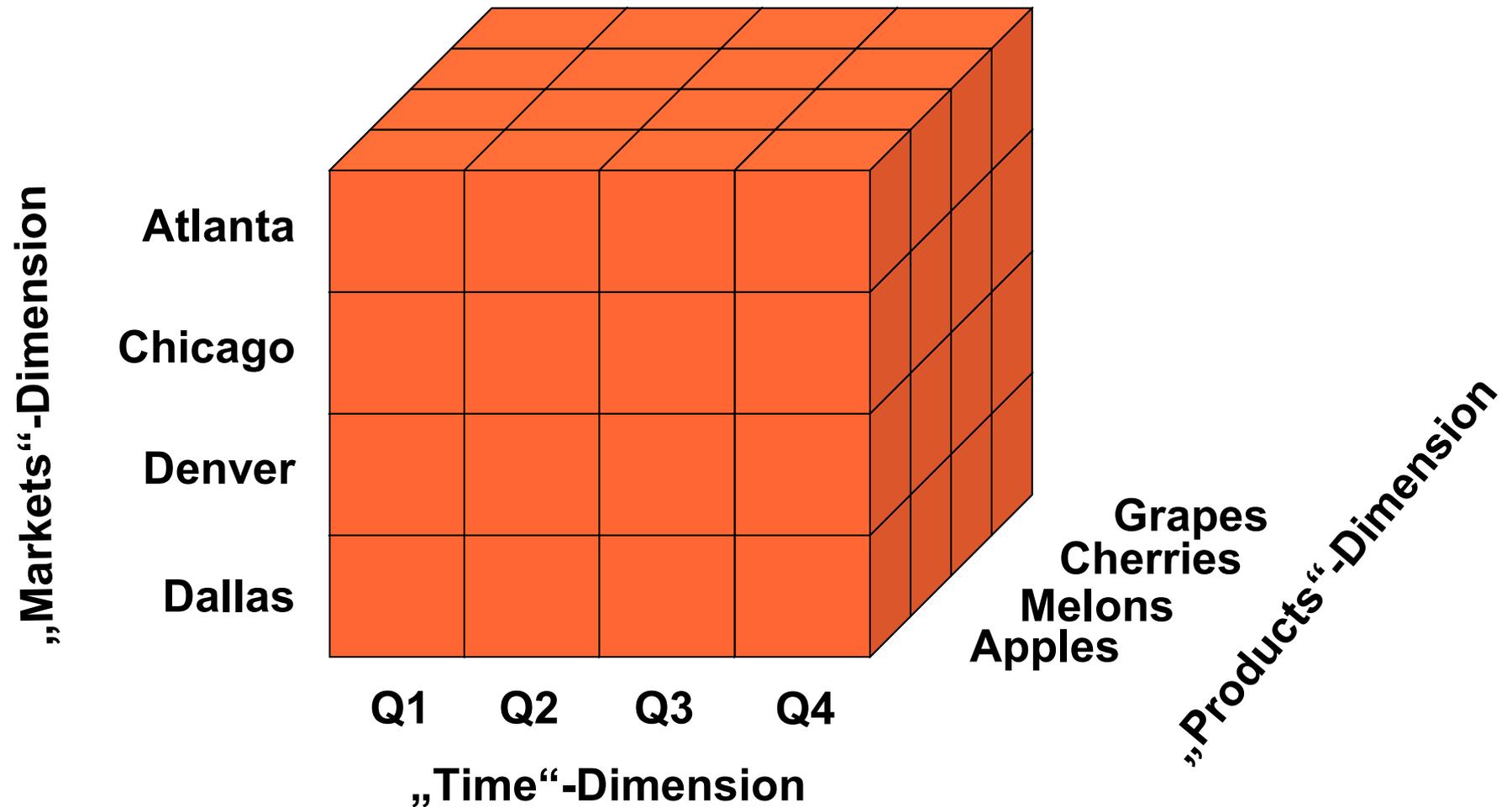
- Basis ist eine selbst referenzierende Tabelle
- Nicht jedes Element der Hierarchie besitzt die gleiche Anzahl von übergeordneten Elementen
- Es können Datenwerte auf aggregierten Ebenen existieren!
- Aggregationen werden zur Laufzeit berechnet!
- UNARY Operator für individuelle Aggregation

# Parent-Child Hierarchie (2)



Employee	Manager
Smith	<none>
Jones	Smith
White	Smith
Block	Jones
Hart	Jones
Knight	Jones
Fox	Hart
Hunt	Hart
Smart	Hunt

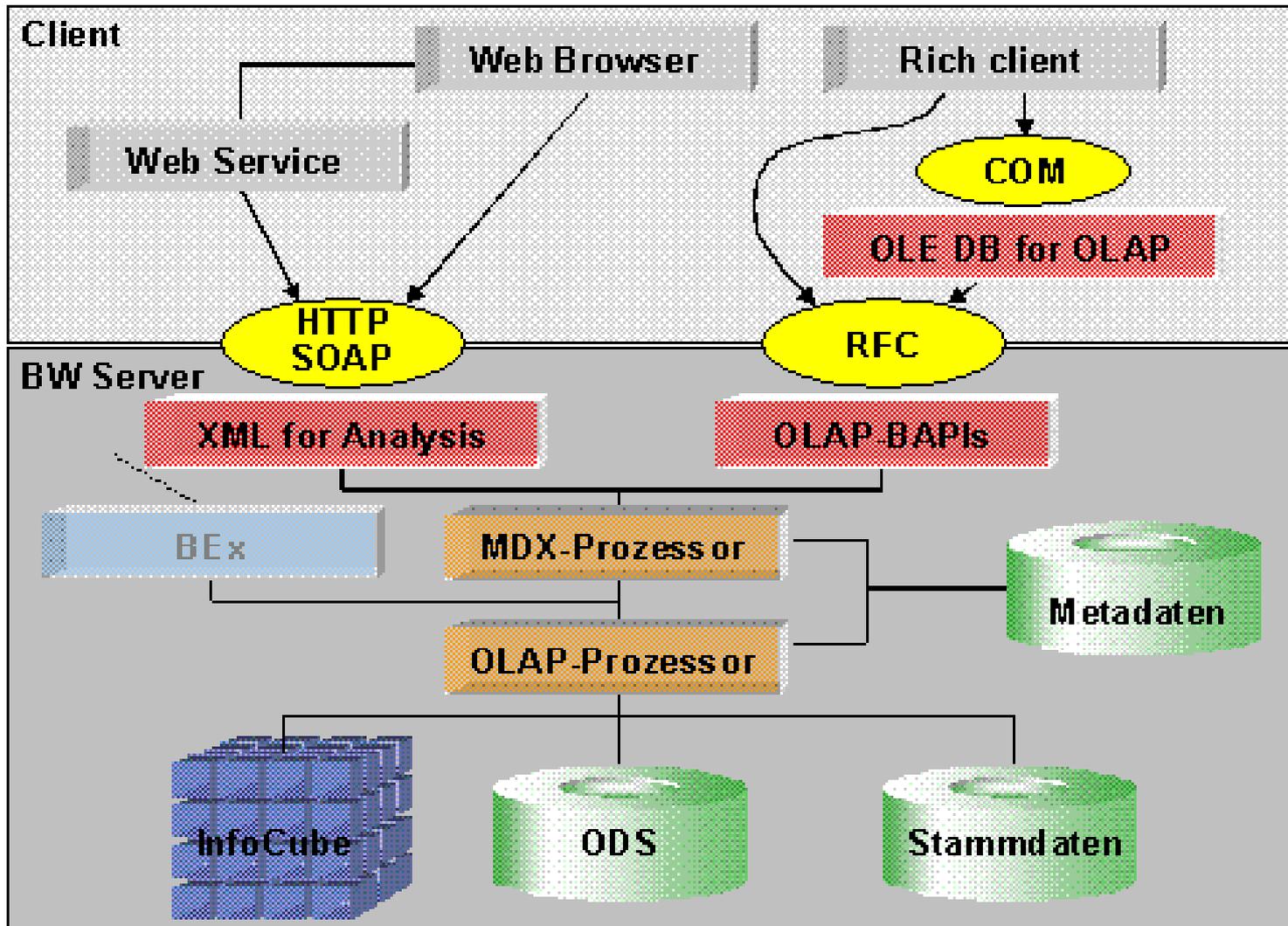
# Der Cube



- Sind **numerische** Daten
- Entsprechen den **Fakten** in der Faktentabelle
- Liegen mit allen Dimensionen auf allen Ebenen in einer Schnittmenge
- Bilden eine **eigene Dimension!**
- Vorsicht: nicht jede Kennzahl ist **kumulativ**
  - Beispiel: Kreditstand (semi-additiv)
  - Beispiel: Durchschnittspreis (gar nicht additiv)

- MDX: Multidimensional Expressions
- MDX ist keine Erweiterung von SQL (!)
  - SQL ist zwei-dimensional: Zeilen und Spalten
  - Gemeinsamkeit: Select ... From ... Where; mehr nicht!
- MDX gibt es nicht nur bei Microsoft
  - Hyperion Essbase
  - SAS
  - Pentaho Analysis Services: Mondrian Project
  - SAP BW
- Real Life: Ohne MDX geht nix!

# MDX und SAP BW



Quelle: SAP-Bibliothek – Zusätzliche Entwicklertechnologien

- Werden zur Abfrage von Cubes verwendet
- SQL Server Management Studio: MDX Editor
- Anwendungen:
  - OLEDB for OLAP
  - ADOMD.NET
- Grundstruktur:
  - SELECT ... FROM ... WHERE
- Problem Ergebnisanzeige:
  - Tabellarische Darstellung: zwei-dimensional (!)

- Ergebnis ist ein skalarer Wert
  - Oder aber: ein Set (Named Sets in Analysis Services)
- Schwerpunkt: berechnete Measures
  - Im Cube
  - In Abfragen

- Punkt-Bestimmung
  - Zwei-Dimensional: (x,y)
  - Drei-Dimensional: (x,y,z)
- Vollständiges Tupel
  - Ein Tupel definiert einen Punkt im Cube durch Angabe von Elementen aller (!) Dimensionen.
- Unvollständiges Tupel
  - Fehlende Dimensionsangaben werden durch die aktuellen Elemente (.CurrentMember) ersetzt.

- Eine Menge von Tupeln gleicher Dimensionalität
- CROSSJOIN Methode ist wichtiges Element in der dynamischen Erzeugung von Sets
  - Beispiel: Darstellung von Elementen der Dimension Zeit und Kunden auf einer Achse bei einer zwei-dimensionalen Ausgabe

- **Berechnete Elemente**
  - Berechnete Measures (z.B. Durchschnittspreis)
  - Aber auch: Elemente auf anderen Dimensionen
- **MDX Script**
  - Umfangreiche Manipulation der Berechnungen bis zur Berechnung einzelner Zellen
- **MDX lässt sich individuell erweitern**
  - Einbinden von .NET Assemblies

- **Microsoft SQL Server 2005 Analysis Services**  
SAMS Publishing  
ISBN: 0-672-32782
- **Professional SQL Server Analysis Services 2005**  
Wrox / Wiley  
ISBN: 0-7645-7918-5
- **MDX Solutions (Second Edition)**  
Wiley  
ISBN: 0-471-74808-0
- **Business Intelligence und Reporting mit SQL Server 2005**  
Microsoft Press  
ISBN: 3-86063-994-3
- Und viele andere....

# Diskussion

---