

Dimensionale Modellierung mit Oracle BI EE und Oracle OLAP - Tipps und Tricks aus der Praxis

Claus Jordan
Trivadis GmbH
Stuttgart

Schlüsselworte:

Business Intelligence
Oracle Business Intelligence Enterprise Edition
Oracle OLAP
Dimensionale Modellierung

Einleitung

Es gibt für eine Problemstellung im BI-Umfeld selten *die eine* richtige Lösung. Dieser Beitrag zeigt anhand einfacher, praxisnaher Beispiele, wie mit *Oracle Business Intelligence - Enterprise Edition* (kurz OBIEE) - mit und ohne *Oracle OLAP Option* im Backend - verschiedene Endanwender-Sichten auf dieselben Daten implementiert werden können.

Aufgrund der Vielzahl von möglichen Sichten sollen nur zwei Dimensionen und deren Kombination untereinander betrachtet werden:

- Kennzahlen wie *Umsatz, Absatz*, usw.
- Zeitabschnitte wie z.B. *Aktuelles Jahr, Vorjahr, Abweichung zum Vorjahr*, usw.

Diese verschiedenen Sichten auf BI relevante Informationen sind notwendig, weil unterschiedliche Endanwendergruppen häufig unterschiedliche Anforderungen bezüglich der benötigten Analysepfade und Auswertungslayouts haben. Es kommt aber auch vor, dass ein und derselbe Endanwender die Daten aus verschiedenen Blickwinkeln analysieren möchte. Die Abbildung 1 zeigt ein Beispiel für eine häufig geforderte Kreuztabellen-Auswertung.

| | Actual Period | Prev. Year | Var. PER-PY | Var. % PER-PY |
|-------------|---------------|--------------|--------------|---------------|
| Measure | | | | |
| Units | 34.950,00 | 27.670,00 | 7.280,00 | 26,31 |
| Sales | 8.400.439,78 | 7.056.994,53 | 1.343.445,25 | 19,04 |
| Costs | 7.787.159,48 | 6.533.568,28 | 1.253.591,20 | 19,19 |
| Margin | 613.280,30 | 523.426,25 | 89.854,05 | 17,17 |
| Margin in % | 7,30 | 7,42 | -0,12 | -1,57 |

Abb. 1: Kreuztabelle – Kennzahlen und Zeitabschnitte als Dimension. On-the-fly Berechnungen sind markiert. Die Reihenfolge der Berechnungen spielt- zumindest für die grün markierten Zellen - eine Rolle

In dieser Kreuztabelle werden für einen ausgewählten Monat in den Zeilen verschiedene Kennzahlen wie *Units, Sales, Costs* dargestellt. In den Spalten werden verschiedenen Zeitabschnitte wie *Actual Period, Previous Year (PY)* gegenüber gestellt. Zwischen den Zeilen- und Spaltenausprägungen werden jeweils Berechnungen (Differenzen und/oder Relationen) dargestellt. Selbstverständlich liegen diese Berechnungen nicht im Quellsystem, auch nicht im Data Warehouse vor, da es sich zumindest

bei den Prozentwerten um nicht aggregierbares Zahlenmaterial handelt und eine Summierung oder ungewichtete Durchschnittsbildung zu falschen Ergebnissen führen würde. Stattdessen muss deren Berechnung *on-the-fly*, d.h. zur Laufzeit während der Abfrage durch den Endanwender, erfolgen.

Wie kann nun mit OBIEE Bordmittel die Kreuztabelle aus Abbildung 1 implementiert werden?

Die Abbildungen 2 und 3 zeigen zwei naheliegende Lösungsansätze für die Implementierung innerhalb der Metadaten.

Lösungsansatz A:
Je Kennzahl-Zeitabschnitt eine eigene Attributspalte

The screenshot shows the OBIEE metadata editor interface. On the left, a tree view shows the metadata structure for 'F03 Ref Facts with Measures and Times'. A red box highlights the 'Sales PER' column in the table definition. A red arrow points from this box to the 'Sales PER' column in the table definition. Below the table definition, a smaller table shows a summary of measures for 'JAN-03'.

| Measure | Actual Period | Target | Var. % PER-PY | Var. % PER-PY |
|-------------|---------------|--------------|---------------|---------------|
| Units | 34.950,00 | 37.000,00 | 7.280,00 | 26,31 |
| Sales | 8.400.440,38 | 8.155.000,00 | 245.440,38 | 29,94 |
| Costs | 7.787.159,19 | 6.533.568,28 | 1.253.590,91 | 19,19 |
| Margin | 613.280,20 | 523.436,25 | 89.843,95 | 17,17 |
| Margin in % | 7,30 | 7,42 | -0,12 | -1,62 |

- Abgeleitete Kennzahlen, z.B. „Margin“ sind einfach in den Metadaten oder im Frontend zu definieren ☺
- Die geforderte Auswertung (siehe Problemstellung) kann so nicht definiert werden ☹
 - Weil eine Trennung von Kennzahl und Zeitabschnitt nicht möglich ist

DOAG 2010 - Dimensionale Modellierung mit OBIEE 9

Abb. 2: Je Kennzahl-Zeitabschnitt-Kombination wird in den OBIEE Metadaten eine eigene Attributspalte definiert

Lösungsansatz B: Kennzahl-Zeitabschnitt als Dimensionsausprägung

| | Actual Period | Prev. Year | Var. PER-PY | Var. % PER-PY |
|-------------|---------------|--------------|--------------|---------------|
| Units | 34.950,00 | 27.670,00 | 7.280,00 | 26,31 |
| Sales | 8.400.439,78 | 7.056.994,53 | 1.343.445,25 | 19,04 |
| Costs | 7.787.159,48 | 6.533.568,28 | 1.253.591,20 | 19,19 |
| Margin | 613.280,30 | 523.426,25 | 89.854,05 | 17,17 |
| Margin in % | 7,30 | 7,42 | -0,12 | -1,57 |

- Hohe Flexibilität beim Aufbau von Auswertungen
 - Trennung von Kennzahl und Zeitreihe möglich ☺
 - Abgeleitete Kennzahlen (Margin, Previous Year) können im Frontend definiert werden
- Formatierung für alle Kennzahlen gleich ☹

Abb. 3: Eine Attributspalte für die Daten, Kennzahlen und Zeitabschnitte als Dimension

Der Lösungsansatz B führt im Gegensatz zu A zum Ziel, d.h. die geforderte Kreuztabellen-Auswertung aus Abbildung 1 kann den Endanwendern zur Verfügung gestellt werden. In den folgenden beiden alternativen Lösungsvorschlägen wird deshalb lediglich der Lösungsansatz B weiter verfolgt.

Lösungsvorschlag 1 (ohne Oracle OLAP)

Dieser Lösungsvorschlag basiert auf rein relationaler Technologie (ROLAP) im Backend. Basis ist ein dimensionales Star- oder Snowflake-Schema. Die Faktentabelle enthält nur die Kennzahlen *Sales*, *Units* und *Costs*.

Schritt 1: Transformieren und Erweitern der Faktentabelle

In diesem vorgelagerten Schritt muss die zugrundeliegende Faktentabelle *in Form* gebracht werden. In diesem Beispiel werden die drei Kennzahlen in dem Attribut *Measure* gespeichert. Gleichzeitig werden zu jedem Datensatz die Vorjahresdaten ermittelt und über das Attribut *Timeshift* spezifiziert (siehe Abbildung 4).

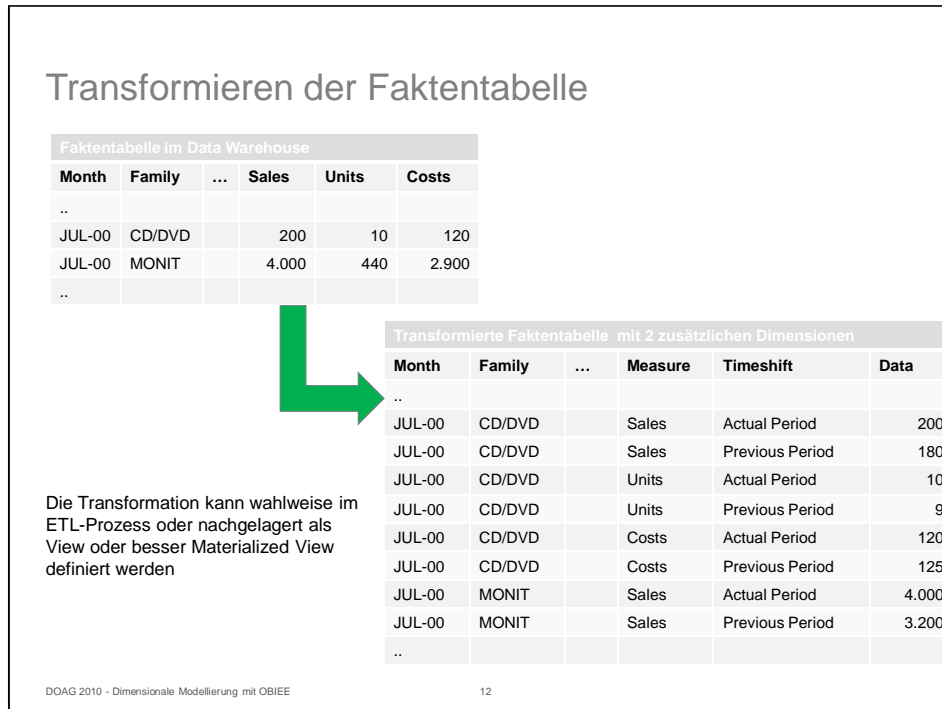


Abb. 4: Physische oder logische Transformation der Faktentabelle (auf künstliche Schlüssel wurde in der Darstellung verzichtet)

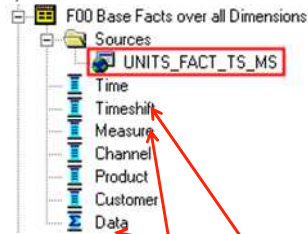
Die neue Faktentabelle bzw. -view kann nun in OBIEE als Datenquelle referenziert werden. Natürlich existieren für die beiden neuen Attribute *Measure* und *Timeshift* entsprechende Dimensionstabellen.

Schritt 2: Definition des *Business Model and Mapping Layer (BMM)* in OBIEE

Mit Hilfe des OBIEE Administration Tools kann im OBIEE Metadatenrepository die neue Faktentabelle bzw. -view sowie die Dimensionstabellen für *Measure* und *Timeshift* in das *Physical Layer* eingefügt werden. Im *BMM* werden dann mehrere sogenannte Sichten auf die o.g. transformierte Faktentabelle definiert. Bereits mit der ersten Sicht (Abbildung 5) kann die geforderte Kreuztabellen-Auswertung (Abbildung 1) erstellt werden. Die weiteren Sichten (Abbildungen 6 bis 8), welche jeweils die erste Sicht mit Hilfe der *FILTER*-Funktion referenzieren, können optional definiert werden um den Endanwendern weitere Alternativen anzubieten.

BMM - Definition (1)

1. Anlegen einer neuen logischen Faktentabelle „F00“.
2. Die Spalten referenzieren die transformierten Faktentabelle



Transformierte Faktentabelle mit 2 zusätzlichen Dimensionen

| Month | Family | ... | Measure | Timeshift | Data |
|--------|--------|-----|---------|-----------------|-------|
| .. | | | | | |
| JUL-00 | CD/DVD | | Sales | Actual Period | 200 |
| JUL-00 | CD/DVD | | Sales | Previous Period | 180 |
| JUL-00 | CD/DVD | | Units | Actual Period | 10 |
| JUL-00 | CD/DVD | | Units | Previous Period | 9 |
| JUL-00 | CD/DVD | | Costs | Actual Period | 120 |
| JUL-00 | CD/DVD | | Costs | Previous Period | 125 |
| JUL-00 | MONIT | | Sales | Actual Period | 4.000 |
| JUL-00 | MONIT | | Sales | Previous Period | 3.200 |

DOAG 2010 - Dimensionale Modellierung mit OBIEE

Abb. 5: Logische Faktentabelle F00 mit Dimension Measure und Timeshift

BMM – optionale Definition (2)



1. Anlegen einer weiteren logischen Faktentabelle „F01“.
2. Die Spalten „Sales“, „Units“ und „Costs“ referenzieren die Spalte „Data“ der ersten logischen Faktentabelle „F00 ..“
3. Margin und Margin % dagegen berechnen sich aus den Spalten „Sales“ und „Costs“ von „F01 ..“

Logical Column - Sales

General | Data Type | Aggregation | Levels

Name: Sales

Belongs to Table: ["GLOBAL_RELATIONAL"."F01_Ref_Facts_with_Measures_in_Columns"]

Sort order column: None

Use existing logical columns as the source

FILTER (GLOBAL_RELATIONAL.F00_Base_Facts_over_all_Dimensions.Data USING GLOBAL_RELATIONAL.D03_MEASURE_DIM.MEASURE_ID = 'SALES')

Logical Column - Margin

General | Data Type | Aggregation | Levels

Name: Margin

Belongs to Table: ["GLOBAL_RELATIONAL"."F01_Ref_Facts_with_Measures_in_Columns"]

Sort order column: None

Use existing logical columns as the source

GLOBAL_RELATIONAL.F01_Ref_Facts_with_Measures_in_Columns.Sales - GLOBAL_RELATIONAL.F01_Ref_Facts_with_Measures_in_Columns.Costs

DOAG 2010 - Dimensionale Modellierung mit OBIEE

14

Abb. 6: Logische Faktentabelle F01 dimensioniert nach Dimension Timeshift

BMM – optionale Definition (3)

The screenshot shows the 'Logical Column - PER' configuration panel. The 'Name' field is 'PER'. It belongs to the table 'GLOBAL_RELATIONAL.F02_Ref_Facts_with_Timeshift_in_Columns'. The 'Use existing logical columns as the source' checkbox is checked. The source filter is: `FILTER(GLOBAL_RELATIONAL.F00_Base_Facts_over_all_Dimensions.Data USING GLOBAL_RELATIONAL.D02_TIMESHIFT_DIM.TS_ID = 'PER')`.

The 'Logical Column - Var. PER-PY' configuration panel shows the 'Name' field as 'Var. PER-PY'. It also belongs to the table 'GLOBAL_RELATIONAL.F02_Ref_Facts_with_Timeshift_in_Columns'. The 'Use existing logical columns as the source' checkbox is checked. The source filter is: `GLOBAL_RELATIONAL.F02_Ref_Facts_with_Timeshift_in_Columns.PER - GLOBAL_RELATIONAL.F02_Ref_Facts_with_Timeshift_in_Columns.PY`.

On the left, a tree view shows the 'F02 Ref Facts with Timeshift in Columns' table with sources: PER, PY, Var. PER-PY, and Var. % PER-PY. Red and green boxes highlight these sources, with arrows pointing to their respective configuration panels.

1. Anlegen der logischen Faktentabelle „F02 ..“
2. Die Spalten „PER“ (Periodenwert) und „PY“ (Previous Year bzw. Vorjahr) referenzieren die Spalte „Data“ der ersten logischen Faktentabelle „F00 ..“
3. Die Abweichung dagegen berechnen sich aus den Spalten „PER“ und „PY“ von „F02 ..“

DOAG 2010 - Dimensionale Modellierung mit OBIEE 15

Abb. 7: Logische Faktentabelle F02 dimensioniert nach Dimension Measure

BMM - optionale Definition (4)

The screenshot shows the 'Logical Column - Sales' configuration panel. The 'Name' field is 'Sales'. It belongs to the table 'GLOBAL_RELATIONAL.F03_Ref_Facts_with_Measures_and_Timeshift_in_Columns'. The 'Use existing logical columns as the source' checkbox is checked. The source filter is: `FILTER(GLOBAL_RELATIONAL.F02_Ref_Facts_with_Timeshift_in_Columns.PER USING GLOBAL_RELATIONAL.D03_MEASURE_DIM.MEASURE_ID = 'SALES')`.

The 'Logical Column - Sales PY' configuration panel shows the 'Name' field as 'Sales PY'. It also belongs to the table 'GLOBAL_RELATIONAL.F03_Ref_Facts_with_Measures_and_Timeshift_in_Columns'. The 'Use existing logical columns as the source' checkbox is checked. The source filter is: `FILTER(GLOBAL_RELATIONAL.F02_Ref_Facts_with_Timeshift_in_Columns.PY USING GLOBAL_RELATIONAL.D03_MEASURE_DIM.MEASURE_ID = 'SALES')`.

The 'Logical Column - Sales Var. % PER-PY' configuration panel shows the 'Name' field as 'Sales Var. % PER-PY'. It belongs to the table 'GLOBAL_RELATIONAL.F03_Ref_Facts_with_Measures_and_Timeshift_in_Columns'. The 'Use existing logical columns as the source' checkbox is checked. The source filter is: `100 * (GLOBAL_RELATIONAL.F03_Ref_Facts_with_Measures_and_Timeshift_in_Columns.Sales - GLOBAL_RELATIONAL.F03_Ref_Facts_with_Measures_and_Timeshift_in_Columns.Sales PY) / GLOBAL_RELATIONAL.F03_Ref_Facts_with_Measures_and_Timeshift_in_Columns.Sales PY`.

On the left, a tree view shows the 'F03 Ref Facts with Measures and Timeshift in Columns' table with sources: Units, Units PY, Units Var. % PER-PY, Sales, Sales PY, Sales Var. % PER-PY, Costs, Costs PY, Costs Var. % PER-PY, Margin, Margin PY, and Margin Var. % PER-PY. Red, green, and blue boxes highlight these sources, with arrows pointing to their respective configuration panels.

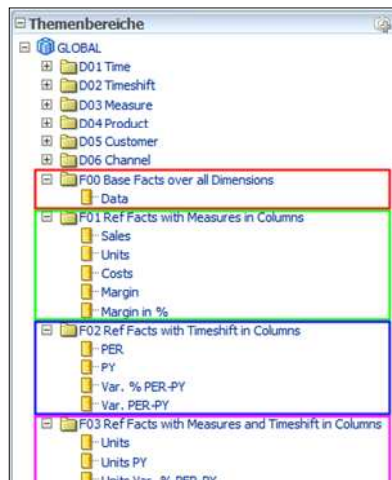
1. Anlegen einer dritten logischen Faktentabelle „F03 ..“
2. Die Spalten „Sales“ und „Sales PY“ referenzieren die Spalte „PER“ bzw. „PY“ der logischen Faktentabelle „F02 ..“
3. Die Abweichung dagegen berechnen sich aus den Spalten „Sales“ und „Sales PY“ von „F03 ..“

DOAG 2010 - Dimensionale Modellierung mit OBIEE

Abb. 8: Logische Faktentabelle F03 nicht dimensioniert nach den Dimensionen Measure und Timeshift

Ergebnis und Bewertung von Lösungsvorschlag 1

Ergebnis: 4 alternative Endanwendersichten
auf die selben Daten



| Transformierte Faktentabelle mit 2 zusätzlichen Dimensionen | | | | | |
|-------------------------------------------------------------|--------|-----|---------|-----------------|------|
| Month | Family | ... | Measure | Timeshift | Data |
| ... | | | | | |
| JUL-00 | CD/DVD | | Sales | Actual Period | 200 |
| JUL-00 | CD/DVD | | Sales | Previous Period | 180 |
| JUL-00 | CD/DVD | | Costs | Actual Period | 120 |
| JUL-00 | CD/DVD | | Costs | Previous Period | 125 |
| ... | | | | | |

- ← 1. Kennzahlen und Zeitabschnitt als Dimensionsausprägung
- ← 2. Zeitreihe als Dimensionsausprägung und Kennzahlen als Attributspalten
- ← 3. Kennzahlen als Dimensionsausprägung und Zeitabschnitt als Attributspalten
- ← 4. Alle Kennzahlen als Attributspalten

Abb. 9: Vier alternative Endanwendersichten bieten größtmögliche Flexibilität beim Erstellen von Auswertungen

Abgeleitet Kennzahlen mit „Calculated Items“

The screenshot shows the SAP BW interface. On the left, the 'Subject Areas' tree is visible, with 'D03 Measure' and 'D03 Measure Dimension' highlighted. The main window displays a 'Compound Layout' for 'rel_F00 Base Facts over all Dimensions'. Below this, a pivot table is shown with the following data:

| Measure | Actual Period | Prev. Year | Var. PER-PY | Var. % PER-PY |
|-------------|---------------|--------------|--------------|---------------|
| Units | 34.950,00 | 27.670,00 | 7.280,00 | 26,31 |
| Sales | 8.400.439,78 | 7.056.994,53 | 1.343.445,25 | 19,04 |
| Costs | 7.787.159,48 | 6.533.568,28 | 1.253.591,20 | 19,19 |
| Margin | 613.280,30 | 923.426,25 | 89.854,05 | 17,17 |
| Margin in % | 7,30 | 7,42 | -0,12 | -1,57 |

The 'Edit Calculated Item' dialog is open, showing the 'Display Label' as 'Margin in %' and the 'Selected' formula as '100 * (SALES=Sales-'COST=Costs) / 'SALES=Sales''. The 'Available' list contains 'Units', 'Sales', and 'Costs'. Red arrows point from the 'New Calculated Item' button to the 'Edit Calculated Item' dialog and from the 'Selected' formula to the 'Margin in %' row in the pivot table. Green arrows point from the 'Units', 'Sales', and 'Costs' items in the 'Available' list to the formula in the 'Selected' field.

Abb. 9: Die in Abbildung 1 geforderte Kreuztabelle ist mit Hilfe der Funktion Calculated Item machbar (siehe rote Markierung)

Bewertung aus Anwendersicht

- Die physische Speicherung der transformierten Faktentabelle erhöht die Anzahl der Datensätze erheblich
→ ggf. Performanceprobleme!!! ☹️
- Analysen aus verschiedenen Perspektiven sind damit möglich 😊
- Abfrage mit den Dimensionen „Timeshift“ und „Measure“ ist möglich.
 - Berechnete Elemente (Calculated Items) im Frontend
 - Definition von Calculated Items sind alternativ bereits in den Metadaten (RPD-Datei) zu definieren (mit Hilfe des neuen Feature CALCULATEDMEMBERS!)

Abb. 10: Bewertung des Lösungsvorschlags 1

Lösungsvorschlag 2 (mit Oracle OLAP)

Dieser Lösungsvorschlag basiert auf einer Mischung aus relationaler und multidimensionaler Technologie. Basis ist ein dimensionales Star- oder Snowflake-Schema. Die zugehörige Faktentabelle im relationalen Data Warehouse enthält nur die Basis-Kennzahlen *Sales*, *Units* und *Costs* (analog Lösungsvorschlag 1).

Mit Hilfe von Oracle OLAP wird ein Cube definiert. Dieser hat zusätzlich zu den für die Endanwender relevanten Dimensionen des relationalen Star- oder Snowflake-Schemas die Dimensionen *Measure* und *Timeshift*.

Schritt 1: Laden der Faktentabelle in Oracle OLAP

Das Laden des Cubes erfolgt im Rahmen eines regelmäßigen ETL-Prozesses, in der Regel nach erfolgreichem Laden der relationalen Faktentabelle. Die Auflösung der beiden Dimensionen *Measure* und *Timeshift* erfolgt durch Rechenregeln innerhalb von Oracle OLAP, welche erst zur Laufzeit, d.h. bei der Abfrage ausgeführt werden. Die Performance ist trotzdem sehr gut, zusätzlicher Speicherplatz wird dafür nicht benötigt.

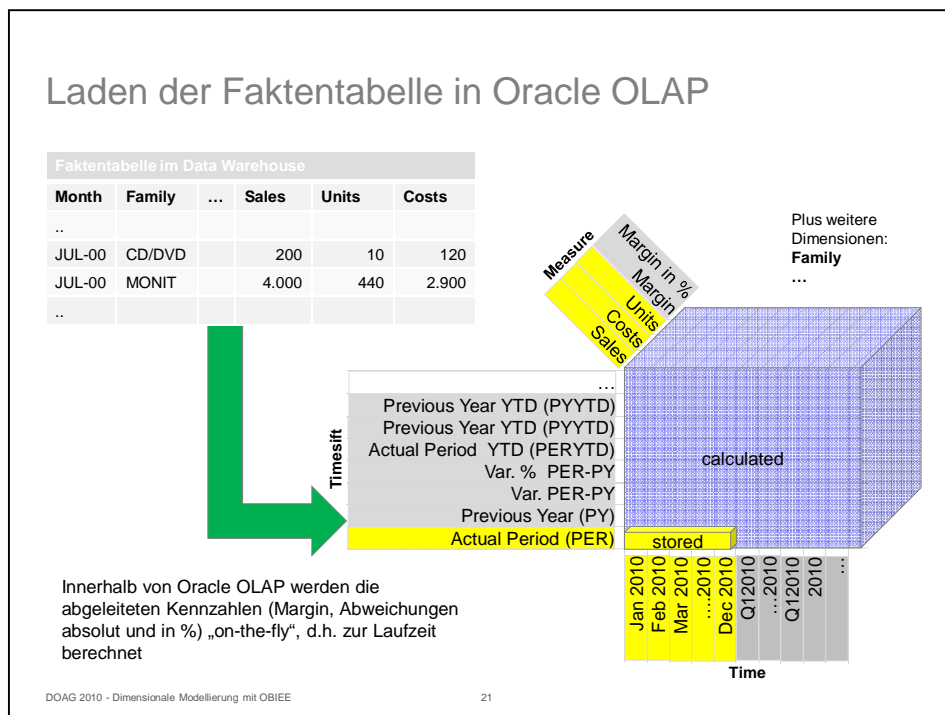


Abb. 11: Laden der Faktentabelle in Oracle OLAP. Die mit *calculated* bezeichneten Cube-Zellen werden erst innerhalb von Oracle OLAP nur teilweise physisch berechnet

Schritt 2: Definition des *Business Model and Mapping Layer (BMM)* in OBIEE

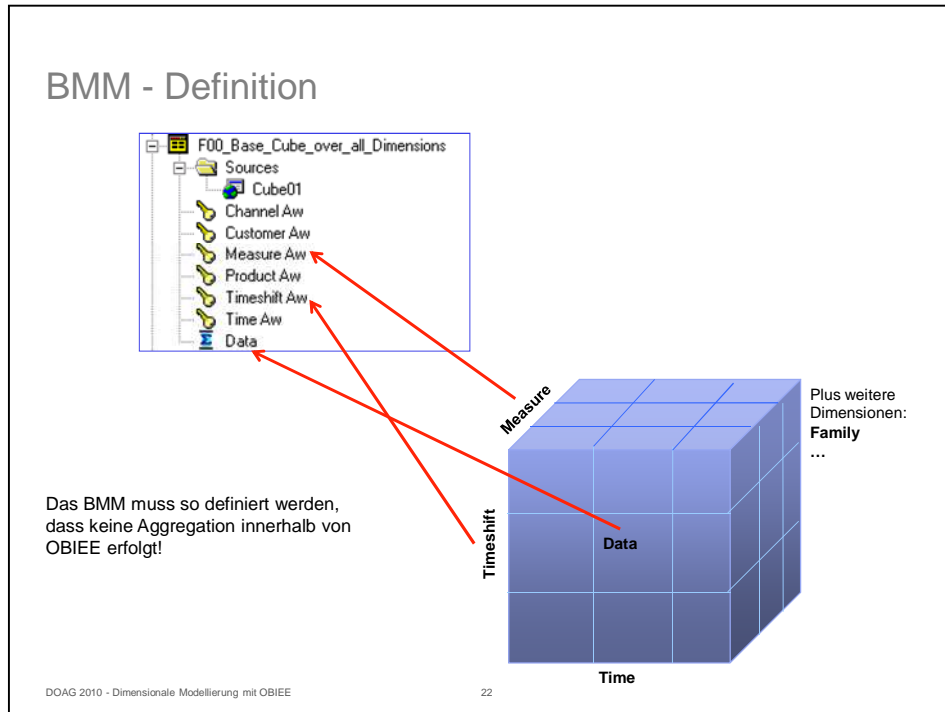


Abb. 12: Die logische Faktentabelle in OBIEE referenziert den OLAP Cube auf allen Hierarchieebenen (Berechnungen erfolgen innerhalb von Oracle OLAP!)

Ergebnis und Bewertung von Lösungsvorschlag 2

Ergebnis: OK!!!

- Im Gegensatz zum relationalen Ansatz stehen sämtliche abgeleiteten Kennzahlen in Kombination zueinander bereits in der Datenquelle zur Verfügung.
- Die Berechnung erfolgt ausschließlich in Oracle OLAP

← Kennzahlen und Zeitreihe als Dimensionsausprägung

Abb. 13: Die in Abbildung 1 geforderte Kreuztabelle ist möglich

Achtung mit Zwischensummen

- Seit OBIEE 11g durch Hierarchien besser ☺

aw_F00_Cube_with_Measure_and_Timeshift_Dimension
Wrong Totals for 46 Measures!!!

| | Actual Period | Prev. Year | Var. PER-PY | Var. % PER-PY |
|--------------------|---------------------|---------------------|---------------------|---------------|
| Region | | | | |
| Asia Pacific | 2.511.959,29 | 1.683.240,47 | 828.718,82 | 49,23 |
| Europe | 1.039.346,29 | 1.028.914,12 | 10.432,17 | 1,01 |
| North America | 4.235.853,90 | 3.821.413,69 | 414.440,21 | 10,85 |
| Gesamtsumme | 7.787.159,48 | 6.533.568,28 | 1.253.591,20 | 61,09 |

aw_F00_Cube_with_Measure_and_Timeshift_Dimension
Right Totals over 11g Hierarchies!!!

| | Actual Period | Prev. Year | Var. PER-PY | Var. % PER-PY |
|-----------------------|---------------------|---------------------|---------------------|---------------|
| D05 Customer | | | | |
| Total Customer | 7.787.159,48 | 6.533.568,28 | 1.253.591,20 | 19,19 |
| Asia Pacific | 2.511.959,29 | 1.683.240,47 | 828.718,82 | 49,23 |
| Europe | 1.039.346,29 | 1.028.914,12 | 10.432,17 | 1,01 |
| North America | 4.235.853,90 | 3.821.413,69 | 414.440,21 | 10,85 |

Abb. 14: Beim Einsatz von Oracle OLAP muss beim Einsatz von Zwischensummen aufgepasst werden. Abhilfe schaffen die neuen Hierarchien in OBIEE 11g

Bewertung aus Anwendersicht

- Zusätzliche Lizenzkosten für Oracle OLAP ☹
- Komplexe Zeitreihenanalysen sind in Oracle OLAP sehr einfach und vor allen Dingen performant zu implementieren
- Verschiedenste Analysen sind nun möglich ☺
- Sehr gute Abfrage-Performance durch MOLAP ☺
- Vorsicht mit den Zwischensummen von OBIEE!
 - Hier aggregiert OBIEE und liefert bei nicht aggregierbaren, berechneten Kennzahlen falsche Ergebnisse (z.B. Prozentwerte werden summiert) ☹ → Abhilfe durch Hierarchien in OBIEE 11g!!!

Abb. 15: Bewertung des Lösungsvorschlags 2

Kontaktadresse:

Claus Jordan
Trivadis GmbH
Industriestrasse 4
D-70565 Stuttgart

Telefon: +49 (0) 711-903 63 230
Mobile: +49-(0) 162-295 96 43
Fax: +49 (0) 711-903 63 259
E-Mail: claus.jordan@trivadis.com
Internet: www.trivadis.com