

alea iacta est – Würfel (nicht nur) am Beispiel von PROC OLAP

Christian Kothenschulte
LBS Westdeutsche Landesbausparkasse
Himmelreichallee 40
48149 Münster
christian.kothenschulte@lbswest.de

Zusammenfassung

In der LBS West wurde 2011 der SAS Enterprise BI Server eingeführt. Bis dahin wurden alle Würfel (mehrdimensionale Ergebnisdarstellung) als MDDB (Multi-Dimensional Database) mit der SAS-Prozedur PROC MDDB erstellt. Nun werden die Würfel mit der SAS-Prozedur PROC OLAP erzeugt.

In dem Beitrag werden zunächst, losgelöst von SAS, der Aufbau, die Verwendungsmöglichkeiten und die Vorteile eines Würfels erläutert. Dabei werden Begrifflichkeiten wie "Slice & Dice" und "Roll-Up & Drill-Down" behandelt. Anschließend wird anhand der SAS-Prozedur PROC OLAP die Umsetzung in SAS demonstriert.

Zwar kann ein Würfel auch ohne Programmcode über das SAS OLAP Cube Studio erzeugt werden, für das Verständnis einiger Begriffe wie "Dimension", "Hierarchie" und "Level" oder das programmgesteuerte Erstellen (vor allem mit Makros/Makrovariablen) sollte man die Prozedur kennen. Der Beitrag bietet hier einen Einstieg. Als Beispieldaten dienen die Datensätze aus der Tabelle SASHELP.PRDSALE. Abschließend werden kurz die Zugriffsmöglichkeiten über den SAS Enterprise Guide, über die SAS Add-Ins für Microsoft Office und das SAS Web Report Studio gezeigt.

Schlüsselwörter: OLAP, MDDB, Würfel, Cube

1 Motivation / Einleitung

Nicht nur bei Bausparkassen existieren eine Vielzahl an Auswertungen mit unterschiedlichen fachbezogenen Inhalten. Häufig wird eine Auswertung von mehreren Organisationseinheiten genutzt. Dabei sind oft unterschiedliche Blickwinkel auf die Ergebnisse notwendig.

Viele AnwenderInnen denken bei (neuen) Auswertungen an die Tabellenform, da diese aufgrund der Verbreitung verschiedener Office-Produkte (z.B. Microsoft Access und Microsoft Excel) die bekannteste ist. Bei dieser immer zweidimensionalen Darstellung von Ergebnissen kann man leicht bestimmte Ausprägungen filtern oder Zwischensummen etc. berechnen. Bei hohen Anzahlen von Spalten und Zeilen gehen jedoch vielleicht wichtige Sichten auf die Ergebnisse oder Zusammenhänge verloren.

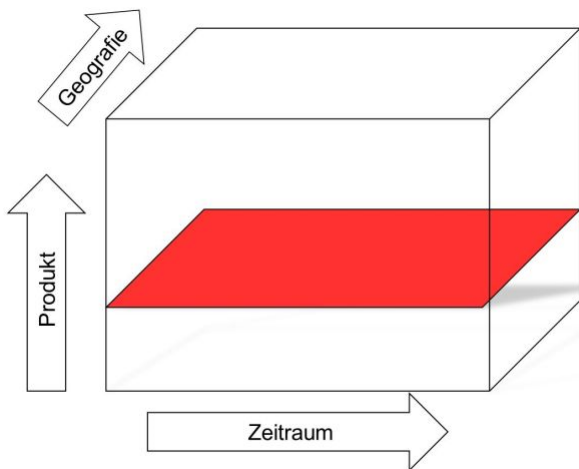
Diese Ausarbeitung versteht sich als Werbung für die Nutzung von mehrdimensionalen Strukturen. Zunächst werden die Theorie und wichtige Fachbegriffe erläutert, um anschließend den Transfer zu SAS und den Umsetzungsmöglichkeit zu machen.

2 Würfel

Ein Würfel (oder auch Cube) wird in der Literatur in der Regel dreidimensional dargestellt. Zum einen können die Würfel so grafisch dargestellt werden, zum anderen hat ein Mensch die Vorstellungskraft für bis zu drei Dimensionen. Bei vier oder mehr Dimensionen wird beides schwierig. Es sei aber explizit hervorgehoben, dass ein Würfel aus mehr Dimensionen bestehen kann und damit auch seine Stärke entfaltet. Eine für fast alle Branchen nachvollziehbare Auswertung ist der Ausweis von Umsatz je Zeitraum (Dimension 1), je Produkt (Dimension 2) und je Vertriebsgebiet (Geografie, Dimension 3). Die folgende Grafik zeigt die Zusammenhänge.

Die in dem Würfel enthaltene Fläche stellt die Produktsicht dar: Ein Produkt über den gesamten Vertriebsweg und den gesamten Zeitraum. Dieses Herausschneiden einer Scheibe wird *Slice* genannt. Denkbar sind ebenfalls die regionale Sicht oder zeitliche Sicht. Ein Würfel im Würfel (ein Zeitraum, ein Produkt, eine Region) kann als Ad-hoc-Sicht bezeichnet werden.

Ein Würfel ist selbstverständlich nicht auf ein Maß (Umsatz) beschränkt. Je nach fachlicher Anforderung können weitere Maße ergänzt werden (zum Beispiel: Anzahl Verkäufe). Dabei muss das Maß nicht zwingend summiert werden. Viele weitere Funktionen sind möglich, zum Beispiel: maximaler Wert, minimaler Wert oder Durchschnitt.



Ein weiterer Begriff, der häufig im Zusammenhang mit *Slice* genannt wird, lautet *Dice*. Damit wird die Rotation eines Würfels bezeichnet, um neue, vielleicht auch überraschende Sichtweisen, Erkenntnisse und Denkanstöße auf die Ergebnisse zu erhalten.

Abbildung 1: Würfel Produktsicht

Die folgende Tabelle zeigt einen weiteren Vorteil von Würfeln auf.

Tabelle 1: Detailliertere Darstellung des Würfels

| | | 2013 | 2013 | | | | |
|-------------|-----------------------------------|------|------|----|----|----|----|
| | | | Q1 | Q2 | Q3 | Q4 | |
| Deutschland | => Drill-Down <= Roll-Up | 425 | Nord | 10 | 11 | 12 | 13 |
| | | | Süd | 20 | 21 | 22 | 23 |
| | | | West | 30 | 31 | 32 | 33 |
| | | | Ost | 40 | 41 | 42 | 44 |

Auf der linken Seite wird der Umsatz (425) für das Jahr 2013 (Dimension Zeit) und Deutschland (Dimension Vertriebsgebiet) summiert dargestellt. Wer den Umsatz weitergehend analysieren möchte, wünscht sich die Aufteilung auf detaillierterer Ebene. Auf der rechten Seite wird der Umsatz je Region und je Quartal aufgezeigt. Dieses Navigieren in den Daten wird *Drill-Down* (und *Roll-Up* als Rückweg) genannt.

3 SAS-Prozedur OLAP

Würfel lassen sich in SAS leicht mit dem SAS OLAP Cube Studio erzeugen. Da dies grafisch unterstützt, vergleichbar mit einem Assistenten, erfolgt, sind tiefere Kenntnisse der Prozedur OLAP nicht erforderlich.

Für die automatisierte Erstellung von Würfeln und deren Parametrisierung (zum Beispiel Auslagerung von Pfaden in Makrovariablen oder das Verwalten einer übergreifenden Dimension Zeit) sind sie jedoch notwendig.

3.1 Beispieldaten

Als Beispieldaten eignet sich die Tabelle PRDSALE aus der Bibliothek SASHELP. Sie enthält zwei Maße (Actual und Predict) und (für dieses Beispiel) drei Dimensionen.

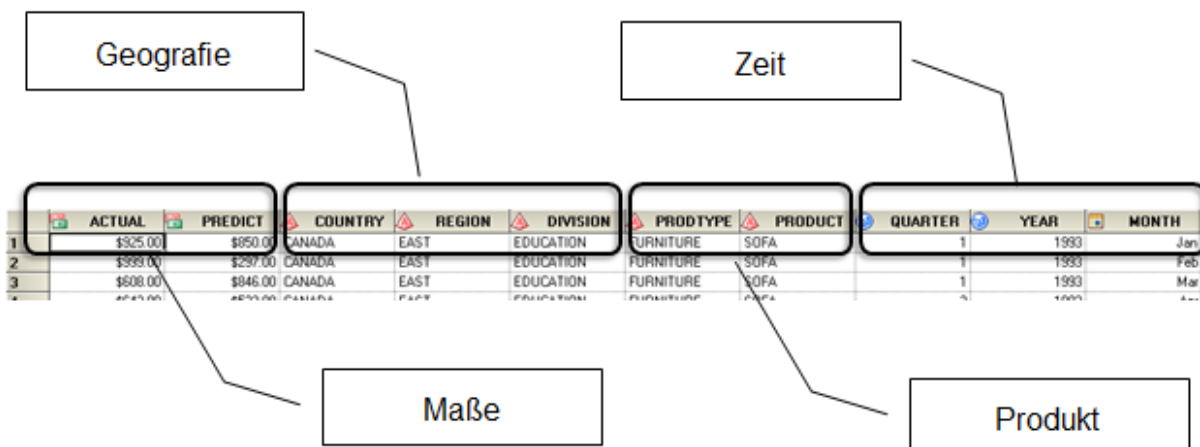


Abbildung 2: Beispieldaten SASHELP.PRDSALE

Aus den Beispieldaten soll der Würfel „Verkaeufe“ entstehen.

3.2 Prozedurkopf

```

PROC OLAP
  CUBE = "[...]/Cubes/Verkaeufe"
  DATA = SASUSER.PRDSALE
  DRILLTHROUGH_TABLE = SASUSER.DETAILS_PRDSALE
  PATH = '[...]\CUBE'
  DESCRIPTION = 'Beschreibung des Würfels Verkaeufe'
;
    
```

In dem Prozedurkopf werden der Name, der Pfad, die Quelltable und eine Beschreibung angegeben. Als sogenannte Drillthrough-Table kann eine (in diesem Beispiel

fiktive) Tabelle angegeben werden, die detaillierte Daten beinhaltet. Hier können später bei Zugriff auf den Würfel (siehe unten) die zugrunde liegenden Sätze ausgegeben werden. Ein Vorteil dabei: Die Quelltable und die Drillthrough-Tabelle müssen nicht in derselben Bibliothek liegen. So können viele AnwenderInnen Zugriff auf den Würfel haben, jedoch nur ausgewählte AnwenderInnen Zugriff auf die Detaildaten haben. Wenn die Quelltable bereits alle Detaildaten enthält, kann sie auch als Drillthrough-Tabelle verwendet werden.

3.3 Metaserver

```
METASVR
HOST      = "[Server]"
PORT      = 8565
OLAP_SCHEMA = "SASApp - OLAP Schema";
```

Die Angaben zum Metaserver sind administrativ vorgegeben oder werden als Makrovariablen ausgelagert, so dass ein Programm in mehreren Umgebungen nutzbar ist.

3.4 Dimensionen / Hierarchien

Eine Dimension besteht aus n Hierarchien. Eine Hierarchie besteht aus n Leveln.

3.4.1 Zeit

```
DIMENSION Zeit
CAPTION      = 'Zeit'
TYPE         = TIME
SORT_ORDER   = ASCENDING
HIERARCHIES  = (
    Zeit
) /* HIERARCHIES */;

HIERARCHY Zeit
ALL_MEMBER   = 'Alle Zeit'
CAPTION      = 'Zeit'
LEVELS       = (
    YEAR QUARTER MONTH
) /* LEVELS */
DEFAULT;

LEVEL YEAR
FORMAT       = 4.
TYPE         = YEAR
CAPTION      = 'Year'
SORT_ORDER   = ASCENDING;

LEVEL QUARTER
FORMAT       = 8.
TYPE         = QUARTERS
CAPTION      = 'Quarter'
```

```

        SORT_ORDER      = ASCENDING;

LEVEL MONTH
  FORMAT      = MONNAME3.
  TYPE       = MONTHS
  CAPTION    = 'Month'
  SORT_ORDER = ASCENDING;

```

Bei der Dimension Zeit ist es sinnvoll, direkt den TYPE mit TIME anzugeben, damit bei Zugriff auf den Würfeln auch relative Filter möglich sind (zum Beispiel die letzten drei Jahre). Die Dimension Zeit hat den folgenden Aufbau:

Jahr => Quartal => Monat

Wird eine zeitliche Sicht auf die Daten ohne den Zwischenschritt des Quartals benötigt, kann unter derselben Dimension eine weitere Hierarchie angelegt werden, in der das Level Quartal nicht angelegt wird.

3.4.2 Produkt

```

DIMENSION Produkt
  CAPTION      = 'Produkt'
  SORT_ORDER  = ASCENDING
  HIERARCHIES = (
    Produkt
  ) /* HIERARCHIES */;

HIERARCHY Produkt
  ALL_MEMBER = 'Alle Produkt'
  CAPTION    = 'Produkt'
  LEVELS     = (
    PRODTYPE PRODUCT
  ) /* LEVELS */
  DEFAULT;

LEVEL PRODTYPE
  FORMAT      = $CHAR10.
  CAPTION    = 'Product type'
  SORT_ORDER = ASCENDING;

LEVEL PRODUCT
  FORMAT      = $CHAR10.
  CAPTION    = 'Product'
  SORT_ORDER = ASCENDING;

```

Die Dimension Produkt hat den folgenden Aufbau: Produkttyp => Produkt

4 Measures / Maße

```
MEASURE PREDICTSUM
  STAT      = SUM
  COLUMN    = PREDICT
  CAPTION   = 'Summe von PREDICT'
  FORMAT    = DOLLAR12.2
  DEFAULT;
```

Als Maß wird die Summe von Predict erzeugt. Weitere Maße mit weiteren Rechenvorschriften sind möglich. Außerdem ist es möglich, eigene Maße zu erstellen, die aus vorhandenen abgeleitet werden.

```
DEFINE MEMBER '[VERKAEUFE].[MEASURES].[DIFF]' AS
  '[MEASURES].[ACTUALSUM] - [MEASURES].[PREDICTSUM]';
```

4.1 Aggregationen

```
AGGREGATION /* Standard */
  /* levels */
  COUNTRY DIVISION MONTH PRODTYPE
  PRODUCT QUARTER REGION YEAR
  / /* options */
  NAME      = 'Standard';
```

Unter Aggregation sind alle im Würfel verwendeten Level anzugeben.

5 Gründe für OLAP

Warum sollte man OLAP verwenden? Die zunächst einfache Antwort lautet: OLAP ist sehr gut für viele Anwendungsfälle geeignet. Immer, wenn man mindestens zwei Blickwinkel auf Daten hat (zum Beispiel Zeit, Produkt, Vertriebsweg, Kunde), kann man die Daten mehrdimensional aufbereiten. Besonders interessant sind Würfel dann, wenn man zunächst Daten auf oberster Ebene betrachten und anschließend weiter ins Detail gehen will.

Durch die technische Aufbereitung des Würfels sind bereits die angegebenen Dimensionen in den enthaltenen Hierarchien vorverdichtet und die Maße im Vorfeld berechnet. Dies führt zu einer hohen Zugriffsgeschwindigkeit bei der Nutzung (Online).

In sehr vielen Fällen wird ein Würfel initial erzeugt und erhält typischer Weise nach einem bestimmten Zeitraum (zum Beispiel Tag oder Monat) zusätzliche neue Daten. Hierbei ist es nicht erforderlich, den Würfel bei jeder Aktualisierung neu zu erzeugen. Dies würde mit zunehmender Datenmenge immer mehr Ressourcen in Anspruch nehmen. Stattdessen kann die Aktualisierung mit der Prozedur OLAP und den Optionen `update_inplace` und `add_data` erfolgen. Dies führt zu einem enormen Performancegewinn (Batch).

Ein weiterer Vorteil ist die einfache Vergabe von Rechten auf Detaildaten. Der Würfel kann Daten bis zu einem bestimmten Detailierungsgrad enthalten. Detailliertere Daten können in der Drillthrough-Tabelle hinterlegt werden (siehe oben). Auf diese Daten, typischerweise auf Einzelsatzebene, erhalten nur bestimmte Anwendergruppen Zugriff. Aber auch innerhalb eines Würfels können die Zugriffsrechte beschränkt werden. So darf beispielsweise der Vertriebsleiter für den Bereich „Education“ die Einzelwerte für „East“ und „West“ sehen, jedoch für den Bereich „Consumer“ nur die Gesamtsumme ohne Aufteilung nach „East“ und „West“. Diese Konfiguration erfolgt in der Management Konsole (Authorization-Manager → Ressourcen-Verwaltung → Nach Speicherort → SASApp → SASApp-OLAP-Schema → Würfel (zum Beispiel Verkäufe) → Dimensionen → zum Beispiel Vertriebsweg → Rechtsklick auf Eigenschaften → Reiter Autorisierung).

Beispiel: Die Formel

```
{[Vertriebsweg].[Alle Vertriebsweg],
 [Vertriebsweg].[Alle Vertriebsweg].[CANADA],
 [Vertriebsweg].[Alle Vertriebsweg].[GERMANY],
 [Vertriebsweg].[Alle Vertriebsweg].[U.S.A.]}
<!--CONDITION-->
{[Vertriebsweg].[Alle Vertriebsweg].[U.S.A.].[CONSUMER],
 [Vertriebsweg].[Alle Vertriebsweg].[U.S.A.].[EDUCATION]}
```

sorgt dafür, dass die Rechtegruppe die jeweiligen Gesamtsummen von Canada, Germany und U.S.A sehen darf, jedoch lediglich U.S.A eine Ebene weiter aufblättern kann:

| | | |
|--|---|--------------|
| Alle Vertriebsweg | | \$360,211.00 |
| | CANADA | \$119,329.00 |
| | GERMANY | \$117,119.00 |
| <input type="checkbox"/> <input checked="" type="checkbox"/> Alle Vertriebsweg | U.S.A. | \$123,763.00 |
| | <input type="checkbox"/> <input checked="" type="checkbox"/> U.S.A. CONSUMER | \$62,913.00 |
| | <input type="checkbox"/> <input checked="" type="checkbox"/> U.S.A. EDUCATION | \$60,850.00 |

Abbildung 3: Rechtestruktur innerhalb eines Würfels

Während bei der angegebenen Formel die Ausprägungen Consumer und Education fest vergeben wurden, kann man die nachfolgenden Ausprägungen auch variabel vergeben (Schlüsselwort „Descendants“):

```
{[Vertriebsweg].[Alle Vertriebsweg],
 [Vertriebsweg].[Alle Vertriebsweg].[CANADA],
 [Vertriebsweg].[Alle Vertriebsweg].[GERMANY],
 [Vertriebsweg].[Alle Vertriebsweg].[U.S.A.],
 Descendants([Vertriebsweg].[Alle Vertriebsweg].[U.S.A.]})}
```

Damit können dann auch Consumer und Education weiter aufgeklappt werden.

6 Zugriffsmöglichkeit

6.1 Add-In für Microsoft Office

Über die Microsoft Office Add-Ins erfolgt das Einlesen eines Würfels direkt in eine Pivot-Tabelle.

| | A | B | C | D |
|----|----------------------|-----------------------|--------------|------------------|
| 1 | Summe von PREDICT | Spaltenbeschriftungen | | |
| 2 | Zeilenbeschriftungen | ± 1993 | ± 1994 | Gesamtergebnis * |
| 3 | ± FURNITURE | \$147.401,00 | \$130.917,00 | \$278.318,00 |
| 4 | ± OFFICE | \$212.810,00 | \$215.167,00 | \$427.977,00 |
| 5 | Gesamtergebnis * | \$360.211,00 | \$346.084,00 | \$706.295,00 |
| 6 | | | | |
| 7 | | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | | | |
| 11 | | | | |
| 12 | | | | |
| 13 | | | | |
| 14 | | | | |

Abbildung 4: Ansicht in Microsoft Excel

6.2 SAS Enterprise Guide

Im SAS Enterprise Guide kann der Würfel per Drag&Drop genutzt werden

| | | Year | ± 1993 | ± 1994 |
|--------------|-----------|----------|-------------------|-------------------|
| | | Measures | Summe von PREDICT | Summe von PREDICT |
| Product type | Country | | | |
| ± FURNITURE | ± CANADA | | \$47.155,00 | \$42.786,00 |
| | ± GERMANY | | \$50.080,00 | \$43.233,00 |
| | ± U.S.A. | | \$50.166,00 | \$44.898,00 |
| ± OFFICE | ± CANADA | | \$72.174,00 | \$70.904,00 |
| | ± GERMANY | | \$67.039,00 | \$71.202,00 |
| | ± U.S.A. | | \$73.597,00 | \$73.061,00 |

Abbildung 5: Ansicht in SAS Enterprise Guide

Wenn mit Daten aus dem Würfel weiter gearbeitet werden soll, kann dies mit der rechten Maustaste und der Auswahl [Segment erstellen] erfolgen:



Abbildung 6: Weiterverarbeitung in SAS Enterprise Guide

6.3 SAS Web Report Studio

Im SAS Web Report Studio kann der Würfel als Kreuztabelle genutzt werden.

Angewendete Filter: Ohne

| Product type | | FURNITURE | | OFFICE | |
|--------------|------|------------------|-------------------|------------------|-------------------|
| | | Summe Von Actual | Summe Von Predict | Summe Von Actual | Summe Von Predict |
| Country | Year | | | | |
| CANADA | 1993 | \$48,564.00 | \$47,155.00 | \$72,456.00 | \$72,174.00 |
| | 1994 | \$49,300.00 | \$42,786.00 | \$76,670.00 | \$70,904.00 |
| GERMANY | 1993 | \$54,246.00 | \$50,080.00 | \$73,158.00 | \$67,039.00 |
| | 1994 | \$46,948.00 | \$43,233.00 | \$71,646.00 | \$71,202.00 |
| U.S.A. | 1993 | \$46,085.00 | \$50,166.00 | \$74,968.00 | \$73,597.00 |
| | 1994 | \$45,482.00 | \$44,898.00 | \$70,814.00 | \$73,061.00 |

Abbildung 7: Ansicht im SAS Web Report Studio

Mit Klick auf die (unterstrichenen) Maße öffnet sich ein weiteres Fenster mit den zugrundeliegenden Daten (Drillthrough-Tabelle, siehe oben).

Literatur

- [1] P. Gluchowski: Grundstrukturen multidimensionaler Datenmodelle. BI-Spektrum, Ausgabe 01-2011, Seite 31ff
- [2] SAS(R) 9.2 OLAP Server: User's Guide
- [3] N. Clausen: OLAP Multidimensionale Datenbanken. ADDISON-WESLEY, München, 1998

