

	Kundennummer BA	Geschlecht	Altersgruppe	Wohnort
1	12345A12345	M	18-21	München
2	12345A12346	M	18-21	München
3	12345A12347	M	18-21	München
4	12345A12348	W	18-21	Berlin
5	12345A12349	W	18-21	Berlin
6	12345A12350	W	18-21	Berlin
7	12345A12351	W	18-21	Berlin
8	12345A12352	W	18-21	Berlin
9	12345A12353	W	18-21	Berlin
10	12345A12354	M	21-25	Berlin
11	12345A12355	W	21-25	Berlin
12	12345A12356	M	21-25	Berlin
13	12345A12357	W	21-25	Koblenz
14	12345A12358	W	21-25	Koblenz

Leitfaden

PowerPivot-Datenmodelle richtig aufbauen

Ein Basic-Workshop

Inhalt

1. Ausgangslage	3
2. Ursachenforschung	3
3. Mögliche Probleme	4
4. Grundlagen.....	5
5. Ergebnis aus dem Stern-Schema	7
6. Wie erstelle ich Dimensionstabellen?	8
7. Zum letzten offenen Problem	8
8. Weitere wichtige Tipps	9
9. Fazit	9
Über den Autor.....	10

Hinweis:

Aufgrund der besseren Lesbarkeit wird in diesem Text nur die männliche Schreibweise verwendet. Diese schließt automatisch die weibliche Form mit ein. Wir bitten alle Leserinnen und Leser um Verständnis.

1. Ausgangslage

Am Anfang ist noch alles gut, die Pivot-Tabellen funktionieren einwandfrei und die Daten werden schnell und gut angezeigt. Nach einiger Zeit fangen aber die ersten Probleme an. Wenn z.B. Aktualisierungen fehlschlagen und/oder irgendwelche Daten, auf die gefiltert wurde, nicht mehr gefunden werden. Oder wenn die Tabellen beim Filtern oder Slicen (Datenschnitte) langsam werden, oder... Was nun? In diesem Fall sollten Sie das zugrundeliegende Datenmodell genauer begutachten und ggf. „tunen“ oder sogar neugestalten.

2. Ursachenforschung

	Kundennummer BA	Geschlecht	Altersgruppe	Wohnort
1	12345A12345	M	18-21	München
2	12345A12346	M	18-21	München
3	12345A12347	M	18-21	München
4	12345A12348	W	18-21	Berlin
5	12345A12349	W	18-21	Berlin
6	12345A12350	W	18-21	Berlin
7	12345A12351	W	18-21	Berlin
8	12345A12352	W	18-21	Berlin
9	12345A12353	W	18-21	Berlin
10	12345A12354	M	21-25	Berlin
11	12345A12355	W	21-25	Berlin
12	12345A12356	M	21-25	Berlin
13	12345A12357	W	21-25	Koblenz
14	12345A12358	W	21-25	Koblenz
15	12345A12359	W	21-25	Koblenz
16	12345A12360	M	21-25	Koblenz
17	12345A12361	M	21-25	Koblenz

Gerne werden große und breite Tabellen in das PowerPivot-Datenmodell geladen, wie z.B. der nebenstehende Screenshot zeigt. Die Tabelle beinhaltet alle Attribute/Felder, die für die Pivot-Tabelle benötigt werden und wird ohne weitere Tabellen genutzt.

Hinweis: Diese Tabelle beinhaltet nur 681 Zeilen, in der Realität sind die Tabellen meist größer (> 500.000 Datensätze). Sollten Sie überwiegend

nur mit solchen kleinen Daten-beständen zu tun haben, werden Sie i.d.R. keine Performance-Probleme haben. Dennoch sollten Sie jetzt weiterlesen und auch das Modell ggf. entsprechend anpassen, denn am Anfang scheint vieles noch klein und übersichtlich, aber nach einigen Wochen und Monaten oder gar Jahren ist aus einer kleinen Auswertung schnell mal eines der wichtigsten Entscheider-Instrumente mit wesentlich mehr Daten als erwartet geworden.

Aus solchen Daten werden dann bspw. folgende Pivot-Tabellen erstellt:

Altersgruppe	Anzahl von Kundenummer BA			Gesamtergebnis
	M	W		
18-21				
21-25				
	Berlin	19	139	158
	Koblenz	122	188	310
	München	27	186	213
	Gesamtergebnis	168	513	681

In diesem Beispiel wurde „einfach“ das Attribut „Kundenummer BA“ in die Werte gezogen, das Geschlecht in die Spalten und der Wohnort in die Zeilen. Auf das Attribut „Altersgruppe“ wurde dann ein Datenschnitt/Slicer erstellt. Alles ist so weit gut, oder?

3. Mögliche Probleme

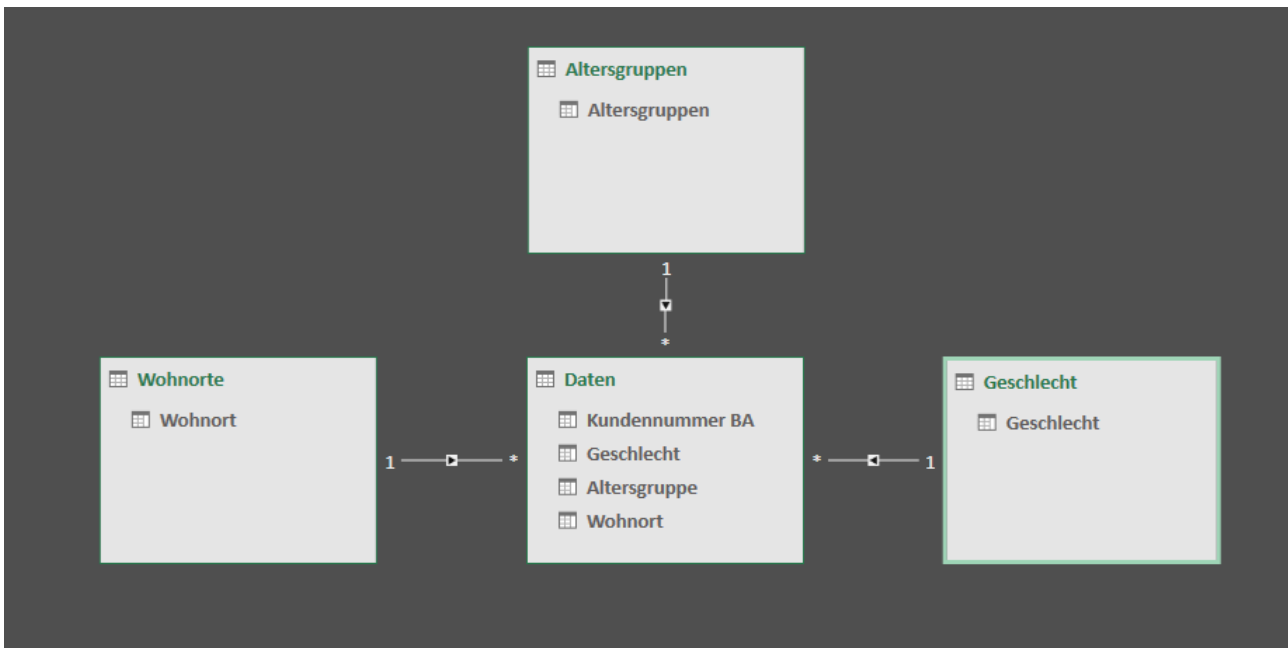
So weit, so gut – doch dahinter steckt mehr. Zum Beispiel auch die folgenden „versteckten“ Probleme:

1. Für die einfachen drei Zeilen „Berlin, Koblenz und München“ musste PowerPivot 681 Zeilen gruppieren, um die Eindeutigkeit zu prüfen und um diese darzustellen – das passiert übrigens bei jedem Bedienen eines Filters, Slicers/Datenschnitts. Und auch beim Ändern von Feldern passiert das immer wieder – Ihr PC und Excel werden nun richtig gefordert!
2. Wie bei 1. gilt dies auch für das Geschlecht!
3. Wie bei 1. gilt dies auch für den Slicer „Altersgruppe“
4. In einigen Excel-Versionen (u.a. Excel 2010) gibt es ein weiteres Problem: Nehmen wir an, Sie haben im Datenschnitt/Slicer die Altersgruppe „21-25“ ausgewählt. Nun gibt es aber keine Person mehr in den Grunddaten, die in diese Personengruppe fällt. Somit gibt es diesen Eintrag auch nicht mehr in der Tabelle. Nun wird die Pivot-Tabelle aktualisiert, dann erhalten Sie Fehlermeldungen und die gesamten Pivots in der Arbeitsmappe werden nicht mehr aktualisiert. Nun müssen Sie zuerst die Ursache feststellen, indem Sie den fehlenden Eintrag identifizieren, den Datenschnitt auf einen anderen Wert legen oder entfernen. Dann erst funktioniert wieder eine Aktualisierung.
5. Bei der Anzahl von „Kundenummer BA“ handelt es sich i.d.R. um ein implizites Measure, d.h. das Feld wurde in die Werte gezogen und Excel muss nun bei jedem Zugriff die Daten automatisch zählen – das kostet Zeit!

Schon in diesem einfachen Beispiel steckt also ein größeres Konfliktpotenzial. Probleme, die mir in größeren Arbeitsmappen aber immer wieder über den Weg laufen – Ihnen sicherlich auch.

4. Grundlagen

Bevor wir nun versuchen, die Probleme zu lösen bzw. präventiv zu verhindern, brauchen wir Grundlagen – diese entnehmen wir aus den Architekturen von Data-Warehouse-Systemen. Ein gutes und performantes Datenmodell ist nach dem **Stern-** bzw. **Schneeflocken-**Schema aufgebaut:



In diesem Beispiel ist die bisherige Daten-Tabelle als sog. **Faktentabelle** verfügbar, d.h. in dieser Tabelle werden lediglich die Zähldaten gespeichert, die den momentanen Zustand wiedergeben (welcher Kunde ist welchem Geschlecht, welcher Altersgruppe und welchem Wohnort zugeordnet – also alles das, was in der Auswertung verfügbar ist). Die drei Tabellen um die Faktentabelle

Wohnort
1 Berlin
2 Koblenz
3 München
4 Köln

herum bezeichnet man als **Dimensionstabellen** (siehe links bspw. die Tabelle „Wohnort“), d.h. in diesen sind nur die eindeutigen Werte, bzw. die Zustände die es geben kann. Diese Dimensionstabellen sind dann mit den Fakten verknüpft. Eine Dimensionstabelle kann, wie dieser Screen zeigt, auch Daten beinhalten, die momentan nicht in den Fakten verfügbar sind. In

diesem Screen ist beispielsweise der Wohnort „Köln“ enthalten. Das bietet die Möglichkeit, Werte vorzusehen, die es mal gab oder geben könnte. Nehmen wir an, dass in einer Dimensionstabelle ein Wert fehlt, der aber in der Faktentabelle vorhanden ist. Dann würden diese Fakten nicht in der Auswertung ausgegeben werden können. Konkret bedeutet das, dass aus der Faktentabelle lediglich das Feld „Kundennummer BA“ zur Zählung genutzt wird. Die restlichen drei Felder dienen nur noch den Verknüpfungen und werden in der Excel-Pivot nicht mehr genutzt.

In dem obigen Beispiel sehen Sie ein Stern-Schema, ein Schneeflocken-Schema würde sich ergeben, wenn weitere Dimensionstabellen an einer anderen Dimensionstabelle angehängt sind. Das lassen wir aber an dieser Stelle aufgrund der Übersichtlichkeit und Einfachheit weg.

Hinweis:

Geübte Anwender werden sicherlich gleich merken, dass in den Fakten eher das Alter abgebildet werden würde und nicht die Altersgruppe. Gleiches gilt auch für das „Geschlecht“, da dies sich meist nicht häufiger ändert (auch wenn das vorkommen kann, hierfür gibt es dann weitere Lösungsmodelle wie z.B. Langsam-Veränderliche-Dimensionen für die Abbildung). Man würde sonst eine Dimensionstabelle anstellen, die die Person widerspiegelt und dann lediglich mit der „Kundennummer BA“ verknüpft ist. Da es sich in diesem Leitfaden aber erstmal um „Basics“ handeln soll, wurde bewusst darauf verzichtet.

5. Ergebnis aus dem Stern-Schema

Erstellt man nun nach einem solchen Modell das Pivot, sehen Sie, dass nur noch das Feld „Kundennummer BA“ aus der Faktentabelle für die Werte genutzt wird. Die Felder für den Slicer / Datenschnitt und für die Spalten- und Zeilen kommen aus den Dimensionstabellen:

Altersgruppen	Anzahl von Kundennummer BA	M	W	Gesamtergebnis
15-21	Berlin	19	139	158
21-25	Koblenz	122	188	310
25-50	München	27	186	213
	Gesamtergebnis	168	513	681

AKTIV ALLE	
In den Bericht aufzunehmende Felder auswählen:	
Suchen	
Altersgruppen	
Daten	<input checked="" type="checkbox"/> Kundennummer BA
	<input type="checkbox"/> Geschlecht
	<input type="checkbox"/> Altersgruppe
	<input type="checkbox"/> Wohnort
Geschlecht	<input checked="" type="checkbox"/> Geschlecht
Wohnorte	<input checked="" type="checkbox"/> Wohnort
Felder zwischen den Bereichen ziehen und ablegen:	
FILTER	SPALTEN
	Geschlecht
ZEILEN	WERTE
Wohnort	Anzahl von K...

Das Ergebnis ist selbstverständlich das gleiche (ausgenommen beim Datenschnitt sehen Sie eine neue Altersgruppe, in der aber keine Daten enthalten sind. Dies soll verdeutlichen, dass Werte aus den Dimensions-Tabellen genutzt werden können, die in den Fakten nicht enthalten sind).

Das ganze obige löst die Probleme 1 – 4 (Mögliche Probleme), da PowerPivot nun nur noch die Zeilen aus den Dimensionstabellen gruppieren muss. Darin sind aber eben nur unterschiedliche enthalten – und davon auch nur wenige. Konkret bedeutet das in diesem Beispiel: Die Gruppierung wurde bei den Wohnorten auf drei reduziert (gegenüber vorher 681). Stellen Sie sich nun mal vor, was das bedeutet, wenn die Tabelle 500.000 Datensätze oder mehr beinhaltet – Excel und Ihr PC werden es Ihnen danken!

6. Wie erstelle ich Dimensionstabellen?

Jetzt ist es aber auch an der Zeit zu erfahren, wie Sie denn nun Dimensionstabellen erstellen. Sie haben i.d.R. zwei Möglichkeiten:

1. Sie fragen die Datenbank gruppiert nach den jeweiligen Dimensionsdaten ab, bspw. so:

```
SELECT DISTINCT Ort FROM Kundentabelle
```

Damit bekommen Sie alle eindeutigen Orte aus der Kundentabelle zurück und können damit die Dimensionstabelle aufbauen – Vorteil hierbei ist, dass die Tabelle immer aktuell ist, es werden also alle Orte zurückgegeben, die auch in den Daten enthalten sind. Ein Nachteil: Sollte ein Ort nicht mehr in der Tabelle enthalten sein, erscheint dieser auch nicht mehr in den Filtern oder Datenschnitten/Slicern ... (siehe Mögliche Probleme Punkt 4)

2. Sie erstellen und pflegen diese Tabelle von Hand. Initial empfiehlt sich aber, dass Sie zuerst einmal die Faktentabelle erstellen. Anschließend erstellen Sie ein Pivot, indem Sie das Dimensionsfeld nur in die Zeilen schieben. Nun kopieren Sie die Daten in eine eigene Tabelle und binden diese in die PowerPivot ein. Vorteil ist, dass diese Tabelle historisch alles enthält und von Hand zu pflegen ist. Nachteil: Auch neue Daten müssen von Hand eingepflegt werden.

Zeilenbeschriftungen ▼				Wohort
Berlin				Berlin
Koblenz				Koblenz
München				München
Gesamtergebnis				

7. Zum letzten offenen Problem

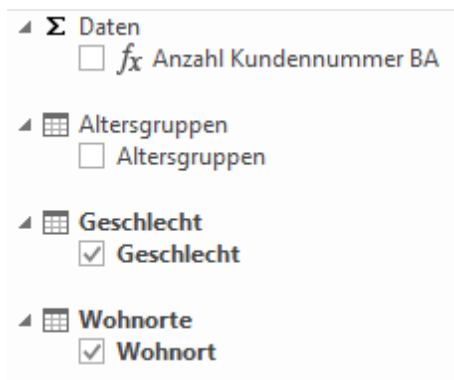
Die „Kundennummer BA“ in unserem Pivot ist weiterhin noch ein implizites Measure. Dazu erstellen wir einfach ein explizites Measure mit folgendem Code:

```
„Anzahl Kundennummer BA:=DISTINCTCOUNT([Kundennummer BA])“.
```

Nun ist eine Berechnung deutlich schneller, da PowerPivot dies übernimmt - und nicht Excel.

8. Weitere wichtige Tipps

Das obige Beispiel lässt noch einige Wünsche offen. Denn als Primär- und Fremdschlüssel-Felder sind größere Texte enthalten. Anstatt nun beispielsweise Berlin mit Berlin zu verknüpfen, empfiehlt es sich, ein Zahlenfeld als Schlüsselfeld zu wählen. In der Dimensionstabelle halten Sie zu dem Schlüssel ein entsprechendes Textfeld mit der Übersetzung. Weiter empfiehlt es sich, nicht genutzte Felder auszublenden. Das Ergebnis in Excel könnte dann beispielsweise so aussehen:



Nun sind nur noch die Dimensionstabellen sichtbar und in unserer Faktentabelle sehen wir nur noch das Measure. Das macht die Bedienung deutlich einfacher und übersichtlicher.

9. Fazit

Ein geschickt aufgebautes Datenmodell macht sich i. d. R. erst später bezahlt. Am Anfang haben Sie gegenüber der klassischen Methode einen nicht unerheblichen Aufwand. Dennoch: Wenn mehr Daten hinzukommen und wenn mehrere Anwender damit arbeiten möchten, machen Sie das Modell schneller und einfacher verständlich. Es empfiehlt sich auch, bestehende Arbeitsmappen umzubauen, selbst wenn das anfänglich viel Arbeit bedeutet. Aber ich kann Ihnen versprechen: Wenn Sie viel mit der Mappe arbeiten, wird sich dieser Aufwand sehr schnell auszahlen.

Über den Autor



Philipp Lenz ist bei der Firma prosozial GmbH als Senior Consultant im Bereich Business Intelligence tätig. Er betreut Data-Warehouse-Landschaften sowie „Self Service BI“-Umgebungen und berät Kunden beim Einsatz in diesen Technologien. Sollten Sie Fragen oder Anmerkungen hierzu haben oder auch Hilfe benötigen, kontaktieren Sie gerne Herrn Lenz: philipp.lenz@prosozial.de

© Alle Rechte vorbehalten prosozial GmbH 2015

Mit dem Copyright zeigen wir an, dass prosozial die Urheberrechte und wirtschaftlichen Nutzungsrechte für von prosozial erstellte Texte, Bilder, Filme, PDF sowie alle weiteren Medien beansprucht. Es handelt sich, wenn dieses Zeichen erscheint, um urheberrechtlich geschützte Werke. Änderungen, Löschungen bzw. Ergänzungen sind unseren Kunden im Rahmen des Servicevertrages für eigene Zwecke erlaubt, fallen aber unter die Verantwortung des jeweiligen Autors und sind als solche zu kennzeichnen. Eine Gewähr für die Richtigkeit, Angemessenheit usw. des geänderten Mediums als Ganzem sowie aller einzelnen Änderungen übernimmt prosozial nicht.