



Einleitung

Excel-Benutzer lieben Zahlen. Vielleicht ist es auch eher so, dass Zahlenliebhaber Excel lieben. Wie dem auch immer sein mag – sobald Sie daran interessiert sind, Erkenntnisse aus irgendeiner Art von Datenmenge zu gewinnen, ist es sehr wahrscheinlich, dass Sie schon viel Zeit mit Excel, Pivottabellen und Formeln verbracht haben.

Im Jahr 2015 wurde Power BI veröffentlicht. Heutzutage kann man mit Fug und Recht sagen, dass Zahlenliebhaber sowohl Power Pivot für Excel als auch Power BI lieben. Beide Tools haben viele Merkmale gemeinsam, nämlich die Datenbank-Engine VertiPaq und die von SQL Server Analysis Services geerbte Sprache DAX.

Um in früheren Excel-Versionen Erkenntnisse aus Zahlen zu gewinnen, musste man zunächst einige Datasets laden und dann Spalten berechnen und Formeln schreiben, um Diagramme zu entwerfen. Dabei gab es einige Einschränkungen: Die Größe der Arbeitsmappe war von Bedeutung, und die Formelsprache von Excel war für große Zahlen nicht gerade ideal. Die neue Engine von Power BI und Power Pivot stellt einen Riesenschritt nach vorn dar. Jetzt können Sie sämtliche Möglichkeiten einer Datenbank und eine hervorragende Sprache (nämlich DAX) nutzen. Doch diese neuen Möglichkeiten bringen auch mehr Verantwortung mit sich. Wenn Sie dieses neue Instrument wirklich nutzen wollen, müssen Sie mehr lernen, nämlich die Grundlagen der Datenmodellierung.

Datenmodellierung ist keine höhere Mathematik, sondern eine Grundfertigkeit, die jeder beherrschen sollte, der sich dafür interessiert, Erkenntnisse aus Daten zu gewinnen. Wenn Sie Zahlen mögen, dann werden Sie auch Datenmodellierung lieben. Es ist also nicht nur eine leicht zu erwerbende Fertigkeit, sondern macht auch unglaublich viel Spaß.

In diesem Buch lernen Sie die grundlegenden Konzepte der Datenmodellierung anhand von praktischen Beispielen kennen, denen Sie sehr wahrscheinlich in Ihrem täglichen Arbeitsleben begegnen. Wir wollten kein kompliziertes Buch über Datenmodellierung schreiben, in dem wir die vielen komplexen Entscheidungen, die Sie treffen müssen, um eine komplexe Lösung zu erstellen, im Detail erläutern. Stattdessen konzentrierten wir uns auf Beispiele aus unserer täglichen Beratertätigkeit. Wann immer ein Kunde uns bat, bei der Lösung eines Problems zu helfen, und wir der Meinung waren, dass dieses Problem häufig auftritt, dann hoben wir es auf. Anschließend haben wir dieses Problemarchiv geöffnet und für jedes dieser Beispiele eine Lösung bereitgestellt, die wir anschließend so geordnet haben, dass sie auch als Schulung zur Datenmodellierung dienen können.

Nachdem Sie die Lektüre dieses Buches beendet haben, werden Sie noch kein Experte für Datenmodellierung sein, haben aber bereits eine größere Sensibilität für das Thema erworben. Wenn Sie sich anschließend wieder Ihrer eigenen Datenbank zuwenden, um zu überlegen, wie Sie einen erforderlichen Wert berechnen können, und Sie dabei auf die Idee kommen, dass eine Änderung des Modells hilfreich sein könnte, dann haben wir mit diesem Buch unser Ziel erreicht. Vor allem sind Sie dann auf dem Weg, ein erfolgreicher Datenmodellierer zu werden. Diesen letzten Schritt – also ein großartiger Datenmodellierer zu werden –, können Sie nur mit viel Erfahrung und nach vielen Fehlern gehen. Leider ist Erfahrung nichts, was man in einem Buch lernen kann.

Zielgruppe

Dieses Buch hat eine sehr breite Zielgruppe. Sind Sie vielleicht ein Excel-Benutzer, der Power Pivot für Excel einsetzt, oder ein Datenforscher, der Power BI nutzt? Oder stehen Sie noch am Anfang Ihrer Karriere als Business-Intelligence-Profi und möchten eine Einführung in die Themen der Datenmodellierung lesen? In all diesen Fällen ist dies das richtige Buch für Sie.

Beachten Sie, dass wir Personen, die ein Buch über Datenmodellierung lesen möchten, nicht in diese Liste aufgenommen haben. Tatsächlich sind wir beim Schreiben dieses Buches davon ausgegangen, dass unsere Leser gar nicht wissen, dass sie Datenmodellierung brauchen. Unser Ziel ist es, Ihnen deutlich zu machen, dass Sie Datenmodellierung erlernen müssen, und Ihnen dann einige Einblicke in die Grundlagen dieser wunderbaren Wissenschaft zu geben. Kurz gesagt, wenn Sie wissen möchten, was Datenmodellierung ist und warum sie eine nützliche Fähigkeit darstellt, dann ist dies das richtige Buch für Sie.

Voraussetzungen

Wir gehen davon aus, dass unsere Leser über Grundkenntnisse in Excel-Pivottabellen verfügen oder Power BI als Berichts- und Modellierungswerkzeug einsetzen. Auch einige Erfahrungen mit der Analyse von Zahlen sind sehr hilfreich. Die Benutzeroberfläche von Excel und Power BI behandeln wir in diesem Buch nicht. Stattdessen konzentrieren wir uns auf Datenmodelle und darauf, wie man sie erstellt und abwandelt, damit sich der Code einfacher schreiben lässt. Das heißt, wir behandeln das »Was« und überlassen Ihnen das »Wie«. Wir wollten keine schrittweise Anleitung schreiben, sondern ein Buch, das anspruchsvolle Themen auf einfache Weise erklärt.

Ein Thema, das wir bewusst nicht in dem Buch behandeln, ist die Sprache DAX. Es wäre unmöglich gewesen, Datenmodellierung und DAX zusammen im selben Buch zu erklären. Wenn Sie bereits mit der Sprache vertraut sind, dann können Sie viel Nutzen aus den vielen DAX-Codebeispielen in diesem Buch ziehen. Sollten Sie DAX hingegen noch lernen müssen, dann lesen Sie *The Definitive Guide to DAX*, den umfassendsten Leitfaden zur dieser Sprache, der sich auch gut an die Themen dieses Buches anschließt.

Aufbau dieses Buches

Das Buch beginnt mit ein einigen einfachen, einleitenden Kapiteln, gefolgt von einer Reihe monographischer Kapitel, die jeweils ein bestimmtes Datenmodell behandeln. Die folgende Aufstellung gibt eine kurze Beschreibung der Kapitel:

Kapitel 1, »Einführung in die Datenmodellierung«, gibt eine kurze Einführung in die Grundprinzipien der Datenmodellierung. Hier zeigen wir, was Datenmodellierung ist, erklären Granularität und definieren die grundlegenden Modelle für Data Warehouses, nämlich Sternschema, Schneeflockenschema, Normalisierung und Denormalisierung.

Kapitel 2, »Header/Detail-Tabellen«, behandelt ein sehr weit verbreitetes Modell, nämlich Header/Detail-Tabellen. Hier finden Sie Erklärungen und Lösungen für Situationen, in denen Sie z. B. Bestellungen und deren einzelne Zeilen in zwei getrennten Faktentabellen unterbringen.

Kapitel 3, »Mehrere Faktentabellen«, beschreibt Situationen, in denen Sie mehrere Faktentabellen haben und einen Bericht erstellen müssen, der sie kombiniert. Hier betonen wir, wie wichtig es ist, das korrekte dimensionale Modell aufzustellen, um die Daten richtig durchsuchen zu können.

Kapitel 4, »Datum und Uhrzeit«, ist eines der längsten des Buches. Es behandelt Zeitinformationsberechnungen. Wir erklären, wie Sie eine korrekte Datumstabelle erstellen und grundlegende Zeitinformationsberechnungen durchführen (YTD, QTA, PARALLELPERIOD usw.). Außerdem zeigen wir einige Beispiele für die Berechnung von Arbeitstagen, die Handhabung von besonderen Zeiträumen im Jahr und den richtigen Umgang mit Kalenderdaten im Allgemeinen.

Kapitel 5, »Historische Attribute«, beschreibt die Verwendung von langsam veränderlichen Dimensionen. In diesem Kapitel werden die Transformationsschritte genauer erläutert, die Sie in Ihrem Modell durchführen müssen, wenn Sie veränderliche Attribute verfolgen müssen. Außerdem erfahren Sie, wie Sie Ihren korrekten DAX-Code für langsam veränderliche Dimensionen schreiben.

Kapitel 6, »Snapshots«, behandelt die faszinierenden Aspekte von Snapshots. Wir erklären, was Snapshots sind, warum und wann sie verwendet werden sollten und wie Sie aus einem Snapshot Werte berechnen. Außerdem stellen wir das äußerst leistungsfähige Modell der Übergangsmatrix vor.

Kapitel 7, »Datums- und Zeitintervalle«, geht noch mehrere Schritte über den Inhalt von Kapitel 5 hinaus. Wiederum geht es um Zeitberechnungen, aber diesmal analysieren wir Modelle, bei denen die in den Faktentabellen gespeicherten Ereignisse eine bestimmte Dauer haben und daher eine besondere Behandlung benötigen, um korrekte Ergebnisse zu liefern.

Kapitel 8, »m:n-Beziehungen«, erklärt, wie Sie m:n-Beziehungen einsetzen können. Solche Beziehungen spielen in jedem Datenmodell eine sehr wichtige Rolle. Wir behandeln reguläre m:n-Beziehungen, kaskadierende Beziehungen und die Verwendung von Zuordnungsfaktoren und Filtern, besprechen ihre Leistung und erklären, wie Sie sie verbessern können.

Kapitel 9, »Unterschiedliche Granularitäten«, geht eingehender auf die Arbeit mit Faktentabellen mit unterschiedlicher Granularität ein. Wir zeigen ein Beispiel mit einer Datenvorhersage, bei dem die Faktentabellen unterschiedliche Granularität aufweisen, und stellen verschiedene Lösungsmöglichkeiten mit DAX-Code und durch Anpassung des Datenmodells vor.

Kapitel 10, »Segmentierungsmodelle«, erläutert verschiedene Segmentierungsmodelle. Wir beginnen mit einer einfachen Segmentierung nach Preis, gehen dann zur dynamischen Segmentierung mithilfe virtueller Beziehungen über und erläutern schließlich die ABC-Analyse mit DAX.

Kapitel 11, »Währungsumrechnung«, behandelt den Umgang mit unterschiedlichen Währungen. Bei der Arbeit mit Wechselkursen ist es wichtig, die Anforderungen zu verstehen und das richtige Modell zu erstellen. Wir untersuchen mehrere Fälle mit unterschiedlichen Anforderungen und stellen für jeden davon die jeweils beste Lösung vor.

Anhang A, »Grundlagen der Datenmodellierung«, ist als Nachschlagewerk gedacht. Die in diesem Buch verwendeten Grundbegriffe werden anhand von Beispielen kurz beschrieben. Wann immer Sie sich unsicher sind, können Sie dorthin vorblättern, um sich zu informieren, und dann zur eigentlichen Lektüre zurückkehren.

Die Komplexität der Modelle und der Lösungen nimmt Kapitel für Kapitel zu, sodass es besser ist, dieses Buch von Anfang an zu lesen, anstatt von Kapitel zu Kapitel zu springen. So können Sie dem natürlichen Fluss der Komplexität folgen und ein Thema nach dem anderen erlernen. Nach Beendigung der Lektüre können Sie das Buch jedoch auch als Nachschlagewerk nutzen. Wenn Sie also eine Lösung für ein bestimmtes Modell benötigen, können Sie direkt zu dem Kapitel springen, in dem es behandelt wird, und dort die Einzelheiten der Lösung einsehen.

Begleitende Inhalte

Um die Arbeit mit diesem Buch zu vereinfachen, haben wir begleitende Inhalte zusammengestellt, die Sie von der folgenden Seite herunterladen können:

https://www.dpunkt.de/datenanalyse_excel

Zu diesen Inhalten gehören die Excel- und Power BI Desktop-Dateien für alle Beispiele in diesem Buch. Für jede Abbildung im Buch gibt es eine eigene Datei, sodass Sie die einzelnen Schritte genau nachvollziehen und immer auf dem gleichen Stand ausgehen können, um die Beispiele selbst auszuprobieren. Bei den meisten dieser Beispiele handelt es sich um Power BI Desktop-Dateien. Wenn Sie sie auf Ihrem eigenen PC nachvollziehen wollen, sollten Sie sich die neueste Version dieses Programms von der Power BI-Website herunterladen.

Schreibweisen und Hinweiskästen

In diesem Buch werden folgende Schreibweisen und Hinweiskästen verwendet:

Fettschrift wird für Text verwendet, den Sie eingeben.

Kursivschrift wird für neue Begriffe, für Datenbankelemente wie Tabellen, Spalten usw., für Elemente der Benutzeroberfläche sowie für URLs verwendet.

Code erscheint in nicht proportionaler Schrift.

Bei Tastaturkürzeln sind die einzelnen zu drückenden Tasten durch ein Pluszeichen verbunden. Beispielsweise bedeutet `Strg` + `Alt` + `Entf`, dass Sie gleichzeitig die Tasten `Strg`, `Alt` und `Entf` drücken müssen.



Hinweis



Tipp



Zusätzliche Informationen



Achtung

Danksagungen

Bevor wir zum Abschluss dieser Einleitung kommen, möchten wir unserer Lektorin Kate Shoup danken, die uns bei der ganzen redaktionellen Arbeit begleitet hat, sowie unserem Fachgutachter Ed Price. Ohne ihre akribische Arbeit wäre dieses Buch viel schwerer zu lesen! Dass dieses Buch weniger Fehler enthält als unser Originalmanuskript, ist diesen beiden Personen zu verdanken. Wenn es immer noch Fehler enthält, so liegt das natürlich ganz an uns.

Errata und Support

Wir haben alle Anstrengungen unternommen, um die Richtigkeit dieses Buches und der begleitenden Inhalte sicherzustellen. Alle Fehler des englischsprachigen Originals, die seit der Veröffentlichung gemeldet wurden, sind auf der Website von Microsoft Press unter der folgenden Adresse aufgeführt:

<https://aka.ms/AnalyzeData/errata>

Mit Anmerkungen, Fragen oder Verbesserungsvorschlägen zu diesem Buch können Sie sich auch in Deutsch an den dpunkt.verlag wenden:

hallo@dpunkt.de

Bitte beachten Sie, dass über unsere E-Mailadresse kein Software-Support angeboten wird. Für Supportinformationen bezüglich der hier verwendeten Microsoft-Produkte besuchen Sie die Microsoft-Website:

<http://support.microsoft.com>