Engineer

Sternschema

★ Eine Faktentabelle mit Dimensionstabellen ringsum

Was es ist

Ein Sternschema ordnet Tabellen für die Analyse an. Eine Tabelle in der Mitte erfasst Geschäftsereignisse wie Verkäufe, Sendungen oder Zahlungen als Zahlen. Darum stehen die Tabellen, die diese Ereignisse beschreiben: welches Produkt, welches Geschäft, welcher Tag, welcher Kunde. Verbindet man in einem Diagramm die mittlere Tabelle mit jeder der anderen, entsteht ein Stern.

Die mittlere Tabelle ist die Faktentabelle; die umliegenden sind Dimensionstabellen. Die Beschreibung der Kimball Group ist knapp: „Sternschemata bestehen typischerweise aus Faktentabellen, die mit zugehörigen Dimensionstabellen verknüpft sind.“ Die Verbindung ist ein gewöhnlicher Schlüssel. Jede Dimensionstabelle hat genau einen Primärschlüssel. Jede Faktenzeile führt ihn als Fremdschlüssel mit, einen für jede Dimension, die das Ereignis beschreibt.

Die Aufteilung folgt der Unterscheidung der Kimball Group zwischen Messung und Kontext. Ein dimensionales Modell wird um die Messereignisse eines Geschäftsprozesses aufgebaut: ein an der Kasse gescannter Artikel, ein versandter Auftrag, eine beglichene Forderung. Die dabei entstehenden Zahlen, etwa Menge und Betrag, sind die Fakten. Alles, was Wer, Was, Wo, Wann, Warum und Wie beschreibt, ist Kontext und steht in den Dimensionen.

Die beiden Tabellentypen haben entgegengesetzte Formen. Auf der detailliertesten Ebene enthält eine Faktentabelle eine Zeile pro Ereignis mit den numerischen Messwerten, einem Fremdschlüssel zu jeder Dimension und manchmal einer Referenz wie einer Rechnungs- oder Belegnummer. Diese Nummer ist alles, was vom Dokument übrig bleibt, sobald Datum, Kunde und Geschäft zu Dimensionen geworden sind. Sie steht deshalb in der Faktenzeile ohne eigene Dimensionstabelle dahinter; die Kimball Group nennt das eine

degenerierte Dimension. Eine Dimensionstabelle ist dagegen breit, flach und bewusst denormalisiert, mit vielen beschreibenden Textspalten. Eine Produktdimension wiederholt Marke, Kategorie und Abteilung in jeder Produktzeile, statt sie in eigene Tabellen auszulagern.

Ein Sternschema ist die relationale Form eines dimensionalen Modells, aufgebaut aus gewöhnlichen Tabellen in einer gewöhnlichen SQL-Datenbank. Die andere von der Kimball Group beschriebene Form ist der OLAP-Würfel: eine mehrdimensionale Datenbank mit denselben Attributen und Fakten, abgefragt mit analytisch mächtigeren Sprachen als SQL, etwa MDX. XMLA stellt dabei ein Zugriffsprotokoll bereit. Würfel werden häufig aus einem Sternschema aufgebaut, als letzter Schritt, bevor Benutzer die Daten sehen, oder als Aggregatschicht. Für beide gelten dieselben Modellierungsregeln. Die Kimball Group nennt sie architekturneutral: Sie gelten für ein Data Warehouse nach Kimball ebenso wie für Bill Inmons Corporate Information Factory mit ihrer Hub-and-Spoke-Architektur.

Das Konzept für eine Entscheidung nutzen

Nutzen Sie „Was ist ein Sternschema?" für eine konkrete Entscheidung. Wir können helfen, das Konzept mit dem System in Verbindung zu bringen, das Sie entwickeln oder einkaufen. Vertriebs-, Betriebs- und Finanzdaten werden zusammengeführt, damit zwei Abteilungen nicht länger mit zwei Antworten in dieselbe Besprechung kommen.

Unsere Leistungen entdecken Kontakt aufnehmen →

Die Annahmen hinter dem Begriff prüfen

Nutzen Sie die Erklärung und die verlinkten Arbeiten, um Fragen für Ihr Projekt zu formulieren. Wir können die Annahmen gemeinsam mit Ihnen prüfen, bevor ein Werkzeug ausgewählt wird. Geografische Datenpipeline 50-mal schneller — 50x schnellere Geodaten-Pipelines

Geografische Datenpipeline 50-mal schneller Kontakt aufnehmen →

Warum es wichtig ist

Die Form vereinheitlicht analytische Fragen. Man wählt die zu summierenden Fakten, verbindet die Faktentabelle mit den beschreibenden Dimensionen, filtert nach einigen Dimensionsspalten und gruppiert nach anderen. Die Kimball Group beschreibt Faktentabellen als Hauptziel von Berechnungen und dynamischen Aggregationen und Dimensionsattribute als Hauptquelle für Abfragebedingungen, Gruppierungen und Berichtsbeschriftungen. Wer dieses Muster einmal gelernt hat, kann ohne Hilfe die nächste Abfrage auf einer neuen Faktentabelle schreiben. Ein Berichtswerkzeug kann sie aus einer Feldliste aufbauen, weil jeder Stern gleich angeordnet ist.

Lesbarkeit ist der zweite Grund. Eine operative Datenbank soll jeden Sachverhalt einmal speichern und konsistent halten. Die Kategorie eines Produkts steht deshalb meist in einer separaten Nachschlagetabelle, die über eine ID verknüpft wird. Das ist das richtige Design für ein System, das Aufträge erfasst, aber das falsche für jemanden, der Umsätze je Kategorie sucht: Zuerst muss diese Person herausfinden, welche Tabellen in welcher Reihenfolge zu verbinden sind. Eine flache Dimension versammelt alle beschreibenden Spalten an einem Ort, in ausgeschriebenen Wörtern. Die Kimball Group warnt vor der Alternative, die Hierarchie einer Dimension in weitere Tabellen zu normalisieren, weil Fachanwender das Ergebnis schwer verstehen und sich darin nur mühsam zurechtfinden.

Der dritte Grund: Ein gut gebauter Stern kann Fragen beantworten, die beim Entwurf niemand gestellt hat. Die Granularität einer Faktentabelle, also die genaue Bedeutung einer Zeile, wird vor jeder Dimension und jedem Fakt festgelegt. Die Kimball Group empfiehlt den Einstieg auf atomarer Ebene, der niedrigsten vom Geschäftsprozess erfassten Stufe. Eine Tabelle mit einer Zeile pro Belegposition lässt sich nach Tag, Geschäft, Marke oder jeder Kombination daraus summieren. Eine Tabelle mit bloßen Tagessummen je Geschäft lässt sich niemals nach Produkt aufschlüsseln, unabhängig von der Formulierung der Frage. In der Terminologie der Kimball Group halten atomare Daten unvorhersehbaren Abfragen stand, während verdichtete Daten die späteren Fragen voraussetzen.

Der Preis ist gewollte Redundanz. Der Abteilungsname wiederholt sich in jeder Produktzeile. Der Ladeprozess, der den Stern aufbaut, muss diese Wiederholungen konsistent halten und festlegen, was bei einer geänderten Beschreibung geschieht. Der Stern ersetzt auch nicht das operative Modell. Ein Ladeprozess baut ihn daraus auf, und er ist nur so aktuell wie dessen letzter Lauf.

Wie es in der Praxis aussieht

Ein Einzelhändler braucht Umsätze nach Produktkategorie und Monat. Der Geschäftsprozess ist der Kassenbeleg, und die Granularität wird als eine Zeile pro Belegposition festgelegt. Die Produktdimension ist eine flache Tabelle. Die Faktentabelle enthält einen Schlüssel pro Dimension, die Belegnummer als degenerierte Dimension und zwei Messwerte.

```
CREATE TABLE dim_product (
    product_key  bigint PRIMARY KEY,   -- surrogate key, owned by the warehouse
    sku          text NOT NULL,        -- the code in the source system
    product_name text NOT NULL,
    brand        text NOT NULL,
    category     text NOT NULL,
    department   text NOT NULL
);

CREATE TABLE fact_sales (
    date_key       integer NOT NULL REFERENCES dim_date,
    store_key      integer NOT NULL REFERENCES dim_store,
    product_key    bigint  NOT NULL REFERENCES dim_product,
    receipt_number text    NOT NULL,   -- degenerate dimension
    quantity       integer NOT NULL,
    sales_amount   numeric(12, 2) NOT NULL
);

SELECT d.year_month, p.category, sum(f.sales_amount) AS sales
FROM fact_sales AS f
JOIN dim_date AS d USING (date_key)
JOIN dim_product AS p USING (product_key)
WHERE p.department = 'Grocery'
GROUP BY d.year_month, p.category
ORDER BY d.year_month, p.category;
```

Die Abfrage hat die Form jeder Sternabfrage: Faktentabelle, ein Join pro Dimension, Filter und Gruppierungen auf Dimensionstext und eine Summe über einen Fakt. Für die Region des Geschäfts kommen ein weiterer Join und eine weitere Spalte in `GROUP BY` hinzu; die Faktentabelle ändert sich nicht. Die Belegnummer bleibt in der Faktenzeile, damit die Positionen eines Warenkorbs weiterhin gemeinsam gezählt werden können. Sie bekommt aber keine Dimensionstabelle, weil es dazu nichts Weiteres zu beschreiben gibt.

Ein Stern ist eine Antwort auf das Versprechen des Angebots für Datenbankentwurf und Datenmodellierung: eine Struktur, die die tatsächlichen Abläufe des Geschäfts abbildet, damit sich später benötigte Fragen weiterhin stellen lassen. Das operative Modell, aus dem er aufgebaut wird, ist die andere Hälfte dieser Arbeit. In ihrer laufenden Arbeit für NextMariner hat die Gründerin eine maritime Domäne in 348 normalisierten Tabellen modelliert, verteilt auf 34 PostgreSQL-Schemas und mit Fachkräften, Unternehmen, Schiffen, Stellen und Bewertungen. Kleine stabile Listen wie Statuswerte, Typen und Kategorien wurden dort zu SMALLINT-Nachschlagetabellen. Das hält die Zeilen kompakt und Joins günstig; echte Fremdschlüssel und Constraints machen die Integrität zur Aufgabe der Datenbank. Das ist die richtige Form für ein System, das Zeile für

Zeile schreibt und korrekt bleiben muss. Ein darauf aufbauendes Sternschema für Berichte geht den umgekehrten Weg: Es macht die Codes in flachen Dimensionen wieder zu Wörtern, weil ein Mensch und nicht die Anwendung es liest.

Wo es schiefgeht

Die Dimensionen werden normalisiert und aus dem Stern wird eine Schneeflocke. Marke, Kategorie und Abteilung aus der Produktdimension in eigene Tabellen auszulagern wirkt auf jemanden mit Erfahrung im operativen Entwurf ordentlich. Die Kimball Group nennt das Ergebnis Schneeflocke und rät davon ab: Fachanwender verstehen es nur schwer und finden sich darin mühsam zurecht, und zusätzliche Joins können Abfragen verlangsamen, während eine flache Dimension genau dieselben Informationen enthält. Halten Sie jede Dimension als eine breite Tabelle und lassen Sie den Ladeprozess die Hierarchie in jeder Zeile wiederholen.

Die Granularität wird auf den ersten Bericht zugeschnitten. Eine Faktentabelle mit Tagesumsätzen je Geschäft beantwortet die Fragen des Dashboards, für das sie gebaut wurde, aber keine unterhalb dieser Ebene. Die erste Frage nach einem einzelnen Produkt oder Warenkorb verlangt einen neuen Ladevorgang aus der Quelle. Noch schlechter ist eine Tabelle, die Tagessummen und Einzelpositionen mischt, weil jede Summe darüber denselben Verkauf doppelt zählt. Legen Sie zuerst die Granularität fest, beginnen Sie auf der vom Prozess erfassten atomaren Ebene und geben Sie jeder Granularität eine eigene Faktentabelle. Separate Zusammenfassungstabellen können anschließend zur Beschleunigung darauf aufsetzen.

Die Dimensionsattribute sind die Codes des Quellsystems. Operative Systeme speichern kompakte Codes. Werden sie direkt in eine Dimension kopiert, entstehen Berichte mit Zeilenbeschriftungen, die nur das Quellsystem versteht. Die Kimball Group erlaubt Codes als Attribute, bezeichnet aber ausführliche Beschreibungen als die nützlichsten Attribute, weil diese später die Berichtsbeschriftungen bilden. Übersetzen Sie beim Laden jeden Code in eine ausgeschriebene Beschreibung und behalten Sie den Code in einer Spalte neben der Beschreibung für alle, die eine Zeile zu ihrer Quelle zurückverfolgen müssen.

Die Faktentabelle sieht aus wie ein Bericht. Es liegt nahe, sie anhand der bereits verwendeten Tabellenkalkulation zu entwerfen, mit einer Spalte pro Monat oder einer Zeile pro Position des Managementberichts. Die Kimball Group

stellt ausdrücklich klar, dass der Entwurf einer Faktentabelle auf der gemessenen physischen Aktivität beruht, nicht auf den daraus abgeleiteten Berichten. Berichtslayouts ändern sich; der Kassenbeleg nicht. Modellieren Sie das Ereignis mit einer Zeile pro Vorkommen auf der festgelegten Granularität und lassen Sie Abfragen die Layouts erzeugen.

Weiterführende Lektüre

- Star Schemas and OLAP Cubes — die Definition, Fakten und Dimensionen als Messung und Kontext sowie die Beziehung von Würfeln zu Sternen.
- Grain — festlegen, was eine Faktenzeile bedeutet und warum die atomare Granularität zuerst kommt.
- Fact Table Structure — numerische Messwerte und Fremdschlüssel in einer Zeile pro Ereignis, von der Aktivität statt vom Bericht aus entworfen.
- Dimension Table Structure — breite, flache, denormalisierte Tabellen, deren beschreibende Attribute Filter, Gruppierungen und Beschriftungen bestimmen.
- Snowflaked Dimensions — warum die Normalisierung einer Dimensionshierarchie ihre Nutzung erschwert.
- Microsoft: Datenzugriff auf mehrdimensionale Modelle — MDX als Abfrage- und Berechnungssprache und XMLA als Schnittstelle, über die ein Client eine MDX-Anweisung senden kann.

Glossar

Für Ihre tatsächlichen Rahmenbedingungen auswählen

Bringen Sie einen repräsentativen Datensatz, die Fragen, die er beantworten muss, und den aktuellen Engpass mit. Wir können vereinbaren, wie Qualität, Durchsatz und Wiederherstellung gemessen werden.

Preise ansehen Kontakt aufnehmen →

Die Begriffe praktisch nutzbar machen

Fragen Sie, wie „Was ist ein Sternschema?“ auf Ihre Organisation zutrifft. Ein konkretes Beispiel aus Ihrem bestehenden System reicht aus, um das Gespräch zu beginnen.

Unsere Leistungen entdecken

welcome@engineer.company +45 50 10 61 84 Kopenhagen, Dänemark

Nehmen Sie diese Seite mit

Kontaktkarte

Mit Sorgfalt und Liebe gestaltet von D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Copyright © 2025 – heute · Engineer ApS · Alle Rechte vorbehalten.

Gedruckt von https://engineer.company/de/glossary/star-schema/