Engineer

Stjärnschema

★ en modell med en faktatabell och dimensionstabeller runt den

Vad det är

Ett stjärnschema är ett sätt att lägga upp tabeller för analys. En tabell i mitten registrerar affärshändelser, som försäljningar, leveranser eller betalningar, i form av siffror. Runt den står tabellerna som beskriver händelserna: vilken produkt, vilken butik, vilken dag, vilken kund. Ritar man det som ett diagram, med ett streck från tabellen i mitten till var och en av de andra, ser det ut som en stjärna.

Tabellen i mitten kallas faktatabellen, och tabellerna runt den kallas dimensionstabeller. Kimball Group beskriver stjärnscheman just så: faktatabeller som är kopplade till de dimensionstabeller som hör till dem. Kopplingen är en helt vanlig nyckel. Varje dimensionstabell har en enda primärnyckel, och varje rad i faktatabellen bär den nyckeln som främmande nyckel, en för varje dimension som beskriver händelsen.

Uppdelningen följer en åtskillnad som Kimball Group gör mellan mätning och sammanhang. En dimensionell modell byggs kring mäthändelserna i en affärsprocess: en vara skannas i kassan, en order skickas, en försäkringsersättning betalas ut. Siffrorna som händelsen ger, som antal och belopp, är fakta. Allt som säger vem, vad, var, när, varför och hur är sammanhang, och det hör hemma i dimensionerna.

De två tabelltyperna har motsatt form. På den mest detaljerade nivån har en faktatabell en rad per händelse, med de numeriska måtten, en främmande nyckel till varje dimension och ibland ett referensnummer, som ett faktura- eller kvittonummer. När dokumentets datum, kund och butik har blivit dimensioner är numret allt som finns kvar av det, så det lagras på faktaraden utan någon dimensionstabell bakom sig. Kimball Group kallar det en degenererad dimension. En dimensionstabell är däremot bred, platt och medvetet denormaliserad, med

många beskrivande textkolumner. En produktdimension upprepar varumärke, kategori och avdelning på varje produktrad i stället för att bryta ut dem i egna tabeller.

Ett stjärnschema är den relationella formen av en dimensionell modell, byggd av vanliga tabeller i en vanlig SQL-databas. Den andra formen som Kimball Group beskriver är OLAP-kuben, en flerdimensionell databas med samma attribut och fakta, som frågas med språk som har mer analytisk kraft än SQL, till exempel MDX, medan XMLA tillhandahåller ett protokoll för åtkomst. Kuber byggs ofta från ett stjärnschema, antingen som ett sista steg innan användarna ser data eller som ett lager av aggregat. Modelleringsreglerna är desamma för båda formerna, och Kimball Group kallar dem arkitekturneutrala: de gäller lika väl för ett analysdatalager i Kimballs anda som för Bill Inmons Corporate Information Factory, som är uppbyggd som ett nav med ekrar.

Använd begreppet i ett beslut

Använd ”Vad är ett stjärnschema?” i ett konkret beslut. Vi kan knyta begreppet till systemet ni bygger eller köper. Siffror från försäljning, drift och ekonomi samlade på ett ställe, så att två avdelningar slutar komma med två olika svar till samma möte.

Se våra tjänster Hör av er →

Granska antagandena bakom ordet

Använd förklaringen och det länkade arbetet för att formulera frågor till projektet. Vi kan undersöka antagandena tillsammans innan ett verktyg väljs. Gjorde pipelinen för geodata 50 gånger snabbare — 50x snabbare datapipelines för geodata

Gjorde pipelinen för geodata 50 gånger snabbare Hör av er →

Varför det spelar roll

Formen gör att analytiska frågor liknar varandra. Välj de fakta som ska summeras, koppla faktatabellen till dimensionerna som beskriver dem, filtrera på några dimensionskolumner och gruppera på andra. Kimball Group beskriver faktatabeller som det främsta målet för beräkningar och dynamiska aggregeringar, och dimensionsattribut som den främsta källan till villkor, grupperingar och etiketter i rapporter. En analytiker som har lärt sig mönstret en gång kan skriva nästa fråga mot en ny faktatabell utan hjälp, och ett rapportverktyg kan bygga den från en lista med fält, eftersom alla stjärnor är uppbyggda på samma sätt.

Läsbarhet är det andra skälet. En driftdatabas är gjord för att lagra varje uppgift en gång och hålla den konsekvent, så en produkts kategori ligger gärna i en egen uppslagstabell som kopplas tillbaka via ett id. Det är rätt design för systemet som tar emot order och fel design för den som letar efter försäljning per kategori och

först måste lista ut vilka tabeller som ska kopplas, och i vilken ordning. En platt dimension samlar alla beskrivande kolumner på ett ställe, utskrivna i ord. Kimball Group avråder från alternativet, att normalisera en dimensions hierarki ut i fler tabeller, eftersom verksamhetens användare har svårt att förstå och navigera i resultatet.

Det tredje skälet är att en välbyggd stjärna kan besvara frågor som ingen ställde när den utformades. Granulariteten i en faktatabell, alltså den exakta betydelsen av en rad, bestäms före varje dimension och varje faktum, och Kimball Group rekommenderar att börja på atomär granularitet: den lägsta nivå som affärsprocessen registrerar. En tabell där varje rad är en kvittorad kan summeras per dag, butik, varumärke eller vilken kombination av dem som helst. En tabell som bara har dagliga summor per butik kan aldrig brytas ned per produkt, hur frågan än formuleras. Med Kimball Groups ord tål atomära data oförutsägbara frågor, medan sammanfattade data förutsätter vilka frågor de ska få.

Priset är medveten redundans. Avdelningsnamnet upprepas på varje produktrad, och laddningen som bygger stjärnan måste hålla upprepningarna konsekventa och avgöra vad som ska hända när en beskrivning ändras. Stjärnan ersätter inte heller driftmodellen; den byggs från driftmodellen genom en laddningsprocess och är aldrig färskare än den laddningens senaste körning.

Så ser det ut i praktiken

En detaljhandelskedja vill se försäljning per produktkategori och månad. Affärsprocessen är kassakvittot, och granulariteten bestäms till en rad per kvittorad. Produktdimensionen är en enda platt tabell, och faktatabellen har en nyckel för varje dimension, kvittonumret som degenererad dimension och två mått.

```
CREATE TABLE dim_product (
    product_key  bigint PRIMARY KEY,   -- surrogatnyckel, ägs av lagret
    sku          text NOT NULL,        -- koden i källsystemet
    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,   -- degenererad 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 = 'Dagligvaror'
GROUP BY d.year_month, p.category
ORDER BY d.year_month, p.category;
```

Frågan har den form som alla stjärnfrågor har: faktatabellen, en koppling per dimension, filter och grupperingar på dimensionstext och en summa över ett faktum. Att lägga till butikens region i rapporten är en koppling och en kolumn till i `GROUP BY`; faktatabellen ändras inte. Kvittonumret står kvar på faktaraden, så att raderna i en kundkorg fortfarande kan räknas ihop, men det får ingen dimensionstabell, eftersom det inte finns något mer att säga om det.

Ett stjärnschema är ett svar på löftet i tjänsten databasdesign och datamodellering: en struktur som speglar hur verksamheten faktiskt fungerar, så att de frågor man behöver svar på senare fortfarande går att ställa. Driftmodellen som stjärnan byggs från är den andra halvan av det arbetet. I sitt pågående arbete för NextMariner har grundaren modellerat en sjöfartsdomän i 348 normaliserade tabeller i 34 PostgreSQL-scheman, med yrkesverksamma, företag, fartyg, jobb och recensioner. Där blev små, stabila uppräkningar som statusar, typer och kategorier uppslagstabeller med SMALLINT, vilket håller raderna kompakta och joinarna billiga, och riktiga främmande nycklar och villkor gör integriteten till databasens uppgift. Det är rätt form för ett system som skriver en rad i taget och måste förbli korrekt. En stjärna som byggs för rapportering ovanpå ett sådant system går åt andra hållet: den gör om koderna till ord igen i platta dimensioner, eftersom läsaren är en människa och inte applikationen.

Var det går fel

Dimensionerna normaliseras, och stjärnan blir en snöflinga. Att bryta ut varumärke, kategori och avdelning ur produktdimensionen i egna tabeller ser prydligt ut för alla som är vana vid att designa driftsystem. Kimball Group kallar resultatet en snöflinga (snowflake) och avråder från det: verksamhetens användare har svårt att förstå och navigera i den, och de extra kopplingarna kan

göra frågor långsammare, medan en platt dimension innehåller exakt samma information. Håll varje dimension som en enda bred tabell och låt laddningen upprepa hierarkin på varje rad.

Granulariteten anpassas efter den första rapporten. En faktatabell med daglig försäljning per butik svarar på den dashboard den byggdes för och på ingenting under den; den första frågan om en enskild produkt eller kundkorg kräver en ny laddning från källan. En tabell som blandar dagliga summor med enskilda rader är ännu värre, eftersom varje summa över den räknar samma försäljning två gånger. Bestäm granulariteten först, börja på den atomära nivå som processen registrerar och ge varje granularitet en egen faktatabell. Sammanfattningstabeller kan sedan läggas ovanpå för snabbhetens skull, som separata tabeller.

Dimensionsattributen är källsystemets koder. Driftsystem lagrar kompakta koder, och kopierar man dem rakt in i en dimension blir det rapporter med radetiketter som bara källsystemet förstår. Kimball Group tillåter koder som attribut men säger att de mest användbara attributen är utförliga beskrivningar, eftersom det är de som hamnar som etiketter i rapporten. Översätt varje kod till ord under laddningen och behåll koden i en kolumn bredvid beskrivningen för den som behöver spåra en rad tillbaka till källan.

Faktatabellen är formad som en rapport. Det är frestande att utforma den efter det kalkylark som verksamheten redan använder, med en kolumn per månad eller en rad per post i ledningsrapporten. Kimball Group är tydlig med att en faktatabells utformning bygger på den fysiska aktivitet som mäts, inte på de rapporter som senare tas fram ur den. Rapportlayouter ändras; kassakvittot gör det inte. Modellera händelsen, en rad per förekomst på den bestämda granulariteten, och låt frågorna ta fram layouterna.

Läs mer

- Star Schemas and OLAP Cubes — definitionen, fakta och dimensioner som mätning och sammanhang, och hur kuber hänger ihop med stjärnscheman.
- Grain — vad en rad i faktatabellen betyder och varför atomär granularitet kommer först.
- Fact Table Structure — numeriska mått och främmande nycklar på en rad per händelse, utformade efter aktiviteten och inte efter rapporten.
- Dimension Table Structure — breda, platta, denormaliserade tabeller vars beskrivande attribut styr filter, grupperingar och etiketter.

- Snowflaked Dimensions — varför en dimension blir svårare att använda när dess hierarki normaliseras.
- Microsoft: Dataåtkomst för flerdimensionella modeller — MDX som fråge- och beräkningsspråk och XMLA som gränssnittet genom vilket en klient kan skicka en MDX-instruktion.

Ordlista

Välj utifrån era faktiska förutsättningar

Ta med ett representativt dataset, frågorna det ska besvara och dagens flaskhals. Vi kan avtala mått på kvalitet, genomströmning och återställning.

Se vad det kostar Hör av er →

Gör begreppen praktiska

Fråga hur ”Vad är ett stjärnschema?” gäller i er organisation. Ett konkret exempel från ert nuvarande system räcker för att börja samtalet.

Se våra tjänster

welcome@engineer.company +45 50 10 61 84 Köpenhamn, Danmark

Ta med dig den här sidan

Visitkort

Omsorgsfullt skapad med kärlek av D.

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

Upphovsrätt © 2025 – nu · Engineer ApS · Alla rättigheter förbehållna.

Utskrivet från https://engineer.company/sv/glossary/star-schema/