Engineer

Stjerneskema

★ en model med én faktatabel og dimensionstabeller omkring den

Hvad det er

Et stjerneskema, på engelsk star schema, er en måde at lægge tabeller an på, når data skal analyseres. En tabel i midten registrerer forretningens hændelser, fx salg, forsendelser eller betalinger, som tal. Rundt om den står de tabeller, der beskriver hændelserne: hvilket produkt, hvilken butik, hvilken dag, hvilken kunde. Tegner man det som et diagram med en streg fra tabellen i midten ud til hver af de andre, ligner det en stjerne.

Tabellen i midten er faktatabellen, og tabellerne rundt om er dimensionstabeller. Kimball Group, der har beskrevet teknikkerne bag dimensionel modellering, forklarer stjerneskemaet som faktatabeller, der er koblet til de dimensionstabeller, som hører til dem. Koblingen er en helt almindelig nøgle. Hver dimensionstabel har én primærnøgle, og hver række i faktatabellen bærer den nøgle som fremmednøgle, én for hver dimension, der beskriver hændelsen.

Opdelingen følger en sondring, Kimball Group gør mellem måling og kontekst. En dimensionel model bygges op om de målbare hændelser i en forretningsproces: en vare, der scannes ved kassen, en ordre, der afsendes, en erstatning, der udbetales. De tal, hændelsen giver, fx antal og beløb, er fakta. Alt det, der fortæller hvem, hvad, hvor, hvornår, hvorfor og hvordan, er kontekst, og det ligger i dimensionerne.

De to slags tabeller har hver sin form. På det mest detaljerede niveau har en faktatabel én række pr. hændelse med de numeriske mål, en fremmednøgle til hver dimension og nogle gange et referencenummer som et faktura- eller bonnummer. Det nummer er alt, hvad der er tilbage af dokumentet, når dets dato, kunde og butik er blevet til dimensioner, så det gemmes i faktarækken uden nogen dimensionstabel bag sig, og Kimball Group kalder det en degenereret dimension. En dimensionstabel er derimod bred, flad og bevidst

denormaliseret, med mange beskrivende tekstkolonner. En produktdimension gentager mærke, kategori og afdeling på hver eneste produktrække i stedet for at lægge dem ud i egne tabeller.

Et stjerneskema er den relationelle udgave af en dimensionel model, bygget af almindelige tabeller i en almindelig SQL-database. Den anden udgave, Kimball Group beskriver, er OLAP-kuben: en multidimensionel database med de samme attributter og fakta, som man forespørger med sprog, der har flere analytiske muligheder end SQL, fx MDX, mens XMLA fungerer som adgangsprotokol. Kuber bygges ofte ud fra et stjerneskema, enten som sidste trin, før brugerne får data at se, eller som et lag af aggregeringer. Modelleringsreglerne er de samme for begge, og Kimball Group kalder dem arkitekturneutrale: de gælder både i et data warehouse efter Kimballs model og i Bill Inmons Corporate Information Factory med dens hub-and-spoke-opbygning.

Brug begrebet i en beslutning

Brug »Hvad er et stjerneskema?« i en konkret beslutning. Vi kan knytte begrebet til det system, I bygger eller køber. Tal fra salg, drift og økonomi samlet ét sted, så to afdelinger ikke længere kommer til det samme møde med hver sit svar.

Se vores ydelser Kontakt os →

Undersøg antagelserne bag ordet

Brug forklaringen og det tilknyttede arbejde til at formulere spørgsmål til jeres projekt. Vi kan undersøge antagelserne sammen, før et værktøj vælges. Har gjort en geodatapipeline 50 gange hurtigere — 50x hurtigere geografiske datapipelines

Har gjort en geodatapipeline 50 gange hurtigere Kontakt os →

Hvorfor det betyder noget

Formen får analytiske spørgsmål til at ligne hinanden. Man vælger de fakta, der skal lægges sammen, joiner faktatabellen med de dimensioner, der beskriver dem, filtrerer på nogle dimensionskolonner og grupperer efter andre. Kimball Group beskriver faktatabeller som det primære mål for beregninger og dynamiske aggregeringer og dimensionernes attributter som den primære kilde til betingelser, grupperinger og etiketter i rapporterne. En analytiker, der har lært mønsteret én gang, kan skrive den næste forespørgsel mod en ny faktatabel uden hjælp, og et rapportværktøj kan bygge den ud fra en liste af felter, fordi alle stjerneskemaer er bygget op på samme måde.

Læsbarhed er den anden grund. En driftsdatabase er designet til at gemme hver oplysning én gang og holde den konsistent, så et produkts kategori ligger typisk i en separat opslagstabel, der joines tilbage via et id. Det er det rigtige design til systemet, der tager imod ordrer, og det forkerte til den, der leder efter salg pr.

kategori og først skal regne ud, hvilke tabeller der skal joines, og i hvilken rækkefølge. En flad dimension samler alle beskrivende kolonner ét sted, skrevet ud i ord. Kimball Group fraråder alternativet, at normalisere en dimensions hierarki ud i flere tabeller, fordi forretningsbrugere har svært ved at forstå og finde rundt i resultatet.

Den tredje grund er, at et velbygget stjerneskema kan besvare spørgsmål, som ingen stillede, da det blev designet. Faktatabellens granularitet, altså præcis hvad én række betyder, fastlægges før nogen dimension eller noget faktum, og Kimball Group anbefaler at starte på atomar granularitet: det laveste niveau, forretningsprocessen registrerer. En tabel med én række pr. bonlinje kan summeres pr. dag, butik, mærke eller enhver kombination af dem. En tabel med kun daglige totaler pr. butik kan aldrig brydes ned på produkt, uanset hvordan spørgsmålet formuleres. Kimball Groups pointe er, at atomare data holder til uforudsigelige forespørgsler, mens sammenfattede data forudsætter de spørgsmål, de skal besvare.

Prisen er bevidst redundans. Afdelingens navn gentages på hver produktrække, og den indlæsning, der bygger stjerneskemaet, skal holde gentagelserne ens og afgøre, hvad der sker, når en beskrivelse ændres. Stjerneskemaet erstatter heller ikke driftsmodellen. Det bygges ud fra den af en indlæsningsproces og er aldrig mere aktuelt end indlæsningens seneste kørsel.

Sådan ser det ud i praksis

En detailkæde vil se salg pr. produktkategori og måned. Forretningsprocessen er kassebonen, og granulariteten fastlægges som én række pr. bonlinje. Produktdimensionen er én flad tabel, og faktatabellen har en nøgle til hver dimension, bonnummeret som degenereret dimension og to mål.

```
CREATE TABLE dim_product (
    product_key  bigint PRIMARY KEY,   -- surrogatnøgle, ejet af warehouset
    sku          text NOT NULL,        -- koden i kildesystemet
    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,   -- degenereret 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 = 'Dagligvarer'
GROUP BY d.year_month, p.category
ORDER BY d.year_month, p.category;
```

Forespørgslen har den form, alle forespørgsler mod et stjerneskema har: faktatabellen, én join pr. dimension, filtre og grupperinger på dimensionernes tekst og en sum over et faktum. Skal butikkens region med i rapporten, er det én join og én kolonne mere i `GROUP BY`; faktatabellen ændres ikke. Bonnummeret bliver i faktarækken, så linjerne fra én indkøbskurv stadig kan tælles sammen, men det får ingen dimensionstabel, for der er ikke mere at sige om det.

Et stjerneskema er ét svar på løftet bag ydelsen databasedesign og datamodellering: en struktur, der passer til, hvordan forretningen faktisk fungerer, så de svar, man får brug for senere, stadig kan hentes. Driftsmodellen, som stjerneskemaet bygges ud fra, er den anden halvdel af arbejdet. I sit nuværende arbejde for NextMariner har grundlæggeren modelleret et maritimt domæne i 348 normaliserede tabeller fordelt på 34 PostgreSQL-skemaer med fagfolk, virksomheder, skibe, jobs og anmeldelser. Dér blev de mange små, stabile lister som statusser, typer og kategorier til SMALLINT-opslagstabeller, hvilket holder rækkerne kompakte og joins billige, og rigtige fremmednøgler og constraints gør integriteten til databasens opgave. Det er den rigtige form til et system, der skriver én række ad gangen og skal forblive korrekt. Et stjerneskema bygget til rapportering oven på sådan et system går den modsatte vej: Det oversætter koderne tilbage til ord i flade dimensioner, fordi dets læser er et menneske og ikke applikationen.

Hvor det går galt

Dimensionerne bliver normaliseret, og stjernen bliver til et snefnug. At lægge mærke, kategori og afdeling ud af produktdimensionen i egne tabeller ser pænt ud for den, der er vant til at designe driftssystemer. Kimball Group kalder resultatet et snowflake-skema og fraråder det: Forretningsbrugere har svært ved at forstå og finde rundt i det, og de ekstra joins kan gøre forespørgslerne

langsommere, mens en flad dimension rummer nøjagtig de samme oplysninger. Hold hver dimension som én bred tabel, og lad indlæsningen gentage hierarkiet på hver række.

Granulariteten bliver valgt, så den passer til den første rapport. En faktatabel med dagligt salg pr. butik besvarer det dashboard, den blev bygget til, og intet under det; det første spørgsmål om et enkelt produkt eller en enkelt indkøbskurv kræver en ny indlæsning fra kilden. En tabel, der blander daglige totaler med enkelte linjer, er endnu værre, fordi enhver sum over den tæller det samme salg to gange. Fastlæg granulariteten først, start på det atomare niveau, processen registrerer, og giv hver granularitet sin egen faktatabel. Opsummeringstabeller kan derefter lægges ovenpå for hastighedens skyld, som separate tabeller.

Dimensionernes attributter er kildesystemets koder. Driftssystemer gemmer kompakte koder, og kopieres de direkte over i en dimension, får rapporterne rækkeetiketter, som kun kildesystemet forstår. Kimball Group tillader koder som attributter, men de mest nyttige attributter er ifølge dem udførlige beskrivelser, fordi det er dem, der ender som etiketter i en rapport. Oversæt hver kode til ord under indlæsningen, og behold koden i en kolonne ved siden af beskrivelsen til den, der skal spore en række tilbage til kilden.

Faktatabellen er formet som en rapport. Det er fristende at designe den ud fra det regneark, forretningen allerede bruger, med en kolonne pr. måned eller en række pr. linje i ledelsesrapporten. Kimball Group er klar i mælet: En faktatabels design hviler på den fysiske aktivitet, der måles, ikke på de rapporter, der senere trækkes ud af den. Rapportlayouts skifter, men kassebonen gør ikke. Modellér hændelsen, én række pr. forekomst på den fastlagte granularitet, og lad forespørgslerne levere opstillingen.

Læs mere

- Star Schemas and OLAP Cubes — definitionen, fakta og dimensioner som måling og kontekst, og hvordan kuber hænger sammen med stjerneskemaer.
- Grain — hvad én række i faktatabellen betyder, og hvorfor den atomare granularitet kommer først.
- Fact Table Structure — numeriske mål og fremmednøgler i én række pr. hændelse, designet ud fra aktiviteten og ikke rapporten.
- Dimension Table Structure — brede, flade, denormaliserede tabeller, hvis beskrivende attributter styrer filtre, grupperinger og etiketter.

- Snowflaked Dimensions — hvorfor det gør en dimension sværere at bruge at normalisere dens hierarki.
- Microsoft: Dataadgang til multidimensionelle modeller — MDX som forespørgsels- og beregningssprog og XMLA som grænsefladen, hvorigennem en klient kan sende en MDX-sætning.

Ordliste

Vælg ud fra jeres faktiske vilkår

Medbring et repræsentativt datasæt, de spørgsmål det skal besvare, og den aktuelle flaskehals. Vi kan aftale målinger af kvalitet, gennemløb og gendannelse.

Se, hvad det koster Kontakt os →

Gør begreberne praktiske

Spørg, hvordan »Hvad er et stjerneskema?« gælder i jeres organisation. Et konkret eksempel fra jeres nuværende system er nok til at begynde samtalen.

Se vores ydelser

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

Tag denne side med dig

Visitkort

Lavet med kærlighed af D.

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

Copyright © 2025 – nu · Engineer ApS · Alle rettigheder forbeholdes.

Udskrevet fra https://engineer.company/da/glossary/star-schema/