Engineer

Stjerneskjema

★ en modell med én faktatabell og dimensjonstabeller rundt den

Hva det er

Et stjerneskjema er en måte å legge opp tabeller for analyse på. Én tabell i midten registrerer forretningshendelser, som salg, forsendelser eller betalinger, i form av tall. Rundt den står tabellene som beskriver hendelsene: hvilket produkt, hvilken butikk, hvilken dag, hvilken kunde. Tegner man det som et diagram, med en strek fra tabellen i midten til hver av de andre, ser det ut som en stjerne.

Tabellen i midten kalles faktatabellen, og tabellene rundt den kalles dimensjonstabeller. Kimball Group beskriver stjerneskjemaer nettopp slik: faktatabeller koblet til de dimensjonstabellene som hører til dem. Koblingen er en helt vanlig nøkkel. Hver dimensjonstabell har én primærnøkkel, og hver rad i faktatabellen bærer den nøkkelen som fremmednøkkel, én for hver dimensjon som beskriver hendelsen.

Oppdelingen følger et skille Kimball Group trekker mellom måling og kontekst. En dimensjonell modell bygges rundt målehendelsene i en forretningsprosess: en vare skannes i kassen, en ordre sendes, et forsikringskrav utbetales. Tallene hendelsen gir, som antall og beløp, er faktaene. Alt som sier hvem, hva, hvor, når, hvorfor og hvordan, er kontekst, og det hører hjemme i dimensjonene.

De to tabelltypene har motsatt form. På det mest detaljerte nivået har en faktatabell én rad per hendelse, med de numeriske målene, en fremmednøkkel til hver dimensjon og noen ganger et referansenummer, som et faktura- eller kvitteringsnummer. Når dokumentets dato, kunde og butikk har blitt dimensjoner, er nummeret alt som er igjen av det, så det lagres i faktaraden uten noen dimensjonstabell bak seg. Kimball Group kaller det en degenerert dimensjon. En dimensjonstabell er derimot bred, flat og bevisst denormalisert,

med mange beskrivende tekstkolonner. En produktdimensjon gjentar merke, kategori og avdeling på hver produktrad i stedet for å skille dem ut i egne tabeller.

Et stjerneskjema er den relasjonelle formen av en dimensjonell modell, bygd av vanlige tabeller i en vanlig SQL-database. Den andre formen Kimball Group beskriver, er OLAP-kuben, en flerdimensjonal database med de samme attributtene og faktaene, som spørres med språk som har mer analytisk kraft enn SQL, for eksempel MDX, mens XMLA fungerer som en protokoll for tilgang. Kuber bygges ofte fra et stjerneskjema, enten som et siste steg før brukerne ser dataene eller som et lag med aggregater. Modelleringsreglene er de samme for begge formene, og Kimball Group kaller dem arkitekturnøytrale: de gjelder like godt for et datavarehus i Kimball-stil som for Bill Inmons Corporate Information Factory, som er bygd som nav og eiker.

Bruk begrepet i en beslutning

Bruk «Hva er et stjerneskjema?» i en konkret beslutning. Vi kan knytte begrepet til systemet dere bygger eller kjøper. Tall fra salg, drift og økonomi ført sammen, slik at to avdelinger slutter å komme med to ulike svar til det samme møtet.

Se tjenestene våre Ta kontakt →

Undersøk antakelsene bak ordet

Bruk forklaringen og det tilknyttede arbeidet til å formulere spørsmål for prosjektet. Vi kan undersøke antakelsene sammen før et verktøy velges. Gjorde en geodatapipeline 50 ganger raskere — 50x raskere pipelines for geografiske data

Gjorde en geodatapipeline 50 ganger raskere Ta kontakt →

Hvorfor det er viktig

Formen får analytiske spørsmål til å ligne hverandre. Velg faktaene som skal summeres, koble faktatabellen til dimensjonene som beskriver dem, filtrer på noen dimensjonskolonner og grupper på andre. Kimball Group beskriver faktatabeller som hovedmålet for beregninger og dynamiske aggregeringer, og dimensjonsattributter som hovedkilden til betingelser, grupperinger og etiketter i rapporter. En analytiker som har lært mønsteret én gang, kan skrive neste spørring mot en ny faktatabell uten hjelp, og et rapportverktøy kan bygge den fra en liste med felt, fordi alle stjerner er bygd likt.

Lesbarhet er den andre grunnen. En driftsdatabase er laget for å lagre hvert faktum én gang og holde det konsistent, så et produkts kategori ligger gjerne i en egen oppslagstabell som kobles tilbake via en id. Det er riktig design for systemet som tar imot ordrer, og feil design for den som leter etter salg per kategori og først må finne ut hvilke tabeller som skal kobles, og i hvilken

rekkefølge. En flat dimensjon samler alle beskrivende kolonner på ett sted, skrevet ut i ord. Kimball Group fraråder alternativet, å normalisere hierarkiet i en dimensjon ut i flere tabeller, fordi forretningsbrukere synes resultatet er vanskelig å forstå og navigere i.

Den tredje grunnen er at en godt bygd stjerne kan svare på spørsmål ingen stilte da den ble designet. Granulariteten i en faktatabell, altså den nøyaktige betydningen av én rad, fastsettes før noen dimensjon eller noe faktum, og Kimball Group anbefaler å starte på atomær granularitet: det laveste nivået forretningsprosessen registrerer. En tabell med én rad per kvitteringslinje kan summeres per dag, butikk, merke eller en hvilken som helst kombinasjon av dem. En tabell som bare har daglige totaler per butikk, kan aldri brytes ned på produkt, uansett hvordan spørsmålet formuleres. Med Kimball Groups ord tåler atomære data uforutsigbare spørringer, mens oppsummerte data forutsetter spørsmålene de skal få.

Prisen er bevisst redundans. Avdelingsnavnet gjentas på hver produktrad, og innlastingen som bygger stjernen, må holde gjentakelsene konsistente og avgjøre hva som skal skje når en beskrivelse endres. Stjernen erstatter heller ikke driftsmodellen; den bygges fra driftsmodellen gjennom en innlastingsprosess, og den er aldri ferskere enn siste kjøring av den innlastingen.

Slik ser det ut i praksis

En detaljhandelskjede vil se salg per produktkategori og måned. Forretningsprosessen er kassakvitteringen, og granulariteten er fastsatt til én rad per kvitteringslinje. Produktdimensjonen er én flat tabell, og faktatabellen har en nøkkel for hver dimensjon, kvitteringsnummeret som degenerert dimensjon og to mål.

```
CREATE TABLE dim_product (
    product_key  bigint PRIMARY KEY,   -- surrogatnøkkel, eid av varehuset
    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,   -- degenerert dimensjon
    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;
```

Spørringen har den formen alle stjernespørringer har: faktatabellen, én kobling per dimensjon, filtre og grupperinger på dimensjonstekst og en sum over et faktum. Å legge butikkens region til i rapporten er én kobling og én kolonne til i `GROUP BY`; faktatabellen endres ikke. Kvitteringsnummeret blir stående i faktaraden, så linjene i én handlekurv fortsatt kan telles sammen, men det får ingen dimensjonstabell, fordi det ikke er mer å si om det.

Et stjerneskjema er ett svar på løftet i tjenesten databasedesign og datamodellering: en struktur som samsvarer med hvordan virksomheten faktisk fungerer, slik at spørsmålene man trenger svar på senere, fortsatt kan stilles. Driftsmodellen stjernen bygges fra, er den andre halvdelen av det arbeidet. I sitt nåværende arbeid for NextMariner har grunnleggeren modellert et maritimt domene i 348 normaliserte tabeller fordelt på 34 PostgreSQL-skjemaer, med fagfolk, selskaper, skip, stillinger og omtaler. Der ble små, stabile oppregninger som statuser, typer og kategorier til SMALLINT-oppslagstabeller, som holder radene kompakte og join-operasjonene billige, og ekte fremmednøkler og skranker gjør integriteten til databasens jobb. Det er riktig form for et system som skriver én rad om gangen og må forbli korrekt. En stjerne bygd for rapportering oppå et slikt system går motsatt vei: Den gjør kodene om til ord igjen i flate dimensjoner, fordi leseren er et menneske og ikke applikasjonen.

Hvor det går galt

Dimensjonene normaliseres, og stjernen blir et snøfnugg. Å skille merke, kategori og avdeling ut av produktdimensjonen i egne tabeller ser ryddig ut for alle som er vant til å designe driftssystemer. Kimball Group kaller resultatet et snøfnugg (snowflake) og fraråder det: Forretningsbrukere synes det er vanskelig å forstå og navigere i, og de ekstra koblingene kan gjøre spørringer tregere,

mens en flat dimensjon inneholder nøyaktig den samme informasjonen. Hold hver dimensjon som én bred tabell, og la innlastingen gjenta hierarkiet på hver rad.

Granulariteten tilpasses den første rapporten. En faktatabell med daglig salg per butikk svarer på det dashbordet den ble bygd for, og ingenting under det; det første spørsmålet om et enkelt produkt eller en handlekurv krever en ny innlasting fra kilden. En tabell som blander daglige totaler med enkeltlinjer, er enda verre, fordi enhver sum over den teller det samme salget to ganger. Fastsett granulariteten først, start på det atomære nivået prosessen registrerer, og gi hver granularitet sin egen faktatabell. Oppsummeringstabeller kan så legges oppå for fart, som separate tabeller.

Dimensjonsattributtene er kildesystemets koder. Driftssystemer lagrer kompakte koder, og kopierer man dem rett inn i en dimensjon, får man rapporter med radetiketter som bare kildesystemet forstår. Kimball Group tillater koder som attributter, men sier at de mest nyttige attributtene er utfyllende beskrivelser, siden det er de som ender opp som etiketter i rapporten. Oversett hver kode til ord under innlastingen, og behold koden i en kolonne ved siden av beskrivelsen for den som må spore en rad tilbake til kilden.

Faktatabellen er formet som en rapport. Det er fristende å designe den etter regnearket virksomheten allerede bruker, med en kolonne per måned eller en rad per linje i ledelsesrapporten. Kimball Group er tydelig på at utformingen av en faktatabell bygger på den fysiske aktiviteten som måles, ikke på rapportene som senere hentes ut av den. Rapportoppsett endres; kassakvitteringen gjør det ikke. Modeller hendelsen, én rad per forekomst på den fastsatte granulariteten, og la spørringene lage oppsettene.

Les mer

- Star Schemas and OLAP Cubes — definisjonen, fakta og dimensjoner som måling og kontekst, og hvordan kuber henger sammen med stjerneskjemaer.
- Grain — hva én rad i faktatabellen betyr, og hvorfor atomær granularitet kommer først.
- Fact Table Structure — numeriske mål og fremmednøkler i én rad per hendelse, utformet etter aktiviteten og ikke etter rapporten.
- Dimension Table Structure — brede, flate, denormaliserte tabeller der de beskrivende attributtene styrer filtre, grupperinger og etiketter.

- Snowflaked Dimensions — hvorfor det gjør en dimensjon vanskeligere å bruke å normalisere hierarkiet i den.
- Microsoft: Datatilgang for flerdimensjonale modeller — MDX som spørre- og beregningsspråk, og XMLA som grensesnittet en klient kan sende en MDX-setning gjennom.

Ordliste

Velg ut fra deres faktiske rammer

Ta med et representativt datasett, spørsmålene det skal besvare og dagens flaskehals. Vi kan avtale målinger av kvalitet, gjennomstrømning og gjenoppretting.

Se hva det koster Ta kontakt →

Gjør begrepene praktiske

Spør hvordan «Hva er et stjerneskjema?» gjelder i organisasjonen deres. Et konkret eksempel fra dagens system er nok til å starte samtalen.

Se tjenestene våre

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

Ta med deg denne siden

Visittkort

Laget med kjærlighet av D.

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

Opphavsrett © 2025 – nå · Engineer ApS · Alle rettigheter forbeholdt.

Skrevet ut fra https://engineer.company/nb/glossary/star-schema/