Engineer

Schéma en étoile

★ une table de faits entourée de tables de dimensions

Ce que c’est

Un schéma en étoile est une manière d’organiser des tables pour l’analyse. Une table centrale enregistre sous forme de nombres des événements métier, comme des ventes, des expéditions ou des paiements. Autour se trouvent les tables qui décrivent ces événements : quel produit, quel magasin, quel jour, quel client. Dessinée avec un trait entre la table centrale et chacune des autres, cette organisation ressemble à une étoile.

La table centrale est la table de faits, et celles qui l’entourent sont les tables de dimensions. La description du Kimball Group est concise : « Les schémas en étoile se composent généralement de tables de faits reliées aux tables de dimensions associées. » Le lien est une clé ordinaire. Chaque table de dimensions possède une seule clé primaire, et chaque ligne de faits porte cette clé comme clé étrangère, une pour chaque dimension qui décrit l’événement.

Cette séparation suit la distinction du Kimball Group entre mesure et contexte. Un modèle dimensionnel s’organise autour des événements de mesure d’un processus métier : un article scanné en caisse, une commande expédiée, un sinistre indemnisé. Les nombres produits par l’événement, comme la quantité et le montant, sont les faits. Tout ce qui dit qui, quoi, où, quand, pourquoi et comment constitue le contexte et se trouve dans les dimensions.

Les deux sortes de tables ont des formes opposées. Au niveau le plus détaillé, une table de faits contient une ligne par événement, avec les mesures numériques, une clé étrangère vers chaque dimension et parfois une référence telle qu’un numéro de facture ou de ticket. Ce numéro est tout ce qui reste du document une fois sa date, son client et son magasin devenus des dimensions. Il reste donc sur la ligne de faits sans table de dimensions associée : le Kimball Group parle de dimension dégénérée. Une table de dimensions, à l’inverse, est

large, plate et délibérément dénormalisée, avec de nombreuses colonnes de texte descriptif. Une dimension produit répète la marque, la catégorie et le rayon sur chaque ligne de produit au lieu de les séparer dans leurs propres tables.

Un schéma en étoile est la forme relationnelle d'un modèle dimensionnel, constituée de tables ordinaires dans une base SQL ordinaire. L'autre forme décrite par le Kimball Group est le cube OLAP : une base multidimensionnelle contenant les mêmes attributs et faits, interrogée dans des langages dotés d'une puissance analytique supérieure à SQL, comme MDX, XMLA fournissant un protocole d'accès. Les cubes sont souvent construits à partir d'un schéma en étoile, comme dernière étape avant la présentation des données aux utilisateurs ou comme couche d'agrégats. Les règles de modélisation sont les mêmes pour les deux. Le Kimball Group les dit indépendantes de l'architecture : elles s'appliquent aussi bien à un entrepôt de données de type Kimball qu'à la Corporate Information Factory de Bill Inmon, organisée autour d'un noyau et de branches.

Utilisez le concept pour prendre une décision

Utilisez « Qu'est-ce qu'un schéma en étoile ? » pour éclairer une décision réelle. Nous pouvons vous aider à relier ce concept au système que vous construisez ou achetez. Réunir les chiffres des ventes, des opérations et des finances pour que deux services ne présentent plus deux réponses à la même réunion.

Découvrir nos prestations Contactez-nous →

Vérifiez les hypothèses derrière le terme

Appuyez-vous sur l'explication et les réalisations liées pour formuler les questions de votre projet. Nous pouvons examiner les hypothèses avec vous avant le choix d'un outil. Pipeline géographique accéléré de 50 fois — 50x d'accélération des pipelines de données géographiques

Pipeline géographique accéléré de 50 fois Contactez-nous →

Pourquoi c'est important

Cette forme rend les questions analytiques semblables. Choisir les faits à additionner, joindre la table de faits aux dimensions qui les décrivent, filtrer sur certaines colonnes de dimensions et regrouper sur d'autres. Le Kimball Group décrit les tables de faits comme la cible principale des calculs et agrégations dynamiques, et les attributs de dimensions comme la source principale des contraintes de requête, des regroupements et des libellés de rapports. Un analyste qui a appris cette démarche une fois peut écrire seul la requête suivante sur une nouvelle table de faits. Un outil de restitution peut aussi la construire à partir d'une liste de champs, puisque chaque étoile suit la même organisation.

La lisibilité est la deuxième raison. Une base opérationnelle est conçue pour stocker chaque fait une seule fois et en maintenir la cohérence : la catégorie d'un produit se trouve donc généralement dans une table de référence séparée, reliée par un identifiant. C'est le bon choix pour le système qui prend les commandes, mais le mauvais pour une personne qui cherche les ventes par catégorie et doit d'abord déterminer quelles tables joindre et dans quel ordre. Une dimension plate rassemble toutes les colonnes descriptives au même endroit, en toutes lettres. Le Kimball Group met en garde contre l'autre choix, normaliser la hiérarchie d'une dimension dans des tables supplémentaires, car les utilisateurs métier ont du mal à comprendre et à parcourir le résultat.

La troisième raison tient à ce qu'une étoile bien construite peut répondre à des questions qui n'avaient pas été posées lors de sa conception. Le grain d'une table de faits, c'est-à-dire la signification exacte d'une ligne, est déclaré avant toute dimension ou tout fait. Le Kimball Group recommande de commencer au grain atomique : le niveau le plus fin que le processus métier enregistre. Une table contenant une ligne par ligne de ticket peut être agrégée par jour, magasin, marque ou toute combinaison de ces dimensions. Une table qui ne conserve que des totaux quotidiens par magasin ne pourra jamais être détaillée par produit, quelle que soit la formulation de la question. Selon les termes du Kimball Group, les données atomiques supportent les requêtes imprévues, tandis que les données agrégées présupposent les questions qui leur seront posées.

Le prix est une redondance voulue. Le nom du rayon est répété sur chaque ligne de produit, et le traitement de chargement qui construit l'étoile doit maintenir la cohérence de ces répétitions et décider ce qui se passe lorsqu'une description change. L'étoile ne remplace pas non plus le modèle opérationnel : elle en est issue par un traitement de chargement et n'est à jour qu'à la date de la dernière exécution de ce chargement.

À quoi cela ressemble en pratique

Un détaillant veut connaître les ventes par catégorie de produits et par mois. Le processus métier est le ticket de caisse, et le grain est déclaré comme une ligne par ligne de ticket. La dimension produit est une seule table plate. La table de faits contient une clé pour chaque dimension, le numéro de ticket comme dimension dégénérée et deux mesures.

```
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;
```

La requête a la forme commune à toutes les requêtes en étoile : la table de faits, une jointure par dimension, des filtres et regroupements sur le texte des dimensions et une somme sur un fait. Ajouter la région du magasin au rapport demande une jointure de plus et une colonne de plus dans le `GROUP BY` ; la table de faits ne change pas. Le numéro de ticket reste sur la ligne de faits, ce qui permet encore de compter ensemble les lignes d'un même panier. Il ne reçoit toutefois aucune table de dimensions, puisqu'il n'y a rien de plus à en dire.

Une étoile est une réponse à la promesse de la prestation de conception et de modélisation de bases de données : une structure qui correspond au fonctionnement réel de l'activité pour que les questions nécessaires plus tard puissent encore être posées. Le modèle opérationnel dont elle est issue constitue l'autre moitié de ce travail. Dans son travail actuel pour NextMariner, la fondatrice a modélisé un domaine maritime en 348 tables normalisées, réparties sur 34 schémas PostgreSQL et couvrant les professionnels, entreprises, navires, offres d'emploi et avis. Les petites listes stables, comme les statuts, types et catégories, y sont devenues des tables de référence SMALLINT, qui gardent les lignes compactes et les jointures peu coûteuses. De véritables clés étrangères et contraintes confient l'intégrité à la base de données. C'est la bonne forme pour

un système qui écrit les lignes une à une et doit rester correct. Une étoile de restitution construite au-dessus d'un tel système va dans l'autre sens : elle transforme à nouveau les codes en mots dans des dimensions plates, parce qu'elle est lue par une personne plutôt que par l'application.

Où les problèmes surviennent

Les dimensions sont normalisées, et l'étoile devient un flocon. Séparer la marque, la catégorie et le rayon de la dimension produit dans leurs propres tables paraît bien ordonné à qui pratique la conception opérationnelle. Le Kimball Group appelle le résultat un flocon et le déconseille : les utilisateurs métier ont du mal à le comprendre et à y naviguer, et les jointures supplémentaires peuvent ralentir les requêtes, alors qu'une dimension plate contient exactement les mêmes informations. Garder chaque dimension dans une seule table large et laisser le chargement répéter la hiérarchie sur chaque ligne.

Le grain est fixé pour le premier rapport. Une table de faits contenant les ventes quotidiennes par magasin répond aux besoins du tableau de bord pour lequel elle a été créée, mais à rien de plus fin. La première question sur un produit ou un panier particulier exige un nouveau chargement depuis la source. Une table qui mélange les totaux quotidiens et les lignes individuelles est pire : toute somme y compte deux fois la même vente. Déclarer d'abord le grain, commencer au niveau atomique enregistré par le processus et donner à chaque grain sa propre table de faits. Des tables de synthèse séparées peuvent ensuite être ajoutées au-dessus pour accélérer les calculs.

Les attributs des dimensions sont les codes du système source. Les systèmes opérationnels stockent des codes compacts. Les copier directement dans une dimension donne des rapports dont seul le système source comprend les libellés des lignes. Le Kimball Group autorise les codes comme attributs, mais indique que les attributs les plus utiles sont les descriptions détaillées, puisque ce sont elles qui deviennent les libellés d'un rapport. Décoder chaque code en mots pendant le chargement et garder le code dans une colonne voisine de la description pour toute personne qui doit retrouver l'origine d'une ligne.

La table de faits prend la forme d'un rapport. Il est tentant de la concevoir à partir du tableur déjà utilisé par l'entreprise, avec une colonne par mois ou une ligne par rubrique du rapport de gestion. Le Kimball Group est explicite : la conception d'une table de faits repose sur l'activité physique mesurée, pas sur les

rapports qui en seront tirés. La mise en page des rapports change ; le ticket de caisse, non. Modéliser l'événement, avec une ligne par occurrence au grain déclaré, et laisser les requêtes produire les présentations.

Pour aller plus loin

- Star Schemas and OLAP Cubes — la définition, les faits et dimensions comme mesure et contexte, et le rapport entre cubes et étoiles.
- Grain — déclarer ce que signifie une ligne de faits et pourquoi le grain atomique vient d'abord.
- Fact Table Structure — mesures numériques et clés étrangères sur une ligne par événement, conçue à partir de l'activité plutôt que du rapport.
- Dimension Table Structure — tables larges, plates et dénormalisées dont les attributs descriptifs servent aux filtres, regroupements et libellés.
- Snowflaked Dimensions — pourquoi normaliser la hiérarchie d'une dimension rend son utilisation plus difficile.
- Microsoft : Multidimensional model data access — MDX comme langage de requête et de calcul, et XMLA comme interface par laquelle un client peut envoyer une instruction MDX.

Glossaire

Choisissez selon vos contraintes réelles

Présentez un jeu de données représentatif, les questions auxquelles il doit répondre et le goulot d'étranglement actuel. Nous pouvons convenir de la manière de mesurer la qualité, le débit et la reprise.

Consulter les tarifs Contactez-nous →

Rendez la terminologie concrète

Demandez comment « Qu'est-ce qu'un schéma en étoile ? » s'applique à votre organisation. Un exemple concret tiré de votre système actuel suffit pour engager la discussion.

Découvrir nos prestations

welcome@engineer.company +45 50 10 61 84 Copenhague, Danemark

Emportez cette page avec vous

Fiche de contact

Créé avec soin et amour par D.

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

Droit d'auteur © 2025 – aujourd'hui · Engineer ApS · Tous droits réservés.

Imprimé depuis https://engineer.company/fr/glossary/star-schema/