



Data Science Academy

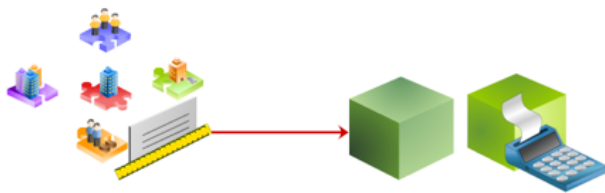
www.datascienceacademy.com.br

Design e Implementação de Data Warehouses

Tipos de Fatos

A tabela fato é a principal tabela do Data Warehouse e ela fica no centro do Star Schema, sendo rodeada por dimensões. A tabela fato armazena o que ocorreu, é o fato propriamente dito. É na tabela fato onde as métricas são armazenadas, junto com as Surrogate Keys que ligam às dimensões que descrevem essa métrica.

Existem 6 tipos de fatos:



- Fato transacional
- Fato consolidada
- Fato agregada
- Fato de snapshot periódico
- Fato de snapshot acumulado
- Fato sem fato

Fatos transacionais são as mais comuns. A maioria dos bilhões de linhas que temos no Data Warehouse são de tabelas de fato transacional. Elas geralmente utilizam métricas aditivas, aquelas métricas que podemos somar por todas as dimensões.

A tabela fato consolidada é bem parecida com a agregada, mas serve para combinar 2 tipos de processos. O que é processo? Área de negócio, área de assunto, processo de negócio. Diferente da tabela fato agregada, que serve para fazer agregações em geral, a fato consolidada é para consolidar duas fatos. No processamento ETL, na hora de carregar a fato, você vai carregar uma, carregar a outra, e misturar as duas. Evidente que a granularidade precisa ser a mesmo. Um exemplo clássico: você tem uma tabela fato venda e depois precisa unir com uma tabela fato venda orçada. É claro que você pode fazer uma dimensão de cenário e já colocar lá se ela é orçada. Mas isso é em um mundo ideal onde você começa seu Data Warehouse do zero sem nenhum problema. E muitas vezes, mesmo que não tenha sido um desleixo, na hora de fazer o levantamento de requisitos, ninguém nunca falou que seria necessário trabalhar com orçamento de venda, então ao invés de criar uma dimensão de cenário e reprocessar tudo, foi criada uma nova tabela fato. São várias situações em que isso pode acontecer. As tabelas fatos consolidadas adicionam uma complexidade extra no processamento do ETL. Você vai ter que reconfigurar todo o processo ou simplesmente fazer um novo.

A tabela fato agregada tem a função de acelerar o desempenho das consultas. Ela serve para agregar dados quando eu não quero analisar no nível do grão.

A snapshot periódica é a famosa foto. Tiramos uma foto do momento atual e salvamos. Ela tem as colunas de entrada e saída, que são opcionais, mas é bom colocar, porque a equipe de negócio sempre pede esse tipo de métrica adicional. Isso é bem comum em redes de varejo onde precisamos usar esse tipo de tabela, onde vai ter somente o consolidado daquele dia.

E qual a diferença de um snapshot acumulado para periódico? O periódico pega o momento no período, tira uma fotografia e insere na tabela fato. O acumulado também é uma fotografia, mas em mais de um momento.

E temos ainda a “Fato Sem Fato” (Factless Fact Table). Uma tradução livre seria fato sem métricas. Ela também é chamada de fato de associação ou de interseção, mas o termo técnico é fato sem fato ou factless fact table. E para que serve? Para fazer uma interseção de dimensões. Às vezes queremos comparar ou cruzar algo somente entre duas dimensões e não tem uma métrica para fazer essas comparações. Essa tabela fato é a exceção, só é usada quando se precisa fazer uma interseção entre as dimensões. Dois exemplos de fato sem fato: frequência de aluno ou venda com promoção.

Nem todos esses tipos de fatos serão usados no dia a dia, mas é sempre bom saber que existem, para quando houver uma necessidade específica.

Referências:

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling

https://www.amazon.com.br/Data-Warehouse-Toolkit-Definitive-Dimensional-ebook/dp/B00DRZX6XS/ref=sr_1_2?ie=UTF8&qid=1521159195&sr=8-2&keywords=kimball

Dimensional Modeling Techniques

<https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/>