

Modelagem Dimensional

Prof. Leandro Correia



Modelagem Dimensional

- Simula um ambiente dimensional em um ambiente relacional;
- Objetivo de promover uma melhor performance nas consultas;
- Facilidade de consultas ad-hoc pelo próprio usuário;
- As atualizações são realizadas apenas através de cargas;
- Modelagem Esquema-Estrela (*Star Schema*) e Floco de Neve (*Snow Flake*).

Dimensões e Fatos

A modelagem multidimensional se baseia nos conceitos de DIMENSÃO e FATO:

- **Dimensão:** Informações descritivas sobre os dados analisados (descrição do produto, nome do cliente, data, etc). Determina o ângulo pelo qual se quer analisar os dados.
- **Fatos:** Medidas sobre os dados analisados. Para cada cruzamento das dimensões são gerados fatos que quantificam os dados (média venda, valor do faturamento, total do estoque, etc).

Dimensões e Fatos

Granuralidade

- A granularidade define o nível de detalhamento dos dados armazenados;
- Mudar a granularidade de um modelo dimensional é extremamente custoso, logo, a definição da granularidade é um dos aspectos mais importantes de um projeto de Data Warehouse;
- Aspectos influenciados pela definição da granularidade:
 - Consultas que podem ser atendidas;
 - Volume de dados no Data Warehouse;
 - Periodicidade de carga.

Modelando Dimensões

Normalização

- Foco em redução de redundância e na eficiência de operações de atualização (Insert, Update, Delete);
- Os modelos tradicionais acabam sendo muito complexos e de difícil entendimento;
- Necessidade de múltiplos Joins, trazem problemas de performance e excesso de consumo de recursos.

Modelando Dimensões

Normalização

- Os modelos dimensionais desnormalizam de forma significativa o modelo de dados relacional;
- A desnormalização provoca redundância, o que privilegia a performance de consultas em detrimento do consumo de espaço;
- Ficam dificultadas as alterações de dados, por isso a necessidade de criação de um processo de carga do dados no DW.

Modelando Dimensões

Chaves Artificiais (*Surrogate Keys*)

- Usadas para controlar o histórico dos dados nas dimensões
- Permitem a alteração de chaves de negócio sem comprometer a integridade do modelo multidimensional
- Não têm (e não podem ter) significado para o negócio

ID_PRODUTO	COD_PRODUTO	DESC_PRODUTO	...
1	315	Leite Condensado	...
2	315	Leite Condensado	...
3	316	Detergente	...
4	317	Sabão em pó	...
5	317	Sabão em pó	...

Modelando Dimensões

Controle de Versão

- Como a manutenção do histórico da informação é um dos pontos mais significativos da modelagem dimensional, é necessário que exista algum mecanismo de controle de versão dos dados nas dimensões.
- Existem pelo menos 3 formas de controle de versão em um Data Warehouse:
 - Sobrescrever os dados
 - Inserir novos registros
 - Adicionar novos campos

Modelando Dimensões

Controle de Versão

- Sobrescrever os dados:
 - Consiste em atualizar a informação diretamente no registro original;
 - Não mantém histórico já que somente a última posição permanece armazenada na dimensão.

ID_PRODUTO	COD_PRODUTO	DESCRICAO PRODUTO	...
1	315	Leite Condensado Diet	...

ID_PRODUTO	COD_PRODUTO	DESCRICAO PRODUTO	...
1	315	Leite Condensado Light	...

Modelando Dimensões

Controle de Versão

- Inserir novos registros:
 - Consiste em inserir um novo registro para cada nova versão dos dados;
 - Mantém histórico da informação já que tanto o novo registro quanto os registros anteriores permanecem armazenados na dimensão;
 - É necessário que exista algum mecanismo que indique a vigência do registro, de forma que a versão corrente possa ser identificada.

Modelando Dimensões

Controle de Versão

- Inserir novos registros (controle por data de vigência):

ID_PRODUTO	COD_PRODUTO	DESC_PRODUTO	DATA_INICIO	DATA_FIM
1	315	Leite Condensado Diet	01/01/2011	13/03/2011
2	315	Leite Condensado Light	13/03/2011	<u>null</u>

Modelando Dimensões

Controle de Versão

- Inserir novos registros (controle por flag de situação):

ID_PRODUTO	COD_PRODUTO	DESC_PRODUTO	ATIVO
1	315	Leite Condensado Diet	Não
2	315	Leite Condensado Light	Sim

Modelando Dimensões

Controle de Versão

- Criação de novos campos:
 - Consiste em criar novos campos, durante o processo de carga (ETL), para acomodar as novas versões da informação;
 - Custo de manutenção elevado.

Modelando Dimensões

Hierarquias

- Agrupamentos de conceitos correlatos em uma única dimensão;
- Aplica-se às dimensões para que os fatos sejam agrupados em diferentes níveis de detalhamento;
- Tipos de Hierarquia:
 - Simples: um nó tem somente um pai
 - Múltipla: um nó pode ter vários pais

Modelando Dimensões

Hierarquias

- Exemplo: País → Região → UF → Cidade
- As tabelas País, Região, UF e Cidade são armazenadas separadamente no banco de dados fonte (normalizado);
- No Data Warehouse, elas compõem uma única tabela, a dimensão Geografia.

Modelando Dimensões

Hierarquias

- A navegação através dos níveis ocorre através das operações de *drilling*:
 - Drill Down: detalha a informação;
 - Drill Up: sumariza a informação;
 - Drill Across: muda de uma fato para outro com dimensões em conformidade;
 - Drill Through: busca um detalhe menor do que aquele colocado na fato.

Modelando Dimensões

- Dimensão Tempo
 - Dimensão especial que recebe um tratamento diferente das demais;
 - Permite a implementação de hierarquia de tempo:
 - Dia, Mês, Trimestre, Semestre, Ano
 - Pode ser carregada manualmente e somente uma vez (não faz parte das cargas periódicas);
 - Presente na maior parte dos modelos multidimensionais.

Modelando Fatos

- Tipos de Fatos

- Fato Aditivo:

- Pode ser livremente usado para a operação de adição em **qualquer dimensão**;
 - O resultado é sempre válido, ou seja, possui significado semântico;
 - Total de vendas é um exemplo de fato de aditivo.

- Fato semi-aditivo:

- Não pode ser adicionado ao longo de **pelo menos uma** dimensão.

- Fato não-aditivo:

- Não pode ser adicionado ao longo de **nenhuma** dimensão pois perderá o seu valor semântico;
 - Margem de lucro é um exemplo de fato não-aditivo.

Modelando Fatos

- Agregações
 - São agrupamentos dos fatos, ou seja, resultados obtidos através da redução da granularidade a partir tabela de fatos;
 - Pode-se criar agregações a partir de qualquer atributo de uma dimensão. Normalmente são definidas de acordo com as hierarquias das dimensões;
 - Os resultados das consultas são pré-calculados, o que melhora muito o desempenho;
 - Deve-se avaliar o tempo de carga e o espaço ocupado em disco.

Modelo Multidimensional

O modelo multidimensional pode ser implementado em bancos de dados relacionais de 2 formas:

- *Star Schema* (Esquema Estrela)
 - Dimensões desnormalizadas
 - Alta performance
 - Forte consumo de espaço em disco
- *Snow Flake* (Floco de Neve)
 - Dimensões normalizadas
 - Performance menor
 - Menor consumo de espaço em disco

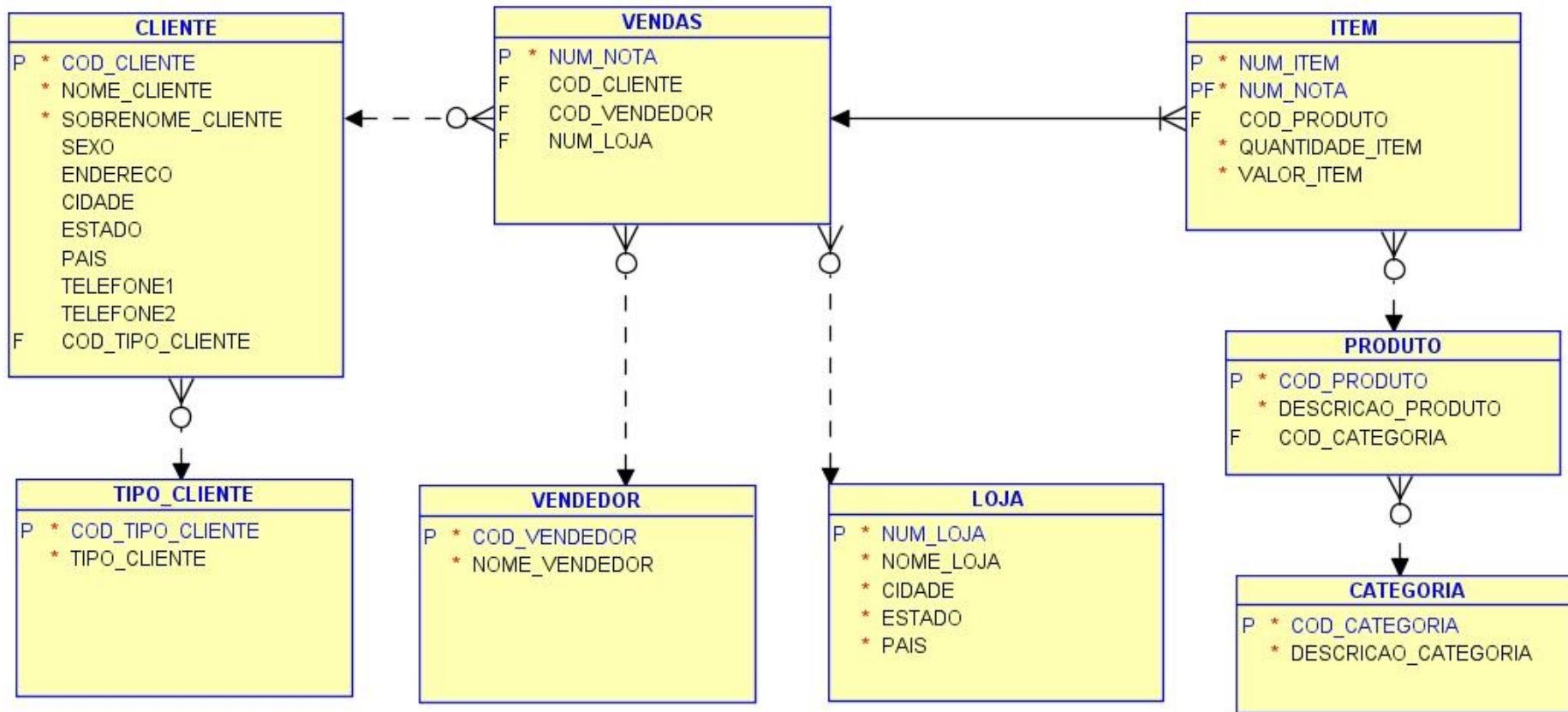
Modelo Multidimensional

Star Schema (Esquema Estrela)

- Consiste em exportar as chaves primárias das dimensões para a tabela de fato.
- A chave primária da tabela de fato será definida pela composição das chaves exportadas das dimensões.

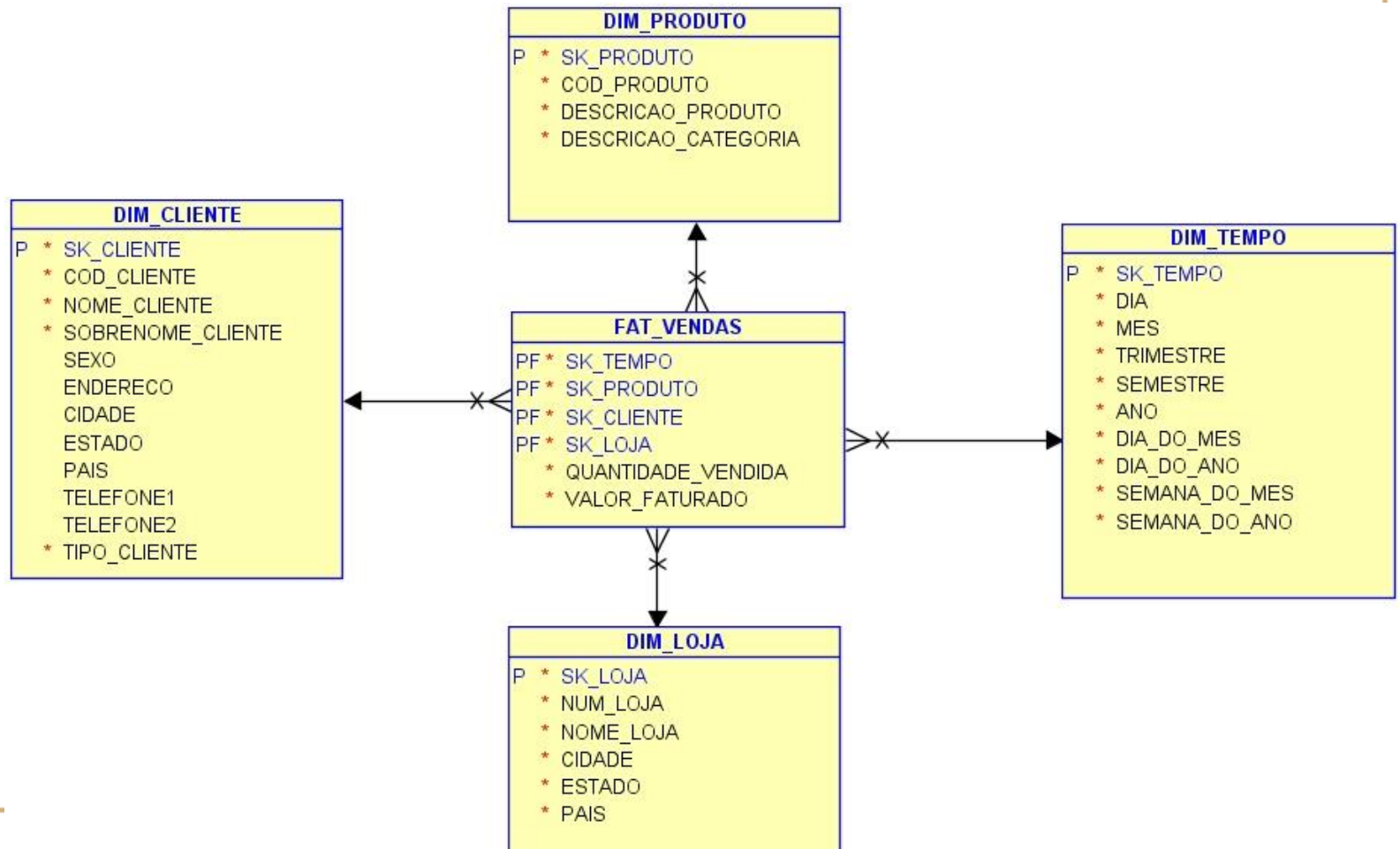
Modelo Multidimensional

- Modelo relacional normalizado:



Modelo Multidimensional

- Star Schema



Modelo Multidimensional

- Snow Flake (Floco de Neve)
 - Consiste em normalizar o Star Schema visando economizar espaço em disco;
 - O grau de normalização deve ser determinado com cautela para não prejudicar significativamente a performance das consultas.

Modelo Multidimensional

- Snow Flake

