

Data Science  
Academy

Data Science  
Academy

# Formação Desenvolvedor Microsoft para Data Science e Inteligência Artificial

Power BI Avançado para Análise de Dados



# Capítulo 4

## Modelagem de Dados





# Conteúdo

- Tabelas e Relacionamentos
- Relacionamentos Ativos e Inativos
- Direção de Filtro Cruzado
- Tabela Calendário
- Sistemas OLTP e OLAP
- Star Schema e Snowflake Schema
- Configurações de Metadados



# Nomenclaturas

- Um dataset é uma estrutura de dados armazenada no Power BI.
- Podemos alterar a estrutura dos dados na fase de Modelagem.
- Um relatório/visualização necessita de um dataset.
- Um dataset pode ser utilizado para fornecer dados a vários relatórios.
- Um dataset pode ter dados de várias e diferentes data sources (fontes de dados).
- Um data source pode oferecer diferentes tipos de conexões.



# OLTP

- Visa gerenciar dados transacionais como matrículas, cancelamentos, pagamentos, recebimentos, etc e fazem referências a outros dados como alunos, instrutores, cursos, formações, etc.
- Características: muitos usuário “comuns”, suporte a transações, altamente normalizado, atualizável e acrescentável, **cargas de trabalho com gravações intensas e leituras moderadas**, consultas a centenas/milhares de registros, modelo relacional com formato tabular.



# Desafios da Abordagem OLTP

- Podem não ser adequados para manipular agregações em grandes volumes de dados.
- Pode exigir um maior processamento para realizar consultas em dados altamente normalizados.
- Armazenamento dos registros por tempo indeterminado e/ou o armazenamento excessivo de dados podem levar a uma redução do desempenho de consulta.



# OLAP

- Organiza bancos de dados para dar suporte a consultas sem prejudicar os sistemas transacionais.
- Projetados para **cargas de trabalho de leitura intensa e pouca gravação.**
- Características: poucos usuário com conhecimento do negócio, processamento de consultas complexas, consultas a milhões de registros, tamanho físico maior que o OLTP.



# Design e Implementação de DWs

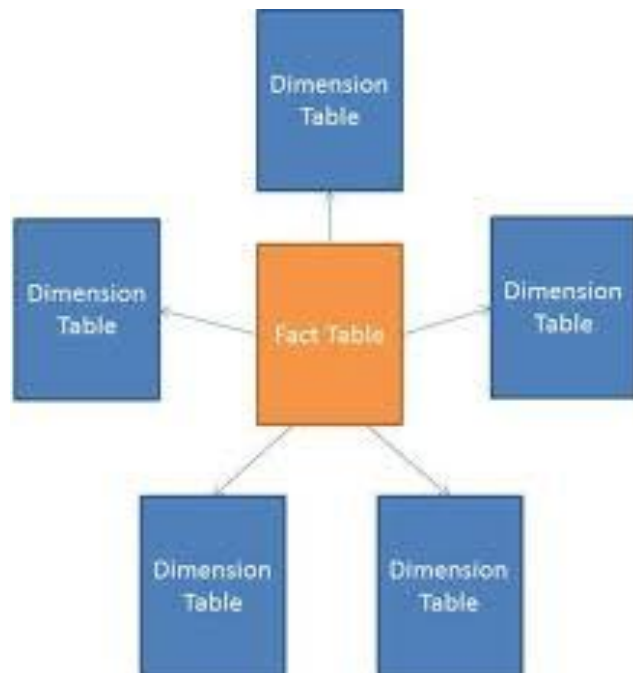
- Fases de Design de um DW
- Modelagem de Negócio
- Modelo Lógico e Dimensional
- Normalizações, Star Schema, Snowflake Schema
- Tabelas Fato e Dimensão
- Hierarquias de Dimensões
- etc...

<https://www.datascienceacademy.com.br/course?courseid=design-e-implementacao-de-data-warehouses>





# Esquema Estrela





# Esquema Estrela (Star Schema)

- Organiza dados em tabelas de fatos e dimensões.
- Tabelas-fato contém dados quantitativos normalmente gerados em um sistema transacional e carregados no Datawarehouse.
  - Exemplos: Tabela-fato para análise de vendas; Tabela-fato para análise de desempenho de vendedores, etc.
- Tabelas-dimensões contém as informações na qual os dados da tabela-fato poderão ser filtrados, agregados, segmentados.
  - Exemplos: Clientes, Fornecedores, Estados e Municípios, Produtos, etc.



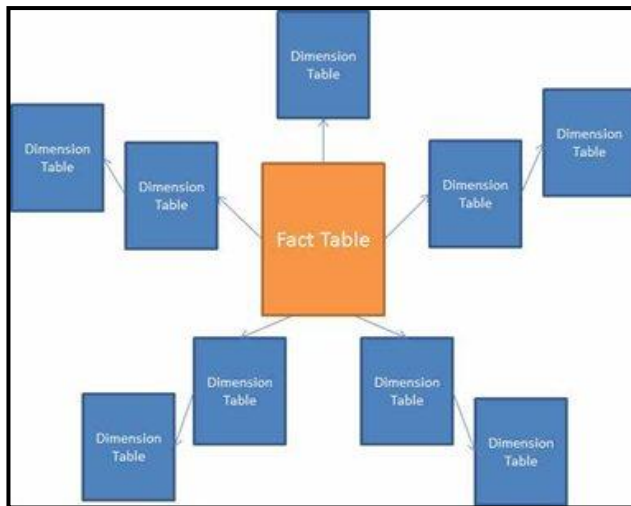
# Esquema Estrela (Star Schema)

- Possui uma (ou mais) tabelas-fato referenciando várias tabelas-dimensões.
- A quantidade de linhas na tabela-fato é bem superior a quantidade de linhas das tabelas-dimensões.
- Devido a sua estrutura, possibilita ganho de performance para execução de consultas:
  - Número menor de joins do que em modelos altamente normalizados;
  - Possibilidade de uso de dados desnormalizados.



# Esquema Floco de Neve (Snowflake Schema)

- Tabelas-Dimensões podem ser normalizadas e relacionadas com outras tabelas-dimensões
  - Tradeoff entre tamanho do banco de dados e complexidade das consultas.





# Modelo de Dados

<b>Visuals</b>	Gráficos de Linha, Colunas, Pizza, Heatmaps, Mapas
<b>Calculations</b>	DAX: Medidas, Tabelas e Colunas Calculadas
<b>Data Model</b>	Tabelas, Colunas, Tipos de Dados, Propriedades, Categorias, Hierarquias, Relacionamentos
<b>Query</b>	Data Sources, Tabelas, Colunas, Transformações

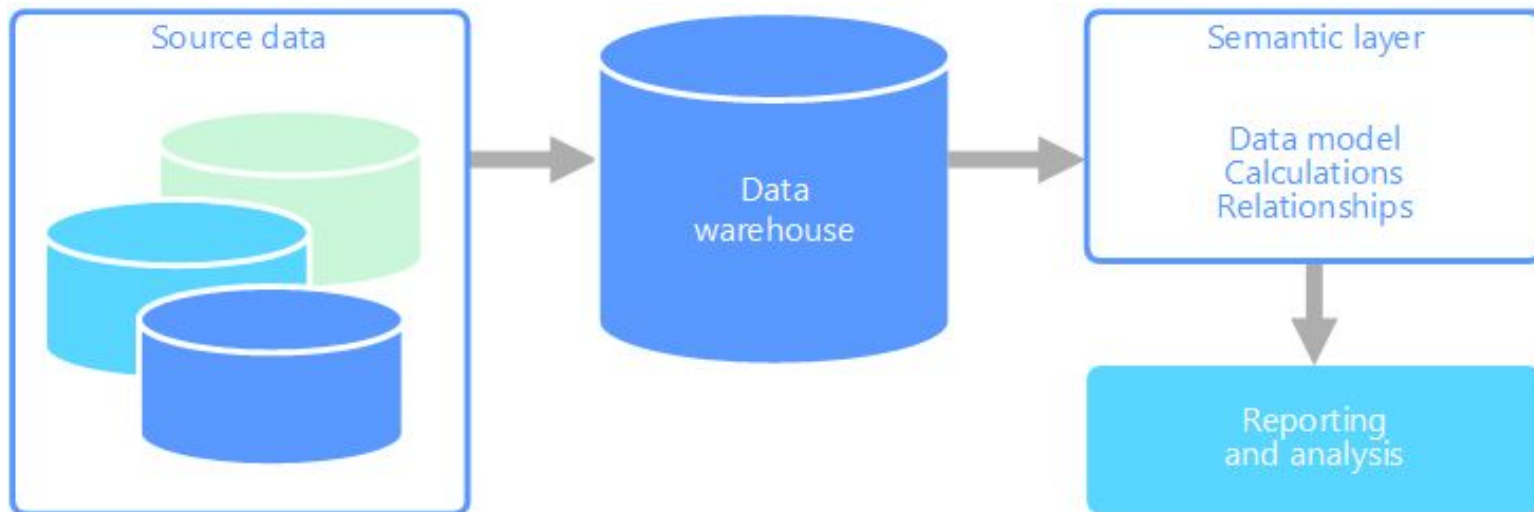


# Modelo Semântico

- Modelo conceitual que fornece um nível de abstração do esquema do banco de dados (ou de outras fontes de dados).
- Objetivo é que o usuário (analista de dados, analista de negócio) não precise conhecer estruturas, nomenclaturas e termos técnicos utilizados internamente pela equipe de TI.
- A utilização de nomes mais óbvios e amigáveis e alinhado com a linguagem de negócio facilita para que os próprios usuários realizem suas consultas no dataset.



# Modelo Semântico



# Obrigado.

# Tenha uma excelente jornada de aprendizagem!



Data Science  
Academy