



Data Science Academy



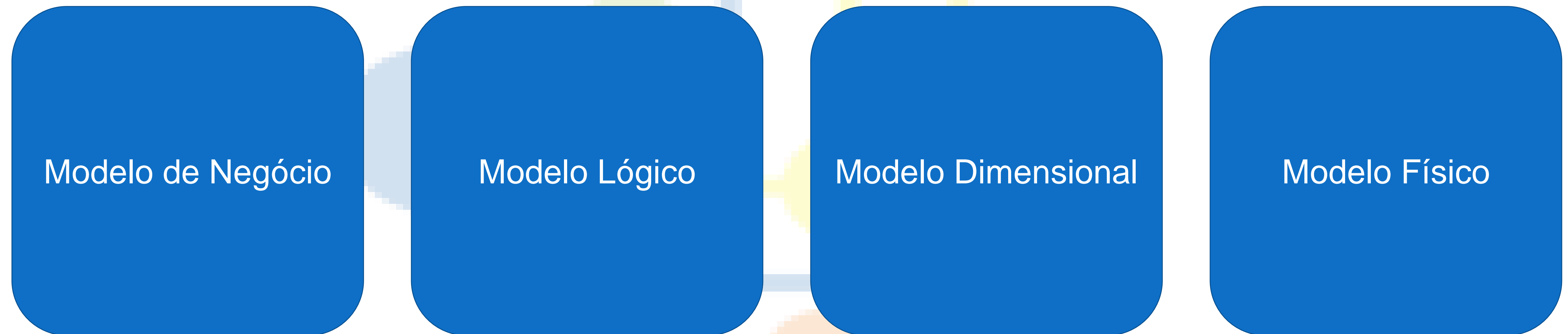
Design e Implementação de Data Warehouses

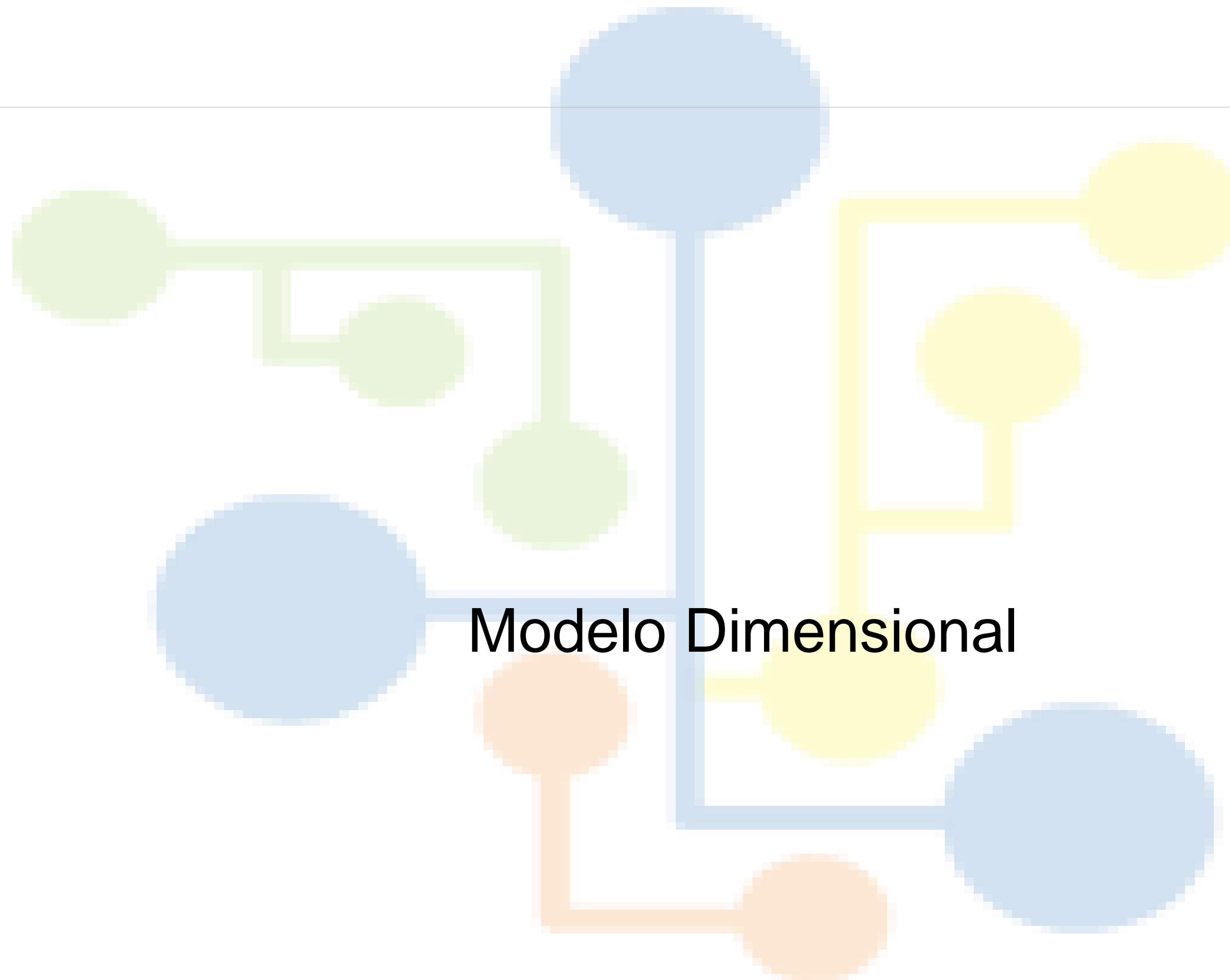


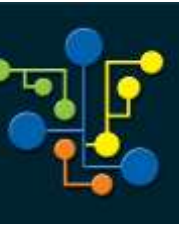
Data Science Academy

Modelo Dimensional





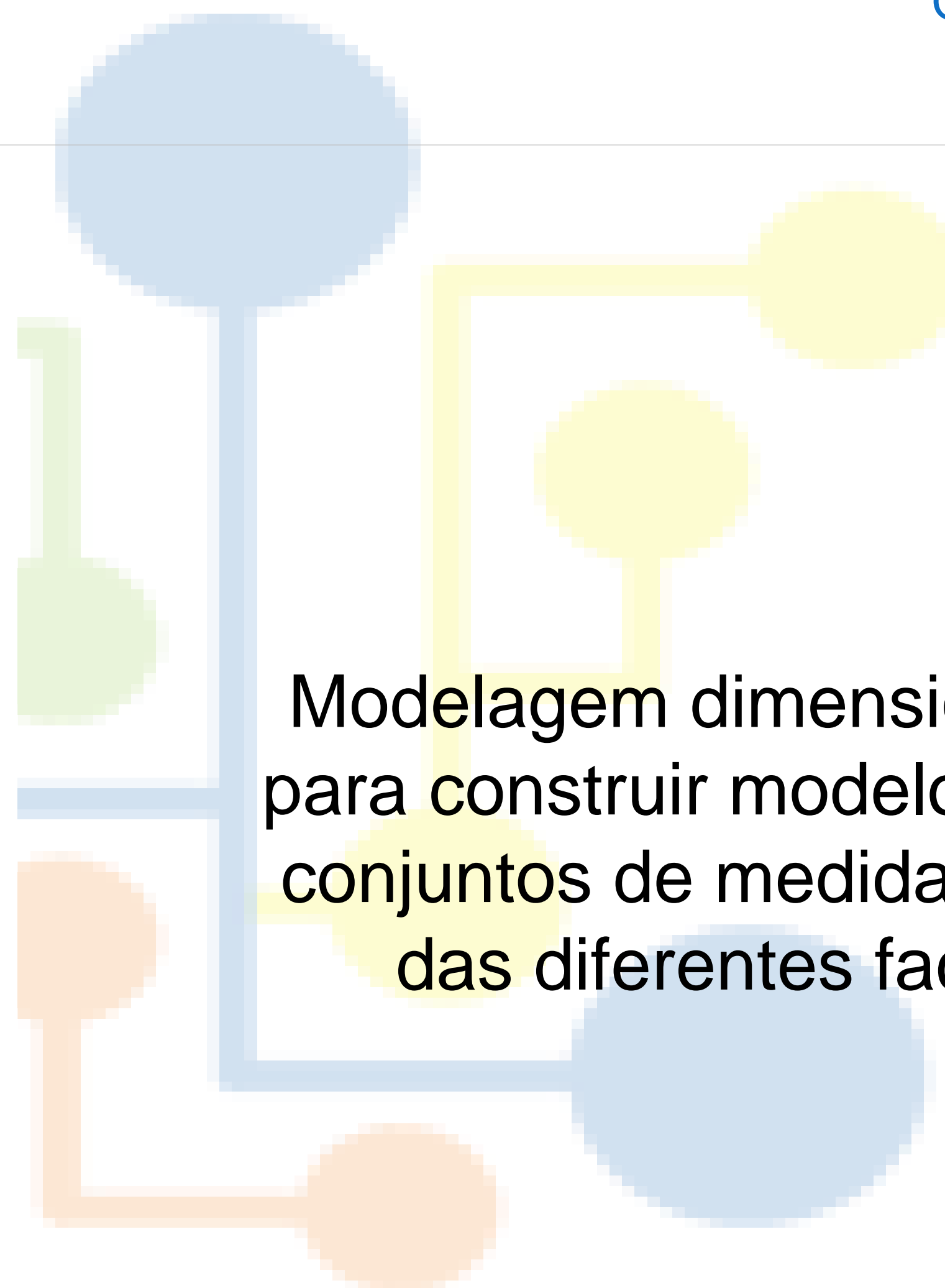
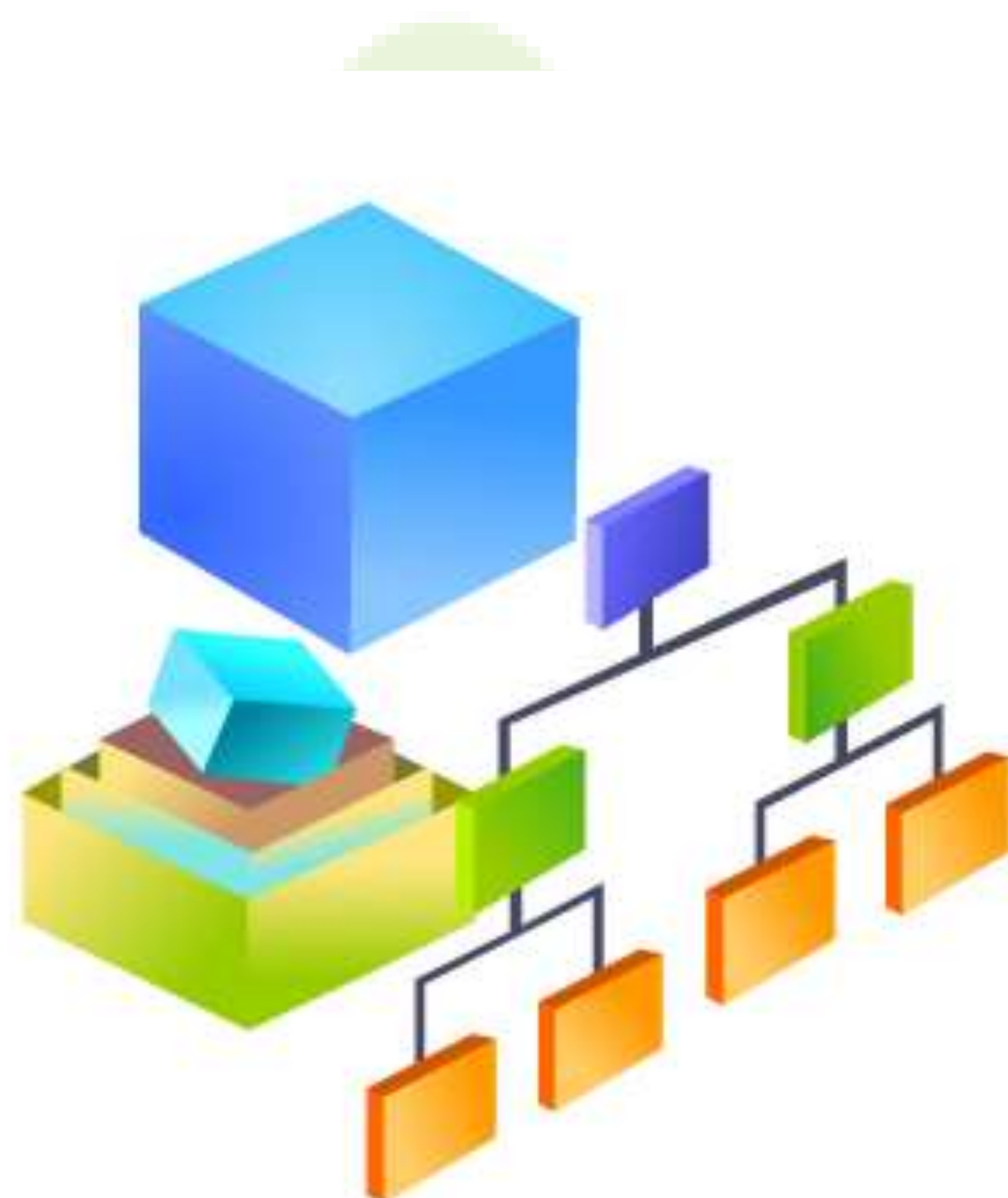




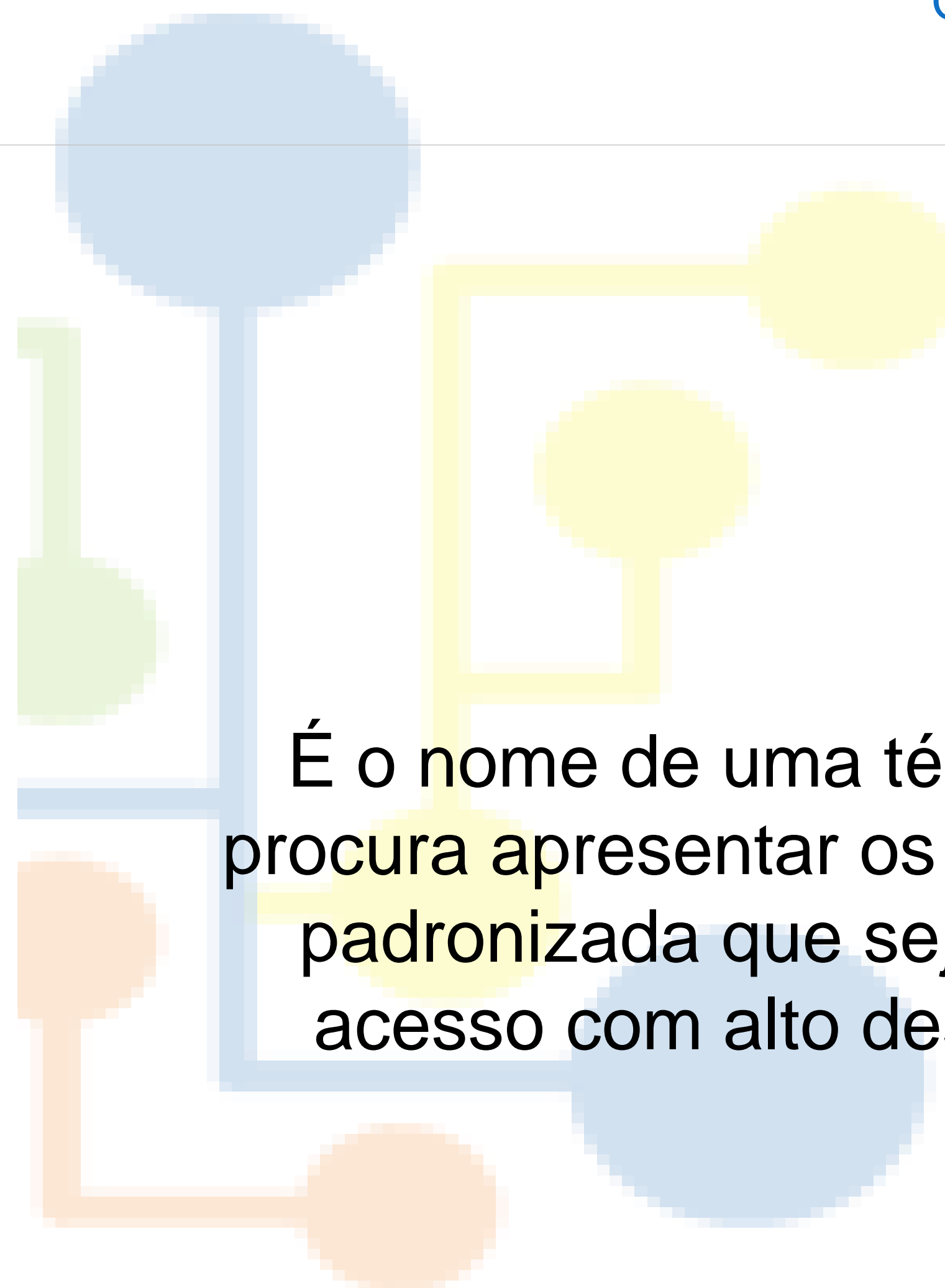
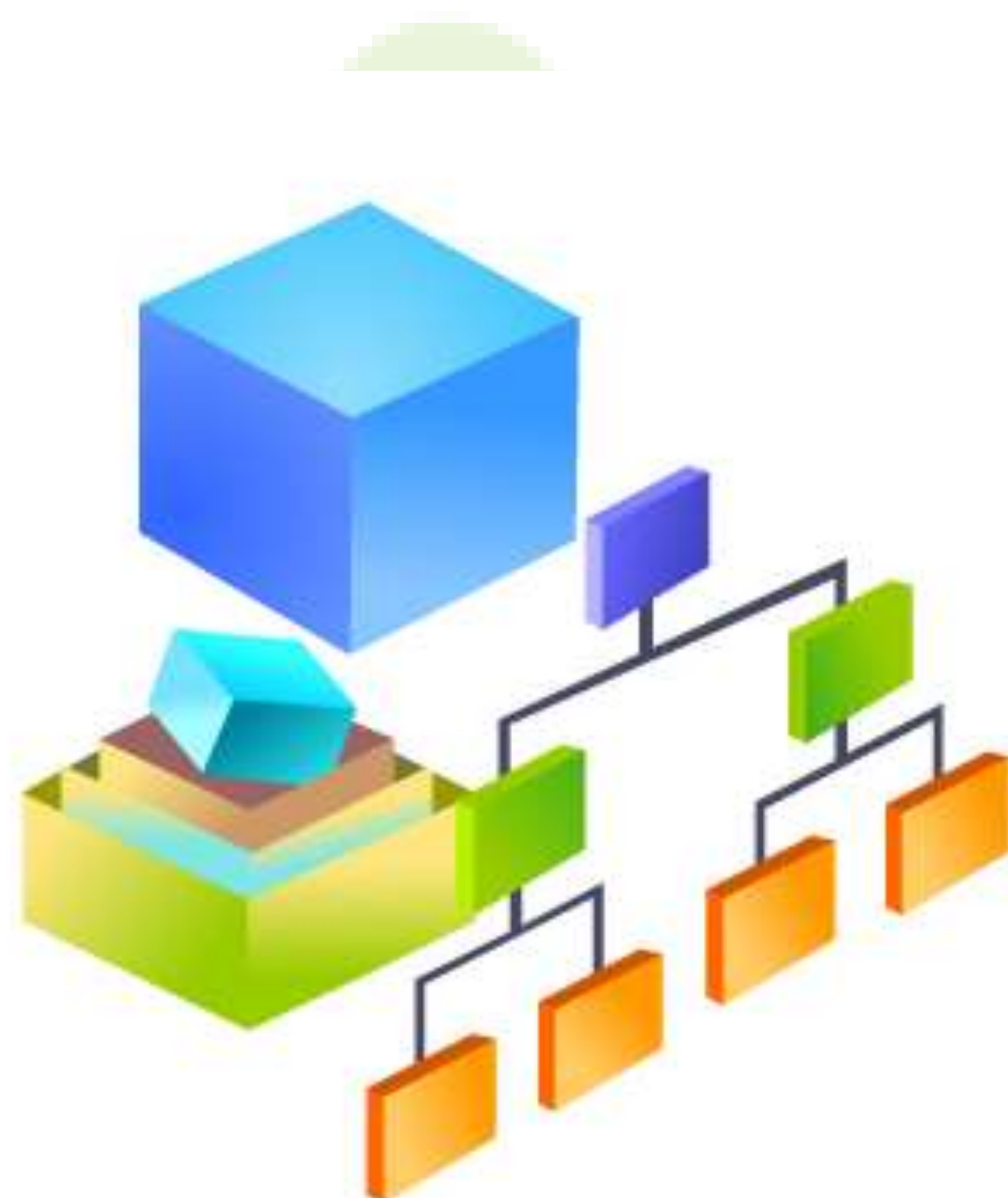
Data Science Academy

O Que é Modelagem Dimensional?

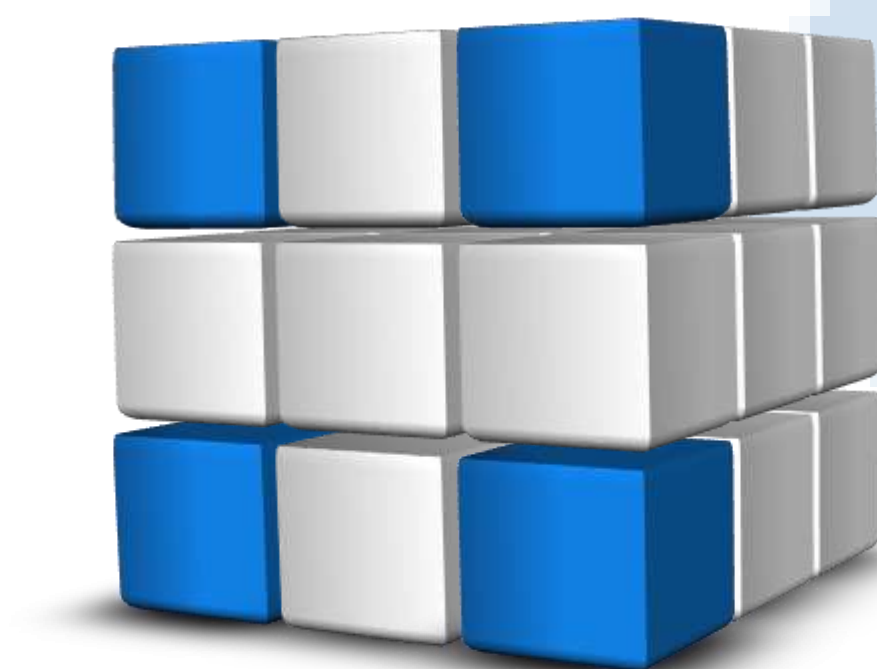




Modelagem dimensional é uma técnica para construir modelos de negócio como conjuntos de medidas descritas através das diferentes faces do negócio.

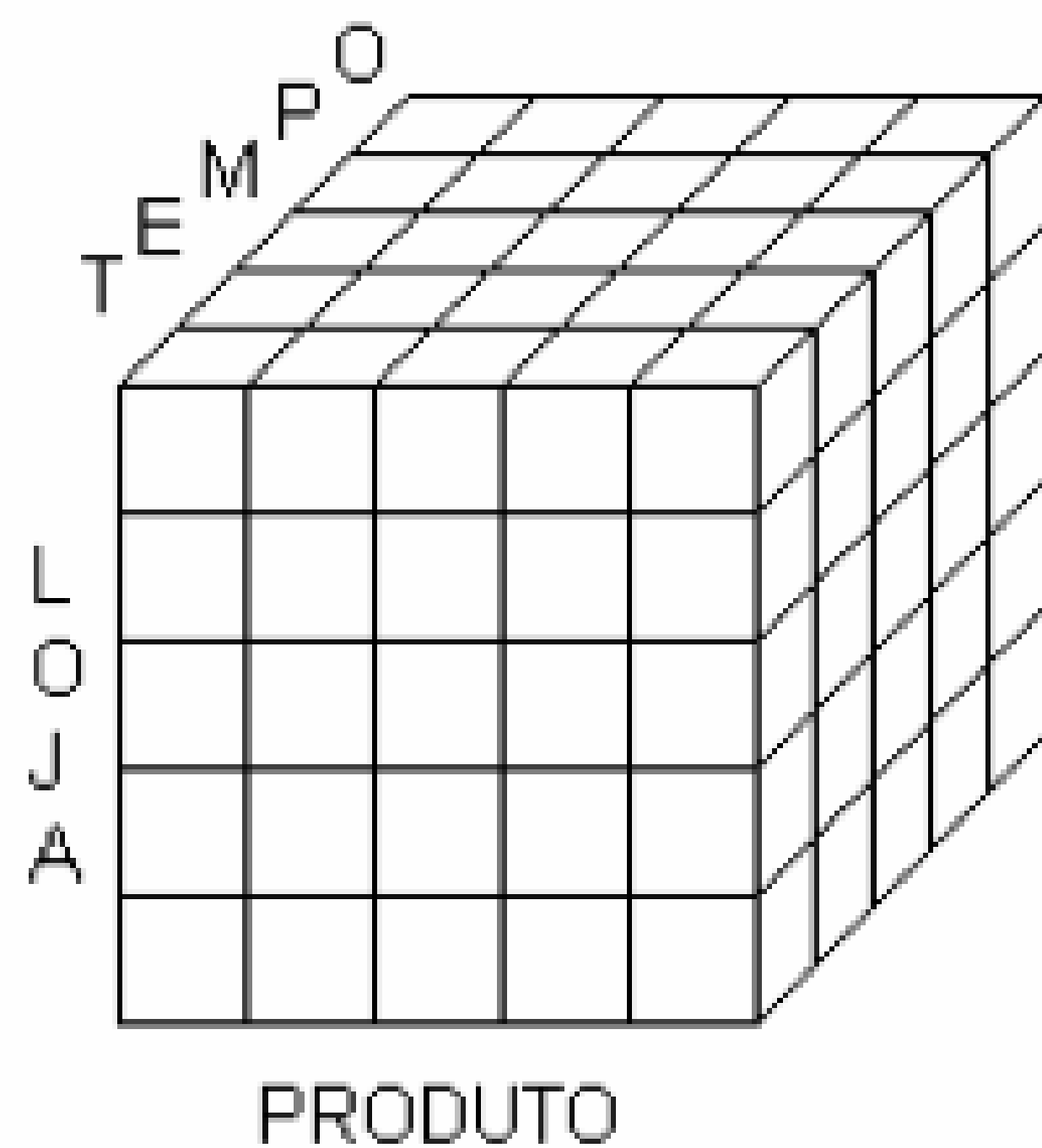


É o nome de uma técnica de projeto que procura apresentar os dados numa estrutura padronizada que seja intuitiva e permita acesso com alto desempenho [KIM98].



A origem do termo *dimensional* está relacionada com a ideia de que os dados devem ser agrupados de maneira a formar um cubo, ou hipercubo, que seria a estrutura padrão para visualizar os dados.



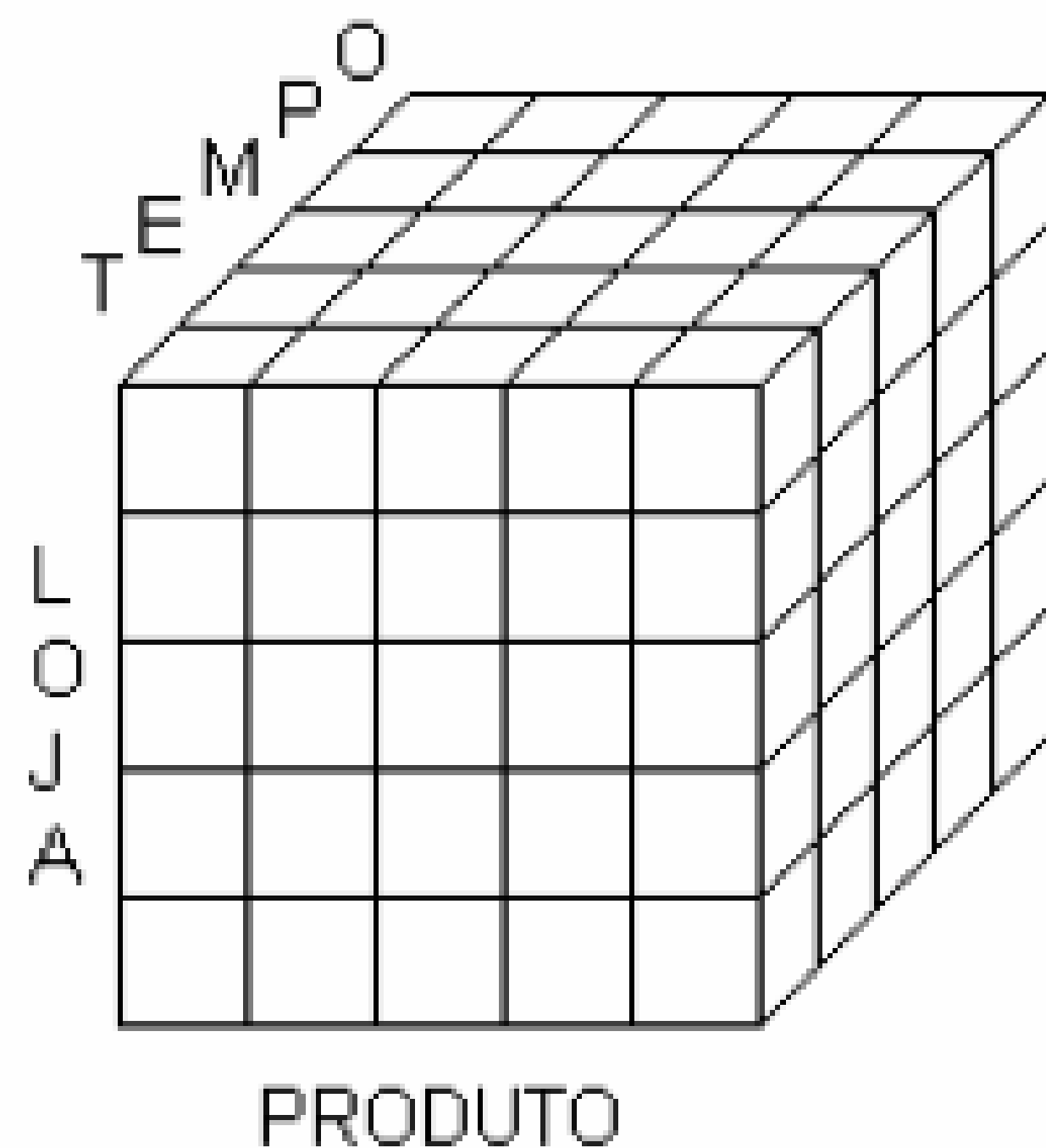


“Vendemos produtos em várias lojas e avaliamos nosso desempenho ao longo do tempo”

Foram vendidos 100 kg de Feijão na filial da Zona Oeste durante o mês de Fevereiro de 2018.

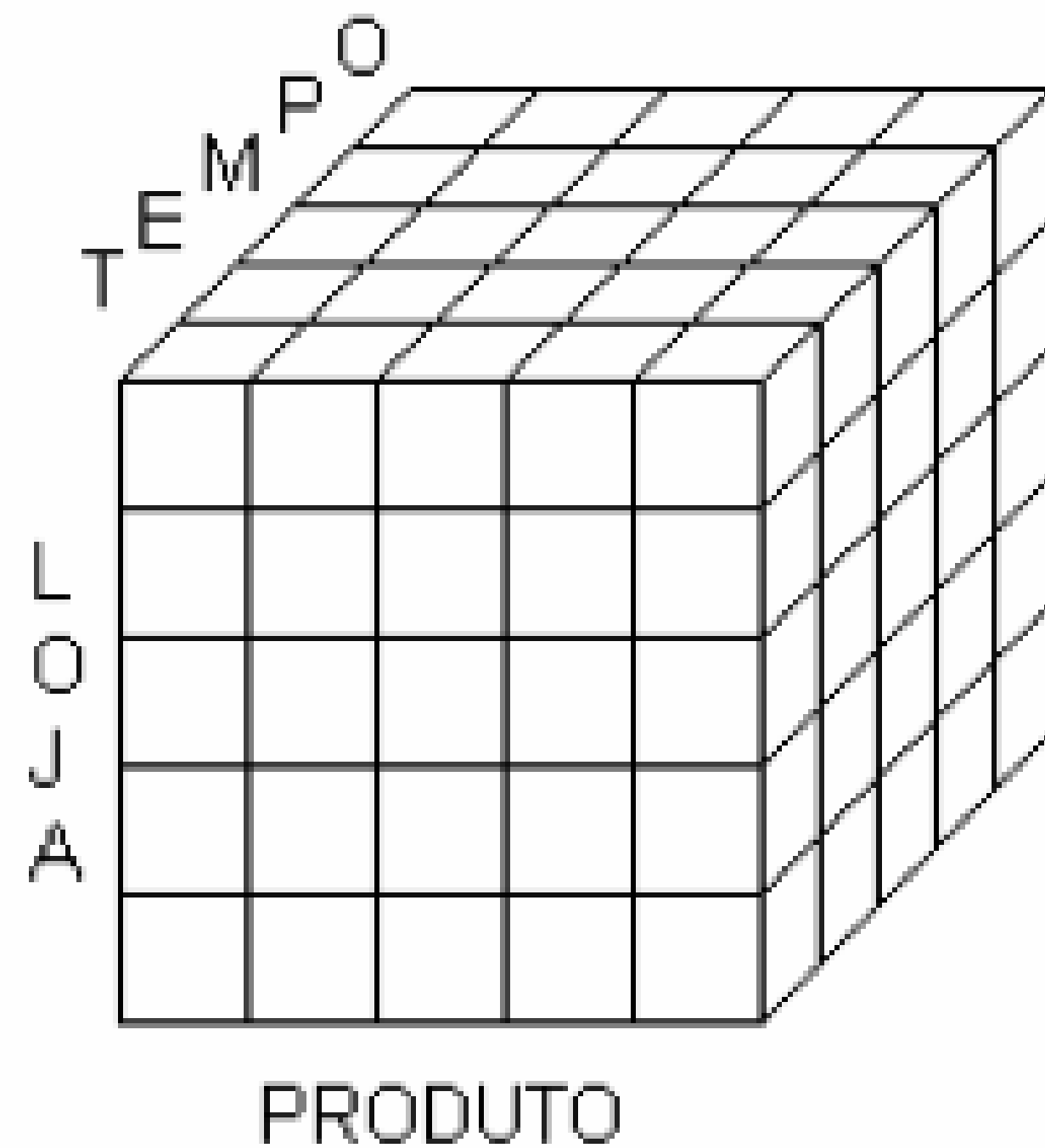


No modelo dimensional, cada ponto interno ao cubo contém as medições para uma combinação de Produto, Mercado e Tempo.

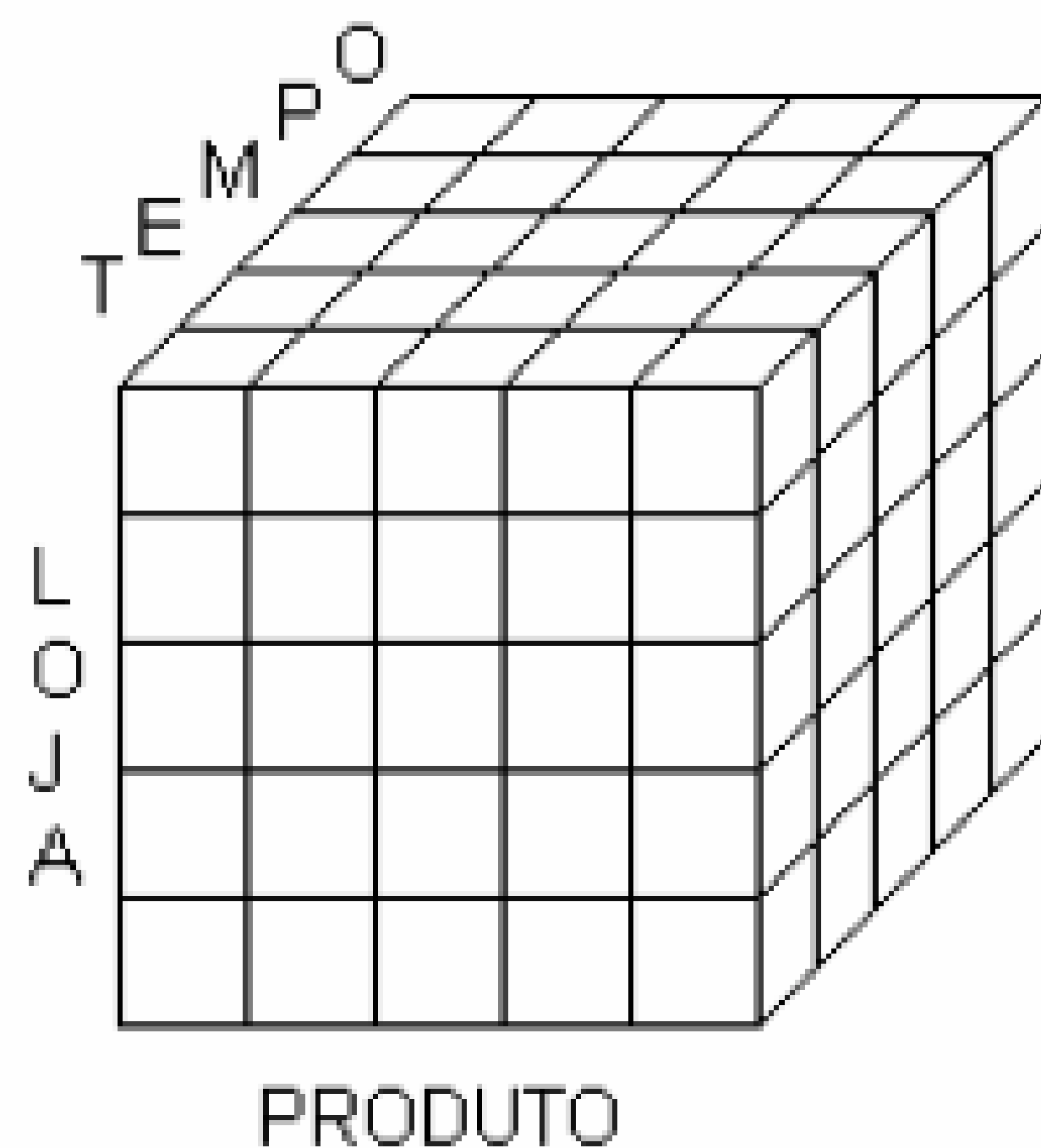


Slice and Dice: ou fatiamento do cubo, é a restrição das coordenadas nas dimensões de acordo com critérios definidos em cima de atributos das dimensões. Por exemplo: visualizar somente os dados dos mercados de pequeno porte.





Agregação: permite reduzir a dimensionalidade de um cubo ou de uma fatia de cubo. Por exemplo: ao visualizar a receita total de cada mercado por mês estamos agregando os dados de receita na dimensão produto, eliminando-a do cubo.

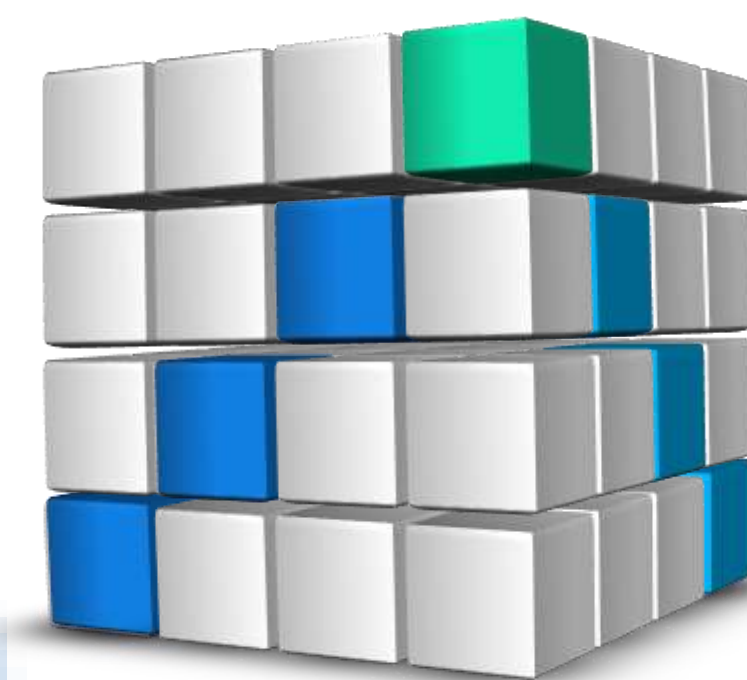


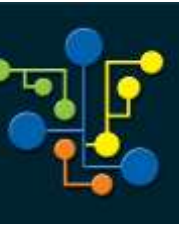
Drill-up/Drill-down: é a navegação entre níveis de agregação, de acordo com hierarquias existentes nas dimensões. Exemplo: o usuário pode começar visualizando dados totais de receita para os mercados e fazer um drill-up para visualizar os dados totais por bairro, agregando os mercados próximos.





O cubo é uma abstração poderosa para visualizar os dados, mas ele é restrito a três dimensões. Um modelo dimensional não pode, portanto, ser definido como sendo um cubo.





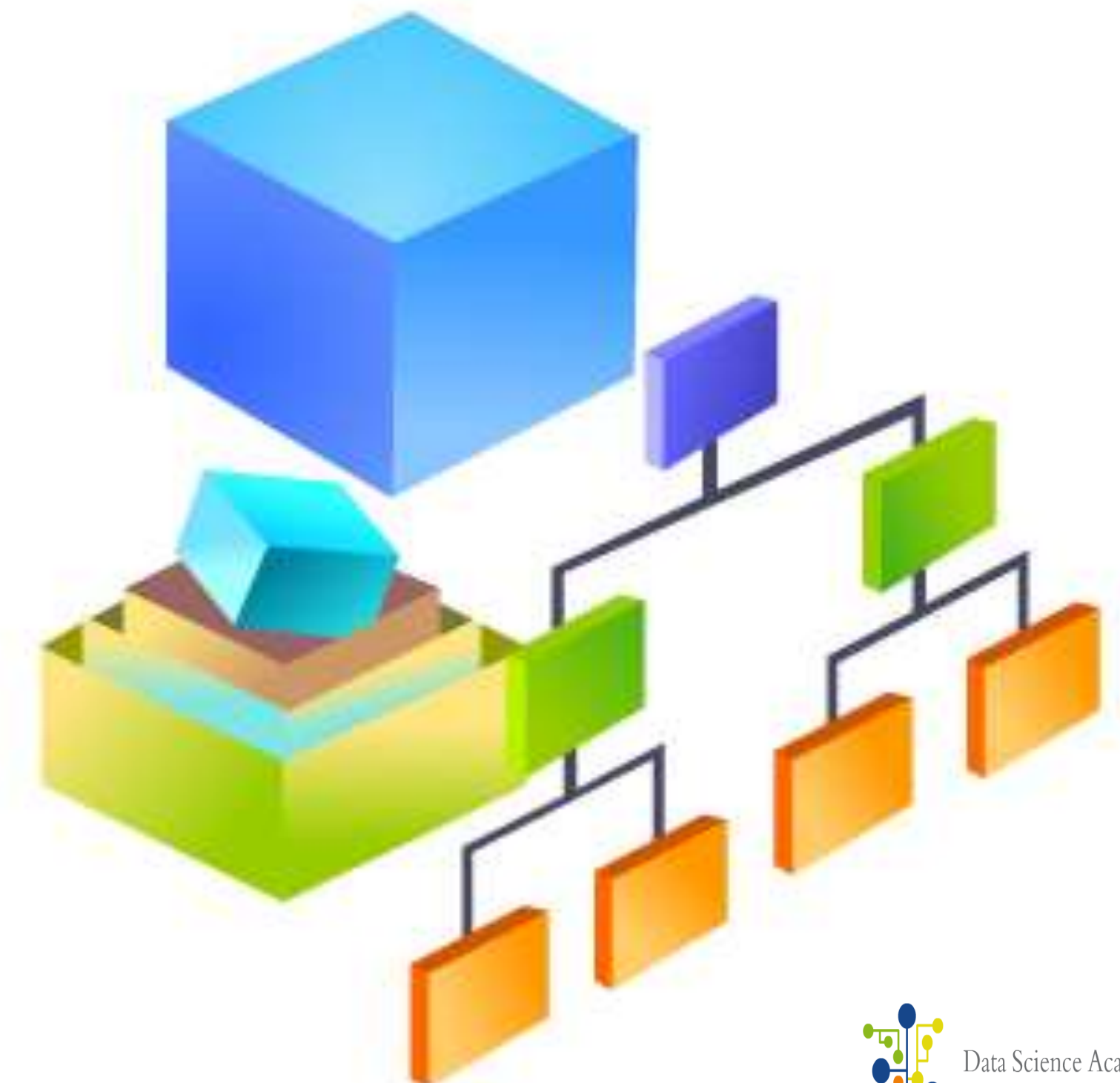
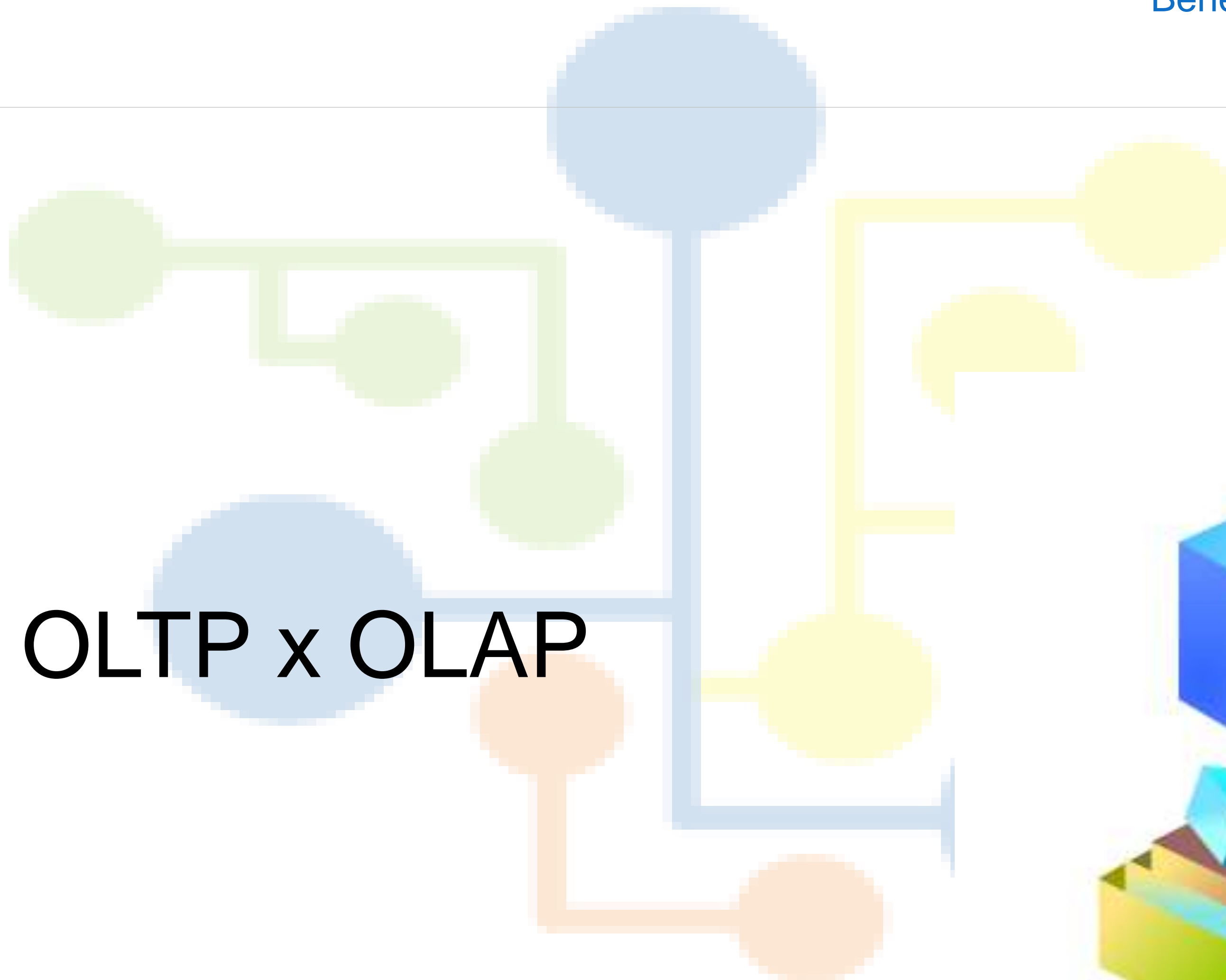
Data Science Academy

Benefícios da Modelagem Dimensional



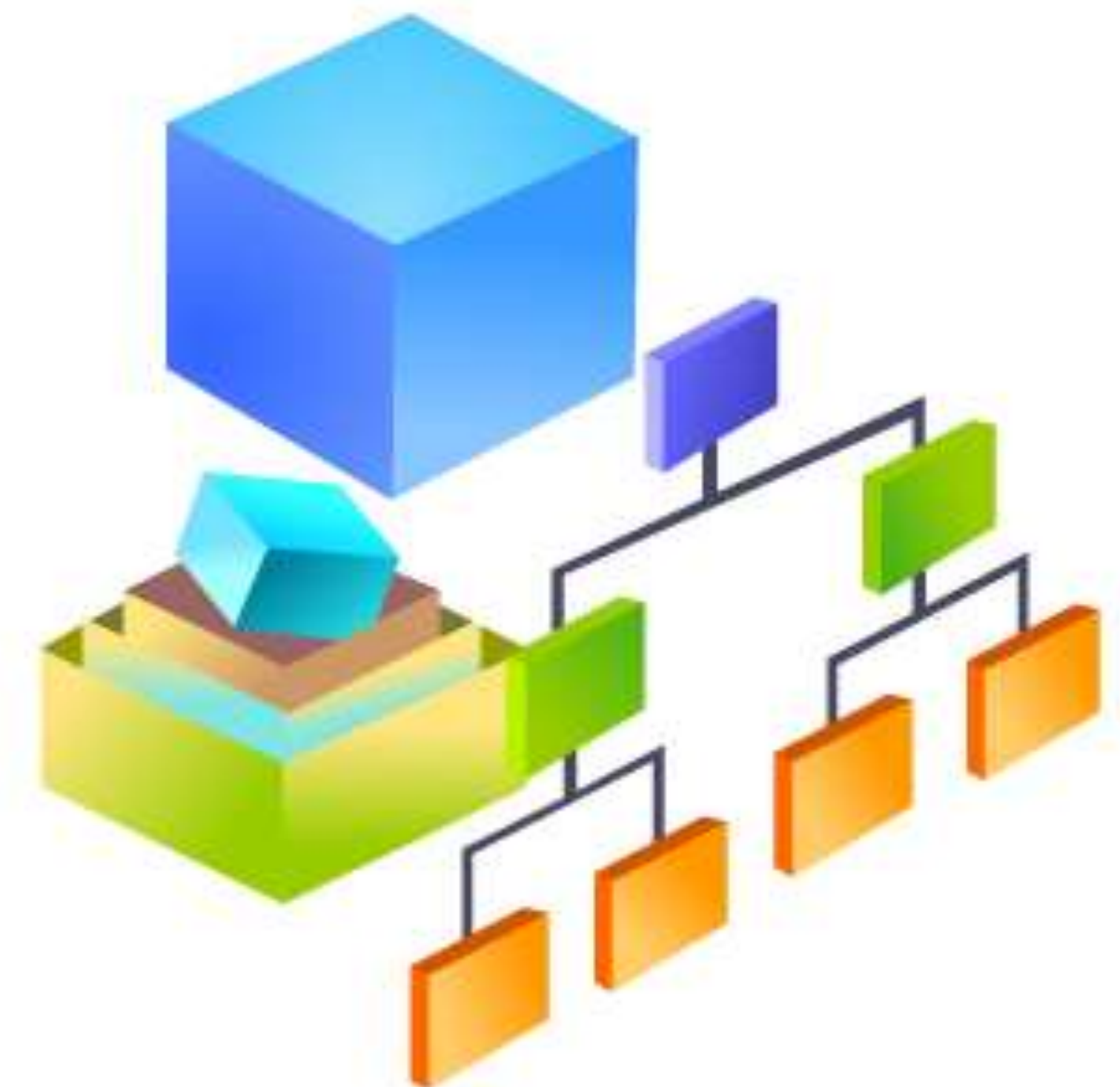


OLTP x OLAP





- Suporte a análise multidimensional
- Cria um design que tem foco na performance
- Permite otimização
- Design extensivo
- Permite acesso de várias ferramentas de análise

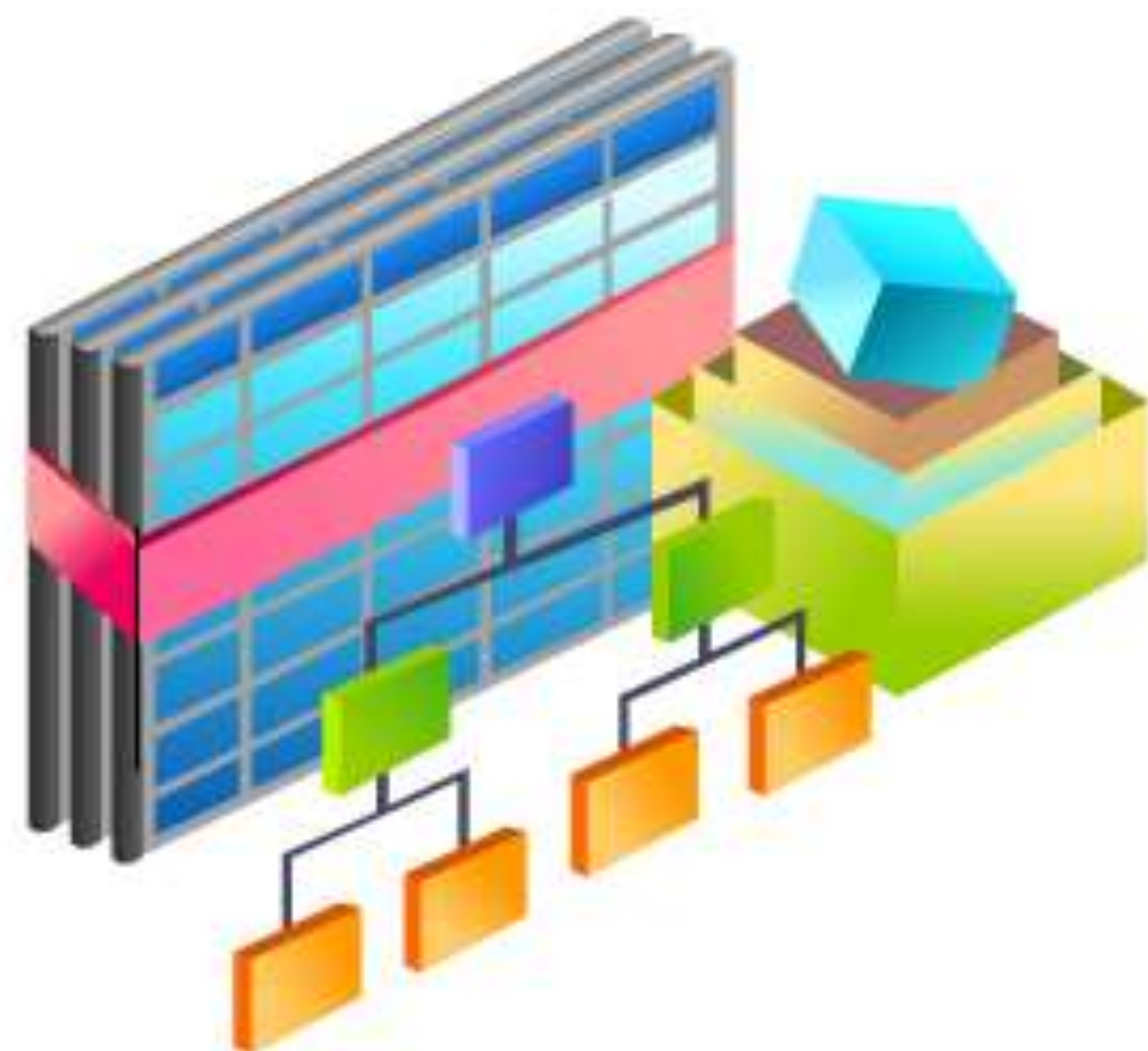




Data Science Academy

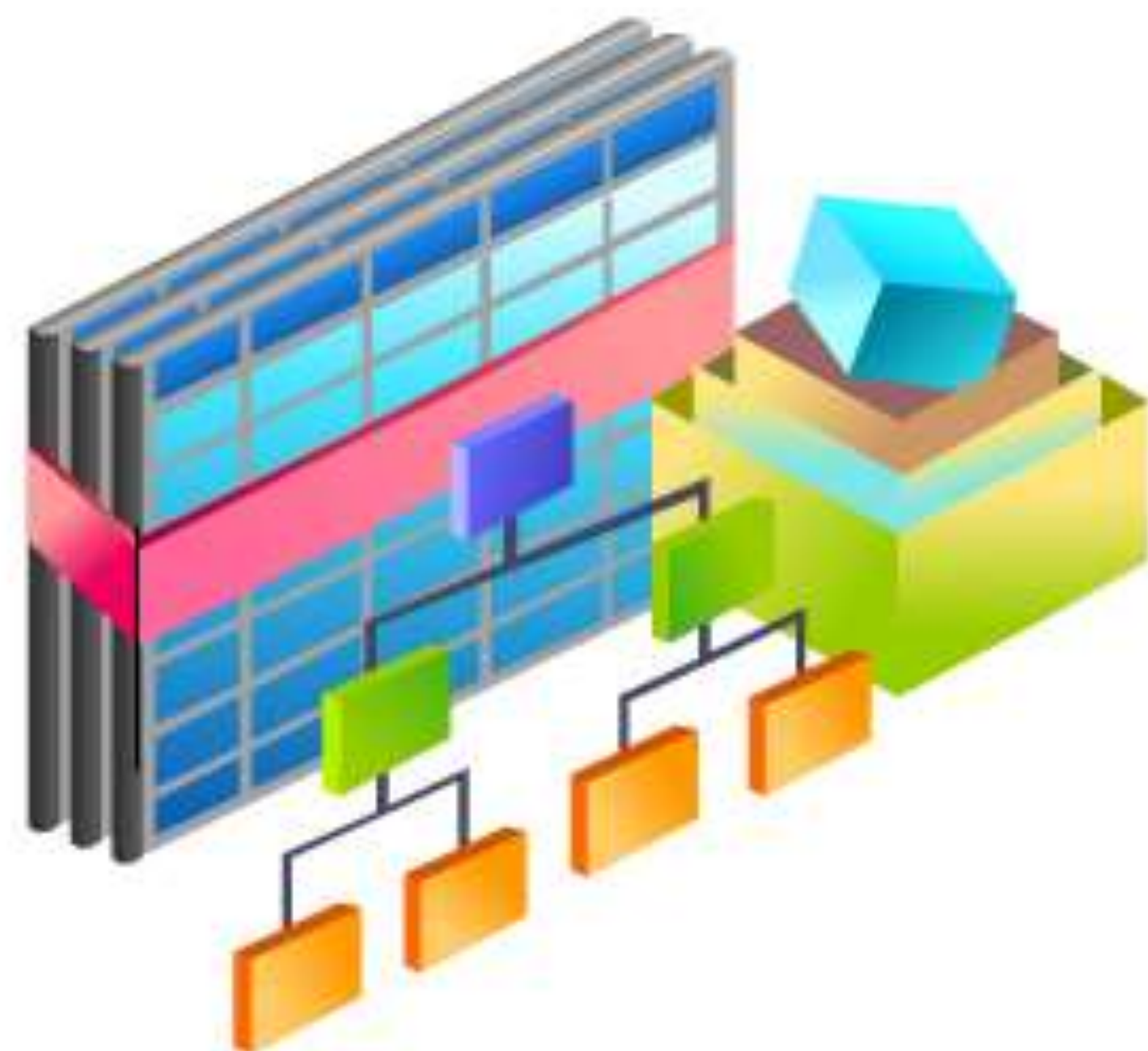
O Que são Dimensões?



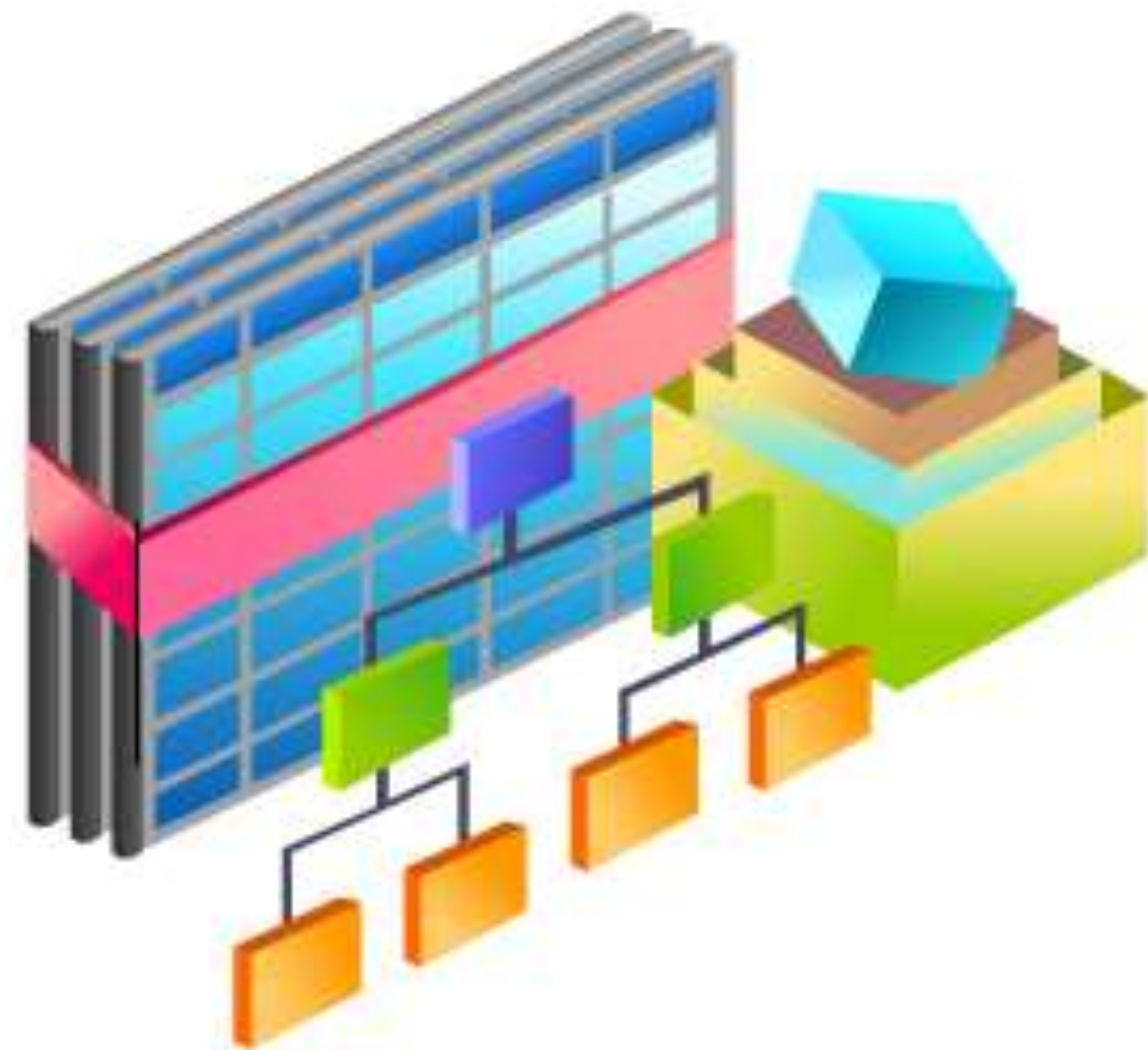


Uma dimensão é um conjunto de objetos que descrevem e classificam os fatos através de seus atributos.

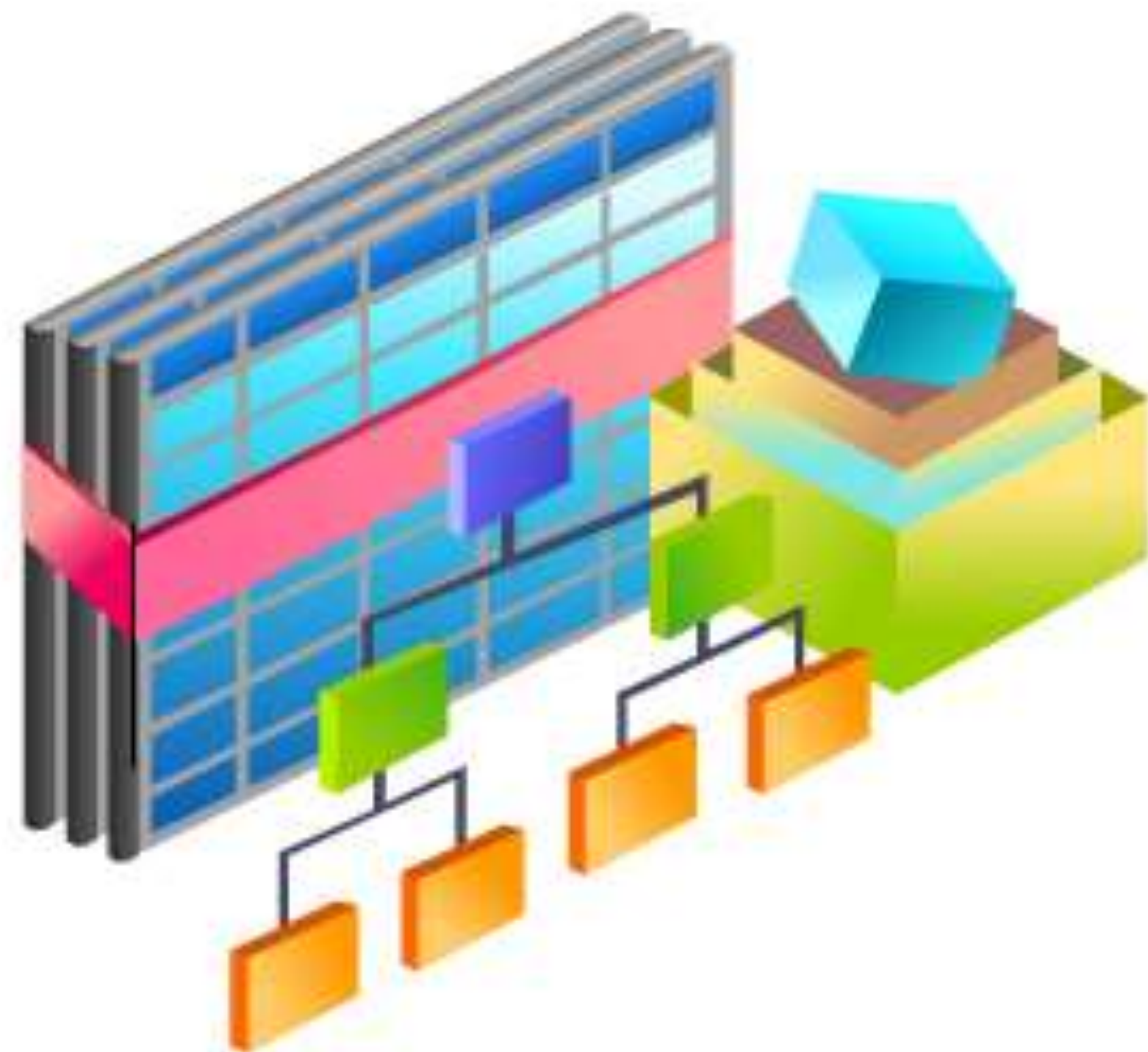




São os atributos e as hierarquias das dimensões que permitem realizar operações interessantes sobre os fatos, como as operações de fatiamento, agregação e navegação hierárquica descritas anteriormente.



Em geral, os atributos de dimensão são dados textuais.

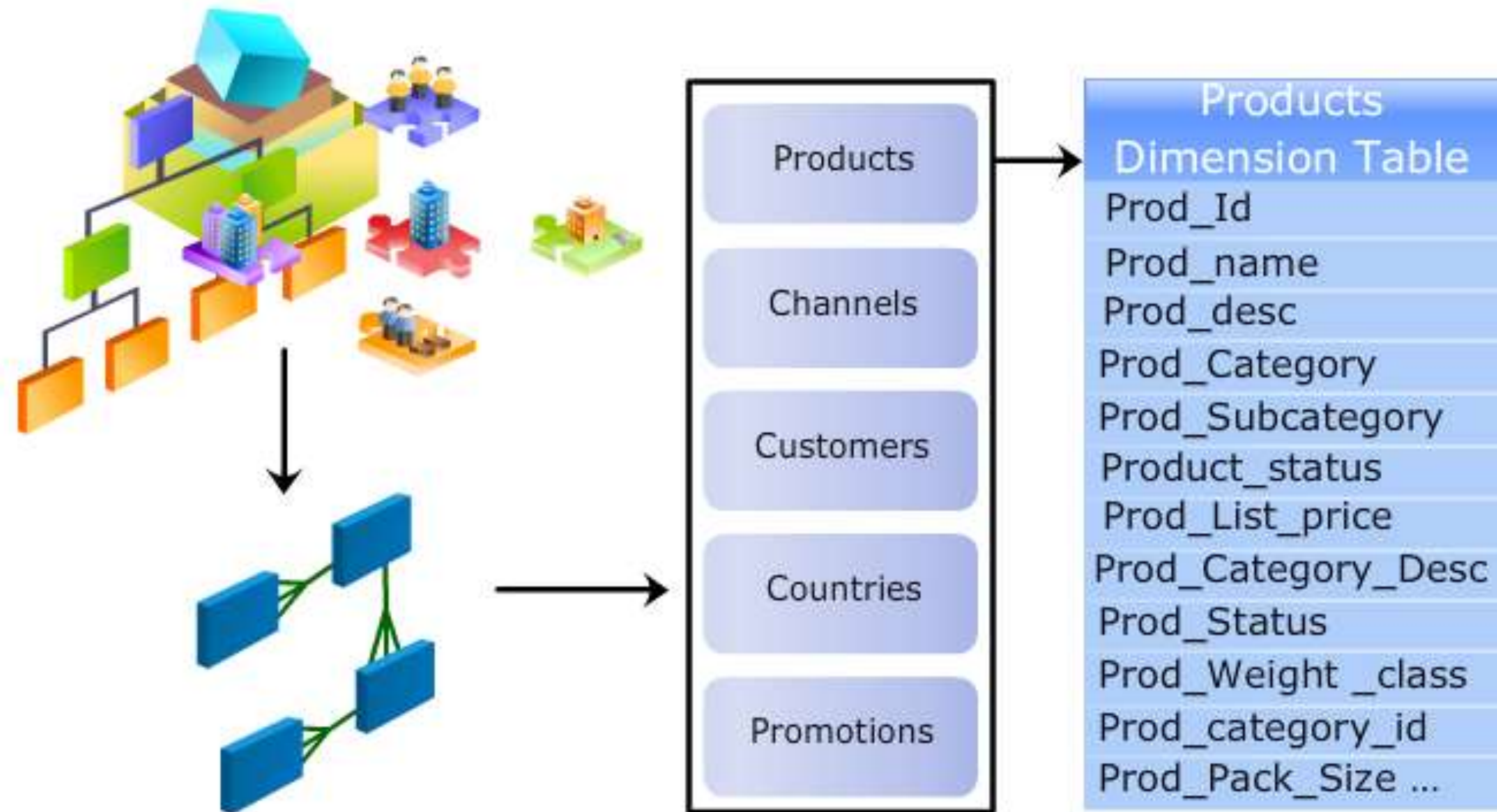


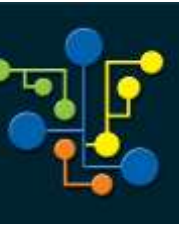
- Contém informações textuais que representam atributos de negócio.
- Contém dados que são relativamente estáticos.
- São relacionados com a tabela FATO através de chaves estrangeiras.



Business dimensions

Dimension tables

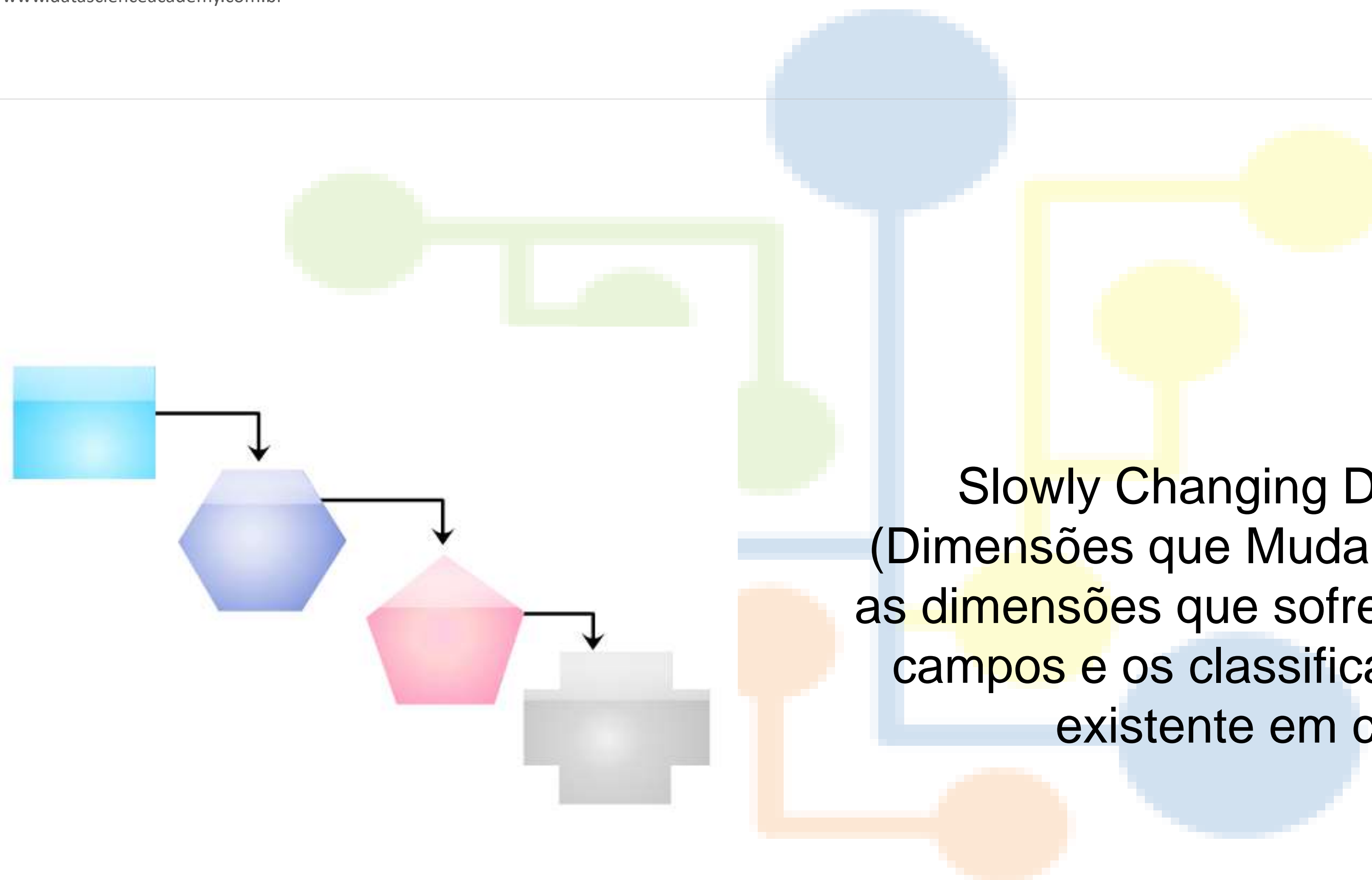




Data Science Academy

Slowly Changing Dimensions





Slowly Changing Dimensions ou SCDs
(Dimensões que Mudam Lentamente) retratam
as dimensões que sofrem atualizações em seus
campos e os classifica pelo tipo de mudança
existente em cada uma delas.





Registro Original

Chave Dimensão	Código	Nome Setor	Nome Responsável
0001	86	Departamento de Vendas	Maurício de Souza

SCD Tipo 1

Chave Dimensão	Código	Nome Setor	Nome Responsável
0001	86	Departamento de Vendas	Pedro Mariano

SCD Tipo 2

Chave Dimensão	Código	Nome Setor	Nome Responsável
0001	86	Departamento de Vendas	Maurício de Souza
0002	86	Departamento de Vendas	Pedro Mariano

SCD Tipo 3

Chave Dimensão	Código	Nome Setor	Nome Responsável Anterior	Nome Responsável Atual
0001	86	Departamento de Vendas	Maurício de Souza	Pedro Mariano



Data Science Academy

Tipos de Dimensões





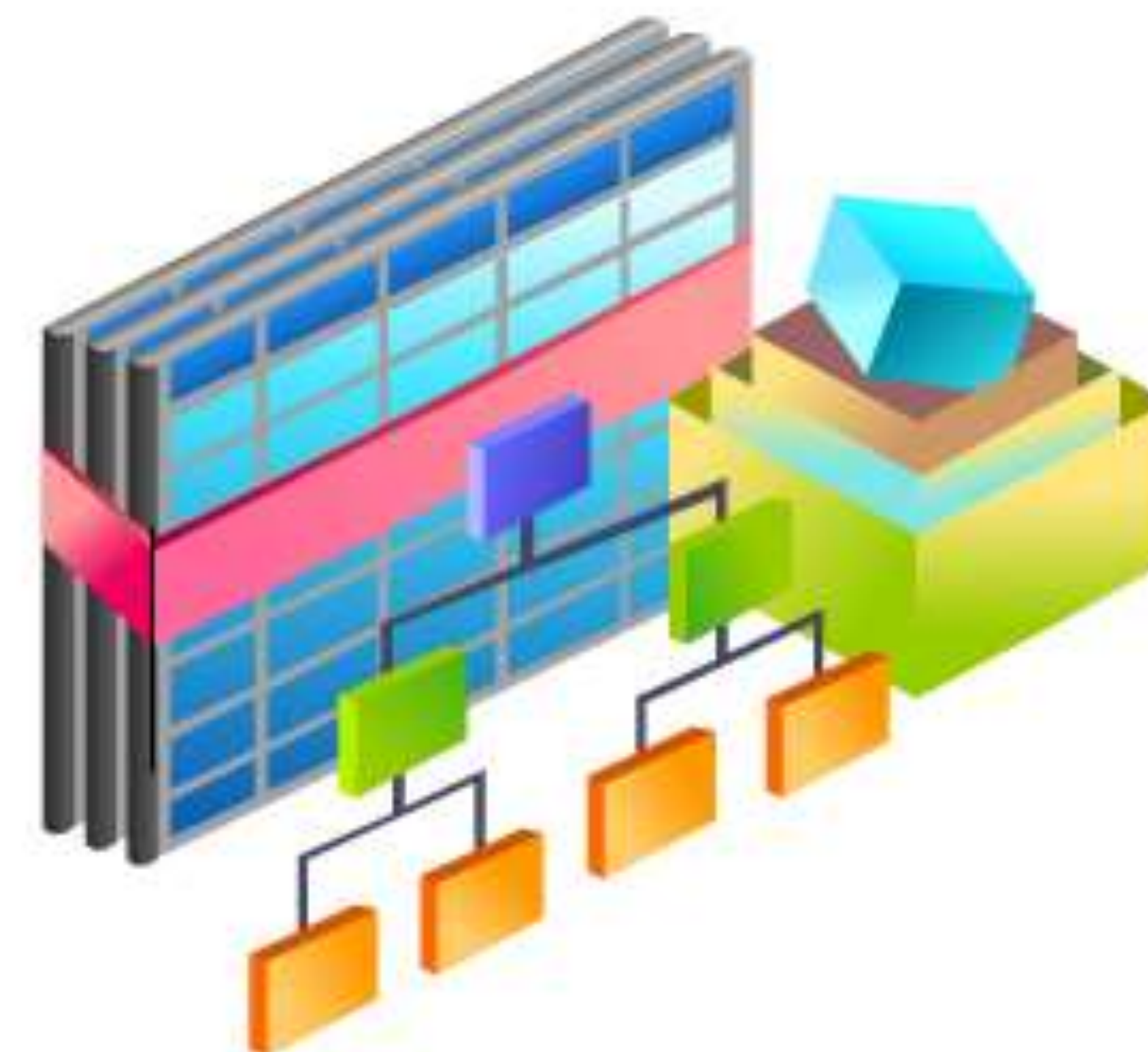
Data Science Academy

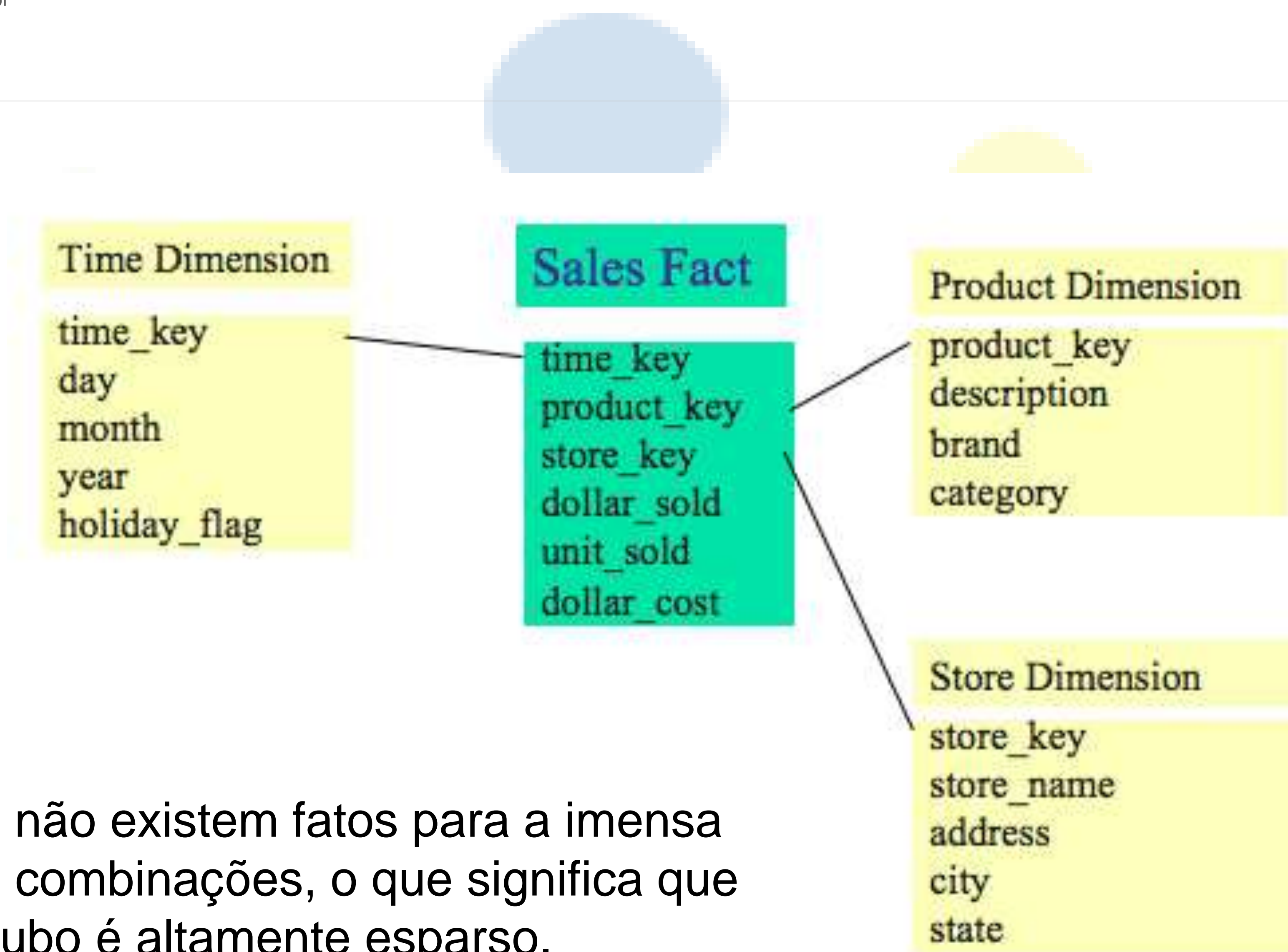
O Que são Fatos?





Os fatos são medições do negócio.
Geralmente são dados numéricos e aditivos,
ou seja, podem ser agregados por soma,
média ou outras funções.





Na prática, não existem fatos para a imensa maioria das combinações, o que significa que o cubo é altamente esparsos.



A tabela fato armazena o que ocorreu, é o fato propriamente dito, por isso ela tem esse nome, porque é o fato ocorrido. A tabela fato está sempre ligada a duas ou mais dimensões e não existe tabela fato com menos de duas dimensões.





Data Science Academy

Fatos Aditivos, Semi-Aditivos e Não-Aditivos





Tudo que a empresa for mensurar é uma métrica. Na maioria das vezes, a métrica vai ser o que o usuário quer medir. Pode ser também chamada de quantificador ou medida.

Alguns chamam de KPI (Key Performance Indicator), mas KPI também pode ser considerado um cálculo entre duas métricas.

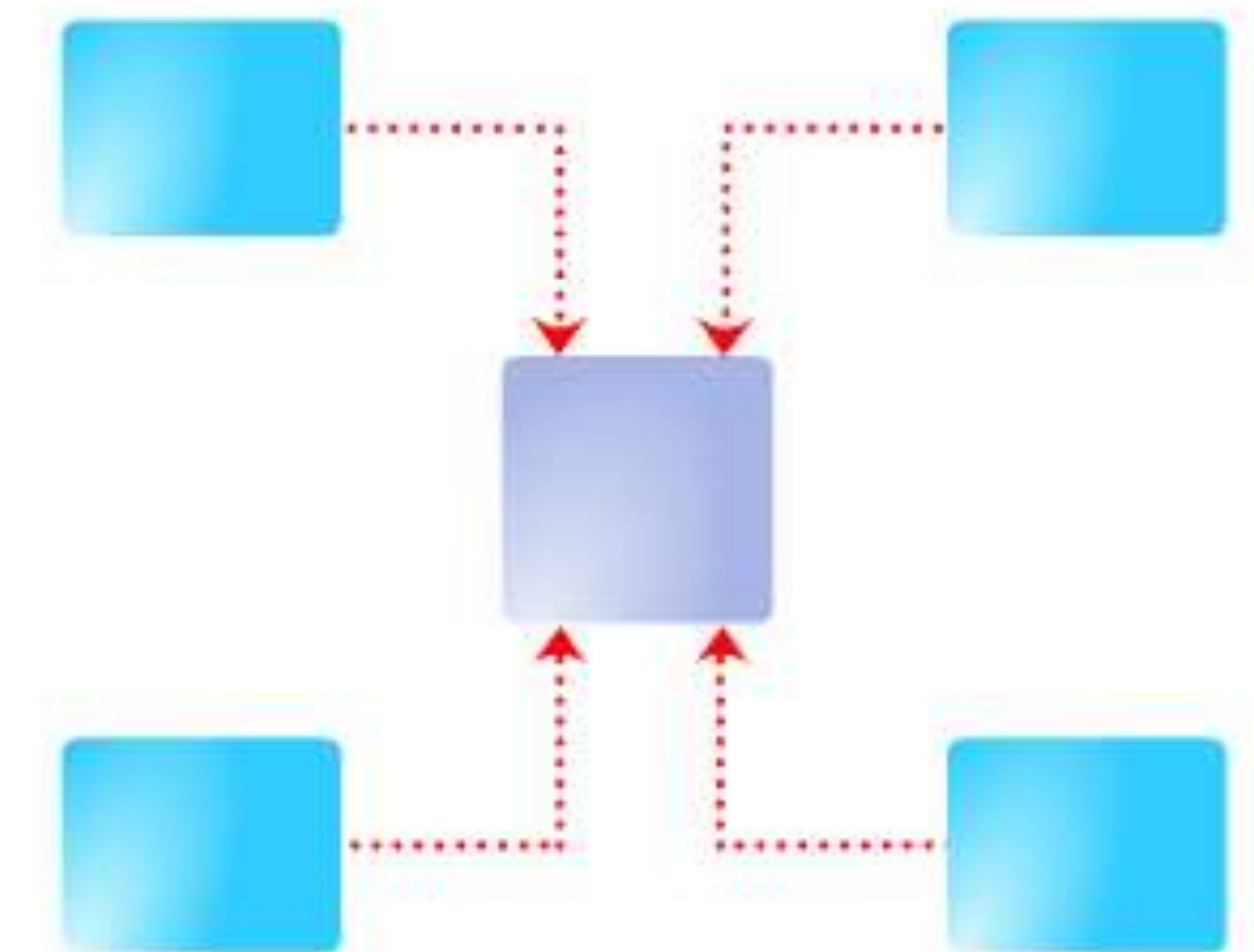
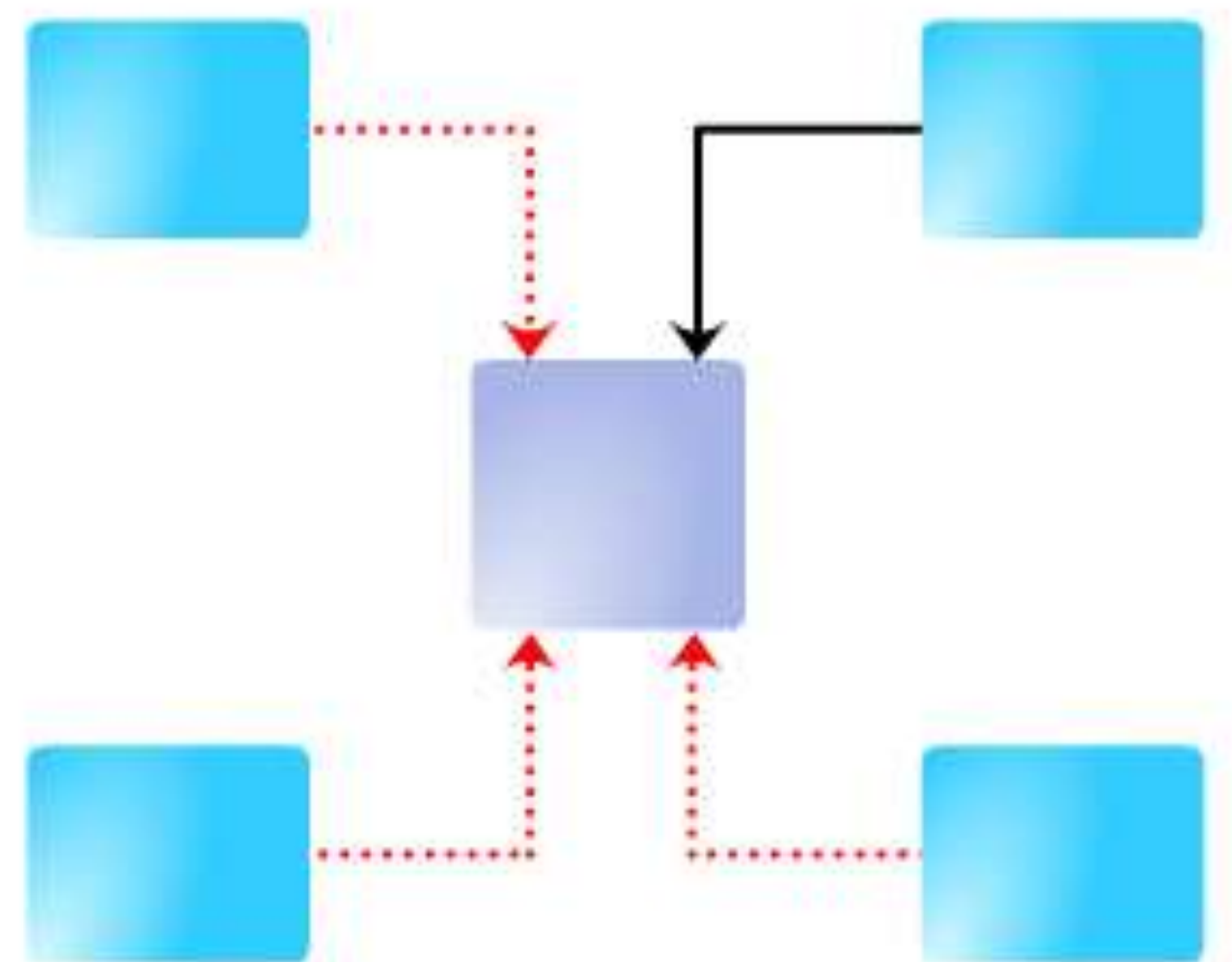
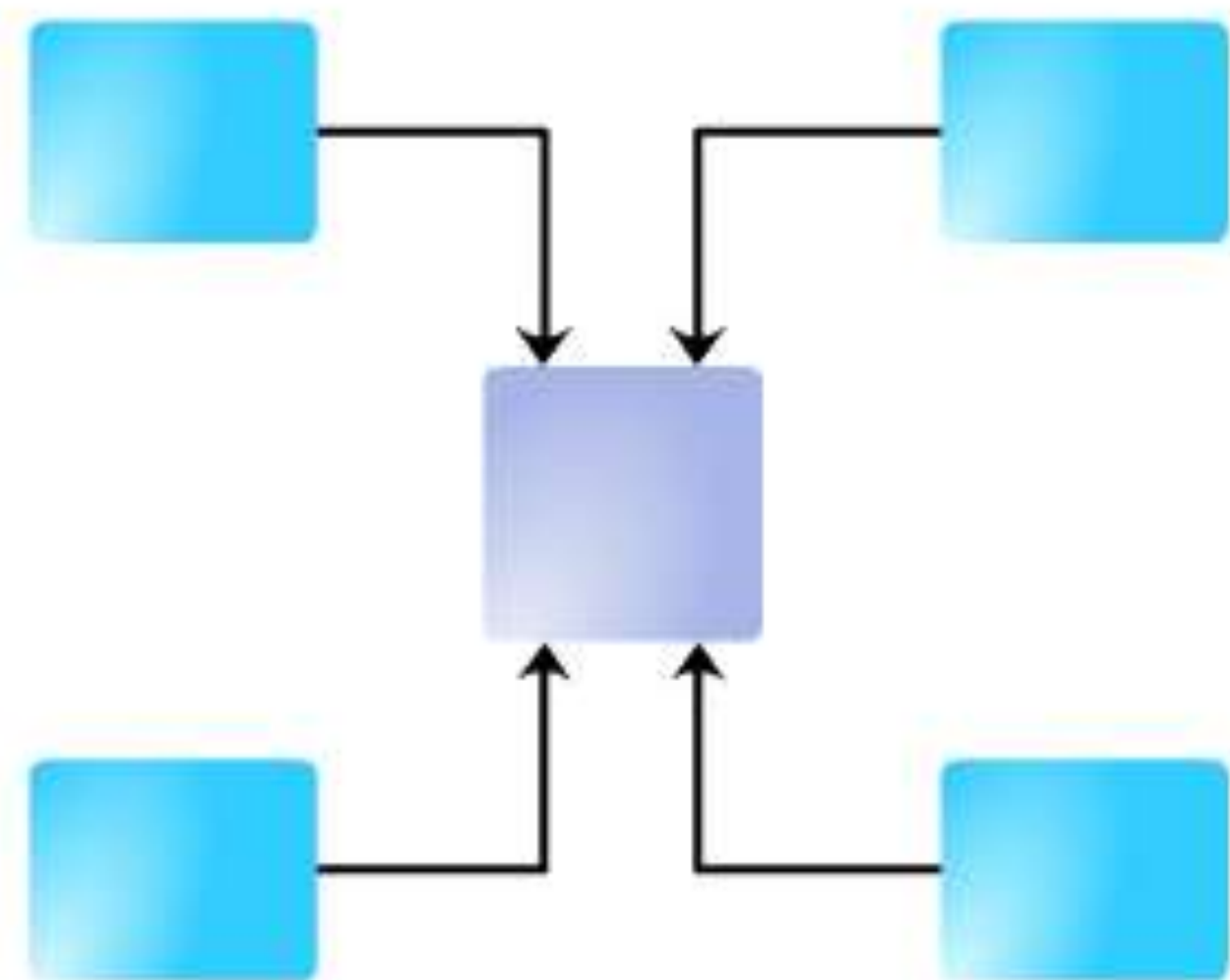




Aditivos
Adicionados através de todas
as dimensões

Semi-Aditivos
Adicionados através de
algumas dimensões

Não-Aditivos
Não podem ser adicionados
através das dimensões



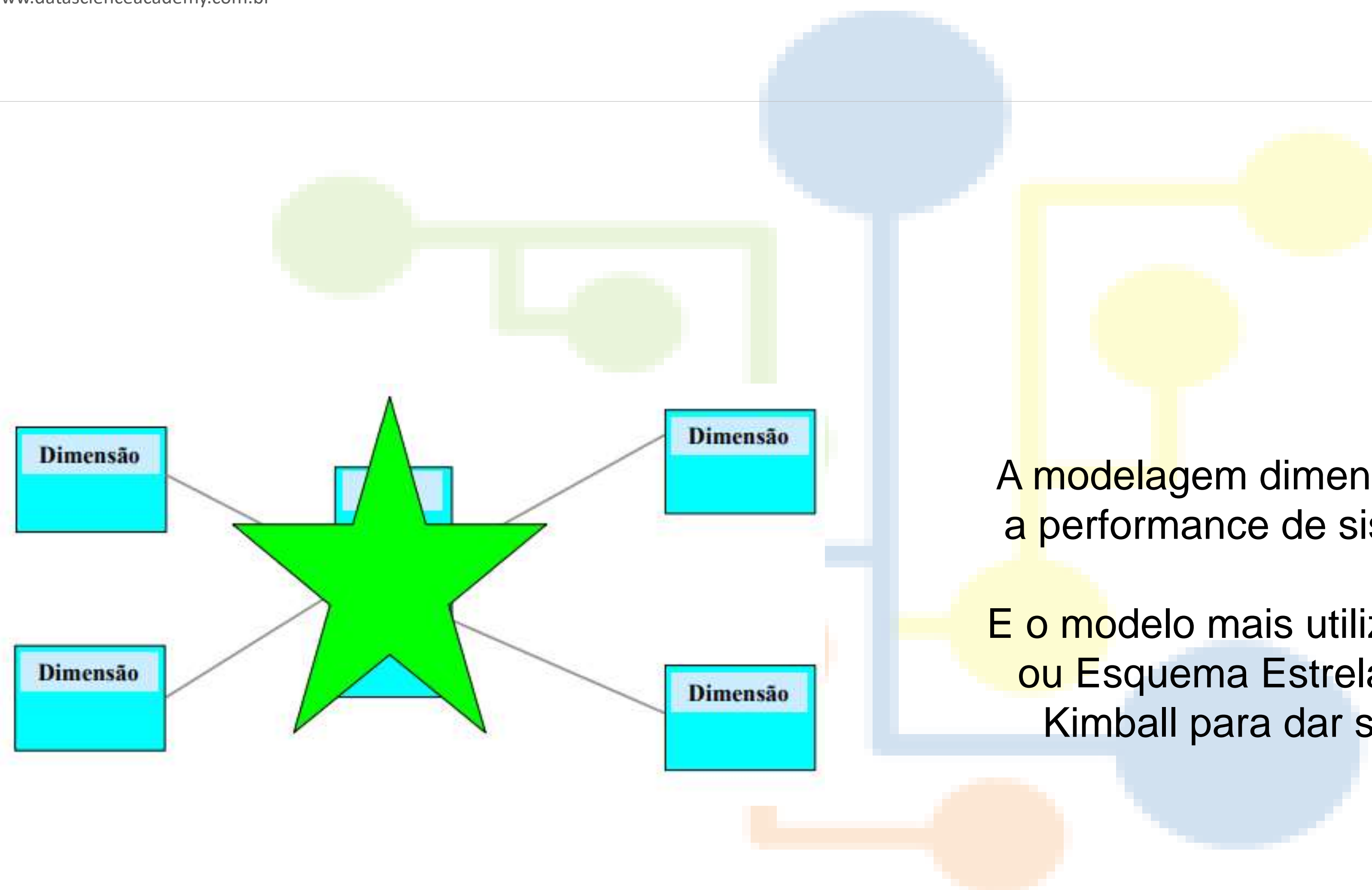




Data Science Academy

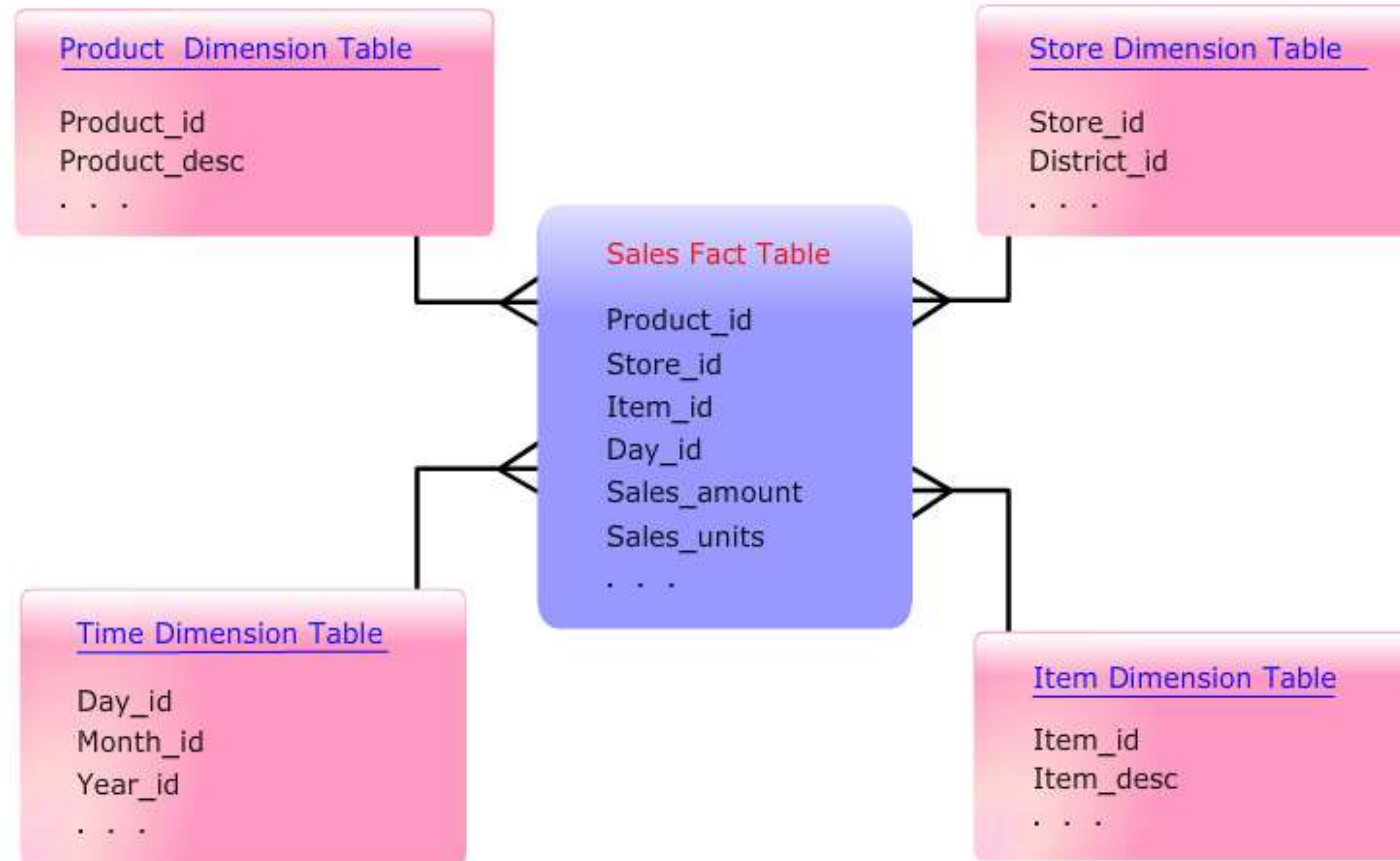
Star Schema





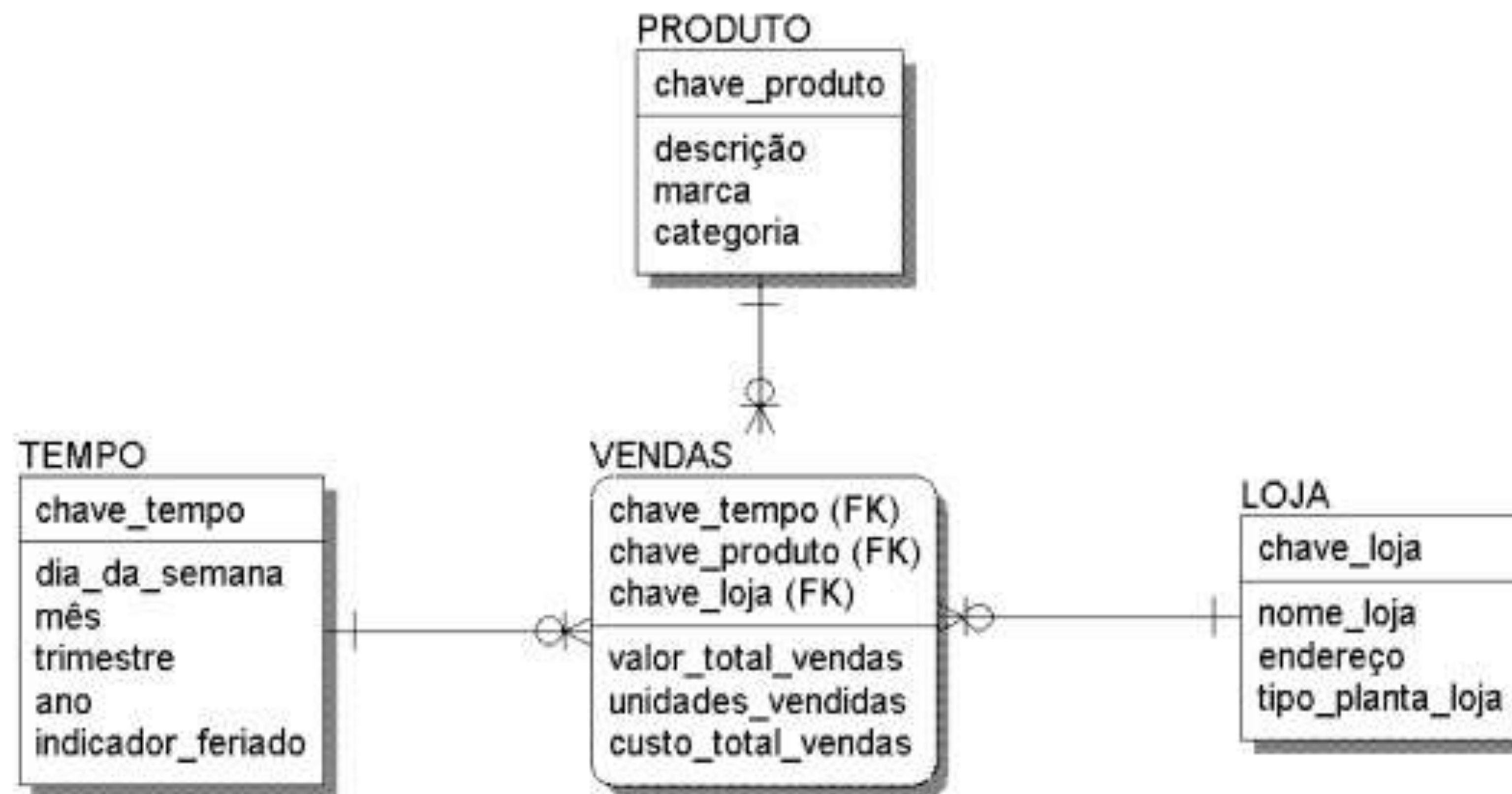
A modelagem dimensional é utilizada para melhorar a performance de sistemas voltados para consulta.

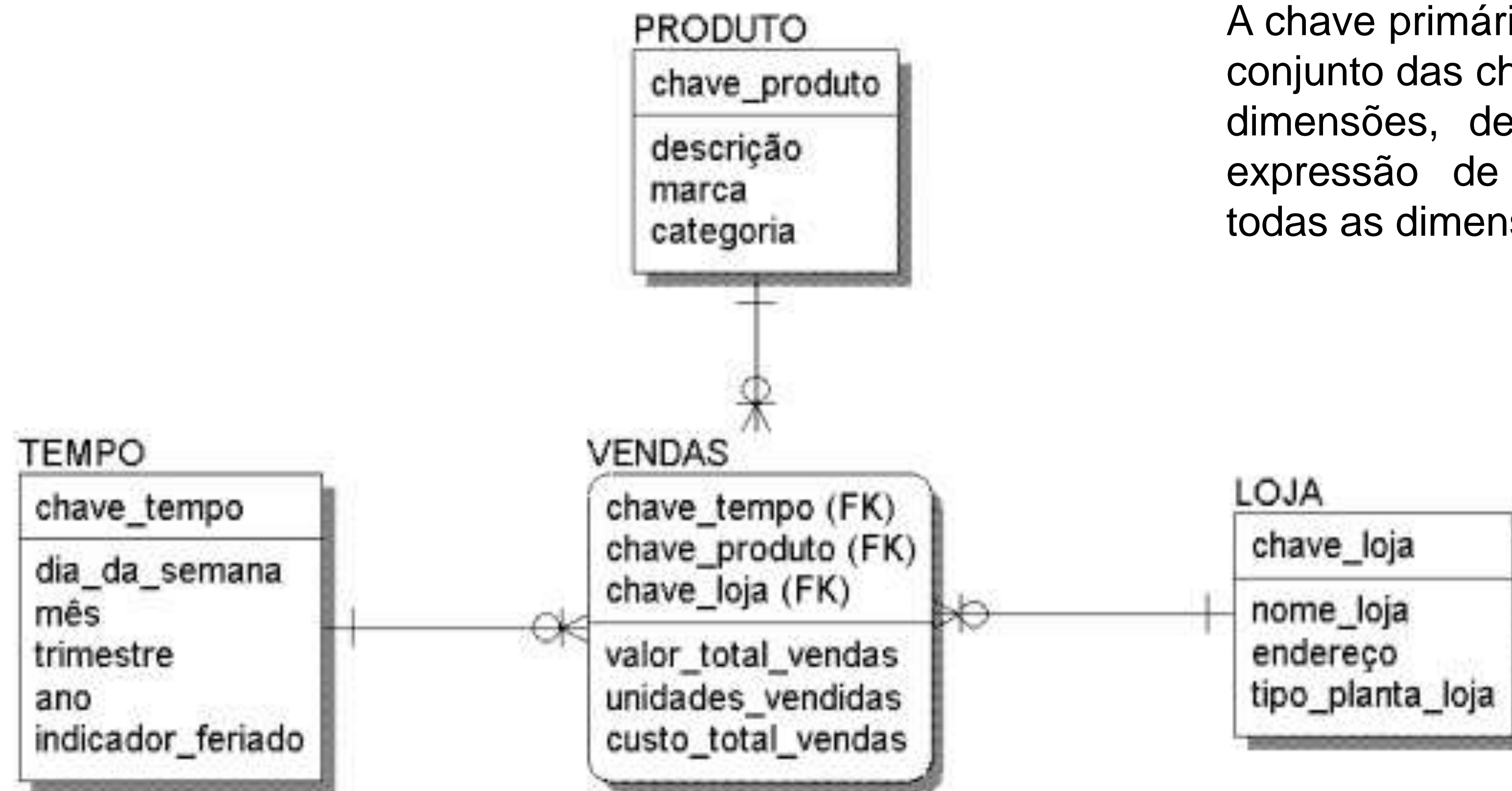
E o modelo mais utilizado para isso é o Star Schema ou Esquema Estrela, que foi idealizado por Ralph Kimball para dar suporte à tomada de decisão.



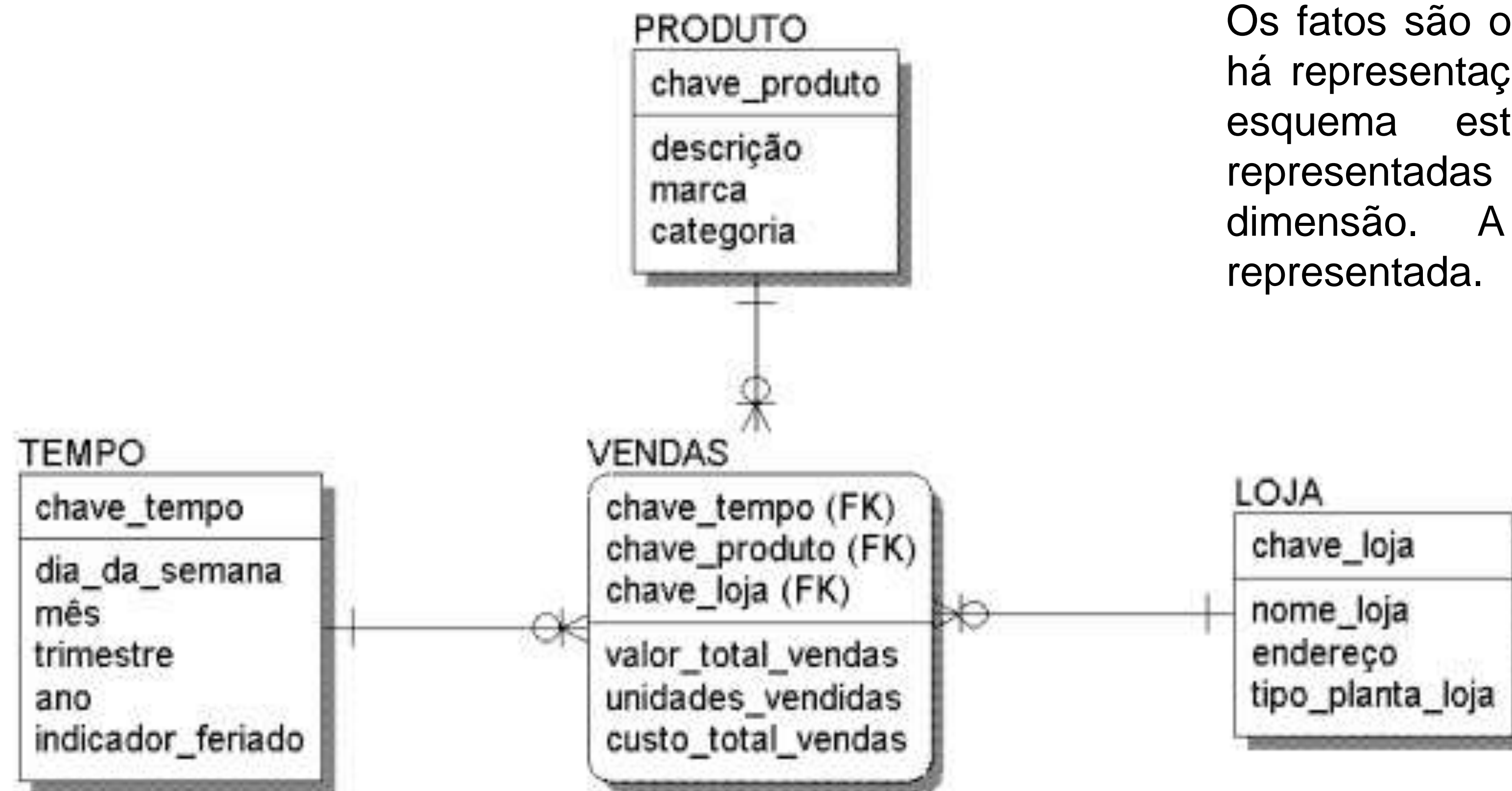


O Star Schema utiliza a abordagem relacional com algumas importantes restrições para representar o modelo dimensional. São usados os mesmos componentes de um diagrama ER lógico, como entidades, relacionamentos, chaves primária e estrangeira, cardinalidade.





A chave primária da tabela de fatos é formada pelo conjunto das chaves estrangeiras provenientes das dimensões, de tal forma que ela é sempre a expressão de um relacionamento *n-ário* entre todas as dimensões.

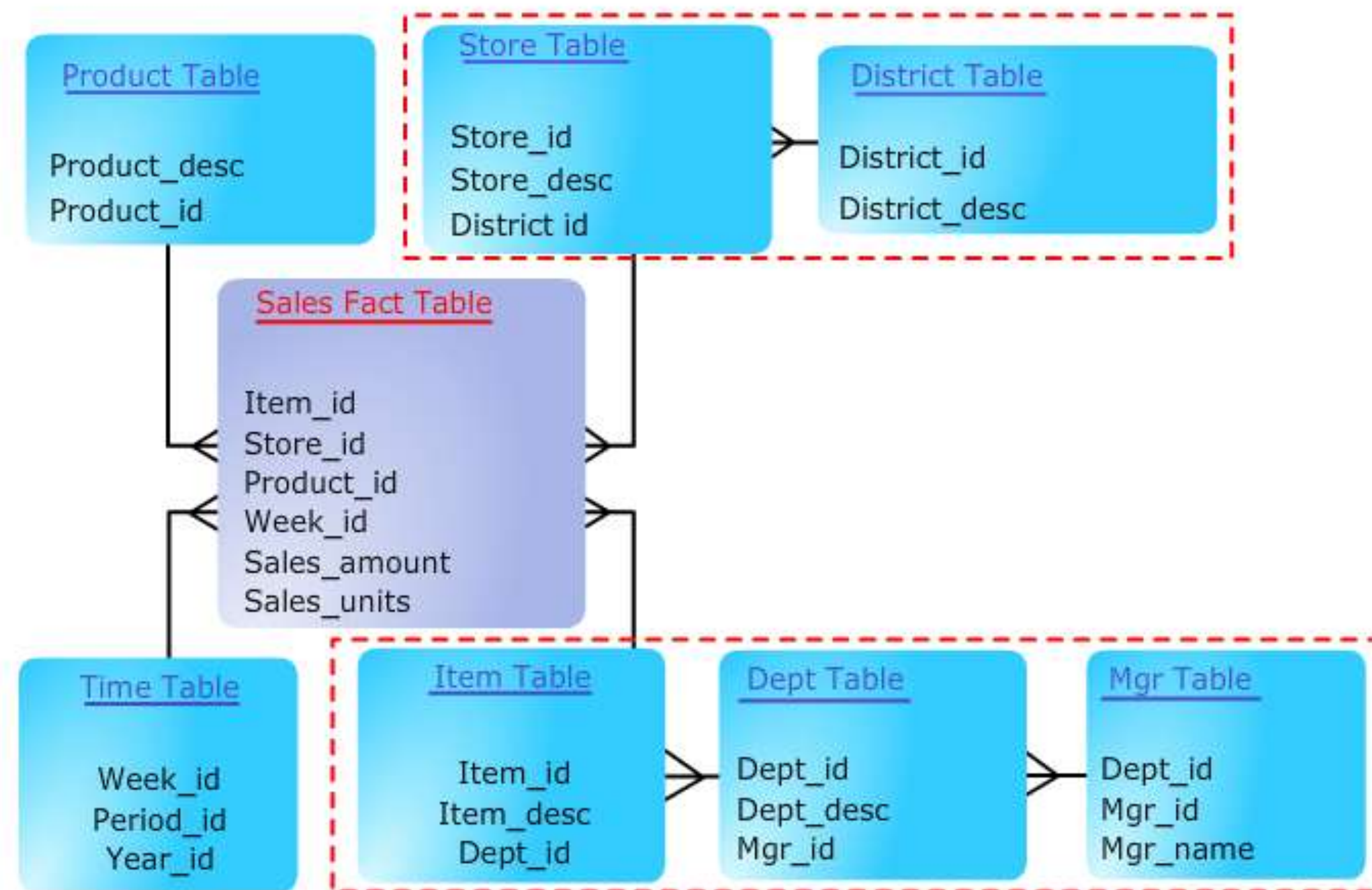


Os fatos são os atributos da tabela de fatos. Não há representação explícita para as hierarquias no esquema estrela. Elas são implicitamente representadas pelos atributos das tabelas de dimensão. A aditividade também não é representada.



SnowFlake Schema

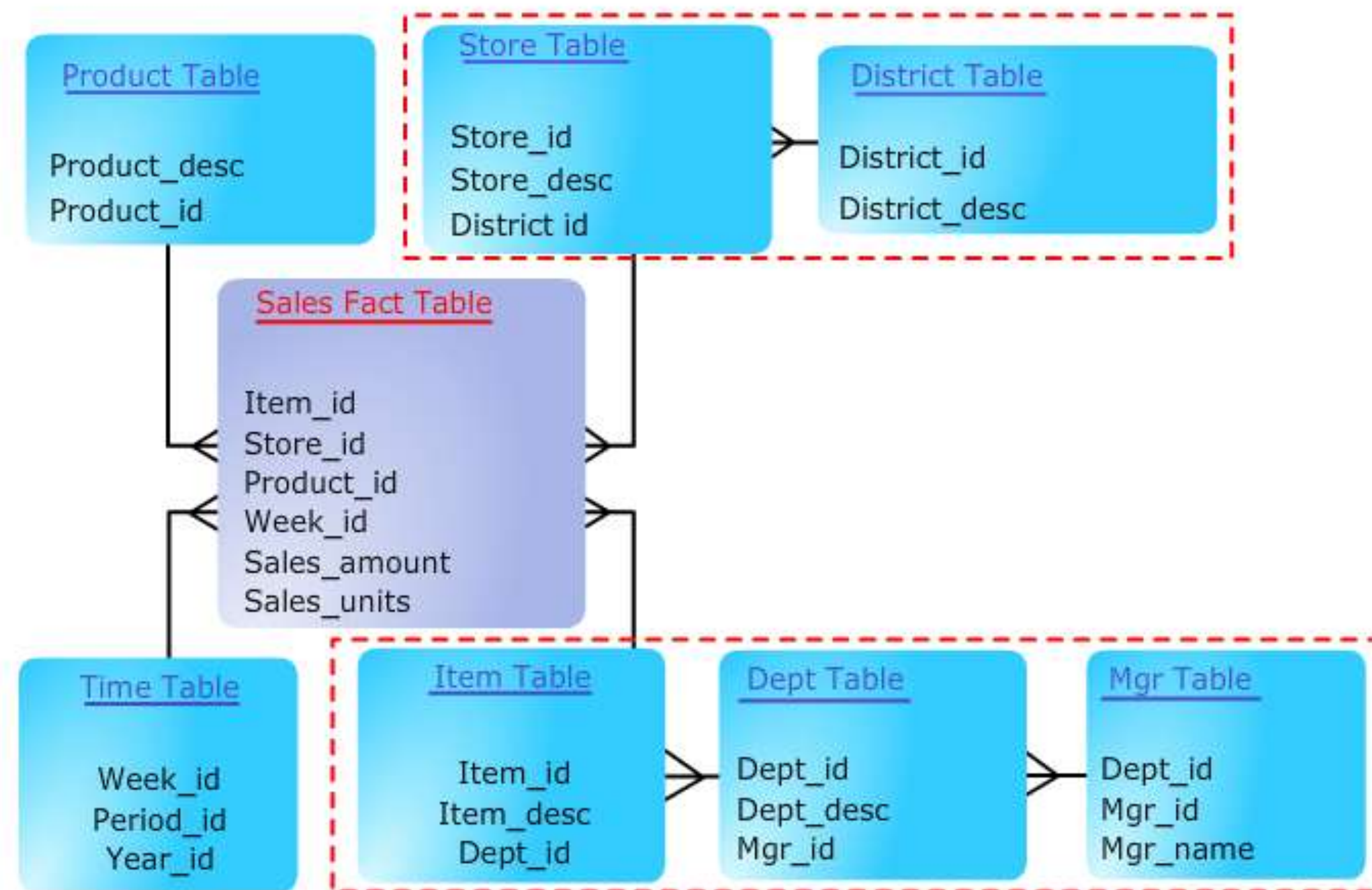




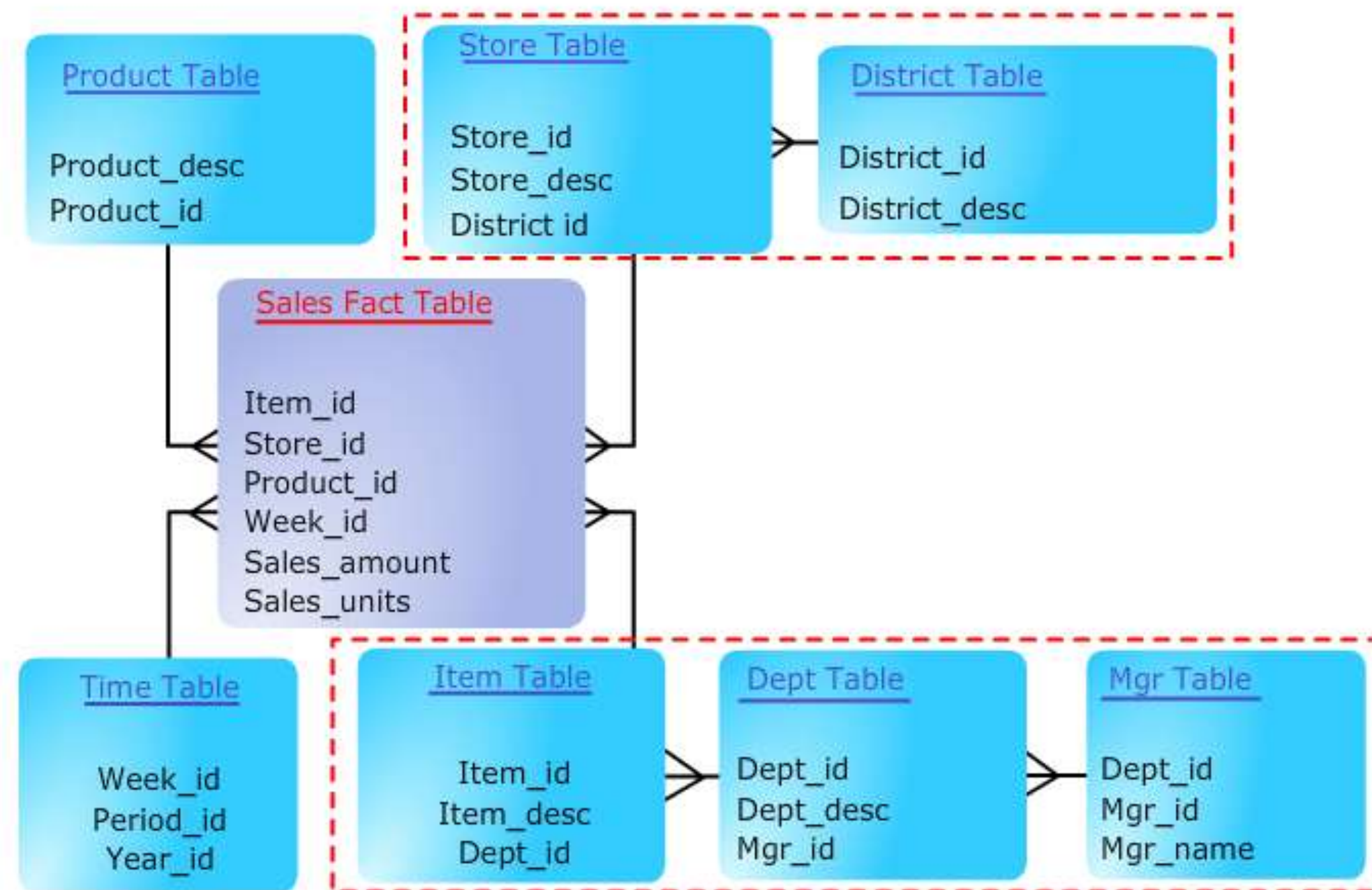
Existe uma variação do esquema estrela conhecida com *esquema floco-de-neve* (*snowflake schema*), que faz uma normalização nas tabelas de dimensão.



Normalização?



Existe uma variação do esquema estrela conhecida com *esquema floco-de-neve* (*snowflake schema*), que faz uma normalização nas tabelas de dimensão.



- É mais flexível a mudanças
- Possui metadados mais complexos
- O processo ETL pode ser mais rápido
- Degrada performance de queries




Data Science Academy

Dados Normalizados X Dados Desnormalizados







The background features a faint, stylized diagram of a data warehouse schema. It consists of several interconnected nodes represented by circles in light blue, light green, light yellow, and light orange. The nodes are connected by lines, forming a complex network that illustrates the relationships between different data tables or dimensions in a normalized or denormalized structure.

Dados Normalizados (OLTP) **x** **Dados Desnormalizados (OLAP)**



Banco Transacional (OLTP)

DW

Tabela Cliente

ID_cliente
Nome_cliente
ID_localidade

Tabela Localidade

ID_localidade
Nome_localidade

Dimensão Cliente

ID_cliente
Nome_cliente
Nome_localidade



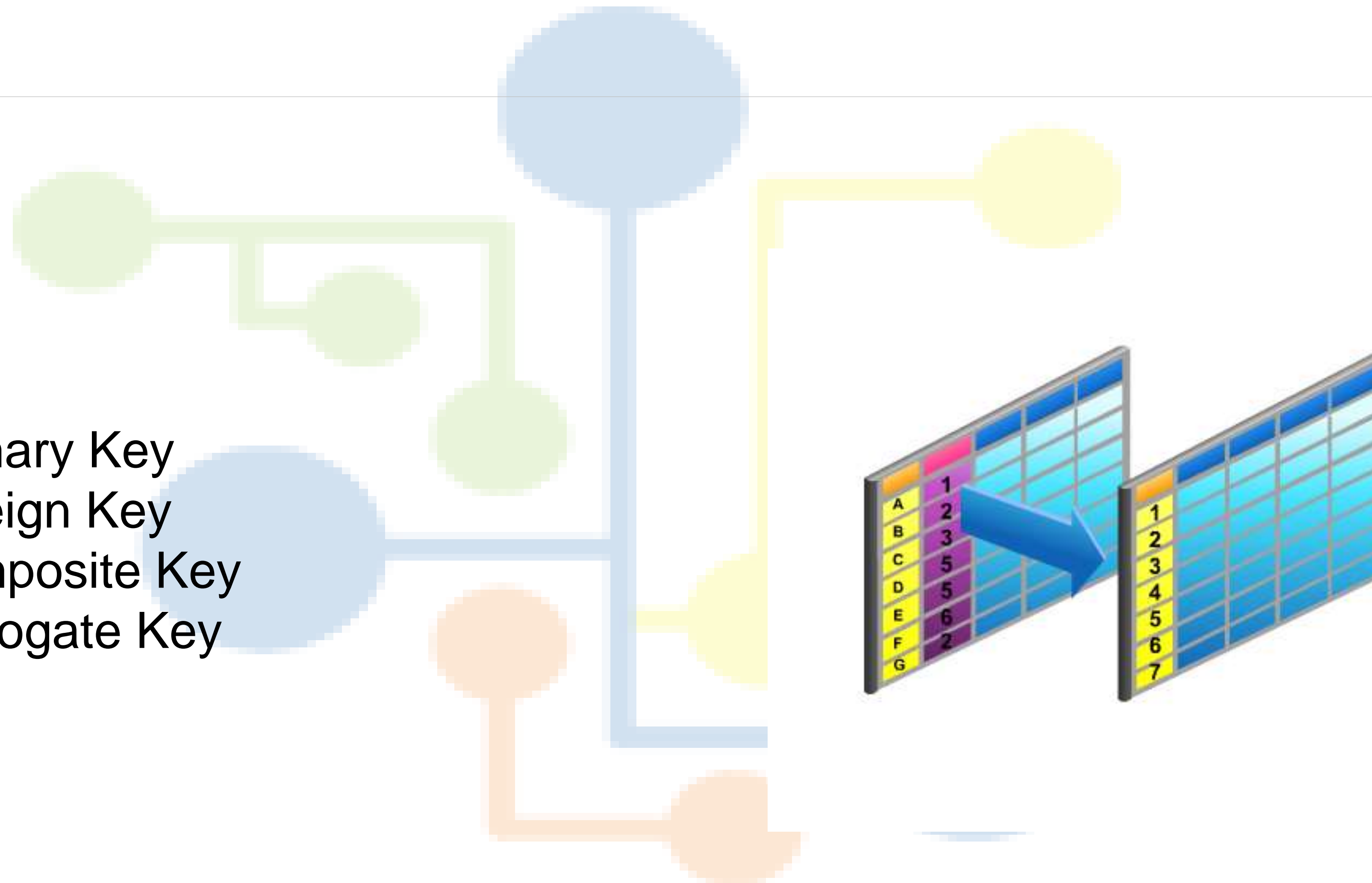
Data Science Academy

Tipos de Chaves





- Primary Key
- Foreign Key
- Composite Key
- Surrogate Key





Surrogate Keys

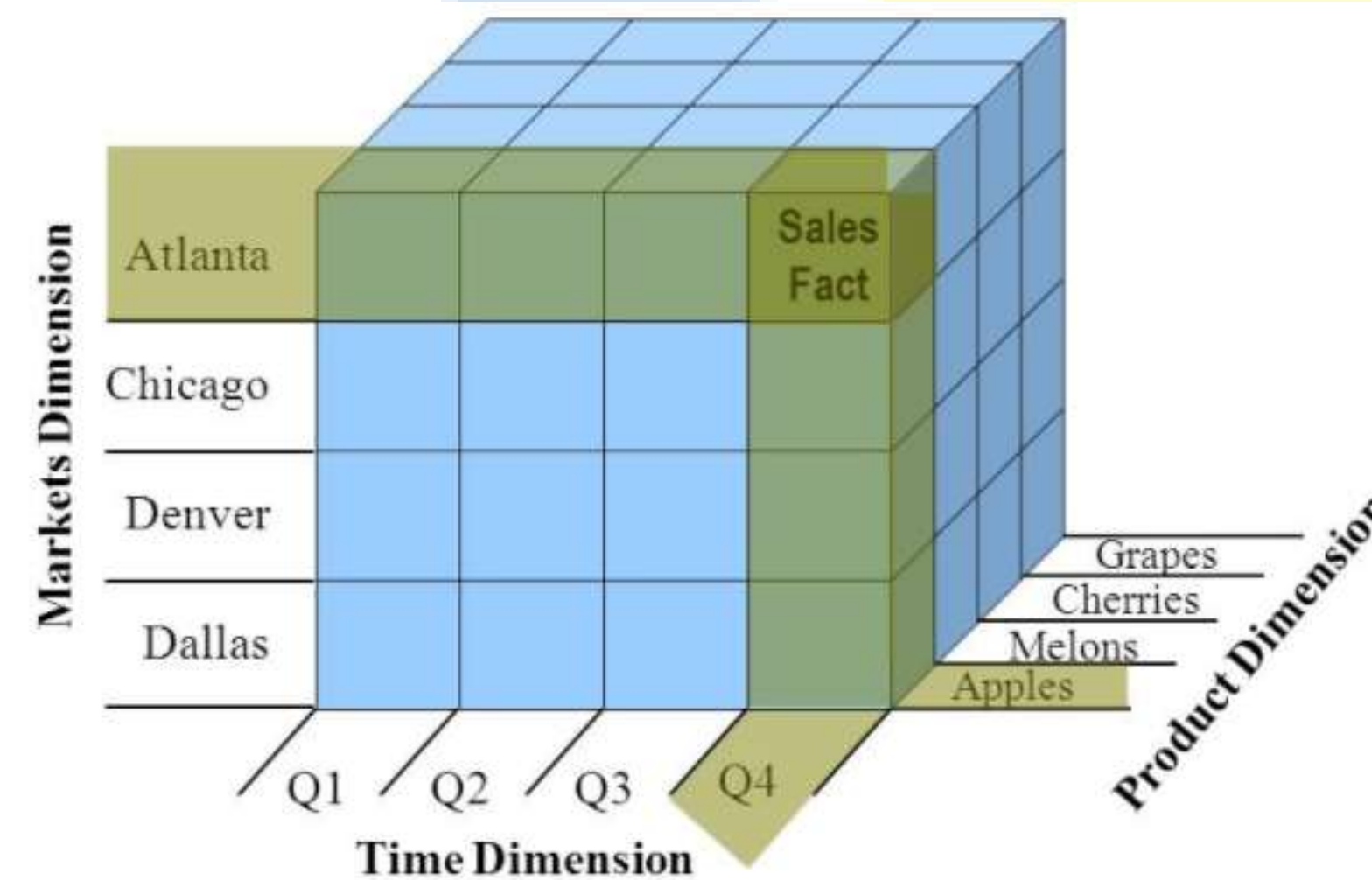
- Ajudam a manter a estabilidade, através da neutralidade.
- Evitam manutenção custosa de tabelas, especialmente das tabelas fatos.
- Chaves naturais podem ter problemas de unicidade, ausência, tamanhos exagerados.
- Chaves artificiais ficam transparentes (invisíveis) para os usuários, servindo apenas como ligação entre dimensões e fatos.
- Campos naturais poderão ser indexados, tornando as consultas amistosas.
- Se produzidas automaticamente, deve-se ter cuidado no processo de preparação (ETL), especialmente nos reprocessamentos.
- A desvantagem das chaves artificiais é que a tabela fato não pode ser consultada diretamente, pois os campos de filtro estarão armazenados nas dimensões.



Data Science Academy

Dimensão Tempo

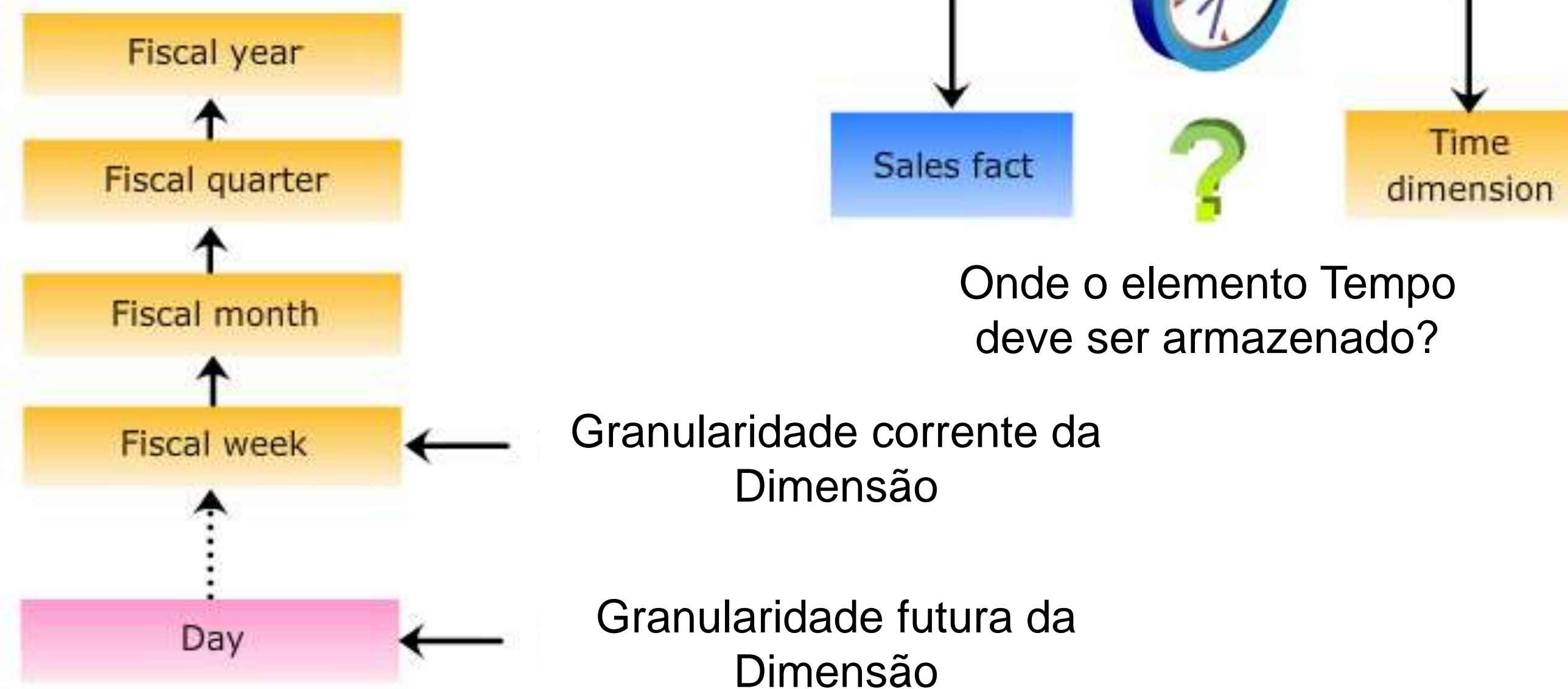




- Qual a diferença no total de vendas em dias úteis e finais de semana?
- Qual o número de caixas de cerveja vendidas no Carnaval deste ano?
- Qual a taxa de ocupação do hotel no mesmo período do ano passado?
- Quantos itens são produzidos por hora e por máquina?
- Quantas mercadorias são retornadas por mês?



- A dimensão Tempo é crítica em um DW
- Precisamos definir o nível de granularidade para esta dimensão





Data Science Academy

Hierarquia de Dimensões





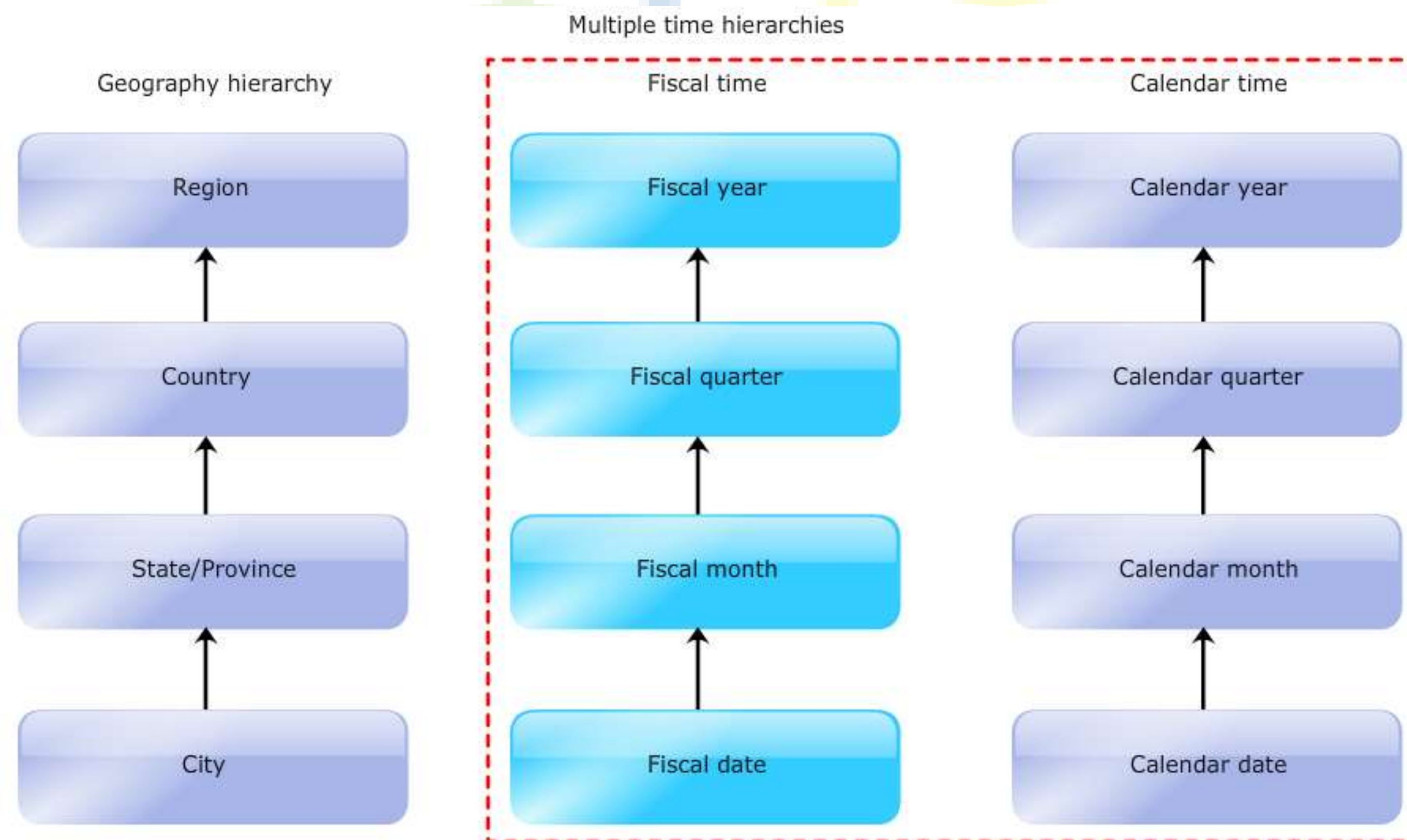
Uma dimensão pode ter múltiplas hierarquias além de outros atributos descritivos.

- Exemplo para uma empresa atacadista:

- Geografia física:
 - » cep, cidade, estado, região, país
- Geografia de vendas:
 - » território, região, zona
- Geografia de Distribuição:
 - » primária, secundária

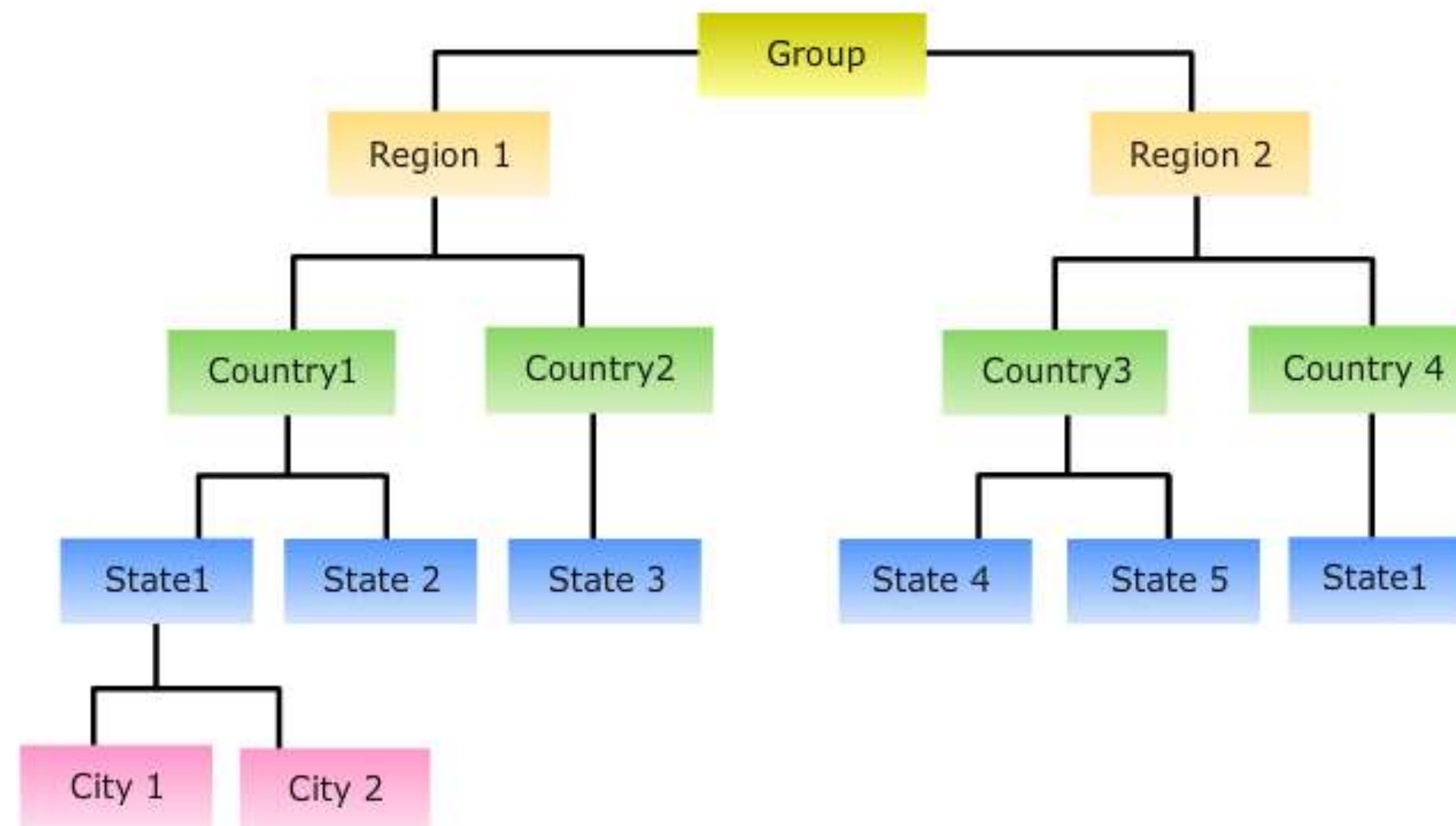


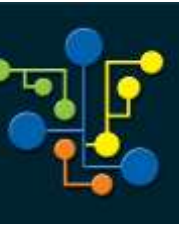
- Podemos ter diferentes hierarquias de Tempo (como ano calendário e ano fiscal).
- Os níveis de granularidade não precisam ser exatamente os mesmos.





- Hierarquias são importantes para operações de Drill-up/Drill-down.





Data Science Academy

Regras de Ouro Para a Modelagem Dimensional





Passos para a Modelagem Dimensional

- Definição da área de negócio
- Definir processos dentro da área de negócio
- Definição da granularidade desejada para os dados de cada processo
 - Considerar volumes e dificuldades de se obter o nível desejado
- Definição dos atributos e hierarquia das dimensões
 - Considerar hierarquias múltiplas
- Definição das métricas das tabelas de Fatos
 - Observar valores aditivos, semi-aditivos e não aditivos



Como Preparar o Modelo Dimensional

- Todo Fato deve ter uma Dimensão de Tempo associada.
- Um Fato deve ter, sempre, a mesma granularidade.
- Relacionamento muitos-para-muitos é um Fato.
- Usar Surrogate Key nas Dimensões.
- Criar Conformed Dimensions.
- Faça ou use um modelo de dados convencional ER como ponto de partida para o trabalho de modelagem dimensional:
 - Observe os relacionamentos 1:N existentes. Eles podem sugerir dimensões.
 - Observe as entidades fortes. Elas também podem sugerir dimensões.
 - Observe as entidades que expressam documentos como Nota Fiscal, Pedido, Ordem de Compra, etc. Elas podem sugerir fatos.
 - Observe os relacionamentos M:N. Na sua interseção, pode haver valores numéricos. Isto sugere fatos.
 - Observe os atributos que estarão nas tabelas de dimensões.
- Analise a relação de hierarquias entre esses atributos de dimensão. Atente para os relacionamentos M:N entre eles. Isto pode definir granularidade.





Como Preparar o Modelo Dimensional

- As tabelas FATOS, tipicamente, armazenam dados, valores atômicos ou agregados obtidos a partir destes.
- As métricas das tabelas FATOS são normalmente aditivas em certas dimensões.
- As tabelas FATOS possuem chaves que as conectam às diferentes DIMENSÕES que as circundam. Essa conexão se dá num nível de granularidade compatível entre elas (FATO e DIMENSÃO).
- As tabelas DIMENSÃO armazenam os valores de filtro, check, acesso e textos que caracterizam os dados trabalhados.
- As tabelas FATOS são normalmente normalizadas.
- As tabelas DIMENSÕES são normalmente desnormalizadas (Esquema Estrela).
- A granularidade combinada da tabela FATO com a de suas tabelas DIMENSÕES determina o número de linhas das tabelas do projeto.



Muito Obrigado!

É um prazer ter você aqui.
Tenha uma excelente jornada de aprendizagem.



Oportunidade



Disponibilidade



Conhecimento