

Introdução a Modelagem Dimensional

Formação Power BI Analyst

Juliana Mascarenhas

Tech Education Specialist DIO / Owner @Simplificandoredes
e @SimplificandoProgramação

Mestre em modelagem computacional | Cientista de dados

@in/juliana-mascarenhas-ds/

Objetivo Geral

- Entender o que é modelagem dimensional
- O que é Cubo multifacetado?
- Principais modelos dimensionais
- Sistemas Transacionais e Analíticos
- Comparação entre transacional e dimensional
- Criação de esquemas transactional e dimensional

Etapa 1

Desmistificando a Modelagem Dimensional

// Power BI Analyst

Modelagem Dimensional

Por que? Demandas

- Performance
- Escalabilidade
- Disponibilidade

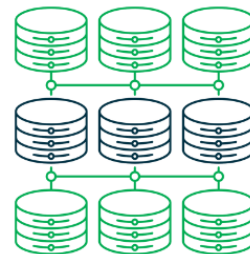


Modelagem Dimensional

Sistemas de Dados

- Modelo transacional – tradicional
- Modelo dimensional – analítico

OLTP



OLAP

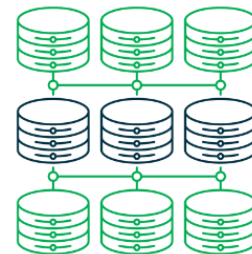


Modelagem Dimensional

Transacional

- Fim específico
- Cenário otimizado
- Suporte a operação
- SGBDs

OLTP



OLAP

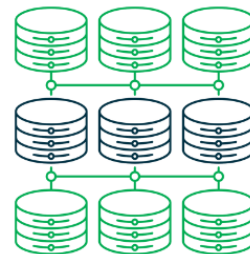


Modelagem Dimensional

Analítico

- Permite redundâncias
- Esquema flexível
- Foco em analyses
- Modelo em cubo

OLTP



OLAP



Modelo em cubo?

Table

Columns

Rows

Rows = records
Columns = attributes



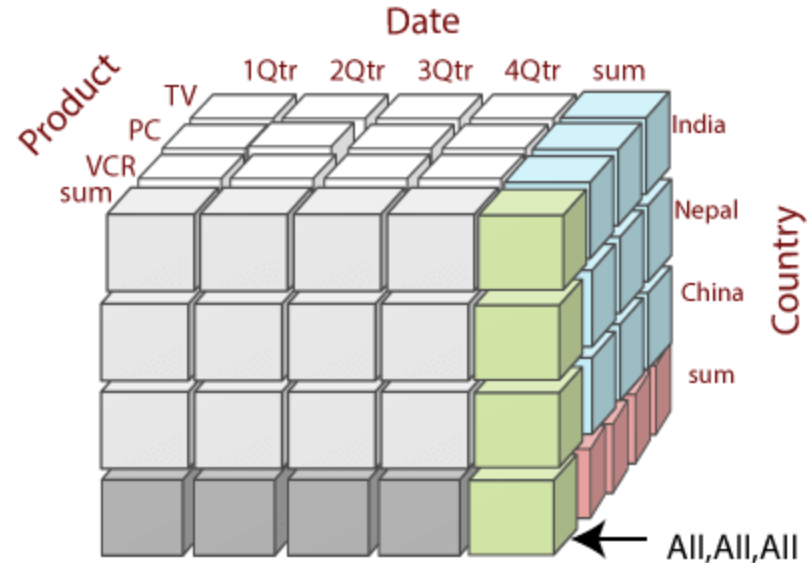
Modelo em cubo?

Table

Columns

Rows

Rows = records
Columns = attributes

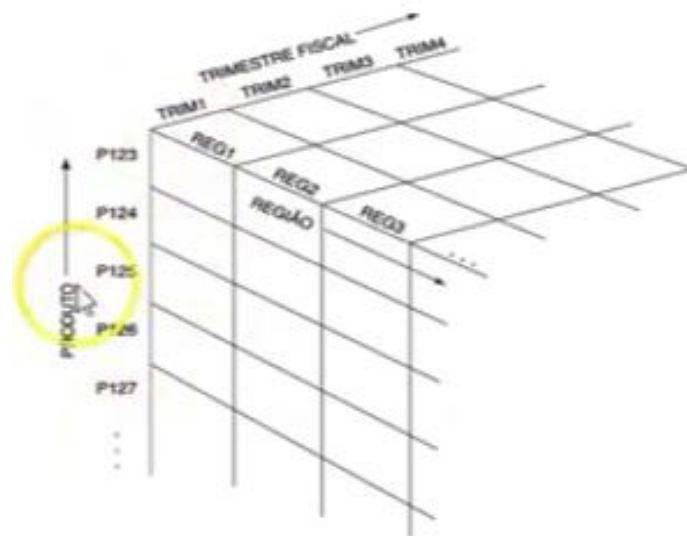


Fonte: <https://www.javatpoint.com/data-warehouse-what-is-data-cube>

Modelo em cubo?

Tabular - Relacional

	REGIÃO			
	REG1	REG2	REG3
PRODUTO	P123			
	P124			
	P125			
	P126			
	⋮			



Cubo - Dimensional

Formalizando o Cubo



Características

- Eixos: representa os components do esquema
- Interseção: Representam medidas e dados do context
- Visão consolidada do contexto
- Análises de perspectivas distintas

Qual é melhor?



Transacional

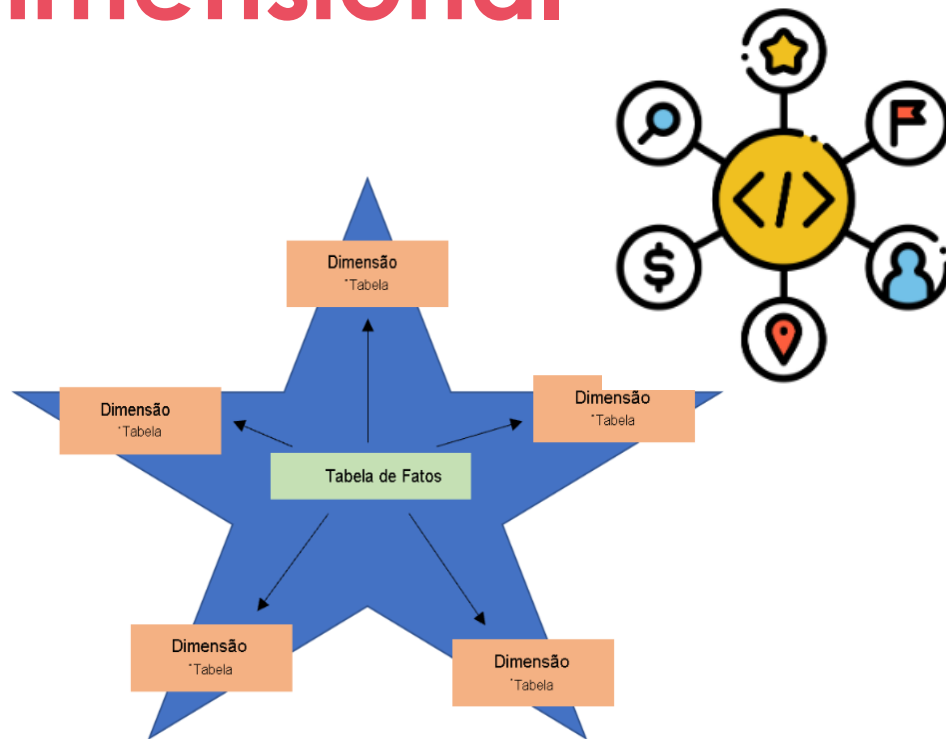
- Sistemas de vendas
- Alta disponibilidade
- Confiança na estrutura e restrições

Analítico

- Análise x Disponibilidade
- Consolidar informações
- Análises dos dados

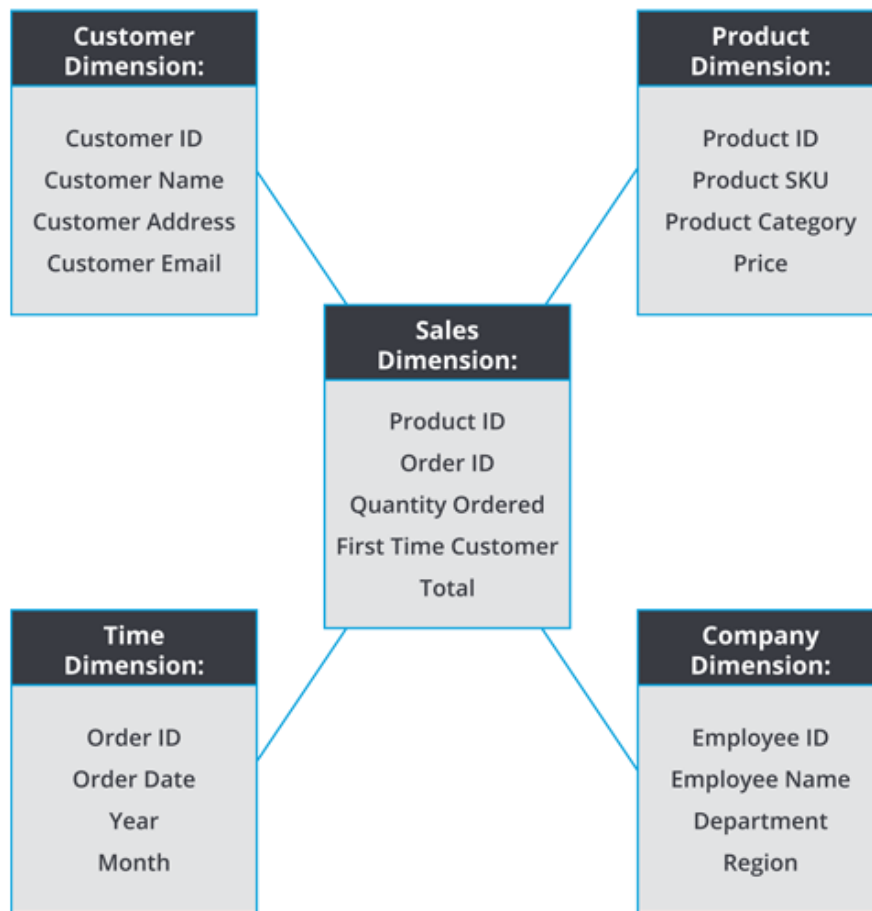
Coexistem – cada sistema possui sua particularidade

Modelagem Dimensional



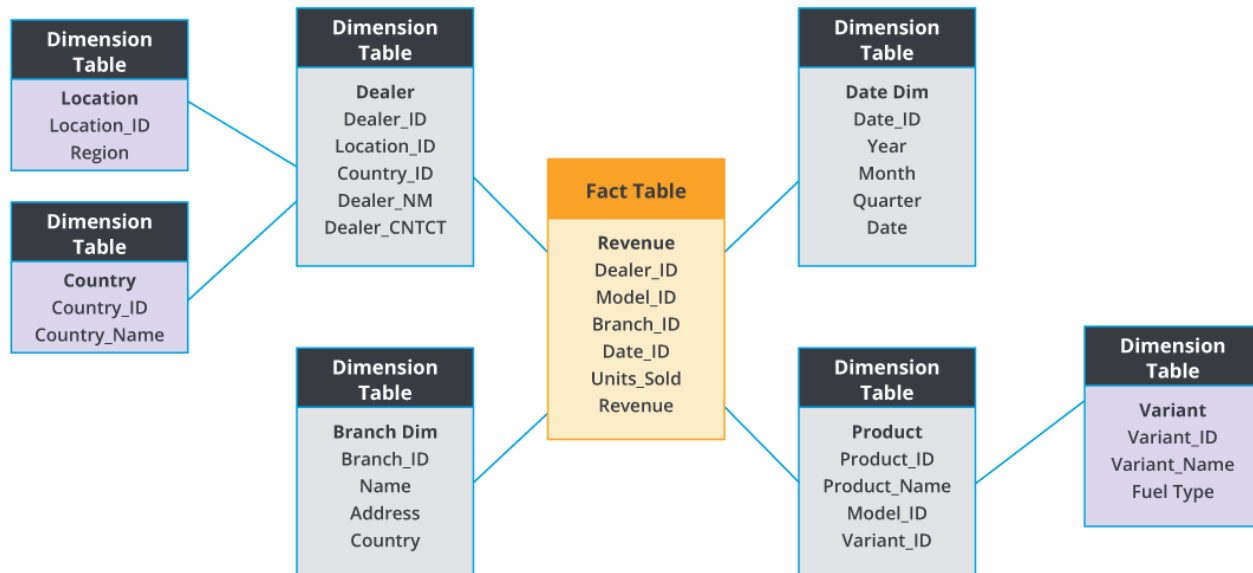
Modelos

Quais existem?



Modelos

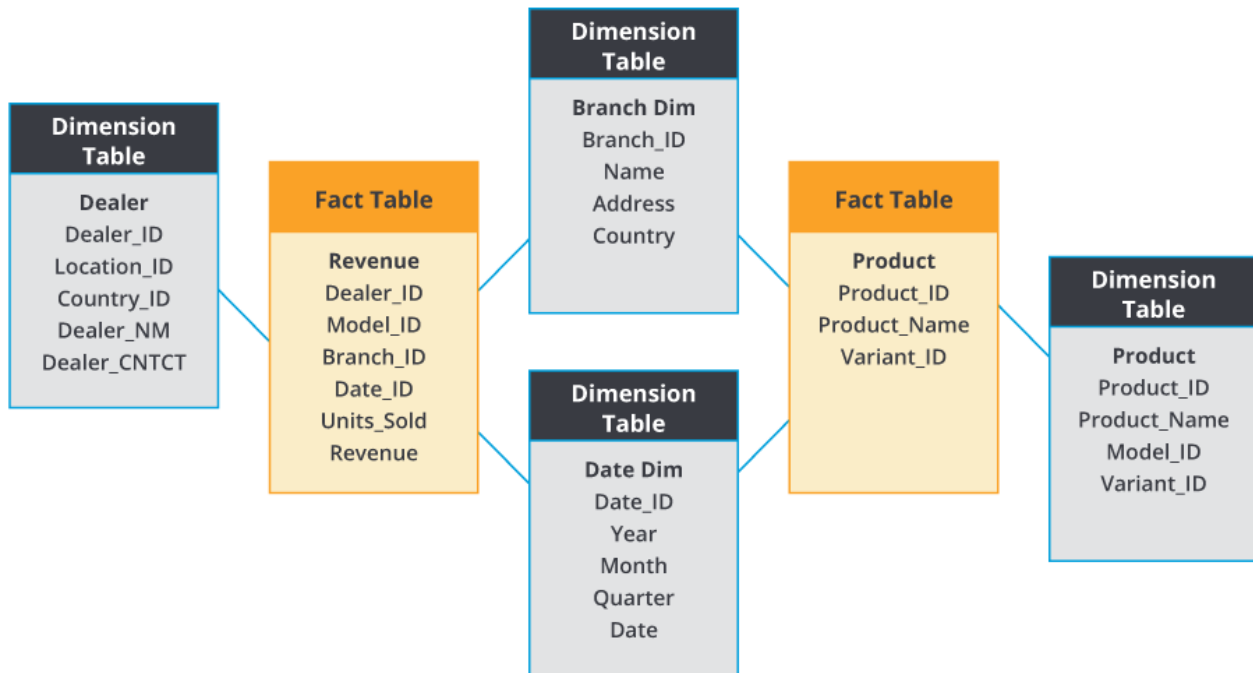
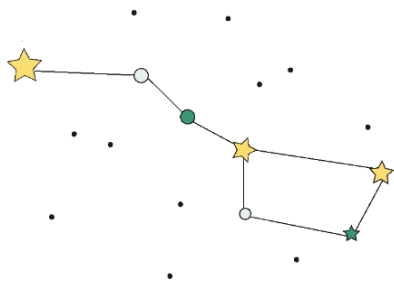
Quais existem?



Example of Snowflake Schema

Modelos

Quais existem?



Example of Galaxy Schema

Dúvida cruel...

Os banco de dados com modelos relacionais, permitem fazer análises e responder perguntas?

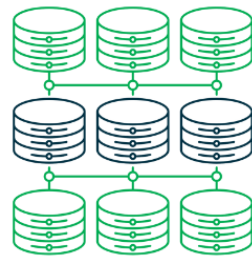
R: SIM!



OLTP



OLAP



Etapa 2

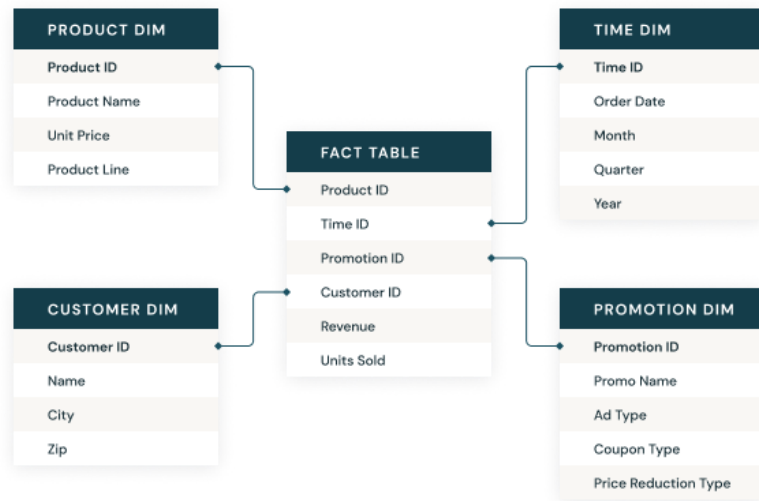
Explorando brevemente os Modelos de dados Dimensionais

// Power BI Analyst

Modelo Estrela

- Mais difundido dos modelos
- Star schema
- Conexão das tabelas em forma de estrela
- Dois tipos de tabelas: Fato e Dimensão

Star schema



Modelo Estrela

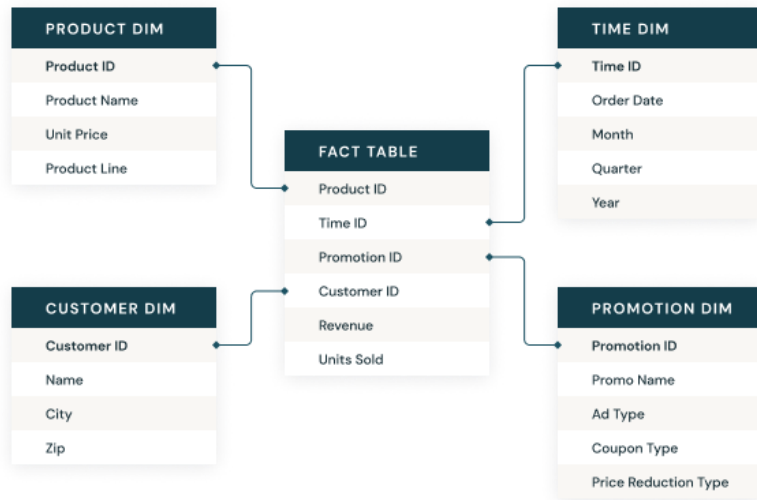
Se relaciona com as demais por meio de junções

- Tabela Fato – Principal
- Chave Artificial
- Tabela Dimensão - Detalhes

PKs simples e dados exclusivos

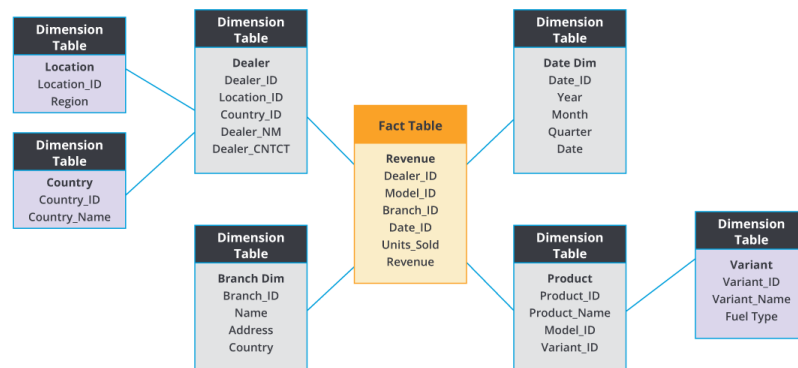
Mapeamento com desnormalização

Star schema



Modelo Snowflake

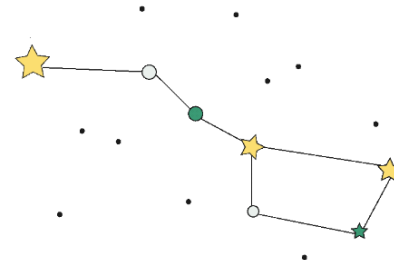
- Variação do modelo estrela
- Tabelas dimensão podem ter junções entre si
- Impondo um nível de normalização



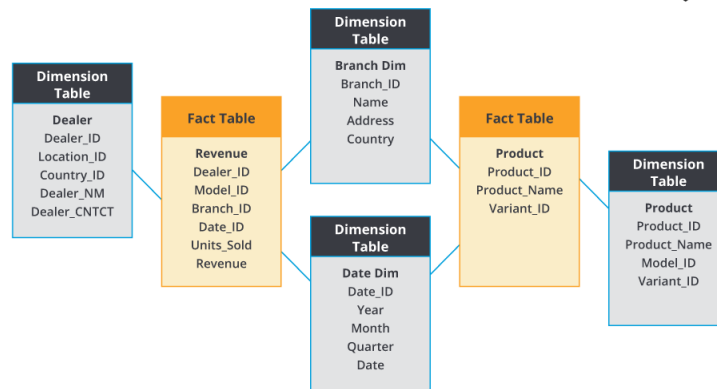
Example of Snowflake Schema

A normalização pode comprometer o desempenho do data warehouse

Constelação / Galáxia



- Múltiplas Estrelas são relacionadas entre si
- Tabelas fatos podem compartilhar tabelas dimensão
- Integra diversos assuntos em um mesmo contexto



Example of Galaxy Schema

Fonte: <https://streamsets.com/blog/schemas-data-warehouses-star-galaxy-snowflake/>

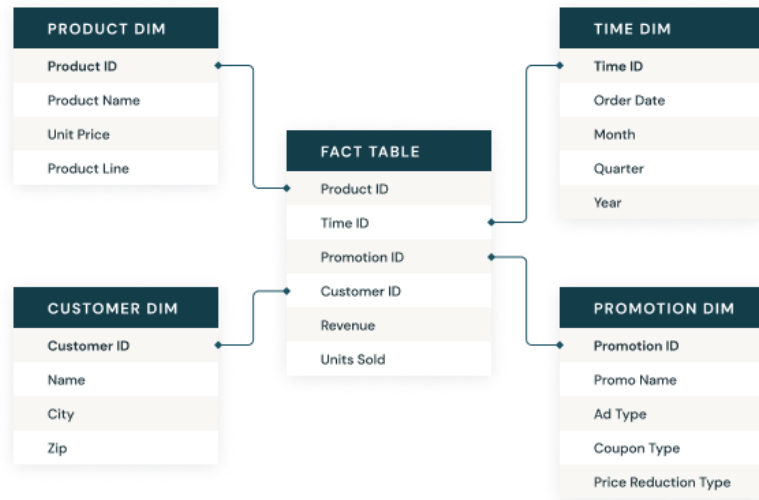
Tabelas Fato e Dimensão

Aspectos que compõem o contexto analisado

- Tabela Fato – Principal
- Tabela Dimensão - Detalhes

Detalhes de um aspecto específico

Star schema



Etapa 3

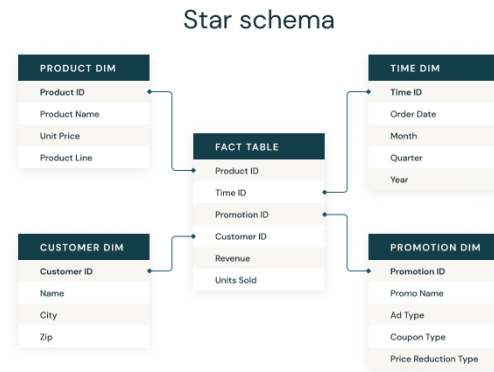
O que é e como Definir a Granularidade dos Dados?

// Power BI Analyst

Granularidade

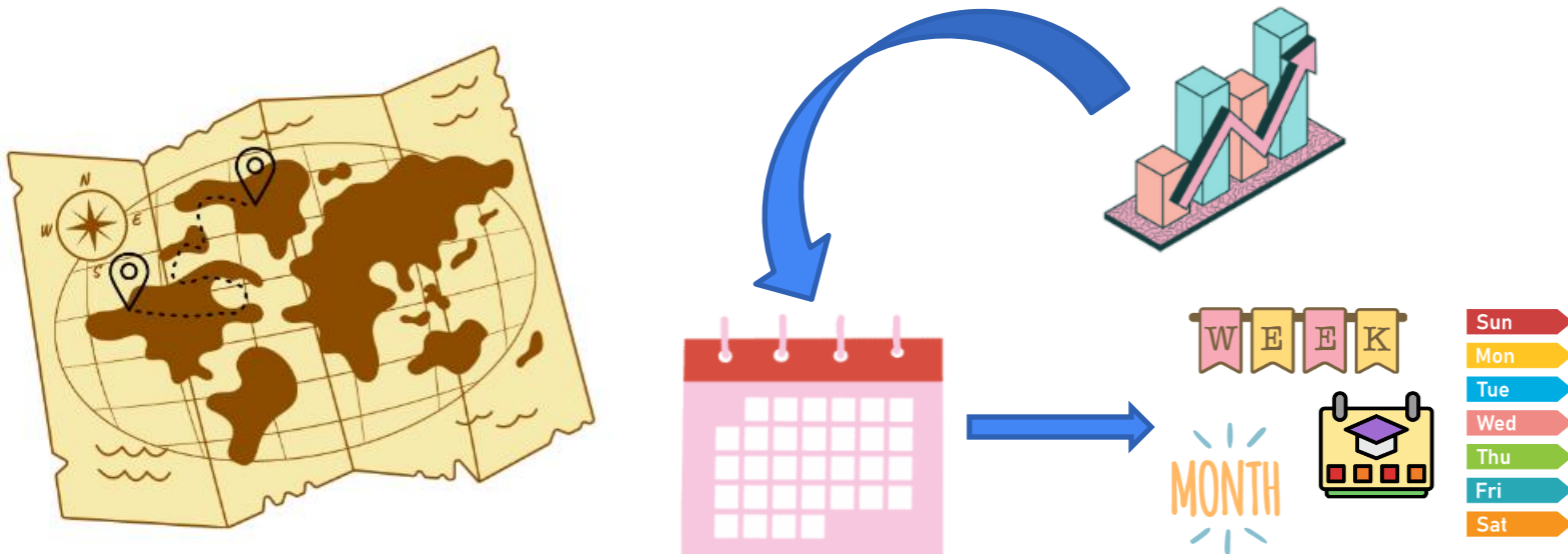
Como definir a granularidade?

- Grão ou granularidade
- Nível de detalhamento dos dados na tabela fato
- Maior o grão – menos detalhes
- Trade-off na definição



Granularidade - Exemplo

Como definir a granularidade?



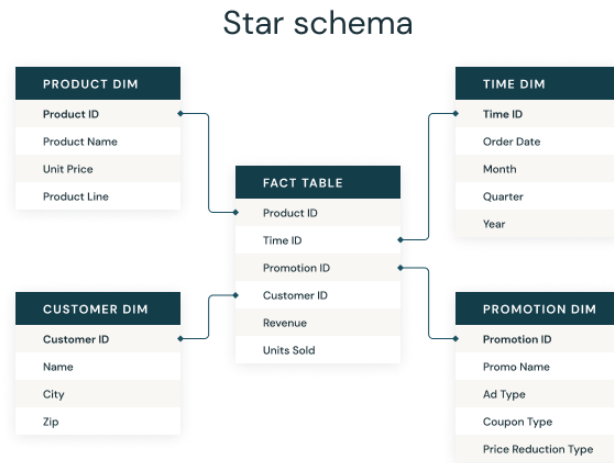
Granularidade

Ponto de Atenção!



Granularidade x processamento

Não adianta modelo perfeito sem poder de processamento



Etapa 4

Chave Artificial – Surgate Key

// Power BI Analyst

Chave Artificial

O que isso? Como uso esse trem?

- Chave relacionada ao modelo dimensional
- Substitui a identificação sem sobreposição
- Facilita a identificação de registros no modelo atual



Chave Artificial

O que isso? Como uso esse trem?

Identificador numérico

1	Código do Canal de Distribuição	Descrição
2	Z01-XPTO-QWERTY1231	CANAL DE DISTRIBUIÇÃO 001
3	X02-XPTO-QWERTY1232	CANAL DE DISTRIBUIÇÃO 002
4	Q03-XPTO-QWERTY1233	CANAL DE DISTRIBUIÇÃO 003
	W04-XPTO-QWERTY1234	CANAL DE DISTRIBUIÇÃO 004



Hands On! Construir esquemas de dados

*“Falar é fácil.
Mostre-me o código!”*

Linus Torvalds

Etapa 4

Modelo Dimensional – Slowly Changing Dimensions

// Power BI Analyst

Slowly Changing Dimensions

Relacionados as mudanças temporais dos dados

Tipo de dimensão

- SCD-0
- SCD-1, 2,
- SCD-3, 4, 5
- SCD-6 [1,2,3]

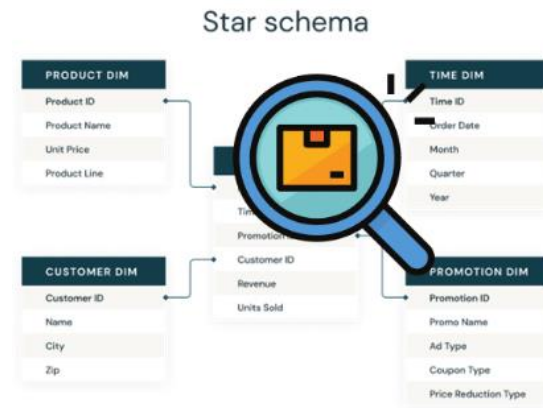


Slowly Changing Dimensions

Relacionados as mudanças temporais dos dados

Tipo SCD-0

- Não há modificação
- Modo passive
- TRUNCATE TABLE – sem histórico

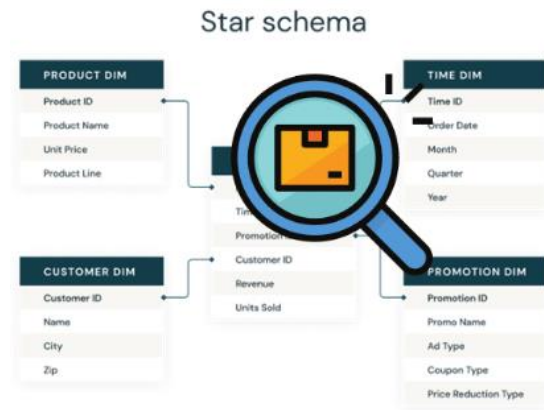


Slowly Changing Dimensions

Relacionados as mudanças temporais dos dados

Tipo SCD-1

- Atualização dos valores
- Sem rastreamento de mudanças
- UPDATE ou INSERT




Slowly Changing Dimensions

CustomerKey	CustomerCode	MaritalStatus	Gender	Designation	StartDate	EndDate	IsCurrent
11011	AW00011011	M	M	Professional	2012-01-01	9999-12-31	Yes
11012	AW00011012	M	F	Management	2012-01-01	9999-12-31	Yes

Fonte: <https://www.sqlshack.com/implementing-slowly-changing-dimensions-scds-in-data-warehouses/>

Tipo SCD-2

- Preocupação com histórico
- Modos distintos de rastrear as mudanças




CustomerKey	CustomerCode	MaritalStatus	Gender	Designation	StartDate	EndDate	IsCurrent
11011	AW00011011	M	M	Professional	2012-01-01	9999-12-31	Yes
11012	AW00011012	M	F	Management	2012-01-01	2021-06-01	No
11013	AW00011012	M	F	<u>Snr. Management</u>	2021-06-01	9999-12-31	Yes

Slowly Changing Dimensions

Tipo SCD-3

- Novos atributos (colunas) são criados
- Manter o estado de um atributo específico

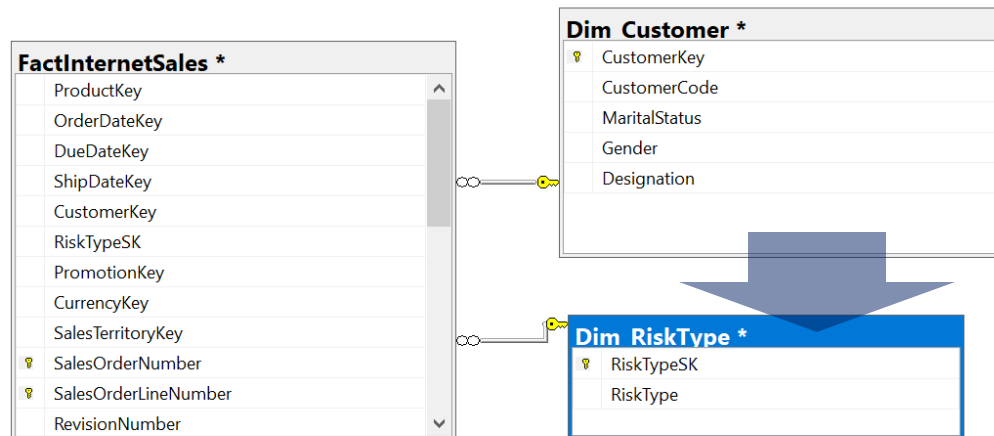


CustomerKey	CustomerCode	MaritalStatus	Gender	Designation	PreviousDesignation
11011	AW00011011	M	M	Professional	NULL
11012	AW00011012	M	F	Snr. Management	Management

Slowly Changing Dimensions

Tipo SCD-4

- Manutenção do histórico com Tabela de histórico
- Mesma estrutura



Fonte: <https://www.sqlshack.com/implementing-slowly-changing-dimensions-scds-in-data-warehouses/>

Slowly Changing Dimensions

Tipo SCD-6

- Junção dos tipos 1, 2 e 3 = 6 (somatório)



CustomerKey	CustomerAlternateKey	Gender	MaritalStatus	Occupation	CurrentOccupation	CurrentDate	EndDate	IsCurrent
11000	AW00011000	M	M	Professional	Snr. Management	2019-01-01	2020-06-15	No
11001	AW00011001	M	S	Professional	Professional	2021-01-01	9999-12-31	Yes
11002	AW00011000	M	M	Management	Snr. Management	2020-06-16	2021-01-01	No
11003	AW00011000	M	M	Snr. Management	Snr. Management	2021-01-01	9999-12-31	Yes

Hands On! **Aplicar Slowly Changing Dimensions**

*“Falar é fácil.
Mostre-me o código!”*

Linus Torvalds

Links úteis

- <https://streamsets.com/blog/schemas-data-warehouses-star-galaxy-snowflake/>
- <https://www.databricks.com/glossary/star-schema>
- <https://www.softwaretestinghelp.com/data-warehouse-modeling-star-schema-snowflake-schema/>
- <https://www.sqlshack.com/implementing-slowly-changing-dimensions-scads-in-data-warehouses/>
- <https://www.javatpoint.com/data-warehouse-what-is-data-cube>