

MBA em Inteligência Artificial

Big Data e Visualização de Dados

Prof. Diego Nogare



Faculdade de Computação e Informática
Mackenzie

Diego Nogare



Former MVP Artificial Intelligence
Microsoft Regional Director
Diretor no PASS.org
Chief Data Officer @ Lambda3
Membro notável na I2AI
Mestre em IA

www.diegonogare.net
www.livrosdonogare.com.br





DADOS ESTRUTURADOS E NÃO ESTRUTURADOS



■ Categoria do T-SQL

Baseado no padrão ANSI – American National Standards Institute – criado na década de 70

DML – Data Manipulation Language

TCL – Transaction Control Language

DDL – Data Definition Language

DCL – Data Control Language

■ Formato de Armazenamento

Normalizado: Se utiliza de relacionamentos das tabelas para evitar dados redundantes – Ambientes OLTP

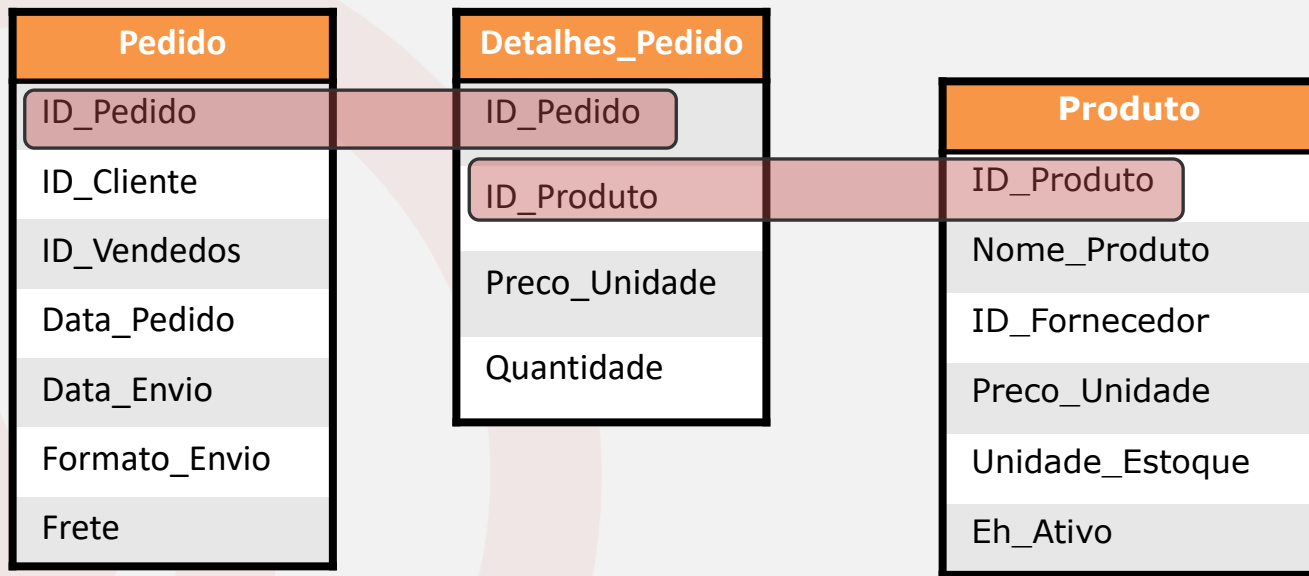
Desnormalizado: Não faz relacionamento das tabelas para evitar processamento na hora da consulta, mantendo os dados redundantes – Ambientes OLAP

■ OLTP vs. OLAP

OLTP (On-line Transaction Processing) são sistemas que se encarregam de registrar todas as transações contidas em uma determinada operação organizacional. Por exemplo: sistema de transações bancárias que registra todas as operações efetuadas em um banco, caixas de multibanco, reservas de viagens ou hotel on-line, Cartões de Crédito.

OLAP (On-line Analytical Processing) é a capacidade para manipular e analisar um grande volume de dados sob múltiplas perspectivas. As aplicações OLAP são usadas pelos gestores em qualquer nível da organização para lhes permitir análises comparativas que facilitem a sua tomada de decisões diárias

Banco de Dados Relacional



OLTP

Crie uma nova conta...

Endereço de e-mail

Senha

Confirme a senha

Data de nascimento

Mês ▼ Dia ▼ Ano ▼ ?

science SQAGTO

Digite as duas palavras:

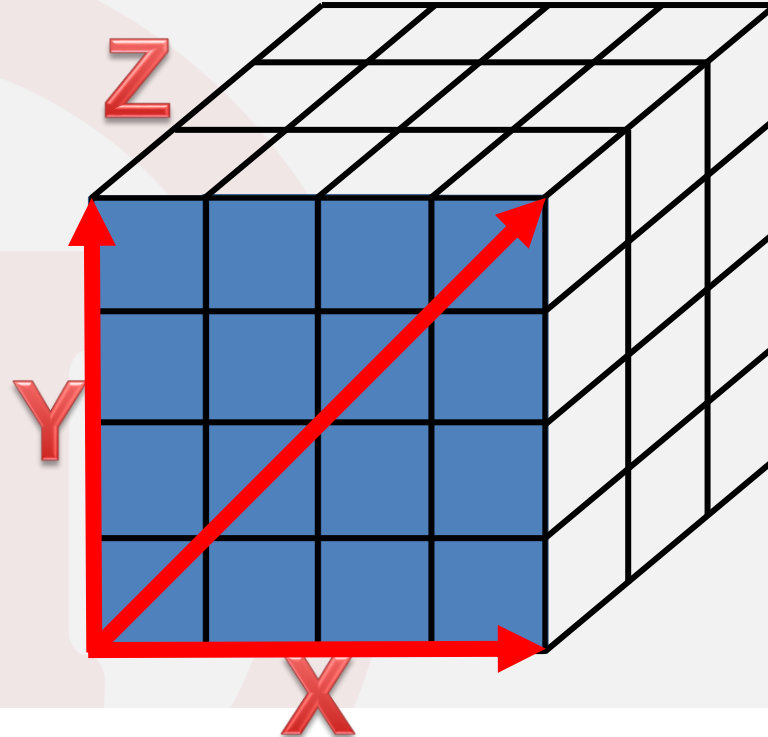
reCAPTCHA™
stop spam.
read books.

Registre-se

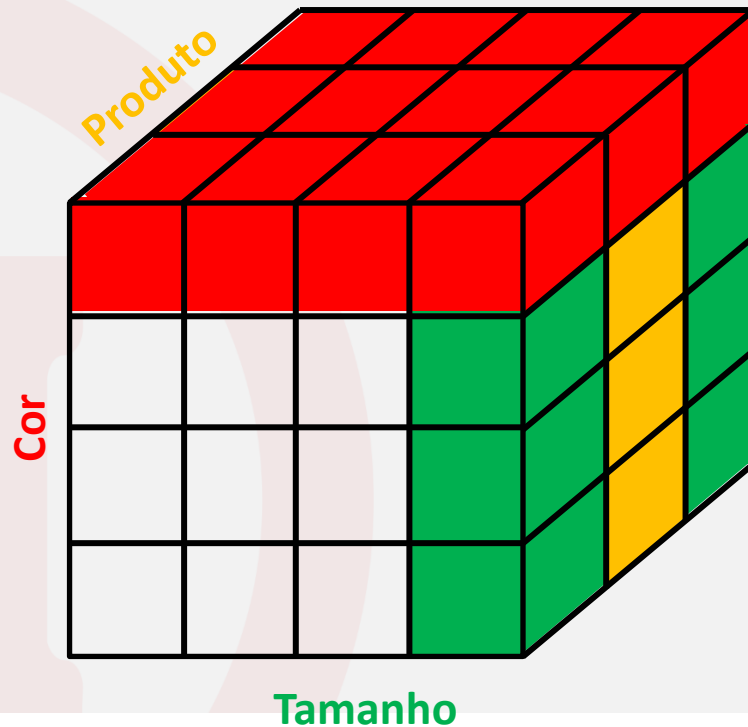
1.000.000



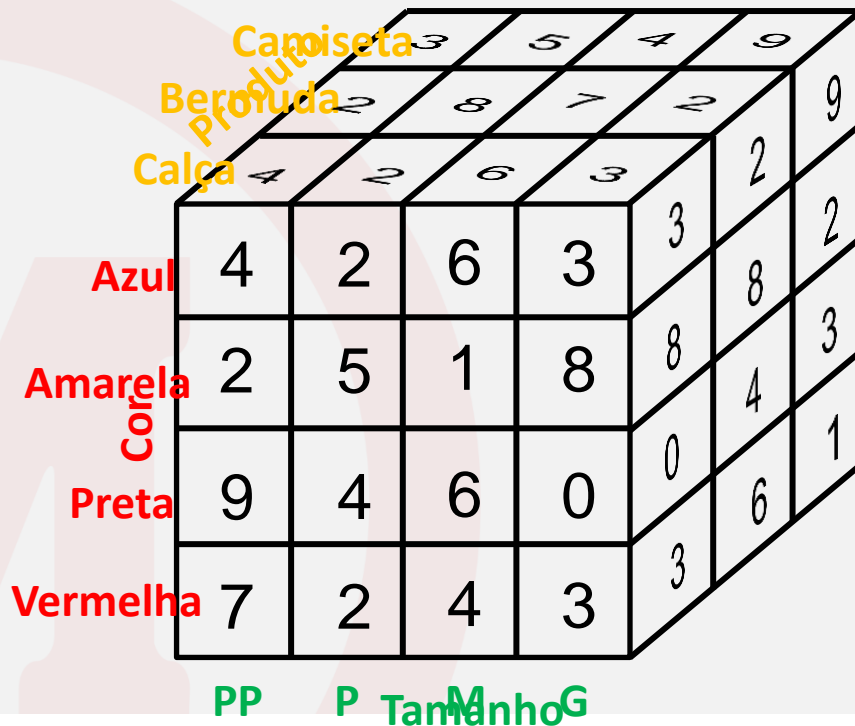
OLAP



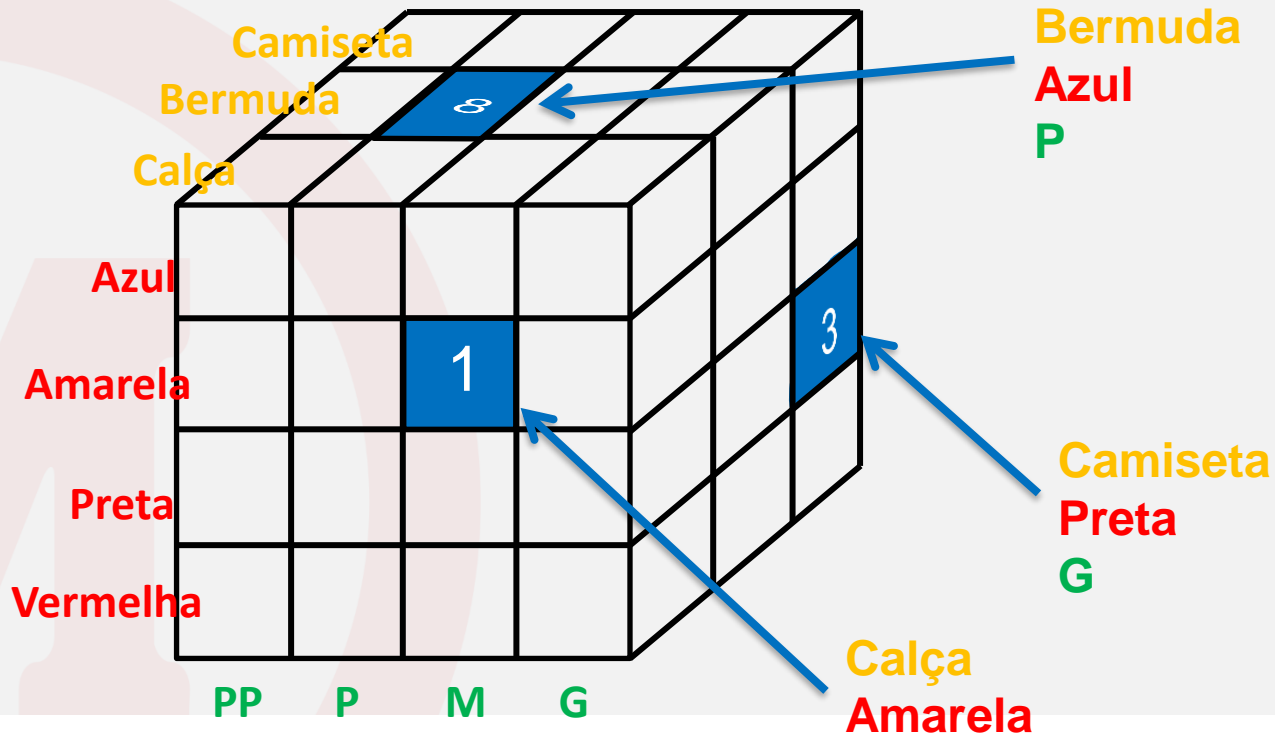
OLAP



OLAP



OLAP



Azul = 55

3	5	4	9
2	8	7	2
4	2	6	3
4	2	6	3
2	5	1	8
9	4	6	0
7	2	4	3
2	8	4	6
9	2	3	1
3	0	6	1

Calça = 66

4	2	6	3
2	5	1	8
9	4	6	0
7	2	4	3

Modelagem multidimensional

Focada em responder questões de negócios com base em medidas que são agregadas por dimensões

Medidas são fatos calculados sobre dimensões de negócios

Dimensões são formas das quais as medidas são agregadas:

Exemplo:

Total de faturamento por equipe de vendas
Lucro de uma linha de produto
Quantidade de pedidos por produto
Custo do produto
Total de faturamento por cliente
Lucro por região
Total de faturamento por cidade

- Agrupa dimensões relacionadas, dentro de tabelas dimensões
- Agrupa as medidas dentro de tabela fato
- Relaciona as tabelas dimensões com as tabela fato

Star Schema

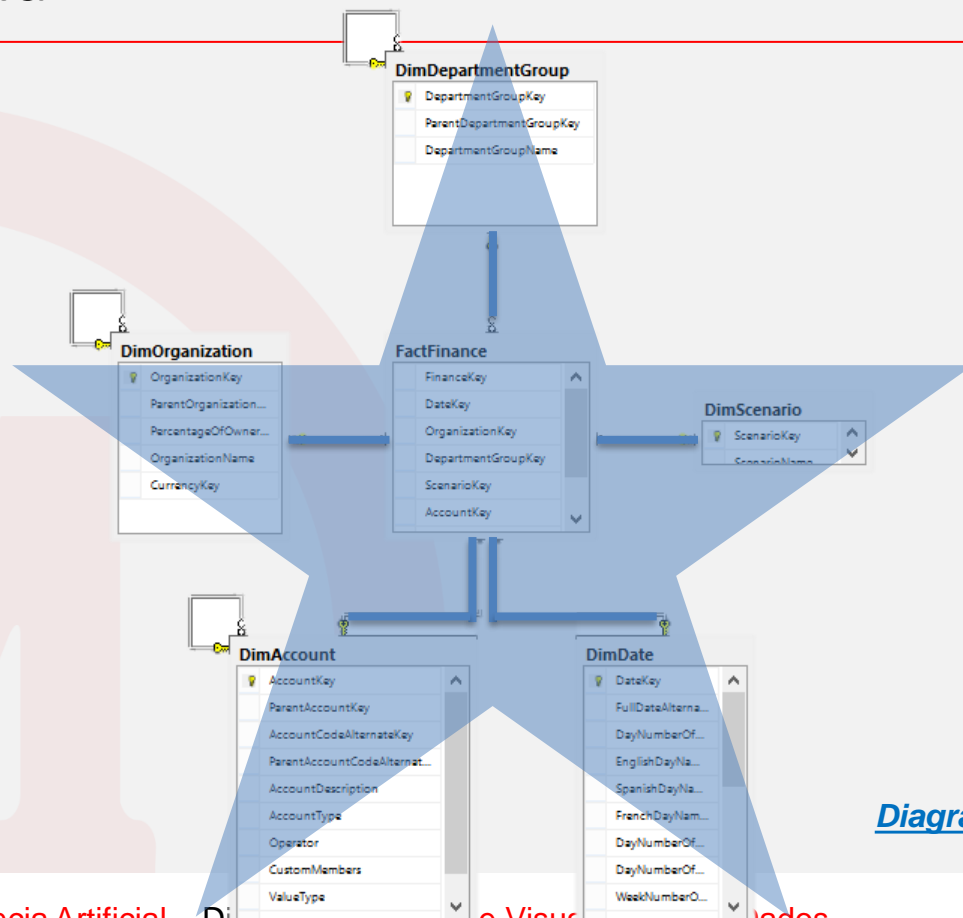


Diagrama: Finance

Snowflake Schema

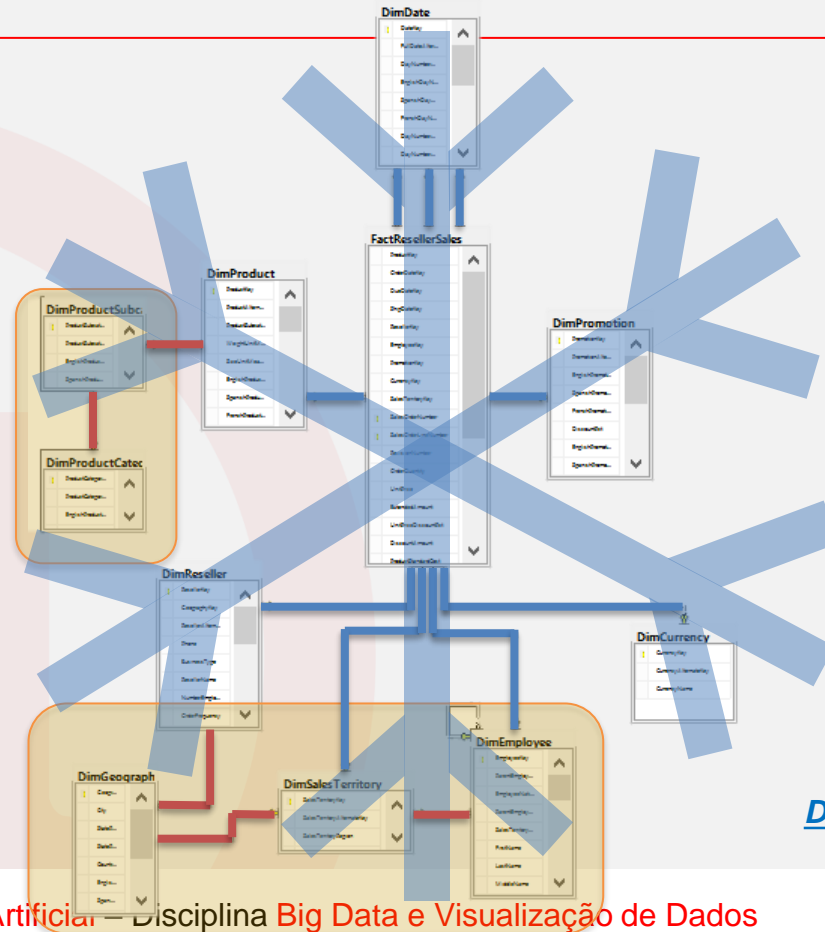
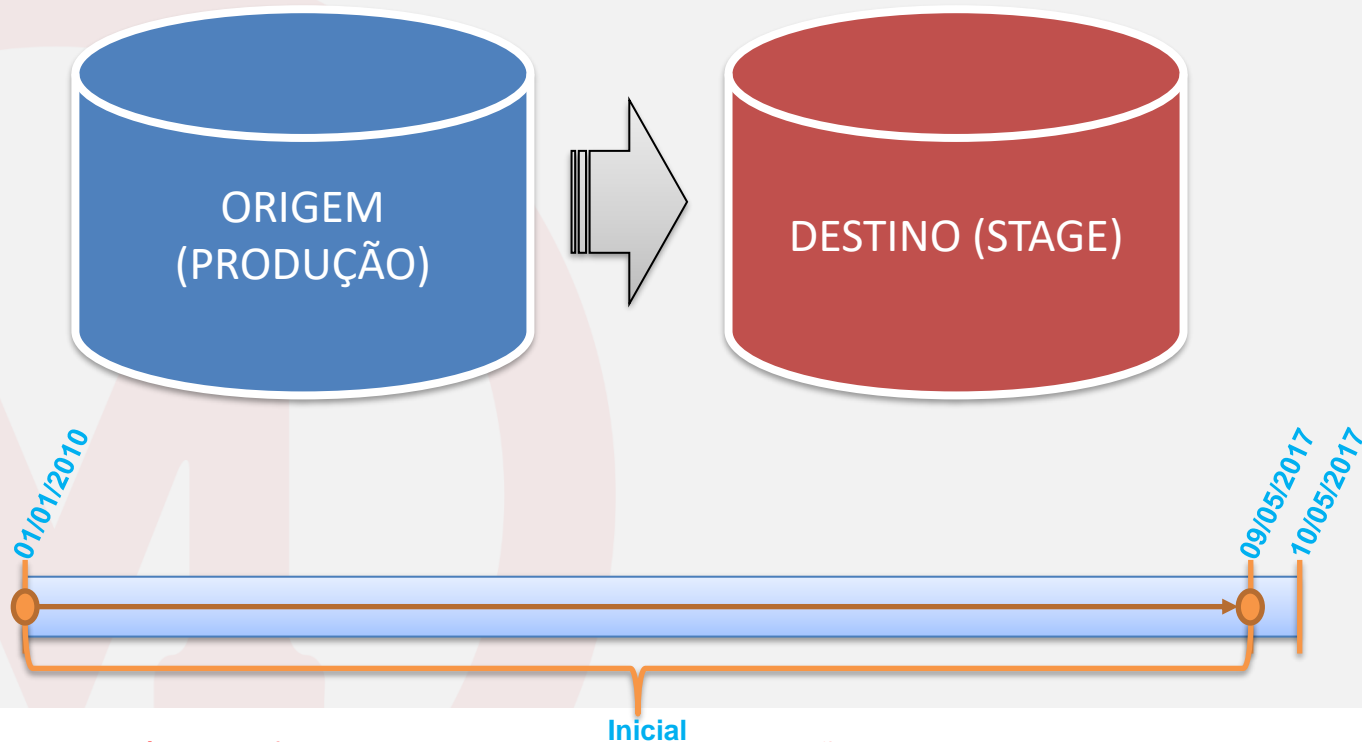
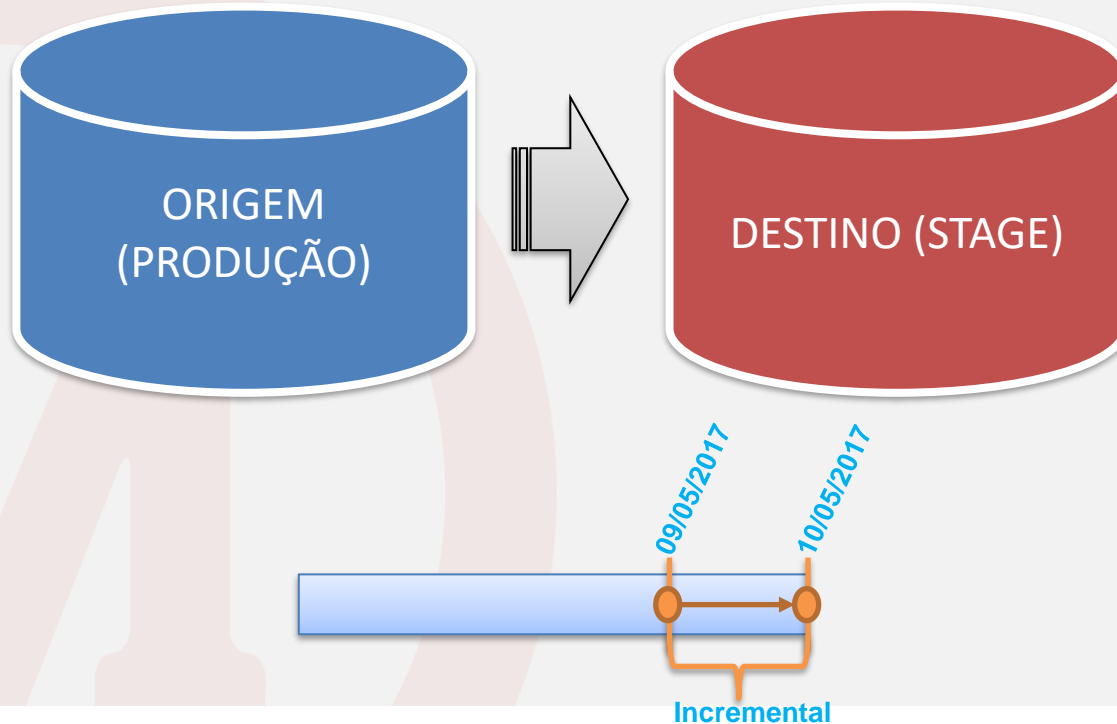


Diagrama: Reseller Sales

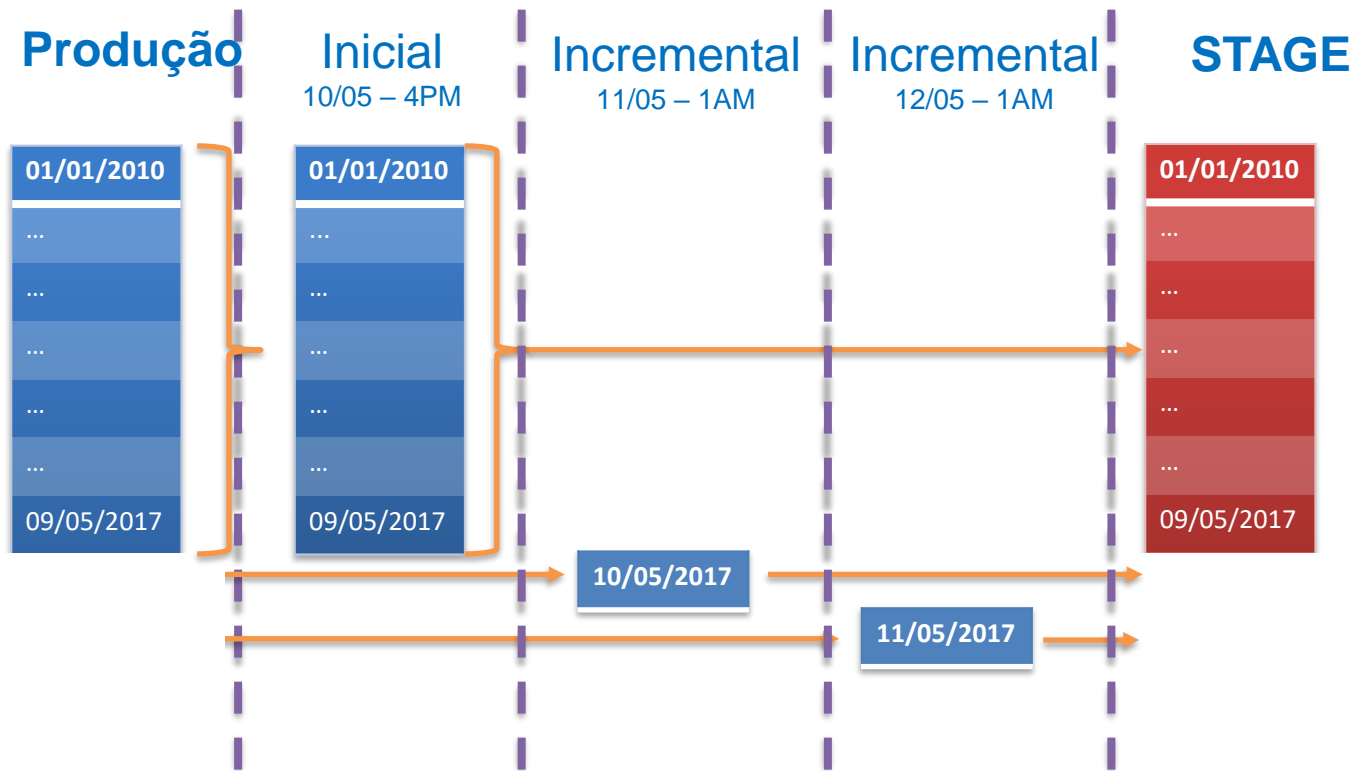
■ Primeira carga



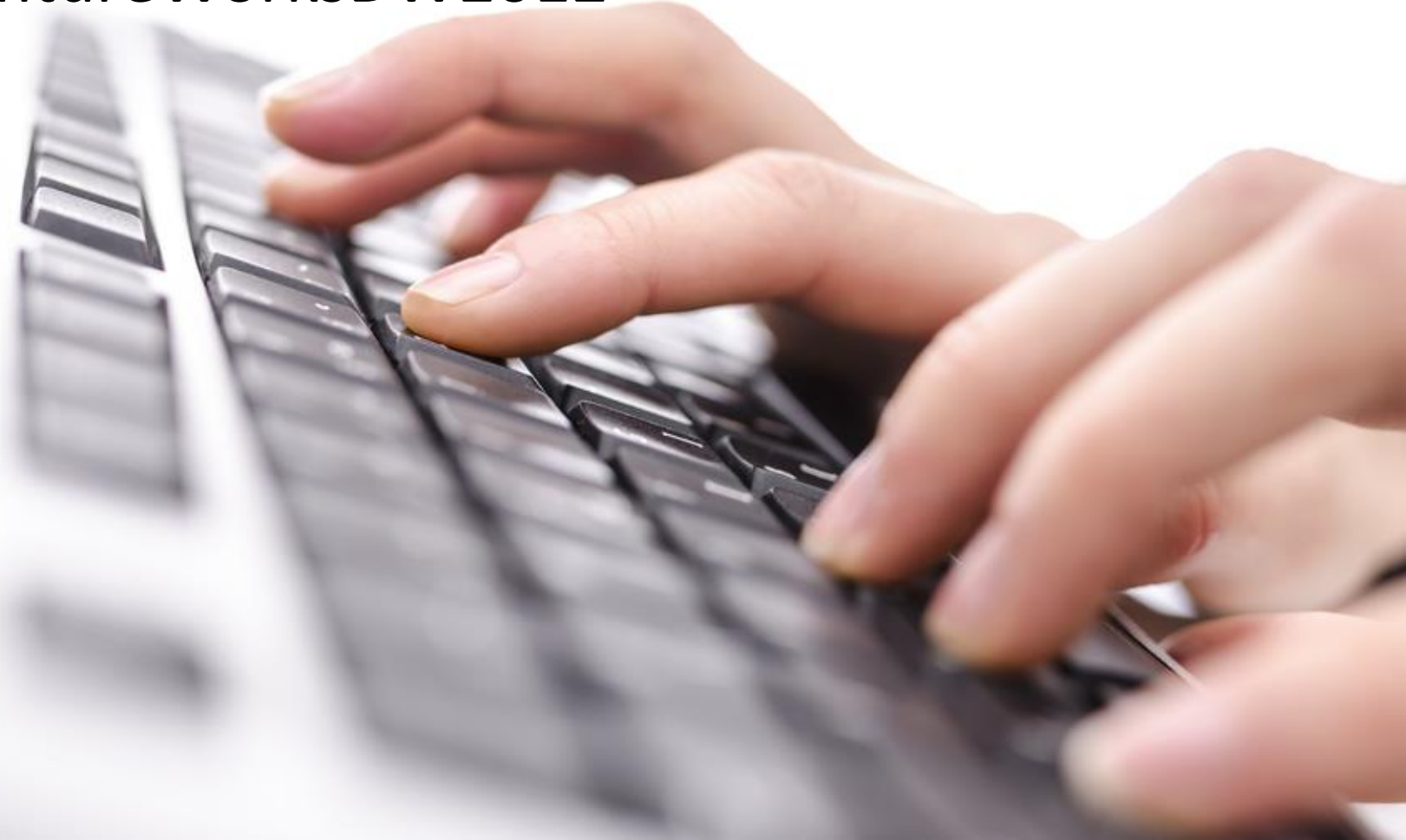
Carga Incremental



Timeline de atualização



AdventureWorksDW2012



Measures

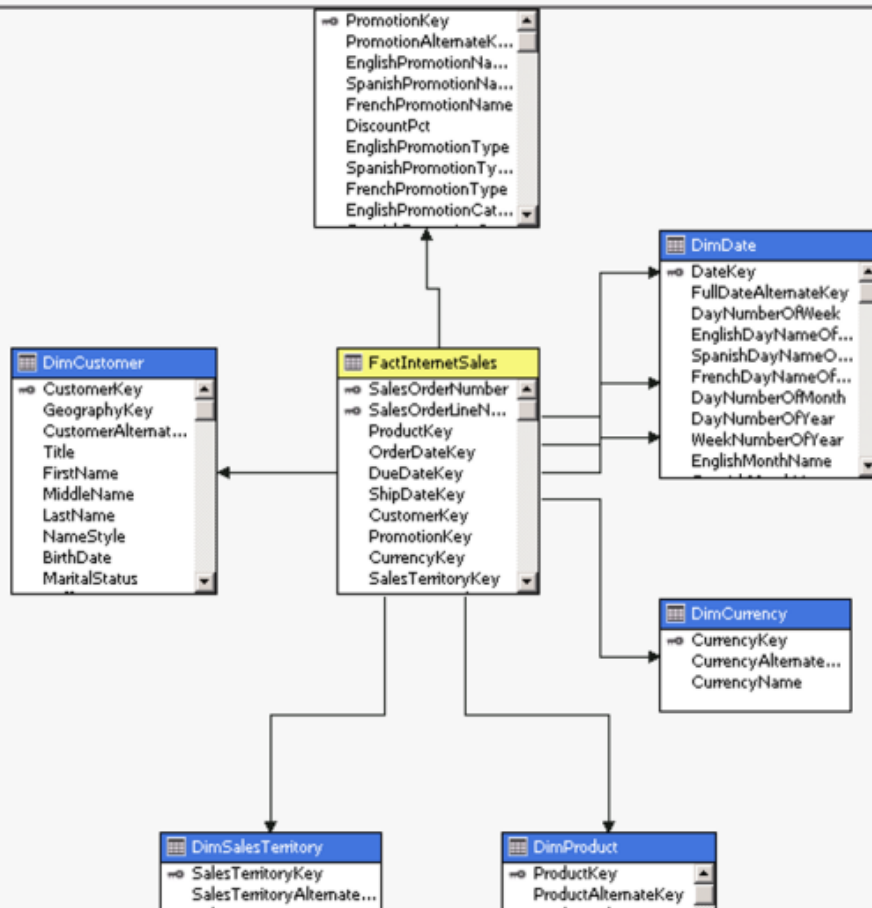
Adventure Works DW2012

Fact Internet Sales

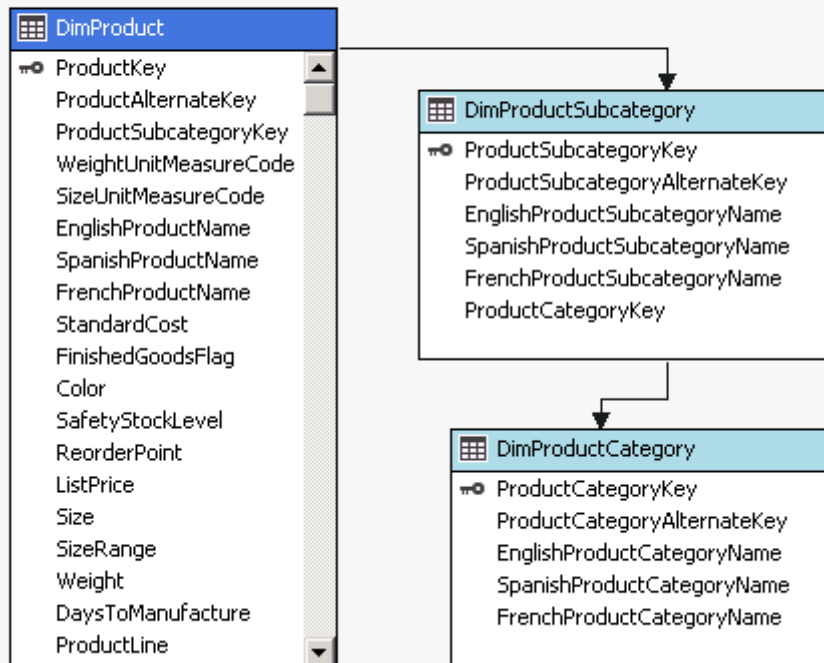
Dimensions

Adventure Works DW2012

- Dim Promotion
- Order Date
- Dim Sales Territory
- Dim Customer
- Dim Currency
- Dim Product
- Ship Date
- Due Date



Data Source View





Quanto (em dinheiro) cada categoria de produto vendeu?

Apresente estes números por país ordenado por Ano

Referências



Nogare, D. **Do Banco de Dados Relacional à Tomada de Decisão**, (2014), MVTech





Faculdade de Computação e Informática
Mackenzie