



A aula interativa do Módulo 2 - Bootcamp Engenheiro de Dados começará em breve!

Atenção:

- 1) Você entrará na aula com o microfone e o vídeo DESABILITADOS.**
- 2) Apenas a nossa equipe poderá habilitar seu microfone e seu vídeo em momentos de interatividade, indicados pelo professor.**
- 3) Utilize o recurso Q&A para dúvidas técnicas. Nossos tutores e monitores estarão prontos para te responder e as perguntas não se perderão no chat.**
- 4) Para garantir a pontuação da aula, no momento em que o professor sinalizar, você deverá ir até o ambiente de aprendizagem e responder a enquete de presença. Não é necessário encerrar a reunião do Zoom, apenas minimize a janela.**

Armazenamento de Dados

Segunda Aula Interativa

Prof. Ricardo Brito Alves

Nesta aula



- ☐ Tópicos da Disciplina e Temas Interessantes.
- ☐ Desafio.
- ☐ Ensaio Pentaho.

Tópicos da Disciplina e Temas Interessantes

Sistema de Arquivos



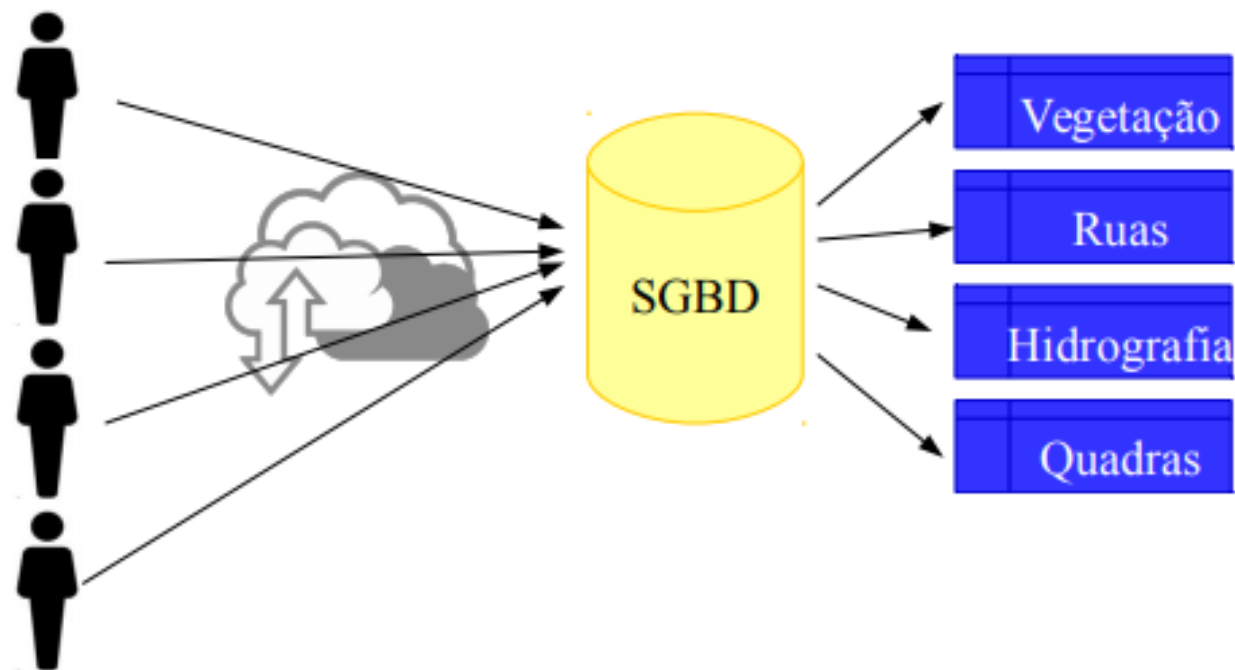
Banco de Dados



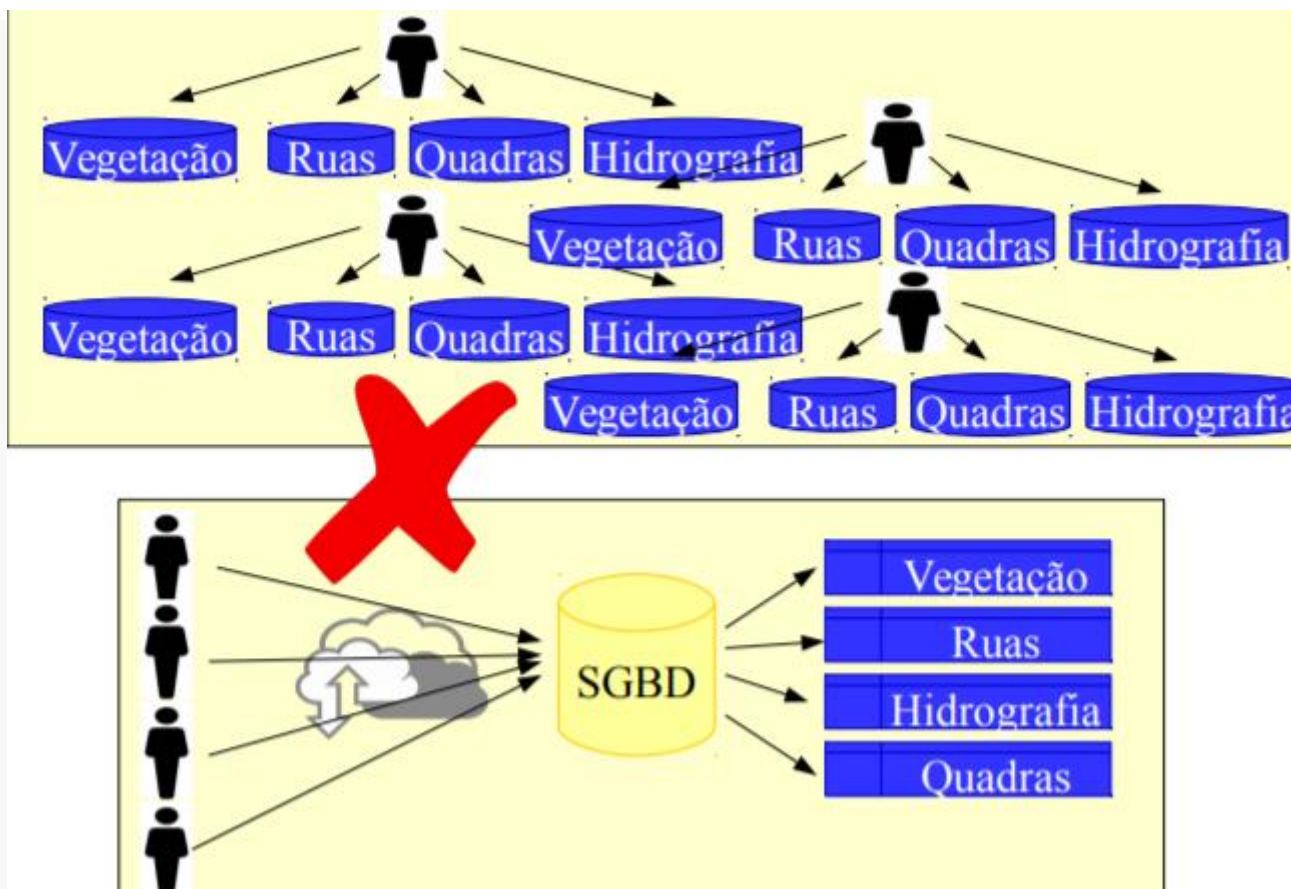
- [Chu, 1985] – Um banco de dados é um conjunto de arquivos relacionados entre si.
- [Date, 2000] – Um banco de dados é uma coleção de dados operacionais armazenados usados pelas aplicações de uma determinada organização.

Sistema Gerenciador de Banco de Dados

Um sistema de gerência de banco de dados (SGBD) é um conjunto de programas que permite a criar e manter um banco de dados.



Sistema Gerenciador de Banco de Dados



Alguns SGBDs

IGTI



SGBD Relacional



PROPRIETARIO

CPF	NOME	RUA	NUMERO	BAIRRO
08940256	JOÃO DA SILVA	SAO JOAO	180	CENTRO
03727298	HENRIQUE CARDOSO	IMIGRANTE	1700	VILA 12
97260089	JOSÉ DE SOUZA	SAO JOAO	35	CENTRO

LOTE

NUMERO	PROPRIETARIO_CPF	AREA_TOTAL	AREA_CONST
00001	08940256	400.000	0
00003	03727298	150.000	75.00
00039	03727298	500.000	0

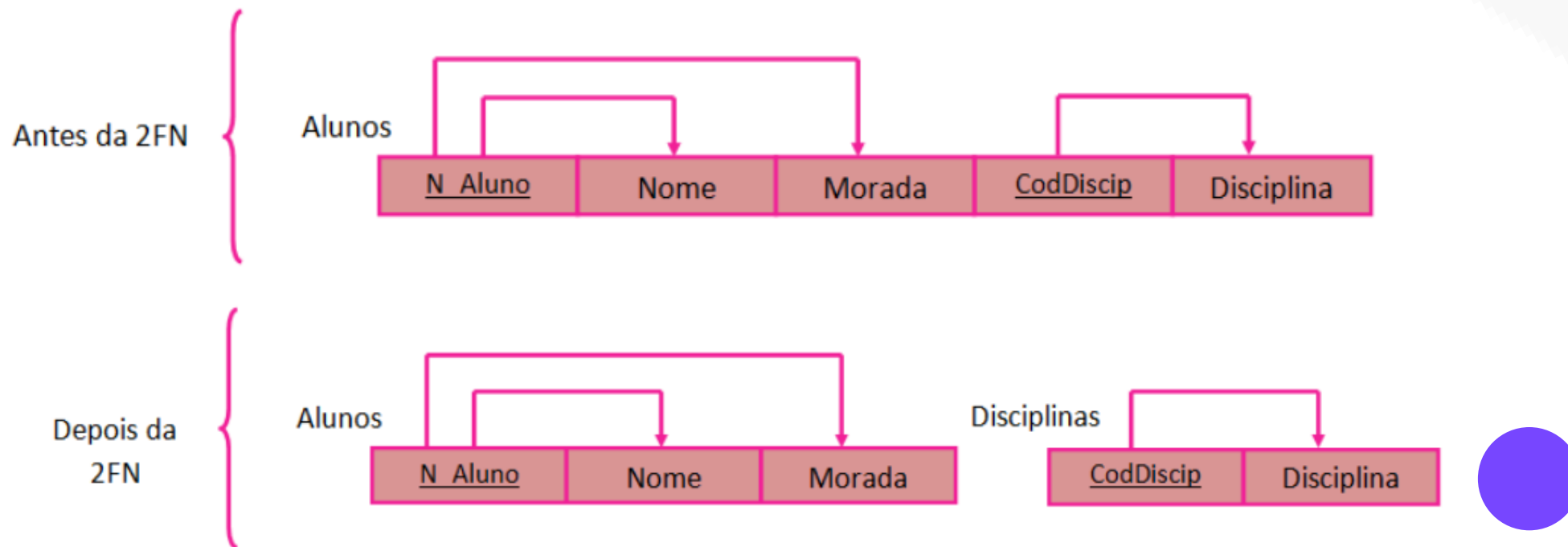
Normalização



CodCliente	Cliente	Morada	Encomendas			
			N_Enc	Data	Produto	Quantidade
C01	Aníbal	Lisboa	1	2010-05-25	Ananás	10
			5	2010-05-30	Cebolas	20
C02	Belmiro	Braga	3	2010-05-26	Bananas	30

CodCliente	Cliente	Morada	N_Enc	Data	Produto	Quantidade
C01	Aníbal	Lisboa	1	2010-05-25	Ananás	10
C01	Aníbal	Lisboa	5	2010-05-30	Cebolas	20
C02	Belmiro	Braga	3	2010-05-26	Bananas	30
C03	Casimiro	Coimbra	2	2010-05-25	Tomates	50
C03	Casimiro	Coimbra	4	2010-05-26	Cebolas	20

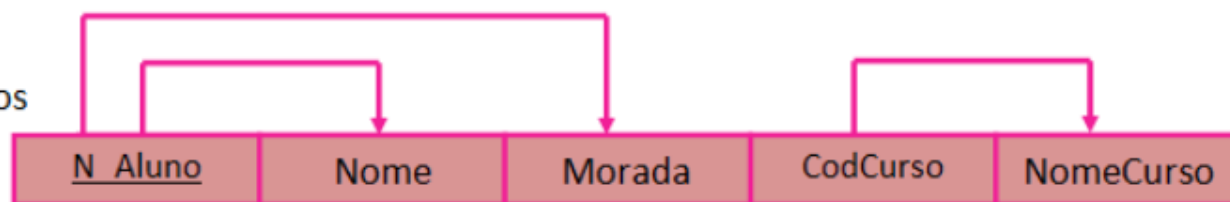
Normalização



Normalização

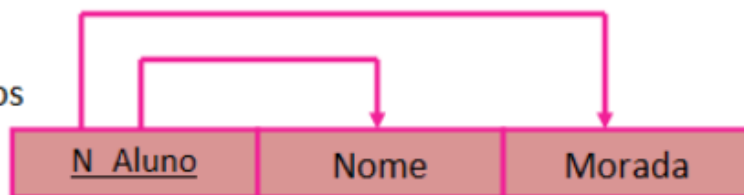
Antes da 3FN

Alunos

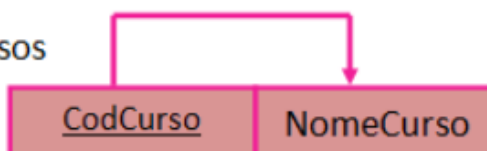


Depois da 3FN

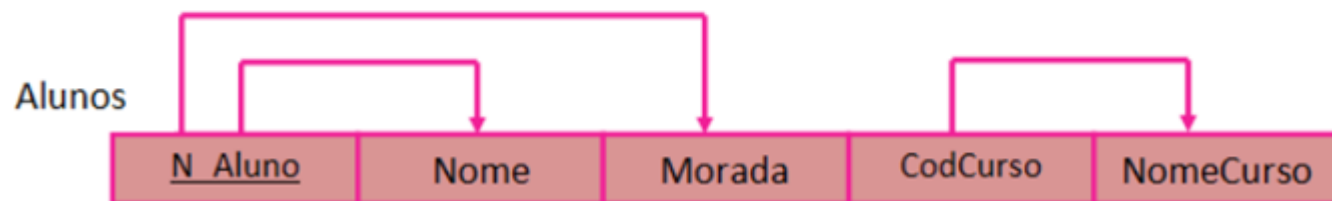
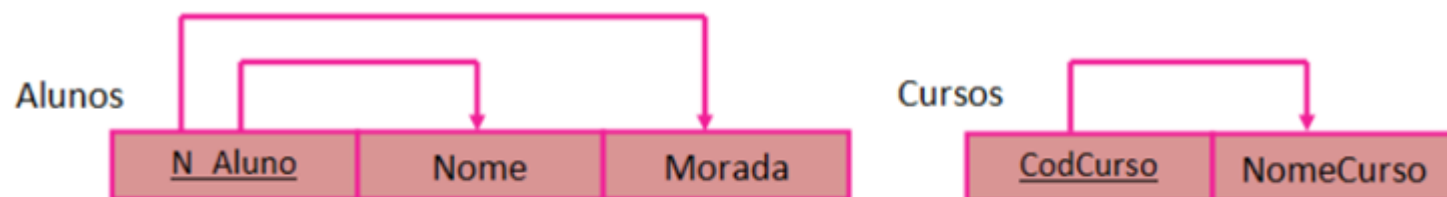
Alunos



Cursos

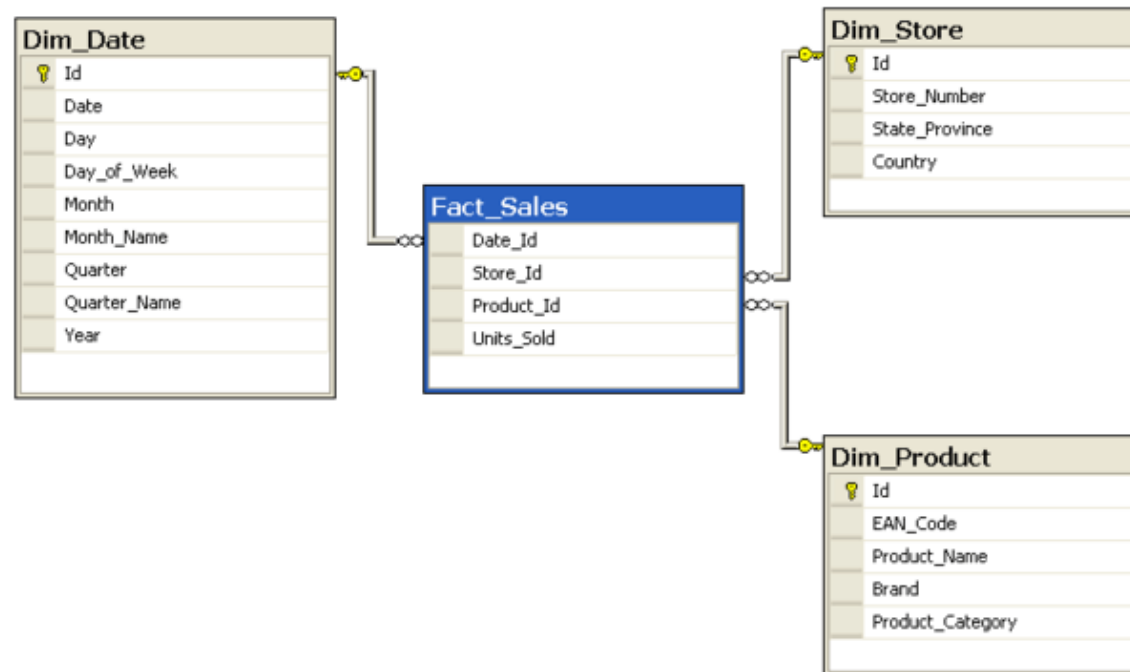


Desnormalização



Modelo Star Schema

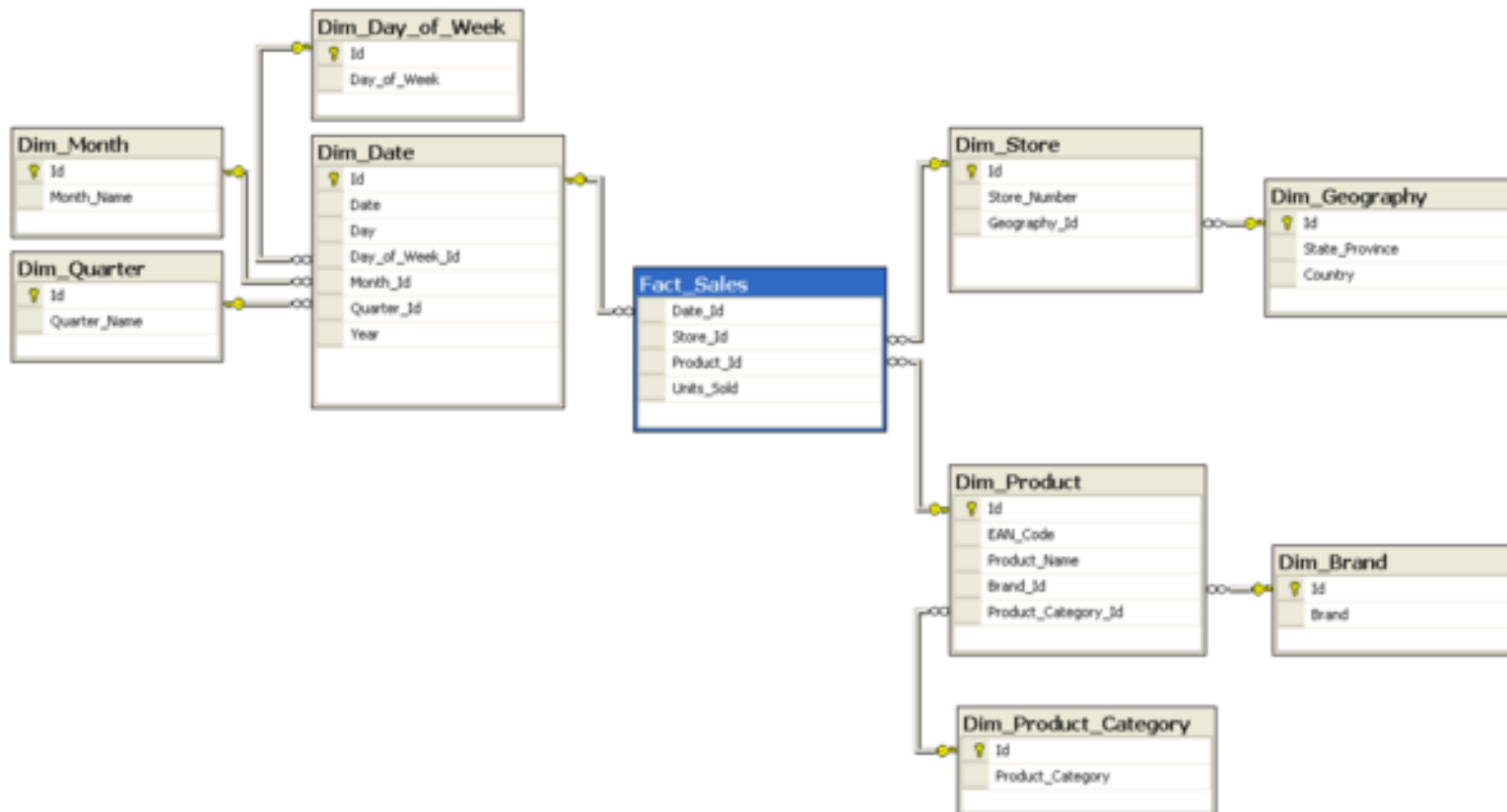
- ✓ Considerando um banco de dados de lojas, produtos e um data warehouse para executar relatórios de vendas agrupados por loja, data, categoria ou marca do produto. Se esse data mart estiver usando um esquema em estrela, teria a seguinte aparência:



Modelo SnowFlake



- ✓ O mesmo cenário no esquema SnowFlake seria estruturado da seguinte maneira:



Considerações sobre os Modelos Dimensionais

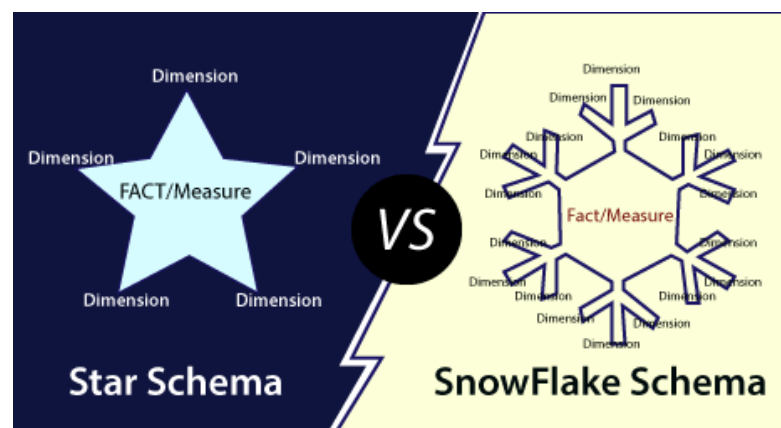


✓ Modelos Star Schema *(mais usado)*

- Dimensões Desnormalizadas.
- Voltado para acessos com performance.
- Hierarquias achatadas.
- Mais simples e mais fácil navegação.
- Utiliza mais espaço repetindo as mesmas descrições ao longo de toda a tabela.

✓ Modelo Snowflake

- Normalizado
- Hierarquias mantidas.
- Muitas tabelas → Muitas Junções – 1:N.
- Reduz o espaço de armazenamento dos dados dimensionais, mas acrescenta várias tabelas, deixando-o mais complexo.
- Acesso mais lento que no StarSchema.



Ranking Banco de Dados



<https://db-engines.com/en/ranking>

Rank			DBMS	Database Model	Score		
Aug 2020	Jul 2020	Aug 2019			Aug 2020	Jul 2020	Aug 2019
1.	1.	1.	Oracle +	Relational, Multi-model i	1355.16	+14.90	+15.68
2.	2.	2.	MySQL +	Relational, Multi-model i	1261.57	-6.93	+7.89
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model i	1075.87	+16.15	-17.30
4.	4.	4.	PostgreSQL +	Relational, Multi-model i	536.77	+9.76	+55.43
5.	5.	5.	MongoDB +	Document, Multi-model i	443.56	+0.08	+38.99
6.	6.	6.	IBM Db2 +	Relational, Multi-model i	162.45	-0.72	-10.50
7.	↑ 8.	↑ 8.	Redis +	Key-value, Multi-model i	152.87	+2.83	+8.79
8.	↓ 7.	↓ 7.	Elasticsearch +	Search engine, Multi-model i	152.32	+0.73	+3.23
9.	9.	↑ 11.	SQLite +	Relational	126.82	-0.64	+4.10
10.	↑ 11.	↓ 9.	Microsoft Access	Relational	119.86	+3.32	-15.47
11.	↓ 10.	↓ 10.	Cassandra +	Wide column	119.84	-1.25	-5.37
12.	12.	↑ 13.	MariaDB +	Relational, Multi-model i	90.92	-0.21	+5.96
13.	13.	↓ 12.	Splunk	Search engine	89.91	+1.64	+4.03
14.	↑ 15.	↑ 15.	Teradata +	Relational, Multi-model i	76.78	+0.81	+0.14
15.	↓ 14.	↓ 14.	Hive	Relational	75.29	-1.14	-6.51

NoSQL



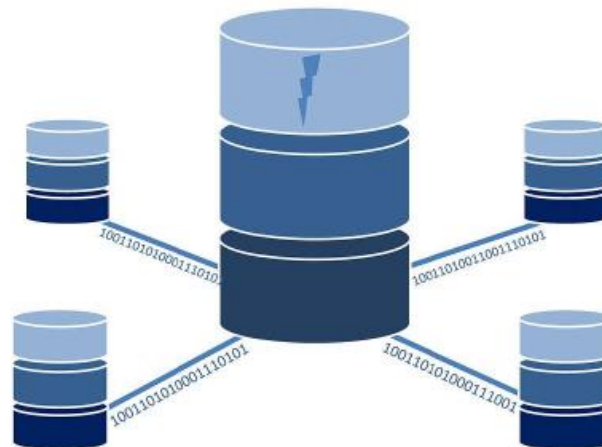
- ✓ NoSQL (originalmente se referindo a "no SQL": "não SQL" ou "não relacional", posteriormente estendido para Not Only SQL - Não Somente SQL) é um termo genérico que representa os bancos de dados não relacionais.
- ✓ Bancos de dados NoSQL são cada vez mais usados em big data e aplicações web de tempo real.



NoSQL – Banco de Dados de Documentos



- ✓ Armazenam chave/valor.
- ✓ O valor é um documento estruturado e indexado, com metadados.
- ✓ Valor (Documento), pode ser consultado.
- ✓ JSON: JavaScript Object Notation.
 - ✓ Feito para troca de dados.
 - ✓ Mais compacto e legível que XML.



NoSQL - JSON



Estrutura nome/valor, entre aspas duplas



Dados separados por vírgula



Chaves separam objetos



Vetores entre colchetes

MongoDB



- ✓ Open source.
- ✓ Multiplataforma.
- ✓ Escalável.



Relacional	MongoDB
Banco de Dados	Banco de Dados
Tabela	Coleção
Linha	Documento
Coluna	Campo

MongoDB - Comandos



```
db.posts.insert([
  {nome:"André",postagem:"Produtos caros", data:"12-01-2019",idade:25},
  {nome:"Ricardo",postagem:"Produtos caros", data:"14-07-2019", idade:12}])

#idade menor ou igual a 12
> db.posts.find({postagem:"Produtos caros",idade: {$lte: 12}})
{ "_id" : ObjectId("5d0911600ee1100c307004da"), "nome" : "Ricardo", "postagem"
: "Produtos caros", "data" : "14-07-2019", "idade" : 12 }
```

- **\$lt: menor que**
- **\$lte: menor ou igual que**
- **\$ne: diferente de**
- **\$in: contém**
- **\$nin: não contém**

MongoDB - Comandos



```
> db.posts.find()
```

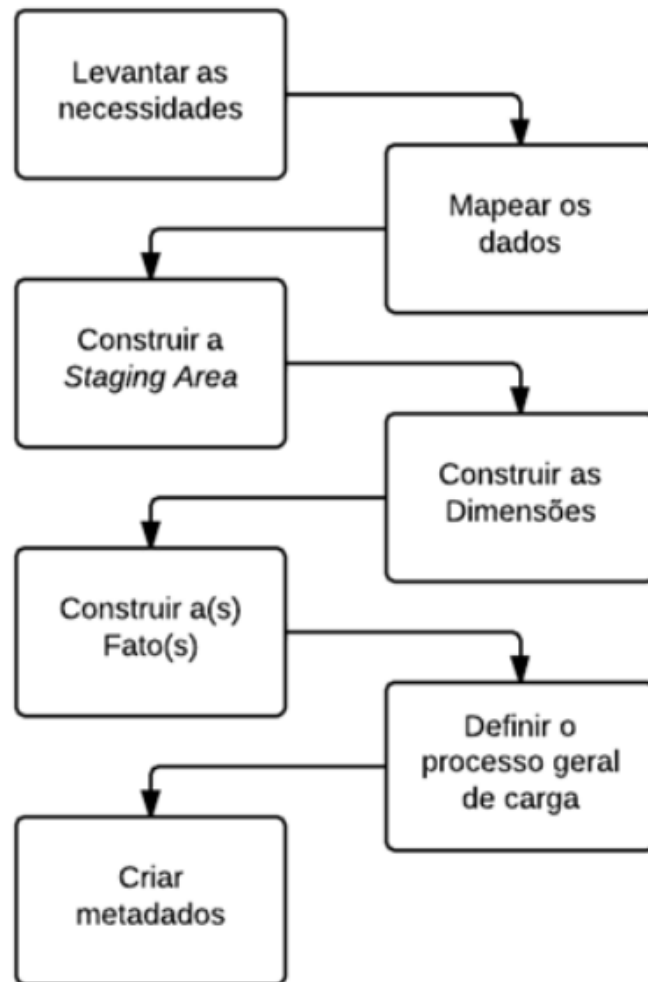
```
{ "_id" : ObjectId("5d090bc10ee1100c307004d4"),  
  "nome" : "José", "postagem" : "Bons Produtos!",  
  "data" : "31-06-2019" }
```

```
{ "_id" : ObjectId("5d090cd10ee1100c307004d5"),  
  "nome" : "Antonio", "postagem" : "Minha bike  
quebrou", "data" : "26-05-2019" }
```

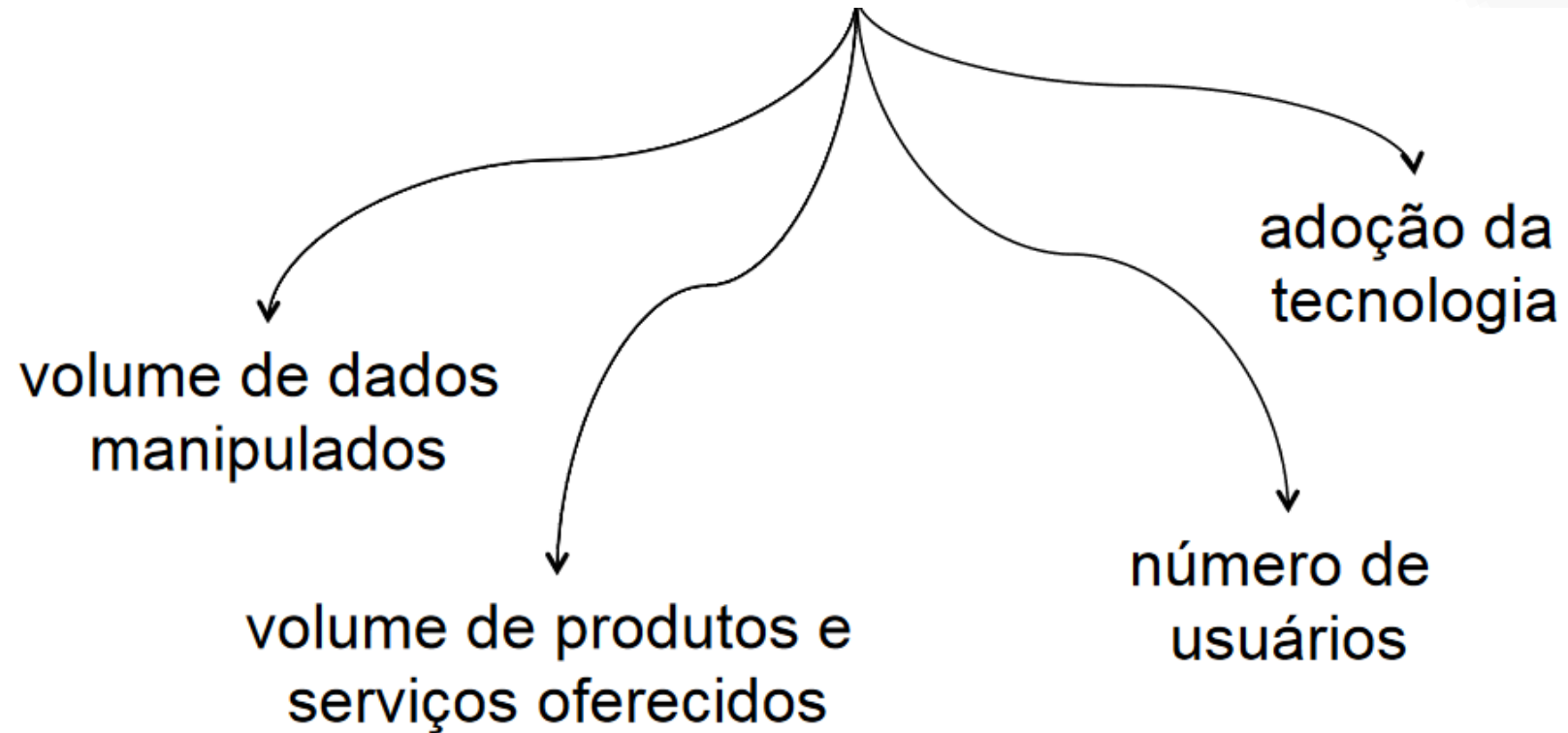
```
{ "_id" : ObjectId("5d090cd10ee1100c307004d6"),  
  "nome" : "Maria Silva", "postagem" : "Encontrei  
tudo que procurava", "data" : "14-06-2019" }
```

```
{ "_id" : ObjectId("5d090cd10ee1100c307004d7"),  
  "nome" : "Lucas Andrade", "postagem" : "Ótimo  
atendimento!", "data" : "12-04-2019" }
```

Etapas na Construção de um DW



Arquitetura



Tuning em Banco de Dados

Em TI, Tuning refere-se basicamente ao conceito de propor e aplicar mudanças visando otimizar o desempenho na recuperação ou atualização de dados. Em curtas palavras, Tuning é sinônimo de otimização.

Para fazer um bom trabalho de Tuning, é necessário executar criteriosamente os seguintes processos:

- Entender o problema;
- Elaborar o diagnóstico;
- Aplicar as dicas e técnicas de otimização (que se aplicam ao diagnóstico elaborado).

Tuning em Banco de Dados

1- Planejamento de performance:

Definição e configuração do ambiente em que o BD será instalado, considerando-se os seguintes itens: Hardware, Software, Sistema Operacional e Infraestrutura de rede.

2- Tuning de instância e BD:

Ajuste de parâmetros e configurações do BD (atividades que fazem parte do trabalho de um DBA).

3- SQL Tuning:

Otimização de instruções SQL.

5 exemplos de BI



Copa do Mundo 2014: Alemanha campeã

A utilização de BI foi um dos fatores responsáveis pela vitória da seleção alemã na Copa do Mundo de 2014. Ele serviu como ferramenta para impulsionar a produtividade do time.

Com o BI, foi possível analisar dados relevantes para o desempenho dos jogadores, tais como:

- número de passes;
- velocidade em campo;
- finalizações;
- defesas;
- penalidades.

Com base nas informações obtidas, foi possível identificar quais atletas tinham melhor rendimento e, assim, escalar o melhor time titular.

5 exemplos de BI

Toyota USA

Há alguns anos, a Toyota enfrentava, nos Estados Unidos, um problema com a gestão de seus dados. Os computadores da empresa geravam grandes quantidades de relatórios, mas nenhum direcionamento.

Com isso, a administração da Toyota era incapaz de exercer controle sobre todos esses documentos e, como consequência, não conseguia adotar estratégias e tomar decisões data-oriented.

A Toyota, então, decidiu implementar o Business Intelligence. O resultado foi impressionante.

Em pouco tempo, a empresa conseguiu otimizar o custo de fabricação dos automóveis e os fluxos de trabalho dos funcionários. **O retorno sobre o investimento é o que mais impressiona: 506%.**

5 exemplos de BI

FIAT

Na época o Uno era um dos principais carros da FIAT Automóveis. Para o lançamento da versão 2011, a empresa tinha um grande desafio: deixar o carro com “a cara do consumidor”.

Para alcançar esse objetivo, a FIAT realizou pesquisas de mercado e levantou dados sobre os seus seguidores nas redes sociais.

Depois, a empresa recorreu ao Business Intelligence para analisar tudo o que foi coletado nas pesquisas.

O sucesso de críticas e de vendas superou as expectativas. O novo Uno recebeu vários prêmios, dentre eles o de Carro do Ano em 2011 pela revista Auto Esporte.

5 exemplos de BI

GASMIG

A Companhia de Gás de Minas Gerais (GASMIG) apresentava sérias dificuldades em gerenciar seus relatórios de faturamento e de vendas.

A GASMIG viu como solução implantar ferramentas de criação e de gerenciamento de banco de dados. Dessa forma, poderiam usar uma análise inteligente desses dados e aplicar na melhoria dos processos de negócio.

Com isso, foi possível organizar melhor todas as demandas, otimizar a execução das tarefas e produzir relatórios com mais qualidade.

5 exemplos de BI

TNT

A TNT é uma das maiores transportadoras de carga expressa do mundo.

Presente em 61 países e com mais de 56.000 funcionários, a TNT tem como um de seus desafios manter a excelência de seus serviços. Para isso, identificou a necessidade de instalar um sistema avançado para análise de dados.

A partir disso, a TNT conseguiu prever certas falhas ou situações que poderiam prejudicar suas entregas.

Com as informações que o BI ajudou a fornecer, foi possível criar inúmeras estratégias de enfrentamento desses problemas e indicadores importantes para a otimização dos recursos e das atividades da empresa.

Internet of Things

IGTI



Inteligência Empresarial



Poluição Sonora.

Otimizar coleta de lixo.

Controle de tráfego.

Controle de distribuição de energia elétrica.

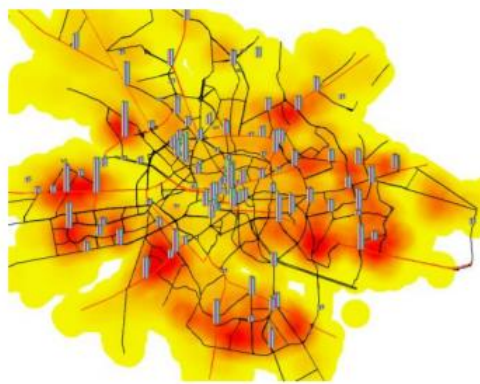
Segurança pública.



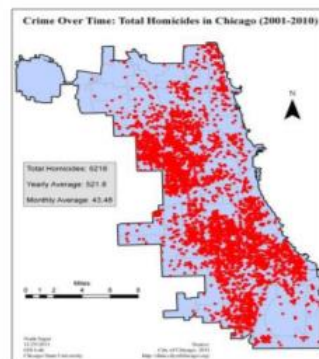
Inteligência Empresarial



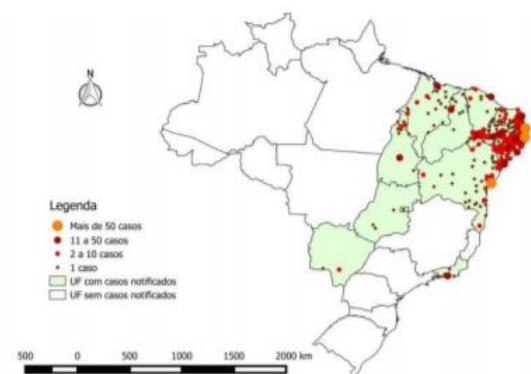
Análise de Rotas



Geomarketing



Análise de Crimes

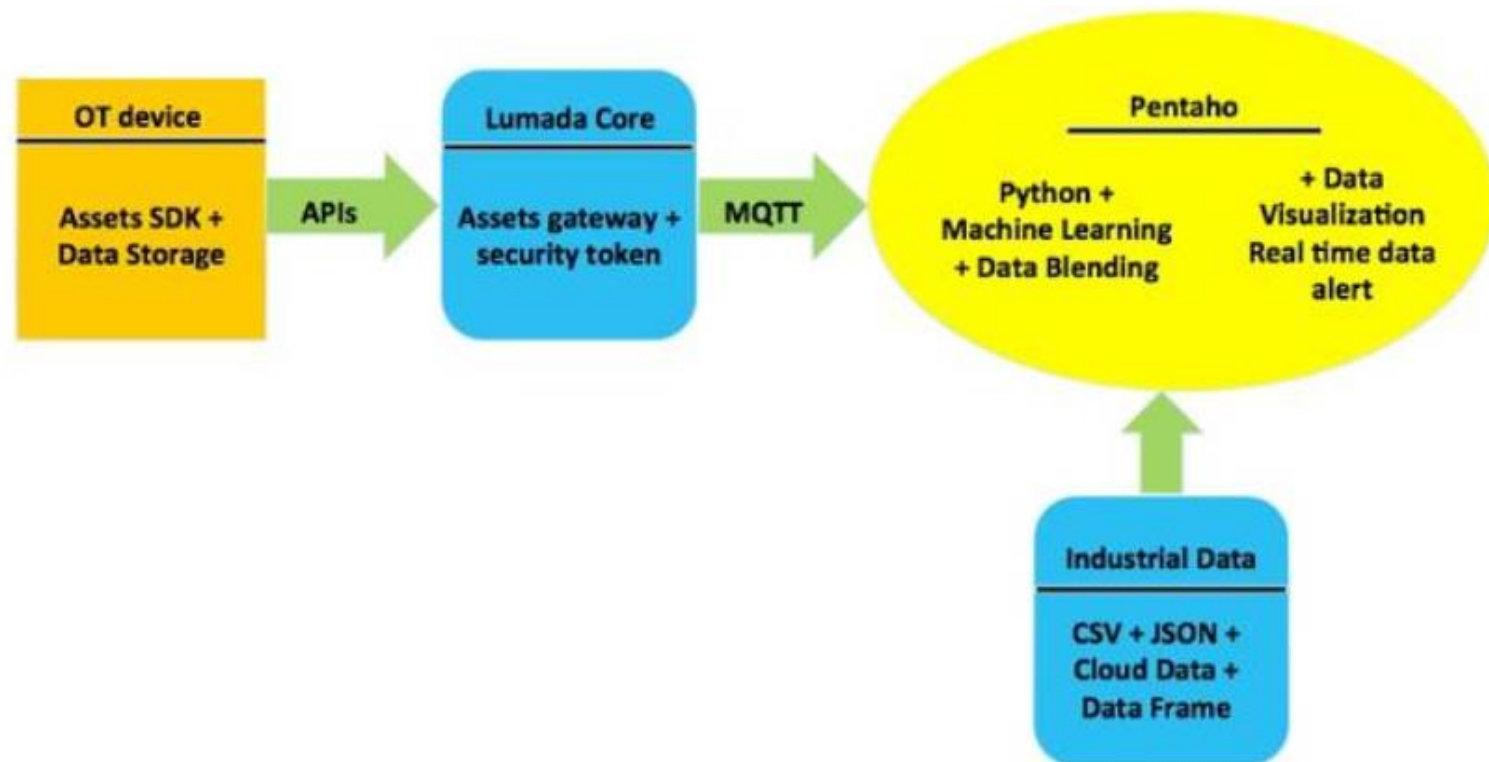


Análise de Doenças



Análise de Desastres

Lumada- Integration with Pentaho

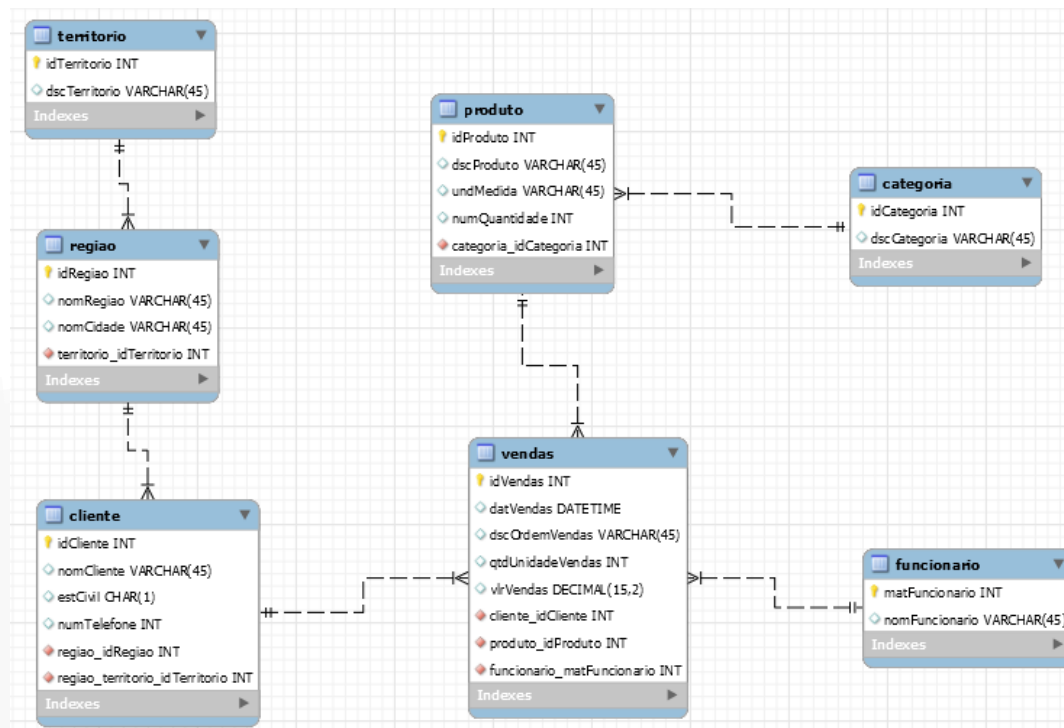


Desafio

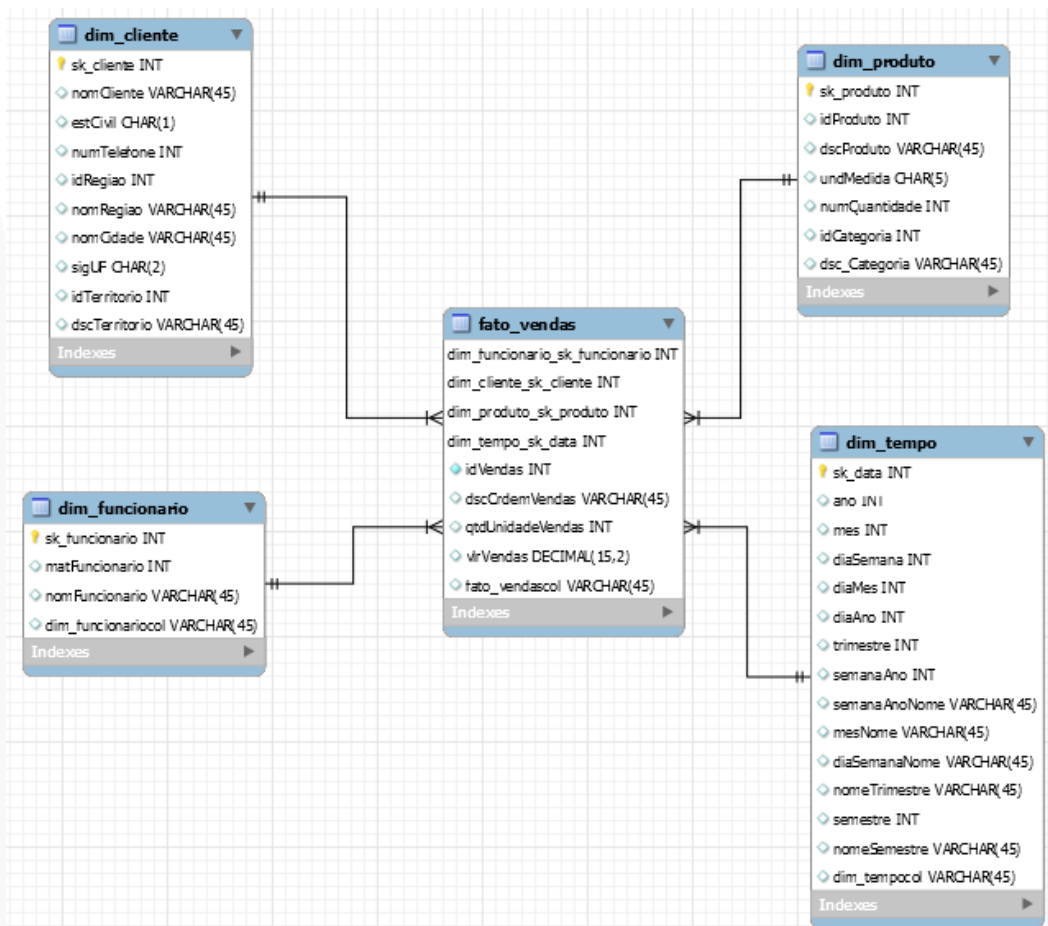
Desafio



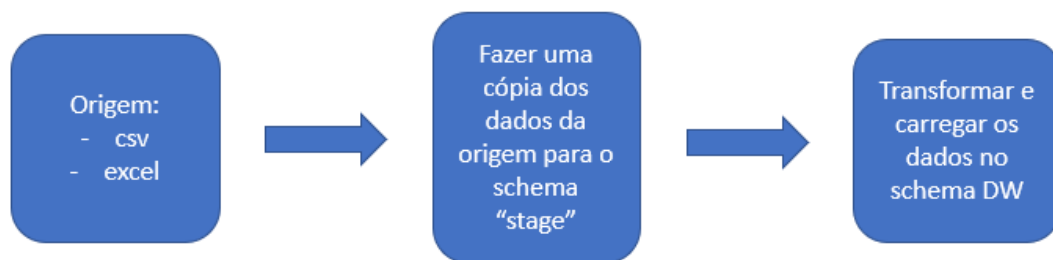
- Categoria.csv
- Cliente.csv
- Funcionario.csv
- Produto.xlsx
- Regiao.csv
- Territorio.csv
- Vendas.xlsx



Desafio



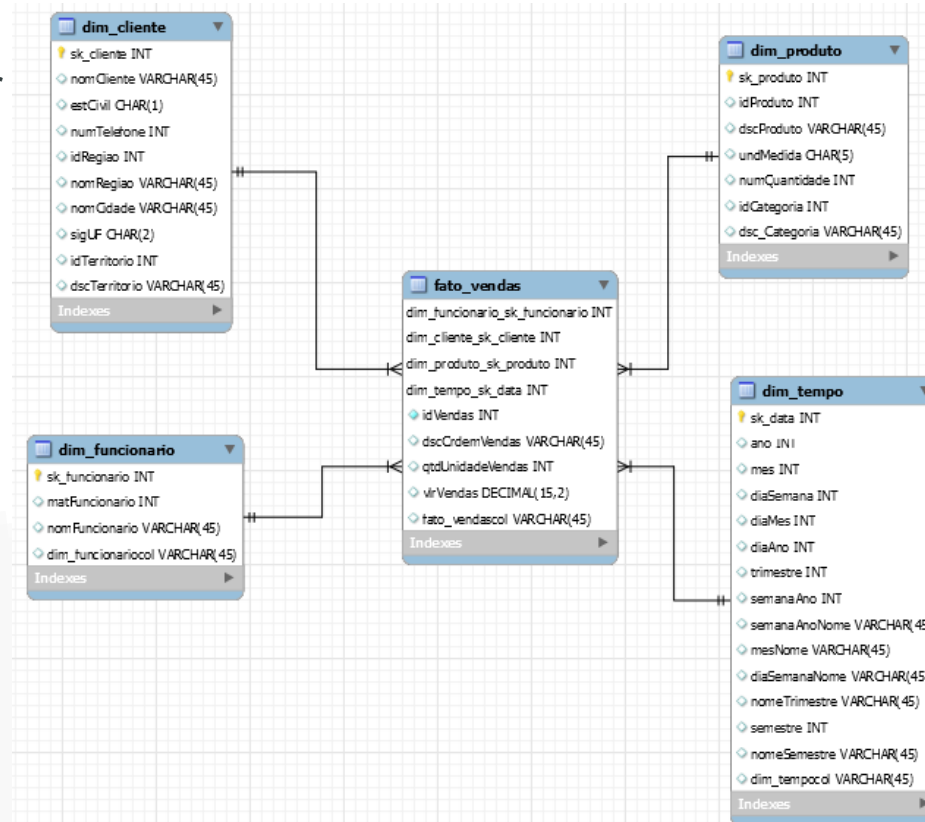
Trabalho Extra



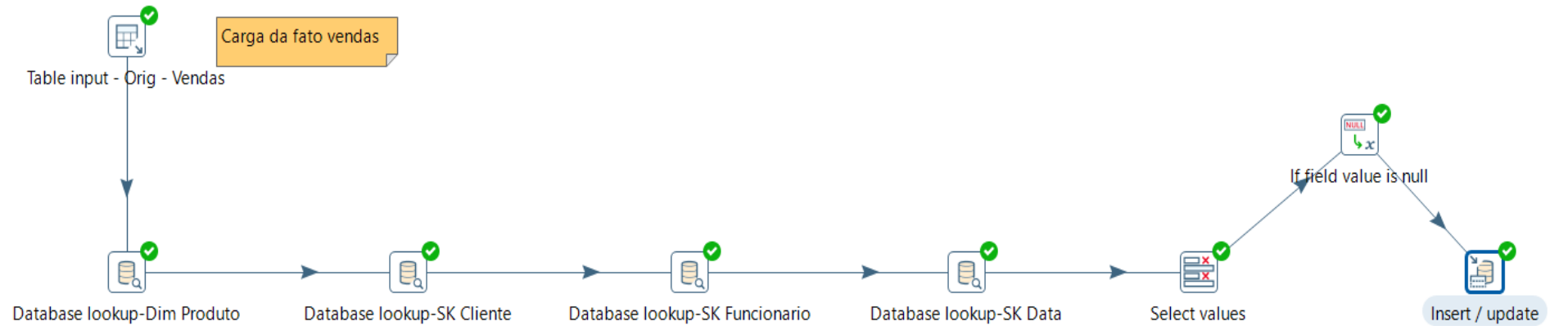
Desafio



- Origem.ktr
- Dimensao_Data.ktr
- Dimensao_Produto.ktr



Desafio



Ensaio no Pentaho



Trabalhando com a passagem de parâmetros e Jobs.



E-mail

IGTI



Início



Informações pessoais



Dados e personalização



Segurança



Pessoas e compartilhamento



Pagamentos e assinaturas

Bem-vindo, I

Gerencie suas informações, privacidade e segurança

Privacidade e personalização







Veja os dados na sua Conta do Google e escolha qual atividade será salva para personalizar sua experiência no Google




[Gerenciar seus dados e a personalização](#)

E-mail



-  Início
-  Informações pessoais
-  Dados e personalização
-  **Segurança**
-  Pessoas e compartilhamento
-  Pagamentos e assinaturas

 Encontrar um dispositivo perdido

[Gerenciar dispositivos](#)

[Gerenciar acesso](#)

Acesso a app menos seguro

Sua conta está vulnerável porque você permite que apps e dispositivos que usam tecnologias de login menos seguras a acessem. Para manter sua conta segura, o Google desativará essa configuração automaticamente se ela não estiver sendo usada. [Saiba mais](#)



 Ativado

[Desativar o acesso \(recomendado\)](#)

Fazer login em outros sites



Fazer login com o Google

Você usa sua Conta do Google para fazer login em 2 sites e apps >

Ensaio no Pentaho



Trabalhando com a Jobs para o Desafio.

