



## A aula interativa do Módulo 3 – Bootcamp Desenvolvedor Business Intelligence em breve!

### Importante:

- 1) Não se esqueça de acessar a aula com seu e-mail cadastrado no ambiente de aprendizagem do IGTI e seu nome completo. Entrou com os dados errados? Saia da sala e entre novamente com os dados corretos.
- 2) Para sua frequência ser computada, responda a enquete no ambiente de aprendizagem, no horário indicado pelo professor. A enquete ficará disponível por 10 minutos.
- 3) Utilize o Chat para interagir com os colegas durante a aula, a ferramenta Raise Hands para pedir a palavra, e, em caso de dúvidas sobre o conteúdo, utilize o Q&A (perguntas e respostas) para que o professor tutor possa respondê-las.

# Informações importantes

- Acesse a aula com o seu nome completo e com o mesmo e-mail utilizado no cadastro do ambiente de aprendizagem do IGTi. Entrou com os dados errados? É só sair da sala e entrar com os dados corretos!
- As presenças das Aulas Interativas são computadas através de uma enquete, que será realizada no ambiente de aprendizagem do IGTi. Para sua frequência ser computada, quando solicitado pelo professor, você deverá ir até a seção “Enquete de presença da Aula Interativa”, localizada na Área Acadêmica da disciplina, e responder a enquete. Essa seção estará logo abaixo da que contiver o link para a Aula Interativa em questão. A enquete ficará no ar por 10 minutos e sua nota estará disponível ao término desse tempo.
- Utilize o chat para interagir com os colegas durante a aula interativa! Em caso de dúvidas sobre o conteúdo, é só postá-las no Q&A que o tutor irá respondê-las! Lembre-se que você ainda pode esclarecer as suas dúvidas nos fóruns disponibilizados no Ambiente de Aprendizagem.

# Informações importantes

- Ah! E se você não conseguir assistir a aula interativa, não se preocupe! Sua gravação ficará disponível no Área Acadêmica, juntamente com os slides utilizados pelo professor, em até 24 horas úteis após o término da aula. Você também poderá realizar a atividade de reposição para recuperar os pontos de presença!
- Se você precisar solicitar prorrogação e/ou 2ª oportunidade para entrega de atividades, saiba que isso é realizado somente mediante a apresentação de atestado médico ou de óbito de parentes de 1º grau.
- Para melhor experiência nas aulas interativas, sugerimos que você baixe o aplicativo do Zoom no seu computador.

# Aplicações em ETL

Segunda Aula Interativa

Prof. Ricardo Brito Alves

# Nesta aula



- ☐ Trabalho Prático.
- ☐ Tópicos da Disciplina e Temas Interessantes.

# Trabalho Prático

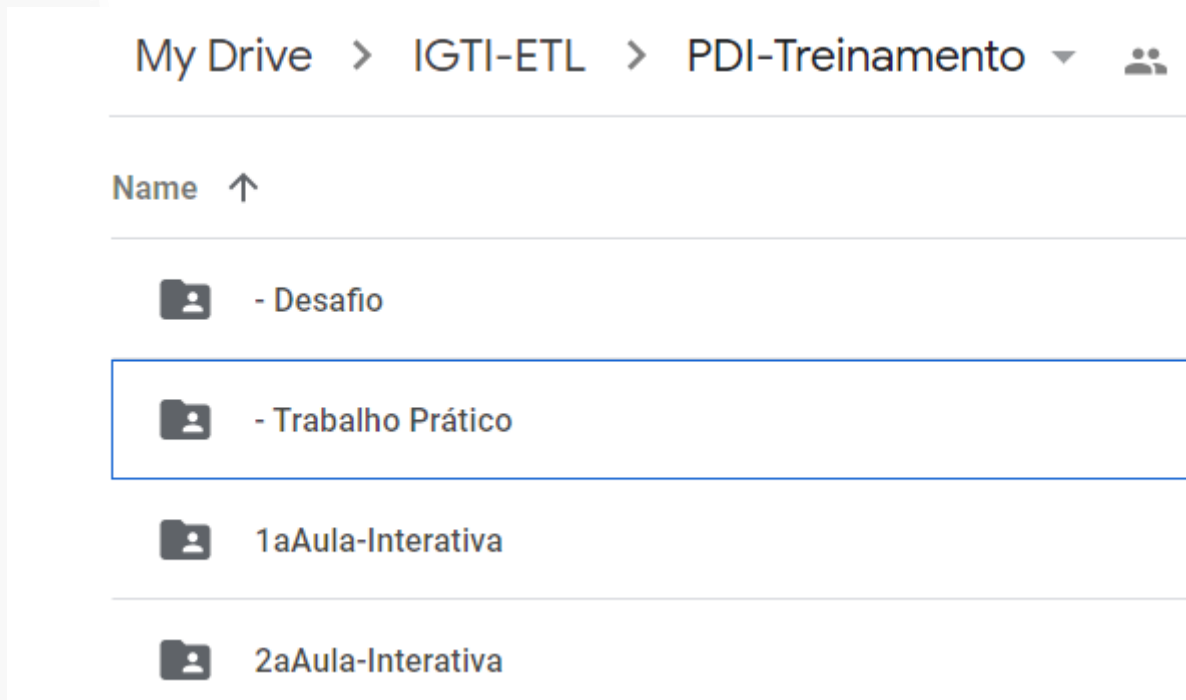
---



# Material



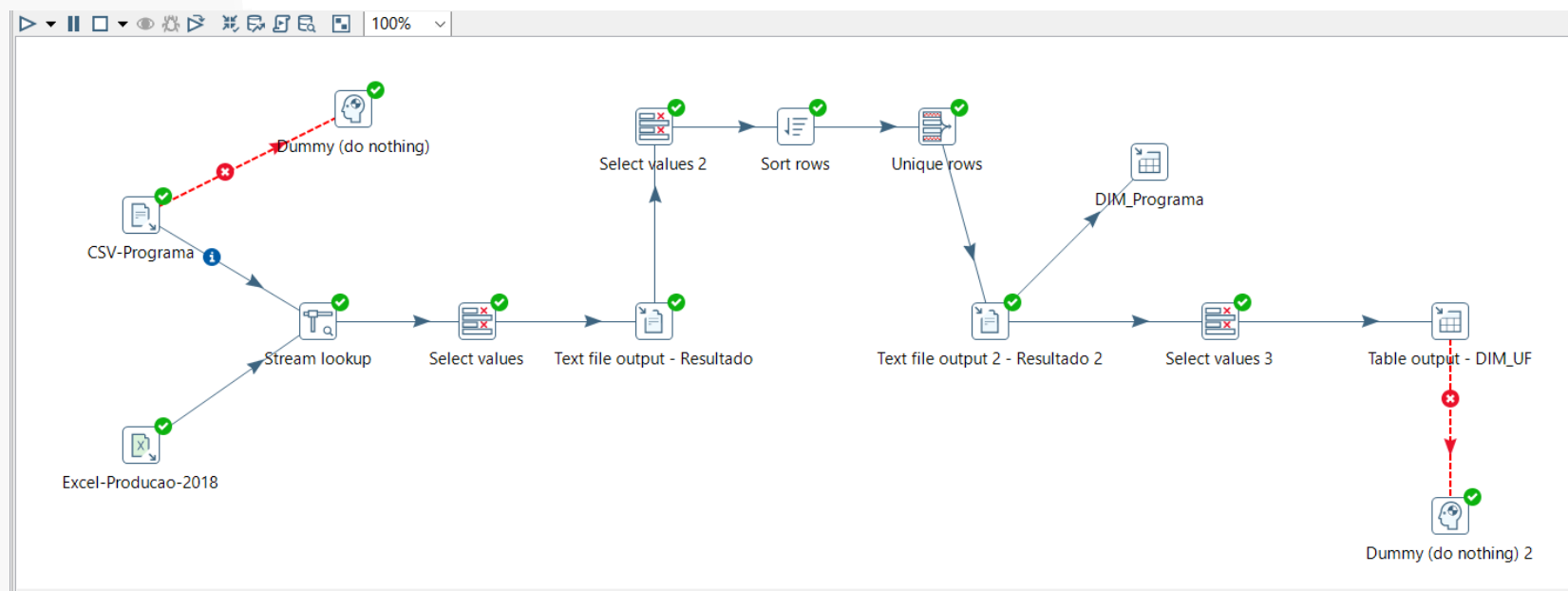
<https://drive.google.com/drive/folders/1AJutZfoazPGdPdGQsE8Sf6jIPlqsdw-f>



# Trabalho Prático



- Producao-2018-Bibliografica-Artigo.xls.
- Programa-2018-Capes.csv.





# **Tópicos da Disciplina e Temas Interessantes**

---

# Ensaio no Pentaho



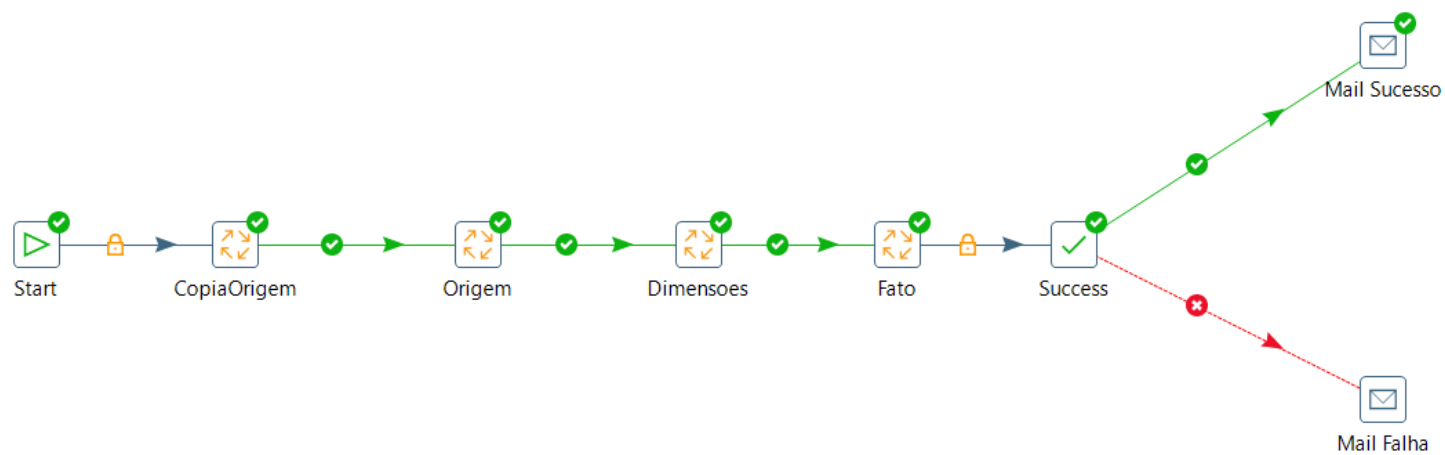
Trabalhando com a passagem de parâmetros e Jobs.



# Ensaio no Pentaho



Trabalhando com a Jobs para o Desafio.



# E-mail

IGTI

Gmail



Início



Informações pessoais



Dados e personalização



Segurança



Pessoas e compartilhamento



Pagamentos e assinaturas

Bem-vindo, I

Gerencie suas informações, privacidade e seguran

## Privacidade e personali- zação







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 >

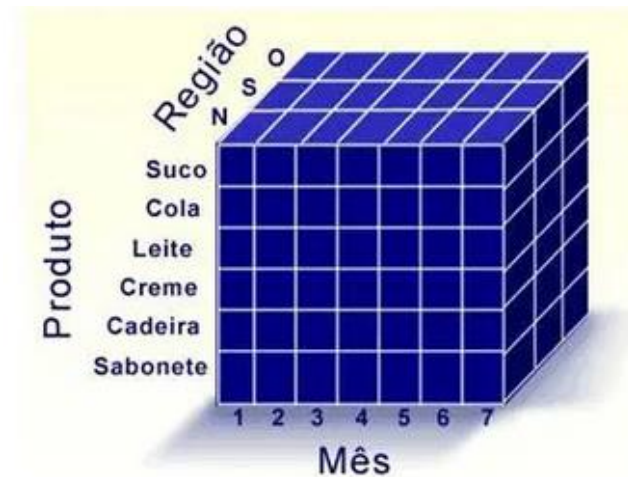
## OLAP - On-Line Analytical Processing (Processamento Analítico On-Line)



OLAP é um conceito de interface com o usuário que proporciona a capacidade de ter ideias sobre os dados, permitindo analisá-los profundamente em diversos ângulos.

As funções básicas do OLAP são:

- Visualização multidimensional dos dados.
- Exploração.
- Rotação.
- Vários modos de visualização.



# Sistema de Arquivos





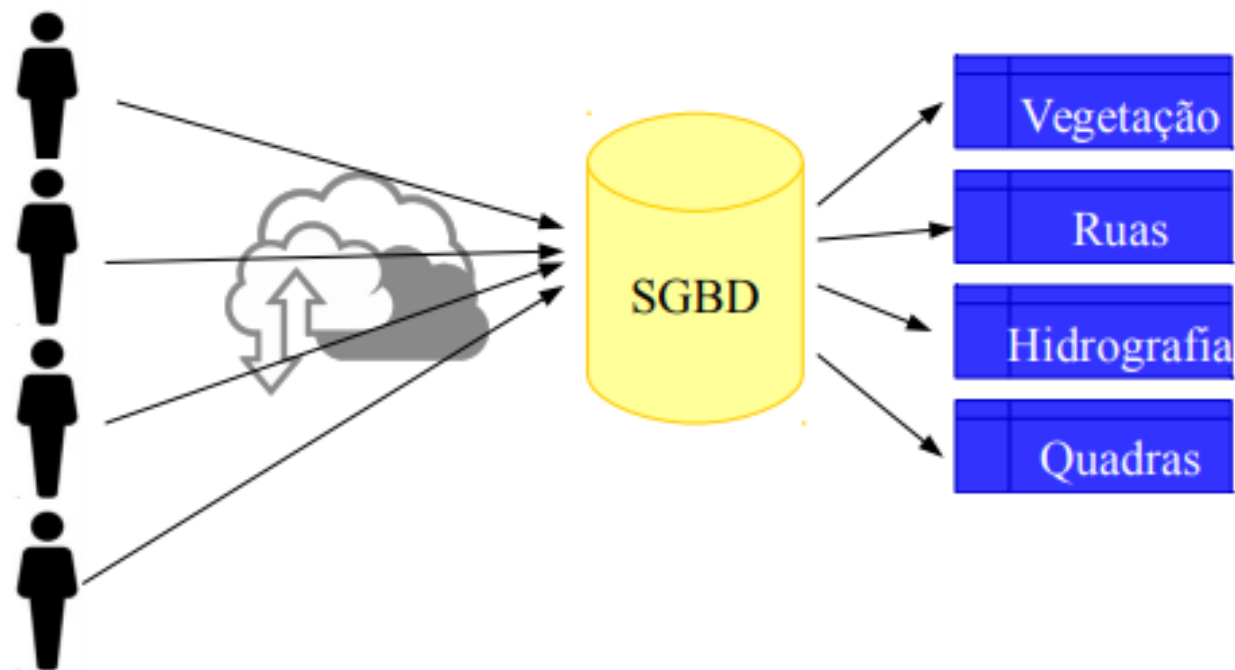
# Banco de Dados



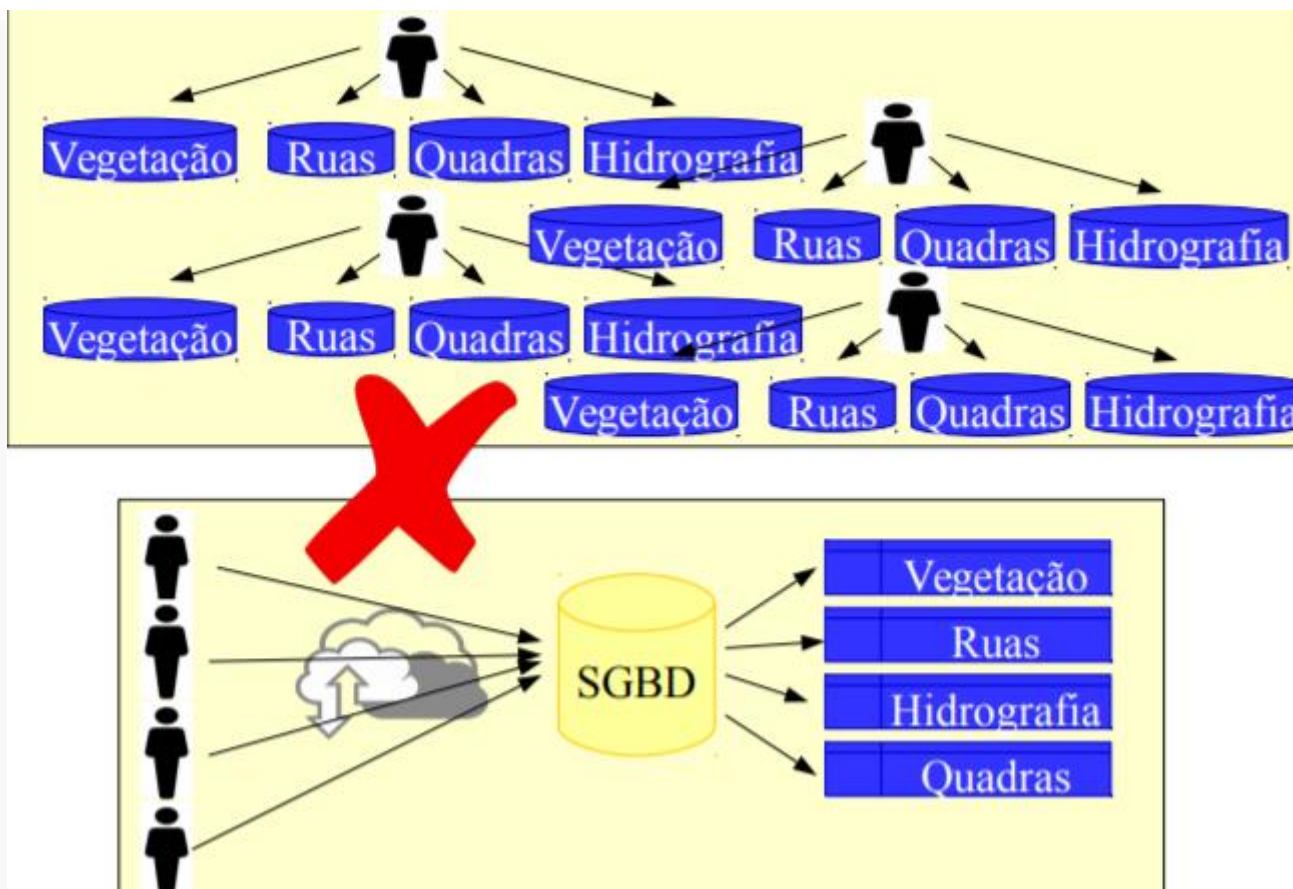
- [Chu, 1985] – Um banco de dados é um conjunto de arquivos relacionados entre si.
- [Date, 2000] – Um banco de dados é uma coleção de dados operacionais armazenados usados pelas aplicações de uma determinada organização.

# 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

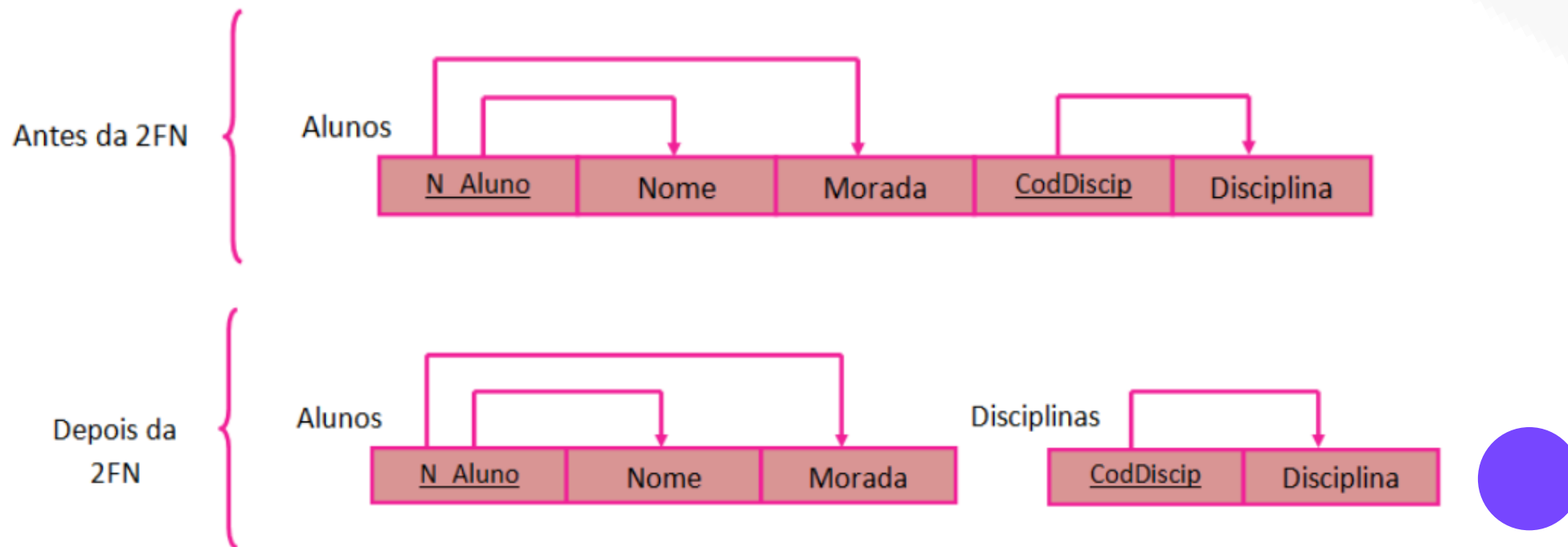
# Normalização



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

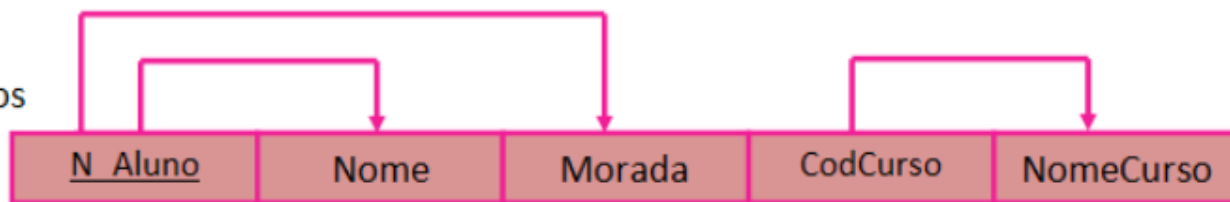




# Normalização

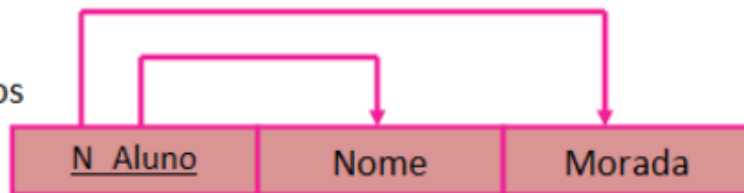
Antes da 3FN

Alunos

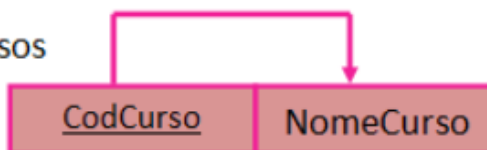


Depois da 3FN

Alunos

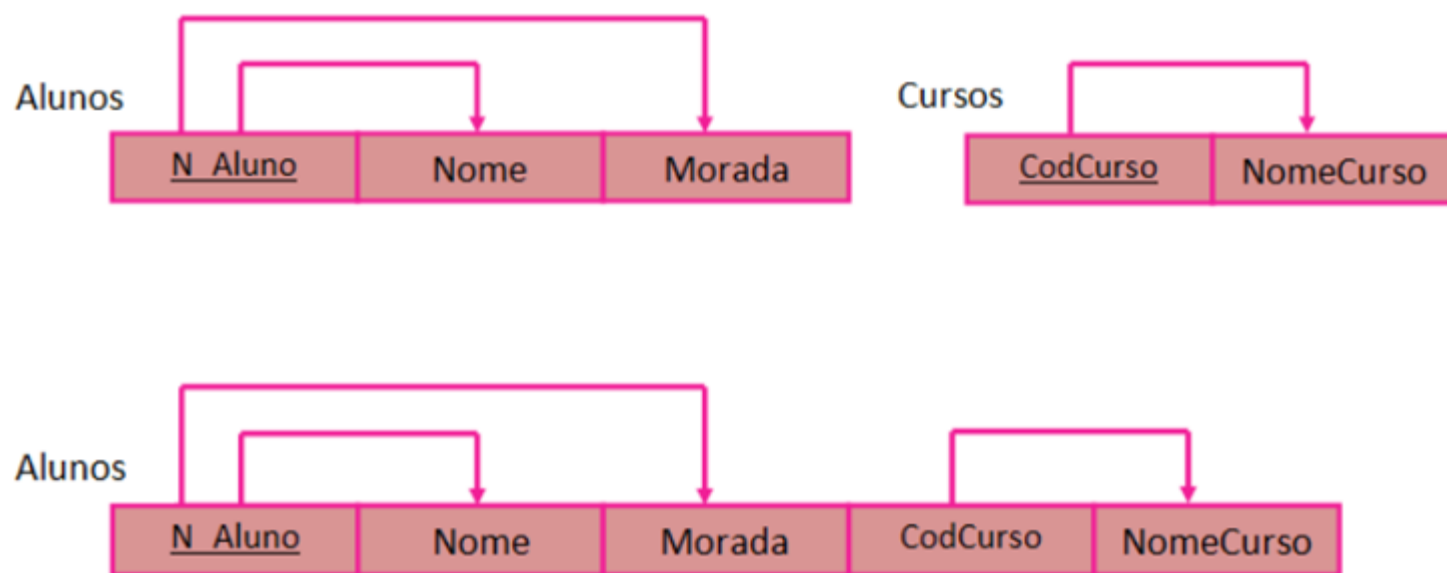


Cursos



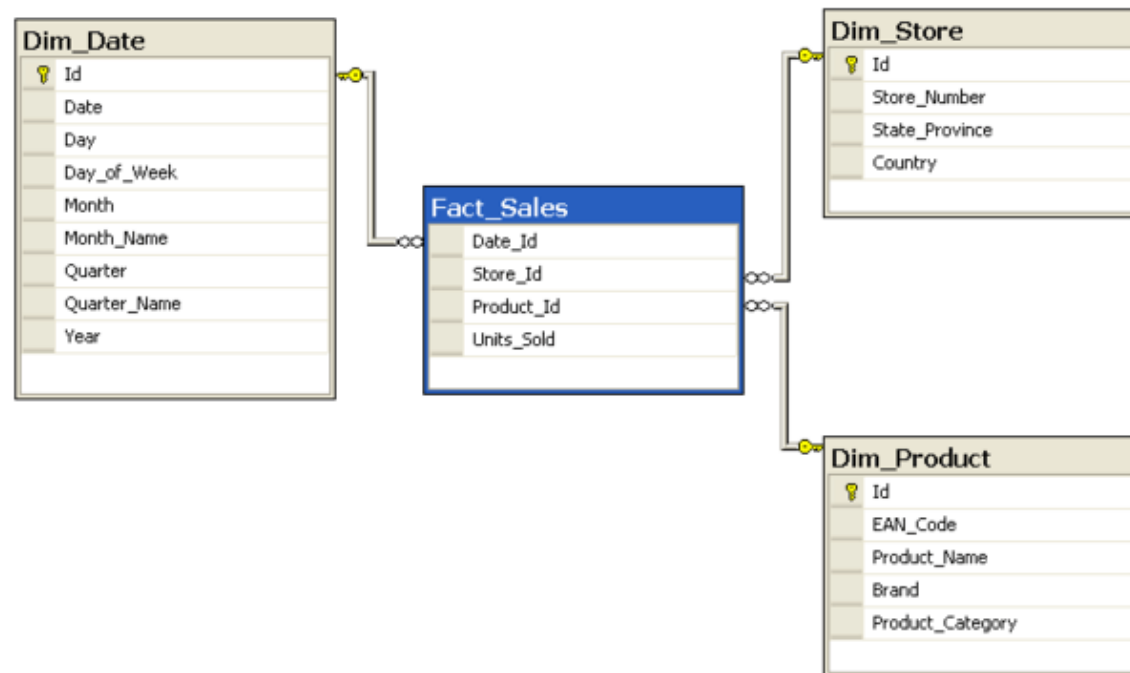
# Desnormalização

- ✓ Dimensão Alunos em um modelo Star Schema



# Modelo Star Schema

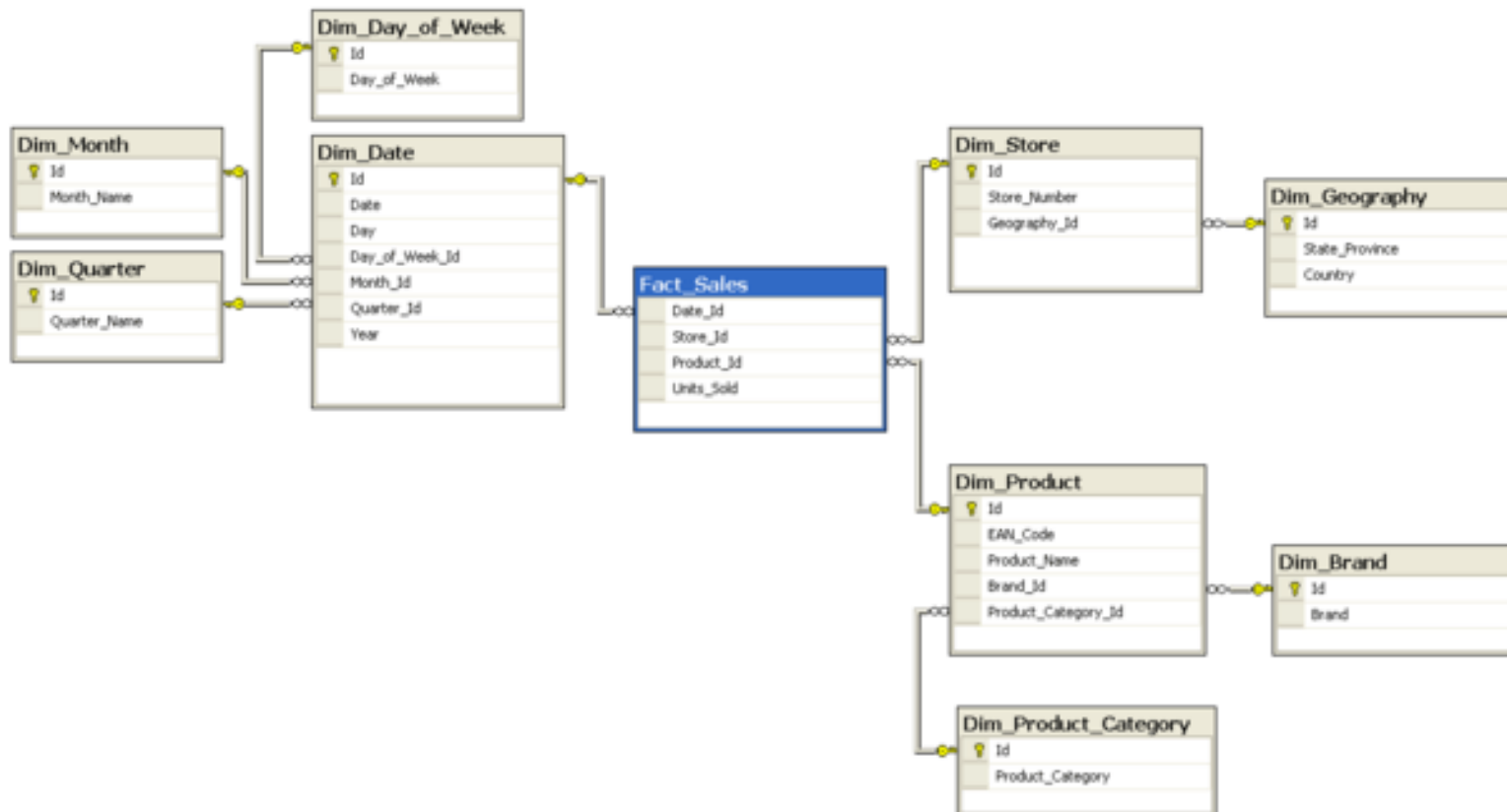
- ✓ Considerando um banco de dados de lojas, produtos e um data warehouse para executar relatórios de vendas agrupados por loja, data, ou categoria ou marca do produto. Se esse data mart estiver usando um esquema em estrela , teria a seguinte aparência:



# Modelo SnowFlake



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



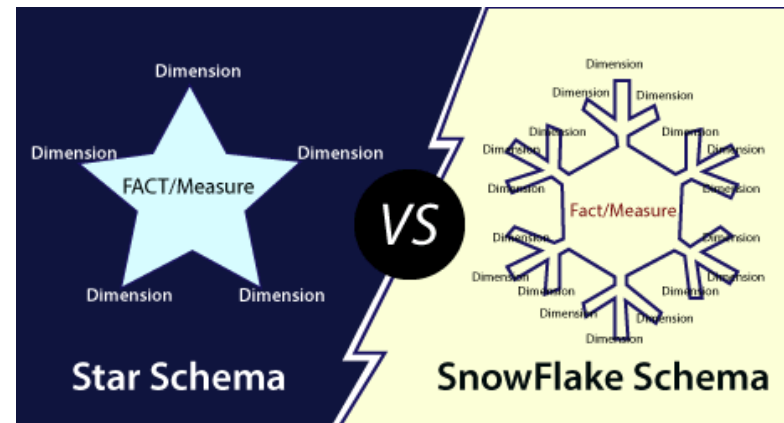
# Considerações sobre os Modelos Dimensionais

## ✓ 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 complex.
- **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.

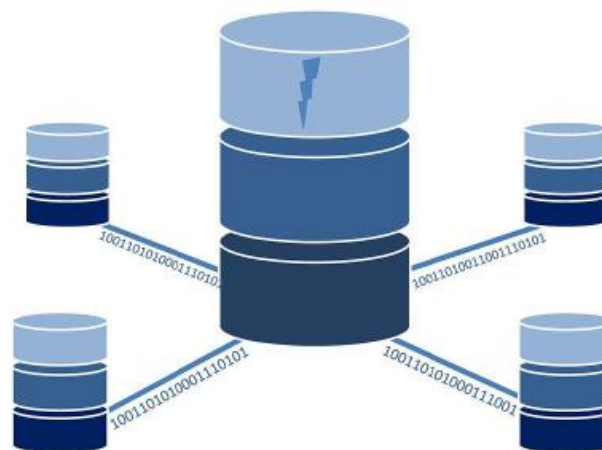




# NoSQL – Banco de Dados de Documentos



- ✓ 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



IGTI

- ✓ 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" }
```