

# BigData

## VANTAGENS

- Interpretar grandes quantidades de dados
- Analisar dados não estruturados
- Interpretar tendências de eventos

## DESVANTAGENS

- Privacidade dos dados
- Segurança dos dados
- Discriminação dos dados



An abstract graphic on the left side of the slide, consisting of a complex network of light blue lines and dots on a dark blue background. The lines are of varying thickness and form a dense, interconnected web that resembles a circuit board or a data network. Some lines are straight, while others are jagged or curved. Small dots are scattered throughout the network, some appearing as bright light sources.

05

**ETL**

Samuel Reine Alves Ribeiro

E - EXTRACT

T - TRANSFORM

L - LOAD

ETL é o nome de um processo de gerenciamento de dados, dividido em três partes: extração, transformação e carregamento.

Esse processo é muito comumente utilizado para a organização de **Data Warehouses**, principalmente quando o fluxo de dados a ser gerenciado é muito grande. O processo de ETL ajuda não somente na organização de dados, mas também na otimização do funcionamento de um sistema.

# Data Warehouse



Empresa

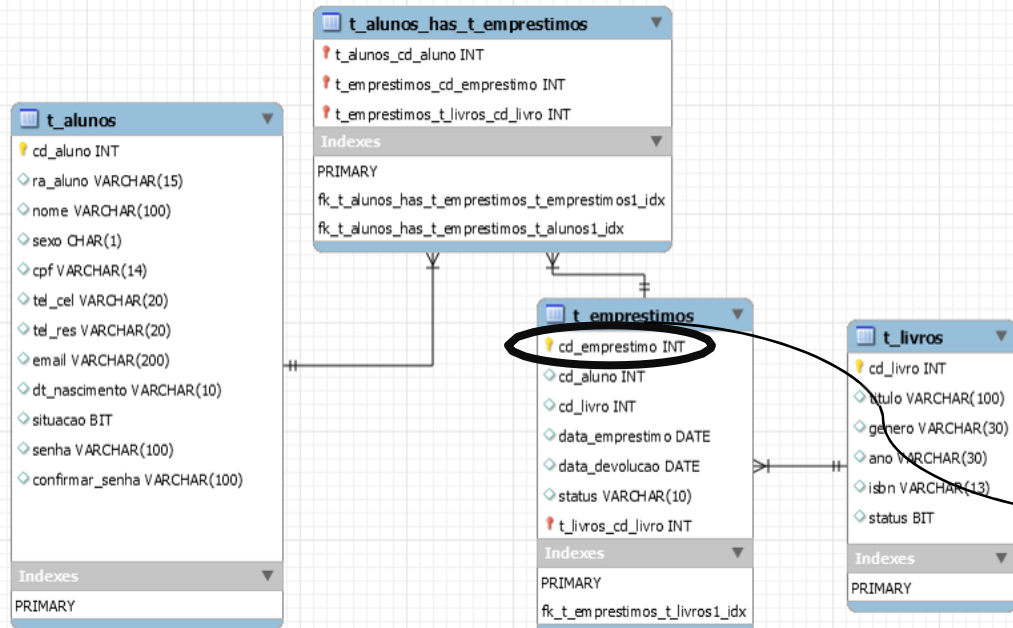


Banco de Dados

Os bancos de dados armazenam de maneira eficiente dados relacionados a uma empresa ou sistema: histórico de vendas de um colaborador, dados de um cliente (e-mail, telefone, endereço, CPF, etc), dados de fornecedores.

Os bancos de dados também possuem um sistema para acesso às suas informações armazenadas: SGDB – Sistema de Gerenciamento do Banco de dados.

# Data Warehouse



Os sistemas gerenciais dos banco de dados geralmente trabalham com uma tabela de dados. Cada item da tabela de dados pode conter uma planilha, ou uma lista, ou uma própria tabela, em relação a si mesmo.

id	dt	id_produto	id_vendedor	id_cliente	qt
1	2014-06-01 13:31:39.383	3	10	1	6
2	2008-07-18 13:31:39.383	7	8	1	9
3	2010-07-06 13:31:39.383	5	10	7	14
4	2015-05-10 13:31:39.383	7	2	10	12
5	2016-11-09 13:31:39.383	3	1	2	9
6	2010-10-27 13:31:39.383	4	2	7	13
7	2016-03-02 13:31:39.383	6	2	3	4
8	2012-10-07 13:31:39.383	5	4	10	5
9	2007-12-06 13:31:39.383	2	11	7	10
10	2015-03-05 13:31:39.383	3	7	9	13
11	2009-09-21 13:31:39.383	7	7	4	6
12	2008-02-06 13:31:39.383	3	7	3	11

# Data Warehouse

Com o crescimento das empresas, as tabelas de dados começam a se tornar cada vez mais complexas e as listas armazenadas nos bancos de dados não são mais suficientes: precisamos de relatórios a partir dessas listas. Precisamos analisar os dados armazenados no nosso banco de dados. Isso requer sempre **PODER DE PROCESSAMENTO**.

Tabela de vendas

id	dt	id_produto	id_vendedor	id_cliente	qt
1	2014-06-01 13:31:39.383	10	1	6	
2	2008-07-18 13:31:39.383	7	8	1	9
3	2010-07-06 13:31:39.383	6	10	7	14
4	2015-05-10 13:31:39.383	7	2	10	12
5	2016-11-09 13:31:39.383	3	1	2	9
6	2010-10-27 13:31:39.383	4	2	7	13
7	2016-03-02 13:31:39.383	6	2	3	4
8	2012-10-07 13:31:39.383	5	4	10	5
9	2007-12-06 13:31:39.383	2	11	7	10
10	2015-03-05 13:31:39.383	3	7	9	13
11	2009-09-21 13:31:39.383	7	7	4	6
12	2008-02-06 13:31:39.383	3	7	3	11

Relatório para o RH

Mes_Referência	Colaborador	Cargo	Total_Vendas	Salario	Comissoes	Salario_do_mes
Jan/2017	Vitor Festeiro	DBA	R\$ 1.862.400,00	R\$ 30.000,00	R\$ 27.936,00	R\$ 57.936,00
Jan/2017	Marcelino Pinga	CEO	R\$ 914.900,00	R\$ 40.000,00	R\$ 13.723,50	R\$ 53.723,50
Jan/2017	Saturnino Besta	Coordenador de IT	R\$ 1.049.500,00	R\$ 30.000,00	R\$ 15.742,50	R\$ 45.742,50
Jan/2017	Rolando Rocha	Analista de IT	R\$ 1.707.400,00	R\$ 20.000,00	R\$ 25.611,00	R\$ 45.611,00
Jan/2017	Salvador das Dores Fortes	Gerente de IT	R\$ 600.500,00	R\$ 30.000,00	R\$ 9.007,50	R\$ 39.007,50
Jan/2017	Email Suarez	Gerente de Vendas	R\$ 999.500,00	R\$ 20.000,00	R\$ 14.992,50	R\$ 34.992,50
Jan/2017	Vespertino Cedo	Gerente Administrativo	R\$ 1.239.900,00	R\$ 15.000,00	R\$ 18.598,50	R\$ 33.598,50

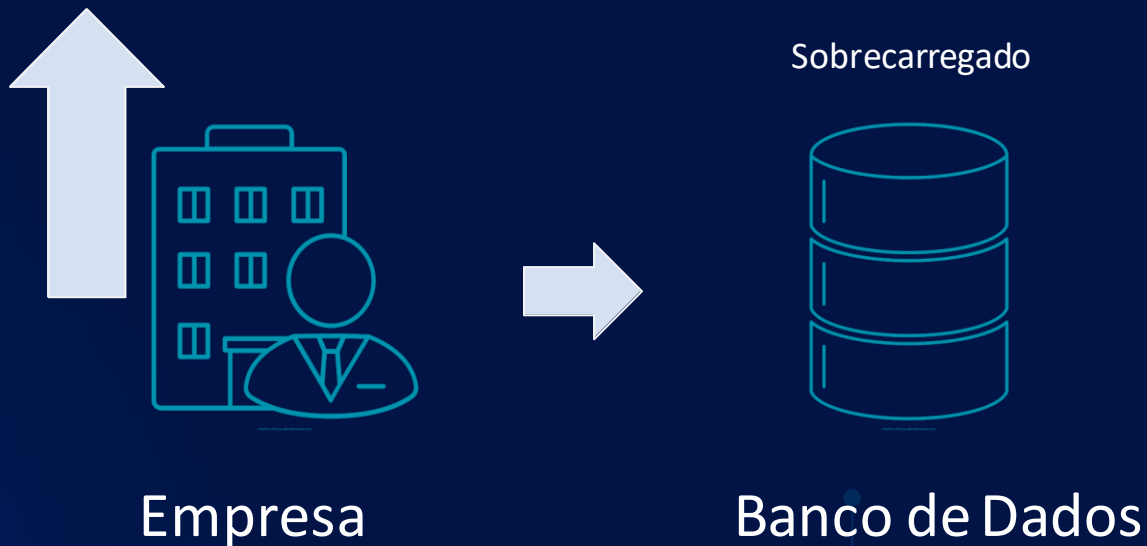
Relatório para o Comercial

Mes_Referência	Produto	Meta	Vendas	% meta atingida
Jan/2017	CPU Nasa 10Ghz	R\$ 1.000.000,00	R\$ 1.034.000,00	103.40 %
Jan/2017	HD SSD 1TB	R\$ 1.000.000,00	R\$ 1.012.000,00	101.20 %
Jan/2017	Monitor 32 Pol Full HD	R\$ 500.000,00	R\$ 537.000,00	107.40 %
Jan/2017	Mouse Gammer 25 botões	R\$ 100.000,00	R\$ 98.800,00	98.80 %
Jan/2017	Notebook Gamer Pro	R\$ 12.000.000,00	R\$ 13.025.000,00	108.54 %
Jan/2017	RAM DDR4 4GB Powerturbo	R\$ 300.000,00	R\$ 206.000,00	68.67 %
Jan/2017	Teclado Gammer 350 teclas	R\$ 100.000,00	R\$ 175.500,00	175.50 %

# Data Warehouse

Se a empresa cresce muito, o processamento do banco de dados pode começar a ser muito requerido e, com o volume cada vez maior de dados a ser armazenado, gerenciado, processado e analisado, o sistema da empresa pode acabar por ficar sobrecarregado e lento.

Como o banco de dados fica dentro do servidor da empresa, melhorar o poder de processamento do servidor poderia resolver o problema da sobrecarga do sistema. Porém, essa solução pode ser muito onerosa.



# Data Warehouse & ETL

Qual a solução para o problema de armazenamento, gerenciamento e análise de dados com o crescimento de uma empresa? Uma das soluções mais comuns é chamada de CRESCIMENTO VERTICAL COM DATA WAREHOUSE E ETL. Consiste em manter o banco de dados originais, com as listas e tabelas anteriores, e ao mesmo tempo criar um novo banco de dados, chamado de Data Warehouse no qual serão carregados os dados tratados (informações) provenientes do banco de dados original.

## Banco de produção - Relacional

Vendas					
id	dt	id_produto	id_vendedor	id_cliente	dt
1	2014-06-01 13:31:39.363	3	10	1	4
2	2008-07-18 13:31:39.363	7	8	1	9
3	2010-01-06 13:31:39.363	5	10	7	14
4	2010-09-18 13:31:39.363	7	2	10	12
5	2016-11-09 13:31:39.363	3	1	2	9
6	2010-10-27 13:31:39.363	4	7	7	13
7	2010-03-02 13:31:39.363	6	2	3	4
8	2012-10-01 13:31:39.363	9	4	10	5
9	2007-12-04 13:31:39.363	2	11	7	10
10	2015-03-05 13:31:39.363	3	7	9	13
11	2009-09-21 13:31:39.363	7	7	4	6
12	2009-02-05 12:31:39.363	3	7	3	11

Colaboradores					
id	id_gestor	nm	du_juogo	id_salario	pc_comissao
1	0	Marcos Pires	CEO	40000.00	1.50
2	1	Brazil Soares	Gerente de Vendas	20000.00	1.50
3	1	Feliciano Triste	Assistente de Vendas	7000.00	1.50
4	1	Estelinda Damiana	Assistente de Vendas	6000.00	1.50
5	1	Assisencio das Neves	Assistente de Vendas	8000.00	1.50
6	8	Vicentini Celso	Gerente Administrativo	15000.00	1.50
7	9	Assisencio da Silva	Caixa	7000.00	1.50
8	5	Neuquim da Luz	Assistente Administrativo	7000.00	1.50
9	0	Joel Cincha	Assessor Técnico	15000.00	1.50
10	0	Salvador dos Santos Farias	Gerente de IT	30000.00	1.50
11	9	Vitor Soares	DBA	20000.00	1.50
12	9	Salvador Beata	Coordenador de IT	30000.00	1.50

Produtos					
id	nm	vl	ic_ativo	meta	custo
1	Mouse Gamer 25 botões	200.00	1	100000.00	100.00
2	Teclado Gamer 350 teclas	300.00	1	200000.00	195.00
3	Monitor 32 Pol Full HD	1000.00	1	600000.00	490.00
4	RAM DDR4 4GB Power Turbo	500.00	1	300000.00	325.00
5	CPU Razer 10GHz	2000.00	1	1000000.00	1300.00
6	HD SSD 1TB	2000.00	1	1000000.00	1300.00
7	Notebook Gamer Pro	25000.00	1	1200000.00	16250.00

## Banco de dados original

# ETL

Novo Banco de dados, com informações tratadas – Data Warehouse

## Banco datawarehouse

### folha\_pagamentos

Mes Referência	Colaborador	Cargo	Total_Vendas	Salario	Comissao	Salario_br_jeje
Jan/2017	Vitor Soares	DBA	R\$ 1.952.400,00	R\$ 20.000,00	R\$ 27.336,00	R\$ 57.036,00
Jan/2017	Marcos Pires	CEO	R\$ 514.300,00	R\$ 40.000,00	R\$ 13.725,00	R\$ 55.725,00
Jan/2017	Salvador Beata	Coordenador de IT	R\$ 1.049.500,00	R\$ 30.000,00	R\$ 15.742,50	R\$ 45.742,50
Jan/2017	Ricardo Rocha	Analista de IT	R\$ 1.707.400,00	R\$ 25.000,00	R\$ 25.611,00	R\$ 45.611,00
Jan/2017	Salvador dos Santos Farias	Gerente de IT	R\$ 400.000,00	R\$ 30.000,00	R\$ 5.007,50	R\$ 35.007,50
Jan/2017	Estel Celso	Gerente de Vendas	R\$ 995.500,00	R\$ 20.000,00	R\$ 54.900,00	R\$ 54.900,00
Jan/2017	Vicentini Celso	Gerente Administrativo	R\$ 1.239.300,00	R\$ 15.000,00	R\$ 18.586,50	R\$ 23.586,50

### acompanhamento metas

Mes Referência	Produto	Meta	Vendas	% meta atingida
Jan/2017	CPU Razer 10GHz	R\$ 1.000.000,00	R\$ 1.034.000,00	103,40 %
Jan/2017	HD SSD 1TB	R\$ 1.000.000,00	R\$ 1.012.000,00	101,20 %
Jan/2017	Monitor 32 Pol Full HD	R\$ 500.000,00	R\$ 537.000,00	107,40 %
Jan/2017	Mouse Gamer 25 botões	R\$ 100.000,00	R\$ 98.800,00	98,80 %
Jan/2017	Notebook Gamer Pro	R\$ 12.000.000,00	R\$ 13.025.000,00	108,54 %
Jan/2017	RAM DDR4 4GB Power Turbo	R\$ 300.000,00	R\$ 206.000,00	68,67 %
Jan/2017	Teclado Gamer 350 teclas	R\$ 100.000,00	R\$ 175.500,00	175,50 %

# Data Warehouse & ETL

O ETL então vai extrair os dados do banco de produção de dados original (Extract), transformar nos relatórios que eu preciso, analisa-os (Transform) e carregar para o novo banco de dados, chamado Data Warehouse as informações já prontas para consulta (Load).

## Banco de produção - Relacional

### Vendas

id	dt	id_cliente	id_produto	id_vendedor	id_loja
1	2014-06-01 13:31:29.363	3	10	1	6
2	2009-07-18 13:31:29.363	7	8	1	9
3	2010-07-06 13:31:29.363	5	10	7	14
4	2010-05-10 13:31:29.363	7	2	10	12
5	2010-11-08 13:31:29.363	3	1	3	9
6	2010-10-27 13:31:29.363	4	2	7	13
7	2010-03-02 13:31:29.363	6	2	3	4
8	2012-10-01 13:31:29.363	5	4	10	5
9	2007-12-06 13:31:29.363	2	11	7	10
10	2010-03-01 13:31:29.363	3	7	9	13
11	2009-09-21 13:31:29.363	7	7	4	6
12	2009-02-05 13:31:29.363	3	7	3	11

### Colaboradores

id	id_empresa	nm	dt_nascimento	id_unidade	id_salario	id_comissao
1	NULL	Marcelino Pires	CSO	40000.00	1.50	
2	1	Enal Soares	Gerente de Vendas	20000.00	1.50	
3	1	Feliciano Sales	Assistente de Vendas	7000.00	1.50	
4	1	Estelina Clemente	Assistente de Vendas	8000.00	1.50	
5	1	Alencar dos Santos	Assistente de Vendas	9000.00	1.50	
6	1	Vespertino Cado	Gerente Administrativo	15000.00	1.50	
7	5	Alencar da Silva	Cabeleireiro	7000.00	1.50	
8	5	Joel Costa	Assistente Administrativo	7000.00	1.50	
9	6	Salvador dos Santos	Gerente de IT	30000.00	1.50	
10	9	Vitor Farias	CSO	20000.00	1.50	
11	9	Salvador dos Santos	Coordenador de IT	30000.00	1.50	

### Produtos

id	nm	id_categoria	id_subcategoria	id_preco	id_estoque
1	Mouse Gamer 25 botões	200.00	1	100000.00	130.00
2	Teclado Gamer 350 teclas	300.00	1	200000.00	190.00
3	Monitor 32 Pol Full HD	1000.00	1	800000.00	480.00
4	RAM DDR4 4GB PowerTurbo	800.00	1	300000.00	320.00
5	CPU i7 10GHz	2000.00	1	1000000.00	1300.00
6	HD SSD 1TB	2000.00	1	1000000.00	1300.00
7	Notebook Gamer Pro	25000.00	1	1200000.00	76250.00

Banco de dados original

# ETL

Novo Banco de dados, com informações tratadas – Data Warehouse

## Banco datawarehouse

### folha\_pagamentos

Mes_Preferencia	Colaborador	Cargo	Total_Vendas	Salario	Comissao	Salario_Alt_Jan
Jan/2017	Vitor Farias	CSO	R\$ 1.940.400,00	R\$ 30.000,00	R\$ 27.500,00	R\$ 57.500,00
Jan/2017	Marcelino Pires	CSO	R\$ 914.900,00	R\$ 40.000,00	R\$ 13.723,50	R\$ 53.723,50
Jan/2017	Salvador dos Santos	Coordenador de IT	R\$ 1.049.500,00	R\$ 30.000,00	R\$ 15.742,50	R\$ 45.742,50
Jan/2017	Rolando Rocha	Analista de IT	R\$ 1.707.400,00	R\$ 20.000,00	R\$ 25.611,00	R\$ 45.611,00
Jan/2017	Salvador dos Santos	Gerente de IT	R\$ 600.000,00	R\$ 10.000,00	R\$ 8.307,50	R\$ 38.307,50
Jan/2017	Enal Soares	Gerente de Vendas	R\$ 999.000,00	R\$ 20.000,00	R\$ 14.952,50	R\$ 34.952,50
Jan/2017	Vespertino Cado	Gerente Administrativo	R\$ 1.239.900,00	R\$ 15.000,00	R\$ 18.598,50	R\$ 33.598,50

### acompanhamento metas

Mes_Preferencia	Produto	Meta	Vendas	% meta atingida
Jan/2017	CPU i7 10GHz	R\$ 1.000.000,00	R\$ 1.034.000,00	103,40 %
Jan/2017	HD SSD 1TB	R\$ 1.000.000,00	R\$ 1.012.000,00	101,20 %
Jan/2017	Monitor 32 Pol Full HD	R\$ 500.000,00	R\$ 537.000,00	107,40 %
Jan/2017	Mouse Gamer 25 botões	R\$ 100.000,00	R\$ 98.800,00	98,80 %
Jan/2017	Notebook Gamer Pro	R\$ 12.000.000,00	R\$ 13.025.000,00	108,54 %
Jan/2017	RAM DDR4 4GB PowerTurbo	R\$ 300.000,00	R\$ 206.000,00	68,67 %
Jan/2017	Teclado Gamer 350 teclas	R\$ 100.000,00	R\$ 175.500,00	175,50 %

# Data Warehouse

## Vantagens

- Alta Performance.
- Fácil de se consultar e analisar os dados, já que no DW as informações já são tratadas.
- Reaproveitamento de processamento: uma vez transformados, os dados já ficam prontos para consulta e análise.

## Desvantagem

- Alto custo gerencial para tratar e transformar os dados. Pode sair caro para pequenas empresas.

An abstract background pattern of glowing blue lines and dots, resembling a circuit board or data network, set against a dark blue gradient.

05

# Data Mining

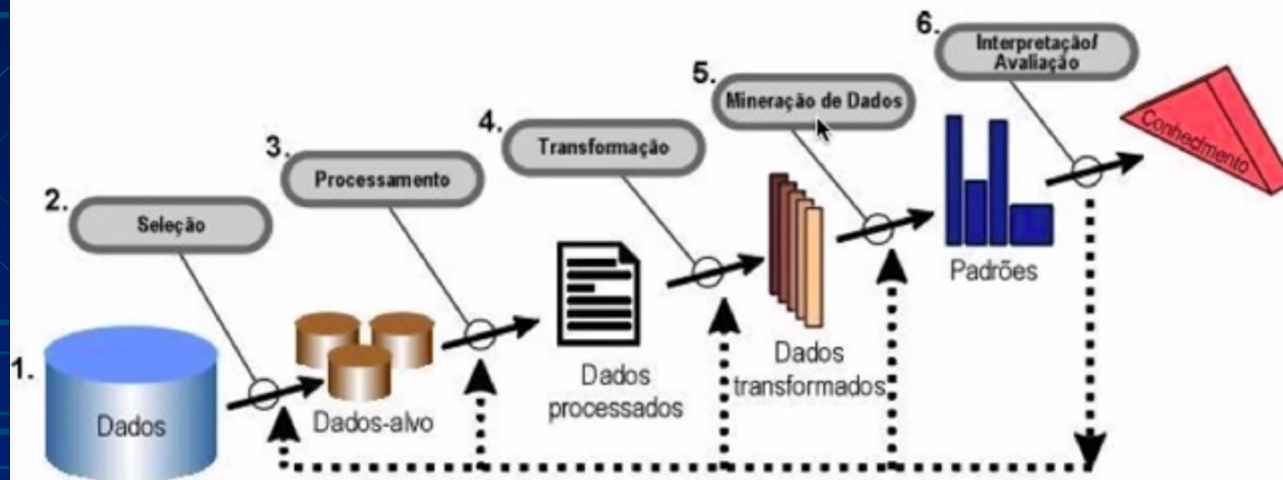
Breno Palma Miele Aniceto

# O que é data mining?

Data mining é formada por um conjunto de ferramentas e técnicas que através do uso de algoritmos de aprendizagem ou classificação baseados em redes neurais e estatística. Estes são capazes de explorar um conjunto de dados, extraíndo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento.



# Etapas do Processo




# Para que serve?


Buscamos com a MD (Mineração de Dados) conhecer melhor os clientes, seus padrões de consumo e motivações. Data Mining resgata em organizações grandes o papel do dono atendendo no balcão e conhecendo sua clientela.



# Veja algumas aplicações possíveis do data mining em diversos segmentos:

- Identificação do perfil dos possíveis compradores de determinado produto, para captação de clientes;
  - Decisão sobre alocação de produtos nas prateleiras de acordo com o perfil de consumo dos clientes;
  - Identificação de atividades suspeitas e prevenção a fraudes;
  - Análise de competências dos candidatos a uma vaga;
  - Previsão de volume de pacientes em determinada categoria.
- 
- A decorative graphic at the bottom of the slide consists of a network of thin, light-blue lines forming a circuit-like pattern. These lines connect to small, solid blue dots of varying sizes, scattered across the lower half of the image.

# Algumas ferramentas:

- Redes neurais
  - Indução de regras
  - Árvores de decisão
  - Visualização
- 

An abstract graphic on the left side of the slide, consisting of a complex network of light blue lines and dots on a dark blue background. The lines are of varying thickness and form a circuit-like pattern, with some dots appearing as small glowing points.

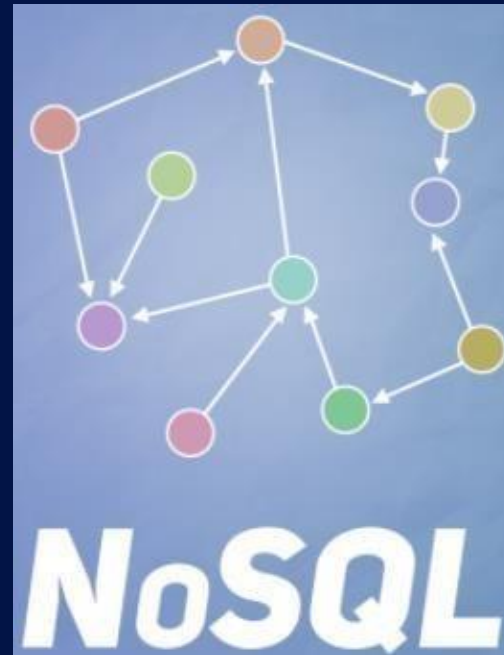
06

# NoSQL

Carlos Livius da Silva

# O que é?

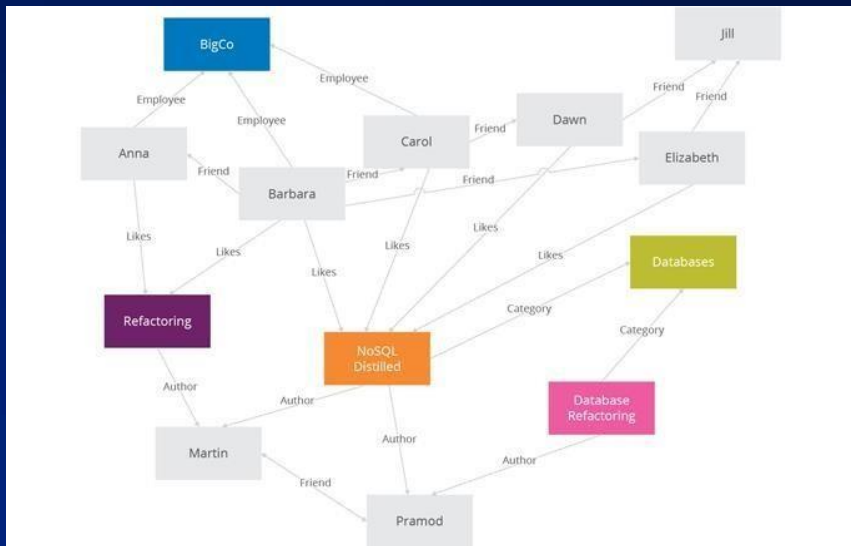
- NoSQL - Banco de Dados Não Relacional
- Para que serve?
- Como surgiu?



# Quais seus tipos e funcionamentos?

Bancos  
de dados gráficos:

Banco de dados  
Documentos:

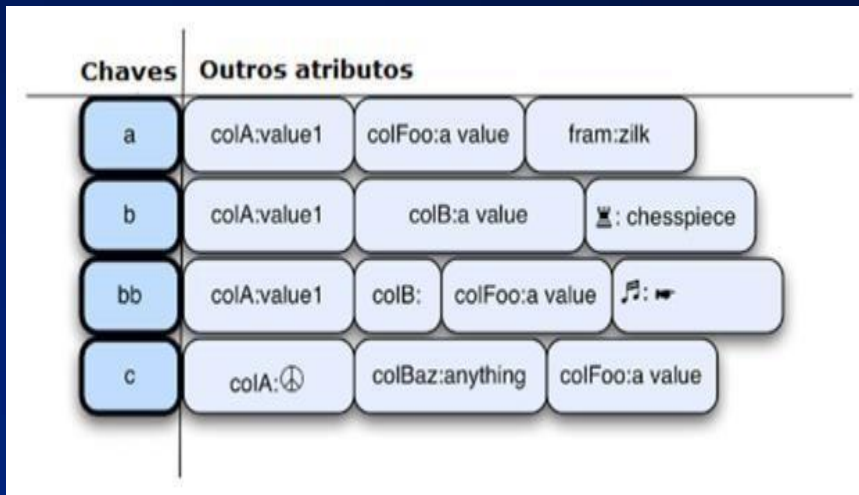


```
{
  "id": 55,
  "Pais": "Brasil",
  "Regiao": "América do Sul",
  "Populacao": 201032714,
  "PrincipaisCidades": [
    {
      "NomeCidade": "São Paulo",
      "Populacao": 1182876,
    },
    {
      "NomeCidade": "Rio de Janeiro",
      "Populacao": 6323037,
    }
  ]
}
```

# Quais seus tipos e funcionamento?

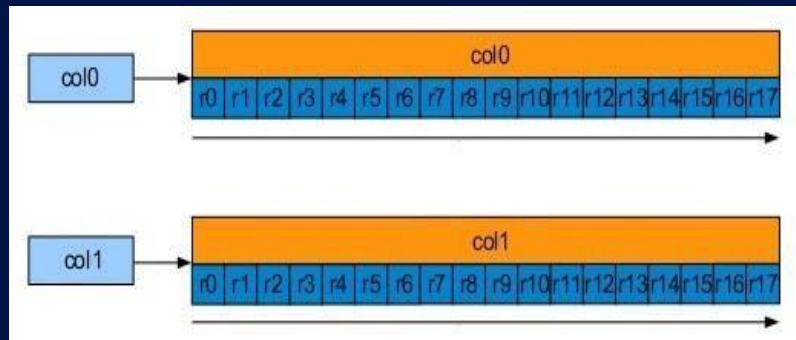
Bancos de dados

Chave-Valor:



Banco de dados

Coluna:



# SQL vs NoSQL

SQL

Table:

	id	name	createdAt	updatedAt
	Filter	Filter	Filter	Filter
1.	1	Ashton	2018-08-12 1...	2018-08-12 1...
2.	2	Jokeem	2018-08-12 1...	2018-08-12 1...
3.	3	Uma	2018-08-12 1...	2018-08-12 1...

NoSQL →



Users Collection



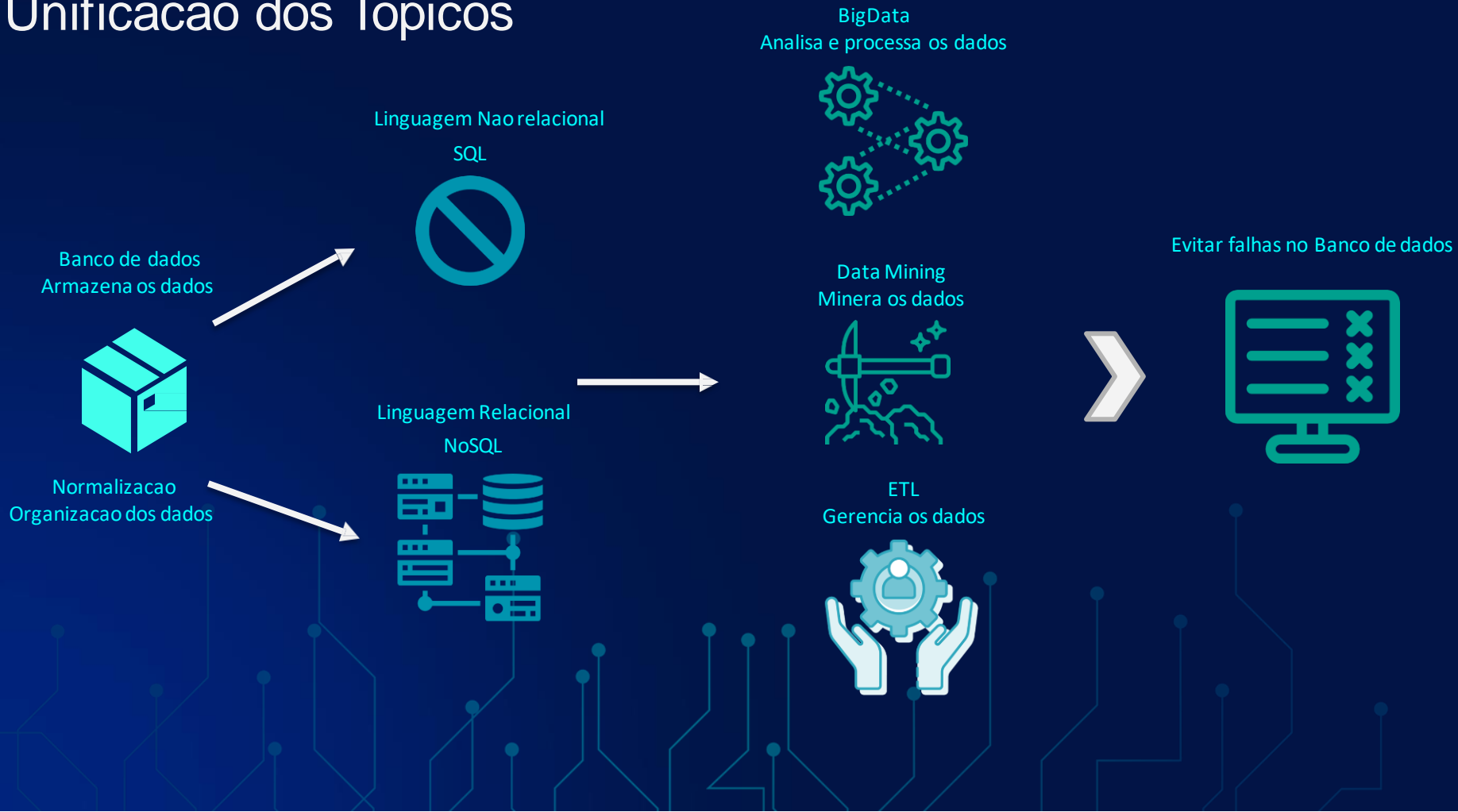
# Vantagens

- Escalabilidade
- Altamente funcional
- Alta performance
- Flexibilidade

# Desvantagens

- Recente
- Pouco reconhecido
- Suporte técnico
- Difícil instalação

# Unificacao dos Topicos



# Referências

- ALVES, Gustavo Furtado de Oliveira. **O que é um banco de dados?**. DICAS DE PROGRAMAÇÃO. Disponível em: <https://dicasdeprogramacao.com.br/o-que-e-um-banco-de-dados/>. Acesso em 14 set 2020
- FIA. **Big Data: o que é, como aplicar, a importância e exemplos**. Disponível em: <<https://fia.com.br/blog/big-data/amp/>>. Acesso em: 13 set. 2020
- ORACLE. **O Que é Big Data?**. Disponível em: <<https://www.oracle.com/br/big-data/what-is-big-data.html>>. Acesso em: 07 set. 2020
- NASCIMENTO, R. **Afinal, o que é Big Data?**. Disponível em: <<http://marketingpordados.com/analise-de-dados/o-que-e-big-data-%f0%9f%a4%96/>>. Acesso: 10 set. 2020
- PATEL, N. **Big Data: O Que é, Ferramentas e Como Aplicar na Sua Empresa**. Disponível em: <<https://neilpatel.com/br/blog/big-data-o-que-e/>>. Acesso em: 07 set 2020
- Josué. "O que é DATA WAREHOUSE? Você precisa de um?". Youtube. 26 de set. de 2017. <<https://www.youtube.com/watch?v=Q81zwSmaJo0>>. Acessado em 11 de set. de 2020
- MACORATTI, José Carlos. MACORATTI.NET. **SQL – Consultas T-SQL básicas**. Disponível em: <[http://www.macoratti.net/14/05/sql\\_tsq3.htm](http://www.macoratti.net/14/05/sql_tsq3.htm)>. Acesso: em 12 set 2020
- DEVMEDIA. **Guia completo de SQL**. Disponível em <<https://www.devmedia.com.br/guia-completo-de-sql/38314>>. Acesso: em 11 set 2020

# Referências

- [MICROSOFT. CREATE PROCEDURE \(Transact - SQL\)](https://docs.microsoft.com/pt-br/sql/t-sql/statements/create-procedure-transact-sql?view=sql-server-ver15). Disponível em <<https://docs.microsoft.com/pt-br/sql/t-sql/statements/create-procedure-transact-sql?view=sql-server-ver15>>. Acesso em: 11 set 2020
- Cetax. "DATA MINING: O QUE É, CONCEITO E DEFINIÇÃO". Blog Cetax. abr. 2016. Disponível em: <<https://www.cetax.com.br/blog/data-mining/>> . Acesso em 10 de set. de 2020
- AMAZON . **O que é NoSQL?**. Disponível em: < <https://aws.amazon.com/pt/nosql/> >. Acesso em: 15 set 2020
- GEEKHUNTER, Penélope Victória. **Banco de dados NoSQL: um manual prático e didático**. Disponível em: < [https://blog.geekhunter.com.br/banco-de-dados-nosql-um-manual-pratico-e-didatico/#O\\_que\\_sao\\_bancos\\_de\\_dados\\_NoSQL](https://blog.geekhunter.com.br/banco-de-dados-nosql-um-manual-pratico-e-didatico/#O_que_sao_bancos_de_dados_NoSQL) > Acesso em: 14 set 2020
- CIÊNCIAS E DADOS, Davi Matos. **Top 6 NoSQL Databases**. Disponível em: < <http://www.cienciaedados.com/top-6-nosql-databases> > Acesso em: 15 set 2020
- MONGODB, Lauren Schaefer. **What is NoSQL**. Disponível em: < <https://www.mongodb.com/nosql-explained> > Acesso em: 14 set 2020
- MARQUES FERNANDES, Henrique Marques Fernandes. **Banco de Dados Relacional (SQL) e Não Relacional (NoSQL) – O que são, para que servem e qual a diferença?**. Disponível em: < <https://marquesfernandes.com/tecnologia/banco-de-dados-relacional-sql-e-nao-relacional-nosql-o-que-sao-para-que-servem-e-qual-a-diferenca/> > Acesso em: 15 set 2020
- Sehn Körting. Thales. "How data mining works". Youtube. set. 2014. Disponível em: < <https://www.youtube.com/watch?v=W44q6gszdgY&t=125s> >