
FIA/P GRADUAÇÃO





✉ proftadeu.kanashiro@fiap.com.br

TADEU KANASHIRO

PROFESSOR
E GERENTE DE ENG. ANALÍTICA

- **Experiência profissional:** Mais de 15 anos de atuação na área de Tecnologia, focado em análise e estruturação de dados.
- **Áreas de atuação:** Big Data, Analytics Engineering, BI e Database Marketing.
- **Segmentos:** Saúde, setor imobiliário, farmacêutica, fintech, financeiro, Internet, telecomunicações, educação e filantropia.
- **MBA:** Big Data (Data Science).
- **Graduação:** Sistemas de Informações e Gestão Financeira.

AULA 0

INTRODUÇÃO AO **CONTEXTO** **DE DADOS** E À LINGUAGEM **SQL**



A economia digital



A jornada do dado

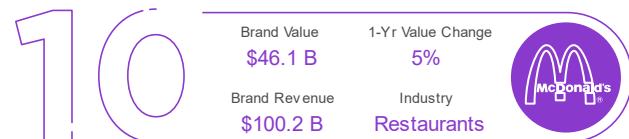
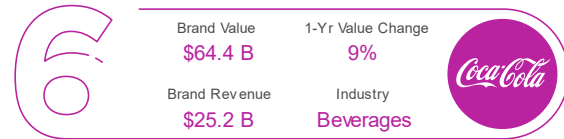
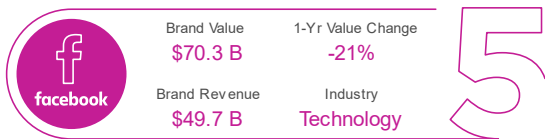
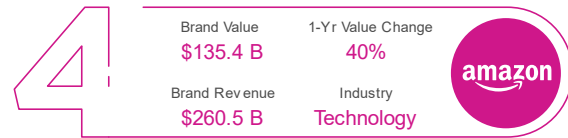
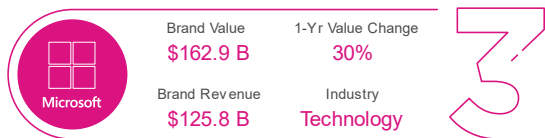
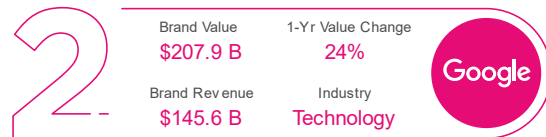
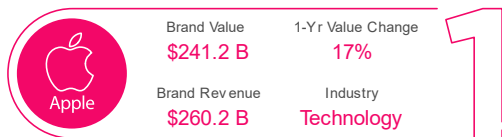


Introdução à Linguagem SQL



A ECONOMIA **DIGITAL**





The
2019
World's Most
Valuable
Brands

AS MARCAS

MAIS VALIOSAS DO MUNDO

- Lista da Forbes das 100 marcas mais valiosas do mundo (2019).
- Seis são empresas de tecnologia.
- A maioria dos produtos do Google e do Facebook é gratuita.

FONTE: <https://www.forbes.com/powerful-brands/list/> (FORBES, 2019)



2017, THE ECONOMIST



The Economist

MAY 6TH-12TH 2017

Crunch time in France

Ten years on: banking after the crisis

South Korea's unfinished revolution

Biology, but without the cells

The world's most valuable resource



Data and the new rules
of competition

WALK OUT
WITH NO CHECKOUT
**JUST
WALK
OUT**
SHOPPING

amazon



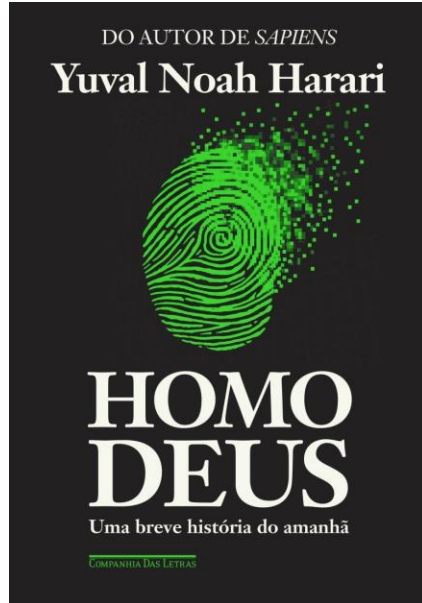
...Assim que tive **oportunidade**,
fui logo conhecer uma **loja**...



POSE ESTIMATION



DATAÍSMO



“Facebook or Google will be able to understand you better than you understand yourself because they will have the data... understand exactly how you feel and why you feel the way that you feel...”

FONTE: <https://www.youtube.com/watch?v=4ChHc5jhZxs> (HARARI, 2017)



EXEMPLO

DE DATAÍSMO



- Um **algoritmo** estimou que Angelina Jolie tinha um risco de **87%** de desenvolver **câncer** de mama.
- Em 2013, Angelina Jolie decide se submeter a uma **mastectomia** dupla.
- “**Important decisions** about your health, about your body, during your lifetime, will not be taken by you on the basis of your feelings, they **will be taken by algorithms**” (HARARI, 2017).

FONTE: <https://www.youtube.com/watch?v=4ChHc5jhZxs> (HARARI, 2017)



CULTURA DE DATA-DRIVEN

- “Building a data-driven enterprise **is not just** about encouraging the use of **data in decision-making...**”
- “**With this data**, or this type of insight, how could we fundamentally **change** the value **propositions** for our **customers?**”
- “How can we deliver **new value** propositions?”

FONTE: Build a Data-Driven Enterprise (GARTNER, 2018)

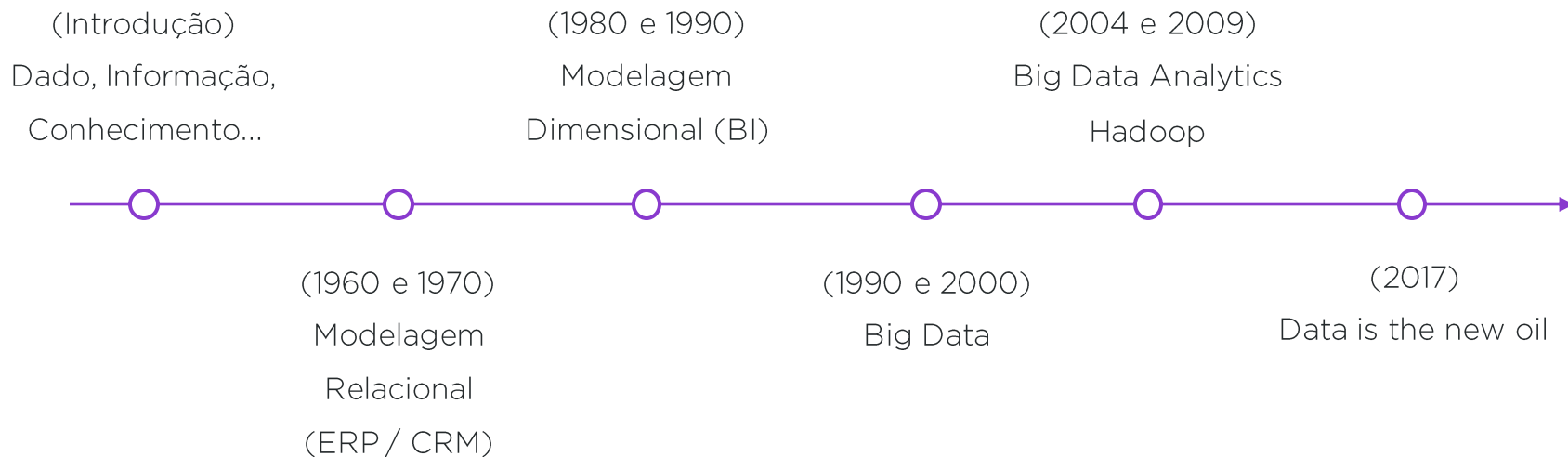


A JORNADA **DO DADO**



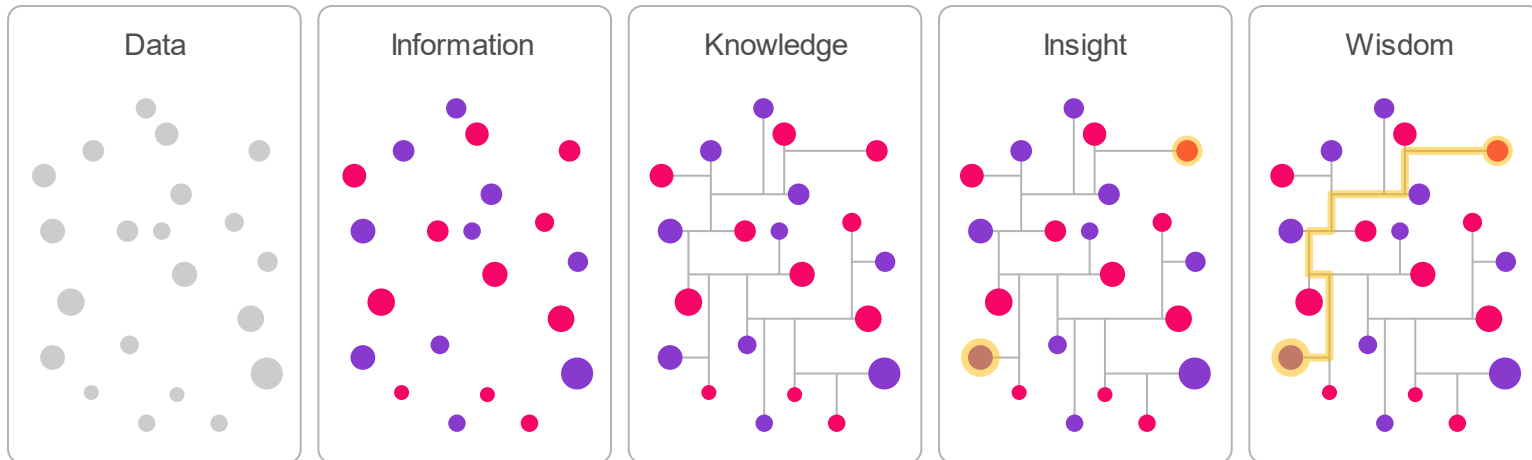
“ORDEM CRONOLÓGICA”

DO DADO

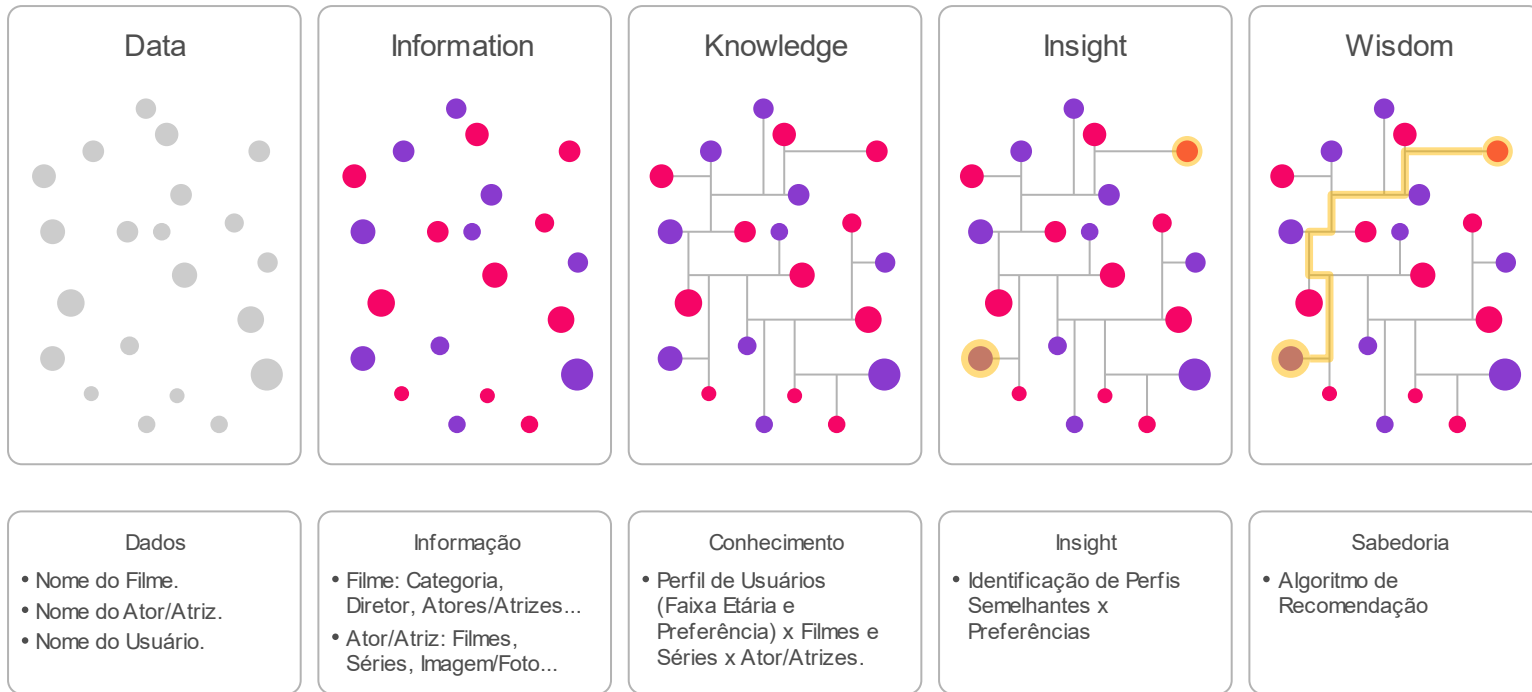


INTRODUÇÃO:

DADO, INFORMAÇÃO E CONHECIMENTO.



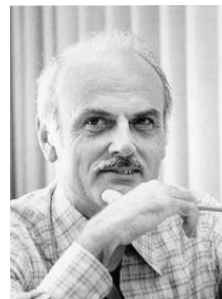
EXEMPLO NETFLIX



MODELAGEM RELACIONAL (ERP / CRM)

(1960 – 1970)

- Ted Codd e Peter Chen (1960 ~ 1970)
- Foco: Garantir as transações do sistema (CRUD)
- Normalização dos Dados (Transacionais)
- 1FN: Remover os Atributos Multivalorados e Compostos
- 2FN: Remover Dependências Parciais
- 3FN: Remover Dependências Transitivas



NORMALIZAÇÃO DOS DADOS

(1960 – 1970)

TABELA: CLIENTE				
CÓDIGO	NOME	SEXO	D_NASC	TELEFONE
0001	JOÃO	M	20/02/1950	(11) 5050-4040
0001	JOÃO	M	20/02/1950	(11) 9-7878-9090
0002	MARIA	F	19/08/1967	(11) 3535-1111
0003	JOSÉ	M	23/12/1970	(11) 5055-6060
0003	JOSÉ	M	23/12/1970	(11) 9-2222-3535
0003	JOSÉ	M	23/12/1970	(11) 9-7676-3434
0004	JOANA	F	26/03/1968	(11) 5040-7766

TABELA: CLIENTE			
CÓDIGO	NOME	SEXO	D_NASC
0001	JOÃO	M	20/02/1950
0002	MARIA	F	19/08/1967
0003	JOSÉ	M	23/12/1970
0004	JOANA	F	26/03/1968

TABELA: TELEFONE		
CÓDIGO	COD_CLIENTE	TELEFONE
1	0001	(11) 5050-4040
2	0001	(11) 9-7878-9090
3	0002	(11) 3535-1111
4	0003	(11) 5055-6060
5	0003	(11) 9-2222-3535
6	0003	(11) 9-7676-3434
7	0004	(11) 5040-7766

MODELAGEM DIMENSIONAL (BI)

(1980 – 1990)

- Inmon e Kimball (1980 ~ 1990)
- Foco: Retornar as consultas informacionais para a tomada de decisão
- Modelo Estrela e Modelo Floco de Neve
- Fatos (Indicadores e Medidas)
- Dimensões (Atributos e Hierarquia)



ERP / RDS

TABELA: ESTADO
TABELA: MUNICIPIO
TABELA: BAIRRO
TABELA: TIPO_END
TABELA: LOGRADOURO
TABELA: CLIENTE
TABELA: TELEFONE
TABELA: TIPO_TEL

TABELA: PRODUTO
TABELA: ITEM_PEDIDO
TABELA: PEDIDO
TABELA: PDV

TABELA: CATEGORIA
TABELA: MARCA
TABELA: MODELO

DATA MART / DATA WAREHOUSE

TABELA: CLIENTE

TABELA: PRODUTO

TABELA: VENDEDOR

TABELA: VENDA

TABELA: PDV

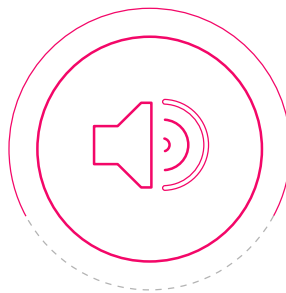
TABELA: PERIODO

TABELA: REGIÃO

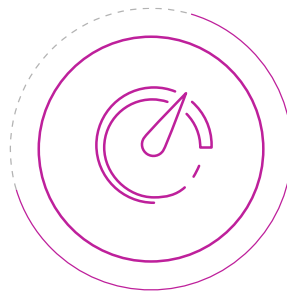


BIG DATA

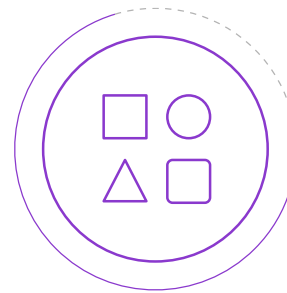
(1990 – 2000)



Volume



Velocity



Variety

“Os **3Vs** do Big Data...”

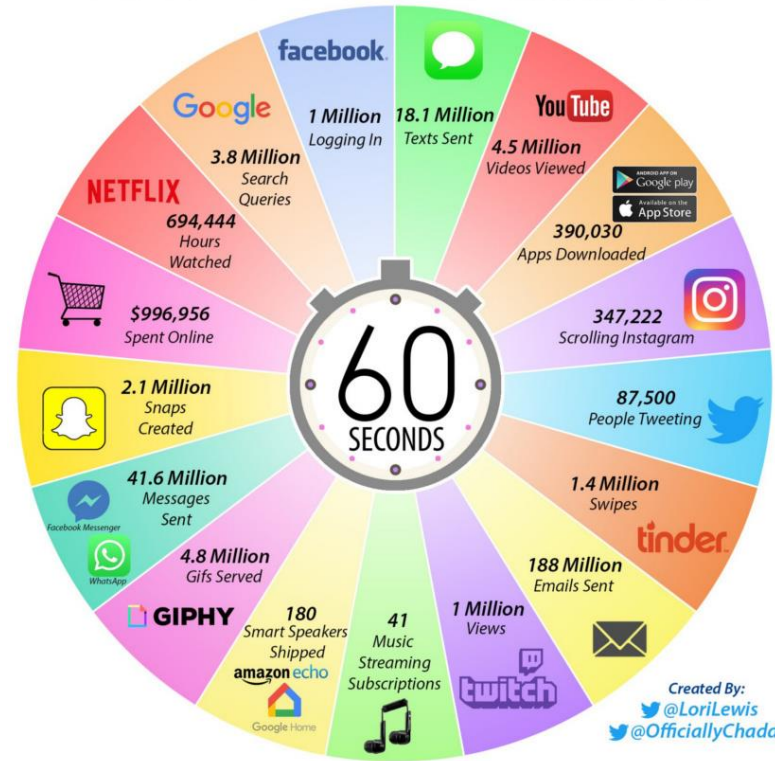
FONTE: <https://www.kdnuggets.com/2018/01/exclusive-interview-doug-laney-big-data-infonomics.html>



2018 *This Is What Happens In An Internet Minute*



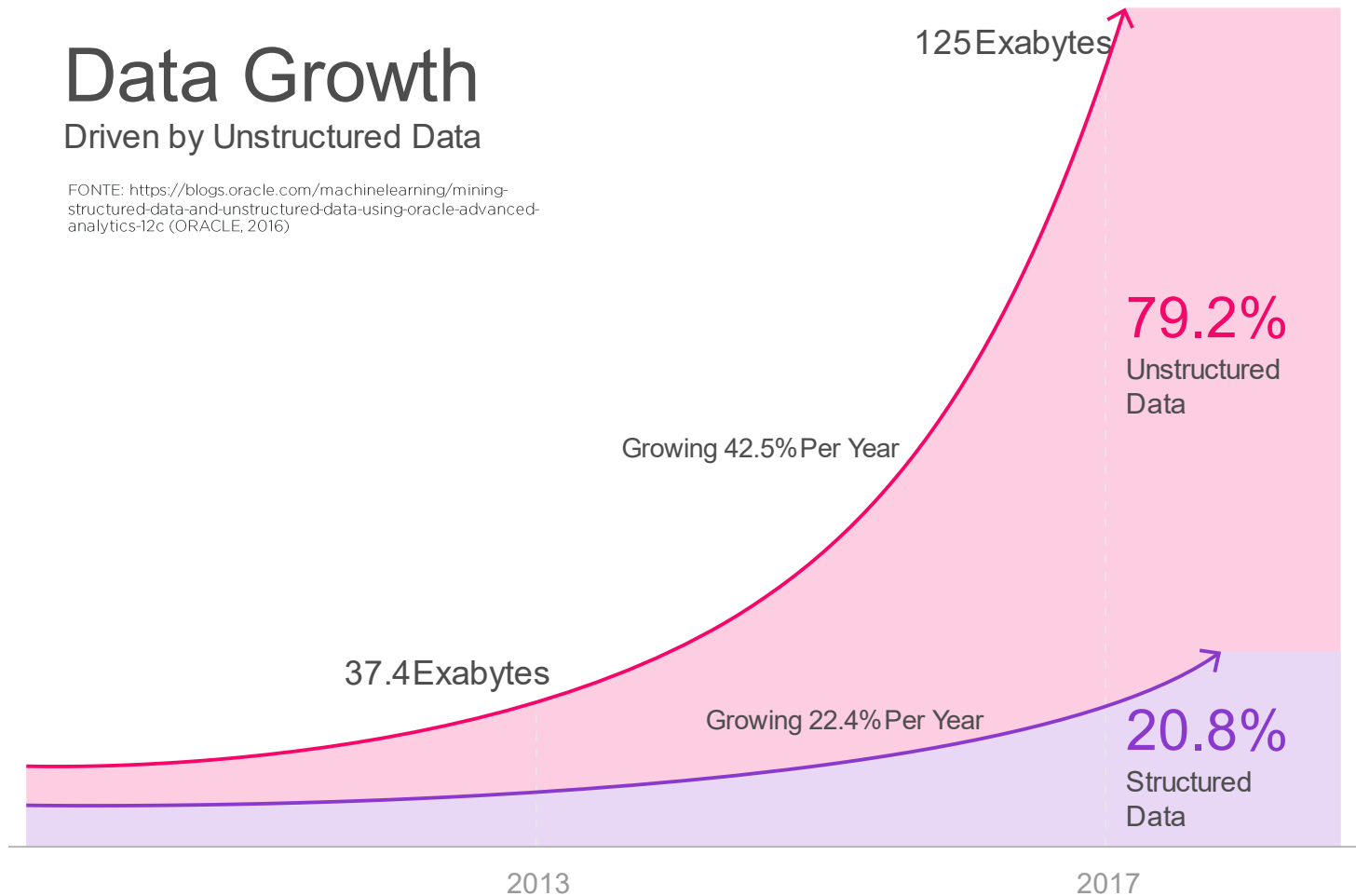
2019 *This Is What Happens In An Internet Minute*



Data Growth

Driven by Unstructured Data

FONTE: <https://blogs.oracle.com/machinelearning/mining-structured-data-and-unstructured-data-using-oracle-advanced-analytics-12c> (ORACLE, 2016)



BIG DATA ANALYTICS

(2004 – 2009)

- Clive Humby (2006)
- “Dados são o novo petróleo...”



“...É valioso, **MAS se não for refinado, não pode ser realmente utilizado**. Ele precisa ser transformado em gás, plástico, produtos químicos etc. para criar uma entidade valiosa que impulse a atividade lucrativa; os **dados DEVEM ser** divididos e **analisados para que tenham valor**.”

FONTE: <https://www.theguardian.com/technology/2013/aug/23/tech-giants-data>



BIG DATA ANALYTICS

(2004 – 2009)



“Noto que muitas pessoas pensam que um programa de **fidelidade** significa **tornar** os **clientes mais fiéis**. E isso não é verdade, não é isso o que acontece. O que acontece é que você começa a **entender** o **cliente**, oferecendo produtos que ele realmente quer e pelos quais ele realmente é **apaixonado**.”



FONTE: <https://www.youtube.com/watch?v=dd5kWLs-998>



IAN MACLAURIN (CEO DA TESCO)

(2004 – 2009)



“O que me assusta é que você, em três meses, conhece muito mais sobre os meus clientes, do que eu, em 30 anos.”





(2004)

- Google (2004)
- Apache (2006)
- Cloudera (2008)
- Doug Cutting (2009)
- Hortonworks (2011)
- Processamento Massivo Paralelo
 - **Definição:** É utilizar múltiplos (dois ou mais) processadores, simultaneamente, para executar uma determinada tarefa.
 - **Objetivo:** Aumento de desempenho para reduzir o tempo de execução de uma determinada tarefa.
 - Basicamente, consiste em quebrar as tarefas em partes menores.



QUEBRA DE PARADIGMAS

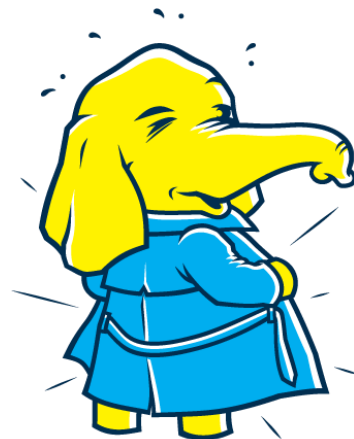
(2009)

Arquitetura Vertical

- Aumenta a memória e CPU;
- Há limitação de escalabilidade.

Arquitetura Horizontal

- Confiança menor no hardware e maior no software;
- Crescimento horizontal para escalar.



**MY HADOOP IS
BIGGER
THAN YOURS...**



ECOSSISTEMA

(2009)



BIG DATA WEEK 2019

SAIBA MAIS



2017, THE ECONOMIST



The Economist

MAY 6TH-12TH 2017

Crunch time in France

Ten years on: banking after the crisis

South Korea's unfinished revolution

Biology, but without the cells

The world's most valuable resource



Data and the new rules
of competition

INTRODUÇÃO **A LINGUAGEM SQL**



PESQUISA ANUAL DA O'REILLY

- 983 respostas.
- 45 países.
- Linguagens mais utilizadas pelos entrevistados.

70% SQL.

57% R.

54% Python.

2016 Data Science Salary Survey

Tools, Trends, What Pays (and What Doesn't)
for Data Professionals

John King & Roger Magoulas

O'REILLY



RANKING DOS BANCOS

MAIS POPULARES DB-ENGINES

- Menções em websites de busca (exemplos: Google e Bing).
- Frequência dos debates técnicos em fóruns (exemplos: Stack Overflow e DBA Stack Exchange).
- Número de ofertas de empregos e profissionais em redes sociais (exemplos: LinkedIn e Twitter).

360 systems in ranking, November 2020

Rank			DBMS	Database Model	Score		
Nov 2020	Oct 2020	Nov 2019			Nov 2020	Oct 2020	Nov 2019
1.	1.	1.	Oracle +	Relational, Multi-model f	1345.00	-23.77	+8.93
2.	2.	2.	MySQL +	Relational, Multi-model f	1241.64	-14.74	-24.64
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model f	1037.64	-5.48	-44.27
4.	4.	4.	PostgreSQL +	Relational, Multi-model f	555.06	+12.66	+63.99
5.	5.	5.	MongoDB +	Document, Multi-model f	453.83	+5.81	+40.64
6.	6.	6.	IBM Db2 +	Relational, Multi-model f	161.62	-0.28	-10.98
7.	↑ 8.	↑ 8.	Redis +	Key-value, Multi-model f	155.42	+2.14	+10.18
8.	↓ 7.	↓ 7.	Elasticsearch +	Search engine, Multi-model f	151.55	-2.29	+3.15
9.	9.	↑ 11.	SQLite +	Relational	123.31	-2.11	+2.29
10.	10.	10.	Cassandra +	Wide column	118.75	-0.35	-4.47
11.	11.	↓ 9.	Microsoft Access	Relational	117.23	-1.02	-12.84
12.	12.	↑ 13.	MariaDB +	Relational, Multi-model f	92.29	+0.52	+6.72
13.	13.	↓ 12.	Splunk	Search engine	89.71	+0.30	+0.64
14.	14.	↑ 15.	Teradata +	Relational, Multi-model f	75.60	-0.19	-4.75
15.	15.	↓ 14.	Hive	Relational	70.26	+0.71	-13.96
16.	16.	16.	Amazon DynamoDB +	Multi-model f	68.89	+0.48	+7.52
17.	17.	↑ 25.	Microsoft Azure SQL Database	Relational, Multi-model f	66.99	+2.59	+39.37
18.	18.	↑ 19.	SAP Adaptive Server	Relational	55.39	+0.23	+0.10
19.	19.	↑ 20.	SAP HANA +	Relational, Multi-model f	53.58	-0.66	-1.53
20.	↑ 21.	↑ 22.	Neo4j +	Graph	53.53	+2.20	+3.00
21.	↓ 20.	↓ 17.	Solr	Search engine	51.82	-0.66	-5.96
22.	22.	↓ 21.	HBase +	Wide column	47.11	-1.25	-6.73
23.	23.	↓ 18.	FileMaker	Relational	46.66	-0.73	-9.07
24.	24.	↑ 27.	Google BigQuery +	Relational	35.08	+0.67	+9.64
25.	25.	↓ 24.	Microsoft Azure Cosmos DB +	Multi-model f	32.50	+0.49	+0.52
26.	26.	↓ 23.	Couchbase +	Document, Multi-model f	30.55	+0.22	-1.44
27.	27.	↑ 28.	Memcached	Key-value	25.75	-0.35	+0.63
28.	28.	↑ 33.	InfluxDB +	Time Series	24.96	+0.81	+5.02
29.	29.	↑ 31.	Amazon Redshift +	Relational	22.99	-0.15	+1.21
30.	30.	↓ 26.	Informix	Relational, Multi-model f	22.57	-0.14	-3.62

FONTE: <https://db-engines.com/en/ranking> (DB-ENGINES, 2020)



SQL

HISTÓRIA DO SQL

- O nome original da linguagem era **SEQUEL** (Structured English Query Language) ou Linguagem de Consulta Estruturada.
- Desenvolvida pela **IBM**, no início dos anos **70**, como forma de **interface** para o sistema de banco de dados relacional.
- Contudo, o **primeiro SGBD**, lançado no mercado, com uma implementação da linguagem SQL, em 1979, foi o **Oracle** (que, na época, se chamava Relational Software).

PADRONIZAÇÃO DO SQL

- Em **1986**, o American National Standards Institute (**ANSI**) definiu um padrão para a linguagem SQL, homologando-a para trabalhar com bancos de dados relacionais.
- Em **1987**, foi também aceito pelo International Standards Organization (**ISO**).
- Após essas padronizações, diferentes **SGBDRs**, tais como Oracle, SQL Server, MySQL, DB2, entre outros, passaram a utilizar o **SQL como linguagem-padrão**.



PADRONIZAÇÃO DO SQL

- Embora o SQL seja um padrão, há diferenças nas implementações dessa linguagem nos diversos SGBDs do mercado.
- **Cada fabricante** de software **cria** extensões (com **sintaxes**, **funções**, **recursos** etc.) específicas de seus sistemas, formando diferentes “dialetos” e “versões” da linguagem.
- É possível escrever aplicações utilizando apenas o SQL padrão ANSI (com os comandos comuns à maioria dos SGBDs).
- Porém, na prática, muitas vezes, faz-se uso das sintaxes proprietárias para aproveitar recursos específicos ou conseguir um melhor desempenho.



EXEMPLO DE DIFERENCIAÇÃO DE SINTAXE

- Sintaxe para trazer a primeira linha da tabela TB_CLIENTE em dois SGBDs diferentes:

TB_CLIENTE			
CÓDIGO	NOME	SEXO	D_NASC
0001	JOÃO	M	20/02/1950
0002	MARIA	F	19/08/1967
0003	JOSÉ	M	23/12/1970
0004	JOANA	F	26/03/1968



SELECT TOP 1 * FROM TB_CLIENTE



SELECT * FROM TB_CLIENTE WHERE ROWNUM = 1



CARACTERÍSTICAS DA LINGUAGEM SQL

- As **instruções padronizadas** seguem a mesma nomenclatura e formato para os **diferentes** tipos de **SGBDs**, respeitando as particularidades de cada um.
- A **migração** de um SGBD para outro **não requer grandes mudanças**.
- Quando ocorre a migração de um SGBD, a adaptação dos profissionais é facilitada com a redução de tempo e de custos para treinamentos, pois as **instruções possuem nomes e funcionalidades iguais**.
- Portabilidade entre as plataformas.



CARACTERÍSTICAS DA LINGUAGEM SQL

- As queries (**consultas**) do SQL têm uma **estrutura semelhante** à linguagem natural (em **inglês**).
- As instruções de **manipulação de dados** sempre **trabalham com tabelas** e/ou retornam resultados em forma de tabelas (com **linhas** e **colunas**).
- É uma **linguagem não procedural**, isto é, as sentenças em SQL declaram a tarefa a ser realizada (“**o que**”) **sem** a necessidade de especificar os procedimentos passo a passo que o sistema deve seguir para executá-la (“**como**”).
- **Independência física dos dados** e dispositivos: **não especifica** o método de acesso aos dados nem **como estão armazenados** (os SGBDs possuem um “otimizador” que se encarrega da maneira de recuperar os dados).



SUBCONJUNTOS DA LINGUAGEM SQL

Os comandos da linguagem SQL podem ser divididos em grupos:

DML
Data Manipulation Language
Manipulação de Dados

DDL
Data Definition Language
Definição de Estruturas e Objetos

DCL
Data Control Language
Controle da Base de Dados



DML: DATA MANIPULATION LANGUAGE.

- Engloba os **comandos** que correspondem às quatro funções **essenciais** de **manipulação** de **dados**:

Inclusão: inserir linhas em tabelas (comando **INSERT**).

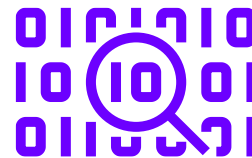
Atualização: de dados em tabelas (comando **UPDATE**).

Exclusão: de linhas das tabelas (comando **DELETE**).

Recuperação (consulta): retornar dados (comando **SELECT**).

- Inclui também comandos para controle de transação:

COMMIT e ROLLBACK, entre outros.



DDL: DATA DEFINITION LANGUAGE.

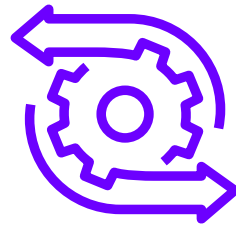
- Para **definição de estruturas** e objetos do banco de dados.
- Principais comandos:

Criação de estruturas de objetos (comando **CREATE**).

Alteração de estrutura de objetos (comando **ALTER**).

Eliminação de estrutura de objetos (comando **DROP**).

Exclusão física de linhas de tabelas (comando **TRUNCATE**).

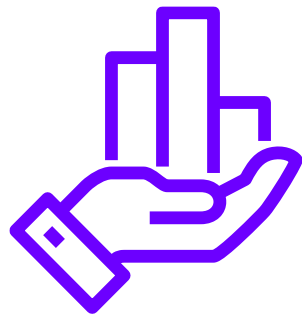


DCL: DATA CONTROL LANGUAGE.

- Para controle da base de dados.
- Basicamente, controle de acesso e permissão de usuários.
- Principais comandos:

Conceder privilégios aos usuários (comando GRANT).

Revogar privilégios dos usuários (comando REVOKE).



PRINCIPAIS COMANDOS

DDL	DML	DCL
CREATE ALTER DROP	INSERT UPDATE DELETE	GRANT REVOKE
RENAME	SELECT	
TRUNCATE	COMMIT ROLLBACK	



BIBLIOGRAFIA BÁSICA

- **TAYLOR**, A. SQL para Leigos. Porto Alegre: Bookman, 2016.
- **PUGA**, S; **FRANÇA**, E; **GOYA**, M. Banco de Dados: Implementação em SQL, PL/SQL e Oracle 11g. São Paulo: Pearson Universidades, 2013.
- **LIVE SQL**. Learn and share SQL: Running on Oracle Database 19c. Disponível em: <<https://livesql.oracle.com/apex/f?p=590:1000>>. Acesso em: 23 nov. 2020.





Copyright © 2024 Prof. Tadeu Kanashiro e Prof. André Santos

Todos direitos reservados. Reprodução ou divulgação total ou parcial deste documento é expressamente proibido sem o consentimento formal, por escrito, do Professor (autor).

