

Notas de Leitura em Sistemas de Bases de Dados

# 01 >> Dados, Informação e Conhecimento

Orlando Belo

Departamento de Informática, Escola de Engenharia, Universidade do Minho  
PORTUGAL

2020

01



# Resumo

Nesta unidade de ensino iniciaremos o estudo aos **sistemas de dados**, contextualizando a sua **emergência e utilidade no mundo real**. Além disso, abordaremos também alguns dos conceitos mais fundamentais que todos aqueles que estudam ou trabalham no domínio dos sistemas de bases de dados devem conhecer e saber utilizar. Em particular, apresentaremos e discutiremos conceitos como **dados, informação e conhecimento**. Numa fase mais avançada, trabalharemos o domínio dos **sistemas de gestão de bases de dados**, dando particular ênfase a aspetos relacionados com as estruturas de armazenamento de dados com que estes sistemas trabalham: as **bases de dados**.



# Conteúdo

- Introdução.
- O “cerco” da informação.
- Dados, informação e conhecimento.
- Bases de Dados.
- Sistemas de Gestão de Bases de Dados.



“Information has gone from scarce to superabundant. That brings huge new benefits, but also big headaches”

Kenneth Cukier, in “Data, data everywhere – a special report on managing information”,  
The Economist, February 27th, 2010



# Introdução

- Tudo o que nos rodeia contribui com algum tipo de **informação** que nos ajuda a desenvolver as nossas ações e tarefas mais mundanas.
- A forma como a informação é ou pode ser utilizada depende muito da **forma como se apresenta e da forma como sabemos lidar** com ela.
- Um mesmo elemento de informação poderá ter **interpretações e utilizações diferentes em contextos diferentes por pessoas diferentes.**



# O “Cerco” da Informação

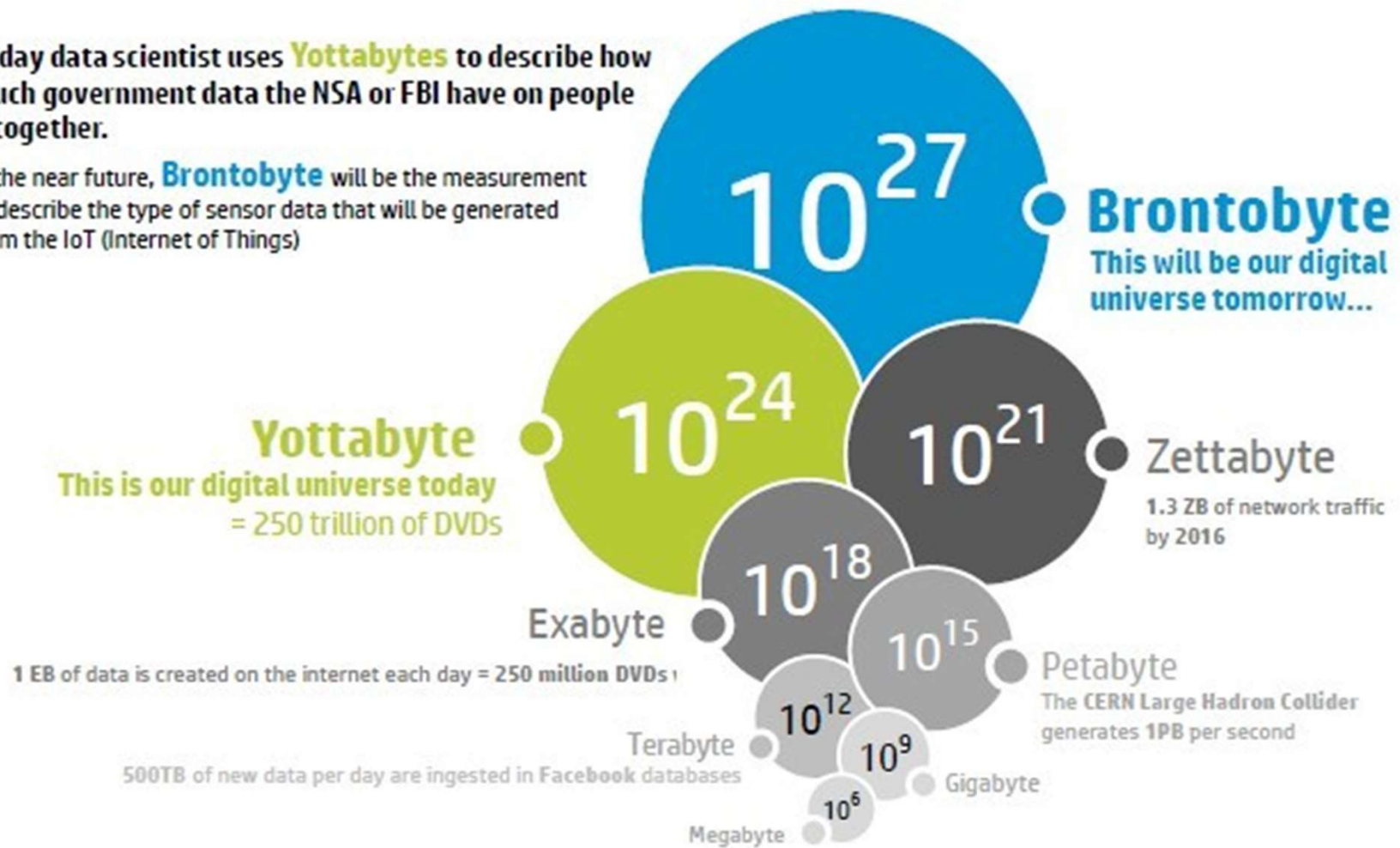
- Facilmente constatamos que em **qualquer lugar e em qualquer altura** estamos “cercados” por um **volume enorme de elementos de informação**.
- Porém, nós **sabemos ser seletivos**, escolhendo aquilo que nos interessa num dado contexto e numa determinada altura.
- O importante é que se consiga no meio desta “panóplia enorme” de elementos de informação **aceder e utilizar da melhor maneira a informação** que mais se adequa às nossas necessidades do momento.

...mas isso não é fácil. Mesmo nada fácil.



Today data scientist uses **Yottabytes** to describe how much government data the NSA or FBI have on people altogether.

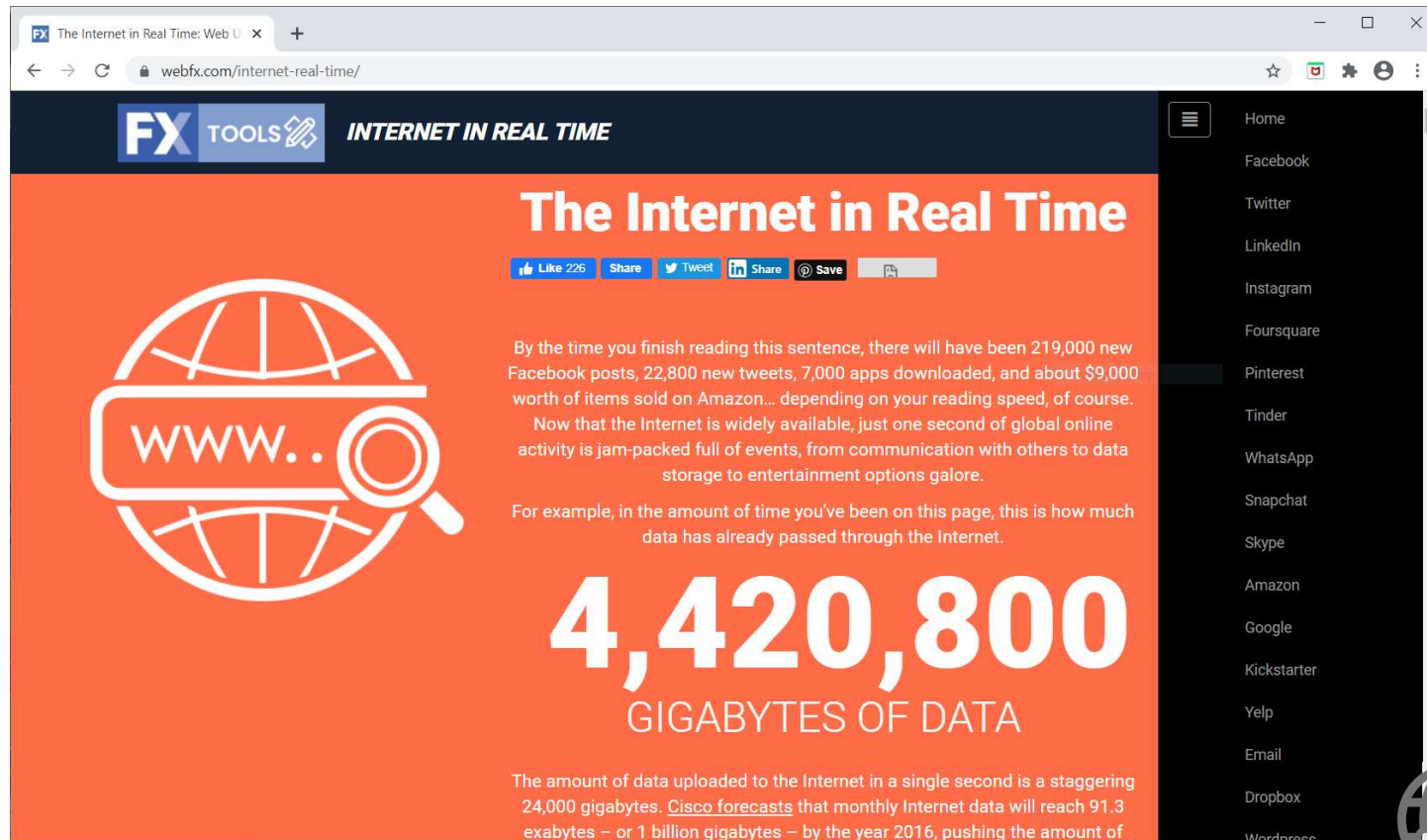
In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



Fonte: [http://www.theregister.co.uk/2012/12/04/hp\\_discover\\_autonomy\\_vertica\\_big\\_data/](http://www.theregister.co.uk/2012/12/04/hp_discover_autonomy_vertica_big_data/)



# Uma Curiosidade em Tempo Real



Fonte: <https://www.webfx.com/internet-real-time/>



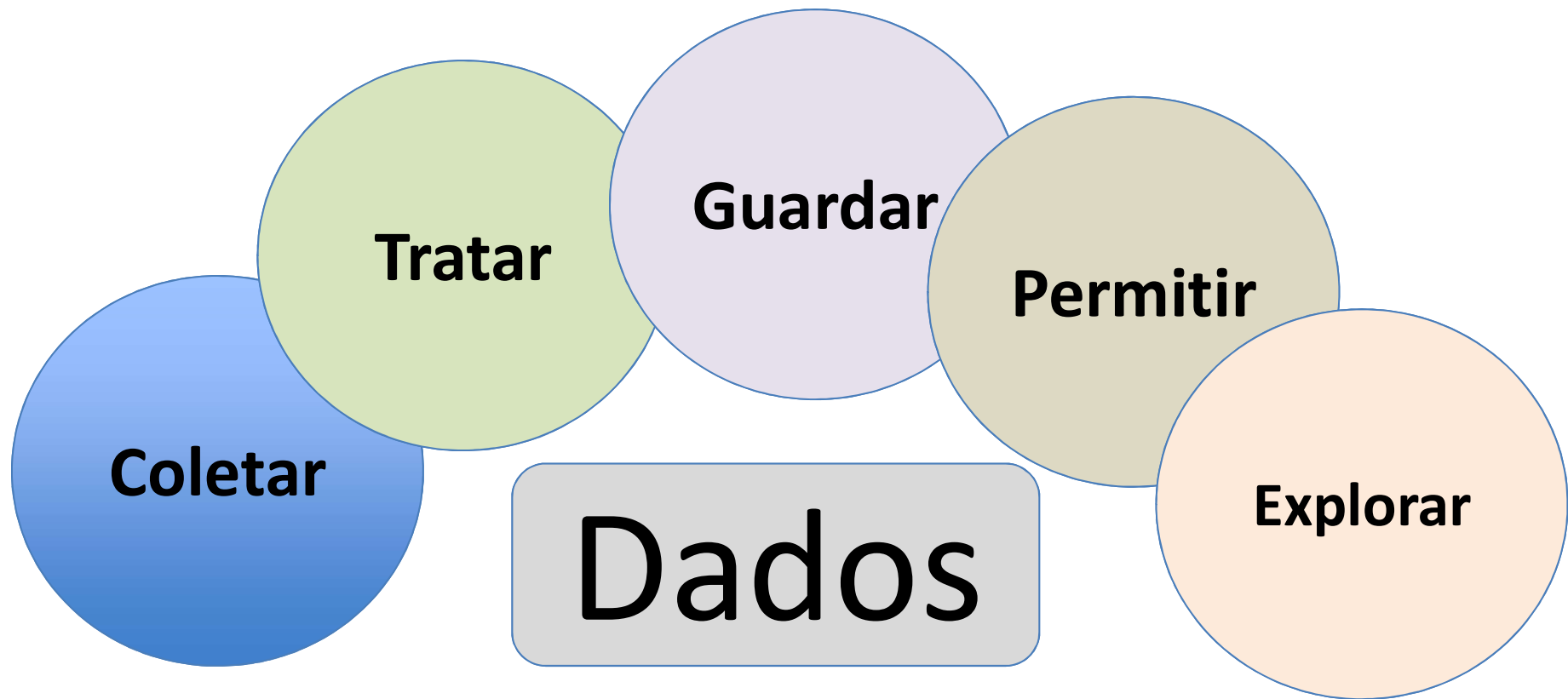


# Os Dados

- Os dados relativos a um dado problema ou situação devem ter algum tipo de “**arranjo**” para que a sua gestão e exploração possam ser efetivas e contribuïrem para os tornar **disponíveis e com utilidade**.
- Este “arranjo” deve ser o resultado, de alguma forma, de uma ponderação dos vários objetivos relacionados com os dados que **vamos coletar, tratar, armazenar e explorar**.  
(Obs.: não esquecer, obviamente, os aspetos financeiros associados)



# Tarefas Sobre os Dados



# Um Dashboard



# Conceitos Base

## Sabedoria

*Entendimento dos efeitos e resultados da aplicação de conhecimento.*

## Conhecimento

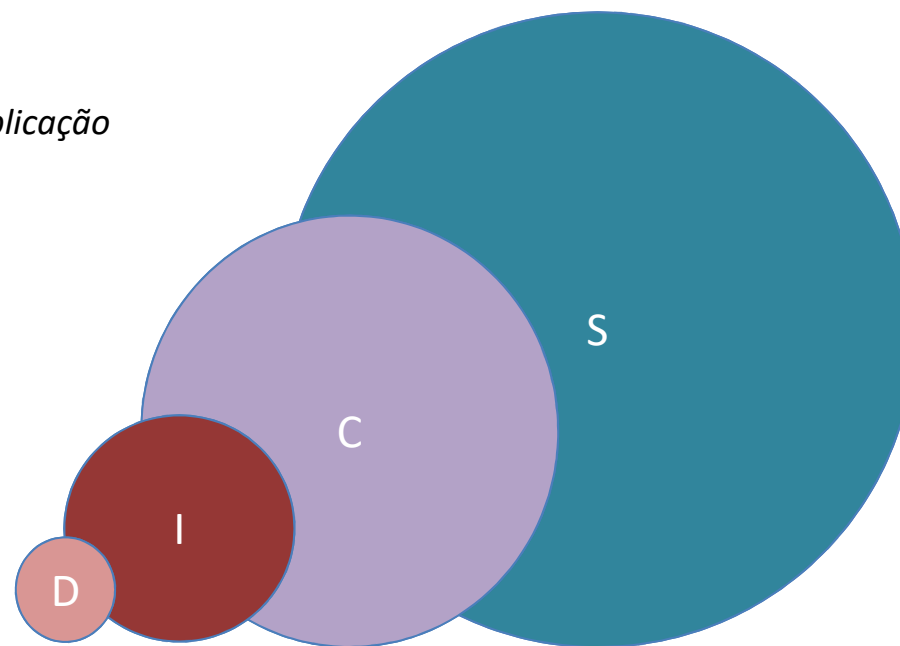
*Aplicação da informação que se possui.*

## Informação

*Relacionamento dos dados disponíveis.*

## Dados

*Factos.*



# Conceitos Base - Dados

- **Dados** - *Data* – um conjunto de factos discretos acerca de um dado conjunto de eventos, que não apresentam qualquer tipo de estrutura ou organização em particular; são os elementos mais básicos da informação – os seus “átomos”.
  - Hoje venderam-se 156 guarda-chuvas.
  - O livro de ‘Bases de Dados’ tem 750 páginas.



# Conceitos Base - Informação

- **Informação** - *Information* – aparece ao se definir o contexto dos dados e costuma responder a questões como ‘O quê?’, ‘Porquê?’, ‘Onde?’, ‘Quando?’ ou ‘Quem?’; pode ser vista como um pacote de dados estruturados que lhe acrescenta significado, dando-lhes um contexto e significado.
- Apesar de uma manhã radiosa, hoje à tarde, pelas 15:00, de forma repentina, choveu imenso.
- Para o teste de ‘Bases de Dados’ só é necessário estudar os primeiros 21 capítulos.



# Conceitos Base - Conhecimento

- **Conhecimento** – *Knowledge* – aplicação de experiência adquirida e observações acerca da informação disponível permitem responder às questões ‘Como?’
  - As pessoas quando saem de casa com sol, raramente trazem consigo um guarda-chuva.
  - Se entendermos bem a matéria dos capítulos 5,6 e 7, não precisamos de estudar os capítulos 12, 13 e 18.



# Conceitos Base - Sabedoria

- Sabedoria - *Wisdom* – aborda as questões ‘Porquê?’ utilizando de forma mais profunda conhecimento contextual
  - Mesmo com estas condições meteorológicas não vale a pena aumentar o stock de guarda-chuvas de forma extraordinária. Estamos no verão.
  - Dados os objetivos da unidade curricular de ‘Bases de Dados’, o professor não colocará no teste as matérias dos capítulos 20 e 21. Tais matérias integram-se usualmente na área de ‘Sistemas Distribuídos’.





# Conceitos Base

- Apesar das muitas tentativas realizadas até hoje na definição dos conceitos ‘Dados’, ‘Informação’, ‘Conhecimento’ e ‘Sabedoria’, ainda não temos uma visão completa e clara acerca do que cada um deles e daquilo que envolve, bem como o tipo de relacionamentos que mantêm entre si.
- As fronteiras das definições de cada um destes conceitos não estão ainda bem definidas, tendo zonas, digamos, bastante difusas.



# As Bases de Dados

*“A database is an organized collection of data. It is the collection of schemes, tables, queries, reports, views and other objects. The data is typically organized to model aspects of reality in a way that supports processes requiring information, such as modelling the availability of rooms in hotels in a way that supports finding a hotel with vacancies.”*

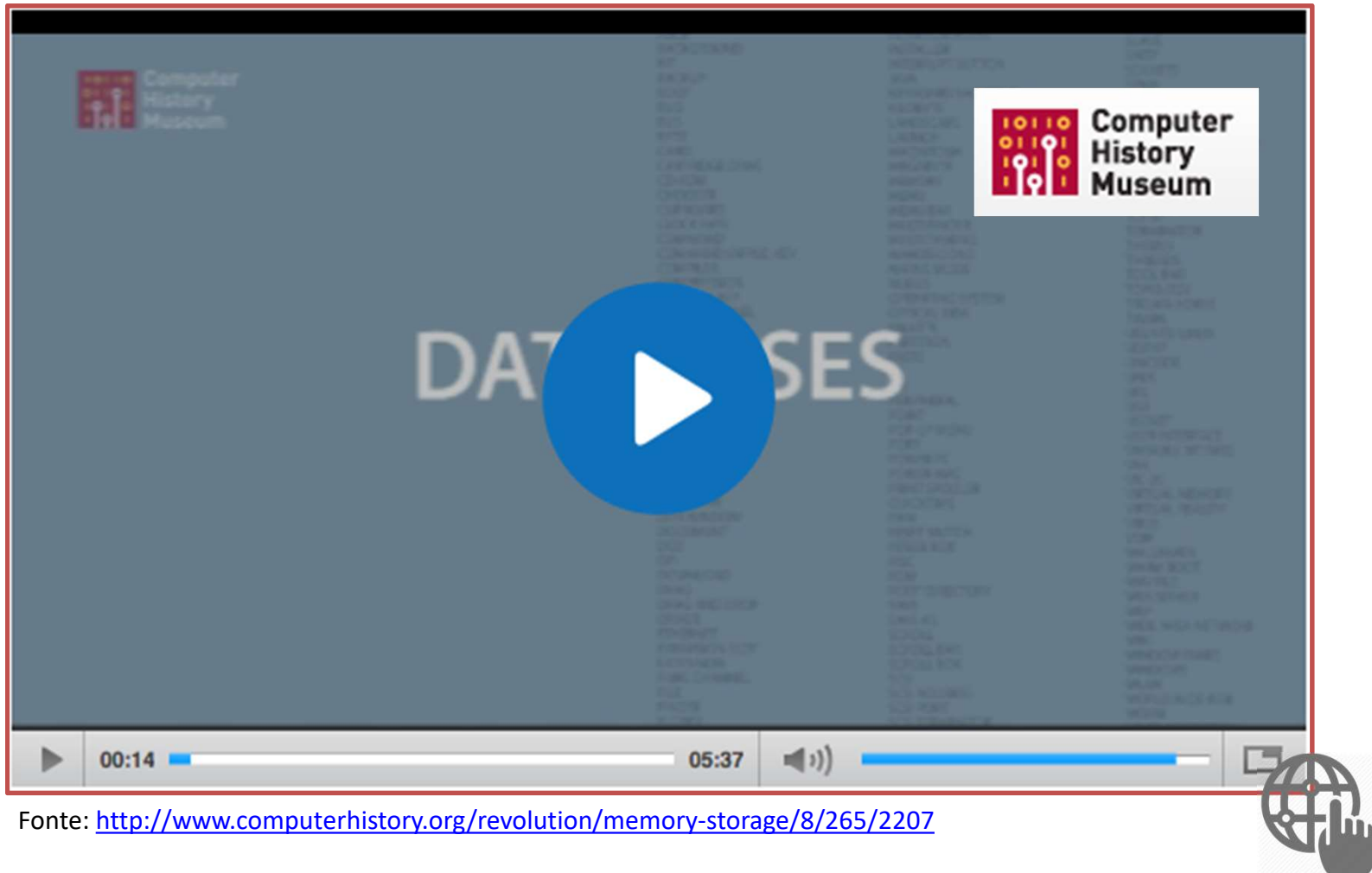
Fonte: <https://en.wikipedia.org/wiki/Database>



# As Bases de Dados



# A História das Bases de Dados



Fonte: <http://www.computerhistory.org/revolution/memory-storage/8/265/2207>



# Um “Pequeno” Problema

- Como é que um aluno de um curso guarda e gere os seus diversos elementos de estudo (livros, apontamentos, blocos de exercícios, etc.) ao longo dos vários semestres do seu curso?



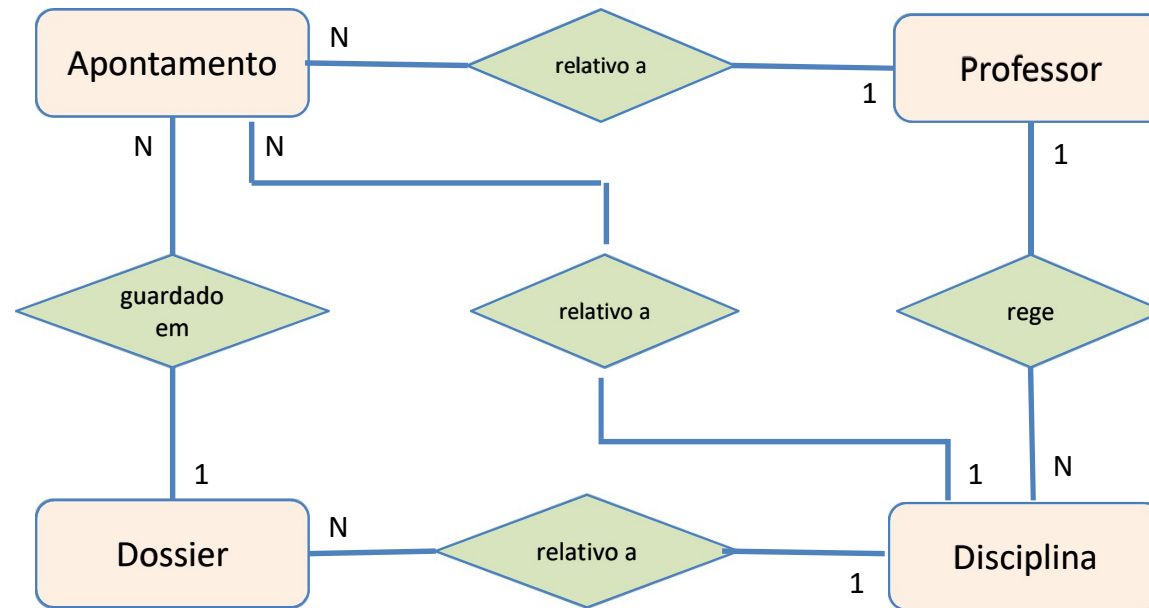
# Alguns Requisitos

- Numa breve **análise ao problema** em causa “o aluno” decidiu enumerar alguns dos **requisitos** que queria ver satisfeitos pela sua futura bases de dados, nomeadamente:
  - (R1) Os apontamentos tirados em cada aula de uma disciplina deveria ter um registo próprio, no qual constasse, por exemplo, uma referência à disciplina, ao professor, à data de realização, entre outras coisas.
  - (R2) Cada um desses registos deveria também estar em que dossier é que os apontamentos de cada disciplina são guardados.
  - (...)

Requisitos?



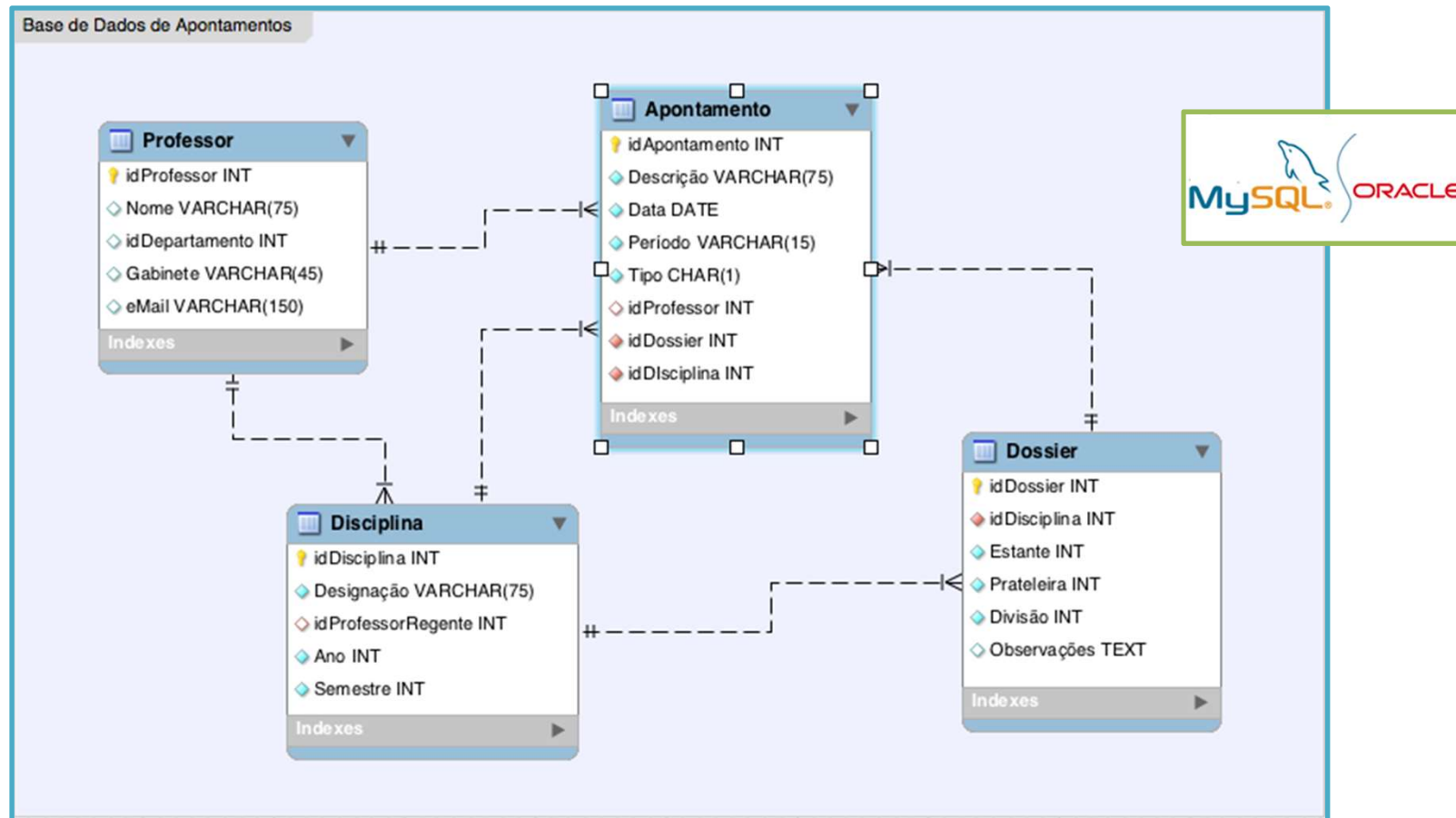
# Um Possível Modelo



Um Esquema Conceptual



# Um Esquema Lógico





# Um Script de Criação

```
SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0;
SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0;
SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='TRADITIONAL,ALLOW_INVALID_DATES';

CREATE SCHEMA IF NOT EXISTS `Apontamentos` DEFAULT CHARACTER SET utf8 COLLATE utf8_general_ci ;
USE `mydb` ;

-----
-- Table `mydb`.`Professor`
-----
CREATE TABLE IF NOT EXISTS `mydb`.`Professor` (
  `idProfessor` INT NOT NULL,
  `Nome` VARCHAR(75) NULL,
  `idDepartamento` INT NULL,
  `Gabinete` VARCHAR(45) NULL,
  `eMail` VARCHAR(150) NULL,
  PRIMARY KEY (`idProfessor`))
ENGINE = InnoDB;

-----
-- Table `mydb`.`Disciplina`
-----
CREATE TABLE IF NOT EXISTS `mydb`.`Disciplina` (
  `idDisciplina` INT NOT NULL,
  `Designação` VARCHAR(75) NOT NULL,
  `idProfessorRegente` INT NULL,
  `Ano` INT NOT NULL,
  `Semestre` INT NOT NULL,
  PRIMARY KEY (`idDisciplina`),
  INDEX `fk_Disciplina_Professor_idx` (`idProfessorRegente` ASC),
  CONSTRAINT `fk_Disciplina_Professor`
```

# SQL



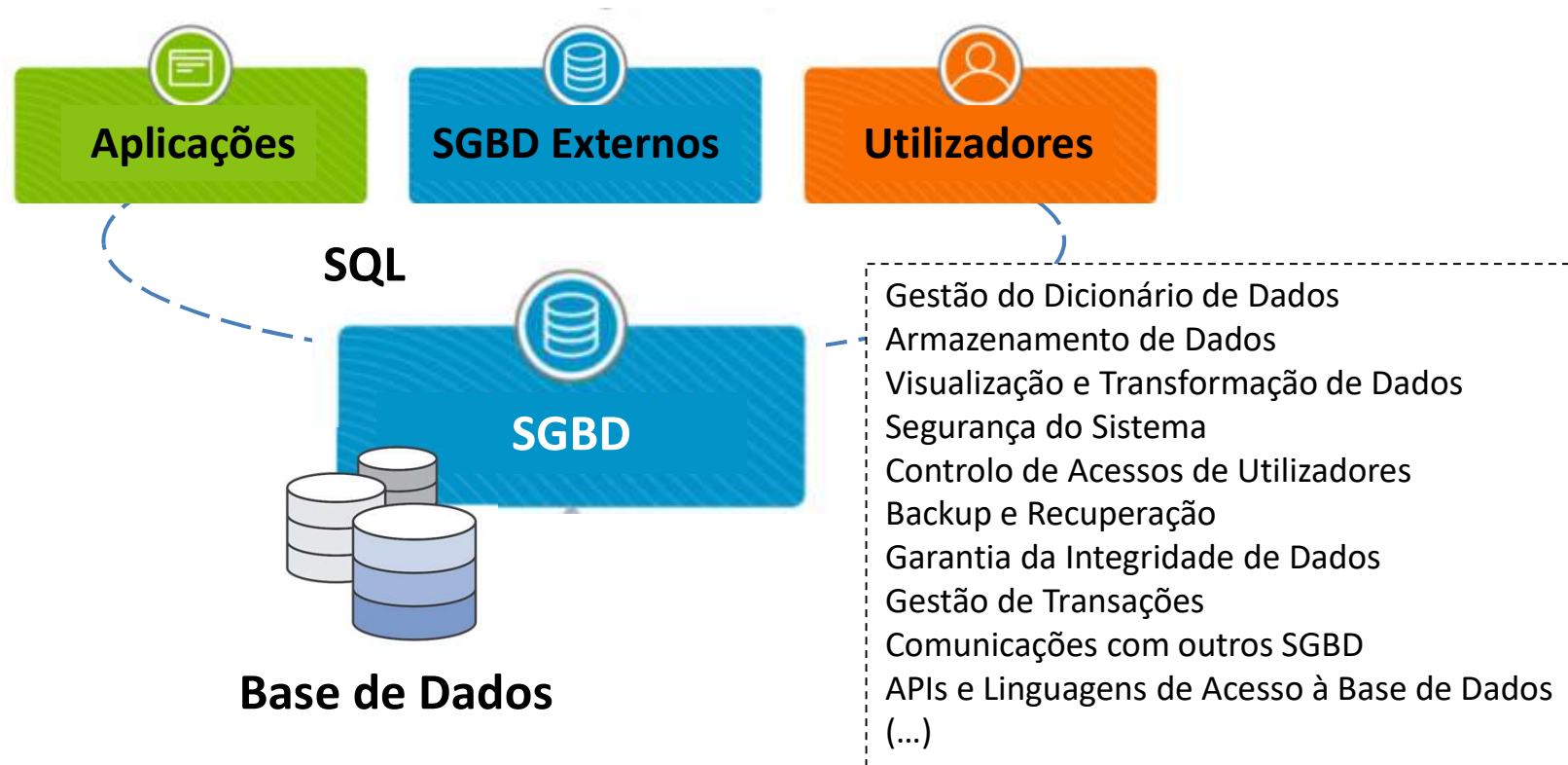
# Sistemas de Gestão de Bases de Dados (SGBD)

*“A database management system (DBMS) is a **computer software application** that interacts with the user, other applications, and the database itself to capture and analyze data. A general-purpose DBMS is designed to allow the definition, creation, querying, update, and administration of databases.”*

Fonte: <https://en.wikipedia.org/wiki/Database>



# Arquitetura Básica de um SGBD



# Vantagens de um SGBD

- Controlo de redundância de dados.
- Garantia de integridade de dados.
- Partilha de dados entre utilizadores.
- Aplicação de standards.
- Satisfação rápida de queries.
- Gestão e segurança integrados.
- Transações e controlo de concorrência.
- Cópias de segurança e recuperação de dados.
- (...)



# O Mercado dos SGBD

359 systems in ranking, October 2020

Rank			DBMS	Database Model	Score		
Oct 2020	Sep 2020	Oct 2019			Oct 2020	Sep 2020	Oct 2019
1.	1.	1.	Oracle +	Relational, Multi-model ⓘ	1368.77	-0.59	+12.89
2.	2.	2.	MySQL +	Relational, Multi-model ⓘ	1256.38	-7.87	-26.69
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model ⓘ	1043.12	-19.64	-51.60
4.	4.	4.	PostgreSQL +	Relational, Multi-model ⓘ	542.40	+0.12	+58.49
5.	5.	5.	MongoDB +	Document, Multi-model ⓘ	448.02	+1.54	+35.93
6.	6.	6.	IBM Db2 +	Relational, Multi-model ⓘ	161.90	+0.66	-8.87
7.	↑ 8.	7.	Elasticsearch +	Search engine, Multi-model ⓘ	153.84	+3.35	+3.67
8.	↓ 7.	8.	Redis +	Key-value, Multi-model ⓘ	153.28	+1.43	+10.37
9.	9.	↑ 11.	SQLite +	Relational	125.43	-1.25	+2.80
10.	10.	10.	Cassandra +	Wide column	119.10	-0.08	-4.12
11.	11.	↓ 9.	Microsoft Access	Relational	118.25	-0.20	-12.93
12.	12.	↑ 13.	MariaDB +	Relational, Multi-model ⓘ	91.77	+0.16	+5.00
13.	13.	↓ 12.	Splunk	Search engine	89.40	+1.51	+2.57
14.	14.	↑ 15.	Teradata +	Relational, Multi-model ⓘ	75.79	-0.61	-2.95
15.	15.	↓ 14.	Hive	Relational	69.55	-1.68	-15.19

Fonte: <http://db-engines.com/en/ranking>



# Referências

- Connolly, T., Begg, C., **Database Systems: A Practical Approach to Design, Implementation, and Management**, 6th edition, Pearson, January, 2014.
- Date, C., **An Introduction to Database Systems**, 8th Edition, Pearson, July, 2003.
- Garcia-Molina, H., Ullman, J., Widom, J., **Database Systems: The Complete Book**, 2nd Edition, Pearson, June, 2008.
- Stonebraker, M., Hellerstein, J., **What Goes Around Comes Around. In Readings in Database Systems**, 2004.
- M. Stonebraker et al, **One Size Fits All: An Idea Whose Time Has Come and Gone**, 2005. >
- Halevy, A., Norvig, P., Pereira, F., **The Unreasonable Effectiveness of Data**, IEEE Intelligent Systems, 2009.



# Referências Web

- [http://www.theregister.co.uk/2012/12/04/hp\\_discover\\_autonomy\\_vertica\\_big\\_data/](http://www.theregister.co.uk/2012/12/04/hp_discover_autonomy_vertica_big_data/)
- <https://www.webfx.com/internet-real-time/>
- <https://www.linkedin.com/pulse/20140722053326-3033591-the-journey-from-data-to-knowledge-and-wisdom/>
- <https://www.youtube.com/watch?v=FR4QleZaPeM>
- <http://www.computerhistory.org/revolution/memory-storage/8/265/2207>
- <https://en.wikipedia.org/wiki/Database>
- <http://db-engines.com/en/ranking>

