Sistemas de Informação e Bases de Dados
2020/2021

Aula 01: Apresentação e Introdução às Bases de Dados

Prof. Paulo Carreira

Class Outline

- ☐ DB Systems Overview and Course Presentation

  - ☐ Data, Information, and Knowledge

  - ☐ Information Systems

  - ☐ Fundamentals of Databases and Database Systems

  - ☐ Client-Server Architecture

  - ☐ Goals & Syllabus

- ☐ Course Logistics

- ☐ Course Policies

- ☐ Important Dates

- ☐ Q & A

Course Overview and Presentation

Database Systems

Are ubiquitous

- banking, hospitality, health care, real estate, retail, pharma industry, automotive, etc.

- Any application has to keep information

Information Systems and Databases

Goal

Learn how to design and implement large-scale (big) information systems using relational databases that are maintainable

Data, Information, and Knowledge

Data

When are data considered information?
When is information considered knowledge?

Data

Definition

Facts about an object or abstract concept that are not yet processed (i.e., they are *without* any interpretation or analysis)

Ex. *"The price of the oil barrel today is $80"*

Information

Definition

Result of interpreting data (give meaning to data)
envisioning a specific use or activity

Ex. "*The average monthly price of the oil barrel as risen from $70 to $80 over the last quarter*"

Knowledge

Definition

The **combination** of **information** and **experience** that can be translated into value (for a individuals or businesses)

"Whenever the price of the oil barrel drop 10 dollars fuel drops 10 cents per litter, therefore I should take the chance to fill-up"

Can data be lost?   Can information be lost?   Can knowledge be lost?

Can Data, Information, or Knowledge be lost?

▶ Can Data be lost?

Yes. If the support of data is lost.

▶ Can Information be lost?

No. It can be derived (processed) from data.

▶ Can Knowledge be lost?

Yes. If the ability (processes) to act is lost.

Data *vs.* Information *vs.* Knowledge

(Data is the <u>recording</u>, **Information** is the <u>message</u>, **Knowledge** is the <u>ability</u> to use it)

Information Systems

In a broad scope, the term Information Systems (IS) is a **scientific field of study** that addresses the range of **strategic, managerial and operational activities** involved in the **gathering, processing, storing, distributing and use of information, and its associated technologies, in society and organizations**.

The term information systems is also used to describe an **organizational function that applies IS knowledge** in industry, government agencies and not-for-profit organizations.

Information Systems often refers to the interaction between algorithmic processes and technology. This interaction can occur within or across organizational boundaries. **An information system is not only the technology an organization uses, but also the way in which the organizations interact with the technology and the way in which the technology works with the organization's business processes.**

Information Systems are distinct from Information Technology (IT) in that an Information System <u>has an</u> Information Technology component that interacts with the processes components

http://en.wikipedia.org/wiki/Information_systems

⚠ **Fundamental Idea**: not all components of the information system are digital
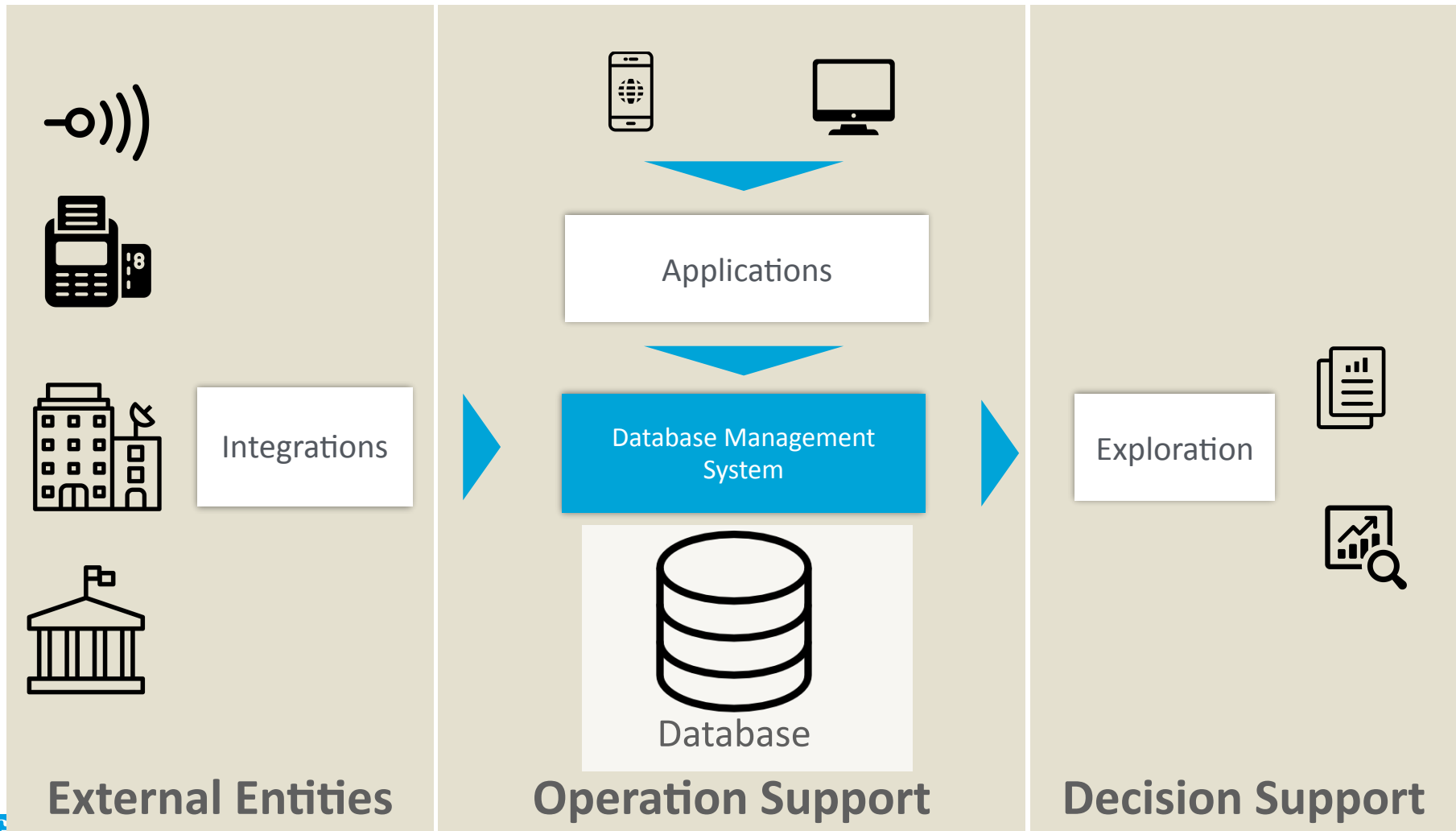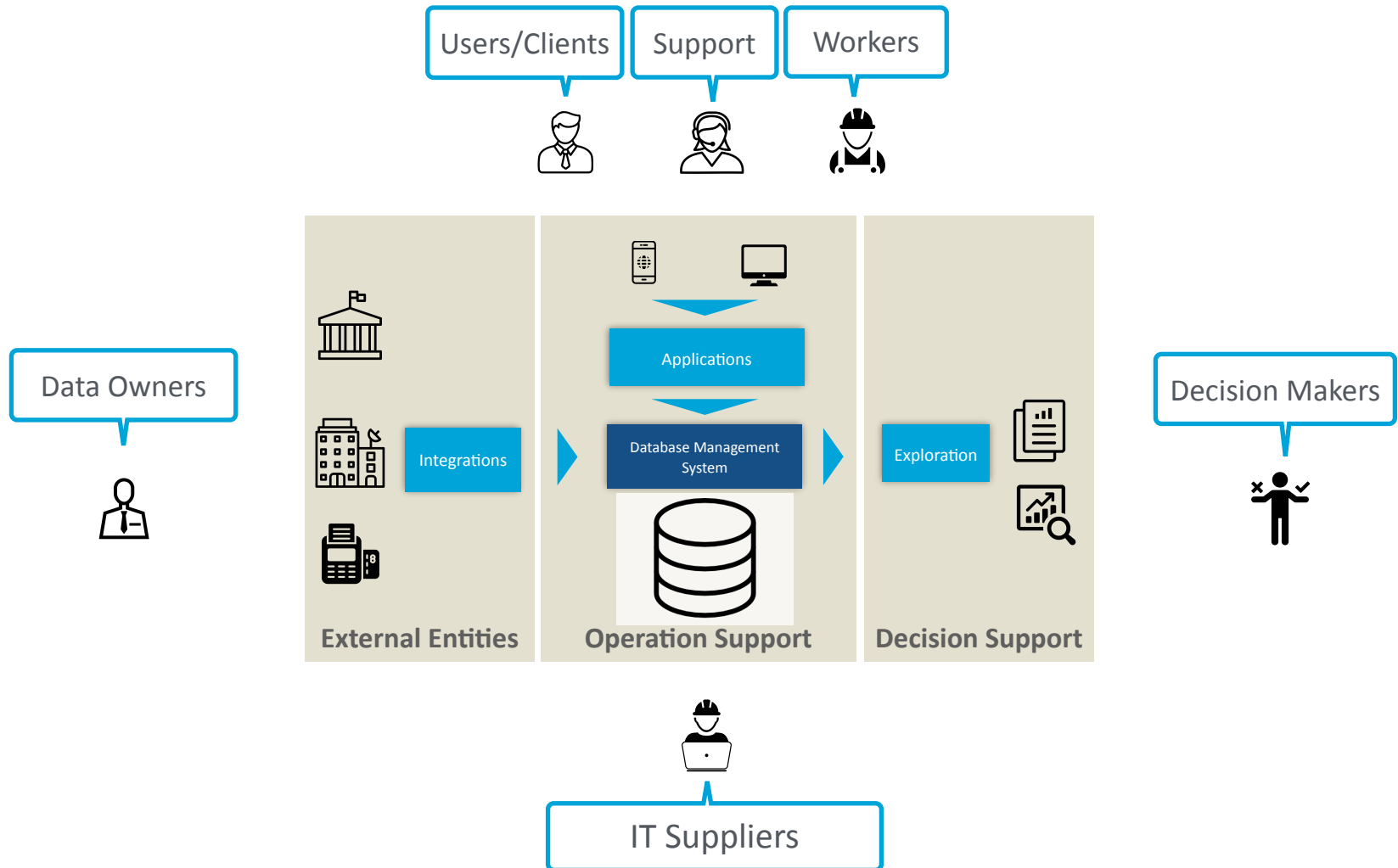
Information System (IS)

Definition

An Information System is a system acquires, stores and processes data into information to create value for an organisation or individual

Informally, an Information System transforms data into Information; the organisation transforms Information into Knowledge

Blueprint of an Information System

External Entities

Operation Support

Decision Support

Integrations

Applications

Database Management System

Database

Exploration

Actors of an Information System

Data Concepts

The Data Hierarchy

Structural Realm

- Database
- File
- Record
- Field

Physical Realm

- Byte
- Bit

Student File | Course File

Grade File

sid	name	bdate
399044	John	12/Feb/1975
59432	Mary	01/Jan/1981
64029	Paul	02/Nov/1982

sid	name	bdate
59432	Mary	01/Jan/1981

sid	name	bdate
59432	Mary	01/Jan/1981

01001101 (Letter 'M'in ASCII)

0

Conceptual Realm

- Entities
- Attributes

Data Organisation Terms and Concepts

Structural

- **Database**: Group of related files
- **File**: Group of records of same type
- **Record**: Group of related fields
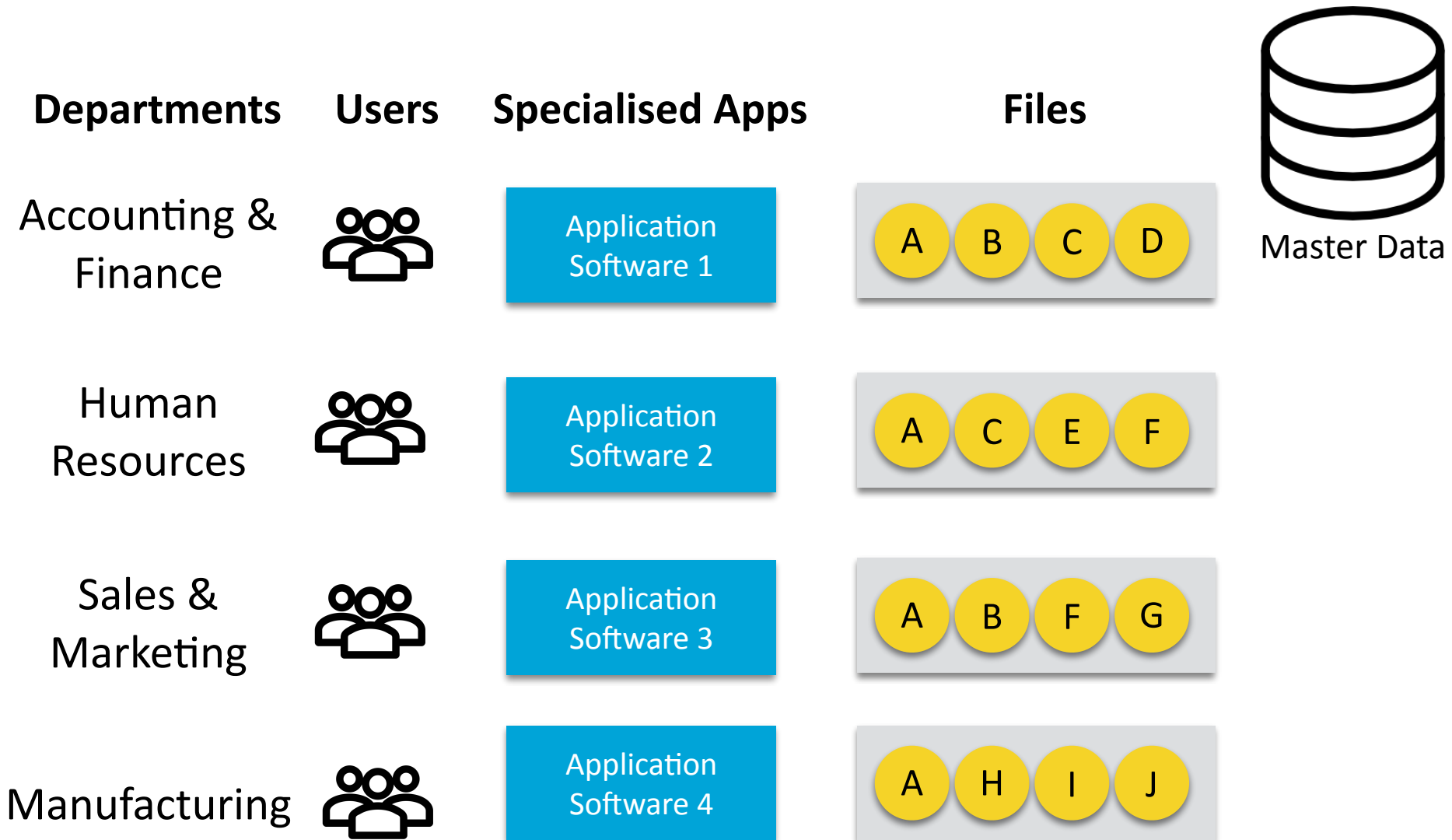- **Field**: Group of characters as word(s) or number(s)

Conceptual

- **Entity**: Person, place, thing on which we store information

- **Attribute**: Each characteristic, or quality, describing entity

Physical

- **Bit**: Smallest unit of information

- **Byte**: A sequence of bit (typically 8) that resents a character (a number, a digit, or punctuation)

Traditional File Environment

Legacy Data Environment

Departments	Users	Specialised Apps	Files	
Accounting & Finance		Application Software 1	A B C D	Master Data
Human Resources		Application Software 2	A C E F	
Sales & Marketing		Application Software 3	A B F G	
Manufacturing		Application Software 4	A H I J	

Life-Cycle of Legacy Applications

1. Applications are deployed independently by each department

2. Each Application manages its own files

3. As years go by, the organisations end up with:

   - Separate applications

   - Written in multiple languages

   - By distinct suppliers and engineering teams

4. These applications do not integrate well, are costly to maintain and evolve

Problems with Legacy Application Environments

- **Program-data coupling**. Data are kept in the context of each software application (applications "own" the files); data is affected if the application is changed (this hampers app maintainability and evolvability).

- **Lack of data sharing and integration**. Files maintained separately by different departments leading to data redundancy and inconsistency (incoherencies).

- **Lack of flexibility.** Data cannot be explored independently from the applications.

Conway's Law

Conway 1967

"Organisations which design systems are constrained to produce designs which are copies of the communication structures of these organisations" — M. Conway, 1967

Yourdon & Constantine 1979

"The structure of any system designed by an organization is isomorphic (*reflect*) to the structure of the organization (*that produces them*)" — Yourdon and Constantine, Structured Design, 1979
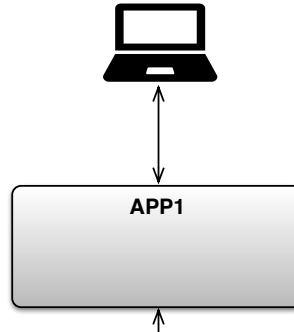
Advantages of DBMSs

Database Management System (DBMS)

Definition

A Database Management System (DBMS) is software package that manages a Database allows data to be easily accessed, manipulated and updated

DBMS

APP1

Solves the problems of legacy implementations!

Centralised Storage

▸ A unified and centralised data repository that serves multiple applications

▸ Decouples data from the applications (data and schema are independent form the applications)

▸ Avoids data redundancy because there is no more need for separate copies

▸ Enables centralised data security management

Centralised storage

Advantages

▶ Data and schema are independent from the application

- Schema can be shared

- Data can be shared

▶ Constraints on data spanning multiple applications can be shared

▶ Clients can connect and disconnect and the data lives on

Client-Server Operation

Read

Create
Update
Delete

Query Requests

APP

Server

DBMS

APP
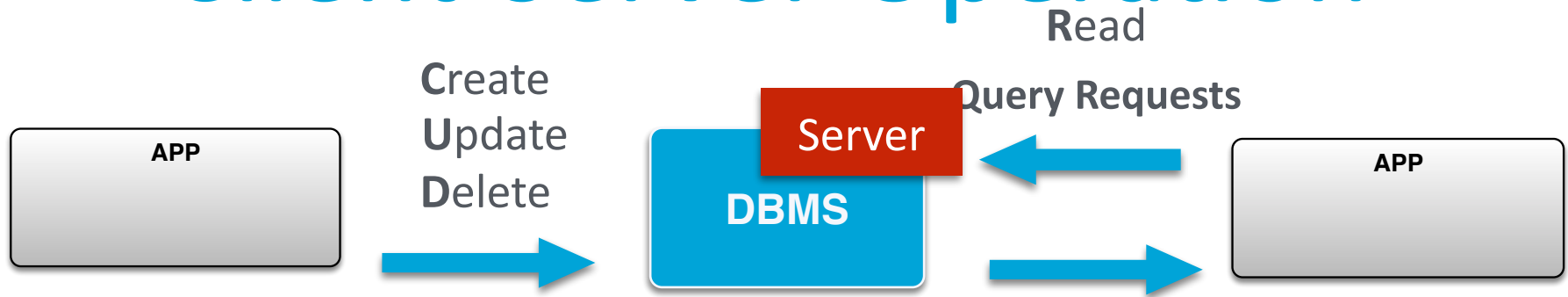
Storage

- ▶ **Client-Server**: Clients (applications) can connect to a central repository and disconnect at anytime; data li...

- ▶ **Data Independence**: Applications are decoupled from the data structures used to represent data.

  - Data can be moved to new storage device <u>without any change to the application software</u>

- ▶ **Data Quality**: Integrity Constraints on data are kept independently from the applications to actively promoting data consistency and integrity

⚠ Data are self-describing (though schemas)

Some Technical Features of DBMSs

▶ **Intelligent I/O**: group multiple data requests and try to satisfy them all in one (or as few as possible) access to storage

▶ **Caching**: results to common queries and often accessed data are kept in memory

▶ **Query optimisation**: depending on statistics and characteristics of data distinct algorithms are used; moreover these algorithms are determined automatically

▶ **Multi-user and Concurrency Control**: Gives distinct users (or apps) the illusion that they have exclusive access to data ensuring that no updates leave data inconsistent

▶ **Undo and Recovery Mechanisms**: Applications can easily undo changes to thousands or even millions of records

These features are fundamental for the performance of data intensive applications and cannot be easily implemented by each application

Course Goals & Syllabus

Goals

1. Model and criticise large databases designs using Entity-Association diagrams

2. Be able to derive correct and non-redundant database designs of databases from Entity-Association diagrams

3. Formulate correct and maintainable SQL Queries

4. Understand and fix issues related to database performance and use indexes to optimize Query performance

5. Be able to write robust Database Applications

6. Apply correctly the main concepts of data exploration

Syllabus

1. Introduction to SQL

2. Entity-Association Modelling

3. Translating EER to SQL

4. Advanced SQL Querying

5. Indexes and Database optimisation

6. Database Application Development

7. Data Exploration

8. Big Data and advanced topics

Course Logistics

Classes, Labs, Project

Lectures

- Introduces foundational concepts

- Examples

- Real-world situations

Labs

- Exercises

- Support to the development of the project

Project Assignment

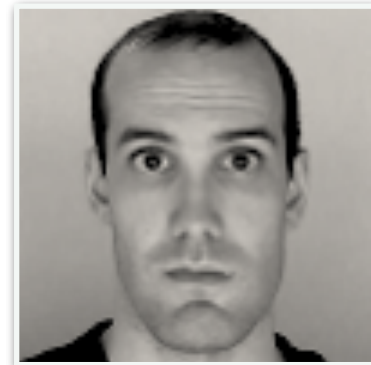- Autonomous work that aims at brings all concepts together

Faculty

Prof. Paulo Carreira

Prof. Francisco Regateiro

T.A. Inês Filipe

T.A. João Granado

Evaluation

2 Tests + Project

- ☐ 2 Tests: 60% (30%T1 + 30%T2; with m.g. 9,5/20,0)

- ☐ Projecto: 40% (10%E1 + 10%E2 + 20%E3; com m.g. 9,5/20,0)

Recovery Exam (REP)

- ■ Students that fail the tests have a second chance

Worker-Student

- ■ Are not required to be in a project group

- ■ May choose not to do the project (2 Tests=100%)

Grades of previous years

- ■ Are discarded
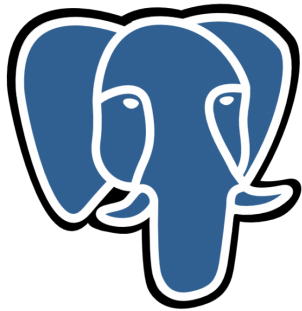
Labs & Project Groups

Labs

- Exercise guides are published a few days before the lab

- Make sure to prepare them

Project Groups

- 3 students — no exceptions

- Students are responsible for finding the colleagues for the group

- After the deadline for group registrations, students without group <u>will be assigned a group (randomly)</u>

Vehicle Software

Course Policies

Erasmus Students

- <u>All course</u> material is available in English

- The <u>course book is available</u> in English

- Lectures will be in English

- Project Assignments and Exam will be main English

- Online announcements will be <u>in Portuguese</u>

- Participation in project discussions is mandatory

E-mail Policy

- Exclusively with to deal with bureaucratic issues

- Should be preferably sent by the class delegate

- <u>Questions related to the course syllabus **will not be** answered by e-mail</u>

- <u>Questões already answered in the course FAQ **will not** be answered by e-mail</u>

Office Hours

- **Principal Lecturer**: Foundational/theoretical aspects and organisation issues

- **Lab faculty**: Issues related to the laboratories and to the project assignment (including interpretation/implementation)

- There will be no office hours in the weeks of project discussions

- There will be no office hours during the academic break periods

- There will be no office hours after last examination period

Office Hours (cont)

- The calendar with the office hours and zoom links for each faculty member is available online at the course web page

- Office hours will be run over Zoom

- Exceptionally, (due postgraduate classes, classes on another campus, department appointments) office hours may have to adjusted

Office hours calendar of Professor Paulo is always up to date at this link: http://bit.ly/2t1ihpe

Calendário Dúvidas Prof Paulo Carreira

⚠ Changes to schedule of OH that with more than 48H will not be announced in Fénix!

Project

- The organisation and indentation of the code are subject to evaluation

- The report must be simple, concise, and clean

- Project files must be submitted according to the format and organisation specified in the Project Assignment

⚠ Penalties will apply to grades of those groups who fail to comply with these requirements!

Project Submission

▶ Projects are to be submitted online using the Fénix system until the 23:59:00.000 of the due date (that is: 1 **minute before 00:00**)

▶ Regardless of reasons invoked, no submissions will be accepted **by any other means** ou **or after the due date and time** as specified above!

Groups (or respective Students) must ensure that the submission is made well in advance to avoid last minute problems (such as delays of the performance of the Fénix project submission itself)

Project Discussions

▶ Project discussions will occur in the <u>last week</u> pf the semester

▶ **WARNING**: Students that do <u>not show up to discuss their project</u> will be evaluated with **<u>zero</u>** (0/20) regardless of the reasons invoked

NOTE: In the last week of classes there will be no theoretical lectures since this will be the week for projects discussions/defences

Grading Review Sessions

- Place: **Sala Polivalente 0.17 no Informática II**

- The rules of the ***Grading Review Sessions*** are published in the FAQ & Policies section in course web page

- Dates are announced in Fénix and in the slides

Student are responsible for reserving the date for the Grading Review Session on their own calendars

⚠ Any e-mails to faculty requesting a ad-hoc grading review will be ignored!

Test/Exam Registrations

- Make sure to register for the test/exam

- Strictly on the dates announced

- There will be no last minute Test/Exam registrations!

NOTE: Students who do not register for the exam must report to the exam/text room 15 min before the start time without guarantee that they can take the test. E-mails with requests for late registrations will be ignored.

Important Dates

Important Dates — Projects

Project groups registrations

Openning for project groups registrations	25/Sep/20
Closing date for project group registrations	**05/Oct/20**

Project

Project assignment E1 released	25/Sep/20
Assignment E1 due (23:59)	**16/Oct/20**
Assignment E2 due (23:59)	**13/Nov/20**
Project assignment grading released - phase 1	20/Nov/20
Project review week - phase 1	23-27/Nov/20
Assignment E3 due (23:59)	**11/Dec/20**
Project defense/discussion week	14-18/Dec/20
Project assignment grading due - phase 2	04/Jan/21
Project review week - phase 2	**04-08/Jan/21**

Important Dates — Tests

Test TE1

Openning for test TE1 enrollment	04/Nov/20
Closing for test TE1 enrollment	**18/Nov/20**
SIBD TE1 (09:30-11:00)	**20/Nov/20**
SIBD TE1 grading due	30/Nov/20
SIBD TE1 grading review session (10:00 - 12:00 & 14:30 - 17:30)	02/Dec/20

Test TE2

Openning for test TE2 enrollment	29/Nov/20
Closing for test TE2 enrollment	**13/Dec/20**
SIBD TE2 (11:00-12:30)	**15/Dec/20**
SIBD TE2 grading due	04/Jan/21
SIBD TE2 grading review session (10:00 - 12:00 & 14:30 - 17:30)	06/Jan/21

Relevant Dates — Exams

Recovery (REP) Exam

Openning for exame REP enrollment	10/Jan/21
Closing for test REP enrollment	**24/Jan/21**
SIBD Exame REP (11:30-14:30)	**26/Jan/21**
SIBD REP Exam grading due	01/Feb/21
SIBD REP Exam grading review session (10:00 - 12:00 & 14:30 - 17:30)	02/Feb/21

Special Exam Season

Exam SIBD EE	-

Paulo Carreira

Database Specialist,
Researcher,
Entrepreneur

linkedin.com/in/pjcarreira

paulo.carreira@tecnico.ulisboa.pt

Q & A