IPCA ESCOLA SUPERIOR DE TECNOLOGIA

Project Guide 01

Decision Support Systems (DSS), 2023-24

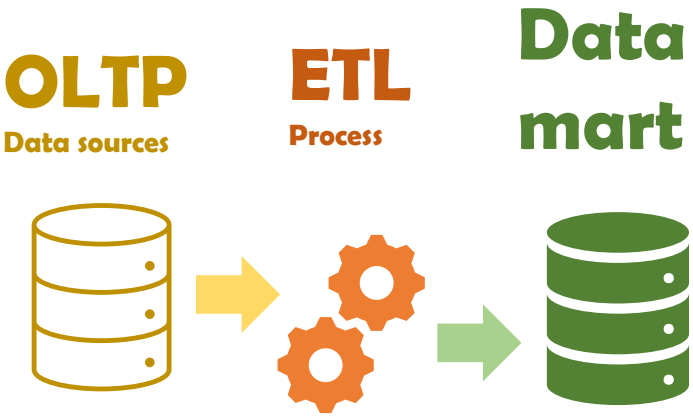
Degrees on Computer Systems Engineering and
Medical Informatics Engineering

Introduction

- The goal of project 01 is the implementation of a data mart based on an operational database.

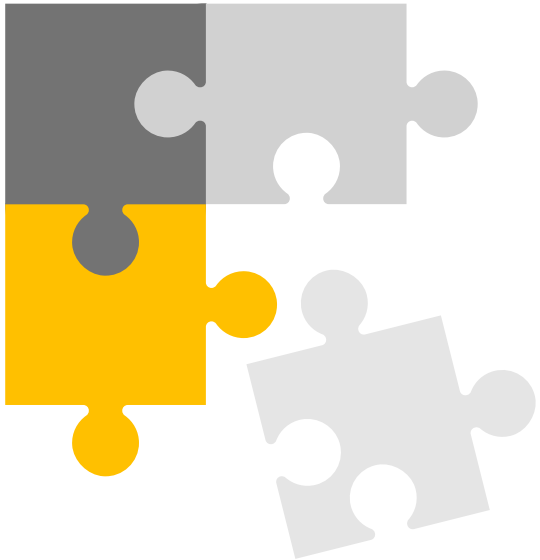
- The team should undertake the following tasks
  - Business requirements and data profiling
  - Dimensional modeling of the data mart
  - Implementation of ETL Process

- OLTP data source
  - **Transfermarkt** database – airports and flights
  - Available on PostgreSQL server: **postgres2.ipca.pt**



Project team and submission

- The project must be carried out in teams of **three members**
  - The project grade is individual, based on the discussion of the project
- The project must be submitted on the Moodle platform as a **zipped file** that must include
  - (1) the project report and (2) the ETL file
  - It should be named **DSS_###_P01**, where ### corresponds to the team number. It also applies to (1)
  - The report must be submitted in PDF format and according to the **template** available on Moodle
- Projects submitted after the deadline may not be evaluated



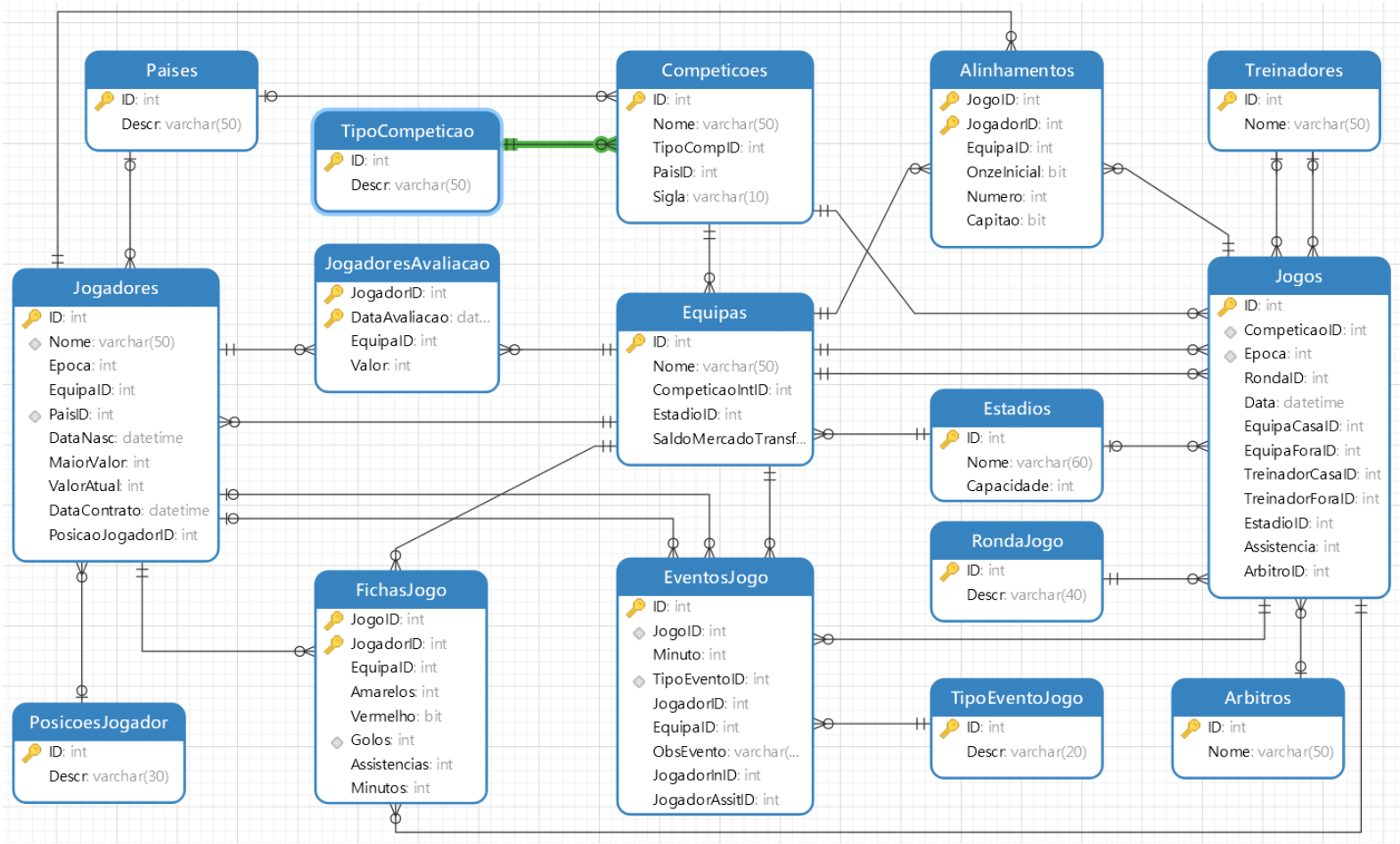
Transfermarkt database

- The data was downloaded from Kaggle (https://www.kaggle.com/datasets/davidcariboo/player-scores)
- The original data was scraped from Transfermarkt (https://www.transfermarkt.co.uk/)
- The database was created as part of an academic project by three students [1]:
  - The project was implemented in SQL Server
  - The database was migrated to PostgreSQL database

[1] J. P. Gomes, P. Carvalho, J. Carreira. Projeto 01 de Bases de Dados Avançadas. MEI, IPCA, 2024.

Table_name	Rows
Alinhamentos	69 909
Arbitros	2 335
Competicoes	43
Equipas	426
Estadios	2 280
EventosJogo	518 713
FichasJogo	1 432 321
Jogadores	30 302
JogadoresAvaliacao	440 571
Jogos	51 499
Paises	180
PosicoesJogador	13
RondaJogo	116
TipoCompeticao	11
TipoEventoJogo	4
Treinadores	5 553

TransferMarkt DB



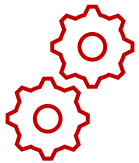
Tasks to be developed

Data Profiling

- Analysis of the data schema
- Study of the data content

Dimensional modeling

- Identify goals and objectives
- Develop the DW matrix
- Identify Facts and Dimensions
- Include the ER Model and data description maps



Extract, Transform, Load (ETL)

- Implement the ETL process
- Document the transformations
- Run the ETL jobs
- Make a brief summary of the data mart content

Final remarks

The main concepts and techniques about dimensional design, covered during the classes, should be put into practice.

The evaluation of projects will take into account, among other things, the application of skills, team commitment and the matching of the solution to the project objectives.

Thank you!