



Ciências
ULisboa

Tecnologias de Processamento de dados - 2019/2020

Project definition and outline

Project goal: To start, plan, implement and use for analysis a Data warehouse.

Student Groups of 3 or 4 from the same lab class

Phase I (deadline: March 23)

- Select one data set from a publicly available repository. Check the options at the T01 lecture. **This selection must be aproved by the couse responsible**
- Identify the data elements for Dimensions and Facts and possible external data to enrich the dimensions.
- Define the star schema of the proposed data warehouse and estimate its size and growth over time
- Identify the main usage of the data system and propose different analysis queries for the system

Phase II (deadline: April 19)

- Identify data issues.
- Define an ETL workflow
- Implement the Data warehouse: Build the dimensions and facts. Identify hierarchies and fact granularity
- Follow the workflow and load the Data warehouse
- Do a critical assesement of the work and compare it to the expectations of phase I, both in structure as in size

Phase III (deadline: May 24)

- Identify bottlenecks (long queries, long dimensions, difficult to analyse data)
- Use the system by executing analysis queries and report your findings
- Define dashboards and KPIs (if meaningful)