



Ciências
ULisboa

Tecnologias de Processamento de dados - 2019/2020

Phase I (deadline: March 30)

Write a report with the answers to each of the following points

1. Describe your original data set. Identify the most important information. Identify missing or incomplete data, and identify possible strategies to use (or discard) them
2. Describe what other data sets (if any) are going to be used for the construction of the data warehouse. How will they complement the existing information
3. Identify and characterize the facts table. Identify and characterize the grain of each element. If more than one fact table is present identify the grain for each of them.
4. Identify all the possible dimensions with the data set and the accessory data found. Here we just need to identify the Dimensions, not define their structure.
5. Identify how many data marts are going to be used and what types of business processes might use them. If more than one data mart is present, define the Bus matrix and identify their position in a feasibility/value matrix
6. Define the star schema of the proposed data warehouse
 - a. For each dimension identify and describe its columns
 - b. Identify and characterize the measures of the facts table
 - c. Estimate its size and growth over time (in number of lines in tables)
7. Identify the main usage of the data system and propose different analysis queries for the system. It's just necessary to write them down in plain language

Please note that no code whatsoever is required, nor any data work in Excel or otherwise.

The report can be written in Portuguese or English and should not be longer than 7 pages. It is to be submitted on the course's Moodle **until 23:59 of March 30**. On the first page it must contain the **group number, the Student Id's and names, and an estimation of how many hours each student contributed to this phase**