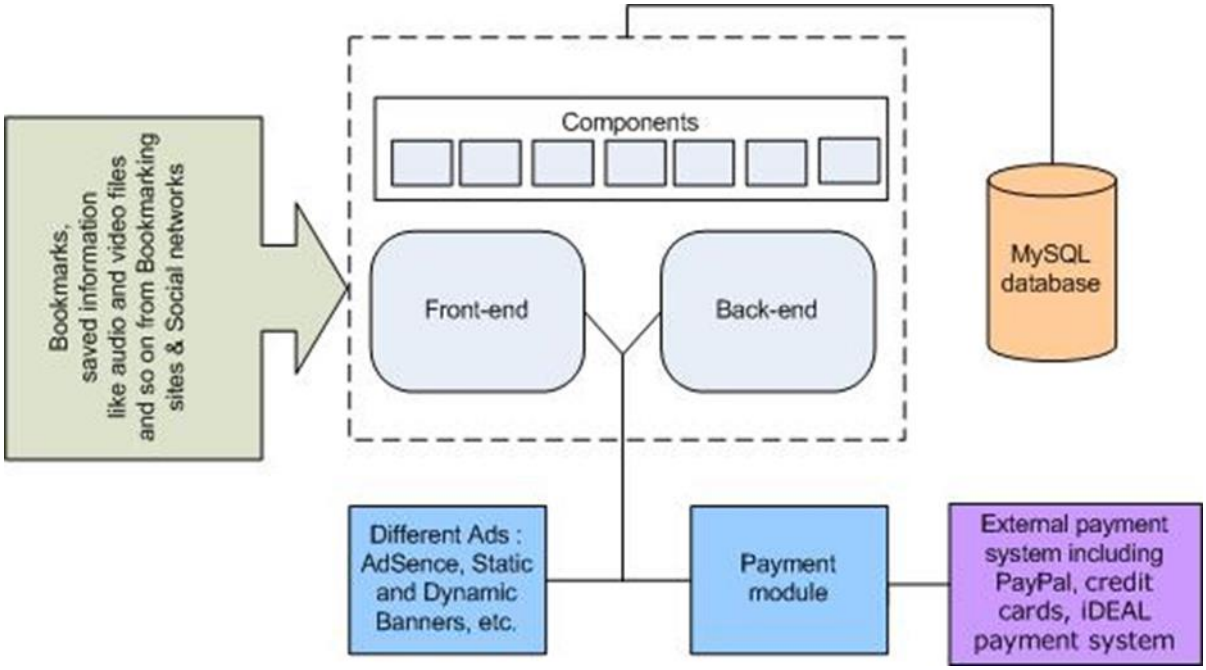
Final Project

Argument

Your are an employee of an international software company which is selling a Web CMS.

"*A content management system (Web CMS) is a bundled or stand-alone application to create, deploy, manage and store content on Web pages. Web content includes text and embedded graphics, photos, video, audio, and code (e.g., for applications) that displays content or interacts with the user. A Web CMS may catalog and index content, select or assemble content at runtime, or deliver content to specific visitors in a requested way, such as other languages. Web CMSs usually allow client control over HTML-based content, files, documents, and Web hosting plans based on the system depth and the niche it serves*."

The architecture of the system is described as follows:



(Search Internet for more information about how a CMS works and the functionality of each block).

The CMS is adapted to the client, so any new project requires always changes in some modules. The company has different development teams in south Europe, which make the changes depending on the client's country.

Your boss is concerned because in the last 3 years there have been many major faults in the system, and he wants to know why. A major fault is defined as a fault that causes an extra investment of effort to fix errors, of more than 20% of the effort investment of the original development.

He knows that you took the course of Data Engineering and he asks you to find the causes of these faults. You asked him for a data description of the projects of the last years, and he gave you the attached file with the following data:

- Country: of the development team
- FAULT: if it was found a major fault
- Effort: effort in person/months of the original development
- OS: operating system related with the errors
- Module, Component: affected by the changes to fix the errors
- BackEnd, FrontEnd: indicating if they were also affected by the changes to fix the errors

Final Project description

1. Apply data mining to this problem, and find which the causes of the major faults are.
2. Describe what you have done in each CRISP-DM phase and cycle. Use the attached GUIDE FOR REPORTS.