Universitat Politècnica de Catalunya
Facultat d'Informàtica de Barcelona

Bachelor Degree in Informatics Engineering
Computer Engineering Specialization
Degree Final Project
Thesis management course

Performance analysis and optimization of a combustion simulation

Second assignment: Time planning

Author
Guillem Ramírez Miranda

Director
Marta Garcia Gasulla

Co-director
David Vicente Dorca

Tutor
Julian David Morillo Pozo

GEP Tutor
Joan Sardà Ferrer

11th October 2020

Contents

List of Figures

List of Tables

1 Description of tasks

1.1 Project management

Project management is a compulsory task for any project.

1.1.1 Meetings

Regular meetings are contemplated to keep track of the progress of tasks and to make decisions. It is planned to do approximately 2.5 hour of meetings per week. Time estimated: 45h

1.1.2 Context and scope of the project

Definition of context and scope of the project. Time estimated: 30h

1.1.3 Time Planning

Definition of the tasks and the time planning of the project. Time estimated 10h

1.1.4 Budget and sustainability

Compute the monetary costs of the project. Definition of the sustainability report. Time estimated 10h

1.1.5 Integration and final document

Review, fix and integrate the last 3 tasks into a final document that groups all the information related to the project management. Time estimated 15h

1.2 Hands-on

These tasks are intended to get introduced to the application analysed and ensure we are analysing the correct thing.

1.2.1 Contact with application developers

We will meet with the application developers where they will introduce us to the application module. The following points will be worked:

- Basic explanation of the science between the module and the input case.
- Localization of the code and the input case.
- Explanation of the compilation process of the code.

- Explanation of how to run the code.
- Download the code and the input case
- Compilation of the code
- First runs of the code.

Time estimated: 10h

1.2.2 Study the compilation process

It is very important to understand the compilation process and the different options that the build chain allows. This is a simple task but helps a lot in minimizing the possible errors we can encounter adding or modifying the code in the future.
Time estimated: 15h

1.2.3 Run and test the application

Every user in a HPC machine has its own environment. As the code is in development process and I may need to modify it an make my own installations thus some things from the application may differ. It is a critical task to perform some runs and verify with the application developers that we are not having errors, we are running the correct thing and the intended thing to analyse.
Time estimated: 15h

1.3 Performance analysis

1.3.1 Trace extraction

Once everything it is sanity-checked, the environment is ready and we are familiarized with the application the next step is to extract traces of the executions. As we will use the POP method we will need traces from executions from 1 process to 48 process and from 1 node to 16 nodes. This process is long and needs to operate with data.
Time estimated: 30h

1.3.2 Modelfactors

Once we have the traces we have to use basic analysis tool in order the extract the POP metrics from the executions. This process require a few hours to process the traces but nothing compared to extracting them.
Time estimated: 10h

1.3.3 Focus of analysis

In this stage of the project we will study the results of the modelfactors. From the results we will obtain insights into what is reducing the performance of the application. With this, we will use the other BSC tools to find the bottlenecks. Once the bottlenecks are found we will need to quantize the real impact on the performance to know which bottleneck is better to optimize. This bottleneck will be called the "Focus of analysis".
Knowing the time estimated is tricky as it depends on the easy is to identify the bottlenecks and the analysis in general, but, based in previous expiriences we estimate a total of 35h

1.3.4 Feedback to application developers

At this point of the project, a presentation and a conclusions of the analysis will be prepared and presented to the application developers.
Estimated time: 15h

1.4 Optimization

1.4.1 Design implementation

Based on the results of the previous section, we will decide what we will do to attack the bottleneck. The decision and the design it is an unknown for now. It is expected to be a total of 30h long.

1.4.2 Implementation

This phase is the great unknown of the project since we do not know what we are going to optimize or what we are going to do to optimize it. It is also the main task of the project and the one that will surely take the most hours. However, we cannot know how long it will take, since for now what we are going to do is unknown. Based on previous experiences and the time limit of the project we expect to invest 100h in this phase.

1.5 Testing implementation

It is mandatory to test and ensure that the modifications the code still lead to a correct program. A proper testing suite in collaboration with the application developers will be done.
Time estimated: 30h

1.5.1 Evaluation of the improvements

Once the job is done it is very important to evaluate the performance of the application with the improvements. This task consists of gathering

traces and timing data from both versions, the one which is improved and the original one and elaborating a final conclusions of how is performing the optimization.
Time estimated: 40h

1.6 Final milestone

It is necessary to write the documentation before ending the project. Two sub-tasks are expected:

- Memory redaction. Time estimated: 70h
- Presentation preparation. Preparation of the final presentation of the work, this includes the material and the training for the presentation. Time estimated : 15h

1.7 Task dependencies and summary of tasks

Table 1 shows a summary of tasks and it's dependencies. Almost each tasks depends on its predecessor making this project really sequential.

1.8 Task resources

1.8.1 Human resources

The BSC researcher will be in charge of the project and will be the main responsible of all tasks.
Other human resources are:

- The project director that is responsible of tracking the status of the project and giving feedback and suggestions to the researcher.
- The application developers are also human resources needed as they are in charge of introducing the researcher to the program (T1) and giving support if any problem with the application is encountered.

1.8.2 Material resources

- Dell Latitude 7490. Laptop that the project author will use for all the tasks.
- MareNostrum 4 supercomputer used for tasks T2, T3 and T4.
- Control versioning server used for keeping track of the changes to the code and the documentation. Tasks T1, T3.7, T4 and T6
- A Mailserver used for communication with the project director and the application developers.

Id	Name	Time (h)	Dependencies
T1	Project managment	110	
T1.1	Meetings	45	
T1.2	Context and scope of the project	30	T1.1
T1.3	Time planning	10	T1.2
T1.4	Budget and sustainability	10	T1.3
T1.5	Integration and final document	15	T1.4
T2	Hands-on	40	
T2.1	Contact with application developers	10	
T2.2	Study the compilation process	15	T2.1
T2.3	Run and test the application	15	T2.2
T3	Performance analysis	90	
T3.4	Trace extraction	30	T2.3
T3.5	Modelfactors	10	T3.4
T3.6	Focus of analysis	35	T3.5
T3.7	Feedback to application developers	15	T3.6
T4	Optimization	200	
T4.1	Design implementation	30	T3.6
T4.2	Implementation	100	T4.1
T4.3	Testing implementation	30	T4.2
T4.4	Evaluation of the improvements	40	T4.3
T5	Final milestone	85	
T5.1	Memory redaction	70	T4.3
T5.2	Presentation prepration	15	T5.1
Total			525

Table 1: Summary of tasks

- LaTeXused for writing documentation. Tasks T1, T3.7, T4 and T6.
- Ganttproject for the Gantt diagram tracking. Task T1.3.
- A text editor for writing the code, the documentation and in general manipulating files. All the tasks involved.
- Office supplies.

2 Gantt

The start of the project is expected by the 13/9/2020 which is the approximate start date of the thesis management course. The project is intended to end before the 29 of January that is the date of defence.
Figure 1 shows the gantt diagram of the project.

3 Risk management: alternative plans and obstacles

- **MareNostrum4 not available**: The unavailability of the machine can lead to delays in almost all tasks. In general when the machine becomes unavailable in 1 day the issue is solved. We can expect a delay of maximum a 5% of the expected time. If a catastrophic event happens to the machine and it becomes unavailable during a notable period of time it is contemplated to move the study to a similar machine that BSC disposes of. This implies an increment of 20h to move the project and adapt to the new environment.
- **Misinterpretation of a metric during the analysis**: This can lead to small delays of maximum 5h. From previous experiences we can expect to not affect the project timing and this risk will be lesser as the researcher is more experienced.
- **Incorrect runs**: This risk affects the project for maximum 2h as it is fast to detect and fast to fix.
- **Machine noise**: As this risk is contemplated and it is solved by gathering the proper number of samples it is already contemplated on the duration of the tasks and won't affect the timing of the project.

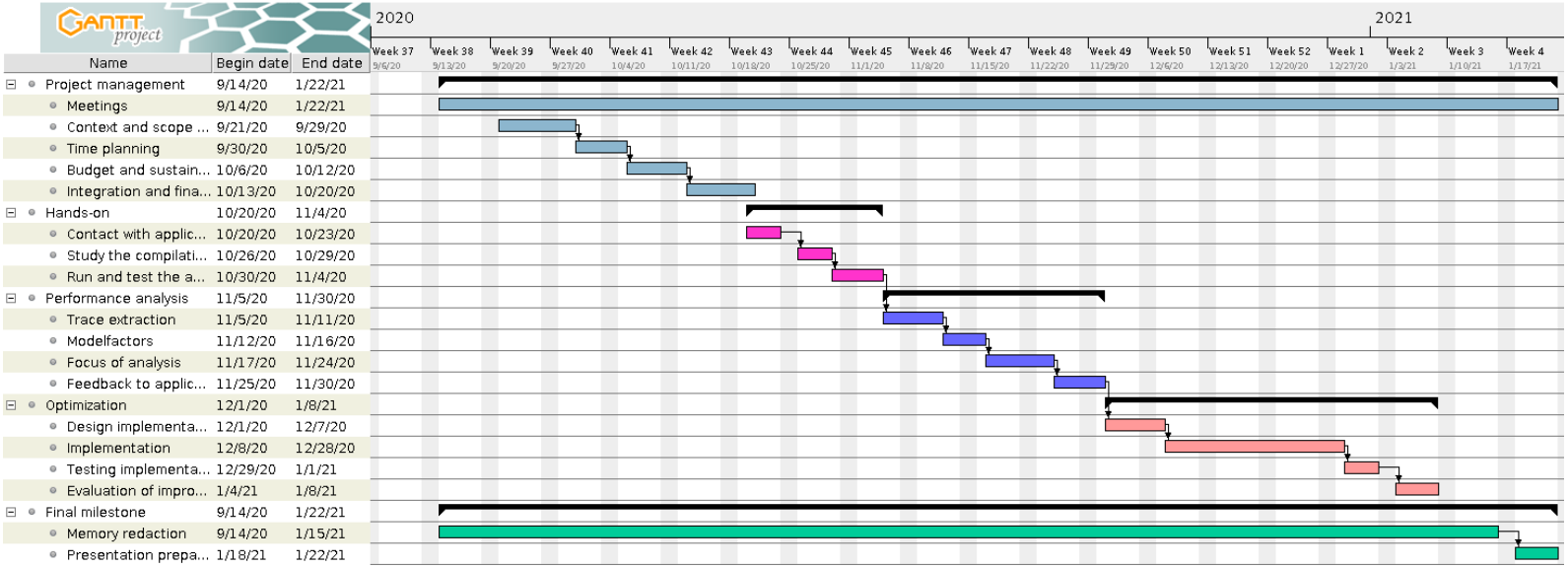

Figure 1: Gantt diagram. Own compilation.