WIP Title: Dynamic resource allocation in the cloud for compute heavy tasks in a containerized environment

Elia Ravella

July 22, 2022

Contents

Part I

The Problem, the State of the Art and Current Available Solutions

1   Introduction

2   Containerized Environment and High Performance Computing

3   State of the Art

3.1   Shifter

3.2   SLURM

3.3   Kubernetes

3.4   Serverless Approach

4   The Problem

This section highlights the problems of the currently available solutions: the focusing on scaling through replication rather than on resources size, and the problem of having a dynamical *in two senses*, both resource- and replication-wise, computational layer

Part II

Design and Testing Phase

5   MapNCloud Original Architecture

Here I talk about the original deployment of the MapNCloud service. I plan to add a subsection explaining in detail the tech stack.

6   Problems Addressed

1. database choice and API modification

2. queue monitoring

3. resizable backend containers

4. cloud provider integration

At the end of this section I will present the "final" design draft

7   Testing and Validation

HERE I will introduce the "diffusion analysis" to justify the test parameters

1. CouchDB testing

2. RabbitMQ testing

3. Cloud providers options, pros and cons

4. technological limitations (docker-compose, load balancers)

I will also present the real "final" Architecture that will be deployed here, with cloup provider's technological names and services

Part III
Implementation

8 Frontend

9 Backend

10 Database

11 Messaging Middleware

12 Computational Layer

12.1 Renderino

Part IV
Conclusion and Bibliography