POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

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

Tesi di Laurea Magistrale in
Computer Science and Engineering

Author: **Elia Ravella**

Student ID: 967243
Advisor: Prof. Raffaela Mirandola
Co-advisors: Name Surname, Name Surname
Academic Year: 2021-22

Abstract

Keywords: Cloud, Containers, Dynamic infrastructure

Contents

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

1.1. Introduction

1.2. Containerized Environment and High Performance Computing

1.3. State of the Art

1.3.1. Shifter

1.3.2. SLURM

1.3.3. Kubernetes

1.3.4. Serverless Approach

1.4. 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

2 | Design and Testing Phase

2.1. 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.

2.2. 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

2.3. 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

3 Implementation

3.1. Frontend

3.2. Backend

3.3. Database

3.4. Messaging Middleware

3.5. Computational Layer

3.5.1. Renderino

Bibliography