

---

# **Orchestra**

**EMBL-EBI Microbiome Informatics Team**

**Aug 23, 2021**



**CONTENTS:**

<b>1</b>	<b>What is Orchestra?</b>	<b>1</b>
<b>2</b>	<b>Arhitecture</b>	<b>3</b>
<b>3</b>	<b>Installation</b>	<b>5</b>
<b>4</b>	<b>Configuration</b>	<b>7</b>
<b>5</b>	<b>Job execution</b>	<b>9</b>
<b>6</b>	<b>Indices and tables</b>	<b>11</b>



## WHAT IS ORCHESTRA?

It is an orchestration system that can coordinate the execution of analyses jobs within a set of distributed heterogeneous compute clusters. It act as an abstraction layer to hide the complexities of distributed clusters and provide a unified interface to interact with and monitor the service.



**ARCHITECTURE**



**INSTALLATION**

TODO



## CONFIGURATION

### Clusters Configuration

The first step is to configure the Clusters that Orchestra will interact with. Orchestra keeps the clusters in it's database.

#### Cluster Model

- name
- cluster\_type (what are we going to store here?)
- status (active, inactive)
- messenger (which queue system, GCP pub/sub by default)
- messenger\_queue (queue identifier)

Each cluster needs to have a queue in place. GCP Pub/Sub is the queue system supported by default.

#### Job Types

Each job will have a job type, which at the moment is a free text field. This is used by the demon to infer which log parser to use.



## JOB EXECUTION

The cli of conductor is the interface to lunch jobs at the moment of writing this.

```
conductor run job.yaml
```

```
name: "Name of the job"  
cluster: "Cluster name or id"  
job_type: "Job type"  
script_path: "Full path to the script"
```

conductor will report back the job id.

Each job will be scheduled to be executed.



## INDICES AND TABLES

- `genindex`
- `modindex`
- `search`