
Orchestra

EMBL-EBI Microbiome Informatics Team

Aug 23, 2021

CONTENTS:

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

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