

# Summary of *vcluster* Command

Vafa Andalibi

January 20, 2019

Virtual cluster is a tool that can be used to build a cluster from selection of the nodes, submit parallel jobs to them and retrieve the results of those jobs to the local machine. The full documentation for the *vcluster* tool as well as corresponding examples are available online<sup>1</sup>. In the following sections, the main assumptions of the tool are described:

## The virtual cluster

A virtual cluster is built based on a series of nodes in *yaml* configuration file of the Cloudmesh V4. In the current implementation it is assumed that the selected nodes are single machines not cloud clusters. Moreover, virtual cluster connects to nodes using SSH and its configuration file which contains the path to the private key. It is also assumed that the public key of the user is already copied in the remote nodes.

Accordingly, the two main limitations are:

- How to automate the authentication (LDAP, OAuth)?
- How to handle the cluster nodes like Chameleon cloud nodes or Azure nodes?

## The job

The job is expected to be a executable file, e.g. a script or binary, which takes one of the followings as an input:

- Command line arguments
- An accompanied file, e.g. a data file
- Both command line arguments and a file

For instance, suppose there is a big data file on which the executable works and you can specify which index/part of the data you want the script to work on. In this case it is possible to specify which part/index of data should be executed by which node using the command line argument.

---

<sup>1</sup><https://cloudmesh-community.github.io/cm/vcluster.html>

## Configurations

The `vcluster` command also saves its temporary configurations, including current virtual cluster, submitted jobs and required information to check/retrieve the results, in a yaml file located in the *workspace* folder. Should this be changed to use MongoDB instead of a local file?