COMP90024 2021S1 Project2 Report

**ABSTRACT**

**This the report for the COMP90024 Project2. In this project we are going to analyze twitter data base on Melbourne Research Cloud. In section 1, we will introduce how to create an instance, build a running environment on the instance, and start applications on instances.**

# User guide

## Instances assignment

We will allocate a total of four instances with a capacity of 100GB. Except for the first one that will be used to process the front-end, the remaining three will deploy couchdb and establish a cluster. One of these three will be the master of the cluster, and it will be Deploy Tweet harvester to collect tweets and upload them to the cluster.

|  |  |  |
| --- | --- | --- |
| Instance Number | Function | Volume (GB) |
| 1 | Frontend | 100 |
| 2 | Cluster Master, Tweet harvester | 100 |
| 3 | Cluster node | 100 |
| 4 | Cluster node | 100 |

## Deploy instances

In this step, we use Ansible to set dependencies, security groups, volumes, and images, also dynamically generate instances in MRC through these.

In the beginning, we install the dependencies required to run the script task (such as Python-pip, openstack sdk and update pip) and get the image we need from openstack. Then define the volume required for each instance (these volumes will be used to store data such as tweets) and define the security rules for each instance (We will open all ports of the instance for emergency and convenience.). Finally, use all the above settings to create four instances.

|  |  |  |
| --- | --- | --- |
| Hosts | Role | Description |
| localhost | Instance-common | Install pip, update pip, and openstack sdk dependencies. |
| Instance-images | Retrieve all images from openstack and select the right one we use. |
| Instance-volumes | Set the volume required for each image. |
| Instance-security-groups | Create security groups and their rules |
| Instance-creation | Create instances on MRC and save the IP address of each instance to the inventory folder. |

### 1.1.1 How to run

Run ./launch-instances.sh in ubuntu system, make sure you have located the location of this file.

## Configure environments

In this step, we configure the environment of instances deployed in step 1.1, this contains adding proxy, assignment of volume, docker configuration, and git initialization.

First, we add proxy to instance to allow it access network outside LAN of Unimelb. (or it can’t install dependencies from network), then reboot instance also wait a *reboot\_timeout: 6000* to make sure proxy works. After that install some dependencies we may use late. Second, mount volume we define in step 1.1 to corresponding instance. Third, to reduce the risk of compatibility, we will run our worker program in docker container, so we need to install docker on our instances and add proxy for it for the same reason as above. Fourth, to make it easier to get the latest version of our worker program and other codes, we decided to install the GitHub repository in the instance.

|  |  |  |
| --- | --- | --- |
| Hosts | Role | Description |
| instances | environment-installation | Add proxy to instances and reboot, then install dependencies. |
| environment-mount | Assign volume to each instance. |
| environment-setup-docker | Add proxy for docker and then restart docker |
| environment-git-clone | Setup GitHub repository in instances. |

### How to run

Run ./setup-environment.sh in ubuntu system, make sure you have located the location of this file.

## Start applications

In this step, we deploy couchdb on every instance used for building cluster, then run cluster making command on the instance which will be the master node of cluster. After cluster established, run python script to harvest corresponding tweets from twitter API to cluster.

|  |  |  |
| --- | --- | --- |
| Hosts | Role | Description |
| dbServers | deploy-couchdb | Deploy couchdb on every instance used for building cluster. |
| dbMaster | deploy-cluster | Create cluster and make this host the master, also active CORS. |
| deploy-harvester | Start Tweets harvester program, it harvest corresponding tweets to couchdb cluster we set up. |

### 1.3.1 How to run

Run ./deploy.sh in ubuntu system, make sure you have located the location of this file.