

How To Set Up a Gen3 Data Commons Using Helm Charts

Center for Translational Data Science (CTDS), University of Chicago
 Open Commons Consortium (OCC)
 Australian BioCommons
 New Zealand eScience Infrastructure (NeSI)



- Introduction
- Helm Charts on Desktop - Center for Translational Data Science (CTDS), University of Chicago
- Helm Charts on AWS - Center for Translational Data Science (CTDS), University of Chicago
- Helm Charts on GCP - Open Commons Consortium (OCC)
- Cloud Automation on OpenStack - New Zealand eScience Infrastructure (NeSI)
- Open Discussion

Introduction

Robert Grossman, Center for Translational Data Science, University of Chicago

Welcome to the Bi-Monthly Gen3 Community Forum

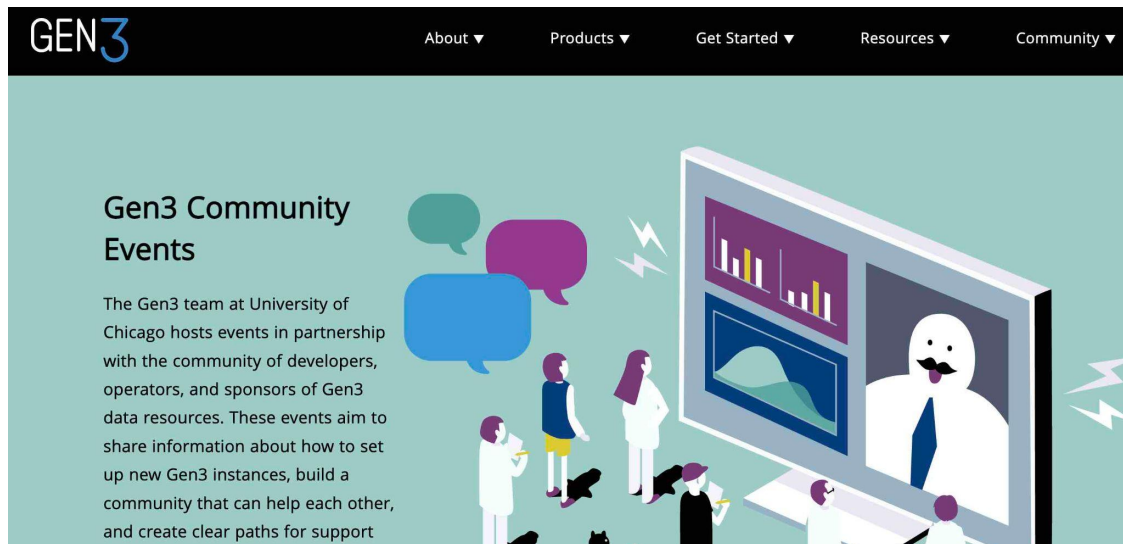
Gen3 is a data platform for building data commons and data ecosystems.

The Gen3 platform consists of open-source software services that support the emergence of healthy data ecosystems by enabling the interoperation and creation of cloud-based data resources, including data commons and analysis workspaces. Gen3 aims to accelerate and democratize the process of scientific discovery by making it easy to manage, analyze, harmonize, and share large and complex datasets in the cloud.

[Experience Demo](#)

[Get Started](#)



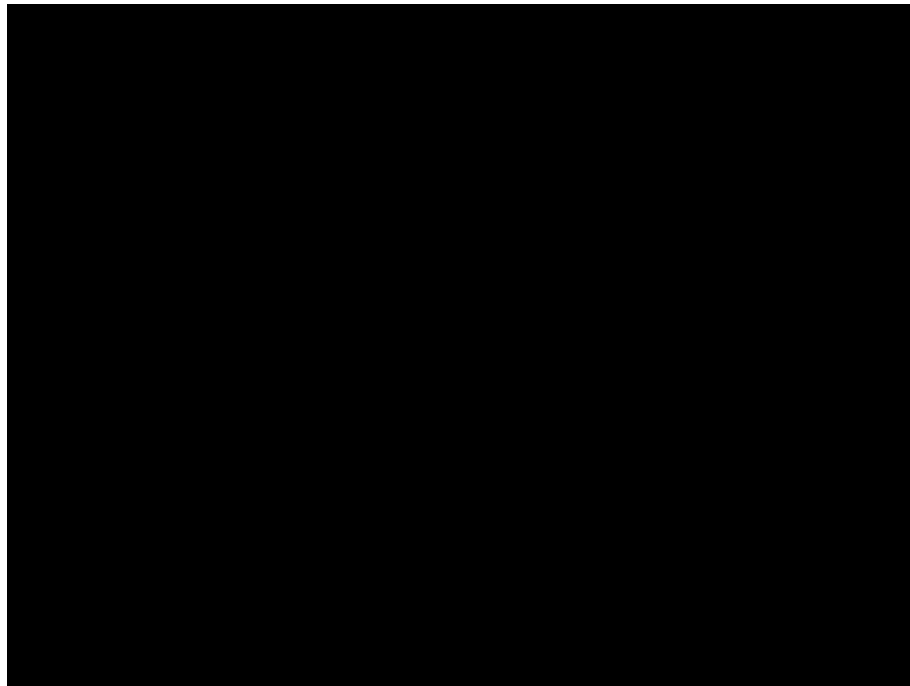


- Make it easier for third parties to set up, operate commons.
- Make it easier for third parties to contribute to the open source Gen3 software base.
- Build a community of researchers using Gen3 to explore and analyze data.

Helm Charts for Desktop

Jawad Qureshi, Center for Translational Data Science, University of Chicago

Helm Charts for Desktop



Helm Charts for *AWS*

Jawad Qureshi, Center for Translational Data Science, University of Chicago

Helm Charts for AWS

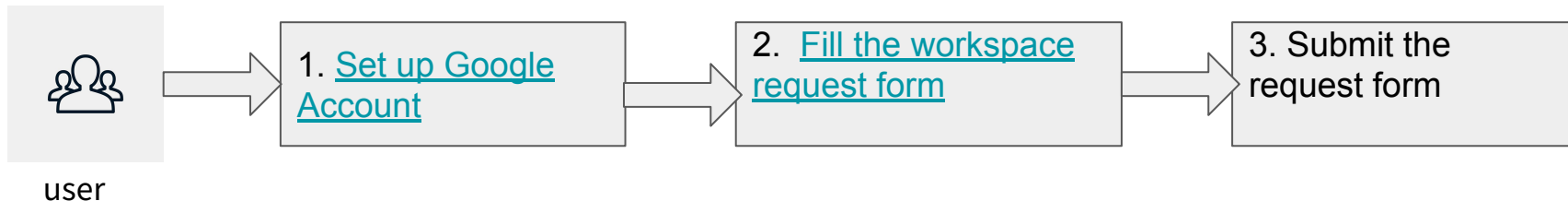


Helm Charts for GCP

Plamen Martinov, Mikisha Patel, Urvi Sheth; Open Commons Consortium

Helm Charts for GCP Use Case

- Enable students to start Gen3 resources with \$300 GCP credits and billing tight to their own account automatically while using Gen3 resources with OCC security boundary. For more information go to <https://pandemicresponsecommons.org/blog/>

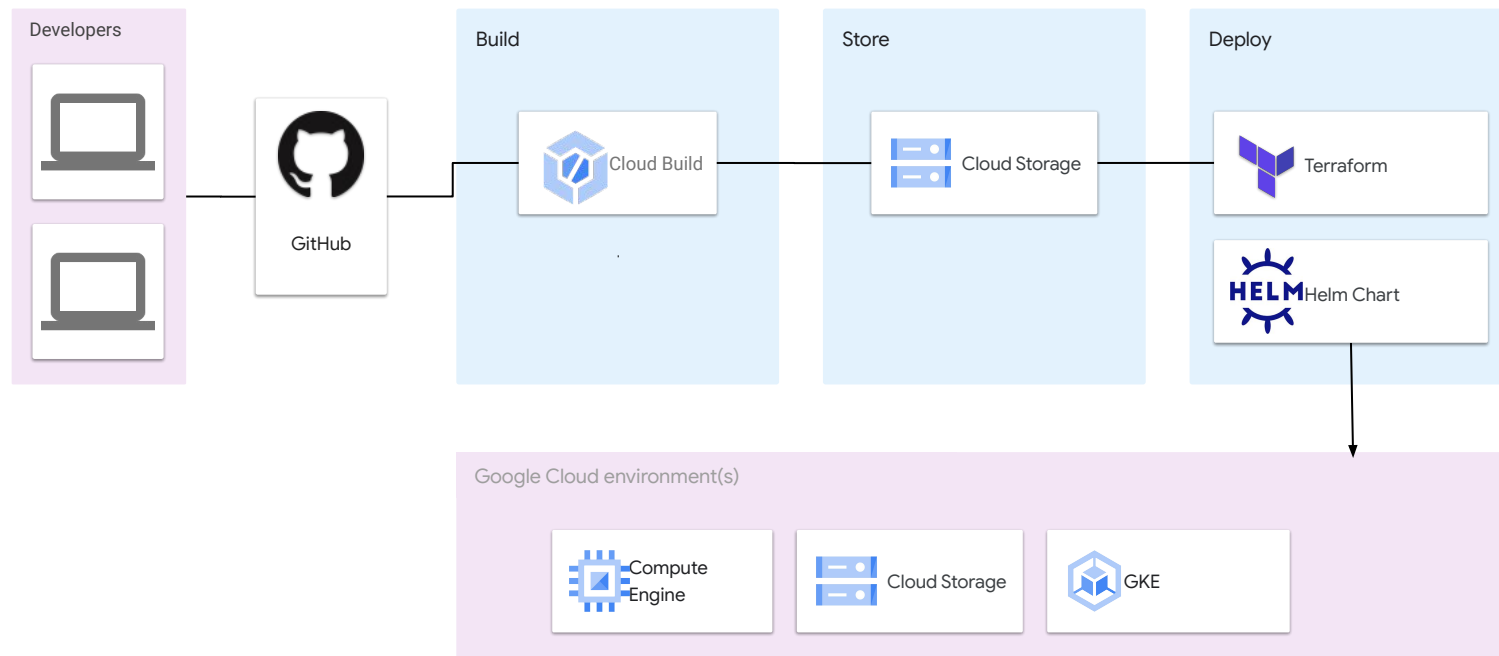


- Enable organizations natively on GCP to use Gen3 within their own secure boundary with ease of setup.

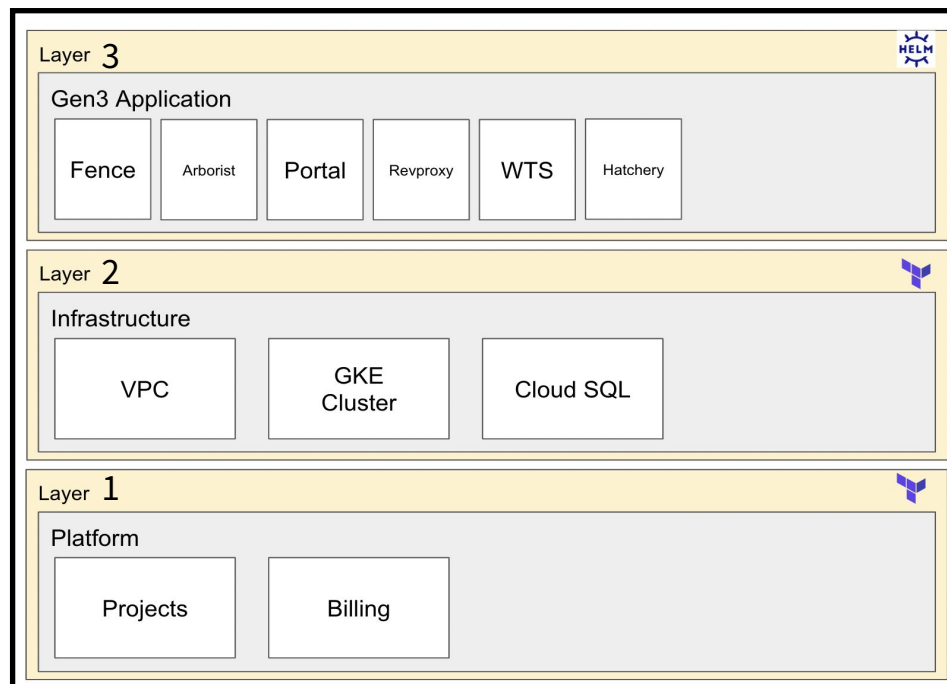
Design Overview



Infrastructure pipeline



Build a Gen3 Project in GCP



Demo

Cloud Automation Deployment for OpenStack

Somesh Nistala, Eirian Perkins; New Zealand eScience Infrastructure

About Aotearoa Genomic Data Repository

Data Repository for Taonga Species -

<https://data.agdr.org.nz/>

This drove the need to deploy GEN3 on NeSI own Cloud system... on premise deployment

GEN3 is used for the Aotearoa Genomics Data Repository and Rakeiora project (prototype application)

- <https://data.agdr.org.nz/>
- <https://rakeiora.data.nesi.org.nz/login> (URL will change)



Kākāpō¹



Snapper²



**genomics
aotearoa**

On-premise Deployments

1. Docker-compose deployment for submitting data

<https://repo.data.nesi.org.nz/>

- Prototyping first in Docker-compose
- HPC platform

2. Kubernetes deployment for the application

<https://data.agdr.org.nz/>

- Data storage on NeSI storage via Globus
- FlexiHPC platform (Openstack)

genomics aotearoa Aotearoa Genomic Data Repository

Projects Exploration Metadata Dictionary Open To Collaborate About Support Login

kākāpō chick photo by Dianne Mason, 2009 CC2.0

Aotearoa Genomic Data Repository

The Aotearoa Genomic Data Repository has been jointly developed by Genomics Aotearoa and NeSI to provide a secure place for the New Zealand research community to store and share genomic data. Version 1.0 of the repository is up and running, but we are still adding functionality, so you will see ongoing improvements as we implement new features.

Projects

Submit Data

28 Projects

35 Experiments

2301 Files

0% 25% 50% 75%

KAKAPO AGDR00019 LAMPREY TREVALLY others

Dictionary v0.2.1-9-g520a239 Submission v2021.03 Portal vdevelop-4.3

Aotearoa Genomic Data Repository is hosted on NeSI. NeSI's privacy policy

Gen3 Changes

- Creating and assigning (Minting) new Digital Object Identifiers for projects visible on the Discovery page
- Deployment NeSI own metadata-service to easily support biocultural (BC) and traditional knowledge (TK) labels and notices
 - BC/TK information must be dynamically retrieved per project

A screenshot of a project page on the Gen3 platform. The page title is "Complete genomes of two extinct New Zealand passerines". It is divided into "Project Details" and "Summary Information".

Project Details

Description Human intervention, pre-human climate change (or a combination of both), as well as genetic effects, contribute to species extinctions. While many species from oceanic islands have gone extinct due to direct human impacts, the effects of pre-human climate change and human settlement on the genomic diversity of insular species and the role that loss of genomic diversity played in their extinctions remains largely unexplored. To address this question, we sequenced whole genomes of two extinct New Zealand passerines, the hula (*Heteralocha acutirostris*) and South Island kōkako (*Callaeas cinereus*). Both species showed similar demographic trajectories throughout the Pleistocene. However, the South Island kōkako continued to decline after the last glaciation, while the hula experienced some recovery. Moreover, there was no indication of inbreeding resulting from recent mating among closely-related individuals in either species. This latter result indicates that population fragmentation associated with forest clearing by Māori may not have been strong enough to lead to an increase in inbreeding and exposure to genomic erosion. While genomic erosion may not have directly contributed to their extinctions, further habitat fragmentation and the introduction of mammalian predators by Europeans may have been an important driver of extinction in hula and South Island kōkako.

Investigator Name Michael Knapp
Investigator Affiliation University of Otago
Contact michael.knapp@otago.ac.nz
Funding Source FORMAS (2015-676) to L.D.; the Swiss National Science Foundation to N.D. [P25KPF3_165031 and P300PA_177845]; University of Otago FBRF grants to N.D., M.K. and B.C.R.

Summary Information

Species from this study *Heteralocha acutirostris*, *Callaeas cinereus*

Experiments 1
Raw Files 24
Processed Files Not available
Application Form [Apply for access](#)

Citation Dussex Nicolas, von Seth Johanna, Knapp Michael, Kardalsky Olga, Robertson Bruce C. and Daley Love 2019 Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction Biol. Lett. 15:20190491-20190491

Project ID TAONGA-AGDR0004

A screenshot of a metadata panel titled "Traditional Knowledge, Biocultural Labels, and Notices". It features a "Southwest App" header and a list of labels with corresponding icons:

- BC Non-Commercial (BC NC) with a crossed-out dollar sign icon
- TK Outreach (TK O) with a hand icon
- TK Attribution (TK A) with a double-headed arrow icon
- TK Verified (TK V) with a checkmark icon

The "Project ID" is shown as "TAONGA-TESTPROJECT". A red "Example" watermark is overlaid on the panel.

K8s Deployment

We started from gen3OnK8s.md deployment approach

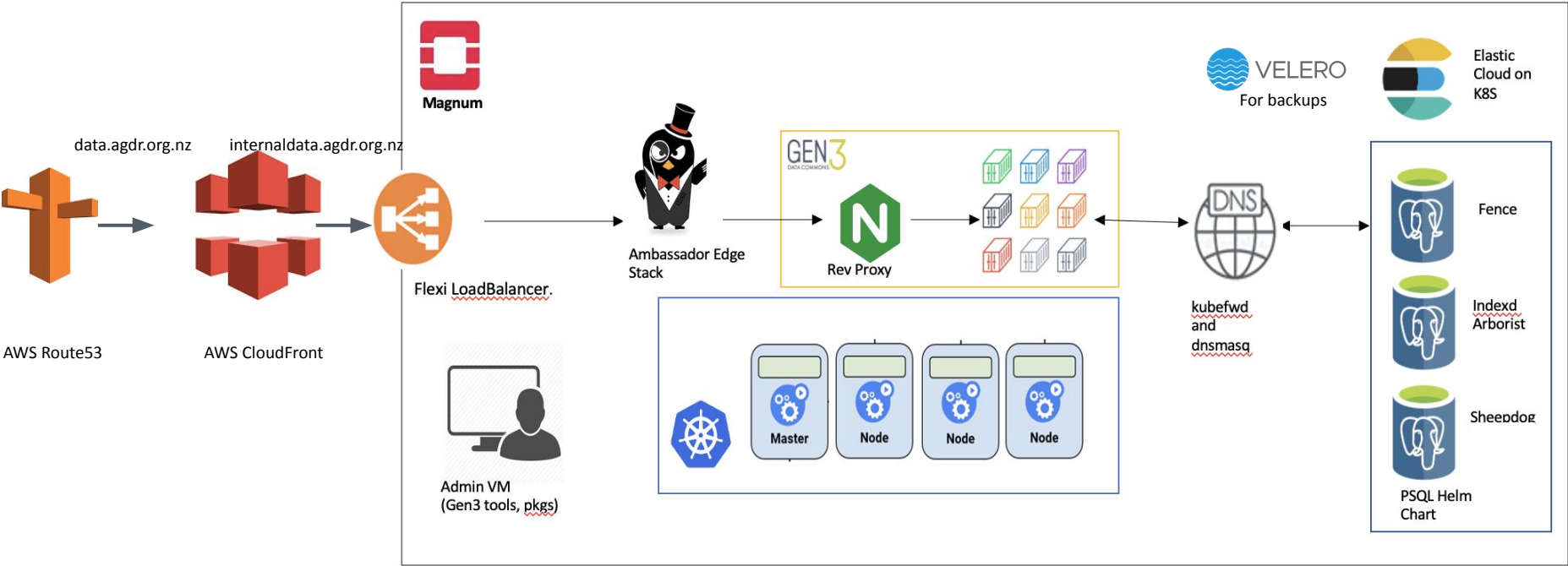
- <https://github.com/uc-cdis/cloud-automation/blob/master/doc/gen3OnK8s.md>

To deploy the solution on NeSi cloud:

- use Kubernetes deployment files present under cloud-automation/kube/services at master
 - uc-cdis/cloud-automation
- gen3 scripts cloud-automation/gen3 at master
 - uc-cdis/cloud-automation,
- but excluded the scripts present under cloud-automation/tf_files at master
 - uc-cdis/cloud-automation .

The required Kubernetes infrastructure to deploy the gen3 solution has been created using openstack core automation or using UI.

K8s Architecture



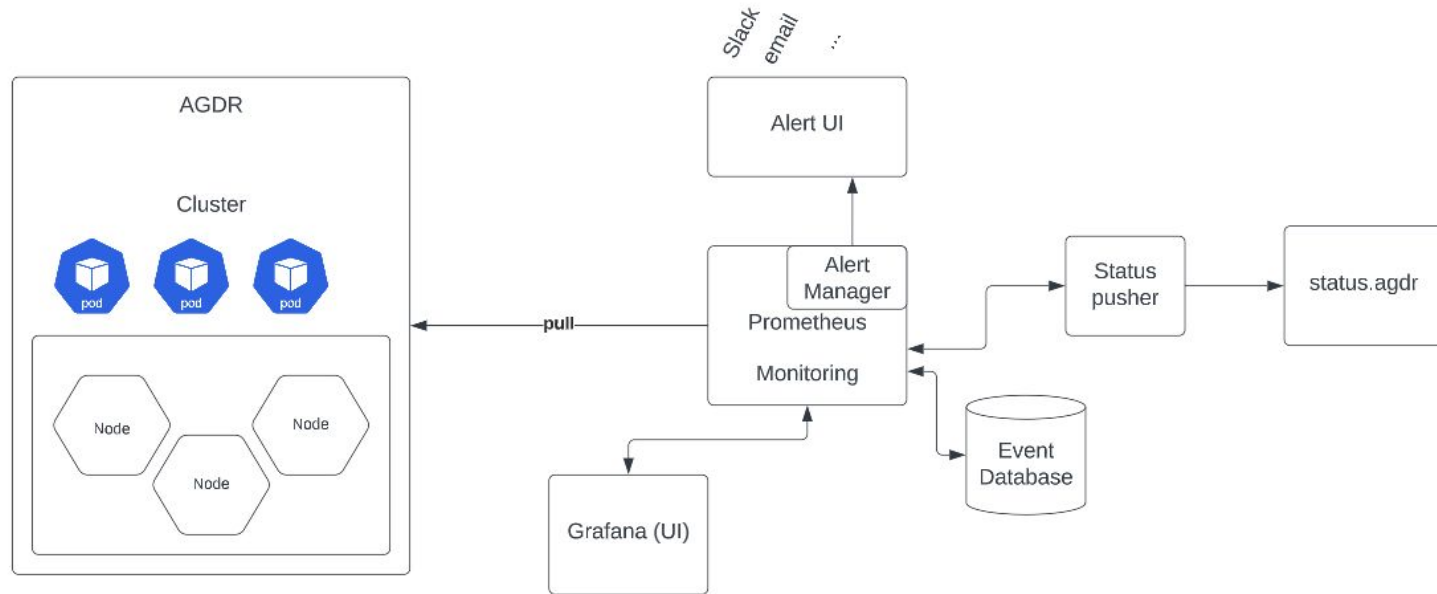
Deployment Steps

1. Setup Ubuntu instance as Gen3 admin VM
2. Customize configurations
 - Adjust **cloud-automation** to deploy on the OpenStack e.g decouple specific AWS configs
 - Adjust **cloud-automation** repo the scripts to custom service, AGDR metadata-service, Images, adjust nginx rev proxy
 - Updated fence, user.yaml , db credentials, etc in Gen3Secrets, etlMappings, manifest.json,etc in cdis-manifest, and other configs
3. Deploy k8s cluster on Flexi Infrastructure using on OpenStack CLI
4. Whitelist Kube API server access
5. Configure DNSMASQ packages and **kubefwd** to enable access to k8s from adminVM
6. Deploy Postgresql Helm charts, configure fence_db, arborist_db, indexd_db and metadata_db database
7. Roll Gen3 deployment
8. Deploy Ambassador Edge Stack & Configure AWS route 53
9. Configure AWS CloudFront for internal access during the maintenance windows
10. Deploy ECK stack for logging
11. Deploy Velero for the k8s backup and recovery

Step3:

```
openstack coe cluster create ProdAgdrGen3 \  
--cluster-template AGDR_NoFlotIp \  
--keypair xxxxxxxx \  
--node-count 2 \  
--master-count 3 \  
--master-flavor m3.medium \  
--flavor m3.xlarge \  
--fixed-network xxxxxxxxxxxxxxxxxxxxxx \  
--fixed-subnet xxxxxxxxxxxxxxxxxxxxxx \  
--labels monitoring_enabled=false \  
--labels floating_ip_enabled=false \  
--labels master_lb_enabled=true \  
--labels master_lb_floating_ip_enabled=true \  
--labels auto_healing_enabled=false \  
--labels auto_scaling_enabled=true \  
--labels min_node_count=2 \  
--labels max_node_count=5 \  
--labels  
admission_control_list="NodeRestriction,NamespaceLifecycle,L  
mitRanger,ServiceAccount,ResourceQuota,TaintNodesByCondition,  
Priority,DefaultTolerationSeconds,DefaultStorageClass,Storage  
ObjectInUseProtection,PersistentVolumeClaimResize,MutatingAdm  
issionWebhook,ValidatingAdmissionWebhook,RuntimeClass"  
--merge-labels
```

Monitoring System (in progress)



Documentation

- Not infrastructure independent
 - Lots of trial and error to make it work with OpenStack
- Troubleshooting to improve

- Gen3 Website: <https://gen3.org/>
- Gen3 Helm Documentation: <https://github.com/uc-cdis/gen3-helm>
- Gen3 User Forum: <https://forums.gen3.org/>
- Gen3 on Slack:
<https://docs.google.com/forms/d/e/1FAIpQLSczyhhOXeCK9FdVtpQpelOHYnRj1EAq1rwwnm9q6cPAe5a7ug/viewform>
- Email support: support@datacommons.io

Questions?

