EMBASSY: Research use cases

Erik van den Bergh
Cloud application architect

evdbergh@ebi.ac.uk
The Pan-Cancer Analysis of Whole Genomes Project (PCAWG)

The International Cancer Genome Consortium (ICGC)
Enabling rapid cloud-based analysis of thousands of human genomes via Butler, S. Yakneen et al, 2017 - https://doi.org/10.1101/185736
Butler runs on EMBASSY
Butler data NFS exports

- Important part of Pancancer is delivering high-throughput data
- EMBASSY allows us to export data straight from the archive filers as NFS
- Backed by Isilon, so load balanced
- 1PB of storage in use at the moment
- Data does not leave premises
  - Compliant with policy
Marine metagenomics
Use-case: metagenomics assembly pipeline

- Pipeline to assemble **all** metagenomics studies in the ENA in a standardised way
  - But focusing on Marine environments first
- Workflow of QC and gene prediction tools
  - Ported to CWL as part of jump to EMBASSY
A few iterations...

- The non-CWL solution:
  - LSF (OpenLava) on the cloud
  - Terraform + Ansible
Bioimaging

• Science demonstrator for EOSC Pilot project
• Goal:
  • “To analyse publicly available image data from genome-scale RNAi screens and small molecule screens to gain insights into cellular functions of human genes and try to identify inhibitors of these functions”
• Previously running on EBI cluster
  • Openlava cloud setup was re-used
• Currently still in initial phase
Phenome and Metabolome aNalysys (Phenomenal)

- VRE for medical metabolomics research
- Built on Open Source tools
- Developed with large-scale in mind
Quickstart

New to Galaxy, start here!
First time users can watch the example workflow tutorials on YouTube and run the example workflows with the instructions given in these tutorials:

- Fluxomics
- LC-MS
- NMR
- Statistics

You can also try our interactive tours.

Support

- Learn Galaxy (official documentation)
- YouTube tutorials
- How to get an account
- Request help (ticket)
- Online training offered by EMBL-EBI
- More about setting up a private instance
Architecture layers

Containers
C. Orchestrator
Container Engine
VMs OS
Hardware/IaaS

K8s package manager
HELM

Container Orchestration
kubernetes
docker

Infrastructure Provisioning
TERRAFORM

Software Provisioning
CoreOS

CiscoCloud / mantl
kubernetes-incubator / kargo
mcapuccini / KubeNow

From: P. Moreno, 2017
Kubernetes @ EBI

- Current exploration towards providing container orchestration and management
- Openshift initial project now running for 1yr
- ‘Vanilla’ Kubernetes-as-a-service being explored as part of Elixir IS
- Will most likely be hosted on top of EMBASSY