Barcelona Techno Week 2023

Cloud for scientific computing

Dave Morris
Institute for Astronomy
Edinburgh University

Cloud for scientific computing

… isn’t a specific ‘thing’

Cloud is an infrastructure level component

just another part of the stack

Linux for scientific computing

Storage for scientific computing

Network for scientific computing

Cloud for scientific computing

D.Morris
Institute for Astronomy,
Edinburgh University

Cloud for scientific computing

(*) based on personal experience

Using cloud for science 2000...2023

From cloud advocate in early 2000’s to developing science platforms in 2023

- Technical development
- Financial environment
- Science user cases

D.Morris
Institute for Astronomy,
Edinburgh University

Technical development 2000…2015

(*) based on personal experience

Limited options

Few academic cloud providers

Technically, commercial cloud is widely available

Difficult to propose projects using commercial cloud

- Cloud is see as impermanent, transitory
- Ethics issues, funding issues

Technical development 2000…2015

(*) based on personal experience

Using commercial cloud myself since early 2000's

Quick and easy virtual machines
for prototypes and experiments

Invaluable learning experience
root account and public IP address
impossible within an institute

Hosting project infrastructure
- Svn repository
- Issue tracker
- Package repository

< £10 / month,
not worth the paperwork
to claim

D.Morris
Institute for Astronomy,
Edinburgh University

Technical development 2015...2020

(*) based on personal experience

...2015 Using KVM and libvirt to manage VMs manually.

- Shell scripts to automate create and delete

2015 Openstack becomes mainstream

2017 Institute begins to experiment with on-premises cloud

- Early adopter – learning as we go.
- **Huge** learning curve
  - Openstack was complicated and thinly documented
  - Documentation covers *how* to adjust reciprocal flange, but nothing explained *why* you would want to.

Technical development 2015...2020

(*) based on personal experience

2015 Docker becomes mainstream

Early adopters - exactly what we had been waiting for.

Docker still has security concerns

System admins wary of installing Docker

Physical hypervisor considered a high value target

Virtual machines provide isolation

D.Morris
Institute for Astronomy,
Edinburgh University

Technical development 2015...2020

(*) based on personal experience

2018 Kubernetes becomes mainstream

System admins happier about running Docker

Using Openstack VMs to provide isolation

- Openstack cloud
- Docker containers
- Kubernetes orchestration

Even now, few system admins taking the extra step to
Run Kubernetes in rootless containers

D.Morris
Institute for Astronomy,
Edinburgh University

Technical development 2020...2023

(*) based on personal experience

2020 Science platform for astronomy project

Openstack cloud

Docker containers

Kubernetes orchestration

(*) Issues with integration of Kubernetes and Openstack
mostly documentation but some bugs too.

Machine learning use cases

Data access becomes limiting factor

D.Morris
Institute for Astronomy,
Edinburgh University

Docker containers

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment …2015
(*) based on personal experience

Traditional model for grant applications

Specific section in the proposal for hardware

Guess how much hardware the project will need

(*) before you write the software

Cloud is seen as impermanent / transitory

(*) grant has fixed lifetime, what happens after ?

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2015…2020

(*) based on personal experience

Institutes begin to invest in on-premises cloud compute

Proposals can include quotes for on-premises cloud

(*) hardware specifics are SEP

Data persistence is backed by institute policy

(*) ongoing maintenance is paid for by the institute

Financial environment 2015...2020

(*) based on personal experience

Institutes begin to invest in on-premises cloud compute

Shaky start – everyone is learning

System admins learning how to manage Openstack

Projects learning how to estimate requirements

Institute learning how to manage billing

- Who pays for an undeleted VM ?
- What detail goes on the bill ?

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2020...

(*) based on personal experience

Science and Technology Funding Council
(STFC IRIS)

National level federated cloud compute

European Open Science Cloud
(EOSC)

Proposals have to justify NOT using cloud

Data persistence is backed by national policy

(*) Petabyte scale storage facilities available

Commercial cloud is still seen as impermanent / transitory

(*) Projects have a duty of care to curate the data

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2020...

(*) based on personal experience

National level federated cloud compute — Science and Technology Funding Council (STFC IRIS)

- Annual resource request process
  - Need to estimate resource requirements a year in advance
- Finite compute resources
  - Resource allocation optimized to maximize use
  - Spare cycles allocated to batch processing
  - Limited head room for dynamic scaling

ROYAL OBSERVATORY EDINBURGH

Next - Science users

Science users

Dark matter

Gravitational lensing

Gravity waves

Researchers want to think about science

Supernova

Fast radio bursts

Exoplanets

operating systems

filesystems

Researchers don’t want to have to learn about cloud

databases

container orchestration

datacenters

virtual machines

D.Morris
Institute for Astronomy,
Edinburgh University

Science users

Quick poll – how do you use cloud compute?

Everybody uses ‘apps’ in the cloud

Online email (GoogleMail etc)

Online documents (GoogleDocs, Overleaf etc)

Online identity (ORCID etc)

(*) No issues with private data in the cloud

Online code (GitHub etc)

Everything I have written in the last 20 years is public on GitHub

No, I don’t have a backup.

D.Morris
Institute for Astronomy,
Edinburgh University

Science users

Quick poll – how do you use cloud compute?

Mainly engineers(*) interacting with cloud-compute

- Openstack
- Ansible
- Docker
- Kubernetes
- Helm

Scary stuff for science users

Dynamic fast moving, transitory

Quick turn around – delete and re-create

DevOps

‘cattle not pets’

(*) they may be academic staff, but working as engineers

Cloud providers

Quick poll – who is using your service ?

1/3 Researchers interacting directly with cloud-compute

…. ….

2/3 Project level

Team of people with a range of skills

Cloud compute is handled by software engineers

Researchers interact with science interface

(*) Based on data from Cambridge HPC Research Computing Services

D.Morris
Institute for Astronomy,
Edinburgh University

Cloud providers

Quick poll – how are they using your service ?

1/3 Single instance virtual machines

Created and operated manually

…. ….

2/3 Full stack orchestrated system

Automated creation and deployment

Kubernetes, ClusterAPI etc.

(*) Based on data from Cambridge HPC Research Computing Services

Science platforms

Science platforms

NAVO - NASA Astronomical Virtual Observatories

‘login and do science in < 10 min’

JupyterLab, Linux terminal, Python shell

Data services stay on-premises

Portal is on AWS – but the users don’t know

Designed to be platform agnostic

Partners are using different cloud platforms

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

Rubin Observatory science platform (RSP)

https://data.lsst.cloud/

Deployed using Kubernetes running on Google Cloud

Using components from a number of other science platforms

IVOA standards to integrate data access between the components

https://ivoa.net/

Data access services from OpenCADC in Canada.

https://github.com/opencadc

User interface components from Firefly developed by Caltec-IPAC .

https://github.com/Caltech-IPAC/firefly

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

Rubin Observatory science platform (RSP)

https://data.lsst.cloud/

Deployed using Kubernetes running on Google Cloud

Using components from a number of other science platforms

IVOA standards to integrate data access between the components

https://ivoa.net/

Data access services from OpenCADC in Canada.

https://github.com/opencadc

User interface components from Firefly developed by Caltec-IPAC .

https://github.com/Caltech-IPAC/firefly

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

Rubin Observatory science platform (RSP)

https://data.lsst.cloud/

Deployed using Kubernetes running on Google Cloud

Wide Field Astronomy Unit (Edinburgh University)

Deploying RSP on STFC funded Openstack cloud

Contributing changes back to upstream project

CfA - Center for Astrophysics Research and Development (Harvard & Smithsonian)

Pulumi - platform agnostic infrastructure as code

Using new tools to make it easier to deploy the Rubin Science Platform on a variety of different cloud platforms.

D.Morris
Institute for Astronomy,
Edinburgh University

The end ...