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 users
- Science platforms

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

- Source repository
- Issue tracker
- Package library

< £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

Technical development 2015...2023

(*) based on personal experience

2018 Kubernetes becomes mainstream

System admins happier about running Docker

Using Openstack VMs to provide isolation

Standard deployment

- 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 2000...2023

(*) based on personal experience

2000...2015

Very few academic cloud providers
HPC and grid were the main players

2015 Openstack becomes mainstream

2015 Docker becomes mainstream

2018 Kubernetes becomes mainstream

Everyone is using ‘cloud’

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2000…2023

(*) based on personal experience

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2000…2015

(*) based on personal experience

Traditional model for grant applications

Specific section in project proposals for hardware

Commercial cloud is seen as impermanent / transitory

- Project has fixed lifetime - what happens after ?

Moore’s law helps sustain legacy systems

- Funding for new projects helps to support general IT services
- Data from old projects is a fraction of the size of new projects

Financial environment 2015…2020

(*) based on personal experience

Institutes begin to invest in on-premises cloud compute

Project proposals may include costs for on-premises cloud

Funding bodies require data management *beyond* the end of the project

Data persistence is backed by institute policy

Issuing a DOI implies a commitment by the institute

Financial environment 2015...2020

(*) based on personal experience

Institutes begin to invest in on-premises cloud compute

2017 Edinburgh university Eleanor Openstack system

Shaky start – everyone was learning

- System admins learning how to manage Openstack
- Projects learning how to use cloud-computing
- Finance learning how to manage billing

2023 Eleanor Openstack system is standard part of research services

- Generic cloud compute available to staff and students

Financial environment 2020...2023

(*) based on personal experience

National level federated cloud compute

European Open Science Cloud
(EOSC)

UK Science and Technology Funding Council
(STFC IRIS)

New projects have to justify why they are not using cloud

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2020...2023

(*) based on personal experience

European federated cloud compute (EOSC)

Openstack deployments at key sites

Generic cloud compute for a range of projects

European level identity provider

OpenID-Connect delegation to institute level authentication

Project level storage systems

Rucio distributed storage

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2020...2023

(*) based on personal experience

UK national federated cloud compute

Openstack deployments at key sites

- Generic cloud compute for a range of projects

National level identity provider

- OpenID-Connect delegation to institute level authentication

National level storage provider

- Peta byte Ceph object store
- Peta byte Rucio distributed storage

D.Morris
Institute for Astronomy,
Edinburgh University

Financial environment 2020...2023

(*) based on personal experience

UK national federated cloud compute

Annual resource request process

- Need to estimate resource requirements a year in advance
- Time for service providers to purchase and deploy the hardware

Finite compute resources

- Resource allocation optimized to maximize use
- Limited head room for dynamic scaling

Potential opportunity to use commercial cloud for scale-out ?

D.Morris
Institute for Astronomy,
Edinburgh University

Science use cases

D.Morris
Institute for Astronomy,
Edinburgh University

Science use cases

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

Researchers

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)

Online code (GitHub etc)

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

No, I don’t have a backup.

(*) No issues with private data in the cloud

D.Morris
Institute for Astronomy,
Edinburgh University

Researchers

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

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

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

Data analysis for astronomy

- New missions are producing peta bytes of data

Compute co-located with the data

- “bring the code to the data”

Common structure

- JupyterHub notebook service
- Docker containers for custom code
- User storage for results
- Hosted on generic cloud compute

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

NAVO - NASA Astronomical Virtual Observatories

‘login and do science in < 10 min’

JupyterLab, Linux terminal, Python shell

Shared collaboration space

Data services stay on-premises (for now)

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

Designed to be platform agnostic

Partners may be using different cloud platforms

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

CANFAR - Canadian Advanced Network for Astronomical Research

2013 – Openstack VM based system

Users interact directly with the Openstack portal

Expert users only

2021 – K8s based system on Openstack

Users interact with the CANFAR portal

Expert users configure project specific science applications

Normal users just use science applications

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

Rubin Observatory science platform (RSP)

https://data.lsst.cloud/

Initial data preview (simulated data)

“a production environment to support Data Preview releases”

Deployed using Kubernetes running on Google Cloud

D.Morris
Institute for Astronomy,
Edinburgh University

Science platforms

Rubin Observatory science platform (RSP)

Other projects are already porting RSP to other platforms

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

Cloud computing 2000...2023

(*) based on personal experience

D.Morris
Institute for Astronomy,
Edinburgh University

Cloud computing 2000...2023

(*) based on personal experience

2000...2015

Very few academic cloud providers

2015 Openstack becomes mainstream

2015 Docker becomes mainstream

2018 Kubernetes becomes mainstream

Everyone is using ‘cloud’

D.Morris
Institute for Astronomy,
Edinburgh University

Cloud computing 2000...2023

(*) based on personal experience

2015 Openstack becomes mainstream

2015 Docker becomes mainstream

2018 Kubernetes becomes mainstream

Everyone is using ‘cloud’

2020 Data becomes too big

Commercial cloud makes sense

D.Morris
Institute for Astronomy,
Edinburgh University

Dave Morris
Institute for Astronomy
Edinburgh University

dmr@roe.ac.uk