

Bringing the Computing to the Data

William O'Mullane, Dave Morris, John Hoar

Gaia Science Operations Centre
European Space Astronomy Centre
European Space Agency

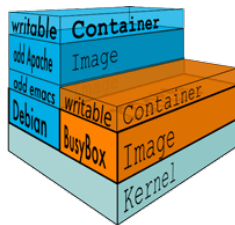
ADASS XXV October 26th 2015 CAASTRO, Sydney, Australia

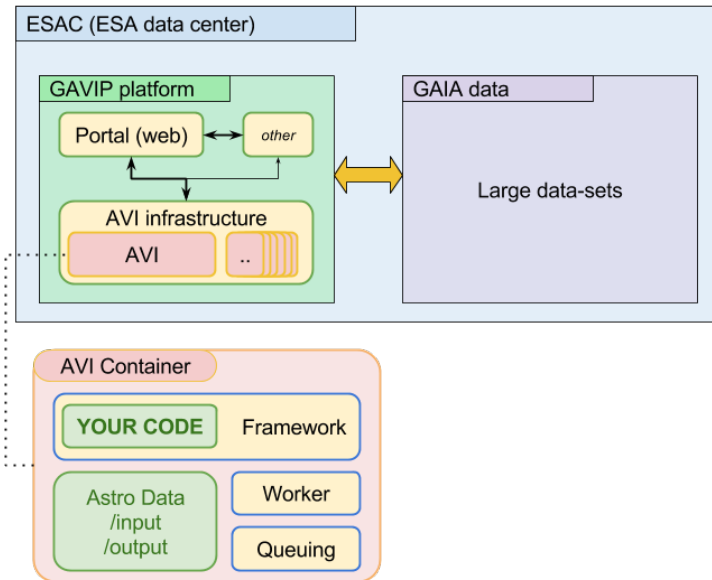
- Doing a few virtualization/container projects in ESAC and ROE
- Wondered what others are doing
- Offered to do ADASS survey for this talk - first ADASS survey talk ?
- Survey sent to ADASS mailing list.
- Will explain a little about the technology and what we are doing
- Will try to summarise the general feeling from the survey

17 textual responses

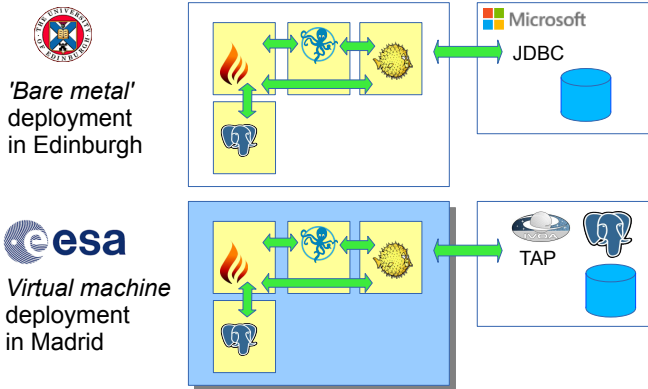
- Institute and project goal
- Which virtualization/container technologies are you using ?
- Are you looking some way of 'bringing the code to the data' ?
- Do you intend to allow third party code to run in your data centre ? and code handling
- What role do you see technologies like Docker playing in your solution ?
- How do you deal with the configuration and deployment of virtual machines and containers ?
- How widely has virtualization been adopted within your project/department/institute ?
- Has your project/department/institute adopted any other ideas from the DevOps community ?
- Success and issues ..

- Packaged piece of software with complete file system it needs to run!
- Portable unit - standard shape and connectors
- Boots in fraction of a second
- Lightweight - 10-100 times containers per machine w.r.t to VMs.
- Docker leading the pack now for containers - others exist ..
Solaris Zones FreeBSD Jails



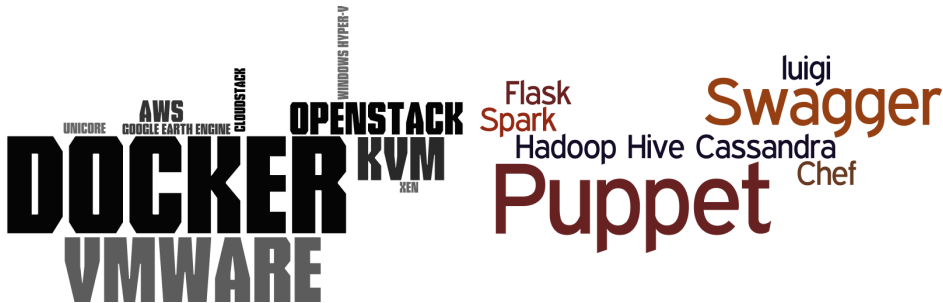


Two different deployment environments



Also doing maven builds and testing with containers

- 1 ASTRON - LOFAR processing
- 2 AUI inc - ALMA
- 3 Australian Astronomical Observatory - Archive access
- 4 CADAC - Big Storage/Processing
- 5 CEA Saclay - Svom:GRB mission
- 6 ESA/ESAC - Gaia AVI
- 7 ESO - ALMA almascience.org
- 8 Gemini Observatory - DP Arc
- 9 Google - remote sensing
- 10 Institute of 4D Technologies -FLARECAST
- 11 Johns Hopkins University - SciServer sciserver.org
- 12 LERMA Observatoire de Paris - Theory
- 13 NAOC - China VO and Astroinformatics
astrocloud.china-vo.org
- 14 Observatoire de Paris - General
- 15 Space Telescope Science Institute - Astro research
- 16 University of Edinburgh - Euclid
- 17 University of Melbourne and Swinburne - study cloud impact Astro
- 18 U of Maryland - ALMA Science Products



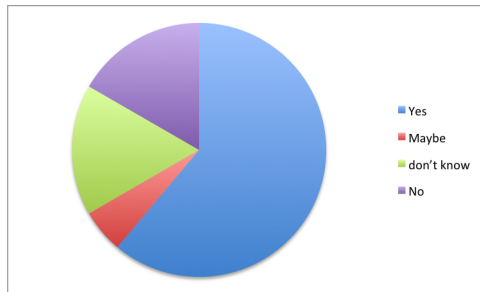
Have not listed mentioned languages .. less pertinent with virtualization.
Java mentioned 4 times ok it is a VM .. but **C++** and **Python** also mentioned many times.

Role of technologies like Docker .. Vital, key component, central to approach, allows isolation/robustness, main method for consistent execution, see after testing, maybe(2), none yet(3), none(3), VM is fast enough.

btw. Containers are not just for 3rd party code.

Collocation of external user code close to data.

- Yes(6). Python to container/setting up VMs to connect to RDS storage
- Almost a necessity
- Should allow to integrate new prediction algorithms
- Essentially yes
- Yes in a simple sense
- Yes anticipate data to grow faster than network bandwidth
- We do this every day (but I think they mean their own code)
- User code: No.
- Had some vague idea yes.



How do we ensure repeatable consistent deployments ..

- Docker repo of containers
- bamboo, docker compose
- software repo and VM naming convention
- git repos
- struggling with that
- build our own AWS image
- What is the question..
- **we don't**

- The number of VMs has been over 230
- most machines virtual(2)
- For services its a key part of our infrastructure
- Very common
- key part of our architecture/infrastructure (5) to ridiculous level.
- ingrained in our work
- widely adopted after some reluctance
- becoming part of the culture
- More than an experiment, less than a key part ...
- None
- Not relevant
- no real virtualization yet

Mainly these are for infrastructure not 3rd party code.

What is DevOps?

- Historically developers and admins/operations have been separate.
- Traditional methods don't scale to thousands of nodes in the cloud.
- DevOps brings operations and development together.

Why ask about DevOps?

- Interesting to see how people are using Docker.
 - As a stand-alone solution for a specific problem.
 - As part of a move towards a DevOps based process.

What results did we get?

- Only a few respondents knew what DevOps is.
- Of those who had, only two were looking to do more.

Conclusion?

- Not really of interest to the astronomy community.
- Interesting to see how things evolve in the future.

- Excellent! great success so far.
- **It works well for some things, not so well for others.**
- We have been successful in adopting virtualization (not getting user code to data yet though)
- Cloud good - containers flaky
- Not perfect, but good enough.
<http://astrocloud.china-vo.org>
- improves workflow
- satisfied at this point
- look very promising
- Still testing
- Early days - a lot of work left
- Security!(2) solves `rm -rf`
- struggle with the scale we need
- Concern integrating with existing tools - but ok.
- takes a lot of time to setup/maintain all the containers
- takes quite some time to move code from dev to container
- **transition - training - lack of skills**
- though all dependencies shipped within a container, still dependencies e.g. databases, other containers

- first ADASS survey .. need a better distribution mechanism
- Modest response
- Virtualization is reasonably ubiquitous (selection bias ?)
- Can we say containers are emerging as the technology this year ??

PLEASE RAISE YOU HAND IF YOU ARE USING OR
EXPERIMENTING WITH CONTAINERS
PLEASE RAISE YOU HAND IF CONTAINERS HAVE BEEN
REJECTED AS RUBBISH ALREADY

Thank You !

- first ADASS survey .. need a better distribution mechanism
- Modest response
- Virtualization is reasonably ubiquitous (selection bias ?)
- Can we say containers are emerging as the technology this year ??

**PLEASE RAISE YOU HAND IF YOU ARE USING OR
EXPERIMENTING WITH CONTAINERS**

PLEASE RAISE YOU HAND IF CONTAINERS HAVE BEEN
REJECTED AS RUBBISH ALREADY

Thank You !

- first ADASS survey .. need a better distribution mechanism
- Modest response
- Virtualization is reasonably ubiquitous (selection bias ?)
- Can we say containers are emerging as the technology this year ??

**PLEASE RAISE YOU HAND IF YOU ARE USING OR
EXPERIMENTING WITH CONTAINERS**

**PLEASE RAISE YOU HAND IF CONTAINERS HAVE BEEN
REJECTED AS RUBBISH ALREADY**

Thank You !

- first ADASS survey .. need a better distribution mechanism
- Modest response
- Virtualization is reasonably ubiquitous (selection bias ?)
- Can we say containers are emerging as the technology this year ??

**PLEASE RAISE YOU HAND IF YOU ARE USING OR
EXPERIMENTING WITH CONTAINERS
PLEASE RAISE YOU HAND IF CONTAINERS HAVE BEEN
REJECTED AS RUBBISH ALREADY**

Thank You !