



Deploying advanced computing infrastructures with ATLAS Open Data

Giovanni Guerrieri - Università degli Studi di Trieste - ATLAS Week Outreach Parallel Session - 14/10/21







- Software as a Service (SaaS) and Infrastructure as a Service (IaaS) have reshaped the way of data handling, analysis, storage, and sharing; particularly in multinational collaborations.
- We explore how a SaaS + IaaS approach can be adapted to modest scenario, using virtual machines and containers, for educational purposes; the target audience of this products proposal are trainers and small/medium institutions SysAdmin.
- We are using the current ATLAS Open Data (OD) and analysis examples.
- A set of prototypes is in place. They are based on Infrastructure as Code (IaC) tools such as <u>terraform.io</u> for cloud instance creation, and <u>docker</u> containers to obtain Jupyter{Lab,Hub}based environments for multi-user scenarios.

ATLAS Open Data

Users, settings and resources:

Educators use ATLAS OD resources to complement diverse HEP and data-analysis training programs.

Actually in use: institutional resources (e.g. in-house computers), free (e.g. MyBinder, CoLab) or commercial cloud (e.g. AWS, GCP...) to run hands-on sessions

Students bring their computers and use a VM to run long-term _____ projects, like a thesis or a university course, based on or profiting from a Jupyter UI

An easy setup is crucial!

Giovanni Guerrieri - Università degli Studi di Trieste - ATLAS Week Outreach Parallel Session - 14/10/21







What and Who



What's the idea

Develop a series of tested and production-ready recipes to effortlessly deploy analyses infrastructures, e.g. Jupyter{Lab,Hub}-based environments for multi-user scenarios.

Who will benefit from it

- Intermediate tier/local experts who need to deploy infrastructure for public purposes:
 - Classes
 - Workshops
 - Online seminars
 - etc.
- Entities who want to create *ad hoc* infrastructures:
 - Schools
 - Universities
 - Research groups
 - Privates

jupyterhub Example Notebook	Logout Control Panel
File Edit View Insert Cell Kernel Widgets Help	Trusted Kernel O
🖴 🕂 🦗 🙆 🏠 🛧 🔸 🕅 Run 🔳 C 🇭 Code 🗸 📼	Memory: 398 MB
In [1]: import ROOT	
In [1]: import ROOT	
WEICOME DO ORDYNOOI 0.21/00	
In []:	



laaS+SaaS



A concrete proposal

- Automated deployment of Jupyterub + single user notebooks on a cloud instance.
- Terraform IaC recipes ready to use in cloud infrastructures:
 - CERN OpenStack: up and running
 - AWS: up and running
 - GCP: work in progress
- Containers that can replace or enhance the existing ATLAS OD single-user VM
- Combinations of containers using <u>JupyterHub Docker Spawner</u> to deploy multiple-user JupyterHub with same single-user configuration



Pros

Useful features

- Docker containers construction
 - JupyterHub
 - ROOT
 - ML environments
 - Notebooks variety
- Infrastructure optimization
 - Spawner choice
 - Secure connection
 - Data persistance
 - External/shared volumes
 - Minimal spawn time

\checkmark
COOL Don't lose your work
COOL Share big datasets
COOL Setting everything up it's a matter of minutes

COOL Flexibility



open data

WORK IN PROGRESS



SaaS using Docker

Development, test and deployment of the containers, making use of the GitLab CI/CD infrastructure at CERN

- The containers are deployed in a registry
- At the user level: containers are managed with Docker app, hosting and using several environments (analogous to use <u>VMs+VirtualBox</u>)

https://hub.docker.com/r/atlasopendata/root_notebook



A dedicated hub for containers and recipes



Summary

- Combination IaaS + SaaS comes handy when developing an asymptotical "single-click" solution deployment for small and mediumsize institutions or individual trainers.
- Collection of containers that we can seamlessly use in a single-user mode (e.g. a laptop) or a multi-user solution like a JupyterHub.
- Solutions are designed to work standalone or in combination for a complete IaC + data analysis suite.

Resources

- Automated JupyterHub deployment on AWS
 - https://gitlab.cern.ch/atlas-open-data-iac-qt-2021/aws_automated_jh_deployment
- Automated JupyterHub deployment on OpenStack@CERN
 - https://gitlab.cern.ch/atlas-open-data-iac-qt-2021/automated_jh_deployment
- Notebook Images (with CI)
 - <u>https://gitlab.cern.ch/atlas-open-data-iac-qt-2021/notebooks-images</u>
- Quick, local JupyterHub deployment with Docker
 - https://gitlab.cern.ch/atlas-open-data-iac-qt-2021/orchestrateddocker-notebook-suite/-/tree/master/jupyterhub-local

A dedicated hub for containers and recipes



Summary

- Combination IaaS + SaaS comes handy when developing an asymptotical "single-click" solution deployment for small and mediumsize institutions or individual trainers.
- Collection of containers that we can seamlessly use in a single-user mode (e.g. a laptop) or a multi-user solution like a JupyterHub.
- Solutions are designed to work standalone or in combination for a complete IaC + data analysis suite.

≡ Tags	Hc	ome About Apps Data News/Blog Software EN
Jupyter Notebooks	C++/Python frameworks	Virtual Machine(s)
	Starre Selipsk Freide Sorre Selipsk Starres 3 do	



SATLAS

Thanks!

SATLAS

Backup



https://my.cloud.infn.it









CERN Firewall preventing "conda" packages installation

No JupyterHub container