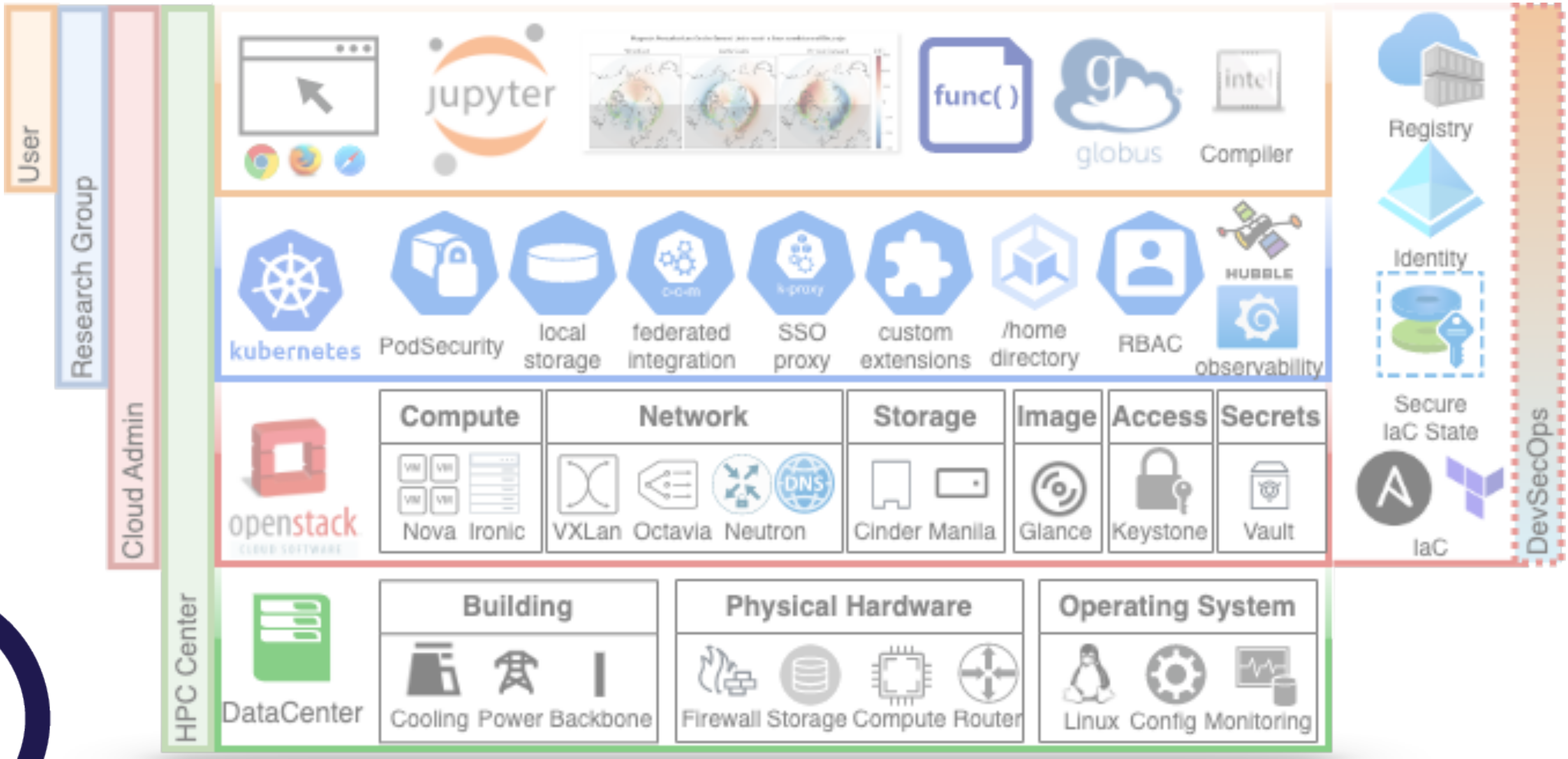
AUSTRIAN DATALAB AND SERVICES
TECHNISCHE UNIVERSITÄT WIEN
TU Graz
WU WIRTSCHAFTSUNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS
universität innsbruck
universität wien
MEDIZINISCHE UNIVERSITÄT WIEN
BOKU
JKU JOHANNES KEPLER UNIVERSITY LINZ



# Cloud Native Labs & OpenScience

Constanze Roedig, January 2023

# Disclaimer and Copyright Note

Icons, Symbols, Logos and similar copyrighted material used below belongs to the original authors.

The lectures are vendor neutral and usage of a particular tool does not imply recommendation or endorsement.

We are very grateful for the contributions from various University's departments, companies and our funding agencies.

In case of any oversights, please help us improve by raising the issue via email to support@austrianopencloudcommunity.org

# Stats

Funding Agency: BMBWF + participating universities

Start: 06.2020

End: 12.2024

Who should benefit: all Austrian universities

Which parts are effected: research, teaching, ZID

Goal: create an Austrian Open Cloud Community

PI: self organized scrum team (formal TU Wien)

Partners:

TU WIEN TECHNISCHE UNIVERSITÄT WIEN

TU Graz

WU WIRTSCHAFTSUNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

universität innsbruck

universität wien

MEDIZINISCHE UNIVERSITÄT WIEN

BOKU

JKU JOHANNES KEPLER UNIVERSITY LINZ

# Team

Carina Urbanke

Mojib Wali

Ariel Simulevski

Vlad Popescu-Vifor

Thomas Lahmer

Peter Kandolf

Thomas Weber

Franz Marko

Adam McCartney

Constanze Roedig

Elias Wimmer

# Agenda

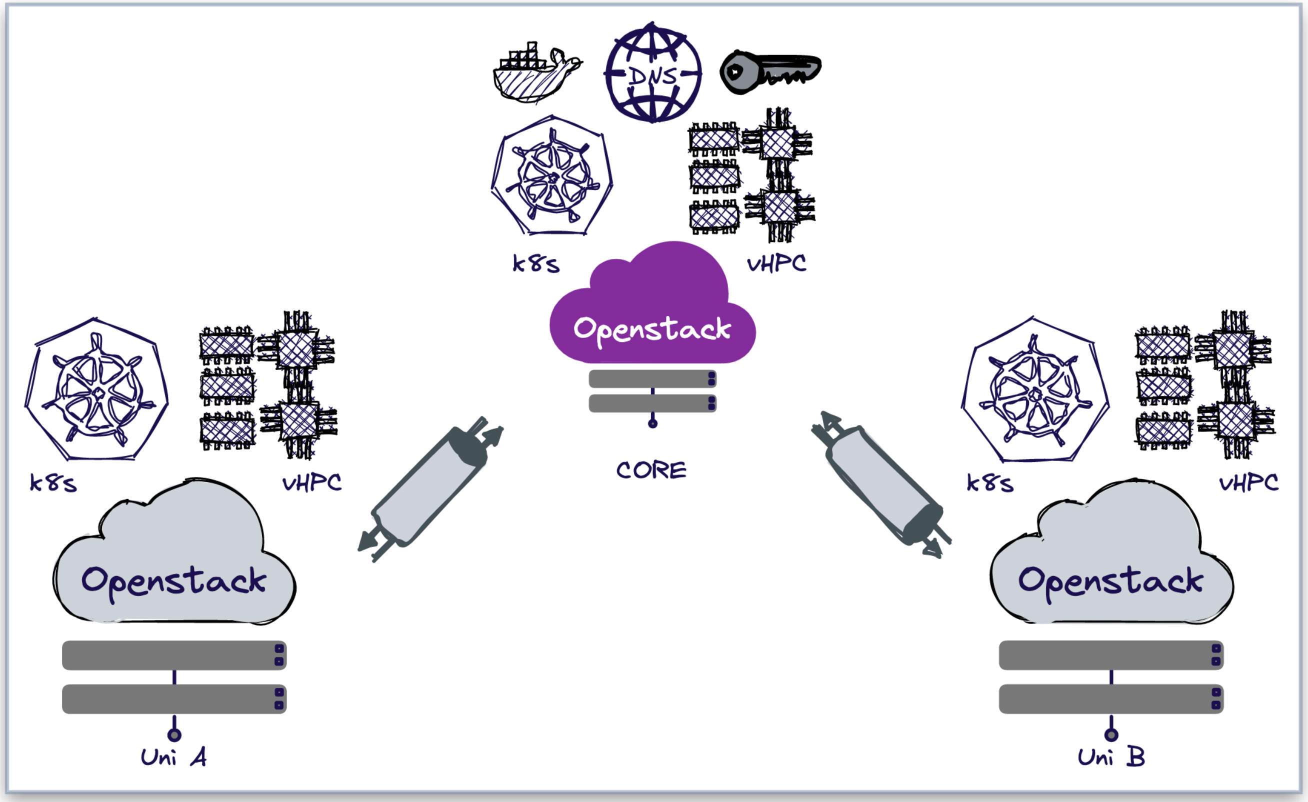
Focus: Use Cases for the Austrian Open Cloud Community

1) OpenScience with GeoSpace example (collaboration with JHU-APL)
   a) Our version of Jupyter to make this possible
   b) How we designed a Kubernetes-based platform to run this

2) Teaching Cloud Native, DevOps and Security
   a) Hands On Collaborative Approach
   b) Capture the flag /scavenger hunt type of gamified exams

3) Community Aspects: how you can get involved
   a) BluePrints and OpenSource

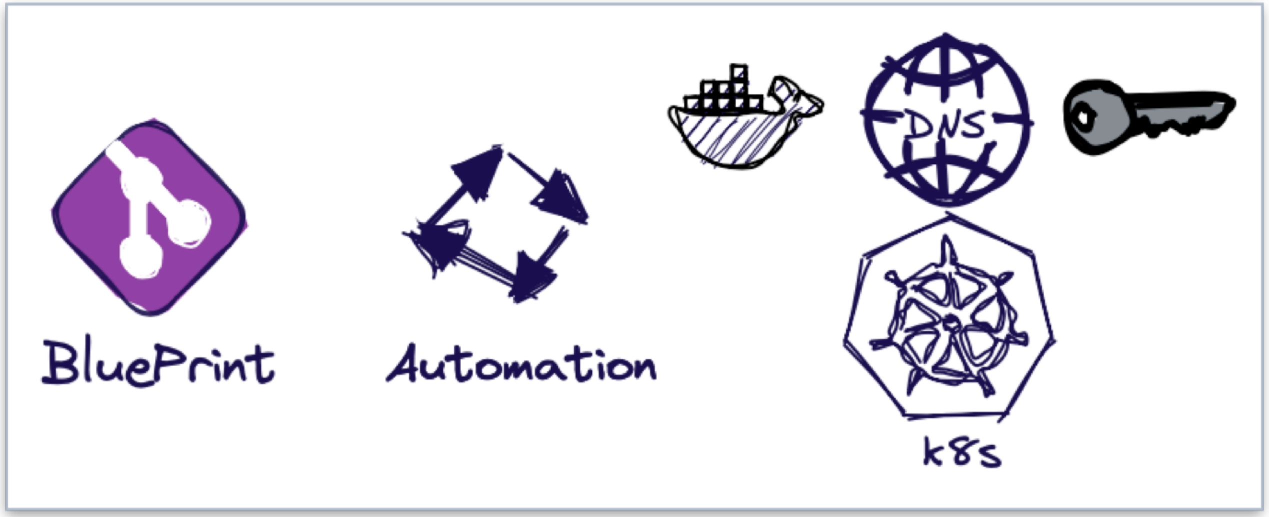
# Goal - converged cloud infrastructure



## We look for answers to:

What technologies? How to connect them? How to make it secure?
How to share data/users? How to include LMS? How to federate everything? etc.

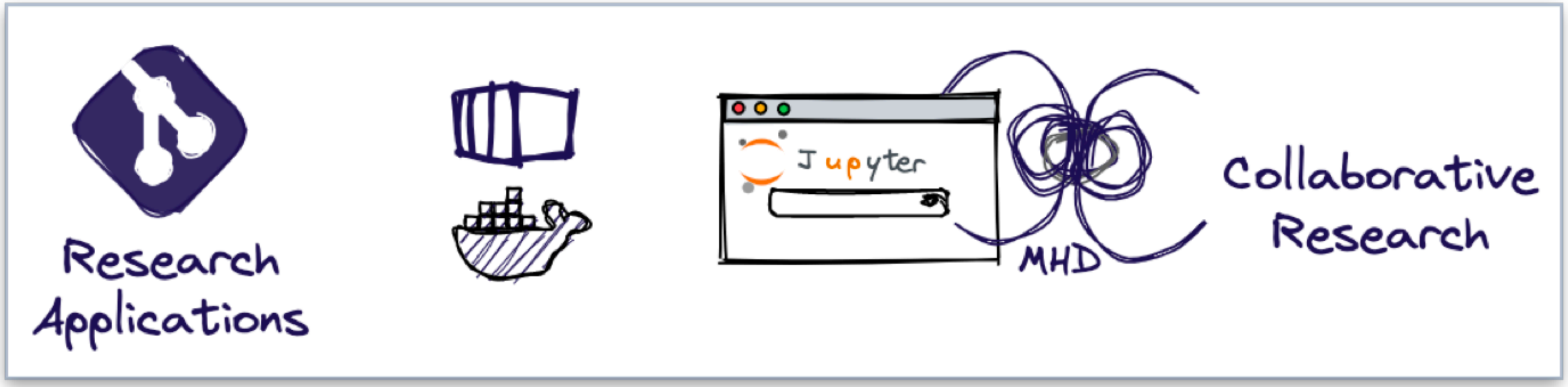
# Goal - secure collaboration



## How can I share my program, data, algorithm?

Enable an interactive version of a research paper including data and algorithms, you could also allow people to run. Make sure your results survive.

# Goal - OpenScience & FAIR

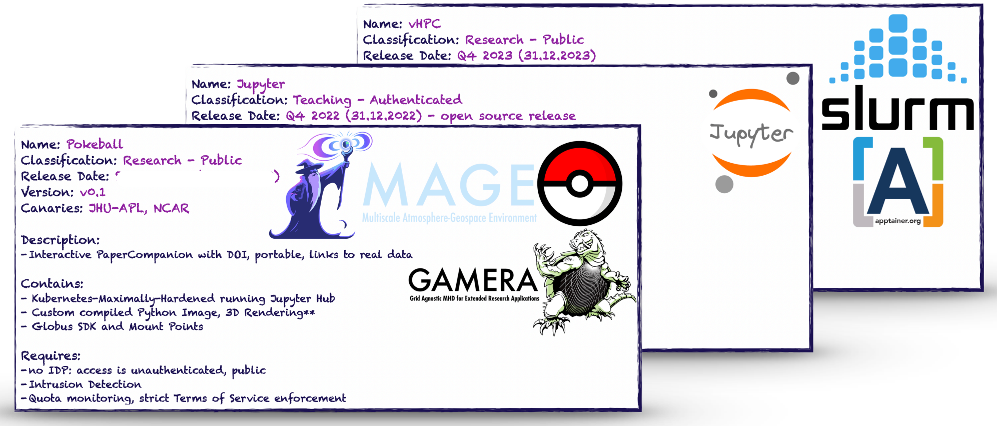


## Enable cooperative development in research

develop a common code base in your research community and gain visibility,
benefit from a reproducible environment and interoperability

# Interactive High performance compute???

HPC code and results can be hard to "publish" and "teach"



Name: vHPC
Classification: Research - Public
Release Date: Q4 2023 (31.12.2023)

Name: Jupyter
Classification: Teaching - Authenticated
Release Date: Q4 2022 (31.12.2022) - open source release

Name: Pokeball
Classification: Research - Public
Release Date:
Version: v0.1
Canaries: JHU-APL, NCAR

Description:
-Interactive PaperCompanion with DOI, portable, links to real data

Contains:
- Kubernetes-Maximally-Hardened running Jupyter Hub
- Custom compiled Python Image, 3D Rendering**
- Globus SDK and Mount Points

Requires:
-no IDP: access is unauthenticated, public
-Intrusion Detection
-Quota monitoring, strict Terms of Service enforcement

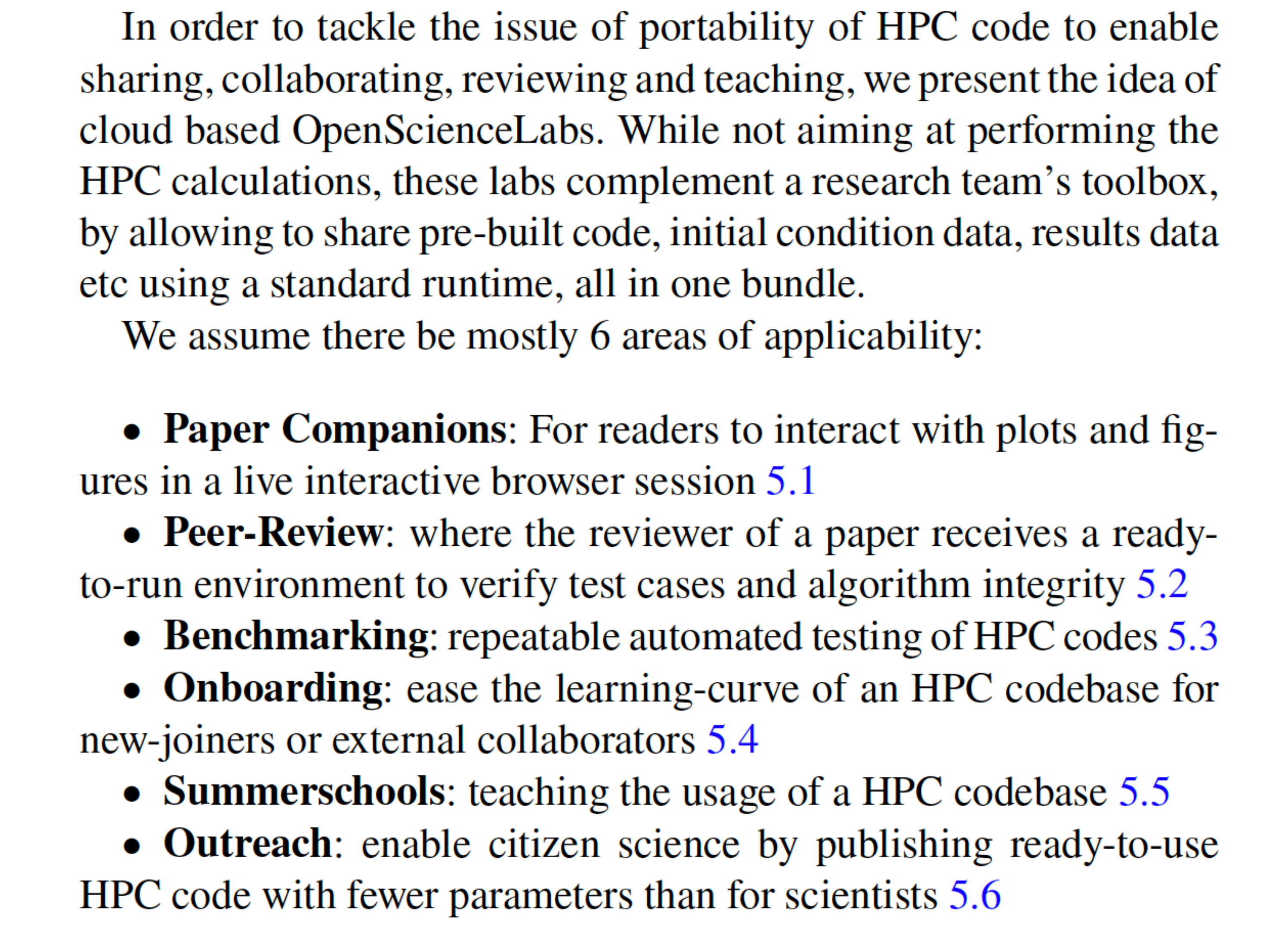
slurm
apptainer.org
Jupyter
MAGE
Multiscale Atmosphere-Geospace Environment
GAMERA
Grid Agnostic MHD for Extended Research Applications

How to be FAIR

- Highly optimized -> Hard to share binaries -> Reproduce
- Specific dependencies -> hard to compile/build
- Often decades of niche knowledge inside -> hard to understand
- Maintainers not incentivized for FAIR and OpenSource

Who cares?

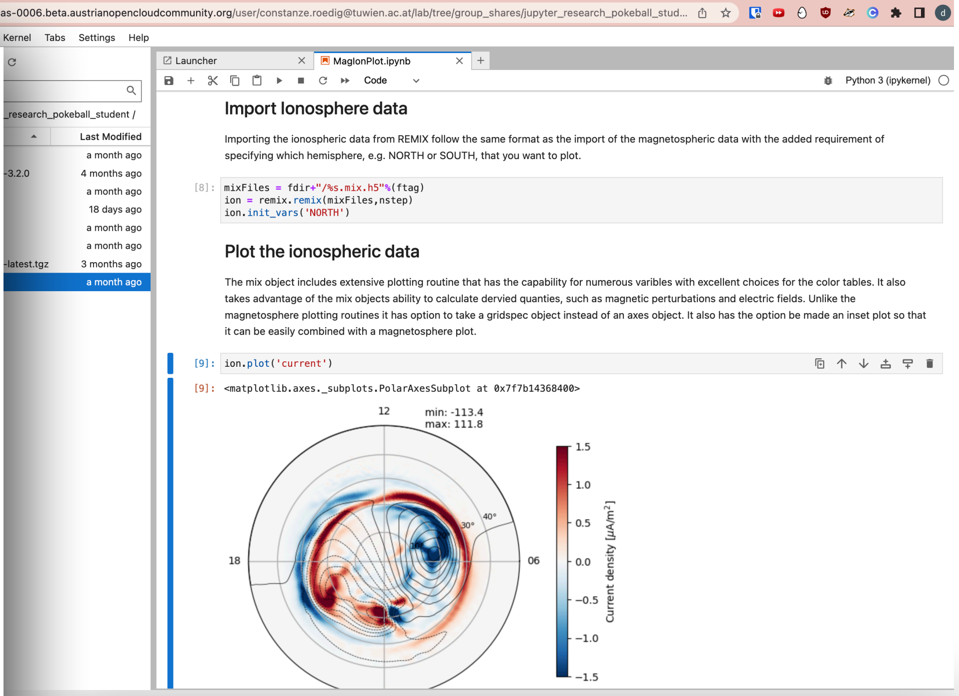
# 1) OpenScienceLabs for HPC



In order to tackle the issue of portability of HPC code to enable sharing, collaborating, reviewing and teaching, we present the idea of cloud based OpenScienceLabs. While not aiming at performing the HPC calculations, these labs complement a research team's toolbox, by allowing to share pre-built code, initial condition data, results data etc using a standard runtime, all in one bundle.

We assume there be mostly 6 areas of applicability:

- **Paper Companions**: For readers to interact with plots and figures in a live interactive browser session 5.1
- **Peer-Review**: where the reviewer of a paper receives a ready-to-run environment to verify test cases and algorithm integrity 5.2
- **Benchmarking**: repeatable automated testing of HPC codes 5.3
- **Onboarding**: ease the learning-curve of an HPC codebase for new-joiners or external collaborators 5.4
- **Summerschools**: teaching the usage of a HPC codebase 5.5
- **Outreach**: enable citizen science by publishing ready-to-use HPC code with fewer parameters than for scientists 5.6



### Import Ionosphere data

Importing the ionospheric data from REMIX follow the same format as the import of the magnetospheric data with the added requirement of specifying which hemisphere, e.g. NORTH or SOUTH, that you want to plot.

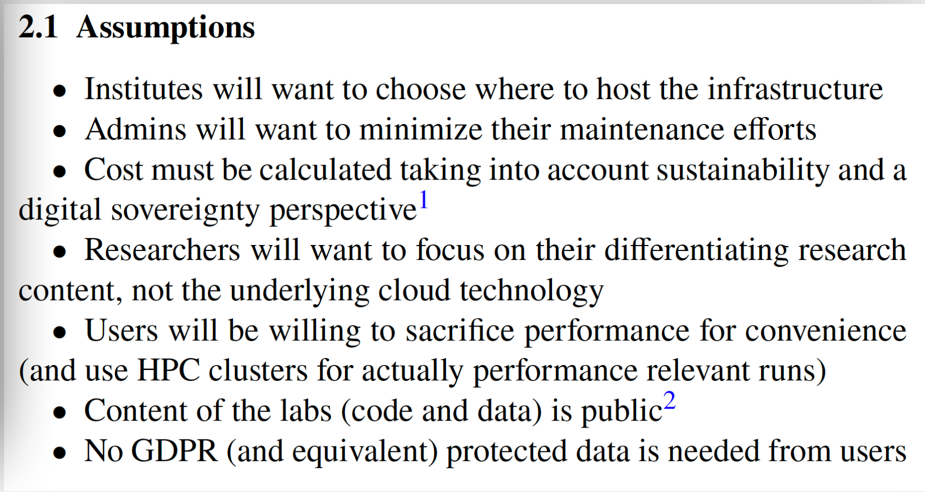
```
[8]: mixFiles = fdir+"/%s.mix.h5"%(ftag)
     ion = remix.remix(mixFiles,nstep)
     ion.init_vars('NORTH')
```

### Plot the ionospheric data

The mix object includes extensive plotting routine that has the capability for numerous varibles with excellent choices for the color tables. It also takes advantage of the mix objects ability to calculate dervied quanties, such as magnetic perturbations and electric fields. Unlike the magnetosphere plotting routines it has option to take a gridspec object instead of an axes object. It also has the option be made an inset plot so that it can be easily combined with a magnetosphere plot.

```
[9]: ion.plot('current')
[9]: <matplotlib.axes._subplots.PolarAxesSubplot at 0x7f7b14368400>
```

12
min: -113.4
max: 111.8
18
06
10°
20°
30°
40°
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Current density [μA/m²]



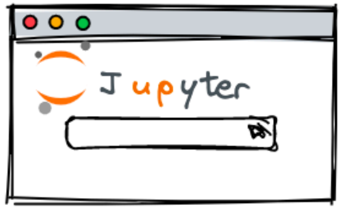
## 2.1 Assumptions

- Institutes will want to choose where to host the infrastructure
- Admins will want to minimize their maintenance efforts
- Cost must be calculated taking into account sustainability and a digital sovereignty perspective[1]
- Researchers will want to focus on their differentiating research content, not the underlying cloud technology
- Users will be willing to sacrifice performance for convenience (and use HPC clusters for actually performance relevant runs)
- Content of the labs (code and data) is public[2]
- No GDPR (and equivalent) protected data is needed from users

# Interlude:
# what is Jupyter?

Jupyter => open standards & web services
for interactive computing across all programming languages

JupyterLab => web-based interactive development environment
for notebooks, code, data science, scientific computing,
computational journalism, and machine learning

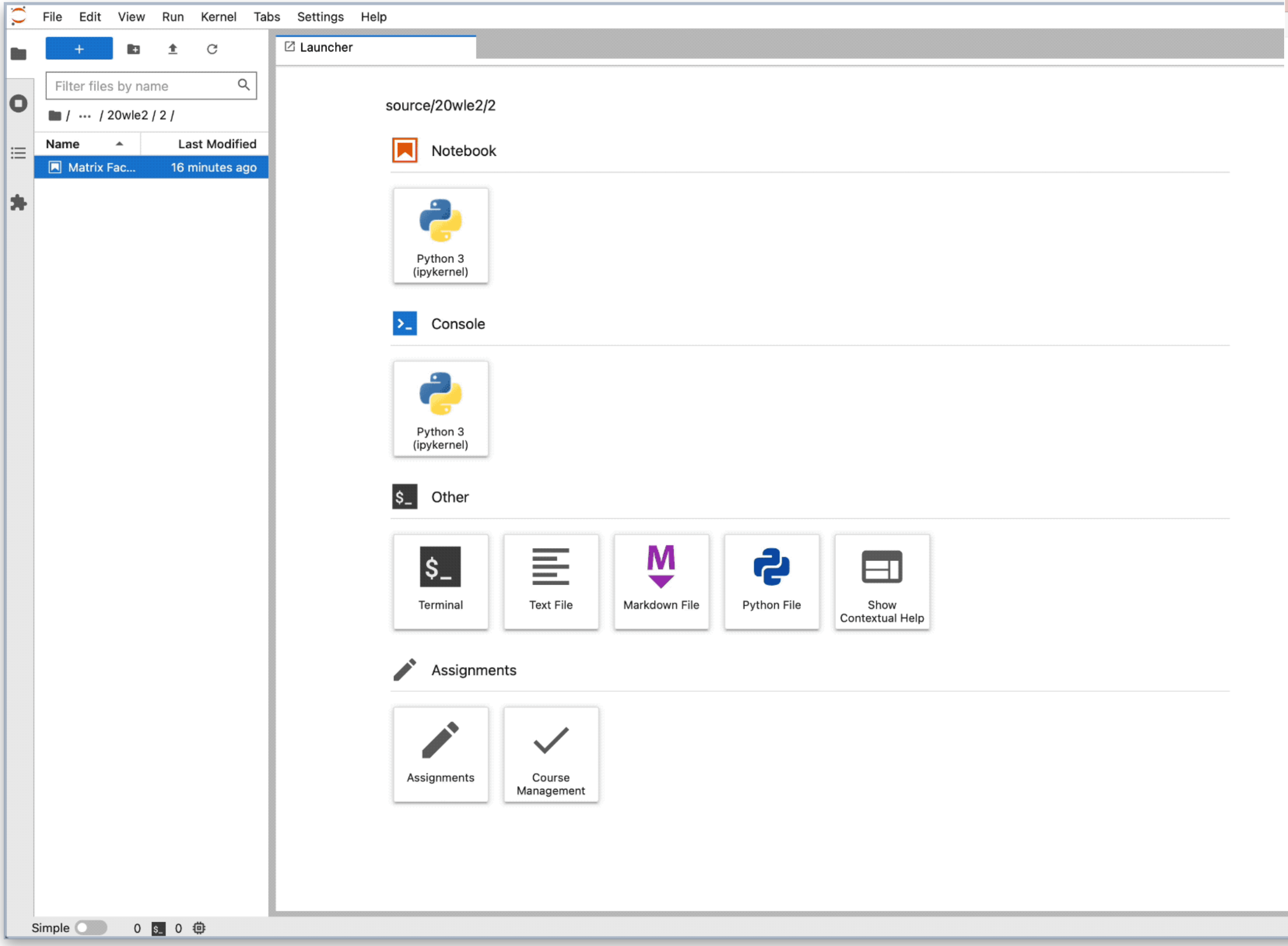


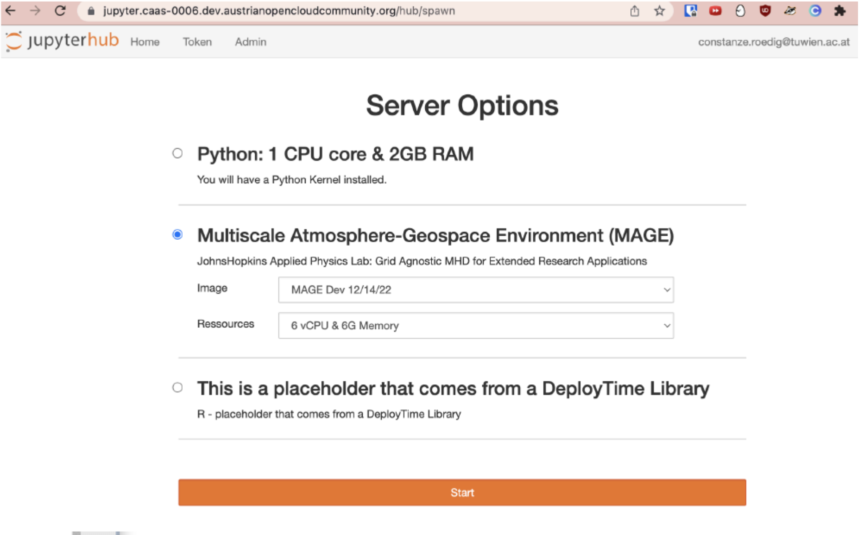
JupyterHub => multi-user version
for companies, classrooms and research labs

# For whom is Jupyter?

- classrooms, research labs and companies doing:
  - data science, scientific computing,
  - computational journalism, machine learning, etc.
- EXECUTE, VISUALISE & SHARE code and data in a web-based environment

# What is Jupyter?




jupyter.caas-0006.dev.austrianopencloudcommunity.org/hub/spawn
jupyterhub
Home
Token
Admin
constanze.roedig@tuwien.ac.at

## Server Options

○ Python: 1 CPU core & 2GB RAM

You will have a Python Kernel installed.

● Multiscale Atmosphere-Geospace Environment (MAGE)

JohnsHopkins Applied Physics Lab: Grid Agnostic MHD for Extended Research Applications

Image: MAGE Dev 12/14/22

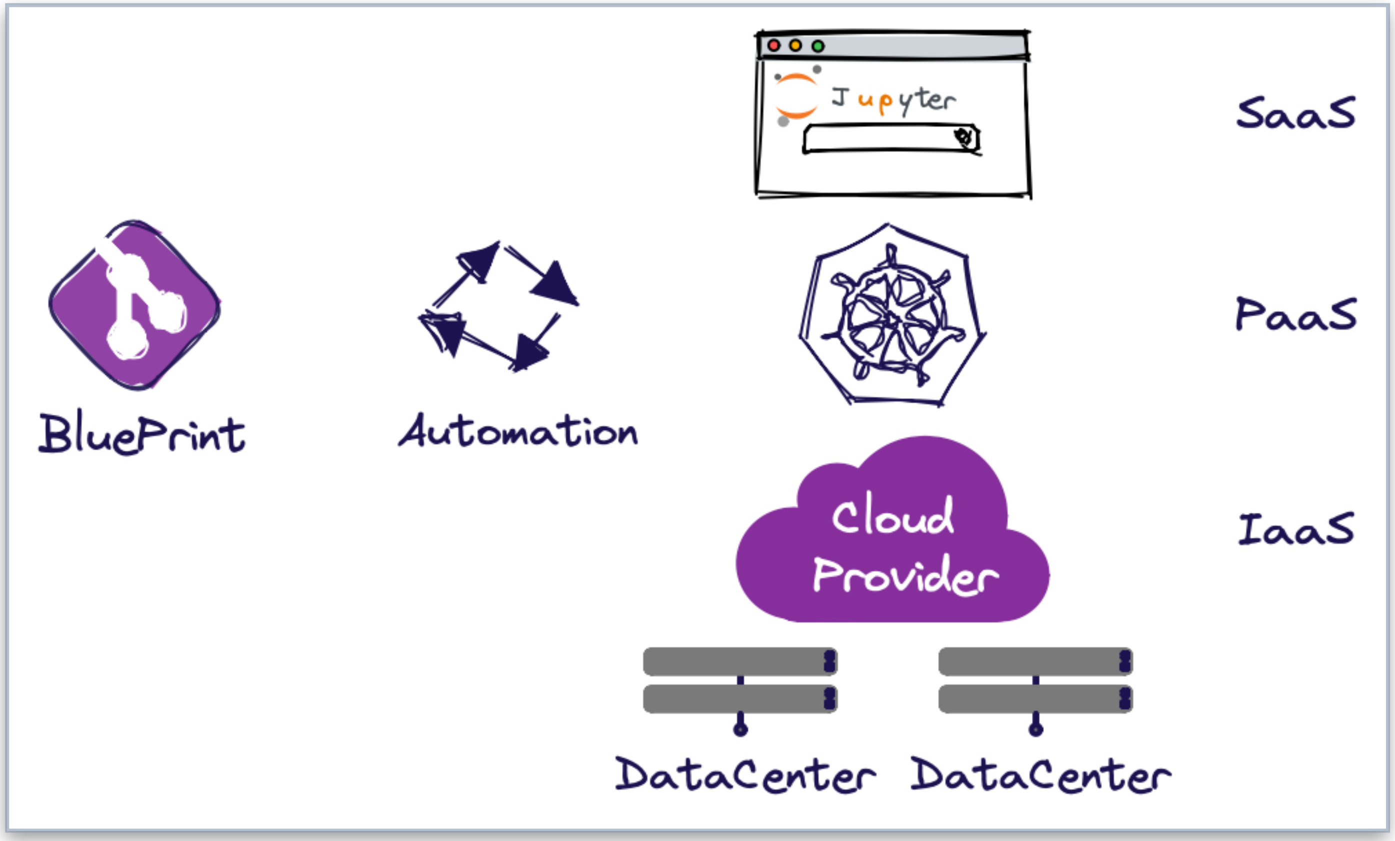
Ressources: 6 vCPU & 6G Memory

○ This is a placeholder that comes from a DeployTime Library

R - placeholder that comes from a DeployTime Library

Start

- Browser based
- Multi-user
- Publish Application (image)
- Run on CPU/GPU/ MEM and DISK depending on quota



# What do we mean by Open Science Labs?

Definition: A distributed virtual runtime, where Infrastructure, Platforms and Applications can be elastically consumed 'as a service', while the customer provides the content.

Key paradigm features: automation, standard APIs, self-service, shared ownership, community driven

# OpenScienceLab: what is it?

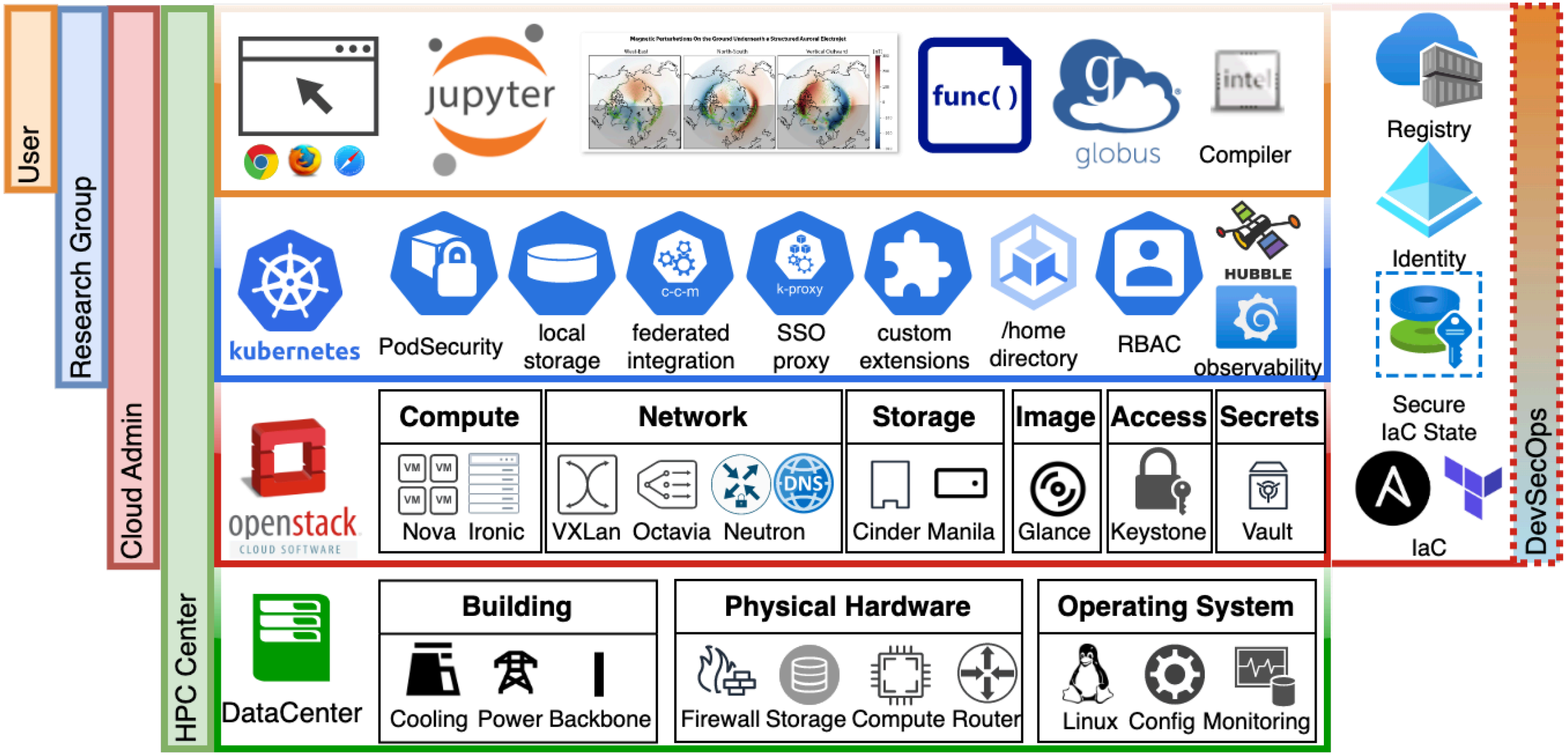


**Figure 1.** Layers and Components to run an OpenScienceLab on HPC center premises. Users access their containerized HPC code via the web-browser wrapped into a Jupyter-Extension and have access to visualization, data (via e.g. Globus), their compilers and their own pre-built code (orange). Each Research Group self-manages their tenant (blue), which contains an opinionated kubernetes thus abstracting low-level IT components from the users. Each tenant is ephemeral and of dynamic size. The cloud service provider (CSP) (red) is assumed to be Openstack for on-premises and provides unified access to virtualized infrastructure. These three layers are managed, maintained and modified using DevSecOps practices and tooling (some of which use public cloud commodity services to easily achieve multi-region redundancy). In the case of on-premises, the (green) layer depicts the usual data center, where the (red) cloud would form a sub-set of. The green layer can be exchanged for a public cloud provider or a commercial hosting provider, if needed.

# Profile: OpenScienceLab for HPC

Name: Pokeball ^1
Classification: Research - Public
Release Date: 2023
Version: v0.1
Canaries: JHU-APL, NCAR

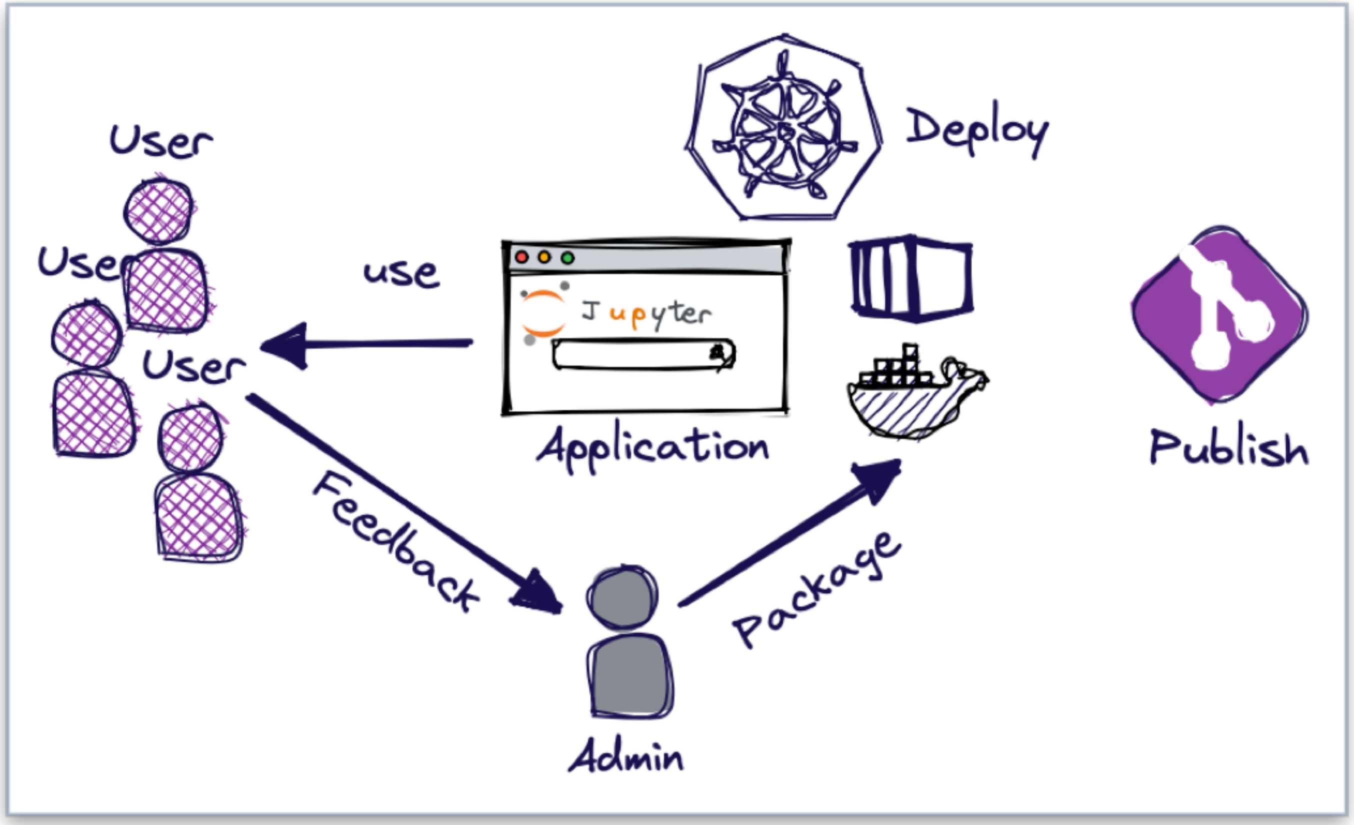
Description:
- Interactive PaperCompanion with DOI, portable, links to real data

Contains:
- Kubernetes-Maximally-Hardened running Jupyter Hub
- Custom compiled Python Image, 3D Rendering**
- Globus SDK and Mount Points

Requires:
- Access is public
- Intrusion Detection**
- Quota monitoring, strict Terms of Service enforcement

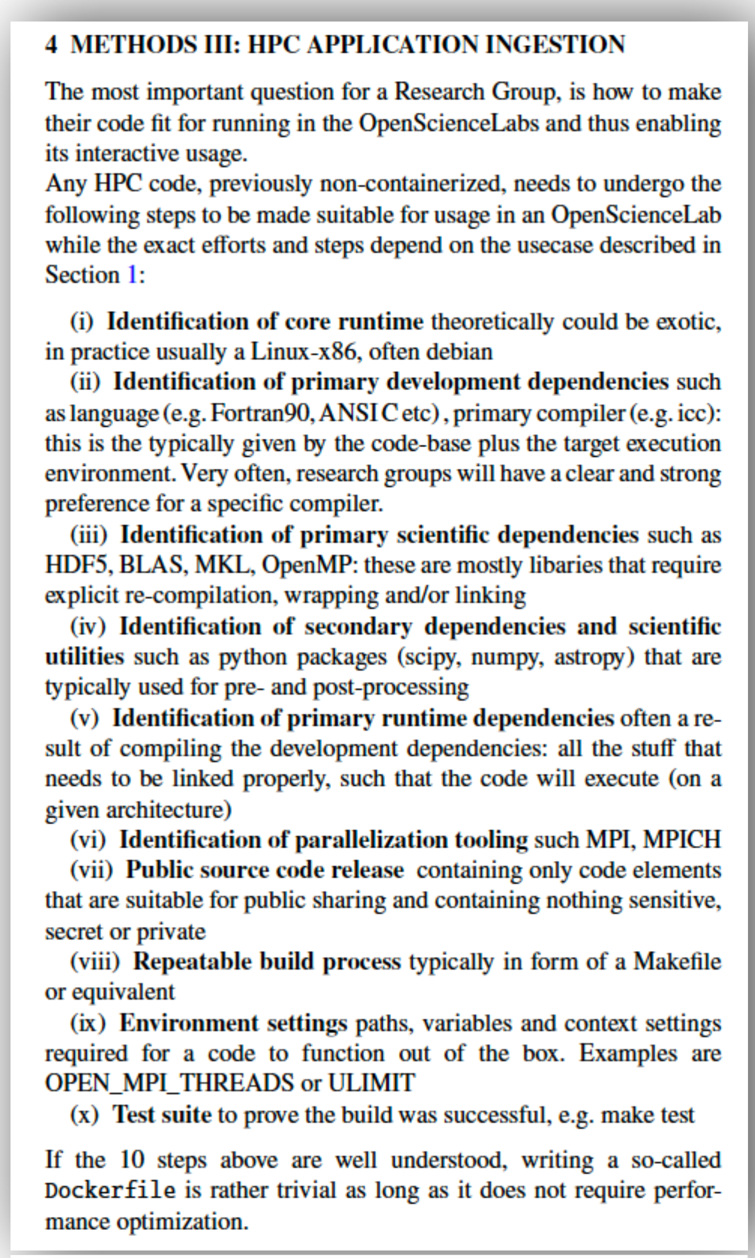
**planned for beta
^1 Nintendo: please dont sue us

# OpenScience- User Perspective



## How to collaborate on an HPC code?

1. Containerize HPC code properly
2. Create the lab cluster & test it
3. Publish as required and add interactive content



### 4 METHODS III: HPC APPLICATION INGESTION

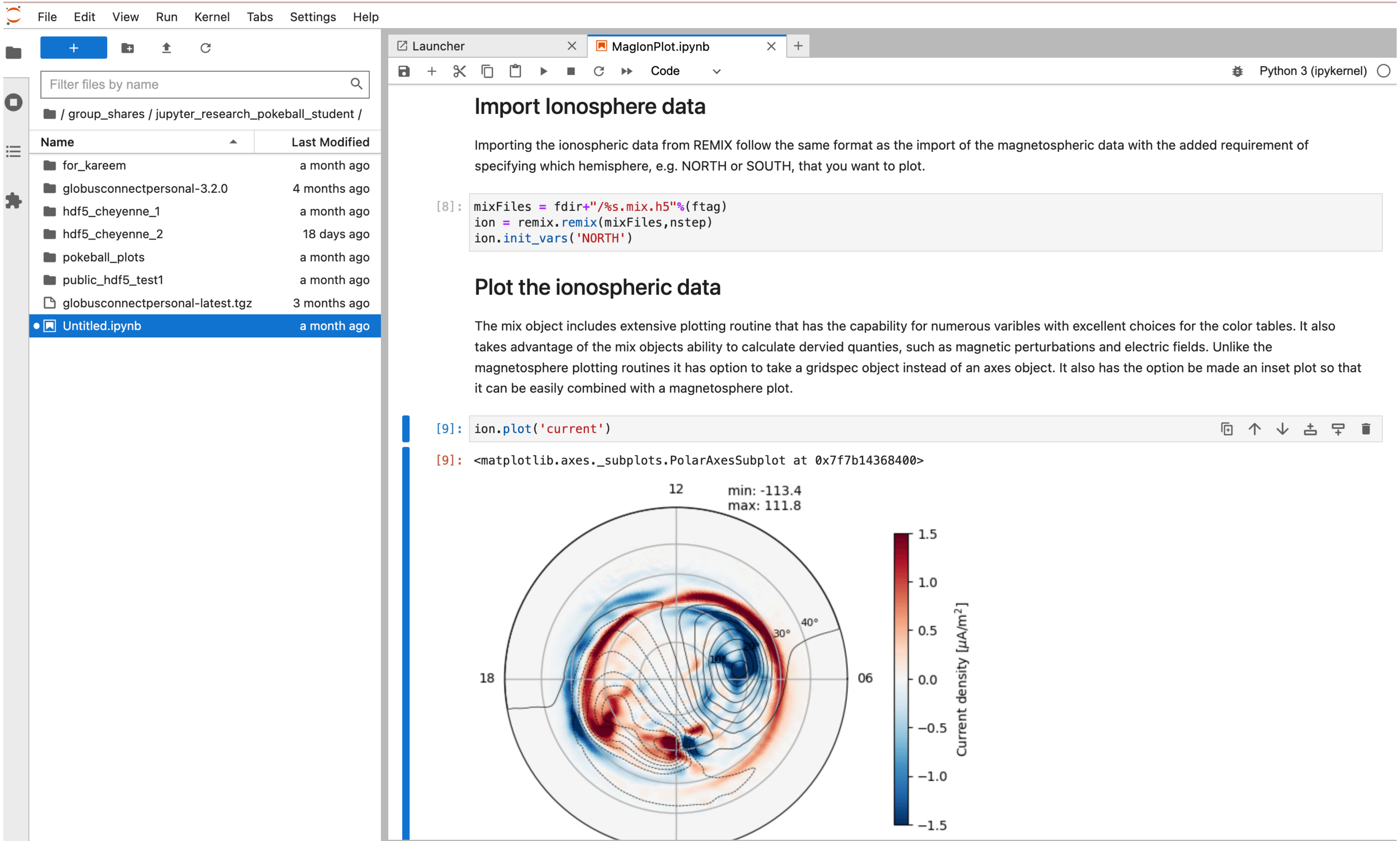
The most important question for a Research Group, is how to make their code fit for running in the OpenScienceLabs and thus enabling its interactive usage.
Any HPC code, previously non-containerized, needs to undergo the following steps to be made suitable for usage in an OpenScienceLab while the exact efforts and steps depend on the usecase described in Section 1:

(i) **Identification of core runtime** theoretically could be exotic, in practice usually a Linux-x86, often debian
(ii) **Identification of primary development dependencies** such as language (e.g. Fortran90, ANSI C etc), primary compiler (e.g. icc): this is the typically given by the code-base plus the target execution environment. Very often, research groups will have a clear and strong preference for a specific compiler.
(iii) **Identification of primary scientific dependencies** such as HDF5, BLAS, MKL, OpenMP: these are mostly libaries that require explicit re-compilation, wrapping and/or linking
(iv) **Identification of secondary dependencies and scientific utilities** such as python packages (scipy, numpy, astropy) that are typically used for pre- and post-processing
(v) **Identification of primary runtime dependencies** often a result of compiling the development dependencies: all the stuff that needs to be linked properly, such that the code will execute (on a given architecture)
(vi) **Identification of parallelization tooling** such MPI, MPICH
(vii) **Public source code release** containing only code elements that are suitable for public sharing and containing nothing sensitive, secret or private
(viii) **Repeatable build process** typically in form of a Makefile or equivalent
(ix) **Environment settings** paths, variables and context settings required for a code to function out of the box. Examples are OPEN_MPI_THREADS or ULIMIT
(x) **Test suite** to prove the build was successful, e.g. make test

If the 10 steps above are well understood, writing a so-called `Dockerfile` is rather trivial as long as it does not require performance optimization.

AUSTRIAN DATALAB
AND SERVICES

# Demo OpenScience



| Name | Last Modified |
|---|---|
| for_kareem | a month ago |
| globusconnectpersonal-3.2.0 | 4 months ago |
| hdf5_cheyenne_1 | a month ago |
| hdf5_cheyenne_2 | 18 days ago |
| pokeball_plots | a month ago |
| public_hdf5_test1 | a month ago |
| globusconnectpersonal-latest.tgz | 3 months ago |
| Untitled.ipynb | a month ago |

## Import Ionosphere data

Importing the ionospheric data from REMIX follow the same format as the import of the magnetospheric data with the added requirement of specifying which hemisphere, e.g. NORTH or SOUTH, that you want to plot.

```
[8]: mixFiles = fdir+"/%s.mix.h5"%(ftag)
     ion = remix.remix(mixFiles,nstep)
     ion.init_vars('NORTH')
```

## Plot the ionospheric data

The mix object includes extensive plotting routine that has the capability for numerous varibles with excellent choices for the color tables. It also takes advantage of the mix objects ability to calculate dervied quantities, such as magnetic perturbations and electric fields. Unlike the magnetosphere plotting routines it has option to take a gridspec object instead of an axes object. It also has the option be made an inset plot so that it can be easily combined with a magnetosphere plot.

```
[9]: ion.plot('current')
```

```
[9]: <matplotlib.axes._subplots.PolarAxesSubplot at 0x7f7b14368400>
```

# Demo OpenScience

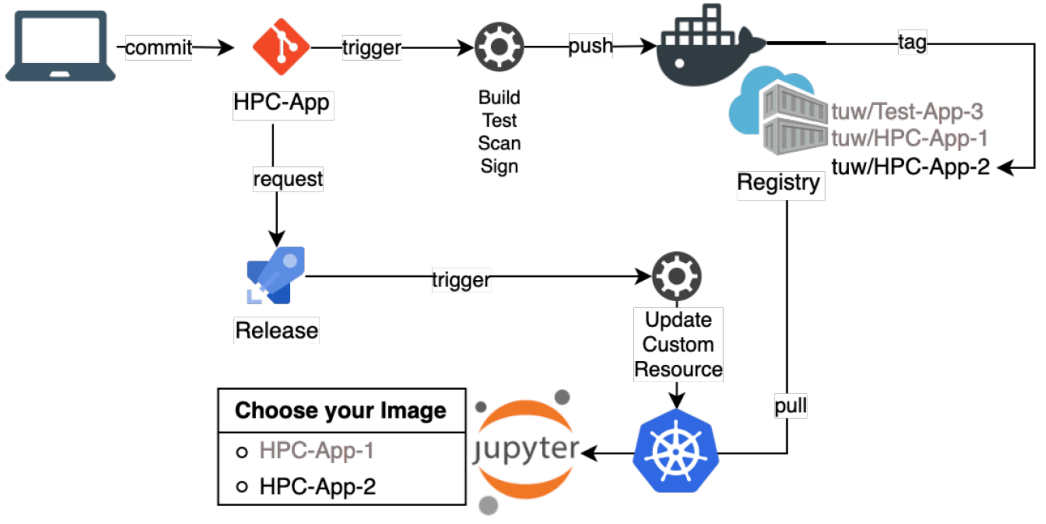
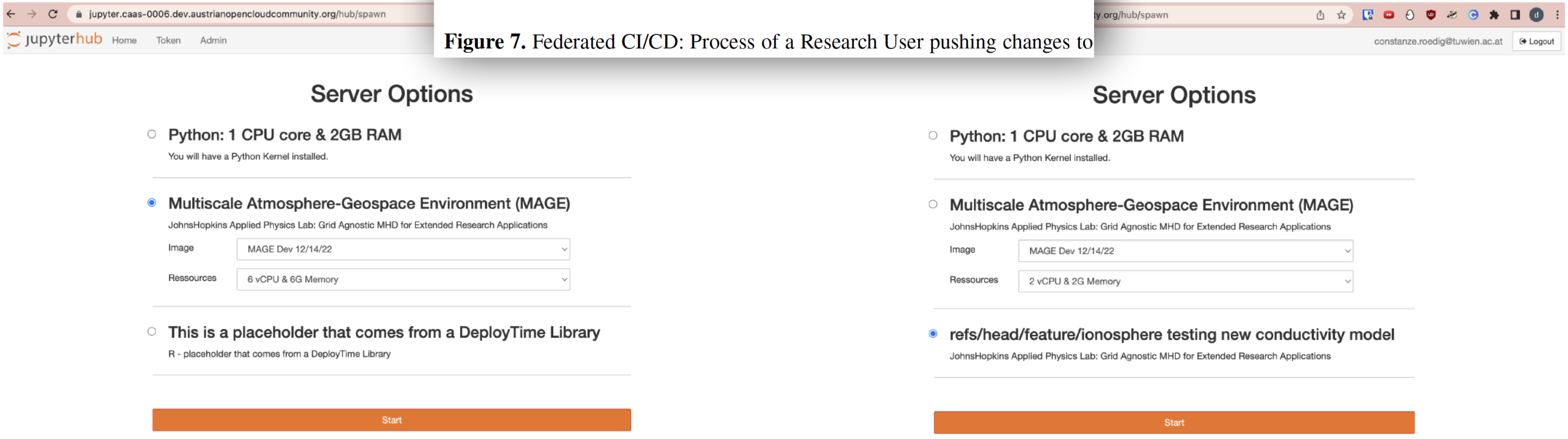


**Figure 7.** Federated CI/CD: Process of a Research User pushing changes to


jupyter.caas-0006.dev.austrianopencloudcommunity.org/hub/spawn
jupyterhub
Home
Token
Admin
ty.org/hub/spawn
constanze.roedig@tuwien.ac.at
Logout

## Server Options

○ Python: 1 CPU core & 2GB RAM

You will have a Python Kernel installed.

● Multiscale Atmosphere-Geospace Environment (MAGE)

JohnsHopkins Applied Physics Lab: Grid Agnostic MHD for Extended Research Applications

Image: MAGE Dev 12/14/22

Ressources: 6 vCPU & 6G Memory

○ This is a placeholder that comes from a DeployTime Library

R - placeholder that comes from a DeployTime Library

Start

## Server Options

○ Python: 1 CPU core & 2GB RAM

You will have a Python Kernel installed.

○ Multiscale Atmosphere-Geospace Environment (MAGE)

JohnsHopkins Applied Physics Lab: Grid Agnostic MHD for Extended Research Applications

Image: MAGE Dev 12/14/22

Ressources: 2 vCPU & 2G Memory

● refs/head/feature/ionosphere testing new conductivity model

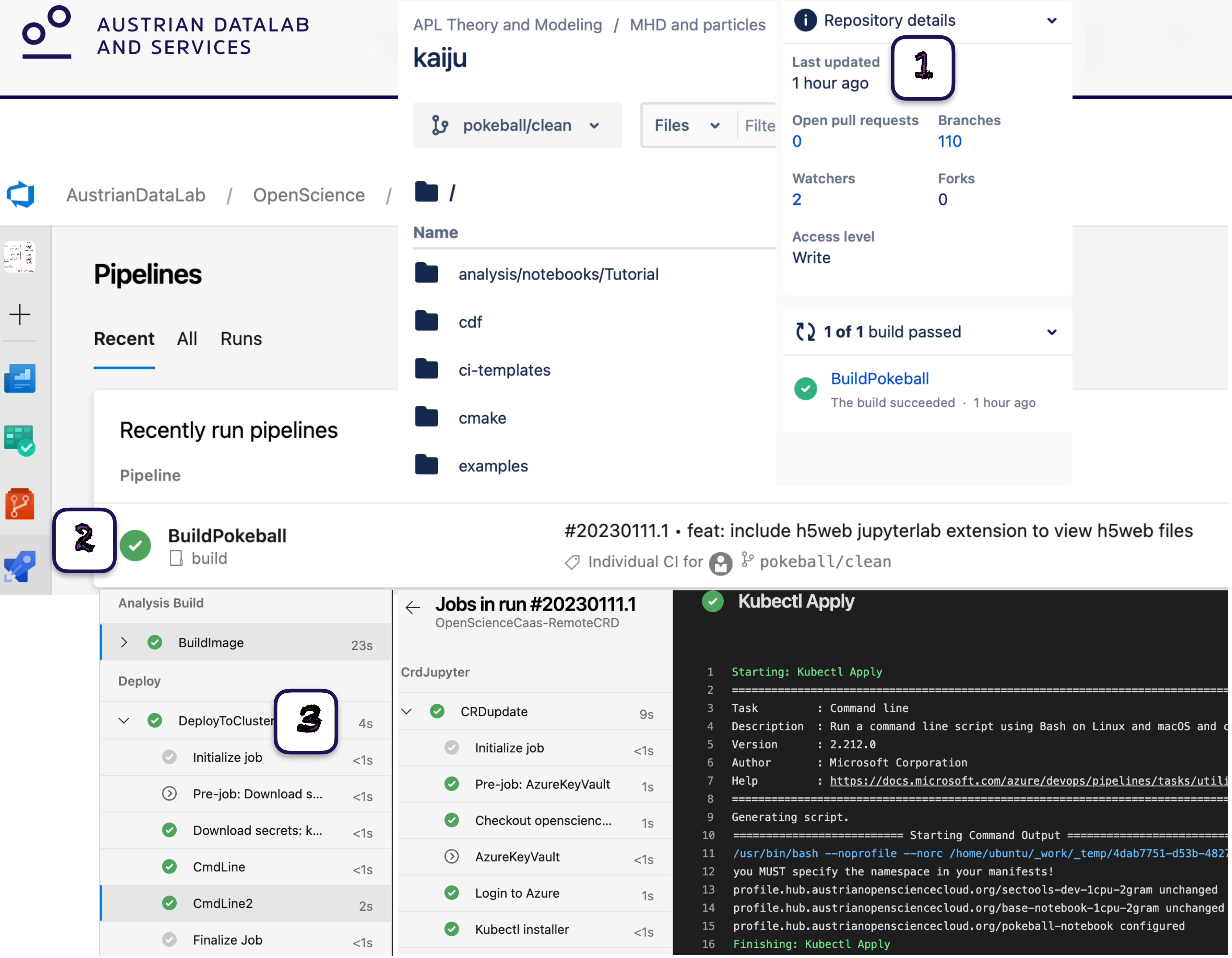
JohnsHopkins Applied Physics Lab: Grid Agnostic MHD for Extended Research Applications

Start

**Figure 8.** Updating Custom Resource Definitions to share a pre-built model interactively with other researchers

AUSTRIAN DATALAB AND SERVICES

# Demo OpenScience


APL Theory and Modeling / MHD and particles

kaiju

pokeball/clean

Files

Filte

/

Name

- analysis/notebooks/Tutorial
- cdf
- ci-templates
- cmake
- examples

Repository details

| | |
|---|---|
| Last updated<br>1 hour ago | |
| Open pull requests<br>0 | Branches<br>110 |
| Watchers<br>2 | Forks<br>0 |
| Access level<br>Write | |

1 of 1 build passed

BuildPokeball
The build succeeded · 1 hour ago

AustrianDataLab / OpenScience /

Pipelines

Recent All Runs

Recently run pipelines

Pipeline

BuildPokeball
build

#20230111.1 • feat: include h5web jupyterlab extension to view h5web files
Individual CI for pokeball/clean

Analysis Build

| | |
|---|---|
| BuildImage | 23s |

Deploy

| | |
|---|---|
| DeployToCluster | 4s |
| Initialize job | <1s |
| Pre-job: Download s... | <1s |
| Download secrets: k... | <1s |
| CmdLine | <1s |
| CmdLine2 | 2s |
| Finalize Job | <1s |

Jobs in run #20230111.1
OpenScienceCaas-RemoteCRD

CrdJupyter

| | |
|---|---|
| CRDupdate | 9s |
| Initialize job | <1s |
| Pre-job: AzureKeyVault | 1s |
| Checkout opensciencc... | 1s |
| AzureKeyVault | <1s |
| Login to Azure | 1s |
| Kubectl installer | <1s |

Kubectl Apply

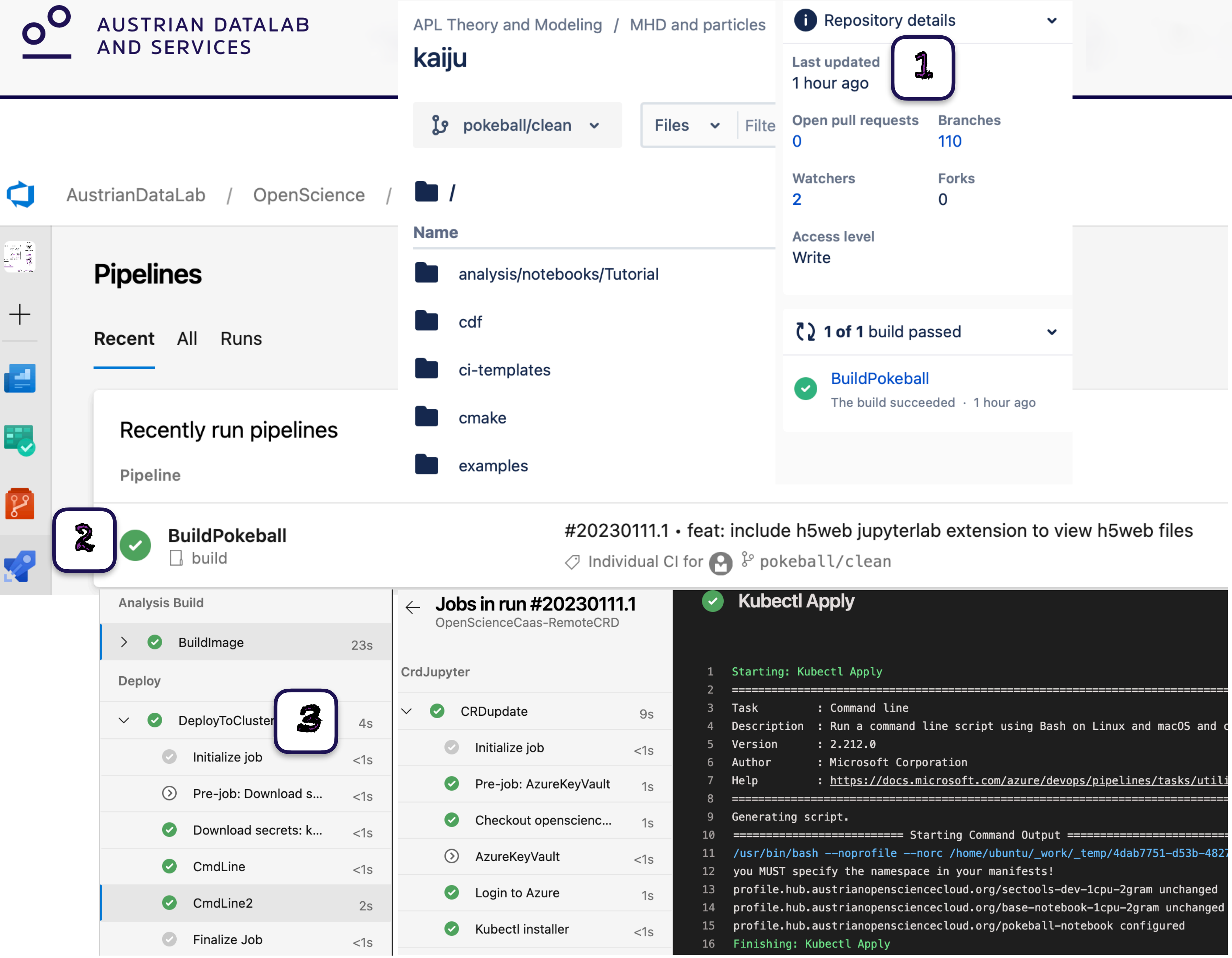
```
Starting: Kubectl Apply
==============================================================================
Task         : Command line
Description  : Run a command line script using Bash on Linux and macOS and c
Version      : 2.212.0
Author       : Microsoft Corporation
Help         : https://docs.microsoft.com/azure/devops/pipelines/tasks/utili
==============================================================================
Generating script.
========================== Starting Command Output ===========================
/usr/bin/bash --noprofile --norc /home/ubuntu/_work/_temp/4dab7751-d53b-4827
you MUST specify the namespace in your manifests!
profile.hub.austrianopensciencecloud.org/sectools-dev-1cpu-2gram unchanged
profile.hub.austrianopensciencecloud.org/base-notebook-1cpu-2gram unchanged
profile.hub.austrianopensciencecloud.org/pokeball-notebook configured
Finishing: Kubectl Apply
```

Steps:

1. Research User commits code
2. Build Pipeline triggers a publication of the new image
3. Deploy Pipeline triggers and makes the new image available in a cluster

AUSTRIAN DATALAB AND SERVICES

# Demo OpenScience



APL Theory and Modeling / MHD and particles

kaiju

pokeball/clean | Files | Filte

/

| Name |
|---|
| analysis/notebooks/Tutorial |
| cdf |
| ci-templates |
| cmake |
| examples |

Repository details

1

Last updated: 1 hour ago

Open pull requests: 0

Branches: 110

Watchers: 2

Forks: 0

Access level: Write

1 of 1 build passed

BuildPokeball
The build succeeded · 1 hour ago

AustrianDataLab / OpenScience /

Pipelines

Recent | All | Runs

Recently run pipelines

Pipeline

2

BuildPokeball
build

#20230111.1 • feat: include h5web jupyterlab extension to view h5web files

Individual CI for pokeball/clean

| Analysis Build | |
|---|---|
| BuildImage | 23s |
| **Deploy** | |
| DeployToCluster | 4s |
| Initialize job | <1s |
| Pre-job: Download s... | <1s |
| Download secrets: k... | <1s |
| CmdLine | <1s |
| CmdLine2 | 2s |
| Finalize Job | <1s |

3

Jobs in run #20230111.1
OpenScienceCaas-RemoteCRD

| CrdJupyter | |
|---|---|
| CRDupdate | 9s |
| Initialize job | <1s |
| Pre-job: AzureKeyVault | 1s |
| Checkout opensciencc... | 1s |
| AzureKeyVault | <1s |
| Login to Azure | 1s |
| Kubectl installer | <1s |

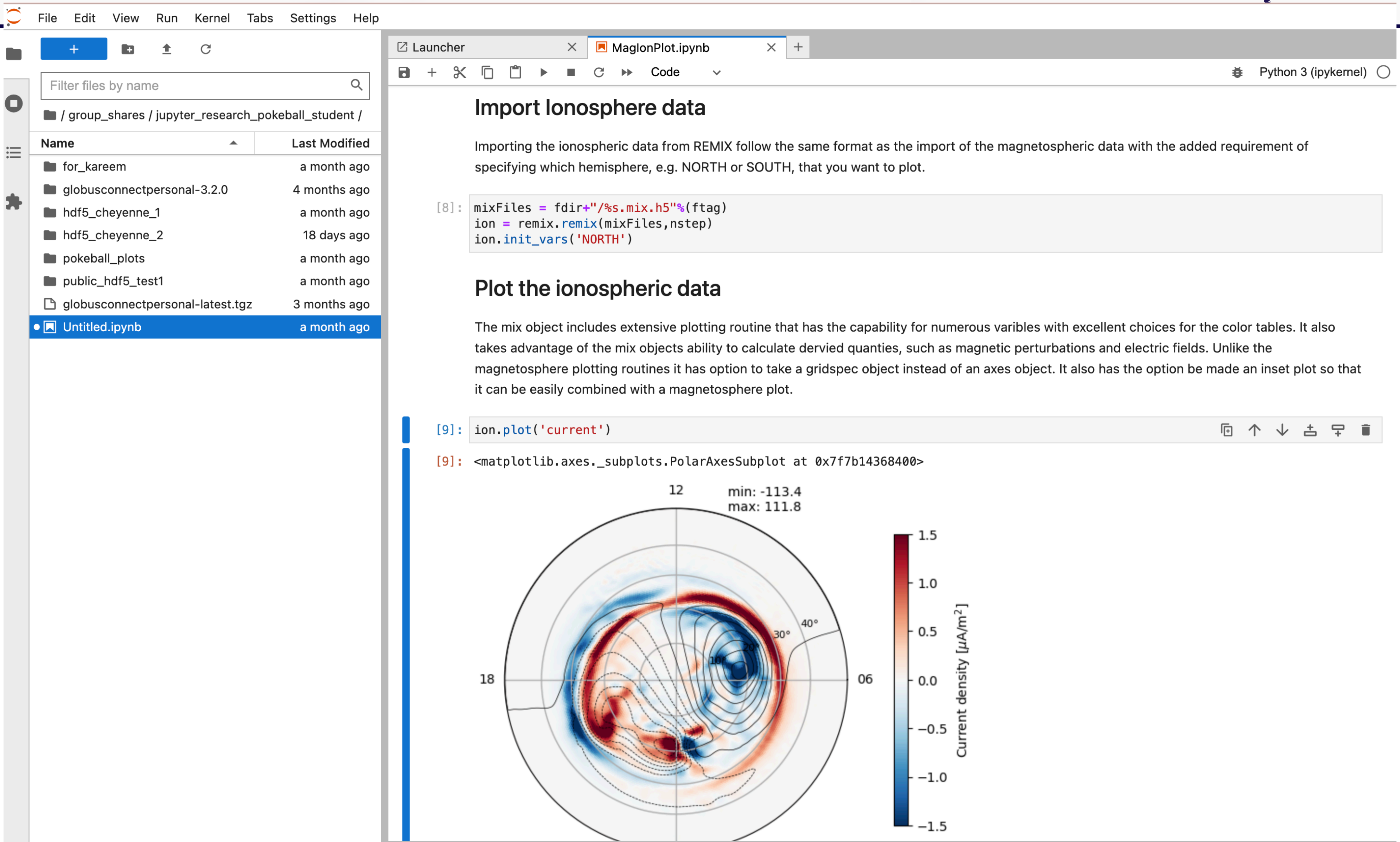
Kubectl Apply

```
Starting: Kubectl Apply
==============================================================================
Task         : Command line
Description  : Run a command line script using Bash on Linux and macOS and c
Version      : 2.212.0
Author       : Microsoft Corporation
Help         : https://docs.microsoft.com/azure/devops/pipelines/tasks/utili
==============================================================================
Generating script.
========================== Starting Command Output ===========================
/usr/bin/bash --noprofile --norc /home/ubuntu/_work/_temp/4dab7751-d53b-4827
you MUST specify the namespace in your manifests!
profile.hub.austrianopensciencecloud.org/sectools-dev-1cpu-2gram unchanged
profile.hub.austrianopensciencecloud.org/base-notebook-1cpu-2gram unchanged
profile.hub.austrianopensciencecloud.org/pokeball-notebook configured
Finishing: Kubectl Apply
```

Steps:

1. Research User commits code
2. Build Pipeline triggers a publication of the new image
3. Deploy Pipeline triggers and makes the new image available in a cluster

# Demo OpenScience



/ group_shares / jupyter_research_pokeball_student /

| Name | Last Modified |
|---|---|
| for_kareem | a month ago |
| globusconnectpersonal-3.2.0 | 4 months ago |
| hdf5_cheyenne_1 | a month ago |
| hdf5_cheyenne_2 | 18 days ago |
| pokeball_plots | a month ago |
| public_hdf5_test1 | a month ago |
| globusconnectpersonal-latest.tgz | 3 months ago |
| Untitled.ipynb | a month ago |

## Import Ionosphere data

Importing the ionospheric data from REMIX follow the same format as the import of the magnetospheric data with the added requirement of specifying which hemisphere, e.g. NORTH or SOUTH, that you want to plot.

```
[8]: mixFiles = fdir+"/%s.mix.h5"%(ftag)
     ion = remix.remix(mixFiles,nstep)
     ion.init_vars('NORTH')
```

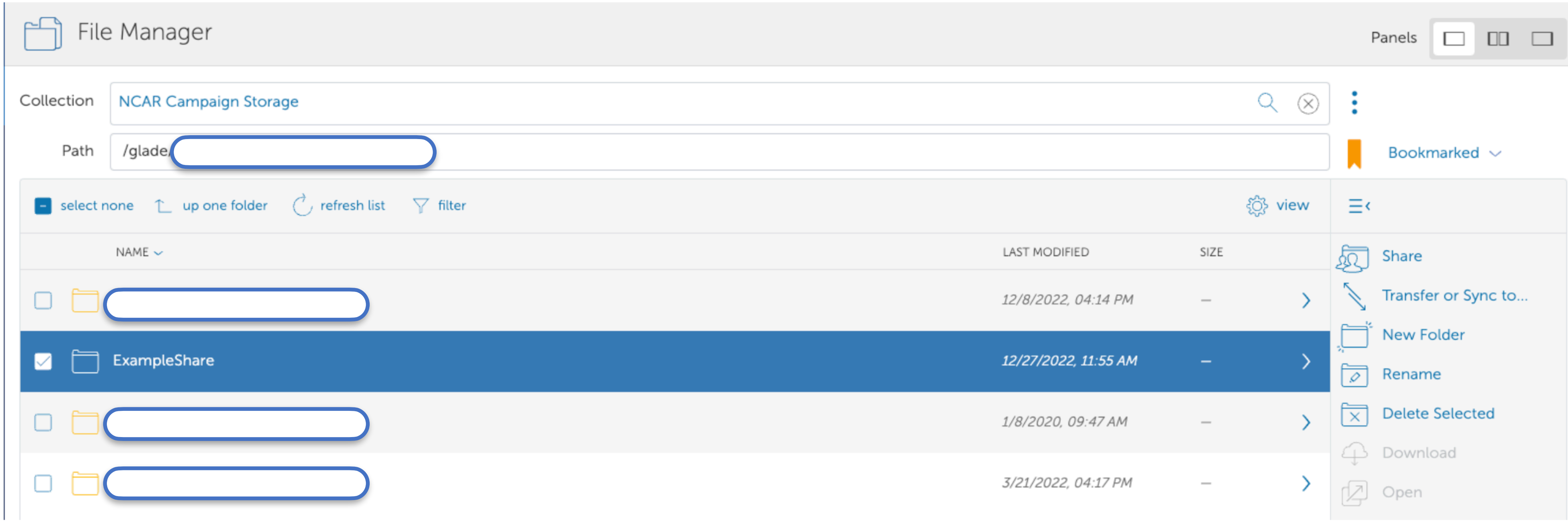
## Plot the ionospheric data

The mix object includes extensive plotting routine that has the capability for numerous varibles with excellent choices for the color tables. It also takes advantage of the mix objects ability to calculate dervied quanties, such as magnetic perturbations and electric fields. Unlike the magnetosphere plotting routines it has option to take a gridspec object instead of an axes object. It also has the option be made an inset plot so that it can be easily combined with a magnetosphere plot.

```
[9]: ion.plot('current')
```

```
[9]: <matplotlib.axes._subplots.PolarAxesSubplot at 0x7f7b14368400>
```

To share a directory with other Globus users first create the directory on the NCAR filesystem you wish to share with other users. Via the Globus website use your bookmarks to navigate to the file system you just created and then select the share option to create a new guest collection. The image below show how to share the directory /glade/[redacted]ExampleShare as a guest collection.



Collection: NCAR Campaign Storage

Path: /glade/[redacted]

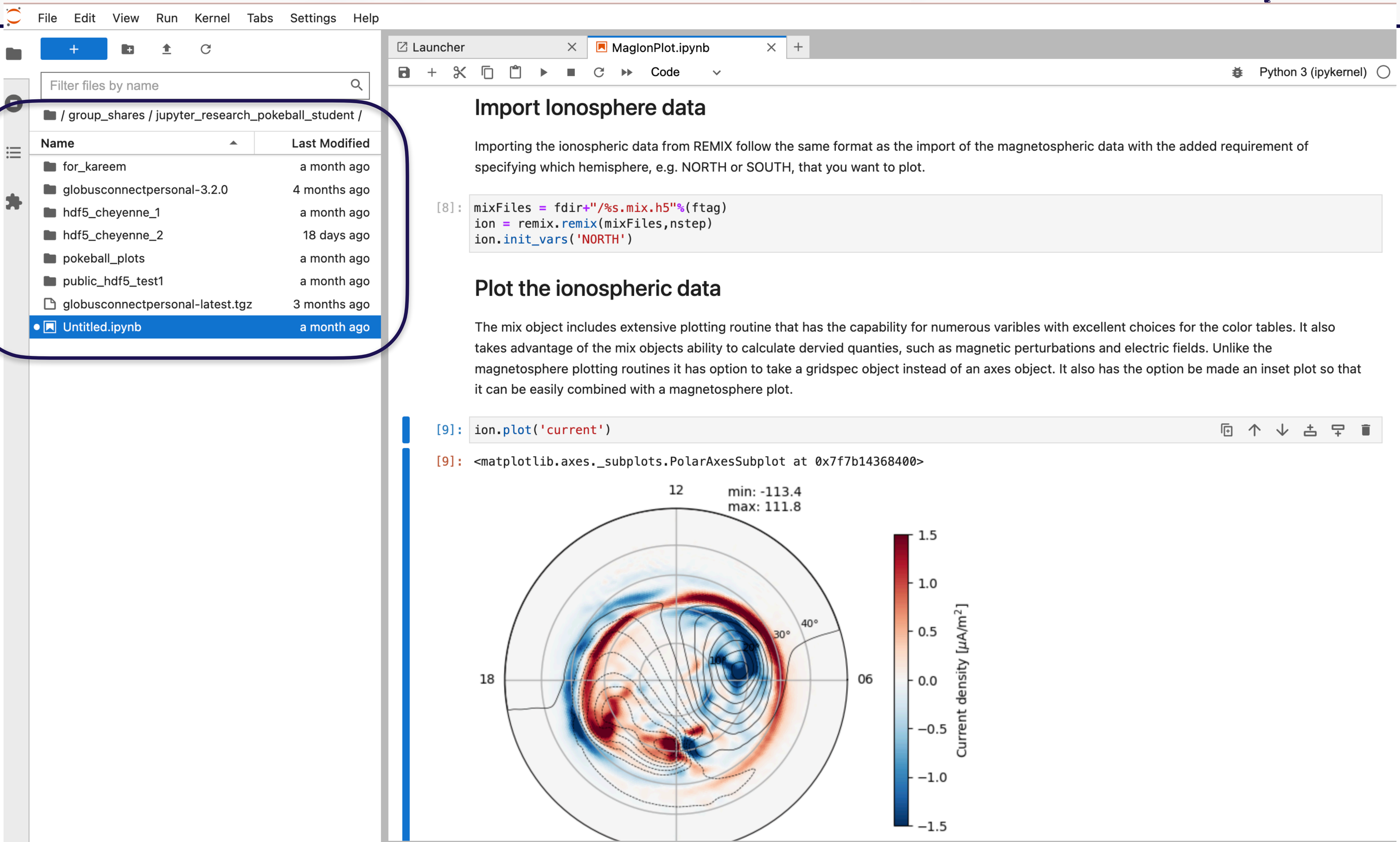
| | NAME | LAST MODIFIED | SIZE |
|---|---|---|---|
| □ | [redacted] | 12/8/2022, 04:14 PM | – |
| ☑ | ExampleShare | 12/27/2022, 11:55 AM | – |
| □ | [redacted] | 1/8/2020, 09:47 AM | – |
| □ | [redacted] | 3/21/2022, 04:17 PM | – |

Once you click share you will need to choose the option for creating a new guest collection and then you'll need to provide the required display name and meta data. If the data is related to a publication you can provide the DOI in the information link available in *view more fields* option on the web page.

Once you create the collection you will have option to add permissions for sharing the data with other users. There are currently for levels of sharing, specific users, groups, all global users, and public. The web site will then provide a link that you can use for sharing with other users.

AUSTRIAN DATALAB AND SERVICES

# Demo OpenScience



/ group_shares / jupyter_research_pokeball_student /

| Name | Last Modified |
|---|---|
| for_kareem | a month ago |
| globusconnectpersonal-3.2.0 | 4 months ago |
| hdf5_cheyenne_1 | a month ago |
| hdf5_cheyenne_2 | 18 days ago |
| pokeball_plots | a month ago |
| public_hdf5_test1 | a month ago |
| globusconnectpersonal-latest.tgz | 3 months ago |
| Untitled.ipynb | a month ago |

## Import Ionosphere data

Importing the ionospheric data from REMIX follow the same format as the import of the magnetospheric data with the added requirement of specifying which hemisphere, e.g. NORTH or SOUTH, that you want to plot.

```
[8]: mixFiles = fdir+"/%s.mix.h5"%(ftag)
     ion = remix.remix(mixFiles,nstep)
     ion.init_vars('NORTH')
```

## Plot the ionospheric data

The mix object includes extensive plotting routine that has the capability for numerous varibles with excellent choices for the color tables. It also takes advantage of the mix objects ability to calculate dervied quantities, such as magnetic perturbations and electric fields. Unlike the magnetosphere plotting routines it has option to take a gridspec object instead of an axes object. It also has the option be made an inset plot so that it can be easily combined with a magnetosphere plot.

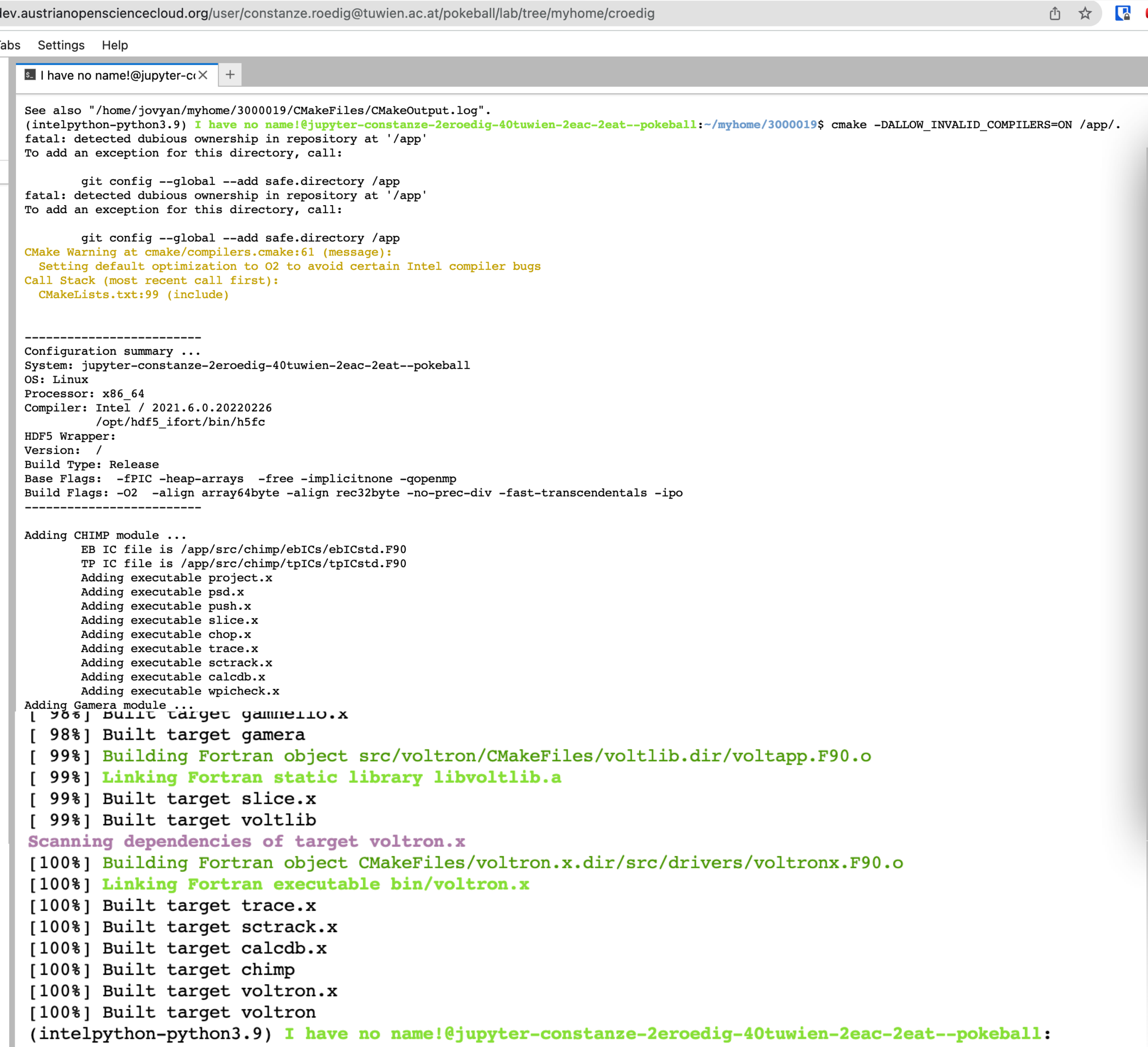
```
[9]: ion.plot('current')
```

```
[9]: <matplotlib.axes._subplots.PolarAxesSubplot at 0x7f7b14368400>
```

min: -113.4
max: 111.8

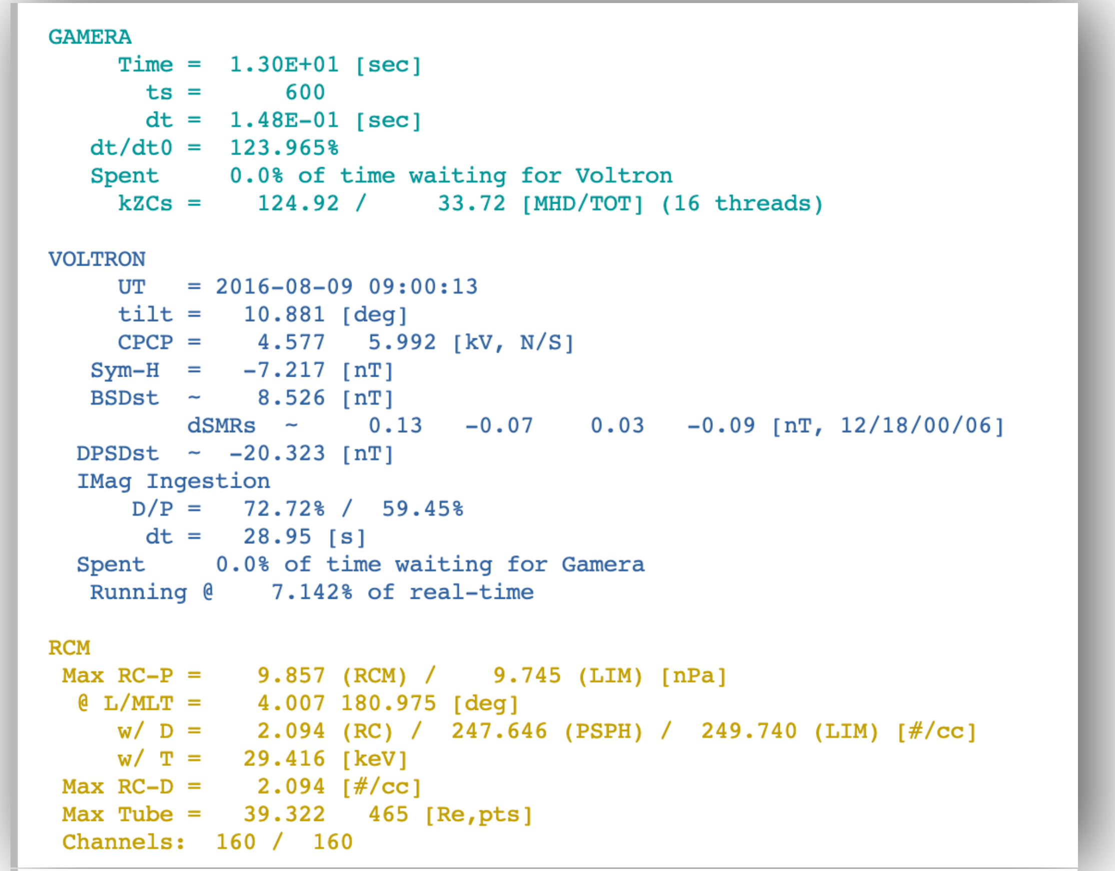
AUSTRIAN DATALAB AND SERVICES

# Demo OpenScience


lev.austrianopensciencecloud.org/user/constanze.roedig@tuwien.ac.at/pokeball/lab/tree/myhome/croedig
abs
Settings
Help
I have no name!@jupyter-c

```
See also "/home/jovyan/myhome/3000019/CMakeFiles/CMakeOutput.log".
(intelpython-python3.9) I have no name!@jupyter-constanze-2eroedig-40tuwien-2eac-2eat--pokeball:~/myhome/3000019$ cmake -DALLOW_INVALID_COMPILERS=ON /app/.
fatal: detected dubious ownership in repository at '/app'
To add an exception for this directory, call:

        git config --global --add safe.directory /app
fatal: detected dubious ownership in repository at '/app'
To add an exception for this directory, call:

        git config --global --add safe.directory /app
CMake Warning at cmake/compilers.cmake:61 (message):
  Setting default optimization to O2 to avoid certain Intel compiler bugs
Call Stack (most recent call first):
  CMakeLists.txt:99 (include)


-------------------------
Configuration summary ...
System: jupyter-constanze-2eroedig-40tuwien-2eac-2eat--pokeball
OS: Linux
Processor: x86_64
Compiler: Intel / 2021.6.0.20220226
          /opt/hdf5_ifort/bin/h5fc
HDF5 Wrapper:
Version:  /
Build Type: Release
Base Flags:  -fPIC -heap-arrays  -free -implicitnone -qopenmp
Build Flags: -O2  -align array64byte -align rec32byte -no-prec-div -fast-transcendentals -ipo
-------------------------

Adding CHIMP module ...
        EB IC file is /app/src/chimp/ebICs/ebICstd.F90
        TP IC file is /app/src/chimp/tpICs/tpICstd.F90
        Adding executable project.x
        Adding executable psd.x
        Adding executable push.x
        Adding executable slice.x
        Adding executable chop.x
        Adding executable trace.x
        Adding executable sctrack.x
        Adding executable calcdb.x
        Adding executable wpicheck.x
Adding Gamera module ...
```

```
[ 98%] Built target gamhello.x
[ 98%] Built target gamera
[ 99%] Building Fortran object src/voltron/CMakeFiles/voltlib.dir/voltapp.F90.o
[ 99%] Linking Fortran static library libvoltlib.a
[ 99%] Built target slice.x
[ 99%] Built target voltlib
Scanning dependencies of target voltron.x
[100%] Building Fortran object CMakeFiles/voltron.x.dir/src/drivers/voltronx.F90.o
[100%] Linking Fortran executable bin/voltron.x
[100%] Built target trace.x
[100%] Built target sctrack.x
[100%] Built target calcdb.x
[100%] Built target chimp
[100%] Built target voltron.x
[100%] Built target voltron
(intelpython-python3.9) I have no name!@jupyter-constanze-2eroedig-40tuwien-2eac-2eat--pokeball:
```



```
GAMERA
     Time =  1.30E+01 [sec]
       ts =       600
       dt =  1.48E-01 [sec]
   dt/dt0 =  123.965%
   Spent      0.0% of time waiting for Voltron
     kZCs =    124.92 /     33.72 [MHD/TOT] (16 threads)

VOLTRON
     UT    = 2016-08-09 09:00:13
     tilt =    10.881 [deg]
     CPCP =     4.577    5.992 [kV, N/S]
   Sym-H   =   -7.217 [nT]
   BSDst   ~    8.526 [nT]
          dSMRs   ~      0.13    -0.07     0.03    -0.09 [nT, 12/18/00/06]
  DPSDst  ~  -20.323 [nT]
  IMag Ingestion
      D/P =    72.72% /  59.45%
       dt =    28.95 [s]
  Spent     0.0% of time waiting for Gamera
   Running @     7.142% of real-time

RCM
 Max RC-P =     9.857 (RCM) /     9.745 (LIM) [nPa]
  @ L/MLT =     4.007 180.975 [deg]
    w/ D =     2.094 (RC) /  247.646 (PSPH) /  249.740 (LIM) [#/cc]
    w/ T =    29.416 [keV]
 Max RC-D =     2.094 [#/cc]
 Max Tube =    39.322    465 [Re,pts]
 Channels:   160 /   160
```

# Agenda

Focus: Use Cases for the Austrian Open Cloud Community

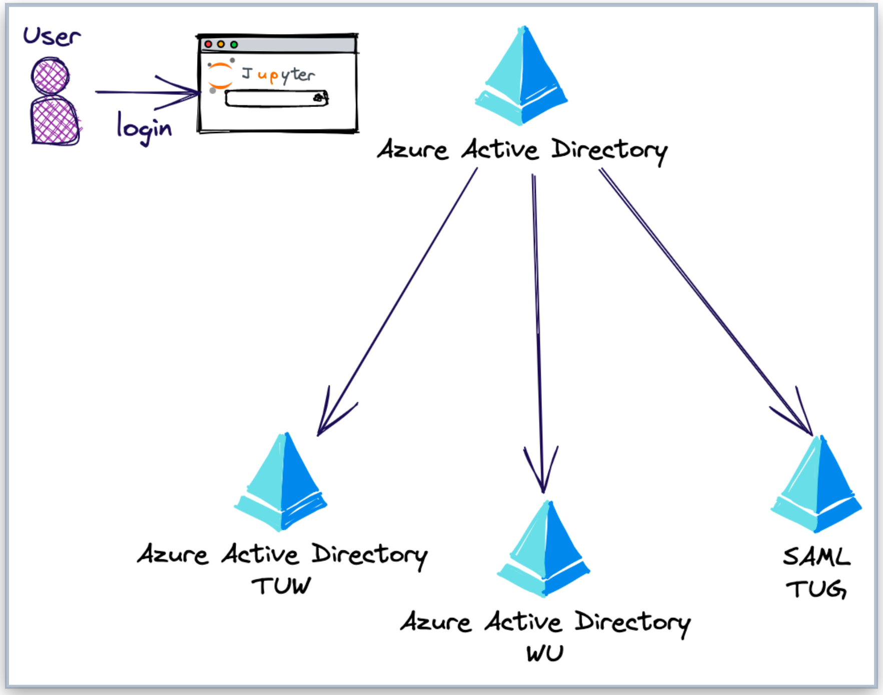
1) OpenScience with GeoSpace example
   a) Our version of Jupyter to make this possible
   b) How we designed a Kubernetes-based platform to run this

2) Teaching Cloud Native, DevOps and Security
   a) Hands On Collaborative Approach
   b) Capture the flag /scavenger hunt type of gamified exams

Community Aspects: how you can get involved

# Jupyter as Web Runtime with strong Controls

- We adapted Jupyter to include custom and strong security and access restrictions which we can dynamically enforce
- Kubernetes Operator for quota/network/application access etc

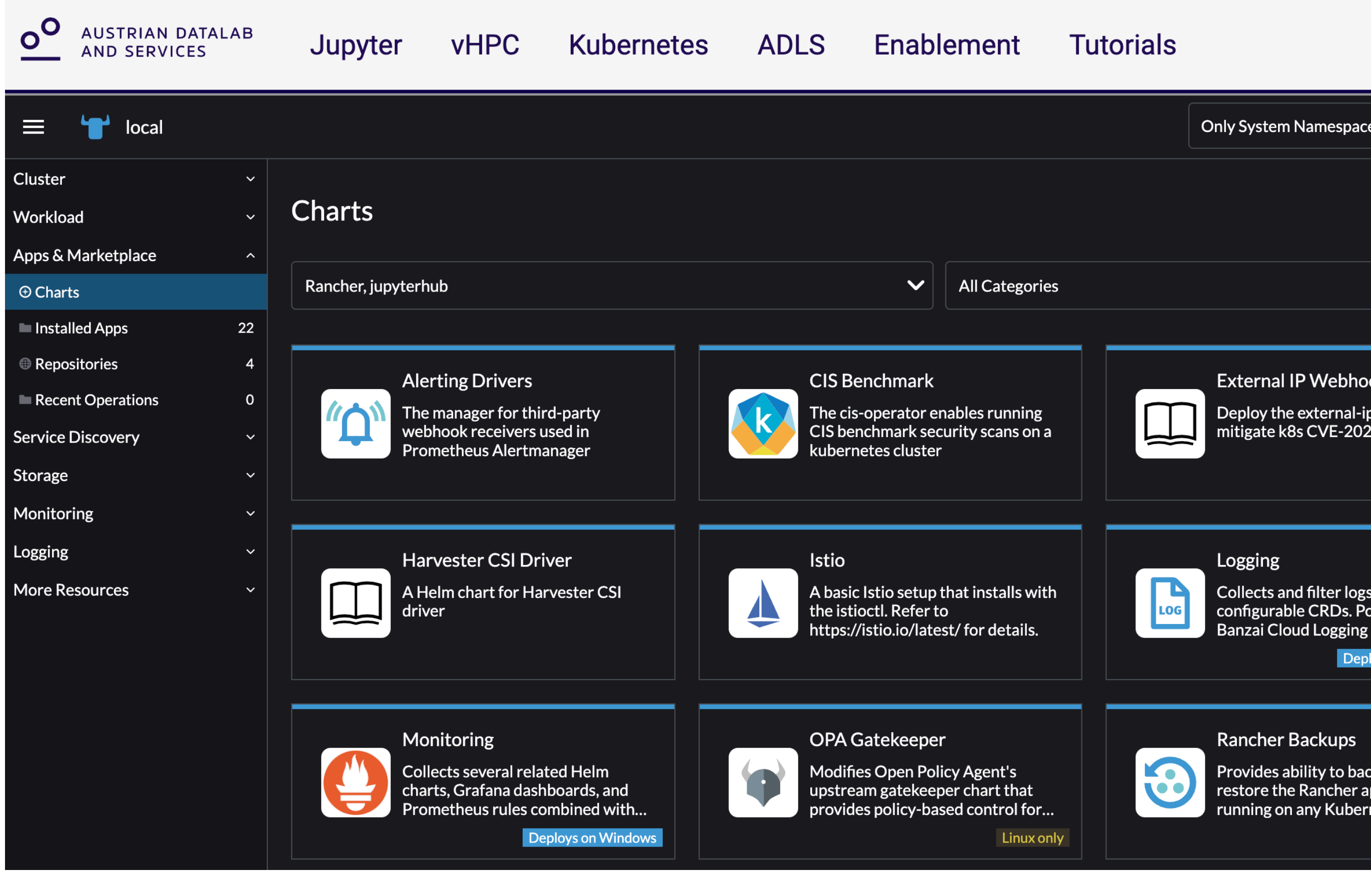


Future Work:

- Automated IAM for remote pipelines
- User management self-service
- Extensions such as 3D Rendering/GPU

# What is Kubernetes | k8s ?

- Orchestrates containers
- Highly extensible and scalable
- web-based, multi-user, interactive computing environment
- across all programming languages



Charts

Rancher, jupyterhub

All Categories

**Alerting Drivers**
The manager for third-party webhook receivers used in Prometheus Alertmanager

**CIS Benchmark**
The cis-operator enables running CIS benchmark security scans on a kubernetes cluster

**External IP Webho**
Deploy the external-i mitigate k8s CVE-202

**Harvester CSI Driver**
A Helm chart for Harvester CSI driver

**Istio**
A basic Istio setup that installs with the istioctl. Refer to https://istio.io/latest/ for details.

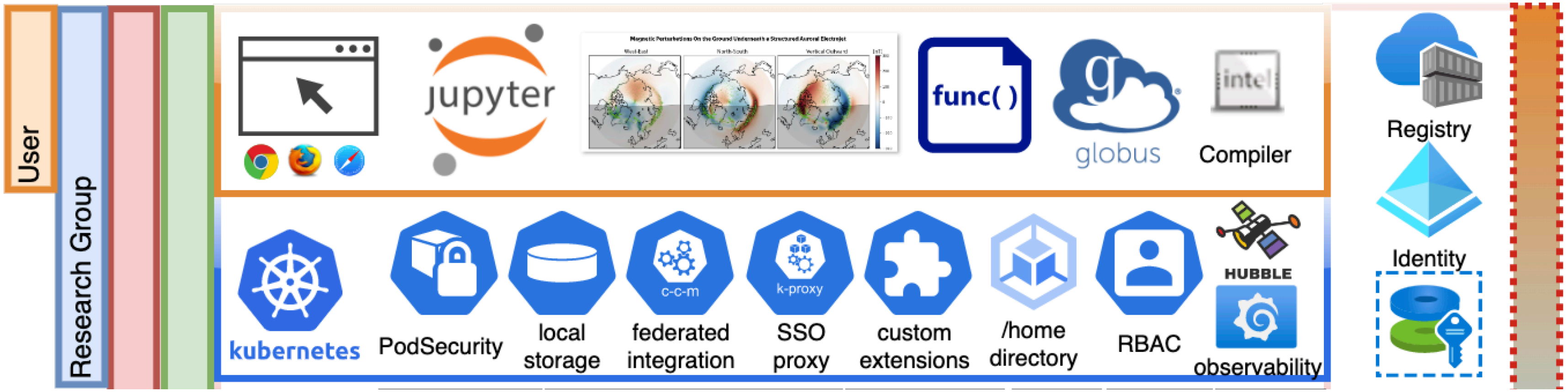
**Logging**
Collects and filter logs configurable CRDs. Po Banzai Cloud Logging
Depl

**Monitoring**
Collects several related Helm charts, Grafana dashboards, and Prometheus rules combined with...
Deploys on Windows

**OPA Gatekeeper**
Modifies Open Policy Agent's upstream gatekeeper chart that provides policy-based control for...
Linux only

**Rancher Backups**
Provides ability to bac restore the Rancher a running on any Kuber

# What's special about our k8s?

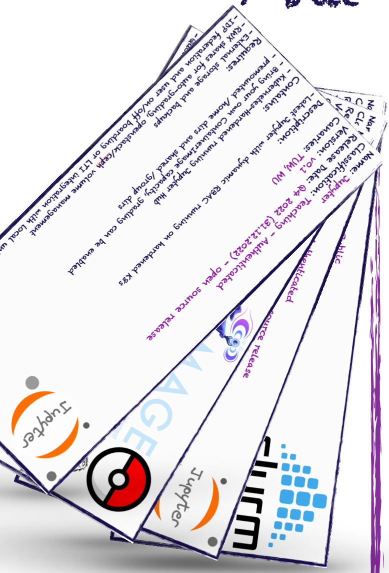


- Given the public audience of OpenScience, our clusters are built for "compromise" :
  - Short life spans (of everything)
- Externalized storage capable of single and shared drives across many clusters (security still being further improved, currently supports only public data)
- Designed for usability in low-cost, low-maintenance academic environment

# Security is a crucial factor

## In order to be "open", we need:

1. Network-segregation for each Lab
   -> no-one can "hide out" in a uni's infrastructure
2. To access data and share it, but
   -> can not use it to "move across" Labs
3. Application runtime is open
   -> but "lower layers" are off-limits

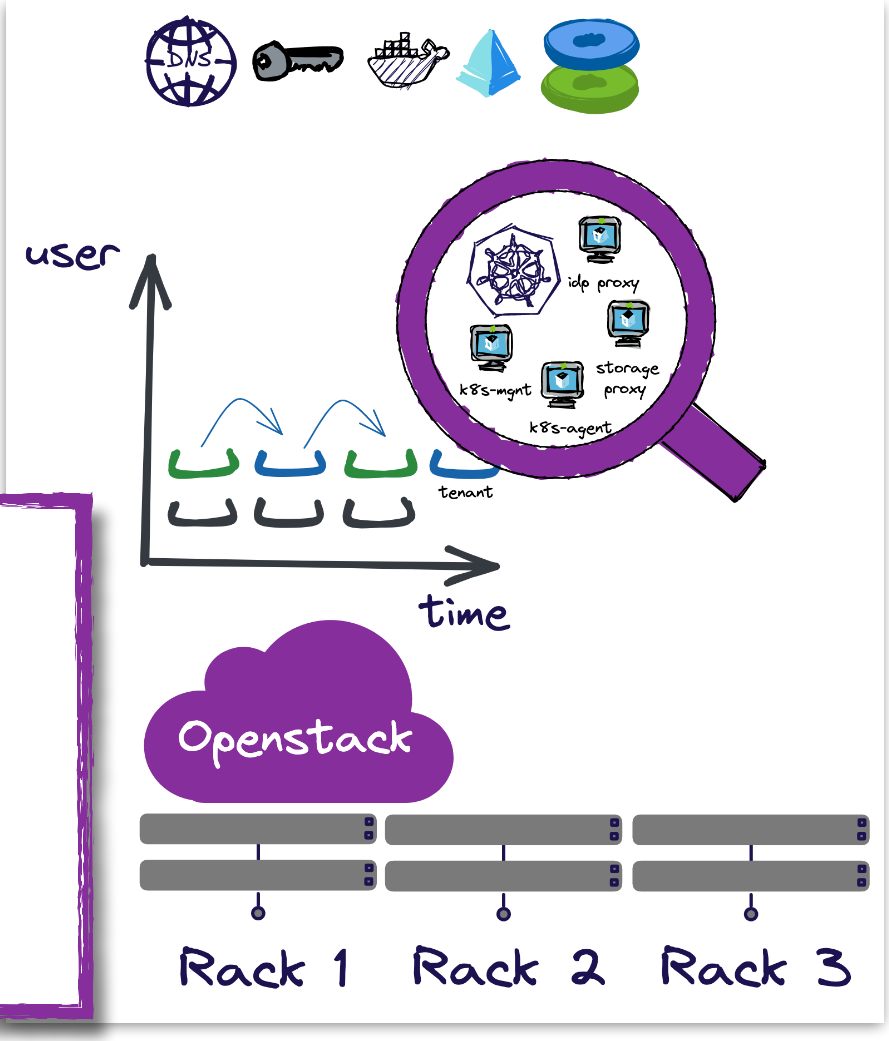


First (penetration) tests show

1. Network of each Lab is truly isolated

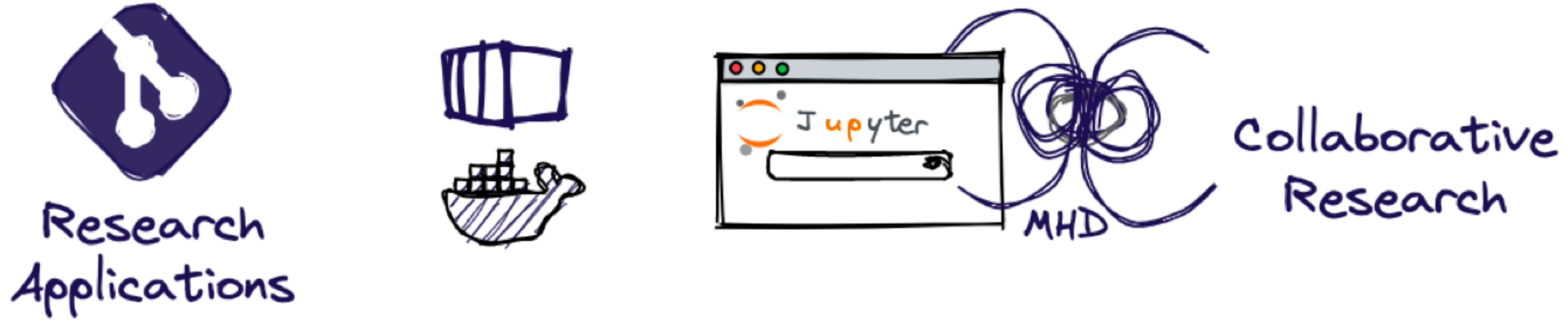
2. Labs usable for public and non-PII data

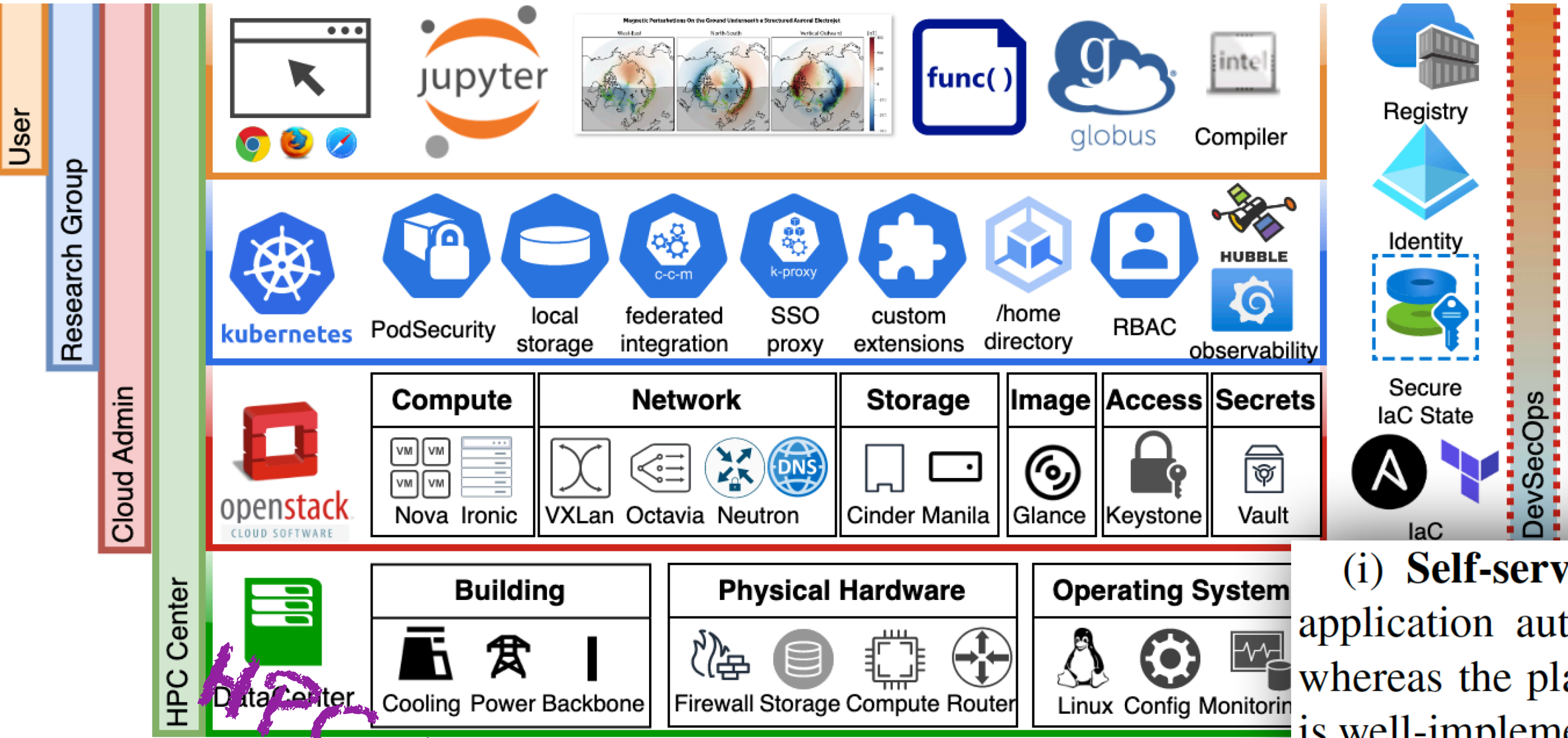
3. Breaking out of Jupyter Labs requires significant effort, skill, luck, time



# Summary

## OpenScienceLabs





- Institutes will want to choose where to host the infrastructure
- Admins will want to minimize their maintenance efforts
- Cost must be calculated taking into account sustainability and a digital sovereignty perspective[1]
- Researchers will want to focus on their differentiating research content, not the underlying cloud technology
- Users will be willing to sacrifice performance for convenience (and use HPC clusters for actually performance relevant runs)
- Content of the labs (code and data) is public[2]
- No GDPR (and equivalent) protected data is needed from users

(i) **Self-service orientation and scalability**: the application and application automation is in clear ownership of the researchers, whereas the platform ownership is with the admins. If automation is well-implemented, researchers can self-service use a tenant of the platform without incurring maintenance effort on the admin side.

(ii) **Train the trainer programs**: given that the research users are only exposed to a very small subset of the platform complexity, we can train senior researchers and other tech-savvy users in a reasonable time to become trainers themselves

HPC Teams
Get in touch

# Agenda

Focus: Use Cases for the Austrian Open Cloud Community

1) OpenScience with GeoSpace example
   a) Our version of Jupyter to make this possible
   b) How we designed a Kubernetes-based platform to run this

2) Teaching Cloud Native, DevOps and Security
   a) Hands On Collaborative Approach
   b) Capture the flag /scavenger hunt type of gamified exams

Community Aspects: how you can get involved

# 1. Fundamentals of cloud, security and DevOps

Status: alpha
Version: v0.1
Release Date: Summer Term 2023 (01.03.2023)
Canaries: TUW, UIBK, WU*, BOKU*

Goals:
- provide overview of current best practice, relevant (anti)-patterns and theory
- provide free access to infrastructure to practice safely what it means to "fail fast"
- provide experience of collaboratively maintaining a live service

Challenges:
- Scalability, Legal, Logistics, Accreditation

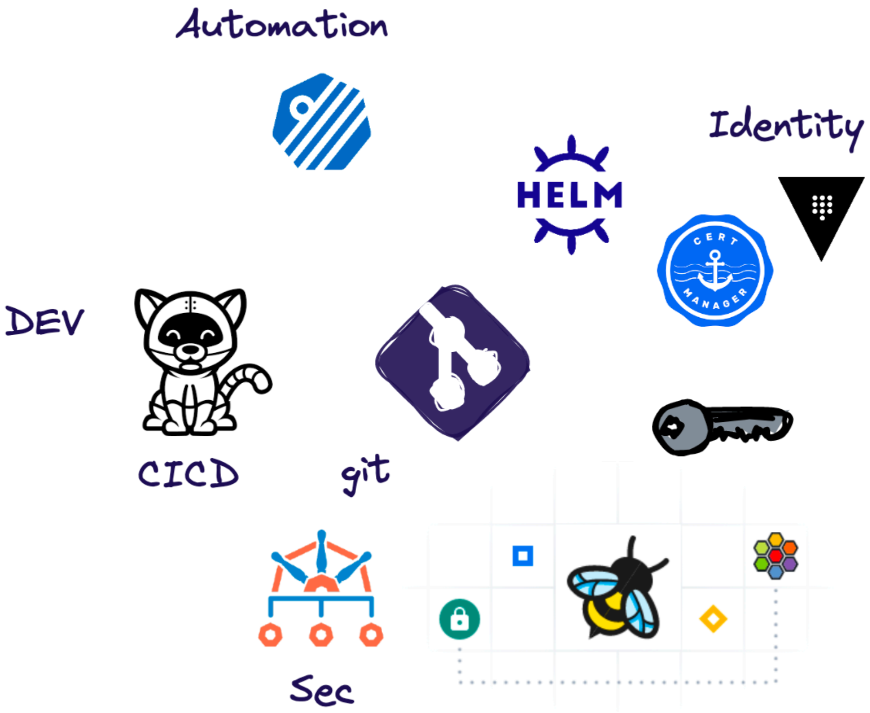
# Profile: Buildsystem

Name: Buildsystem
Classification: Teaching - Authenticated
Release Date: Q2 2023 (1.03.2023)
Version: v0.1
Canaries: TUW, UIBK, WU, BOKU

Description: Cloud Native DevSecOps Tooling

Contains:

- All components one needs for a secure supply chain
- Reference application integration
- Rootless versions of upstream applications

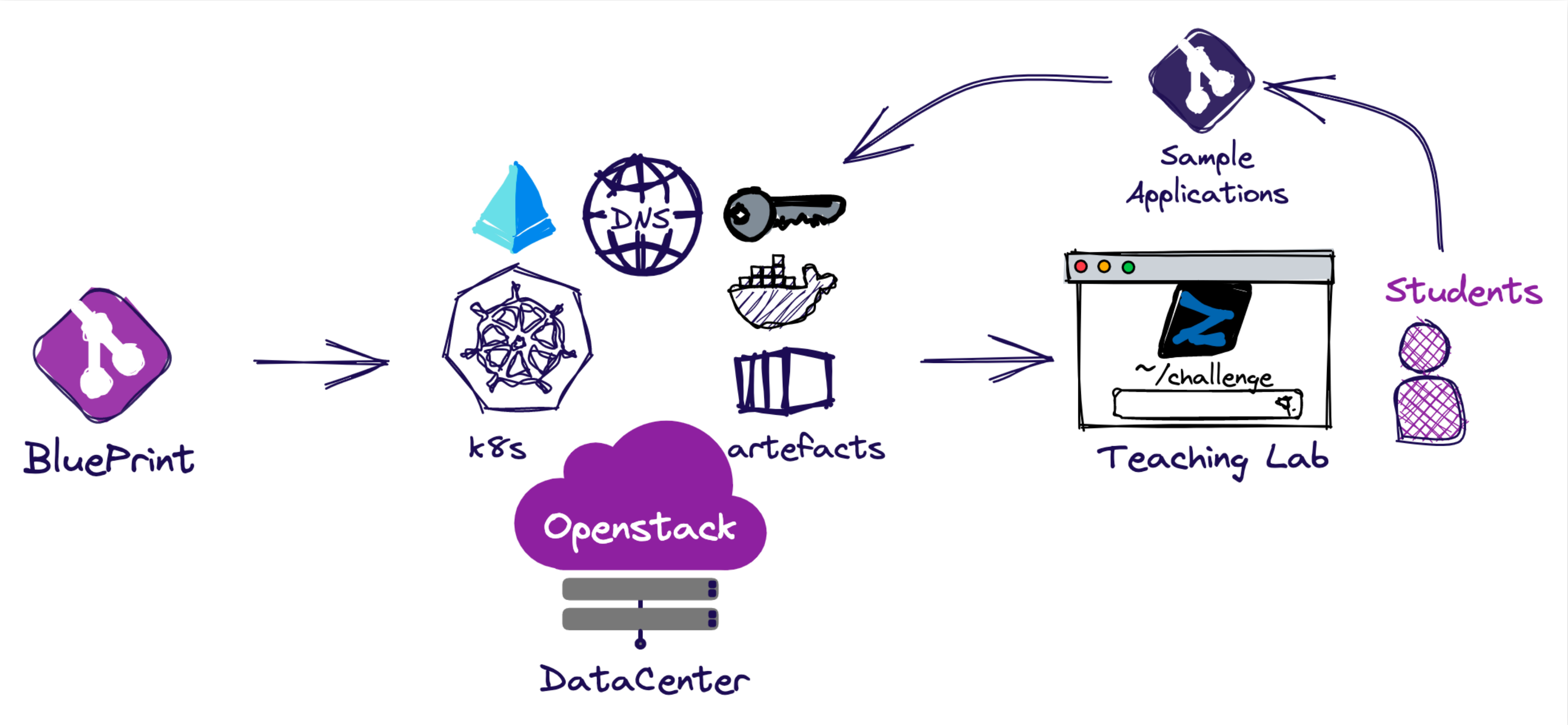


Requires:

- Our reference app : Pac-Man as a Service
- The core cluster profile for Harbor
- External APIs like github/slack/jira etc

```
Cloud: ADLS-Rancher
Zone: nova
Host: pacman-rancher-5b58bbb4f-
s9rf5
Time: 20                Level: 1
Score: 0                Lives: ♥ ♥ ♥
```

# Virtual Labs & CTF like Exams

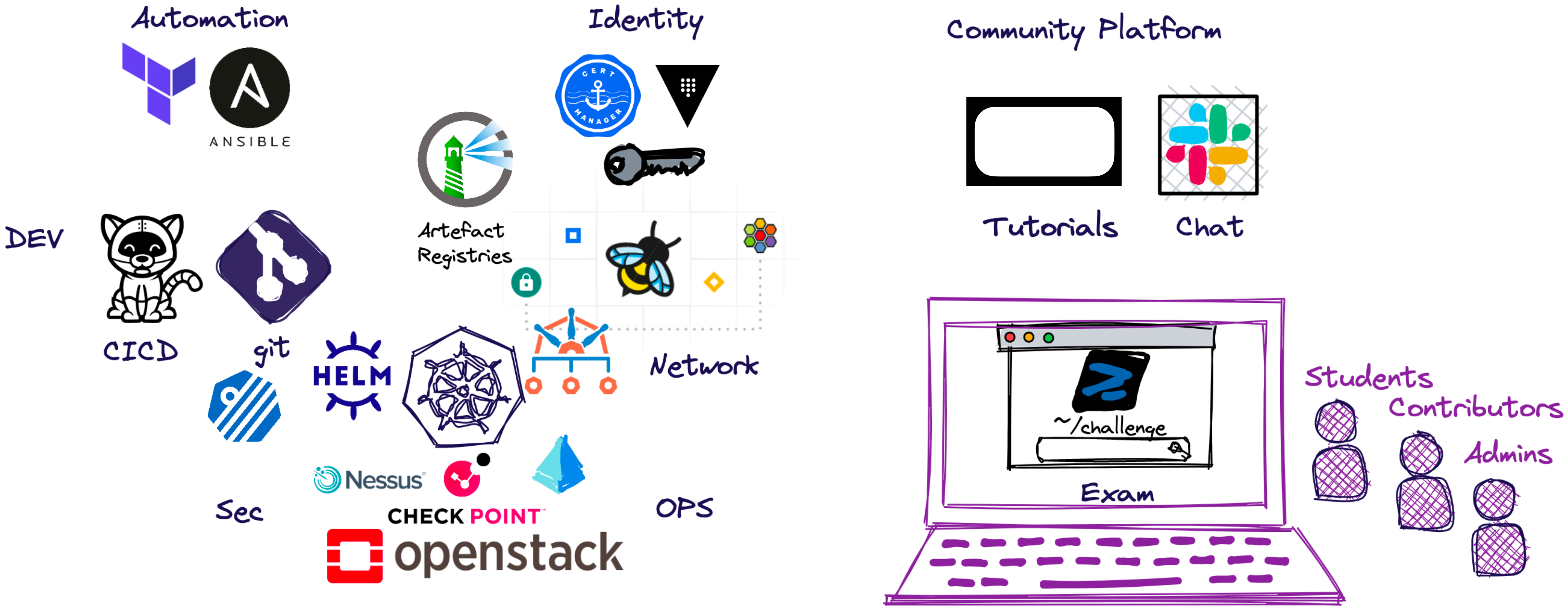


## On demand cloud infrastructure provides:

Hosting various emulated or virtualized IT components

Examples: kubernetes inside a container (hosted on a full scale kubernetes)
Exams in CTF style on ephemeral infrastructure

# Virtual Labs & CTF Like Exams



## Develop a fully functional mini service (as team)

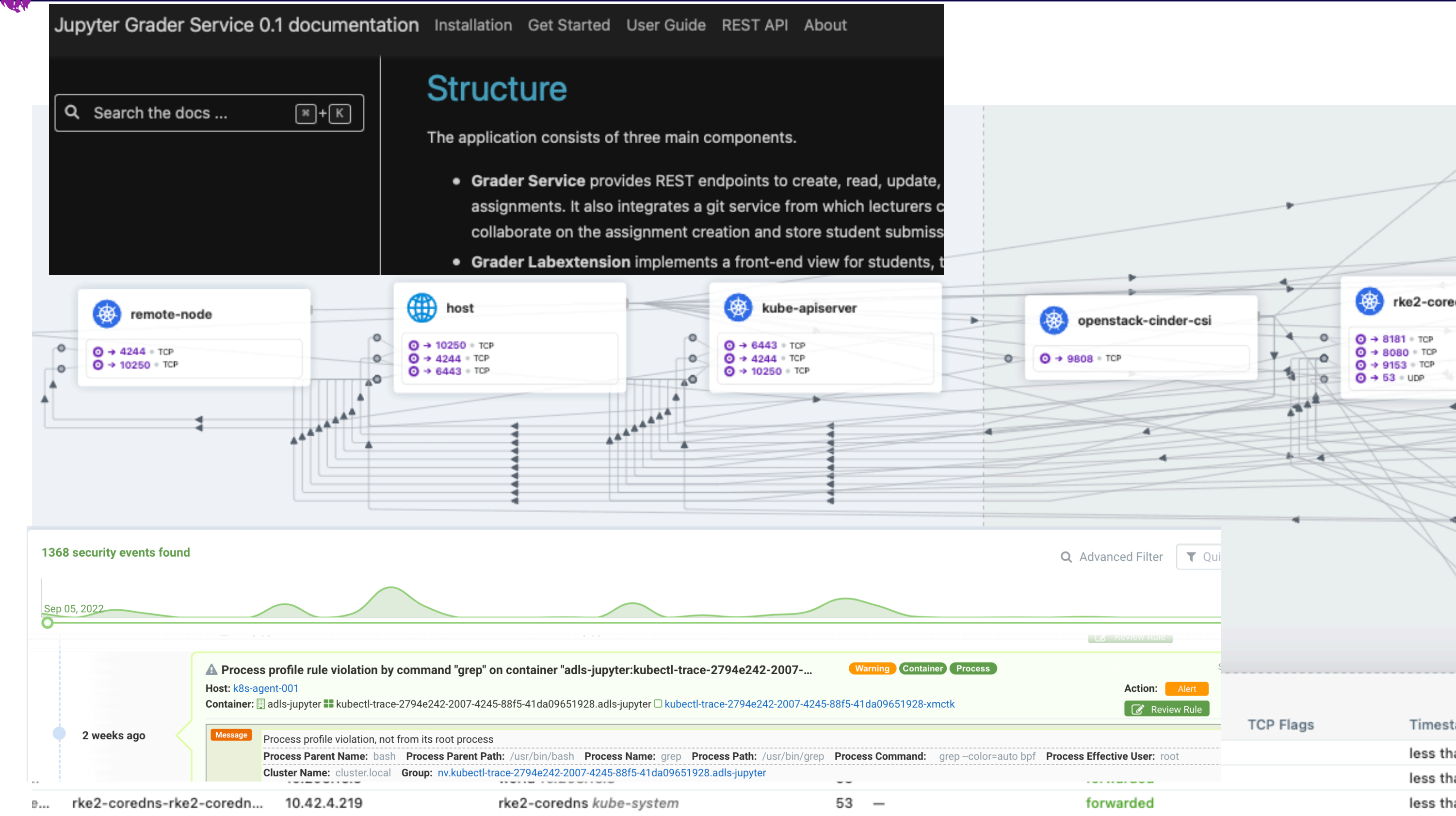
Learn how cloud native works by working on a end 2 end open source** cloud stack

Cover full cycle of relevant topics: collaboration, architecture patterns, licensing, feedback-loops, vulnerabilities, observability, SLAs, code of conduct, incident response

** exceptions may apply

# Pedagogical Approach

8 hrs per week



Jupyter Grader Service 0.1 documentation Installation Get Started User Guide REST API About

Search the docs ...

## Structure

The application consists of three main components.

- **Grader Service** provides REST endpoints to create, read, update, assignments. It also integrates a git service from which lecturers c collaborate on the assignment creation and store student submiss
- **Grader Labextension** implements a front-end view for students, t

1368 security events found

Advanced Filter Qui

Sep 05, 2022

2 weeks ago

⚠ Process profile rule violation by command "grep" on container "adls-jupyter:kubectl-trace-2794e242-2007-... Warning Container Process

Host: k8s-agent-001

Container: adls-jupyter kubectl-trace-2794e242-2007-4245-88f5-41da09651928.adls-jupyter kubectl-trace-2794e242-2007-4245-88f5-41da09651928-xmctk

Action: Alert

Review Rule

Message Process profile violation, not from its root process

Process Parent Name: bash Process Parent Path: /usr/bin/bash Process Name: grep Process Path: /usr/bin/grep Process Command: grep --color=auto bpf Process Effective User: root

Cluster Name: cluster.local Group: nv.kubectl-trace-2794e242-2007-4245-88f5-41da09651928.adls-jupyter

TCP Flags Timest

less tha

less tha

e... rke2-coredns-rke2-coredn... 10.42.4.219 rke2-coredns *kube-system* 53 — forwarded less tha

## From scratch

build a concept from first principle

## Best practise

In "your startup" : solve a real life problem
Collaborate in a (local) group
Work on "your startup service"

# 1. Key choices in pedagogical approach

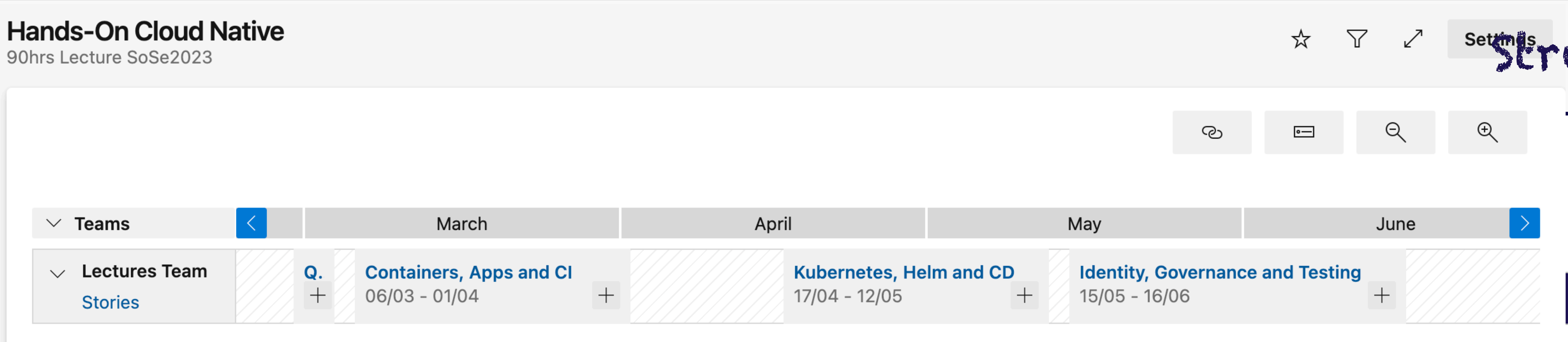
Goals:

- provide overview of current best practice, relevant (anti)-patterns and theory
- provide free access to infrastructure to practice safely what it means to "fail fast"
- provide experience of collaboratively maintaining a live service

Desired outcomes:

- Exposure to real-life like work situation (overabundance of "stack")
- Reliance on "active creation/invention" rather than indoctrination
- Create a playful atmosphere

# Structure



Hands-On Cloud Native

90hrs Lecture SoSe2023

|  |  |  |  |  |
|---|---|---|---|---|
| Teams | March | April | May | June |

| Lectures Team | Q. | Containers, Apps and CI | Kubernetes, Helm and CD | Identity, Governance and Testing |
|---|---|---|---|---|
| Stories |  | 06/03 - 01/04 | 17/04 - 12/05 | 15/05 - 16/06 |

|  |  |  |  |
|---|---|---|---|
| **Theoretical introduction** | blackboard, demo or interview with someone from the industry applying the concept on a daily basis | beginning of each session | lecturer an/or experts from the field<br>“default” lecture hall is Vienna<br>but guest locations are planned |
| **Exercise persentation** | the objectives are stated<br>the virtual lab setup explained | end of theory session | lecturer, experts from the field |
| **Work on exercises** | students work on the tasks in groups<br>Support from staff is avail<br>Community support is highly encouraged | During the entire 4 hrs, this is fully up to the students how they achieve their learning, just like in agile: the “how” is up to every one | Students (in groups) : a startup is formed to develop a service<br>This group should stay constant |
| **Offsite work and chat** | A community chat and other community tools will be available during the entire semester | 24/7 | students<br>staff will be avail on "best effort” |
| **Exam** | - Collaborative: each “startup” open-sources their fully functional application<br>- Individual: play 3x 4 hr CTF that is graded by writeup, not by flag submission | Presentation of Lessons Learnt = last week<br>last week of each block (CTF) | students |

# Agenda

Focus: Use Cases for the Austrian Open Cloud Community

1) OpenScience with GeoSpace example
   a) Our version of Jupyter to make this possible
   b) How we designed a Kubernetes-based platform to run this

2) Teaching Cloud Native, DevOps and Security
   a) Hands On Collaborative Approach
   b) Capture the flag /scavenger hunt type of gamified exams

Community Aspects: how you can get involved

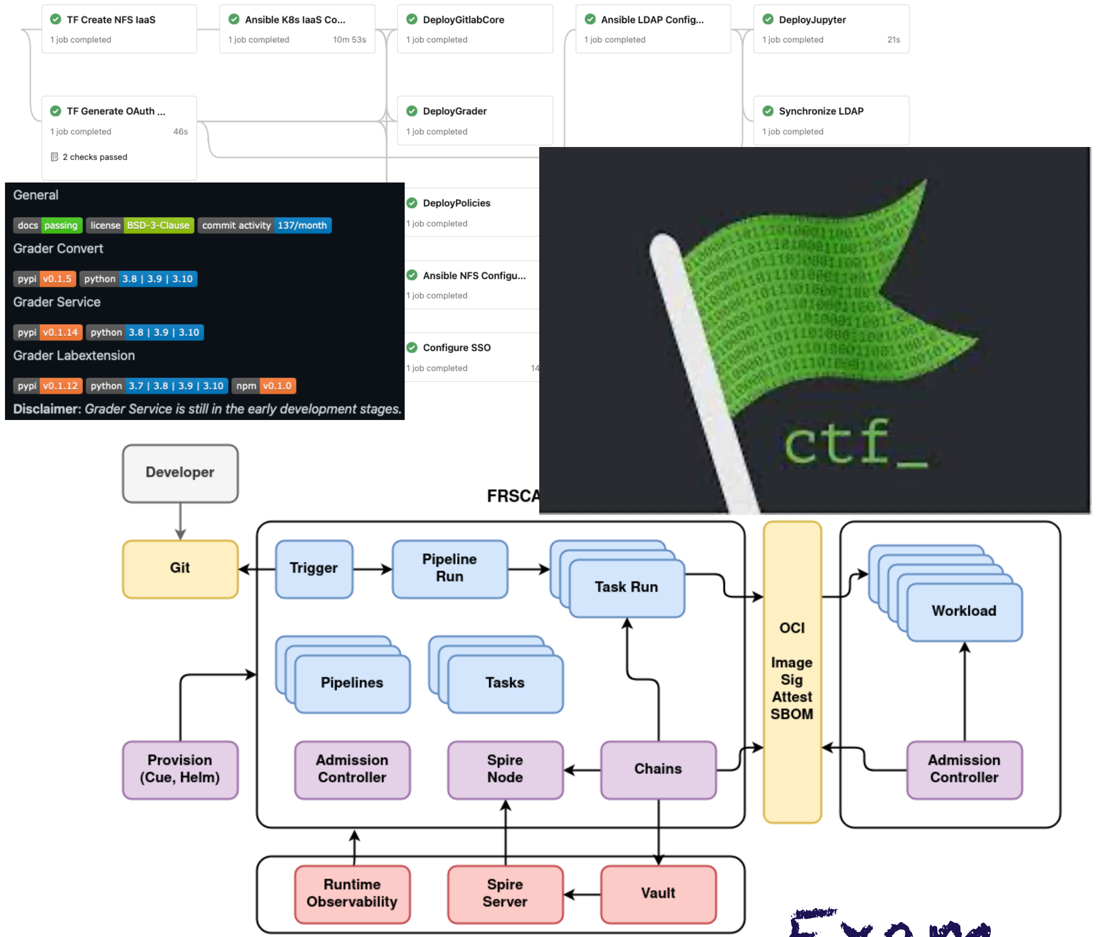
AUSTRIAN DATALAB AND SERVICES

# Selection and Grading

```
Select top XX sum(flag)

FROM: *
GROUP BY: uni

If (uni in {TUW, UIBK})
Then
  enroll direct
Else
  co-enroll @TUW
```

## Entrance qualifier

Series of hands-on challenges:
- operating system
- network
- git, scripting and logic



TF Create NFS IaaS
1 job completed

Ansible K8s IaaS Co...
1 job completed 10m 53s

DeployGitlabCore
1 job completed

Ansible LDAP Config...
1 job completed

DeployJupyter
1 job completed 21s

TF Generate OAuth ...
1 job completed 46s
2 checks passed

DeployGrader
1 job completed

Synchronize LDAP
1 job completed

DeployPolicies
1 job completed

Ansible NFS Configu...
1 job completed

Configure SSO
1 job completed 14

General

docs passing | license BSD-3-Clause | commit activity 137/month

Grader Convert

pypi v0.1.5 | python 3.8 | 3.9 | 3.10

Grader Service

pypi v0.1.14 | python 3.8 | 3.9 | 3.10

Grader Labextension

pypi v0.1.12 | python 3.7 | 3.8 | 3.9 | 3.10 | npm v0.1.0

**Disclaimer:** *Grader Service is still in the early development stages.*

## Exam

-Collaborative:
  student group "startup" open-source
  their fully functional (mini)application
-Individual: Play a 4hr CTF

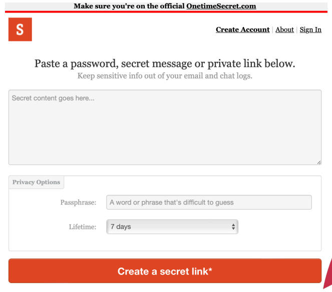
# Profile: Exam

Name: Exam
Classification: Internal - Authenticated
Release Date: TBD
Version: v0.1
Canaries: TBD

Description: Exam tools

Contains:
- Applications we need for hosting an exam
- Auxiliary: One Time Secret, Minio , Harbor , gitlab
- CTForge**

Requires:
- Identity Integration
- ..
- ..



Make sure you're on the official OnetimeSecret.com

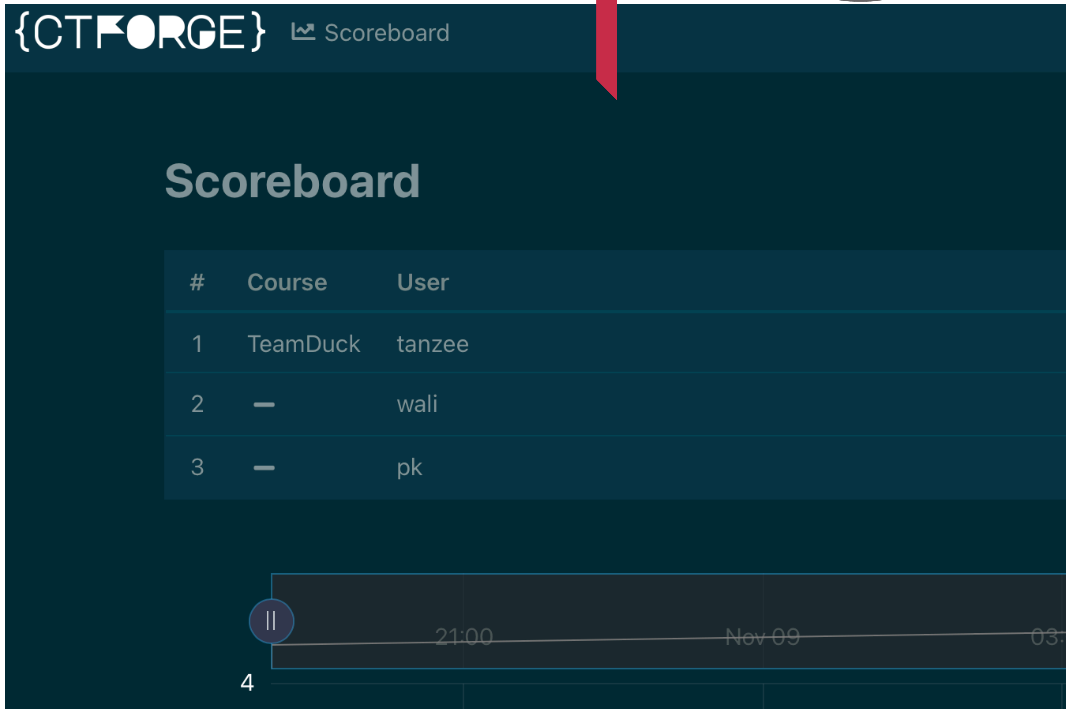
Paste a password, secret message or private link below.
Keep sensitive info out of your email and chat logs.

Secret content goes here...

Privacy Options

Passphrase: A word or phrase that's difficult to guess

Lifetime: 7 days

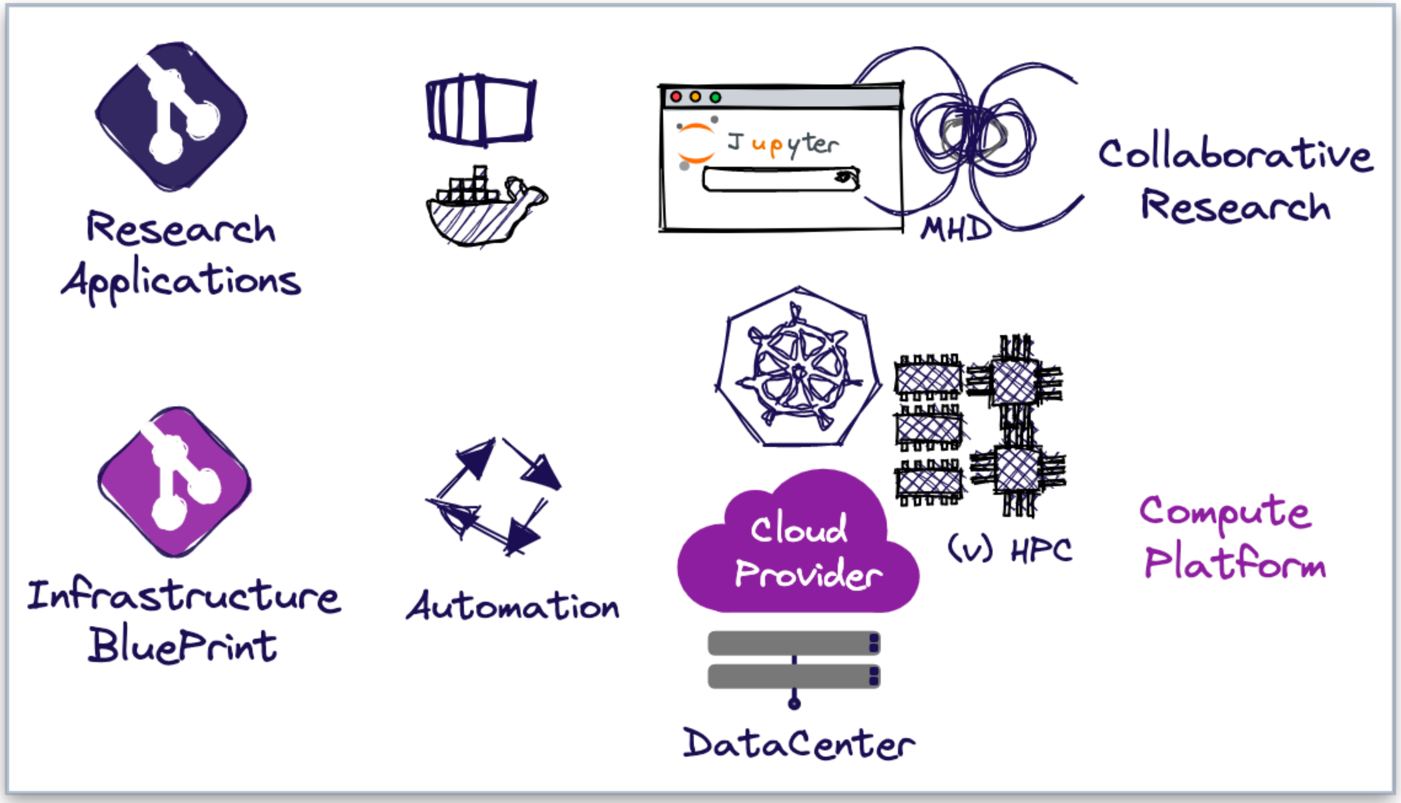


Scoreboard

| # | Course | User |
|---|---|---|
| 1 | TeamDuck | tanzee |
| 2 | — | wali |
| 3 | — | pk |

**Collaboration with TUW Informatics, Security & Privacy

# Goal - open community



## Establish a self supporting community

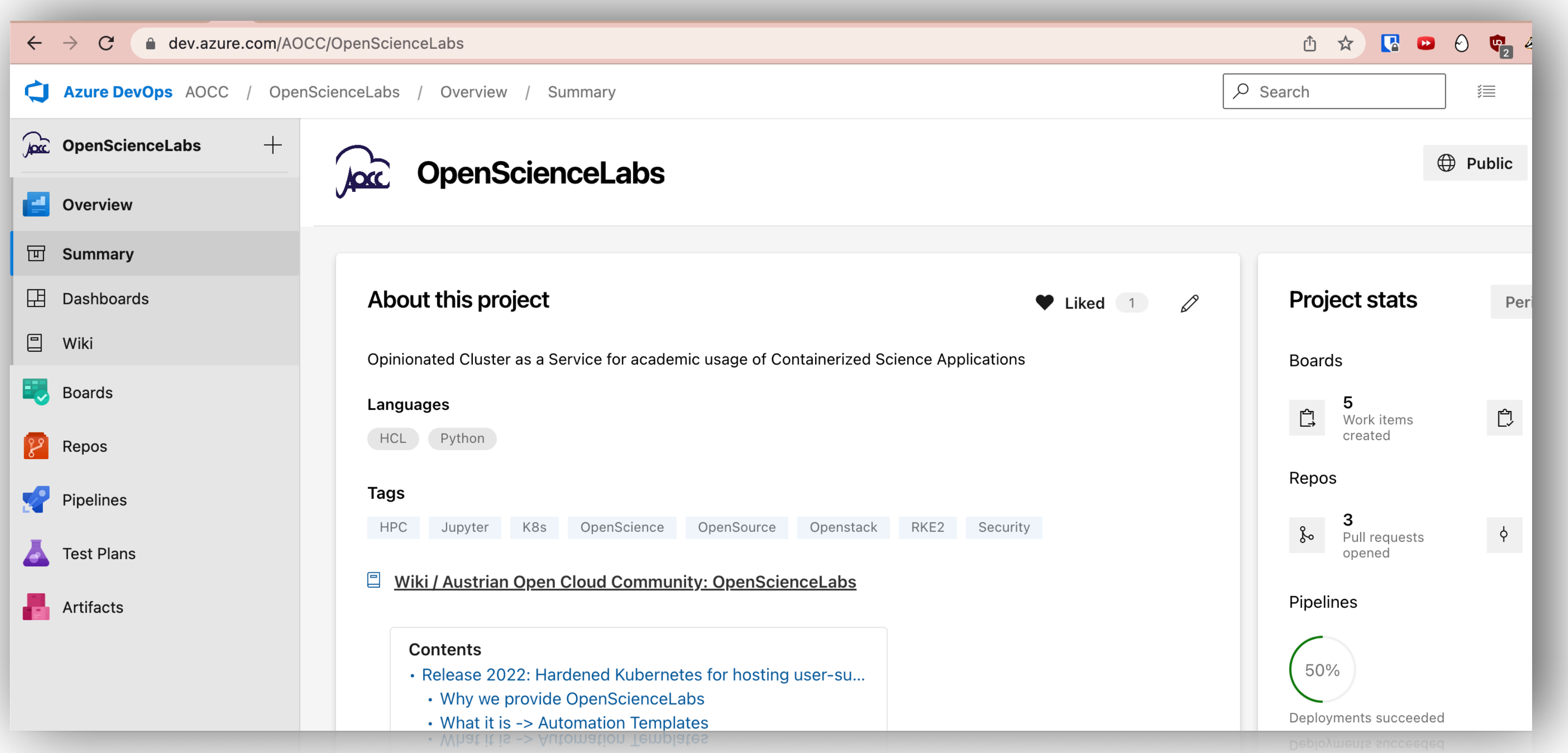
benefit and give back to the community, gain visibility, share your ideas and results, gain new insights, use state of the art technologies

AOCC

# Austrian Open Cloud Community

Starting 2023 we emphasize the community aspect of OpenScience and OpenSource
Thus, the project will work fully in the open

# BluePrints now OpenSourced



dev.azure.com/AOCC/OpenScienceLabs

## OpenScienceLabs

### About this project

Opinionated Cluster as a Service for academic usage of Containerized Science Applications

Languages: HCL, Python

Tags: HPC, Jupyter, K8s, OpenScience, OpenSource, Openstack, RKE2, Security

Wiki / Austrian Open Cloud Community: OpenScienceLabs

Contents
- Release 2022: Hardened Kubernetes for hosting user-su...
  - Why we provide OpenScienceLabs
  - What it is -> Automation Templates

### Project stats

Boards: 5 Work items created

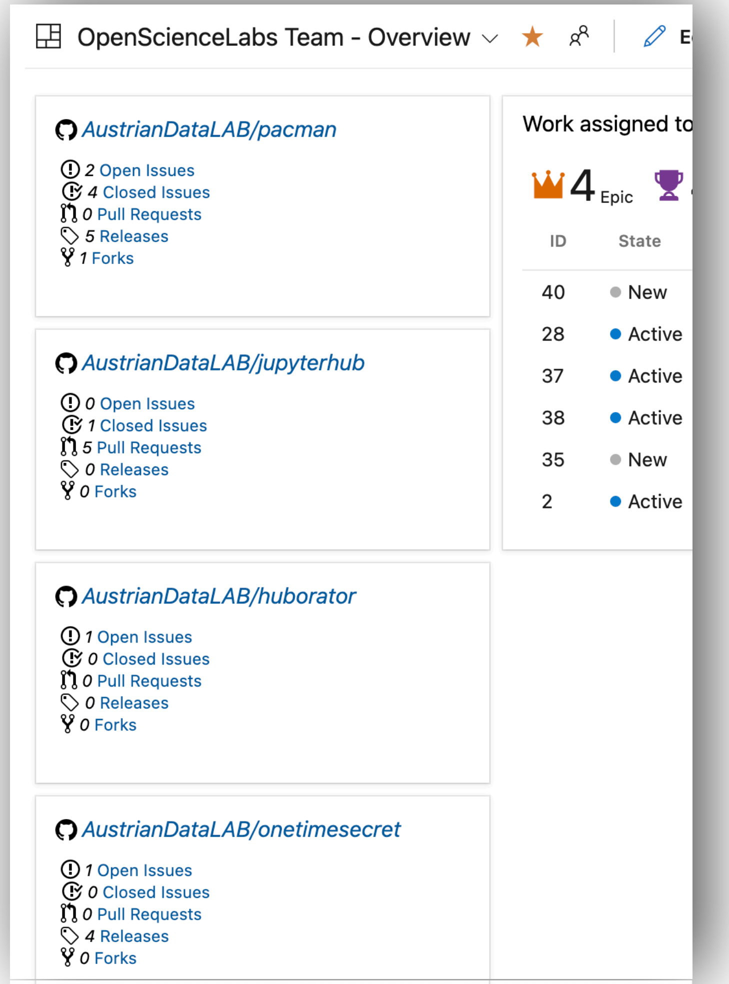
Repos: 3 Pull requests opened

Pipelines: 50% Deployments succeeded

## OpenSourcing the first "batch"

Over NewYear, we made a lot of our work public. We are now working on Documentation , (re) testing and working samples
Also: Automated overnight testing and other CSPs are in the works

# Priority: Working Examples



OpenScienceLabs Team - Overview

AustrianDataLAB/pacman

- 2 Open Issues
- 4 Closed Issues
- 0 Pull Requests
- 5 Releases
- 1 Forks

AustrianDataLAB/jupyterhub

- 0 Open Issues
- 1 Closed Issues
- 5 Pull Requests
- 0 Releases
- 0 Forks

AustrianDataLAB/huborator

- 1 Open Issues
- 0 Closed Issues
- 0 Pull Requests
- 0 Releases
- 0 Forks

AustrianDataLAB/onetimesecret

- 1 Open Issues
- 0 Closed Issues
- 0 Pull Requests
- 4 Releases
- 0 Forks

Work assigned to

4 Epic

| ID | State |
|---|---|
| 40 | ● New |
| 28 | ● Active |
| 37 | ● Active |
| 38 | ● Active |
| 35 | ● New |
| 2 | ● Active |

To achieve portability, we want to emphasize the ease of "setting it up".
We wish to provide sample and poster-child content, maybe even live-labs for admins

Also: individual workshops with other institutes

Offer consulting like interactions with interested institutions

## Sample Content & Consulting

BluePrints Repos: https://dev.azure.com/AOCC/OpenScienceLabs

Webpage: https://webportal.dev.austrianopencloudcommunity.org

Chat: https://join.slack.com/t/aocc-public/shared_invite/zt-1mq6yjnet-YqXMln8G

Email: support@austrianopencloudcommunity.org

LinkedIN: https://linkedin.com/in/croedig

Public Demos: Wednesday 3PM weekly

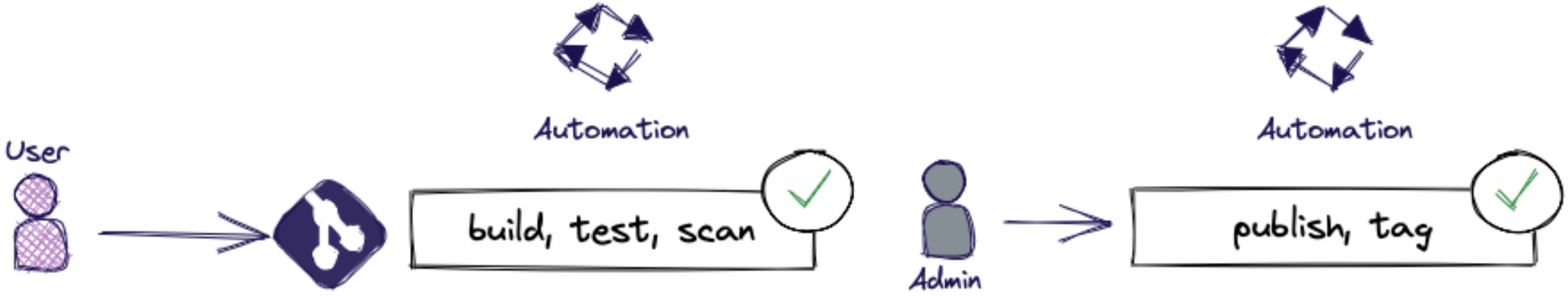
https://tuwien.zoom.us/j/91356277954?pwd=cHpsYW9POXNoL05yU3JLLzVPWm9yQT09

AOCC

Get in touch
Give feedback
Learn with us :)

# Public image-stack for all Users

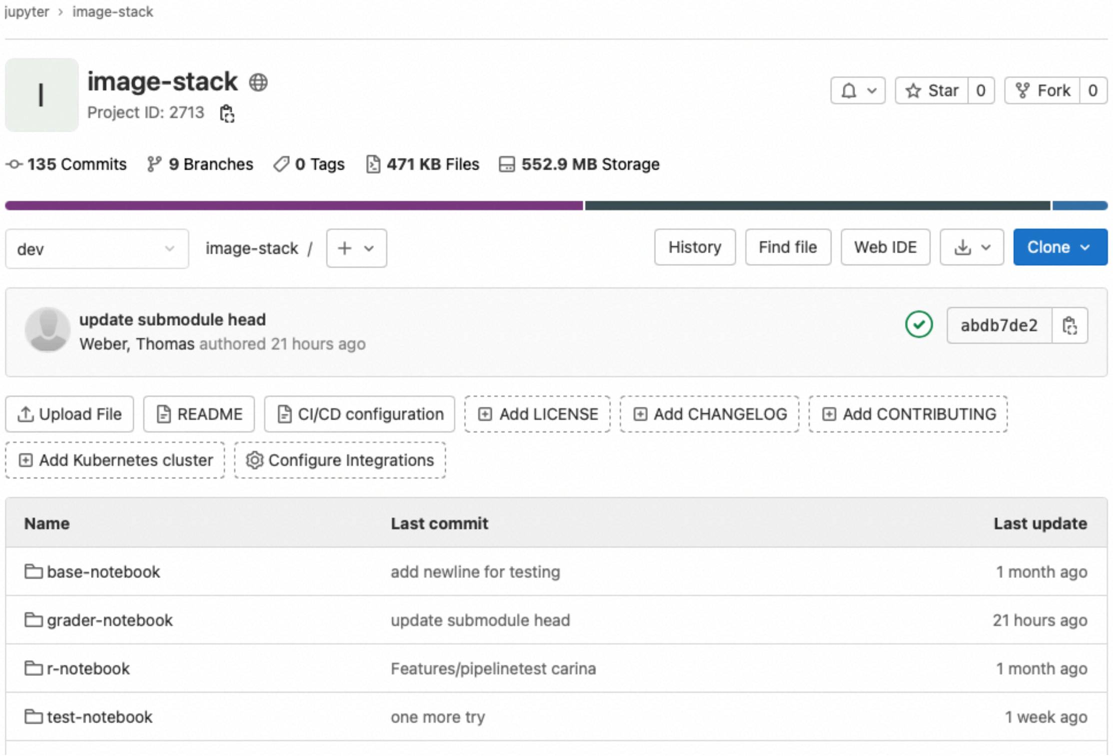
- For UseCase I: a teaching lab in Jupyter
- Community contributions
- Take something from the image stack, add to it/modify existing images
- Focus on teaching, dont worry about underlying layers



# Build System | Enablement

## Public image-stack for all Users

- Users can create Merge Requests for features, or adapt existing ones
- Pipeline handles build, scanning
- Users can add tests for their code
- Admins verify contribution and can trigger publishing to registries – security measure
- Easy handling for users



jupyter › image-stack

image-stack

Project ID: 2713

135 Commits · 9 Branches · 0 Tags · 471 KB Files · 552.9 MB Storage

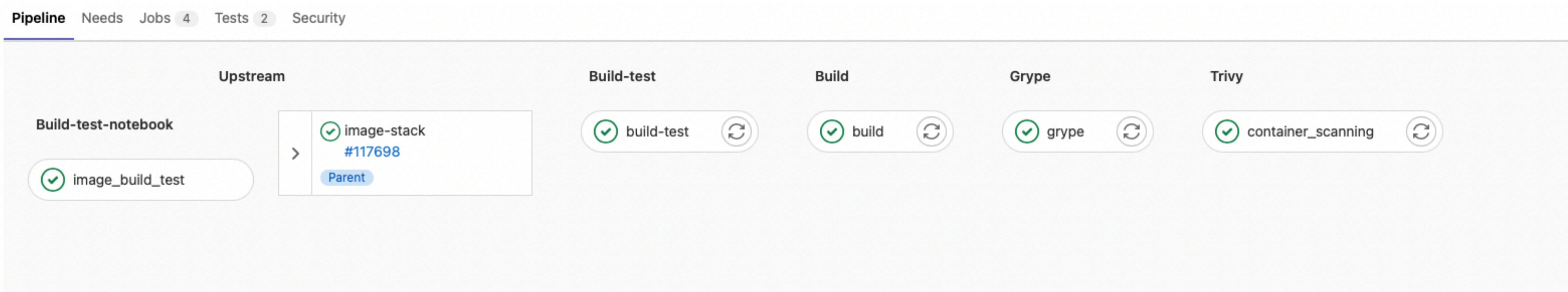
dev · image-stack /

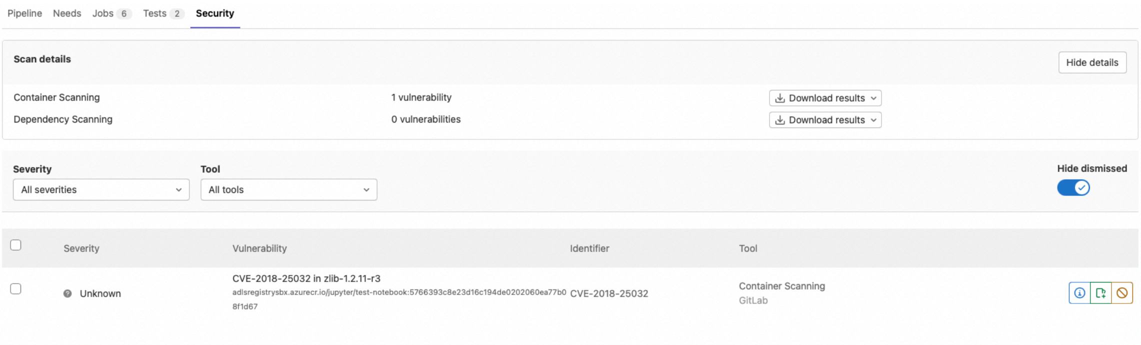
update submodule head
Weber, Thomas authored 21 hours ago · abdb7de2

| Name | Last commit | Last update |
|---|---|---|
| base-notebook | add newline for testing | 1 month ago |
| grader-notebook | update submodule head | 21 hours ago |
| r-notebook | Features/pipelinetest carina | 1 month ago |
| test-notebook | one more try | 1 week ago |

# Pipeline Output

- Check the outcome -> helps to fix something, make modifications





Scan details

Hide details

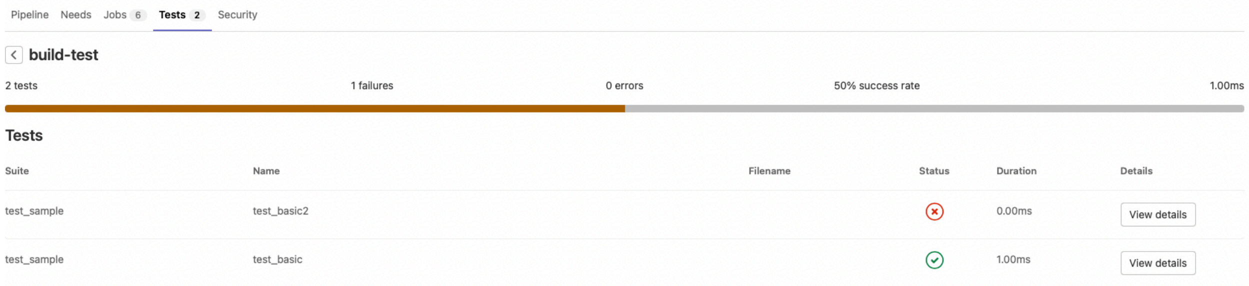
| | | |
|---|---|---|
| Container Scanning | 1 vulnerability | Download results |
| Dependency Scanning | 0 vulnerabilities | Download results |

Severity: All severities

Tool: All tools

Hide dismissed

| □ | Severity | Vulnerability | Identifier | Tool |
|---|---|---|---|---|
| □ | Unknown | CVE-2018-25032 in zlib-1.2.11-r3<br>adlsregistrysbx.azurecr.io/jupyter/test-notebook:5766393c8e23d16c194de0202060ea77b08f1d67 | CVE-2018-25032 | Container Scanning<br>GitLab |



build-test

2 tests | 1 failures | 0 errors | 50% success rate | 1.00ms

Tests

| Suite | Name | Filename | Status | Duration | Details |
|---|---|---|---|---|---|
| test_sample | test_basic2 | | ☒ | 0.00ms | View details |
| test_sample | test_basic | | ✓ | 1.00ms | View details |