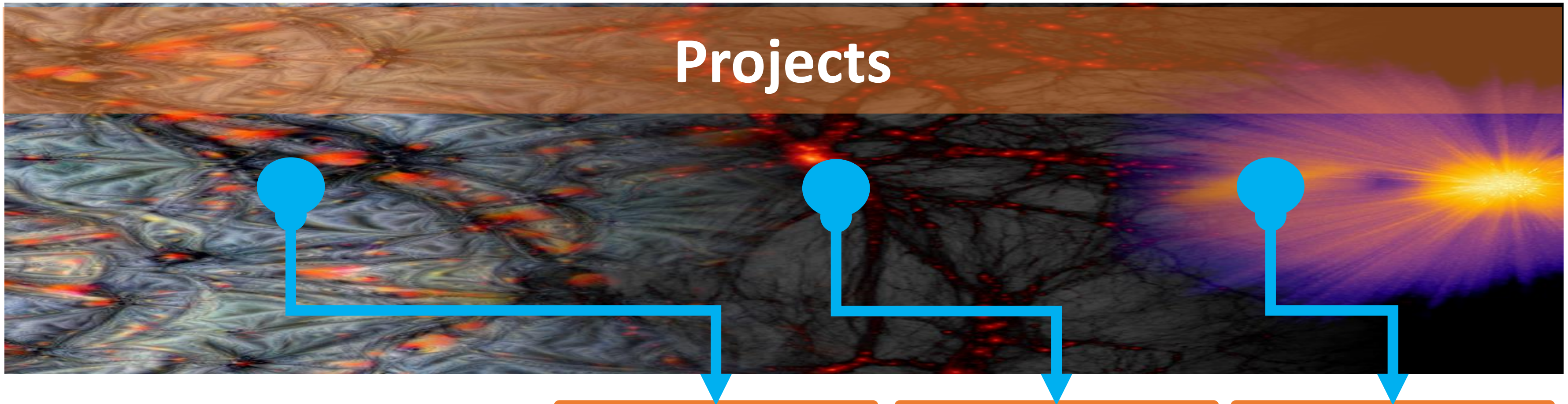
AIP Leibniz-Institut für Astrophysik Potsdam

Optimizing Scientific Workflows

Integrating REANA with PUNCH Infrastructure

Dr Arman Khalatyan /Analysis Facilities Workshop / 19. 06 2024

Research interests



HPC
BigDATA
ML
MPI
OpenMPI
Visualisation

HESTIA: High-resolution Environmental Simulations of The Immediate Area

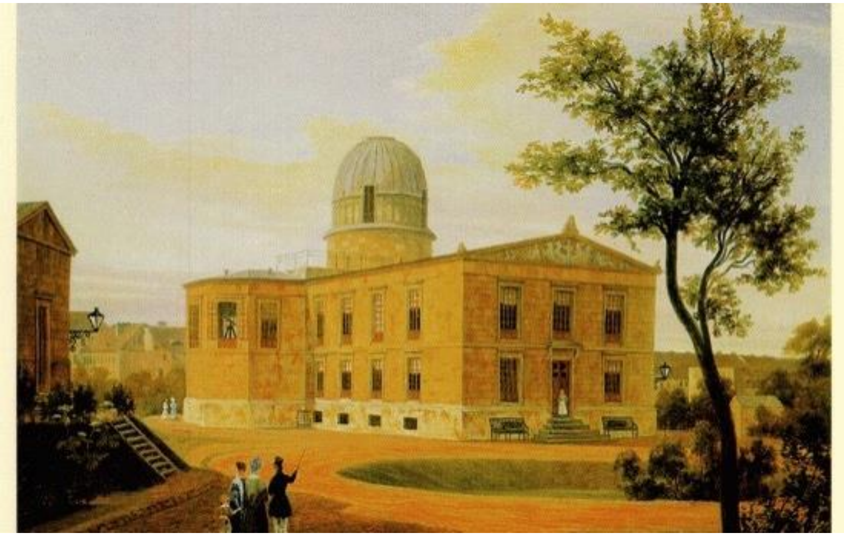
StarHorse: Photo-astrometric distances, extinctions, and astrophysical parameters for Gaia stars brighter than G = 18

colab.aip.de cloud.aip.de vr.aip.de

Nationale Forschungs-Daten Infrastruktur NFDI

PUNCH4NFDI

From Berlin to Babelsberg



The Berliner Sternwarte in Berlin-Dorotheenstadt, today Berlin-Mitte.

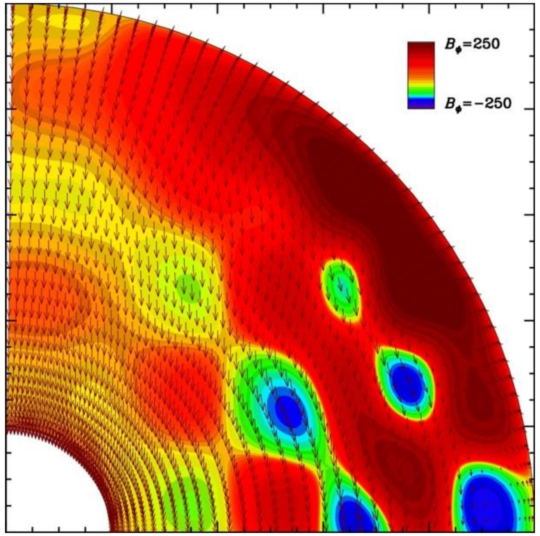
The new Sternwartengebäude in Babelsberg, built 1913.

The Berliner Sternwarte, founded in 1700, moved to 1913 Babelsberg because the growing city made scientific observations difficult. Light pollution and vibrations from traffic being the main reasons

Changing Names

From the Zentralinstitut für Astrophysik (1969) after the wall came down the Astrophysikalische Institut Potsdam (1992) was founded. In 2011 ithe AIP was renamed to **Leibniz-Institut für Astrophysik Potsdam (AIP),** to emphasize the membership with the Leibniz-Gemeinschaft.

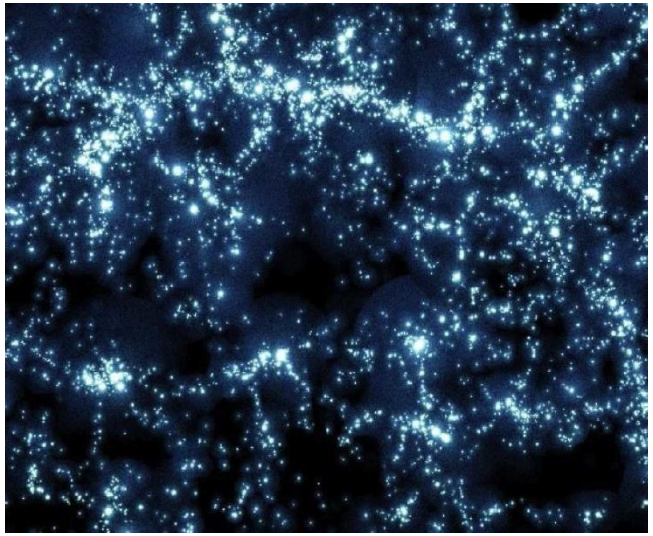
Research Areas



Cosmic Magnetic Fields

Research on solar, stellar and galactic magnetic fields and magnetohydrodynamic (MHD) mechanisms.

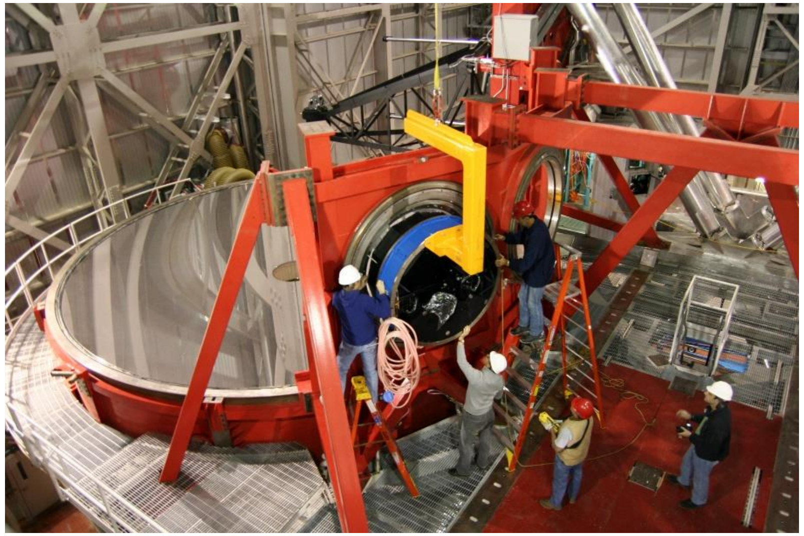
https://www.aip.de



Extragalactic Astrophysics

Active galaxies and quasars. Galactic archaeology and extragalactic research based on high resolution simulations.

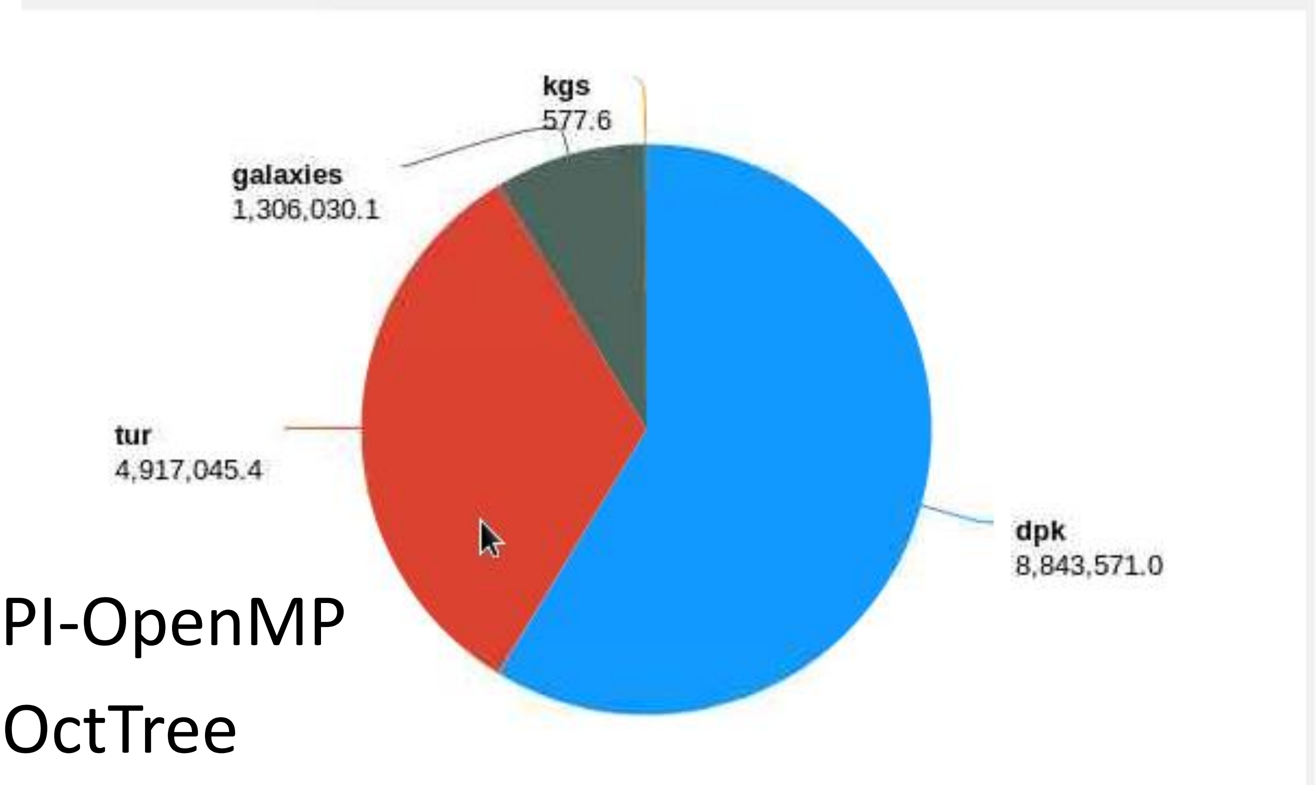
Research technology and infrastructure



„Development of Research technology and infrastructure“ ensures the scientific endeavors of AIP and also its participation in international astronomical projects. AIP has construction workshops and labs for instrumentation, especially with fiber optics, ex: AIP is lead institute of the **4MOST** project a fibre-fed spectroscopic survey facility on the VISTA telescope, op-2021)

Who is using most of the CPU time?

- Cosmology:
  - MHD+Gravity+Gasdynamics
  - Starformation, Cosmic Rays,BH…
- Magneto-hydrodynamics: MHD
- Data processing from telescopes



- Adaptive unstructured mesh, MPI-OpenMP
- Magneto-hydrodynamics: AMR-OctTree
- Data processing from telescopes: python, c, java, other

Languages

AIP/Clusters

- Python
- C/C++
- Fortran
- IDL
- Java
- Perl
- R
- ?Cuda?

Energy Efficiency across Programming Languages How Do Energy, Time, and Memory Relate?
Rui Pereira et al 2017
https://doi.org/10.1145/3136014.3136031

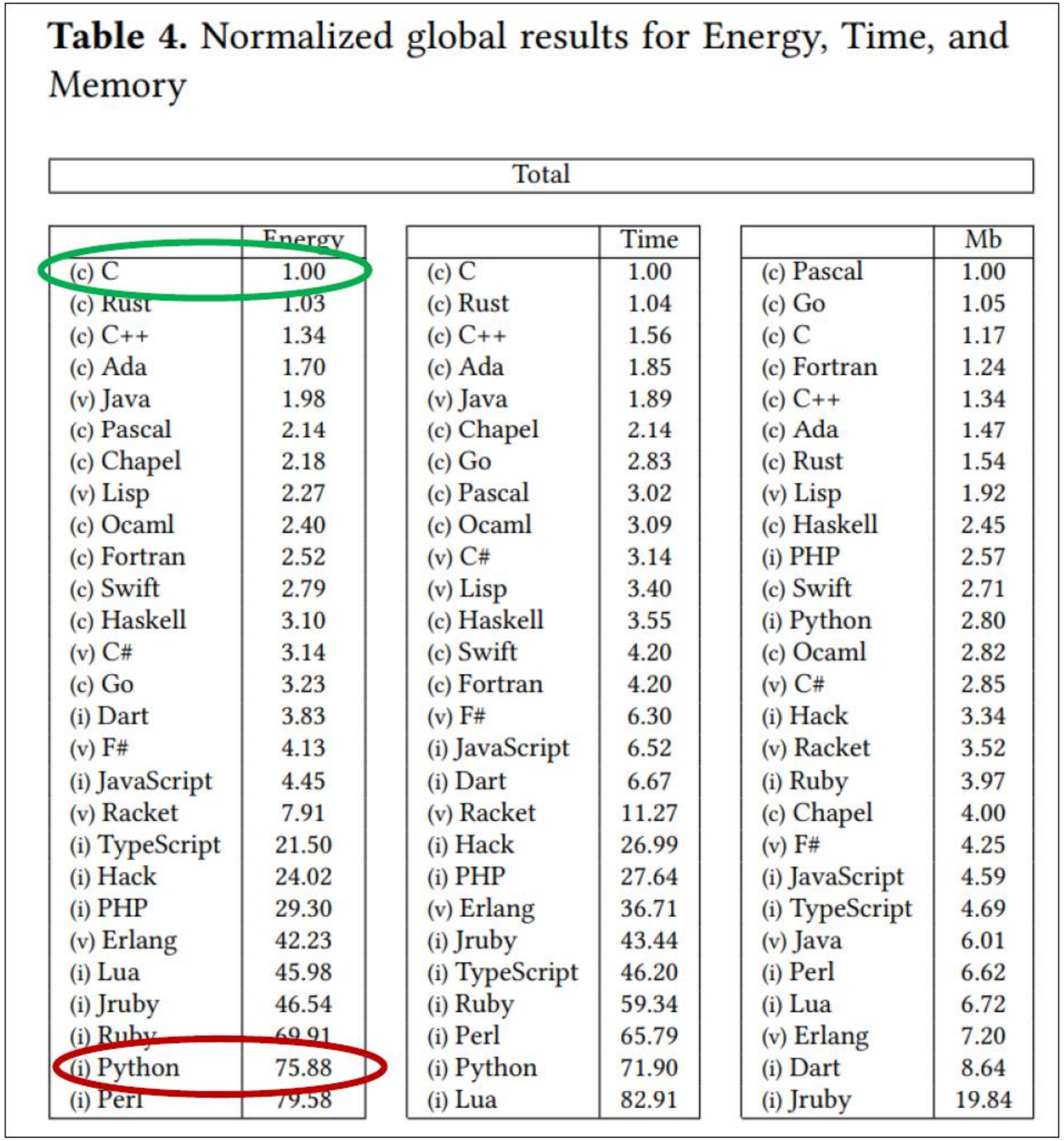


Table 4. Normalized global results for Energy, Time, and Memory

Total					
	Energy		Time		Mb
(c) C	1.00	(c) C	1.00	(c) Pascal	1.00
(c) Rust	1.03	(c) Rust	1.04	(c) Go	1.05
(c) C++	1.34	(c) C++	1.56	(c) C	1.17
(c) Ada	1.70	(c) Ada	1.85	(c) Fortran	1.24
(v) Java	1.98	(v) Java	1.89	(c) C++	1.34
(c) Pascal	2.14	(c) Chapel	2.14	(c) Ada	1.47
(c) Chapel	2.18	(c) Go	2.83	(c) Rust	1.54
(v) Lisp	2.27	(c) Pascal	3.02	(v) Lisp	1.92
(c) Ocaml	2.40	(c) Ocaml	3.09	(c) Haskell	2.45
(c) Fortran	2.52	(v) C#	3.14	(i) PHP	2.57
(c) Swift	2.79	(v) Lisp	3.40	(c) Swift	2.71
(c) Haskell	3.10	(c) Haskell	3.55	(i) Python	2.80
(v) C#	3.14	(c) Swift	4.20	(c) Ocaml	2.82
(c) Go	3.23	(c) Fortran	4.20	(v) C#	2.85
(i) Dart	3.83	(v) F#	6.30	(i) Hack	3.34
(v) F#	4.13	(i) JavaScript	6.52	(v) Racket	3.52
(i) JavaScript	4.45	(i) Dart	6.67	(i) Ruby	3.97
(v) Racket	7.91	(v) Racket	11.27	(c) Chapel	4.00
(i) TypeScript	21.50	(i) Hack	26.99	(v) F#	4.25
(i) Hack	24.02	(i) PHP	27.64	(i) JavaScript	4.59
(i) PHP	29.30	(v) Erlang	36.71	(i) TypeScript	4.69
(v) Erlang	42.23	(i) Jruby	43.44	(v) Java	6.01
(i) Lua	45.98	(i) TypeScript	46.20	(i) Perl	6.62
(i) Jruby	46.54	(i) Ruby	59.34	(i) Lua	6.72
(i) Ruby	69.91	(i) Perl	65.79	(v) Erlang	7.20
(i) Python	75.88	(i) Python	71.90	(i) Dart	8.64
(i) Perl	79.58	(i) Lua	82.91	(i) Jruby	19.84

Data Scales

- Cosmological simulations: >5PB(active)

- Observations(preserve/provide service):
  - GAIA: until now about 500TB, serving 200TB (DB) , soon >1PB (+2025)
  - Applaus: 200TB (photo plates archive)
  - 4MOST: GAIA+?
  - Pepsi: 250TB+
  - Stella- robotic telescope: 150TB+
- Data sizes:
  - 1k up to +32GB single files

Note:
for **5PB** one need to get also backup:

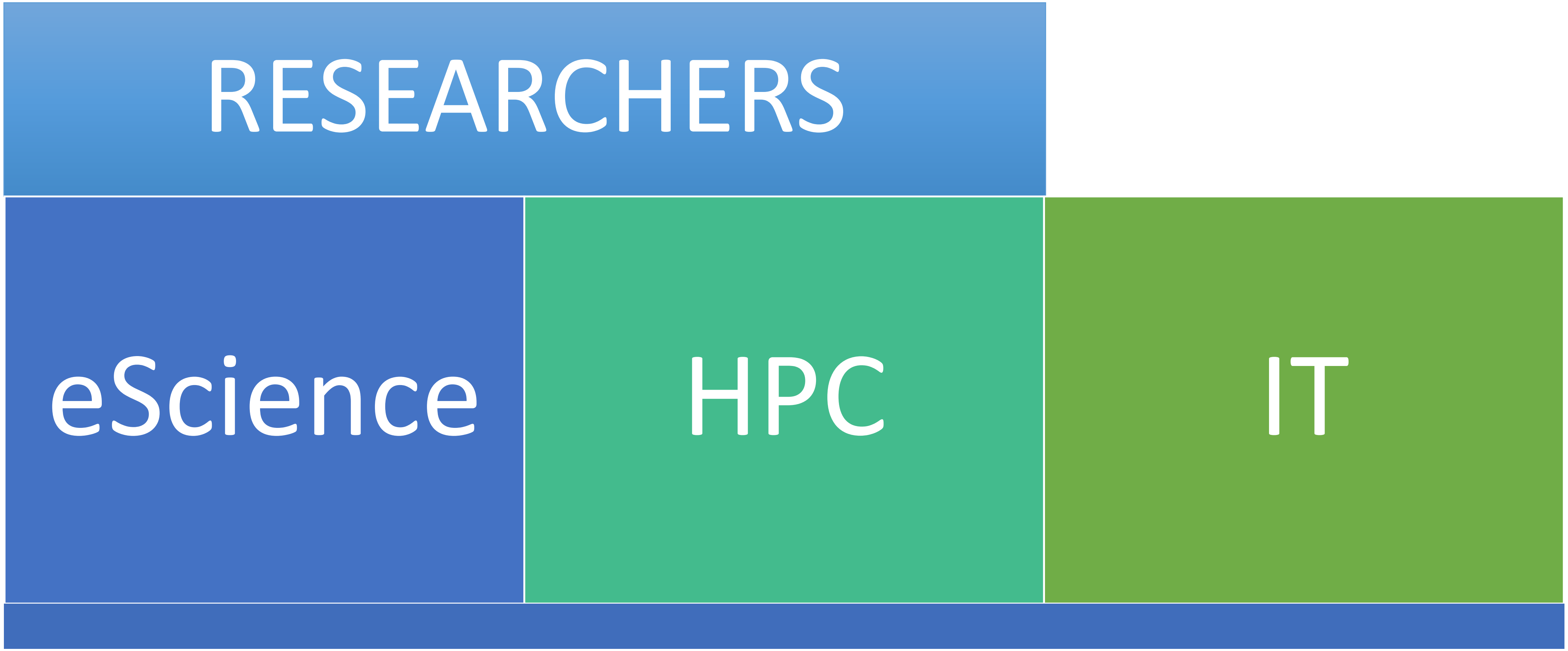
- 5PB-**300000**€ storage system
- **50000**€ For backup system
- power consumption: 500TB - 0.4kWh, **5y** to keep 5PB up and running is about: **160000**€ for power consumption

HPC admins: Users software

Jobqueue

Interactive

eScience+HPC+IT



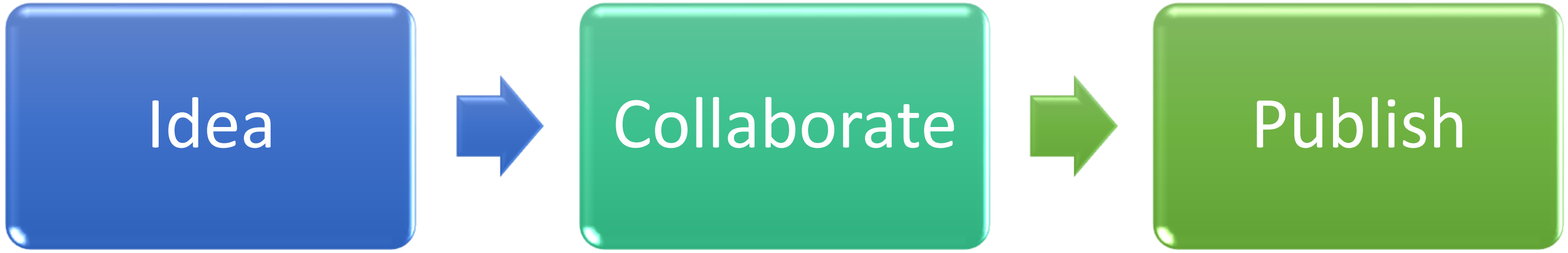
RESEARCHERS

eScience

HPC

IT

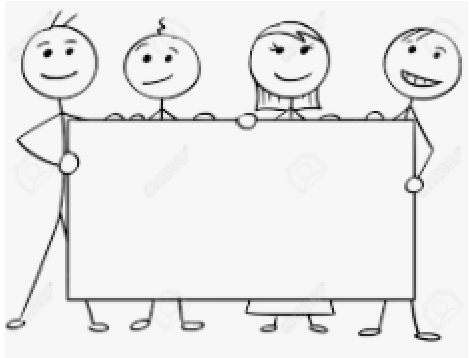
Scientific life (top to down)

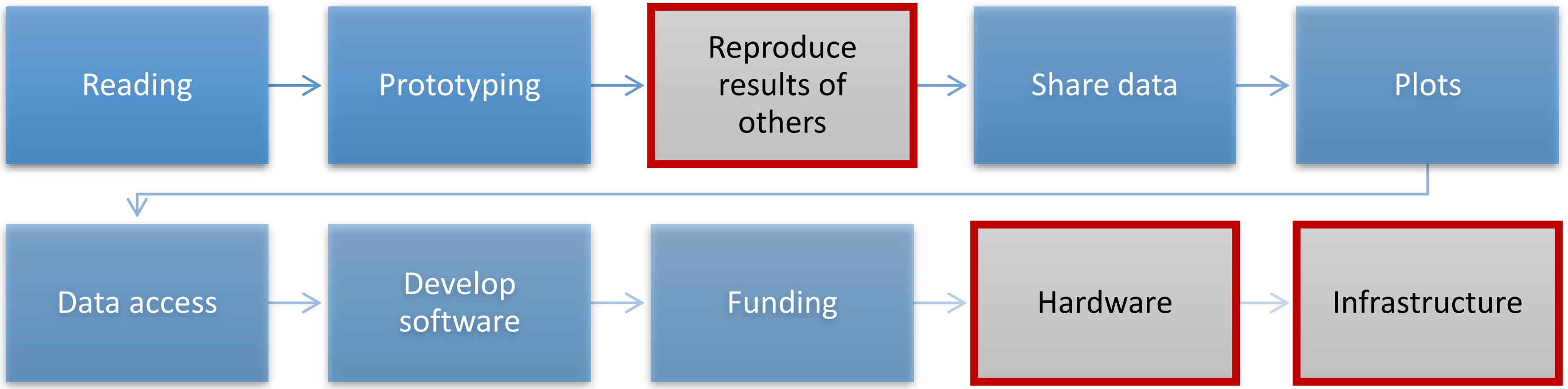


Idea

Collaborate

Publish





Reading

Prototyping

Reproduce results of others

Share data

Plots

Data access

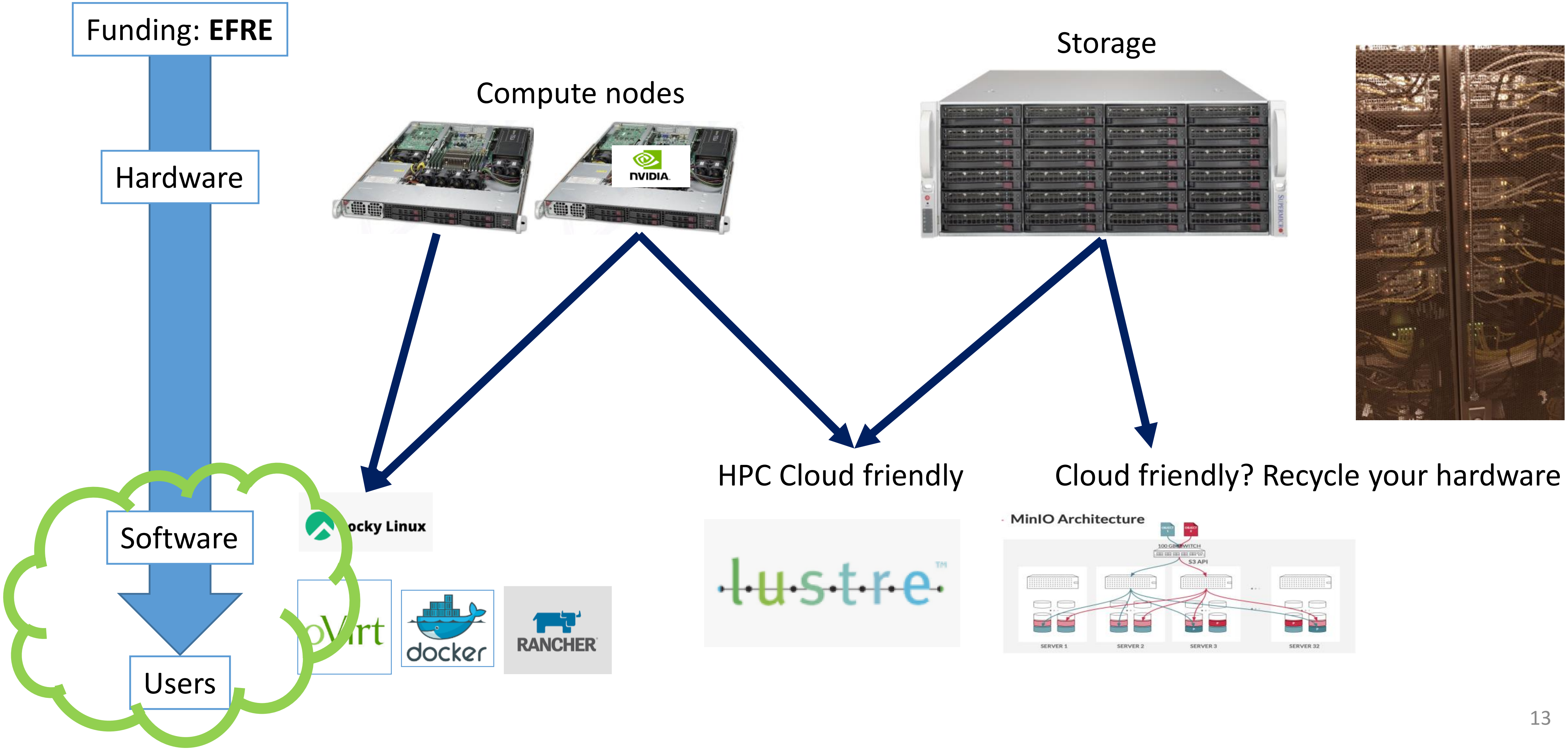
Develop software

Funding

Hardware

Infrastructure

Infrastructure (down to top)



Funding: **EFRE**

Hardware

Software

Users

Compute nodes

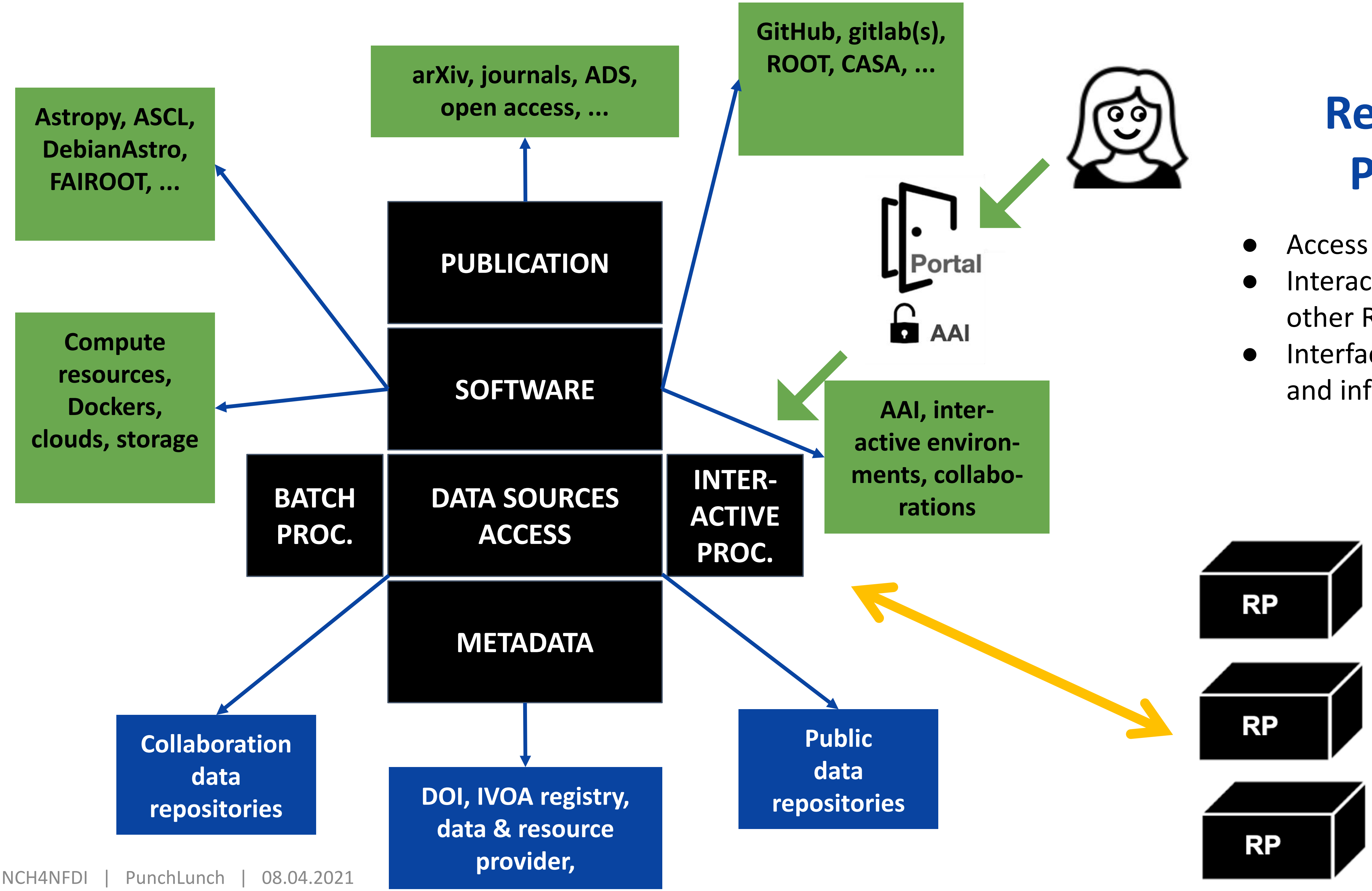
Storage

HPC Cloud friendly

Cloud friendly? Recycle your hardware

Digital Research Product

- Access via portal
- Interaction with other RPs
- Interfaced to tools and infrastructures



PUBLICATION

SOFTWARE

BATCH PROC.

DATA SOURCES ACCESS

INTER-ACTIVE PROC.

METADATA

Astropy, ASCL, DebianAstro, FAIROOT, ...

arXiv, journals, ADS, open access, ...

GitHub, gitlab(s), ROOT, CASA, ...

Compute resources, Dockers, clouds, storage

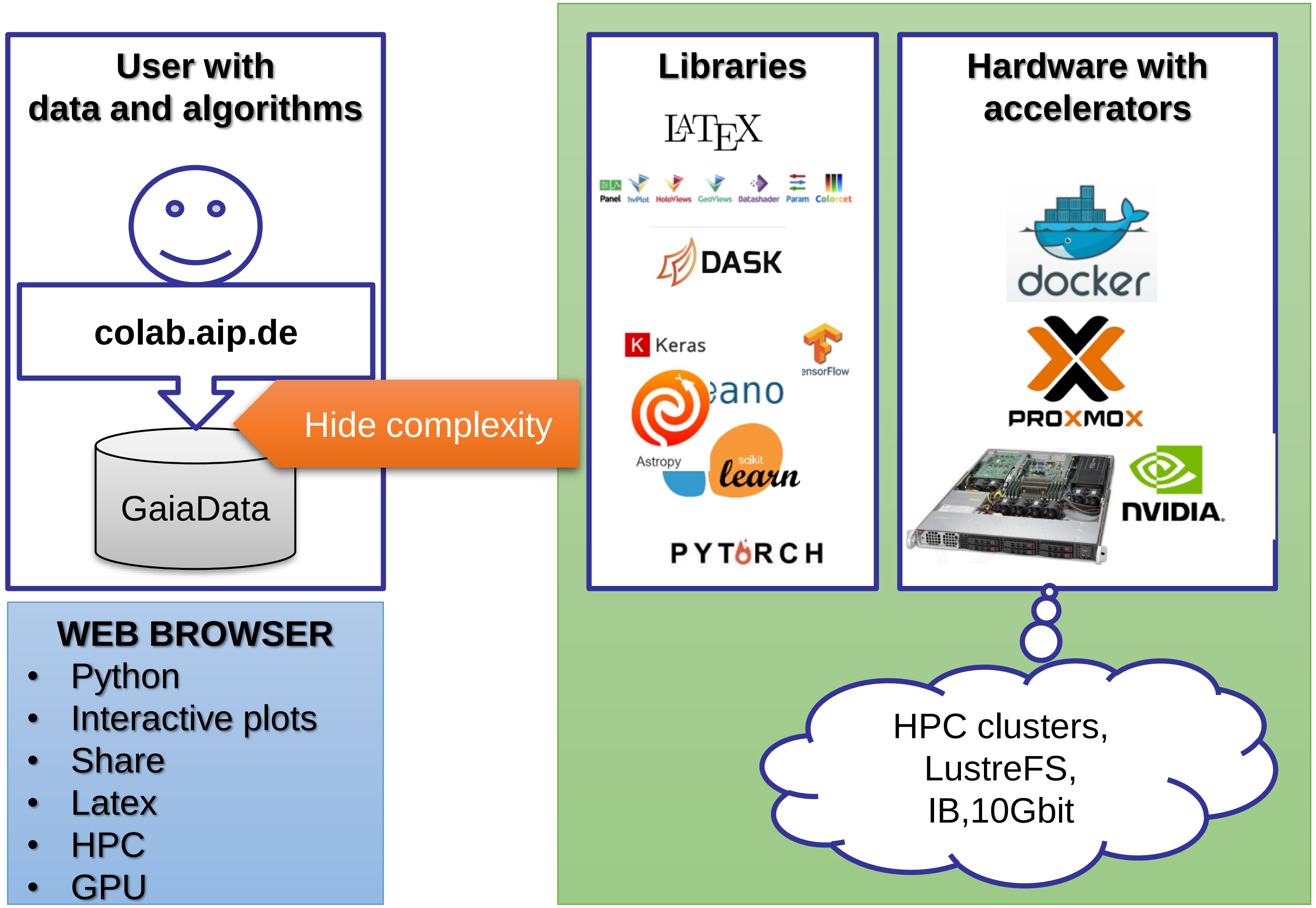
AAI, interactive environments, collaborations

Collaboration data repositories

DOI, IVOA registry, data & resource provider,

Public data repositories

What we did at AIP before 2023?



User with data and algorithms

colab.aip.de

GaiaData

Hide complexity

WEB BROWSER
- Python
- Interactive plots
- Share
- Latex
- HPC
- GPU

Libraries

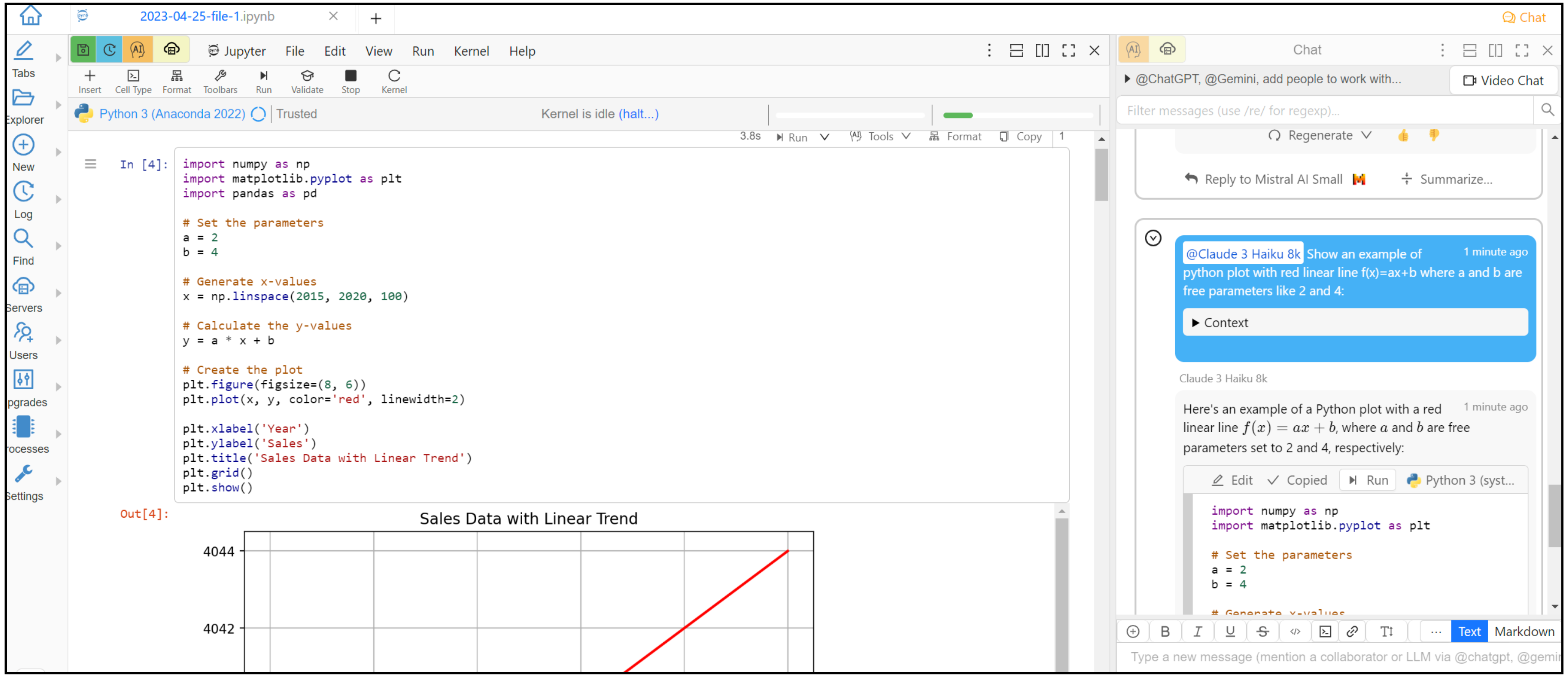
Hardware with accelerators

HPC clusters, LustreFS, IB,10Gbit

Since 2016,
single docker

COCALC
Collaborative Calculation and Data Science
by Sagemath, Inc.

CoCalc Integrated Tool: LLM as an assistant



```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Set the parameters
a = 2
b = 4

# Generate x-values
x = np.linspace(2015, 2020, 100)

# Calculate the y-values
y = a * x + b

# Create the plot
plt.figure(figsize=(8, 6))
plt.plot(x, y, color='red', linewidth=2)

plt.xlabel('Year')
plt.ylabel('Sales')
plt.title('Sales Data with Linear Trend')
plt.grid()
plt.show()
```

@ChatGPT, @Gemini, add people to work with...

@Claude 3 Haiku 8k Show an example of python plot with red linear line f(x)=ax+b where a and b are free parameters like 2 and 4: — 1 minute ago

Claude 3 Haiku 8k

Here's an example of a Python plot with a red linear line $f(x) = ax + b$, where a and b are free parameters set to 2 and 4, respectively: — 1 minute ago

```
import numpy as np
import matplotlib.pyplot as plt

# Set the parameters
a = 2
b = 4

# Generate x-values
```

Type a new message (mention a collaborator or LLM via @chatgpt, @gemi

- ai.aip.de hosting local LLM at AIP
- Based on **ollama**
- **And more…**

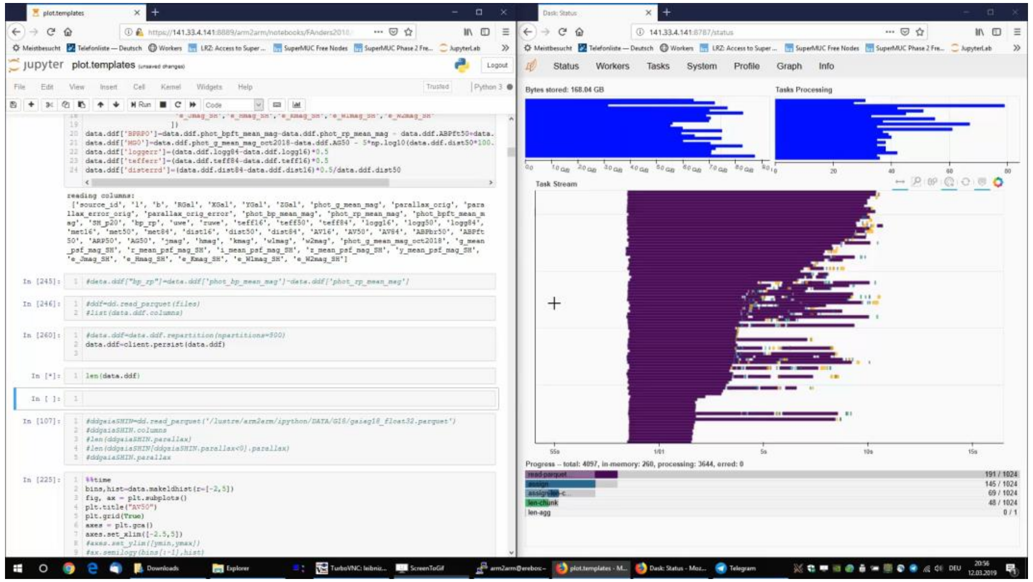
- colab.aip.de: 250 users, over 900+ projects over 6 years
- Ai.aip.de: over 75 users in 3 weeks

The whole complexity is obscured from the users

Users want the all tools in one place
- Data+LaTex+Code
- Collaborators to share
- Article versions
- Cluster access
- Easy publishing for the demo notebooks

Possible solution:
- dask+Kubernetes
- CoCalc project

CoCalc is a web-based cloud computing and course management platform for computational mathematics. Part of the Sage project, it supports editing of Sage worksheets, LaTeX documents and Jupyter notebooks.



```
                          'e_Jmag_SH','e_Hmag_SH','e_Kmag_SH','e_W1mag_SH','e_W2mag_SH'
                         ])
data.ddf['BPRP0']=data.ddf.phot_bpft_mean_mag-data.ddf.phot_rp_mean_mag - data.ddf.ABPft50+data.
data.ddf['MG0']=data.ddf.phot_g_mean_mag_oct2018-data.ddf.AG50 - 5*np.log10(data.ddf.dist50*100.
data.ddf['loggerr']=(data.ddf.logg84-data.ddf.logg16)*0.5
data.ddf['tefferr']=(data.ddf.teff84-data.ddf.teff16)*0.5
data.ddf['disterrd']=(data.ddf.dist84-data.ddf.dist16)*0.5/data.ddf.dist50
```

```
reading columns:
 ['source_id', 'l', 'b', 'RGal', 'XGal', 'YGal', 'ZGal', 'phot_g_mean_mag', 'parallax_orig', 'para
llax_error_orig', 'parallax_orig_error', 'phot_bp_mean_mag', 'phot_rp_mean_mag', 'phot_bpft_mean_m
ag', 'SH_p20', 'bp_rp', 'uwe', 'ruwe', 'teff16', 'teff50', 'teff84', 'logg16', 'logg50', 'logg84',
'met16', 'met50', 'met84', 'dist16', 'dist50', 'dist84', 'AV16', 'AV50', 'AV84', 'ABPbr50', 'ABPft
50', 'ARP50', 'AG50', 'jmag', 'hmag', 'kmag', 'w1mag', 'w2mag', 'phot_g_mean_mag_oct2018', 'g_mean
_psf_mag_SH', 'r_mean_psf_mag_SH', 'i_mean_psf_mag_SH', 'z_mean_psf_mag_SH', 'y_mean_psf_mag_SH',
'e_Jmag_SH', 'e_Hmag_SH', 'e_Kmag_SH', 'e_W1mag_SH', 'e_W2mag_SH']
```

```
In [245]: #data.ddf["bp_rp"]=data.ddf['phot_bp_mean_mag']-data.ddf['phot_rp_mean_mag']
```

```
In [246]: #ddf=dd.read_parquet(files)
          #list(data.ddf.columns)
```

```
In [260]: #data.ddf=data.ddf.repartition(npartitions=500)
          data.ddf=client.persist(data.ddf)
```

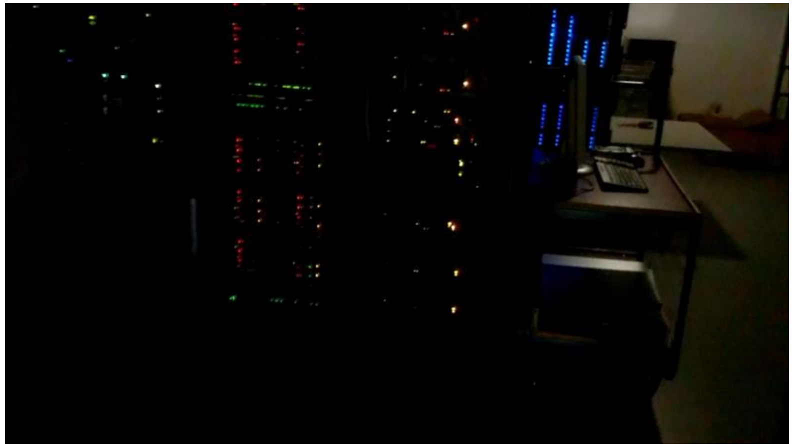
```
In [*]: len(data.ddf)
```

```
In [ ]:
```

```
In [107]: #ddgaiaSHIN=dd.read_parquet('/lustre/arm2arm/ipython/DATA/G18/gaiag18_float32.parquet')
          #ddgaiaSHIN.columns
          #len(ddgaiaSHIN.parallax)
          #len(ddgaiaSHIN[ddgaiaSHIN.parallax<0].parallax)
          #ddgaiaSHIN.parallax
```

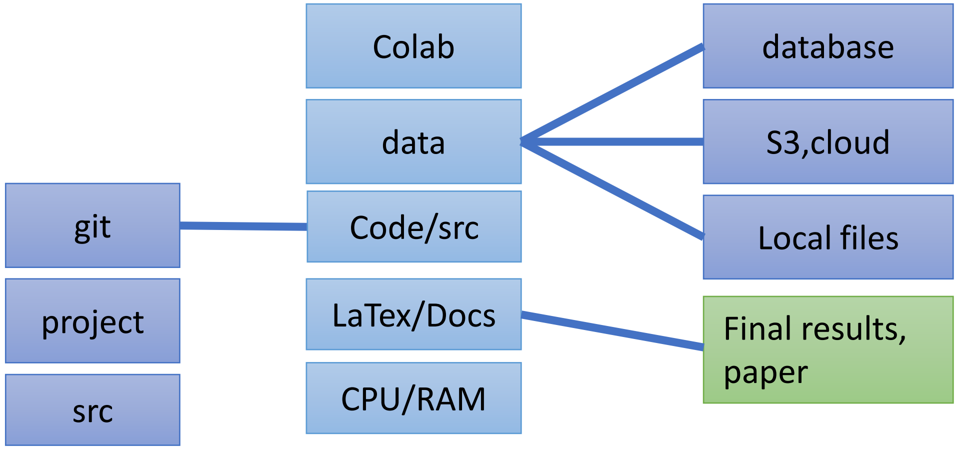
```
In [225]: %%time
          bins,hist=data.make1dhist(r=[-2,5])
          fig, ax = plt.subplots()
          plt.title("AV50")
          plt.grid(True)
          axes = plt.gca()
          axes.set_xlim([-2.5,5])
          #axes.set_ylim([ymin,ymax])
          #ax.semilogy(bins[:-1],hist)
```

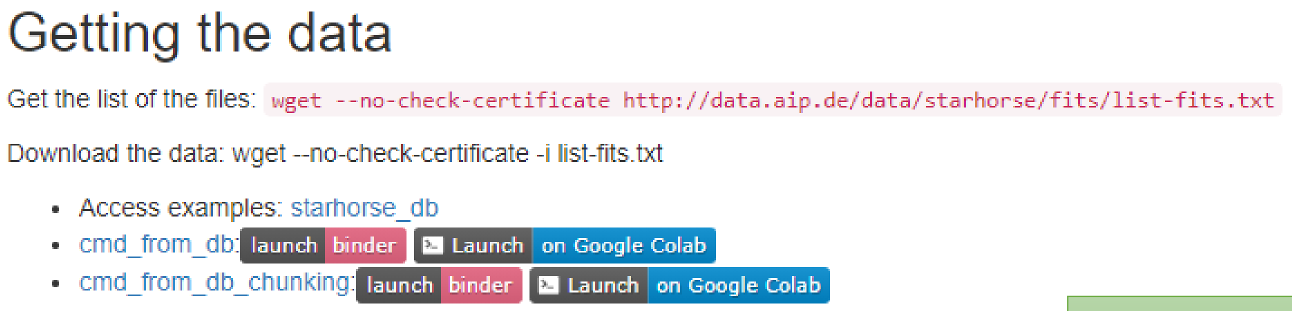
```
Progress -- total: 4097, in-memory: 260, processing: 3644, erred: 0
read-parquet...   191 / 1024
assign            145 / 1024
assign-len-c...    69 / 1024
len-chunk          48 / 1024
len-agg             0 / 1
```



Global Workflow of StarHorse team

A Bayesian code to estimate the photo-astrometric distances, extinctions, and astrophysical parameters for Gaia DR2 stars
F.Anders et al. (2019)





Getting the data

Get the list of the files: `wget --no-check-certificate http://data.aip.de/data/starhorse/fits/list-fits.txt`

Download the data: wget --no-check-certificate -i list-fits.txt

- Access examples: starhorse_db
- cmd_from_db:
- cmd_from_db_chunking:

https://data.aip.de/projects/starhorse2019.html

6 weeks 3000 cores get
400 000 000 Stellar parameters

What we learn from notebooks+jupyter?

- No versioning (even py codes are not versioned)
- No git (it is somehow possible but no one does this)
- No share
- No modularity
- 2-3 years cant run, for got the parameters in the cell.
- Astronomers during prototyping are writing terrible codes.

What about kubernetes?

- in Astrophysics infrastructure we are still in the same stage as “**Docker Inc.” was in 2014**.

Why?

- It was complex
- Rapid development in the Industry
- No LTS

Situation is matured in 2021:

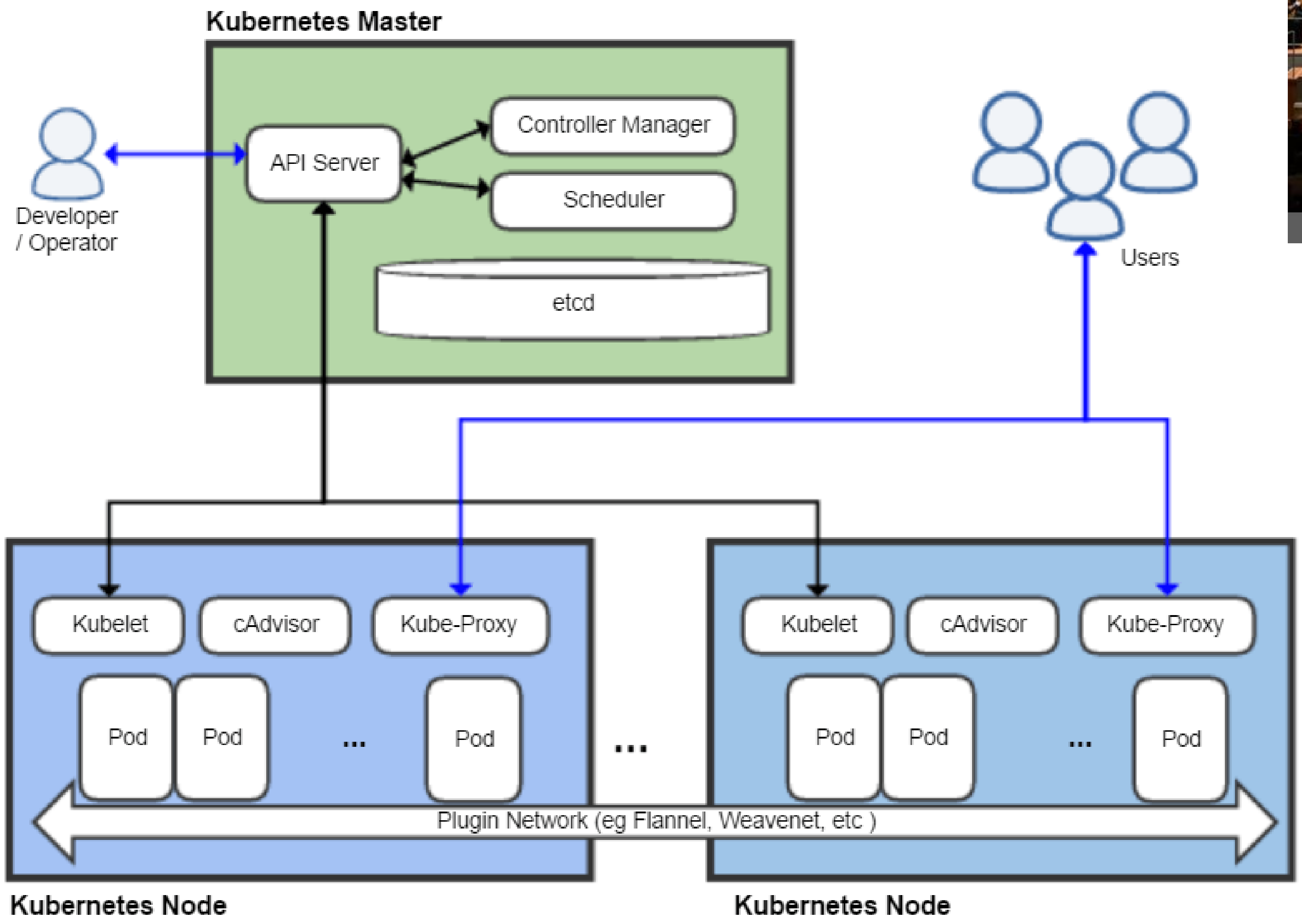
- Because of https://www.cncf.io/

CLOUD NATIVE
COMPUTING FOUNDATION

- We are ready to adopt some Infa from industry into to scientific life

USERS ARE WRITING BASH SCRIPTS TO ORCHESTRATE THEIR CONTAINERS?!
PERL SCRIPTS!
SWARM!
MAYBE CHECK THAT THING GOOGLE HAS BEEN RUNNING IN PROD FOR A DECADE?
KUBERNETES
docker

Kubernetes



Berliner Philharmoniker

Are we special?
Users images are so huge they are filling local host disks

Microservices: Reproducible science

Use Cases at AIP

- **Colab.aip.de**
  - Quotas
  - Project isolation
- **Data analysis pipelines with versioning**
  - Reproducible science
  - Pipeline versioning
    - GaiaDR1,2,3
    - RAVEDR1-6
    - StarHorse-18,19,20
- **Publish papers with interactive plots**
  - like binder
  - Example: distill.pub by google
- **Dynamically Scalable webpages**
- **gitlabs @ aip:** CI integration

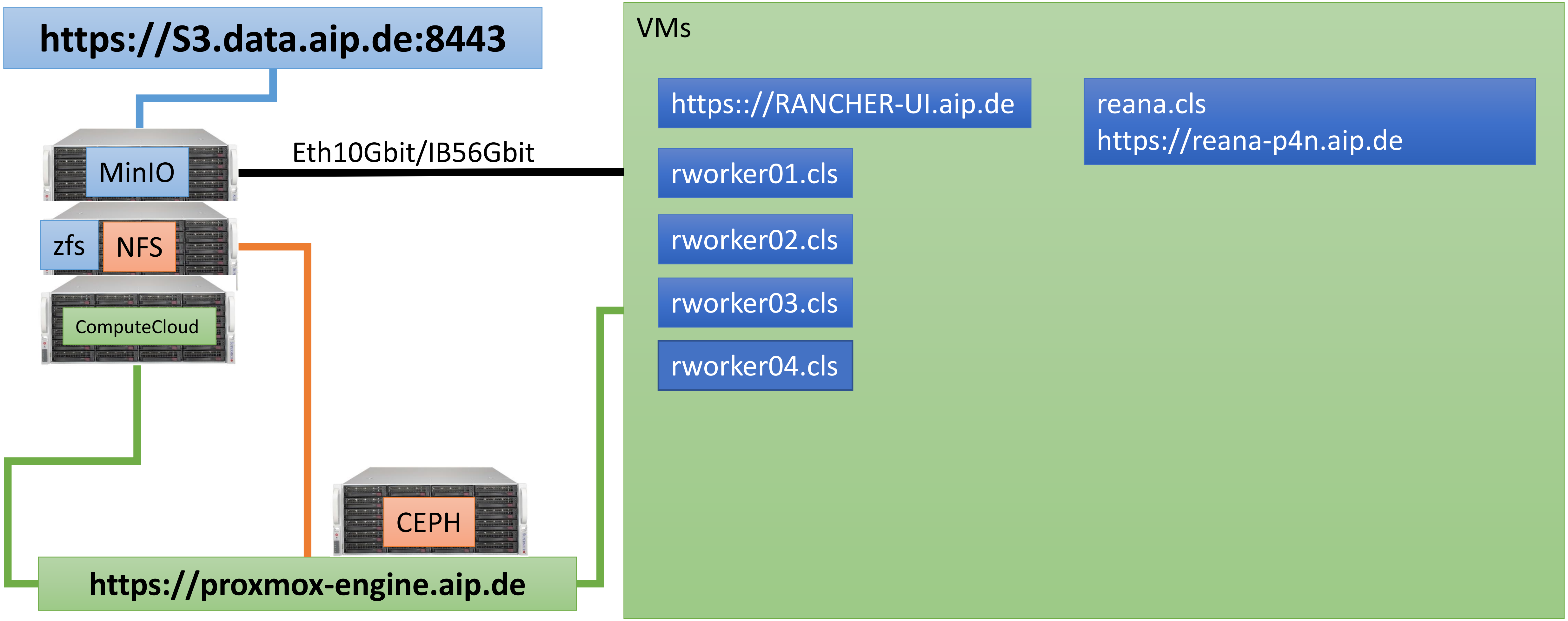
Pros:

- ***Direct GIT integration***
- ***Scalability***
- Modularity
- Distributed development
- Integration
- Save resources/power

Concerns:

- Complexity
- Design
- ***Testing, debugging***
- ***Inter-service call latency***

REANA: Infra



https://S3.data.aip.de:8443

MinIO

zfs

NFS

ComputeCloud

Eth10Gbit/IB56Gbit

CEPH

https://proxmox-engine.aip.de

VMs

https:://RANCHER-UI.aip.de

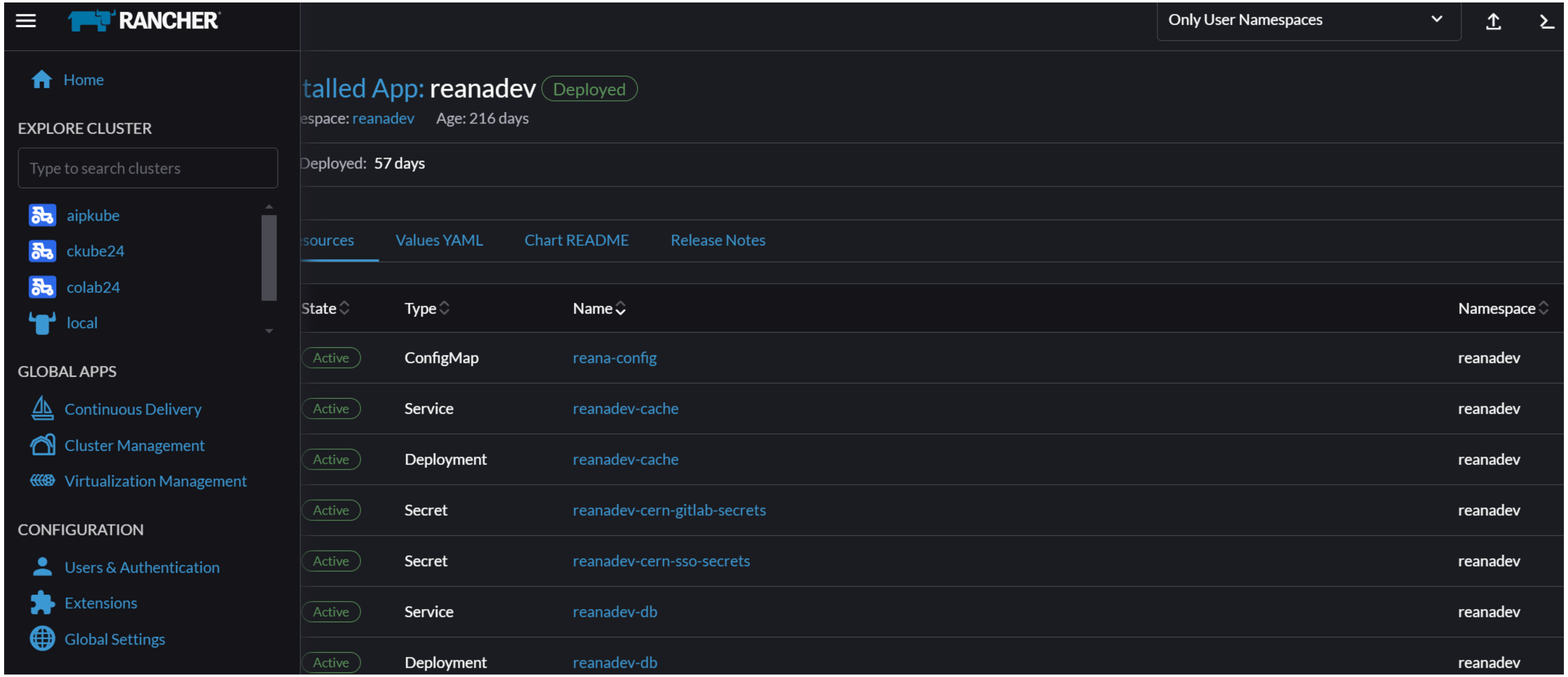
rworker01.cls

rworker02.cls

rworker03.cls

rworker04.cls

reana.cls
https://reana-p4n.aip.de



talled App: reanadev Deployed

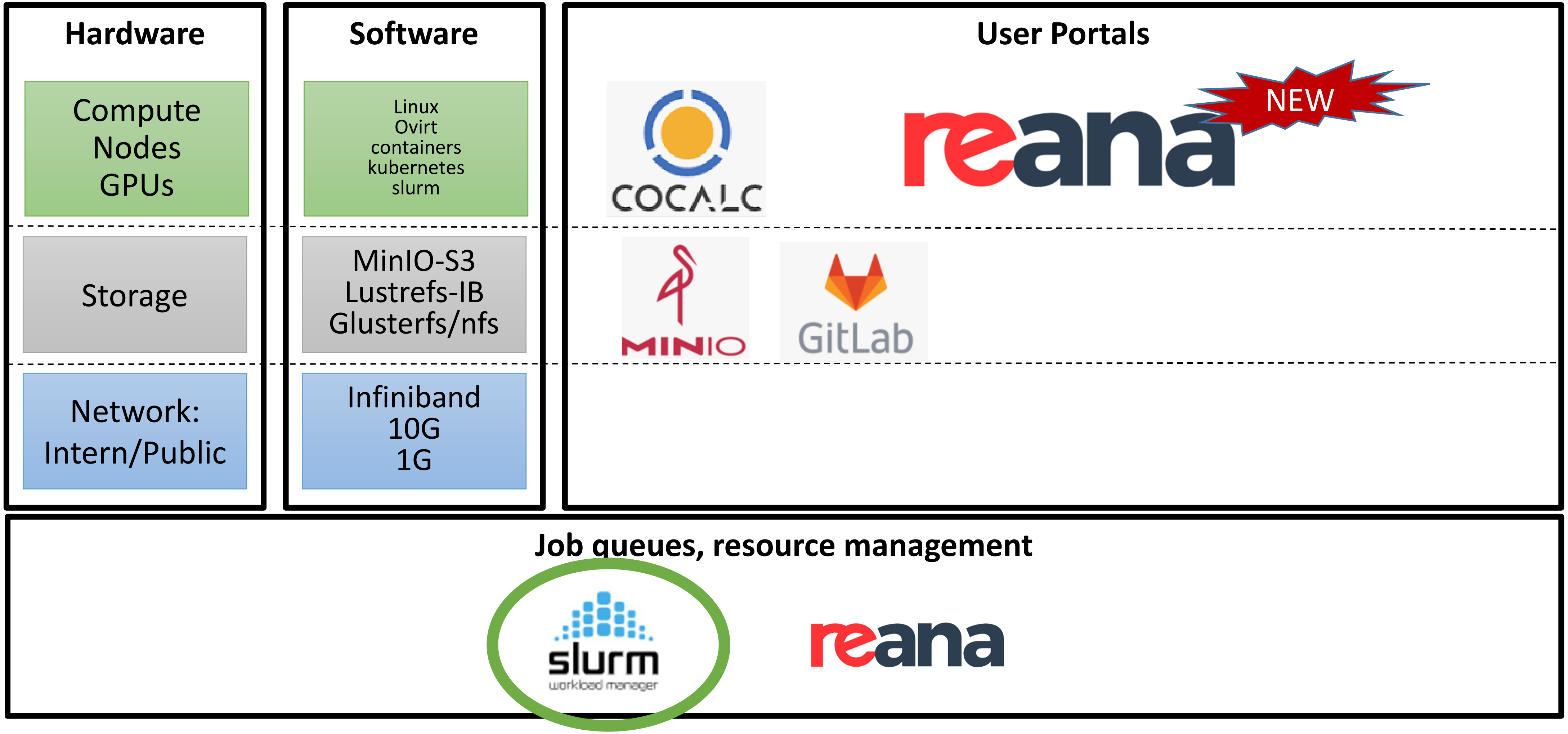
espace: reanadev Age: 216 days

Deployed: 57 days

State	Type	Name	Namespace
Active	ConfigMap	reana-config	reanadev
Active	Service	reanadev-cache	reanadev
Active	Deployment	reanadev-cache	reanadev
Active	Secret	reanadev-cern-gitlab-secrets	reanadev
Active	Secret	reanadev-cern-sso-secrets	reanadev
Active	Service	reanadev-db	reanadev
Active	Deployment	reanadev-db	reanadev

```
helm install reanadev24  reanahub/reana  --create-namespace -n reanadev24  -f values.yaml
```

SAAS,IAAS and PAAS



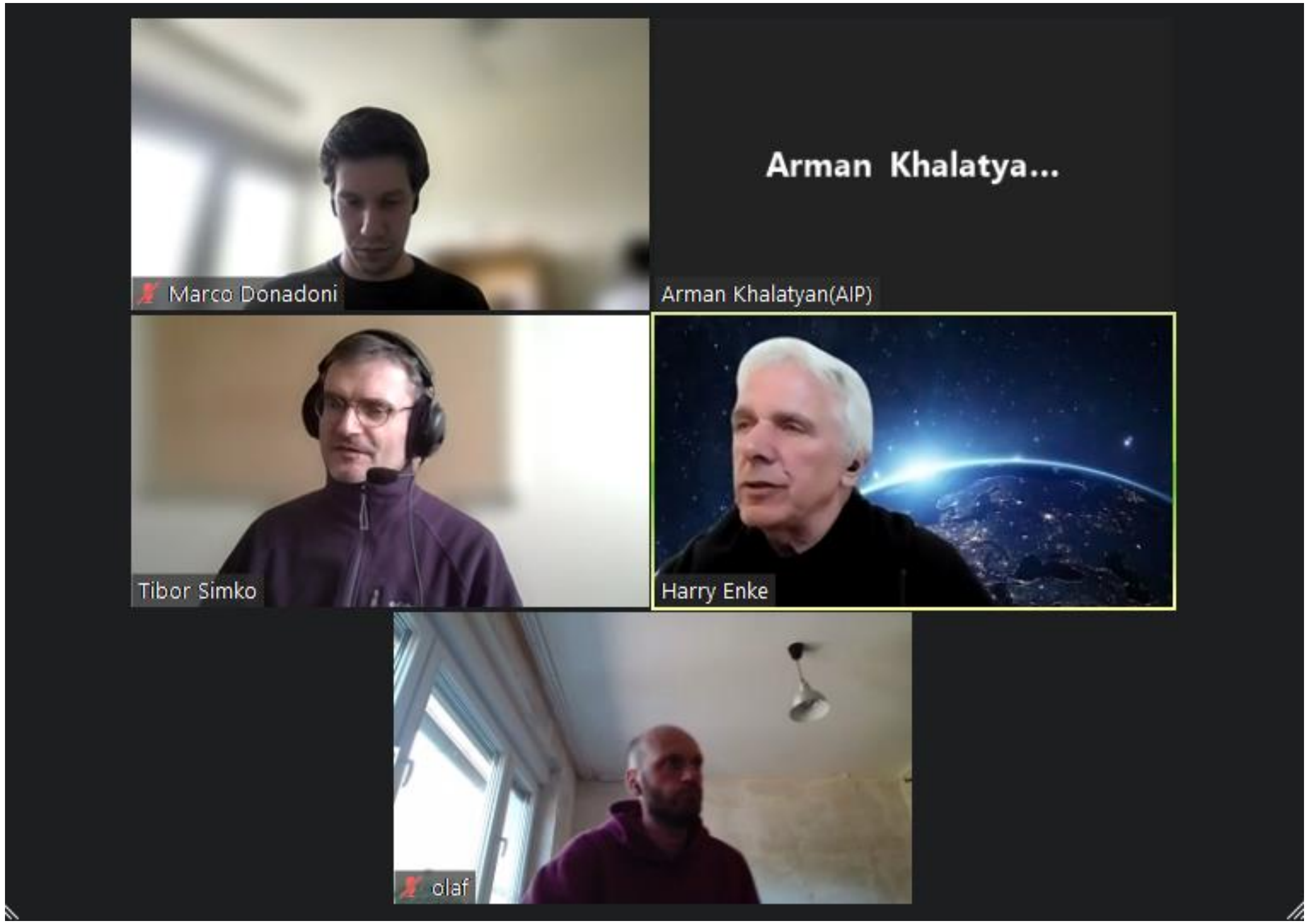
Hardware	Software	User Portals
Compute Nodes GPUs	Linux Ovirt containers kubernetes slurm	COCALC reana NEW
Storage	MinIO-S3 Lustrefs-IB Glusterfs/nfs	MINIO GitLab
Network: Intern/Public	Infiniband 10G 1G	

Job queues, resource management

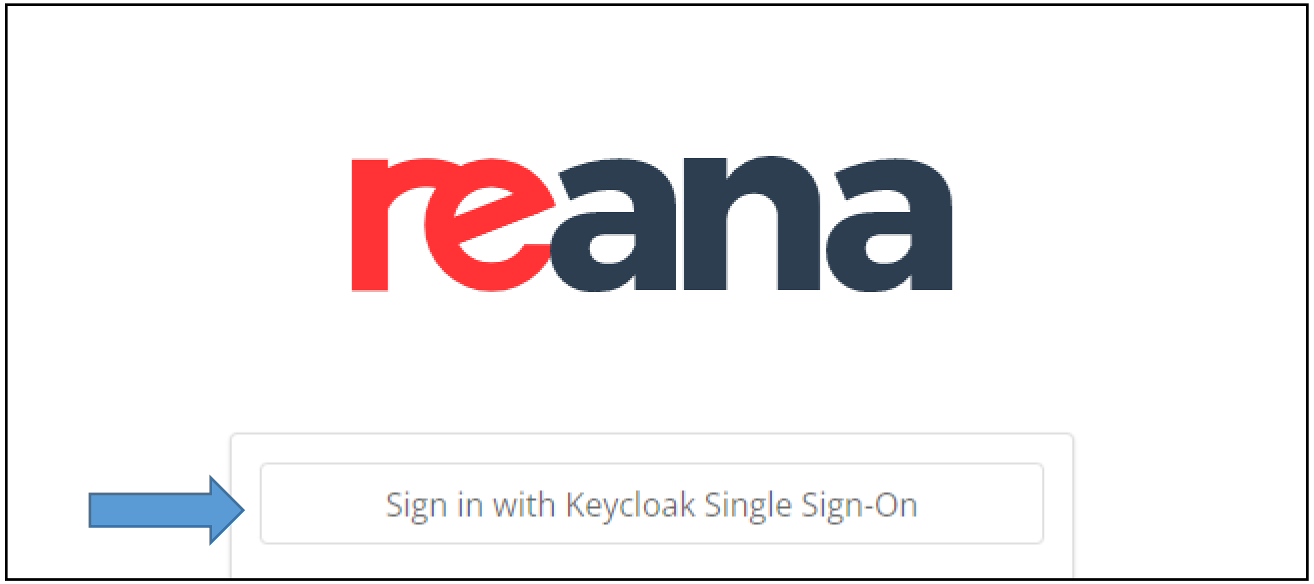
slurm workload manager

reana

REANA and AIP discussion rounds



REANA: in Action

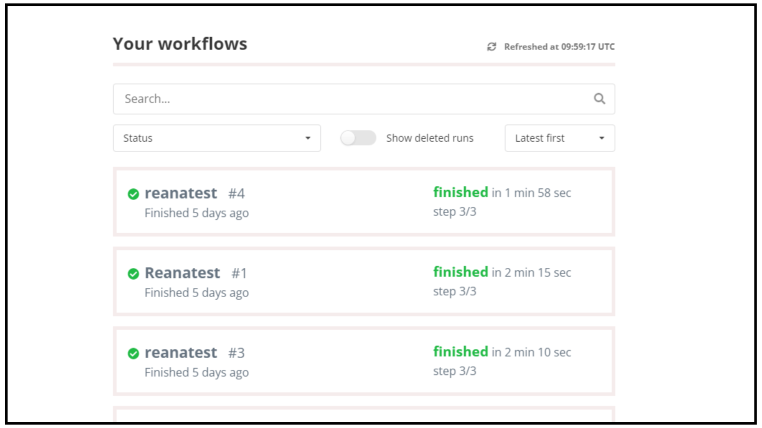


https://reana-p4n.aip.de/

Use **reana-client** from terminal

Connect to gitlab

https://gitlab-p4n.aip.de/arm2arm/reanatest

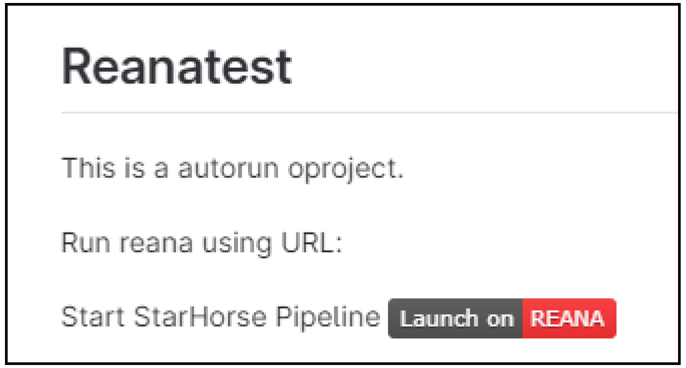


reanatest #4
Finished 5 days ago
finished in 1 min 58 sec
step 3/3

Reanatest #1
Finished 5 days ago
finished in 2 min 15 sec
step 3/3

reanatest #3
Finished 5 days ago
finished in 2 min 10 sec
step 3/3

Launch from URL



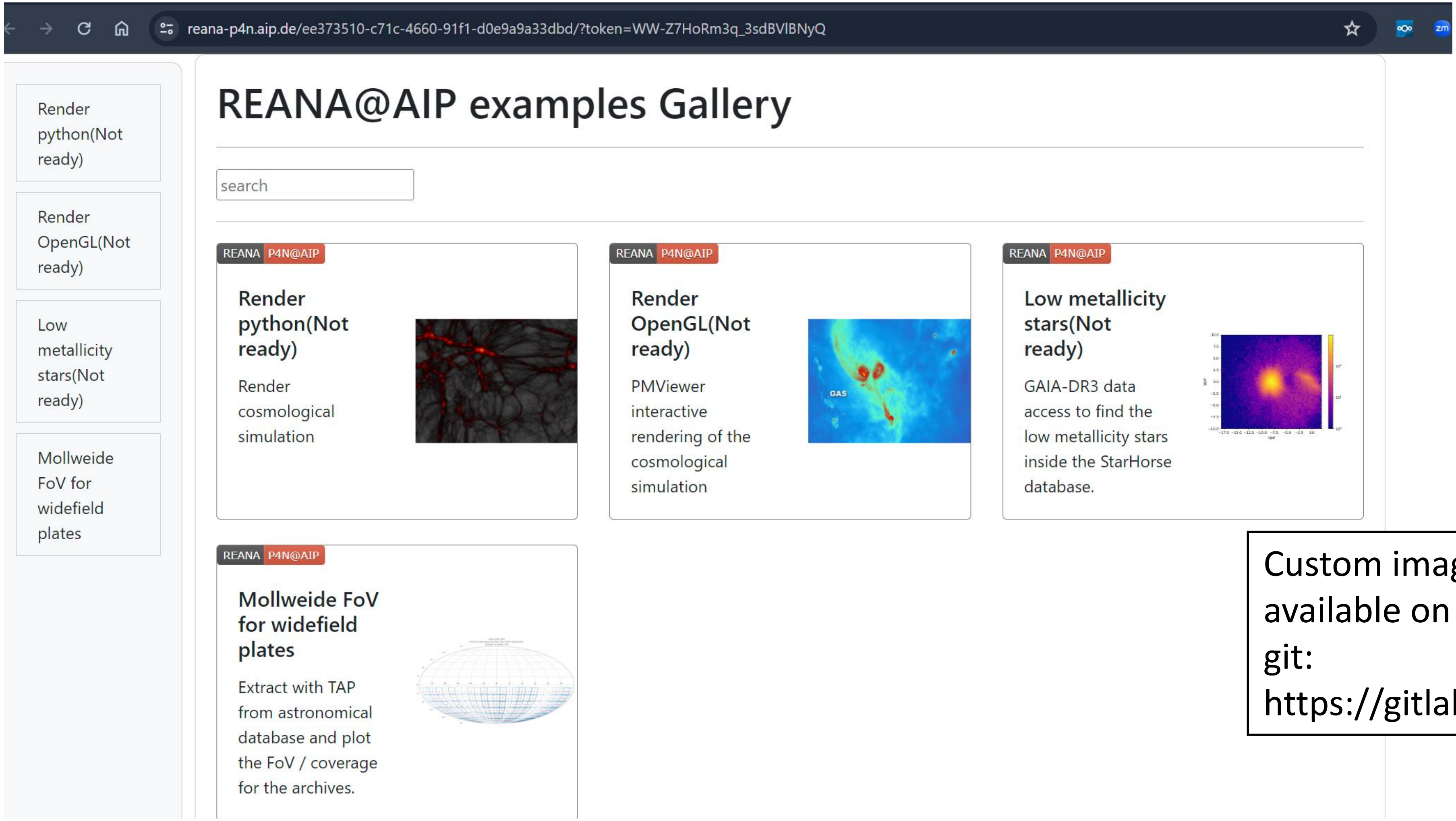
Reanatest

This is a autorun oproject.

Run reana using URL:

Start StarHorse Pipeline

REANA hosting arbitrary webpage



reana-p4n.aip.de/ee373510-c71c-4660-91f1-d0e9a9a33dbd/?token=WW-Z7HoRm3q_3sdBVlBNyQ

- Render python(Not ready)
- Render OpenGL(Not ready)
- Low metallicity stars(Not ready)
- Mollweide FoV for widefield plates

REANA@AIP examples Gallery

search

REANA P4N@AIP

Render python(Not ready)

Render cosmological simulation

REANA P4N@AIP

Render OpenGL(Not ready)

PMViewer interactive rendering of the cosmological simulation

REANA P4N@AIP

Low metallicity stars(Not ready)

GAIA-DR3 data access to find the low metallicity stars inside the StarHorse database.

REANA P4N@AIP

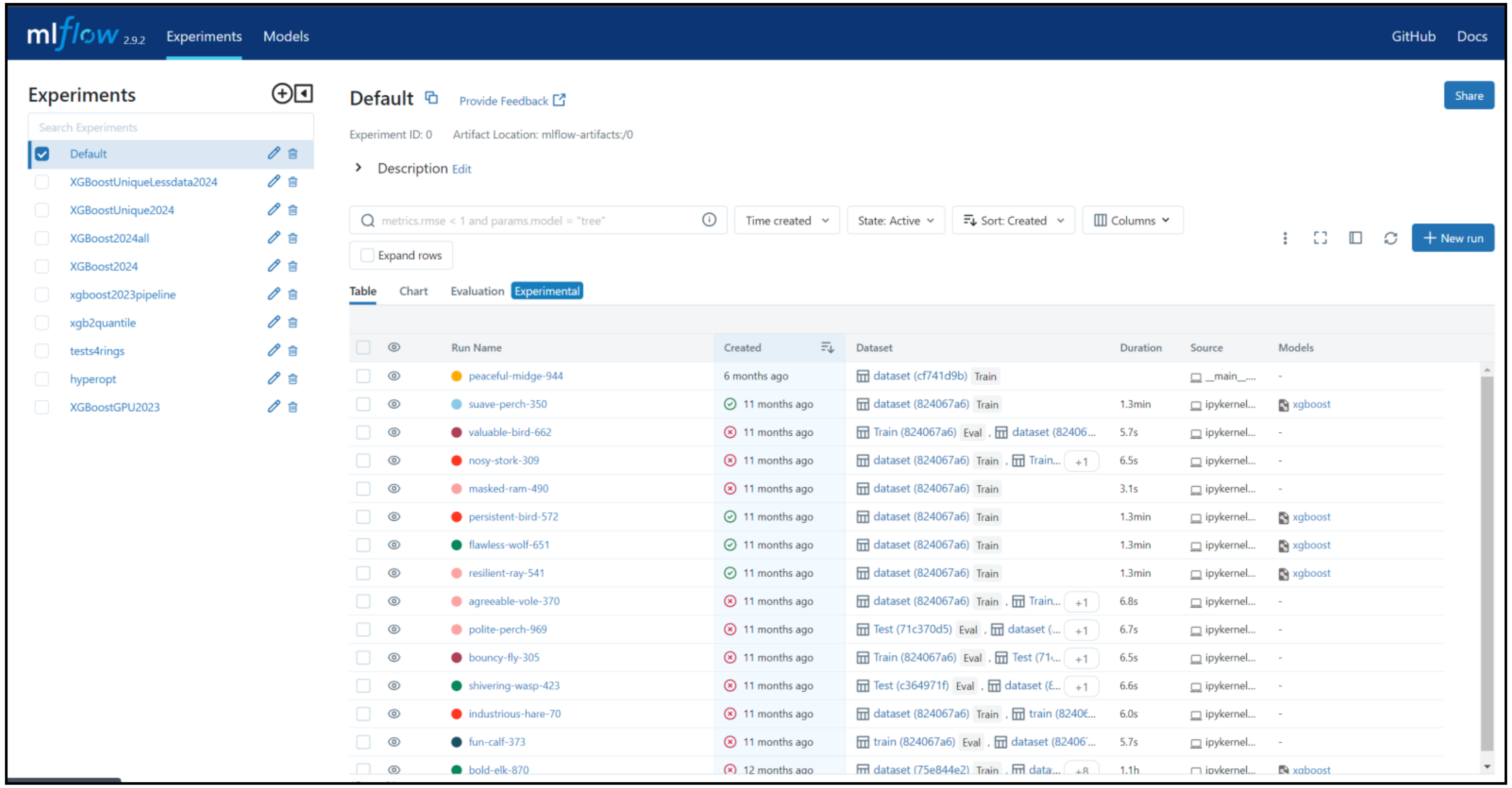
Mollweide FoV for widefield plates

Extract with TAP from astronomical database and plot the FoV / coverage for the archives.

Custom images are available on NFDI4PUNCH git: https://gitlab-p4n.aip.de/

reana-jailbreak?

MLFlow: as a ML experiments server



Experiments

- ☑ Default
- □ XGBoostUniqueLessdata2024
- □ XGBoostUnique2024
- □ XGBoost2024all
- □ XGBoost2024
- □ xgboost2023pipeline
- □ xgb2quantile
- □ tests4rings
- □ hyperopt
- □ XGBoostGPU2023

Default

Experiment ID: 0 Artifact Location: mlflow-artifacts:/0

Description Edit

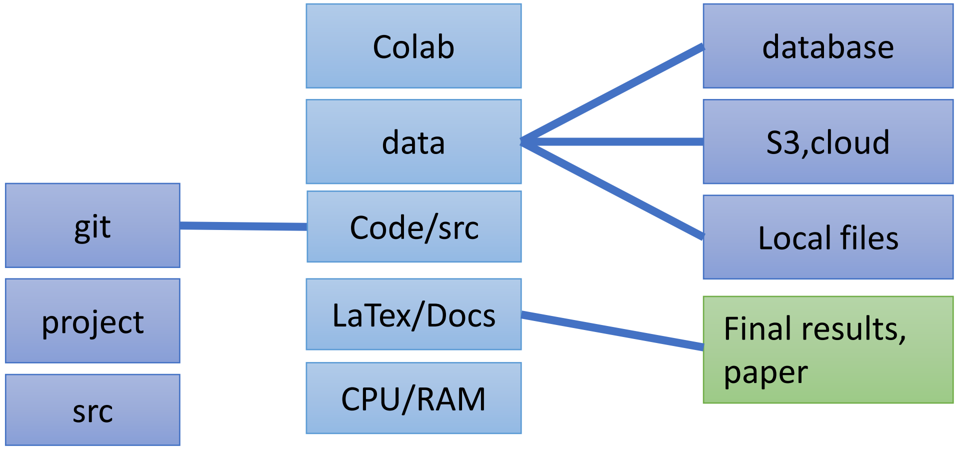
metrics.rmse < 1 and params.model = "tree"

Run Name	Created	Dataset	Duration	Source	Models
● peaceful-midge-944	6 months ago	dataset (cf741d9b) Train		__main__....	-
● suave-perch-350	✓ 11 months ago	dataset (824067a6) Train	1.3min	ipykernel...	xgboost
● valuable-bird-662	☒ 11 months ago	Train (824067a6) Eval , dataset (82406...	5.7s	ipykernel...	-
● nosy-stork-309	☒ 11 months ago	dataset (824067a6) Train , Train... +1	6.5s	ipykernel...	-
● masked-ram-490	☒ 11 months ago	dataset (824067a6) Train	3.1s	ipykernel...	-
● persistent-bird-572	✓ 11 months ago	dataset (824067a6) Train	1.3min	ipykernel...	xgboost
● flawless-wolf-651	✓ 11 months ago	dataset (824067a6) Train	1.3min	ipykernel...	xgboost
● resilient-ray-541	✓ 11 months ago	dataset (824067a6) Train	1.3min	ipykernel...	xgboost
● agreeable-vole-370	☒ 11 months ago	dataset (824067a6) Train , Train... +1	6.8s	ipykernel...	-
● polite-perch-969	☒ 11 months ago	Test (71c370d5) Eval , dataset (... +1	6.7s	ipykernel...	-
● bouncy-fly-305	☒ 11 months ago	Train (824067a6) Eval , Test (71... +1	6.5s	ipykernel...	-
● shivering-wasp-423	☒ 11 months ago	Test (c364971f) Eval , dataset (8... +1	6.6s	ipykernel...	-
● industrious-hare-70	☒ 11 months ago	dataset (824067a6) Train , train (8240...	6.0s	ipykernel...	-
● fun-calf-373	☒ 11 months ago	train (824067a6) Eval , dataset (82406...	5.7s	ipykernel...	-
● bold-elk-870	☒ 12 months ago	dataset (75e844e2) Train , data:... +8	1.1h	ipykernel...	xgboost

- Can we deploy this within the reana?
- Answer: yes
- Security concerns…
- I loved:
  - Streamlit
  - Mlflow
  - Panel
  - React native page

Coming Soon:
Global Workflow of StarHorse team

Transferring spectroscopic stellar labels to 220 million Gaia DR3 XP stars with XGBoost



git

project

src

Colab

data

Code/src

LaTex/Docs

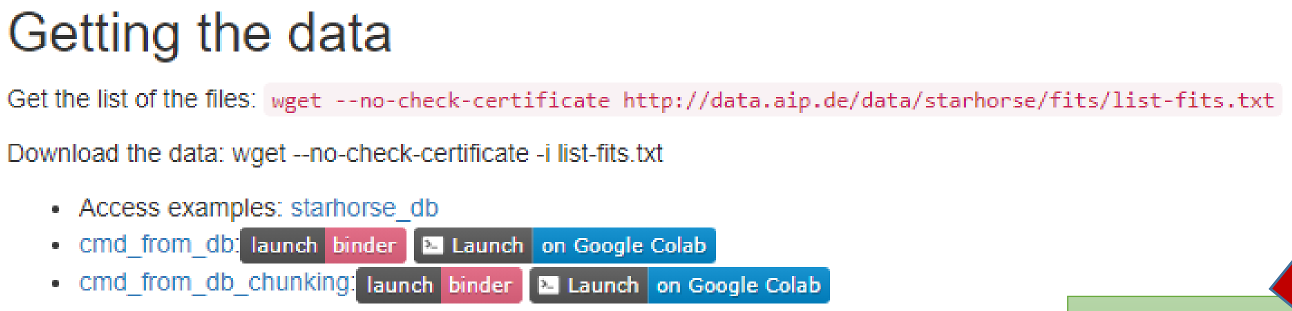
CPU/RAM

database

S3,cloud

Local files

Final results, paper



Getting the data

Get the list of the files: `wget --no-check-certificate http://data.aip.de/data/starhorse/fits/list-fits.txt`

Download the data: wget --no-check-certificate -i list-fits.txt

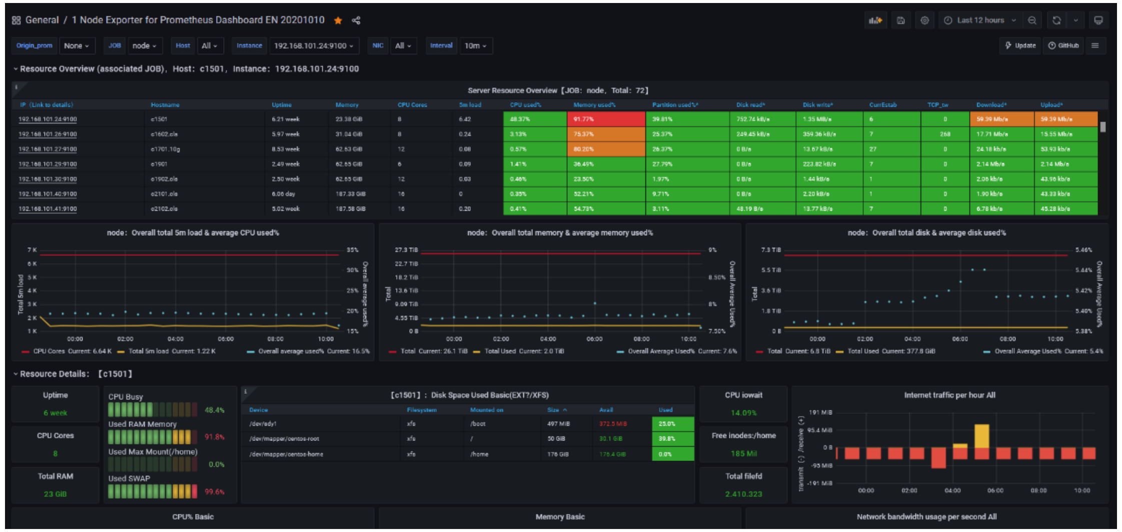
- Access examples: starhorse_db
- cmd_from_db: launch binder | Launch on Google Colab
- cmd_from_db_chunking: launch binder | Launch on Google Colab

1 A100
1 week
Preparation 2 years

~~6 weeks 300 get~~
~~**400 000 00** rameters~~

Monitoring is important!

Cluster Monitoring (Grafana)



General / 1 Node Exporter for Prometheus Dashboard EN 20201010

Origin_prom: None | JOB: node | Host: All | Instance: 192.168.101.24:9100 | NIC: All | Interval: 10m

Resource Overview (associated JOB), Host: c1501, Instance: 192.168.101.24:9100

Server Resource Overview【JOB: node, Total: 72】

IP (Link to details)	Hostname	Uptime	Memory	CPU Cores	5m load	CPU used%	Memory used%	Partition used%*	Disk read*	Disk write*	CurrEstab	TCP_tw	Download*	Upload*
192.168.101.24:9100	c1501	6.21 week	23.38 GiB	8	6.42	48.37%	91.77%	39.81%	752.74 kB/s	1.35 MB/s	6	0	59.39 Mb/s	59.39 Mb/s
192.168.101.26:9100	c1602.cls	5.97 week	31.04 GiB	8	0.24	3.13%	75.37%	25.37%	249.45 kB/s	359.36 kB/s	7	268	17.71 Mb/s	15.55 Mb/s
192.168.101.27:9100	c1701.10g	8.53 week	62.63 GiB	12	0.08	0.57%	80.20%	26.37%	0 B/s	13.67 kB/s	27	0	24.18 kb/s	53.93 kb/s
192.168.101.29:9100	c1901	2.49 week	62.65 GiB	6	0.09	1.41%	36.49%	27.79%	0 B/s	223.82 kB/s	7	0	2.14 Mb/s	2.14 Mb/s
192.168.101.30:9100	c1902.cls	2.50 week	62.65 GiB	12	0.03	0.46%	23.50%	1.97%	0 B/s	1.44 kB/s	1	0	2.06 kb/s	43.96 kb/s
192.168.101.40:9100	c2101.cls	6.06 day	187.33 GiB	16	0	0.35%	52.21%	9.71%	0 B/s	2.20 kB/s	1	0	1.90 kb/s	43.33 kb/s
192.168.101.41:9100	c2102.cls	5.02 week	187.58 GiB	16	0.20	0.41%	54.73%	3.11%	48.19 B/s	13.77 kB/s	7	0	6.78 kb/s	45.28 kb/s

node: Overall total 5m load & average CPU used% — CPU Cores Current: 6.64 K; Total 5m load Current: 1.22 K; Overall average used% Current: 16.5%

node: Overall total memory & average memory used% — Total Current: 26.1 TiB; Total Used Current: 2.0 TiB; Overall Average Used% Current: 7.6%

node: Overall total disk & average disk used% — Total Current: 6.8 TiB; Total Used Current: 377.8 GiB; Overall Average Used% Current: 5.4%

Resource Details: 【c1501】

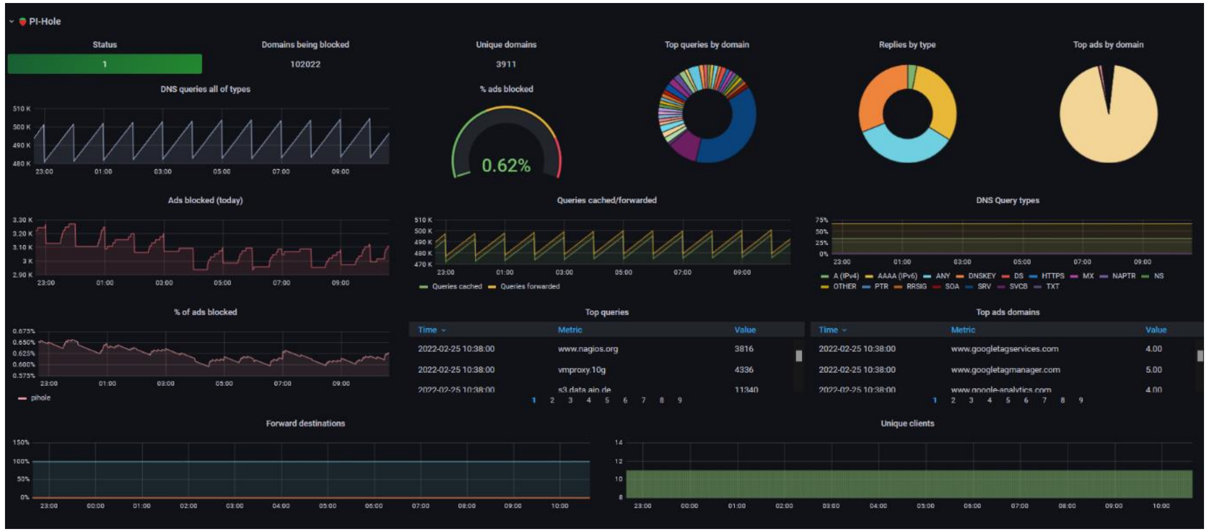
Uptime: 6 week | CPU Cores: 8 | Total RAM: 23 GiB

CPU Busy 48.4% | Used RAM Memory 91.8% | Used Max Mount(/home) 0.0% | Used SWAP 99.6%

【c1501】: Disk Space Used Basic(EXT?/XFS)

Device	Filesystem	Mounted on	Size	Avail	Used
/dev/sdy1	xfs	/boot	497 MiB	372.5 MiB	25.0%
/dev/mapper/centos-root	xfs	/	50 GiB	30.1 GiB	39.8%
/dev/mapper/centos-home	xfs	/home	176 GiB	176.4 GiB	0.0%

CPU iowait: 14.09% | Free inodes:/home: 185 Mil | Total filefd: 2,410,323



Pi-Hole

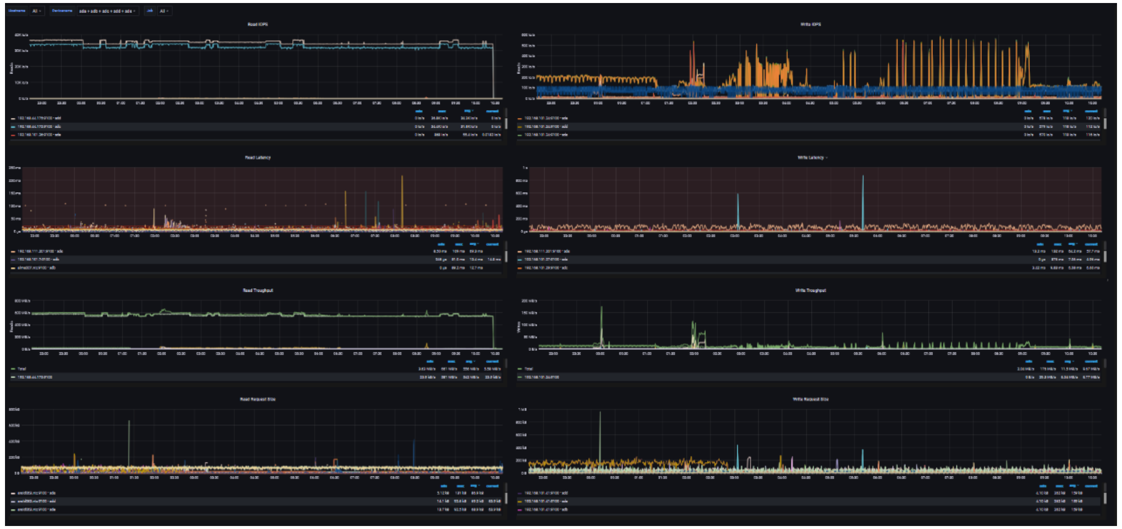
Status: 1 | Domains being blocked: 102022 | Unique domains: 3911 | % ads blocked: 0.62%

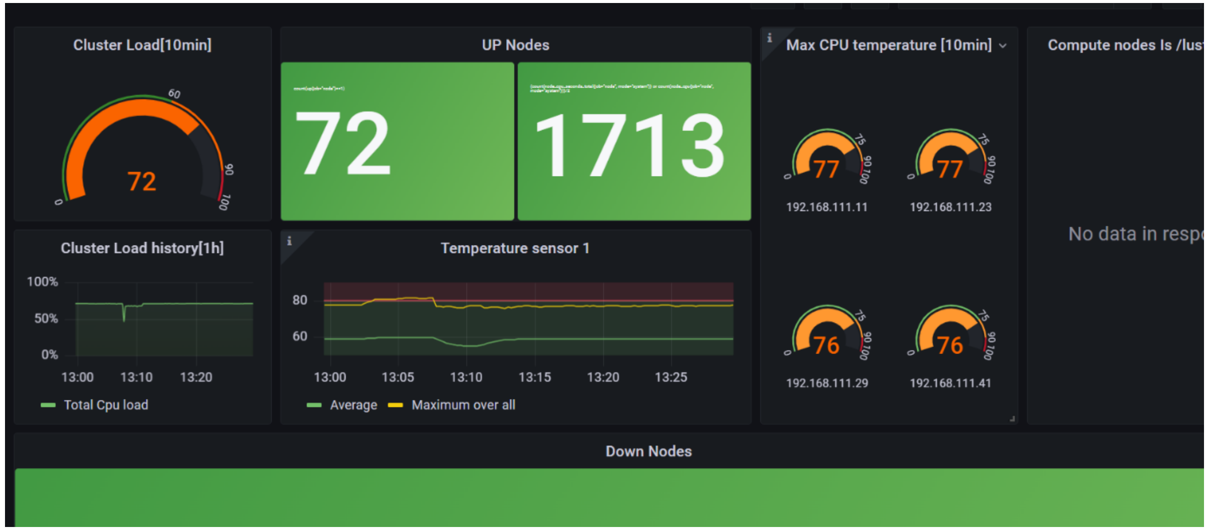
Top queries

Time	Metric	Value
2022-02-25 10:38:00	www.nagios.org	3816
2022-02-25 10:38:00	vmproxy.10g	4336
2022-02-25 10:38:00	s3.data.aip.de	11340

Top ads domains

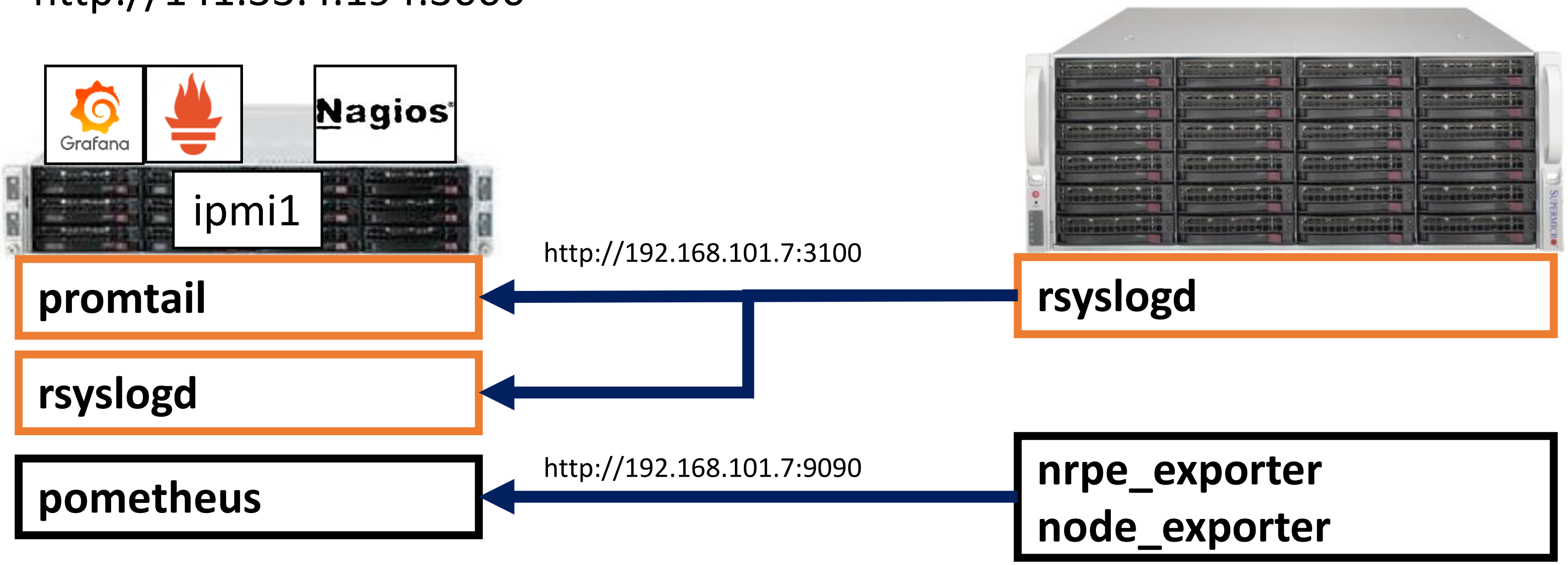
Time	Metric	Value
2022-02-25 10:38:00	www.googletagservices.com	4.00
2022-02-25 10:38:00	www.googletagmanager.com	5.00
2022-02-25 10:38:00	www.google-analytics.com	4.00

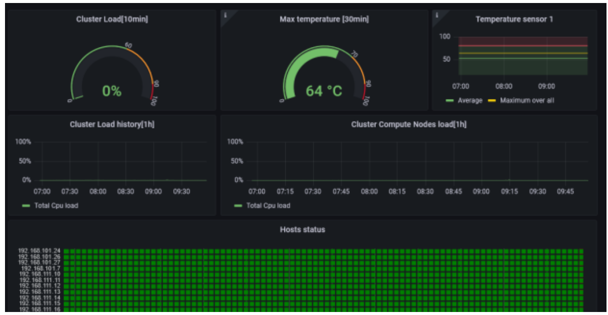


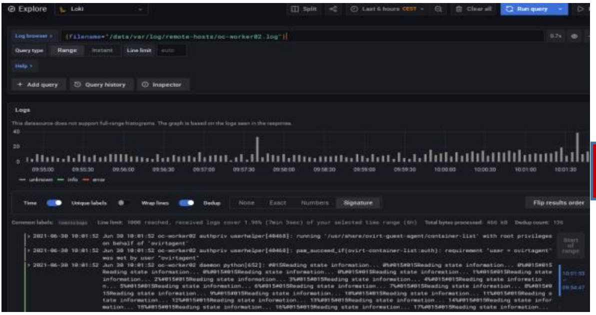


Monitoring stack

http://141.33.4.194:3000





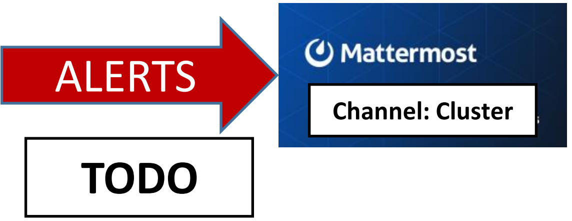


```
{filename="/data/var/log/remote-hosts/oc-worker02.log"}
```

This datasource does not support full-range histograms. The graph is based on the logs seen in the response.

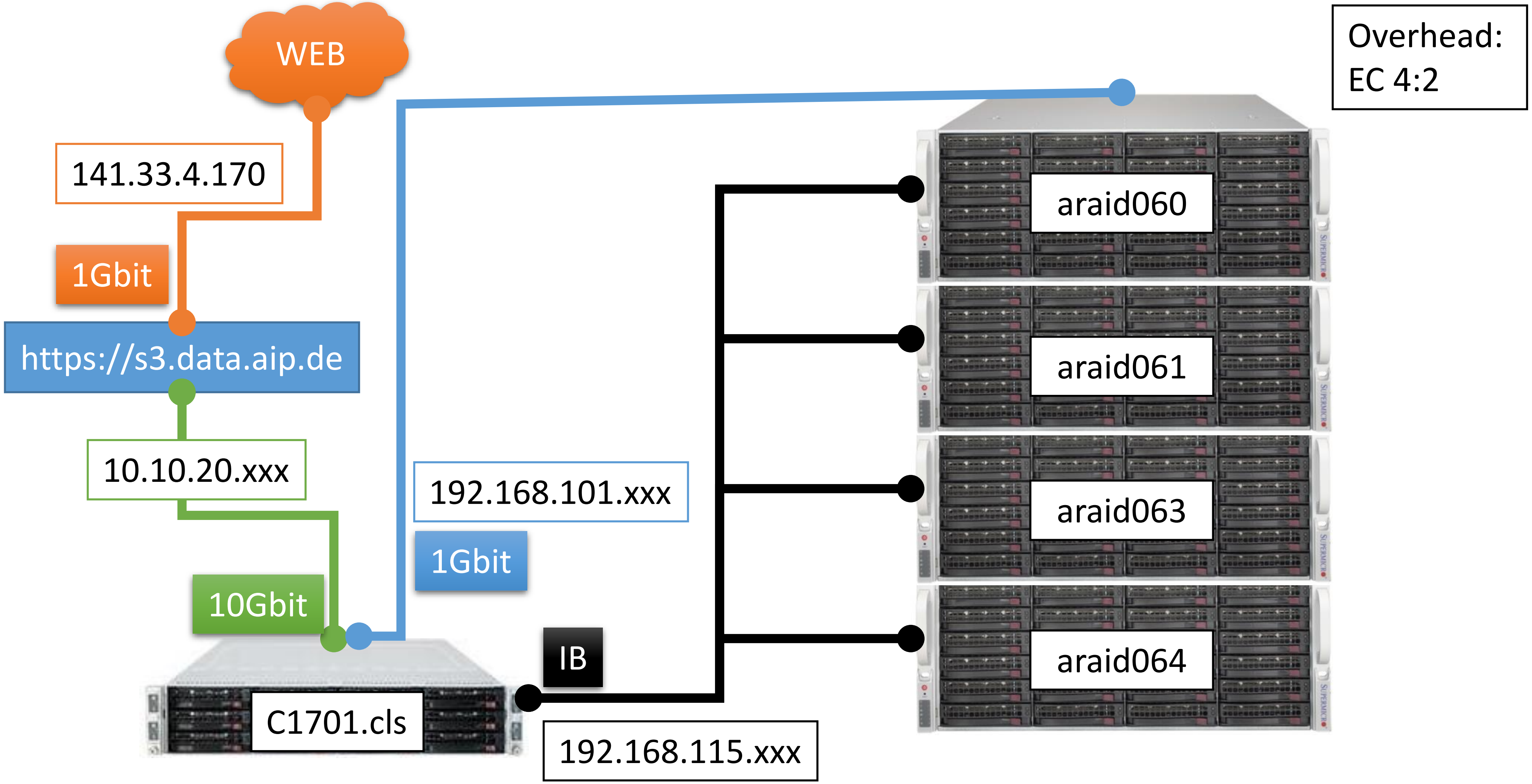
Common labels: remotelogs Line limit: 1000 reached, received logs cover 1.98% (7min 3sec) of your selected time range (6h) Total bytes processed: 456 kB Dedup count: 136

```
2021-06-30 10:01:52 Jun 30 10:01:52 oc-worker02 authpriv userhelper[40468]: running '/usr/share/ovirt-guest-agent/container-list' with root privileges
                    on behalf of 'ovirtagent'
2021-06-30 10:01:52 Jun 30 10:01:52 oc-worker02 authpriv userhelper[40468]: pam_succeed_if(ovirt-container-list:auth): requirement "user = ovirtagent"
                    was met by user "ovirtagent"
2021-06-30 10:01:52 Jun 30 10:01:52 oc-worker02 daemon python[652]: #015Reading state information... 0%#015#015Reading state information... 0%#015#015
                    Reading state information... 0%#015#015Reading state information... 0%#015#015Reading state information... 1%#015#015Reading state
                    information... 2%#015#015Reading state information... 3%#015#015Reading state information... 4%#015#015Reading state informatio
                    n... 5%#015#015Reading state information... 6%#015#015Reading state information... 7%#015#015Reading state information... 8%#015#0
                    15Reading state information... 9%#015#015Reading state information... 10%#015#015Reading state information... 11%#015#015Reading s
                    tate information... 12%#015#015Reading state information... 13%#015#015Reading state information... 14%#015#015Reading state infor
                    mation... 15%#015#015Reading state information... 16%#015#015Reading state information... 17%#015#015Reading state information...
```

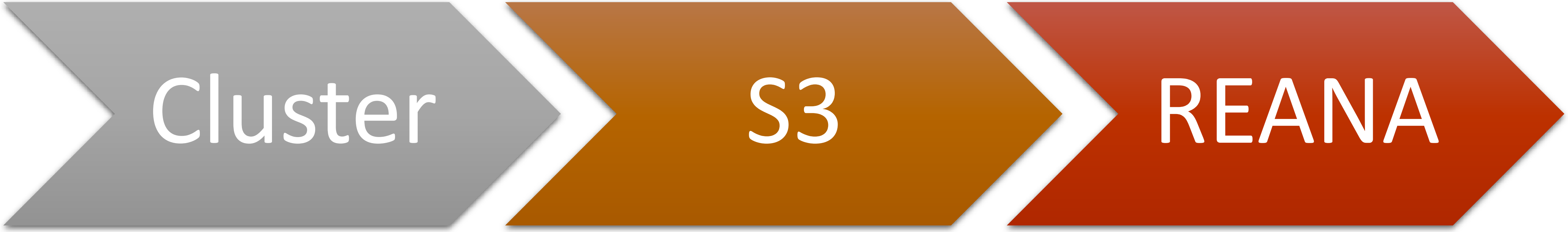


Data management

S3 storage at AIP: MinIO network



Moving data from clusters?



REANA as a main ingredient for NFDI4PUNCH

- Helmholz-AAI is integrated and working w/o problems
- Registered users: >50 users within 3 months
- We are looking for stability tests to announce at AIP
- Gitlab container registry as a main source for containers
- Dev steps:
    - Actively developing HT_Condor integration(Manuel)
    - Any SLURM backend(Arman,Elena)
    - Merging to REANA basecode(Tibor,Marco and team…)
- What is still missing:
    - Workflow shares
    - Data to(from) Storage workflows
    - Easy token management
    - Not implemented the LustreFS integration

~~Use CLOUD everywhere!!!~~

USE STUDY SHARE IMPROVE

I ♥ Free Software

ilovefs.org

Free Software Foundation Europe: fsfe.org

Elena will demonstrate how to use REANA with S3 storage and more

Questions?