



EINSTEIN
TELESCOPE



cnrs



Gravitational Wave Data Analysis on the ESCAPE Virtual Research Environment

A. Iess, CNRS LAPP

XV ET Symposium, Bologna

The ESCAPE Open Collaboration





The ESCAPE Virtual Research Environment (VRE)

The **Virtual Research Environment** was developed by the VRE Team at CERN as part of the ESCAPE Project, under EU Horizon 2020 Grant Agreement no. 824064.

References

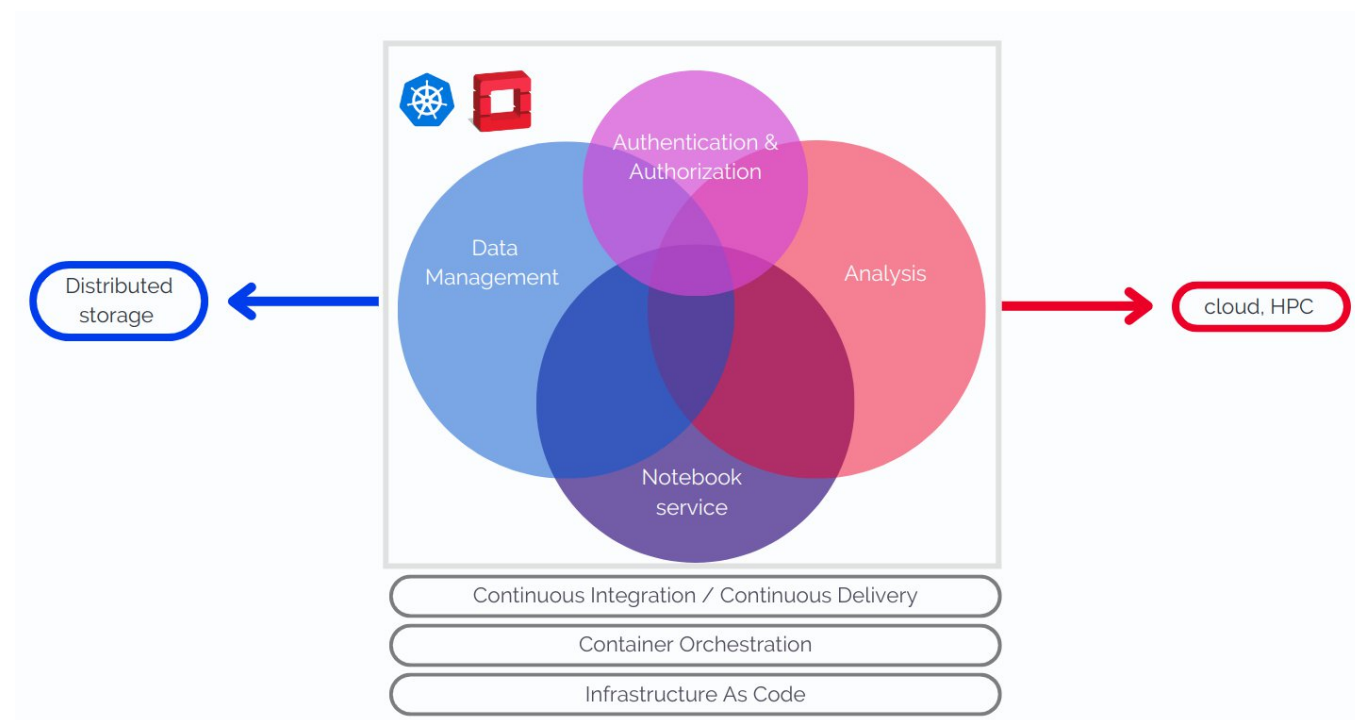
- CHEP2024 proceedings: E. Garcia-Garcia, G. Guerrieri, et al. <https://arxiv.org/pdf/2503.02483>.
- CHEP2024 contribution: <https://indico.cern.ch/event/1338689/contributions/6010696/>
- VRE Documentation: <https://vre-hub.github.io>
- Github Profile: <https://github.com/vre-hub>

CERN VRE Hub: <https://jhub-vre.cern.ch>

Mattermost Channel: <https://mattermost.web.cern.ch/escape/channels/vre-support>

VRE components

- Federated AAI
- ESCAPE Datalake for federated distributed storage .
- Computing cluster supplying the processing power to run full analyses.
- JupyterHub Interface with containerised environments.





ESCAPE AAI

- ESCAPE AAI is based on INDIGO Identity and Access Management (IAM).
- Request an account and wait for approval.
- Add escape group.
- Support Usr+pwd, JSON Web Token OIDC, x.509 certificates.

All ESCAPE Virtual Research Environment (VRE) services and resources are federated through the ESCAPE IAM service.



Welcome to **escape**

Sign in with your escape credentials

 Username

 Password

Sign in

[Forgot your password?](#)

Or sign in with

 Google

 eduGAIN

Not a member?

Apply for an account

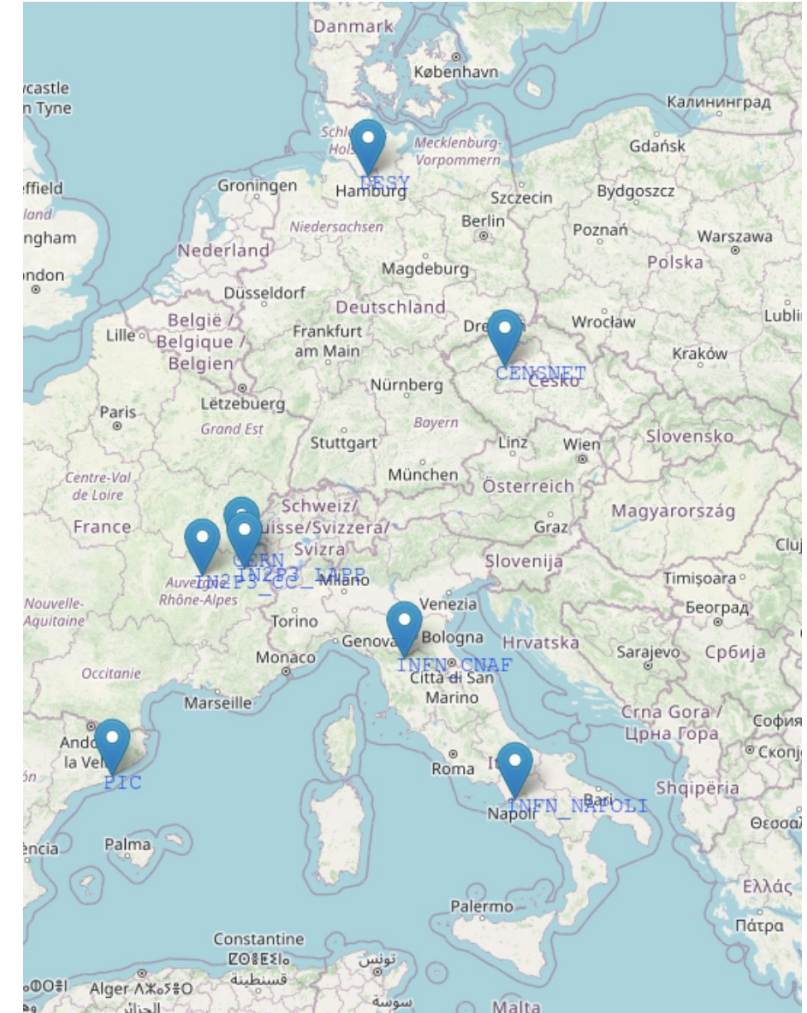
Register an account with eduGAIN

[Info and Privacy Policy](#)

You have been successfully authenticated as
**CN=IESS Alberto lkm3sm8x@cnrs.fr,O=Centre national de la
recherche scientifique,C=FR,DC=tcs,DC=terena,DC=org**
This certificate is not linked to any account in this organization

Rucio for Federated Distributed Data Management

- Distributed Rucio Storage Elements (RSEs).
- Rucio distributed data management system (upload, replication rules, transfer).
- Interaction through docker container or rucio client.
- Authentication through X.509 certificates and OIDC tokens.



The ESCAPE VRE offers a JupyterHub interface:

- AAI (credentials, x.509, OpenID)
- Environments encapsulated in Docker images and run as containers.
- Rucio, REANA plugins.
- VREs at [CERN](#) and at LAPP ([EOSC](#), [internal](#)).
- [Documentation](#).

You can select an environment and run a notebook interactively.

Server Options

- **Default environment**

Based on a scipy notebook environment with a python-3.11 kernel, the Rucio jupyterlab extension and the Reana client installed.

- **ROOT Higgs 2024 environment**

ROOT v6.32.04, and a python-3.11 kernel.

- **ROOT environment**

Legacy ROOT v6.26.10 as well as a ROOT C++ and a python-3.8 kernel.

- **VIRGO - WDF environment**

Contains the full WDF v2.2.3 environment and a Python 3.11 kernel.

- **Python 3.11 environment**

quay.io/jupyter/scipy-notebook:python-3.11 image

- **Default environment - python 3.9**

Same environment as the default one except for a python-3.9 kernel installed. This environment will be deprecated soon.

- **Default environment - python 3.8**

Same environment as the default one except for a python-3.8 kernel installed. This environment will be deprecated soon.

- **KM3Net Science Project environment**

Contains gammapy=1.1, km3lrf and km3net-testdata libraries - Python 3.9 kernel.

- **KM3NeT and CTA combined analysis environment**

Contains gammapy=0.17 and astropy - python 3.9 kernel.



Environments

Alternatively, add an environment with a pull request, which will be reviewed and eventually merged by the VRE team.

README

MIT license

Docker automatic build and publish passing

Environments

VRE user environment images for workflows and notebooks available in the VRE JupyterHub service.

VRE user environments

Custom user environments

If you want to extend or modify any image, find below some suggestions:

We strongly recommend that every user environment image uses the latests `vre-singleuser-py311` image as a base layer (the `FROM` command in the `Dockerfile` - see below).

- Check the latest version of the `vre-singleuser-py311` image [here](#).
- To customize an image please have a look to the [Startup Docker Stack documentation Hooks](#).
- Create one folder per environments. The directory would need to contain a valid `Dockerfile`. Once your PR is merged, the CI will detect the changes in the repository and will trigger the building of the image.
- Add a `README.md` file describing the environment and its usage.
- We strongly recommend to add the following lines to any new user environment (thus in a `Dockerfile`), so that the image would be labelled with the building date.

```
# Beginning of the suggested Dockerfile
FROM ghcr.io/vre-hub/vre-singleuser-py311:sha-5106e39
LABEL maintainer=<your_name>
ARG BUILD_DATE
LABEL org.label-schema.build-date=$BUILD_DATE
...
```

vre-hub / environments

Code Issues 3 Pull requests 9 Actions Projects 1 Wiki Security Insights

environments Public

Edit Pins Watch 4 Fork 1 Star 2

main 16 Branches 0 Tags

Go to file Add file Code

About

VRE user environment images for workflows and notebooks

python docker containers jupyterhub root jupyterlab dask

Readme MIT license Activity Custom properties 2 stars 4 watching 1 fork Report repository

Releases

No releases published Create a new release

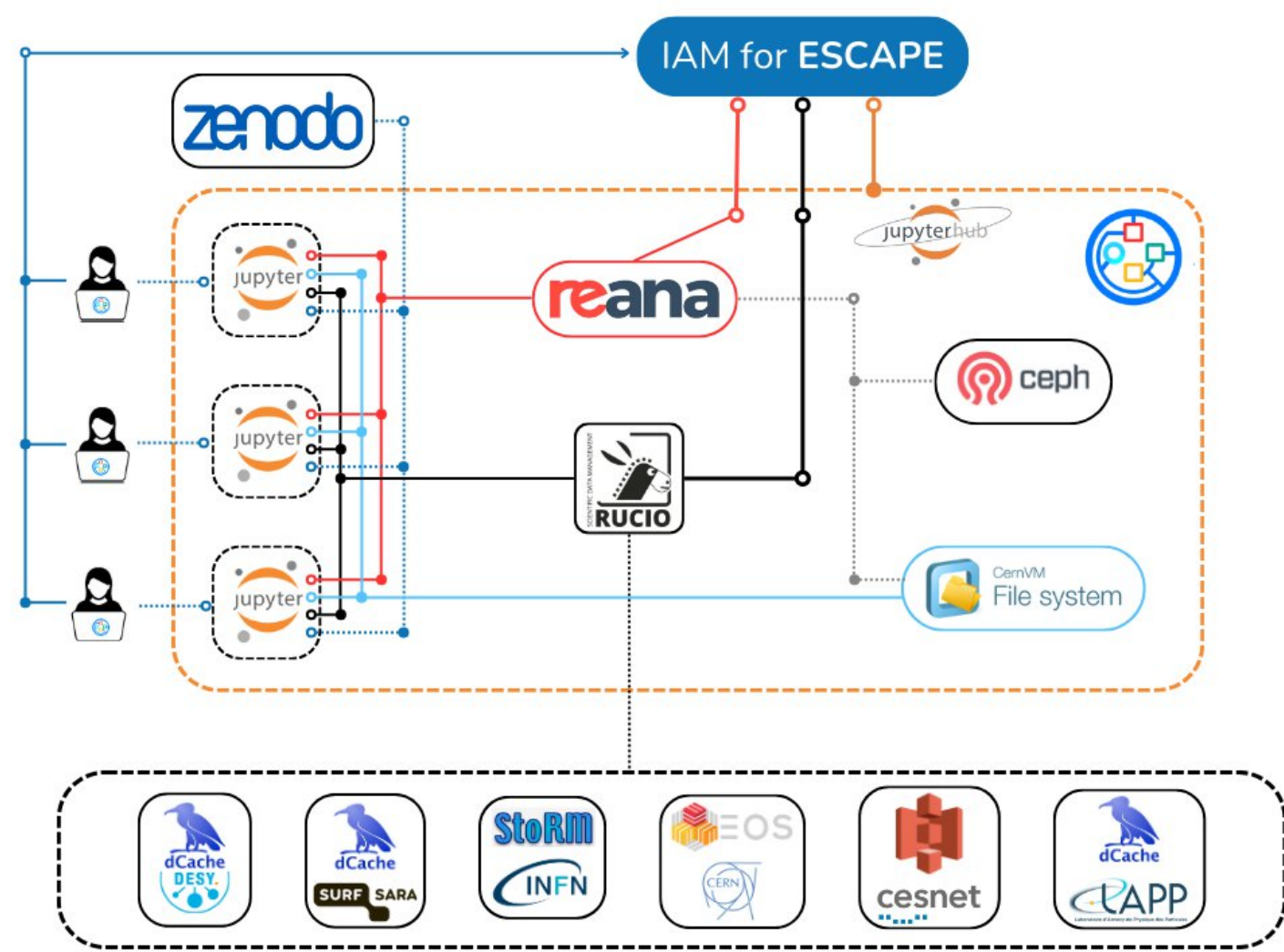
Packages 25

vre-singleuser-intertwin hub atlas-ntuples + 22 packages

Contributors 6

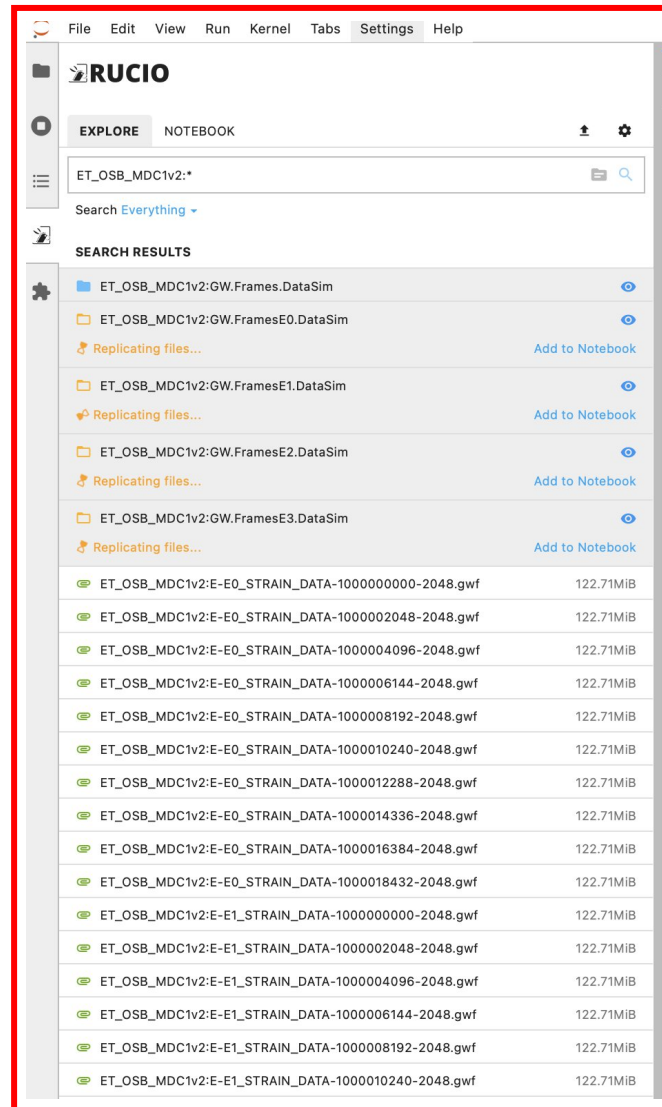
GravitationalWaveSurfer	Feat: add WDF python 3.11 env (#60) ✓	ba497d3 · 15 hours ago	66 Commits
.github	do not trigger second stage, what makes CI fail, if the o...	6 months ago	
atlas-ntuples	change name of repo	2 years ago	
vre-singleuser-combined_ana_km3net_cta	some changes in readme for cta env (#32)	2 years ago	
vre-singleuser-dask-root	adding image with rucio extension and dask (#41)	2 years ago	
vre-singleuser-interTwin	VRE - interTwin DL integration (#45)	8 months ago	
vre-singleuser-itwinai	Itwinai - Rucio jlab extension integration (#50)	5 months ago	
vre-singleuser-km3irf	Km3irf env (#34)	2 years ago	
vre-singleuser-microomega	WIP: migrate microomega docker file to gh and test buil...	2 years ago	
vre-singleuser-py311	Copy correctly vomsdir content (#56)	3 months ago	
vre-singleuser-py38	update py38 base image in agreement with latest chan...	2 years ago	
vre-singleuser-root	Include rsync in packages (#58)	2 months ago	
vre-singleuser-wdf	Feat: add WDF python 3.11 env (#60)	15 hours ago	
vre-singleuser	Trigger build of singleuser (#37)	2 years ago	
AUTHORS.md	some changes in readme for cta env (#32)	2 years ago	
LICENSE	Update LICENSE	2 years ago	
README.md	update environment information	6 months ago	

The Full Picture

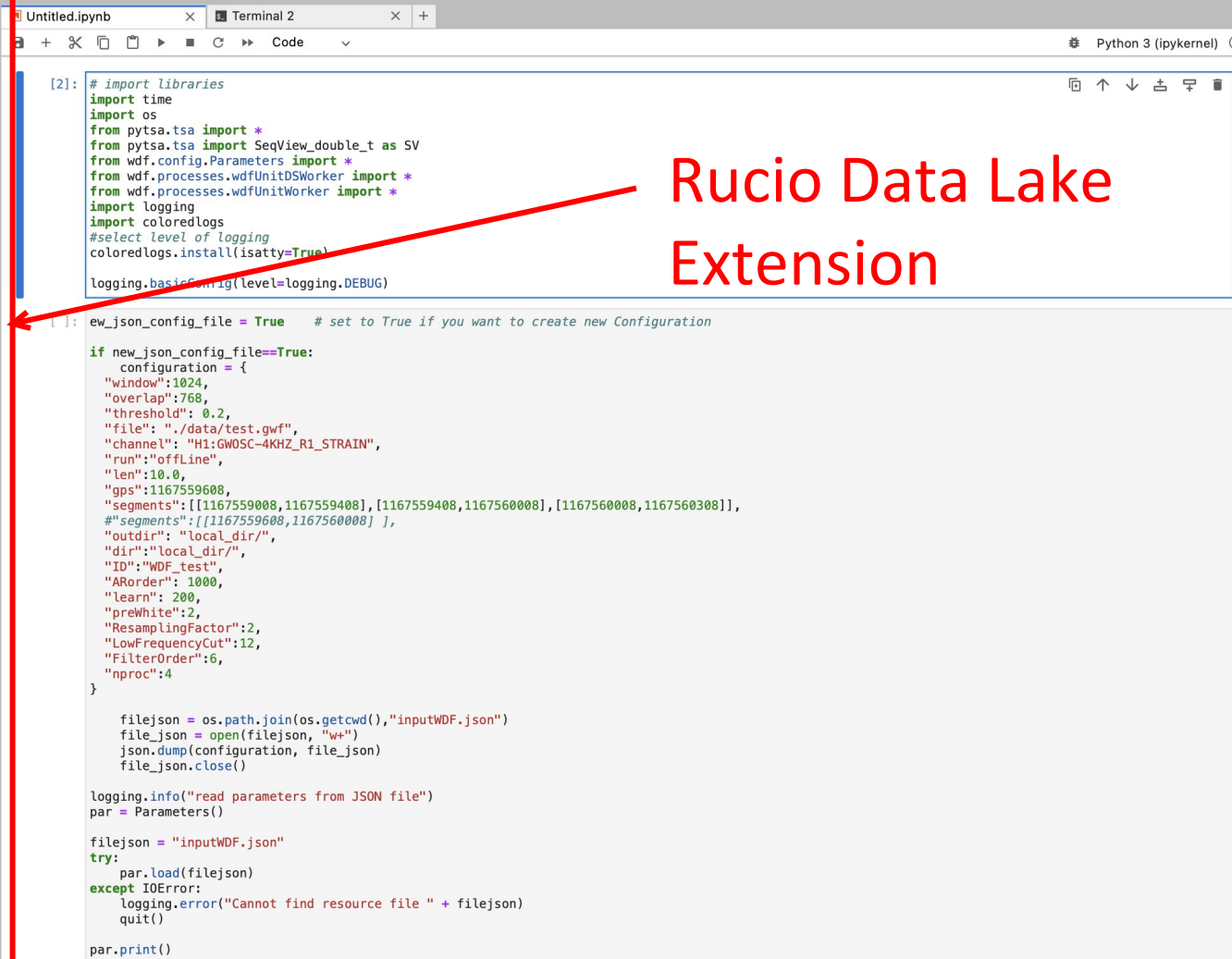


Running an analysis on the ESCAPE VRE

The JupyterHub interface with Rucio extension facilitates access to data, allowing direct upload to Jupyter Notebooks!



The Rucio interface shows a search for 'ET_OSB_MDC1v2:'. The results list several data files, each with a size of 122.71MiB. The files are organized into folders and individual entries, with a 'Replicating files...' status for some. A red box highlights the search results area.



The Jupyter Notebook shows Python code for the Rucio Data Lake Extension. The code imports libraries like 'pytsa', 'wdf', and 'logging'. It then defines a configuration for a WDF test, including parameters like 'window', 'overlap', 'threshold', 'file', 'channel', 'run', 'len', 'gps', 'segments', 'outdir', 'dir', 'ID', 'AROrder', 'learn', 'preWhite', 'ResamplingFactor', 'LowFrequencyCut', 'FilterOrder', and 'nproc'. The code also includes a function to load parameters from a JSON file.

Rucio Data Lake Extension



Running an analysis on the ESCAPE VRE

Example: notebook running multiprocessing Wavelet Detection Filter pipeline for burst signal detection (Cuoco et al. 2018, Cuoco et al. 2001)

File Edit View Run Kernel Git Tabs Settings Help

+

+

↑

↺

Filter files by name

/ ... / ET / E1:STRAIN_1000000000 /

Name	Last Modified
ARcoeff-AR300-...	29 days ago
LVcoeff-AR300-...	29 days ago
parametersUsed...	33 minutes ago
WDFTriggers-E1...	7 seconds ago

test_gw_anal × WDFTriggers × compose.yarr × Terminal 2 × reana.yaml × Snakefile × pipeline.py × downloader.p × param_gener × co

Code git

Ready Python

```
[15]: wdf=wdfUnitDSWorker(par,fullPrint=1)

[*]: # ---- Run multiprocessing wdf (without state vectors) ----
with mp.Pool(par.nproc) as p:
    wdf=wdfUnitDSWorker(par,fullPrint=1)
    p.map(wdf.segmentProcess, par.segments)
    p.close()

#wdf.segmentProcess(par.segment)

2025-05-27 00:20:51 jupyter-albertoiess root[2088] INFO Analyzing segment: 1000002048.0-1000004096.0 for channel E1:STRAIN downsampled at 4096Hz
2025-05-27 00:20:51 jupyter-albertoiess root[2089] INFO Analyzing segment: 1000004096.0-1000006144.0 for channel E1:STRAIN downsampled at 4096Hz
2025-05-27 00:20:51 jupyter-albertoiess root[2087] INFO Analyzing segment: 1000000000.0-1000002048.0 for channel E1:STRAIN downsampled at 4096Hz
2025-05-27 00:20:51 jupyter-albertoiess root[2090] INFO Analyzing segment: 1000006144.0-1000008192.0 for channel E1:STRAIN downsampled at 4096Hz
2025-05-27 00:20:51 jupyter-albertoiess root[2088] INFO Load AR parameters
2025-05-27 00:20:51 jupyter-albertoiess root[2087] INFO Load AR parameters
2025-05-27 00:20:51 jupyter-albertoiess root[2090] INFO Load AR parameters
2025-05-27 00:20:51 jupyter-albertoiess root[2089] INFO Load AR parameters
2025-05-27 00:20:51 jupyter-albertoiess root[2088] INFO Estimated sigma= 2.48899e-23
2025-05-27 00:20:51 jupyter-albertoiess root[2090] INFO Estimated sigma= 2.48976e-23
2025-05-27 00:20:51 jupyter-albertoiess root[2089] INFO Estimated sigma= 2.49079e-23
2025-05-27 00:20:51 jupyter-albertoiess root[2087] INFO Estimated sigma= 2.49165e-23
2025-05-27 00:21:01 jupyter-albertoiess root[2087] INFO Starting detection loop
2025-05-27 00:21:02 jupyter-albertoiess root[2088] INFO Starting detection loop
2025-05-27 00:21:02 jupyter-albertoiess root[2089] INFO Starting detection loop
2025-05-27 00:21:03 jupyter-albertoiess root[2090] INFO Starting detection loop
```


Running an analysis on the ESCAPE VRE

The .csv event trigger file generated by running Wavelet Detection Filter pipeline on the ESCAPE VRE

<div> test_gw_anal × WDFTriggers × compose.yam × Terminal 2 × reana.yaml × Snakefile × pipeline.py × downloader.p × param_gener × config.yaml × </div>									
Delimiter: ,									
	gps	gpsPeak	duration	EnWDF	snrMean	snrPeak	freqMin	freqMean	freqMax
1	1000000002.1875	1000000002.3349609	0.249755859375	0.5090096328597434	0.3304358309852966	1.5626154136668986	64.0	265.15384615384613	492
2	1000000002.5625	1000000002.590332	0.249755859375	0.5150112307452843	0.32948690698212135	1.3899681034144495	88.0	302.53846153846155	532
3	1000000005.5625	1000000005.7119141	0.242919921875	0.5260332433303789	0.38063813106407723	1.5293947286777794	56.0	251.69230769230768	488
4	1000000005.6875	1000000005.7119141	0.249755859375	0.5084603101105939	0.3091762745273994	1.623545971548338	72.0	248.30769230769232	488
5	1000000009.0625	1000000009.2770996	0.249755859375	0.5321581103463993	0.348202901526786	1.735695712904407	60.0	251.46153846153845	548
6	1000000009.125	1000000009.2770996	0.248046875	0.515084243562685	0.32840622654898083	1.7356709540608741	76.0	260.15384615384613	548
7	1000000012.375	1000000012.6000977	0.249755859375	0.5812086311386826	0.32150536887667847	1.855986213599196	72.0	270.6923076923077	560
8	1000000012.4375	1000000012.6000977	0.24560546875	0.5355878675796901	0.3224722419252462	1.8164082288500996	56.0	258.9230769230769	556
9	1000000012.5	1000000012.6000977	0.219482421875	0.5283963368159178	0.31108966417654255	1.8323184161283326	56.0	278.46153846153845	556
10	1000000014.8125	1000000015.032959	0.235595703125	0.6082811202060369	0.3924666378908149	2.2688403490803326	60.0	256.2307692307692	500
11	1000000014.875	1000000015.032959	0.24169921875	0.6437043479061679	0.4213356644600656	2.2444956126386786	68.0	289.3076923076923	524
12	1000000014.9375	1000000015.032959	0.241943359375	0.6161986213791932	0.4098125173637369	2.227456621365669	64.0	282.53846153846155	500
13	1000000015.0	1000000015.032959	0.249755859375	0.5287611498376082	0.3541891159173277	2.2030764128831555	76.0	286.0769230769231	548
14	1000000015.25	1000000015.3498535	0.249755859375	0.5485104527401753	0.34145573075477914	1.6717384709040886	100.0	252.3846153846154	452
15	1000000019.6875	1000000019.7192383	0.249755859375	0.5225201461512395	0.3410631793993525	1.8502531300496663	68.0	269.0769230769231	504
16	1000000019.75	1000000019.8903809	0.243896484375	0.5007517928950685	0.3286354440597249	1.540387518531809	56.0	228.69230769230768	496
17	1000000021.0625	1000000021.2419434	0.24072265625	0.5262832802322703	0.3412362372352981	2.013707989954107	28.0	251.84615384615384	544
18	1000000021.4375	1000000021.6296387	0.249755859375	0.5581610663642006	0.32336850379768367	1.3170298836277476	60.0	219.46153846153845	428
19	1000000021.5	1000000021.6813965	0.248779296875	0.598113337899521	0.3634672927054459	1.5205310713493543	48.0	209.84615384615384	436
20	1000000021.5625	1000000021.6813965	0.238037109375	0.5097732791391584	0.3210342210444631	1.4873077394455962	52.0	212.0	392

Running an analysis on the ESCAPE VRE



For complex reproducible analyses workflows, you can dispatch to REANA cluster.

- The REANA JupyterLab extension allow users to interact with the workflow management system from within JupyterLab.
- Reana - Rucio integration to directly upload files from a Rucio RSE to the Reana workspace. No need to download files locally from Rucio and then upload them to the Reana workspace.

✓ test_pygwb_vre #10
Finished a month ago
finished in 1 min 51 sec step 7/7

⚙️ Engine logs
> Job logs
📁 Workspace
📄 Specification

```

2025-04-23 12:58:14,447 | reana-workflow-engine-snakemake | MainThread | INFO | Workflow spec received: Snakefile
2025-04-23 12:58:14,541 | snakemake.logging | MainThread | WARNING | Building DAG of jobs...
2025-04-23 12:58:14,573 | snakemake.logging | MainThread | WARNING | Using shell: /usr/bin/bash
2025-04-23 12:58:14,574 | snakemake.logging | MainThread | WARNING | Provided cluster nodes: 300
2025-04-23 12:58:14,574 | snakemake.logging | MainThread | WARNING | Singularity containers: ignored
2025-04-23 12:58:14,577 | snakemake.logging | MainThread | WARNING | Job stats:
job                count
-----
all                 1
download_data      4
pygwb_combine      1
run_pygwb          2
total              8

2025-04-23 12:58:14,577 | snakemake.logging | MainThread | WARNING | Select jobs to execute...
2025-04-23 12:58:14,598 | snakemake.logging | MainThread | INFO |
2025-04-23 12:58:14,599 | snakemake.logging | MainThread | INFO | [Wed Apr 23 12:58:14 2025]
2025-04-23 12:58:14,599 | snakemake.logging | MainThread | INFO | rule download_data:
output: E-E1_STRAIN_DATA-1000000000-2048.gwf

```



Running an analysis on the ESCAPE VRE

Nice example:

https://github.com/GeorgySk/et_mdc_pygwb/tree/master

More to come...

The screenshot displays the GitHub repository page for `GeorgySk / et_mdc_pygwb`. The repository is public and has 1 branch and 0 tags. The file list shows the following files and their commit history:

File	Commit Message	Time
<code>.env</code>	Initial commit	3 months ago
<code>README.md</code>	Update README.md	3 months ago
<code>Snakefile</code>	update paths to the MDC data	3 months ago
<code>compose.yaml</code>	Initial commit	3 months ago
<code>config.yaml</code>	Initial commit	3 months ago
<code>parameters.ini</code>	Initial commit	3 months ago
<code>reana.yaml</code>	Initial commit	3 months ago

The README section is titled "How to run a parallel pygwb pipeline on the VRE-Reana cluster with ET MDC data". It states: "The scripts on this repository are configured to run on the Reana VRE instance. However, the pipeline can be run on any Reana instance by adapting the reana credentials." The "Setup" section mentions: "For setting up the certificates and the accounts, see [How to run a serial pygwb pipeline on Reana cluster with ET MDC data](#)."

Conclusions

- The ESCAPE VRE facilitates the development of end-to-end physics workflows, providing researchers with access to an infrastructure and to the digital content necessary to produce and preserve a scientific result in compliance with FAIR principles.
- Lower the entry barrier for young researchers through JupyterHub and straightforward data access, while allowing experienced researchers to run complex workflows through REANA.
- Open-source and modular, in order to make it easily reproducible by partner institutions.
- Experienced team, willing to help setting up and customizing for diverse scientific communities.
- More in upcoming talks on MADDEN and ETAP OSCARS funded projects.

EXTRA SLIDE: REANA

Reproducible analysis platform for containerised data analysis pipelines on remote compute clouds.

- Supported workflow systems: [CWLSerial](#), [Snakemake](#), [Yadage](#)
- Supported compute backends: [HTCondor](#), [Kubernetes](#), [Slurm](#)
- Supported source code and storage systems: GitLab, CVMFS, EOS

