Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
terabit

Status of the ET Site Data Repository

Alessio Fiori

INFN – Sezione di Pisa, Pisa research unit

In collaboration with:

M. Di Giovanni (GSSI & INFN - LNGS)

F. Fidecaro (Università di Pisa & INFN Pisa)

E. Mazzoni (INFN Pisa)

M. Razzano (Univerità di Pisa & INFN Pisa)

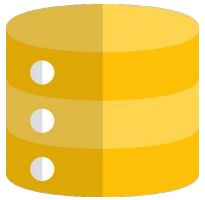
ET: Scienza e Tecnologia in Italia

Assisi, 20-23 Feb 2024

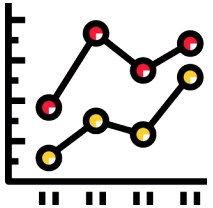
Introduction

Icons from https://www.flaticon.com/

- Services dedicated to the **ET site characterization group**



Repository of seismic data



Analysis infrastructure

- Development started October 2023, prod instance deployed January 2024
- Collaboration between **INFN**, **INGV** and **GSSI**
- Work carried on within the **TeRABIT project** (https://www.terabit-project.it/)

Talk by M. Campanella
Wed, 21 Feb

- Preliminary presentation at SPB session of the 2nd ET Annual Meeting (https://indico.ijclab.in2p3.fr/event/9686/contributions/32239/attachments/22212/31689/afiori_etm_nov2023_v2.pdf)

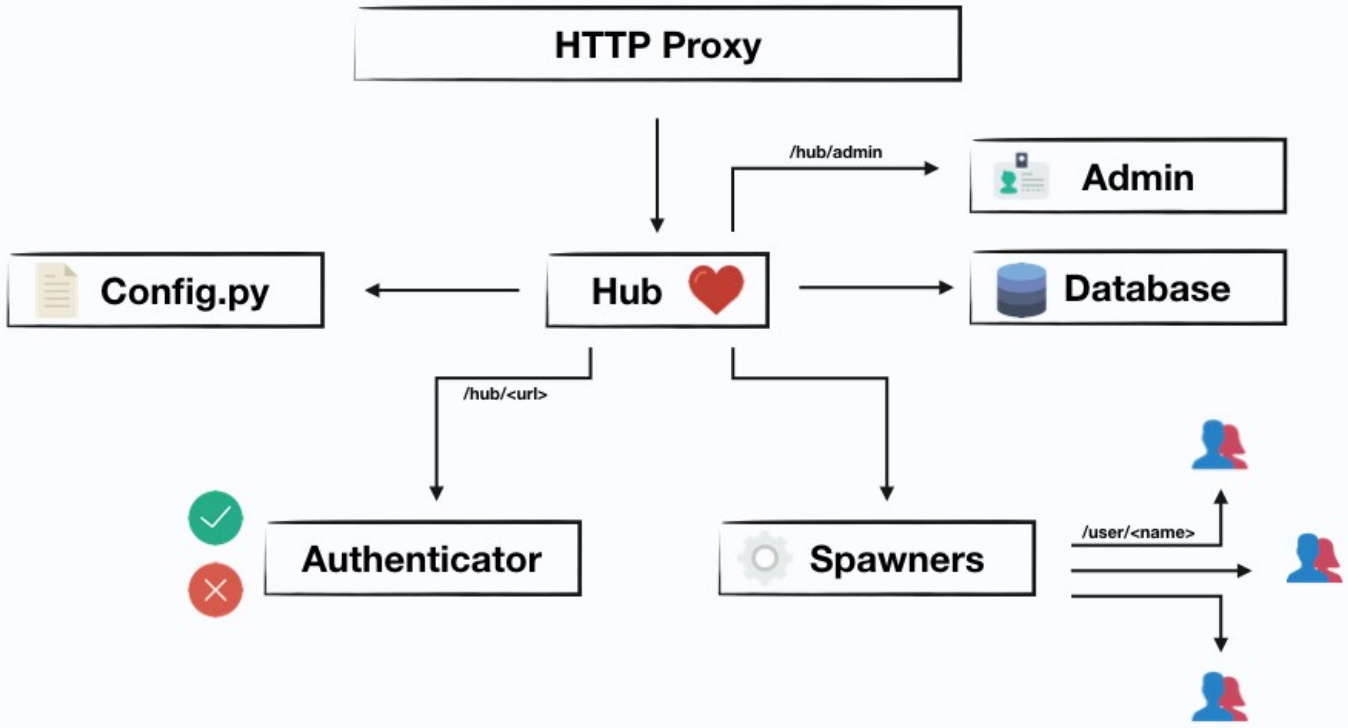
Overview

- Previously hosted at the **Green Data Center @ University of Pisa**
- **New solution:** virtual machine instantiated via **VMWare vCenter @ INFN Pisa**
- **Configuration:**
  - **AlmaLinux OS 9**, open source, compatible with Red Hat 9 (standard distro in high-energy experiments, https://linux.web.cern.ch/)
  - **8 CPUs, 16 GiB memory** (scalable in case of larger usage)
  - **11 TB GPFS disk space** for data and user work dirs
- For improved flexibility and security, we use a fully containerized approach with **podman** (Red Hat built-in container tool, **compatible with docker**)

JupyterHub

FDSN Web Server

The JupyterHub infrastructure



All icons where obtain on Flaticon (https://www.flaticon.com/packs/essential-collection)

Login with ET Members Database credentials

Singleuser servers live inside docker containers

From https://jupyterhub.readthedocs.io/en/stable/

The singleuser server

- Login specifications:

  USER The ETMD username

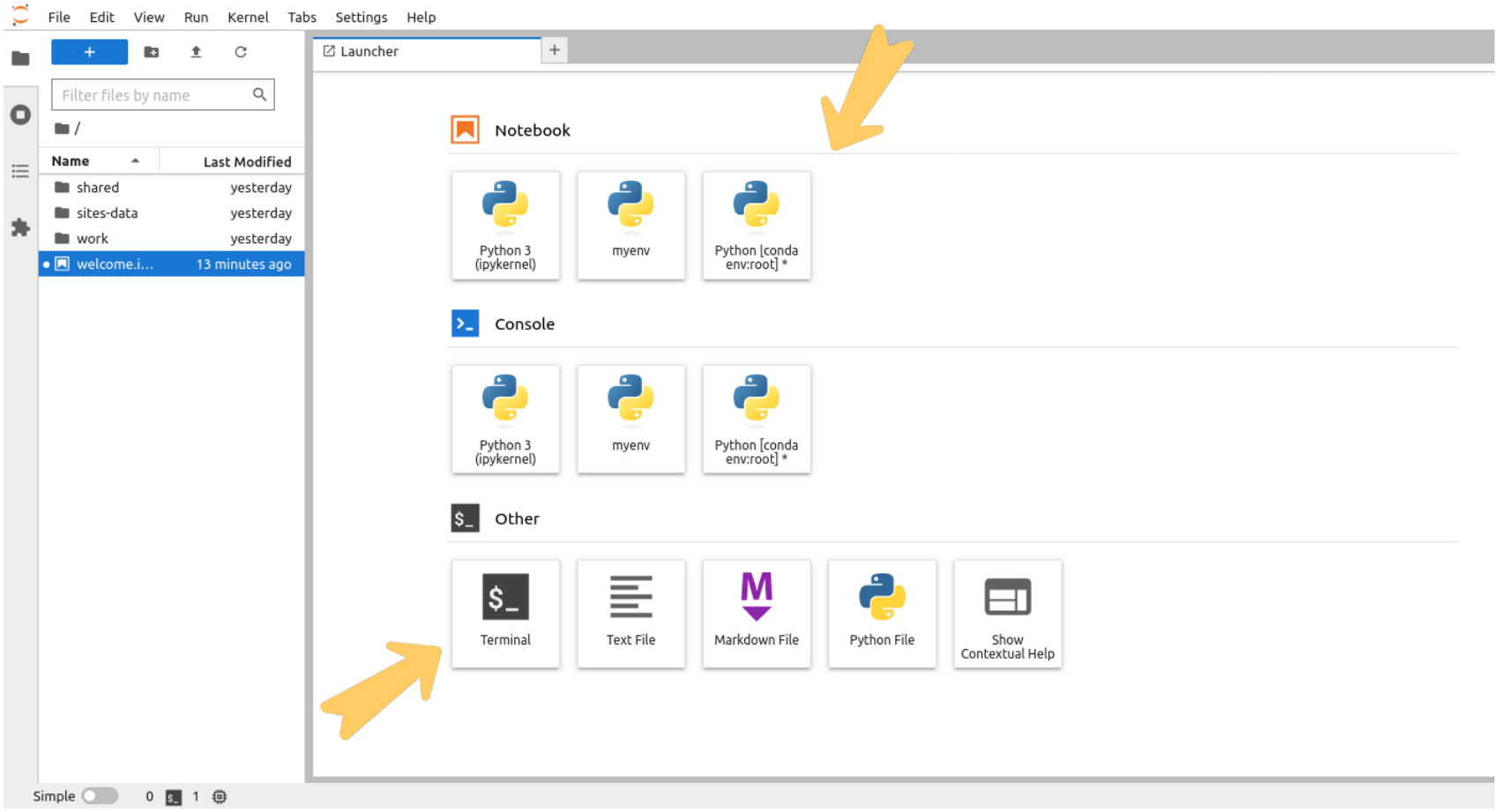
  UID 2000 + ETMD ID number

  GID Common GID required for I/O operations to the GPFS volume

- When the container is created, a bash start script creates the user and configures its home directory
- **The user's home directory is NOT persited by default**
- List of persisted directories:

Link	Location	Permissions	Description
work	/mnt/user-work/username	700	This is your work directory. Store your code and data here. You are the owner of this directory and its content is only visible to you.
sites-data	/mnt/sites-data	755	This is the location where the data are stored. All users have read-only access to the content of this directory.
shared	/mnt/shared	777	This is a shared directory. All users have read-write access to this directory. You should not put any sensitive data inside it.

JupyterHub web interface



- **Conda** preinstalled for all users
- Users can create custom environments
- **ObsPy 1.4.0** already installed in the root env

Remote SSH (new feature)

jupyter

- Enabled OpenSSH service inside the singleuser containers (available since Monday, Feb 19, 2024)
- **Authentication via public RSA key** to be stored under $HOME/work/.ssh/authorized_keys

```
$ mkdir $HOME/work/.ssh
$ echo "PUT YOUR KEY HERE WITH QUOTES" >> $HOME/work/.ssh/authorized_keys
```

- **Disabled username/password authentication**
- Server accessible via a dedicated port (20000 + UID) stored in the **$SSH_PORT** env

```
$ scp -P PORT /path/to/local/file USER@etsitedata.pi.infn.it:/home/USER/work/
```

- Full documentation on the wiki: https://wiki.et-gw.eu/SPB/ETSiteData#Remote_SSH

FDSN Web Service with SeisComP

- Python-based software for data acquisition and analysis
- **Why SeisComP?**
  - Modular package with **several preinstalled tools**
  - Easy installation and configuration
  - Detailed documentation (https://www.seiscomp.de/doc/)
- The **FDSN module** enables HTTP(S) queries to **download data and station info remotely** (https://www.seiscomp.de/doc/apps/fdsnws.html)
- Federated network of digital seismographs (https://www.fdsn.org/webservices/)

gempa SeisComP development services and trainings

GFZ Helmholtz-Zentrum Potsdam

Status of the FDSN Web Service

SeisComP

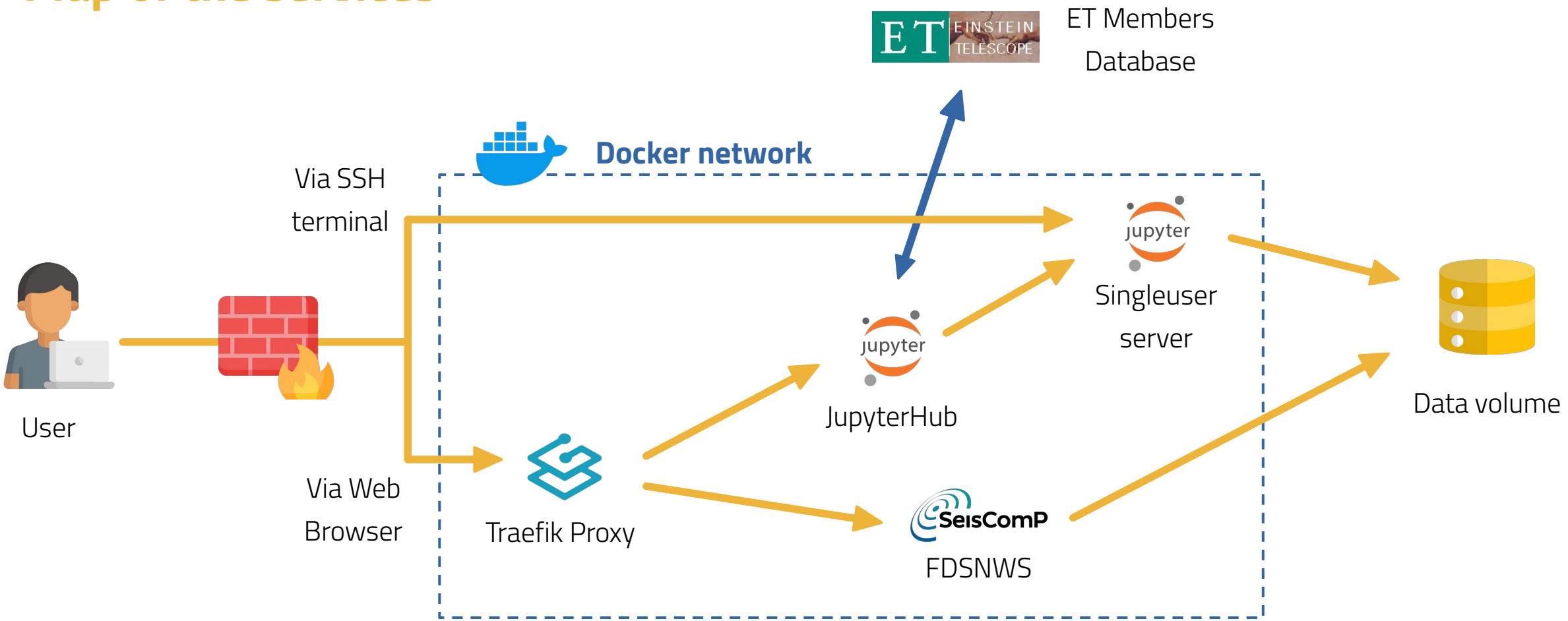
- Structured data archive (SDS) prepared by M. Di Giovanni
- Overview of Sos Enattos stations in a readme file on the TDS (ET-00074A-24) and in this wiki page (https://wiki.et-gw.eu/SPB/ETSiteData#Data_availability)
- **Service currently offline** (waiting for data policy)

```
archive
  + year
    + network code
      + station code
        + channel code
          + one file per day and location,
```

Site	Network	Stations	JupyterHub root path
Sos Enattos	ET (permanent)	SOE0, SOE1, SOE2, SOE2T, SOE3, P2, P3	/mnt/sites-data/sosenattos/ETSCOPE
Sos Enattos	7P (INGV, temporary, seismometers)	P1**, P2**, P3**, WP1-9	/mnt/sites-data/sosenattos/ETSCOPE/temporary_arrays
Sos Enattos	ET (GSSI, temporary, geophones)	P2, P3, BR	/mnt/sites-data/sosenattos/ETSCOPE/temporary_arrays
Sos Enattos	ACNT (temporary, weather + seis.)	SOE0, SOE2, SOE3, SURF0, UNDG1-7	/mnt/sites-data/sosenattos/ETSCOPE/temporary_arrays
Terziet & Cottessen	NL	TERZ, CTSN	/mnt/sites-data/NLData/TerzData /mnt/sites-data/NLData/CTSNData

Map of the services

Icons from https://www.flaticon.com/



Status

Service	Version	Status	Comments
jupyter	v4.0.2	Deployed (prod)	Available here: https://etsitedata.pi.infn.it/jupyter/
SeisComP	v6.1.1	Ready (offline)	Will be available here: https://etsitedata.pi.infn.it/fdsnws/ Data policy to be discussed

- **Useful links**
  - Documentation https://wiki.et-gw.eu/SPB/ETSiteData
  - Sos Enattos stations https://apps.et-gw.eu/tds/?content=3&r=18666
  - GitLab project https://gitlab.et-gw.eu/etsitedata

Thank you

alessio.fiori@pi.infn.it