2nd Einstein Telescope Annual Meeting
Nov 14 – 17, 2023

Migration of the Sardinia data repository to INFN Pisa

Alessio Fiori

- **Terabit Network for Research and Academic Big Data in ITaly**

- Integrating and enhancing Italian digital research infrastructures:

  - Network (GARR-T)

  - High-performance computing (PRACE-Italy)

  - Cloud-edge distributed computing (HPC-BD-AI)

- Data transfer via optical fiber up to 1 Tb/s

- 41M € fund by the National Recovery and Resilience Plan

- **12M € investment to link research institutes in Sardinia with the Italian network**

- **Crucial for ET data distribution from Sos Enattos**

https://www.terabit-project.it/
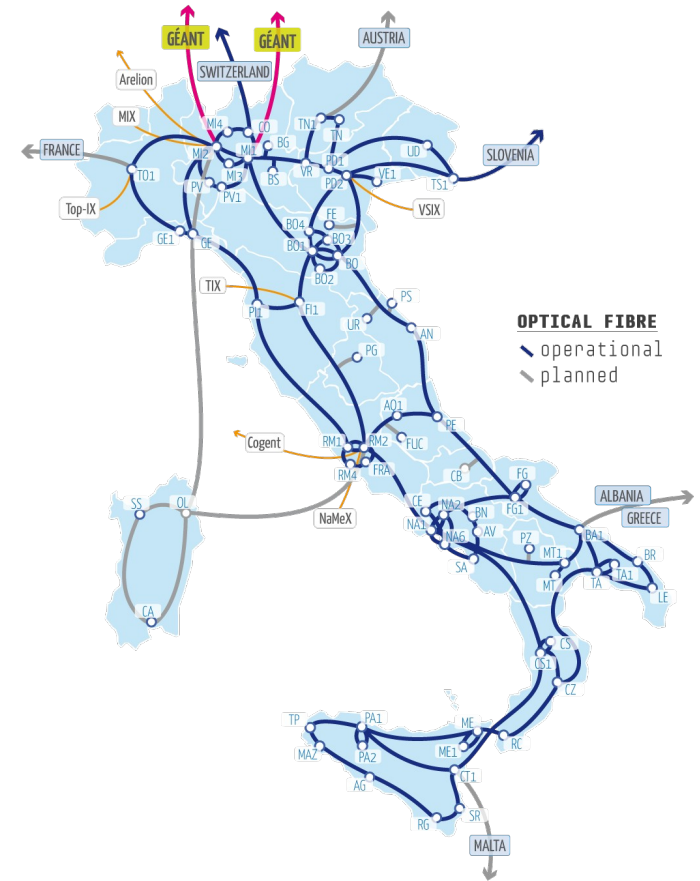
Current repo configuration

- **Virtual machine hosting site data and JupyterLab user space**

- Hosted in Green Data Center @ University of Pisa since 2019

- Microsoft Hyper-V machine with the following specs:

  - Centos7 OS

  - 16 CPU cores Intel Xeon 5120 (28 thread/core)

  - 5 TB hard disk

  - 32 Gb RAM

- User requested via Google Form and updated periodically

- See A. Fiori et al. @ SPB Workshop, ET Symposium 2023 (link)

Client

Google Form

User DB

Host VM

Mail   Account

jupyter

Issues (reported and not)
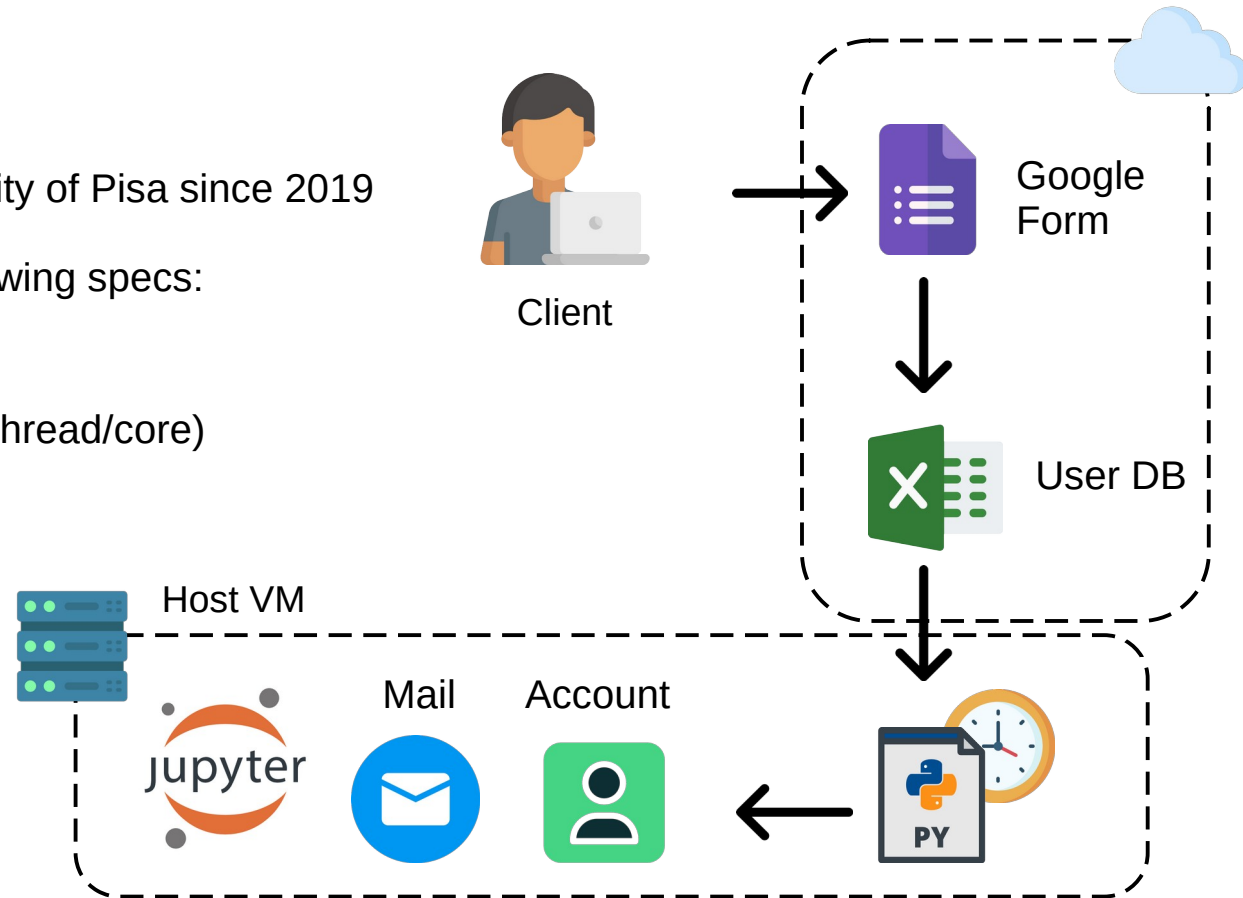
- **Frequent service shortages reported by users**

  - We have no control over VM status (we are not system admins @ Green Data Center)

  - Issues solved by contacting the system administrator in Pisa

- **Data disk often found in read-only safe mode**

  - We always need to stop services, remount disk and restart

- **Centos7 approaching end of life (scheduled June 2024)**

  - Centos9 is Stream, i.e. continuous development (bad for version control)

- **PROPOSED SOLUTION: migration to INFN Pisa**

Migration to INFN Pisa

- **VMWare vCenter available at the Computing Center @ INFN Pisa**

- INFN Pisa programmed to join the Italian INFN Cloud federation in 2024

- **Configuration:**

  - **AlmaLinux OS 9**, open-source distro compatible with Red Hat 9, future standard in high-energy experiments (CERN, FermiLab)

  - High availability of CPU and memory resources

  - ~10 TB **scalable GPFS disk space** for site data and user work dirs

  - Fully containerized services for more flexibility (built-in docker emulator in AlmaLinux9)

- **Enhanced security:** SSH key pair authentication, disabled username:password

- We expect a gradual transition

- **JupyterHub v4.0 deployed with docker-compose**

  - Same configuration as before

  - One base environment with NumPy, SciPy, ObsPy, etc. periodically updated

  - Pre-existing conda installation (users can create their own envs)

  - PEM authentication (same users as the host machine)

- **Persisting user space and data** mounted on single-user JupyterLab servers

- **TLS connection via Traefik router**

- **No interaction with host VM**

Services: web application
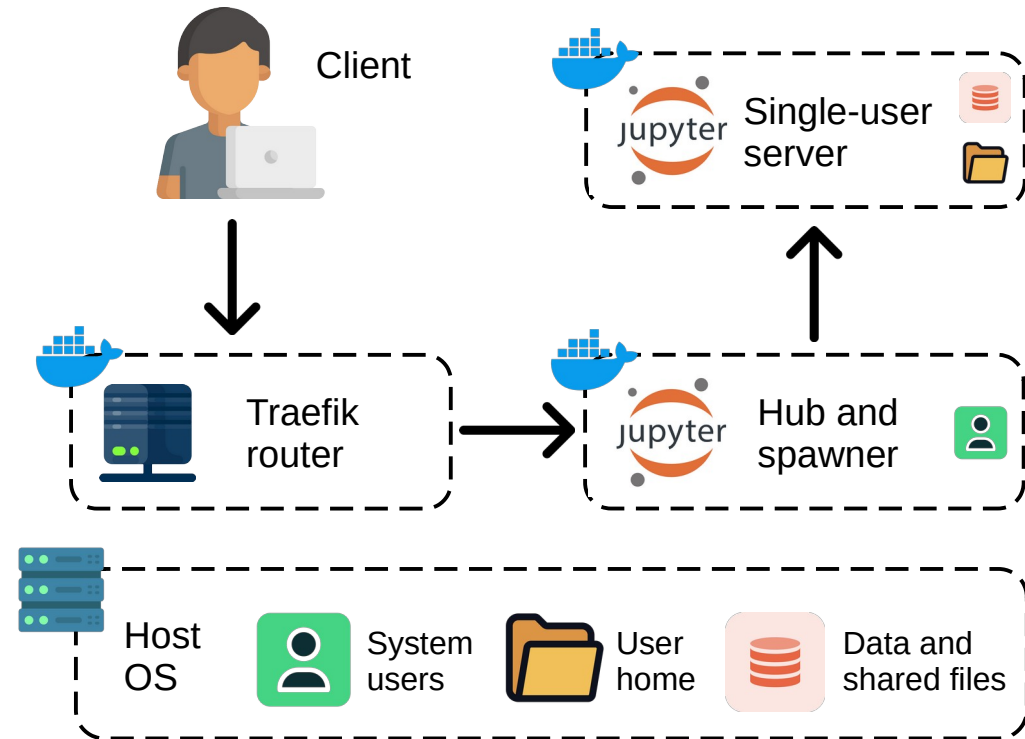
- **Python Flask web app**:
  - **Request form** for new users
  - Accessible via authentication against ETMD (temporary solution based on Java response)
  - **Email SMTP client** for mail notices to users
  - System user management (for admins)
- Possibly hosting other resources and services
- Runs inside a Docker container
- TLS connection enabled via Traefik router
- **Status: debugging**

ET Repository account request

This form is for requesting an account on the ET site repository and on the ET github repository (https://github.com/et-sw). Once you have filled the form, your account will be created and you will receive your credentials via email.

First Name
John

Last Name
Doe

Institution
University of Nowhere

Email
j.doe@nowhere.com

GitHub account (optional)
Enter your GitHub username

Submit

HOME
REQUEST AN ACCOUNT
GO TO JUPYTERHUB
DOCUMENTATION
CONTACTS

Services: FDSN web service

- **Interactive service for seismic data distribution**

- Proposed by colleagues at INGV Pisa to distribute Sardinia data

- **Status:**

  - Under development @etrepo.df.unipi.it

  - Matteo Di Giovanni (INGV Pisa) is organizing Sos Enattos data and metadata on the repo

- **Prospects:**

  - We will sync data with INFN Data Center when the metadata are ready

  - Production FDSN web service @ INFN Pisa when ready

Summary and status

- **Virtual machine already up and running** @ INFN Pisa

  - Temporary domain: *etsrv1.pi.infn.it*

  - We can directly manage resources and services

  > **Proposed name:**
  > *etsitedata*

- Currently migrating system users and home dirs

- Fully containerized **JupyterHub server tested and working**

  - Available at https://etsrv1.pi.infn.it:8443

- **Web application** is being updated to enable user management on AlmaLinux 9

  > **Are there any other required services?**

- We will work on **FDSN web service** in the near future

- **Suggestions and feedback are appreciated**