Sviluppo del PoC di computing model per piccoli/medi esperimenti di astroparticelle

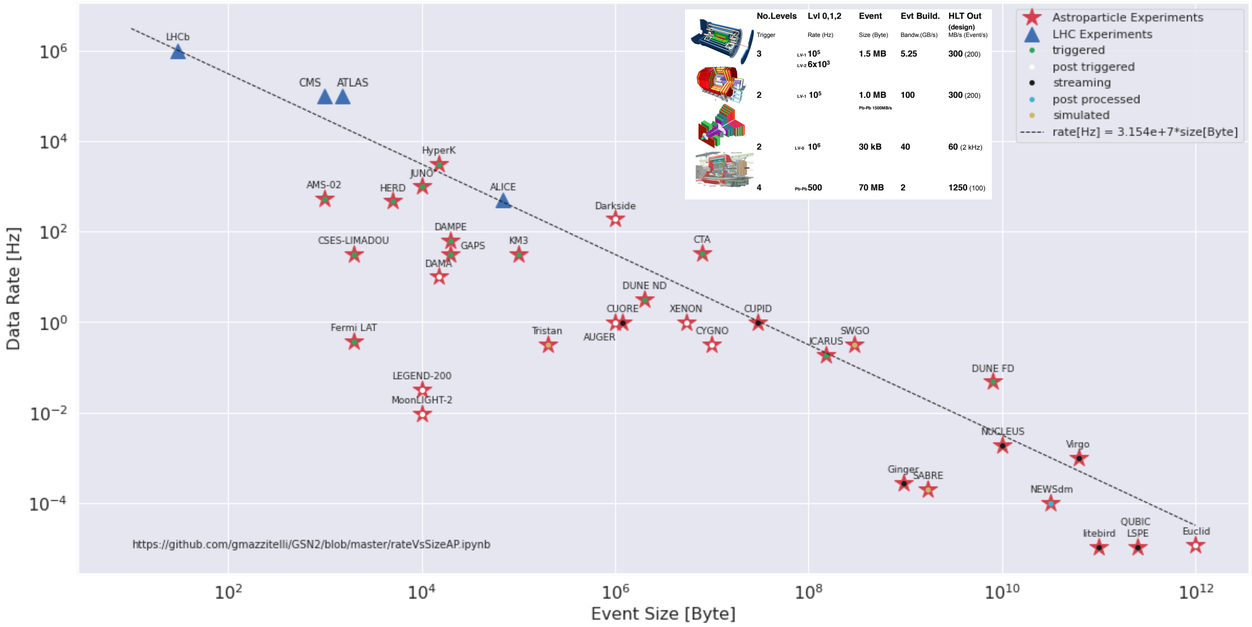
G. Mazzitelli et al. - Workshop sul Calcolo nell'INFN: La Biodola, 26 - 30 maggio 2025

agenda

- astro-particle use case requirements and futures
- PNRR opportunity
- the CYGNO PoC
- the PSGE/PNRR - Pipeline optimization for space and ground based experiments project
- future application and challenges
- conclusion

astroparticle experiments

exploiting CYGNO experience



bigger rather than faster!

astroparticle experiments are characterised by having a **different throughput** respect to typical HEP experiments, anyhow following a scaling law that underline how are anyway demanding in the overall process.

astroparticle experiments features:

- **unique** and **unrepeatable** data (ex. ultra high cosmic events) constraint on uptime/dead-time
- data could be acquired in **difficult and extreme conditions** (ex. space, under water ice, etc) conditioning the possibility of interventions and changes in the setup
- **templates and montecarlo** are needed not only to evaluates systematic but also to identify “candidates” of events. (ex OG, cosmic ray shower, etc) with large request of computing resources
- for many experiment data need to often to be **re-calibrated and reconstructed many times** with discontinuity and peak in the usage of computing resources

CSN2 2023 Computing Survey https://github.com/gmazzitelli/GSN2/blob/master/rateVsSizeAP.ipynb

spoke2 - WP3 - PSGE

Flagship 2.3.6 - PSGE (Pipeline optimization for space and ground based experiments)

Spoke 2

WP2.1: Theoretical Physics

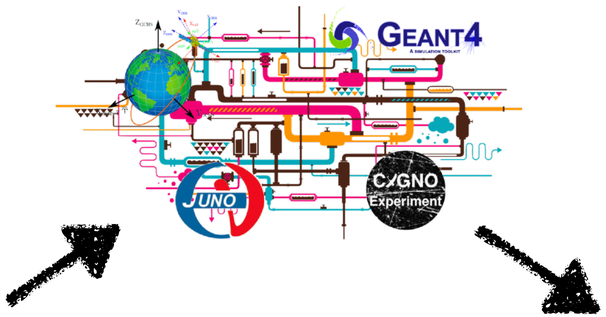
WP 2.2: Experimental High Energy Physics

WP 2.3 Astroparticle Physics and Gravitational waves

WP 2.4 Boosting Computational performance

WP 2.5 Architectural Support

WP 2.6 Cross–domain Initiatives

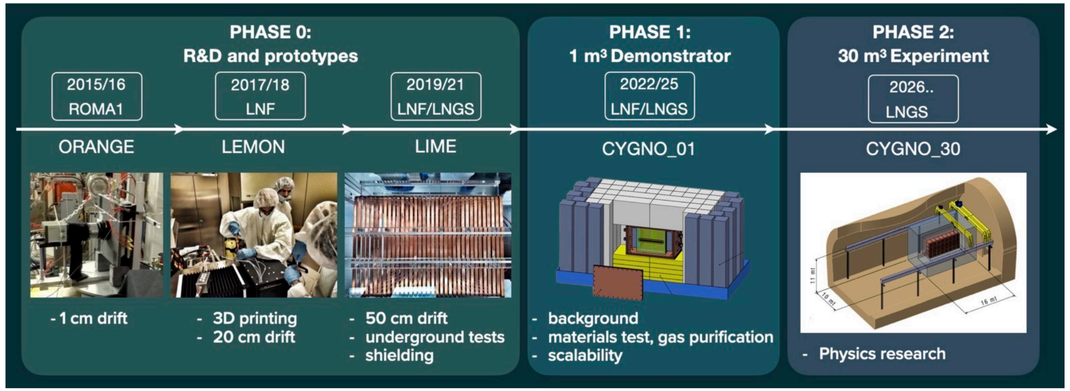


- Space Gravity Missions, that requires order of 100vCPU divided in O(5) VM to test the integration of current gravity mission scripts in portable and optimised containers, and to optimize them to run in cloud environment
- JUNO and similar experiments, that requires order of 100vCPU divided in O(5) VM to test the integration of JUNO da processing in portable and optimised containers
- **CYGNO and small/medium experiments in the astroparticle physics community,** that requires order of 800vCPU divided in O(200) VM to test the integration of CYGNO CM, data processing in portable and optimised containers and deploy it for the astroparticle community offering set of standard tools for data management and data processing by means of batch system on cloud.
- GEANT4 simulations in HPC environments, that requires 50000 core/hours on HPC CPU to test and deploy a GEANT4 based BoGEMMS-HPC framework for multi-treading and multi-node simulation
- - 100TB disk, 200TB tape

Tutte le risorse sono state assegnate dal RAC, 800+100 per CSN2

R&D/CYGNO use case

- **objective**: CYGNO, leveraging advances in CMOS technology, aims to develop a large gaseous **Time Projection Chamber** (TPC) for the detection of directional **Dark Matter** and **Solar Neutrinos**.

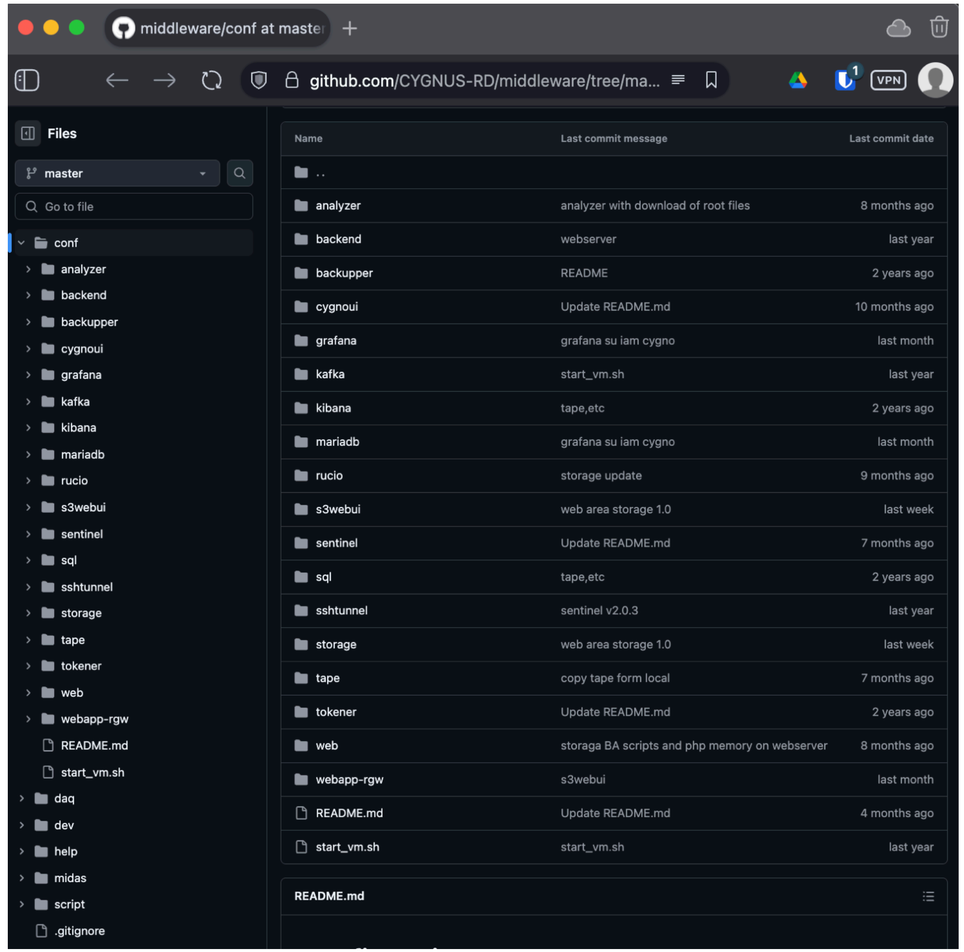


- **requirements**: the computing model requirements align with those typical of **small- to medium-scale** astro-particle experiments, including:
  - DAQ throughput of ~10 Mb/s, along with slow controls
  - Data management (storage, replicas, tape, etc), online reconstruction / reprocessing (recalibration and reconstruction)
  - Comprehensive analysis and simulation tools/facilities
- **approach**: utilizing the INFN Cloud and a **containerized solution,** ensuring **portability, reproducibility, scalability, isolation and security**. Additionally, it integrates **heterogeneous** resources to efficiently handle peak demand.

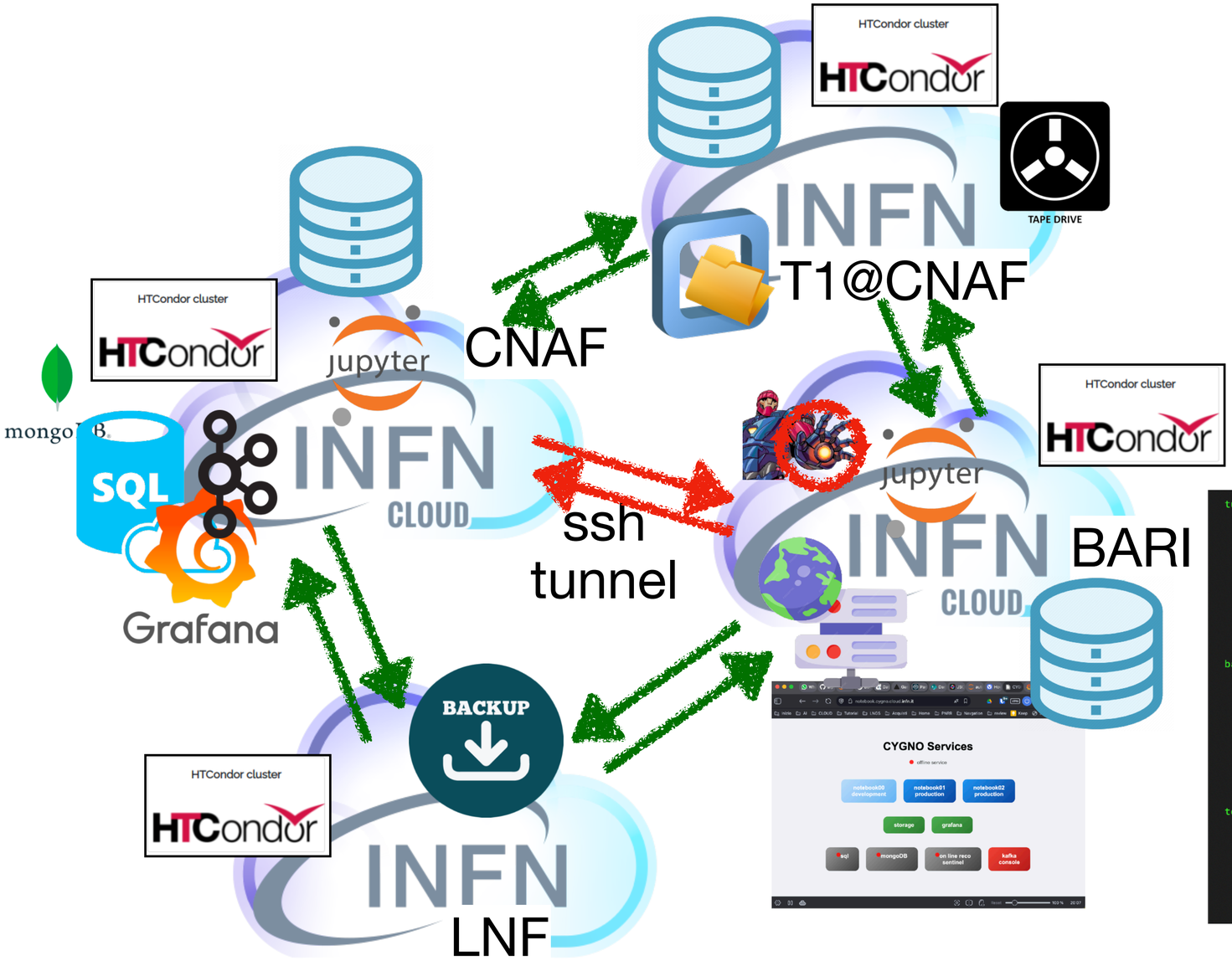
https://github.com/CYGNUS-RD/middleware/

approach

- developing a set of **Docker Compose** services for information exchange, deployed on VMs across **various INFN Cloud sites**.
- **decoupling services into modular units** (e.g., token, rclone, CVMFS, ssh tunnels, DB, tapes, etc.)
- providing a Jupyter-based **user interface** for analysis and simulation development, offering access to experiment **S3 storage** and their **heterogeneous computing resources**.



interconnections & data flow



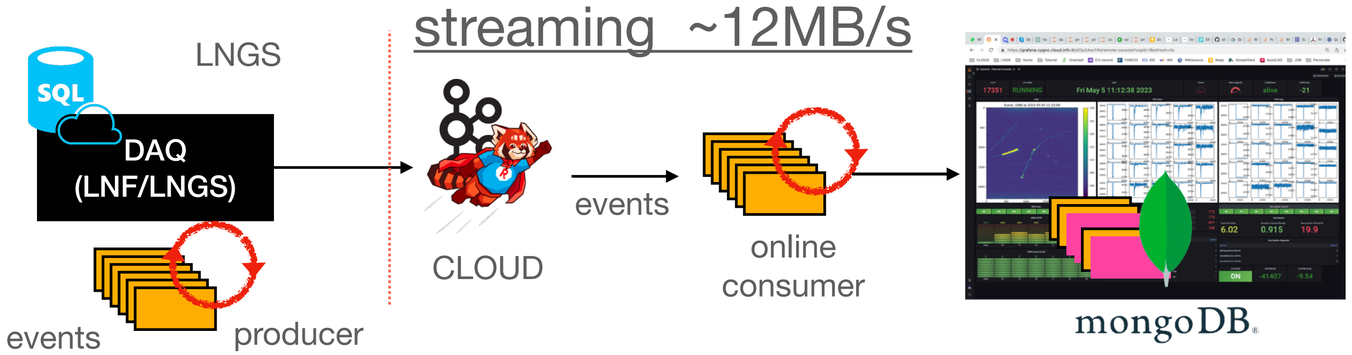
- ensuring communication via ssh tunnels among sites
- ensuring access to heterogeneous resources
- data backup on tape@CNAF
- DB backup on S3 storage
- notebook backup @LNF

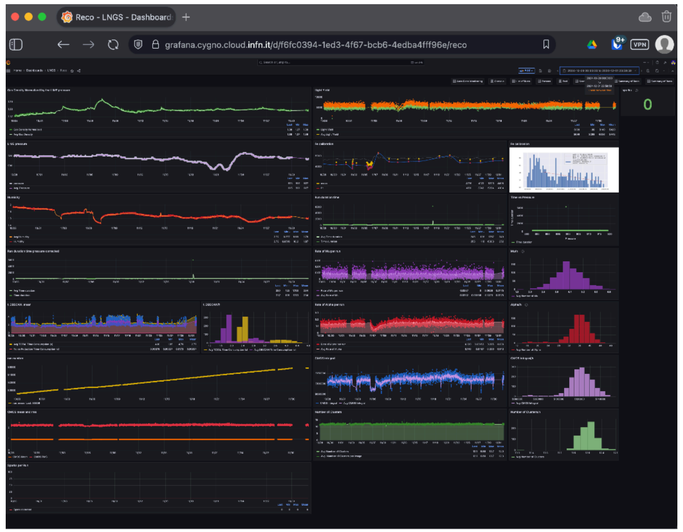
```
tunnel2:
  restart: always
  image: gmazzitelli/sshtunnel
  container_name: tunnel_kafka
  environment:
     REMOTE_PORT: 80
     REMOTE_APP_NAME: localhost
     LOCAL_PORT: 80
     USER: mazzitel
     REMOTE_IP: 131.154.98.101
  volumes:
     - /root/.ssh/daq_id:/root/.ssh/id_rsa

backupper:
  restart: always
  image: gmazzitelli/backupper:v0.1
  container_name: backupper
  environment:
     BCK_START_NAME: 'grafana'
     IAM_CLIENT_ID: ${IAM_CLIENT_ID}
     IAM_CLIENT_SECRET: ${IAM_CLIENT_SECRET}
  volumes:
     - /var/lib/docker/volumes/grafana_dbdata/_data/:/root/data/
     - /tmp/:/tmp/

tokener:
  image: gmazzitelli/tokener:v0.2
  container_name: tokener
  environment:
    REFRESH_TOKEN: ${REFRESH_TOKEN}
    IAM_CLIENT_SECRET: ${IAM_CLIENT_SECRET}
    IAM_CLIENT_ID: ${IAM_CLIENT_ID}
    IAM_TOKEN_ENDPOINT: ${IAM_TOKEN_ENDPOINT}
    SCOPES: ${IAM_SCOPES}
```

```
COMMAND                  NAMES
"/run.sh"                grafana
"/opt/startup.sh"        backupper
"/opt/startup.sh"        tunnel_sentinel
"/opt/startup.sh"        tokener
"/opt/startup.sh"        tunnel_kafka
"docker-entrypoint.s…"   mango_db
"/docker-entrypoint.…"   pma_mdb
"/opt/startup.sh"        backupper_sql
"docker-entrypoint.s…"   mariadb
```

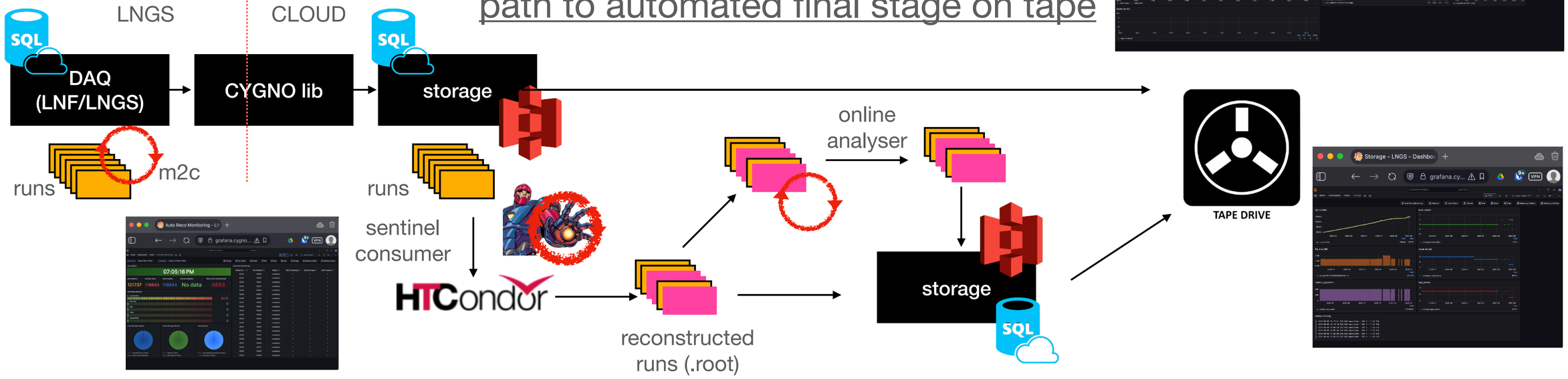
data pipeline

streaming/data management - reconstruction/online analysis

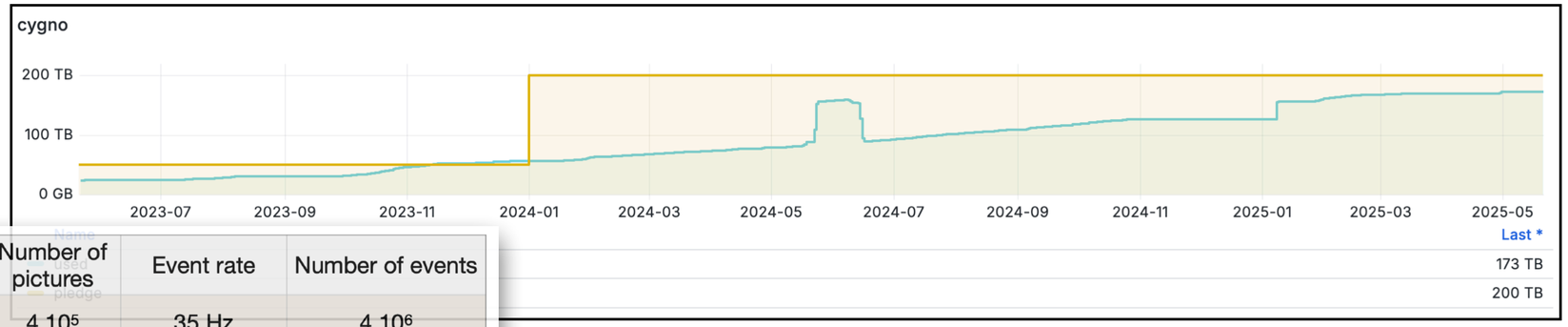




full data pipeline: reconstruction/online analysis (~10 minutes)
path to automated final stage on tape



proto runs storage/tape data flow



	Time slot	Number of pictures	Event rate	Number of events
RUN 1: No-shielding	3 Nov 2022 - 15 Dec 2022	4 10^5	35 Hz	4 10^6
RUN 2: 4 cm Cu shielding	15 Feb 2023 - 15 March 2023	4.5 10^5	3.5 Hz	5 10^5
RUN 3: 10 cm Cu shielding	5 May 2023 - 16 Nov 2023	1.6 10^6	1.5 Hz	7.3 10^5
RUN 4: 10 cm Cu + 40 cm water shielding	30 Nov 2023 - 31 March 2024	2 10^6	1.0 Hz	6 10^5
RUN 5: 10 cm Cu shielding (neutron flux measurements)	17 May 2024 - 1 Dec 2024	12 10^6	1.5 Hz	5.4 10^6

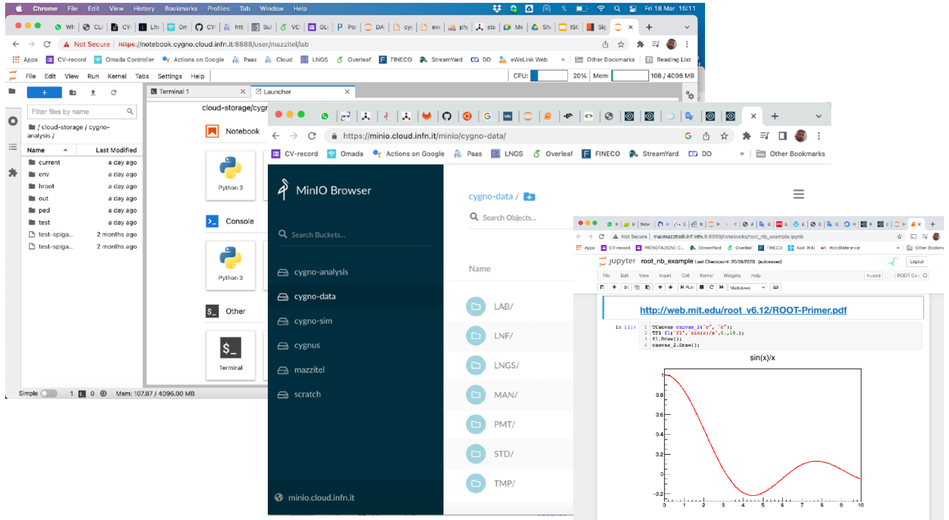
Special data takings

AmBe for Nuclear Recoils	2-4 Aug 2023	2 10^5	0.04 Hz of NR	2.5 10^3 NR
^{241}Am for Electron Recoils	7-16 Nov 2023	7 10^5	50 Hz	10^6
AmBe for Nuclear Recoils	5-15 Dec 2024	6 10^5	0.04 Hz of NR	7.0 10^3 NR

notebook v2

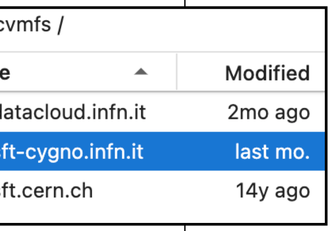
generalizing CYGNO WS analysis and simulation UI

Working Station for CYGNO experiment

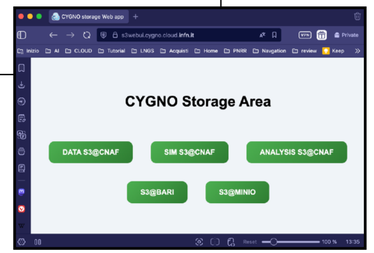


- **upgrading** base UI OS/python/libs
- **detaching process** in multiple services composed docker
- **detaching UI**, by installing all needed programs and experiment software on CVMFS datacloud.infn.it and sft-cygno.infn.it

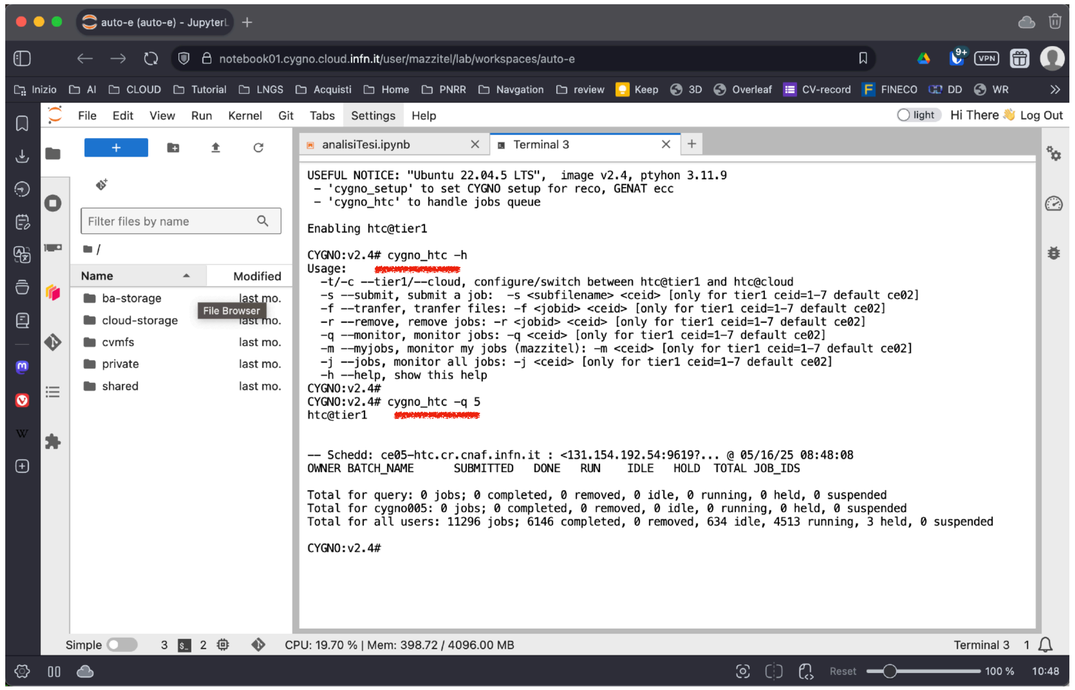
	V1	V2	comments
jupyterhub image	Centos 7 python 3.6	Ubuntu 22.04.5 LTS python 3.11.9	Thanks to G. Vino
base software	image based	CVMFS	Thanks to WP6
libraries (python/C)	image based	CVMFS	
experiment software	image based	CVMFS	
configuration	image based	CVMFS	
notebook storage	local (VM)	distributed (CANF/BARI)	work in progress Tanks to WP2
experiment storage	MINIO (backbone)	MINIO/S3@Ba/ S3@CNAF	
HTC	token based (limited to cloud)	token based	TIER1/on demand HTCondor on cloud
storage webapp	minio	rclone/rgw	



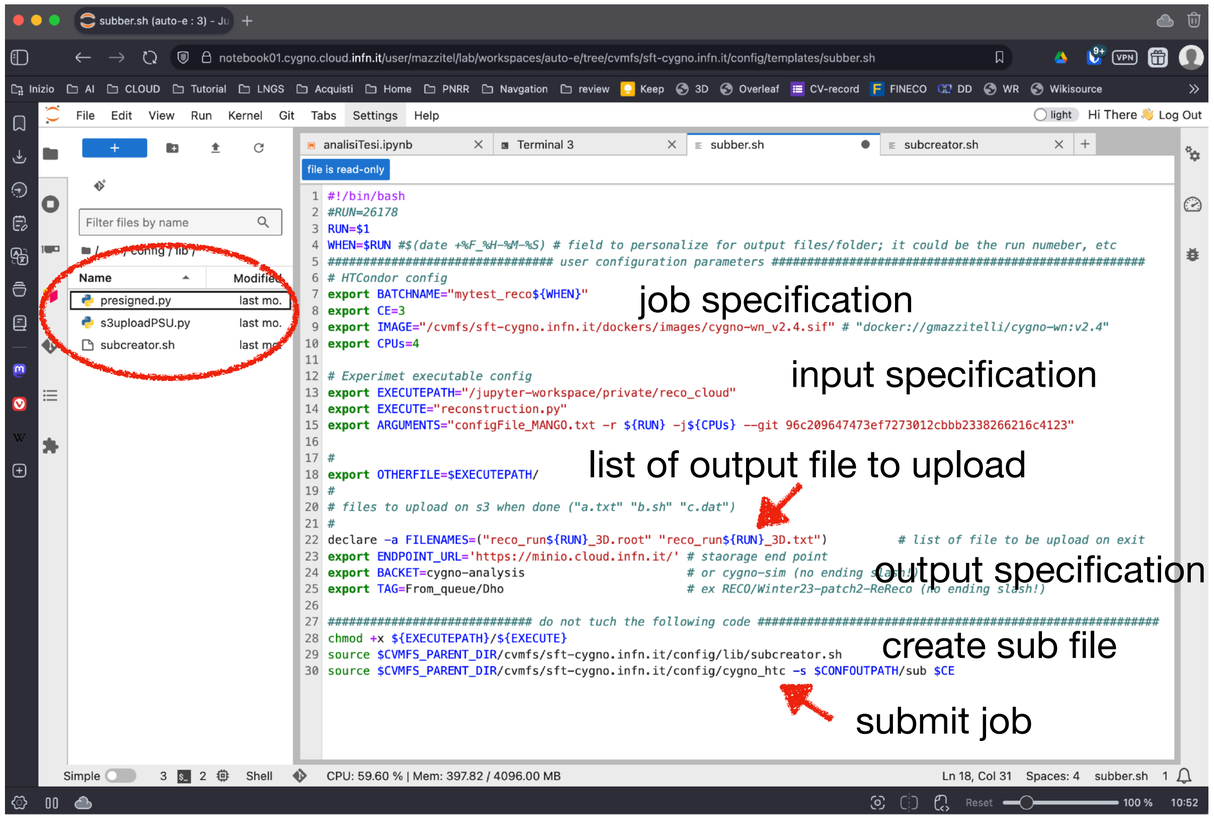
https://github.com/gmazzitelli/notebook00/releases/tag/v2.4



HTCondor v2



- ensuring long life job submission increase the **token lifetime**
- **pushing large file on storage** as soon as possible in order to be not ‘cleaned’ because of large size
- generating **prisigned url**/sub file within templates script for the experiment community



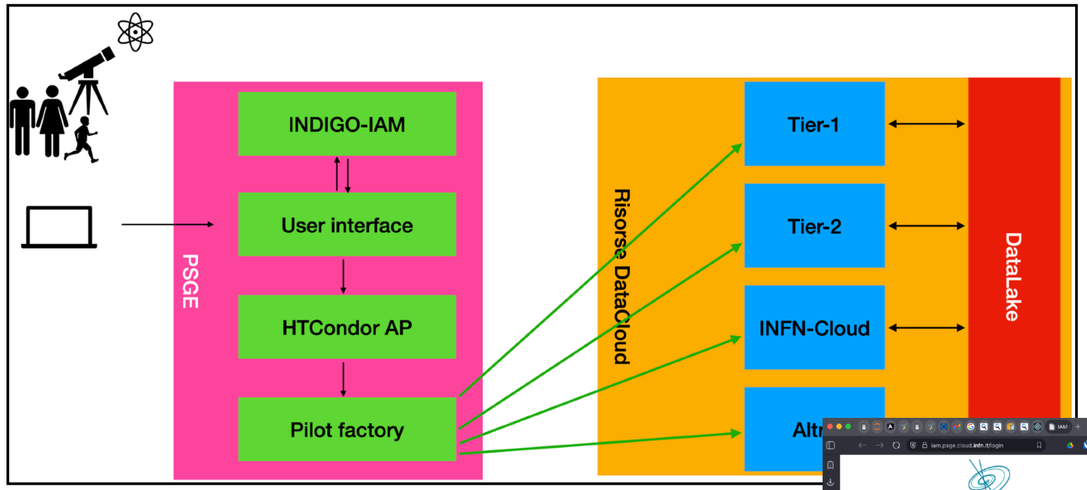
- experiment software and **executing environment** loaded from CVMFS

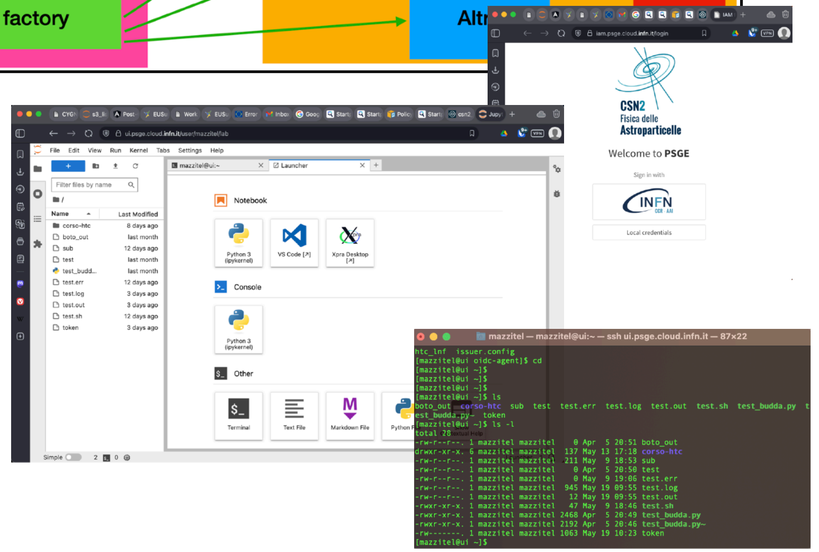
https://github.com/CYGNUS-RD/mycondor?tab=readme-ov-file#submit-job-to-cnaf

PSGE

shared pledge for the astro-particle community

- **objective:** to handle small pledge requests and any potential calculation spikes
- **ID&Auth: https://iam.psge.cloud.infn.it/**
  Admin GdL Calcolo CSN2
- **UI: https://ui.psge.cloud.infn.it/** - notebook, VS code, remote desktop, remote ssh login
- **HTCondor** tier2-ce2.lnf.infn.it (https://github.com/ICSC-Spoke2-repo/PSGE/blob/main/ASTROCM/GuideJobSubmission.md)
- grafana **monitoring**: group usage share
- **storage access:** StoRM-WebDAV/Third Party Copy/RUCIO



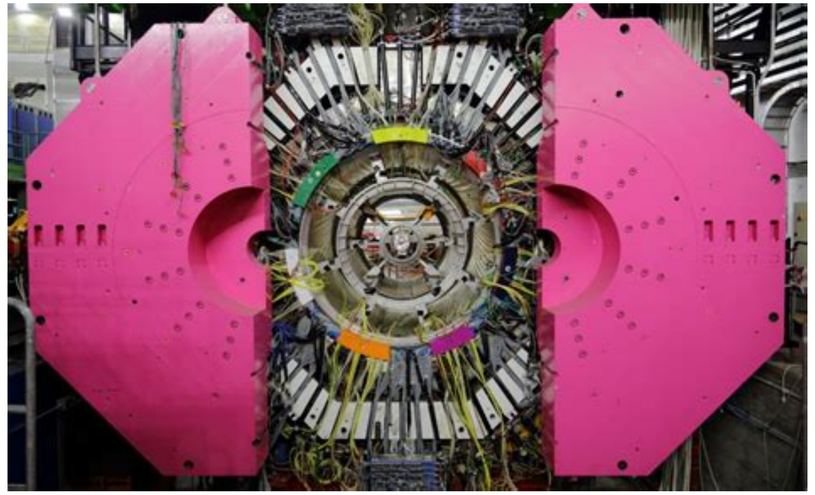


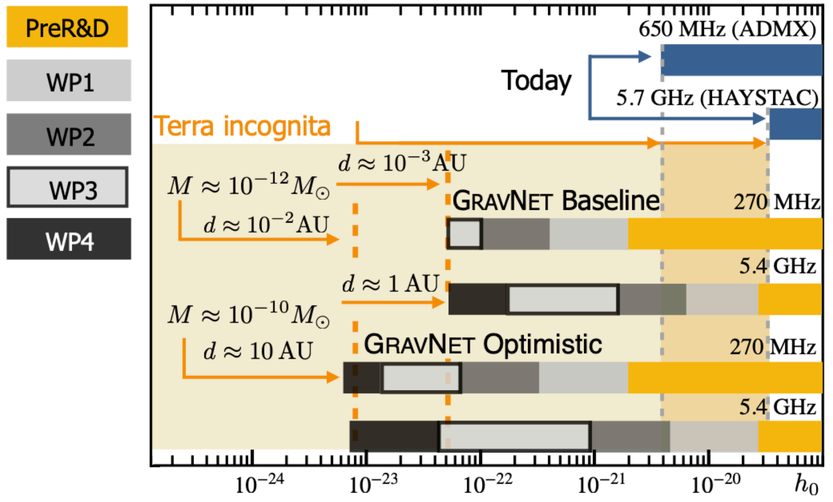
- ideazione e implementazione C. Pellegrino e I. Abritta Costa

CYGNO04/FLASH/GravNet

A Global Network for the Search for High Frequency Gravitational Waves

- **CYGN04** is under installation at LNGS, **with a O(10) increase of the throughput,** preliminary test are ongoing within the infrastructure designed.
- **FLASH/GravNet objective:** develop the **technique** for HFGW detector (MHz-GHz) to search **primordial black hole mergers, ultra-light dark matter interactions, early-universe cosmological events**
- **requirements**: throughput O(100-300MB/s); timing (ns); international collaboration with online/ofline requirement of common data reconstruction, analysis and simulation.
- The idea is to **exploit the work done** and preliminary test are ongoing





INFN Cloud users meeting

https://agenda.infn.it/event/42123/

- conclusions: define a set of **common packages**, available on the platform and adaptable to experiment resources; **integrate services** like RUCIO, CVMFS, Tier1/Cloud HTC access, etc., as soon as possible, creating **middleware standardization and shared know-how**;
- the Cloud continues to be the way to provide resources for **R&D** and the development of scientific/ management services for experiments, **effectively generating guidelines**.
- **CSN2**, due to the quantity and plurality of use cases, certainly remains **a key community for the development of the INFN Cloud**.

conclusion

- the development carried out with **CYGNO PoC** has allowed to:
  - **successfully manage** data acquisition for the CYGNO experiment;
  - **create a set of typical applications** for a computing model for small to medium-sized experiments (and perhaps also for the development of the infrastructure itself and a user community).
- the generalization of the CYGNO PoC is followed by both the PNNR/PSGE project and the **CSN2 common queue**, as well as future implementations for **challenging projects** like CYGNO04 and GravNet.
- the **support from WP6 and WP2** has been crucial for the development and implementation of applications.
- Engagement with the **user community** is growing, by sharing experiences, solutions, and future needs.