data handling of CYGNO experiment using INFN-Cloud solution

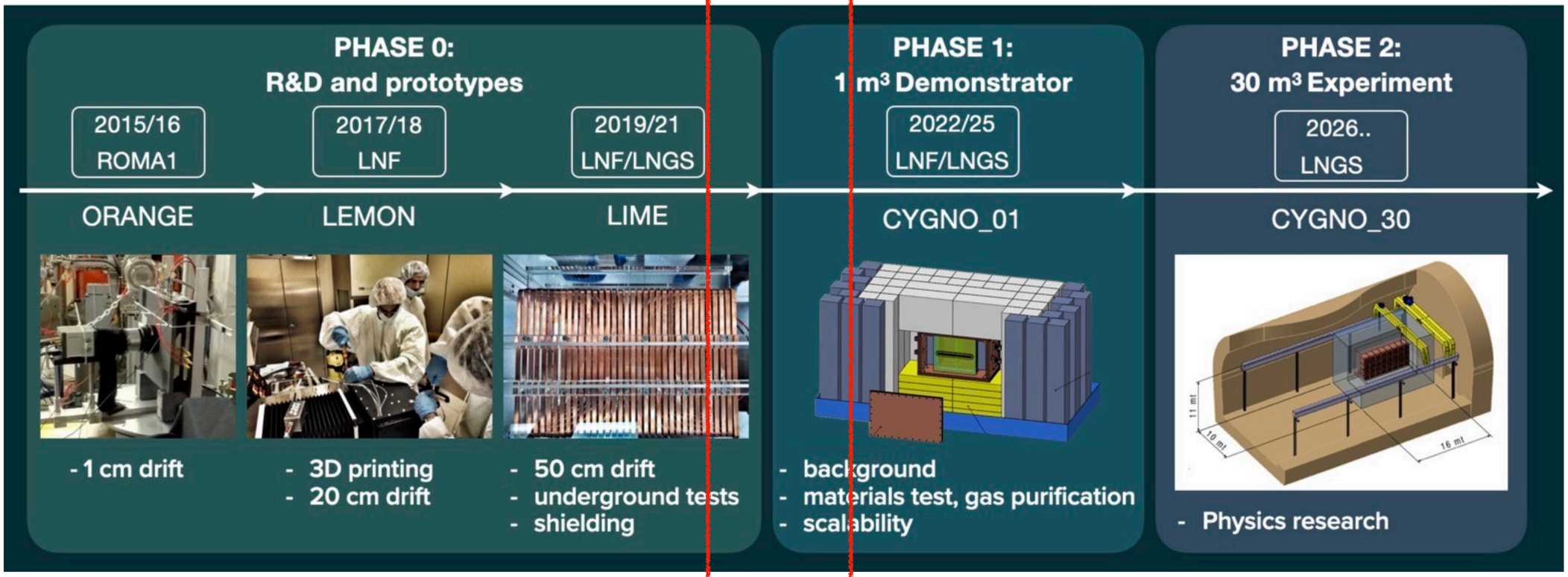
I.A. Costa, G. Mazzitelli on behalf of the INFN Cloud services and the CYGNO experiment

8–12 May 2023, Norfolk, Virgina USA

why (scientific objective)

CYGNO a large TPC for dark matter and neutrino study

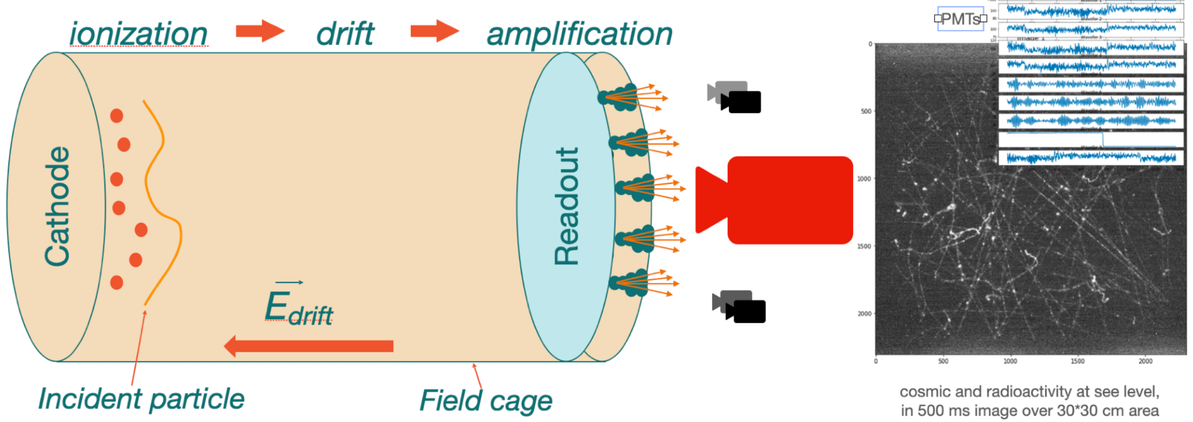
exploiting the progress in **commercial scientific Active Pixel Sensors (APS)** based on CMOS technology to realise a large **gaseous Time Projection Chamber** (TPC) for **Dark Matter and Solar neutrino search.**



low-energy (1–100 keV) nuclear recoils
10^5 beta/gamme rejection

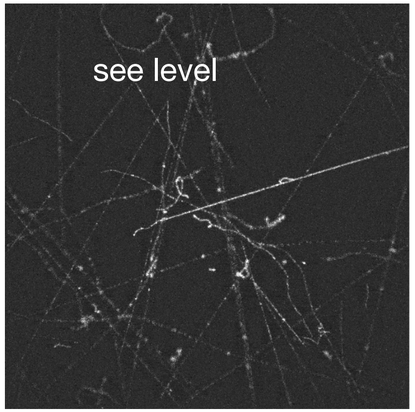
CYGNO prototype under test at LNGS

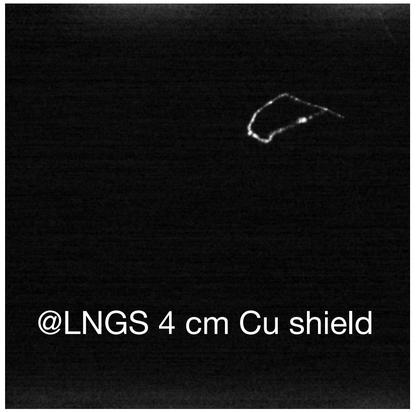
validating montecarlo expectation and testing HW/SW



cosmic and radioactivity at see level,
in 500 ms image over 30*30 cm area

- testing data taking, calibration and reconstruction, and analysis algorithms
- comparing data and monte-carlo full simulation
- validating ancillary system like gas system, DAQ, computing infrastructure

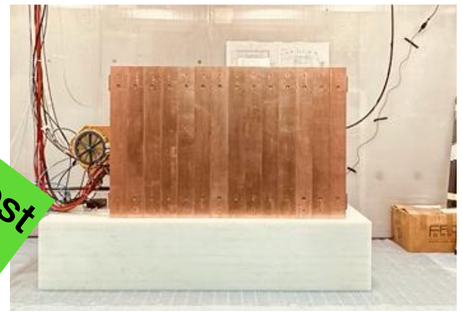




CYGNO project objective

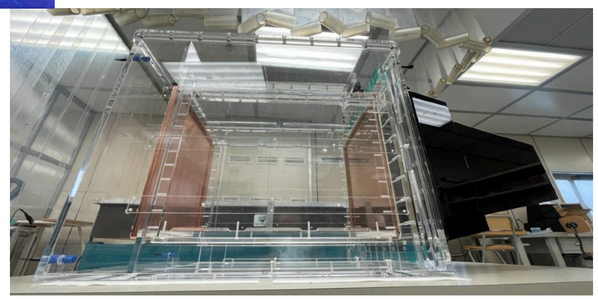
demonstrate the technique and feasibility of large scale detector

prototype



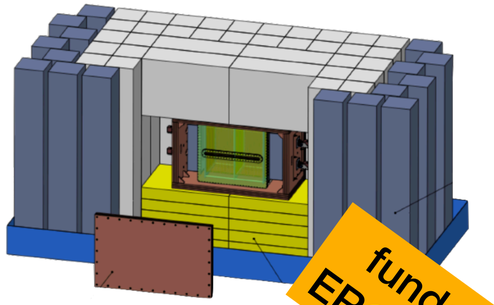
under test at LNGS

x 1 …



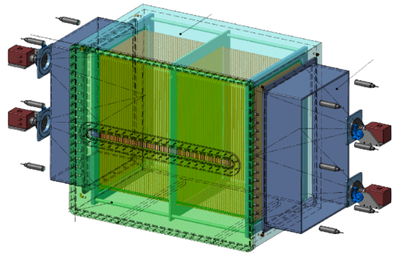
5*10 litres, 1 camera
10 MB/event 0.2—>0.01 Hz

CYGN04



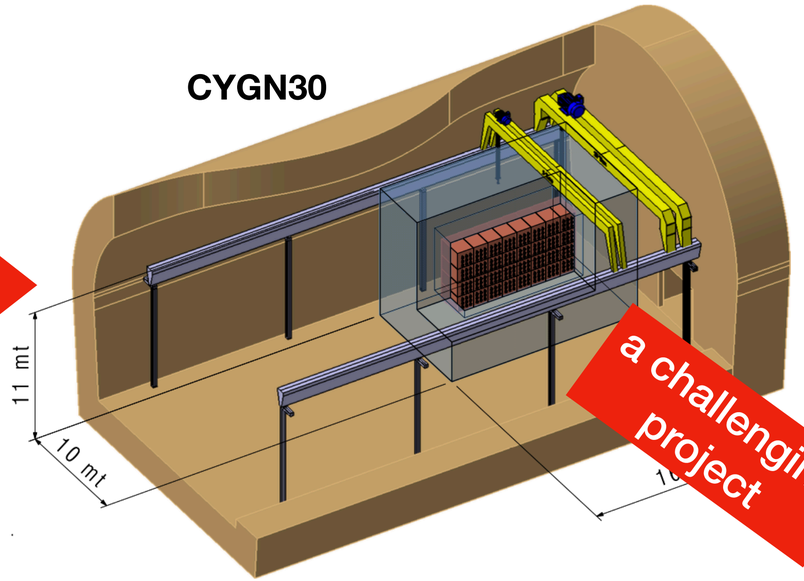
funded by ERC-INITIUM

x 2 …



1*10^3 litres, 4 cameras
45 MB/event (Hz ?)

CYGN30



a challenging project

x 2 …

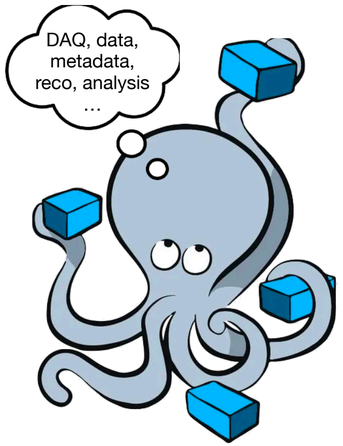
x N …

3*10^4 litres, > 100 cameras
4-5 GB/event (Hz ?)

the middleware CYGNO project

data management and online data validation and qualification

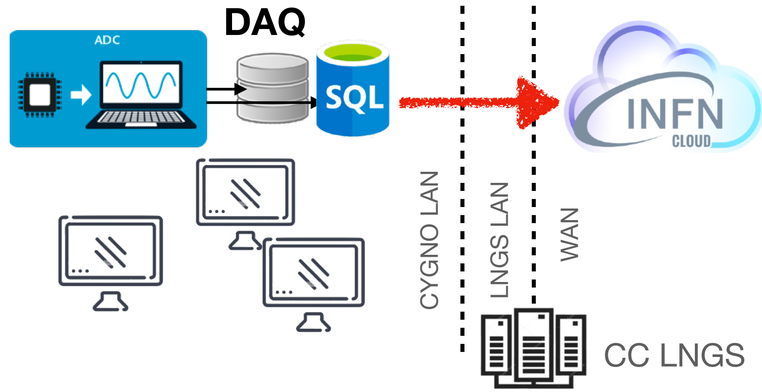
- experiment **data management**;
- experiment front end **metadata** production and management;
- slow/fast remote **experiment monitor** without access to LAN DAQ (shift workers from all over the world);
- online data **reconstruction** and **pre-analysis**;
- online data **validation** and **qualification**;
- high level/back end metadata production and management, **alarms and warnings** dispatcher also via discord experiment channel.



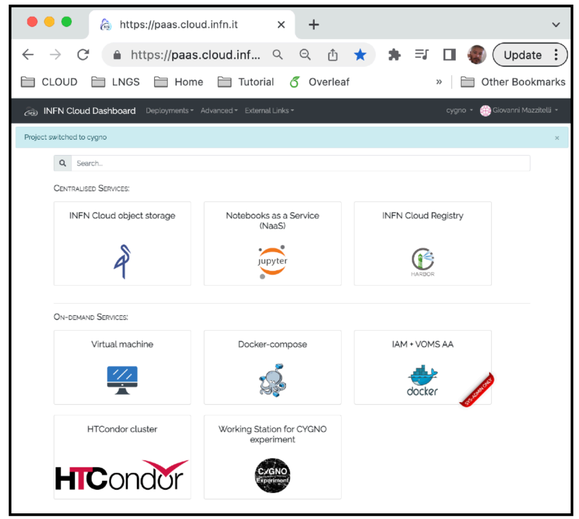
the INFN-Cloud infrastructure

data management and online data validation and qualification

- the CYGNO project is hosted in the **underground laboratory of LNGS** where it is recommended to have only the **minimum setup** necessary to collect data on a local buffer
- many experiments in the past decide to host their computing infrastructure in **CC of LNGS**
- In 2020 started the **INFN-Cloud project**, offering many services at PaaS/SaaS level, optimal to host our computing model, ensuring the characteristics of scalability, safety, reliability etc.
- in collaboration with the INFN-Cloud we **integrate and develop a sets of tools** for data management, analysis and simulation available at **user level** and accessible and exploitable to all the CYGNO international collaborators



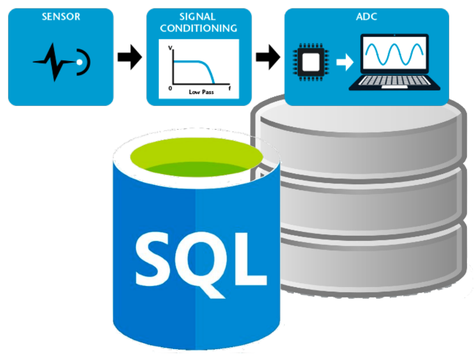
CYGNO-INFN cloud dashboard



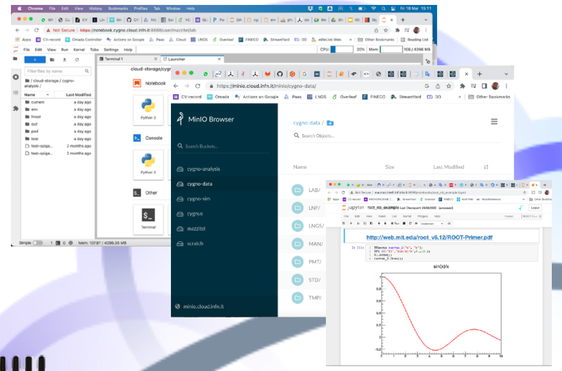
CYGNO… computing model

LNGS/LNF/…

DAQ



WEB GUI

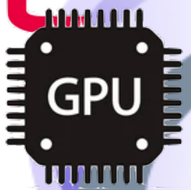


BACH

ANALISYS

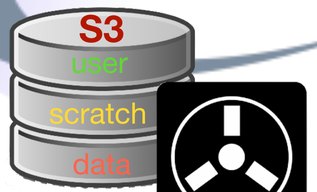
HPC

RESOURCES



INFN CLOUD

SIMULATION



TAPE DRIVE

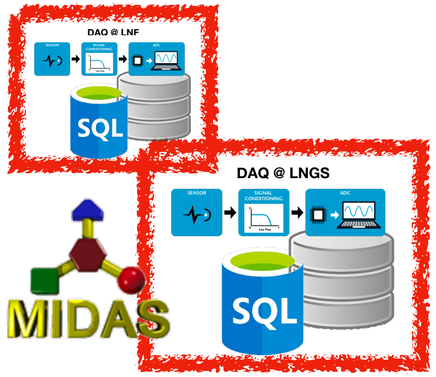
STORAGE/BACKUP

METADATA &
SLOW CONTROL

logical units, “composed” services

docker Compose

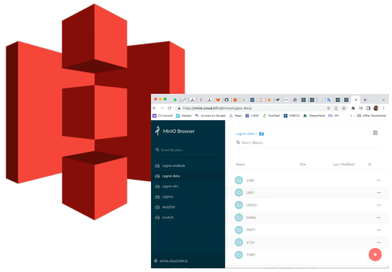
test and development setup at LNF



production setup at LNGS

Mariadb replica for metadata
sql.cygno.cloud.infn.it

S3 storage
minio.cloud.infn.it

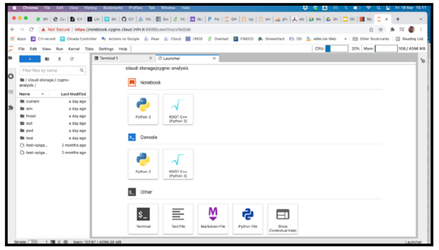


messaging
kafka.cygno.cloud.infn.it

Identity and Access Management
iam.cloud.infn.it

data and metadata monitor
grafana.cygno.cloud.infn.it

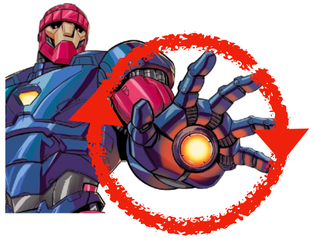
analysis and simulation web interfaces
notebook01.cygno.cloud.infn.it
notebook02.cygno.cloud.infn.it



batch queues
condor01.cygno.cloud.infn.it
condor02.cygno.cloud.infn.it

backup
tape.cygno.cloud.infn.it

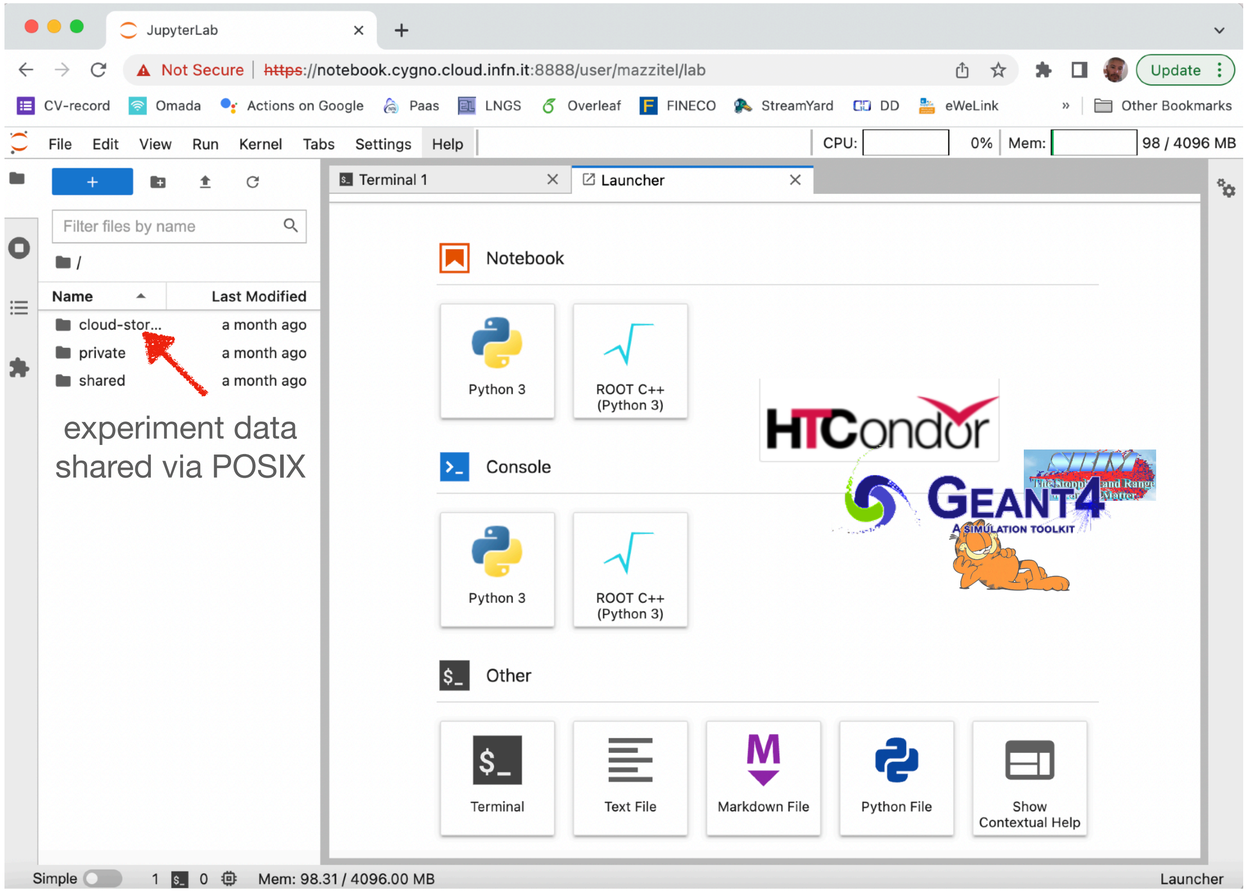
pre analysis and data quality
sentinel.cygno.cloud.infn.it



the user interface and services

multi-user platform integrated with INDIGO IAM authentication and authorisation, batch system, analysis and simulation software

- the tool is base on “**Dynamic On Demand Analysis Service (DODAS)**” project that allows the integration of cloud storage for persistence services with **analysis** (python/root/ecc) and **simulation software** (GEANT/GARFILD/ecc).
- **notebooks/consoles** for scripting in python and root; **terminals; editor; data access via POSIX** (FUSE simulated)
- **batch system on demand**: from the interface the experiment HTCondor queues can be reach to submit and control job
- **user interface** and **work node** software running on the queues is managed by the experiment and can be easily update on user request.



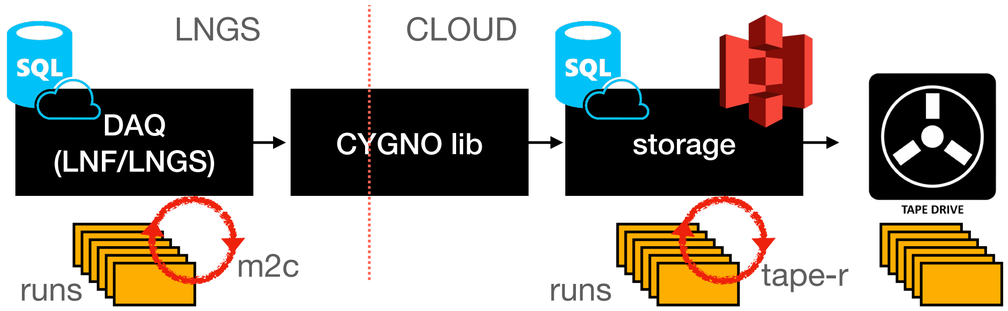
WHAT NEXT

integrate **CVMFS:** scalable, reliable and low-maintenance **software distribution** service

data management

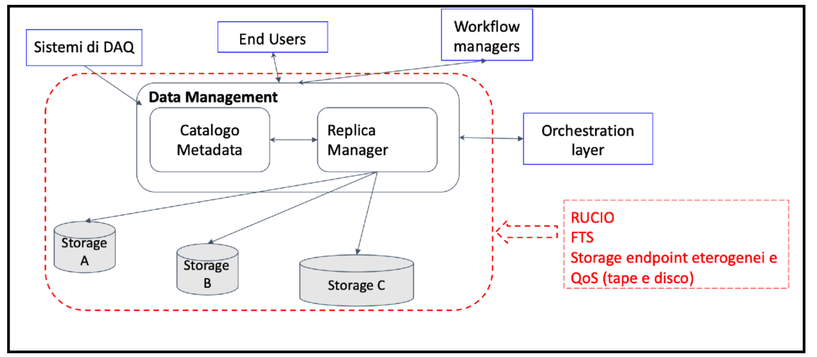
the “tape-r”

- data by means **m2c process**, bunched in runs, are copied on **S3 object storage**, as well as **metadata**, locally stored and replicated on cloud MariaDB;
- a few second after the run is close is available for **full reconstruction** on the cloud HTcondor queue and can be **download** with various tools (web, rest api, POSIX, ecc);
- the **“tape-r” process** replicate **data** on tape and update **metadata** of the run status**;**
- TAPE **@CNAF token based access** in the next future is going to be integrated in **RUCIO** as cloud services for more complete and generalised data management system



WHAT NEXT

data replica dashboard



data reconstruction pipeline

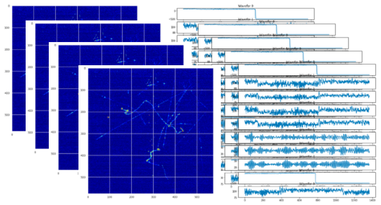
offline/online process

data preparation
zero suppression
noise filters
optical corrections

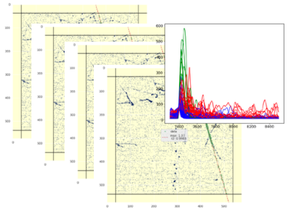
reconstruction
clustering &
superclustering
PMT info evaluation

sCMOS images

PMT waveforms

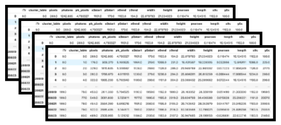


RAW events
calibration source, radioactivity,
cosmic rays, Dark Matter wind…



data “hits”

EVENTs data sets

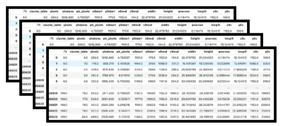


clusters features
PMTs features

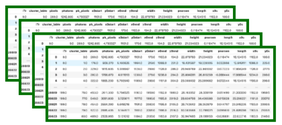
noise filter: median filter
optical correction: vignetting, optical distortion
superclustering: Geodesic Active Contour (GAC)

post-processing
calibrations
high-level analysis

multi-EVENTs data sets



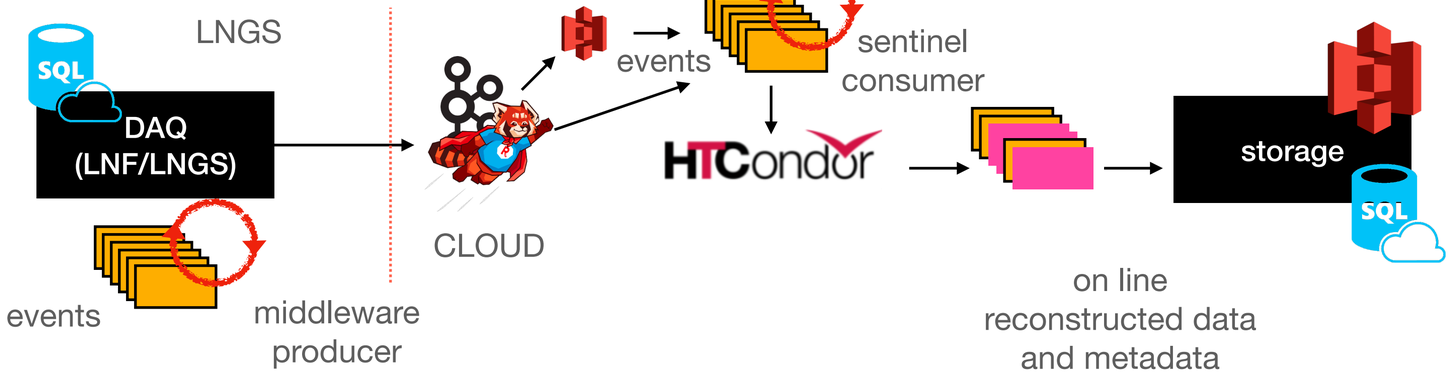
calibrated data sets



user analysis
physical outcomes

online data reconstruction

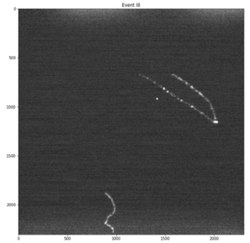
the “sentinel”

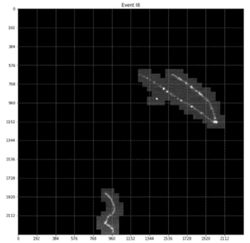


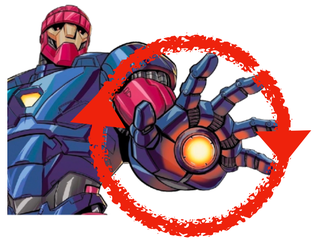
- parallel to run data management, sigle events are send to cloud by means of **kafka producer**
- the **sentinel** process consume data **parallelising the events reconstruction** on the HTCondor queues
- data and metadata are the stored and **presented** for on line motoring

WHAT NEXT

optimise/scale architecture to completely be able to provide online reconstruction
implement **data compression** (triggerless ML/ GPU algorithms are under study)

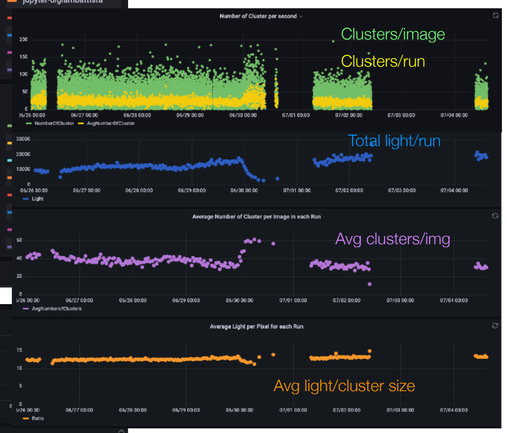


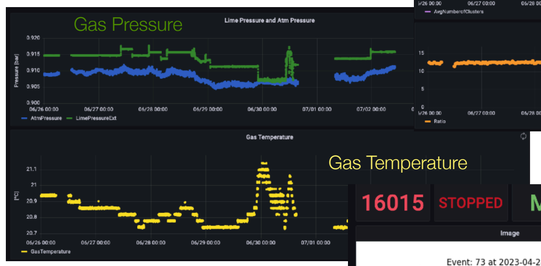




resources monitor

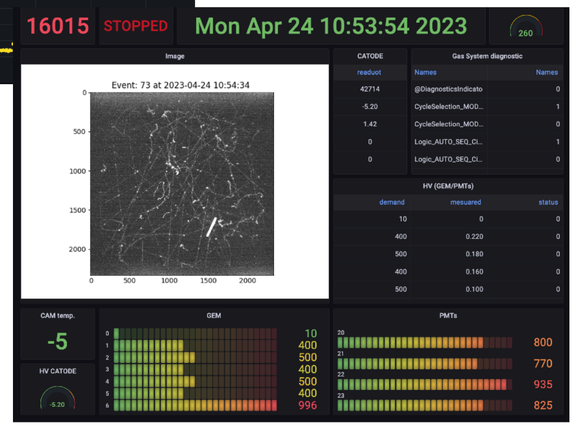
data quality monitor





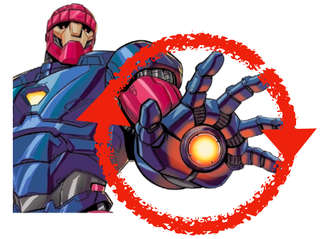
slow control monitor

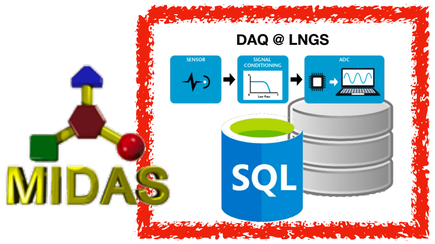
remore console



online experiment monitoring

the r-console



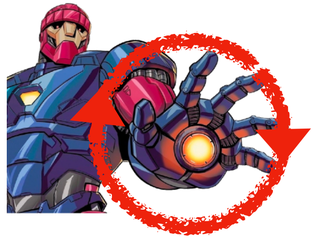


remote console
producer LNGS

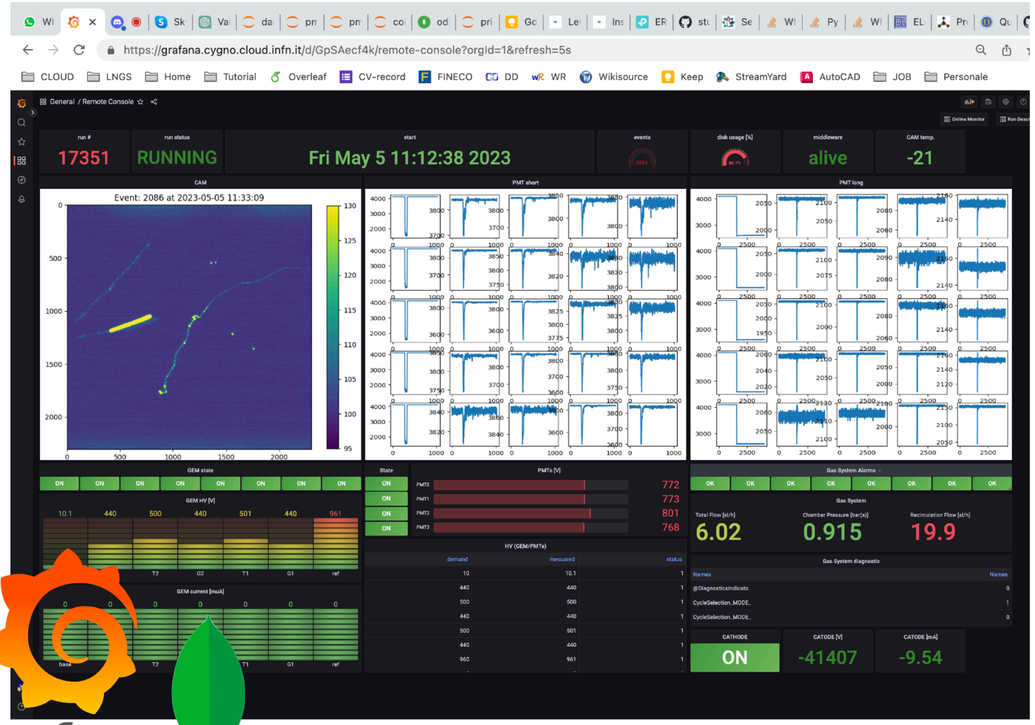
CLOUD

json ODB

json data
images



console
consumer

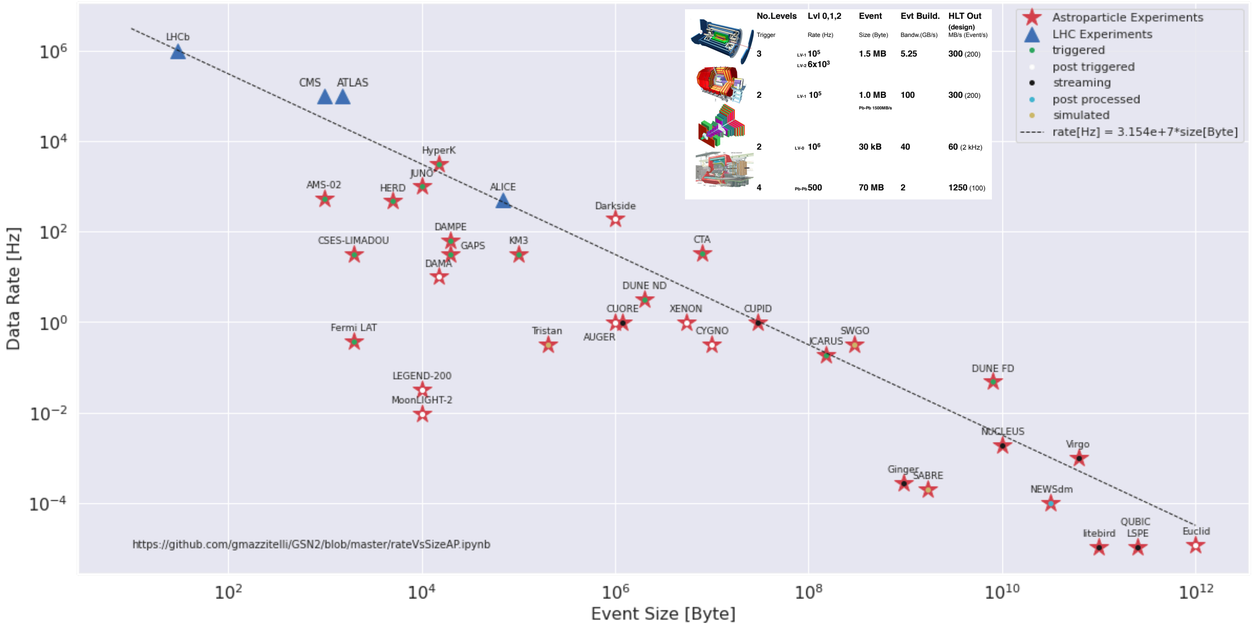


Grafana

mongoDB.

astroparticle experiments

exploiting CYGNO experience



bigger rather then faster!
astroparticle experiments are characterised by having a **smaller 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** whit discontinuity and peak in the usage of computing resources

conclusion

- CYGNO project started and a **technical run** is on going to test all the needs for the full **demonstrator** starting in 2025. if successful a **full scale detector** for physics will follow, who's characteristics will be challenging also from the computing point of view;
- a setup based on the **INFN-Cloud** of full **computing services and data handling tools** for CYGNO experiment has been setup, is **running and show appropriate performance** - see also other technical talk @CHEP23;
- **optimisation and further tools** (orchestration, parallelisation ecc) also are under implementation also due to the support of Next Generation EU program (PNRR - Piano Nazionale di Ripresa e Resilienza) and the Italian center for super computing (ISCS) - see other talks @CHEP23;
- the CYGNO use case is one of the seed that can be easily **generalised** to develop the **computing model of many small/medium experiments** in the astroparticle Italian community, reducing resources requests, costs, energy and environments impact, improving security, ecc. ecc.

thank you!

I. A. Costa and G.Mazzitelli on behalf of
CYGNO/Users: D. Amaro, R. Antonietti, E. Baracchini, L. Benussi, S. Bianco, F. Borra, A. Calanca, C. Capoccia, M. Caponero, D. S. Cardoso, G. Cavoto, R. J. de Cruz Roque, I. A. Costa, E. Dané, G. Dho, E. Di Marco, G. D'Imperio, F. Di Giambattista, R. R. M. Gregorio, F. Iacoangeli, E. Kemp, H. P. Lima Júnior, G. S. P. Lopes, G. Maccarrone, R. D. P. Mano, D.J.G. Marques, G. Mazzitelli, A. G. Mc Lean, P. Meloni, A. Messina, M. Migliorini, C.M.B. Monteiro, R. A. Nóbrega, I. F. Pains, E. Paoletti, L. Passamonti, F. Petrucci, S. Piacentini, D. Piccolo, D. Pierluigi, D. Pinci, A. Prajapati, F. Renga, F. Rosatelli, A. Russo, J.M.F. dos Santos, G. Saviano, N. Spooner, R. Tesauro, S. Tomassini, S. Torelli
INFN-Cloud: S Stanlio, M Antonacci, C. Duma, D. Ciangottini, D. Spiga, C. Pellegrino

This work is partially supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU and by ERC-INITIUM-818744