Running Fermi-LAT analysis on Cloud: the experience with DODAS with EGI-ACE Project

Cristina Duma (INFN-CNAF), Daniele Spiga (INFN Perugia)
Diego Ciangottini (INFN Perugia), **Isabella Mereu (INFN Perugia)**
Luciano Gaido (INFN Torino), Marica Antonacci (INFN Bari)
Sara Cutini (INFN Perugia), Stefano Nicotri (INFN Bari)

ISGC 2022, March 23

Outline

- Fermi Gamma-ray Space Telescope
- Time domain astrophysics: FLT analysis
- EGI-ACE project
- DODAS
- FLT results
- Lessons learnt and next steps
- Summary

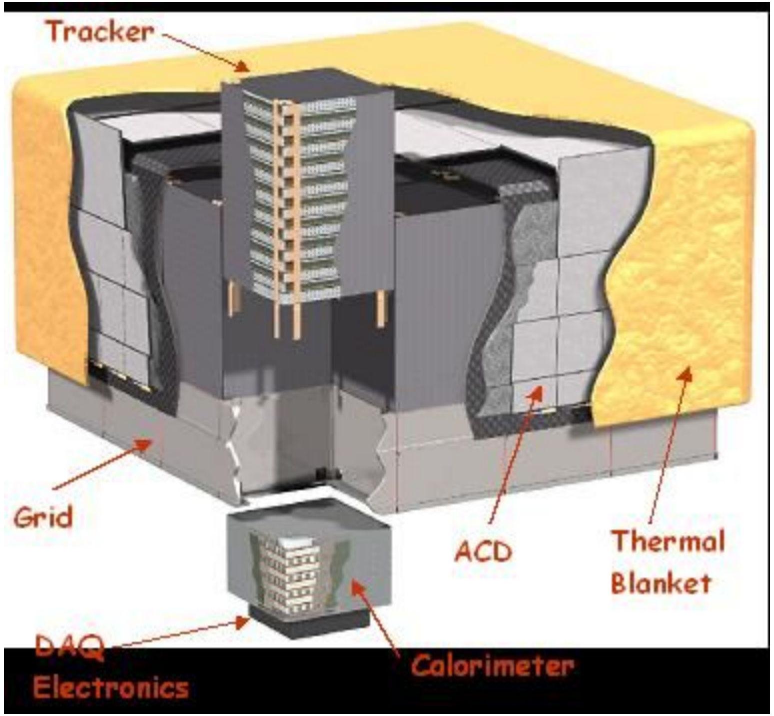
Fermi Gamma-ray Space Telescope

Large Area Telescope (LAT):
- 20 MeV to more than 300 GeV
- observes 20% of the sky at any instant
- Scan the whole sky daily

Gamma-ray Burst Monitor (GBM):
- 8 keV to 40 MeV
- observes entire unocculted sky

Launch: June 11 2008,NASA,
Orbit: circular, 565 km altitude

Fermi-LAT



Large Area Telescope (LAT): pair conversion detector

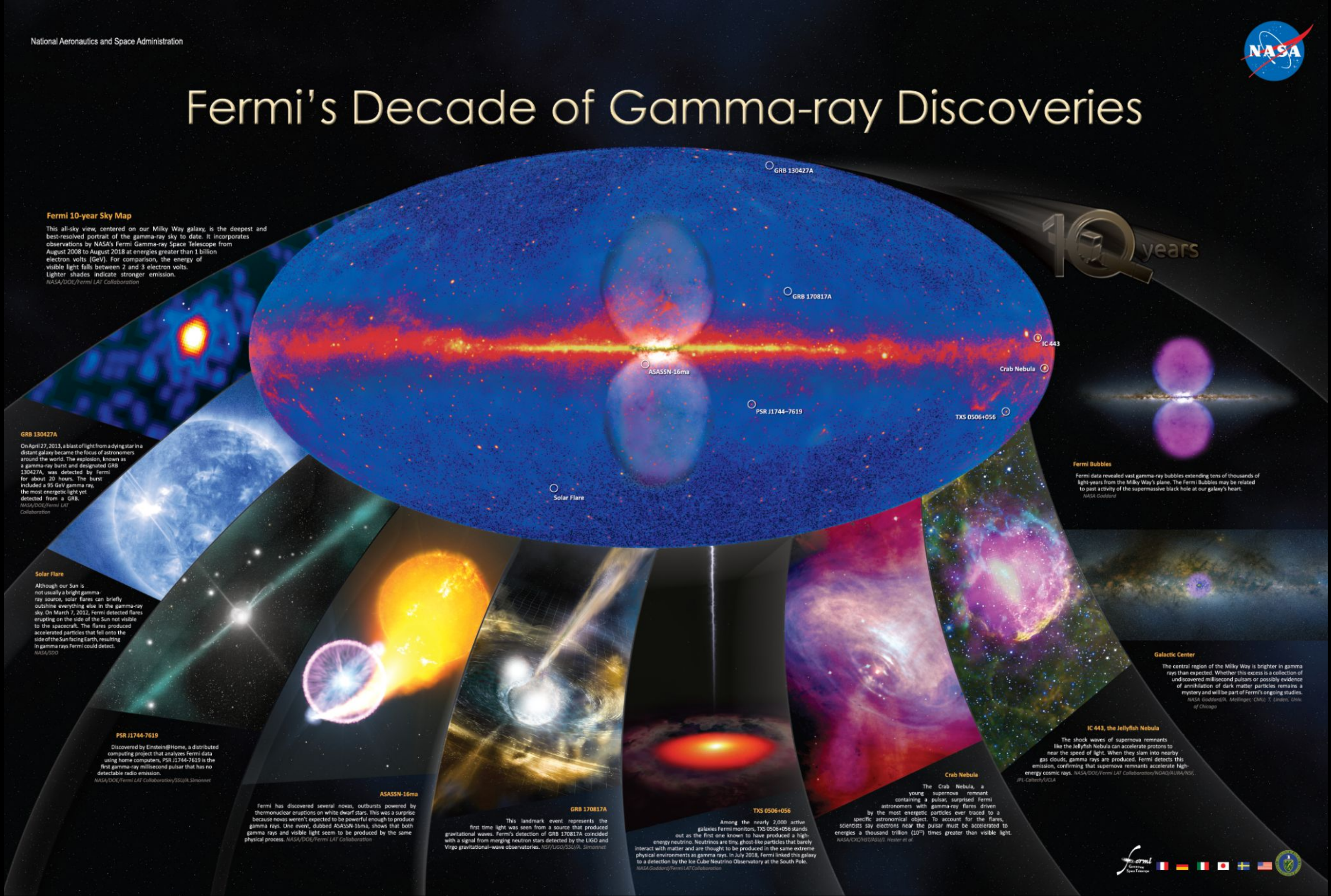
Energy Range: ~20 MeV to >300 GeV

FOV: ~20% of the sky at any instant

LAT trigger rate: can approach 10 kHz

On-board analysis reduction: ~400 events per second

Of these ~2-5 Hz are astrophysical photons


National Aeronautics and Space Administration
NASA
Fermi's Decade of Gamma-ray Discoveries
10 years
GRB 130427A
GRB 170817A
IC 443
Crab Nebula
ASASSN-16ma
PSR J1744–7619
TXS 0506+056
Solar Flare
Fermi 10-year Sky Map
This all-sky view, centered on our Milky Way galaxy, is the deepest and best-resolved portrait of the gamma-ray sky to date. It incorporates observations by NASA's Fermi Gamma-ray Space Telescope from August 2008 to August 2018 at energies greater than 1 billion electron volts (GeV). For comparison, the energy of visible light falls between 2 and 3 electron volts. Lighter shades indicate stronger emission.
NASA/DOE/Fermi LAT Collaboration
GRB 130427A
On April 27, 2013, a blast of light from a dying star in a distant galaxy became the focus of astronomers around the world. The explosion, known as a gamma-ray burst and designated GRB 130427A, was detected by Fermi for about 20 hours. The burst included a 95 GeV gamma ray, the most energetic light yet detected from a GRB.
NASA/DOE/Fermi LAT Collaboration
Solar Flare
Although our Sun is not usually a bright gamma-ray source, solar flares can briefly outshine everything else in the gamma-ray sky. On March 7, 2012, Fermi detected flares erupting on the side of the Sun not visible to the spacecraft. The flares produced accelerated particles that fell onto the side of the Sun facing Earth, resulting in gamma rays Fermi could detect.
NASA/SDO
PSR J1744-7619
Discovered by Einstein@Home, a distributed computing project that analyzes Fermi data using home computers, PSR J1744-7619 is the first gamma-ray millisecond pulsar that has no detectable radio emission.
NASA/DOE/Fermi LAT Collaboration/SSU/A.Simonnet
ASASSN-16ma
Fermi has discovered several novas, outbursts powered by thermonuclear eruptions on white dwarf stars. This was a surprise because novas weren't expected to be powerful enough to produce gamma rays. One event, dubbed ASASSN-16ma, shows that both gamma rays and visible light seem to be produced by the same physical process. NASA/DOE/Fermi LAT Collaboration
GRB 170817A
This landmark event represents the first time light was seen from a source that produced gravitational waves. Fermi's detection of GRB 170817A coincided with a signal from merging neutron stars detected by the LIGO and Virgo gravitational–wave observatories. NSF/LIGO/SSU/A. Simonnet
TXS 0506+056
Among the nearly 2,000 active galaxies Fermi monitors, TXS 0506+056 stands out as the first one known to have produced a high-energy neutrino. Neutrinos are tiny, ghost-like particles that barely interact with matter and are thought to be produced in the same extreme physical environments as gamma rays. In July 2018, Fermi linked this galaxy to a detection by the Ice Cube Neutrino Observatory at the South Pole.
NASA Goddard/Fermi LAT Collaboration
Crab Nebula
The Crab Nebula, a young supernova remnant containing a pulsar, surprised Fermi astronomers with gamma-ray flares driven by the most energetic particles ever traced to a specific astronomical object. To account for the flares, scientists say electrons near the pulsar must be accelerated to energies a thousand trillion (10^{15}) times greater than visible light.
NASA/CXC/HST/ASU/J. Hester et al.
IC 443, the Jellyfish Nebula
The shock waves of supernova remnants like the Jellyfish Nebula can accelerate protons to near the speed of light. When they slam into nearby gas clouds, gamma rays are produced. Fermi detects this emission, confirming that supernova remnants accelerate high-energy cosmic rays. NASA/DOE/Fermi LAT Collaboration/NOAO/AURA/NSF, JPL-Caltech/UCLA
Galactic Center
The central region of the Milky Way is brighter in gamma rays than expected. Whether this excess is a collection of undiscovered millisecond pulsars or possibly evidence of annihilation of dark matter particles remains a mystery and will be part of Fermi's ongoing studies.
NASA Goddard/A. Mellinger, CMU; T. Linden, Univ. of Chicago
Fermi Bubbles
Fermi data revealed vast gamma-ray bubbles extending tens of thousands of light-years from the Milky Way's plane. The Fermi Bubbles may be related to past activity of the supermassive black hole at our galaxy's heart.
NASA Goddard

Time domain astrophysics: particular Long-Term Transient Sources (FLT)

Fermi ⟶ monitors the full sky regularly

⟶ microsecond time resolution

Time-domain astrophysics for the nonthermal Universe is an integral aspect of all Fermi science.

Timescales that the Fermi instruments address range from sub-millisecond-long transients, like Terrestrial Gamma-ray Flashes, through day- and weeks-long transients from novae and AGN.

→1FLT: First Gamma-ray Long-Term Transient sources catalogue

GOAL: Identify transient and variable sources with monthly time scale that were not yet reported in other Fermi-LAT catalogs → possibly confused with the background over multi-year integration times

FLT Analysis Details (1)

14 years of data → **Billions** of astrophysical photons →~150 Gigabyte

Monthly time intervals + 15 days shifted monthly time intervals

FLT

1FLT: first decade of Fermi-LAT data 120 months + 120 15-day shifted months

iFLT: incremental, following 3 years→ 36 x 36 months

312 independent full sky analyzed

FLT Analysis Details (2)

Each sky was divided in 192 circular regions of interest (ROIs) centered on points defined by a HEALPix (https://healpix.sourceforge.io/) → ROIs (partially overlap) include events in cones of 15° radius about the center of the pixel.

To identify the candidate sources (seeds), we used a wavelet transform analysis (Damiani et al. 1997) using the *PGWave* tool (released within the official *Fermitool*)

We performed the analysis independently on each ROI and month deleting the seeds located at the border of the ROI (distance from the ROI center >10°) in order to avoid edge effects.

Collected roughly 1000 seeds per sky

FLT Analysis Details (3) → COMPUTING NEEDS

Maximum likelihood analysis is performed for each seed not-positionally coincident with known gamma-ray sources

→ **i.e. 200 seeds analyzed per months**

To perform the analysis, we used:

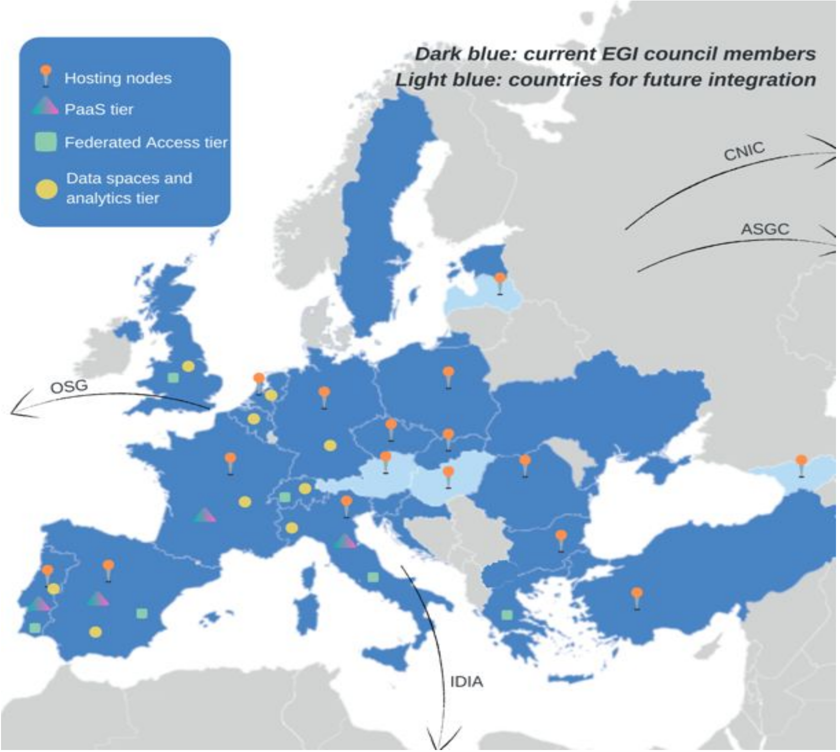
- the Fermitools package available from the Fermi Science Support Center: a suite of tools provided by the Fermi mission
- Fermipy software, a python package that facilitates analysis of data from the LAT with the Fermitools

FLT Analysis Details (4) → COMPUTING NEEDS

Maximum likelihood analysis is a procedure very time consuming **~0.5h per seed**→Using a single workstation: **1300 days of computing.**

Overall FERMI-LAT does not require a extremely large computing / storage resources however a dynamic and elastic system would allow to enable portability and to manage peak of usage - This is why we decided to try to port our system in cloud.

EGI-ACE (EGI Advanced Computing for EOSC)



- EGI Foundation as coordinator
  - 33 participating partners, 23 third-parties
- Consortium at a glance
  - 16 partners to deliver cloud, HTC and HPC resources
  - 12 user access and platform solution providers
  - 21 providers from 13 communities to deliver data spaces
  - 12 federation service providers

\+ Early adopters; Partner infrastructures

57% of the budget for service provisioning

EGI-ACE objectives

Implement the **Compute Platform of the European Open Science Cloud** and contribute to the **EOSC Data Commons** by delivering integrated computing, platforms, data spaces and tools as an integrated solution that is **aligned with major European cloud federation projects and HPC initiatives**.

Objective 1. Deliver the European Open Science Cloud Compute Platform and expand the supply-side

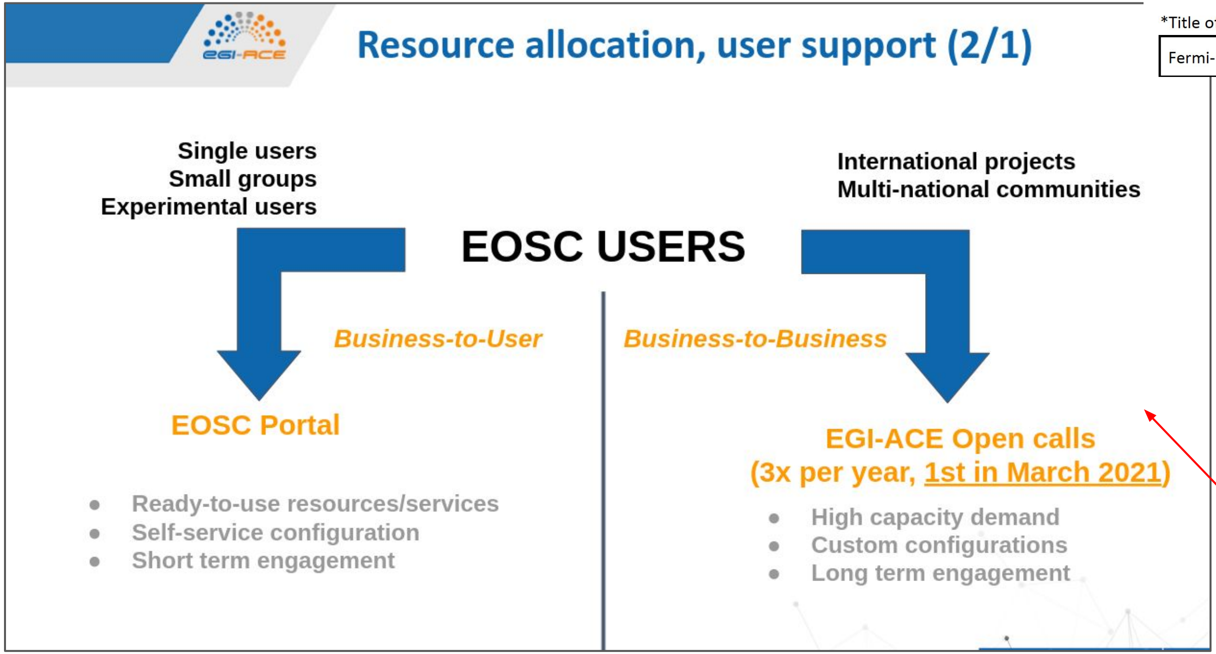
Objective 2: Contribute to the implementation of the EU Data Strategy and the EOSC Data Commons to support the Green Deal, Health and Fundamental Research

Objective 3: Integrate the EOSC Compute Platform with the EOSC Portal and the EOSC Core

Objective 4: Contribute to the realization of a global Open Science Cloud

Objective 5: Expand the demand-side of EOSC across sectors and disciplines

From user perspectives



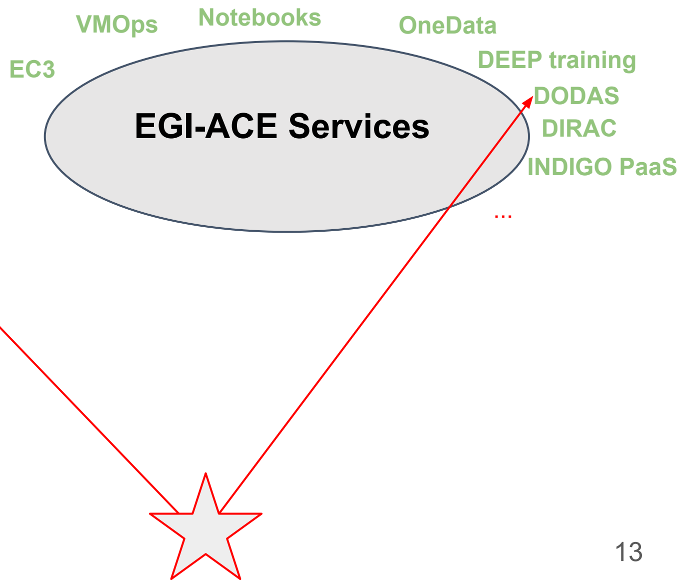
Use case submission form

v1.0

Please submit to egi-ace-call@mailman.egi.eu

*Title of the use case

Fermi-LAT data analysis and interface with italian mirror data archive at SSDC



Dynamic On Demand Analysis Service: DODAS

- A solution designed with the goal to enable users **to create and provision infrastructure deployments, automatically and repeatedly**, on "any cloud provider" **with almost zero effort**.
- Implement the ***Infrastructure as Code*** **paradigm**: driven by a templating engine to specify high-level requirements . Declarative approach **allows to describe "What" instead of "How"**
  - Let the underlying system to abstract providers and automatically instantiate and setup the computing system(s)

Allows to instantiate **on-demand container-based clusters**

- Everything is natively integrated with Kubernetes

- **Resources abstraction**
- **Automation**
- **Multi Cloud support**
- **Federated Authentication**

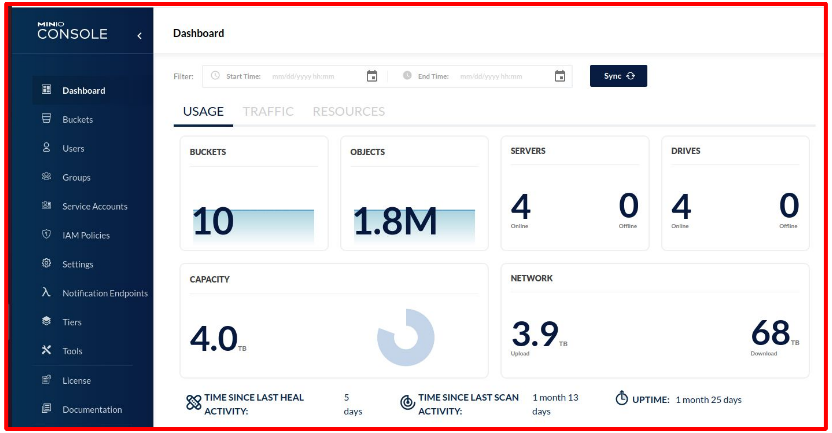
- **Batch System as a Service**
  - HTCondor batch system
  - HTCondor federation solutions
- **Big Data, interactive/quasi interactive analysis**
  - Spark, Jupyter..
- **Services for data access**
  - Caches

Integrating DODAS and Fermi-LAT

kubernetes

DODAS has been used to provide a all-in-one system with

- HTCondor batch on-demand
  - Support Token based Authentication to allow remote job submission
  - User tailored runtime environment
  - Ready to support a cluster federation
- MinIO as a Service
  - Deployed as MinIO Kubernetes Operator
  - Fully integrated with runtime environment (read and write)
  - Ready to support replicas to distributed clusters

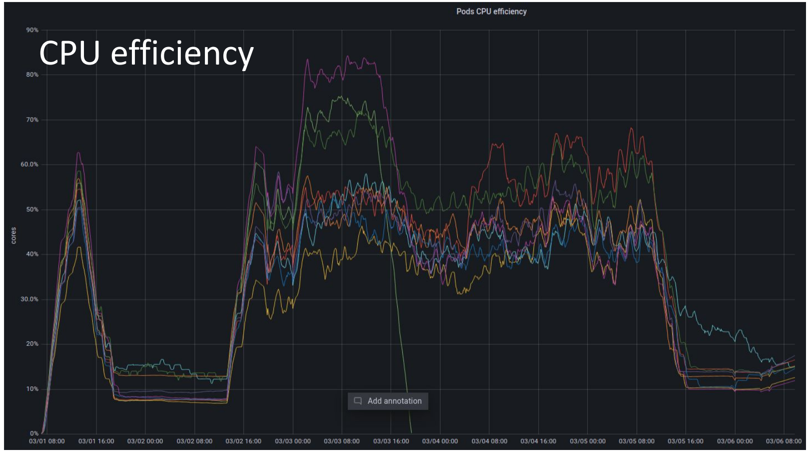


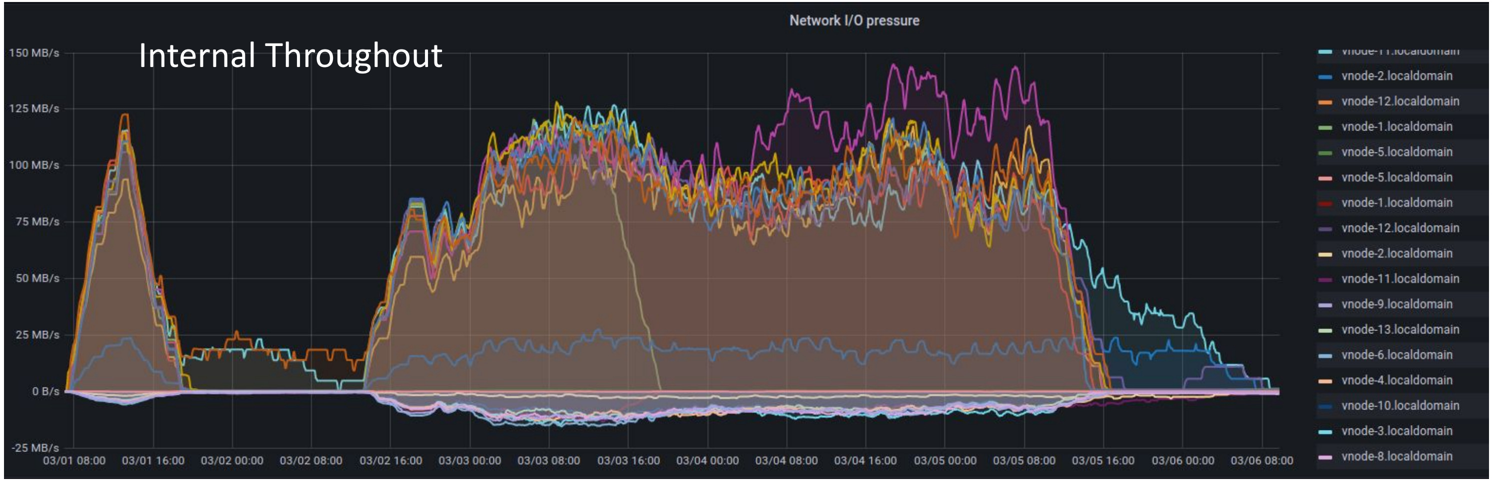
MinIO Console

The current status

A DODAS based system for FERMI-LAT has been successfully deployed at **INFN-CLOUD-BARI**

- 20000 TB MinIO storage and local scratch area
- 150 cores, mostly for Workernodes

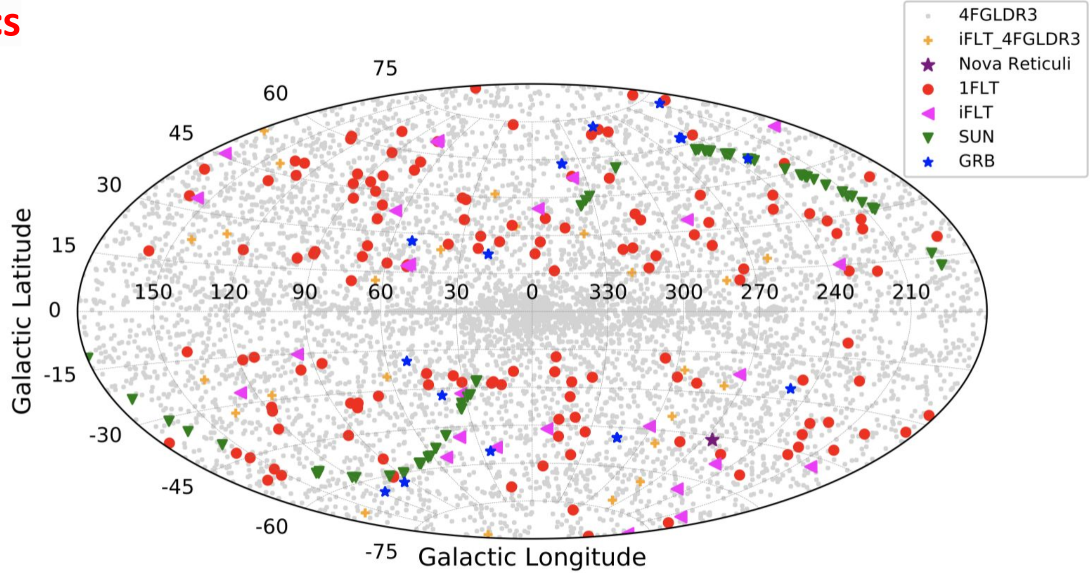




Physics achievements: FLT Results

New results in time-domain astrophysics

- Discovery of **new sources** with a gamma-ray variability in the monthly time scale that were not yet reported in other Fermi-LAT catalogs **→15 per year**
- 72 Sun detections
- 14 GRBs detected



L. Baldini *et al* 2021 *ApJS* 256 13
Mereu, I., Cutini, S., Cavazzuti, E., & Tosti, G. (ICRC 2021).

Lessons learnt and next steps

The current experience has been very positive:

- Porting our analysis in Cloud via DODAS is opening new opportunities either to exploit the distributed resources in a transparent manner but also to reduce our computation time
    - “from years to days”
- Staging data close to the computing resources has the great advantage that allows to improve the efficiency
    - MinIO offers various capabilities to synch multiple instances but need to work on this to better
- Looking for exploiting the cluster federation capabilities of DODAS
    - EGI-ACE project will provide us resource @ INFN-CLOUD-CNAF where we will deploy a twisted cluster for FERMI-LAT

Summary

We submitted a proposal to EGI-ACE open call which has been selected

- DODAS Service has been identified as suitable service matching our needs
    - Enable FERMI-LAT workflow processing for data analysis purposes
    - Allow the porting of the system in Cloud
    - The current model doesn't show any limitation and is allowing us to develop analysis model with fast turnaround
- Since the system is promising we are looking forward to
    - further extension in terms of cluster federation
    - Enhance the data management of the system