Computing en ATLAS y Kubernetes para servicios

Jornadas Tecnológicas del IFIC

Valencia – 17 septiembre 2025

Helena Burriel Navarro (1), Carles Acosta-Silva (2), Pablo Collado Soto (3), Jose del Peso (3), Andreu Pacheco Pages (2), Javier Sánchez Martínez (1), José Francisco Salt Cairols (1), Miguel Villaplana Pérez (1), Roberto Uzum (1), Santiago Gonzalez De La Hoz (1), Emma Torró Pastor (1), Jose Enrique García Navarro (1), Josu Cantero García (1)

(1) Institut de Física Corpuscular (IFIC), University of Valencia and CSIC, Valencia, Spain
(2) Institut de Física d'Altes Energies (IFAE) and Port d'Informació Científica (PIC)
(3) Departamento de Física Teórica y CIAFF, Universidad Autónoma de Madrid, Spain

About me...

Contents

ATLAS Experiment@CERN

WLCG

Spanish sites at WLCG

Spanish Tier2 for ATLAS (IFAE, IFIC, UAM)

Further challenges: High Luminosity

Running jobs per slot

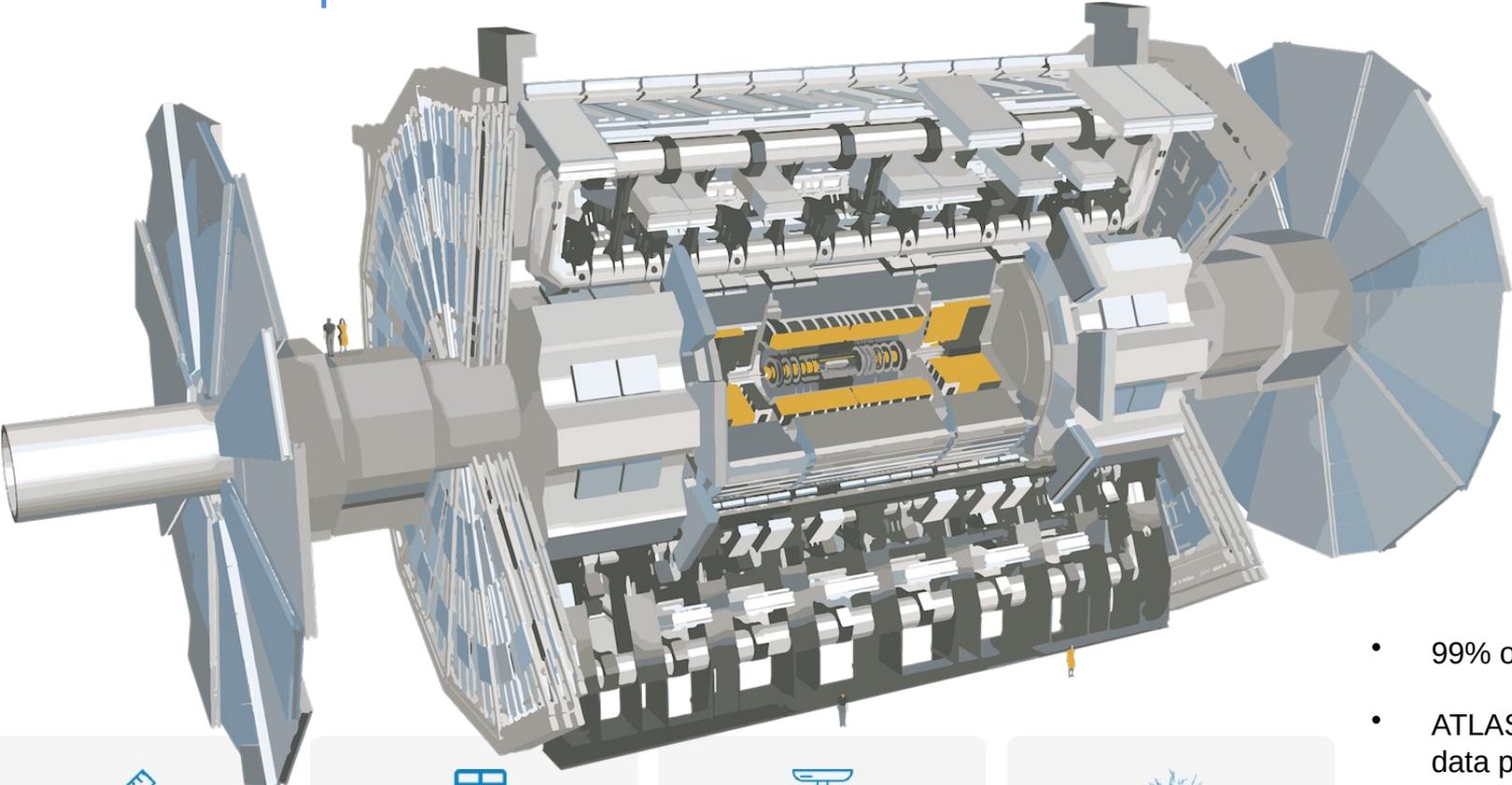Efficiency

Processed events for run3

Storage

New project: Kubernetes + Ceph

Analysis Facility @ IFIC

ATLAS Tier2

The ATLAS Experiment


- 99% of data is not stored

- ATLAS records over 10,000 TB of data per year

- ATLAS processes 25,000 TB of data every week

46 meters long & 25m of diameter

100 meters below ground

7000 tonnes

1000000000 collisions / second

5

Wolrdwide LHC Computing Grid (WLCG)

- ATLAS Experiment data acquisition, storage and distribution infrastructure provides the scientific community with datasets and tools for analysis

- around 170 centers in 42 countries

- 1 EXABYTE of storage

- Tiered structure
  - Tier0 at CERN: raw data
  - Tier1: ~10 sites with data & calibration information
  - Tier2: ~140 sites with distributed data for scientific analysis

- Spain is one of the 10 Tier1 sites: Barcelona PIC
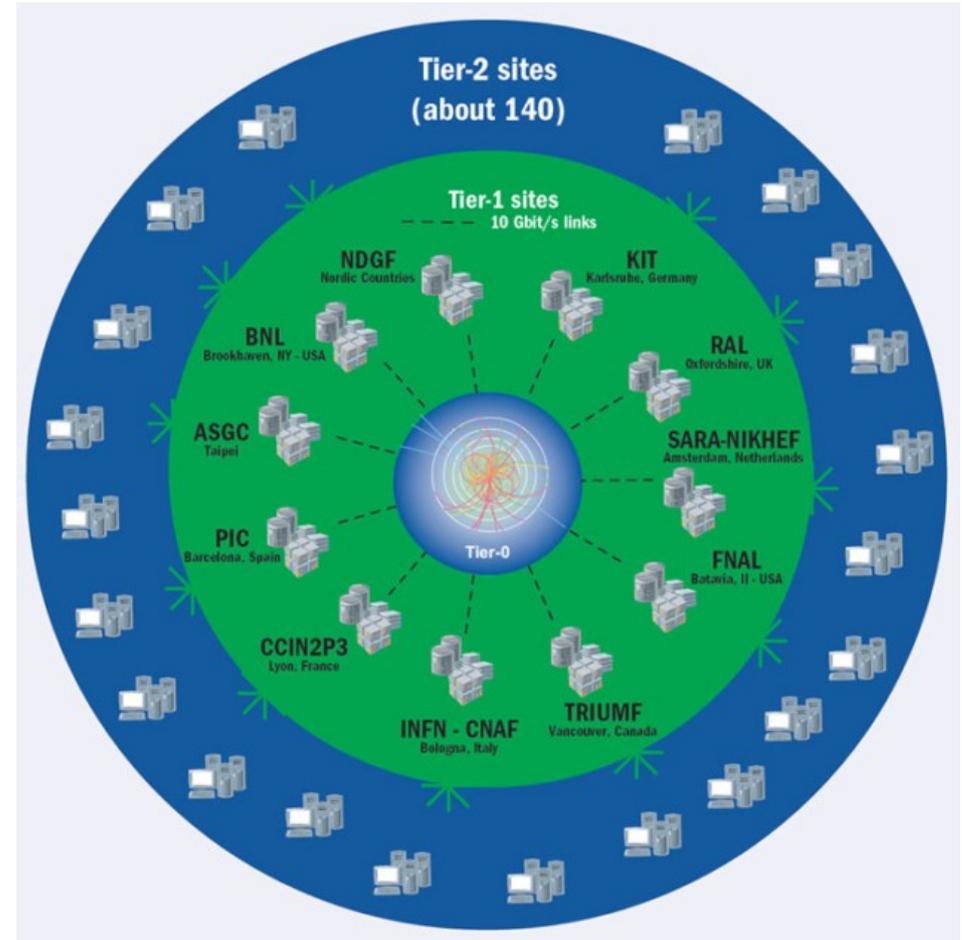
Spanish computing sites for the LHC (WLCG-ES sites)

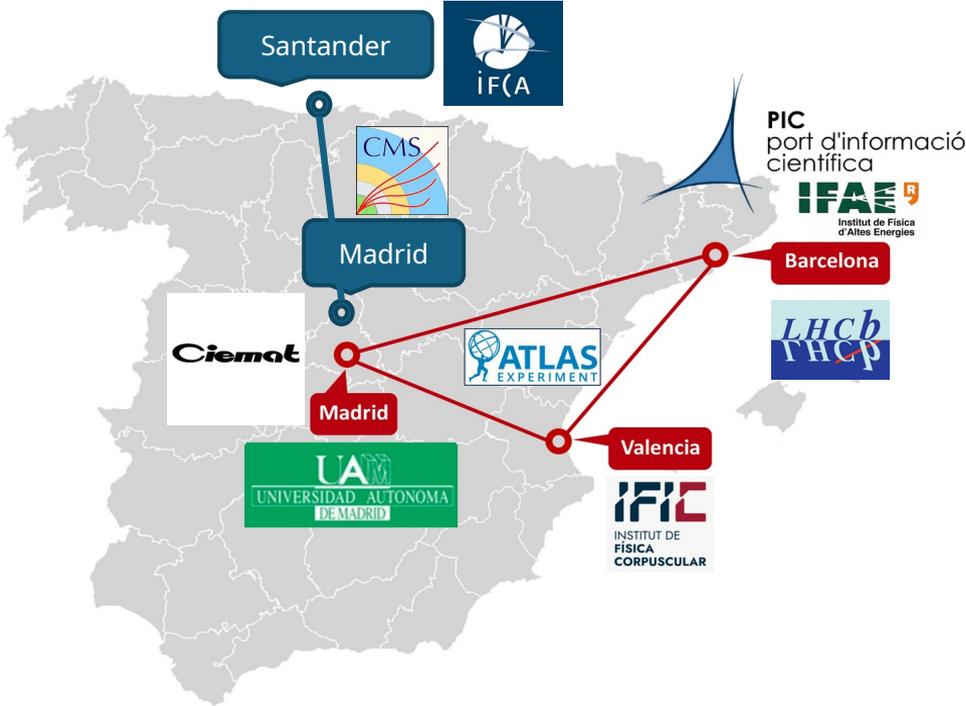Map of the Spanish contribution to ATLAS and other CERN experiments

Composed by Tier 1 and Tier 2 centers

- Tier 1 (ATLAS, CMS, LHCb):
  - PIC-Barcelona

- Federated Tier2s
  - 60% IFIC-Valencia
  - 25% IFAE-Barcelona
  - 15% UAM-Madrid
  - 75% Ciemat-Madrid
  - 25% IFCA-Santander

- What is GRID Computing?
  - Distributed computing over multiple nodes **across the globe**.
  - Provides a **shared pool of resources** (CPU, storage, etc.) for **high-demand tasks** (derivation, analysis, simulation).

- Computing resources are **available 100% of the time.**

7

IFAE

IFAE has 25% of the Spanish ATLAS Tier2 resources

- IFAE is co-located with the Spanish Tier1 at PIC sharing the resources dynamically. The site uses 12% of the PIC batch system: **1566 cores** (20798 HS06) and **2.55 PB** of disk space based on dCache.

2025 Pledged resources:
CPU (HS23): 24750
Disk (TB): 2550

2026 Pledged resources:
CPU (HS23): 27525
Disk (TB): 3062.5

IFIC — INSTITUT DE FÍSICA CORPUSCULAR

IFIC has 60% of the Spanish ATLAS Tier2 resources

- IFIC dedicates 61 computing servers with **1943 cores** with no hyper-threading, 15 storage servers with **6.4 PB** of disks, and several machines for file transfer control, information systems and management

2025 Pledged resources:
CPU (HS23): 59400
Disk (TB): 6120

2026 Pledged resources:
CPU (HS23): 66060
Disk (TB): 7350

UAM

UAM has 15% of the Spanish ATLAS Tier2 resources

- UAM dedicates 21 computing servers with **576 cores**, 44 storage nodes with capacities ranging from 11 TB to 120 TB totaling **1.48 PB** and over 10 machines dealing with transfer metadata, GRID job submission and other auxiliary tasks

2025 Pledged resources:
CPU (HS23): 14850
Disk (TB): 1530

2026 Pledged resources:
CPU (HS23): 16515
Disk (TB): 1837.5

The Spanish ATLAS Tier2 offers 3% of the ATLAS Tier2 resources, will increase up to 5% in 2026
Pledge for Spanish ATLAS Tier2 2026: **12150** TB Disk, **110100** HS23 CPU
IFIC has increased storage this year with **10 new storage servers up to 6.4 PB** of data

8

High Luminosity challenges in the near future

https://hilumilhc.web.cern.ch/content/hl-lhc-project
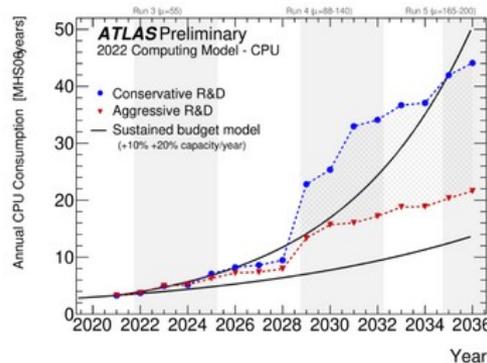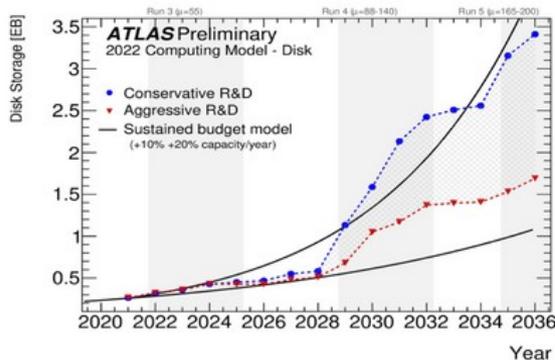
Correct accounting is key to plan future use of resources, e.g. high luminosity:

- Reduce cost

- Future planning: the future runs will be **x10** more demanding (**x10 data generation**)

- Plan for new solutions to avoid **x10** more time of scientific analysis

International collaboration with high standards on availability, storage, processing, monitoring and updates
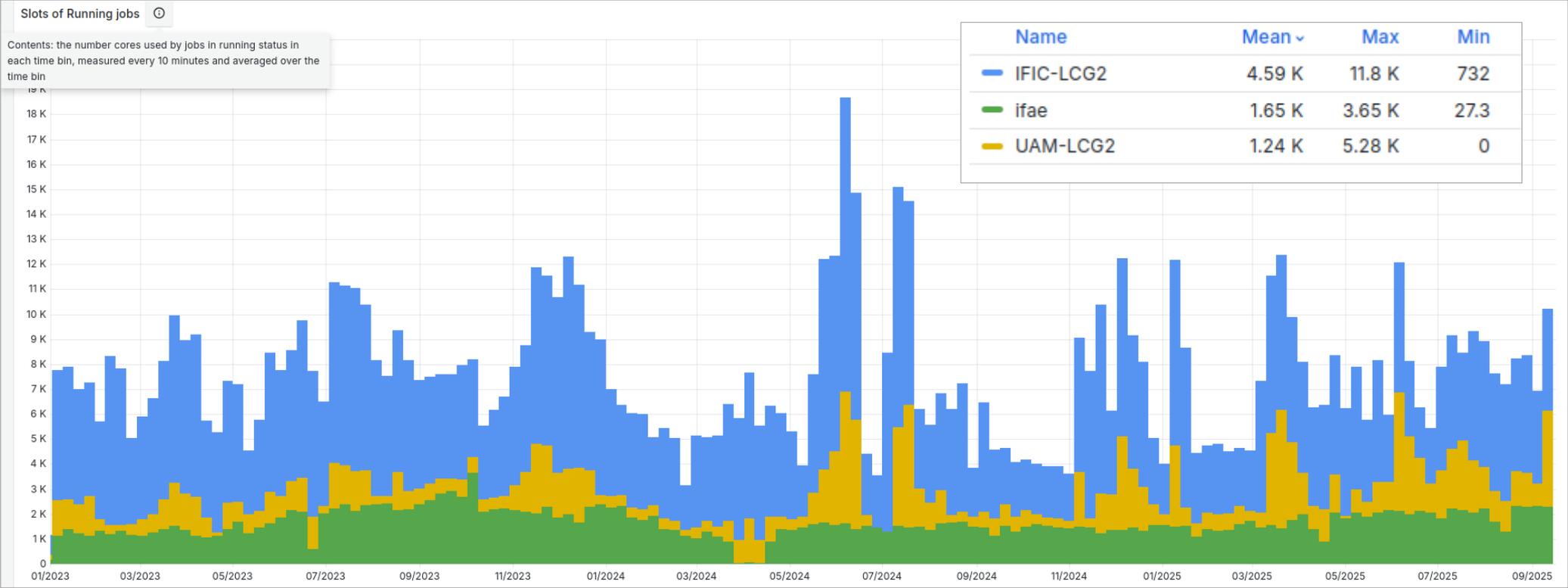
Used metrics and meanings for the accounting of the resources and activity:

- Efficiency: amount of computational resources used

- Availability: % of time that the site is available

- Reliability: % of time that the site is available without considering scheduled downtime

If a service is failing, the whole site is considered to be failing, even when most services are available

Activity accounting

Slots of running jobs

Slots of Running jobs

Contents: the number cores used by jobs in running status in each time bin, measured every 10 minutes and averaged over the time bin

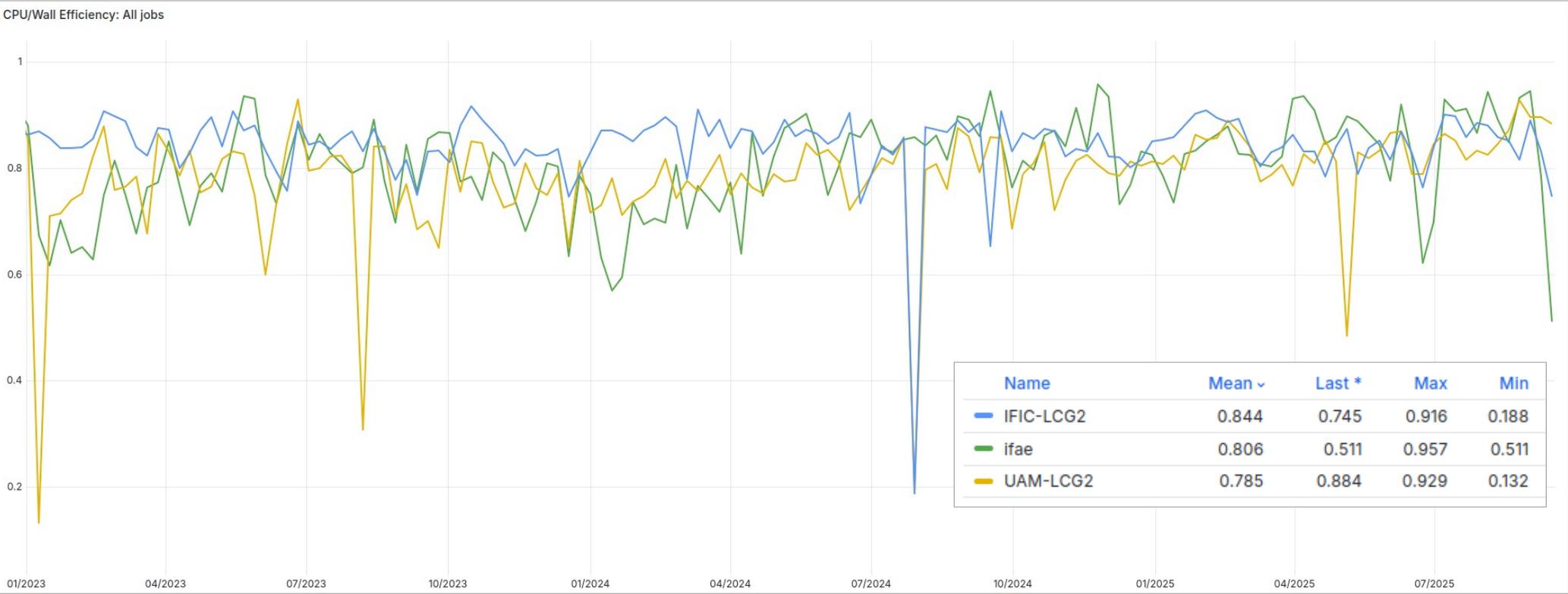| Name	Mean ⌄	Max	Min
IFIC-LCG2	4.59 K	11.8 K	732
ifae	1.65 K	3.65 K	27.3
UAM-LCG2	1.24 K	5.28 K	0

Slots are provided by the batch processing system for an efficient and fair use of the computing resources

Slots for single CPU usage (1 core) or multiple CPU usage (8 cores), with 2GB of memory per CPU

1 CPU - 2 GB / 8 CPU - 16 GB

11

CPU/Wall Efficiency: All jobs

Amount of CPU used for the jobs

CPU/Wall Efficiency: All jobs

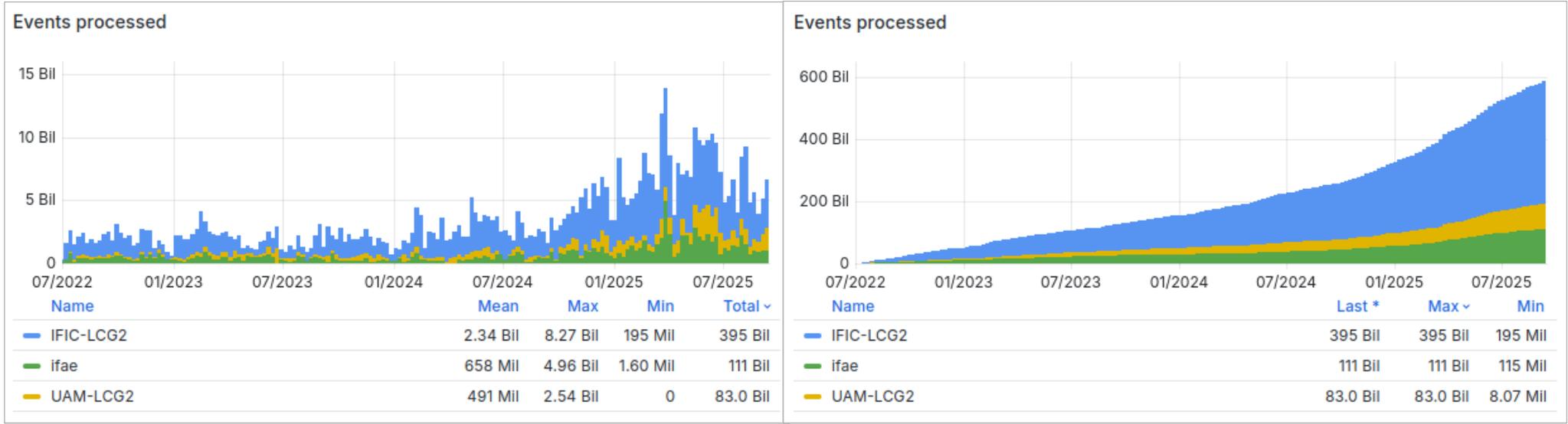| Name	Mean ⌄	Last *	Max	Min
IFIC-LCG2	0.844	0.745	0.916	0.188
ifae	0.806	0.511	0.957	0.511
UAM-LCG2	0.785	0.884	0.929	0.132

This is for CPU efficiency (processed jobs, failed jobs, canceled jobs)
The 0.844 means not wasting computing resources

Input/output operations are not taken into account

12

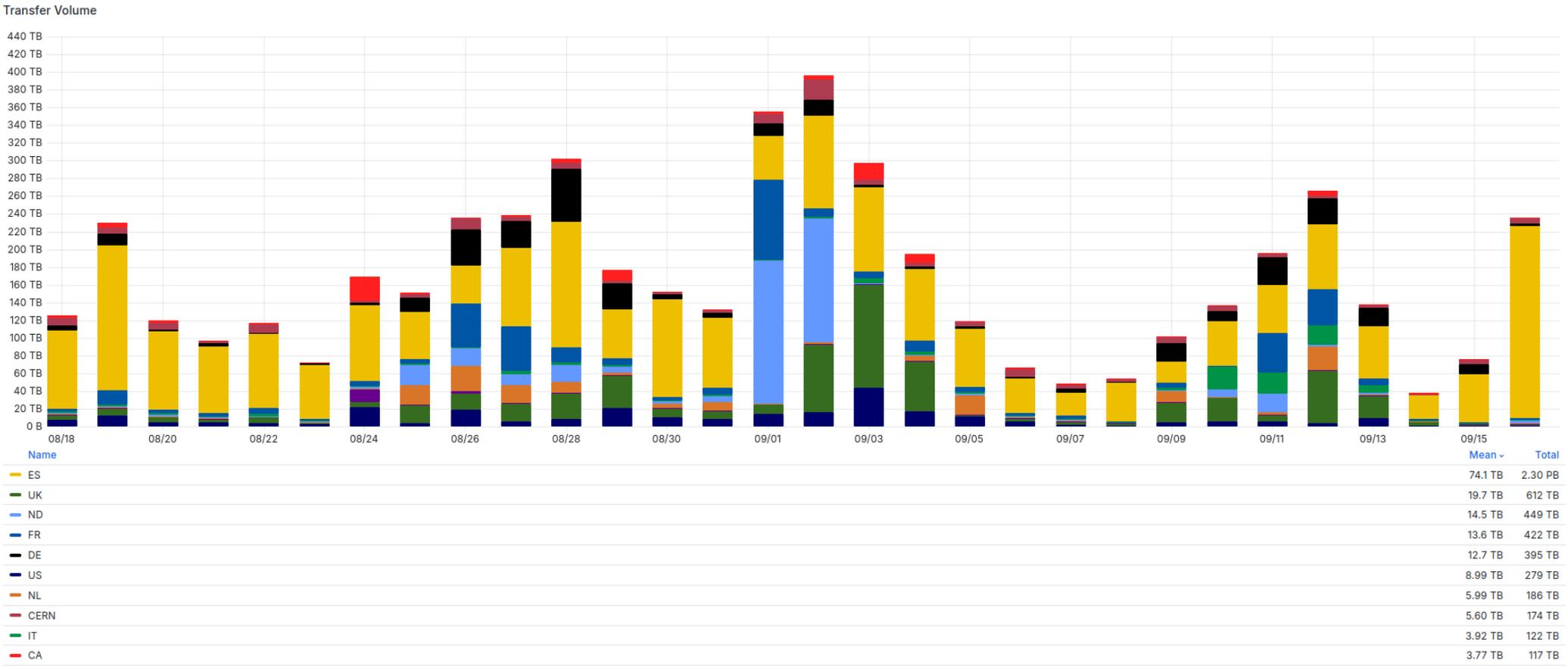# Events processed in RUN3 (time & accumulative)   Storage related and events

Processed events since the beginning of the run3 (5-july-2022) total: **395 Billion at IFIC**
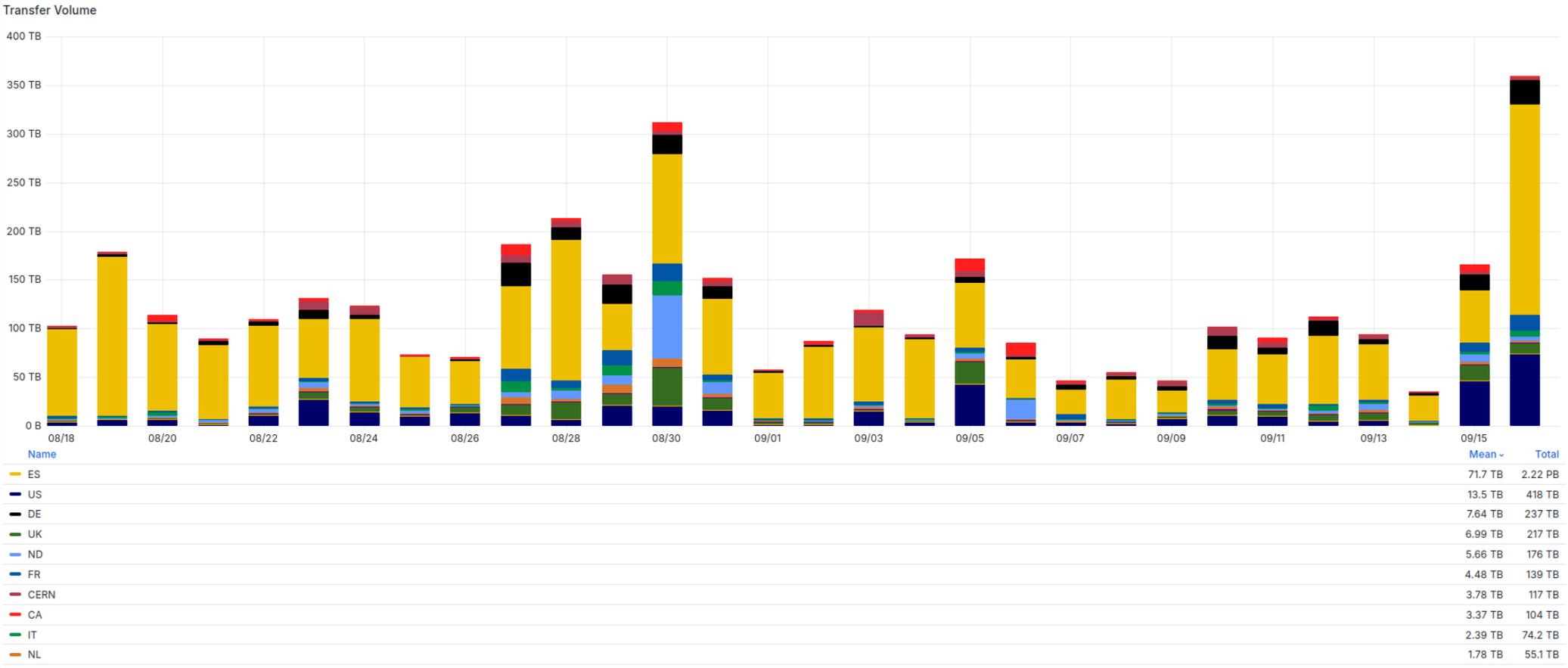
In Bytes: **116.4 PB** processed by the three sites (**73.4 PB IFIC** – 28.6 PB IFAE – 14.4 PB UAM)

In Files: **65 Million files** processed (**39.9 Million files IFIC** – 12.7 Million files IFAE – 12.4 Million UAM)
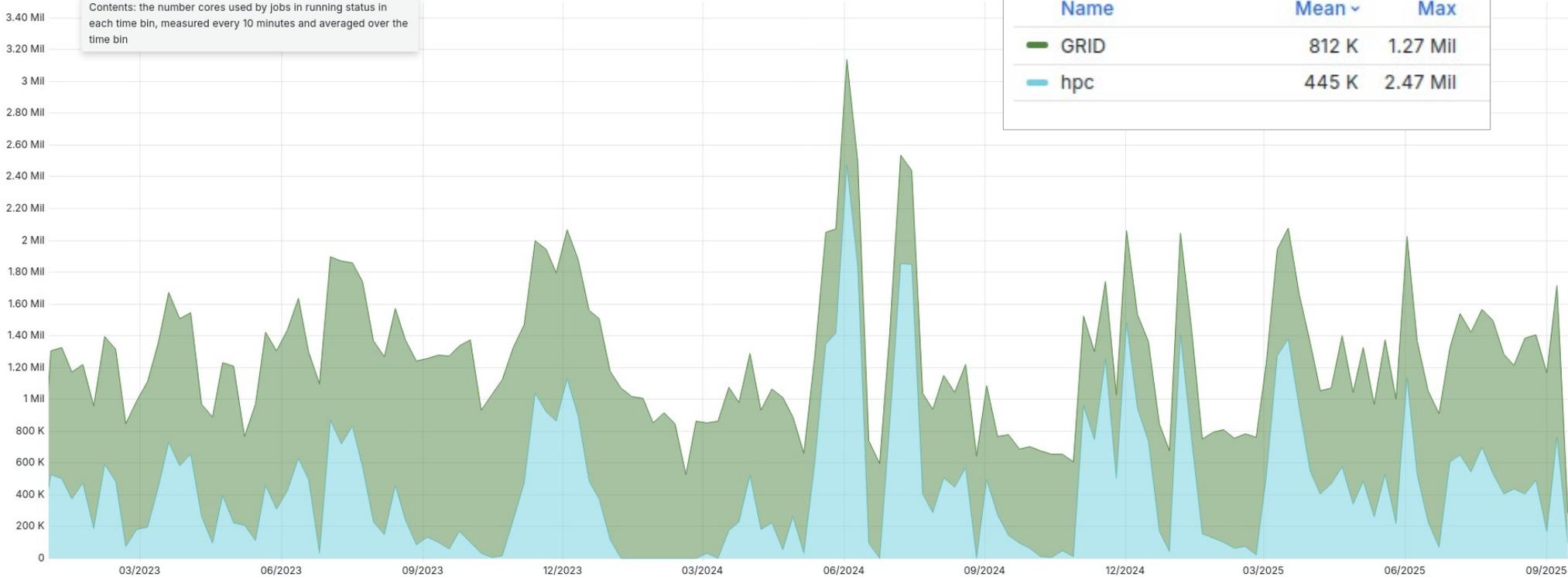
13

Daily Transfers (Destination: IFIC) ~4.16 PB month

Transfer Volume

Name	Mean ↓	Total
ES	74.1 TB	2.30 PB
UK	19.7 TB	612 TB
ND	14.5 TB	449 TB
FR	13.6 TB	422 TB
DE	12.7 TB	395 TB
US	8.99 TB	279 TB
NL	5.99 TB	186 TB
CERN	5.60 TB	174 TB
IT	3.92 TB	122 TB
CA	3.77 TB	117 TB

Daily Transfers (Source: IFIC) ~3.5 PB month

HPC (High Performance Computing)

Slots of Running jobs by Resource Type ⓘ

Contents: the number cores used by jobs in running status in each time bin, measured every 10 minutes and averaged over the time bin

Name	Mean ⌄	Max
GRID	812 K	1.27 Mil
hpc	445 K	2.47 Mil

HPC is dedicated to MonteCarlo simulation

70% IFIC+UAM (Tier2) 30% MareNostrum (HPC) (IFAE HPC is accounted in PIC Tier1)

Increase in HPC jobs are due to the change from MareNostrum 4 to MareNostrum 5 (higher number of job cores)

16

New Project:
Kubernetes + Ceph

Kubernetes

Kubernetes es una plataforma de orquestación de contenedores que automatiza la implementación, escalabilidad y gestión de aplicaciones en contenedores.
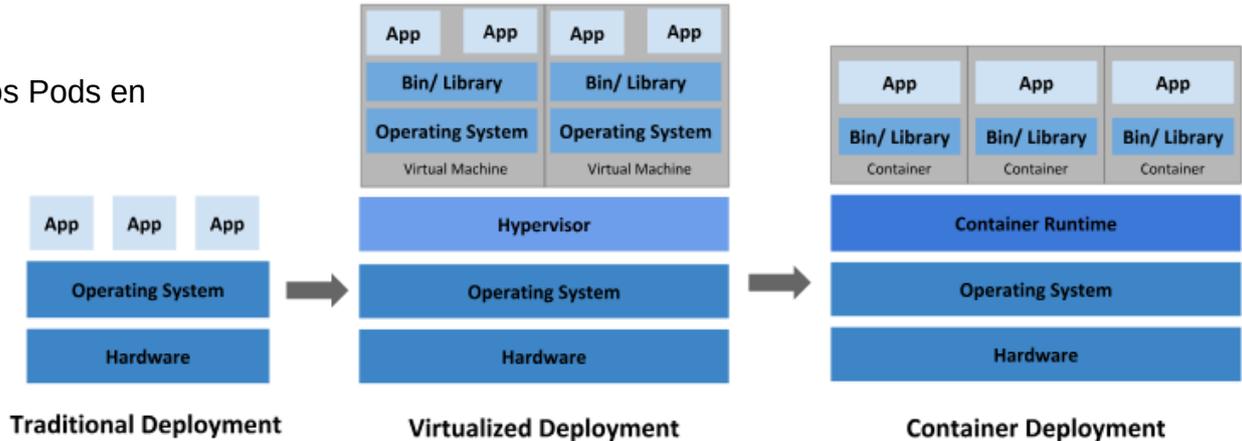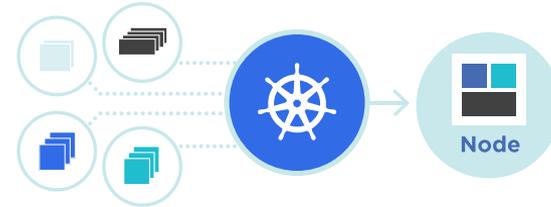
Arquitectura Básica:

- **Nodos**: Máquinas físicas o virtuales que ejecutan contenedores.

- **Pods**: Unidades más pequeñas de ejecución que contienen uno o más contenedores. Comparten almacenamiento y red.

- **Control Plane**: Gestiona el estado del cluster y toma decisiones globales.

- **Node Components**: Ejecutan y gestionan los Pods en cada nodo de trabajo.

Ventajas de Kubernetes:

- Escalabilidad automática.

- Resiliencia.

- Gestión centralizada.

Problema:

Kubernetes no maneja almacenamiento persistente por sí solo, **necesita provisión de almacenamiento persistente**

Traditional Deployment   **Virtualized Deployment**   **Container Deployment**

Ceph

Ceph es una plataforma de almacenamiento distribuido que proporciona almacenamiento en bloque, objeto y sistema de archivos, todo en una única solución.

Arquitectura:

- **Monitores** (MON): Controlan el estado del clúster.

- **OSD** (Object Storage Daemons): Gestionan el almacenamiento de datos reales.

- **MDS** (MetaData Server): Gestionan los metadatos del sistema de archivos, como la estructura de directorios y archivos (si se usa CephFS).

Ventajas de Ceph:

- Escalabilidad horizontal.

- Alta disponibilidad y tolerancia a fallos.

- Flexible para almacenamiento de objetos, bloques y archivos.

Actualmente lo estamos utilizando como **almacenamiento central** de la infraestructura del CPD.

Proporciona almacenamiento de bloques (discos) a máquinas virtuales y físicas.

Lo usamos en **AFS**, **Correo**, **IBOX** y **máquinas virtuales**

Se mantienen **3 copias** para todos los objetos.

- 5 servidores de disco (OSD)

- 3 servidores de monitorización (MON) mantienen el mapa del cluster

- interfaces de 10 gpbs

Integración Kubernetes + Ceph: casos de uso

Ceph como solución a la **provisión de almacenamiento persistente** en Kubernetes.

Kubernetes:

- Despliegue de microservicios en contenedores.

- Automatización de la infraestructura de aplicaciones.

- Escalabilidad en aplicaciones.

Ceph:

- Almacenamiento de grandes volúmenes de datos en clústeres distribuidos.

- Almacenamiento para aplicaciones de alta disponibilidad.

- Backup y recuperación de datos a gran escala.

Propósito de integrar Kubernetes + Ceph:

Prototipo para la creación de infraestructuras de despliegue de contenedores automatizada con persistencia de datos altamente escalable y redundante.

Esto permitirá poder utilizar contenedores para los servicios (en sustitucuón de máquinas virtuales):
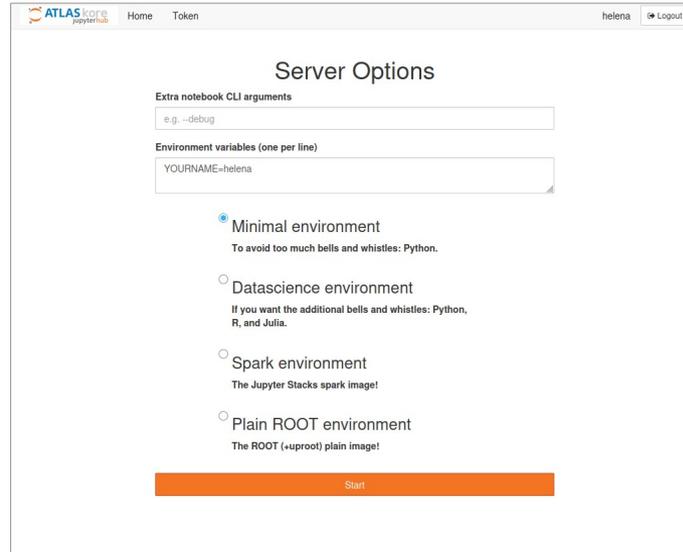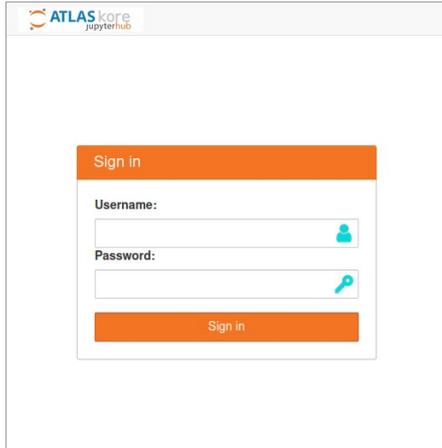
· Servidores web de proyectos.

· Aplicaciones con persisencia, p. ej. **IBOX**, **GitLab.**

· Almacenamiento para las aplicaciones **Flask** a desarrollar**.**

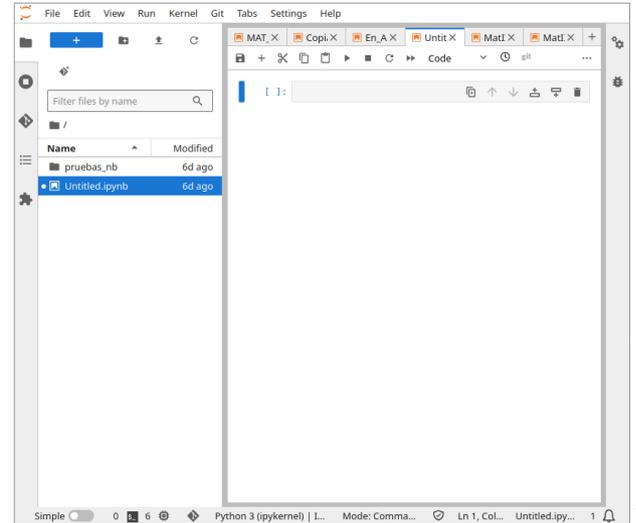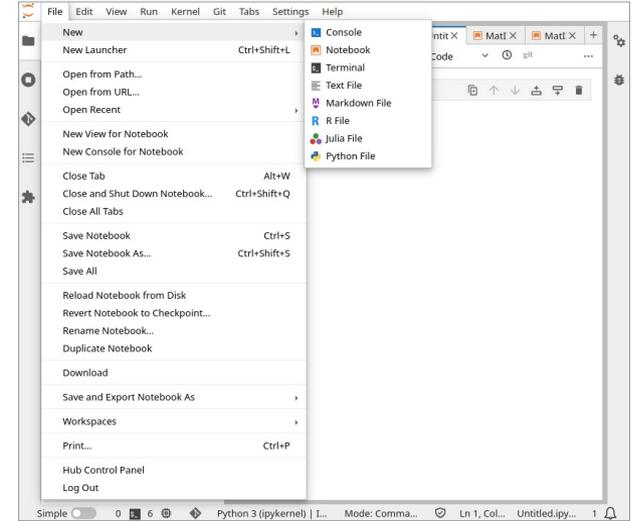· Extender GitLab a **CI/CD** (integración continua).

… y mucho más.

Estamos en fase de desarrollo del prototipo

One more thing

Analysis Facility @ IFIC

https://kore.ific.uv.es/

References

https://hilumilhc.web.cern.ch/content/hl-lhc-project

https://monit-grafana.cern.ch/goto/HlQSOP_SR?orgId=17

https://kubernetes.io

https://ceph.io/en/

https://kore.ific.uv.es/

Thanks!