Infrastructures for computing and AI

ASFAE
ASTROFÍSICA Y FÍSICA
DE ALTAS ENERGÍAS

Computación Avanzada para el procesado intensivo de Big Data en ATLAS: **ASFAE/2022/006**

ARTEMISA: **ASFAE/2022/024**

Helena Burriel Navarro
Matías Salinero Delgado

Instituto de Física Corpuscular

March 2024, Alicante

IFIC INSTITUT DE FÍSICA CORPUSCULAR

CSIC CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

VNIVERSITAT DƏ VALÈNCIA

Computación Avanzada para el procesado intensivo de Big Data en ATLAS
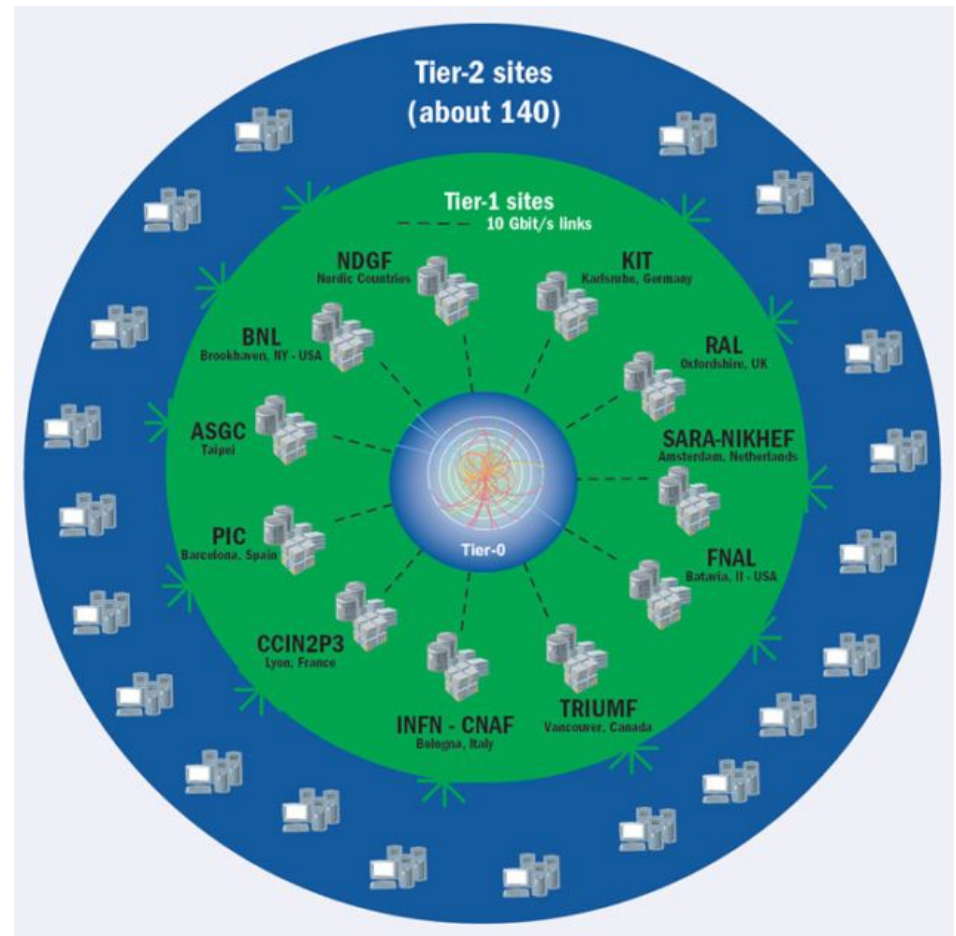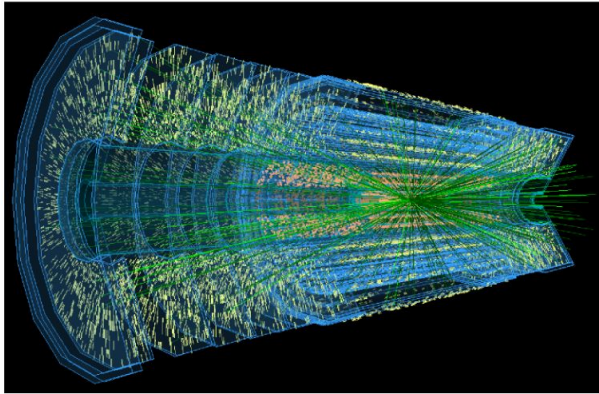
ASFAE/2022/006

LHC computing model

Spanish computing sites for the LHC (WLCG-ES sites)

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

- Federated Tier2s
  - 60% IFIC-Valencia
  - 25% IFAE-Barcelona
  - 15% UAM-Madrid
  - 75% Ciemat-Madrid
  - 25% IFCA-Santander
  - 50% USC-Santiago
  - 50% UB-Barcelona*

  *UB Tier2: Now it is working as a Tier3 where MC simulation is running

- LHC sites pledges in the last 5 years:

  https://wlcg-cric.cern.ch/core/federation/list/

- **Integrated in the WLCG project (World Wide LHC Computing GRID) and strictly following the experiments computing models.**
- **We represented the 4-5% of the total Tier-2s and 5% of the total Tier-1s resources, with the budget reduction now the 3% for Tier2s and 4% for Tier1!!!!!**
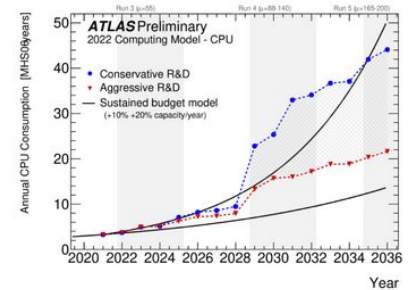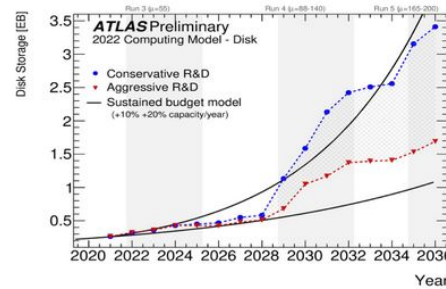
LHC HL-LHC Plan and estimation

The original schedule for HL-LHC had pileup = 200 already in 2027, a large jump.

Revised schedule from January 2022:

- First full year of HL-LHC running at PU=140 is 2030.
- PU=200 reached only in Run 5

⇨ Nevertheless, x10 more events, bigger and more complex events (x5) -> **Unprecedented challenge!**

- **ATLAS Flat Budget projections** from 2022 show that future needs
  - Optimization (both speed and flexibility) of the experiment (e.g.reconstruction, simulation) and non-experiment (e.g. generation) software
  - Optimization of the available hardware infrastructure usage
  - Storage is the most expensive resource to deploy and operate

ATLAS Tier-2 @ IFIC
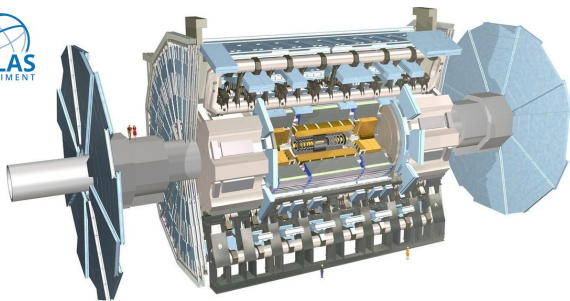
- Computing infrastructure for ATLAS Tier2 is dedicated to storage and data processing for the ATLAS experiment at LHC at CERN
- IFIC dedicates 61 computing servers with **4384 cores**, 15 storage servers with **4 PBytes** of disks, and several machines for file transfer control, information systems and management
- International collaboration with high standards on availability, storage, processing, monitoring and updates

- At the top of availability and reliability ranks: IFIC Tier2 is a Nucleus
  - Nucleus are Tier2s with a big amount of storage and very good network connection, passing job production on to smaller Tier2s (Satellites)
  - ~13 PB of ATLAS data stored in the Spanish Tier 1 and Tier 2 centers -> >4 PB stored at IFIC!
- Now we have RedIRIS-Nova at 100 Gbps -> IFIC's WAN connectivity increased to 100 Gbps
- IFIC Tier 2 by the numbers:
  - We have processed more than 300 Bill. events in these last 5 years
  - Steady state of more than 5.000 running job slots since 2019, typically using 2GB per job slot
  - Mainly running with either 8 or 1 cores ("multi-core" or "single-core") per job, depending on type of job

- **IFIC has 60% of the Spanish ATLAS Tier2 resources**
  - Pledges (C-RRB 2022/04):
  - **Current status:**
    - CPU (HS06): 44768 (28% come from machines > 5 years old)
    - Disk (TB): 3402 (60% of of storage resources at IFIC older than 6 years!)

CPU delivered by Spain to WLCG

CPU work (HS06.hours) delivered by WLCG-ES + BSC
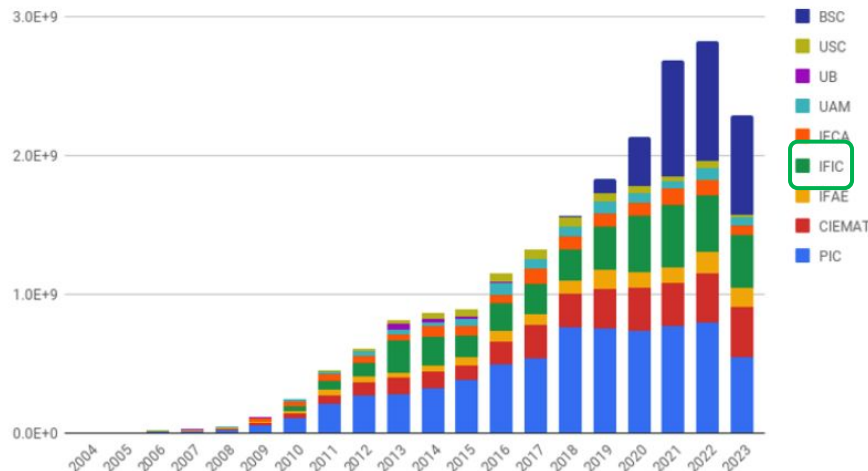
Contribution by site to CPU work delivered in 2004-2023

BSC 14.4%
USC 3.0%
UB 0.5%
UAM 3.9%
IFCA 5.5%
IFIC 17.0%
IFAE 5.8%
PIC 35.3%
CIEMAT 14.4%
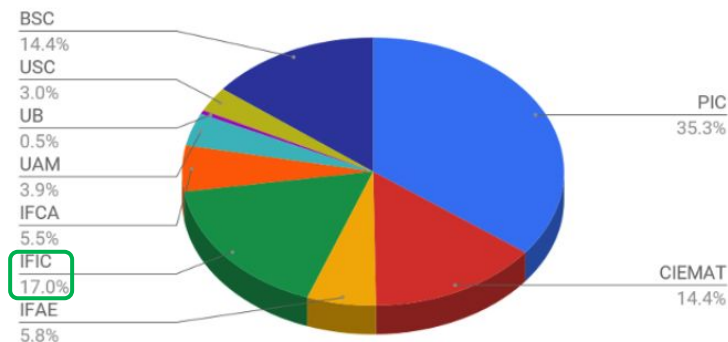
~2000 Million hours delivered during 2004-2023

(20x10^9 HS06.hours; average CPU core power ~10 HS06)

HPC resources

❖ **Use of the Mare Nostrum4 (HPC) by ATLAS Tier-1 and Tier-2s:**
  ➢ Using ARC-CE at PIC, IFIC and UAM to interconnect Mare Nostrum and ATLAS production system.
  ➢ Only simulation workflow validated - singularity containers, pre-placed at MareNostrum's GPFS.
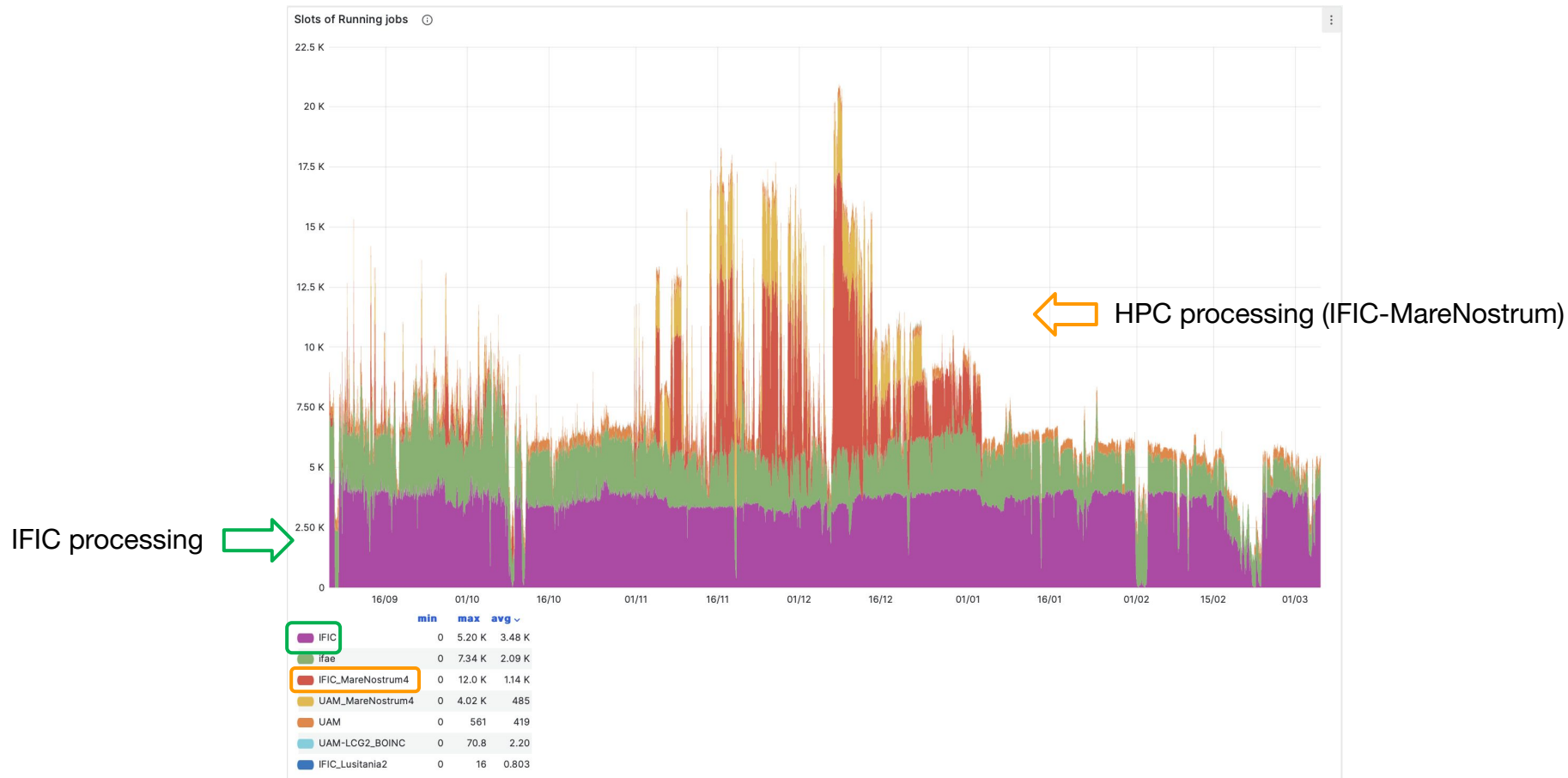 ➢ Mare Nostrum accepts only SSH protocol for job submission and data transfer.
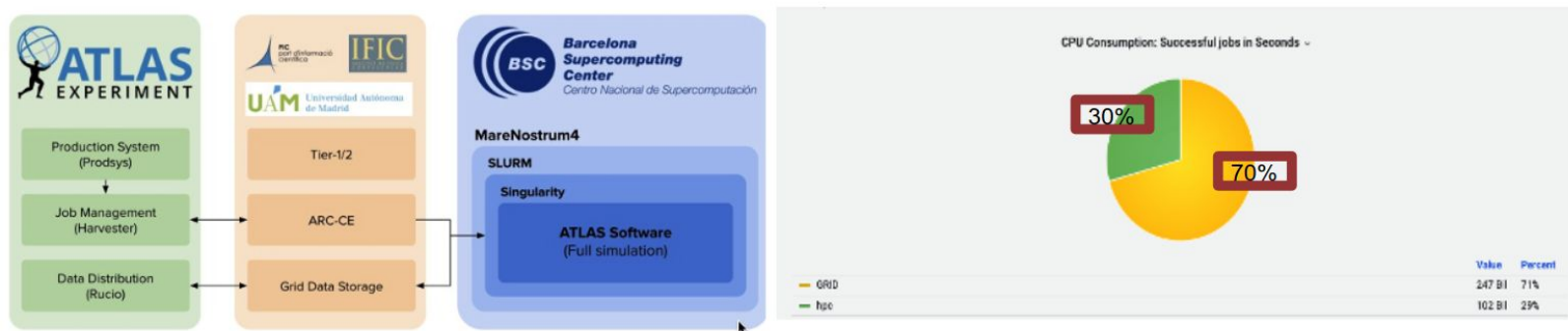
  ➢ Proportion of HS06 (s) provided by GRID resources (yellow) and the MareNostrum 4 HPC (green) in total contribution to the ATLAS computing by the Spanish cloud.
  ➢ 30 million hours approved at Mare Nostrum4 every year by ATLAS through Spanish gateways, which corresponds to 50% of the simulation jobs assigned to Spain.

IFIC Tier 2 efficiency running ATLAS jobs

- Running jobs by site the last 6 months

Slots of Running jobs

HPC processing (IFIC-MareNostrum)

IFIC processing

	min	max	avg
IFIC	0	5.20 K	3.48 K
ifae	0	7.34 K	2.09 K
IFIC_MareNostrum4	0	12.0 K	1.14 K
UAM_MareNostrum4	0	4.02 K	485
UAM	0	561	419
UAM-LCG2_BOINC	0	70.8	2.20
IFIC_Lusitania2	0	16	0.803

IFIC Tier 2 efficiency running ATLAS jobs

- IFIC's CPU efficiency has been higher than 90% during the last years
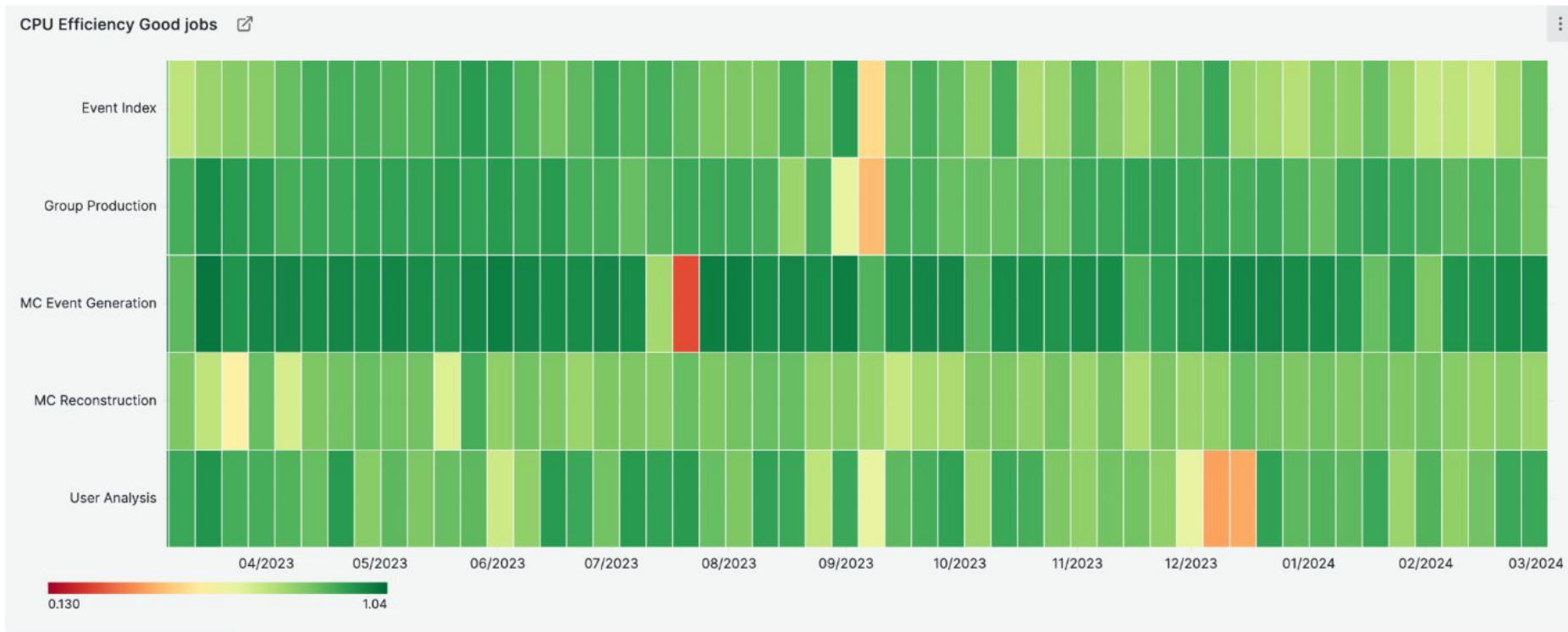
Computing R&D&I activities in Spain

Spain not only contributes computing resources to WLCG, but also participates in computing development projects and in the operation and coordination of the distributed computing infrastructure

Development projects:
- Integration of supercomputing resources (ATLAS)
- Deployment of Analysis Facility (ATLAS)
- Evolution of workload management system (ATLAS)
- Event Index (ATLAS)
- Distributed Physics Analysis (ATLAS)
- Derivation (ATLAS)

Coordination positions:
- Workflow management coordination (ATLAS)
- **Distributed Physics Analysis coordination (ATLAS, IFIC)**
- **Event Index data production and collection coordination (ATLAS, IFIC)**
- WLCG deployment board chair
- **Derivations coordination (ATLAS, IFIC)**
- **LHC Computing Resource Review Board**

IFIC Event Index

❖ A system designed to be a **complete catalogue of ATLAS events**, real and simulated data.

❖ Partitioned architecture, following data flow:
  ➢ **Data Production:** extract event metadata from files produced at Tier-0 or on the Grid
  ➢ **Data Collection:** transfer EventIndex information from jobs to the central servers at CERN
  ➢ **Data Storage**: provide permanent storage for EventIndex data and fast access for the most common queries.
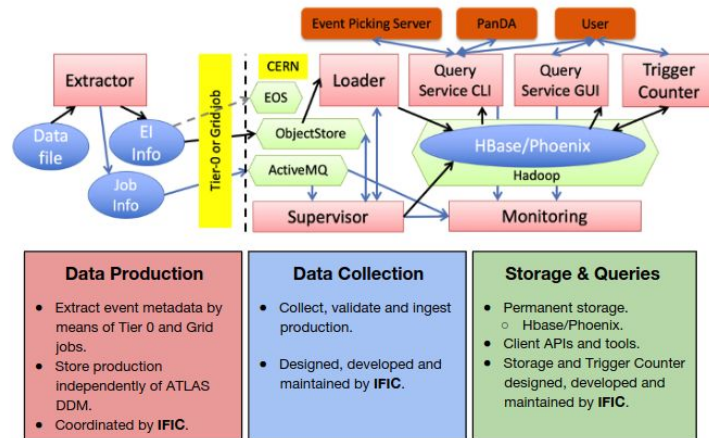
❖ Use Cases:
  ➢ Event Picking.
  ➢ Production consistency checks (Duplicate event and overlap detection).
  ➢ Trigger checks and event skimming
❖ The new system is in operation since last Spring 2022 and performing excellently.

Analysis Facilities

- Guiding principle: Help physicist minimize time-to-insight, enabling iterative exploration of the data:
    - In the future, when processing 10x lumi/evts, avoid physicists 10x waiting time!!
    - Boost productivity and competitiveness of our physics communities
- Typically consist of:
    - Local access to the reduced data samples (e.g. PHYSLITE) with low latency from compute
    - Dedicated storage resources (of order of several 100s of TB).
    - CPU resources used interactively and/or via a batch system (mostly HTCondor).
        - Future User Interface (UI) to be designed in a flexible way, with user-friendly interfaces that do not discourage users.
        - Future implementation of **Jupyter Notebook** instances that will be spawned via a dedicated portal.
    - SW delivery mostly via CVMFS with increasing presence of containers.
    - Expert data/code manager: critical liaison role (not a final user, nor an infrastructure expert, however facilitating technology/access).
    - GPU resources available, but often not dedicated.
        - IFIC: **ARTEMISA** infrastructure (https://artemisa.ific.uv.es)
    - Network:
        - LAN of multiple 25Gbps to support intense data throughput.
        - WAN of 100Gbps connectivity with the WLCG dedicated network for data lake access.

ARTificial Environment for ML and Innovation in Scientific Advanced Computing

ASFAE/2022/024

What is Artemisa?

- GPU-intensive computing infrastructure dedicated to Artificial Intelligence(AI) and Machine Learning(ML) located at IFIC's data centre. Its advanced features and excellent performance enable the accelerated development of projects involving AI areas.

- Capable of ML tasks involving large amounts of data to produce empirical models in physics, chemistry, biology and social studies.

- Provides last generation GPUs, even featuring a 8-GPU A100 server.
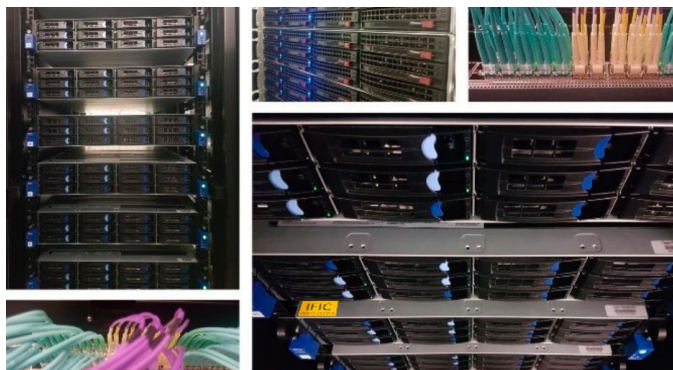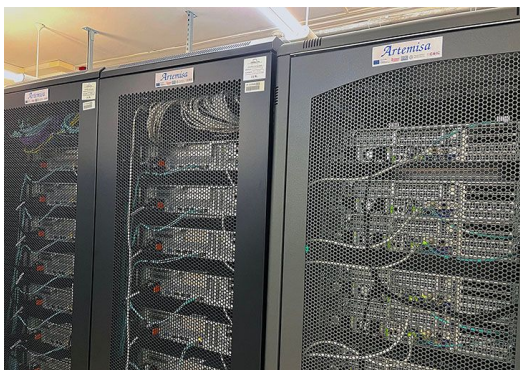
Artemisa was created using funds from the FEDER 2014-2020 Comunidad Valenciana (IDEFEDER/2018/048, bugdet: 1 M€).

Granted a new project within the call "Recuperación y Resiliencia" (ASFAE/2022/024).

Twofold objective:

- Improve support to users and develop further infrastructure

- Increase computing resources by adding new machines

Artemisa (IFIC) is part of the InnDIH (European Digital Innovation Hub) as part of CSIC and Universitat de Valencia.

Financiado por la Unión Europea
NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

GENERALITAT VALENCIANA
Conselleria de Educación, Universidades y Empleo

GVA NEXT
Fondos Next Generation en la Comunitat Valenciana

Who can use Artemisa?

Aimed at researchers with projects that require computing and memory capabilities beyond those provided by conventional computers and participate in public funded projects. Private institutions are also encouraged to apply. to enroll.

Artemisa is particularly suited for:

- Development of machine learning models that require large amounts of data.
- Resolution of optimization problems and/or applications that need distributed computing.
- General studies that would profit from high-performance GPU usage.

There are 3 calls per year (jan-feb, may-jun, sep-oct), 4 months each
Call application is open for a month, accepting projects during that period

Artemisa Projects

Breakdown by region and research area for the projects in Artemisa.
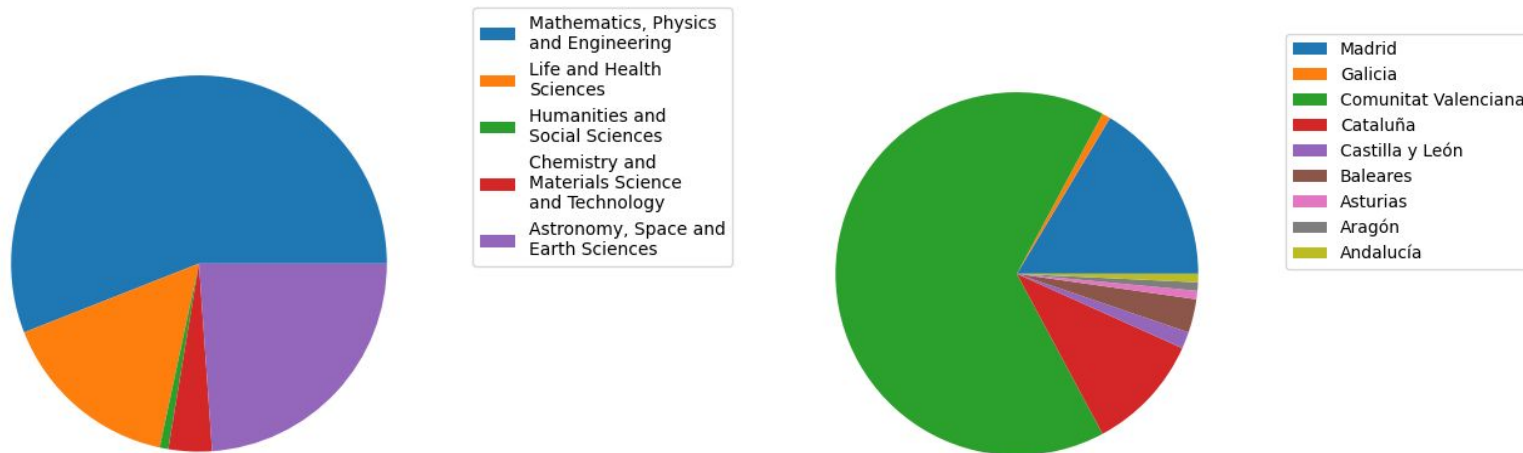
Most of the projects are from "*Comunitat Valenciana*" and "*Mathematics, Physics and Engineering*"

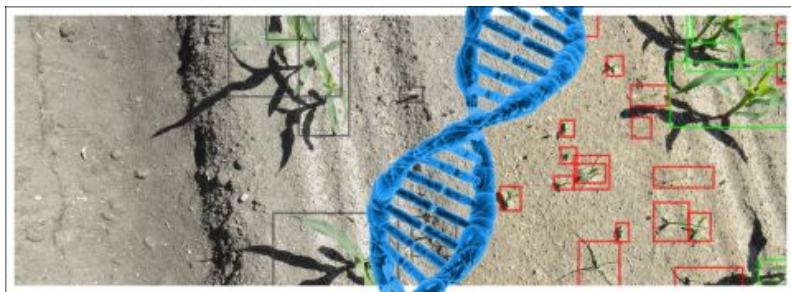but other regions and fields growing.

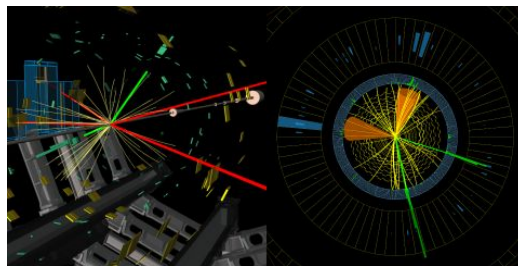# Artemisa Projects

Health

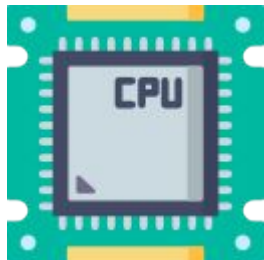**Remote Sensing**
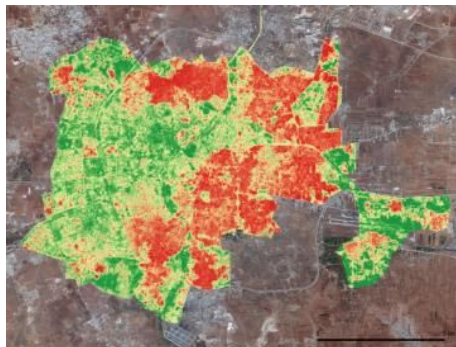
Astronomy
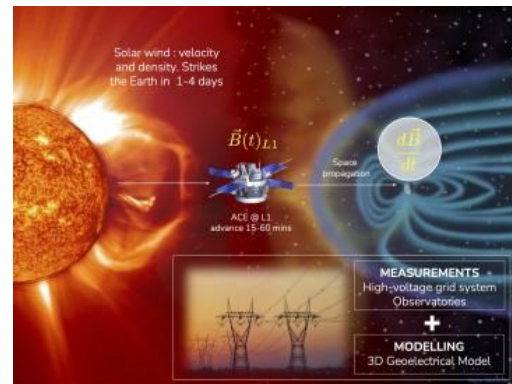
Agriculture

Particle Physics

Fine Arts
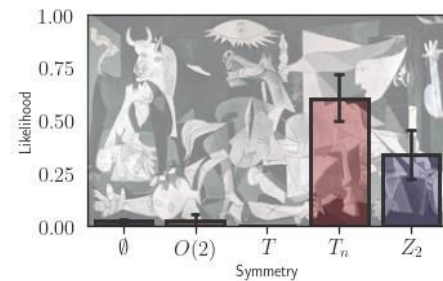
Artemisa Projects

3400 CPU Cores
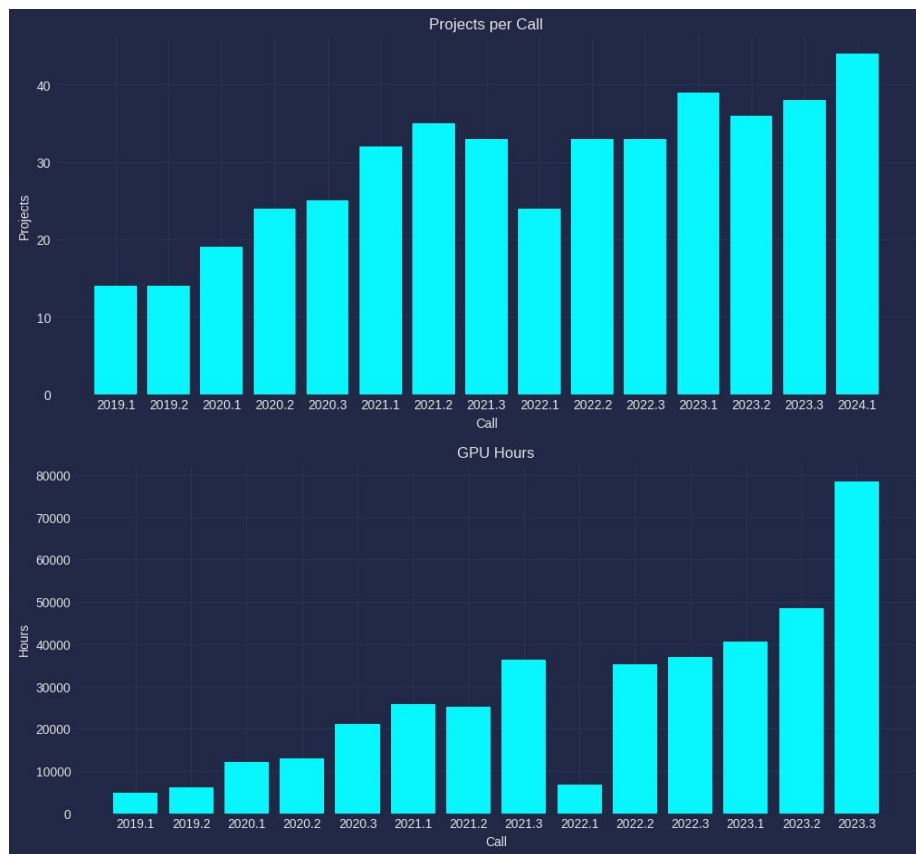
160.000 processing hours
delivered in 2023
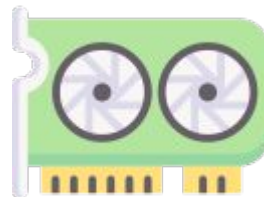
45 GPUs

150.000 processing hours
delivered in 2023

Artemisa Evolution

Projects per Call / GPU Hours

- Steadily increase of projects using Artemisa.

- Artemisa has achieved high engagement and growth rate Registration period for Call 2024.1 closed with:
  - 80% of projects from the previous call were extended.
  - 10+ new projects in this Call.

- GPU processing hours have increased from machine learning tasks and, particularly, deep learning tasks.

Infrastructure

- **User Interfaces (UI) :** entry point for users, provide a working environment where to compile and test their programs. These machines have a GPGPU and access to the storage services.
- **Worker Nodes (WN):** where the batch jobs are executed. Contain high-end CPUs, a larger memory configuration and several configurations of GPGPUs to execute the jobs.
- **Storage Nodes(SN):** store user and project data, accessible from both UIs and WNs.
- All nodes have a 10 Gbps ethernet connection to the IFIC's data center network.

#	Usage	General Characteristics	GPU
2	User interface	2 intel Xeon Gold 6130 (16c), 192 GB RAM	2 GPU NVIDIA
2	Batch	2 Intel Xeon Gold 8160 (24c), 384 GB RAM	1 GPU NVIDIA Tesla V100 32GB
20	Batch	2 Intel Xeon Gold 6248 (20c), 384 GB RAM	1 GPU NVIDIA Tesla V100 32GB
11	Batch	2 AMD EPYC 7532 (32c), 384 GB RAM	1 GPU NVIDIA Ampere A100 40GB
1	Batch	2 Intel Xeon Platinum 8180 (28c), 768 GB RAM	4 GPUs NVIDIA Tesla V100 32GB SMX2 with NVLink
1	Batch	2 AMD EPYC7642 (48c), 512 GB RAM	8 GPUs NVIDIA Ampere A100 40GB SMX2 with NVLink
5	DiskServers	387 TB Lustre	-
3	DiskServers	150 TB Lustre (SSD)	-

Infrastructure

UIs

WNs

mlui01

mlui02

Batch

2 x P100 GPUs (testing)

43 x A100/V100 GPUs
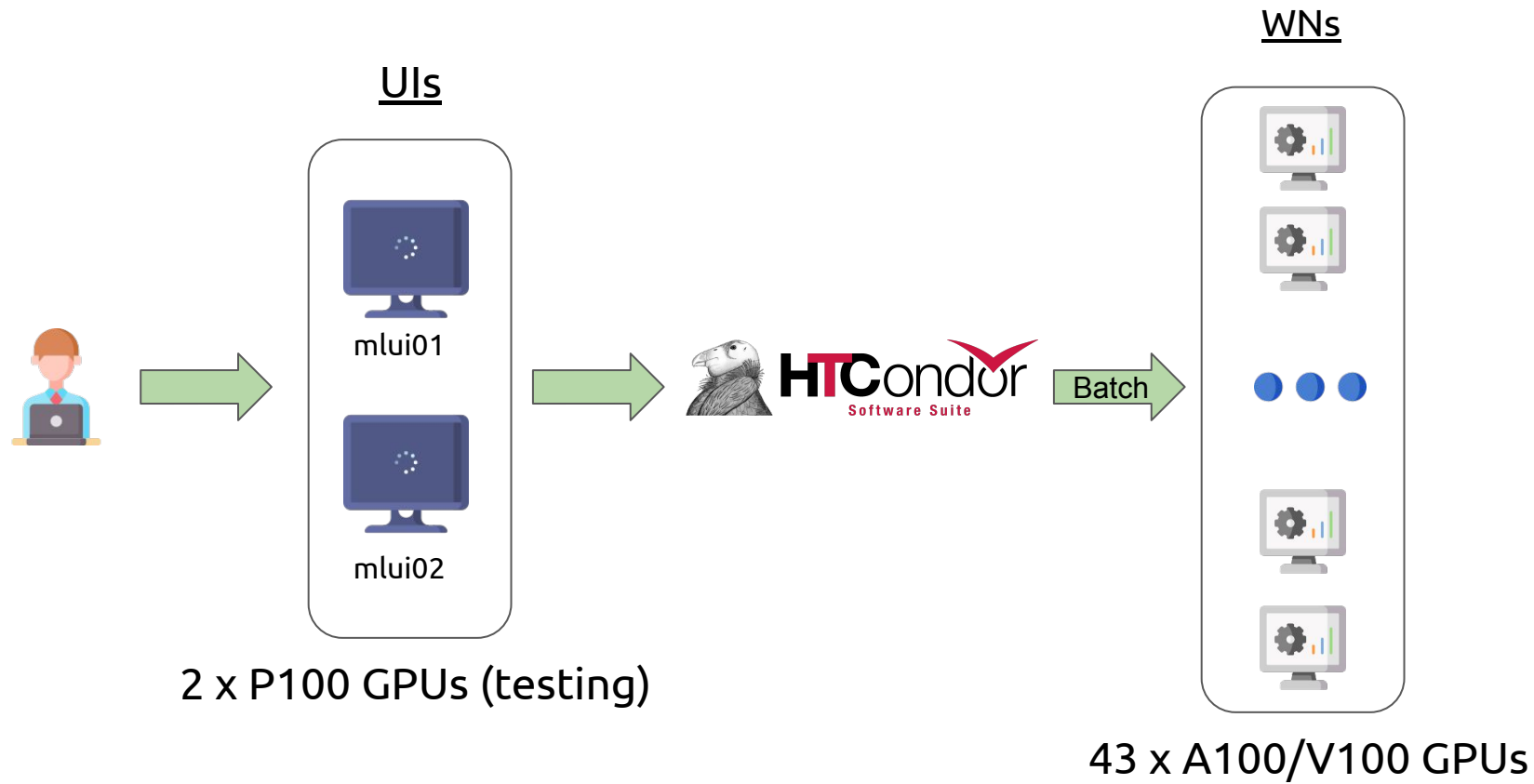
Environment

OS　　　　　Batch management　　Distributed storage

Machine Learning, Deep Learning libraries

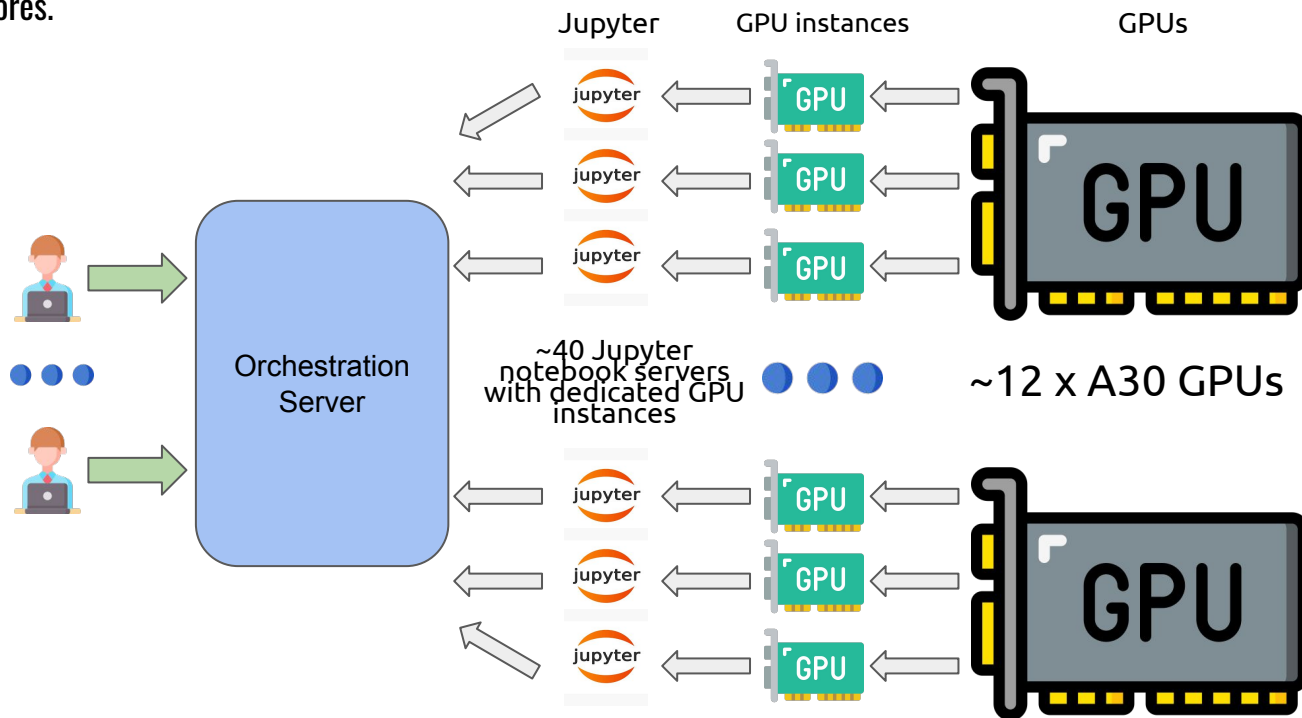## Virtual environments through venv, Anaconda, etc

Work Plan

- Documentation and user experience improvement:

  - Improved Artemisa User Guide, more content is being added covering encountered user needs.

  - Tutorials and seminars for user support and promotion.

  - Improve project management for users and evaluators.

- Infrastructure general developments:

  - Improve system resource monitoring.

  - Keep the infrastructure up-to-date adding servers with state of the art GPUs, even in a high-demand market.

- Upgrading the infrastructure in the short term (2024-2025)

  - New servers dedicated to Artemisa online usage

  - These servers will implement Jupyter Notebooks with Jupyterhub or similar

Future plans

NVIDIA Multi-instance can partition the A30 GPUs into as many as 4 instances, each fully isolated with its own high-bandwidth memory, cache, and compute cores.

Jupyter · GPU instances · GPUs

Orchestration Server

~40 Jupyter notebook servers with dedicated GPU instances

~12 x A30 GPUs

Infrastructures for computing and AI

To summarize the two talks:

Computación Avanzada para el procesado intensivo de Big Data en ATLAS

- ATLAS Tier2 experiment contribution with storage and processing machines
- Further development of the computing infrastructure for next years' pledges (HL-LHC)
- Replacing decaying infrastructure and future points of failure in storage
- Evolution of the Analysis Facilities for more interactive and user friendly use

Artificial Environment for ML and Innovation in Scientific Advanced Computing

- Machine learning focused infrastructure
- Aimed at researchers and also to private institutions
- 3 calls per year
- All fields of science present in past projects
- High use of GPU processing
- Future plans for new GPU servers to allow partitionable interactive GPU usage

THANKS

Highlights

- New member, Matías Salinero, has joined the team to provide support and further develop the infrastructure (contracted through ASFAE grant).

- First contract between a company and IFIC to use Artemisa facility. Agreement signed within the InnDIH (European Digital Innovation Hub) project.

- Projects are obtaining great results by using the infrastructure from inside and outside IFIC

   ‼️ BBVA "Best contribution in Statistics and Operations Research applied to Data Science and Big Data".

   ‼️ Artemisa is even in the acknowledgements of the ATLAS Collaboration.

Artemisa was created using funds from the FEDER 2014-2020 Comunidad Valenciana (IDEFEDER/2018/048, bugdet: 1 M€).

Granted a new project within the call "Recuperación y Resiliencia" (ASFAE/2022/024).

Project members

Jose Enrique García Navarro	PI
Bryan Zaldívar Montero	PI
Francisco Javier Albiol Colomer	
Roberto Ruiz de Austri Bazán	
Verónica Sanz González	

Twofold objective:

- Improve support to users and develop further infrastructure

- Increase computing resources by adding new machines

Artemisa (IFIC) is part of the InnDIH (European Digital Innovation Hub) as part of CSIC and Universitat de Valencia.

Future plans

- Artemisa User Guide is a living document, more content will be added covering encountered user needs.

- Tutorials and seminars for user support and promotion.

- Improve project management for users and evaluators.

- Migrating from CentOS 7 to AlmaLinux 9.

- Improve system resource monitoring.

- Keep the infrastructure up-to-date adding servers with state of the art GPUs, even in a high-demand market.

- Upgrading the infrastructure in the short term(2024-2025)
    - New servers dedicated to Artemisa online usage
    - These servers will implement Jupyter Notebooks with Jupyterhub or similar

Artemisa Support

- Support available by tickets or mail: artemisa-support@ific.uv.es

- ‼ New User Guide available:
  https://artemisa.ific.uv.es/docs/
  - System overview
  - FAQ and best practices
  - Step by step tutorials

- Improved user experience of Intranet:
  - Number of extensions of the project
  - Base project
  - Extend Project link and Help more accessible.

Financiado por la Unión Europea
NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Plan de Recuperación, Transformación y Resiliencia

GENERALITAT VALENCIANA
Conselleria de Educación, Universidades y Empleo

GVA NEXT
Fondos Next Generation en la Comunitat Valenciana