

PIC Status Report - J. Flix [on behalf of PIC team]

PIC port d'informació científica

IBERGRID 2023 Benasque (Spain)

















Spanish WLCG Tier-1 centre \rightarrow ~80% of resources

 \rightarrow Provides ~4% of Tier1 data processing of CERN's LHC detectors ATLAS, CMS and LHCb

¹⁄₄ of the Spanish ATLAS Tier-2 and a Tier-3 ATLAS data analysis facility → ~10% of resources

T2K and DUNE [neutrinos], MAGIC and CTA [gamma-ray astronomy], PAU and EUCLID [cosmology], VIP [instrumentation], opportunistic access to LIGO/VIRGO [gravitational waves], among others...

PIC farm



 \rightarrow Very old hardware switched off last winter (10% HEPScore CPU reduction)

 \rightarrow New purchases:

- 32 servers with AMD EPYC 7452 32-core processors (4096 vcores 11.8 HEPScore/vcore)
- 24 servers with AMD EPYC 7502 32-core processors (3072 vcores 12.8 HEPScore/vcore)
- \rightarrow 100% of compute nodes in dual-stack

 \rightarrow Some compute nodes migrated from CentOS 7 to AlmaLinux 9 (and HTCondor v10)



2x HTCondor-CE v5.1.6-1.el7 2x ARC-CE v.6.17.0-1 (used by ATLAS and LHCb as HPC gateways - see later)

HTCondor setup in PIC is <u>compatible with SciTokens</u>. LHC experiments submitting jobs to our HTCondor-CEs using tokens

18 GPUs available: via JupyterHub and (direct) acces by some VOs, also available through Grid

 \rightarrow 8 GeForce RTX 2080 Ti, 8 Tesla V100-SXM2-32GB, 2 GeForce GTX 1050 Ti

PIC disk storage

- ~18.6 PB running on **dCache 8.2.30**
 - \rightarrow New pools acquired to replace obsolete pools and increase capacity:
 - * 15x SuperStorage SSG-6028R-E1CR24N: 24 HDD SAS*18TB (~360 TB neto) and 2x25Gbps NIC
 - \rightarrow dCache pools in dual-stack
 - \rightarrow TPC enabled for HTTPs and XRootD and token authentication (PIC in DOMA testbeds)
 - \rightarrow dCache DDBB upgraded to PostgreSQL14

StashCache deployed as docker container (OSG repo) for Virgo/Ligo

- \rightarrow 3.2 TB 95% occupancy
- \rightarrow Running XRootD 5.4.2 (OSG 3.6)

xCache deployed (OSG repo) for the CMS experiment

- → 6TB disks (RAID6-175 TB). 48 cores E5-2650L v3 (HT enabled). 128 GB RAM. Bonding active-active 10 Gbps 95% occupancy
- \rightarrow Running XRootD 5.5.1
- \rightarrow Caching *AOD* files off-site, also acting as XCache for the Spanish CIEMAT Tier-2 site





PIC Ceph and NFS

Ceph Nautilus platform deployed on 2019

 \rightarrow 400TB net capacity

 \rightarrow Refurbished old HW adding SSD, memory and network interfaces \rightarrow Cephfs and RBD

Ceph Quincy platform deployed on June 2023

- \rightarrow 2PB net capacity. Pools configured with Erasure Code (k=8,m=2)
- → 4 head nodes (mons, mgr, mds, etc) & 12 OSDs (AMD 7443P, 256G RAM, 50 Gbps, 16HDD 18T, bluestore on NVMe)
- \rightarrow CephFS, RBD and S3 (testing)

* scratch data, RBD dCache images, kubernetes storage

** Ceph Nautilus data being migrated to this platform (90% completed)

NFS

- \rightarrow 200TB with backups
- \rightarrow Single server (SSG-6028R-E1CR24N)
- \rightarrow External snapshots on old servers



port d'informació

Expansion of the new tape library

RETIRE



IBM TS4500 (64 PB capacity):

- \rightarrow 5 frames (L55+D55 + 3xS55) + 10 LT08 drives + 11 LT09 drives
- \rightarrow 4.8 PB capacity installed with cartridges LT07 M8
- \rightarrow 58.4 PB capacity installed with cartridges LT08
- \rightarrow 2x frames and a second robotic arm to be purchased, which will increase system redundancy



port d'informació

This library is growing to host future data

 \rightarrow It hosts new data and data migrated from SL8500 library (<u>finished</u>) \rightarrow Dedicated drives, frames and cartridges installed to handle this

PIC currently runs Enstore 6.3.4-14 (CentOS 7)

 \rightarrow We started testing CTA as a potential replacement

SL8500

200 Gbps network at PIC



Successful network upgrade from **20 Gbps to 200 Gbps (29th June 2021)...** But, the GEANT - REDIRIS (NREN) connection were not migrated at 100 Gbps at that time due to lack of hardware.

- March 2022: the **LHCONE** was upgraded from 10 Gbps to 50 Gbps (used NOTED to load-balance LHCOPN with LHCONE)
- November 2022: the LHCOPN was migrated from 10 Gbps to 100 Gbps (switched off NOTED)
- January 2023: the LHCONE was upgraded from 50 to 70 Gbps
- January 2023: new VLANs for IPv4 and IPv6 CERN-PIC traffic deployed in the LHCOPN
- July 2023: the LHCONE was upgraded from 70 to 100 Gbps



PIC - Port d'Informació Científica @pic_es - Jan 21 ···· Network traffic in&out of PIC a couple of days ago, peaking over 120 Gbps. This is Distributed High Throughput Computing #DHTC in action. Thanks @LaRedIRIS @CSUC_info and @GEANTnews for enabling our high speed #network for #sciencedata!



Integration of BSC CPU resources

PIC port d'informació científica

In 2020 BSC designated LHC computing as a **strategic project**

• Agreement promoted by WLCG-ES community and funding agency

Allocations* of up to a 7% share of MN4 for LHC [100M coreHours max/year]

• ~70M coreHours/year

* Submission of proposals for time allocation every 4 months

Potentially, very **significant contribution** for LHC computing in Spain

• Comparable e.g to all ATLAS+CMS+LHCb simulation needs in the country



Use of the BSC resources through PIC



The LHC experiments have utilized **95 million hours** of resources at the Barcelona Supercomputer Center (BSC) MareNostrum4 HPC facility through services installed at PIC **since 2020**

This corresponds to an average installed capacity of approximately 50 kHS06, representing around **30% of the current grid resources deployed at PIC** for the LHC experiments

The new period, 2023-P2, started in July 1st 2023



Use of the BSC by ATLAS through PIC



Submitting **ATLAS** payloads to BSC from PIC Tier-1 since 2018, <u>in production since 2019</u> Using two **ARC-CEs** at PIC to interconnect MareNostrum and ATLAS production system <u>Only simulation workflow</u> validated - singularity containers, pre-placed at MareNostrum GPFS ~**15 million hours** approved and used at BSC **every year** by ATLAS through PIC gateways **Other gateways available at the Spanish ATLAS Tier-2s**

PIC

port d'informació científica

 $[\]rightarrow$ At CHEP2021 proceedings (link)

Use of the BSC by CMS through PIC



PIC

port d'informació científica

Use of the BSC by CMS through PIC



<u>Current status</u>

- Running in operations (CMS Simulation workflows GEN-SIM)
- Result of the **PIC and HTCondor team collaboration** to use a **shared FS as control path for HTCondor**
- Interaction with BSC execute nodes through the login node, mounting the shared FS through sshfs and sending jobs to the Slurm scheduler via ssh.
 Slurm jobs launch a HTCondor slot that joins the CMS Global Pool
- Using cvmfs_preload to bring cvmfs CMS files to BSC. Two weeks to copy ~37M files (13 TB), at first injection. cvmfsexec used to build the cvmfs file structure
- **Stage-in/out** + **Data Transfer Manager** designed to transfer input and output data from/to PIC (*xRootD server in singularity images*)
- Integrated with WMAgent @ CERN Accounting to APEL ongoing
- A new grant of 7.5 Mhours (typical quarter allocation for CMS)
- → At CHEP2021 proceedings (link)

 \rightarrow At CHEP2023 (<u>link</u>)

PIC

port d'informació científica

Use of the BSC by LHCb through PIC



LHCb used similar technical implementations as ATLAS (**ARC-CE02.PIC.ES**) to exploit BSC resources - submitting grants to BSC as ATLAS and CMS, and **modified DIRAC** for the purpose

LHCb Submit Host Jul 2022 Aug 2022 Oct 2022 Nov 2022 Feb 2023 Mar 2023 Total Sep 2022 Dec 2022 Jan 2028 Percer ce13.pic.es:9619/ce13.pic.es-condor 8.399 11,125 14.649 14,744 14,141 12,499 24.009 14,431 7.968 121,965 34.34% 8.251 11,252 10,358 ce14.pic.es:9619/ce14.pic.es-condor 14,314 13,371 14.006 12,116 24,172 13.654 121,494 34.21% gsiftp://arc-ce02.pic.es:2811/jobs 0% https://arc-ceO2.pic.es:8443/arex 13,640 16,418 8,308 16,836 463 15,714 21,962 9,387 8,929 111.657 31.44% 30,291 38,795 37,274 44,951 28,610 40,329 70,143 37,472 27,255 355.120 8.53% 10.92% 10.50% 12.66% 8.06% 11.36% 19.75% 10.55% 7.67%

PIC JupyterHub



Web browser

JupyterLab session



Terminal session for the users. They can use their own Python environments

PIC JupyterHub: Dask integration

Dask scales data science libraries like Numpy and Pandas to multiple machines

Based on **low level task parallelism**, allowing parallelization of custom codes

Integrated into Jupyter so the user can start and monitor a cluster from the GUI

The started cluster request PIC resources **through HTCondor**

The **adaptive cluster size** can scale up and down based on the workload





CosmoHub



HUB

Build your own Universe

Interactive data analysis of massive cosmological data without any SQL knowledge



Billions of observed and simulated galaxies

Superfast queries means superfast results



Features to make you work faster and easier

Online plotting preview and data download

IberGrid 2023 - PIC status report [J. Flix]

CosmoHub: new Hadoop Cluster

30 nodes:

- 720 cores, 15 TiB RAM
- 60 TB NVMe (for cache)
- 4.3 PB raw storage (2.5 PB usable)

Main use cases:

- CosmoHub query processing
- Euclid mock galaxy catalogs
- HTCondor backfilling, specially suited for ephemeral/adaptive Dask clusters





Convocatoria 2021 de ayudas para la adquisición de equipamiento científico-técnico Clúster Hadoop para la gestión integral de datos científicos masivos" (EQC2021-007479-P) Funding granted: 401kE



The new PIC Science Platform



Coordinating "Línea 8: Computación, big data e inteligencia artificial" of the **Plan Complementario** de Astrofísica y Física de Partículas (2022-2025)



Developing of a science platform to efficiently analyze multi-messenger astronomy data

<u>Expanding the CosmoHub</u> application into a **Science Portal:** will allow access to multi-messenger data and analyze it using different tools: e.g. Jupyter notebooks, SQL, Visualization tools, etc...

Step 5b: Guided Tasks · Run predefined complex functions (Experimental)		
Clustering	⊙ нор	Photo Z
PROOF OF CONCEPT: Please read this first The package used to estimate the galaxy clustering is Treecorr II The 2-point function estimator used is the projected 2-point correlation function Trors are estimated using the jacknift etchnique II Note that this is a proof of concept and it only works for very particular samples: -volume limited samples - "ow" redshift (2 < -0.5)		
RA Column for RA Min. Sep. Min distance in Mpc/h (0.15-50)	Dec Column for Dec Max. Sep. Max distance in Mpc/h (0.15-50) Øc Request	r Column for r

Spanish Supercomputing Network (RES)



National HPC network created in 2007 (BSC coordinator - owns >90% of resources)

In 2019 the Ministry broadened RES scope to include **data services**. PIC applied to an open call to add new data nodes and now is part of the RES since Sep 2020

RES resources are allocated through periodic competitive **<u>open calls</u>**

3 open calls for RES data projects (2021,2022, and 2023)

- <u>7 projects</u> got space allocations at PIC
 - \rightarrow ATLAS, CMS, PAUS, MAGIC
 - \rightarrow Bioimaging (ICFO)
 - \rightarrow Earth Modelling (USC) backup
 - → Genomics (IBB/UAB)

As part of RES, PIC is now in the **RI map** for Spain

Natural forum for HPC-HTC workflows development



PIC enters the ICTS map (2022)



PIC has been included in the new map of the Singular Scientific and Technical Infrastructures (**ICTS**) of Spain, as a node of the Spanish Supercomputing Network (RES)

The inclusion of PIC in the new ICTS map* is a milestone, as it will open up new funding opportunities with other spanish research groups facing data analysis challenges





* More information on IFAE News

New applied Al group

New scientific group at PIC from autumn 2022

Works on deep learning in different fields, aiming to developing synergies

Ongoing work in **cosmology**, **material science**, **bio imaging** and **quantum computing**

Collaboration or interactions with theory, GW and neutrino groups

Teaching of deep learning methods

Involved in **developing infrastructure**, like the **Dask** integration





InCAEM Project

In-situ Correlative Facility for Advanced Energy Materials

Correlative in situ experiments **combining** (S)TEM (Scanning Transmission Electron Microscopy), <u>AFM/STM</u> (Atomic Force Microscopy / Scanning Tunneling Microscopy) and synchrotron radiation

Structure \leftrightarrow Function Operando & in situ Multi-modal & multi-lengthscale analysis

Advanced data analysis: HPC/HTC, deep learning,...

PIC will collaborate with ALBA Synchrotron to build and provide the computing infrastructure for data handling and analysis





https://www.icmab.cat/incaem-workshop-at-alba-synchrotron-on-advanced-materials-imaging















port d'informació científica

IberGrid 2023 - PIC status report []. Flix]

Quantum Spain Project

PIC participates in Quantum Spain project to **deploy a quantum computer in Spain**. Part of the future user support team

Promoted by the Ministry of Economy through the Secretary of State for Digitization and Artificial Intelligence and financed with the Recovery Funds

Budget: €22 million Execution: 01/01/22 – 31/12/25

<u>Goals</u>

- Acquisition and installation of a quantum computer based on superconducting qubits technology
- Create a remote access system in the cloud
- Develop useful quantum algorithms, applicable to real problems

https://quantumspain-project.es/en







port d'informació

Summary



Lot of **work done at PIC to get prepared for future**, which included a major migration to a new tape storage system and a upgrade/re-design of the WAN connectivity

Integration work done to **use a fraction of BSC CPU resources** for ATLAS, CMS and LHCb experiments through PIC gateways

PIC is part of the national **Data services RES nodes** and it is included in the **ICTS** map since March 2022

PIC center getting **reinforced** in the support for **other scientific disciplines**. PIC selected as **one of the four CTA data centers**, **collaboration with ALBA for material sciences**, ...

Increased funding obtained from equipment calls and complementary plans. Additionally, the LHC project is financed with additional MRR funds for the next 3 years



Thanks! Questions? Credits to: E. Acción, V. Acin, C. Acosta, A. Alou, M. Børstad, A. Bruzzese, L. Cabayol, E. Carrasco, J. Carretero, J. Casals, R. Cruz, M. Delfino, J. Delgado, J. Flix, E. J. González, D. Graña, G. Merino, C. Neissner, A. Pacheco, C. Pérez, A. Pérez-Calero, E. Planas, M.C. Porto, B. Rodríguez, P. Tallada, J. Priego, F. Torradeflot

www.pic.es

Acknowledgements

The authors of this work express their gratitude to the PIC and CIEMAT teams for their support in these studies and for deploying novel cache services for the CMS experiment in the Spanish region. This project is partially financed by the Spanish Ministry of Science and Innovation (MINECO) through grants FPA2016-80994-C2-1-R, PID2019-110942RB-C22 and BES-2017-082665, which include FEDER funds from the European Union. It has also been supported by the Ministerio de Ciencia e Innovación MCIN AEI/10.13039/501100011033 under contract PID2020-113614RB-C21, the Catalan government under contract 2021 SGR 00574, and the Red Española de Supercomputación (RES) through the grant DATA-2020-1-0039.





Credits to: E. Acción, V. Acin, C. Acosta, A. Alou, M. Børstad, A. Bruzzese, L. Cabayol, E. Carrasco, J. Carretero, J. Casals, R. Cruz, M. Delfino, J. Delgado, J. Flix, E. J. González, D. Graña, G. Merino, C. Neissner, A. Pacheco, C. Pérez, A. Pérez-Calero, E. Planas, M.C. Porto, B. Rodríguez, P. Tallada, J. Priego, F. Torradeflot

www.pic.es