Updates to the ARNES HPC Cluster in 2025 and Plans for 2026

In 2025, ARNES continued the intensive development and upgrade of its high-performance computing (HPC) infrastructure, further strengthening the capabilities of the national environment for advanced computing. The ARNES computing cluster, in operation since 2009, represents one of the key processing resources for the Slovenian research and education community. With a total computing power of approximately 2.3 petaFLOPS, it is the second most powerful supercomputer in Slovenia and the only one accessible both to beginners and to more advanced users.

The cluster is part of the SLING consortium, and its capacities are available to researchers and students through the Open National Access framework. It is used in educational processes as well as in research activities in the fields of artificial intelligence, knowledge technologies, theoretical physics and particle physics, genomics, computational chemistry, and big data processing.
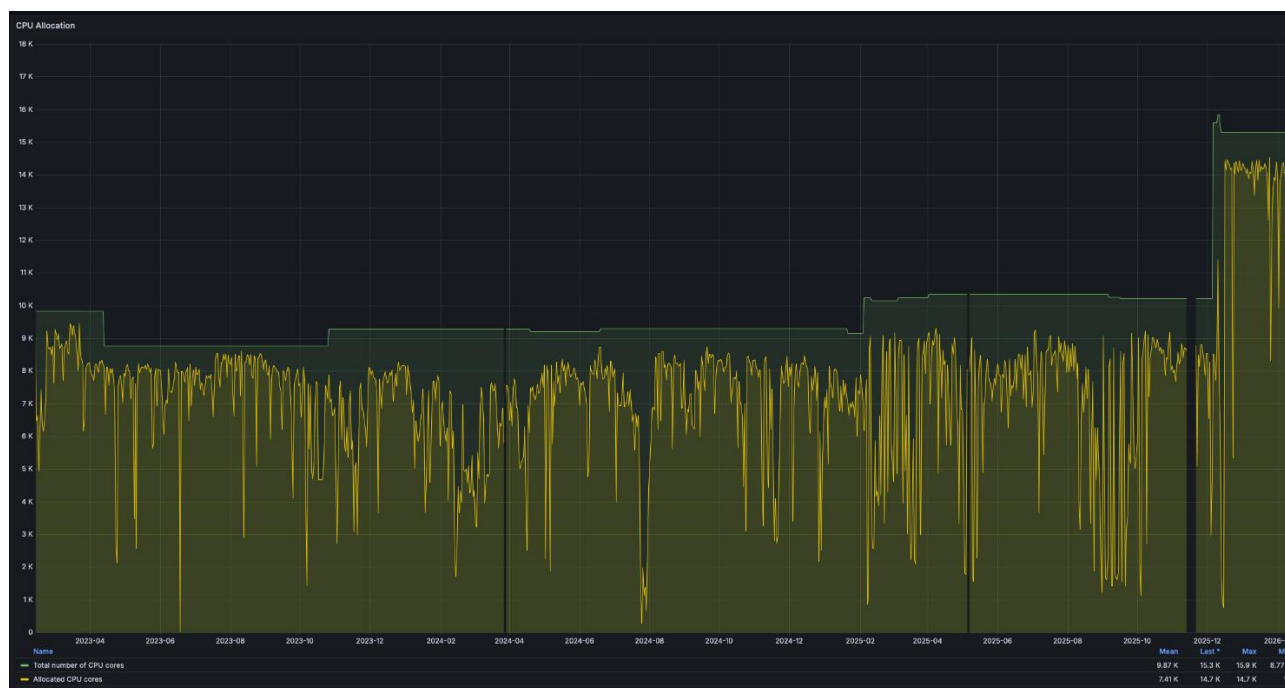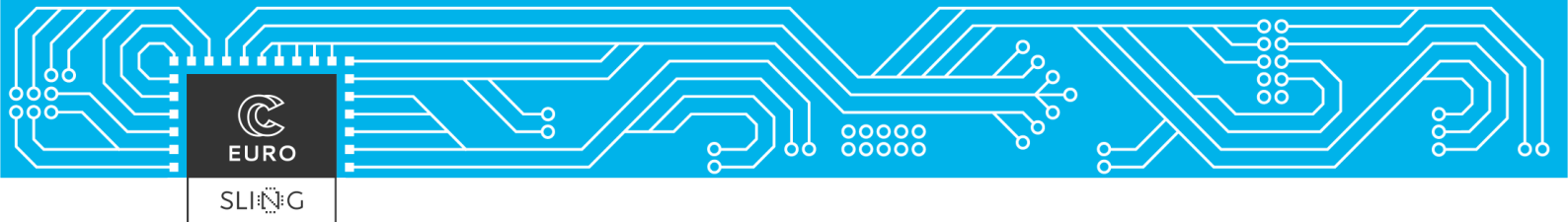
Key Upgrades in 2025

In 2025, ARNES carried out a major hardware renewal, replacing older CPU nodes with new, more powerful units equipped with AMD EPYC 9745 processors and up to 1 TB of RAM. In total, 41 older CPU nodes were replaced with significantly more capable successors. By the end of 2025, the theoretical performance of the CPU part of the cluster reached approximately 0.5 PFLOPS.

A major upgrade was also implemented in the area of GPU accelerators. Three new GPU nodes were added, each equipped with four H100 SXM5 accelerators, providing substantially higher performance compared to the previous generation. At the same time, seven nodes with V100S accelerators were retired. The cluster currently includes 27 GPU nodes with a total of 60 accelerators. Despite the slightly lower number of accelerators, the overall performance of the GPU part of the system increased and reached approximately 1.8 PFLOPS by the end of the year.

ARNES also upgraded the software environment. Compute nodes and the Ceph infrastructure were updated to AlmaLinux 9, while the Ceph storage cluster was gradually upgraded first to version 18 (Reef) and then to version 19 (Squid). Several older disk units were replaced as well, increasing the available storage capacity for users to more than 4 PB of usable space.

Record Usage and Increasing Workload Complexity

The year 2025 was also marked by record utilisation of the supercomputer. In December, users completed 7,761,633 CPU hours, representing a new monthly record. The largest share of resources is currently consumed by the ATLAS project, linked to CERN research and the discovery of the Higgs boson.

REPUBLIC OF SLOVENIA
MINISTRY OF HIGHER EDUCATION, SCIENCE AND INNOVATION

The graph shows the CPU utilisation of the Arnes cluster.

ARNES executes nearly 665,000 jobs per month on its HPC cluster. A clear shift in usage patterns can be observed, as an increasing share of demanding workloads is moving towards GPU resources, largely driven by the rapid expansion of artificial intelligence. Jobs are becoming longer and more complex, and ARNES is continuously adapting the development of its infrastructure to meet these evolving needs.

```
--------------------------------------------------------------------
Cluster Utilization 2025-12-01T00:00:00 - 2025-12-31T23:59:59
Usage reported in TRES Hours
--------------------------------------------------------------------
  Cluster      TRES Name    Allocated      Down PLND Dow      Idle    Planned    Reported
--------- -------------- ----------- ---------- ------- ---------- ---------- -----------
    arnes            cpu     7761637     441522   19656    1476534    1181122    10880470
    arnes       gres/gpu       21318       3226    2056      30848          0       57448


--------------------------------------------------------------------
Cluster Utilization 2026-01-01T00:00:00 - 2026-01-31T23:59:59
Usage reported in TRES Hours
--------------------------------------------------------------------
  Cluster      TRES Name    Allocated     Down PLND Dow      Idle    Planned    Reported
--------- -------------- ----------- -------- -------- --------- ---------- -----------
    arnes            cpu    10163148    22331        0     21661    1179036    11386176
    arnes       gres/gpu       24019      419        0     29130          0       53568
```

In December 2025, users completed 7,761,637 CPU hours on the Arnes cluster, and in January 10,163,148 CPU hours, as shown in the outputs above.

Simplifying Access: JupyterHub and AAI Authentication

An important milestone in 2025 was the successful pilot of a new service based on JupyterHub, enabling users to access JupyterLab and Jupyter Notebook environments directly through a web browser. Authentication is provided via the unified AAI identity system, and the service is connected to the ARNES HPC

cluster, making use of both CPU and GPU resources. The production version of this service will be made available to users in 2026.

Plans for 2026: Stability, European Integration, and Future Capacity

In 2026, ARNES plans further activities in three key areas: ensuring the stability and development of the existing infrastructure, integration into European research infrastructures, and planning future computing capacities.

One of the strategic goals is the establishment of a unified national platform through GPU hardware pooling, which would enable cost-efficient deployment of advanced AI solutions and provide access to state-of-the-art equipment also for smaller research groups.

ARNES will continue simplifying access to computing resources in the direction of "computing capacity as a service" and will further improve backend systems through updated configuration management, an expanded monitoring system, and more advanced error notification mechanisms.

Among the most important technical plans for 2026 are also:

- **the deployment of a production JupyterHub environment,**

- **the preparation of an integration plan for the EOSC node,**

- **the renewal of the central CVMFS system and integration of EESSI,**

- **the implementation of RoCE technology for low-latency inter-node communication,**

- **upgrading the Slurm workload manager,**

- **the optimisation of Ceph storage with the inclusion of NVMe nodes,**

- **the development of a new user management system,**

- **and the exploration of virtualisation options using OpenStack.**

At the same time, ARNES will continue strengthening user support, improving access processes, and maintaining a high level of cybersecurity. Collaboration within the strategic Slovenian AI Factory project (SLAIF) remains particularly important, representing a major step forward in the development of artificial intelligence capabilities in Slovenia.

REPUBLIC OF SLOVENIA
MINISTRY OF HIGHER EDUCATION, SCIENCE AND INNOVATION

EuroHPC
Joint Undertaking