Dedicated Analysis Facility for HEP Experiments

Gábor Bíró[1,2], **Gergely Gábor Barnaföldi**[1], **Péter Lévai**[1]

[1]Wigner Research Center for Physics, 29–33 Konkoly–Thege Miklós Str., H-1121 Budapest, Hungary.

[2]Institute of Physics, Eötvös Loránd University, 1/A Pázmány Péter Sétány, H-1117 Budapest, Hungary.

E-mail: biro.gabor@wigner.hu; barnafoldi.gergely@wigner.hu; levai.peter@wigner.hu

9 March 2023

Abstract. High-energy physics (HEP) provides ever-growing amount of data. To analyse these, continuously-evolving computational power is required in parallel by extending the storage capacity. Such developments play key roles in the future of this field however, these can be achieved also by optimization of existing IT resources.

One of the main computing capacity consumers in the HEP software workflow are detector simulation and data analysis. To optimize the resource requirements for these aims, the concept of a dedicated Analysis Facility (AF) for Run 3 has been suggested by the ALICE experiment at CERN. These AFs are special computing centres with a combination of CPU and fast interconnected disk storage modules, allowing for rapid turnaround of analysis tasks on a dedicated subset of data. This in turn allows for optimization of the analysis process and the codes before the analysis is performed on the large data samples on the Worldwide LHC Computing Grid.

In this paper, the structure and the progress summary of the Wigner Analysis Facility (Wigner AF) is presented for the period 2020-2022.

1. Introduction

The largest detectors of the Large Hadron Collider (LHC) underwent major upgrades during the Long Shutdown 2 (LS2) in the period, 2019-2022 [1]. Detector sensitivity, readout hardware, indeed the associated online and offline softwares were replaced and modernized. The goal of the R&D activities was to enable the experiments to pursue new physics in the Run-3 data taking period (2022-2025) and beyond.

For these aims, efficient data processing was investigated on large data samples from LHC's Run 1 and Run 2. The performance of the Monte Carlo simulations were also tested and optimized for massively parallel event generation on the Worldwide LHC Computing Grid (WLCG). Finally, the aim is to achieve the best computing performance beside keeping the maintenance and operation costs at a reasonable level – despite the age of the existing hardware components.

The Analysis Facility at the Wigner Datacenter (WDC) were established alongside the original structure, inherited by the former CERN's Tier 0 site (Budapest,

Hungary) [2]. Based on the existing hardware, the topology and the modules were further optimized according to the needs of rapid data campaigns of the experimental requirements. The original idea of the offline Analysis Facility (AF) [3] was applied first in large scale at the Wigner AF, where the recent multi-core analysis software framework Hyperloop [4] was also tested.

2. Structure of the Analysis Facility

The Wigner Analysis Facility is the part of the Wigner Scientific Computing Laboratory (WSCLAB), located physically in the WDC. The majority of the Wigner AF's hardware is built from the legacy hardware of the Budapest Tier 0 computing center, mostly AMD Opteron 6276 CPUs [5]. The main purpose of the analysis facility is to efficiently process a considerable amount, $\mathcal{O}$(PB) of data on a daily basis while being able to scale up the resources, $\mathcal{O}$(15%) per year. For this reason, it is essential to have a modular design for both the storage and compute parts, and to ensure high bandwidth communication between them.

After several hardware tests and bandwidth optimization cycles, a dual rack-based 'cell' has been chosen as scalable unit of the Wigner AF. Such a standalone working unit is composed by compute, frontend, storage, and service elements, as illustrated in Fig. 1. Each of the 8 *compute chassis* includes 4 dual processor machines, totaling 1024 threads per cell. The cells process the analysis jobs submitted through a dedicated interface called VO Box [6], which serves as an entry point to the AF from the global WLCG system. The jobs then passed to the HTCondor [7] and HTCondor-CE [8] servers which distribute them among the connected worker nodes.

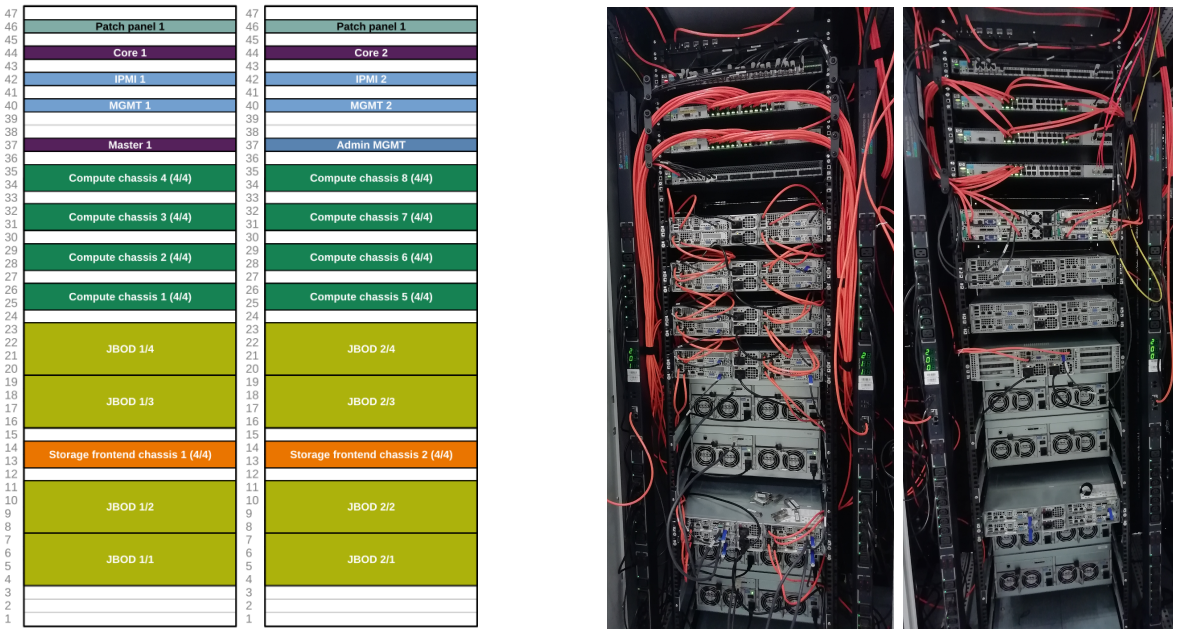

Figure 1. The structure of the a single cell in the Wigner Analysis Facility.

The storage element of a cell consists of a JBOD chassis with 24 disks, controlled by the machines of the frontend chassis through XRootD [9] and EOS [10, 11] services and daemons. The collection of such File Storage Server (FST) nodes is managed by the Management Server. Each of the mentioned server machines with management services is provided with a trusted grid server certificate to ensure the seamless connection to the WLCG infrastructure.

The OS level orchestration of the machines is achieved through the Metal-As-A-Service (MAAS) data centre automation developed by Canonical [12]. The co-location of compute, storage, and network nodes in the same cell serves the purpose of assuring a fast data transmission required by the analysis workflow. The high-speed internal communication between the nodes is ensured by HP ProCurve 6600-24XG (J9265A) switches [13]. Utilizing the SFP+ 10 GbE ports, a high bandwidth of 10 Gbps is achieved within a cell and also between other cells.

Two chassis are maintained in the first working cell uniquely for special purposes. For future developments, a compute chassis is dedicated to machines with graphic accelerator cards (GPUs), while the machines of another compute chassis serve management roles.

3. Computing capacity and network connectivity

The Analysis Facility concept was tested within the CERN's ALICE experimental framework with 4 cells (8 racks in total) at the WSCLAB, comparable to a mid-sized Tier 2 site [2]. The site is located at the KFKI campus, Budapest, Hungary and it is part of the Wigner Datacenter, which is connected to the GEANT network by new devices with 100 Gbps-capable link. The total storage and computing power of the site is summarized in Table 1.

Table 1. The summary of the total resources of the Wigner AF

Total Storage Size	Total Computing Resource
32 FST node	Queues for single-core and multi-core jobs
24×3 TB raw capacity per FST node	128 worker nodes
Total raw capacity: $\sim$2.2 PB	32 vCPU, 64 GB RAM, for each node
Usable capacity with RAID-1: $\sim$1.1 PB	4096 logical cores in total

3.1. Benchmarking the computing resources

In parallel with the hardware installation, a set of performance and optimization tests were performed. The execution of the first, *pilot* jobs occurred in late 2020, still during the hardware and software setup phase of the AF, while the production period with a high job success rate started in February, 2021. The performance test of the 8-rack setup with realistic analysis workload was performed in September 2021, repeated in February 2022 (see Figure 2). The I/O rate increase, therefore the performance of

the AF shows that an optimization of the analysis framework software is essential to fully utilize the capabilities of the underlying hardwares. It can be seen well, that the optimized structure results in the same I/O and +20% analysis throughput for the single core jobs, while for octa-core ones +10% I/O with almost the same analysis throughput.

Our tests show that the theoretical throughput (including the data I/O overhead) of the current setup has a peak at 1.1 PB/day. In average, the Wigner Analysis Facility performed ~ 4 MB/s analysis performance and 18 MB/s and 43 Mb/s I/O rate for the single and octa core jobs, respectively.

Figure 2. Estimated I/O rate and analysis throughput for single- and octa-core jobs.

4. Conclusions

The presented Wigner Analysis Facility is one of the first instances of the throughput-specialized computing facilities that will become increasingly common among high-energy physics experiments in the future. This has been optimized for the specific task tested with the analysis framework and data provided by the CERN ALICE Experiment.

Keeping in mind the infrastructural costs as well (electricity, High Throughput Computing hardware), the Wigner AF can provide a maintainable and scalable solution to the future computational challenges, such as gravitation wave analysis (LIGO/VIRGO) and nuclear databases within the EUPRAXIA project. This knowledge and the site is also open for other large-scale collaborations.

5. Acknowledgements

The research was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under the contract numbers OTKA K135515, and 2019-2.1.6-NEMZ_KI-2019-00011, 2022-4.1.2-NEMZ_KI-2022-00009, and 2022-4.1.2-

NEMZ_KI-2022-00008. The authors would like to express their gratitude to Ádám Pintér, József Kadlecsik and the technical staff of the Wigner Datacenter for the setup of the Analysis Facility hardware. We appreciate the support of the WLCG management.

References

[1] "LHC Machine". In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[2] "The Wigner ALICE Analysis Facility". In: (2021). URL: https://cds.cern.ch/record/2791181.

[3] Schwarz, Kilian et al. "The ALICE Analysis Facility Prototype at GSI". In: *EPJ Web Conf.* 214 (2019), p. 08027. DOI: 10.1051/epjconf/201921408027. URL: https://doi.org/10.1051/epjconf/201921408027.

[4] Raquel Quishpe et al. "Hyperloop – The ALICE analysis train system for Run 3". In: *9th Large Hadron Collider Physics Conference*. Sept. 2021. arXiv: 2109.09594 [physics.ins-det].

[5] *AMD Opteron 6276*. Accessed: 02. 03. 2023. URL: https://www.amd.com/en/products/cpu/6276.

[6] *WLCG VOBOX*. Accessed: 02. 03. 2023. URL: https://twiki.cern.ch/twiki/bin/view/LCG/WLCGvoboxDeployment.

[7] *HTCondor*. Accessed: 02. 03. 2023. URL: https://research.cs.wisc.edu/htcondor/.

[8] *HTCondor-CE*. Accessed: 02. 03. 2023. URL: https://htcondor.github.io/htcondor-ce/.

[9] Fabrizio Furano and Andrew Hanushevsky. "Scalla/xrootd WAN globalization tools: Where we are". In: *Journal of Physics: Conference Series* 219.7 (2010), p. 072005. DOI: 10.1088/1742-6596/219/7/072005. URL: https://doi.org/10.1088/1742-6596/219/7/072005.

[10] Andreas J. Peters and Lukasz Janyst. "Exabyte scale storage at CERN". In: *J. Phys. Conf. Ser.* 331 (2011). Ed. by Simon C. Lin, p. 052015. DOI: 10.1088/1742-6596/331/5/052015.

[11] Geoffray Adde et al. "Latest evolution of EOS filesystem". In: *Journal of Physics: Conference Series* 608 (2015), p. 012009. DOI: 10.1088/1742-6596/608/1/012009. URL: https://doi.org/10.1088/1742-6596/608/1/012009.

[12] *Metal as a service*. Accessed: 02.03.2023. URL: https://maas.io/.

[13] *HP ProCurve 6600-24XG (J9265A)*. Accessed: 02. 03. 2023. URL: https://support.hpe.com/hpesc/public/docDisplay?docId=emr_na-c01832969.