Enabling HPC systems for HEP: The INFN-CINECA Experience

Tommaso Boccali

INFN Sezione di Pisa

L.go B. Pontecorvo 3, 56127 Pisa (ITALY)
E-mail: [*Tommaso.Boccali@cern.ch*](mailto:Tommaso.Boccali@cern.ch)

Stefano Dal Pra[1]
INFN-CNAF
Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)
E-mail: [*dalpra@infn.it*](mailto:dalpra@infn.it)

Stefano Zani
INFN-CNAF
Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)
E-mail: [*stefano.zani@cnaf.infn.it*](mailto:stefano.zani@cnaf.infn.it)

Lucia Morganti
INFN-CNAF
Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)
E-mail: [*lucia.morganti@cnaf.infn.it*](mailto:lucia.morganti@cnaf.infn.it)

Daniele Cesini
INFN-CNAF
Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)
E-mail: [*daniele.cesini@cnaf.infn.it*](mailto:daniele.cesini@cnaf.infn.it)

Vladimir Sapunenko
INFN-CNAF
Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)
E-mail: [*vladimir.sapunenko@cnaf.infn.it*](mailto:vladimir.sapunenko@cnaf.infn.it)

Daniele Spiga
INFN
Sezione di Perugia, Via Alessandro Pascoli 23c, 06123 Perugia (ITALY)
E-mail: [*daniele.spiga@pg.infn.it*](mailto:daniele.spiga@pg.infn.it)

Diego Ciangottini
INFN
Sezione di Perugia, Via Alessandro Pascoli 23c, 06123 Perugia (ITALY)
E-mail: [*diego.ciangottini@pg.infn.it*](mailto:diego.ciangottini@pg.infn.it)

---

[1]   Speaker

Francesco Noferini

INFN

Sezione di Bologna, Viale Carlo Berti Pichat 6, 40127 Bologna (ITALY)

E-mail: *Francesco.Noferini@bo.infn.it*

Concezio Bozzi

INFN

Sezione di Ferrara, Via Saragat 1, 44122 Ferrara (ITALY)

E-mail: *concezio.bozzi@cern.ch*

Stefano Perazzini

INFN

Sezione di Bologna, Viale Carlo Berti Pichat 6, 40127 Bologna (ITALY)

E-mail: *stefano.perazzini@bo.infn.it*

Andrea Valassi

CERN

Esplanade des Particules 1, P.O. Box 1211, Geneva 23 (SWITZERLAND)

E-mail: *andrea.valassi@cern.ch*

Federico Stagni

CERN

Esplanade des Particules 1, P.O. Box 1211, Geneva 23 (SWITZERLAND)

E-mail: *federico.stagni@cern.ch*

Alessandro De Salvo

INFN

Sezione di Roma1, P.le Aldo Moro 2, 00185 Roma (ITALY)

E-mail: *Alessandro.DeSalvo@roma1.infn.it*

Alessandra Doria

INFN

Sezione di Napoli, Via Cintia, 80126 Napoli (ITALY)

E-mail: *alessandra.doria@na.infn.it*

Luca dell'Agnello

INFN-CNAF

Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)

E-mail: *luca.dellagnello@cnaf.infn.it*

Gaetano Maron

INFN-CNAF

Viale Carlo Berti Pichat, 6/2, 40127 Bologna (ITALY)

E-mail: *gaetano.maron@cnaf.infn.it*

Abstract   In this report we want to describe a successful integration exercise between CINECA (PRACE Tier-0) Marconi KNL system and LHC processing. A production-level system has been deployed using a 30 Mhours grant from the 18th Call for PRACE Project Access; thanks to CINECA, more than 3x the granted hours were eventually made available. Modifications at multiple levels were needed: on experiments' WMS layers, on site level access policies and routing, on virtualization. The success of the integration process paves the way to integration with additional local systems, and in general shows how the requirements of a HPC center can coexist with the needs from data intensive, complex distributed workflows.

1 Introduction

This document reports the study and work performed to integrate a set of High Performance Computing (HPC) resources provided by the A2 Partition of the Marconi supercomputer, managed by CINECA [1], in order to make them able to run ordinary payloads such as those distributed through the WLCG [2] Grid Infrastructure to typical High Throughput Computing (HTC) centers.

HPC machines are designed to maximize performances of jobs using a large number of concurrent threads spanning several nodes and submitted by local users, whereas Grid jobs are made of single or eight core applications submitted by remote users and are intended to run on individual machines managed by several computing centers, designed with the goal of maximizing the overall job throughput from the employed machines rather than the execution speed of the individual jobs. The four High Energy Physic experiments running the LHC accelerator at CERN are the main players in the WLCG Grid. Italian researchers have historically been major contributors in the design, construction and operations of the LHC detectors, with funding from the Istituto Nazionale di Fisica Nucleare (INFN [3]). Moreover, Italy hosts 10 WLCG computing centers providing approximately 10% of the total resources, with a Tier-1 site located in Bologna (Italy), at INFN-CNAF. The need for computing power by the LHC experiments is expected to increase in the next decade, at a faster rate than what a fixed model budget could sustain, even considering the more optimistic scenario of CPU power growth allowed by the Moore law [4]. For this reason, looking for alternative sources of computing power has become a topic of crucial interest to investigate, and finding ways to enable on HPC facilities the capabilities needed to run WLCG payloads have been the motivation for the present work.

2 The three parties involved

There are three main actors involved in this work, whose tight interaction was key to success.

- INFN-CNAF, a WLCG Tier-1 site providing GRID and local resources for a total of nearly 40K CPU cores, 42 PB of disk and 95 PB of tape storage; it supports more than 30 research activities funded by INFN. Typical payloads running in the center are single core or eight-core jobs requiring up to 3GB RAM and up to 20GB swap per thread. The center was designed to sustain the data intensive usage of the LHC experiments, and for this reason it provides compute nodes having high bandwidth network access to local and remote storage resources. Furthermore, the INFN-T1 is connected through a 200Gb/s to LHCOPN, a dedicated high bandwidth network for High Energy Physics connecting CERN and Computing centers dedicated to High Energy Physics research. Finally, a dedicated direct dark fiber 600Gb channel is in place connecting CNAF and CINECA via DCI technology, with sites only 8 km far from each other. In fact, nearly 50% of the regular Compute Nodes managed by CNAF are hosted at CINECA for more than 3

years now and are built up as refurbished hardware from the former Marconi A1 partition. More are expected to be deployed in the upcoming months.

- CINECA, a member of the Partnership for Advanced Computing in Europe (PRACE) and a Tier-0 facility, hosts the HPC system Marconi, ranked no. 19 in the top500.org [5] at the time when this work started. The Marconi A2 partition counts 3600 nodes, equipped with 1 Xeon Phi 7250 (KNL) at 1.4 GHz, with 96 GB of on board RAM (out of which, 10GB reserved to the OS by administrators). Each CPU has 68 physical cores, with 4-way hyperthreading. The total 244,800 cores together can perform up to 11 PFlops. Each node has a small SSD local disk, not available for user activities. Temporary user disk space is provided by a large GPFS storage connected via Omni-Path [6]. Outbound connectivity is not allowed from these Compute Nodes. User jobs are submitted to a Slurm [7] Batch System arbitrating access to the resources. Authorization to use resources at CINECA can be requested within PRACE [8], by participating in a call system which grants CPU-hours after a review process. The Italian LHC Community successfully applied to the "18th PRACE Project Access Call for Proposals", and was assigned a grant of 30 Million CPU hours on the Marconi A2 partition (Project 2018194658). It is worth noticing that each Slurm job is assigned with a whole node as a minimum and accounting is evaluated accordingly, by counting the time elapsed on a node multiplied by the number of physical cores on the node (68, in our case). The grant was requested for specific physics studies, but also for initiating an handshaking procedure with CINECA in view of a larger utilization of CINECA's current and future systems for HEP computing. Indeed, the INFN-CNAF Tier-1 and the next CINECA HPC (Leonardo, a pre-exascale system expected by 2021) will be partially co-located in a new facility, and any expertise acquired in running LHC workflows in the current system would be precious in planning the Leonardo infrastructure.
- The italian groups of the four LHC experiments, ALICE, ATLAS, CMS, LHCB, supported by CERN colleagues, have a large and more than decennal expertise in the operations of the Grid facilities for job submission and data management, mainly located at CERN and conforming to the standards of the WLCG community.

3 Marconi A2 and WLCG requirements

A quick recap of how the pilot model adopted in the WLCG community works is described here. Pilots are "general purpose jobs", centrally submitted by a Workload Management System (WMS) owned and managed by each VO (experiment) toward several Grid sites. At the Grid site, the Computing Element (CE) is the endpoint contacted, and must be reachable via its public interface by the WMS. The HTCondor-CE [9] authenticates the client, validates its authorizations, and submits the pilot job to the local batch system (Slurm in our case) using local credentials uniquely mapped to the remote Grid client. The batch system dispatches the jobs to the KNL nodes and executes them. Upon start, the pilot jobs call back their WMS and receive payloads, which are executed in a Singularity [10] container, whose image is available from CVMFS. Input and output data are usually retrieved from or uploaded to storage resources using SRM/Xrootd and the Data Management Systems of the Experiment. These storage resources can be local at the site or remote.

The typical setup and the straightforward operating mode of a HPC cluster do not permit to directly run LHC workflows. A number of constraints must be relaxed, while others must be satisfied by adding a thin infrastructural layer in order to enable access to needed additional resources; all of these changes have been discussed and agreed with the CINECA administrators. Here follows a list of these constraints and the actions needed to deal with them.

- **Inbound connectivity and local submission.** Only local users can submit jobs to the Slurm batch system of the Marconi cluster, which is not network reachable from outside. The only way to enable Grid job submission is to add a CE which can be contacted by Grid submitters. These are authenticated by their X509 proxy certificate and authorized as members of a Virtual Organization by a VO Membership Service (VOMS). The CE can submit jobs to the local Slurm Batch System on behalf of the Grid user using one of a set of local Unix usernames (pool

accounts) which is uniquely mapped with the Grid counterpart. CINECA agreed to host a HTCondor Computing Element and allow connections from a set of CNAF and CERN networks.

- **Outbound connectivity and payload factories.** HPC jobs do not need outbound connectivity. LHC jobs however need to contact upon start a "payload factory" (a component of the WMS) managed by the VO and located at CERN, which assigns the pilot job with the actual task (payload) to execute. They might also need to contact a Storage Manager Service to upload their final output. For this reason, outbound connectivity is required. CINECA administrators agreed to enable outbound connections from Marconi nodes toward CERN and CNAF networks.

- **Read access to external shared file systems.** CVMFS is an http-based network filesystem where WLCG jobs have read access and can find needed libraries or other data they could need. File system contents are centrally managed by the experiments. To enhance performances and reduce network traffic, a local cache (squid) is kept locally at the centre. The CVMFS solution has been considered useful also outside the HEP community, and was already considered by CINECA.

- **Singularity.** Singularity is a software package providing a "lightweight" containerization technology which can work in user space. Several HEP communities adopt this solution to provide their jobs with a well-known and uniform environment. Singularity images are usually distributed via CVMFS. This makes it possible to start a job satisfying bare minimum requirements to start on the node and then launch the "second stage" inside a container. The only requirement on the Marconi nodes is to install the singularity package.

- **Access to CNAF and external storage resources.** Not strictly needed, but strongly useful in order not to overload CINECA's GPN, a partial 40 Gbit/s connection was established on the Infinera private link; this enabled jobs to directly access Input and Output data on storage available at CNAF through a fast dedicated channel, using the Xrootd protocol. For other sources, not directly reachable due to the routing limitations, an Xrootd proxy has been deployed at CNAF to cache the input files and then serve the executing processes. This was done in collaboration with the XDC EU Project [11]. The net result is that all remote files were accessible by the KNL nodes from the cache or directly. Output files are shipped to the CNAF storage system using SRM or Xrootd protocols, and as such registered in the Data Catalog of the Experiment.

4 What to run

Having satisfied the above requirements, Grid jobs can be submitted to the HTCondor-CE from CNAF or from CERN. These in turn are translated to Slurm jobs, locally submitted by the CE as whole node jobs. This is a major difference with respect to the ordinary workflow of the LHC experiments, who typically adopt single core or eight-core payloads. Moreover, the whole node itself is much different than a typical Worker Node adopted in WLCG, as can be seen on table [1]. The color codes are according to the "severity" of the limitation as detailed in [12].

There are two main topics for the LHC experiments to work on:

1. How to make best use of a whole machine having a lower than usual core to memory ratio. In this regard, two strategies are in principle possible:
   a. Running the number of ordinary payloads which best fit the available memory. Which ones depends on how heavily the needed IO impacts on the performances: simulation jobs have the lowest IO footprint and as such are a first promising use case.
   b. Running multi threaded payloads after having selected an optimal number of threads in order to maximize the efficiency without exhausting the available memory

2. Adapting or configuring the Workload Management System to optimize the payload distribution on this kind of computing resources.

A typical standard Marconi A2 node is configured with	A typical WLCG node has
A KNL CPU: 68 or 272(HT) cores, x86_64, rated at ~¼ the HS06 of a typical Xeon	1/2 Xeon-level x86_64 CPUs: typically 32-64 cores, O(10 HS06/thread) with HT on
96 GB RAM, with ~10 to be reserved for the OS: only 0.3 GB/thread if all 272 HT threads are used	2GB/thread. Setups with 3 or 4 are nowadays more common
No external connectivity	Full outbound external connectivity, with remote sw accessed via CVMFS mounts
No local disk (large scratch areas via GPFS/Omnipath)	O(20) GB/thread local scratch space
Access to batch nodes via SLURM; Only Whole nodes can be provisioned, with 24 h lease time	Access via a CE. Single thread and 8 thread slots are the most typical; 48+ hours lease time
Access granted to individuals	Access via pilots and late binding; VOMS AAI for end-user access

Table 1. Differences between Marconi A2 nodes and typical WLCG nodes

Every one of the four LHC experiments manages its own WMS to distribute their own payloads and workflows. Thus, after having enabled Grid submissions to the resources of Marconi A2, the next step is that of selecting the right payloads to run and how to configure or adapt their WMS to use at best these ones.

5 Planning, tests and results from the experiments

5.1 ALICE

ALICE performed its initial tests by performing local manual submissions. This was useful both to test and validate the O2 software framework [13] developed for ALICE Run-3, and to measure performances when executing different multithreaded simulations as payload. The simulations are performed with pythia8 with Pb-Pb collisions at 5.5 TeV per nucleon pair and using GEANT4 [14] as particle propagator in the detector material.

Grid submission tests were also performed, by installing at CNAF a dedicated "vobox", which is the HTCondor based user interface that ALICE employs to submit Grid jobs. This one was configured for testing the Marconi A2 resources. Both input and output files are stored at CERN or at CNAF on ALICE storage via Xrootd; the ALICE O2 software is available via CVMFS and runs within singularity.

Using Pb-Pb simulations allowed us to test all the functionalities of the O2 software and, in particular, the parallel propagation of particles in the same event with multi-thread processes, which was useful to reduce RAM usage. Performances have been determined by collecting metrics monitored during the job execution by a custom script: runtime, output size, RAM usage peak.

The results, graphically reported in Figure [1a,1b], proved we are able to reach a number of simulated events up to 160 with a reasonable time per job (within 12 hours) and using all CPU resources available in the node.

The actual conclusion at the end of this first exercise is the demonstration that ALICE is able to run Run-3 simulation jobs in a HPC system if well defined requirements are satisfied, as provided in our case both from the CINECA and CNAF staff.
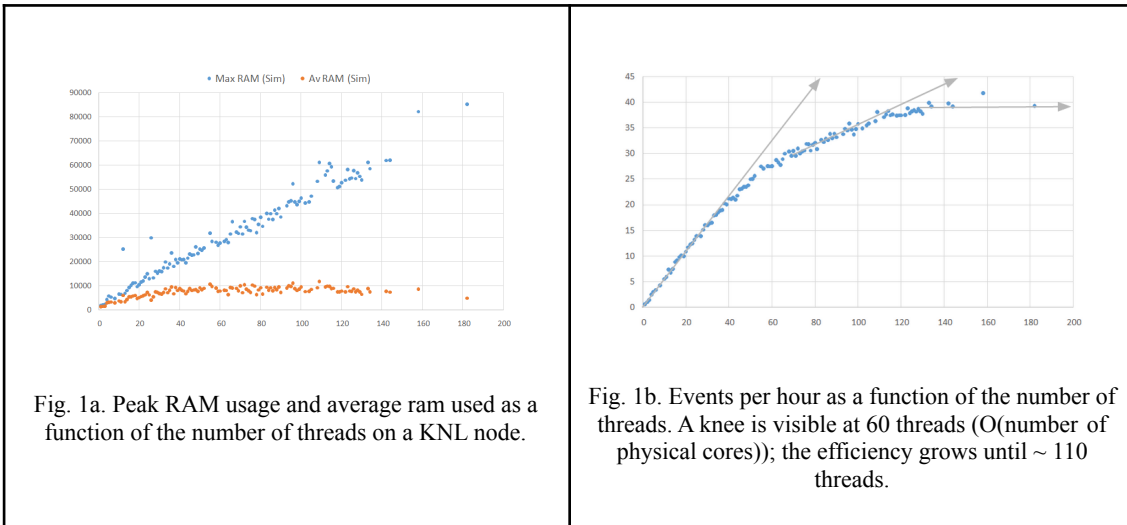
Fig. 1a. Peak RAM usage and average ram used as a function of the number of threads on a KNL node.

Fig. 1b. Events per hour as a function of the number of threads. A knee is visible at 60 threads (O(number of physical cores)); the efficiency grows until ~ 110 threads.

Fig. 1. Left: event/s from ALICE MP simulation on KNL, up to HT4x; Right: throughput and memory utilization (Minimum Bias events) with varying configurations in (*# of Threads per Process, # of Processes per node*).

5.2 ATLAS

The ATLAS Experiment uses the Production and Distributed Analysis (PanDA) system [15] as data-driven WMS, to submit both production and user analysis jobs to the sites. CINECA resources were integrated in Panda as a separate queue of the CNAF site. Multicore jobs with 48 threads, running fast and full Montecarlo simulations, were selected to perform massive submission to Marconi A2, as this size proved to fit well with the available RAM of the nodes.

Fig. 2. Left: Job slots used by ATLAS at Marconi A2, from May 2020 to March 2021; Right: Average CPU efficiency of ATLAS jobs running at CINECA compared with job efficiency at INFN-CNAF

ATLAS was able to run production jobs on CINECA nodes with a peak usage of 12K concurrent slots, and to saturate the available share when a limit was set. The average CPU efficiency was about 85%. For input data retrieval from GRID storage, we tested two job setups: from STORM at CNAF and from EOS at CERN; in both cases gsiftp protocol was used for file transfer. Both configurations showed good performance, with very slight loss of efficiency when reading from CERN, as Montecarlo jobs are not I/O intensive. Success vs failure rate was about 95%, testing period included, which is a very good result, in line with the success rate of the INFN Tier1 site.

5.3 CMS

The CMS Experiment dispatches workloads to its Distributed Computing infrastructure via a WMS based on HTCondor, using a pilot-enabled solution.

As a first step, a GlideinWMS Factory [16] submits "pilots" to every CE we have installed at the O(100) facilities. These are converted into batch jobs according to the matching site-specific technology (SLURM), and are started according to site policies and resources shares as agreed via WLCG.

The pilots, once executed, start a HTCondor STARTD daemon, which connects the node to the CMS Global Pool [17], registering itself as a resource available to CMS. In the case under study, we chose to register Marconi A2 nodes as standard CNAF nodes, in order to experiment with an elastic extension of the italia WLCG Tier-1, instead of creating a brand new site.

CMS HTCondor based matchmaking mechanism allows sites (or even single nodes) to execute specific workloads, via a cherry-picking method. The Marconi resources, when joining the CMS Global Pool, can specify additional requests with respect to CNAF nodes, and in particular we explored the possibilities to:

- ask for low memory jobs, since the standard RAM/thread is lower on KNL than on Xeon nodes at CNAF;
- ask for workload types ("Subtask names") known to need low IO (as simulation and generation);
- veto certain processing types, like analysis jobs on which much less quality assurance is performed before submission;
- ask for resizeable jobs, which are explicitly allowed to be run in a range of thread configurations. This has been chosen in view of further manipulating the payloads, as explained later.

Many combinations of such parameters were tested, starting from very simple ones (only low memory / low IO production jobs), to the most demanding (accept full Monte Carlo production StepChains[18], and end-user analyses). The final configuration, established after the commissioning of the 40 Gbit/s reserved link to CNAF, has demonstrated that CINECA can be used as a standard CMS facility, on which almost all the workflows can be executed. In particular, all analysis jobs were green lighted, and for production jobs only long lasting jobs were discarded; the final cherry picking configuration is indeed

```
"Analysis || ( Production && ExpectedWallTime < 500 minutes)"
```

As the initial grant was extended (see last section), CMS was able to use more than 30 Million CPU hours at CINECA Marconi A2 out of an expected 7.5 Million, with workflow success rates and CPU efficiency close to similar workflows running at standard WLCG sites, as shown in Figs. 3 and 4. The utilization of Xrootd caching proxies showed to be effective, with a limit due to the size of the available disk (20TB), which was shown not adequate when allowing analysis activities to flow over it. Further studies are needed to assess the optimal cache size, and are planned for Marconi100 operations as soon as technically possible.
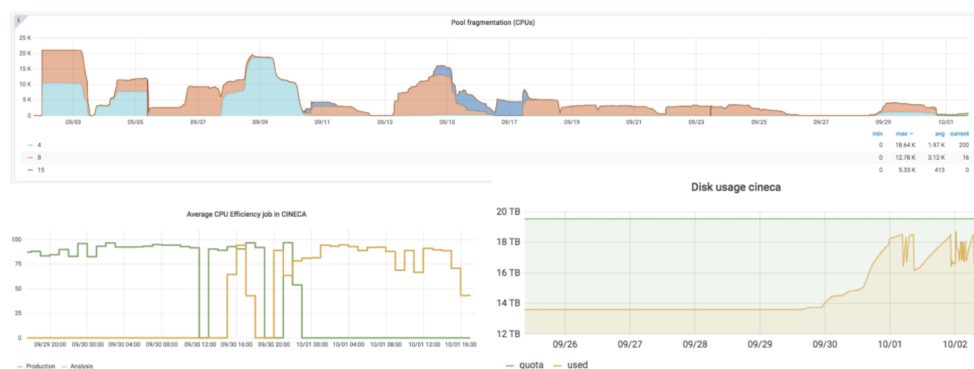
Fig. 3. Performance of CMS processing at CINECA Marconi A2. Top: the utilization of KNL cores, categorized by the multi threading level of the applications (4, 8 and 15 cores per process); top utilization has been in excess of 22k cores. Bottom Left: CPU efficiency of Production and Analysis workflows. Bottom Right: flushing behavior of the 20 TB SSD Xrootd proxy cache at CNAF, used to provide access to all the files in the CMS Xrootd/AAA federation.
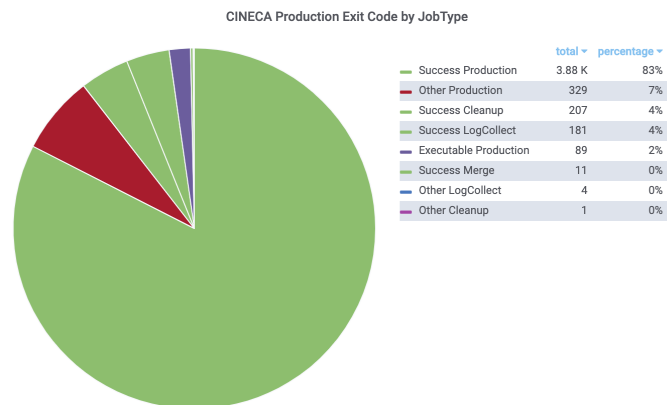
CINECA Production Exit Code by JobType

	total	percentage
Success Production	3.88 K	83%
Other Production	329	7%
Success Cleanup	207	4%
Success LogCollect	181	4%
Executable Production	89	2%
Success Merge	11	0%
Other LogCollect	4	0%
Other Cleanup	1	0%

Fig. 4. Exit code for CMS jobs running at CINECA.

The results obtained include a joint test with the ESCAPE EU Project [19], which has used the CINECA setup in conjunction with its data lake prototype [REF 20]. The press release [REF 21] cites CINECA-A2 as an extension of CNAF-INFN as a first example of a data lake to HPC combined workflow.

5.4 LHCb

In LHCb, Grid jobs are submitted by the DIRAC [22] WMS. Monte Carlo (MC) simulation payloads were used on the CINECA HPC. These are by far the largest consumers of Grid computing power for LHCb, representing nearly 90% of the total. These jobs do not need to retrieve input files and can store the output at CNAF. Until recently, only single core jobs were used. To address this use case, work has been done both at the computing side, to enable multicore submission in DIRAC, and to also on the application software side, to implement the capability to handle concurrent payloads in the same pilot job using a multi-process approach (Figure [6a]). Doing so, it is possible to address the "whole node per job" scenario. Tests of several single-process and multi-process configurations have been performed (Figure [5a,5b]), coming to the conclusion that up to 80 single core jobs can concurrently run, or up to 17 eight core jobs within the limit of the available RAM. The observed performance gain from using multi-core jobs is however nearly negligible. Further details of this work are described in Ref. [23].
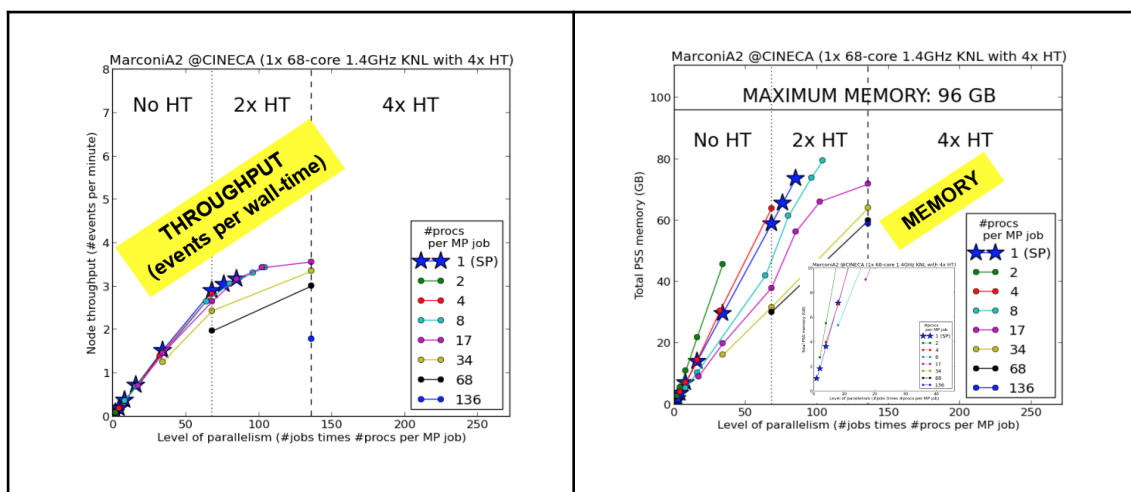

Fig. 5. Throughput in event/s (left) and memory utilization (right) for LHCb MP simulation on KNL, up to 2xHT, for various configurations in (*Number of Processes per Job, Number of Jobs per node*).

After commissioning Multi core job submission in DIRAC, successful production jobs were run in Marconi A2 nodes by Q3 2020. After that, a large-scale exploitation was performed, running (mostly) about 70 single core payloads per node, during Q4 2020 and Q1 2021. The activity is reported by Figure [6b]. During February 2021, CINECA was the largest site outside CERN for MC simulation.

Fig. 6. Left: schematic representation of multi-core job scheduling in DIRAC. Right: number of concurrent LHCB jobs on the CINECA HPC between Q2 2020 and Q1 2021

6 Overall Results

The PRACE project, originally requested for 72 Million core hours, was granted for 30 Million core hours only, from April 2019 to April 2020. After the April deadline, the possibility to use KNL resources was extended by CINECA, and by the end of February 2021 93 Million core hours were used, with all the participating experiments able to reach production level in specific or global activities.

CMS and ATLAS and LHCb were able to (largely) exceed the 7.5 Million core hours as per the grant, while ALICE stopped the activities after a successful proof-of-concept with their workflows. Figure [7] shows the temporal and total utilization of Marconi A2.

Fig. 7. Left: total utilization of the Marconi A2 from April 2019 to February 2021. The brown area shows the amount of remaining grant (30 Million core hours). (right) utilization by experiments as a fraction of the utilized 93 Million core hours.

7 Future directions and conclusions

After the conclusion of the Grant, the Marconi A2 system has been switched off, in order to make space for the newer Marconi100 and Galileo100 CINECA systems [24].

The two new systems, especially Marconi100, are even less standard with respect to Marconi A2. Marconi100 deploys two IBM Power9 CPUs per node, with 4 last generation Nvidia V100 GPGPUs, which provide most of the computing power. The full system consists of 980 nodes, for a total of around 32 PFlops.

INFN and its members in the LHC experiments are planning to experiment on the Marconi100 system; on top of the same solutions deployed for Marconi A2, additional efforts have started in the development, deployment and test aiming to:

- enable PowerPC systems within experiments' Workload Management Systems;
- enable the capability to offer GPUs as a center capability, and to match workflows which can use them;
- validate the system for generic experiment workflows (with a low utilization of the GPU part) and for specific, GPU-heavy, ad-hoc workflows as those from Machine Learning training and inference.

The collaboration with CINECA has already been very fruitful for the LHC experiments, since it has driven a number of R&D needed in order to optimally utilize such a non standard machine; all the experiments reported at recent CERN level reviews the successful utilization of CINECA systems. At the same time, we think continuing with the tests is essential for the experiments in order to pave the way for a greater utilization of future HPC systems, and for CNAF to be prepared for the transition to the future pre-exascale Leonardo system at CINECA.

The present work is a continuation of a series of integration efforts, as reported in [27], [28], [29], [30].

References

1. CINECA Marconi, http://www.hpc.cineca.it/hardware/marconi
2. WLCG, https://wlcg.web.cern.ch/
3. INFN, https://www.infn.it/en
4. T. Boccali, "*Computing models in high energy physics*", Reviews in Physics Volume 4, November 2019, 100034, https://doi.org/10.1016/j.revip.2019.100034
5. top500, https://www.top500.org/
6. Omni-Path, https://newsroom.intel.com/news-releases/intel-architects-high-performance-computing-system-designs-to-bring-power-of-supercomputing-mainstream/#gs.wjj1hy
7. SLURM, https://slurm.schedmd.com/documentation.html
8. PRACE, http://www.prace-ri.eu/
9. HTCondor, https://research.cs.wisc.edu/htcondor/
10. Singularity, https://singularity.lbl.gov/
11. eXtreme DataCloud, http://www.extreme-datacloud.eu/
12. LHC Collaborations, "*Summary of the cross-experiment HPC workshop*", https://indico.cern.ch/event/811997/attachments/1862943/3062278/HPCLHCC.docx.pdf
13. ALICE Collaboration, "Technical Design Report for the Upgrade of the Online-Offline Computing System", CERN-LHCC-2015-006 ALICE-TDR-019, https://cds.cern.ch/record/2011297/
14. S. Agostinelli et al., "GEANT4 - a simulation toolkit", Nucl. Instr. Meth. A, vol. 506, no. 3, pp. 250-303, 2003.

15. T. Maeno et al., Overview of ATLAS PanDA Workload Management J. Phys. Conf. Ser. 331
16. The CMS Collaboration, "*HPC resources integration at CMS*", https://cds.cern.ch/record/2707936/files/NOTE2020_002.pdf
17. J. Balcas et al., "Stability and scalability of the CMS Global Pool: Pushing HTCondor and glideinWMS to new limits", Journal of Physics: Conference Series, Volume 898, Track 3: Distributed Computing
18. Experience in using commercial clouds in CMS, IOP Conf. Series: Journal of Physics: Conf. Series 898 (2017) 052019 doi:10.1088/1742-6596/898/5/052019
19. ESCAPE EU Project, https://projectescape.eu/
20. Bolton, R., Campana, S., Ceccanti, A., Espinal, X., Fkiaras, A., Fuhrmann, P., & Grange, Y. (2020). ESCAPE prototypes a data infrastructure for open science. In *EPJ Web of Conferences* (Vol. 245, p. 04019). EDP Sciences.
21. https://projectescape.eu/news/escape-dios-supporting-cms-experiment-obtaining-maximum-value-its-data-through-high-performance
22. F. Stagni et al., DiracGrid/dirac: v6r20p15 (2018), DOI: 10.5281/zenodo.1451647
23. F. Stagni, A. Valassi, V. Romanovski, "*Integrating LHCb workflows on HPC resources: status and strategies*", CHEP2019, Adelaide, DOI:10.1051/epjconf/202024509002
24. https://www.hpc.cineca.it/hardware/marconi100
25. A. Anysenkov et al., AGIS: Integration of new technologies used in ATLAS Distributed Computing, J. Phys. Conf. Ser. 898
26. GlideinWMS, https://glideinwms.fnal.gov/doc.prd/index.html
27. S. Dal Pra et al., "*Elastic CNAF DataCenter extension via opportunistic resources*", DOI: 10.22323/1.270.0031
28. Boccali, T., Dal Pra, S., Spiga, D., Ciangottini, D., Zani, S., Bozzi, C., ... & Bonacorsi, D. (2020). Extension of the INFN Tier-1 on a HPC system. In *EPJ Web of Conferences* (Vol. 245, p. 09009). EDP Sciences.
29. L. Dell'Agnello et al., "*INFN Tier–1: a distributed site*", The European Physical Journal Conferences 214:08002 · January 2019 DOI: 10.1051/epjconf/201921408002
30. D. Ciangottini et al., "*Distributed and On-demand Cache for CMS Experiment at LHC*", DOI: 10.1109/eScience.2018.00082