Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Heterogeneous computing at INFN-T1

D.Lattanzio*, A.Chierici, D.Michelotto, A.Pascolini, G.Sergi

Conference on Computing in High Energy and Nuclear Physics. October 19-25, 2024 - Kraków

Outline

INFN CNAF

- Introduction
- ARM resources
- RISC-V resources
- ~~Sierra Forest~~

Introduction

INFN CNAF

- At INFN we started a technology tracking program
- At INFN-T1, in farming division, we mainly focused on computing
  - Investigate new processors technologies
  - Power consumption
  - CPU architectures
  - Understand middleware and general software readiness

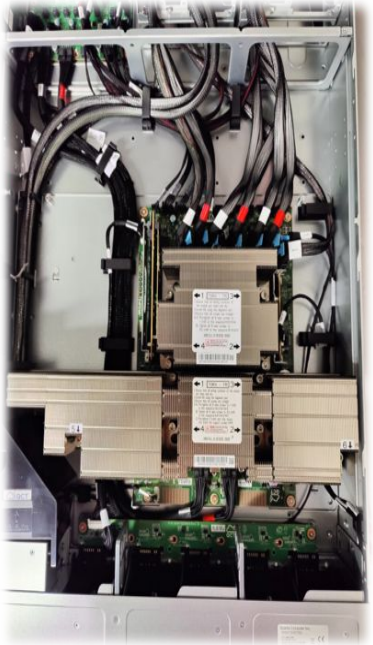
arm RISC-V intel

ARM Resources

ARM resources

- 4x 2U Dual socket ARM – Ampere Altra Max
  - 2x Ampere AltraMax M128-30 128Core 2.8Ghz 250W
  - 1TB RAM
  - 2x NVMe U.3 3.84TB discs
  - 1x Dual port SFP28 ethernet
- 1x GraceHopper superchip (Quanta server)
  - CPU Memory: 480GB
  - GPU Memory: 96GB
  - 2x NVMe disks (1.92TB + 960GB)
  - 1x 100GBit ethernet port
- 1x Grace superchip (Supermicro server)
  - CPU Memory: 480GB
  - 2x NVMe disks (7.68TB + 480GB)
  - 1x SFP28 ethernet port

Some photos



Main differences - Hardware side

- Procurement not so easy
  - Vendors bidded with incredible variation in price
  - Hardware support: no on-site, only on-center
    - This is slowly changing
- Form factor
  - Only 2U, 1 motherboard
  - No possibility for customization (grace/hopper net adapter)
- BMC and firmwares in general don't seem very stable
  - BMC is slow
  - Server takes ages to boot
  - Network installation not always possible

Main differences - Software side

INFN CNAF

- Started with alma8, since puppet build was not available on alma9
- GPFS initially not available for aarch64
  - Now it's available but only with a very recent version, not yet in production at INFN-T1
- Grid middleware still an issue (full availability only with el7)
- Had to make agreements with experiments to run only specific workflows

Some numbers

INFN CNAF

- Ampere altra max
  - Each node provides 3754 hepscore, 14,66 hepscore/core
  - BMC not reliable, couldn't get power consumption
- GRACE
  - The node provides 4459 hepscore, 30,97 hepscore/core
  - The node consumes 1kW average, 4,459 hepscore/watt

As a comparison

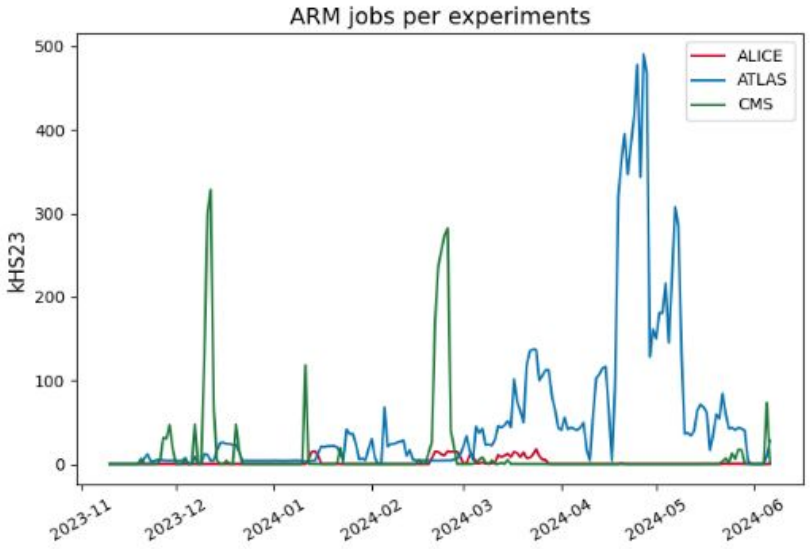
- Leonardo GP
  - The node provides 2880 hepscore, 25,71 hepscore/core (ht off)
  - The node consumes 799W (as per tender doc), 3,6 hepscore/watt

ARM experiment usage at CNAF

INFN CNAF

Current setting (still work in progress)

- Cvmfs available
- Access to external network
- Gpfs client -> now available on ARM but only with a very recent version, not yet in production at INFN-T1
- Condor/GRID -> in production

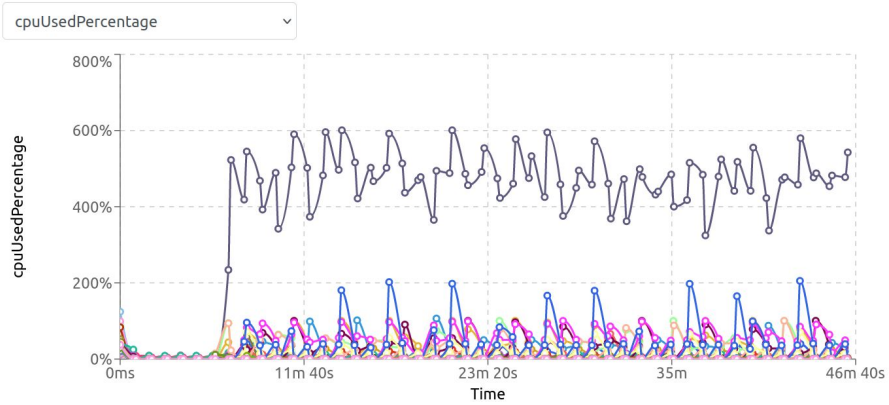


Experiment point of view: ALICE

INFN CNAF

Preliminary tests showed good balancing in using resources: still some instability in ARM builds didn't allow to run a validation so far.

- Data reconstruction



- GRID setting tuned for 8 cores per job
- CPU efficiency consistent with what observed in the GRID node
- Physics validation on the output not yet done

Recent builds on Almalinux9 presented some issues (mainly fixed) → tests going to be resumed

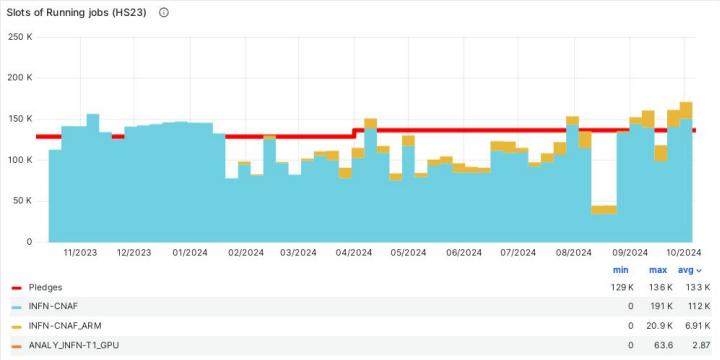
GRID submission @CNAF (GRID submission + aptainer container) → validated

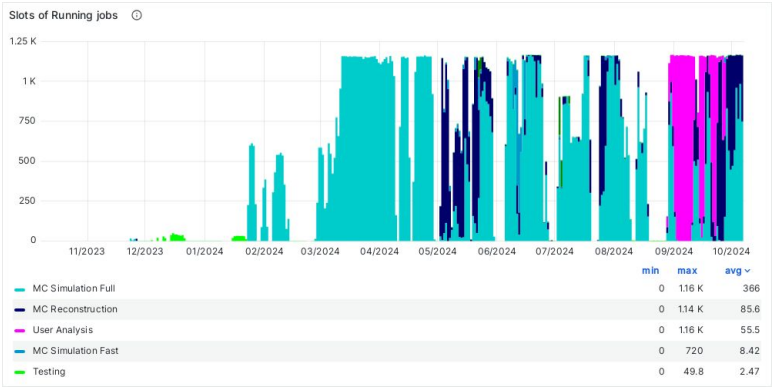
Experiment point of view: ATLAS

INFN CNAF

ATLAS has been running jobs on ARM@CNAF for one year

- initially, only test jobs
  - ATLAS Software already Physics-validated on ARM
  - Technical validation performed at CNAF (HTCondorCE, pilot/PanDA, containers)
- workflows:
  - Full and Fast MC Simulation
  - MC reconstruction
  - Group production
  - User Analysis





Very good performance observed

- Steady use of available resources. On peak: ~12% of ATLAS-dedicated resources at INFN-T1

Experiment point of view: CMS

The ARM nodes at T1_IT_CNAF have been integrated as a sub site of the regular Tier1 and thus accessed via GRID

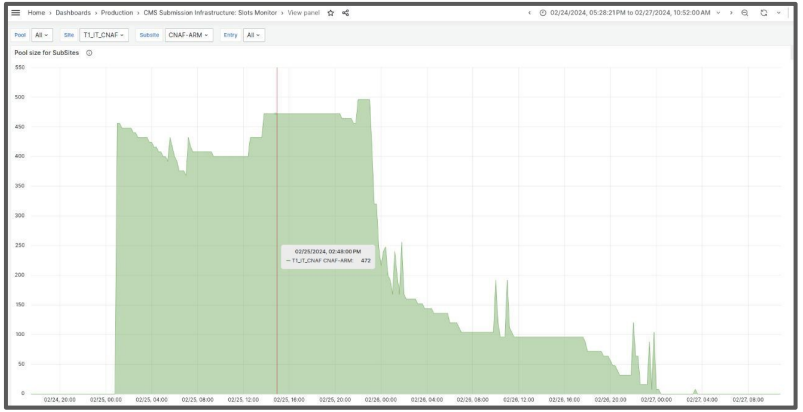
- Minimize the effort on CMS and simplify the site admins life
- At the moment data are accessed via Xrootd protocol. The plan is to provide direct access also via GPFS.

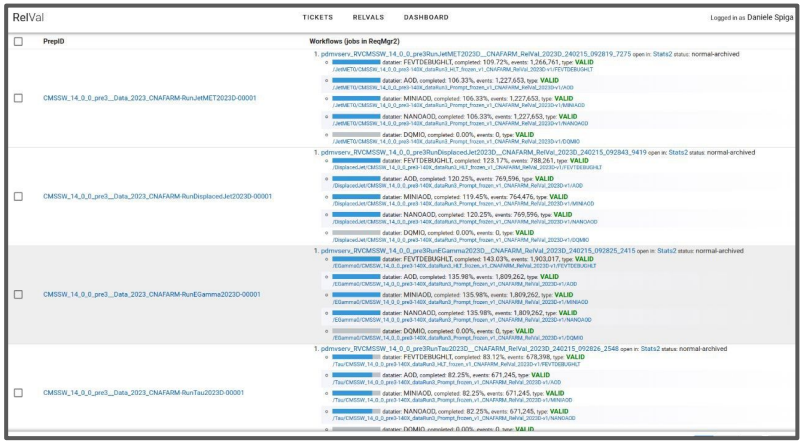
Technical validation fully done.

- Being the first allocation at a Tier site, the CNAF nodes where essential to finally validate that CMS Computing is multi-architectures enabled.

Physics validation in progress:

- most of the subsystems report green light (especially when looking at MC).
- Some discrepancies spotted on DATA require further analysis for a better understanding
- This step has been carried on both at CNAF and at Glasgow temporary allocation





RISC-V Resources

RISC-V Resources

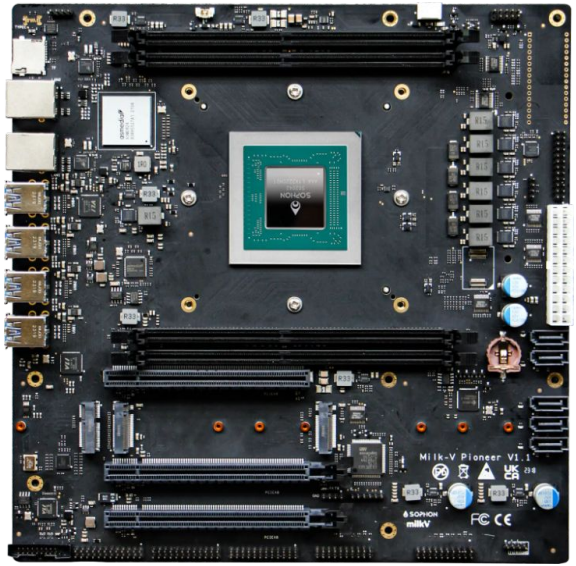
INFN CNAF

- 2x Milk-V Pioneer
  - 64 Core RISC-V CPU up to 2GHz
  - 128GB DDR4 3200
  - 1TB NVMe disk
  - 2x10Gbps network cards
  - OS pre-installed: Fedora 38
- An interesting architecture for the future
  - Not competitive today (cores at the level of Rpi cores), but evolving fast!
  - Open ISA managed by the RISC-V Foundation (riscv.org)

Sophgo SG2042

INFN CNAF

- 64 RISC-V Cores
  - T-Head XuanTier C920s (high performance)
- 2Ghz frequency
- 64MB system cache
- 32x PCIe Gen4 lanes
- good enough to test larger projects like porting CMS full software stack (CMSSW)



Experiment point of view: CMS

INFN CNAF

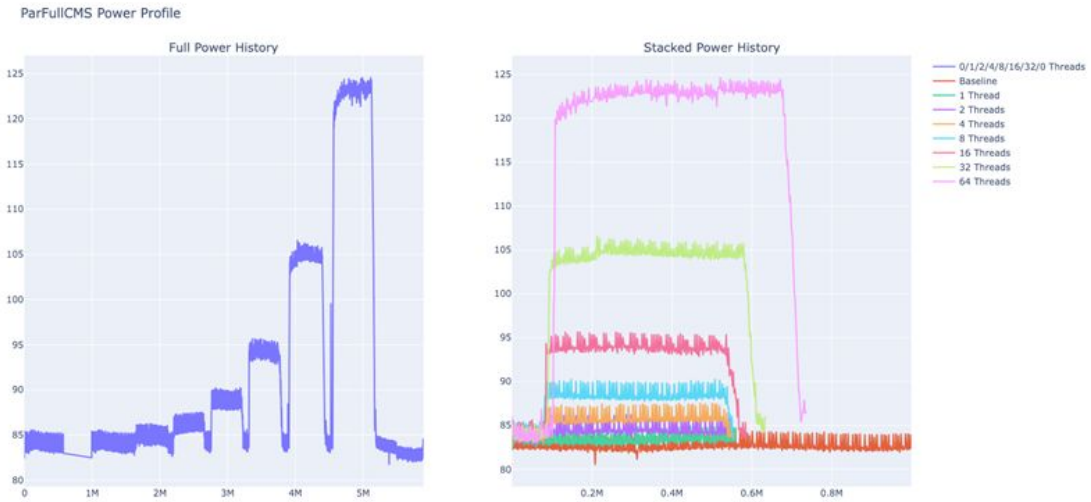
- INFN is deploying compiled codes for tests in: /cvmfs/datacloud.infn.it/repo/riscv64-pioneer/
- Please see the CMS CHEP talk on RISC benchmarking
- The MILK-V performance is comparable to a good desktop pc, but the number of cores is very interesting and can increase more.

Experiment point of view: CMS

INFN CNAF

- db12 CMS benchmark
  - milk-v: 378.3, 5.8 per core
  - e5-2640v3: 248, 3.8 per core
- Milk-v today performs better than a 2016 CPU…
- If we take out the hyperthreading, milk-v is 2-3 times slower than a modern xeon
- 64 real cores, not so common on modern CPUs
- Trade-off between core number and power (take into account power consumption too)

Experiment point of view: CMS



1/2/4/8/16/32/64 threads test of CMS Fast Simulation → power consumption

Thanks for the attention


ICSC
Centro Nazionale di Ricerca in HPC
Big Data and Quantum Computing
Supercomputing
shaping the future