GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

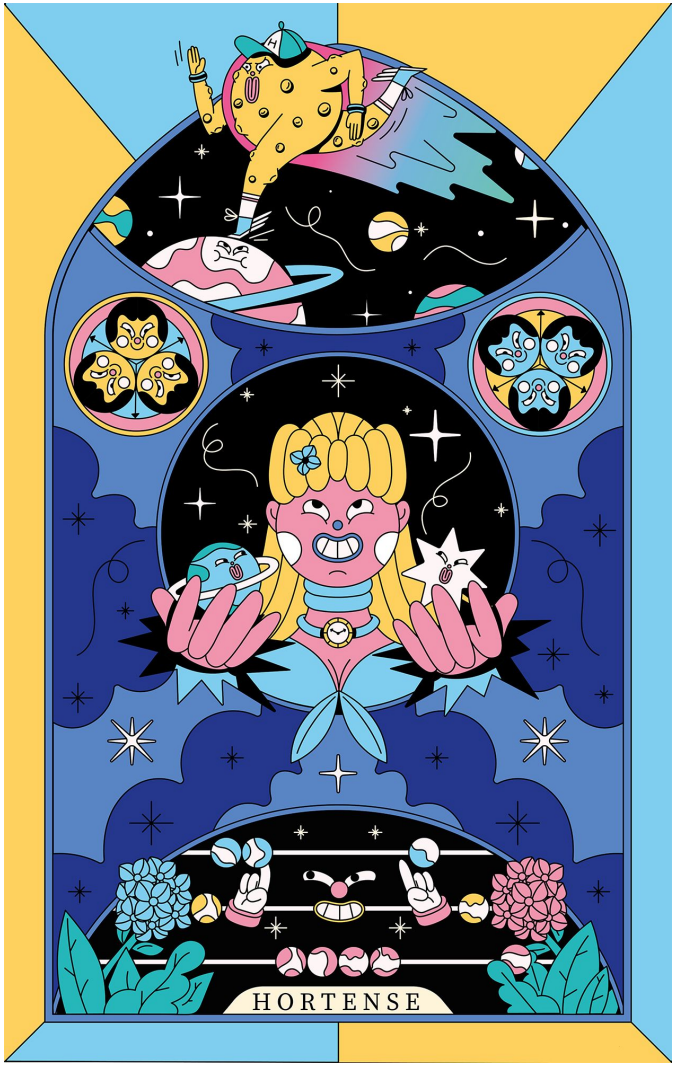
Vlaanderen is supercomputing

VSC Tier-1 Hortense - phase 2 kick-off meeting

compute@vscentrum.be

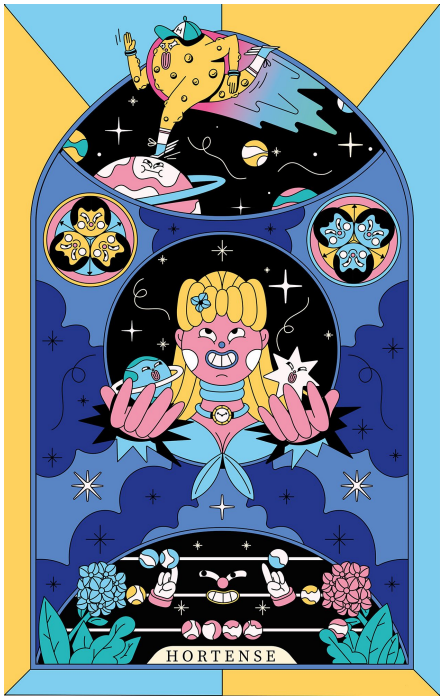
https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html

26 May 2023



Agenda

- Timeline
- **Hortense Tier-1 phase 2**
  - Additional hardware + new partitions
  - Pilot phase for Milan partition
- Impact for future Tier-1 projects and starting grants
- Dedicated interactive + debug partition
- Tier-1 support
- Upcoming maintenance windows



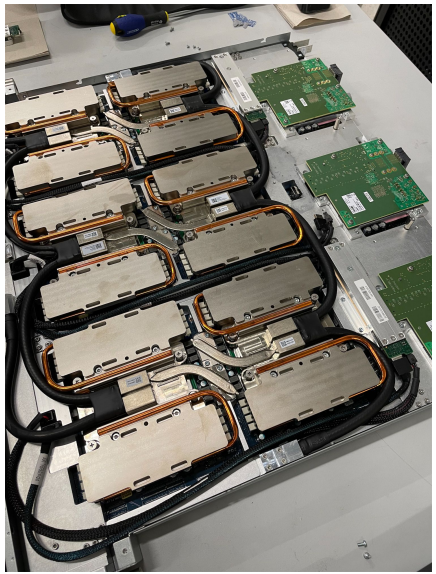
Timeline

- [23 Nov 2021] Hortense phase 1 is ready for testing
- [11 Mar 2022] Hortense phase 1 is ready for production
- [24-28 Oct 2022] Maintenance window (OS updates, preparations for installing phase 2)
- [6-15 March 2023] Maintenance window (preparations for installing phase 2)
- **[2-17 May 2023] Maintenance windows (OS updates + finishing installation of phase 2)**
- **[17 May 2023] Hortense phase 2 is ready for testing**
- *[7 July 2023] (planned) Hortense phase 2 is ready for production*

Hortense phase 2 in pictures

View of phase 2 of Hortense
(Milan nodes + 80GB GPU nodes)

Hortense workernode,
showing on-board
water cooling (no fans!)



Front view of Hortense
Right block: phase 1 (Rome)
Left block: phase 2 (Milan)

Hortense phase 2 in pictures
(cooling infrastructure & piping)

GHENT UNIVERSITY
VLAAMS SUPERCOMPUTER CENTRUM

Pipes for cooling water underneath raised floor (“inside” Hortense)

Pumps in HPC room dedicated for Hortense cooling circuit

Dedicated chillers for Hortense, on roof of UGent datacenter S10, a.k.a. “loopC”

Hortense: hardware

https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html#hardware-details

GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

- Phase 1:
  - 294 AMD Rome CPU nodes with 128 cores + 256GiB RAM + 480 GB local SSD
  - 42 AMD Rome CPU nodes with 128 cores + 512GiB RAM + 480 GB local SSD
  - 20 GPU nodes, each with 48 Rome CPU cores + 256GiB RAM + 4x NVIDIA A100 with 40GB GPU memory
  - 2.7PB shared scratch filesystem + HDR-100 Infiniband interconnect (dual HDR for GPU nodes)
- Phase 2:
  - **+48 AMD Rome CPU nodes** with 256GiB RAM → now 342 nodes in `cpu_rome` partition
  - **+384 AMD Milan CPU nodes**, each with 128 cores + 256GiB RAM + 480GB local SSD + HDR-100 Infiniband
  - **+20 GPU nodes**, each with 48 Rome CPU cores + **512GiB RAM** + 4x A100 **80GB** GPU memory + dual HDR IB
  - Extension of Hortense scratch filesystem: +2.7PB → now **5.4 PB** + doubled in speed (~35GB/s to max. 70GB/s)
  - **Dedicated interactive + debug partition**: 3 oversubscribed nodes (144 virtual cores) with strict per-user limits

Not changed: login nodes - **still AMD Rome CPUs!**

- **Over 100,000 CPU cores now in Hortense!**
  - 384x128 = 49,152 cores in both Rome and Milan partitions + 40x48 = 1,920 CPU cores in GPU nodes
  - ⟹ 100,224 CPU cores in total

Hortense: partitions

https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html#hardware-details

GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

(+) new partition (*) changed partition (#) extended partition

To see nodes per partition: `pbsmon -P`

cpu_rome (#)

342 CPU nodes (HDR100 IB)

each: 128 AMD Rome cores, 256GiB RAM

cpu_rome_512

42 CPU nodes (HDR100 IB)

each: 128 AMD Rome cores, 512GiB RAM

cpu_rome_all

384 AMD Rome CPU nodes

cpu_rome + cpu_rome_512

gpu_rome_a100_40 (+)

20 GPU nodes (dual HDR IB)

each: 48 AMD Rome cores, 256GiB RAM,

4x NVIDIA A100 40GB

gpu_rome_a100_80 (+)

20 GPU nodes (dual HDR IB)

each: 48 AMD Rome cores, 512GiB RAM,

4x NVIDIA A100 80GB

gpu_rome_a100 (*)

40 GPU nodes

gpu_rome_a100_40 +

gpu_rome_a100_80

debug_rome (+)

3 AMD Rome CPU nodes (4x oversubscribed)

each: 12 cores, 1x NVIDIA V100 GPU (16GB),

256GiB RAM, 1x shared NVIDIA P1000 GPU (4GB)

cpu_milan (+)

384 CPU nodes (HDR100 IB)

each: 128 AMD Milan cores, 256GiB RAM

Hortense: interactive + debug partition

https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html#interactive-and-debug-partition

GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

- New: debug + interactive partition (AMD Rome CPUs)

```
module swap cluster/dodrio/debug_rome
qsub -A your_project -l nodes=1:ppn=8 example_job.sh
```

- 3 nodes, each with 12 physical AMD Rome cores (48 virtual),
  1 NVIDIA V100 GPU (16GB) on request + 1 shared NVIDIA P1000 GPU (4GB RAM) always available
- **Can be used by all Tier-1 projects, free of charge (no credits consumed)**
- **4x oversubscription**: performance will degrade under high load!
- Be careful with using available shared GPU: data in GPU memory *may* be read by others!
- **Strict per-user limits for jobs**: max. 8 cores in use, 5 max jobs in queue, max. 3 jobs running
- Recommended for debugging job scripts, interactive desktop session in web portal, etc.

Hortense: Milan partition (pilot phase)

https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html#hardware-details

GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

- Milan partition is currently available for testing (pilot phase)

```
module swap cluster/dodrio/cpu_milan
qsub -A your_project example_job.sh
```

- **Current Tier-1 projects can now use Milan partition for free** - no credits consumed until 7 July 2023
- Differences with Rome partition:
  - Same core count, same amount of RAM (256GiB) as `cpu_rome` partition, same local disk
  - No fat memory nodes (512GiB RAM) or GPU nodes in Milan partition
  - Lower CPU clock speed, better NUMA → HPL: 3-4% *slower*, OpenFOAM: 10-15% faster
- **Login + debug nodes have AMD Rome CPUs**, beware when compiling software to run on Milan partition!

Hortense: software

https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html#software

GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

- OS: Red Hat Enterprise Linux 8.6 (updated in May 2023)
- Infiniband drivers: Mellanox OFED v5.8-2.0.3.1 (updated in May 2023)
  - UCX + OpenMPI modules in central software stack were reinstalled because of update
- Central stack of scientific software via `module` command
  - **Same modules available on both Rome (zen2) and Milan (zen3) CPU partitions**
  - **Same modules available on 40GB and 80GB GPU partitions**
- Slurm v22.05.09 (with Torque frontend wrapper commands - still recommended!)
- Apptainer v1.8 (compatible with Singularity)
- **Be careful with self-installed software when using Milan partition!**

  Compile source code workernode in Milan partition, not on login node or debug node

Future Tier-1 projects

https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html#getting-access

GHENT UNIVERSITY

VLAAMS SUPERCOMPUTER CENTRUM

- Production for Milan partition is planned for 7 July'23

  → end of free usage, credits will be consumed!

- **Future Tier-1 projects will be assigned exclusively to either Rome or Milan partition**
  - Based on resource requirements + software used in project
  - Compute load will be balanced across the Rome/Milan partitions
  - Starting with granted projects from June 2023 cutoff
- Starting grants will get access to both Rome + Milan partitions
- Next cutoff dates for Tier-1 project proposals: 5 June + 2 October 2023

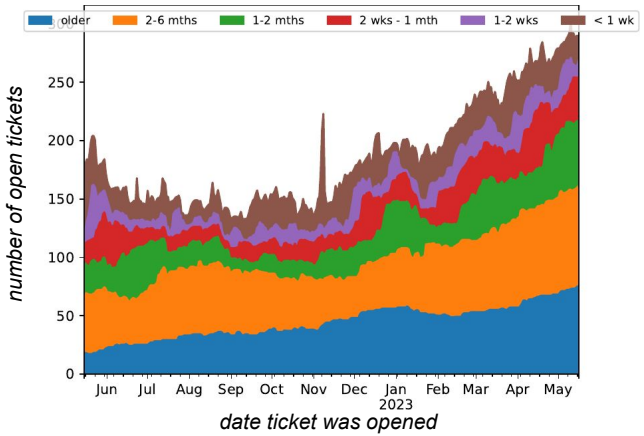
  https://www.vscentrum.be/compute

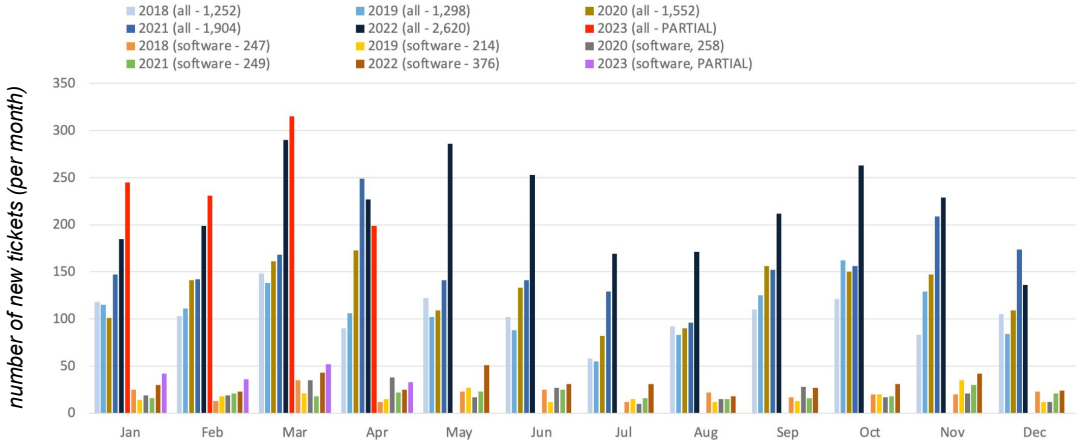
Hortense: getting help

- **For all feedback and questions: contact compute@vscentrum.be**
- Please report problems or unexpected behaviour with:
  - Overall system stability
  - Central scientific software stack
  - Scratch filesystem
  - Unexpected errors in jobs
  - Performance issues
  - Torque frontend job wrappers (qsub, qstat, …)
  - Use of mympirun
- System changes + maintenance will be communicated via:
  - Tier-1 Hortense mailing list: t1-users@lists.ugent.be
  - VSC status page: http://status.vscentrum.be

Hortense: getting help (be patient)

- **HPC-UGent team is struggling to keep up with increasing support demand** (Tier-1 + Tier-2)





- **Be patient** when submitting support questions - we're doing what we can to keep up
- **Help us help you**: read the docs, provide sufficient details (like job IDs, output files, etc.), …
- Feel free to send a reminder in the same mail thread, especially if your work is blocked
- VUB-HPC team is helping with Tier-1 support requests from industry users

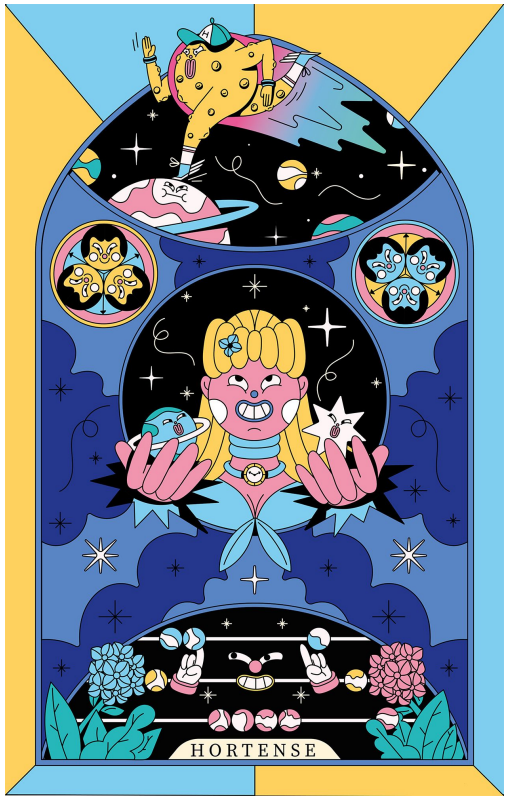
Upcoming maintenance windows

https://status.vscentrum.be/tier1_compute.html

- Tue 11 July 2023 (Flemish holiday): ***test*** of UGent datacenter disaster recovery procedure (DRP)
  - **Temporarily no network access to all infrastructure hosted in UGent datacenter S10**
  - Will be unavailable: Tier-1 login nodes + web portal (+ same for UGent Tier-2), VSC accountpage
  - Should not affect running jobs, unless they require access to outside world (which will not work)
  - Internal network + access to scratch filesystem from jobs will be unaffected
- To be planned: taking *loopC* of S10 cooling infrastructure into production for Tier-1 Hortense

- Will require downtime of Tier-1 Hortense (nodes, storage, access)
- We depend on ATOS+APAC (system vendor + cooling subcontractor)
- We will share more details, including specific dates, when available

Hortense: documentation and support



Documentation: https://docs.vscentrum.be/en/latest/gent/tier1_hortense.html

Status page: https://status.vscentrum.be/tier1_compute.html

For questions or problems: contact VSC support team via email

- compute@vscentrum.be
- **Please mention `[Hortense]` in email subject!**

Mailing list: t1-users@lists.ugent.be (moderated even for list members)

Software installation requests:

- *Please use the HPC-UGent request form!*
- https://www.ugent.be/hpc/en/support/software-installation-request
- **Select Tier-1 Hortense as target system**