university of groningen / center for information technology

High Performance Computing

Site talk - University of Groningen

10th EasyBuild User Meeting
25-27 March 2025

Bob Dröge

About me

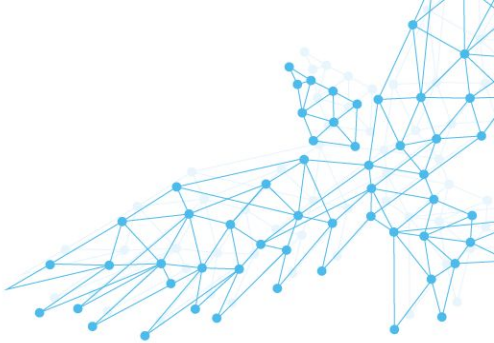
- Computer Science & Mathematics

- Team High Performance Computing (since 2011)
- Center for Information Technology
- University of Groningen, The Netherlands

- HPC user support and training, installing software, HPC system administration

- EasyBuild Maintainer (since June 2021)
- Euclid: WP lead for SDC-NL infrastructure
- EESSI / MultiXscale

university of groningen / center for information technology



Computing at the University of Groningen

university of groningen / center for information technology

Space Exploration

Euclid is designed to explore the evolution of the dark Universe. It will make a 3D-map of the Universe (with time as the third dimension) by observing billions of galaxies out to 10 billion light-years, across more than a third of the sky.

ESA Euclid Telescope

Space Exploration

Euclid is designed to explore the evolution of the dark Universe. It will make a 3D-map of the Universe (with time as the third dimension) by observing billions of galaxies out to 10 billion light-years, across more than a third of the sky.

ESA Euclid Telescope

LOFAR

LOFAR (LOw Frequency ARray) is currently the largest radio telescope operating at the lowest frequencies that can be observed from Earth.

ASTRON

GCC

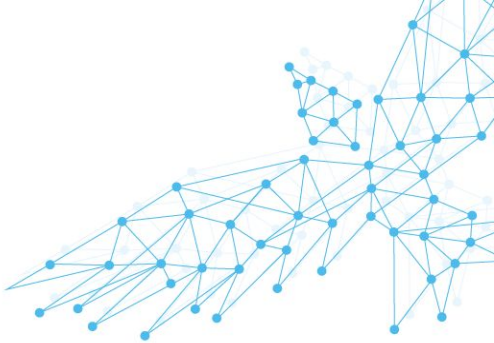
Genomics Coordination Centre, UMCG/UG

>2300 Individual Researchers

100+ Publications/year

2024

1. Adupa, V., Ustyantseva, E., Kampinga, H. H., & Onck, P. R. (2024). Tertiary structure and conformational dynamics of the anti-amyloidogenic chaperone DNAJB6b at atomist [DOI | http]
2. Aguayo, J., Bertoglio, C., & Osses, A. (2024). Distributed parameter identification for the Navier–Stokes equations for obstacle detection. *Inverse Problems*, *40*(1), 015012 [
3. Arrowsmith-Kron, G., Athanasakis-Kaklamanakis, M., Au, M., et al. (2024). Opportunities for fundamental physics research with radioactive molecules. *Reports on Progress* [DOI | http]
4. Castro-Alvarez, S., Sinharay, S., Bringmann, L. F., Meijer, R. R., & Tendeiro, J. N. (2024). Assessment of fit of the time-varying dynamic partial credit model using the posteri *of Mathematical and Statistical Psychology*, page bmsp.12339 [DOI | http]
5. Cotteret, M., Greatorex, H., Ziegler, M., & Chicca, E. (2024). Vector Symbolic Finite State Machines in Attractor Neural Networks. *Neural Computation*, *36*(4), 549–595 [D
6. Edman, L., Sarti, G., Toral, A., Noord, G. V., & Bisazza, A. (2024). Are Character-level Translations Worth the Wait? Comparing ByT5 and mT5 for Machine Translation. *Trans Linguistics*, *12*, 392–410 [DOI | http]
7. Erić, V., Li, X., Dsouza, L., et al. (2024). Observation of Dark States in Two-Dimensional Electronic Spectra of Chlorosomes. *The Journal of Physical Chemistry B*, *128*(15), 35
8. Es Sayed, J., Mukherjee, A., El Aani, S., et al. (2024). Structure–Property Relationships of Granular Hybrid Hydrogels Formed through Polyelectrolyte Complexation. *Macrom*
9. Ferreira, P., Neves, R. P. P., Miranda, F. P., et al. (2024). DszA Catalyzes C–S Bond Cleavage through N ^{5} –Hydroperoxyl Formation. *Journal of Chemical Information and Mo*
10. Gürel, U., Keten, S., & Giuntoli, A. (2024). Bidispersity Improves the Toughness and Impact Resistance of Star-Polymer Thin Films. *ACS Macro Letters*, *13*(3), 302–307 [D
11. Jafari, H., Barts, E., Przybysz, P., et al. (2024). Robust Zeeman-type band splitting in sliding ferroelectrics. *Physical Review Materials*, *8*(2), 024005 [DOI | http]
12. Kasaei, H., Kasaei, M., Tziafas, G., Luo, S., & Sasso, R. (2024). Simultaneous Multi-View Object Recognition and Grasping in Open-Ended Domains. *Journal of Intelligent &*
13. Kleibergen, F. & Kong, L. (2024). Identification robust inference for the risk premium in term structure models. *Journal of Econometrics*, page 105728 [DOI | http]
14. Koopmans, L., Dhali, M. A., & Schomaker, L. (2024). Performance Analysis of Handwritten Text Augmentation on Style-Based Dating of Historical Documents. *SN Computer*
15. König, K., Berengut, J. C., Borschevsky, A., et al. (2024). Nuclear Charge Radii of Silicon Isotopes. *Physical Review Letters*, *132*(16), 162502 [DOI | http]
16. Lan, L., Van Doorn, J. M. H., Wouda, N. A., Rijal, A., & Bhulai, S. (2024). An Iterative Sample Scenario Approach for the Dynamic Dispatch Waves Problem. *Transportation S*
17. Markovitch, O., Wu, J., & Otto, S. (2024). Binding of Precursors to Replicator Assemblies Can Improve Replication Fidelity and Mediate Error Correction. *Angewandte Chem* http]
18. Marx, A. C., Jafari, H., Tekelenburg, E. K., et al. (2024). Nonlinear magnetotransport in MoTe 2. *Physical Review B*, *109*(12), 125408 [DOI | http]
19. Maurits De Roo, C., Sardjan, A. S., Postmus, R., et al. (2024). Reaction of (N4Py)Fe with H $_2$ O $_2$ and the relevance of its Fe(IV)=O species during and after H $_2$ O $_2$ dispropo [DOI | http]
20. Ndung'u, S., Grobler, T., Wijnholds, S. J., Karastoyanova, D., & Azzopardi, G. (2024). Classification of radio galaxies with trainable COSFIRE filters. *Monthly Notices of the R* http]
21. Nelemans, S. (2024). *Asset Pricing with Economic Inequality: a Macrofinancial Approach*. Ph.D. thesis, University of Groningen [DOI | http]
22. Nguyen, H. L., Do, T. N., Zhong, K., et al. (2024). Inter-subunit energy transfer processes in a minimal plant photosystem II supercomplex. *Science Advances*, *10*(8), eadh09
23. Oldenburg, V., Cardenas-Cartagena, J., & Valdenegro-Toro, M. (2024). Forecasting Smog Clouds With Deep Learning: A Proof-Of-Concept [http]
24. P. Neme, N., Jansen, T. L. C., & Havenith, R. W. A. (2024). Cyclopentene ring effects in cyanine dyes: a handle to fine-tune photophysical properties. *Physical Chemistry Che*
25. Sabo, E., Riveni, M., & Karastoyanova, D. (2024). Decentralized Networks Growth Analysis: Instance Dynamics on Mastodon. In H. Cherifi, L. M. Rocha, C. Cherifi, & M. Don *Applications XII*, volume 1144, pages 366–377. Springer Nature Switzerland, Cham. ISBN 978-3-031-53502-4 978-3-031-53503-1. Series Title: Studies in Computational Int
26. Soh, J. H., Jansen, T. L. C., & Palacino-González, E. (2024). Controlling the nonadiabatic dynamics of the charge-transfer process with chirped pulses: Insights from a doubl scheme. *The Journal of Chemical Physics*, *160*(2), 024110 [DOI | http]
27. Tittes, C., Nijland, J., Schoentag, A. M. C., et al. (2024). Development of a genetic system for *Haloferax gibbonsii* LR2-5, model host for haloarchaeal viruses. *Applied and E* [DOI | http]



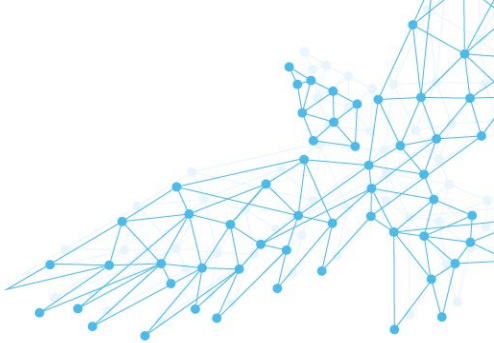
History

History

- 60 Years of Computing
- Started with ZEBRA ("zeer eenvoudig binair rekenapparaat") in 1958
  - 500 FLOPS!
  - For chemists, astronomers and mathematicians

university of groningen / center for information technology

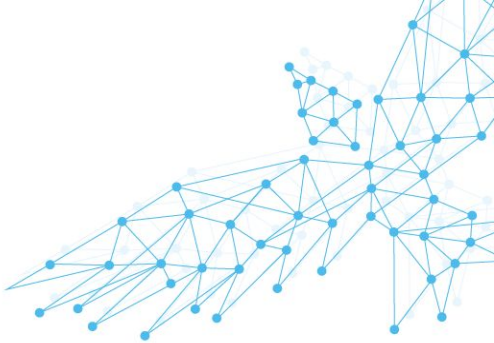
Introduction	*System*	*Top speed in FLOPS*	*Memory in bytes*
1958	Stantec Zebra	500	32.000
1964	Telefunken TR4	50.000	180.000
1972	CDC Cyber 74-16	1.000.000	1.000.000
1978	CDC Cyber 170/760	4.000.000	4.000.000
1989	CDC Cyber 962	5.000.000	60.000.000
1989	Convex C240	200.000.000	250.000.000
1993	Connection Machine CM5	2.000.000.000	500.000.000
1995	Cray J932	6.000.000.000	4.000.000.000
2002	Cray SV1e	60.000.000.000	32.000.000.000
2005	IBM BlueGene/L	34.400.000.000.000	3.000.000.000.000



Now

university of groningen / center for information technology

Today



- Flagship: Compute Cluster Hábrók
- Two OpenStack clouds in two locations:
  - Merlin (Data Centre Eemspoort)
  - Bateleur (Data Centre CBC)
- ILSE Lab
  - The Innovation Lab for Science and Education
    Together with DELL, Intel, Liqid, AMD, NGD Systems, Cornelis Networks, Bossers & Cnossen
- Founders of EESSI
  - European Environment for Scientific Software Installations

university of groningen / center for information technology

Clouds

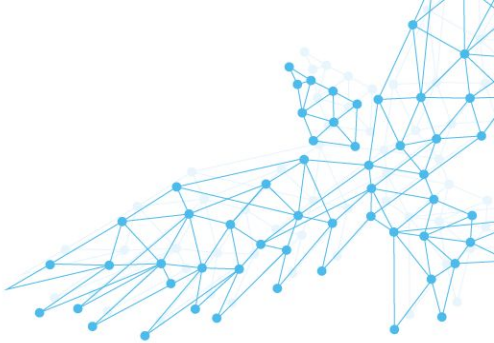
Bateleur:
23542 vCPUs
128.9 TB RAM
1.6 PB VM Storage
Bare-metal focused

Merlin:
1952 vCPUs
8.4 TB RAM
13.6 PB VM Storage
Virtualisation focused

Hábrók

- **~20k CPU Cores**
- **~200 Nodes**
  - Intel Icelake
  - Intel Skylake
  - AMD Zen
  - AMD Zen 3
- **60 NVIDIA GPUs**
  - V100, A100, H100, L40s
- **Omni-Path**
- **>3 PB Storage**

- **Terraform + Ansible**
- **Rocky 8**
- **Slurm**
- **Open OnDemand**
- **EasyBuild**
- **Lmod**
- **ReFrame**
- **Apptainer**
- **XDMOD**



EasyBuild @ University of Groningen

university of groningen / center for information technology

Back in the days…

- Environment Modules
- Shell scripts, manual installations, manually creating module files

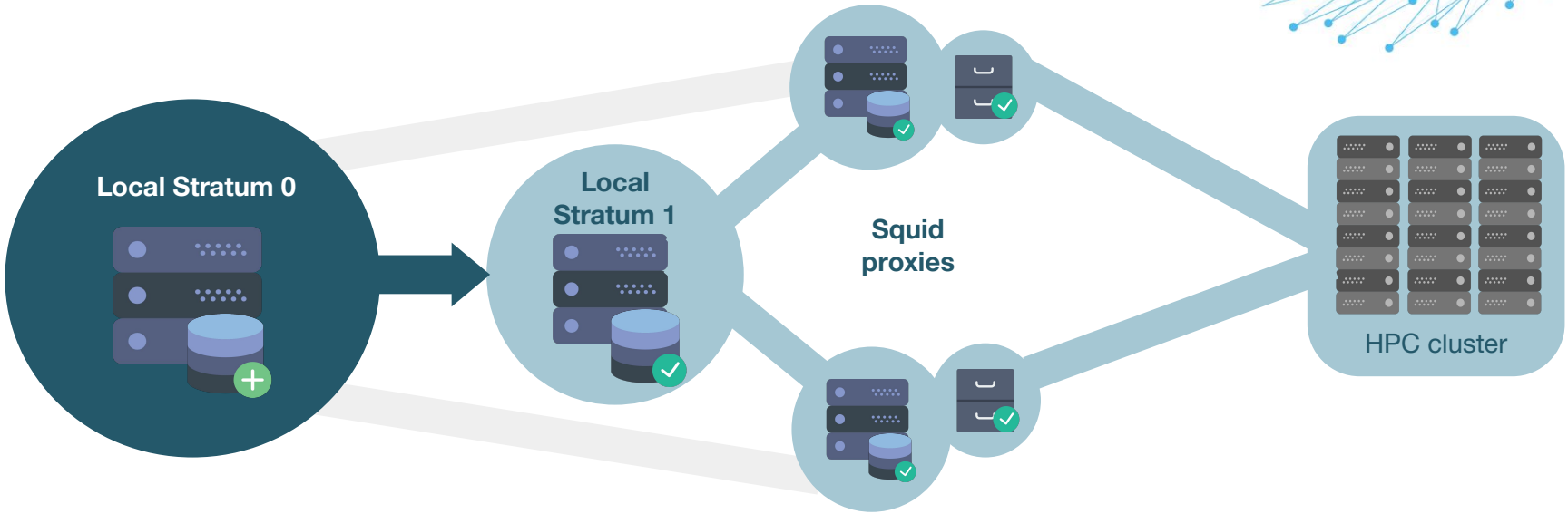
- Colleague learned about EasyBuild at SC14
  - Probably at 1st International Workshop on HPC User Support Tools (HUST-14)?
- Great timing, we were about to set up a new cluster
  - Lmod + EasyBuild
  - Almost no more manual installations
  - Quite minimal build container
  - Initially tried to stick to ~1 toolchain/year, modify other easyconfigs accordingly
    - Too much work…
    - We just install whatever is available now

Current setup

- Rocky 8, Lmod, EasyBuild
- Restricted apps on NFS share, everything else on CVMFS
- Easystack files
- Build container and configuration similar to the EESSI one
  - RPATH enabled (though without filtering `$LD_LIBRARY_PATH`)
  - Install to `/cvmfs` using writable overlay
  - Build tarball at the end, ingest on Stratum 0
  - Transactional ingestions
- It was too early to go EESSI all the way
  - No stable version yet, no GPU support
  - Using an EESSI-like build procedure, without compat layer, built for Rocky 8
  - EESSI is available via a module

university of groningen / center for information technology

Current setup



2 cores
4 GB RAM
500 GB disk

3 cores
16 GB RAM
500 GB disk

3 cores
16 GB RAM
100 GB disk

university of groningen / center for information technology

Current setup

- `/cvmfs/hpc.rug.nl/versions/2023.01/rocky8/x86_64/`
  - `amd/zen`
  - `amd/zen3`
  - `intel/icelake`
  - `intel/skylake_avx512`
  - `generic` -> large binary installations, e.g. MATLAB
- ~2200 modules per architecture -> almost 10k in total
  - ~350GB on CVMFS
- EasyBuild configuration file (environment variables) sets CUDA CC for specific node types
  - E.g. CC 7.0 for Skylake nodes (Nvidia V100)
- Bunch of hooks for site-specific things, e.g. module load messages for restricted modules
- https://gitrepo.service.rug.nl/cit-hpc/habrok/cit-hpc-easybuild

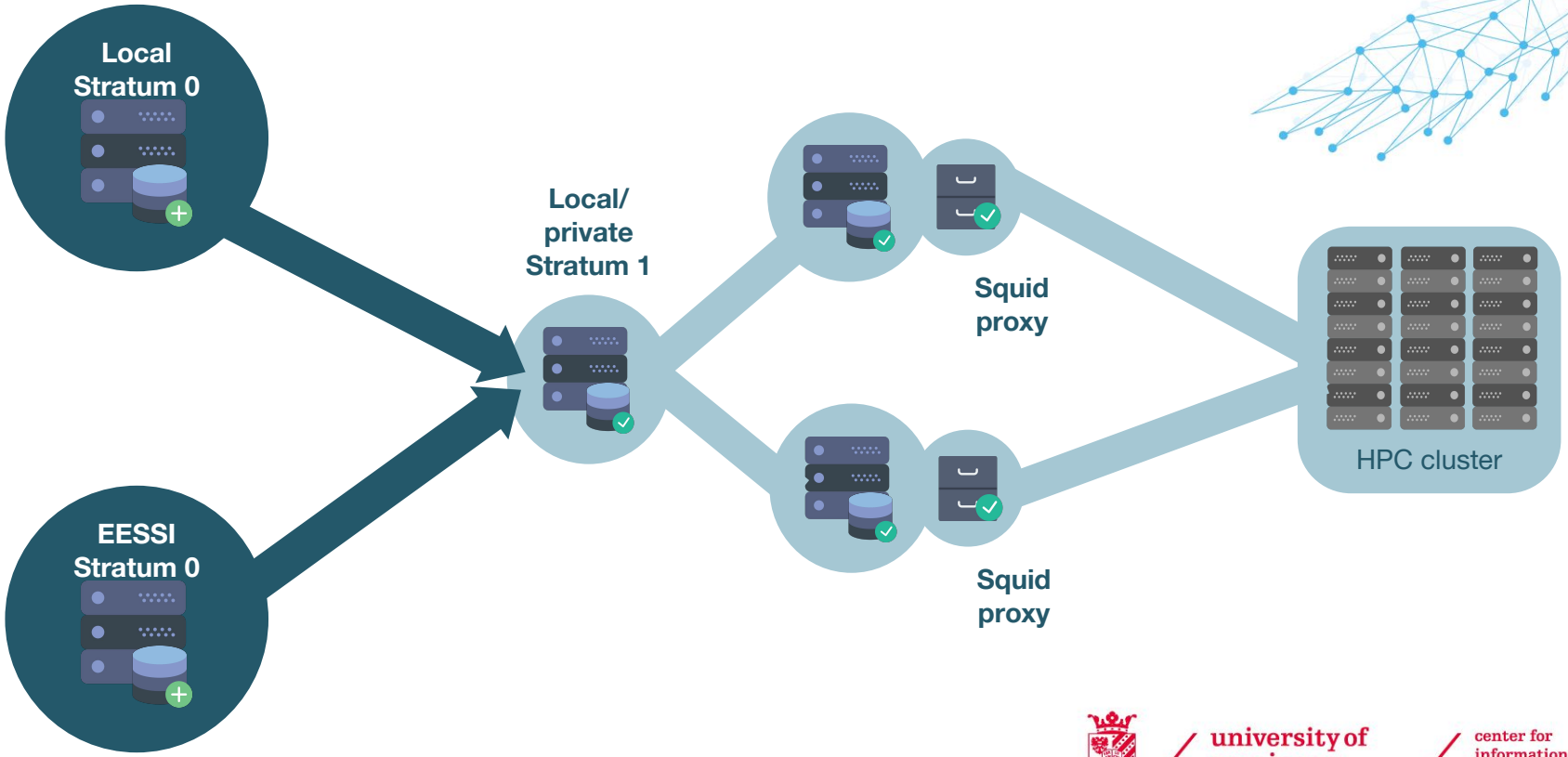
university of groningen / center for information technology

Short-term plans

- Migrate cluster from Rocky 8 to 9
- Create a new version of our local stack
  - Rebuild all/new installations on top of EESSI using EESSI-extend
  - Ingest to our CVMFS repository
  - Use the EESSI build bot for triggering the builds
- Similar kind of setup as the `dev.eessi.io` repository
  - See tomorrow's presentation
- This kind of setup will be shown in an EESSI webinar
  - https://gitlab.com/eessi/support/-/issues/135
  - Probably June 2

university of groningen / center for information technology

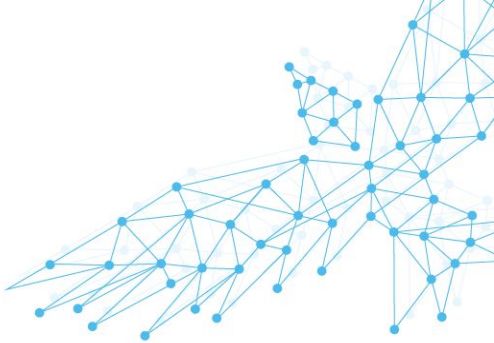
Short-term plans



university of groningen / center for information technology

Feedback

- EasyBuild is a great tool and has made our life much easier
  - May also raise the expectations from users? 😅
  - Excited to make things even easier with EESSI
- Toolchain concept can be hard to explain to users
  - Unclear names (of subtoolchains) don't help
- Python environments can still be a struggle
  - Conda… 😱
  - Mixing libraries from modules and virtual environments
  - Pip installing CUDA libraries
  - We often recommend users to just `pip install` all required libraries themselves in a virtual environment
  - Virtual environments created on one node type may not work on another node (same problem with R libraries)



Questions

university of groningen / center for information technology