

DeiC HPC Type-2 webinar

Dan Søndergaard

Type-2 consortium representative

HPC special consultant @ GenomeDK

June 10th, 2021

Introduction

Type-2 consortium: three universities, three facilities

- GenomeDK (AU)
- Computerome2 (DTU/KU)
- Sophia (DTU)

When do you need type-2?

High-performance or high-throughput computing

- Many jobs varying in size (workflow/pipeline)
- Few, very large jobs (MPI)
- In the order of 100k core hours

Substantial storage needs

- 1+ TB, often in “tens of terabytes”, if not “hundreds of terabytes”

Large amounts of sensitive data

- ISO 27001-compliance

Intended audience

Beginners

- Not afraid of technical documentation
- Some experience with Linux, writing scripts, using the terminal

Experienced

- Comfortable with technical documentation
- Previous HPC experience

Getting access to type-2 facilities

- Ask for resources through your local front office.
- Front office will contact the consortium representative (me).
 - You can wish for a specific facility, but the consortium has the final say.
- When a facility has been decided, user accounts/project will be created on the facility.
- The facility will inform the user about login procedure etc.

Facility specifications

	GenomeDK	Computerome2	Sophia
Hardware	AMD EPYC Rome 7452, 64 kerner i alt, 512 GB RAM, 2 TB SATA-disk	Xeon Gold 6230 Cascade Lake, 40 kerner, 192 GB RAM, 1.9 TB SSD-disk	AMD EPYC 7351, 32 kerner i alt, 128 GB RAM, 1 TB SATA-disk
Netværk	100 Gb/s Infiniband	100 Gb/s Infiniband	100 Gb/s Infiniband
Lager	Parallelt (BeeGFS) filsystem, total kapacitet på 11.5 PB, aggregeret båndbredde på 35 GB/s.	3-niveaus lager, 160 GB/s, i alt 8 PB.	Distribueret CEPH filsystem, total kapacitet 1 PB, aggregeret båndbredde 4 GB/s, BeeGFS-baseret burst buffer, 30 TB for høj I/O.
Køsystem	SLURM	Torque/Moab	SLURM
Node sharing tilladt	Ja	Nej	Nej
Netværkstopologi	Fat-tree	Fat-tree	Fat-tree
Backup	Mulighed for off-site disk-baseret backup.	Mulighed for backup på disksystem.	Mulighed for off-site backup på bånd.
Sikkerhed	To-faktor login. Mulighed for ekstra sikkert miljø (closed zone) der forhindrer læk af data. ISO 27001-compliant.	To-faktor login for alle, SecureCloud-løsning som tilvalg. ISO 27001-compliant.	Generel UNIX og HPC sikkerhed, ikke GDPR-compliant.
Anvendelsesfokus	Bioinformatik, life sciences, data science.	Bioinformatik, life sciences, data science.	Fysik/kemi, simulationer, optimering, MPI.
Understøtter begrænsning af ressourcer brugt af projektet (quota)	Ja	Nej	Ikke endnu
Brugsstatistik	Ja, løbende via kommando (data opdateres dagligt), samt månedlig mail	Ja, løbende via månedlig mail sendt til projektejer	Ikke endnu
Hjemmeside/ dokumentation	genome.au.dk	computerome.dk	dtu-sophia.github.io/docs/

Questions and support

- The user should direct any technical questions to the facility support staff.
- More general type-2 questions, complaints etc. should be directed to the consortium representative (me).

What to be aware of

- The agreement with DeiC *only* includes the hardware specified in the previous slide. If a user runs on other hardware (fat/GPU nodes), the facility may charge the user directly. *This is the user's responsibility!*
- Not all facilities limit projects to their allocated number of core hours, i.e., the user should keep track of usage herself (see the facility specifications).

Questions?