DeiC Throughput HPC

Status and challenges

Dan Søndergaard

HPC special consultant @ GenomeDK

DeiC Throughput HPC representative

What is throughput HPC?

Mostly unattended batch jobs.

Many jobs/many cores.

Large workflows (inhomogeneous jobs).

Requires some level of technical expertise to use (and that’s okay).

Consortium setup

DTU, KU, AU in a consortium.

AU represents the consortium and coordinates ”access” to the facilities.

Offers access to ressources on three different facilities:

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

The available hardware

DeiC has bought *capacity* on specific hardware from the consortium.

Not included in the agreement:

- Backup
- Other node types (fat, GPU) than what is stated in the table
- Any other services provided by the facilities

A surprise to many users!

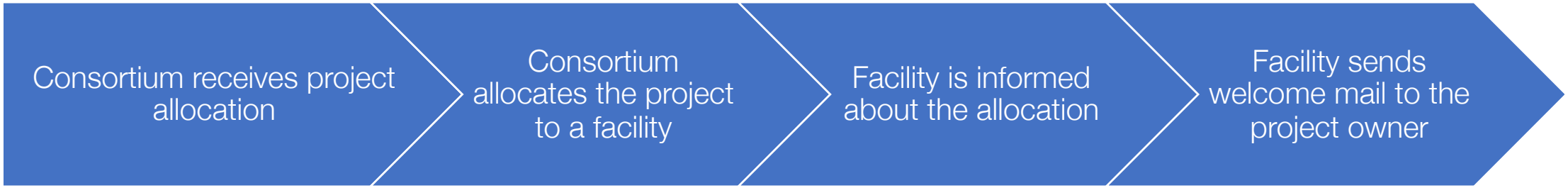
Users may be billed separately if such services are used.

	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 GDRP-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/

Project creation process

Your local front office or DeiC processes applications.

Once an application for access to DeiC Throughput HPC has been accepted, the consortium takes over.

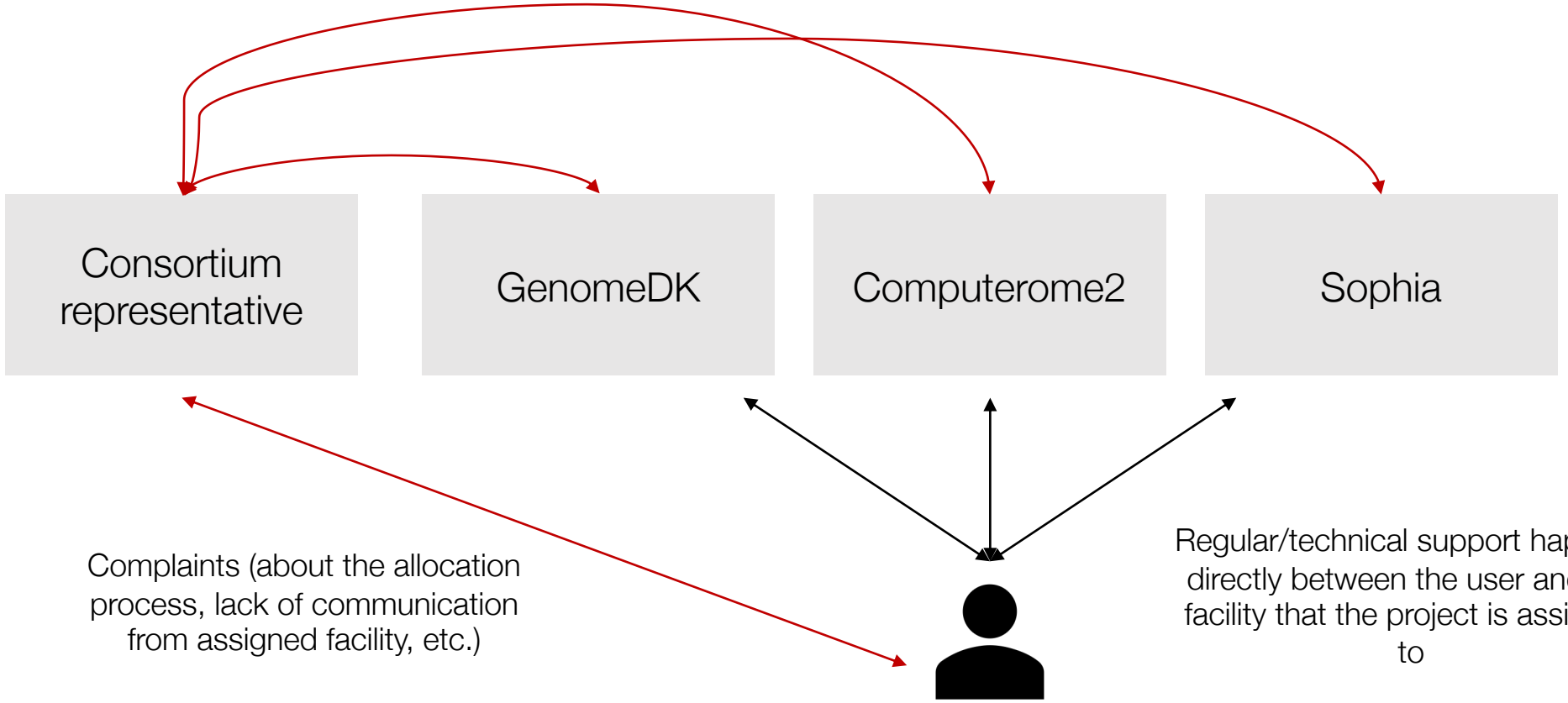


From either front office (local allocation) or DeiC (national allocation).

Based on preferences stated in the application and/or utilization of the facilities.

”Welcome mail” includes instructions for how to access the facility, how other project members get access, setting up two-factor etc.

Support flow



Challenges

Economic stability

Investments are very long-term and capacity is not easily expanded or reduced.

Uncertainty in economy makes planning hardware purchases impossible.

Agreements on a year-to-year basis are impractical. We must aim for 3+ years.

Electricity prices

The surging electricity prices have affected almost everyone in Europe.

High-performance computing and other power-intensive infrastructure is hit particularly hard.

We need to discuss how this is handled on a national level.

Organizational stability

National HPC (in its current form) is still new.

Many critical processes took years to understand, communicate, and agree on.

Bootstrapping a national HPC landscape is a *monumental* task!

We must now *ensure stability* so that we can *learn from our collective experiences* and *fine tune* understandings and communication accordingly.

Security and registry data

Computerome2 and GenomeDK are both ISO 27001-compliant and are used for processing of sensitive data types.

Researchers would like to access data from e.g. DST and Sundhedsdatastyrelsen on these facilities.

Complete lack of transparency regarding what is required of service providers to host this data. *Too much policitics!*

Conclusion

DeiC Throughput HPC offers foundational HPC services to all Danish researchers.

It’s been a long, troubled road, but there’s light ahead ☺

Stability, stability, stability.