CSC

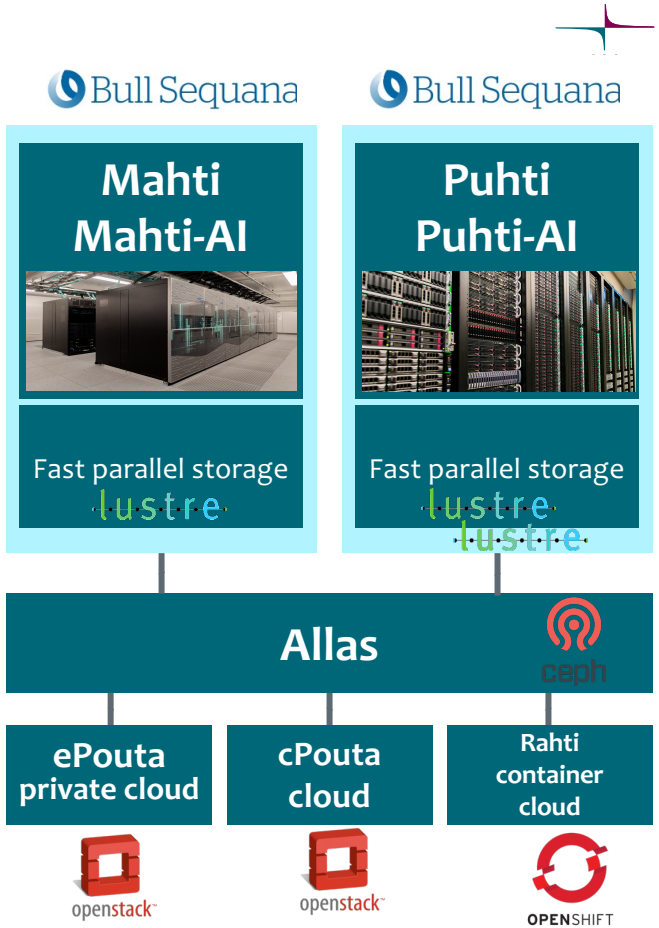
ICT Solutions for Brilliant Minds

Supercomputer storage and updated policies

Supercomputer storage on Puhti

- Lustre storage, for home projappl and scratch
  - 4,8 PiB
  - All disk areas in same volume – with 24 OSTs
  - https://docs.csc.fi/computing/lustre/ for description of Lustre
- Local disks
  - 94 CPU nodes with 1,5 – 5,9 TiB disks
  - 80 GPU nodes with 3,6 TiB disks
  - Login nodes with 2,9 TiB disks

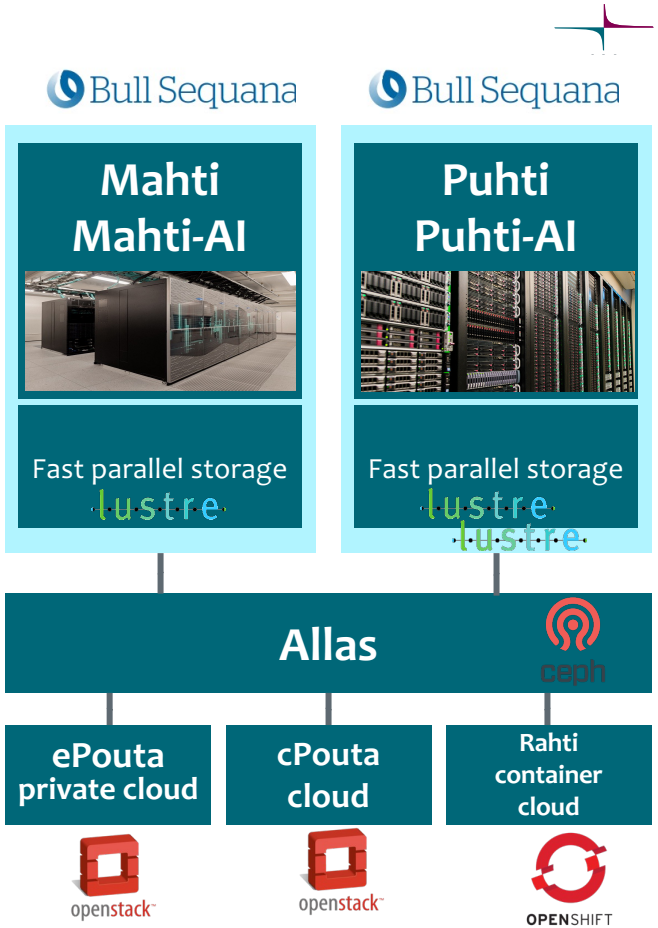
	Owner	Path	Cleaning	backup	Quota
home	Personal	/users/<user-name>	No	No	10 GiB
projappl	Project	/projappl/<project>	No	No	50 GiB (or more)
scratch	Project	/scratch/<project>	**180 days**	No	1 TiB (or more)



Supercomputer storage on Mahti

- Lustre storage, for home projappl and scratch
  - 8,7 PiB
  - Scratch 24 OSTs, home 8 OSTs, projappl 8 OSTs
- Local disks
  - 24 GPU nodes with 3,8 TiB disks
  - Login nodes with 2,9 TiB disks

	Owner	Path	Cleaning	Backup	Quota
home	Personal	/users/<user-name>	No	No	10 GiB
projappl	Project	/projappl/<project>	No	No	50 GiB (or more)
scratch	Project	/scratch/<project>	**Not yet**	No	1 TiB (or more)

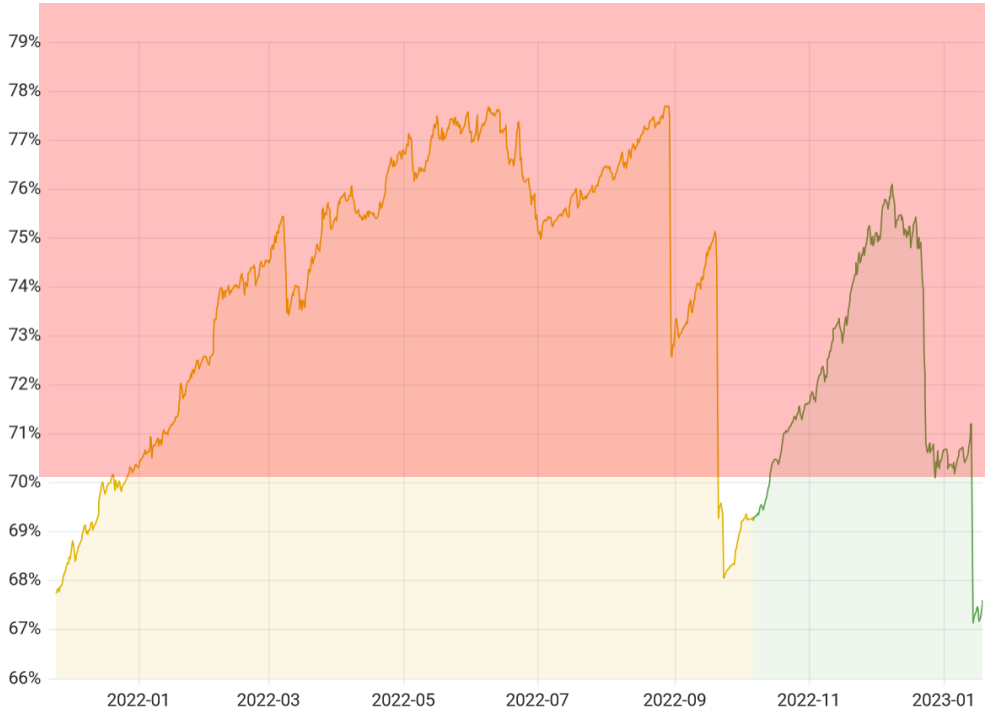


Data retention in Puhti scratch

- Scratch is meant for data that is in active use
  - It is not large enough, or designed for, storing data on a more permanent basis
- Large quotas are given out to projects so that they can store large amounts of data on a temporary basis
- Old data needs to be cleaned (deleted) regularly from scratch
- Main motivations:
  - There is limited space – total quota for all projects is 8+ PiB which simply does not fit
  - Performance is severely impacted when disk fills up

Puhti disk performance and usage level

- Slower performance due to
  - Fragmentation – more difficult to find contiguous space to store data and read it afterwards
  - More files and more complex directory structures (=metadata) makes it slower to locate files
- Puhti has had disk performance issues, need to minimize impact of disk usage
- Even with 12 Month cleaning usage level remained high – going to 6 Month cleaning



Puhti cleaning policy and procedure

- Files that have not been accessed for more than 180 days are identified periodically
  - Accessed includes both read and write access to files
- Lists of files to be deleted from scratch are provided to users
  - /scratch/purge_lists/<PROJECT NAME>/path_summary.txt
- Users then have pre-defined amount of time to manage this data
  - Delete, compress, move, archive
  - https://docs.csc.fi/support/tutorials/clean-up-data/
- Similar cleaning will be implemented in Mahti too, there usage is still reasonable with no impact on performance

Next Puhti cleaning cycle

- **January 16, 2023**: List of files that have not been accessed since July 20, 2022 collected
- **January 17, 2023**: Pre-warning of upcoming data cleaning/deletion process
- **February 1, 2023**: Per-project lists of files to be deleted provided to users. These include files
- **March 15, 2023**: Files still remaining on disk are deleted
- **March 16, 2023**: Next set of files to be deleted are collected (Not accessed since September 17 2022).

How to optimize your I/O

- Data intensive jobs, accessing large amounts of files and doing large amounts of IO operations are not optimized for Lustre
  - https://docs.csc.fi/computing/running/throughput/#inputoutput-efficiency
  - https://csc-training.github.io/csc-env-eff/hands-on/disk-areas/disk-areas-tutorial-fastdisks.html
  - Look at the linked tutorials for info on how and when to use local disks, ramdisks, or other optimizations such as SquashFS for containers.
  - Also installations can be heavy – use Tykky for your Conda environments https://docs.csc.fi/computing/containers/tykky/
- For ML workloads there is specific documentation
  - https://docs.csc.fi/support/tutorials/ml-data/

How to optimize your I/O

- For large files and more traditional HPC applications
  - https://docs.csc.fi/computing/lustre/#best-practices
  - https://docs.csc.fi/computing/lustre/
  - https://docs.csc.fi/support/tutorials/lustre_performance/
  - Use striping – spread out large files over multiple object storage targets (OSTs)
  - Use parallel MPI-IO or other parallel IO libraries
- Ask servicedesk for advice!


facebook.com/CSCfi


twitter.com/CSCfi


youtube.com/CSCfi


linkedin.com/company/csc---it-center-for-science


github.com/CSCfi