Webinar: CSC's new supercomputer Puhti and data storage system Allas
What is CSC upcoming computing and data environment

• Something familiar... but much more powerful
  o Puhti - Supercomputer with Intel CPUs

• Something a bit different
  o Puhti-ai - Supercomputer with GPUs
  o Mahti - Supercomputer with AMD CPUs

• Something new
  o Allas - Large storage system with access and usage possibilities beyond traditional filesystem.
Scientific drivers for new infrastructure

**Large scale simulations**
- For example climate change, space weather, fusion reactors, astronomical phenomena, particle physics

**Mid-scale simulations**
- For example materials science, energy technology, GIS

**Data-intensive computing**
- For example computational econometrics, bioinformatics, language research

**Data-intensive computing for sensitive data**
- For example medical research, register research

**Artificial intelligence**
- For example natural language research, business applications, computer vision

**Internet of Things (IoT) and data streams**
- For example satellites, weather stations, sensor networks

**New challenges**
The new Finnish research infrastructure for data management and computing

Balanced HPC ecosystem for supporting the six drivers

Heterogeneous, workload-optimized node architecture, support for complex workflows, datasets-as-a-service and containerization
Allas - object storage: what it is for?

- Allas is new storage service for all computing and cloud services
  - Meant for data during project lifetime
- Data can be stored and retrieved directly from anywhere in Internet
  - CSC supercomputers
  - Local workstation
  - Measurement devices
  - …
  - At simplest, web browser is enough
- Easy sharing of data outside project
  - Selected data can be shared publicly to Internet
- Arbitrary metadata can be added to data
Technical details
Puhti - computing cluster

Puhti has in total three partitions with 1002 compute nodes

1. CPU partition (Puhti)
   - 682 CPU nodes with in total 27280 cores
   - Floating point performance is 1.8 Petaflops

2. GPU partition (Puhti – AI)
   - 80 nodes with in total 3200 CPU cores and 320 V100 GPUs
   - Floating point performance is 2.7 Petaflops

3. FMI partition (Puhti-FMI)
   - 240 CPU nodes with in total 9600 CPU cores
   - Owned by FMI and operated by CSC - Not available for normal CSC users

• 4 Petabytes work disk for data under active use
Puhti - latest generation technologies

• All nodes equipped with latest generation Intel Xeon Scalable processors
  ○ CPU architecture released April 2019 - formerly codenamed Cascade lake
  ○ 20 cores per CPU, 40 cores per node running at 2.1 GHz
  ○ Supports AVX-512 instructions for vectorized computations - 2x speedup (theoretical) compared to Haswell CPUs in Sisu and Taito per core
  ○ Supports VNNI instructions for AI inference workloads - speedup up to 10x

• Infiniband HDR interconnect between nodes
  ○ First machine in the world!
Puhti-AI

- Puhti-AI
  - In total 80 nodes with 320 GPUs
  - GPUs are latest generation Nvidia GPUs (Volta) with 32 GB of memory per GPU – 2x more than in current Pascal GPUs
  - Very fast network: 2 x 100 Gbps links to each node
  - Each node has a very fast 3.2 TB local NVME disk

- Supported frameworks (plan)
  - Caffe, Tensorflow, Keras, PyTorch
  - Massive multinode workloads with Horovod
  - Higher-level frameworks

- Cloud environments (Pouta & Rahti) also have GPU capacity based on Pascal GPUs
## Technical specifications for nodes

<table>
<thead>
<tr>
<th>CPU partition</th>
<th>CPU</th>
<th>CPU cores @ 2.1GHz</th>
<th>Memory</th>
<th>Number of nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puhti CPU</td>
<td>Xeon Gold 6230</td>
<td>2 x 20 cores</td>
<td>192 GB</td>
<td>532</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@ 2.1GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xeon Gold 6230</td>
<td>2 x 20 cores</td>
<td>384 GB</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@ 2.1GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xeon Gold 6230</td>
<td>2 x 20 cores</td>
<td>384 GB + 3.2 TB NVMe</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@ 2.1GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xeon Gold 6230</td>
<td>2 x 20 cores</td>
<td>768 GB</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@ 2.1GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Xeon Gold 6230</td>
<td>2 x 20 cores</td>
<td>1.5TB</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@ 2.1GHz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puhti-AI CPU</td>
<td>Xeon Gold 6230</td>
<td>4 x V100</td>
<td>384 GB (Host)</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>32 GB</td>
<td>128 GB (GPUs)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.2 TB NVMe</td>
<td></td>
</tr>
</tbody>
</table>
Allas - storage

- **12 Petabytes** of storage space
  - For data stored over the life-time of a project
  - Storage space is **per CSC project**
- Object storage based on open source CEPH
- Data is stored as **objects** within a **bucket**
  - Blobs of data, can be anything (generally, **object** = file)
  - Project can have multiple **buckets**
  - Only one level of hierarchy of buckets (no buckets within buckets)
  - Each object can have an URL and accessed over https e.g. with web browser: [https://a3s.fi/project_id/my_bucket/my_object](https://a3s.fi/project_id/my_bucket/my_object)
**Allas - storage**

- Objects can have metadata
  - Metadata are a key-value pairs, e.g. “content=shoe”

- Objects are replicated so a disk break does not cause data loss.
  - There is no backup i.e. if a file is deleted, it cannot be recovered

- Data cannot be accessed from supercomputers with standard Unix commands (ls, cp, mv, etc.) but one needs to use object storage specific tools (Initially S3 and Swift interfaces).
  - For computations data needs to be typically copied to/from supercomputer with these tools

- Variety of data management features to be built on top of it
Mahti - Supercomputer

- Next generation AMD Rome system CPUs
  - About 200,000 cores in total
  - Especially for large scale simulations, but also for other use cases
- 8 Petabytes of work disk for data under active use
Installation and decommission schedule

• April – July 2019: Puhti and Allas installation and acceptance testing
• August 2019: Puhti and Allas available for customers
  ○ Both CPU and GPU partitions of Puhti
• August 2019: Sisu decommissioned
• Late 2019 – early 2020: Mahti installation and acceptance testing
• End of 2019: Taito decommissioned
• Early 2020: Mahti available for customers
What changes in the new infrastructure?
Access to new systems and data migration

• All users need to apply for new services via CSC customer portal
  o Puhti and Allas access can be applied from early June on

• Users need to transfer their data from current HOME, WRKDIR and project directories to new infrastructure
  o Data should be first migrated to Allas and then to supercomputers when needed
  o Detailed instructions and guidance from CSC will be available

• Current data in Taito and Sisu remains available until end of 2019
Storage in new infrastructure

- Allas is the common storage platform for all systems
  - project duration storage and sharing of data (3-5 years)
    - More long term storage in other services (CSC FAIR data, university services, ...)
  - Storage is provided **per project**, default quota 1 TB
    - Project manager can apply for more quota
  - Used storage space **consumes billing units**
  - No backups
Puhti and Mahti have separate file systems with common structure

- **HOME**: User specific directory for small data. Default quota 10 GB.
- **APPL**: Project specific directory, for example for sharing projects own application codes. Default quota 50 GB.
- **SCRATCH**: Project specific area for temporary data, i.e. intermittent simulation results. Default quota 500 GB.
  - Similar to WRKDIR in current systems (however, WRKDIR was user specific)
  - Used storage space **consumes billing units**
  - **Automatic cleaning**: Files will be deleted 90 days from last access, relevant data should be moved to **Allas**
- **No backups in any of the storage areas**
Running preinstalled applications

- Scientific software offering by CSC remains mostly the same
- Similar **module** system as previously
  - i.e. `module load r-env`, `module load ccdc`, etc.
- Optimum runtime parameters (number of CPU cores etc.) most likely different from current systems
- Similar SLURM batch job queuing system as previously
  - New queues and policies (number and type of nodes, running times)
  - Recommended to write new batch job scripts starting from the templates in CSC user guides
Installing applications by yourself

• New software stack
  o GNU and Intel compilers
  o Various high-performance libraries
  o BullMPI MPI library (based on OpenMPI)

• Applications should be rebuild
  o Configure scripts, Makefiles etc. may need modifications

• Applications should be installed in APPL disk area
  o Easier sharing for the whole project
Developing applications by yourself

- New performance analysis and debugging tools
- Efficient use of Mahti might require further optimisation work e.g. hybrid OpenMP/MPI parallelization
- Own applications should be installed in APPL disk area
  - Easier sharing for the whole project
Training

• Several workshops and webinars about new infrastructure in 2019-2021
  ○ Both basic and advanced level
Timeline of changes

2019

- April: Taito in use; data available
- June: Sisu in use
- July: Puhti and Allas acceptance
- August: Puhti and Allas in production

2020

- January: Mahti acceptance
- January: Mahti in production
Important dates for users

- Early June 2019: request access to **Puhti** and **Allas**
- August 2019: Access to **Puhti** and **Allas** open
- August 16\(^{th}\) 2019: Sisu decommissioned
  - Data can still be accessed via Taito
- January 1\(^{st}\) 2020: Taito decommissioned
- Early 2020: Access to **Mahti** open
- End of 2019: Current data in Taito and Sisu no longer available
Questions?

Up-to-date information about timetables, relevant changes for users etc. at

research.csc.fi/dl2021-utilization