



# CSC computing resources

**Tomasz Malkiewicz, Jura Tarus, Stina Westman and Thomas Zwinger**  
**CSC – IT Center for Science Ltd.**

# Program



- ➡ 10:30 - 12:00 How to utilize new CSC computing resources in your research
  - Supercomputing resources: Sisu and Taito (75')
  - Services for sharing your data (15')
  - Interactive, questions are welcome!
- ➡ 12:00 - 12:30 Round robin / free discussion
  - What needs do you have?
- ➡ 12:30 -> F2F meetings



# CSC at glance



- Founded in 1971
- Owned by Ministry of Education and Culture
- Operates on a *non-profit* principle
- Staff ~260 people
- Facilities in Espoo and Kajaani
- **Free of charge services for higher education institutions in Finland**



# Supercomputing resources



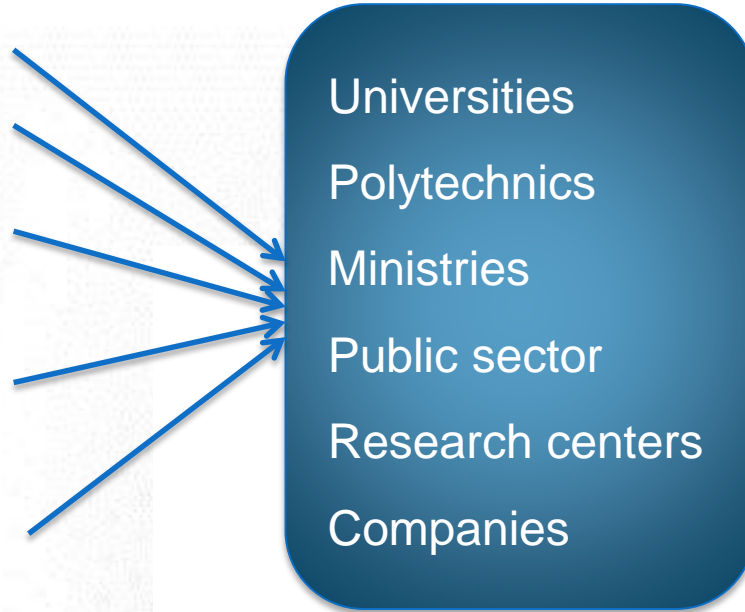
- CSC supercomputers Phase 2
  - *Compute in Sisu* (Cray XC40)
  - *Compute in Taito* (HP cluster)
  - Data storage
  - GPUs and Intel Xeon Phi in Bull, Taito-shell
  - FGI and Cloud
- Training



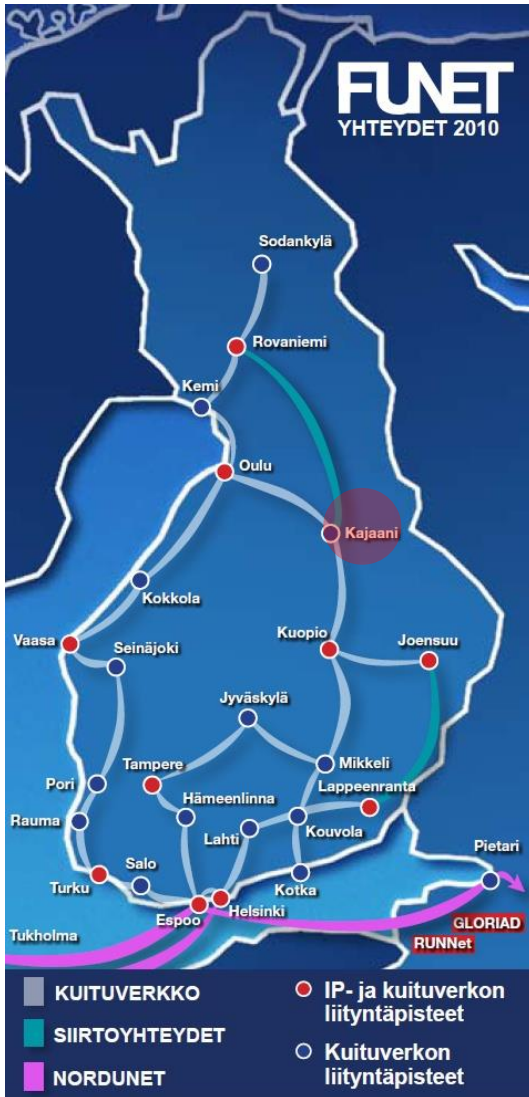
# CSC's Services



- Services for Research
  - Computing Services
  - Application Services
- FUNET Services
- Data Services for Science and Culture
- Information Management Services



# Datacenter CSC Kajaani



# Kajaani HUB room



# Modular Datacenter (MDC)

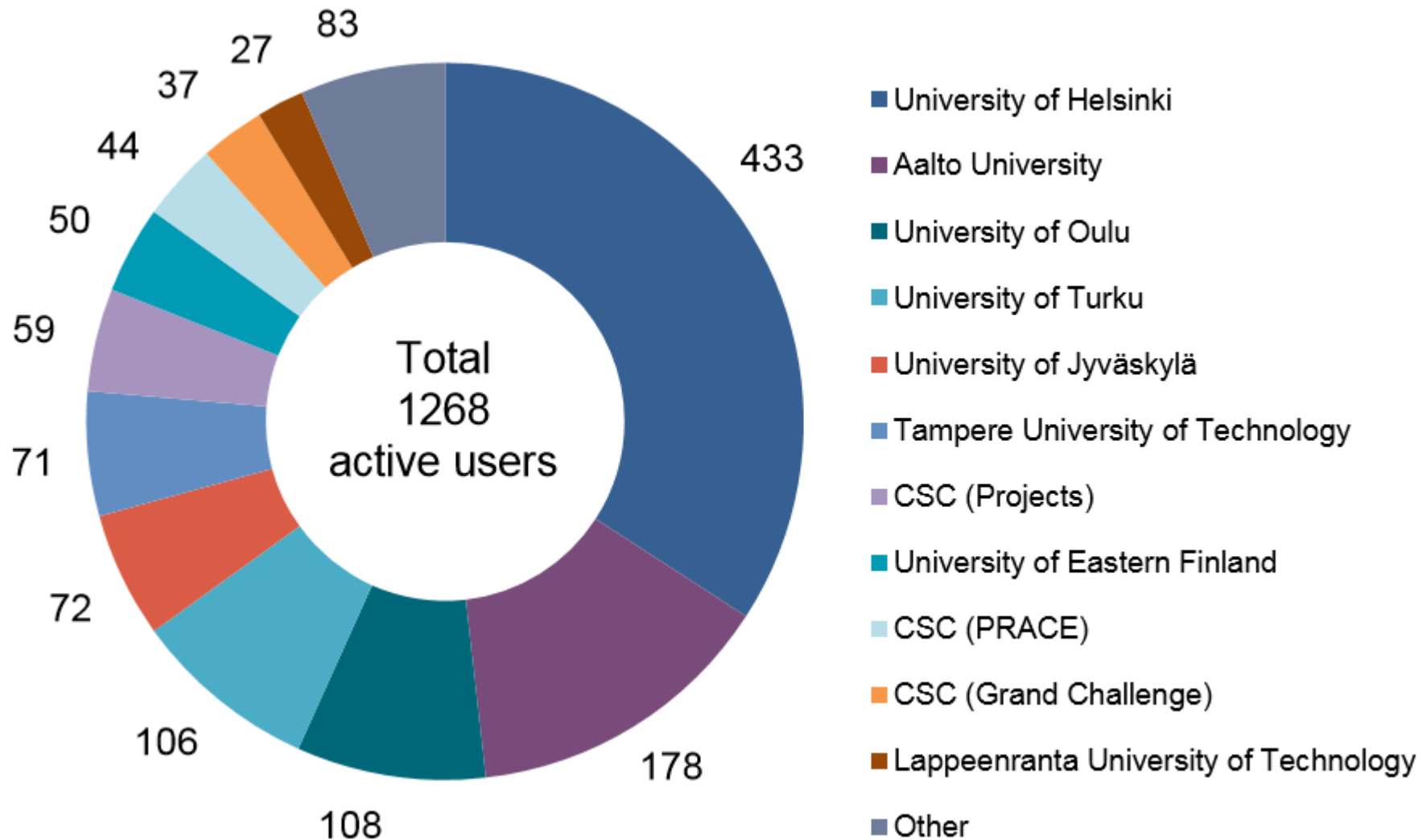


- About 700 active computing projects
  - 3000 researchers use CSC's computing capacity
  - 4250 registered customers
- Haka-identity federation covers all universities and higher education institutes (287 000 users)
- Funet - Finnish research and education network
  - Total of 370 000 end users

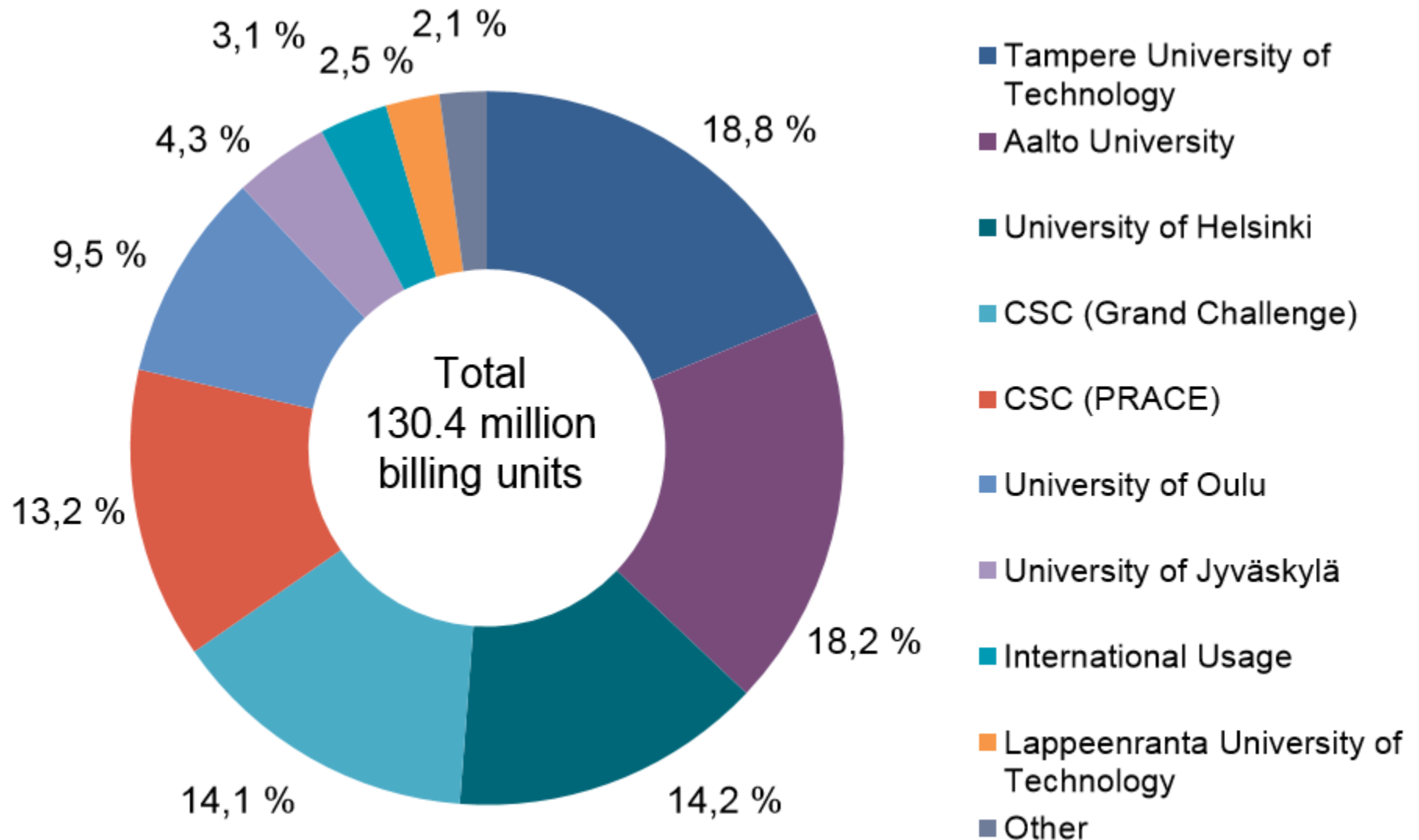


# Users of computing resources by organization

## 1H2014



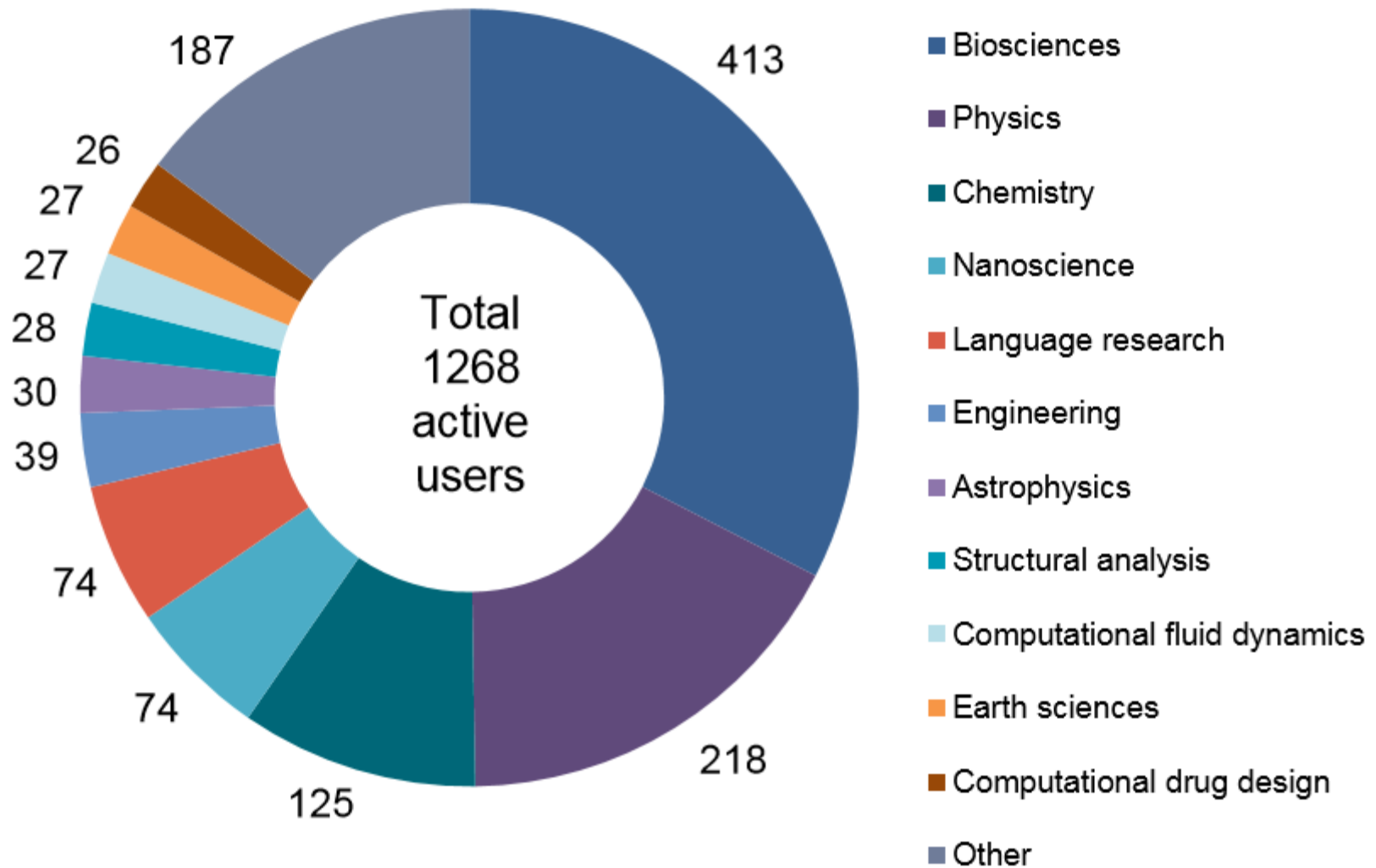
# Computing usage by organization 1H2014



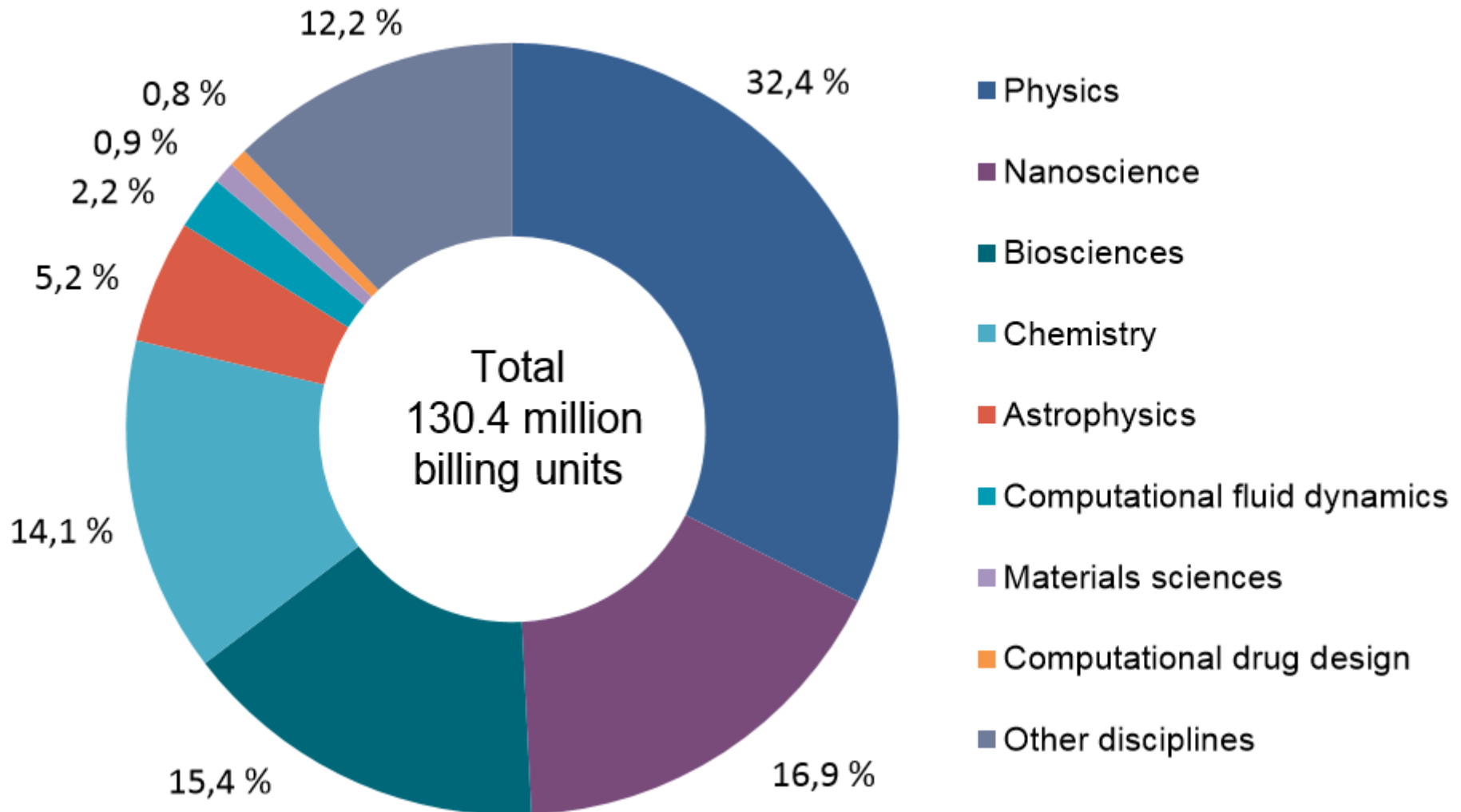
# Users of computing resources by discipline

## 1H2014

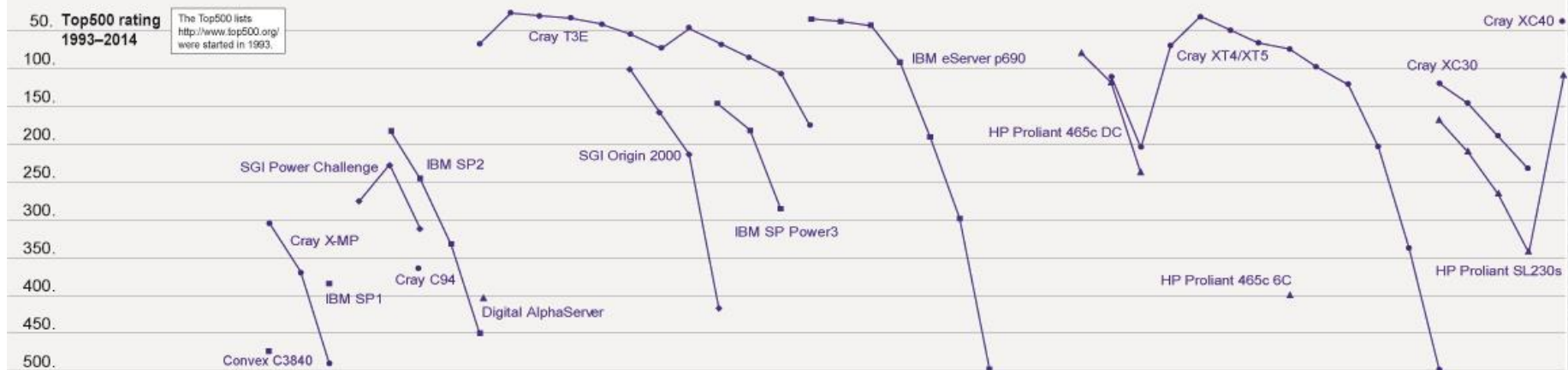
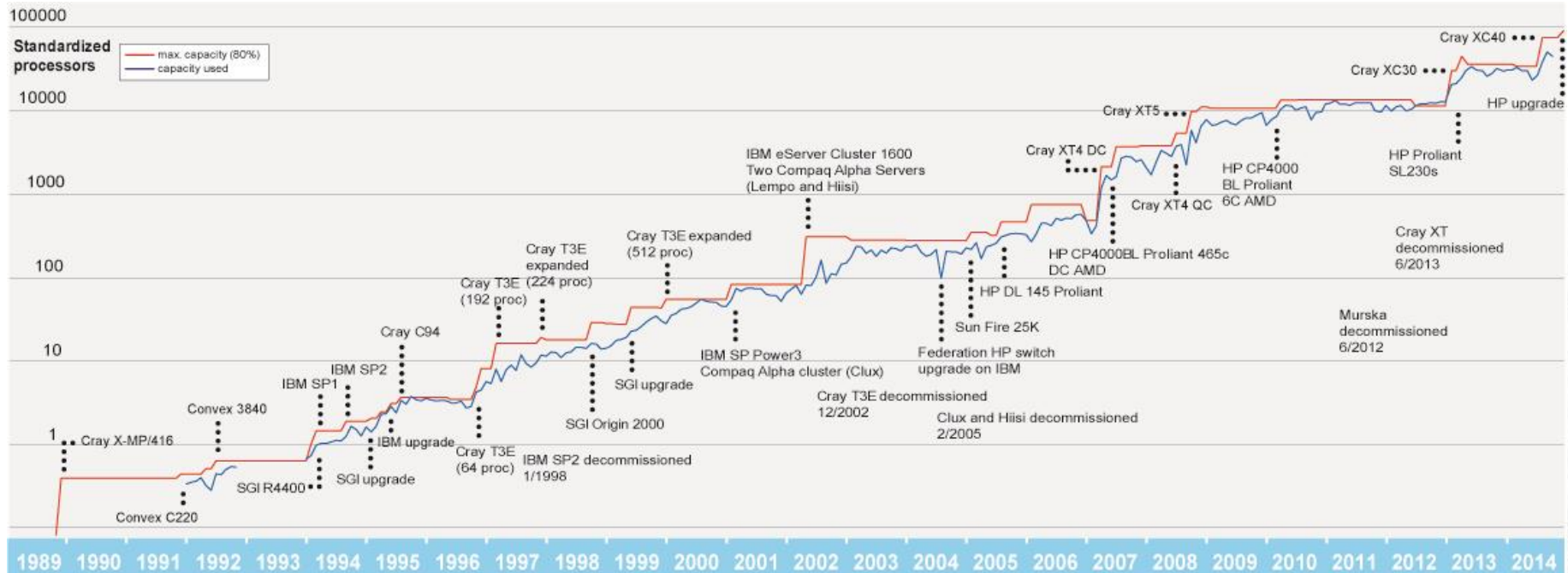
CSC



# Computing usage by discipline 1H2014



# CSC Computing Capacity 1989–2014



# Software and database offered by CSC



- Large selection (over 200) of software and database packages for research <https://research.csc.fi/software>
- Mainly for academic research in Finland
- Centralized national offering: software consortia, better licence prices, continuity, maintenance, training and support

## Services for Research

Home Sciences Computing **Software** News Support Sci

Services for Research → Software → Software Packages

### Software

Software Packages

Programming

Parallel Computing

Code Optimization

Visualization

Source Software

Development at CSC

### Software Package

All software packages in alphabet

Title

Abaqus

ABYSS

Acquis Communautaire Multiling

ADF

afterburner

Ajatella, Miettä, Pohtia, Harkita

ALLPATHS-LG

Amber

ANSYS Academic Research

ANSYS Academic Teaching Intr

ANSYS CFX

ANSYS Fluent

ANSYS ICEM CFD

ARB

ArcGIS

AutoDock

Babel

BEAST



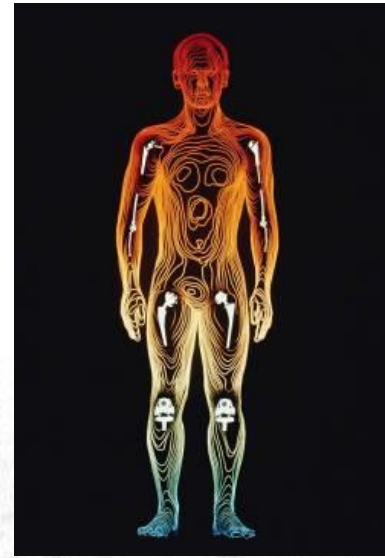
The collage features several scientific computing interfaces and logos. At the top left is the Gromacs website interface. Next to it is the Gaussian.com logo with the text 'THE OFFICIAL GAUSSIAN 03 WEB SITE'. Below these are screenshots of the ADF (Amsterdam Density Functional) software interface, showing molecular models and calculation parameters. At the bottom left is the PERCH logo with the text 'Solutions Ltd'. In the center is a logo for 'PEAK research NMR Software' featuring a fish. To the right of the fish is a screenshot of a presentation slide titled 'CSC presentation' showing a molecular model. At the bottom right is the NWChem logo with the text 'high performance computational chemistry software'.

# New customers

Apply for CSC account:

<https://research.csc.fi/accounts-and-projects>

- Most of CSC services are free for academic researchers,  
but usually a **CSC user account** is required.
  - Basic usage: register as CSC customer via SUI
  - Larger computing resources via an application form
- Benefits
  - A wide selection of scientific programs and databases available at CSC servers.
  - ICT resources and science-aware support  
([servicedesk@csc.fi](mailto:servicedesk@csc.fi))
  - Courses and events covering many areas are organized regularly.
  - Guide books and magazines in PDF.
  - CSC's research and development to improve services.
  - Networks bring together people with similar interests in science and technology.



# HPC PHASE 2 RESOURCES

- SISU, TAITO
- BULL
- STORAGE: DDN (PHASE 3)

# Sisu: Cray XC40 Supercomputer

- For large parallel jobs
- Intel Haswell processor E5-2690 v3 product family; 2,6 GHz (phase 1 Sandy Bridges replaced)
- Cray Aries Interconnect
- 40 512 cores
- 24 cores per node
- 64 GB memory per node



# Sisu Phase 2 features



- AVX-2
  - May need to optimize for wider vectors' size
  - Max 16 flop/cycle
- DDR4
  - Higher bandwidth, lower power consumption
- Max job size increased
  - 400 nodes = 9600 cores
- Native SLURM on the way
  - We might be moving to it at some point

# Running on Sisu Phase 2

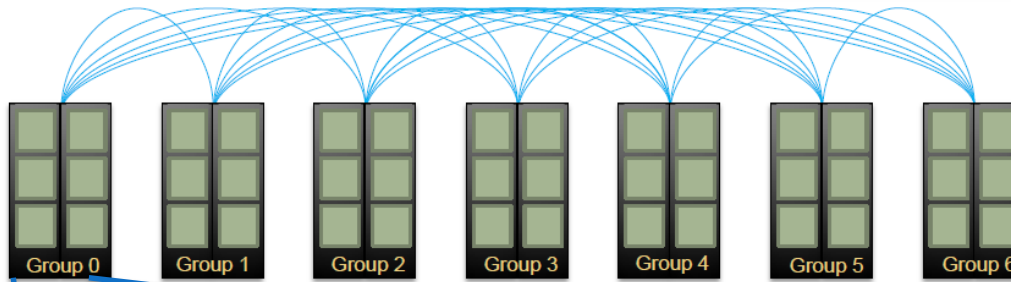
## ➤ Sisu guide

- <https://research.csc.fi/sisu-user-guide>
- Phase 1 binaries (static) may or may not run, **CSC strongly advises to recompile your code (and compare performance)**
- Login nodes are still based on Sandy Bridge (as they were in Phase 1)
  - Cross compiling is required (semi-automatic)
  - Haswell optimized code will not run in login nodes

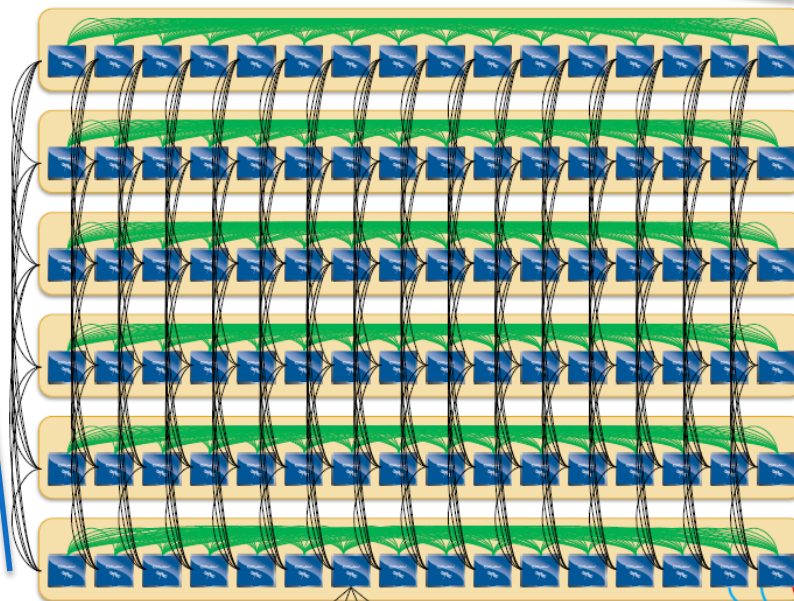
## ➤ Scalability tests required for jobs bigger than 1008 cores

- <https://research.csc.fi/sisu-scalability-tests>
- Large test queue available

# Cray Dragonfly Topology



All-to-all network  
between groups



2 dimensional  
all-to-all network  
in a group



4 nodes connect  
to a single Aries

Optical uplinks to  
inter-group net

CSC presentation

Source:  
Robert Alverson, Cray  
Hot Interconnects 2012 keynote

# Taito: HP Supercluster



- For serial and small parallel jobs
- Intel Sandy Bridge (phase 1) & Intel Haswell (phase 2) processors
- FDR InfiniBand interconnect
- 18 984 cores
- Different memory configurations: 64, 128, 256 GB and 1.5 TB per node



# Taito is a heterogeneous cluster



- Different jobs need different resources
  - Bulk Sandy Bridge compute nodes
  - Largemem Sandy Bridge compute nodes
  - Hugemem Sandy Bridge compute nodes
  - Bulk Haswell compute nodes
  - Taito-shell (Sandy Bridge) for interactive work
- Choosing between Haswell and Sandy Bride nodes
  - `--constraint=snb` or `=hsw`
  - More details:
    - <https://research.csc.fi/taito-constructing-a-batch-job-file#3.1.4>
- Local `/tmp` disk 2 TB on each node
  - reserve only what you need

# One SLURM to serve them all...



- Do old applications run on new CPUs?
  - Should run, CSC **recommends re-compiling**
    - Possible problems due to libraries, not architecture
  - Build your software for both (old and new) architecture
  - Gain depends on architecture
- Batch job scripts need to be updated
  - Number of cores per node: Phase 1: **16**, Phase 2: **24**
  - 128 GB memory in most nodes + 10 with 256 GB
  - Instructions will be available through user guides
  - A parallel job into *either* SB or Haswell (not mixed)

# Preinstalled modules: optimizations



What architectures a code has been optimized for

- (h) only Haswell → runs only on Haswell
- (sh) Haswell and Sandy Bridge → runs optimally on both
- (g) GPGPU aware → needs to be run on taito-gpu.csc.fi
- No entry → optimized for Sandy Bridge: should run on both SB/H, but not optimally on Haswell (or even SB)

```
[GPU-Env ~]$ module avail
```

```
----- /appl/gpu_modulefiles/Compiler/gcc/4.8.2 -----
intelmpi/4.1.3      mkl/11.1.1      openblas/0.2.8
magma/1.4.1 (g)    mvapich2/2.0-GDR (g)    openmpi/1.8.1 (g)

----- /homeappl/appl_taito/gpu_modulefiles/Core -----
StdEnv      cuda/6.0      gcc/4.8.2 (D)    intel/14.0.1    pgi/14.4
cuda/5.5     cuda/6.5 (D)  gcc/4.9.1      pgi/14.3      pgi/14.7 (D)

----- /homeappl/appl_taito/gpu_modulefiles/Linux -----
amber-env/14-cuda (g)    namd-env/2.10b1-cuda (g)    totalview/8.13.0-0
git/1.9.2          python-env/2.7.6 (D)    totalview/8.14.0-21 (D)
gromacs-env/4.6.6-cuda (g)    python-env/3.4.1
```

Where:

(g): built for GPU  
(D): Default Module

# SLURM configuration: Fair usage

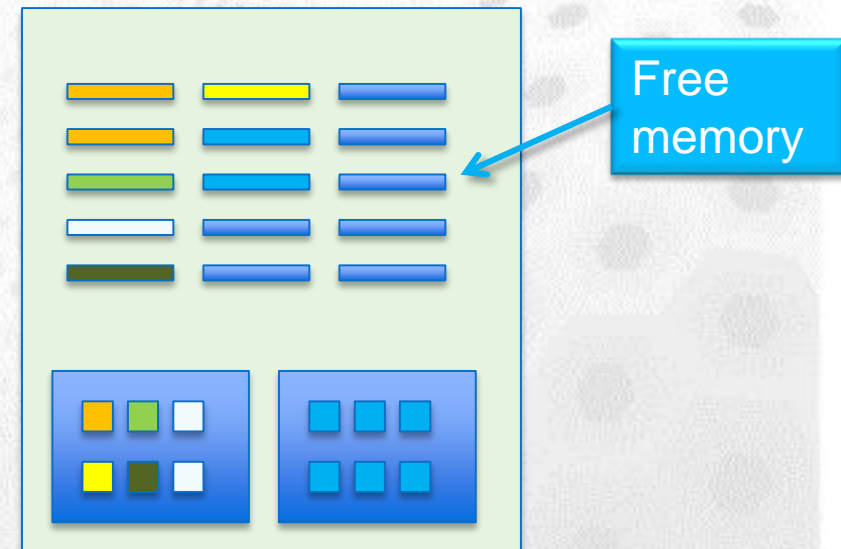
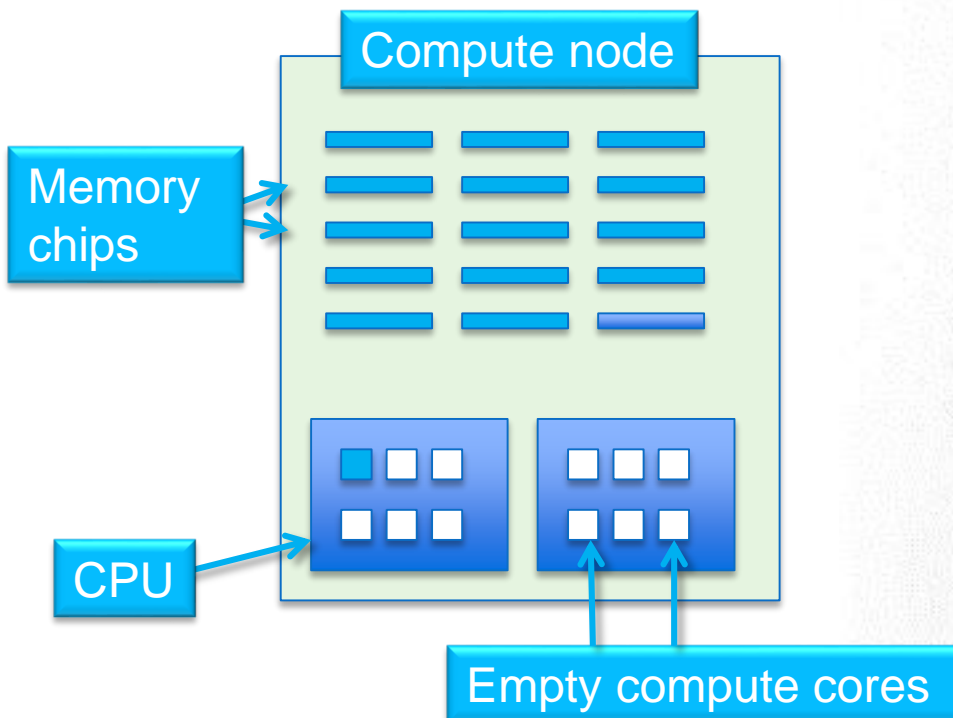


- ➊ SLURM uses fair share: the highest priority jobs go into execution next
  - Priority is decreased by the total amount of resources used in last 2 weeks per user
  - Priority is increased by time spent queueing
  - Backfiller will try to put small jobs into gaps due to current available resources and highest priority job
  - Jobs labeled "Association limit" are not eligible to run (due to too many jobs in queue by the user)
- ➋ *Due to abuse, a maximum limit of jobs in queue now enforced*
- ➌ Chain jobs (--dependency -flag for SLURM) if you need long running time
- ➍ Don't overallocate memory (add this command to your batch script  
`used_slurm_resources.bash` will print requests vs. used at stdout)
  - If you request a full node (-N 1), use `--mem=55000` instead of `--mem-per-cpu=something`
  - If you see abuse or think that the setup is unfair, contact [servicedesk@csc.fi](mailto:servicedesk@csc.fi)
- ➎ SUI has a monitoring tool for your jobs and used resources (*Services -> eServices -> My Project*)

# Batch system: requesting memory



- One job requesting most memory on a compute node.
- Many jobs each requesting little memory.
- All cores usable, some memory left unused (this is the normal case)
- Most cores unusable (this is ok **if** all that memory is **needed** by the job)



# How much resources my job used?

## ➤ After a job: check used/requested resources:

– `sacct -j <jobid> -o MaxRSS, AveRSS, ReqMem, Elapsed`

```
[asillanp@taito]$ sacct -j 2583042 -o jobid,maxrss,averss,reqmem,elapsed
```

JobID	MaxRSS	AveRSS	ReqMem	Elapsed
2583042		512Mc	00:00:04	
2583042.bat+	3300K	3300K	512Mc	00:00:04
2583042.0	74972K	25688K	512Mc	00:00:02

Maximum used memory per core= 74971kB= 73MB = not much

Average used mem per core 25688kB -> multiply with # of cores to get mem per job

```
[asillanp@taito-login4 d.dppc]$ sacct -j 2889733 -o jobid,maxrss,averss,reqmem,elapsed,alloc
```

JobID	MaxRSS	AveRSS	ReqMem	Elapsed	AllocCPUS
2889733		3000Mn	00:02:30	16	
2889733.bat+	7484K	7484K	3000Mn	00:02:30	1
2889733.0	119748K	119748K	3000Mn	00:02:27	16

**Mc** = memory per cpu  
(request with --mem-per-cpu=XXX [in MB])

**Mn** = memory per node  
(request with --mem=XXX [in MB])

## Taito Phase 2

### ➤ Porting strategy

- Compilers, libraries, flags, ...
- Add **AVX-2 flag** when compiling your code
- Flags
  - `icc -xAVX -axCORE-AVX2,CORE-AVX-I`
  - `gcc -march=sandybridge -mtune=haswell`
- CSC ports and optimizes important applications for the new architecture
- Consider testing your code on Sisu, which has Haswell CPUs (or join pilot usage)

- Official opening on **1.10.2014**
- Direct liquid cooled, very energy-efficient
- ***Accelerators and co-processors***
  - 38 NVIDIA K40 nodes = 76 GPUs
    - 12 GB memory per card
  - 45 Intel Xeon Phi (MIC) nodes = 90 Xeon Phis
    - 16 GB memory per card
  - Energy efficient (slow ...) CPU's

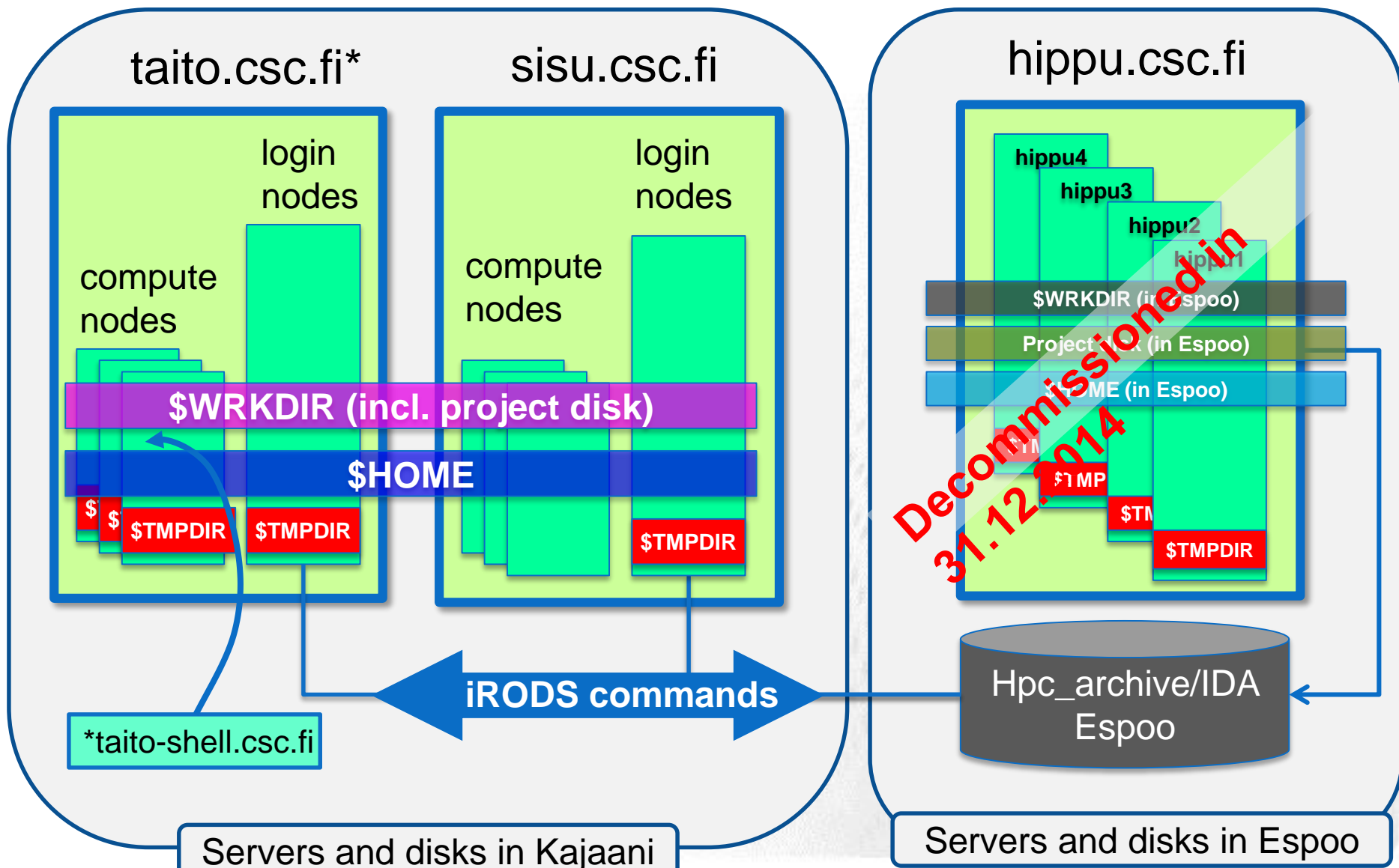
# How to access Bull

- Logically part of Taito
- Accessing the resources
  - Intel Xeon Phi: `ssh taito-mic` (from `taito.csc.fi`)
    - Still in beta phase
  - NVIDIA K40: `ssh taito-gpu.csc.fi`
- <https://research.csc.fi/taito-user-guide>
  - [taito-gpu](#)
  - [taito-mic](#)

## Fast and large storage: DDN Phase 3

- HPC storage used by Sisu and Taito
- System size increased to ~4 PB
  - About 1.9 PB added to the current configuration in early October 2014
  - Aggregate bandwidth > 80 GB/s (previous ~48 GB/s)
- Was made available in autumn 2014

# Directories at CSC Environment



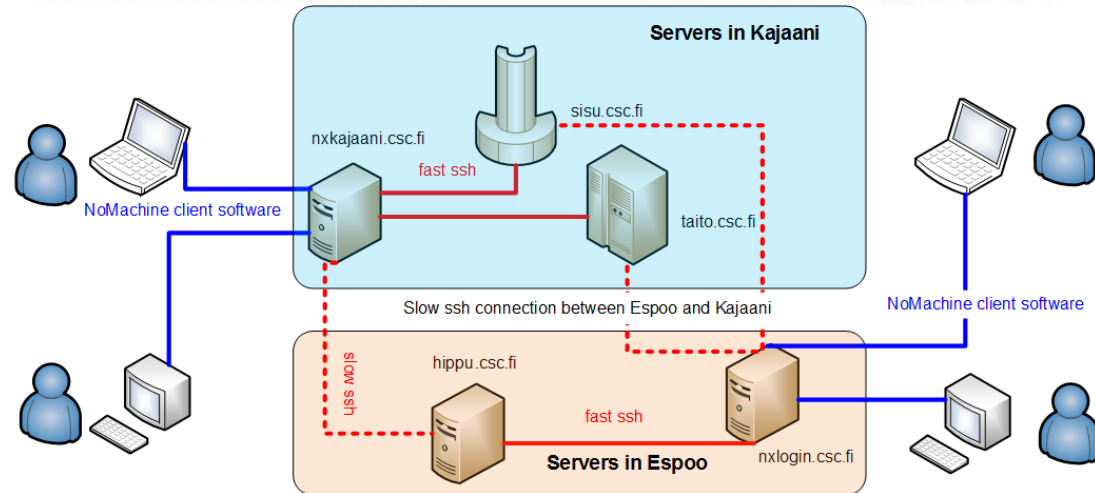
# Disks in total



- *4.0 PB on DDN*
  - \$HOME directory (on Lustre)
  - \$WRKDIR (*not backed up*), soft quota 5 TB / user
  - Up to 100 TB / project
- *HPC Archive*
  - 2 TB / user, common between Sisu and Taito
  - Up to 100 TB / project
- *3 PB disk space through TTA/IDA*
  - 1 PB for Universities
  - 1 PB for Finnish Academy (SA)
  - 1 PB to be shared between SA and ESFRI
  - more could be requested
- */tmp on Sisu and Taito (around 1.8 TB) to be used for compiling codes on login nodes*

# NoMachine Remote Desktop

- **Client connection** between user and gateway
- Good performance even with slow network
- **Ssh** from gateway to server (fast if local)
- Connect to right gateway
  - nxkajaani.csc.fi
  - nxlogin.csc.fi
- Persistent connection
- Suspendable
  - Continue later at another location
- Read the **instructions**...
  - ssh-key, keyboard layout, mac specific workarounds, ...
- Choose an application or server to use (right click)
- Soon possible to pilot also a web client: no installations



# Grand Challenges

- ➔ Normal GC (*call in half a year / year intervals*)
  - CSC resources available for a year
  - <https://www.csc.fi/grand-challenge-proposals>
    - ➔ Currently open call deadline 2.03.2015
- ➔ Remember also PRACE/DECI calls
  - CSC supports the technical aspects of the applications



# Grid computing with Finnish Grid Infrastructure (FGI)



ARC Grid Monitor

2014-05-27 CEST 12:45:37

Processes: ■ Grid ■ Local



Country	Site	CPU	Load (processes: Grid+local)	Queueing
+ Finland	Aesyle (FGI)	72	0+35	0+0
	Alcyone (CMS)	892	156+312	1040+0
	Alcyone (FGI)	892	6+461	19+0
	Asterope (FGI)	192	84+0	10+1
	Celaeno (FGI)	448	172+0	9+0
	Electra (FGI)	672	0+478	0+0
	Jade (HIP)	768	227+541	25+49
	Maia (FGI)	768	360+408	14+0
	Merope (FGI)	1612	0+1319	14+0
	Pleione (FGI)	288	144+0	13+0
	Taygeta (FGI)	360	42+174	15+0
	Triton (FGI)	6972	182+0	2+0
	Usva (CSC/FGI/test)	144	12+0	0+0
<b>TOTAL</b>	<b>13 sites</b>	<b>14080</b>	<b>1385 + 3728</b>	<b>1161 + 50</b>

- In grid computing you can use several computing clusters to run your jobs
- Grids suits well for array job like tasks where you need to run a large amount of independent sub-jobs
- You can also use FGI to bring cluster computing to your local desktop
- FGI: 11 computing clusters, about **10 000** computing cores
- Software: Run Time Environment include applications from all fields, e.g., bioinformatics, chemistry, physics:
  - <https://confluence.csc.fi/display/fgi/Runtime+Environments>

# Using grid



- The jobs are submitted using the ARC middleware (<http://www.nordugrid.org/arc/>)
  - Using ARC resembles submitting batch jobs in Taito or Sisu
- ARC is installed in Hippu and Taito, but you can install it to your local machine too.
  - Setup command in Hippu:
    - `module load nordugrid-arc`
  - Basic ARC commands:

➤ <code>arcproxy</code>	(Set up grid proxy certificate for 12 h)
➤ <code>arcsub <i>job.xrsl</i></code>	(Submit job described in file <i>job.xrsl</i> )
➤ <code>arcstat -a</code>	(Show the status of all grid jobs)
➤ <code>arcget <i>job_id</i></code>	(Retrieve the results of a finished grid job)
➤ <code>arckill <i>job_id</i></code>	(kill the given grid job)
➤ <code>arcclean -a</code>	(remove job related data from the grid)

# Sample ARC job description file



```
&
(executable=runbwa.sh)
(jobname=bwa_1)
(stdout=std.out)
(stderr=std.err)
(gmlog=gridlog_1)
(walltime=24h)
(memory=8000)
(disk=4000)
(runtimeenvironment>="APPS/BIO/BWA_0.6.1")
(inputfiles=
( "query.fastq" "query.fastq" )
( "genome.fa" "genome.fa" )
)
(outputfiles=
( "output.sam" "output.sam" )
)
```

# Getting started with FGI-Grid



1. Apply for a grid certificate from TERENA ( a kind of grid passport)
2. Join the FGI VO (Access to the resources)
3. Install the certificate to Scientists' User Interface and Hippu.
4. Install ARC client to your local Mac or Linux machine for local use)
5. Instructions: *<http://research.csc.fi/fgi-preparatory-steps>*

Please ask help to get started: [helpdesk@csc.fi](mailto:helpdesk@csc.fi)

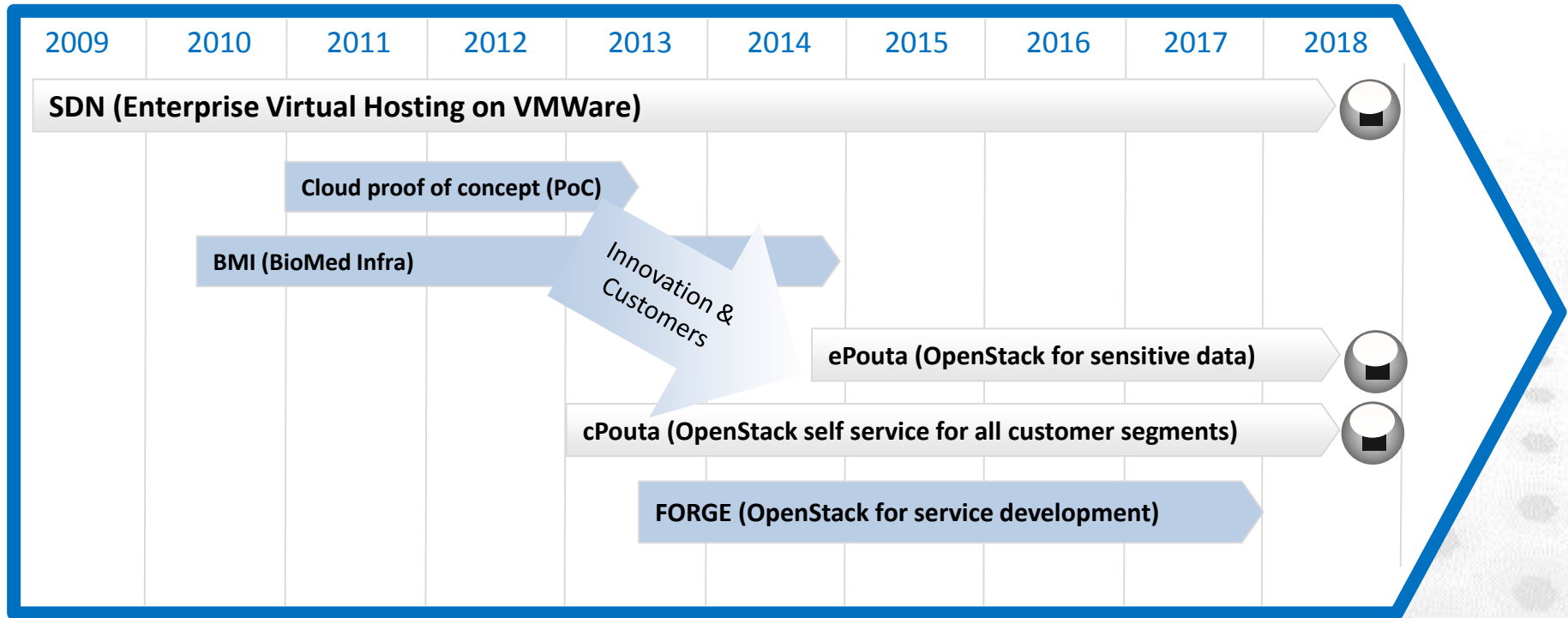
FGI user guide: <http://research.csc.fi/fgi-user-guide>

# Cloud computing: three service models

- **Software as a Service (SaaS)**
  - e.g. Chipster
- **Platform as a Service (PaaS)**
- **Infrastructure as a Service (IaaS)**
  - e.g. Pouta

# CSC IaaS Cloud Roadmap

Goal: Simplify and expand



Key:



What to do?



Decommission



Approved

Single customer segment restriction  
based on funding or service design

Platform not restricted to single  
customer segment (goal/policy decision)

# Cloud service development in 2014

➤ **Pouta** (virtualisation) = CSC cloud brand

<https://research.csc.fi/cloud-computing>

## ➤ **cPouta**

- In Production "Amazon-type" Pouta for Research Communities and Organisations
- Anyone can apply for access

## ➤ **ePouta**

- In development: Enterprise i.e. Biomedinfa –type virtual hosting to collaborate on organisational ICT capacities.
- Development focus: Security features to support biobank (secure) data handling



# Web interface



## Instances

+ Launch Instance

Terminate Instances

<input type="checkbox"/>	Instance Name	IP Address	Size	Keypair	Status	Task	Power State	Actions
<input type="checkbox"/>	oli_test3	192.168.1.19 86.50.168.20	medium   30GB RAM   10GB Disk	keypair-oli	Active	None	Running	Create Snapshot
<input type="checkbox"/>	kalletest	192.168.1.26 86.50.168.22	tiny   1GB RAM   1 VCPU   10GB Disk	keypair-harri	Active	None	Running	Create Snapshot More
<input type="checkbox"/>	lalves_test	192.168.1.26 86.50.168.22	tiny   1GB RAM   1 VCPU   10GB Disk	keypair-harri	Active	None	Running	Create Snapshot More
<input type="checkbox"/>	HarriPerformanceTests_1_4	192.168.1.26 86.50.168.22	tiny   1GB RAM   1 VCPU   10GB Disk	keypair-harri	Active	None	Running	Create Snapshot More
<input type="checkbox"/>	HarriPerformanceTests_1_3	192.168.1.26 86.50.168.22	tiny   1GB RAM   1 VCPU   10GB Disk	keypair-harri	Active	None	Running	Create Snapshot More

```
khappone@pikkulintu:~$ nova list
+-----+-----+-----+-----+-----+-----+-----+-----+
| ID                                         | Name                               | Status | Task State | Power State | Networks |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 781d4a2f-c21c-4dfd-8d58-87428e4c7502    | CT-IFTest1                         | ACTIVE | None       | Running     | CThomas Deployment=10.5.5.10, 86.50.168.30 |
| 7abbe103-c7f0-4db0-87a7-8758aa8c086a    | DS40-server                        | ACTIVE | None       | Running     | csc=192.168.1.32, 86.50.168.64 |
| 21e2f4f3-9c4b-4561-8a4e-2c4c62141237    | Jarin testijärjestelmä            | SUSPENDED | None       | Shutdown    | csc=192.168.1.34 |
| 0532b4d0-9ac6-4e8a-8637-4192f1039039    | PoutaMon                          | ACTIVE | None       | Running     | csc=192.168.1.33, 86.50.168.35 |
| b997c581-e047-4c17-acf4-ee73962f1f71    | lalvesFedCloudTest                | ACTIVE | None       | Running     | csc=192.168.1.2, 86.50.168.7 |
+-----+-----+-----+-----+-----+-----+-----+-----+
khappone@pikkulintu:~$
```

# Command line tools

<https://pouta.csc.fi:8777/v2/csc/servers/0532b4d0-9ac6-4e8a-8637-4192f1039039>  
<https://pouta.csc.fi:8777/v2/csc/flavors/1a0f1143-47b5-4e8a-abda-eba52ae3c5b9>  
<https://pouta.csc.fi:8777/v2/csc/images/>

# REST API

# cPouta – computing in the Cloud

- cPouta
  - Virtual machines on demand
  - Taito hardware
  - Dedicated resources (HPC focus)
- More freedom and flexibility
- More responsibility

# cPouta's use cases



- Enhanced security – isolated virtual machines
- Advanced users – able to manage servers
- Difficult workflows – can't run on Taito
- Complex software stacks
- Ready made virtual machine images
- Deploying tools with web interfaces
- "We need root access"

*If you can run on Taito – run on Taito  
If not – Pouta might be for you*

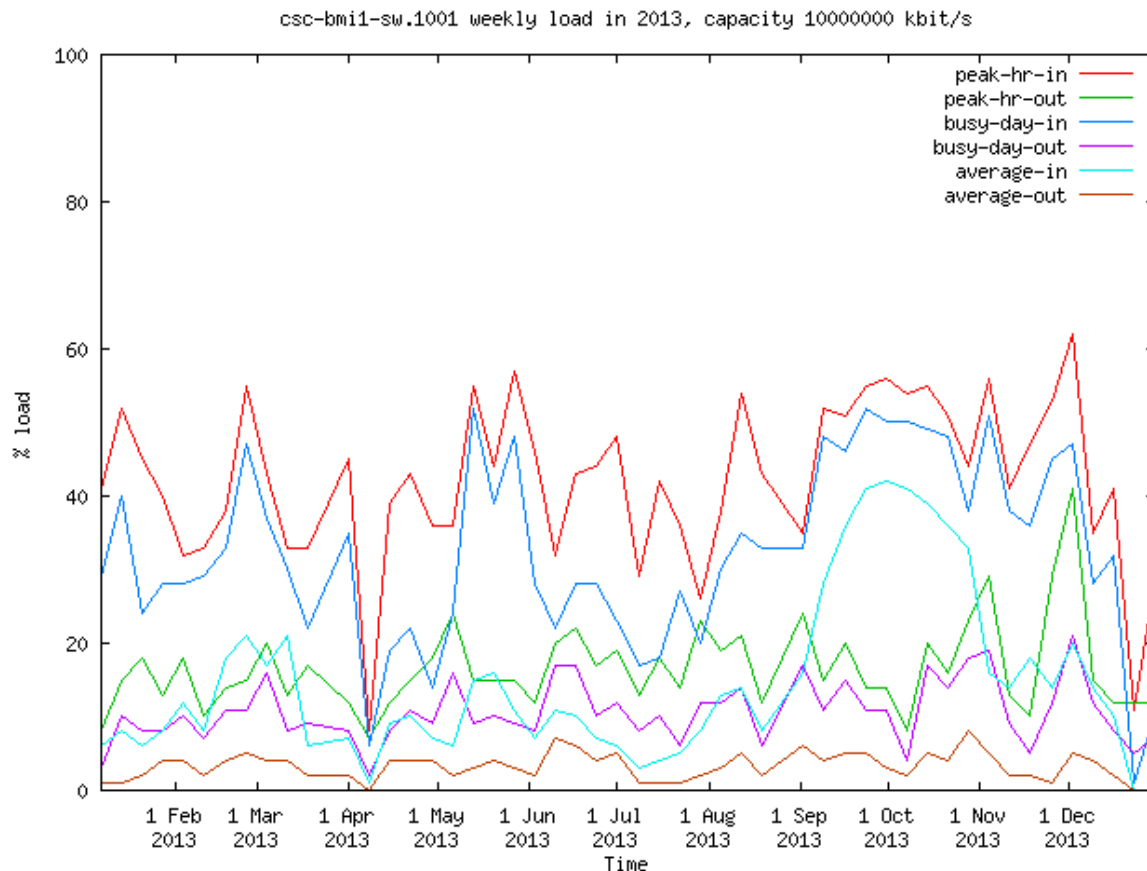
- Pouta user guide: <https://research.csc.fi/pouta-user-guide>

# ePouta

- Renewing the cloud cluster equipment in Espoo in 2015
  - Preparing for the pilot phase
  - Common CSC Open Stack configuration
  - Focus on secure computing and service for organisations
  - Idea: seamless scaling of local resources using a trusted compute center (in Finland)
  - Requires local IT admin contact
  - Funding model and resource allocation policy is still under debate, supported by ELIXIR Finland
  - Vahti L3 (security standard) targetted



# CSC – Meilahti genomics laaS data traffic 2013



**5.8 PB in**  
**1.4 PB out**

Avg. 221 MB/s 24  
hours a day all year  
round

# CSC Cloud pricing – basic package 12 months

CSC

- Includes:
  - 10 000 core-hours
  - 0,5 TB disc space
  - User accounts
  - CSC Service Desk
- Academic: 686 € (alv 0%)
- Commercial: 1005 € (alv 0%)
- Amount of core hours can be increased
- Other services can be added to basic package

**More information: [servicedesk@csc.fi](mailto:servicedesk@csc.fi)**

# Summary of cloud resource @ CSC

- Pouta cloud services in production

<https://research.csc.fi/cloud-computing>

- Support researchers ("Communities") and of organisations ("Enterprise") to integrate to virtualised capacities on the CSC cloud platform
- Further information: [servicedesk@csc.fi](mailto:servicedesk@csc.fi)

# Taito-shell replaces Hippu

- Interactive session on a Taito compute node
  - E.g. run a GUI, run long non-intensive jobs, etc.
- Two 256GB nodes allocated, easy to expand
  - Maximum of 4 cores/128GB per user, no time limit
- Access: **ssh -X taito-shell.csc.fi**
  - Also via drop down menu from nxkajaani
  - Technically a slurm job without dedicated resources
  - Processes killed when logged out
  - Can be left running via **screen** (on Taito) or via nxkajaani (exit with suspend)
- Feedback welcome!
- <https://research.csc.fi/taito-shell-user-guide>

# Courses at CSC



## ➤ **Taito** Phase 2 workshop

– *April 2015 in week 17*

## ➤ **CSC courses:** <http://www.csc.fi/courses>

- Introduction to Linux and Using CSC Environment Efficiently
- Spring School in Computational Chemistry: 10.-13.3.2015
- Pouta training 23.3.2015
- CSC HPC Summer School
- Spring, Autumn, Winter Schools
- Parallel Programming
- Some courses have possibility for remote participation
- Course materials often available from event website for self study

# CSC Phase2 resources' summary



- ***Sisu* supercomputer**
  - General availability since **9.9.2014**
- ***Taito* supercluster**
  - General availability since **19.1.2015**
  - Part of Taito used for *Pouta Cloud*
  - *taito-shell* replacing *Hippu* service
- ***Bull* system**
  - General availability since **1.10.2014**
  - *45 nodes with 2 Intel Xeon Phi coprocessors each*
  - *38 nodes with 2 NVIDIA Tesla K40 accelerators each*
- ***DDN* HPC storage system**
  - *Totaling 4 PB of fast parallel storage*



# Open science and research



- Ministry of Education and Culture initiative for the promotion of information availability and open science
- Goal to make Finland a leading country in openness of science and research
- Scope of the initiative includes publications, research data, and methods
- The aim is to provide researchers with practical knowledge in how they as individuals can implement open science
- Several services offered by the Ministry to researchers

# Services for storing and sharing your data



- Services offered by MINEDU
- Services free of charge to end-users
- Services provided by various partners
- Developed under direction from the initiative
- Recommended by e.g. by the Academy of Finland
- Existing services developed continuously and new services piloted within the initiative

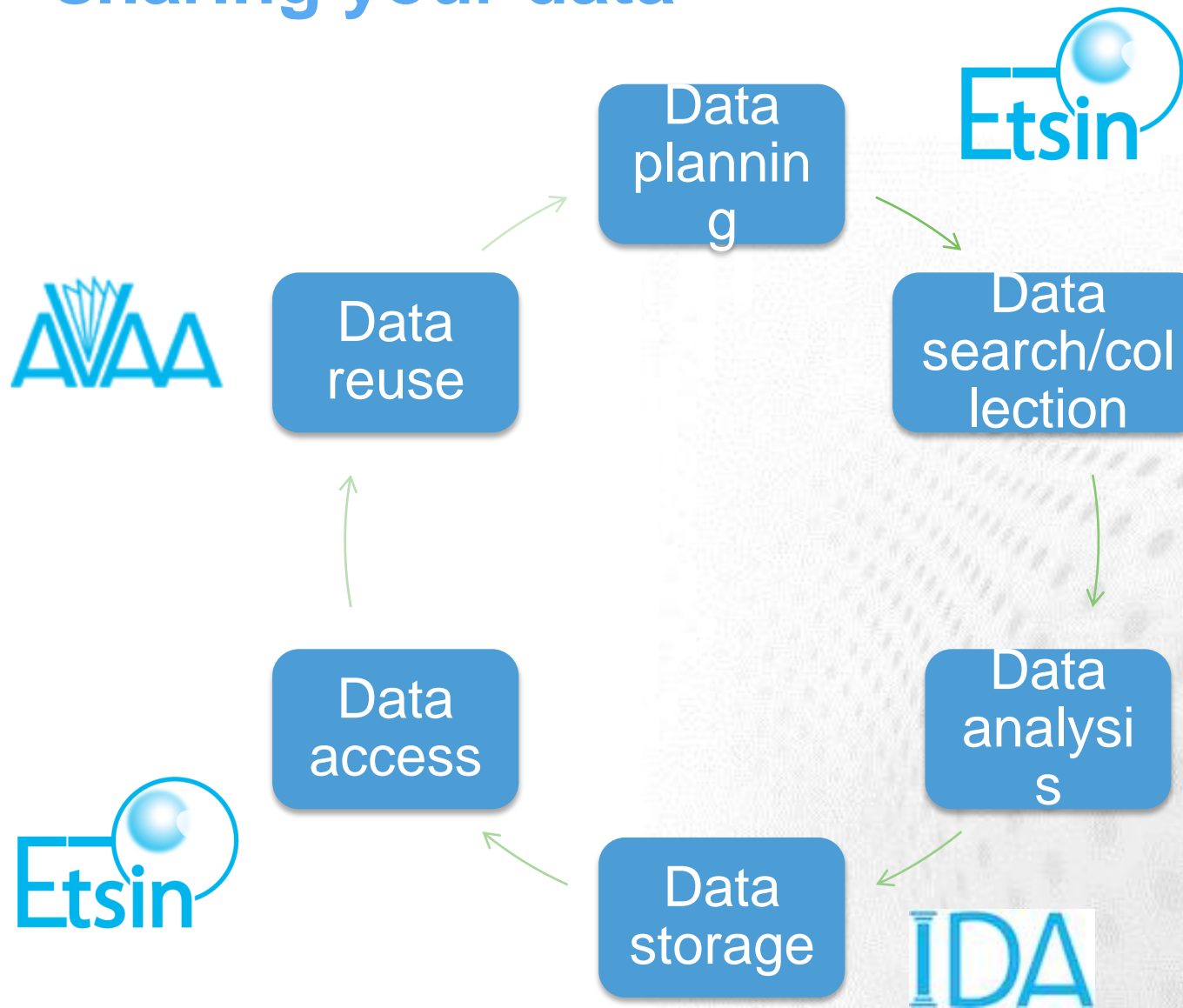
# Services for storing and sharing your data



1. Store data: IDA research data storage service
2. Make data discoverable: Etsin research data finder
3. Share open data: AVAA open data publishing platform

More services and info on [openscience.fi/services](https://openscience.fi/services)

# Services for storing and sharing your data



# IDA research data storage



- iRODS-based storage system for storing research data
- Designed for stable data sets, not for data analysis
  - Lower speed but high safety and availability
- Storage quotas are granted by the Finnish Academy and universities based on applications
  - SA standard quota 2TB per project
  - Aalto total quota 480 TB, contact person Anne Sunikka, Anne Sunikka @ aalto.fi
- Linked to other ATT services:
  - Metadata may be stored in Etsin and data published through AVAA
- More information and application: [tdata.fi/ida](https://data.fi/ida)

# IDA research data storage



## .IDA

- Part of ATT services
- Quotas granted by universities and SA
- Several interfaces (WWW/SUI, network disk, i-commands)
- Internet accessible
- Project based structure
- Flexible sharing within project and to the public

## .HPC archive

- Part of CSC computing environment
- 2 TB default quotas
- Usage with i-commands
- Visible only to CSC environment
- Personal storage area

# IDA research data storage



`.iput file`

`.iget file`

`.ils`

`.icd dir`

`.irm file`

`.imv file file`

`.imeta command`

`.irsync`

`.imkdir`

`.iinit`

move file to IDA

retrieve file from IDA

list the current IDA directory

change the IDA directory

remove file from IDA

move file inside IDA







view and edit metadata



synchronize the local copy with the copy in IDA


create a directory to IDA


initialize your IDA account


# IDA research data storage


- ▶  sisu
- ▶  taito
- ▼  ida
  - ▼  ida
    - ▼  csc
      - ▼  projectx
        - ▼  **foo**
          - ▶  published
- ▶  hpc archive


Name	Type
 data1	application/octet-stream
 data2	application/octet-stream


 Open


 Edit


 Download


 Pack


 Unpack


 Copy

 Cut

 Clear Selection

 Select All

 Delete

 Rename

# Etsin research data finder



- For documenting a dataset as a research output
  - Connects data to a funding project
  - Issues a permanent URN identifier to dataset for citation purposes
  - GUI and REST API
- For sharing data in a controlled way
  - Enables the creation of an application form using Haka authentication
  - Enables linking to open data in IDA and anywhere on the web
- More information: [etsin.avointiede.fi](https://etsin.avointiede.fi)



<https://etsin.avointiede.fi>  
<http://etsin.avointiede.fi/api/3>

[Help](#)[FAQ](#)[Stats](#)[Log in](#)[SEARCH DATASETS](#)[ADD DATASET](#)[ORGANIZATIONS](#)[ABOUT ETSIN](#)[English](#)

## What is Etsin?

Etsin enables you to find research datasets and to make your own datasets discoverable to others.

Etsin offers access to datasets in various fields via a joint metadata model. The descriptive metadata stored in the service includes information on the authors, subject, format and licensing of the data.

Etsin works independently of actual data storage location.

## Is this service for me?

Anyone can use Etsin to search for research datasets. Authenticated users may enter information on their own datasets, and of others' datasets with their permission.

Etsin aims to collect information on datasets from various sources and from various fields. You do not have to be a professional scientist or researcher in order use the service.

## Is all of the data open?

The published metadata on the dataset is open to everyone, but the data owner decides how the data can be accessed and by whom. Nevertheless, sharing your data openly is recommended. You will find more information on the [About Etsin](#) page.

### Search data



### Newest datasets

#### New age -lehti Ultran lukijoiden henkisyys ja henkisyyden harjoittaminen 2011

Tutkimuksen tarkoituksena oli selvittää...

Metadata completeness score: 3/5

[XML](#) [HTML](#)

#### Lasten turvallisuus 2011

Aineisto käsittelee alle 13-vuotiaiden lasten...

Metadata completeness score: 3/5

[XML](#) [HTML](#)

# AVAA open data publishing platform



- Web portal for sharing open research data and related applications, interfaces etc.
  - Suggest your data for publication by contacting CSC servicedesk
- Enables open data downloads from IDA based on user-defined metadata settings
  - URL syntax:  
http://avaa.tdata.fi/openida/dl.jsp?pid=<Identifier.series of the IDA file> e.g.  
<http://avaa.tdata.fi/openida/dl.jsp?pid=urn:nbn:fi:cs-c-ida2015011501152s>
- More information: [avointiede.fi/avaa](http://avointiede.fi/avaa)

## Variables:

Hyytiälä SMEAR II

Meteorology

Gas

Radiation

Soil

Flux

Kumpula SMEAR III

Meteorology

Radiation

Gas

Flux

Värriö SMEAR I

Gas

Radiation

Meteorology

Soil

Tree

Helsinki Hotel Tornii

Flux

From:

2014-09-10

To:

2014-09-11

Shift:

&lt;&lt; Day &gt;&gt;

Make Query

Quality Level:

Any

Averaging:

None

Averaging Type:

None

Arrival Height:

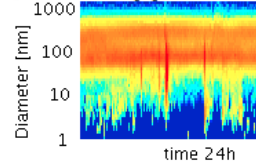
100m

Reload main view

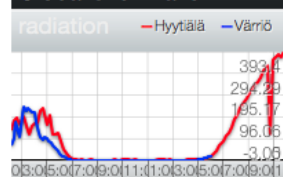
## Temperature 15-16m

CO<sub>2</sub>

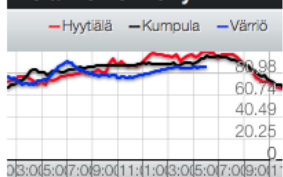
## Size Distribution [m]



## Global shortwave radiation



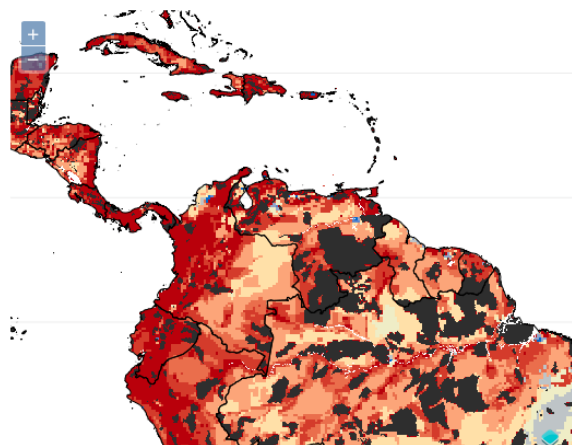
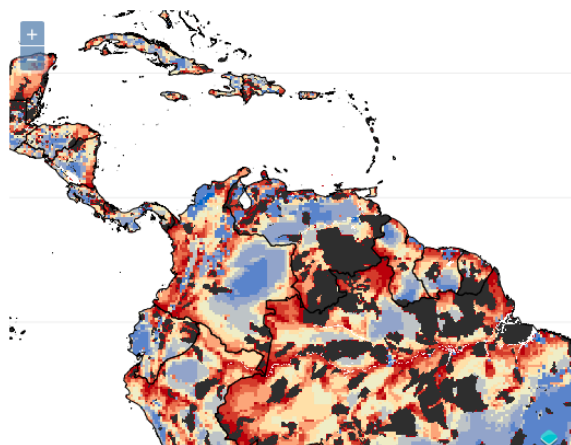
## Relative humidity



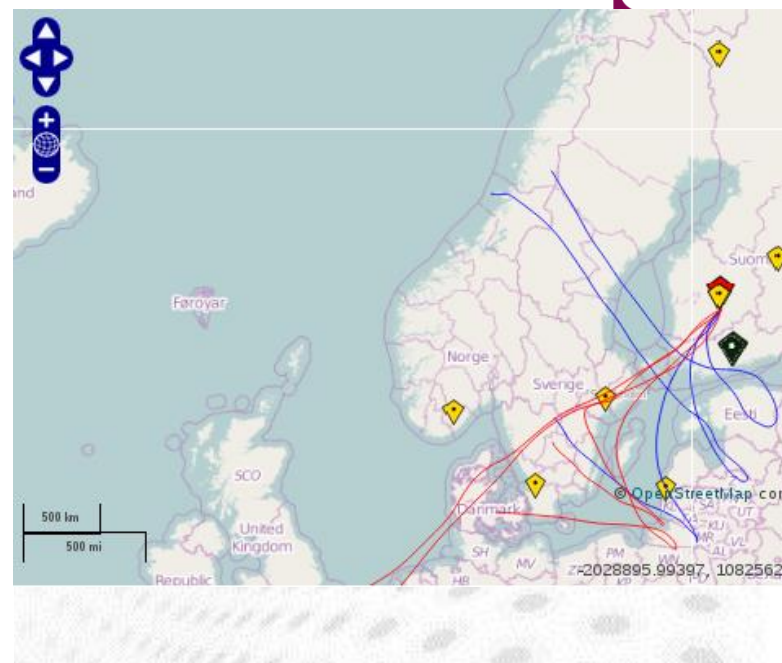
## Ozone concentration



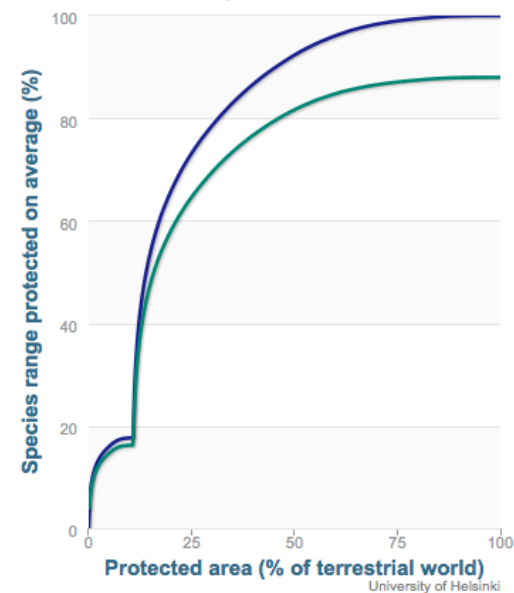
Global present Global future (2040) National present National future (2040) Global present Global future (2040) National present National future (2040)

[Click here to show/hide priority maps](#)

Current PA's Expansion to 17 % 17-25 25-30 30-40 40-50 50-60 60-70 70-80 80-90 90-100 %



## Overall performance



University of Helsinki

# More information



- Detailed information on IDA and interoperability with Etsin & AVAA in IDA FAQ on [tdata.fi/ida](https://tdata.fi/ida)
- Information on all Open science and research initiative's services in the service pages at [openscience.fi/services](https://openscience.fi/services)
- Questions and feedback on IDA, Etsin, AVAA to CSC servicedesk at [servicedesk@csc.fi](mailto:servicedesk@csc.fi)



## Feedback form and Round robin

# Feedback form

- <https://www.webropolsurveys.com/S/57665DDA29516729.par>

- (link also on the seminar home page [www.csc.fi](http://www.csc.fi) →)

- *What are your needs for your research?*
  - How CSC can help?
  - Special libraries/tools?
- How much data you produce that needs processing?
- Courses/training?
- Queue length: 3 (Sisu) / 14 (Taito) days enough?
  - Codes that can't checkpoint?
- Is memory an issue for you?
  - 1.5 TB/nodes usage policy?
- Do you need to move a lot of files? (from where?)
- Interested in GPGPU/Intel Xeon Phi? Which code?

# Participants



- Armi Tiihonen
- Eric Dorval
- Shahab Basiri
- Karmen Dykstra
- Cristian Vasile Achim
- Edward Mutafulungwa
- Sebastian Prepelita
- Enrico Glerean
- Asier Zugarramurdi
- Darko Hric
- Mikko Ervasti
- Patrick Rinke
- Pertti Aarnio
- Simo Tuomisto

- Johan West
- Richard Darst
- Ming Yin
- Riku Linna
- Alexander Karpenko
- Kiran Garimella
- Virginia Estevez
- Kari Nousiainen
- Paavo Niskala
- Jussi Nieminen
- Joonas Piili
- Jaeyoung Kwak
- Juuso Lindgren
- Simppa Akaslompolo
- Arafatul Hasan