CSC computing resources

Pekka Lehtovuori
CSC – IT Center for Science Ltd.

Program

- 10-11 CSC presentation
- 11-11:30 Round robin
- 11:30-> F2F meetings

Outline

- CSC at glance

- CSC supercomputers Phase 2
  - *Sisu* (Cray XC30)
  - *Taito* (HP cluster)

- Other resources available for researchers
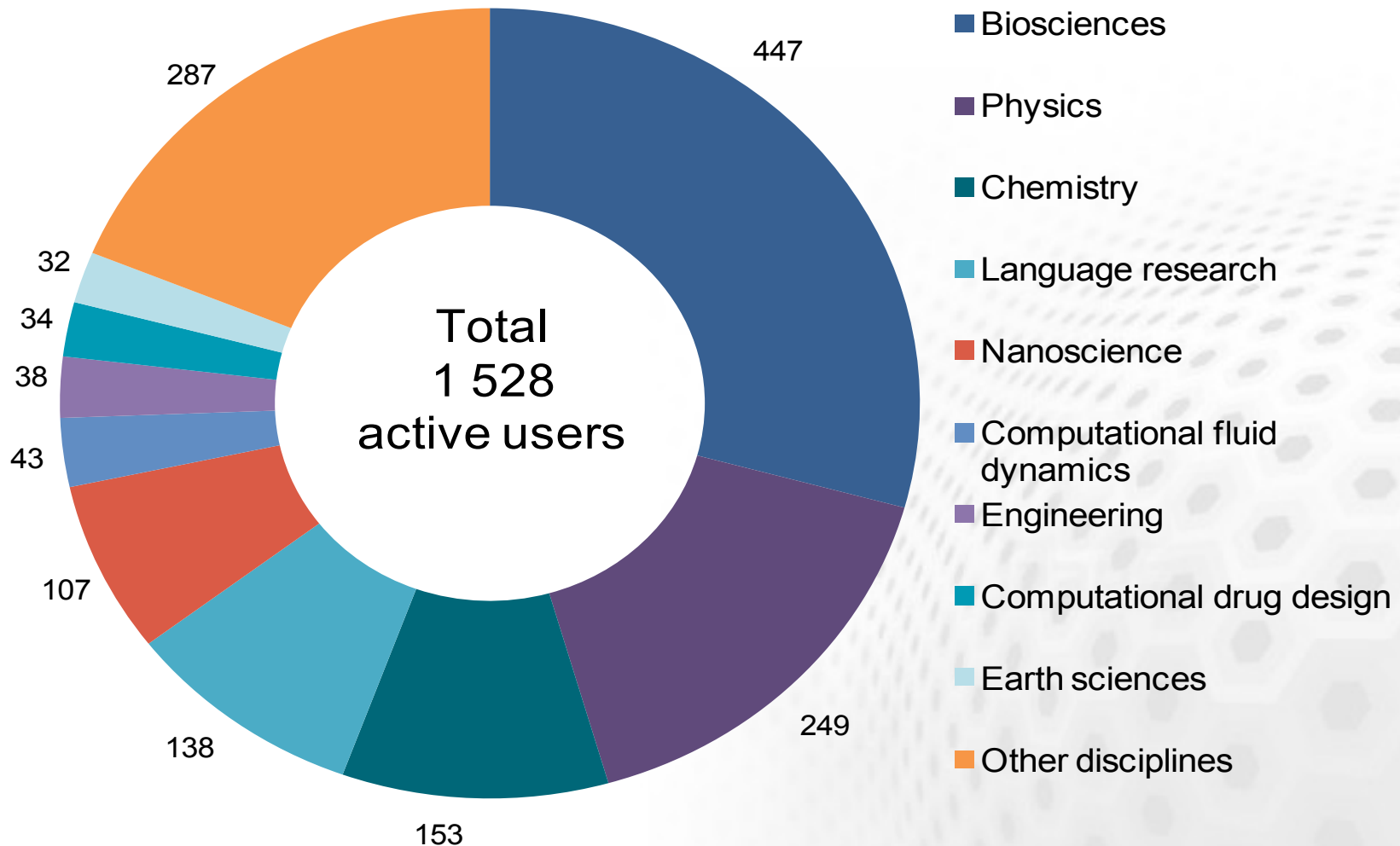
CSC at glance

- Founded in 1971
- Operates on a *non-profit* principle
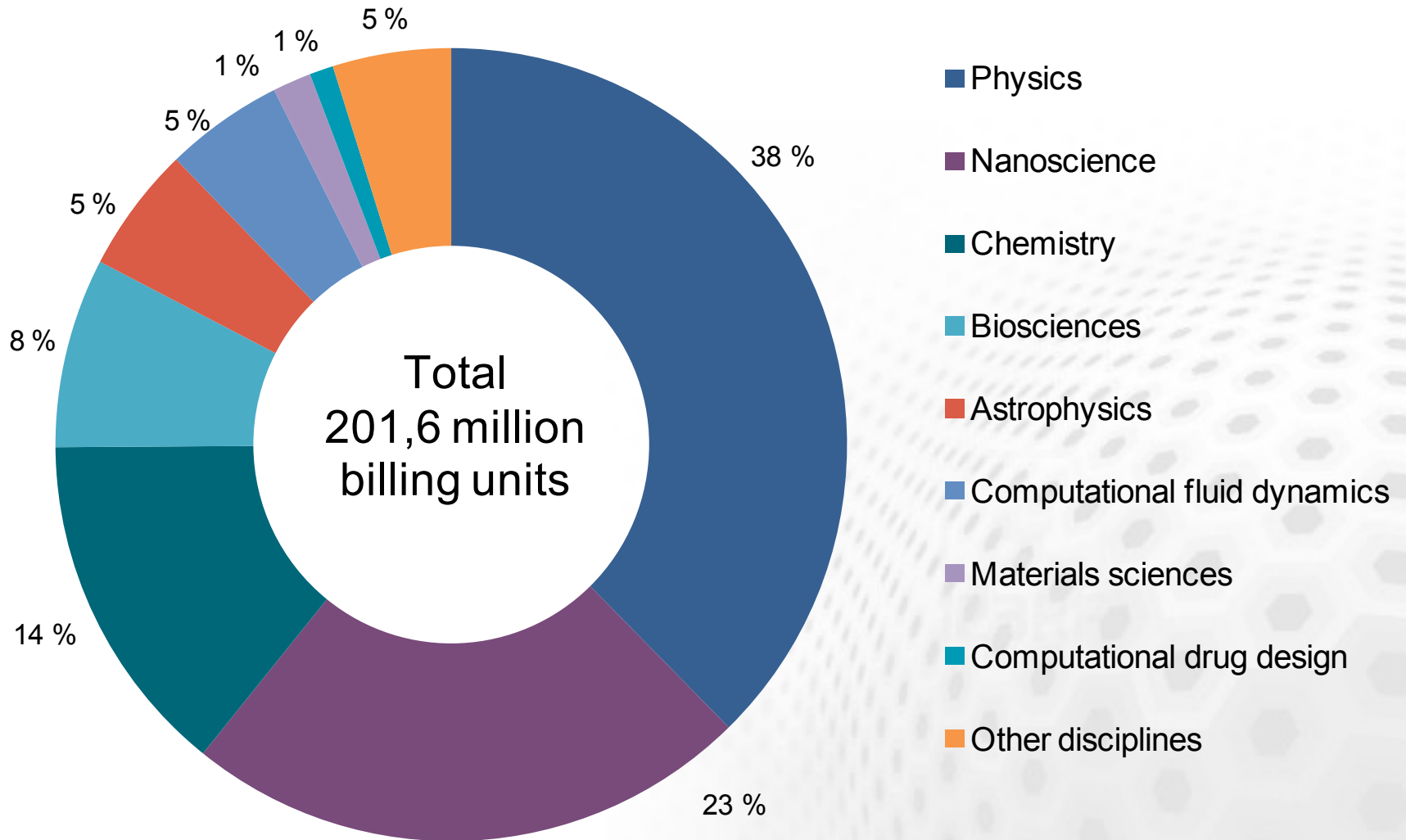- Facilities in Espoo and Kajaani
- Staff ~255 people

Users of computing resources by discipline 2013

Total
1 528
active users
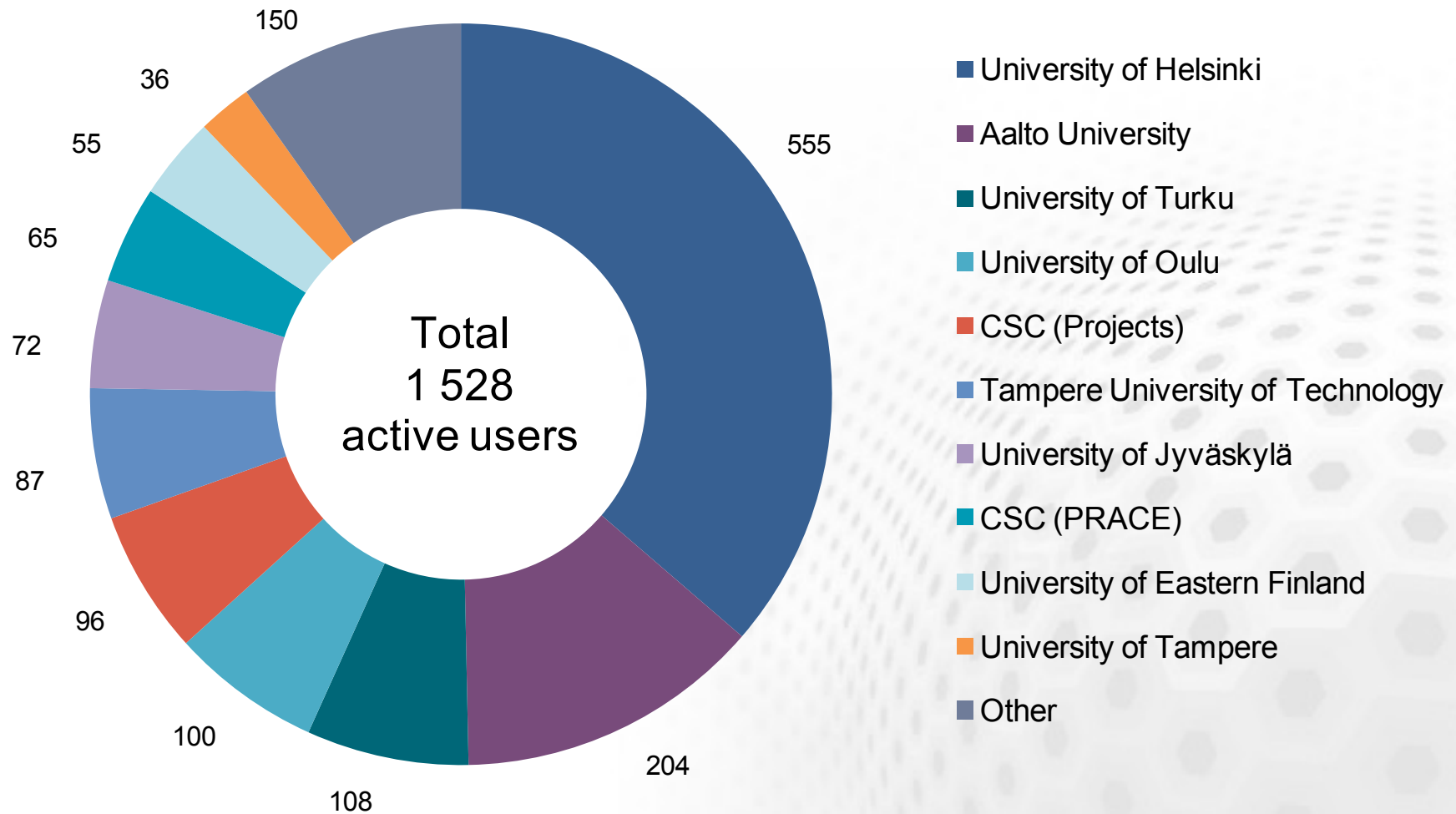
- 447
- 249
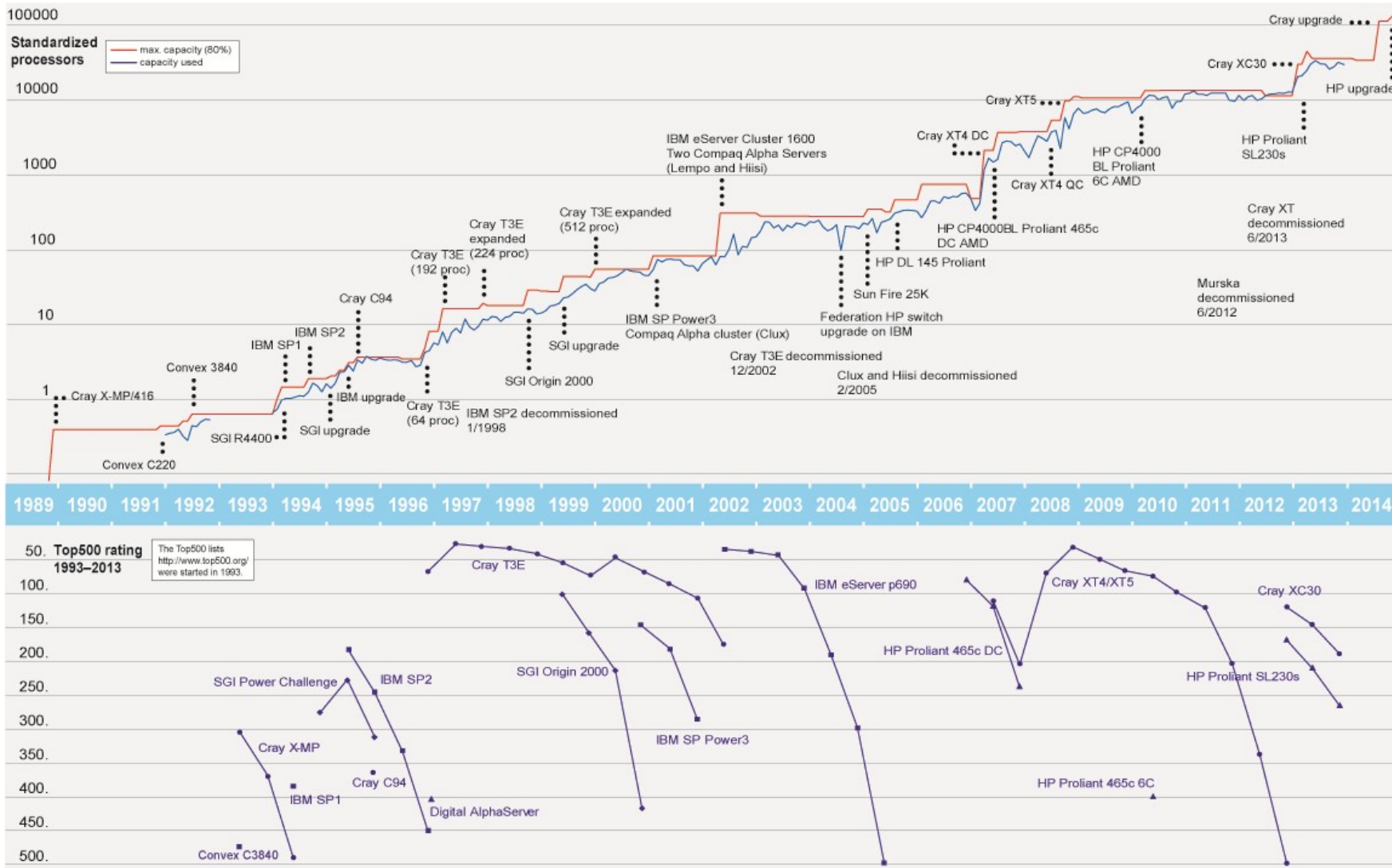- 153
- 138
- 107
- 43
- 38
- 34
- 32
- 287

Legend:
- Biosciences
- Physics
- Chemistry
- Language research
- Nanoscience
- Computational fluid dynamics
- Engineering
- Computational drug design
- Earth sciences
- Other disciplines

Computing usage by discipline 2013

Total
201,6 million
billing units

- 38 % — Physics
- 23 % — Nanoscience
- 14 % — Chemistry
- 8 % — Biosciences
- 5 % — Astrophysics
- 5 % — Computational fluid dynamics
- 1 % — Materials sciences
- 1 % — Computational drug design
- 5 % — Other disciplines

Users of computing resources by organization 2013

Total
1 528
active users

- University of Helsinki — 555
- Aalto University — 204
- University of Turku — 108
- University of Oulu — 100
- CSC (Projects) — 96
- Tampere University of Technology — 87
- University of Jyväskylä — 72
- CSC (PRACE) — 65
- University of Eastern Finland — 55
- University of Tampere — 36
- Other — 150

CSC Computing Capacity 1989–2014

PHASE 2 RESOURCES

- SISU
- TAITO
- DDN (PHASE 3)
- BULL

Sisu: Cray Supercomputer

- Future Intel® Xeon® processor E5-2600 v3 product family
- Cray Aries Interconnect
- ~ 40 000 cores
- 64 GB memory per node

How to prepare?

- Old binaries may run off-hand, **CSC advises to recompile the code**
- OS upgrade in June (login nodes)
  - Major upgrade (software on the same level as on Sisu Phase 2)
  - Anything running on *login nodes* needs to be recompiled
- Porting strategy
  - Under preparation

Sisu

- AVX-2
  - May need to optimize for wider vectors' size

- DDR4
  - Higher bandwidth, lower power consumption

- Max job size likely to increase

- Native SLURM on the way, unlikely to be available after Sisu July hardware update

Sisu (un)availability in summer 2014

- June 2014 software upgrade break
  - Probably 3 days of downtime

- July 2014 hardware upgrade break
  - At least 2 weeks downtime expected

Taito: HP Supercluster

- Intel® Xeon® processor E5-2600 v2 product family & Future Intel® Xeon® processor E5-2600 v3 family
- FDR InfiniBand interconnect
- ~ 17 000 cores
- Different memory per node sizes: 64, 128, 256 GB and 1.5 TB

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 new architecture compute nodes

- 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
  - May run, CSC **recommends re-compiling**
  - 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 may change
  - Memory changes
  - Instructions will be available
- How to submit jobs to either architecture only
  - Specify to which partitions you send your jobs

SLURM configuration: Fair usage

- 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)

- *Due to abuse, a maximum limit of jobs in queue now enforced*

- Chain jobs (--dependency –flag for SLURM) if you need long running time

- 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-core=something)
  - If you see abuse or think that the setup is unfair, contact helpdesk@csc.fi

- SUI has a monitoring tool for your jobs and used resources (*Services -> eServices -> My Project*)

Taito (un)availability in summer 2014

- June 2014 software upgrade break
  - Probably 3 days of downtime

Current Plan for Phase2 Sisu and Taito

- Sisu: planned installation in July-August 2014
  - General availability planned for Q3 2014

- Taito: planned installation in Q4 2014

How to prepare?

- Porting strategy
  - Not much to do at this stage
  - Compilers, libraries, flags, …
  - Preliminary performance data?
  - Add **AVX-2 flag** when compiling your code

Bull

– In pilot/project until end of August 2014

– No guarantee on availability

– 38 NVidia K40 nodes (76 gpus)

- 12 GB memory per card

– 45 Xeon Phi nodes (90 Xeon Phis)

- 16 GB memory per card

– Energy efficient CPU's

How to access (plan)

- **Accessing the resources**
  - **Xeon Phi: ssh taito-mic.csc.fi (TBC)**
  - **Nvidia K40: ssh taito-gpu.csc.fi (TBC)**

Pettu Phase 3

- System size will increase to ~4 PB
  - About 1.9 PB will added to the current configuration
  - Aggregate bandwidth > 80 GB/s (currently ~48 GB/s)
- Available together with Phase2 supercomputers
- Downtime on all systems (~1 day)

Disks in total

- *4.0 PB on DDN*
  - New $HOME directory (on Lustre)
  - $WRKDIR (*not backed up*), soft quota 5 TB / user
- *HPC Archive*
  - 2 TB / user, common between Cray and HP
- *3 PB disk space through TTA/IDA*
  - 1 PB for Univerisities
  - 1 PB for Finnish Academy (SA)
  - 1 PB to be shared between SA and ESFRI
  - more could be requested

- */tmp* (around 1.8 TB) to be used for *compiling codes*

Grid computing with Finnish Grid Infrastructutre (FGI)

ARC Grid Monitor

2012-06-06 CEST 07:43:21

Processes: ■ Grid ■ Local

Country	Site	CPUs	Load (processes: Grid+local)	Queueing
+ Finland	Aesyle (FGI)	72	48+0	77+0
	Alcyone (FGI)	892	0+313	0+0
	Asterope (FGI)	96	0+0	0+0
	Celaeno (FGI)	192	0+133	0+0
	CSC Vuori cluster	3640	0+2565	1+0
	Electra (FGI)	672	0+648	0+0
	Jade	768	600+32	1394+1
	Korundi (UH)	400	0+115	2+239
	Maia (FGI)	768	0+168	0+9
	Merope (FGI)	604	91+143	45+-1
	Pleione (FGI)	240	4+216	35+0
	Taygeta (FGI)	360	215+112	46+3
	Triton (FGI)	2820	173+1202	0+0
	Usva (CSC/FGI/test)	144	108+0	47+0
TOTAL	*14 sites*	*11668*	*1239 + 5647*	*1647 + 251*

ALL

Clusters

Job scheduler
(e.g. SLURM)

Send job (sbatch, qsub...)

User X: Job 1
User X: Job 2
User Y: Job 3
User Z: Job 4

User X

Frontend

Network

Storage

Compute node 1-n

Grids

Storage

Data

User X

Work computer

Grid tools

Send a job

Data

Data

Data

Send a job

Send a job

Grid interface

Lappeenranta cluster

Grid interface

Helsinki cluster

Grid interface

CSC cluster
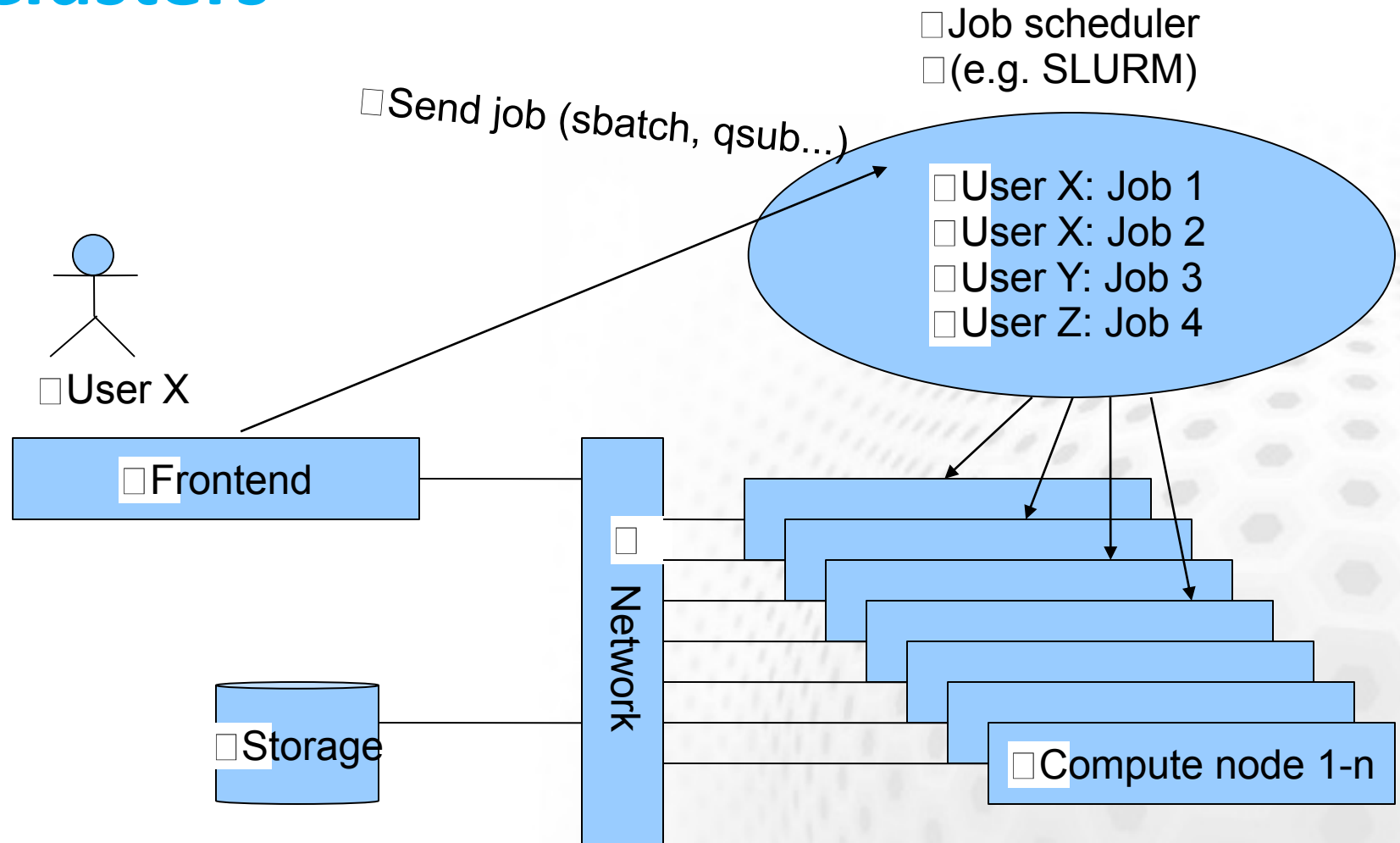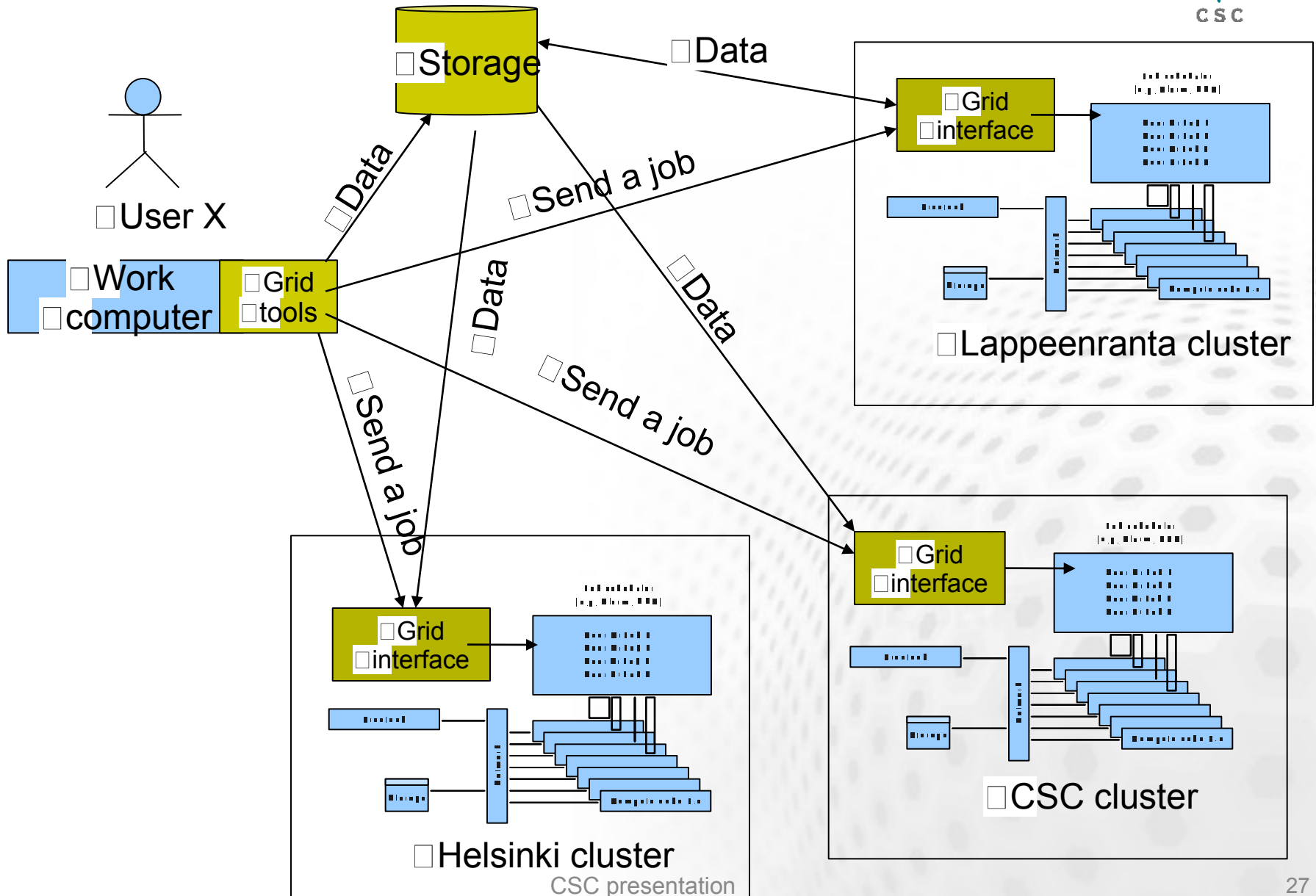
Getting started with FGI-Grid

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

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

Pouta – Computing in the Cloud

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

Web interface

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

Pouta audience

- 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
- "no I really 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*

IDA storage service

- Intended for storing research data, the ultimate goal being to facilitate the exploitation of electronic data in research.

- Secure and user-friendly storage service for data and the associated metadata.

- The integrity of the data to be stored is secured by managing copies and their integrity.

Who can use IDA?

- The IDA service is offered by the Finnish ministry of education and culture, to Finnish universities, universities of applied science, and certain projects of the Academy of Finland.
- Using IDA is free of charge for end-users.
- Storage capacity in total about 3 PB

University Quotas

University	Quota	Quota in TB
Aalto University	0,12	160
University of Helsinki	0,27	420
University of Eastern Finland	0,08	80
University of Jyväskylä	0,07	70
Finnish Academy of Fine Arts	0	
University of Lapland	0,01	10
Lappeenranta University of Technology	0,03	30
University of Oulu	0,09	120
Sibelius Academy	0	
Hanken School of Economics	0,01	10
Tampere University of Technology	0,06	60
University of Tampere	0,06	60
Theatre Academy	0	
University of Turku	0,14	140
University of Vaasa	0,01	10
Åbo Akademi University	0,04	50

Universities for Applied Science total 10 TB

Becoming an IDA user

- Universities: Please contact your local IDA contact person (http://www.tdata.fi/en/idan-kayttajaksi)

- Universities of applied science: Please contact contact@csc.fi

- Academy of Finland: please contact contact@csc.fi

IDA additional quota

- Intended for projects requiring large capacity, e.g. ESFRI projects and projects funded by the Academy of Finland.

- The Ministry of Culture and Education decides on the allocation of this quota.

- 1 PB is reserved for this.

- Applications twice a year

- More information: www.tdata.fi/en/ida

Grand Challenges

- Normal GC *(in half a year / year)*
  - new CSC resources available for a year
  - no bottom limit for number of cores
- Special GC call (mainly for Cray) *(based on your needs)*
  - possibility for short (day or less) runs with the whole Cray
  - Deadline: *May 30th, 2014, at 12:00*
- Remember also PRACE/DECI

CSC Phase2 resources' summary

- ***Sisu*** *supercomputer*
  - Installation planned in *July-August 2014*
  - General availability planned for **Q3 2014**
- ***Taito*** *supercluster*
  - Installation planned in *Q4 2014*
- ***Bull*** system
  - General availability planned for **Q3 2014**
  - *45 nodes* with *2 Intel Xeon Phi coprocessors* each
  - *38 nodes* with *2 NVIDIA Tesla K40 accelerators* each
- ***DDN*** HPC storage system
  - Added *1.9 PB*, in *Q3 2014* totaling *4 PB of fast parallel storage*
  - Supports Cray and HP systems, aggregate bandwidth > *80 GB/s*

Round robin

Tomasz Malkiewicz
CSC – IT Center for Science Ltd.

Round robin

- What are your research interest?
  - How CSC can help?
  - Special libraries/tools?
- Queue length: 3 days enough?
  - Codes that can't checkpoint?
- Is memory an issue for you?
  - 1.5 TB/nodes usage policy?
- Applying for Grand Challenge?
  - Special Grand Challenge?
- Need to move a lot of files? (from where?)
- Interested in GPGPU/MICs? Which code?

Feedback form

- *https://www.webropolsurveys.com/S/9BE41A717A898DE8.par*