CSC computing resources

Jussi Heikonen, Jarno Laitinen, Tomasz Malkiewicz
CSC – IT Center for Science Ltd.

Program

- 14-15 CSC presentation
- 15-15:30 Round robin
- 15:30-> Free discussion / F2F meetings

CSC presentation

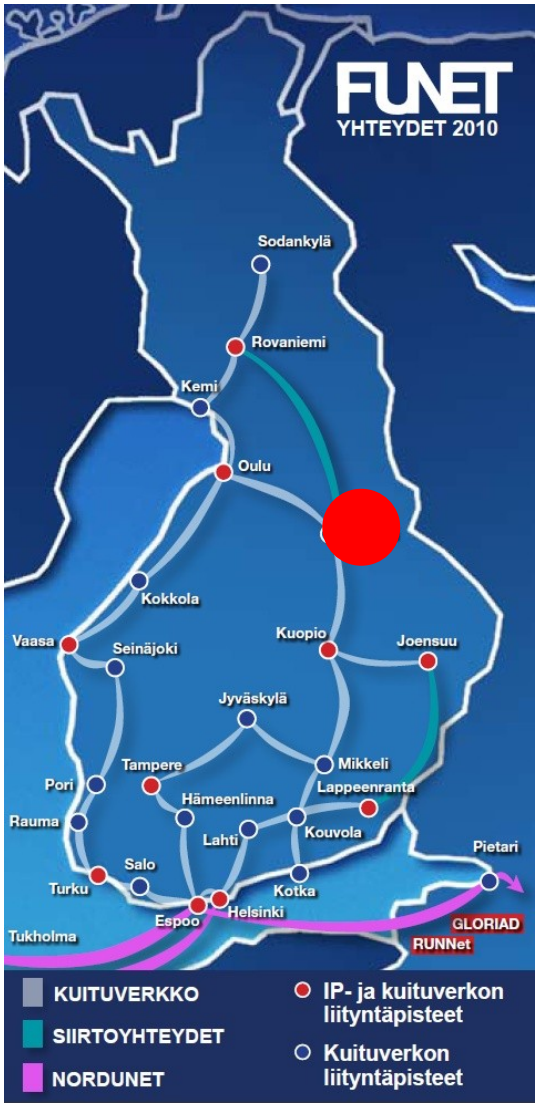
Outline

- CSC at a 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
- Staff ~255 people
- Facilities in Espoo and Kajaani
- Free of charge services for higher education institutions in Finland

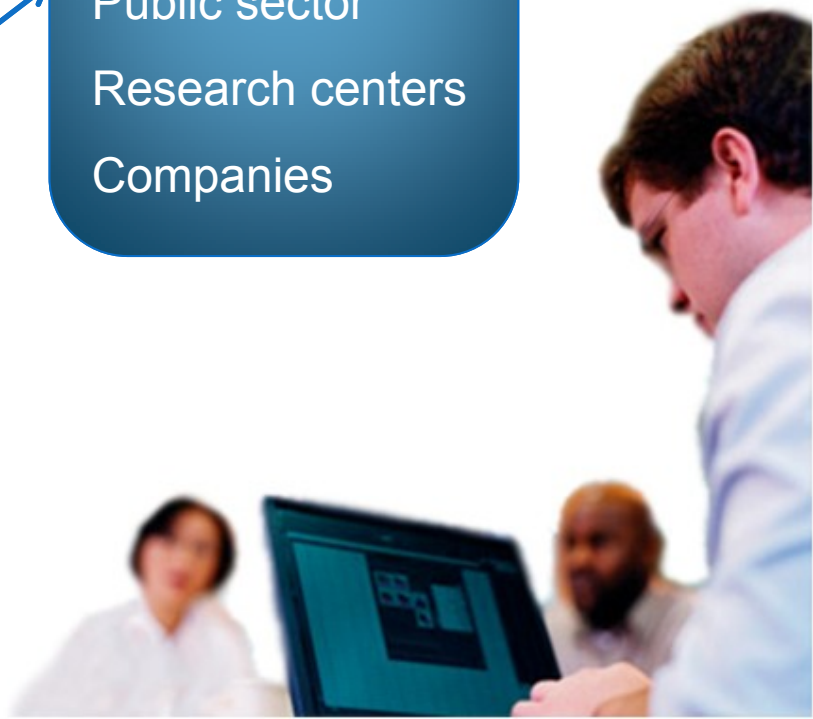
Datacenter CSC Kajaani



CSC's Services

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

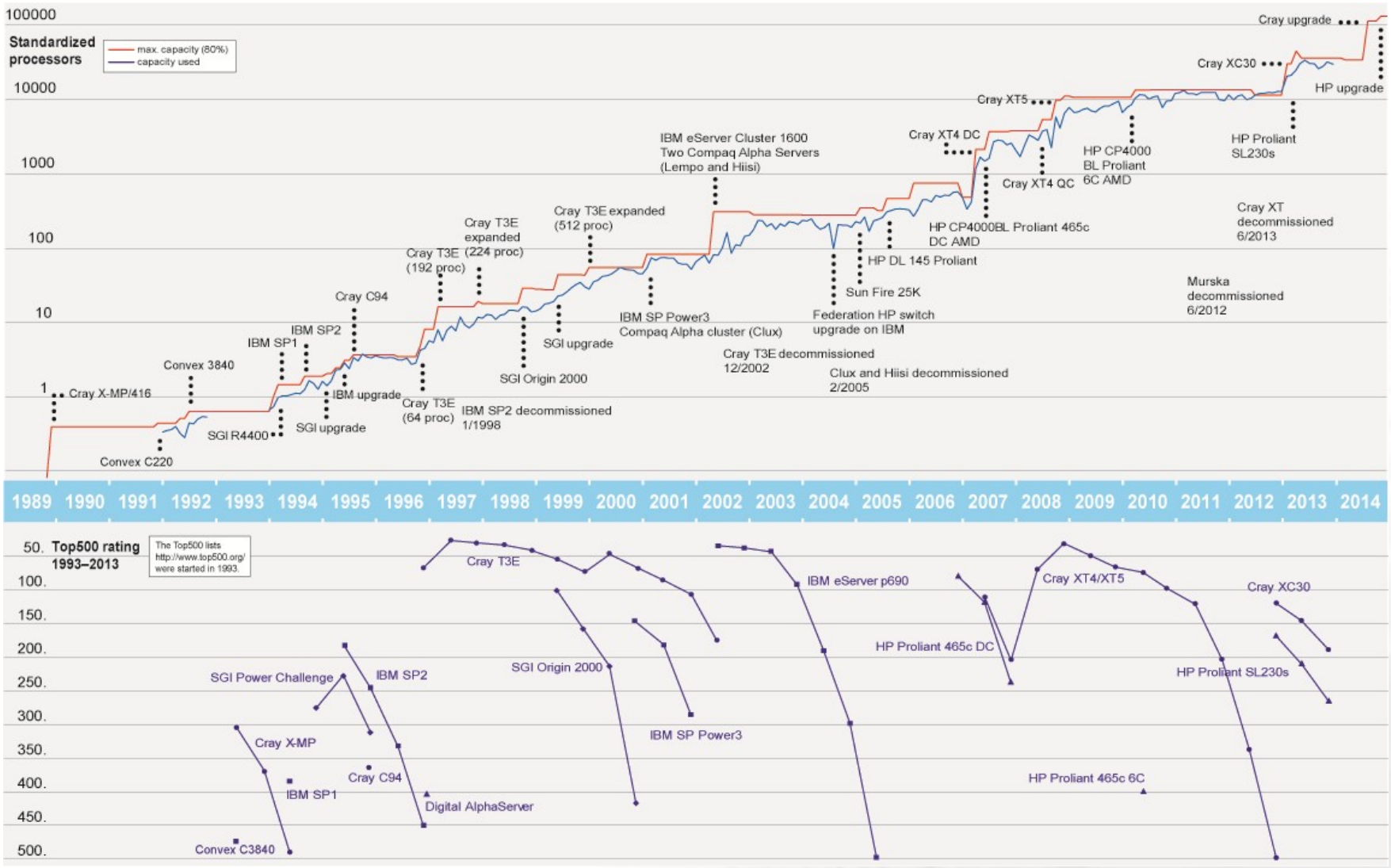
Universities
Polytechnics
Ministries
Public sector
Research centers
Companies



Users

- 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

CSC Computing Capacity 1989–2014



HPC 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 *now* in progress (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 upgrades

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

Taito: HP Supercluster

- Intel® Xeon® processor E5-2600 v2 product family & Future Intel® Xeon® processor E5-2600 v3 family
- FDR InfiniBand interconnect
- ~18 000 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 through user guides
  - Partition CPU architecture can be specified

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*)

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 Intel Xeon Phi nodes (90 Xeon Phis)
  - 16 GB memory per card
- Energy efficient CPU's

How to access (plan)

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

DDN Phase 3

- HPC storage used by Sisu and Taito
- 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), probably in August

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
- *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* (around 1.8 TB) to be used for *compiling* codes on login nodes

Grid computing with Finnish Grid Infrastructure (FGI)

FGI

- 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: 12 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:
    - **arcproxy** (Set up grid proxy certificate for 12 h)
    - **arcsub *job.xrsl*** (Submit job described in file job.xrsl)
    - **arcstat -a** (Show the status of all grid jobs)
    - **arcget job_id** (Retrieve the results of a finished grid job)
    - **arckill job_id** (kill the given grid job)
    - **arcclean -a** (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" )
)
```

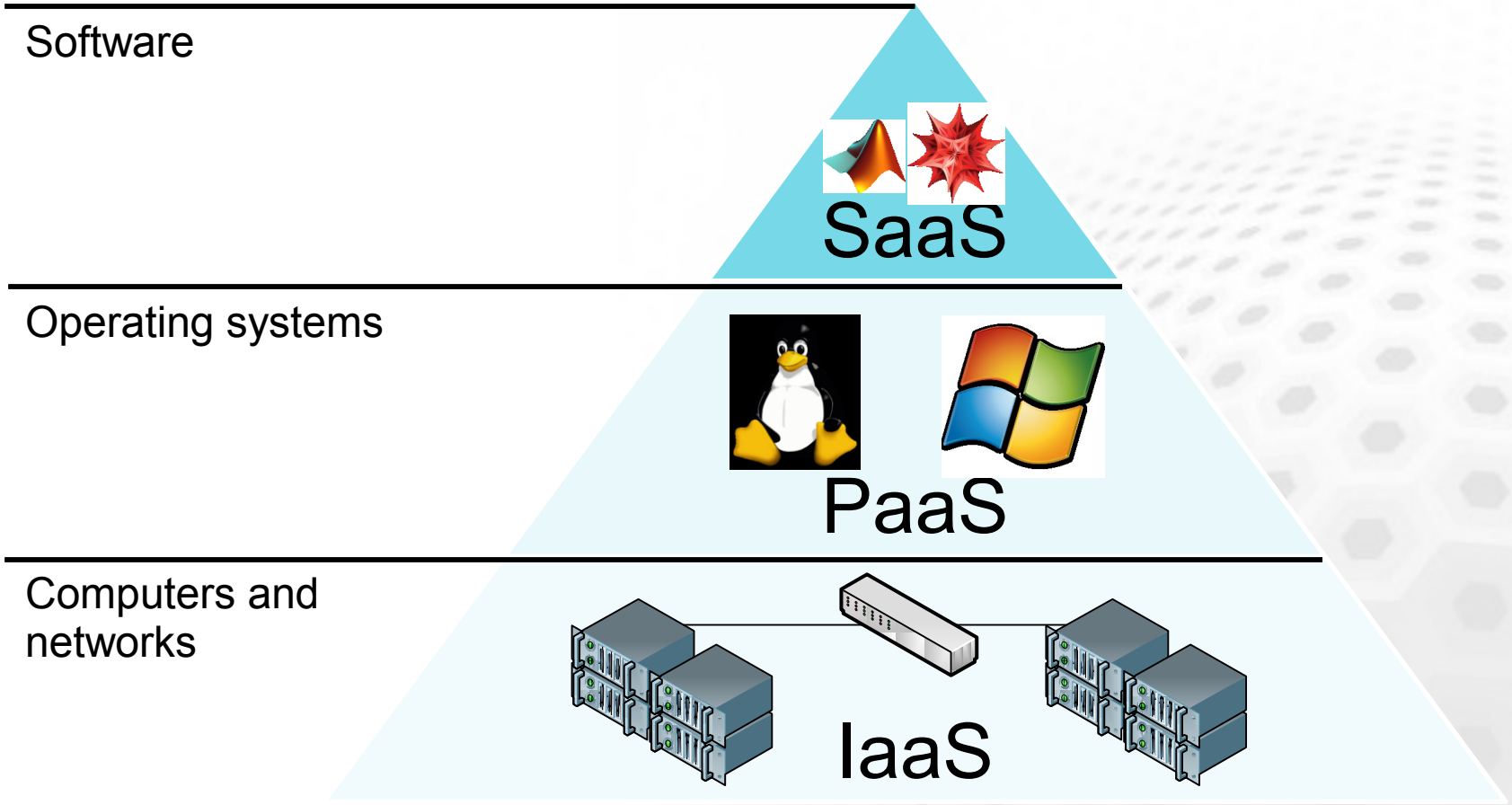
Getting started with FGI-Grid

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

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

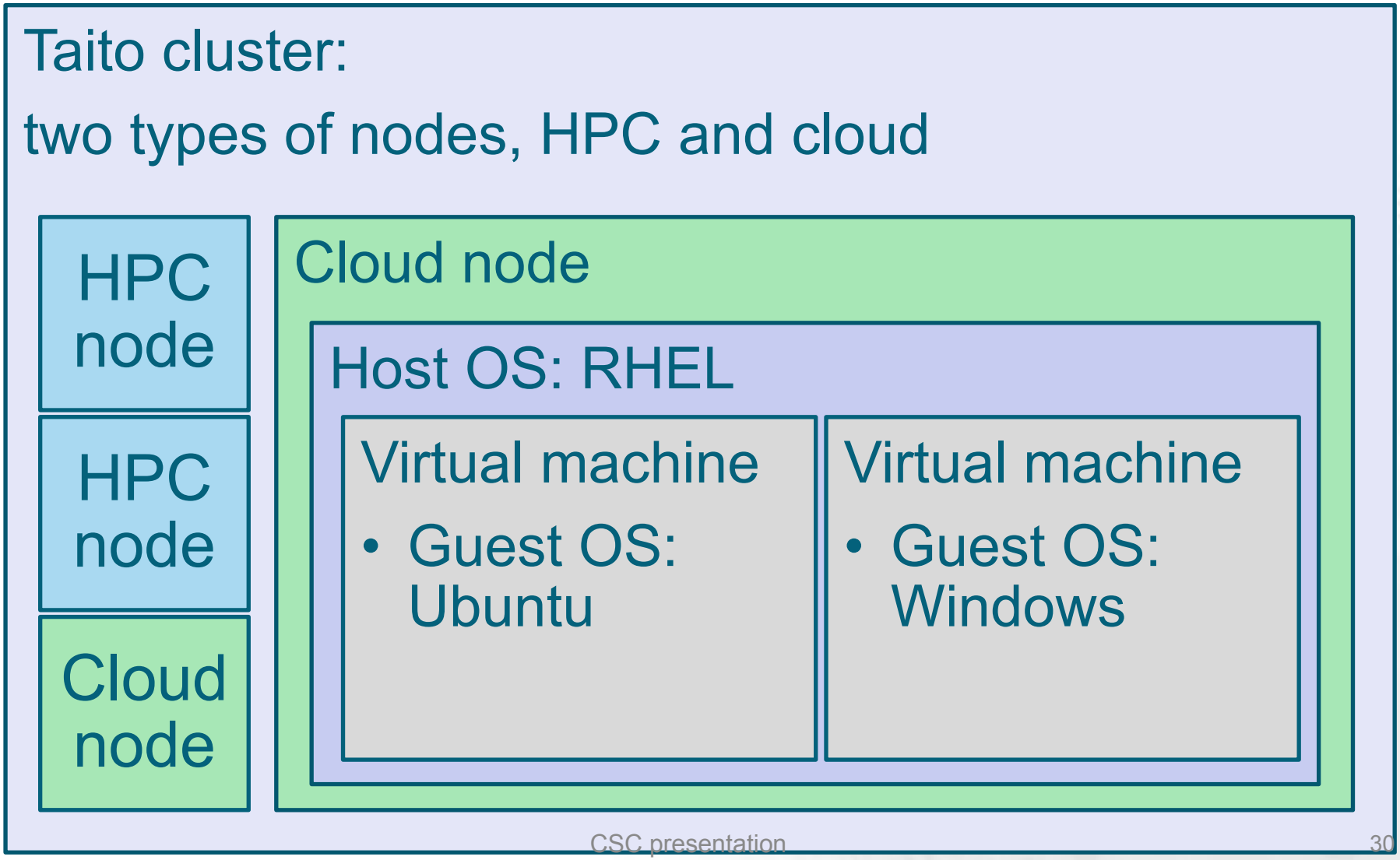
Cloud computing: three service models

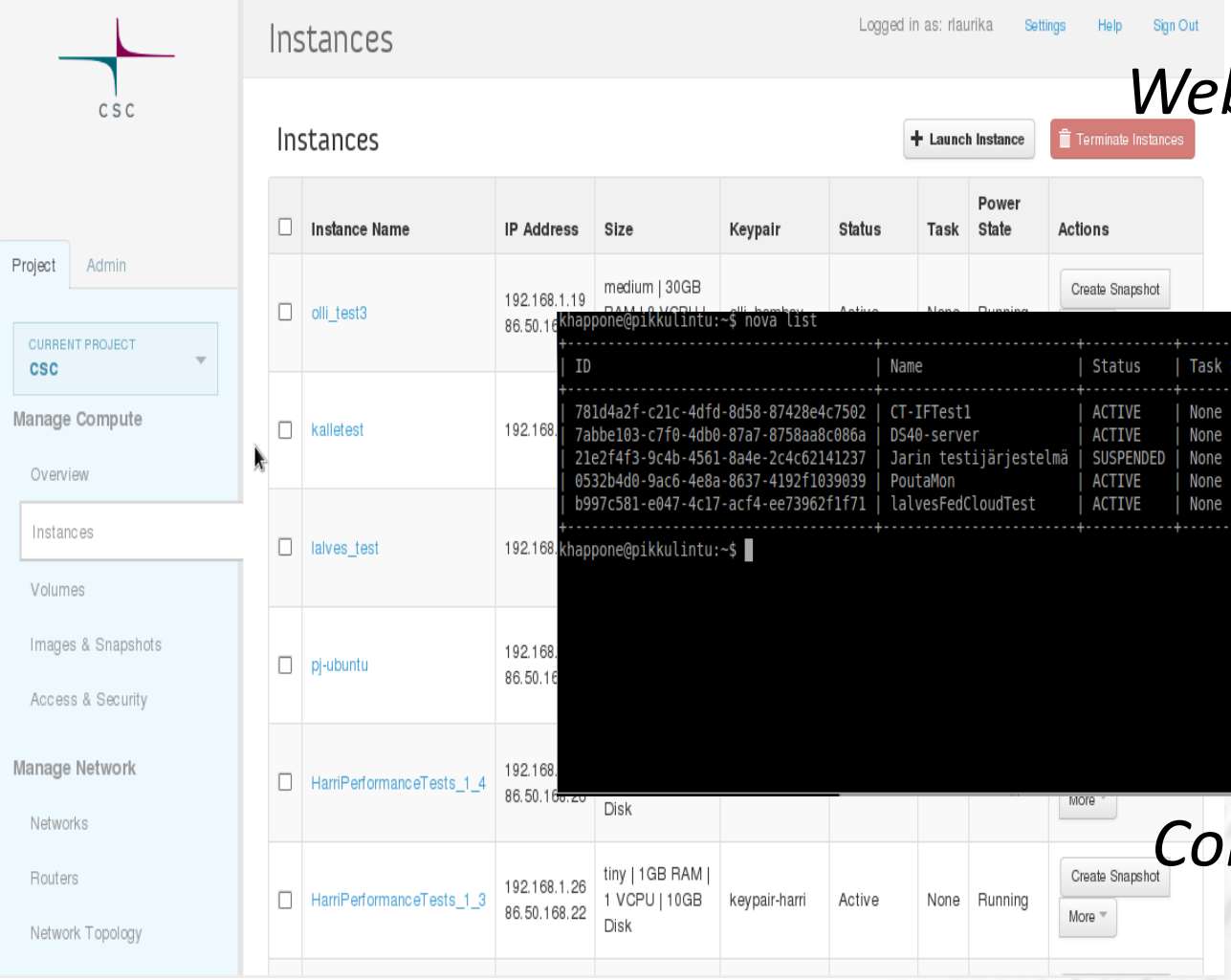


Pouta – Computing in the Cloud

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

Pouta on Taito





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'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
- "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*

Biomedical cloud users

- CSC's cloud cluster in Espoo
  - Built on Biomedinfra project (2010-2013)
  - Services run on CSC IaaS by university IT department for end users (SaaS for end users)
  - Extending existing cluster (computing and storage capacity)
- Users from several institutions, e.g. University of Helsinki, Finnish Institute for Molecular Medicine
- Further information: contact@csc.fi
- *ELIXIR Finland stakeholders' meeting*

ePouta

- Renewing the cloud cluster in Espoo
  - Changes to OpenStack cloud middleware (autumn 2014)
  - Focus on security and organizational customers: service for the IT admins
  - Funding model and resource allocation policy is still open. Hardware needs to be renewed in 2015

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*

Courses

- Sisu Phase 2 workshop
  - Late 2014
- Taito Phase 2 workshop
  - Spring 2015
- CSC courses: *http://www.csc.fi/courses*
  - CSC HPC Summer School
  - Spring, Autumn, Winter Schools
  - Introduction to Linux and Using CSC Environment Efficiently + possibly Pouta training
  - Parallel Programming

Grand Challenges

- Normal GC *(in half a year / year)*
  - New CSC resources available for a year
  - No bottom limit for number of cores
- Remember also PRACE/DECI calls
  - We can help with the technical aspects of the applications

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*
  - Part of Taito used for *Pouta Cloud*
- ***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
  - Adding *1.9 PB*, in *Q3 2014* totaling *4 PB of fast parallel storage*

Feedback form and Round robin

Feedback form

- *https://www.webropolsurveys.com/S/CF54363343815E9B.par*
- *(link also on the seminar home page www.csc.fi →)*

Round robin

- *What are your research interest?*
  - How CSC can help?
  - Special libraries/tools?

- Interested in Cloud service?
- Courses/training?
- Queue length: 3 (Sisu) / 7 (Taito) days enough?
  - Codes that can't checkpoint?
- Is memory an issue for you?
  - 1.5 TB/nodes usage policy?
- Need to move a lot of files? (from where?)
- Interested in GPGPU/MICs? Which code?

- Jiaxin Ling
- Niina Sandholm
- Marjo Hytönen
- Miko Valori
- Emil Ylikallio
- Mabruka Salem
- Preethy Sasidharan Nair
- Anni Evilä
- Laxman Yetukuri
- Christian Benner
- Michael Jeltsch
- Kirsi Järvi
- Minna Brunfeldt
- Kaisa Silander
- Koski Matti
- Mikko Liljeström
- Ville Rantanen
- Risto Lapatto
- Paulina Deptula
- Arnab Bhattacharjee
- Rigbe Gebremichael Weldatsadik
- Vishal Sinha
- Sampsa Hautaniemi
- Tiia Pelkonen
- Tero Hiekkalinna