

GHENT UNIVERSITY

DEPARTMENT ICT

HPC-UGENT USER MEETING 2017

Dr. Ewald Pauwels hpc@ugent.be http://hpc.ugent.be







ROGRAM

- (13h00:Optional tour of datacenter)
- 14h00: Welcoming address
- 14h15: Overview of HPC-UGent, VSC, future plans
- 14h45: Review of user poll results, Q&A
- 15h15: User in the spotlight Pieter Reyniers, LCT
- 15h45: Slots for 1-minute poster presentations ullet
- 16h15 18h00 Networking reception & poster session



ABOUT HPC-UGENT

Part of ICT department, Infrastructure office

Mission

HPC-UGent provides centralised scientific computing services, training, and support for researchers from Ghent University, industry, and other knowledge institutes.



ABOUT HPC-UGENT

Personnel

- User support
- Training
- Infrastructure installation and upkeep (software & hardware) ullet
- Outreach + marketing
- Collaboration with other supercomputing centers



Alvaro Simon Garcia Cloud, user support



Andy Georges Sysadmin, integration





Kenneth Hoste User support



Kenneth Waegeman *Sysadmin, storage*





Always mail hpc@ugent.be

Ewald Pauwels Team lead

Stijn De Weirdt Technical lead

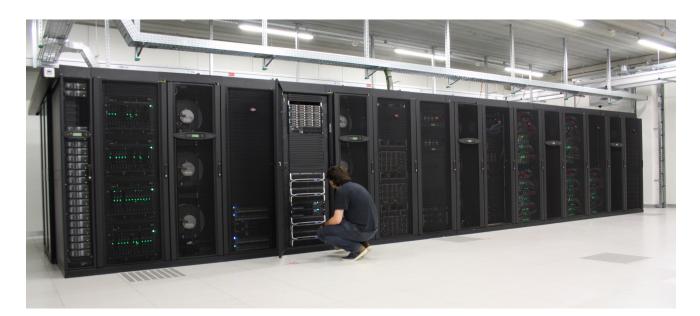




Jens Timmerman Security, sysadmin

Wouter Depypere Sysadmin

HPC-UGENT INFRASTRUCTURE



\$VSC_HOME \$VSC_DATA \$VSC_SCRATCH \$VSC_SCRATCH_KYUKON



GHENT

UNIVERSITY

Compute clusters

	#nodes	CPU	Mem/node	Diskspace/node
Raichu	64	2 x 8-core Intel E5-2670 (Sandy Bridge @ 2.6 GHz)	32 GB	400 GB
Delcatty	160	2 x 8-core Intel E5-2670 (Sandy Bridge @ 2.6 GHz)	64 GB	400 GB
Phanpy	16	2 x 12-core Intel E5-2680v3 (Haswell-EP @ 2.5 GHz)	512 GB	3x 400 GB (SSD, striped)
Golett	200	2 x 12-core Intel E5-2680v3 (Haswell-EP @ 2.5 GHz)	64 GB	500 GB
Swalot	128	2 x 10-core Intel E5-2660v3 (Haswell-EP @ 2.6 GHz)	128 GB	1 TB

Storage 35 TB 702 TB (1 PB) 2017-Q2 2017-Q2 1 PB \$VSC_SCRATCH_PHANPY 35 TB SSD

Network

GbE

FDR InfiniBand

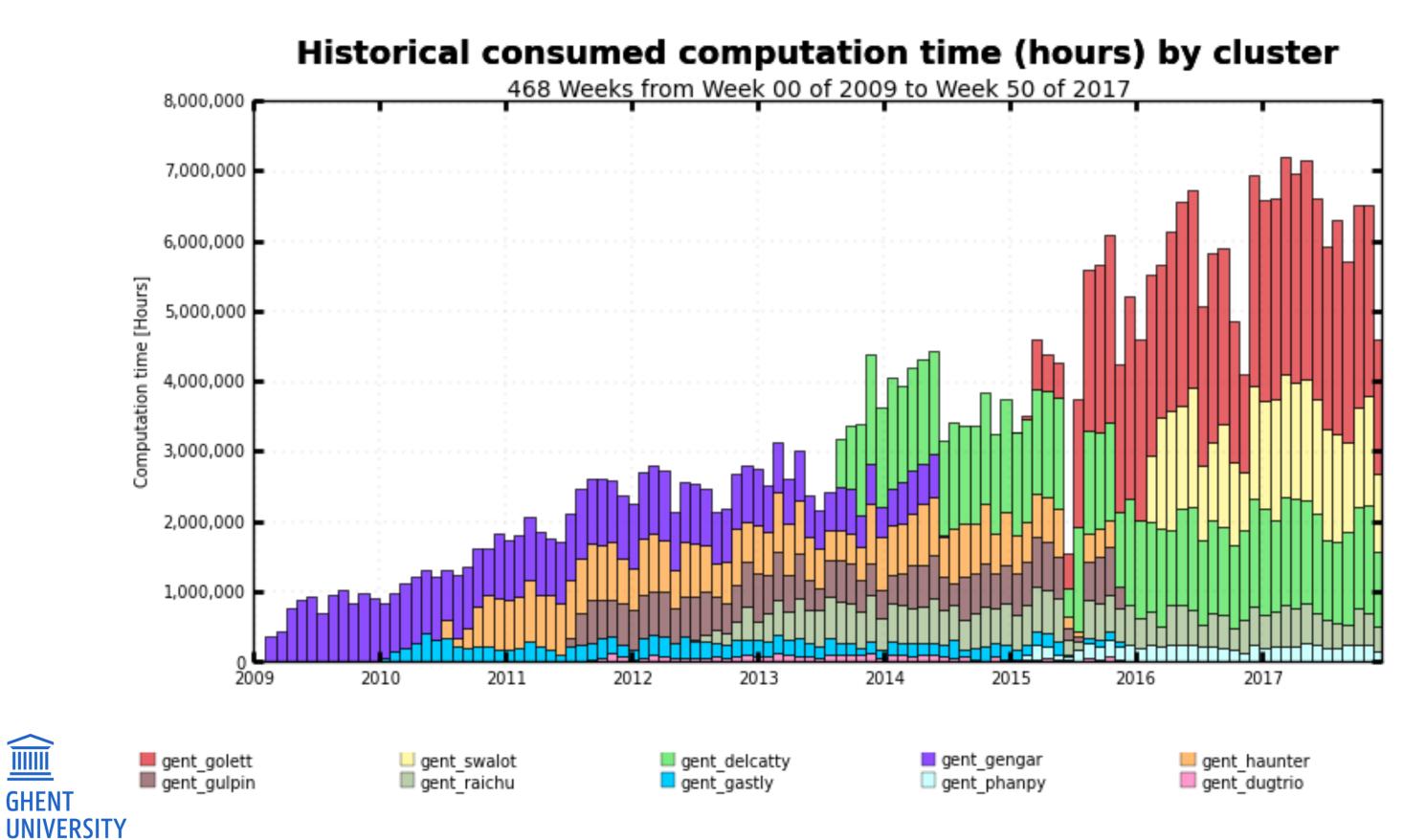
FDR InfiniBand

FDR-10 InfiniBand

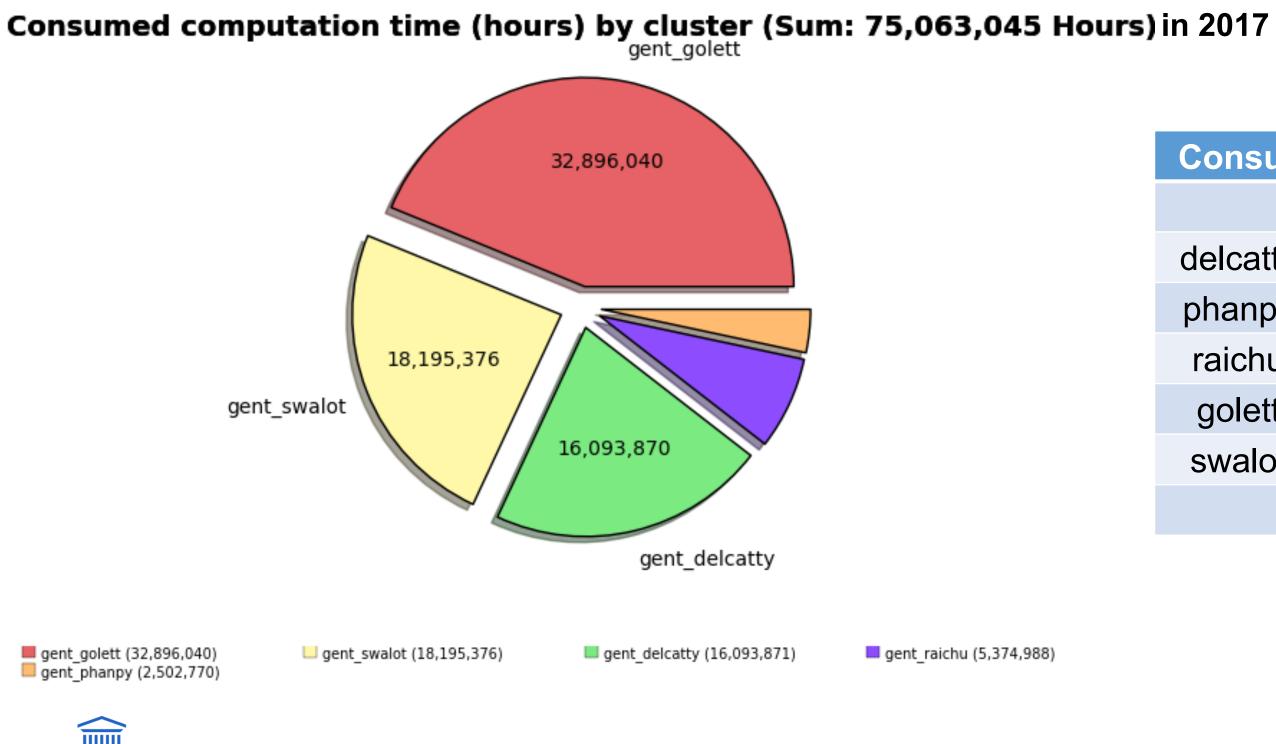
FDR InfiniBand

2016-Q3

GHENT











Consumed compute time in 2017					
	core years	use %			
delcatty	1837	80%			
phanpy	286	78%			
raichu	614	72%			
golett	3755	84%			
swalot	2077	85%			
	8569	82%			

Other

Top 10 users						
	core		core			
	years		years			
vsc40941	477	vsc41326	285			
vsc40944	473	vsc40588	255			
vsc40484	470	vsc40686	248			
vsc40309	398	vsc41224	205			
vsc41948	301	vsc41228	180			

GHENT

UNIVERSITY

34% **Bioinformatics/biostatistics**, biomathematics, biomechanics 3% **Theoretical computer** science 5% **Planetary science Civil engineering** 6%





Chemical technology, process technology 14%

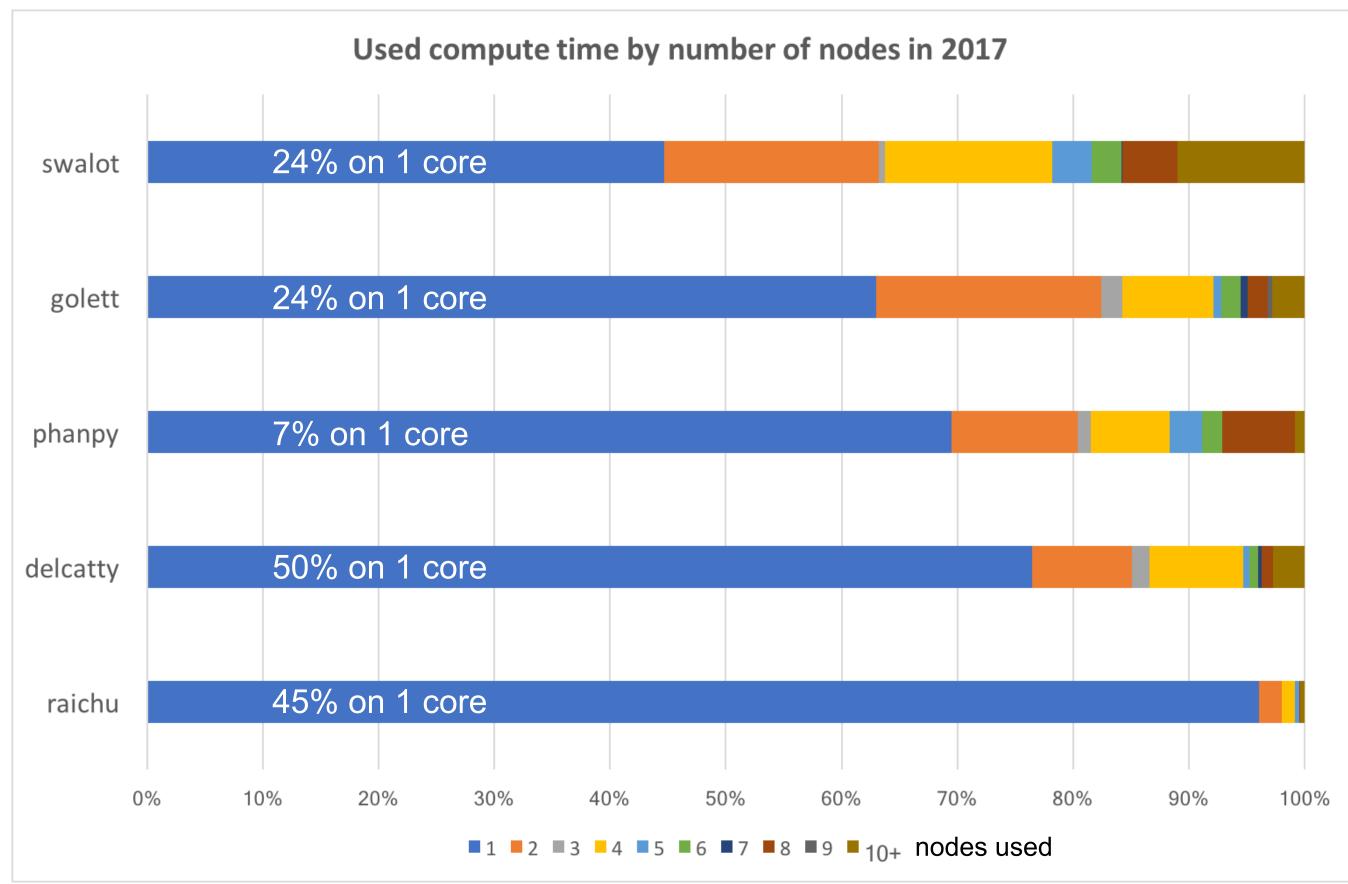
Theoretical chemistry, quantum chemistry 10%

Materials technology 6%

Chemistry, other 3%

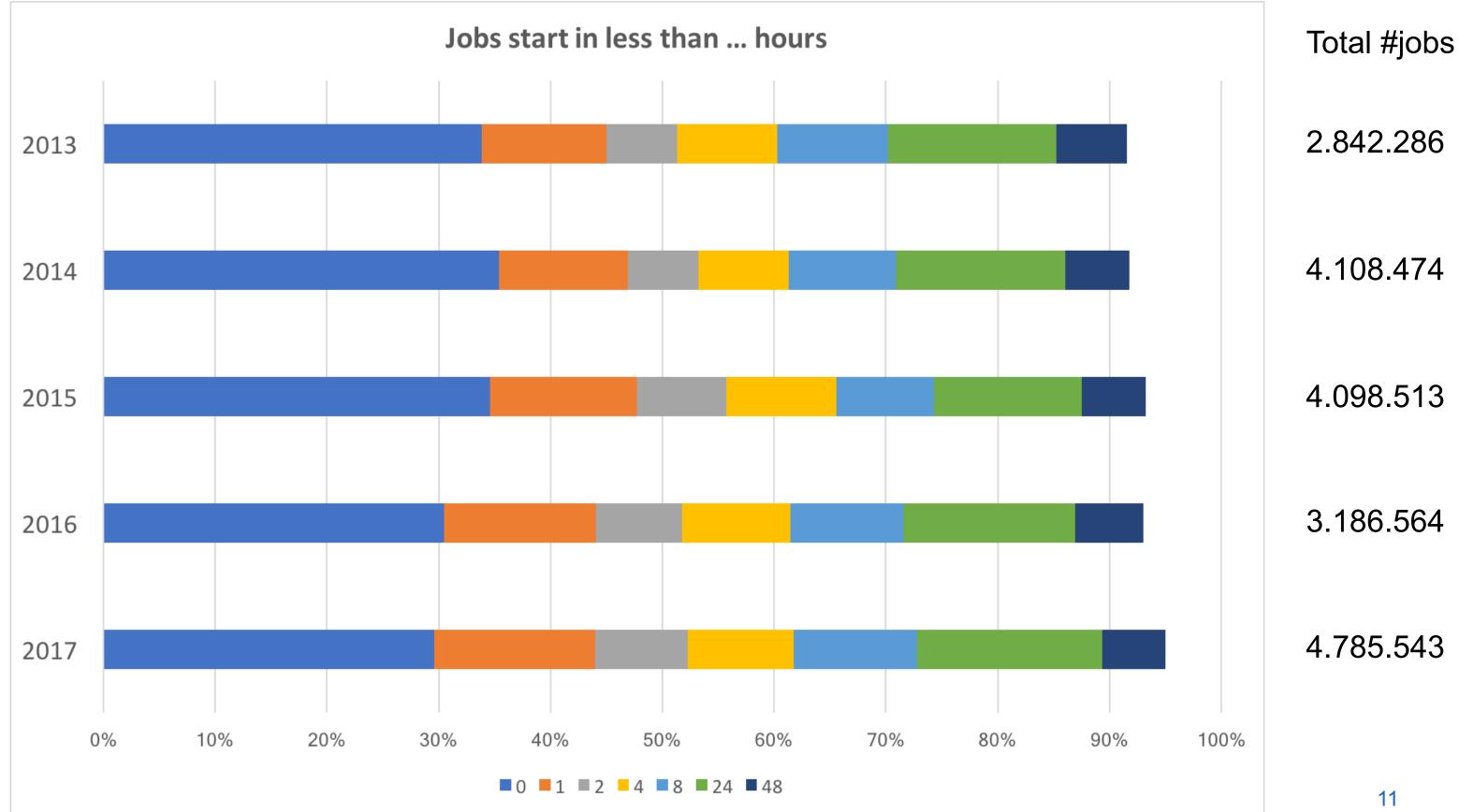
Condensed matter and optical physics 3%

Nanophysics/technology 9%



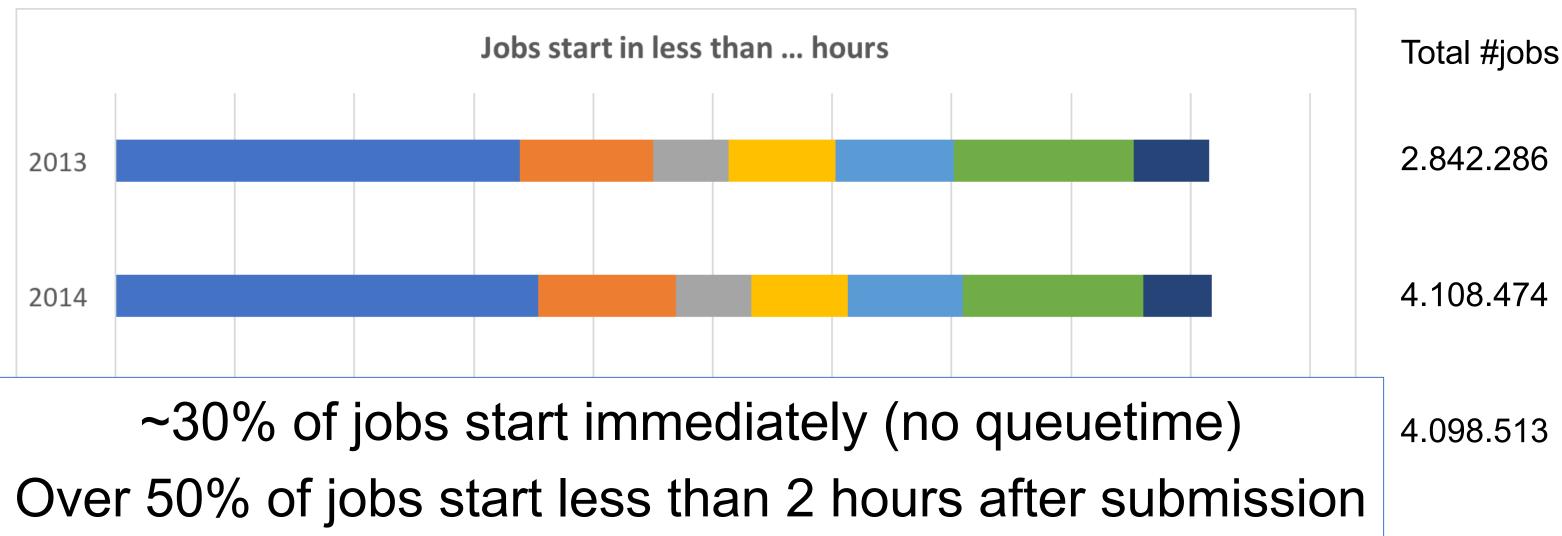






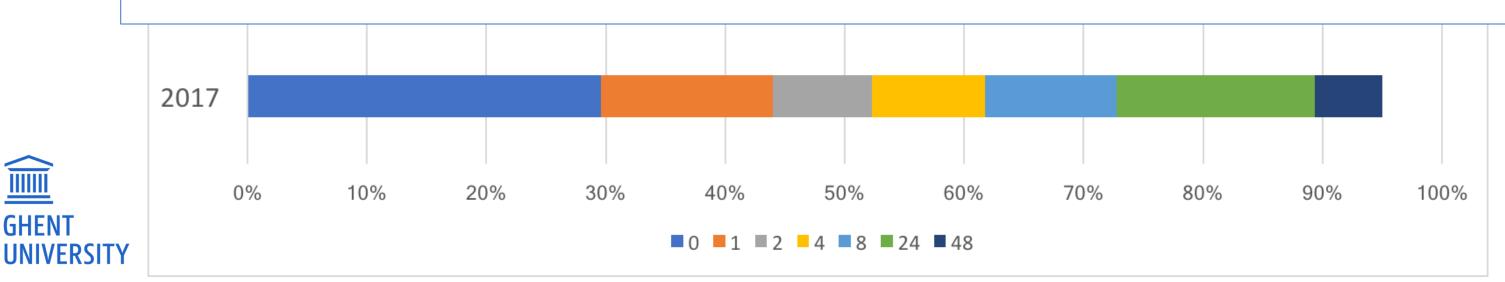






~90% of jobs start in less than 24 hours

GHENT





3.186.564

4.785.543

<u>HPC-UGENT – NEW SERVICES</u>

Extra compute clusters

- 2.2M euro investment
- Production expected by summer 2018

#nodes	CPU	Mem/r
72	2 x 18-core Intel Xeon Gold 6140 (Skylake @ 2.3 GHz)	192 (
	2592 cores in total	
96	2 x 18-core Intel Xeon Gold 6140 (Skylake @ 2.3 GHz)	96 G
	3456 cores in total	



node Disk/node Network

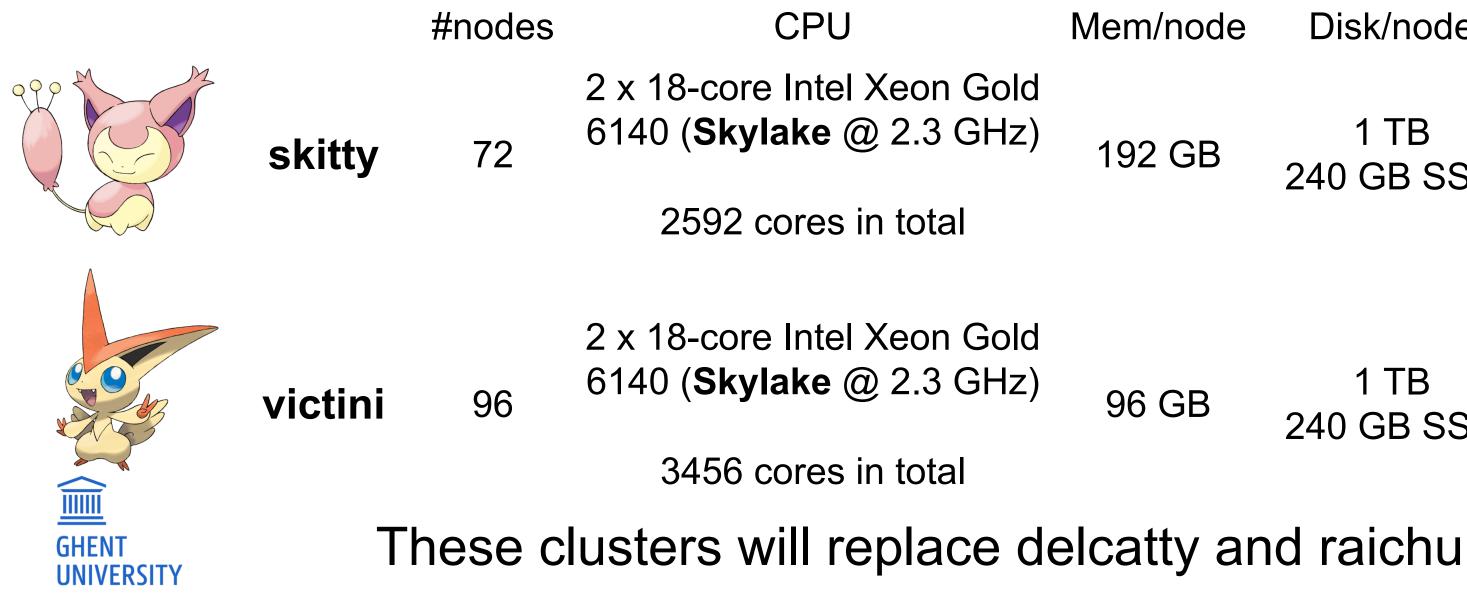
GB 240 GB SSD InfiniBand

GB 1 TB 10 GbE 240 GB SSD

HPC-UGENT – NEW SERVICES

Extra compute clusters

- 2.2M euro investment
- Production expected by summer 2018



Disk/node Network

1 TB EDR 240 GB SSD InfiniBand

1 TB 10 GbE 240 GB SSD

HPC-UGENT – NEW <u>SERVICES</u>

Cloud testbed 'grimer'

- 16 hypervisors 256 CPU cores 200 TB storage (Ceph)
- Reuse of decommissioned muk (old Tier-1) hardware
- For users with specific requirements
 - Public cloud infrastructure
 - Graphical user interface

- Software with wallclock > 72h ? (But checkpointing is far more advisable)
- VM(s) managed by user! Currently free of charge
- Testbed to gather expertise underlying technology may change

Contact hpc@ugent.be if you would like to be pilot



<u>HPC-UGENT – NEW SERVICES</u>

Cloud testbed 'grimer': example application

search tools
RNA-SEQ ANALYSIS
Summarization

🚍 Galaxy / Galaxy @ VIB-UGent

1

8

Quantification using lightweight alignment

Read Mapping FastQ Quality Control

Tools

CHIP-SEQ ANALYSIS

Extract motifs from peak data Binding and Expression Target Analysis (BETA)

Model-based Analysis of ChIP-Seq

CRISPR ANALYSIS

Genome editing

STATISTICAL ANALYSIS

Differential expression

DATA MANIPULATION

Samtools

Bedtools

Deeptools Picard tools

DEFAULT OPERATIONS

<u>Get Data</u>

Send Data

Collection Operations

Text Manipulation

Filter and Sort

Join, Subtract and Group Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments Statistics

Graph/Display Data

Workflows

All workflows



Welcome to the Galaxy Instance of VIB-UGent

Analyze Data Workflow Shared Data - Visualization - Admin Help - User -

How to use Galaxy ?

Tools: In the left panel of this page you will find the list of available tools, you can see this page at any point by clicking 'Analyze Data' in the top menu.

History: The history of your current data analysis is shown in the right panel. A good practice when running multiple analysis in parallel is to create several histories and give each one a unique name. You can click on 'View all histories' button (upper right corner in histories panel) to get an overview and switch between the current histories.

Input Data: Every analysis starts with getting the input data into your current history (right panel). To do this you can upload your own input files or use shared datasets. To upload files from your computer or instruct Galaxy to download files from the web you have to use of the upload tool: Get data (tool panel on the left) -> upload file. Please DO NOT UPLOAD LARGE FILES (~GB), but contact the administrator to create a central data repository! To use available shared data you have to click on Shared data (top menu) -> Data Libraries. You can then browse the available libraries and select the file/s you want to use. By clicking 'to History' button and choosing the desired History name you will import these files and make them available to use as input for future analysis.

Execution: To run a job select the tool from tool panel on the left, then the corresponding interface will be loaded and you will be able to select corresponding input data and (re)define parameters. Please read carefully the labels and help text next to the input fields. After clicking Execute you will be able to see entries for each in your history. The color of a dataset designates the current status of the underlying job

- . Grey: The job is being evaluated to run (new dataset) or is queued. Allow this to complete.
- Yellow: The job is executing.
- Green: successful processing
- Red: The job has failed.

• Light blue: The job is paused. This indicates either a problem with an input (a previous step in the workflow may have failed) or that you have exceeded disk quota set by the administrator of the Galaxy instance you are working on.

Workflows: You can automate your analysis pipeline by using workflows composed of several tools linked by their input/output data. In the Workflows section (upper menu) you can see a list of current workflows and also create your own ones. It is also possible (and highly recommended) to use the shared, and widely tested, workflows available under Shared Data (top menu) -> workflows. To use one of these you first need to import it to your workflows list by clicking on the name and selecting Import. To execute any workflow listed under your workflows first click on it and select run, then choose the input data as with any tool.

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.

<u>IIIIIII</u> GHENT UNIVERSITY





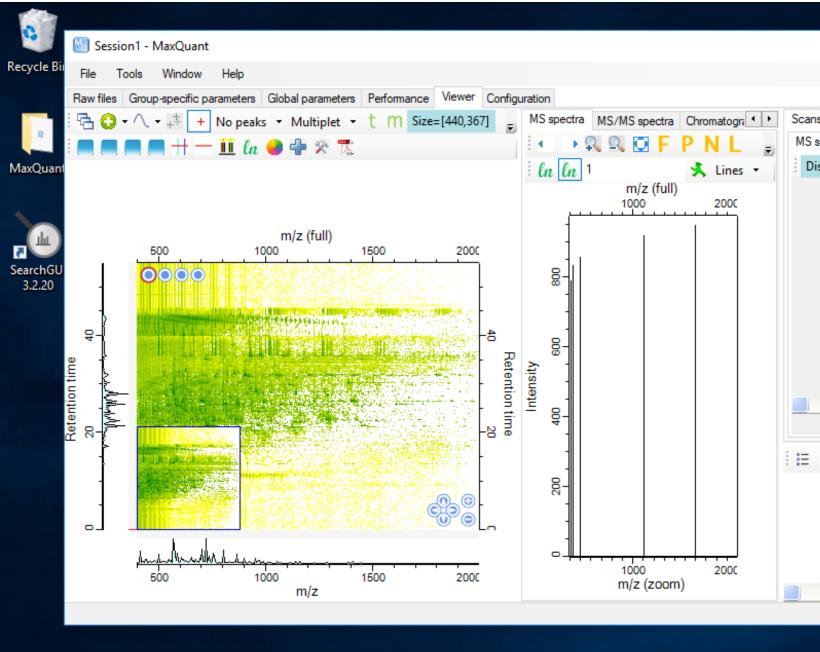
	Using 55.5 GB
History	2 0 🗆
search datasets	8
Salmon strandedness tes 115 shown, 12 <u>deleted</u> , 1 <u>hidd</u>	
50.97 GB	S
<u>127: Salmon on data 1</u> <u>(SAM format)</u>	• / ×
<u>126: Salmon on data 1</u> (Gene Quantification)	• / ×
<u>125: Salmon on data 1</u> (log)	• / ×
<u>124: Salmon on data 1</u> (Quantification)	• / ×
<u>123: Salmon on data 1</u> (SAM format)	● 🖋 🗙
<u>122: Salmon on data 1</u> (Gene Quantification)	● 🖋 🗙
<u>121: Salmon on data 1</u> (log)	● / ×
<u>120: Salmon on data 1</u> (Quantification)	● 🖋 🗙
<u>119: Salmon on data 1</u> (SAM format)	● 🖋 🗙
<u>118: Salmon on data 1</u> (Gene Quantification)	• 🖋 🗙
<u>117: Salmon on data 1</u> (log)	● / ×
<u>116: Salmon on data 1</u> (Quantification)	● 🖋 🗙
<u>115: Salmon on data 1</u> (SAM format)	● / ×
<u>114: Salmon on data 1</u> (Gene Quantification)	● / ×
<u>113: Salmon on data 1</u> (log)	● / ×
<u>112: Salmon on data 1</u> (Quantification)	● 🖋 🗙
😢 110: Salmon on data	@ # x

16

<u>HPC-UGENT – NEW SERVICES</u>

O Type here to search

Cloud testbed 'grimer': example application



e

 $\overline{}$

[]]

Ţ



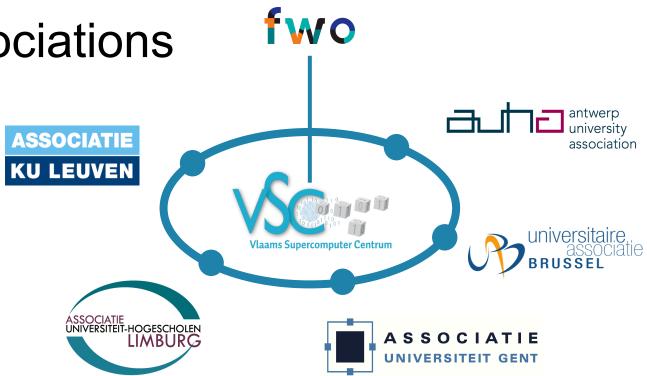


					_)	×	
ns	lde	ntif	icatio	ons	Fea	ature	s	Meta	dat	a	
sca	ns	Μ	S/M	S sc	ans	MS	53 s	cans			
isp	lay :	sel	ecte	d sp	ectr	um					
							_				
							10	0%	~	1	
•	• •		2	Ţ	7						
					- 14						
	_					_					
		_				Ve	ersio	on 1.(5.0	.1 ,;	÷
^	, F	2	¢≈	EN	IG		1:0 10/2	2 2017			
							- A.	- C - C - C - C - C - C - C - C - C - C		- C -	



VSC – Flemish Supercomputer Center

- Partnership between Flemish university associations
- Infrastructure in four hubs
- Managed by FWO \bullet





Mission

The VSC encourages the use of scientific and technical computing in the Flemish academic and industrial landscape. To this end, it offers infrastructure, training and services. In addition, VSC acts as a lever to promote the importance of scientific and technical computing and its added value to society.



ABOUT VSC – GOVERNING BODIES

VSC steering group

Daily (monthly) management of VSC, consensus

UGent representative: Ewald Pauwels

Users committee

Map user needs, advise on Tier1/2 operation, VSC user day UGent representatives: Veronique Van Speybroeck, Marie-Françoise Reyniers Tier1 evaluation committee (non-Flemish experts) Technical evaluation of Tier-1 applications Industrial board

Advise to increase involvement of industry FWO

Funding, final governance





ABOUT VSC - GOALS

- Offer its target audience access to diversified ICT infrastructure that is tailored to the needs of scientific/technical computing.
- 2. Provide a **common user environment** on the computing infrastructure, which is available in the local hubs.
- **3. Support** its users so that they can lift their research and development to a higher level by using scientific/technical computing.
- **4. Inform about the capabilities and achievements** of scientific/technical computing and its potential added value.
- 5. Actively promote scientific/technical computing in Flemish industry and foster the exchange of ideas and expertise between research institutions and industry.
- Offer a diverse and coordinated training program across the VSC consortium to stimulate and advance the uptake of scientific/technical computing in new and existing users.
- 7. Engage and actively participate in **international initiatives** such as PRACE and Horizon 2020, and cooperate with other centers focusing on scientific/technical
- \implies computing.

GHENT UNIVERSITY

ish industry and foster the ons and industry. the VSC consortium to mputing in new and existing

VSC INFRASTRUCTURE

https://www.vscentrum.be/infrastructure/hardware Available hardware

Tier-1

- Our current Tier-1 system is BrENIAC, operated by KU Leuven. The system is aimed at large parallel computing jobs that require a high-bandwidth low-latency interconnect. Compute time is again only available upon approval of a project. See the page on Tier-1 project access and links in that page.
- Our first Tier-1 system is muk, was operated by UGent but is no longer in production.

Experimental setup

• There is a small GPU and Xeon Phi test system which is can be used by all VSC members on request (though a project approval is not required at the moment). The documentation for this system is under development.

→ <u>http://hpc.ugent.be/userwiki/index.php/Tips:Software:GPGPU</u>

Tier-2

Four university-level cluster groups are also embedded in the VSC and partly funded from VSC budgets:

- The UAntwerpen clusters (hopper and leibniz)
- The VUB cluster (hydra)
- The UGent local clusters
- The KU Leuven/UHasselt cluster (ThinKing and Cerebro)

Free of charge

Not free of charge, but heavily discounted



21

Free of charge

VSC INFRASTRUCTURE

Using other VSC infrastructure

- Don't hesitate
- If unsure about pricing, ask KULeuven for quote
- Feel free to ask support at another VSC site, e.g.
 - Error reporting
 - Trouble with credit system
 - Software installation
- Always put <u>hpc@ugent.be</u> in cc



22

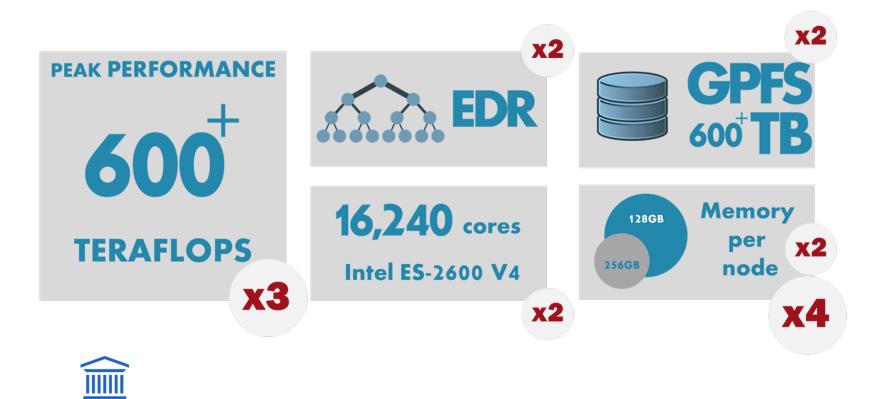
VSC INFRASTRUCTURE – TIER-1

BrENIAC @ KULeuven



Muk @ UGent \rightarrow





GHENT

UNIVERSITY

- 580 nodes
- 128/256 GB RAM ullet
- lacksquare
- 634 TB storage ullet



• 2 Intel Xeon Broadwell \rightarrow 16.240 cores InfiniBand EDR interconnect

TIER-1 ACCESS – STARTING GRANT

https://www.vscentrum.be/en/access-and-infrastructure/tier1-starting-grant

- Purpose = explore, do scaling tests of your software, prepare for project
- 100 node days (= 100 x 28 x 24 = 67.200 core hours)
- Available for 2 months
- Personal grant
- Fast submission procedure, very short proposal
- Constantly reviewed
- Success rate = 100%
- FREE OF CHARGE





TIER-1 ACCESS – PROJECT ACCESS

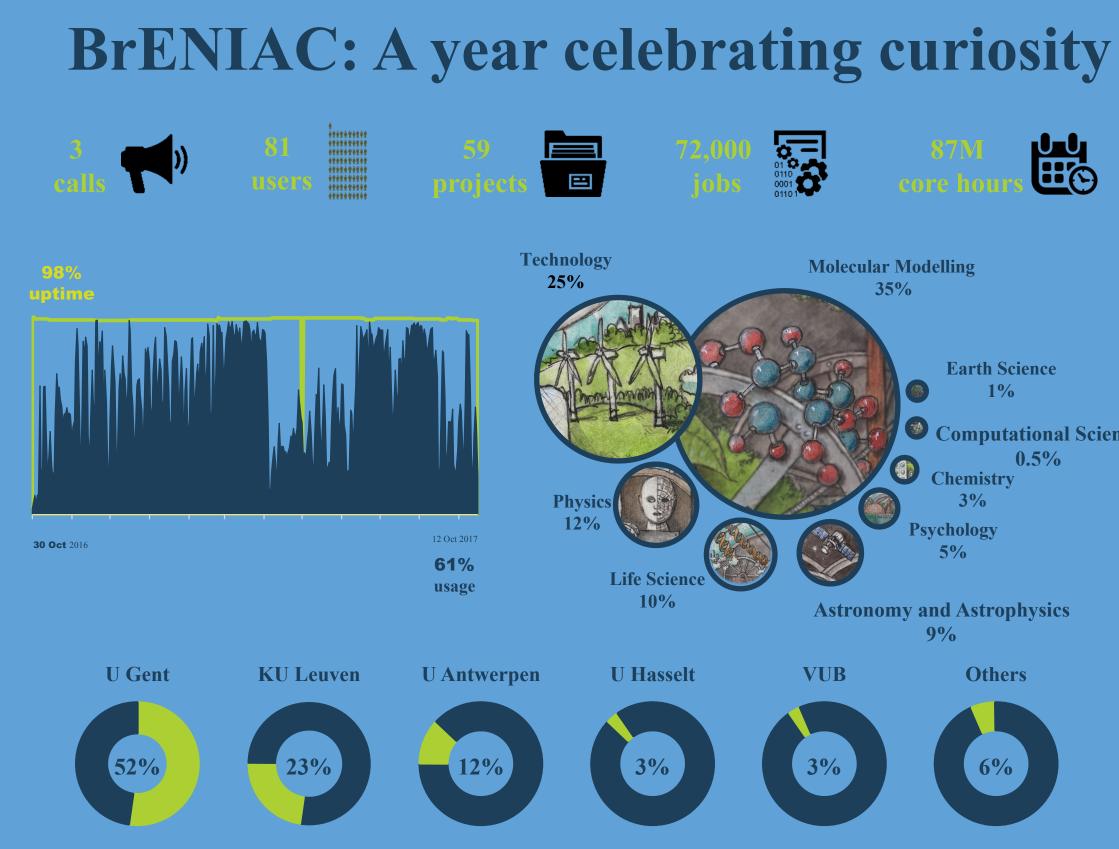
https://www.vscentrum.be/en/access-and-infrastructure/project-access-tier1

- 500 5000 node days (= 336.000 3.360.000 core hours)
- Available for 6 months
- Can be granted to multiple researchers
- Reviewed 3x per year by Tier-1 Evaluation Committee
 - Next deadline = 5 February 2018
- Success depends on quality of your proposal
 - Send your proposal to <u>hpc@ugent.be</u> for prior review
- FREE OF CHARGE





VSC INFRASTRUCTURE TIER-1





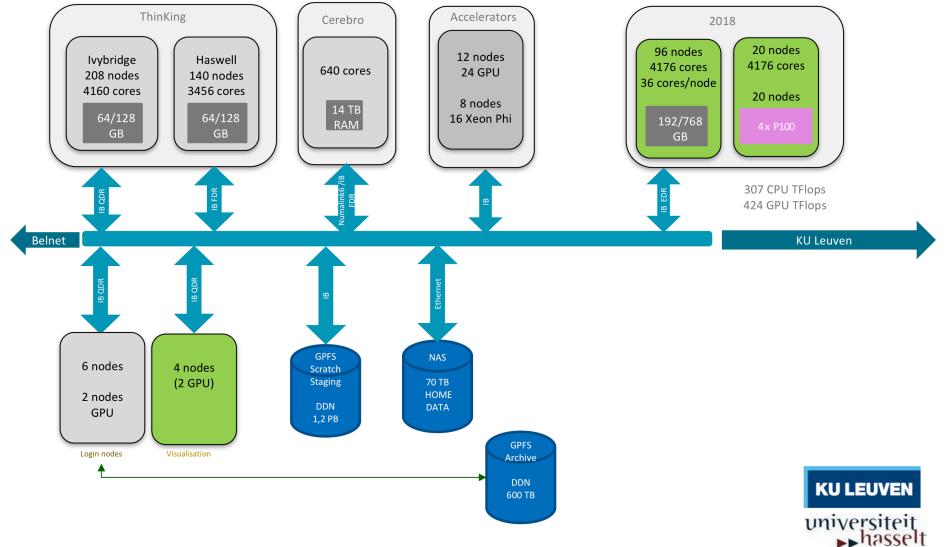


Computational Science

VSC - FUTURE PLANS



20 nodes with each
 4 x Nvidia Tesla P100



- KULeuven will likely open (free) pilot access to GPUs
- Let us know via <u>hpc@ugent.be</u> if you would like to use these GPUs



PUs o use these GPUs

VSC - FUTURE PLANS

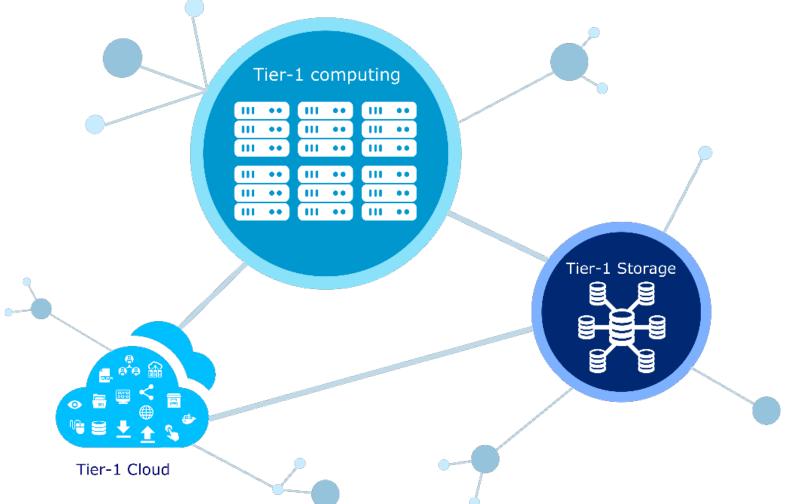
New Tier-1 supercomputing platform

(Subject to final approval by government)

- Intention to be structural ${\color{black}\bullet}$
- Complementary programs ullet
 - CloudStorage very challenging

 - Compute: extending current Tier-1 service
- First developments proposed starting 2018



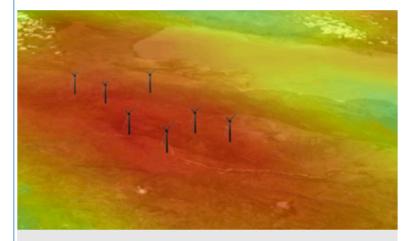


VSC – CALL FOR SUCCESS STORIES

Cases and projects

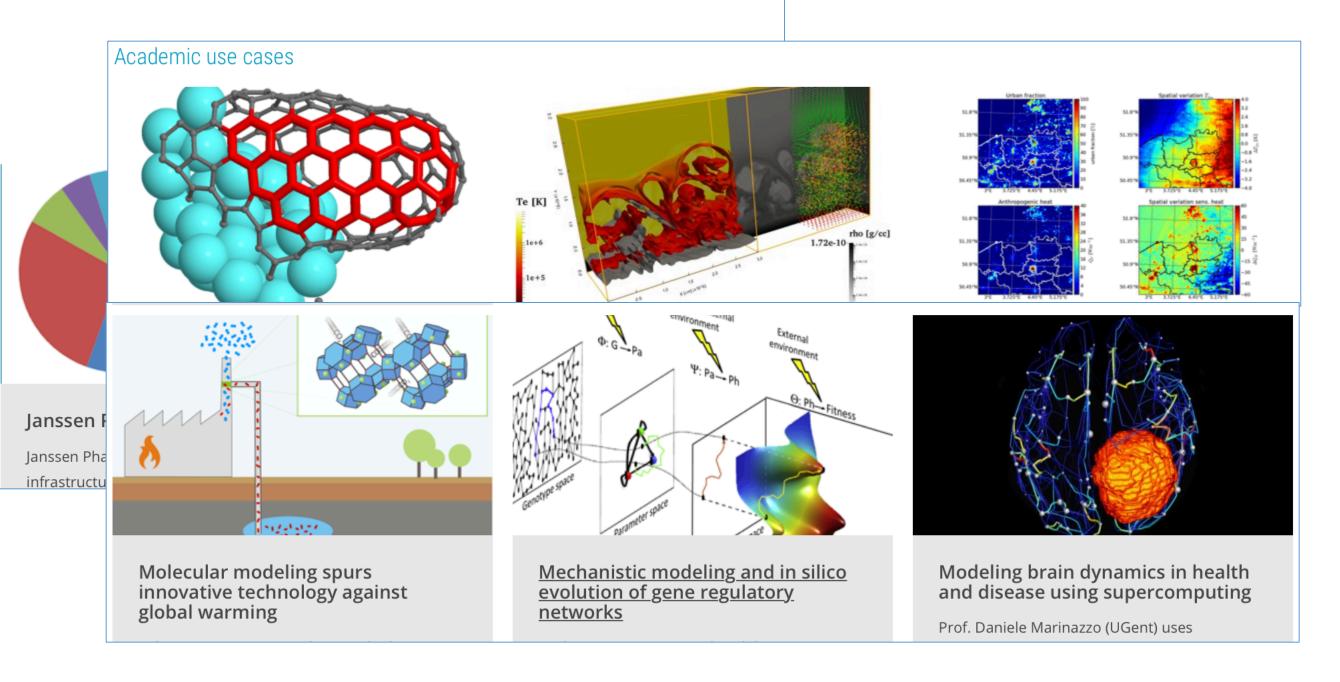
Achievements on our infrastructure

Industrial use cases



Wind energy simulations at 3E NV

3E NV, a Belgian renewable energy consultancy firm, makes use of the VSC infrastructure and





https://www.vscentrum.be/en/project-and-cases

Contact hpc@ugent.be if you could contribute

- 104 respondents
- ~8 questions
- Average completion time: 5 minutes

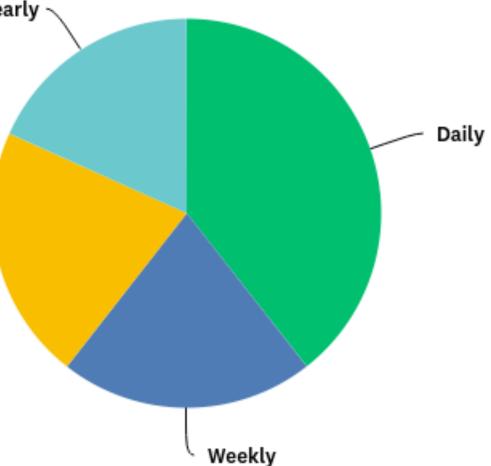


How often do you use the HPC-UGent scientific computing infrastructure?

		Year
	#resp	
On a daily basis	41	
Usually once per week	22	
Once per month	22	Monthly
1-2 times per year or less	19	



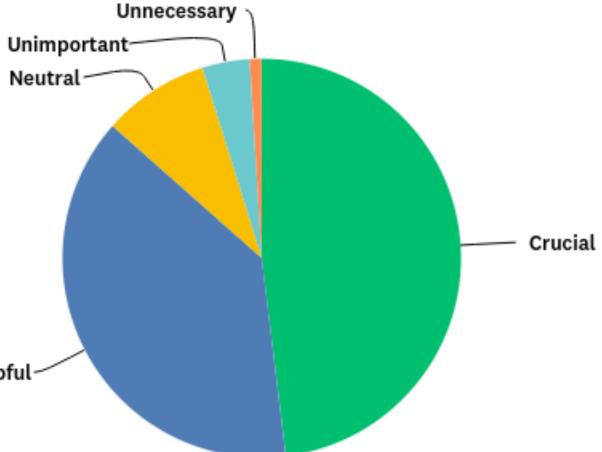




How important is HPC-UGent for your research?

	#resp	
Crucial.	50	Un
I can't do my research without.		Ne
Very helpful.	40	
It allows me to do my research at a faster pace and at a	higher level.	
Neutral	9	
Unimportant.	4	
I can just as well do my research in another way.		Very helpful
Unnecessary.	1	
I don't need it at all to do my research.		





How often / How important?

	How often do you use?						
		Daily	Weekly	Monthly	Yearly		
nt fo ch?	Crucial	32%	7%	7%	3%		
orta sear	Very helpful	7%	14%	13%	5%		
n po	Neutral	1%	0%	2%	6%		
How important for your research?	Unimportant	0%	0%	0%	4%		
Ĭ	Unneccesary	0%	0%	0%	1%		





How would you rate the services that HPC-UGent provides? (compute clusters, login nodes, training, user support, website)

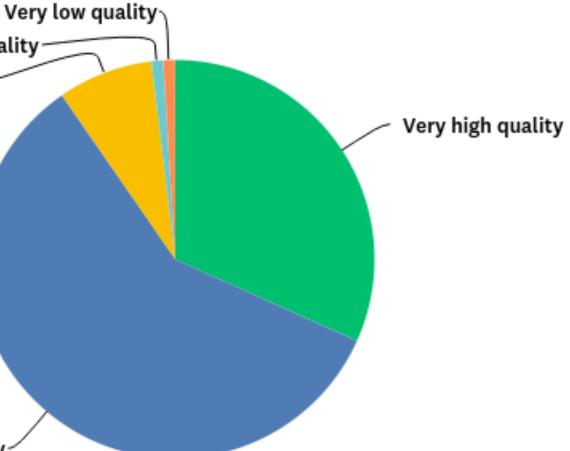
	#resp	Low qualit Meither high or
	-	low quality
Very high quality	33	
High quality	61	
Neither high or low quality	8	
Low quality	1	
Very low quality	1	

High quality-

More than 90% of poll participants rate our services as 'high quality' or above.







How often / Your rating?

	How often do you use?						
L L		Daily	Weekly	Monthly	Yearly		
of ol	Very high quality	15%	8%	7%	2%		
ng c ces	High quality	21%	12%	13%	13%		
ratii ervi	Neither high or low quality	2%	2%	0%	4%		
Your rating of our services?	Low quality	1%	0%	0%	0%		
\succ	Very low quality	0%	0%	1%	0%		





Is there a particular service that stands out or you care to comment about? ©©©

#comments	Service	
16	User Support	excellent, fast, friendly
8	Training	regular, hands-on
6	Infrastructure	diverse, reliable, updated
5	Software	fast installation, version flex
4	Documentation	good, excellent, usable by r





exibility, module system

y newbies

Is there a particular service that stands out or you care to comment about? ©©©

"The clusters in Ghent are still the best managed systems I've been working with. Keep up the nice work guys!"

"The Helpdesk is extremely supportive and helpful! Thank you!"

"The consistent use of the system status page in the User Portal of the VSC website is much appreciated. It is good to have this page as the single point of information for cluster related updates and during upsets."

→ <u>https://www.vscentrum.be/en/user-portal/system-status</u>





Is there a particular service that stands out or you care to comment about? 888

#comments Service

- 5 Infrastructure
- 1 Training
- 1 Documentation

long queues, wallclock, regular downtimes software-specific training sessions

spread out, no overview of basic commands





How could we further improve HPC-UGent services?

- 62 responders \bullet
- 92 suggestions \bullet
- **Detailed follow-up** \bullet in future
- **Recurring suggestions** \bullet in next slides







How could we further improve HPC-UGent services?

Documentation (21)

More structured wiki/website that is SPOI (6)

Documentation better geared towards new users (9)

- Checklist for starters
- Different clusters + storage locations
- Frequently used commands

More examples: scripts, software (3)

Updated documentation (2)

Newsletter





How could we further improve HPC-UGent services?

Documentation (21)

GHENT

UNIVERSITY

GHENT UNIVERSITY							
	ACCESS	INFRASTRUCTURE	USER SUPPORT	TRAINING & LECTURES	INDUSTRY	ABOUT US	
Centralised scientific computi	ıg services, trai	ning, and support for res	earchers and industry				
Support and services			Access to infras	tructure			
→ Documentation	\rightarrow Documentation			\rightarrow Access for staff & academics			
→ Training and lectures			→ Access for industry			1	
 → User support → Publications 			→ Infrastructure ov	erview			
Focus on				News			
	nber 2017			News 12 December 20	17 <u>PRACE</u>	Digest 2017	
		manager VSC and ind	ustry @ FWO			Digest 2017 Ie Learning and Deep	Learning recordin



Home Ghent University Search



ngs

How could we further improve HPC-UGent services?

Documentation (21)

200 YEARS GHENT UNIVERSITY	HIGH	HIGH PERFORMANCE COMPUTING INFRASTRUCTURE							
	ACCESS	INFRASTRUCTURE	USER SUPPORT	TRAINING & LECTURES	INDUSTRY	ABOUT US			
Home > User support > Documentation									
Documentation									
Tutorial									

The HPC-UGent tutorial has been developed by the Flemish Supercomputer Center (VSC), and is tailored towards UGent researchers.

- → Linux edition
- → Mac edition
- → Windows edition

Please note that this tutorial is a constant work in progress. If you find any errors, please report them.



More information

GHENT

UNIVERSITY

- → HPC-UGent userwiki
- → VSCentrum.be user portal





How could we further improve HPC-UGent services?

User experience (21)

Shorter queue times (7)

Way to estimate queue time (4)

Longer wallclock time (4)

http://hpc.ugent.be/userwiki/index.php/User:Checkpointing

contact <u>hpc@ugent.be</u> if you have checkpointing issues

Feedback on efficiency of specific job (4)

Infrastructure (17)

Larger compute power (7)

GPU (4)

Debug infrastructure, remote visualization (2) UNIVFRSITY



How could we further improve HPC-UGent services?

Data (10)

More storage Easier sharing of data • Between VOs VSC DATA SHARED in https://account.vscentrum.be/django/vo/edit \circ Public \rightarrow VMs ? Staging in/out data Archive data of inactive users





How could we further improve HPC-UGent services?

<u>User support (3)</u> More HPC-UGent staff

Policy (4)

Collaboration within VSC

Training (7) More (5)

Security (6) ftp connectivity (5)?

Software (3)





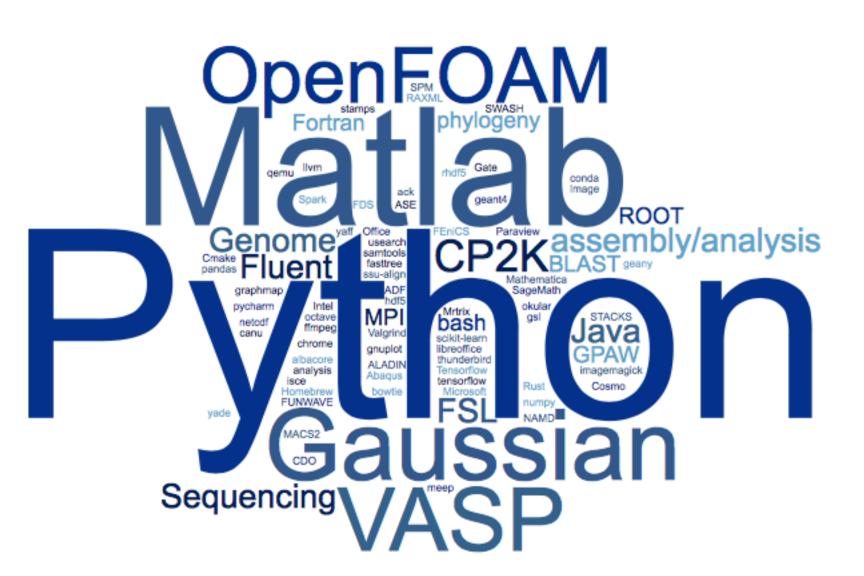
Keep services free of charge

Node sharing, data visible to other users

Form to request software install/update

Which software do you typically use?

Python	24
Matlab	15
R	14
Gaussian	9
VASP	8
OpenFOAM	7
C, C++	7
CP2K	4
Own code	4
FSL	3







Would your research benefit from specific IT hardware or services that HPC-UGent currently does not provide?

	#resp	Specific hard
No	52	Bigdata cl
The current compute platform suffices		Large, sha
I don't know	29	Very high
Yes	22	GPU
I need specific hardware/services		More core
Yes Ves Don't know	No	Other com Longer wa Personnel Additional



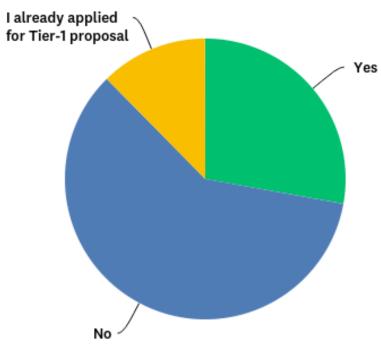
- ware/services suggested cluster with fast I/O ared storage (many TB) memory machine (1 TB)
- es / node
- mpilers than Intel
- allclock time
- el to support coding design
- I personnel for software installs

Would you be interested in applying for a Tier-1 proposal?

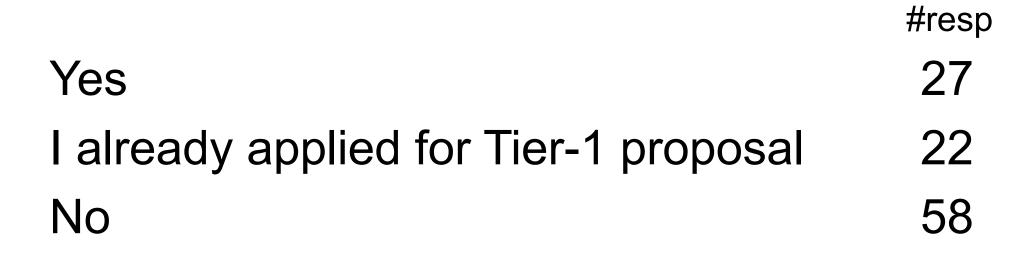
	#resp
Yes	27
I already applied for Tier-1 proposal	22
No	58

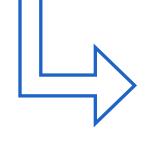






Would you be interested in applying for a Tier-1 proposal?





What is holding you back to submit a Tier-1 proposal? Why do you consider Tier-1 not suitable for your research?

Tier-2 suffices (33) Too much work (1) Lack of experience (3) Gaussian does not scale (1) Don't know what Tier-1 is (3)





I already applied for Tier-1 proposal Yes No

Additional cost (1)

Dr. Ewald Pauwels Scientific coordinator HPC @ Ghent University Vice-coordinator VSC

HPC-UGent

E hpc@ugent.be

www.ugent.be/hpc



RECEPTION: BEERS



Pils 13 Fresh Hop

The Ministry of Belgian Beers Pils (4% Alc. Vol) Zwaluw Siphon Brewing Session IPA (3,3% Alc. Vol)



Huldra

Brouwerij Totem Session IPA (2,7% Alc. Vol)

Kornkråke Double Dry Hopped

Brouwerij Totem Chili IPA (5,3% Alc. Vol)





Cendre

Siphon Brewing Black saison (6.5% Alc. Vol)



Narvi Brouwerij Totem Mandarine Radler (2,1% Alc. Vol)