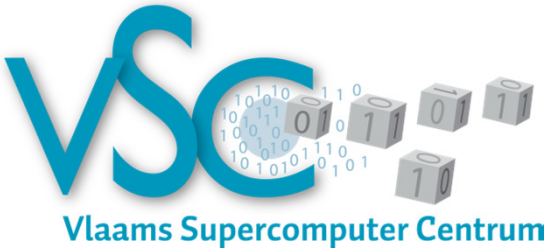
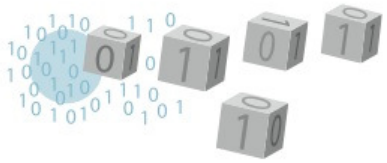


Tier-1b Info Session

UGent 23 June 2016

VSC HPC environment



PRACE

Tier - 0
47 Pf

VSC Vlaams Supercomputer Centrum

Tier - 1
623 Tf

- 16,240 CPU cores
- 128/256 GB memory/node
- IB EDR interconnect

Tier -2
510 Tf

Universiteit Antwerpen — HOPPER/TURING

UNIVERSITEIT GENT — STEVIN

KU LEUVEN universiteit hasselt — THINKING/CEREBRO

Vrije Universiteit Brussel — HYDRA

Tier -3

laptop, workstation

hadoop

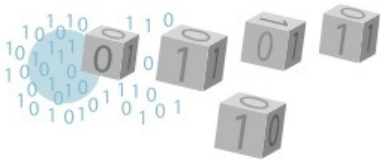
NVIDIA

intel inside Xeon Phi

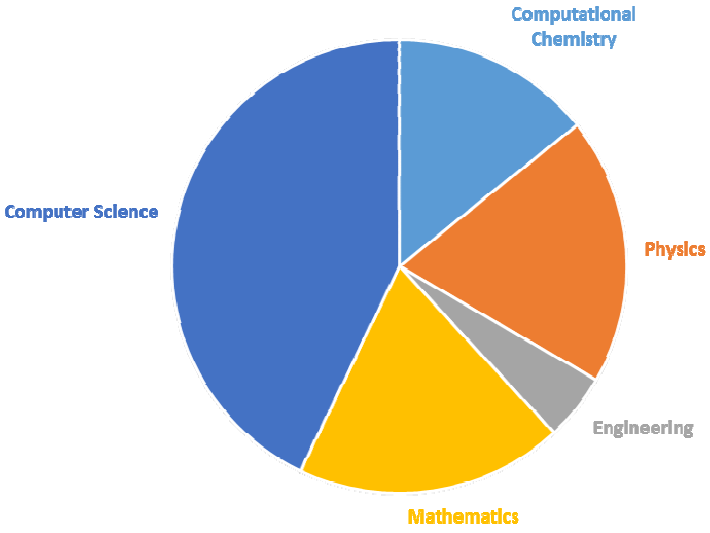
SGI® UV™

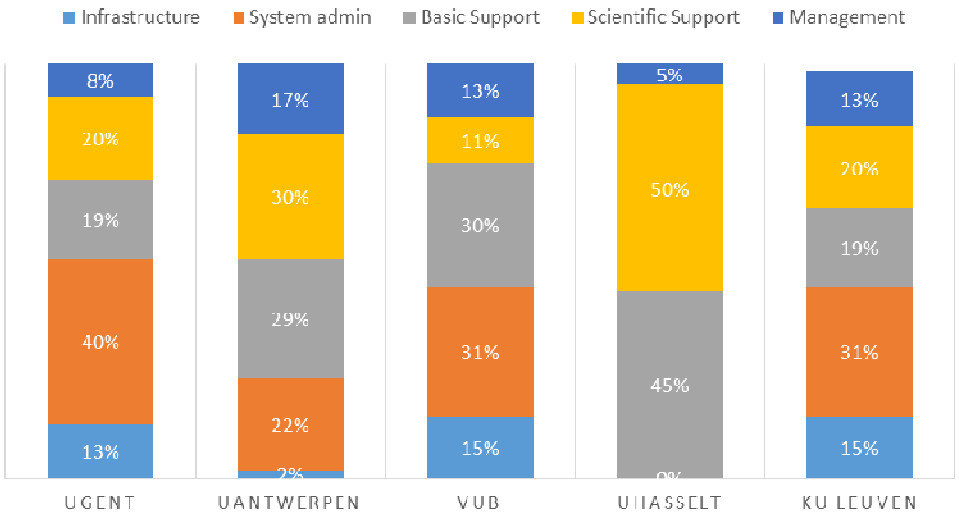
Specialized resources

VSC Staff



28 FTE



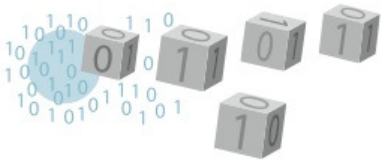


Tier-1 access

Who

Researchers of:

- University and association in the Flemish Community
- Research institute under authority or supervision of a university in the Flemish Community
- Strategic research center



What

- Validated research project
- Experience on Tier1/Tier2
- Technical feasibility
- Responsible for software licenses

When

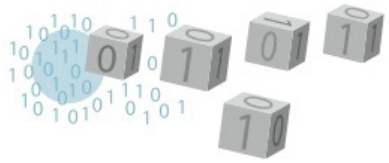
Proposals open continuously

allocation at specified dates

How

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

Are you a Tier-1 candidate?



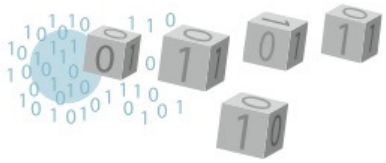
- Do you have already experience with HPC infrastructure ?

- Is your code already running on a HPC infrastructure ?
- Is your code optimized and scalable?

- Do you need lots of compute time ?
- Will it take too long to get the wanted results using Tier-2 ?

- Do you want to simulate more complex, bigger systems?
- Do you need your results faster?

Project applications



Computing resources

- Total node days: 500 – 5,000
- Single job max walltime: 3 days
- Maximum for 6 months
- SCRATCH: 250 GB-100,000 files
- Memory: 128 GB/node 256GB/node

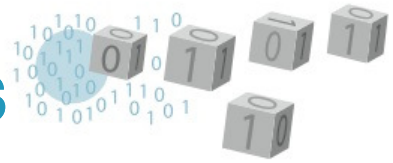
Software

- Software already available at Tier-1
- Software not yet available:
- is suitable for Tier-1
- you have a valid license
  - Installation on Tier-1 @ KUL
  - enough licenses

Funding

- Computer time is FREE

Starting grant/preparatory access



Computing resources

- Allocations are personal
- Total node days: 100
- Maximum for 2 months
- Single job walltime: 3 days
- SCRATCH: 250 GB-100,000 files
- Memory: 128 GB/node 256GB/node

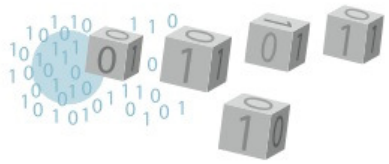
Submissions at any time

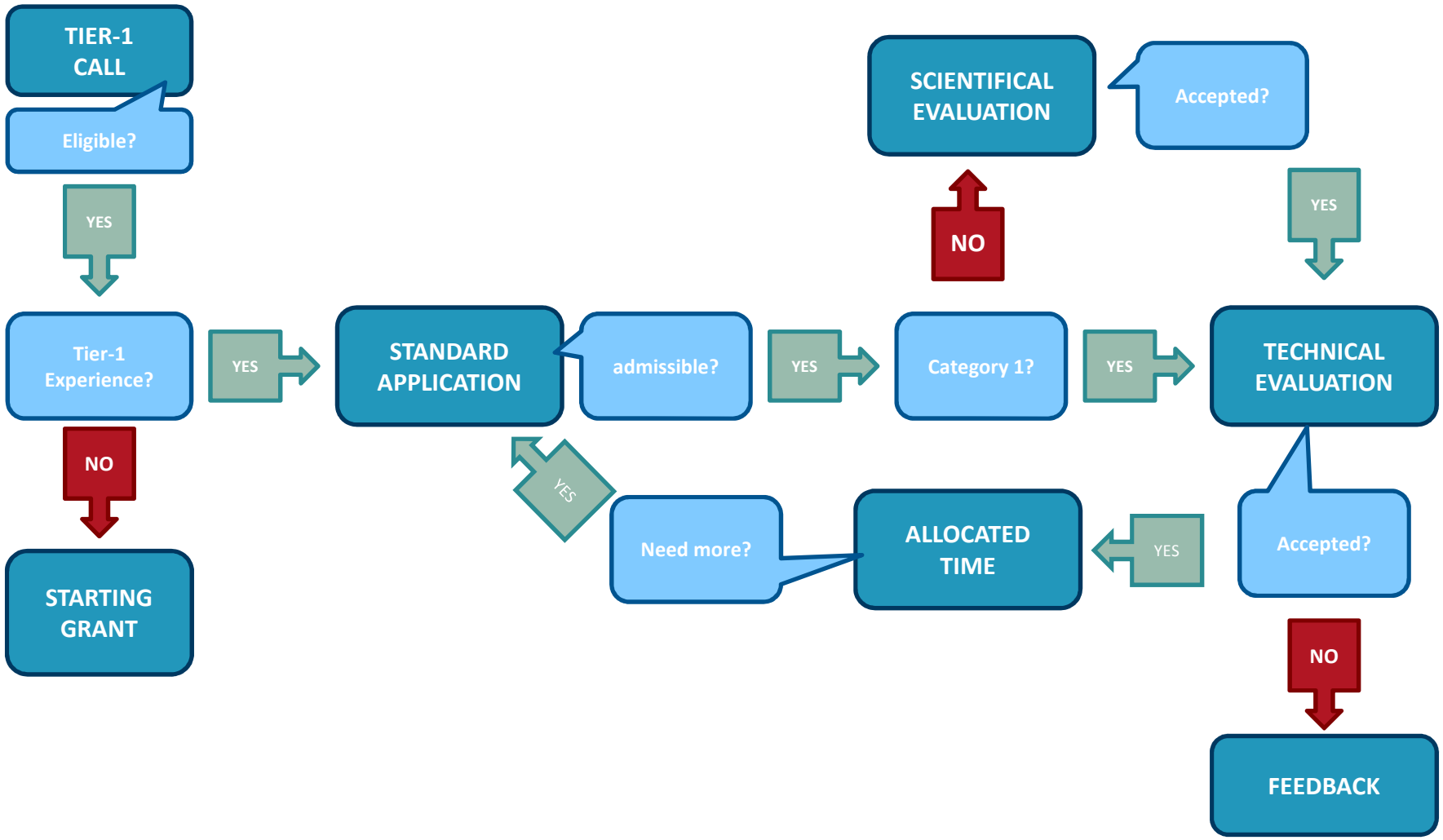
Allocated time free

Fast submission procedure

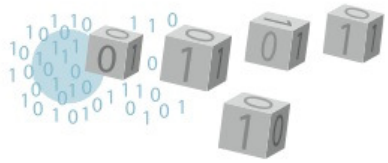
Preparation for project application

Tier-1 application workflow



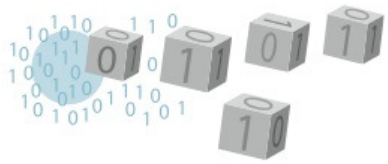


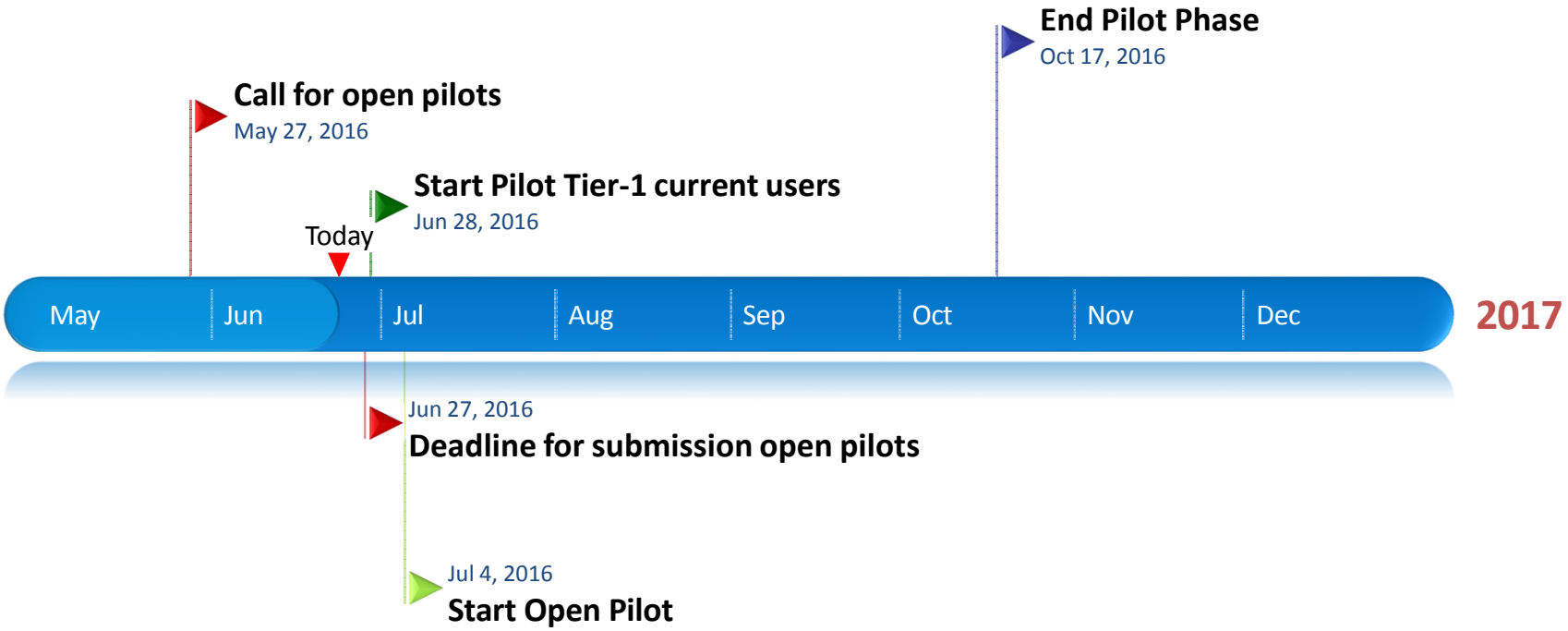
Evaluation committee



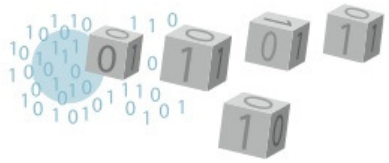
- a Policy Officer of FWO
- at least three experts appointed by the Board of Directors who:
  - are not active within the Flemish Community
  - have wide experience in the field of using large computing capacity

Pilot plan





- Open to all researchers affiliated with a Flemish university or research institute
- Proven HPC experience needed
- More information: https://www.vscentrum.be/assets/1087

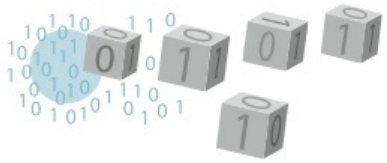


Tier-1b system specs

BrENIAC

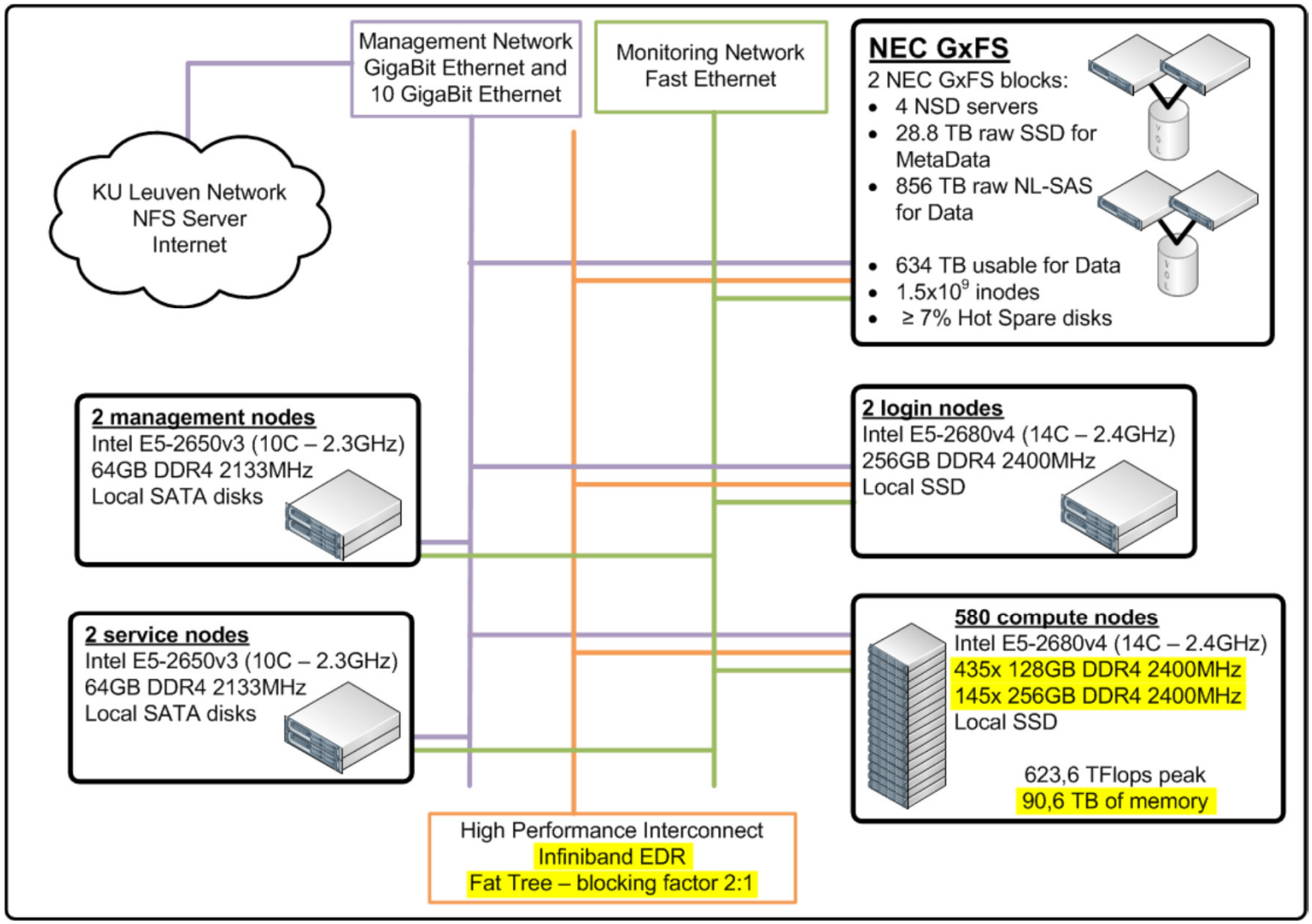
Brain + ENIAC

Top500 list June 2016

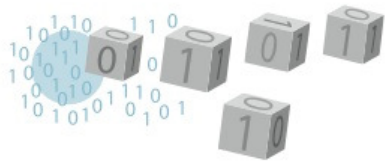


		Sugon				
193	China Unicom China	Lenovo NeXtScale nx360M5, Intel Xeon E5-2660v2 10C 2.2GHz, 10G Ethernet Lenovo	78,000	550.1	1,372.8	
194	NASA/Goddard Space Flight Center United States	**Discover SCU12** - Rackable Cluster, Xeon E5-2697v3 14C 2.6GHz, Infiniband FDR SGI	17,136	548.7	712.9	
195	NASA/Goddard Space Flight Center United States	**Discover SCU11** - Rackable Cluster, Xeon E5-2697v3 14C 2.6GHz, Infiniband FDR SGI	17,136	548.7	712.9	
196	Flemish Supercomputer Center Belgium	**BrENIAC** - NEC HPC1816Rg, Xeon E5-2680v4 14C 2.4GHz, Infiniband EDR NEC	16,128	548.0	619.3	255
197	Hosting Services United States	Cluster Platform DL380p Gen8, Intel Xeon E5-2650v2 8C 2.6GHz, 10G Ethernet Hewlett-Packard	44,832	547.3	932.5	
198	Service Provider United States	Cluster Platform 3000 BL460c Gen8, Xeon E5-2670 8C 2.600GHz, 10G Ethernet Hewlett-Packard	42,848	545.5	891.2	
199	China Unicom China	Lenovo x3950, Xeon E7-8860v3 16C 2.2GHz, 10G Ethernet Lenovo	24,576	545.0	865.1	
200	High Energy Accelerator Research Organization /KEK Japan	**SAKURA** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	49,152	536.7	629.1	247

BrENIAC architecture

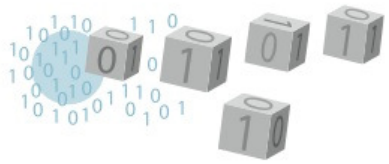


System comparison



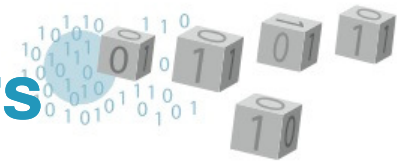
	MUK	BrENIAC	
Total nodes	528	580	
Processor type	Sandy Bridge E5-2670	Broadwell E5-2680v4	
Base Clock Speed	2.6 GHz	2.4 GHz	
Cores per node	16	28	
Total cores	8,448	16,240	x2
Memory per node (GB)	64	435x128/145x256	
Memory per core (GB)	4	4.5/9.1	x2
Local Disk	500 GB SATA	128GB SSD	
Peak performance (TF)	175	623	
Network	Infiniband FDR fat tree non-blocking	Infiniband EDR 2:1	x2

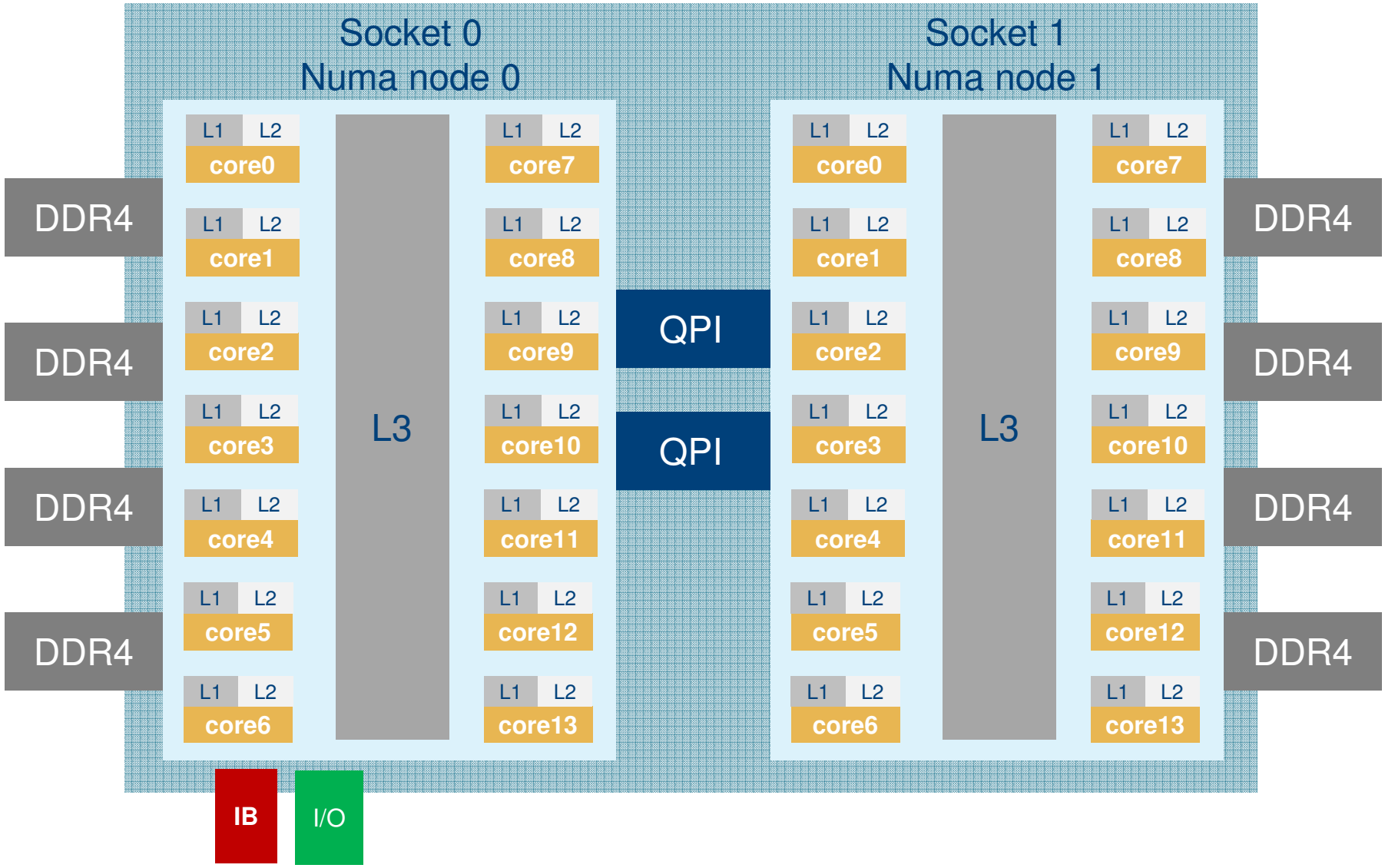
System comparison(2)



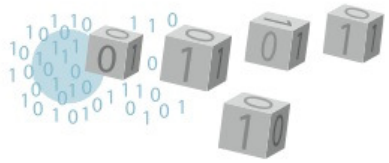
	Swallot	Golett	BrENIAC
Total nodes	128	200	580
Processor type	Haswell E5-2669v3	Haswell E5-2680v3	Broadwell E5-2680v4
Base Clock Speed	2.6 GHz	2.5 GHz	2.4 GHz
Cores per node	20	24	28
Total cores	2,560	4,800	16,240
Memory per node (GB)	128	64	435x128/145x256
Memory per core (GB)	6.4	2.6	4.5/9.1
Local Disk	1 TB SATA	500GB SAS	128GB SSD
Peak performance (TF)	106	192	623
Network	Infiniband FDR	Infiniband FDR-10	Infiniband EDR 2:1

Compute nodes: Broadwell processors

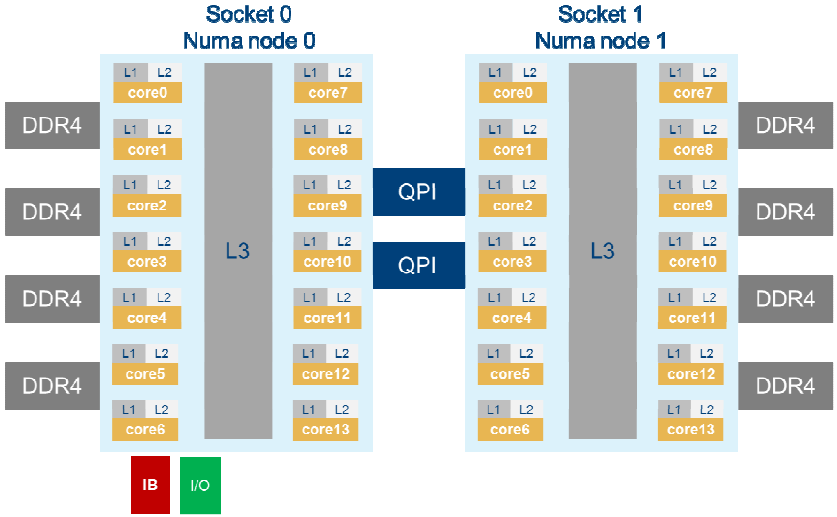




Processor comparison

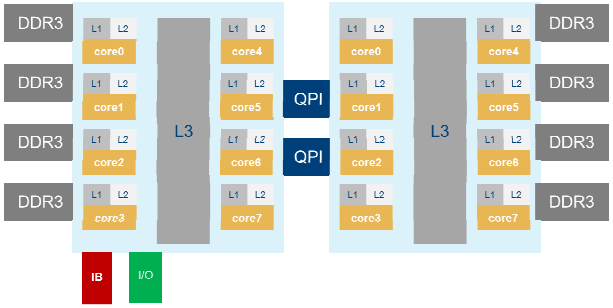


Broadwell



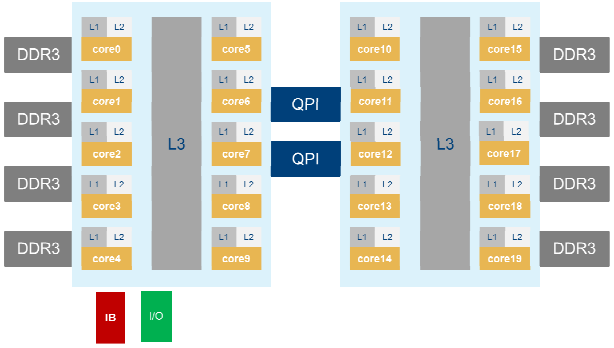
DDR4 2400 MHz
AVX2
16 DP FLOPs/cycle
two 4-wide FMA instructions
Cluster on Die –2 NUMA nodes x socket

Sandy Bridge



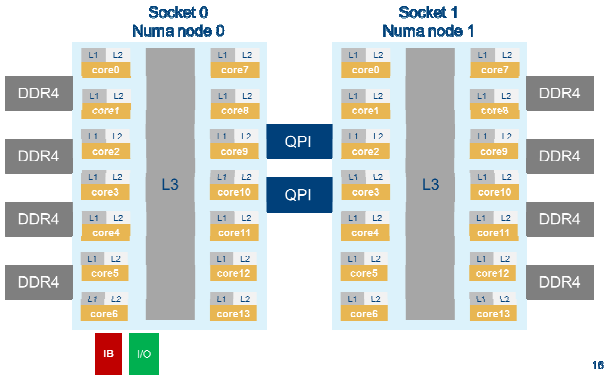
DDR3 1600 MHz
AVX
8 DP FLOPs/cycle
4-wide AVX add +4-wide AVX mult

Haswell 10c



DDR4 2133 MHz
AVX2
16 DP FLOPs/cycle
two 4-wide FMA instructions

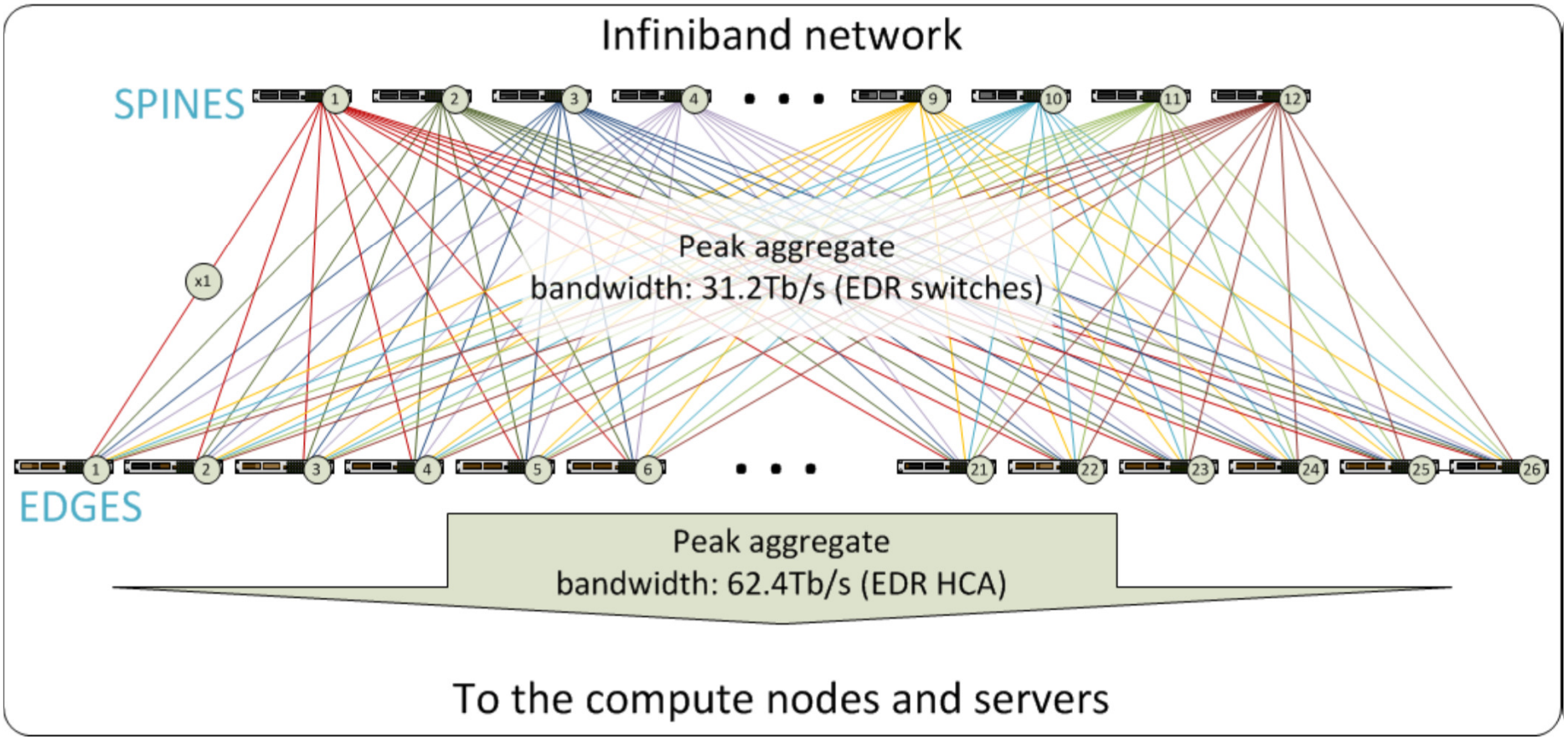
Haswell 12c



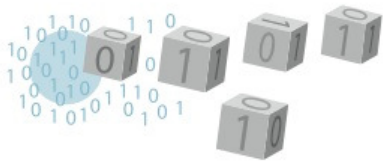
DDR4 2133 MHz
AVX2
16 DP FLOPs/cycle
two 4-wide FMA instructions

Network

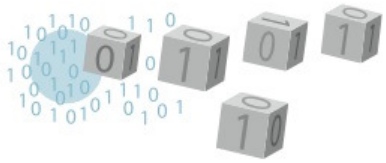
Mellanox Infiniband EDR
Fat tree with blocking factor 2:1



Infiniband Islands



How to start?



- Do you have a VSC-number account?
- Do you have an active Tier-1 project?
- Your VSC-number credentials are known!

- ssh vscXXXXX@login1-tier1.hpc.kuleuven.be

```
login1-tier1.hpc.kuleuven.be - PuTTY
login as: vsc30706
Authenticating with public key "ingrid@office" from agent

Informatie over deze server: login1.tier1.hpc.kuleuven.be
 * cluster: tier1/2016
 * role: login
 * hardware: X10DRi (x86_64)
 * os: CentOS 7.2.1511

: vsc30706@login1 ~ 17:20 $
```

… and you are in!

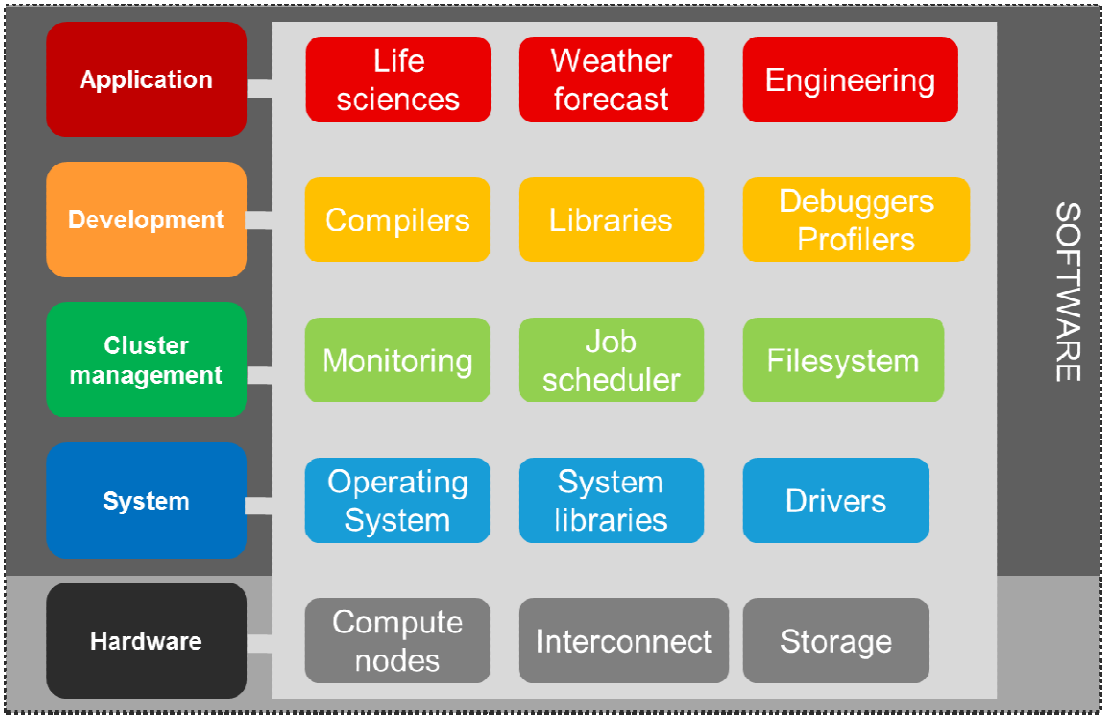
2 login nodes:
login1-tier1.hpc.kuleuven.be
login2-tier1.hpc.kuleuven.be

Adding visualization features?

VSC Common User Environment

Make user's life easy

- Unify VSC cluster environments
- Based in an aggregated layered approach

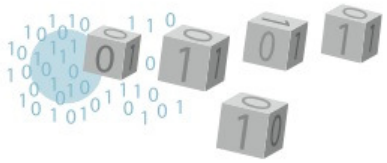


Vlaams Supercomputer Centrum

ASSOCIATIE KU LEUVEN · ASSOCIATIE UNIVERSITEIT GENT · universitaire associatie BRUSSEL · auha · UHasselt

ASSOCIATIE KU LEUVEN · ASSOCIATIE UNIVERSITEIT GENT · universitaire associatie BRUSSEL · auha · UHasselt

CUE base configuration



Application

python

Perl

R

Development

Toolchains

GCC

intel

Cluster management

Adaptive COMPUTING

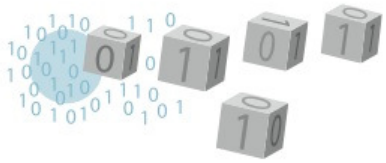
queues
Commands
support scripts
node features

IBM GPFS

System

Module system
Environment variables: $VSC_XXXXI
login.hpc.site.be

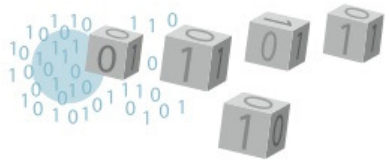
Storage areas



Name	Variable	Type	Access	Backup	Quota
/user/leuven/30X/vsc30XXX	$VSC_HOME	NFS	Global	YES	3 GB
/data/leuven/30X/vsc30XXX	$VSC_DATA	NFS	Global	YES	75 GB
/scratch/leuven/30X/vsc30XXX	$VSC_SCRATCH $VSC_SCRATCH_SITE	GPFS	Global	NO	XXX GB
/node_scratch	$VSC_SCRATCH_NODE	ext4	local	NO	128 GB

- Only data from $VSC_SCRATCH needs to be migrated
- SCRATCH is not removed until the end of the project
- After the project there is 14 days grace

Toolchains

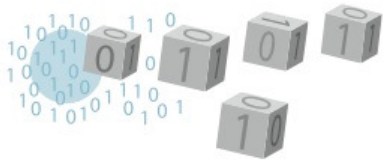


	Intel compilers	Open source
Name	intel	foss
version	2016a	2016a
Compilers	Intel compilers icc, icpc, ifort	GNU compilers gcc, g++, gfortran
MPI Library	Intel MPI	OpenMPI
Math libraries	Intel MKL	OpenBLAS, LAPACK FFTW ScaLAPACK

Intel Toolchain has newer versions!

	VSC	BrENIAC
Icc, icpc, ifort	16.0.1.150 Build 20151021	16.0.3.210 Build 20160415
Intel MKL	11.3.1.150	11.3.3.210
Intel MPI	5.1.2.150	5.1.3.181

Software

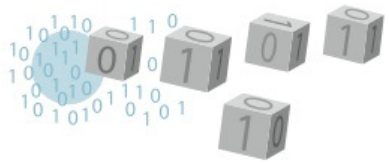


- Base software is already installed

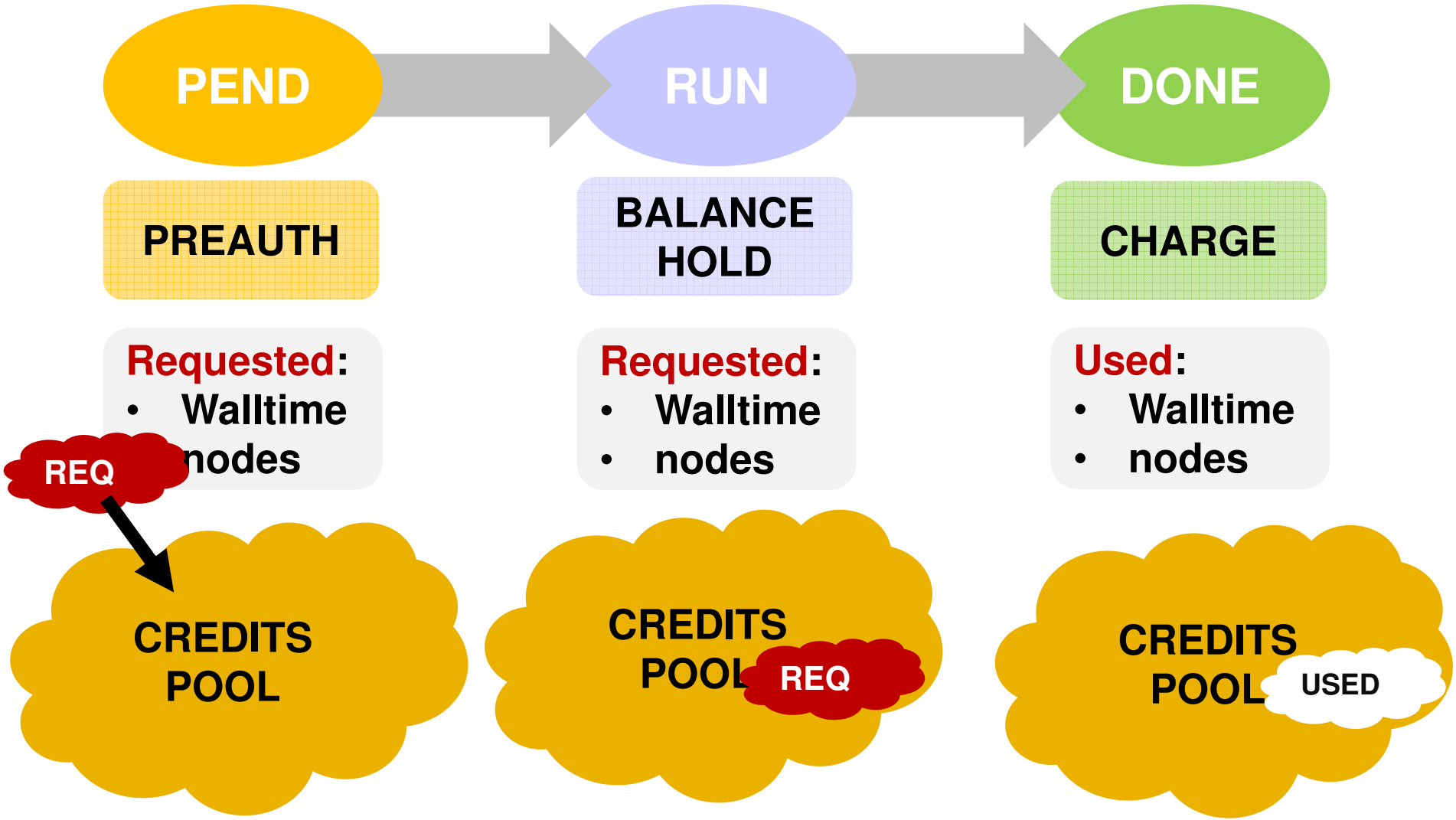
```
------------------------------------------- /apps/leuven/broadwell/2016a/modules/all ---------------------
AllineaForge/6                                  Valgrind/3.11.0-GCCcore-4.9.3
Autoconf/2.69                                   binutils/2.25
Automake/1.15                                   binutils/2.25-GCCcore-4.9.3
Autotools/20150215                              bzip2/1.0.6-GCC-4.9.3-2.25
Bison/3.0.4                                     expat/2.1.1-GCCcore-4.9.3
Bison/3.0.4-GCCcore-4.9.3                       flex/2.5.39
CMake/3.5.2-GCC-4.9.3-2.25                      flex/2.5.39-GCCcore-4.9.3
FFTW/3.3.4-gompi-2016a                          foss/2016a
GCC/4.9.3-2.25                                  gompi/2016a
GCCcore/4.9.3                                   hwloc/1.11.2-GCC-4.9.3-2.25
GDB/7.11.1-GCCcore-4.9.3                        icc/2016.3.210-GCC-4.9.3-2.25
GMP/6.1.0-foss-2016a                            iccifort/2016.3.210-GCC-4.9.3-2.25
GMP/6.1.0-intel-2016a                           ifort/2016.3.210-GCC-4.9.3-2.25
Java/1.8.0_92                                   iimpi/2016.03-GCC-4.9.3-2.25
M4/1.4.17                                       imkl/11.3.3.210-iimpi-2016.03-GCC-4.9.3-2.25
M4/1.4.17-GCCcore-4.9.3                         impi/5.1.3.181-iccifort-2016.3.210-GCC-4.9.3-2.25
NASM/2.11.08-foss-2016a                         intel/2016a
NASM/2.11.08-intel-2016a                        intel/2016a-GCC-4.9.3
OpenBLAS/0.2.15-GCC-4.9.3-2.25-LAPACK-3.6.0     libjpeg-turbo/1.4.2-foss-2016a
OpenMPI/1.10.2-GCC-4.9.3-2.25                   libjpeg-turbo/1.4.2-intel-2016a
Perl/5.22.1-foss-2016a-bare                     libpng/1.6.21-foss-2016a
Perl/5.22.1-intel-2016a-bare                    libpng/1.6.21-intel-2016a
Python/2.7.11-GCC-4.9.3-2.25-bare               libreadline/6.3-GCCcore-4.9.3
R/3.2.5-foss-2016a-bare                         libtool/2.4.6
R/3.2.5-intel-2016a-bare                        ncurses/6.0-GCCcore-4.9.3
SQLite/3.9.2-GCC-4.9.3-2.25                     numactl/2.0.11-GCC-4.9.3-2.25
ScaLAPACK/2.0.2-gompi-2016a-OpenBLAS-0.2.15-LAPACK-3.6.0 zlib/1.2.8
Tcl/8.6.4-GCC-4.9.3-2.25                        zlib/1.2.8-GCCcore-4.9.3
Tk/8.6.4-GCC-4.9.3-2.25-no-X11
```

More to come!

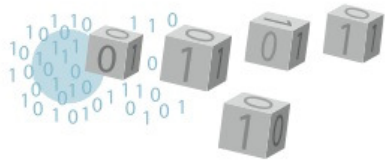
Accounting



Moab Accounting Manager (MAM) = credit card payment

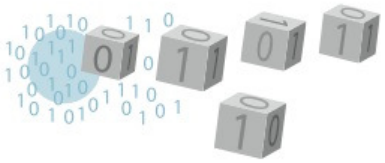


Accounting commands



- Projects : lt1_projectname
- Credits pool = assigned node/days per project
- Module load accounting
  - gbalance -> check your account status
  - gquote -> check how much credits the job will cost
  - gstatement -> detailed overview
  - glsjob -> account information about a job

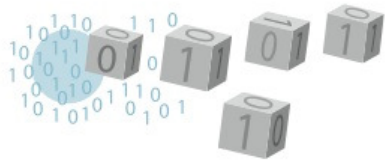
Accounting: gbalance



login.muk.gent.vsc - PuTTY

```
-bash-4.1$ gbalance
  Id   Name                  Balance Reserved Effective CreditLimit Available
---- -------------------- ------- -------- --------- ----------- ---------
 894  lt1_sys                 2630        0      2630           0      2630
 918                          4492        0      4492           0      4492
1550 Account=lt1_bench 9989        0      9989           0      9989
```

Submitting jobs



- Torque/Moab will be used for scheduling
- Single job policy
  - Use worker framework for single core jobs

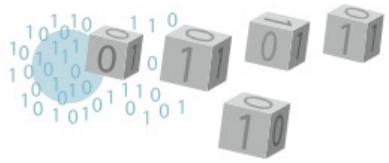
    https://www.vscentrum.be/cluster-doc/running-jobs/worker-framework
- As always: specify resources
  - -lwalltime=4:30:00 (job will last 4h 30 min)
  - -lnodes=2:ppn=28 (job needs 2 nodes and 20 cores per node)
  - -lpmem=4gb (job request 4 GB of memory per core)

Other node features: memory.

- -lfeature=mem128/mem256

qsub –lnodes=10:ppn=28,walltime=1:00:00,pmem=4gb my_job.pbs

Migrating PBS scripts

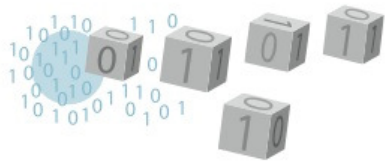


login.muk.gent.vsc - PuTTY

```
-bash-4.1$
#!/bin/bash -l
#PBS -N HPL-2.2-T1-SB-64-1n
#PBS -lnodes=1:ppn=16
#PBS -lwalltime=24:00:00
#PBS -lvmem=55gb


module load intel/2016a
module load scripts
cd $HOME/mydatadir

OUTDIR=$VSC_SCRATCH/myresdir
mkdir -p $OUTDIR
cd $OUTDIR



#execute
mympirun --stats=3 ./xhpl.sb.intel2016a
```

```
#PBS –lnodes=1:ppn=28
#PBS -lwalltime=10:00:00
#PBS –lpmem=4gb
```

Migrating PBS scripts



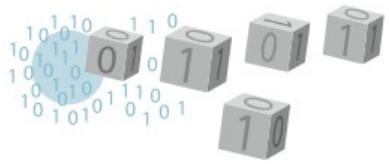
login.muk.gent.vsc - PuTTY

```
-bash-4.1$
#!/bin/bash -l
#PBS -N HPL-2.2-T1-SB-64-1n
#PBS -lnodes=1:ppn=16
#PBS -lwalltime=24:00:00
#PBS -lvmem=55gb


module load intel/2016a
module load scripts
cd $HOME/mydatadir

OUTDIR=$VSC_SCRATCH/myresdir
mkdir -p $OUTDIR
cd $OUTDIR



#execute
mympirun --stats=3 ./xhpl.sb.intel2016a
```

```
#PBS –lnodes=1:ppn=28
#PBS -lwalltime=10:00:00
#PBS –lpmem=4gb
```

```
#PBS –A lt1_project
```

Migrating PBS scripts



login.muk.gent.vsc - PuTTY

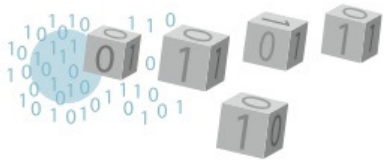
```
-bash-4.1$
#!/bin/bash -l
#PBS -N HPL-2.2-T1-SB-64-1n
#PBS -lnodes=1:ppn=16
#PBS -lwalltime=24:00:00
#PBS -lvmem=55gb


module load intel/2016a
module load scripts
cd $HOME/mydatadir

OUTDIR=$VSC_SCRATCH/myresdir
mkdir -p $OUTDIR
cd $OUTDIR



#execute
mympirun --stats=3 ./xhpl.sb.intel2016a
```

```
#PBS –lnodes=1:ppn=28
#PBS -lwalltime=10:00:00
#PBS –lpmem=4gb
```

```
#PBS –A lt1_project
```

Remove!

Migrating PBS scripts



login.muk.gent.vsc - PuTTY

```
-bash-4.1$
#!/bin/bash -l
#PBS -N HPL-2.2-T1-SB-64-1n
#PBS -lnodes=1:ppn=16
#PBS -lwalltime=24:00:00
#PBS -lvmem=55gb


module load intel/2016a
module load scripts
cd $HOME/mydatadir

OUTDIR=$VSC_SCRATCH/myresdir
mkdir -p $OUTDIR
cd $OUTDIR


#execute
mympirun --stats=3 ./xhpl.sb.intel2016a
```

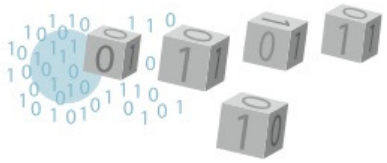
#PBS –lnodes=1:ppn=28
#PBS -lwalltime=10:00:00
#PBS –lpmem=4gb

#PBS –A lt1_project

Not available!

mpirun –np $PBS_NP ./xhpl.sb.intel2016a
export I_MPI_STATS=ipm

Migrating PBS scripts



login.muk.gent.vsc - PuTTY

```
-bash-4.1$
#!/bin/bash -l
#PBS -N HPL-2.2-T1-SB-64-1n
#PBS -lnodes=1:ppn=16
#PBS -lwalltime=24:00:00
#PBS -lvmem=55gb


module load intel/2016a
module load scripts
cd $HOME/mydatadir

OUTDIR=$VSC_SCRATCH/myrecdir
mkdir -p $OUTDIR
cd $OUTDIR


#execute
mympirun --stats=3 ./xhpl.sb.intel2016a
```

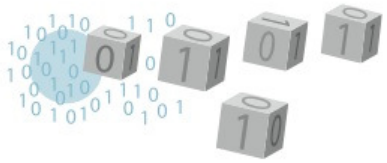
Not available!

```
#PBS –lnodes=1:ppn=28
#PBS -lwalltime=10:00:00
#PBS –lpmem=4gb
```

```
#PBS –A lt1_project
```

```
export I_MPI_STATS=ipm
mpirun –np $PBS_NP ./xhpl.sb.intel2016a
```

Migrating PBS scripts



login.muk.gent.vsc - PuTTY

```
-bash-4.1$
#!/bin/bash -l
#PBS -N HPL-2.2-T1-SB-64-1n
#PBS -lnodes=2:ppn=16
#PBS -lwalltime=24:00:00
#PBS -lvmem=55gb


module load intel/2016a
module load scripts
cd $HOME/mydatadir

OUTDIR=$VSC_SCRATCH/myresdir
mkdir -p $OUTDIR
cd $OUTDIR

#execute
Export OMP_NUM_THREADS=16
mympirun –universe= 2 --stats=3 ./xhpl.sb.intel2016a
```

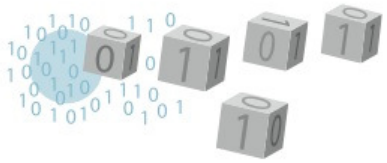
```
#PBS –lnodes=1:ppn=28
#PBS -lwalltime=10:00:00
#PBS –lpmem=4gb
```

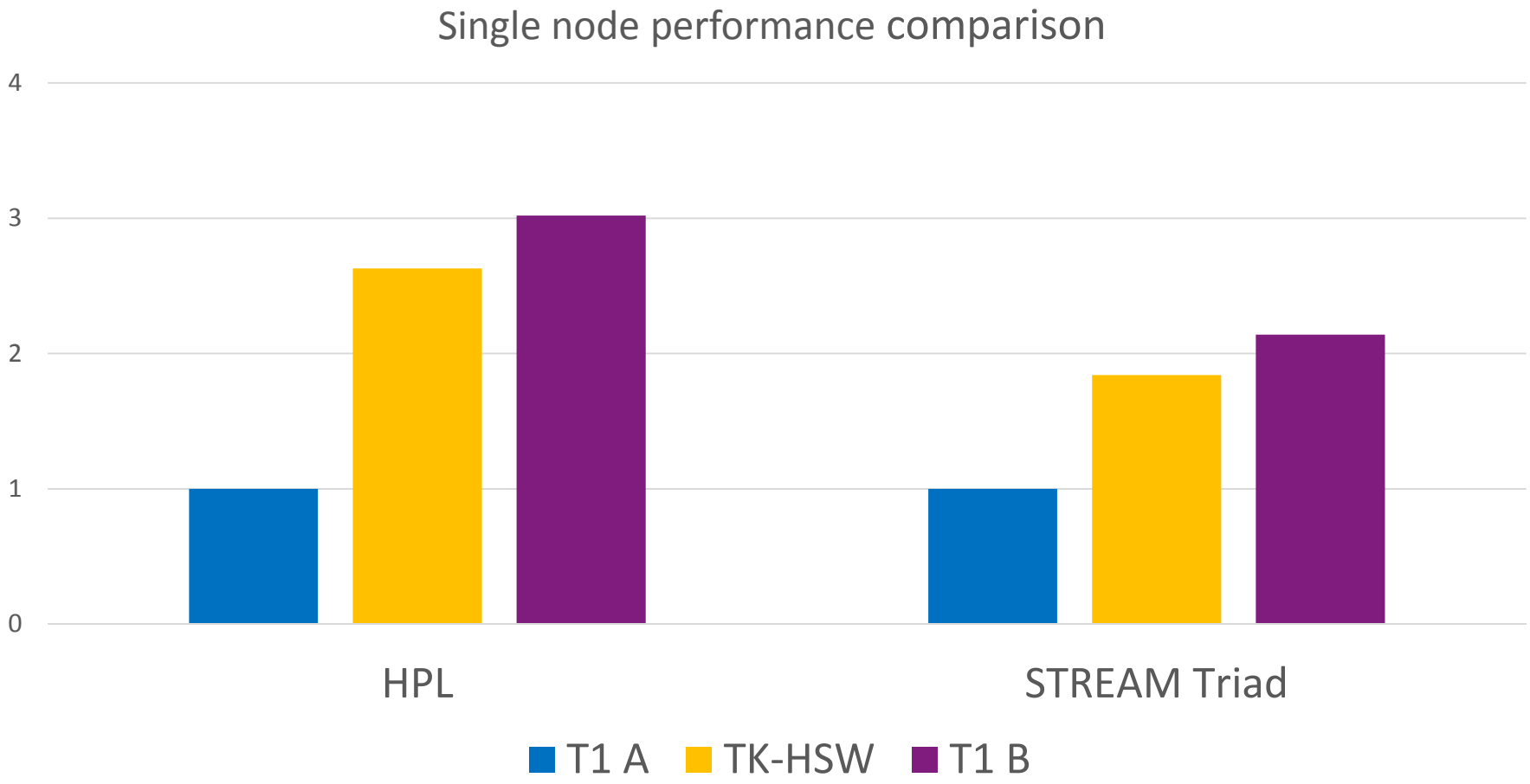
```
#PBS –A lt1_project
```

Hybrid MPI + OpenMP

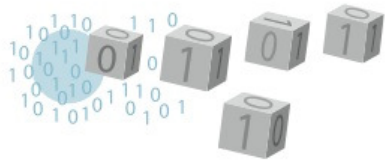
```
export I_MPI_JOB_RESPECT_PROCESS_PLACEMENT=disable
export I_MPI_PIN_DOMAIN=core
export KMP_AFFINITY=compact,1
export I_MPI_STATS=ipm
mpirun –np $PBS_NP ./xhpl.sb.intel2016a
```

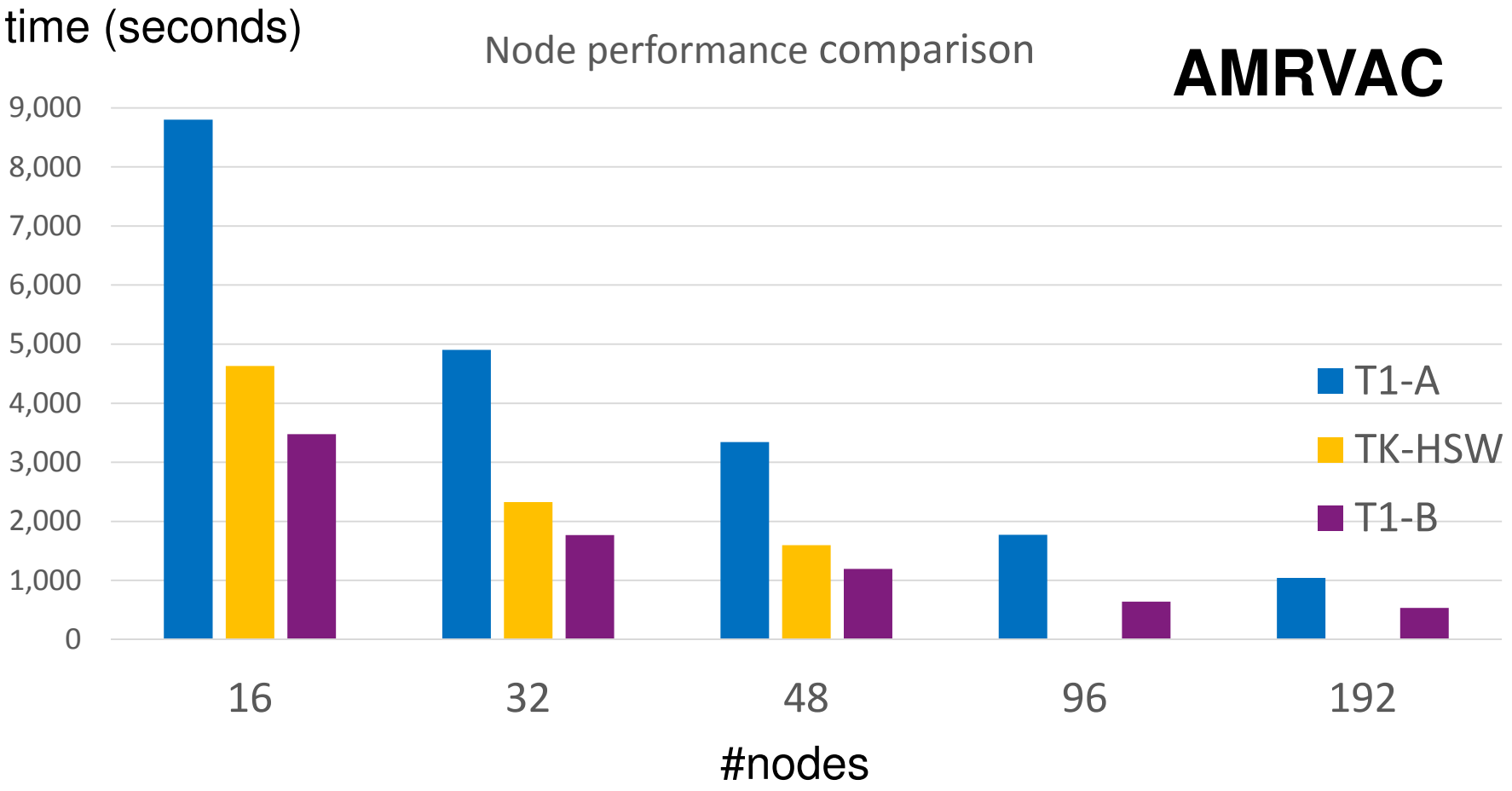
What could you expect?



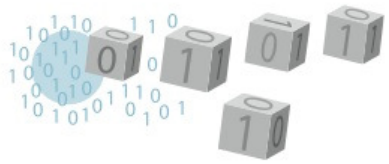


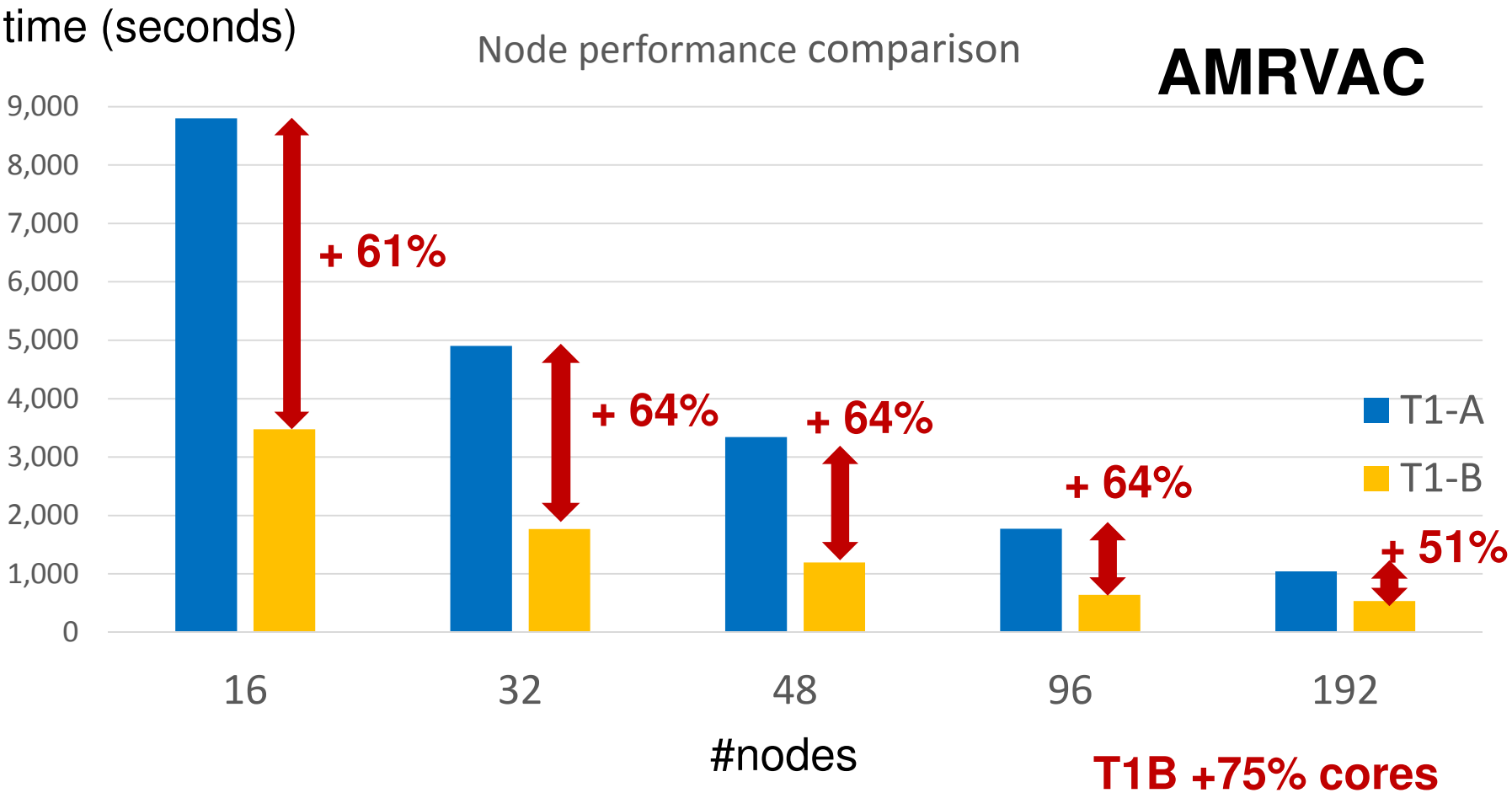
What could you expect?





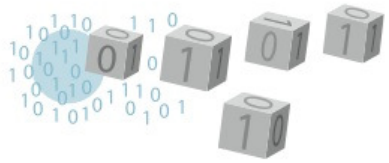
What could you expect?

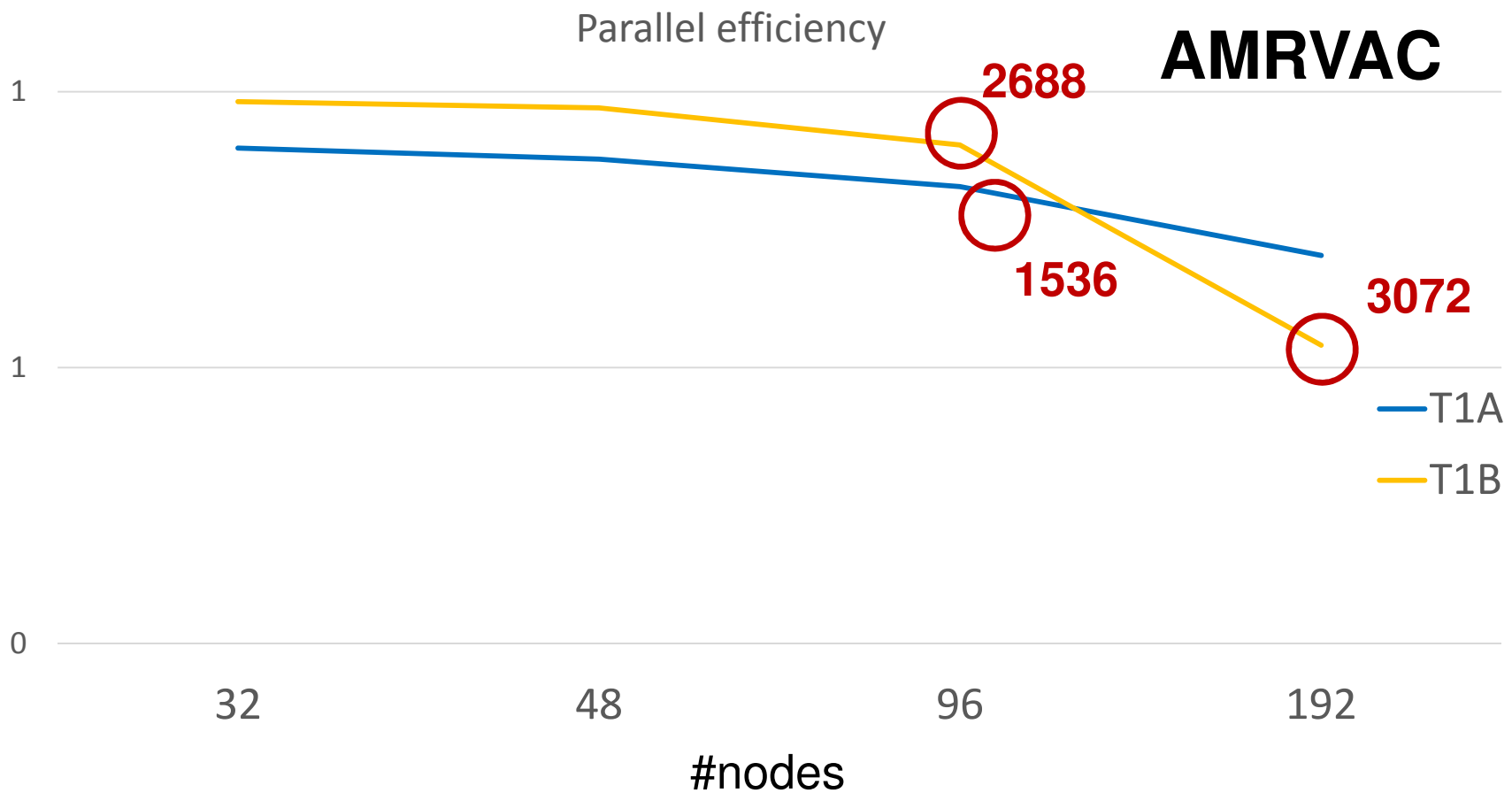




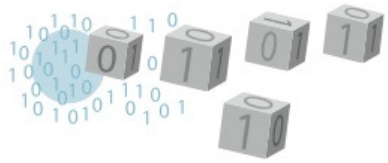
T1B +75% cores
T1A +8% freq
System size

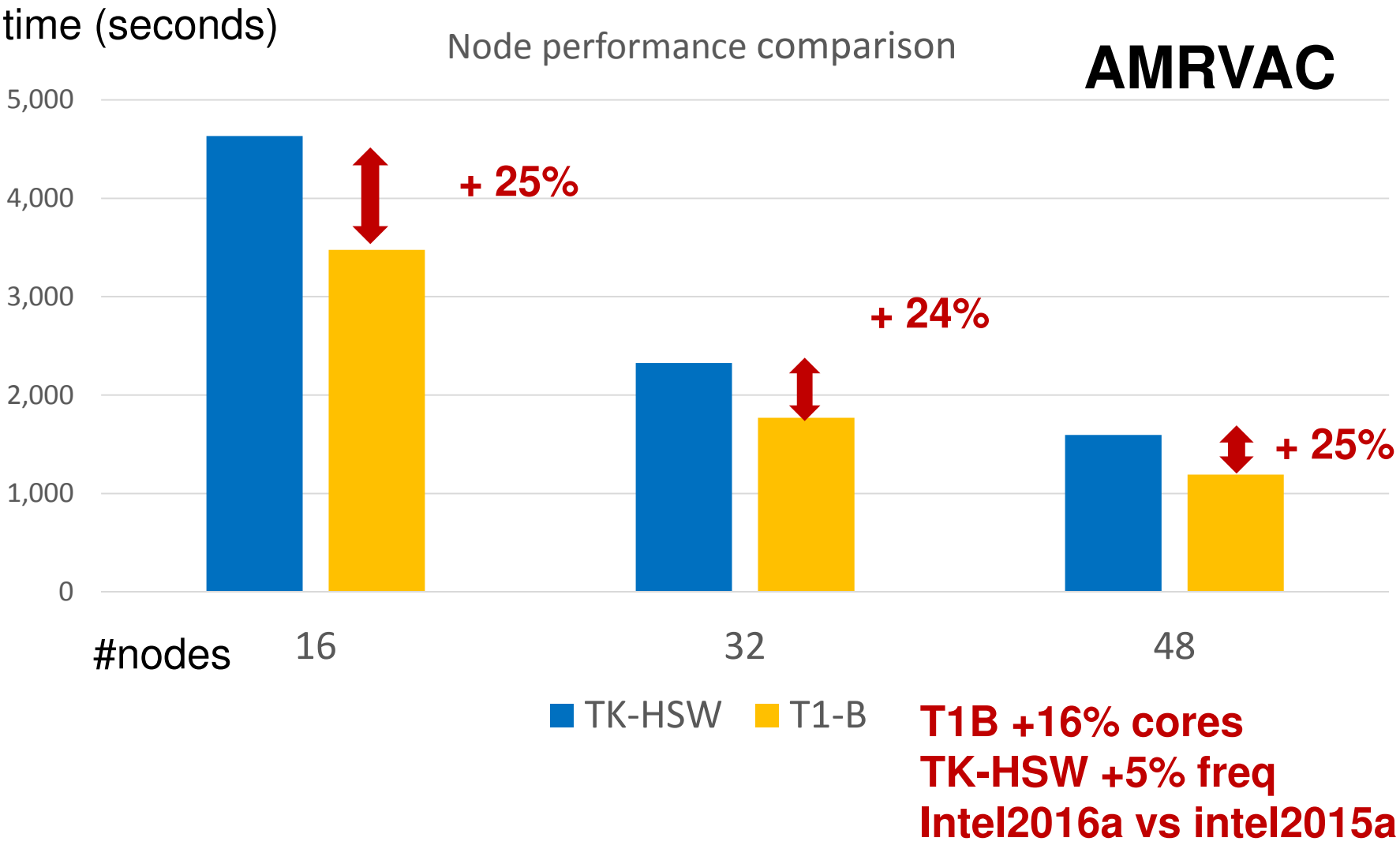
What could you expect?





What could you expect?





Jan Ooghe
Ewald Pauwels
Ingrid Barcena

hpcinfo@kuleuven.be