HP-SEE

HPC cluster at IICT-BAS and HP-SEE infrastructure

www.hp-see.eu

Emanouil Atanassov
Grid Technologies and Applications
Institute of Information and Communication Technologies
Bulgarian Academy of Science
emanouil@parallel.bas.bg

HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities

The HP-SEE initiative is co-funded by the European Commission under the FP7 Research Infrastructures contract no. 261499

Outline

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Objectives of HP-SEE operations
- Regional HPC Infrastructure – present and future
- High Performance Cluster at IICT-BAS
- Obtaining access
- Tools and services
- Conclusions

Objectives of HP-SEE operations

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- **WP5 – Regional HPC operations and interoperation** – will ensure that integrated services will be provided to end users, by deploying the regional HPC infrastructure on top of the existing networking infrastructure, complementary to the existing Grid infrastructure, and fused with end user services.
- Specifically, user community needs in terms of size and availability of HPC resources will be catered for. Current HPC installations in Bulgaria, Romania and Hungary will be integrated at the first stage, followed by the upcoming purchases in Greece and Serbia.
- These will form the backbone of the state-of-the-art regional infrastructure, which will be operated jointly by the project, using a set of operations and management tools that will be supported by WP5.
- The infrastructure will be open for use of the wider user community from the region, not only infrastructure provision countries

Current infrastructure resources

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

Country	Bulgaria	Bulgaria	Romania
Available project month	PM1	PM1	PM1
Administrative Data			
System No / Vendor	1 / HP	2 / IBM	1 / IBM
System Name	HPCG	BG-FEN	InfraGRID
System Short Description	High performance cluster at IICT-BAS	IBM Blue Gene/P	High performance cluster at WUT
Computational Power			
Number of nodes	**36**	2048	50
CPU	Intel Xeon X5560 @2.8Ghz	PowerPC 450 processors (32 bits, 850 MHz)	Intel Xeon E5504 @2.00Ghz
RAM	24 GB per node	4 GB per node	10GB per node
Max number of parallel processes	**576 cores**	**8192 cores**	**400 cores**
Interconnect type	DDR Infiniband		QDR 4x Infiniband
Interconnect latency	2.5 µs		1.1µs
Interconnect bandwidth	20 Gbps		40Gbps
Peak performance (Tflops, double precision)	**3.23**	**27.85**	**2.5**
Achieved performance (Tflops, double precision)	**3**	**23.42**	**2.15**
Operating system	Scientific Linux 5.3 64 bit		
Batch system	torque + maui		
Storage			
Available filesystems	/home, /gscratch		
Total storage on each filesystem	20TB, 10TB		10TB
Maximum throughput on each filesystem	600MB/s, 600 MB/s		500MB/s
Software			
Libraries	BLAS, LAPACK		
Development and application software available	MVIAPICH 1, MVIAPICH 2, OPENMPI, gcc, gfortran, etc.		MPICH, MPICH2, OpenMPI, gcc, gfortran, etc.

Current infrastructure resources

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

Romania	Romania PM1	Hungary	Serbia
PM1		PM1	PM1
3/IBM			
2 / IBM		1/SUN Microsystems	1 / E4
NCIT-Cluster	IBM BladeCenter	NIIFI Supercomputing Center	PARADOX
High performance cluster at UPB		SUN E15K supercomputer cluster at NIIFI's site	High performance cluster at IPB
26 (16xQS22 + 10xLS22)			
48	QS22: 2x IBM PowerXCell 8i @3.2GHz; LS22: 2x AMD Opteron Quad Core 2376 HE @2.3GHz	2	84
28*HS21 Dual Intel Quad-Core Xeon E5504 @2.00Ghz, 4*QS22 Dual IBM PowerXCell 8i @3.2GHz, 20*LS22 Dual AMD Opteron Hex-Core @2.6GHz	32 GB per QS22; 8GB per LS22	US-III+ @1.2GHz, US-IV+ @1.8Ghz	Intel Xeon E5345@2.33GHz
16GB/HS21, 8GB/QS22, 12GB/LS22	368 (32xPPE + 256xSPE + 80xAMD)	158Gbyte, 286GByte	8GB per node
544 cores	Infiniband 4X DDR	144 cores	**672 cores**
4 * Gigabit Ethernet, QDR 4x Infiniband	2.5 µs	2 x Gigabit Ethernet	Gigabit Ethernet
300µs, 1.1µs	20 Gbps	cca 300 µs	50µs
4Gbps, 40Gbps	3.28 for QS22; 0.62 for LS22	1 Gbps	1 Gbps
3.4	2.05 for QS22; 0.39 for LS22	0.5	**6.26**
1.003	Fedora 9 PPC	0.5	**5.25**
Scientific Linux 5.3 64 bit		SUN Solaris 9	Scientific Linux 5.5 64 bit
Sun Grid Engine		Sun Grid Engine	torque + maui
NFS, Lustre FS		UFS, QFS	/home, /scratch, /storage
20TB		6.4 Tbyte	3 TB , 100GB per node, up to 50 TB
400MB/s		NA	500 MB/s, 65 MB/s, 500 MB/s
Blas (ATLAS), Intel MKL, IBM MASS, LAPACK	BLAS, IBM MASS, LAPACK	Sun HPC ClusterTools, PVM3, FFTW, Scalapack, mpiblacs, Blas, Lapack	LAPACK, BLAS, FFTW3, SPRNG, Intel MKL, ScaLAPACK
OpenMPI, gcc (4.1.2 & 4.4.0), gfortran, Intel Compilers 11.0, PGI 7.0.7, SunStudio 12.1, TotalView 8.6, Gaussian03, Code Saturne 2.0, hrm, Matlab	OpenMPI, gcc, gfortran, xlc, xlf	Gaussian03, Gromacs, Sun One Studio Compiler Kit, ddd, Scilab, RasMol, Meep	MPICH, MPICH2, OpenMPI, gcc, gfortran, Intel Compilers (C/C++, Fortran), Portland Group Compilers (C/C++, Fortran), NAMD, CPMD, Paraver, Dimemas etc.

HPC infrastructure – HU

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- SUN Fire supercomputer cluster run by the National Information Infrastructure Development Institute (NIIF);
- The resource currently consists of two SUN Fire (SMP) machines totaling to 216 cores providing 0.6 Tflop/s computational power as well as 20 Tbytes of primary and 40 Tbytes of secondary storage facilities in total.
- Planned:
  - NIIF has managed to establish funding to improve the current compute power and to increase the data storage capacity up to **50 Tflop/s** and **0.5 Pbytes**.
  - The architecture of the four sites has been assembled to fit the most probable expected future usage scenarios and also a procurement was performed in Y2 of 2010, which led to different solutions:
    - one of the resources will facilitate an SMP/NUMA architecture, SGI Ultra Violet
    - The remaining three sites are fat node clusters facilitating an SGI Altrix cluster as well as two HP blade clusters. In the clustered systems Infiniband will be used as the internal interconnect among the different cluster entities, such as compute and data storage elements. On each site a dedicated storage will be used to offer storage space. The biggest HPC cluster will achieve **20 Tflops**.

HPC infrastructure – RO

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- The computing centers at IFIN-HH and UPB:
  - IBM BladeCenter Cluster at IFIN-HH, with an HPL benchmark Rpeak of 4 TFLOPS, contains both IBM PowerXCell 8i and AMD Opteron 2356 processors, counting a total of 368 cores and 592 GB RAM, using Infiniband 4X 10Gbps technology;
  - The Biocomputing cluster, with Rpeak=2,7 TFLOPS, is based on Intel Xeon E5430 (Quad-Core) processors and Myrinet 2000 2Gbps technology.
  - IBM BladeCenter 26 (16xQS22 + 10xLS22) QS22: 2x IBM PowerXCell 8i @3.2GHz; LS22: 2x AMD Opteron Quad Core 2376 HE @2.3GHz 32 GB per QS22; 8GB per LS22 368 (32xPPE + 256xSPE + 80xAMD) Infiniband 4X DDR 2.5 µs 20 Gbps 3.28 for QS22; 0.62 for LS22 2.05 for QS22; 0.39 for LS22
  - Fedora 9 PPC
  - 5.33 TF
- Application software:
  - NAMD, CHARMM, VASP, ATLAS, GotoBLAS, FFTW, GAUSSIAN, Turbomole, Matlab, etc.

HPC Infrastructure – RO - planned

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- IFIN-HH will upgrade the IFIN_BC cluster with 4 Tflops (peak performance, double precision), during the second half of 2011.
- ISS will purchase NVIDIA equipment and install it in a new GPU cluster, getting a peak performance of approximately **80 TFlops**.
- UPB will upgrade with 2 Tflops the NCIT cluster, and will procure GPU equipment based on Fermi NVIDIA architecture.
- UVT resumed the procedure for purchasing a Blue Gene supercomputer with 4096 cores and 11 Tflops in the next few months.

HPC Infrastructure - RS

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- PARADOX high performance cluster consists of 84 worker nodes (2 x quad core Xeon E5345 @ 2.33 GHz with 8GB of RAM). PARADOX is the largest HPC cluster in Serbia. Its computing nodes are interconnected by the star topology Gigabit Ethernet network through three stacked high-throughput Layer 3 switches, each node being connected to the switch by two Gigabit Ethernet cables in channel bonding. In terms of storage resources, PARADOX provides up to 50 TB of disk space to the HP-SEE community.
- tPARADOX training cluster for educational purposes, based on IBM's BladeCenter technology and it consists of IBM BladeCenter H chassis commonly used in high performance computing and different types of Blade servers that cover some of the major CPU architectures currently available: Intel's x86_64 and IBM's POWER and Cell/B.E.
- As part of the 10 Million Euro project for building supercomputing resources in Serbia, the PARADOX cluster will be expanded significantly in 2011.

HPC Infrastructure – GR Planned

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- GRNET has already submitted a proposal to GSRT for the creation of a national HPC center with a budget of around 10 Million Euro. Based on the conclusions of the HellasHPC project and the deliverable that describes the Greek strategy for HPC development the national HPC system is required to have the following technical characteristics
- Computational Power RMax (Linpack) 250-300 TFlops, RPeak: ~350 TFlops
- Number of processing elements 35.000 - 40.000 CPU cores
- Memory Size > 76 TB (at least 2GB per core)
- Storage Size 3 PetaByte
- Interconnect High Speed (>10 Gbit) with low lattency
- Initially a seed resource of about 40 Tflops will be integrated to HPSEE project.

Regional HPC Infrastructure – present and future

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Resources per country – commitments vs achieved

Country	Tflops		
	2010	2011	2012
Greece	0	40	80
Bulgaria	Planned 25/**available 30** Planned 0/ **available 8 GPU**	30+8 GPU	40+20 GPU
Romania	Planned 10/**available 11**	20+100 GPU	30+100 GPU
Hungary	Planned 1 / **available 1**	30	60
Serbia	Planned 0/**available 6**	20	40
OVERALL	**36/available 48+8 GPU**	**140 +108 GPU**	**250+120 GPU**

Current and planned computing power (double precision for CPU and single for GPU)

HPC Cluster at IICT-BAS

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- HP Cluster Platform Express 7000 enclosures with 36 blades BL 280c with dual Intel Xeon X5560 @ 2.8Ghz (total **576** cores), 24 GB RAM per blade
- 8 controlling nodes HP DL 380 G6 with dual Intel X5560 @ 2.8 Ghz, 32 GB RAM
- Non-blocking DDR Interconnection via Voltaire Grid director 2004
- Two SAN switches for redundant access
- MSA2312fc with 48 TB storage, Lustre filesystem
- P2000 G3 with 48 TB storage added last week
- More than 92% efficiency on LINPACK (>3 TFlops, peak performance 3.2TFlops)

HPC Cluster at IICT-BAS

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Extension cluster with 4 GPU cards NVIDIA GTX 295 (each card counts as 2 graphical devices), CPU Intel Core i7 @2.66 Ghz, 12 GB RAM.
- Total number of threads for GPU computing – 4x2x240=1920
- High performance Lustre filesystems:
  - /home – 22 TB
  - /scratch – 7 TB
- Most heavy users: Environmental Modelling, Computational Mechanics, Astronomy, Computational Chemistry, bio-informatics, …

Installed software at IICT

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

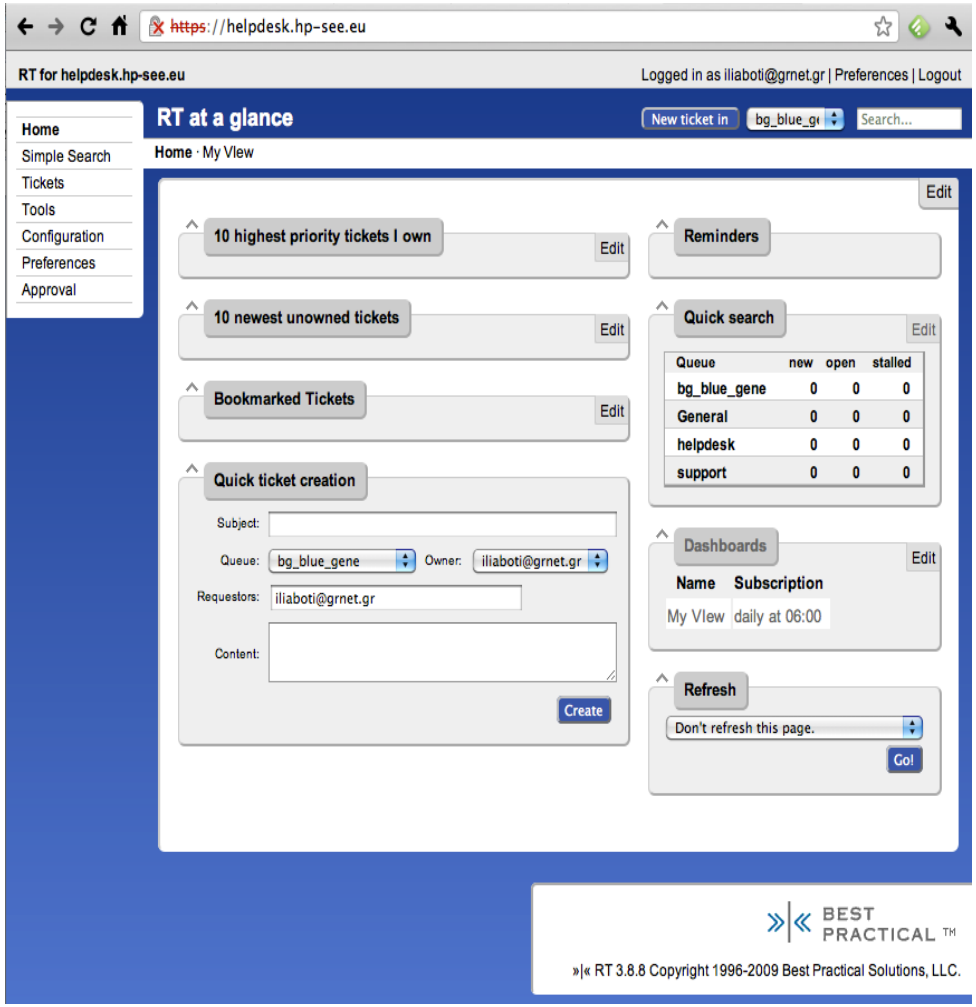
- Torque batch system
- Glite Grid middleware
- Unicore Grid middleware
- NVIDIA GPU computing SDK
- Compilers, MPI(mviapich1, mviapich2, openmpi), debuggers and profilers (MPE, scalasca, mpiP)
- ATLAS, LAPACK, HPL, ScaLAPACK, GotoBLAS, FFTW, SPRNG, MPI (MVIAPICH1/2, OpenMPI), BLACS, BLAS, CUDA, OpenCL, OpenFOAM, octave
- Charm++, GAMESS, GROMACS, NAMD, NWChem, ABINIT, WRF, CMAQ, SMOKE

Helpdesk

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

Helpdesk for HP-SEE users is operational at https://helpdesk.hp-see.eu

Other users of HPC cluster at IICT-BAS should use hpcg-support@bas.bg



Access to Grid

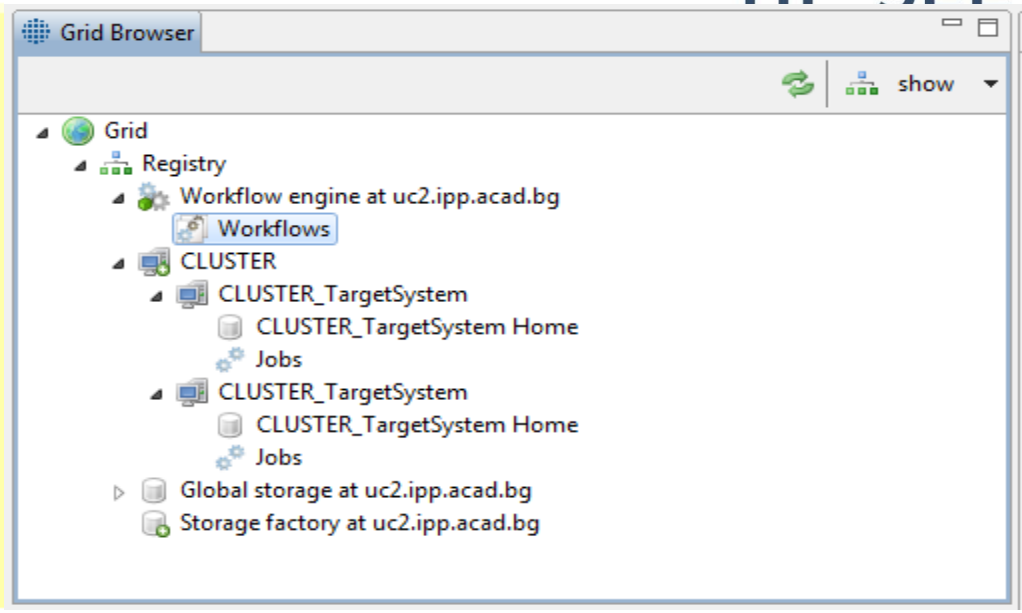
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Obtaining access to the infrastructure of the European Grid Intiative requires:
  - certificate (http://ca.acad.bg)
  - Membership in a virtual organization, that can be:
    - European
    - Regional
    - Bulgarian
- Most of the HPC clusters in the region are also available via Grid middleware
- The gateway node of the HPC cluster at IICT-BAS can serve as Grid User Interface

Unicore usage example

HP-SEE

Site name	Address	Requests served	Status	Message
WORKFLOW	https://uc3.ipp.acad.bg:7700	1637		OK
CLUSTER	https://uc3.ipp.acad.bg:7777	777		OK
REGISTRY	https://uc1.ipp.acad.bg:7778	3117		OK
SERVORCH	https://uc3.ipp.acad.bg:7701	2477		OK




```
[dgeorgiev@wn02 ucc-1.3.1]$ ./bin/ucc connect
You can access 1 target system(s).
[dgeorgiev@wn02 ucc-1.3.1]$ ./bin/ucc run samples/date.u
SUCCESSFUL exit code: 0
/home/dgeorgiev/ucc-1.3.1/./outs/3dc88cd4-0faf-48b6-ab0e-d7360a5031a3.stdout
/home/dgeorgiev/ucc-1.3.1/./outs/3dc88cd4-0faf-48b6-ab0e-d7360a5031a3.stderr
/home/dgeorgiev/ucc-1.3.1/./outs/3dc88cd4-0faf-48b6-ab0e-d7360a5031a3.properties
[dgeorgiev@wn02 ucc-1.3.1]$
```

Access to HP-SEE resources

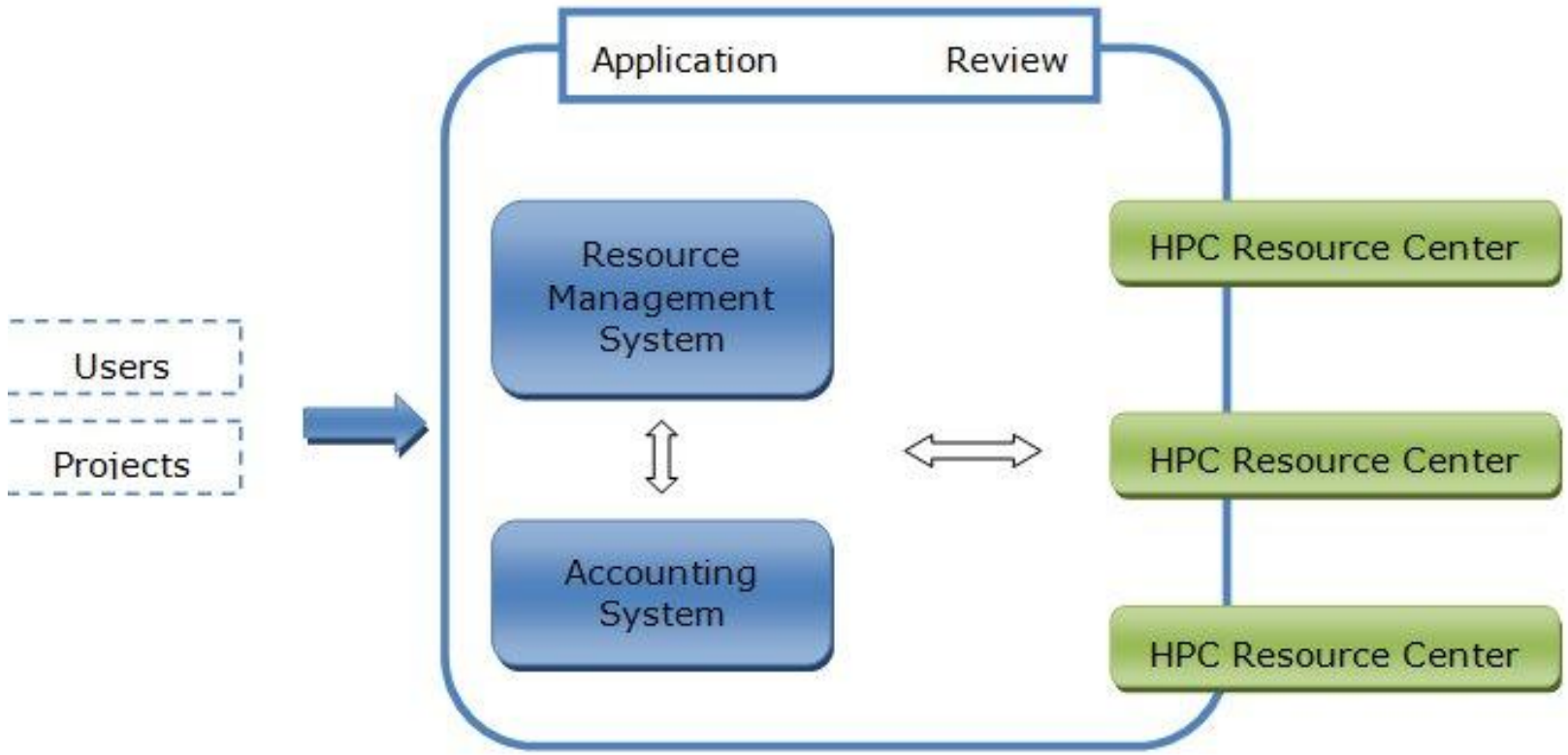


Figure 8. Resource Management

Direct access to HPC cluster at IICT-BAS

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- User that have filled the access form http://www.grid.bas.bg/gta/projects/HP-SEE/form-accountHPCG_bg.doc
- and obtained direct access can log in to the user interface **gw.ipp.acad.bg** with username and password.
- All worker nodes are accessible with ssh without password
- The login node should be used only for compilation and light testing.
- Computational jobs should be submitted to the appropriate queue via **qsub**
- Additional software usually installed under /opt/exp_software

Direct access to HPC cluster at IICT-BAS

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- The /home file system is under raid6 and should be used for permanent storage.
- The /scratch file system should be used for temporary files and directories, especially when high amount of I/O is performed. Files can be removed by an administrator at any time. For each user there is /**scratch/*username***
- The GPU cluster is separate, but can be accessed with same username and password as the main one.

Conclusions

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- The current High Performance infrastructure in the region includes one supercomputer Blue Gene in Sofia and several clusters.
- Several neighboring countries plan significant expansion of their HPC capabilities.
- By sharing of the computational resources scientists from countries that do not posses such resources can develop, deploy and use HPC applications for computationally intensive research.