Russian Academy of Sciences
Program Systems Institute

SKIF MSU Supercomputer

The SKIF MSU supercomputer has been built as a result of implementing one of the projects of the SKIF-GRID scientific and technical program of the Russia and Belarus Union State.

Developers: M.V. Lomonosov Moscow State University, Program Systems Institute of RAS (PSI of RAS), Russian company of T-Platforms.

Location: M.V. Lomonosov Moscow State University.

Lead time: March 2007 - March 19, 2008 (12 months).

At the time of completion of the SKIF MSU supercomputer built on the basis of 625 blades made by T-Platforms its peak performance was 60 Tflops.

Real performance of the system on the Linpack benchmark is 47,04 TFlops, that is 78,4% of the peak performance. This is the best efficiency index among all the systems of the first hundred ranked on the Top500 list of the world's fastest computers built on the basis of Intel Xeon 4-core processors (www.top500.org).

The real performance of the SKIF MSU supercomputer corresponds to the 22d position on the world's current Top500 listing, which has become the absolute record for Russia as of March 2008. The SKIF MSU was also the 7th most powerful system among all the supercomputers used in the world's educational system.

Major features

|	
Peak/Linpack performance	**60 Tflops / 47.17 Tflops**
Coefficient of efficiency	**78.6%**
Number of computing nodes/processors	**625/1250**
Processor type	**4-core Intel® Xeon® E5472, 3,0 GHz**
Total volume of RAM of all computing nodes	**5 512 GB**
Total volume of disk memory of all computing nodes	**16 640 GB**
Type of the system network	**DDR InfiniBand (Mellanox ConnectX)**
Type of managing (auxiliary) network	**Gigabit Ethernet**
Service net	**SKIF ServNet-3 + IPMI**
Formfactor of the node	**blade**
Number of 42U hardware cabinets	**14**
Total occupied area (including cooling subsystem, UPS)	**92 m²**
Total power consumption (including cooling and UPS)	**720 KW**

SKIF MSU Computing Nodes

The SKIF MSU supercomputer is built on the basis of the newest technological solutions and uses a number of Russian developments, created within the bounds of the SKIF and SKIF-GRID supercomputer projects of the Union state.

The basis of the supercomputer is made up with blades of T-Platforms in-house design and production. The density achieved in T-Blade modules is the highest in the industry: 10 blades are held in the chassis that is only 5U high (20 CPU/80 cores/~0.7 Tflops). This is 18% better than that of any analogues. The first blade-solutions in the industry that integrate support for the new Intel 5400 chipset have been implemented in T-Blade modules. Besides, T-Blade modules are compatible with any standard kinds of interconnects and other external devices due to the PCI-Express 2.0 expansion slot.

Chassis with 10 T-Blade modules
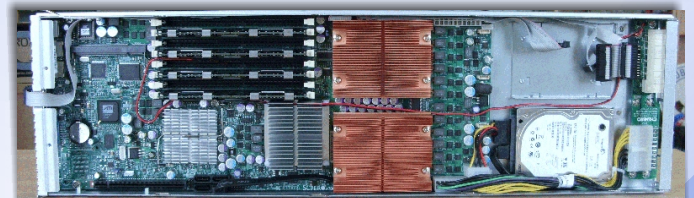
T-Blade module

Features of the T-Blade module

- CPU: 2x Intel XEON "Clovertown/Harpertown"
- RAM: 8 slots 667 / 533MHz FB-DIMM
- Chipset: Intel 5400 (Seaburg)
- HDD: up to two SATA 2.5" discs
- Ethernet: 2 x Gigabit Etherhet
- Expansion: Standard PCI-Express 8x

SKIF MSU supercomputer networks

All the computing nodes in the SKIF MSU are connected with three independent networks:

- System network:
  - Infiniband 4x DDR;
  - 6 x 144-port root switches;
  - 54 x 24-port boundary switches;
  - FBB = 20 Gbit/s;
  - latency = 2.2 microsec.
- Auxiliary network:
  - 2 x 384-port Gigabit Ethernet switches;
  - switching capacity 1536 Gbps;
  - total capacity of packet forwarding — 571 Mpps.
- Service net ServNet-3 for SKIF supercomputers of series 3:
  - two-level architecture;
  - applicable to supercomputers with any number of nodes and chassis (of petaflops range);
  - provides low-level management of computing nodes;
  - provides chassis monitoring and management (measuring and control of rotation speed of fans).

SertvNet-3 service net

ServNet-3 service net of in-house design by the PSI of RAS monitors all the blades and fan modules in every chassis of the supercomputer and supports:

- the following operations with any (randomly chosen) computing node of the supercomputer:
  - switching on/off;
  - status poll [scanning] - ON, OFF, etc.;
  - serial console session;
  - voltage measurements in the power supply and temperature system (at the point where the ServNet T60 card is mounted;
- the following operations with any (randomly chosen) fan of any chassis in the supercomputer:
  - measuring the speed of rotation of the fan;
  - controlling the speed of rotation of the fan.

Further development of monitoring and control functions of the ServNet-3 service net is provided for.

ServNet-T60: blade control card, 80×70 mm
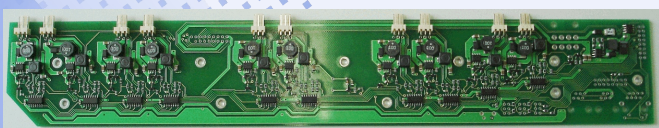
ServNet CMB: Fan Management Board of the chassis, 372×60 mm.

SKIF MSU supercomputer infrastructure

The SKIF MSU supercomputer has a unique information computing engineering infrastructure necessary for reliable round-the-clock functioning of the system.

Data storage and backup subsystem

- ReadyStorage ActiveScale Cluster.
- Number of modules: 12.
- One module footprint: 5 Tbyte.
- Total disk capacity: 60 Tbyte.
- Total throughput: 700 Mbyte/sec.

Uninterruptible power supply 80kVA units

- 10 min. off-line operation.
- Backup [standby] N+1.
- Hot servicing.
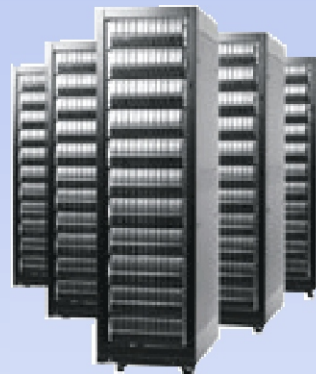
Data storage and backup subsystem

Uninterruptible power supply 80kVA units

Cooling subsystem

- Modular cooling on the level of a row of racks in combination with pressurizing the hot corridor.
- Guaranteed removal of 30 KW heat energy from each rack with computing nodes.
- Guaranteed cooling system redundancy N+1.

Hot corridor

Cold corridor