# IMPLEMENTATION OF JOB SUBMISSION INTERFACE FROM EGEE/WLCG GRID INFRASTRUCTURE TO SKIF SERIES SUPERCOMPUTERS

**V. Edneral V. Ilyin, A. Kryukov, G. Shpiz, L. Shamardin**

*Scobeltsyn Institute of Nuclear Physics,*
*Moscow State Universtiry (SINP MSU),*
*Moscow, 119991, Russia*

The work was done under the contract of the Union State Russia-Belarus # ??-2/07

✓ The scientific program SKIF-GRID [http://skif-grid.botik.ru/] is being implemented underway within the Union State of Russia and Belarus.

✓ The purpose of this program is to build a grid infrastructure capable of efficient usage of SKIF series supercomputers in the areas of nanotechnology, biomedicine, material sciences and other fields of scientific research and technological computations.

✓ The network of SKIF series supercomputers and different research centers are located over the whole territory of the Union State so it was reasonable to choose grid as a technology for giving access to the resources since it is the state of the technology of distributed computations.

The grid technology is successfully used in the research on the world's largest particle accelerator LHC [http://public.web.cern.ch/public/en/LHC/LHC-en.html].

Considering positive experience of using grid technology in the LHC project, the pan-European infrastructure EGEE/WLCG was built [http://public.eu-egee.org, http://lcg.web.cern.ch/LCG/] which is the base computer infrastructure for scientific research in Europe.

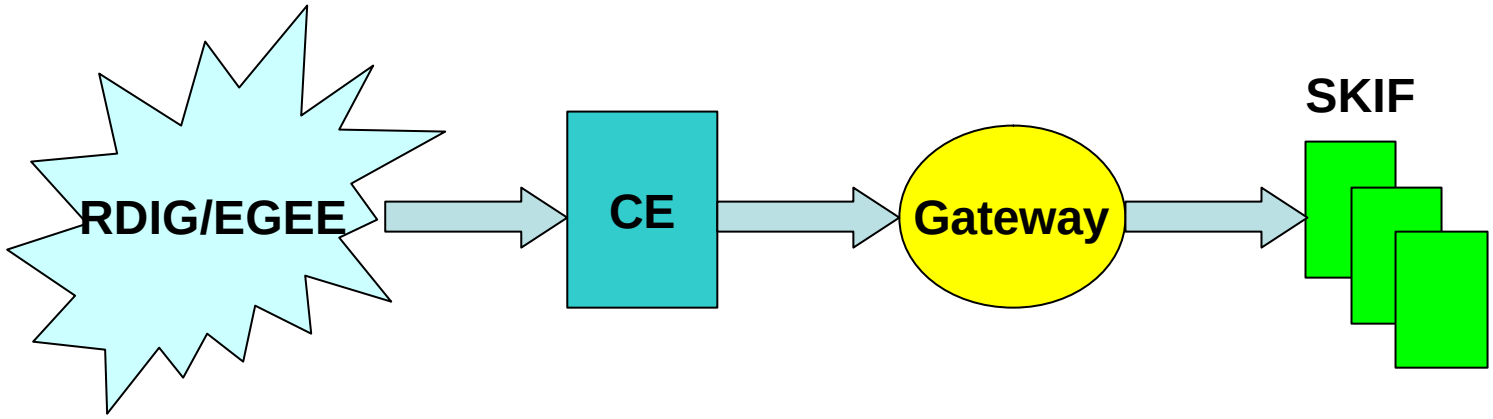
There is a segment of this infrastructure in Russia – Russian Intensive Data Grid (RDIG) [http://egee-rdig.ru/], which unites 15 resource centers and is working in tight cooperation with European infrastructure.

Thus a problem of job submission of parallel jobs (MPI) to SKIF series supercomputers and clusters from EGEE/WLCG grid infrastructure was formulated.

The primary attention in the solution of this problem was paid to the requirement of minimal or even absent intervention to the SKIF cluster operations and its software.

To achieve this goal the interaction with the cluster through the special Gateway based on the gLite Worker Node middleware [http://glite.web.cern.ch/glite/] was chosen.

# The offered scheme allows to split RDIG and SKIF middleware




**The simplified scheme of a SKIF-GRID site**

*Here CE – a computing element (CE gLite), the Gateway on the basis of WN gLite, carrying out transfer of RDIG/EGEE tasks to the super computer SKIF (MPI cluster).*

It is assumed that the computing nodes of the cluster are running the client software of the local batch system. In our particular case we used the **PBS** batch system [http://www.pbspro.com] but it is possible to use the **Torque** system [http://www.clusterresources.com/pages/products/torque-resource-manager.php].

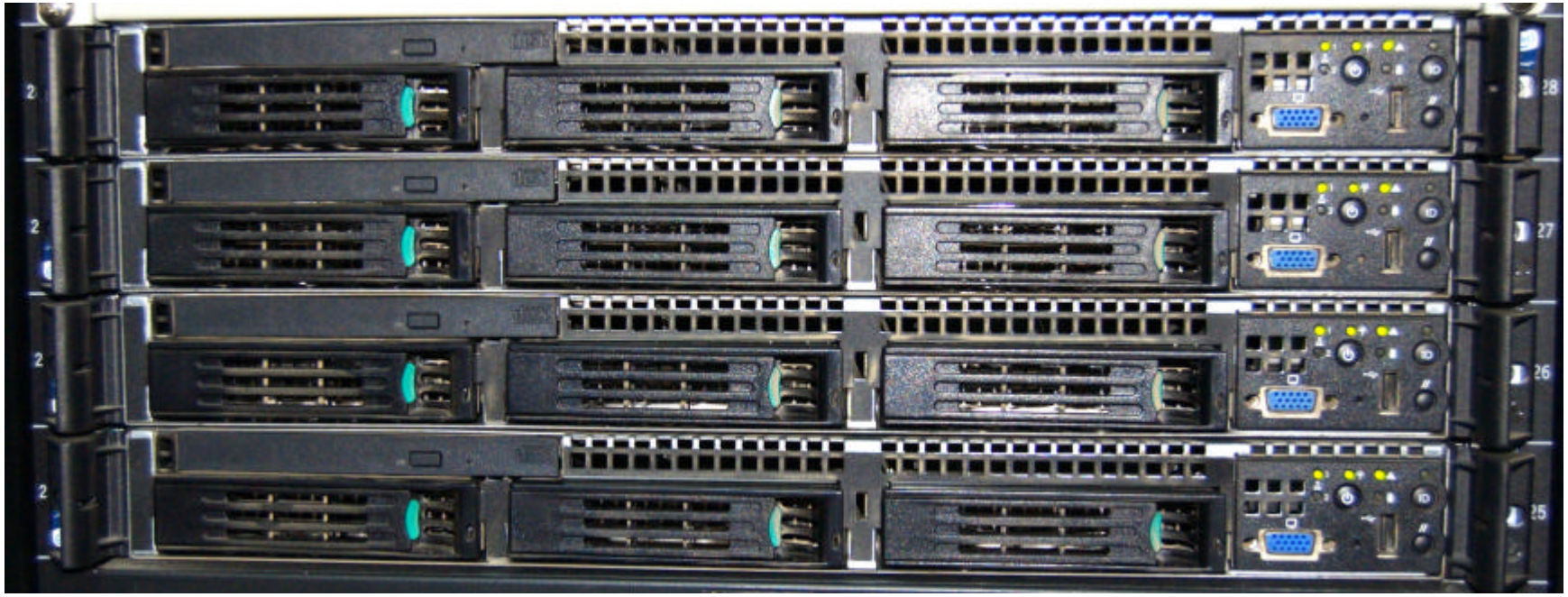
The proposed solution in addition to possibility of job submission from EGEE/WLCG also provides transparency of the informational systems for the grid users and resource brokers. To submit the job, it must be described in a regular JDL file [https://edms.cern.ch/file/722398//gLite-3-UserGuide.pdf] and submitted from the host with the user interface middleware in a regular fashion.

# The testbed of SKIF-GRID site

The site testbed consists of:

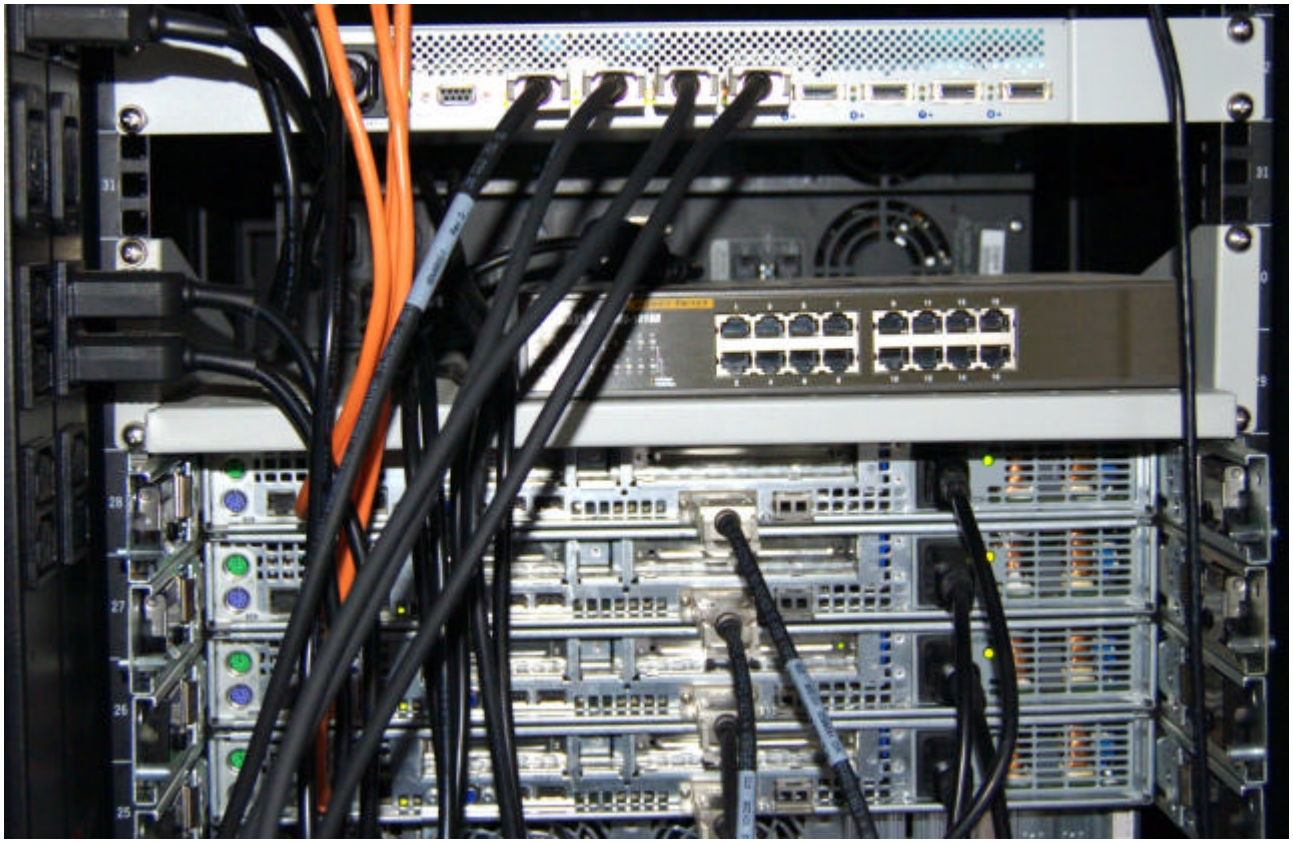
- Computing element (CE gLite)?
- Gateway (WN gLite) between EGEE components and model of the super computer SKIF
- Model of the super computer SKIF on the basis of 16 – kernel MPI cluster:
  - 4 units with 2 double-kernel CPU in everyone
  - processors Xeon 5100 with working frequency of 3 GHz
  - 1 GByte of the RAM on each kernel
  - network connections - Infiniband

# The model of SKIF cluster



4 units with 2 double-kernel CPU in everyone.
Processors Xeon 5100 with working frequency of 3 GHz,
1 Gbyte of the RAM on each kernel.

# The model of SKIF cluster



The upper unit is the Infiniband switch

# Middleware

- Computing element (CE) and Gateway:
  - OS Scientific Linux 4
  - middleware for CE gLite 3 at the CE
  - middleware for WN gLite 3 at the Gateway
  - Set of scripts for transfer of tasks to the super computer SKIF
- Model of the super computer SKIF (MPI cluster):
  - PBS (or Torque) client
  - MPICH (or MPICH-2) library, the version under Infiniband

# Main adjustments

We mount a PBS server and a queues manager (Maui) on the computing element (CE). Following moments should be fixed:

- Maintenance passwordless ssh, rsh connections between the CE, Gateway and nodes of the SKIF cluster for local users (for example for the dteam VO they are dteam001, dteam002, …). In particular, all these users should be known by nodes of the SKIF cluster. For this purpose the file system “/home” of the cluster should be mounted on the CE and the Gateway;

# Main adjustments

- We should have a possibility of starting tasks through a PBS server and a queues manager (Maui) from the CE via the Gateway to nodes of the SKIF cluster (of PBS type). At the cluster nodes PBS clients (MOMs) are established;
- It is necessary to provide impossibility of direct start of EGEE tasks on the SKIF cluster nodes. They should start only on a single WN – on the Gateway.

Details of tuning a PBS server, queues manager (Maui) at the CE, PBS clients (MOMs) at the Gateway and nodes of the SKIF cluster are described in the created set of a documentation.

# At the moment we have had:

- a workable testbed of the SRIF-GRID site
- a set of scripts and documentations for the site
- Results of testing. We send tasks as regular gLite job (JDL and others files) from a standard gLite User Interface and have results in usual way.

# Questions for future

- Is it enough a single gateway for effective loading large SKIF clusters? The described architecture allows a usage of many gateways (gLite WNs).
- Is it enough of EGEE standard monitoring tools for effective management by multi processor systems if it is desirable to take into account information about a cluster structure?
- Debugging tools for users.

# Thanks for attention

GRID 2008. Dubna, on June, 30th
- on July, 4th, 2008