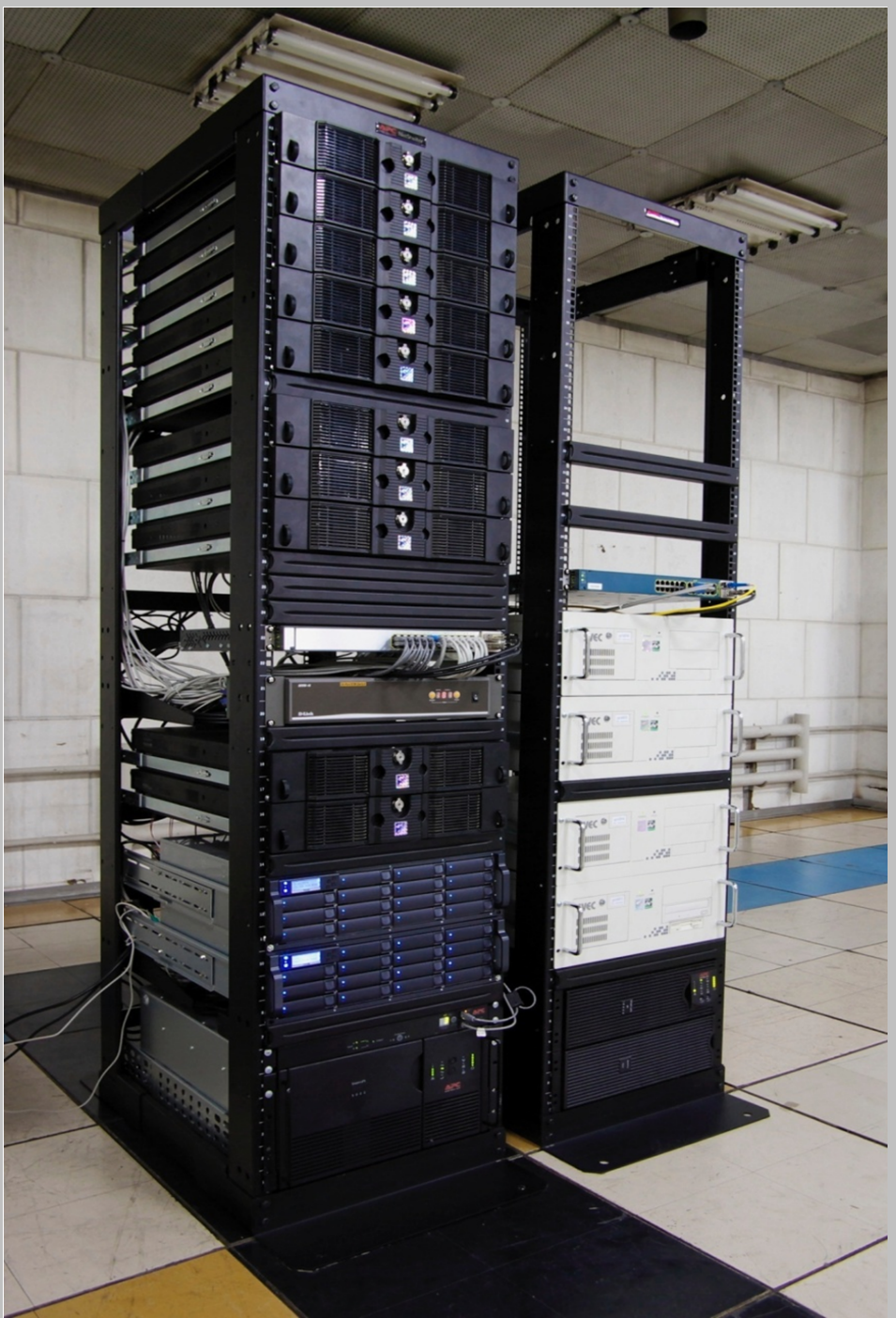
Статус ГРИД-кластера ИЯФ СО РАН. Использование внешних вычислительных ресурсов.

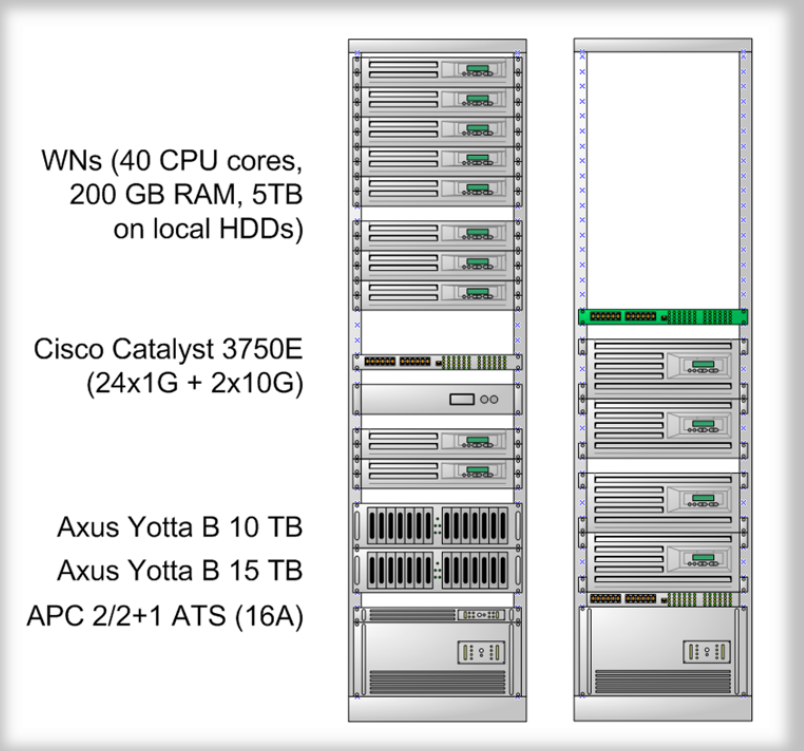
А. Сухарев

Рабочее совещание «Physics & Computing in ATLAS»

МИФИ, 27 января 2011 г.

BINP LCG Farm





CPU: 40 cores (100 kSI2k) | 200 GB RAM
HDD: 25 TB raw (22 TB visible) on arrays
3 TB on PVFS2 (distributed)

Platform: Scientific Linux 5 on x86_64
Virtualization: XEN, KVM

2010 Plans

- Getting OK for all the SAM tests
- Confirm the stability of operations for 1-2 weeks
- Upscale the number of WNs to the production level (up to 32 CPU cores = 80 kSI2k max)
- Ask ATLAS VO admins to install the experimental software on the site
- Test the site for ability to run ATLAS production jobs
- Check if the 110 Mbps SB RAS channel is capable to carry the load of 80 kSI2k site
- Get to production with ATLAS VO

Suspended until 2011Q1: network infrastructural problems are to be solved first.

Future Prospects (2011-...)

- Major upgrade of the BINP/GCF hardware focusing on the storage system capacity and performance
- Up to 550 TB of online (HDD) storage (and 680 CPU cores)
- Switch SAN fabric
- Solving the problem with NSK-MSK connectivity for the LCG site
- Dedicated VPN to MSK-IX seem to be the best solution
- Start getting the next generation hardware this year
- 8x increase of CPU cores density (up to 16 cores/1U)
- Adding DDR IB (20 Gbps) network to the farm
- 8 Gbps FC based SAN
- 2x increase of storage density (up to 10 TB/1U)

Available Computing Resources

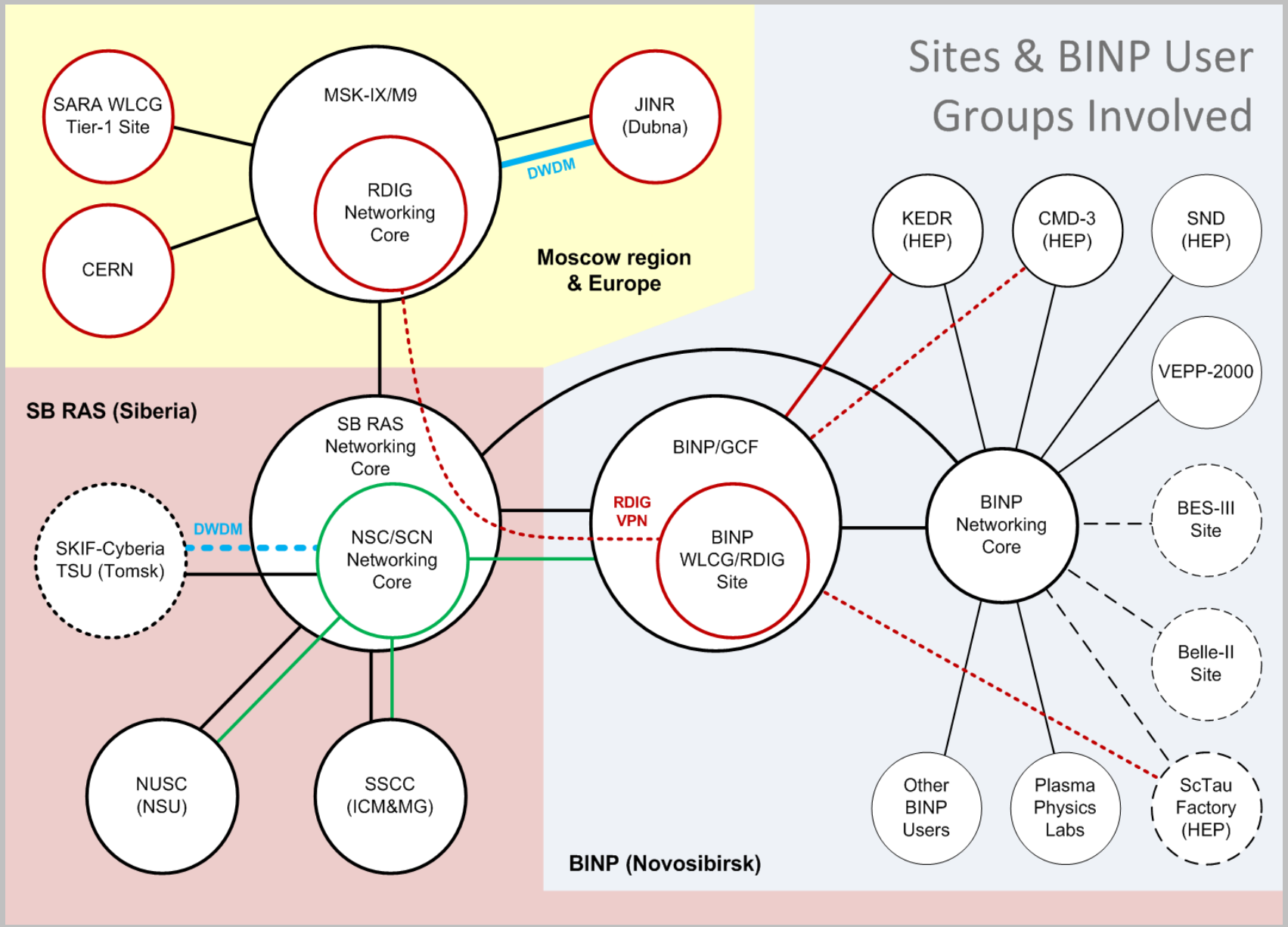
- Siberian Supercomputer Center (SSCC) at the Institute of Computational Mathematics & Mathematical Geophysics (ICM&MG): 17 TFlops combined performance achieved after the recent upgrade (2010Q3)

- Novosibirsk State University (NSU) Supercomputer Center (NUSC): 1280 physical CPU cores, 13.4 TFlops (since 2010Q2)

  10 GbE primary link + 2x auxiliary 1 Gbps links deployed

- SKIF-Cyberia at Tomsk State University: 9 TFlops >> 30 TFlops

External Connectivity Schema for the Research Groups of BINP

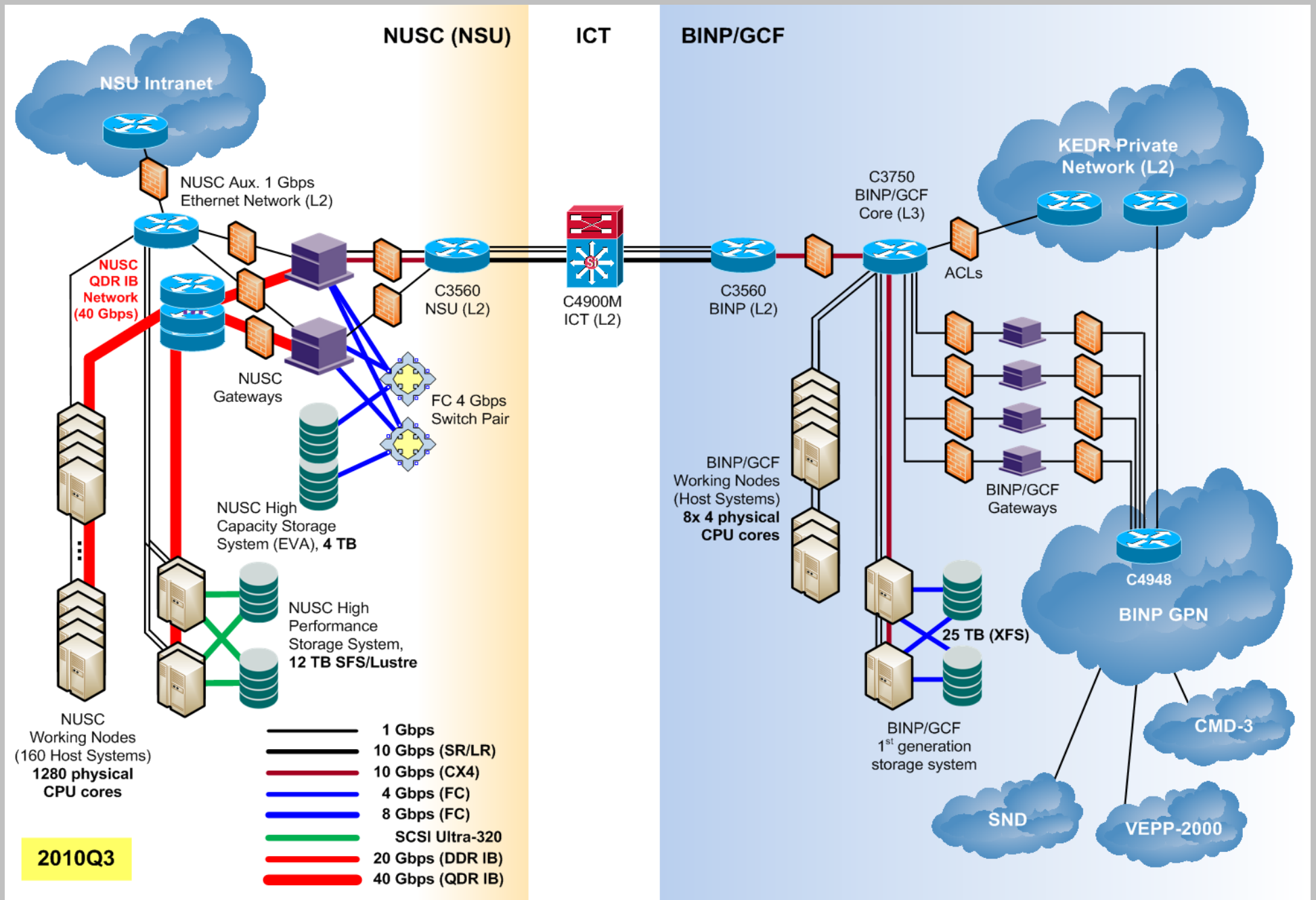


Novosibirsk State University Supercomputer

http://nusc.ru

http://nsu.ru

BINP-NSU Networking Schema



The Task

- Goal: Run common HEP software on a supercomputer
  - Utilization of InfiniBand interconnect
  - Batch system integration
- Method: Virtualization
  - Which platform? (VMware, Xen, KVM)
- Proof of concept: KEDR experiment real data processing and simulation
  - Fixed platform & environment (SLC3 on i686)
  - Already working on BINP/GCF using Xen

Virtualization Platform

- VMware
  - Poor usability (In My Opinion)
  - Need local storage
- Xen
  - Already tested on BINP/GCF (uptime > 2 years)
  - Custom kernel
  - InfiniBand usage causes problems
- KVM
  - No support until September 2010

It is not a problem to convert VM's disk image from a virtualization platform to another.

Xen Test Run (September 2010)

- Several 8-core physical nodes were used exclusively, Xen-specific Linux kernel was installed
- No batch integration
- No InfiniBand utilization
- VM's disk images located on BINP/GCF disk array

1 month stability obtained with up to 120 KEDR VM working completely like common physical KEDR computers

KVM Usage

- No exclusive physical nodes, common kernel, no IB problems
- Batch system integration, i. e. a VM starts as an ordinary batch system job
- All VM's network traffic (including access to disk images) is routed via IB

Technically BINP can make use of all 1280 physical CPUcores of NUSC.

- Recent successful runs:
  - 509 single-core KEDR VMs
  - 512 double-core KEDR VMs (with HyperThreading enabled on physical nodes)

Results

The VM-based solution works fine for the real life example of particle detector experiment:

- Long term VM stability obtained: >1 month at NUSC, >1 year at BINP Grid Farm (lower limits)
- Most of the underlying implementation details are hidden for the users
- No changes were needed for detector offline reconstruction / simulation software and/or its execution environment
- Yet, the solution is not completely automated (KEDR & NUSC batch system integration is not done yet)
- Technically BINP can make use of all 1280 pysical CPU cores (2048 HT cores = ~20 kHEP-SPEC06) at NUSC for the short periods of time (1-2 weeks) and 256 pysical CPU cores (~2.5 kHEP-SPEC06) on a regular basis (prospected typical gLite WN uptime - 1 month)

Next Steps

The solution obtained for KEDR applies to all particle experiments currently running at BINP as well as for LCG.

- Solve network infrastructure problems
- Prepare VMs for gLite WNs to be deployed at NUSC
- Get to production with ATLAS VO using NUSC computing resources
- Expand to other Siberian supercomputers (SSCC)

Thank you