DISTRIBUTED COMPUTING AND BIG DATA AT JINR

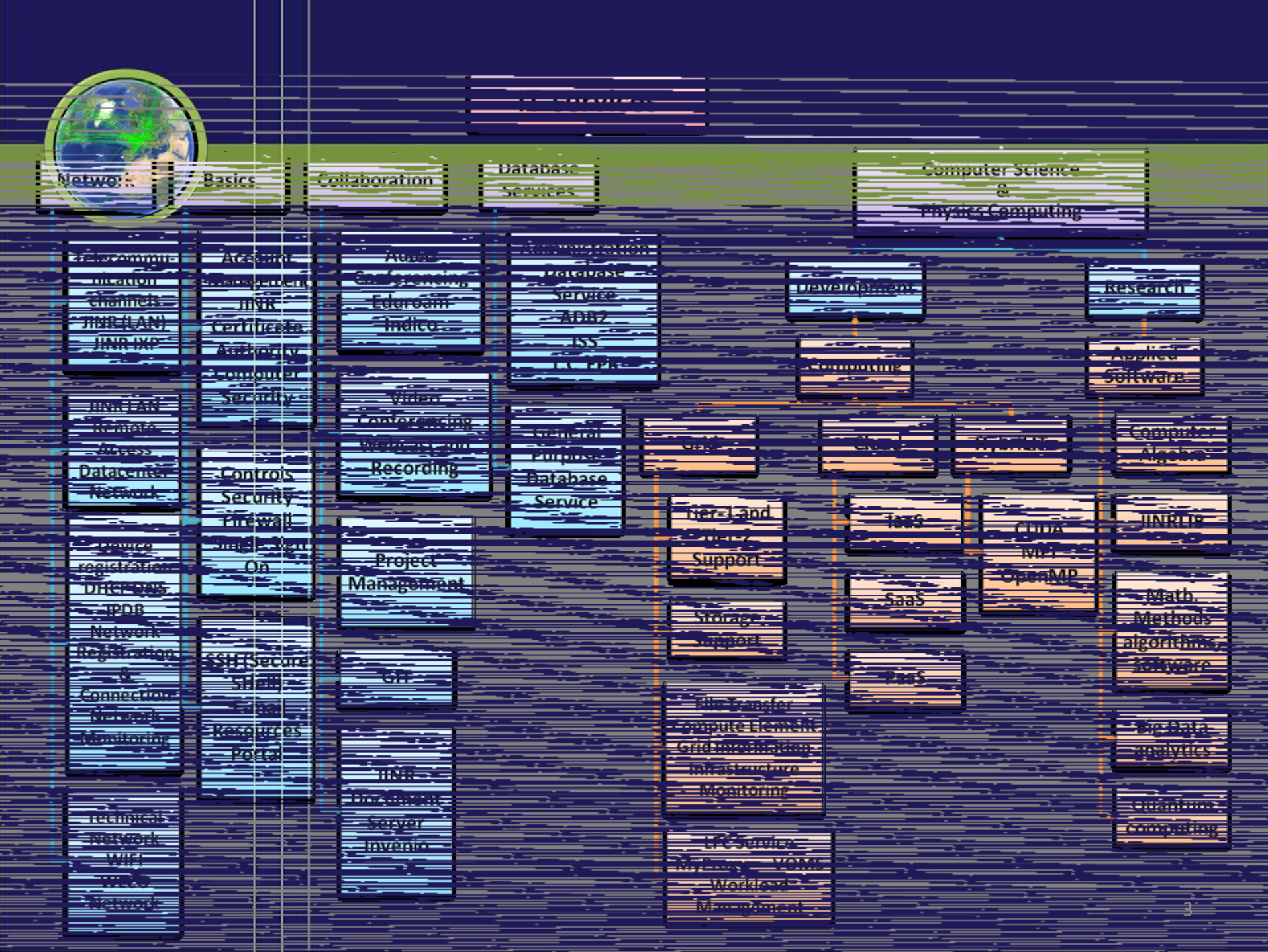
Vladimir Korenkov

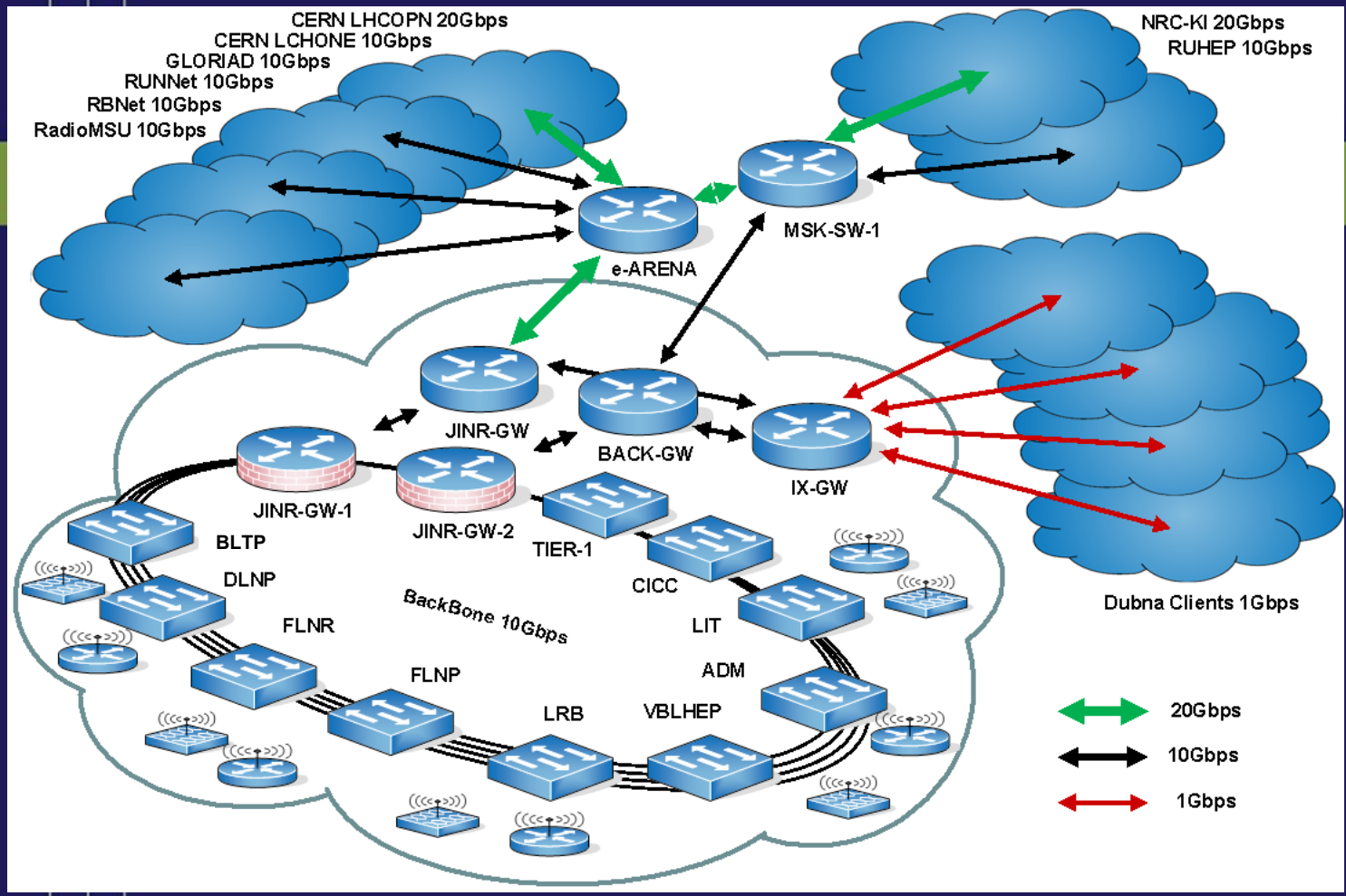
LIT JINR

LIT Fundamentals

- Provide IT services necessary for the fulfillment of the JINR Topical Plan on Research and International Cooperation in an efficient and effective manner
- Building world-class competence in IT and computational physics
- 24/7 support of computing infrastructure and services such availability is called nonstop service

IT-infrastructure is one of the JINR basic facilities


Network
Basics
Collaboration
Database Services
Computer Science & Physics Computing
Telecommunication channels
JINR-(LAN)
JINR-IXP
Account Management
JINR Certificate Authority
Computer Security
Audio Conferencing
Eduroam
Indico
Administration Database Service
ADB2
ISS
IT PPR
Development
Research
Computing
Applied Software
JINR-LAN Remote Access
Datacenter Network
Video Conferencing
Multicast
Recording
General Purpose Database Service
Grid
Cloud
Hybrid
Computer Algebra
Controls
Security
Firewall
Single Sign On
Tier-1 and Tier-2 Support
IaaS
CUDA
MIT
OpenMP
JINR-LIB
Device registration
DHCP-DNS
IPDB
Network Registration & Connection
Network Monitoring
Project Management
SaaS
Storage Support
Math. Methods algorithms, software
SSH(Secure SHell)
Local Resources Portal
PaaS
GIT
File transfer
Compute Element
Grid-Installation
Infrastructure Monitoring
Big Data analytics
Technical Network
WiFi
WAN Network
JINR Document Server
Invenio
Quantum Computing
LFC-Service
Mirrors
VOMS
Worldwide Management



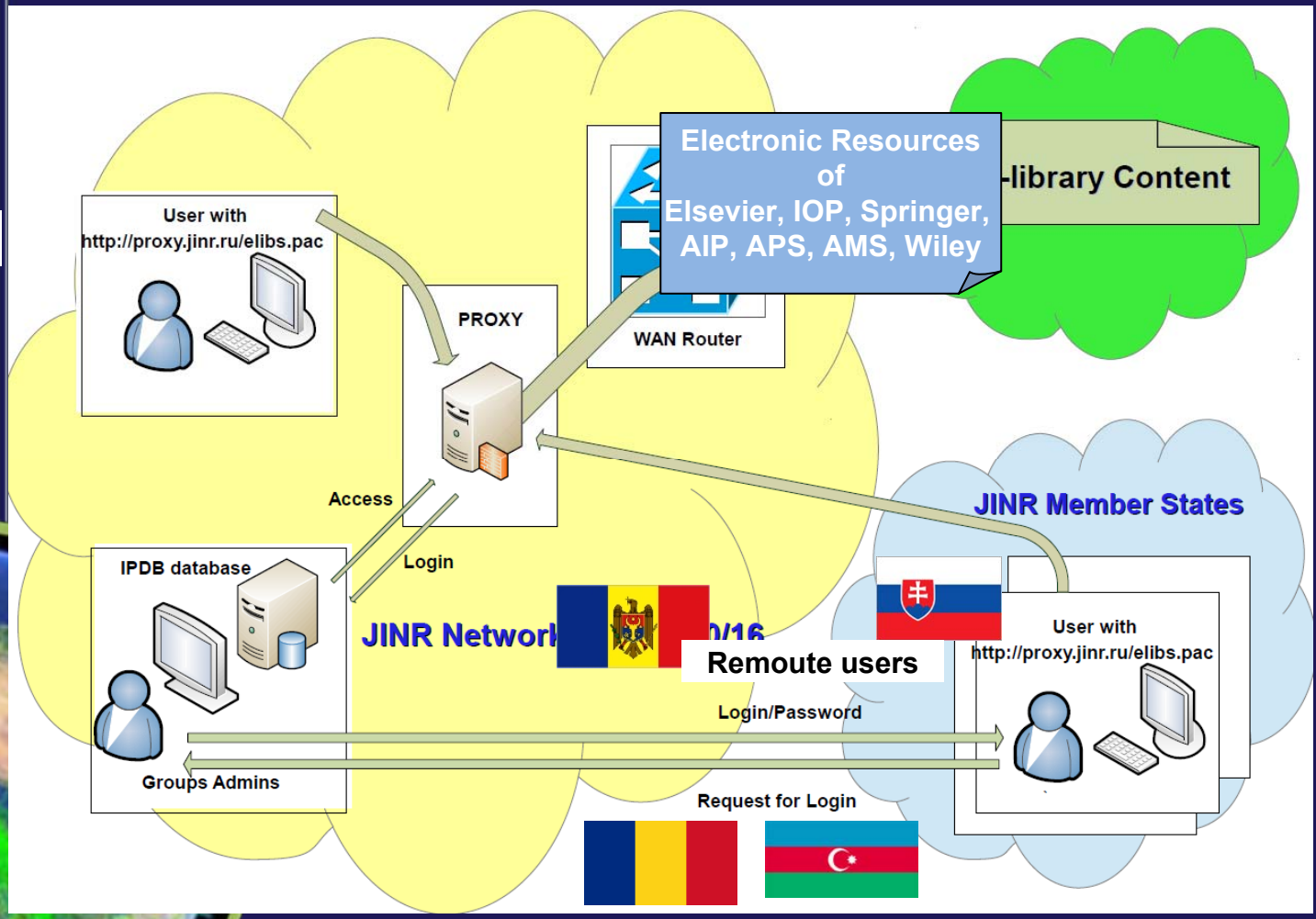
JINR Local Area Network

Comprises **7955** computers & nodes
Users – **4099**, IP – **12568**
Remote VPN users – **864**
E-library**- 1435, mail.jinr.ru-2000**
High-speed transport (10 Gb/s)

Controlled-access at network entrance.
General network authorization system involves basic services (Kerberos,AFS, batch systems, JINR LAN remote access, etc.)
IPDB database - registration and the authorization of the network elements and users, visualization of statistics of the network traffic flow, etc.

Access Service to Electronic Resources of World Publishers

Local users

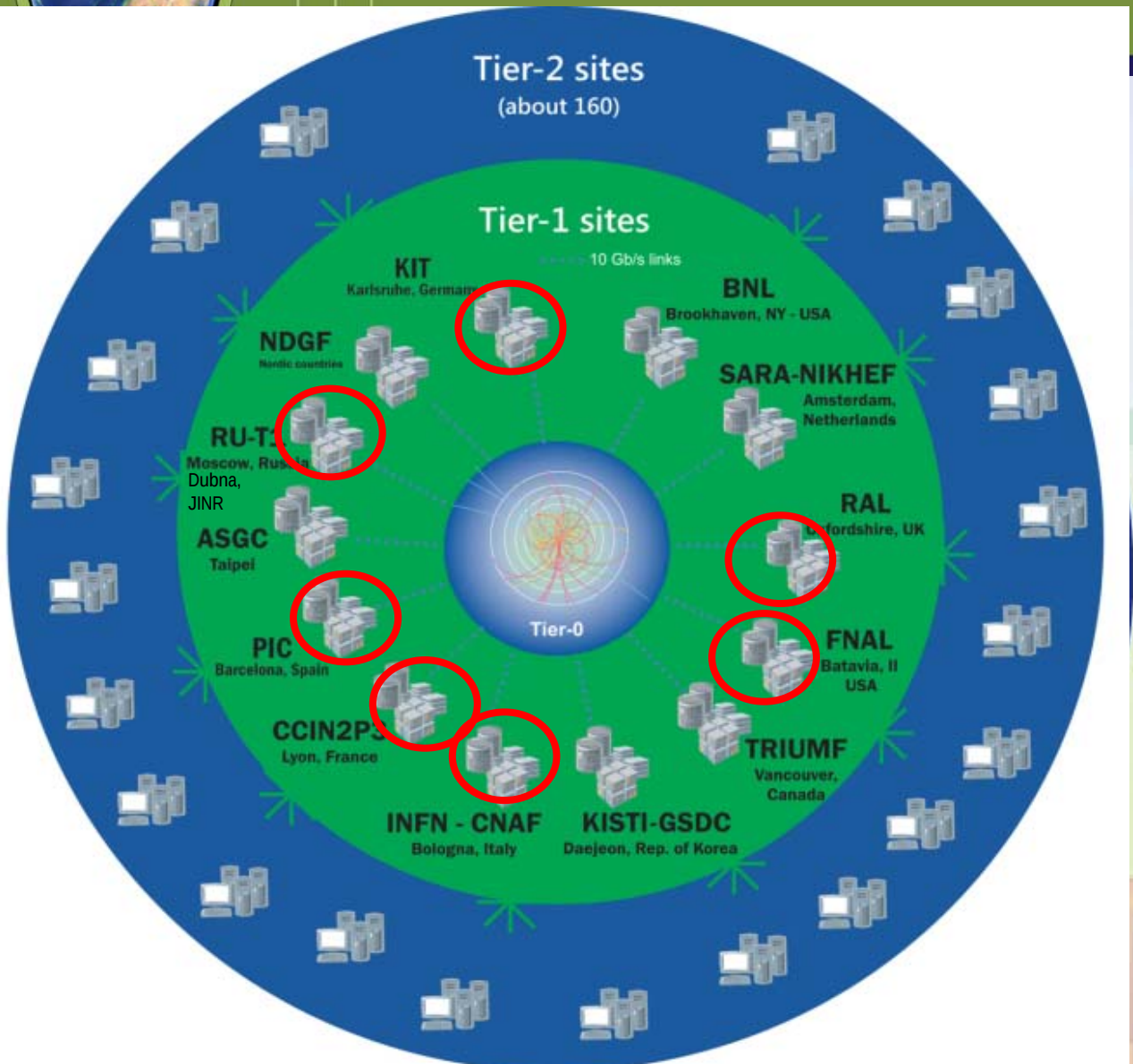


Total e-library users: 1375

Remote JINR users from Member States :

Republic of Azerbaijan - 24
Slovak Republic - 39
Republic of Moldova – 6 (+3)
Romania – 37, Bulgaria -1 (+8), Georgia-1(+7)

LHC Computing Model

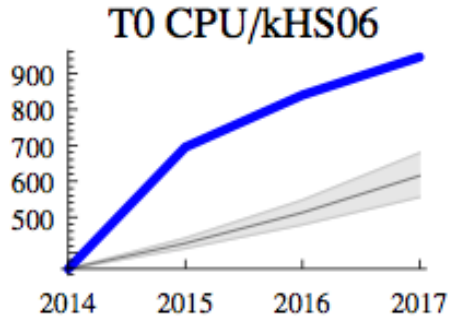


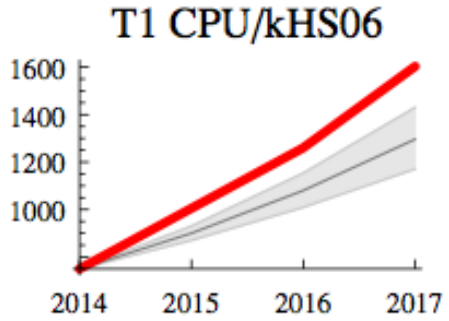
Tier-0 (CERN):
- Data recording
- Initial data reconstruction
- Data distribution

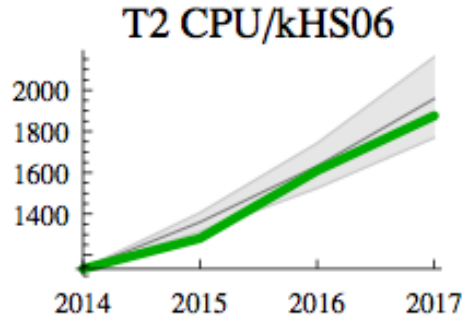
Tier-1 (14 centres):
- Permanent storage
- Re-processing
- Analysis
- Simulation

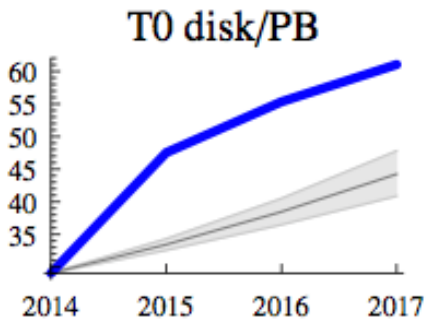
Tier-2 (>200 centres):
- Simulation
- End-user analysis

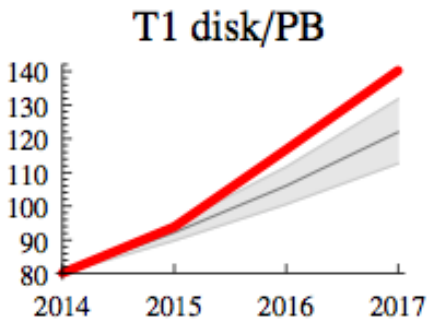
Combined (sum of experiments)

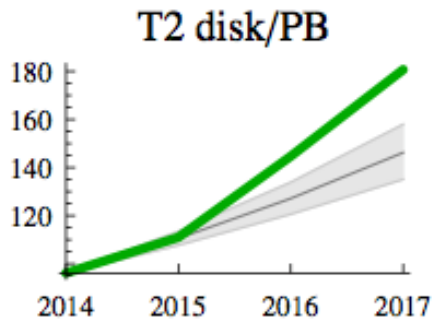


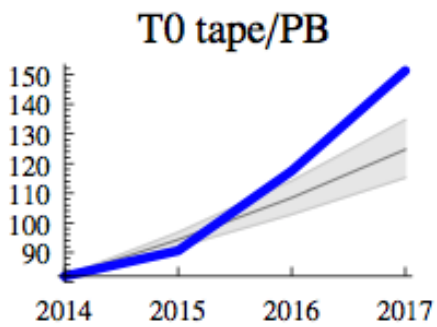


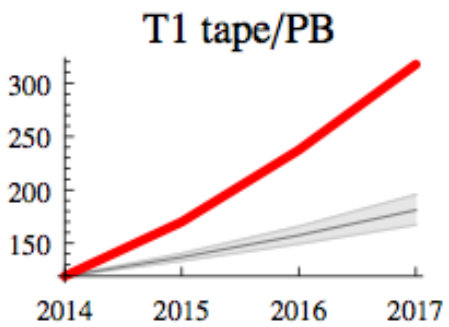


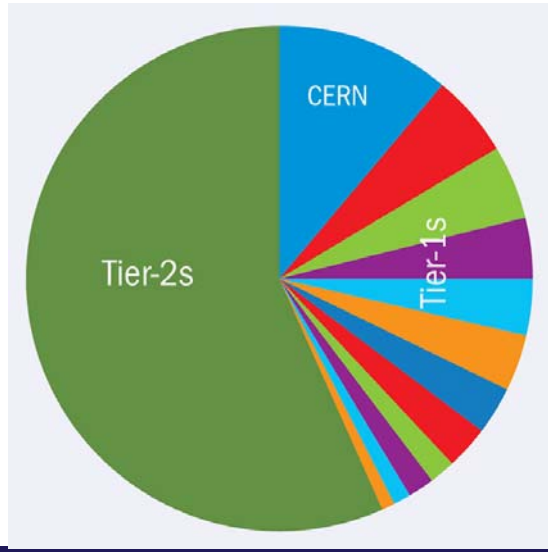












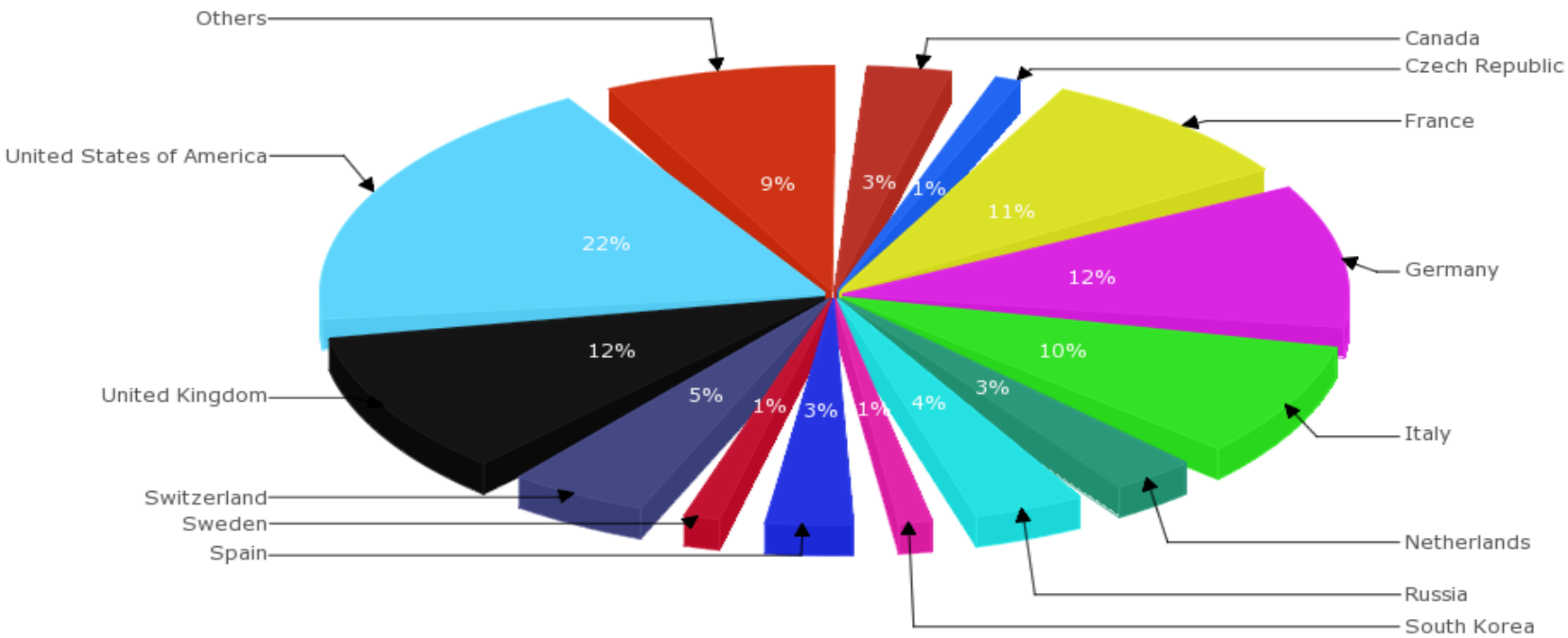
Starting from 2014 pledges

Country Normalized CPU time 2014-2015

CESGA 'EGI View' : / normcpu / 2014:10-2015:5 / COUNTRY-VO / lhc (x) / GRBAR-LIN / l

2015-

COUNTRY Normalised CPU time (kSI2K) per COUNTRY



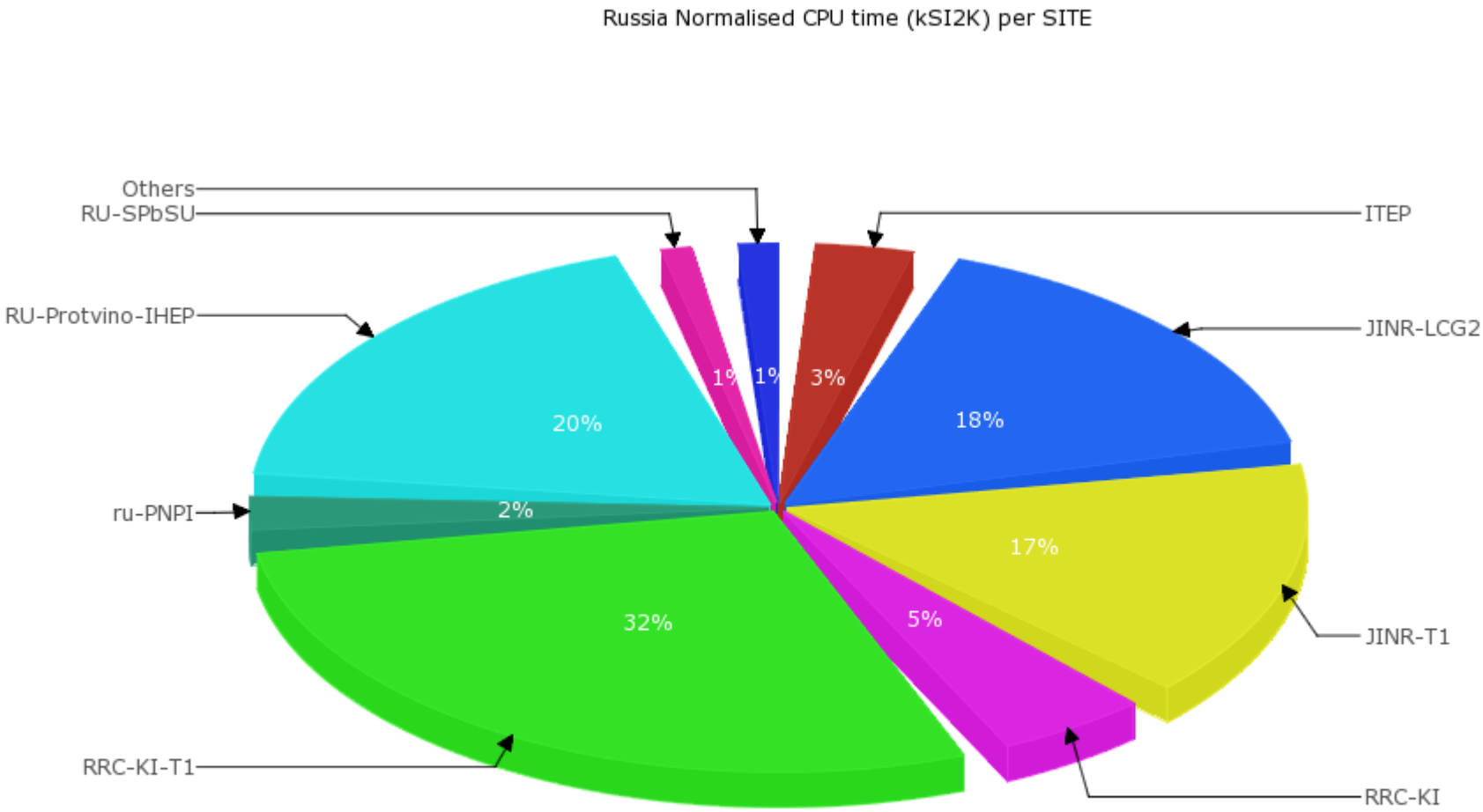
All Country - 29,663,288,664 **Russia- 1,162,595,524**

Job 920,138,350 **30,034,918**

RDIG Normalized CPU time (2015)

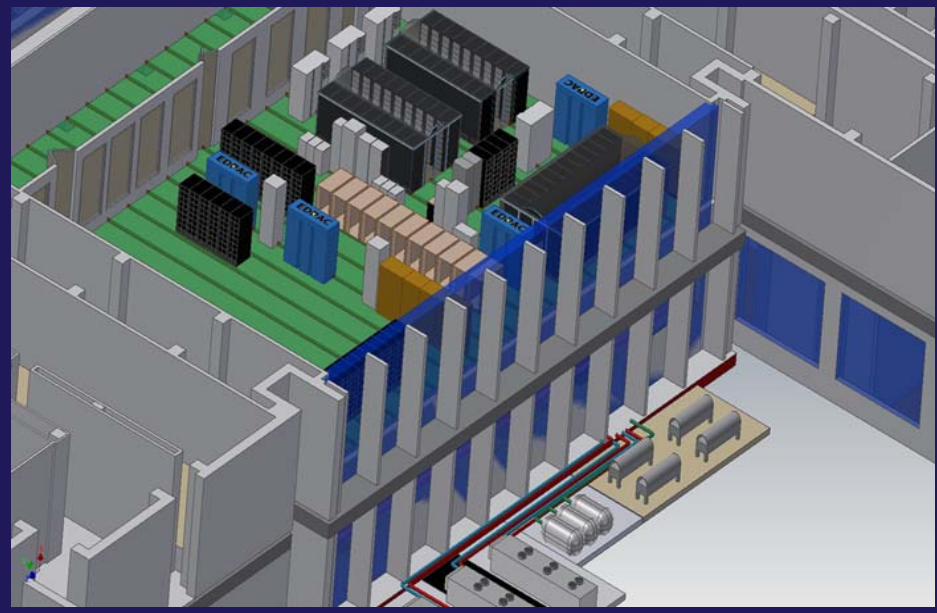
View': / normcpu / 2015:5-2015:6 / SITE-DATE / all (x) / GRBAR-LIN / l

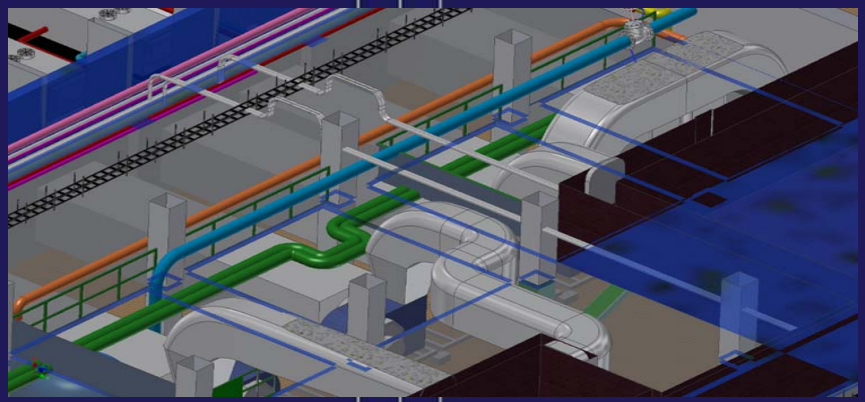
2015-(

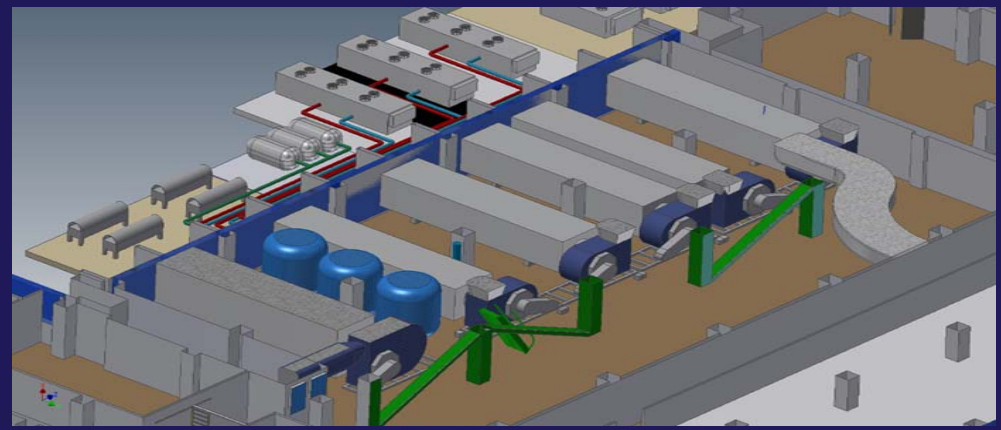


Creation of CMS Tier1 in JINR

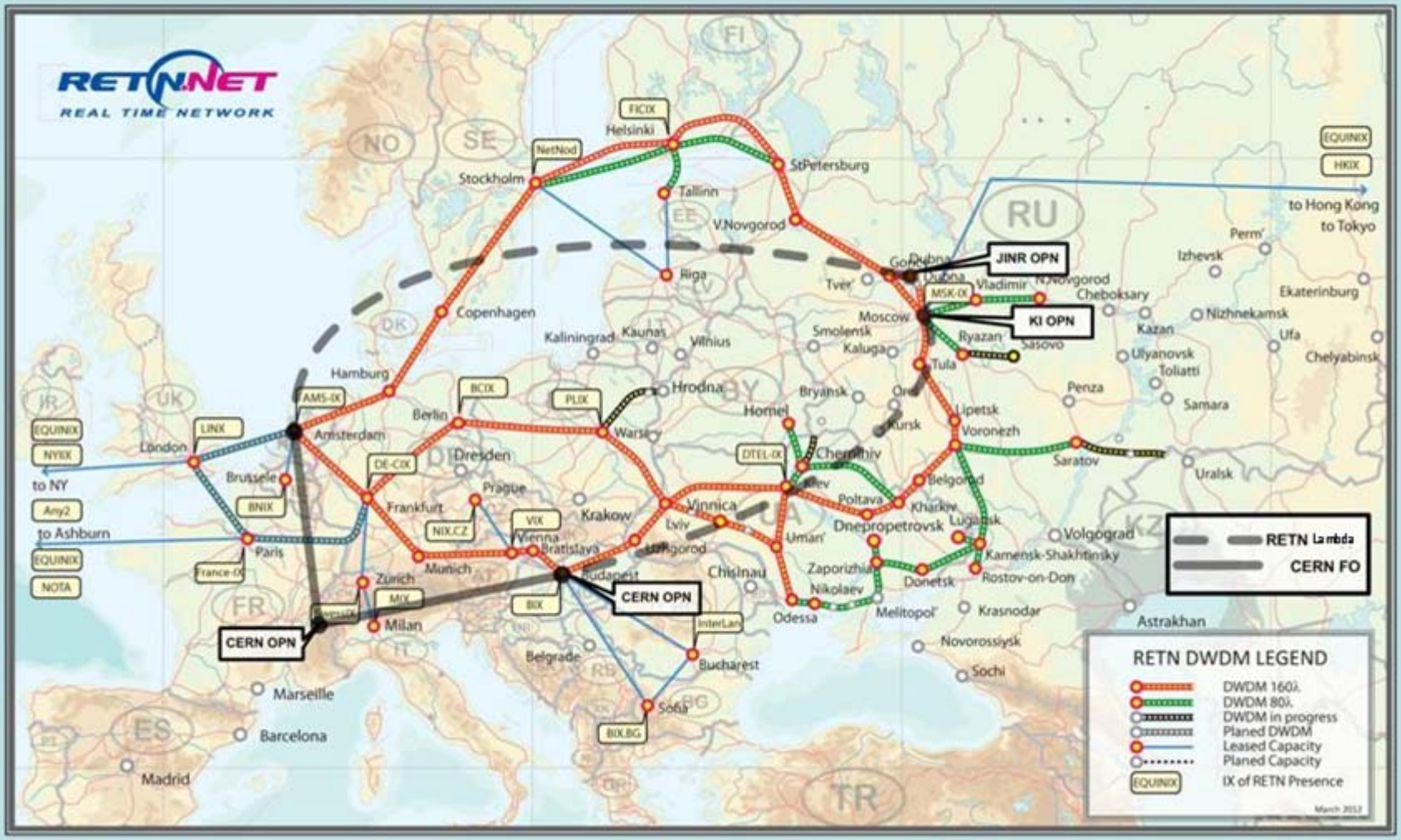
- Engineering infrastructure (a system of uninterrupted power supply, climate - control);
- High-speed reliable network infrastructure with a dedicated reserved data link to CERN (LHCOPN);
- Computing system and storage system on the basis of disk arrays and tape libraries of high capacity;
- 100% reliability and availability.



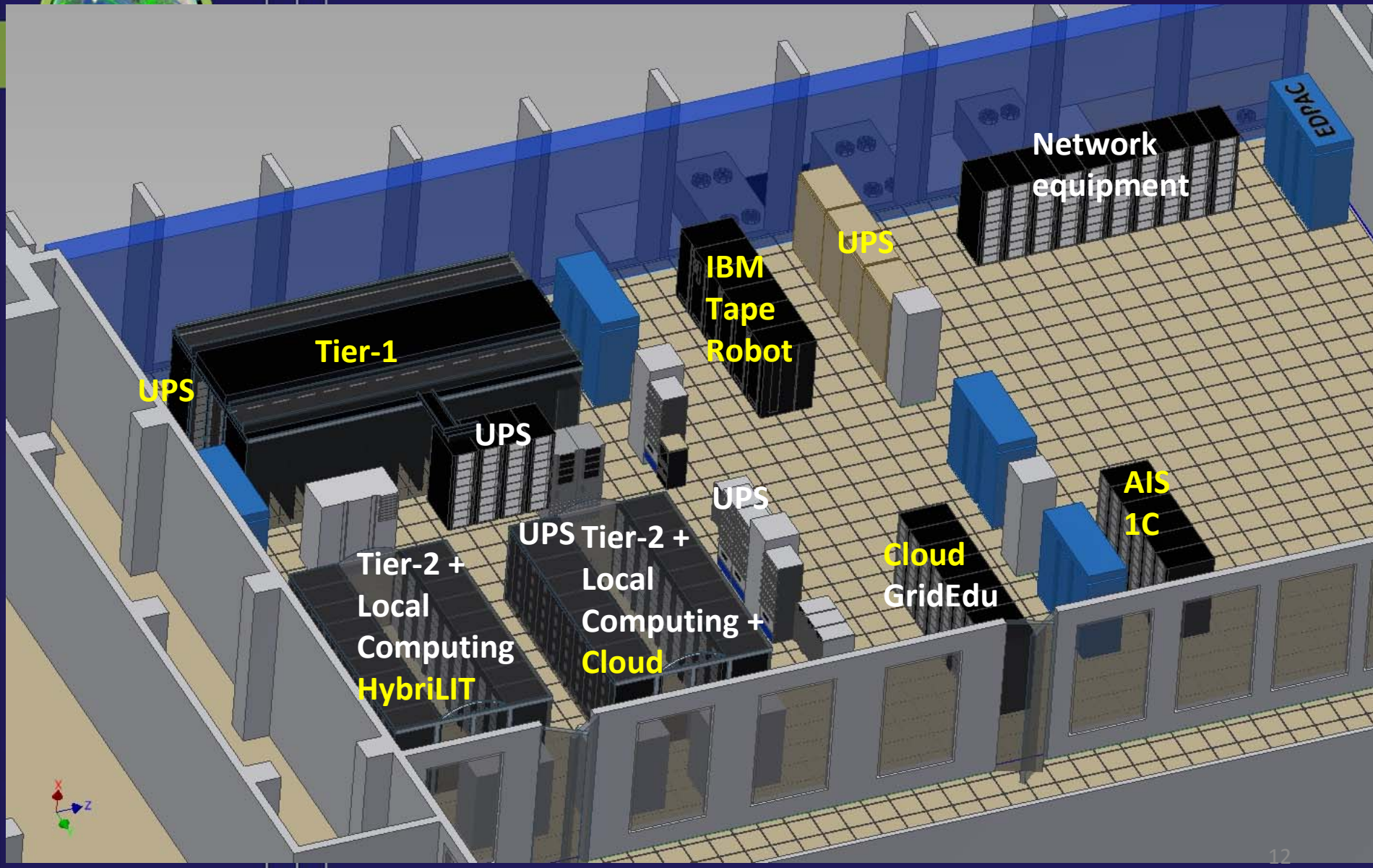




JINR Tier1 Connectivity Scheme



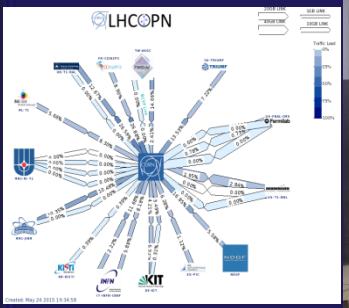
JINR Computing Centre Status

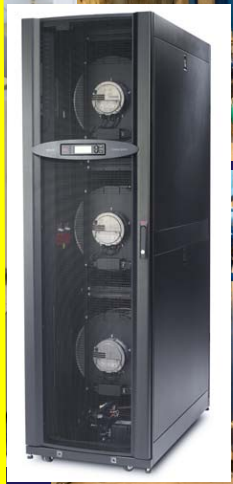


Tier-1 Components

March 2015

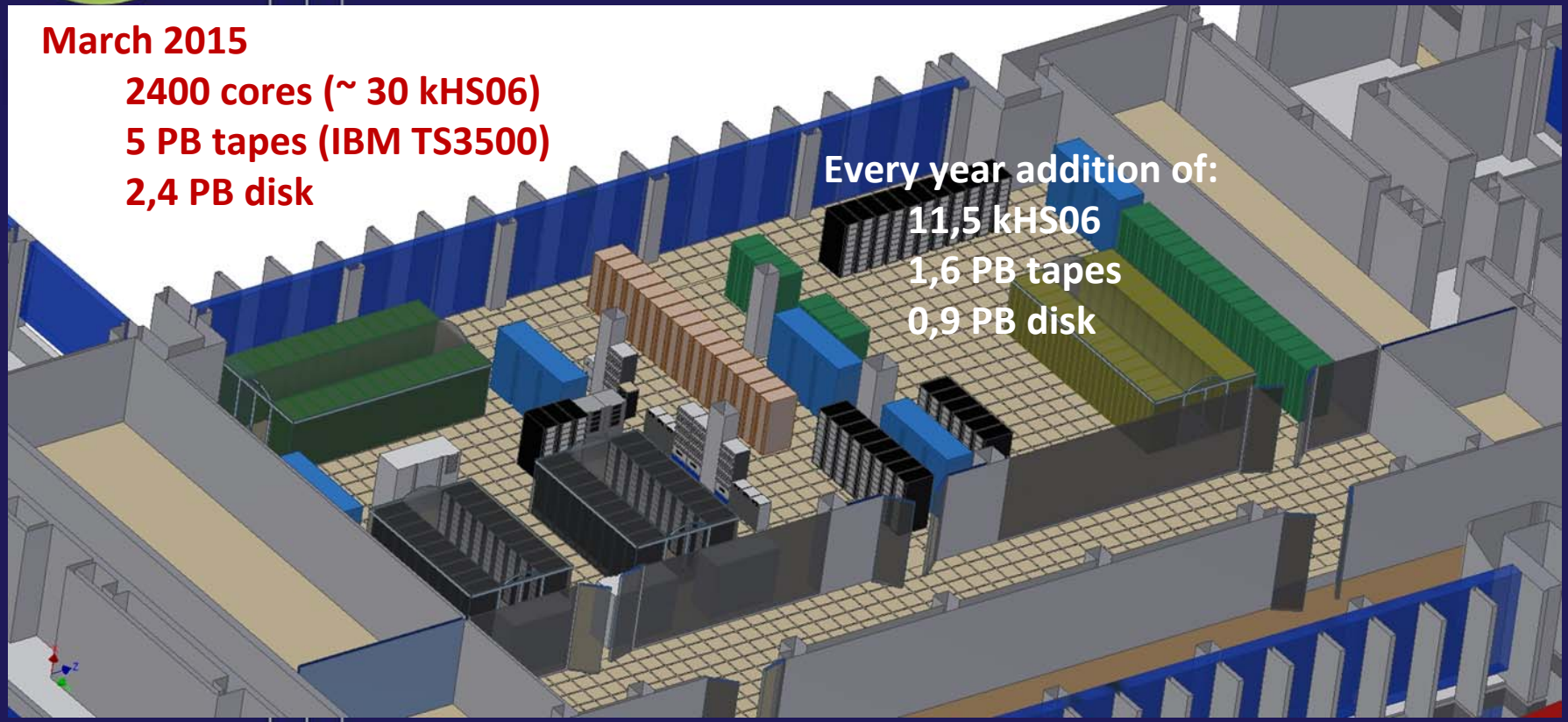
- LHCOPN
- 2400 cores (~ 30 kHS06)
- 5 PB tapes (IBM TS3500)
- 2,4 PB disk
- Close-coupled, chilled water cooling InRow
- Hot and cold air containment system
- MGE Galaxy 7000 – 2x300 kW energy efficient solutions 3Ph power protection with high adaptability





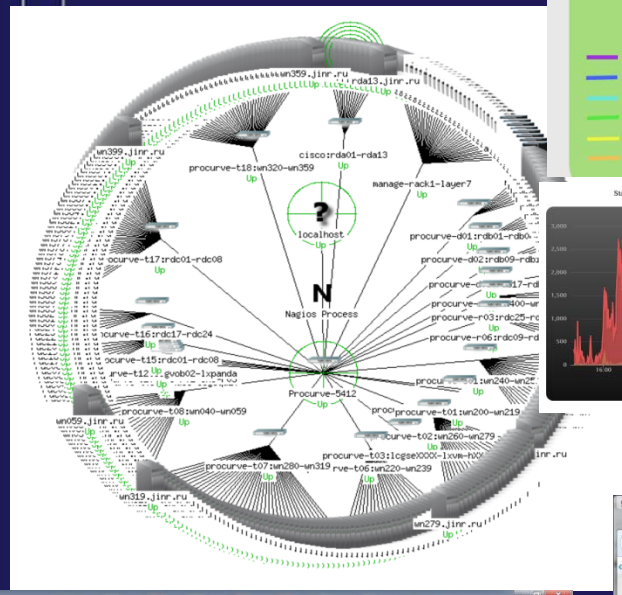
Inauguration of Tier1 CMS center in LIT JINR

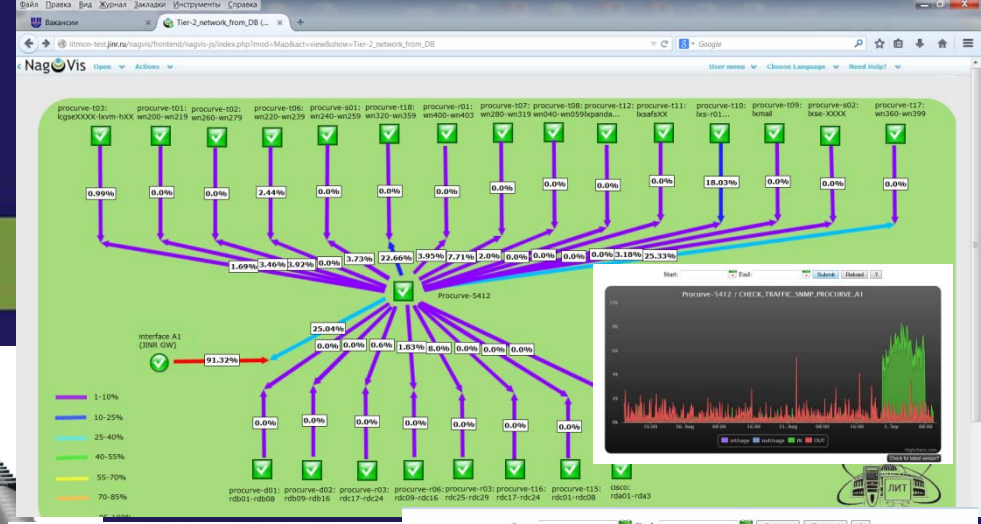
Tier-1 CMS Development

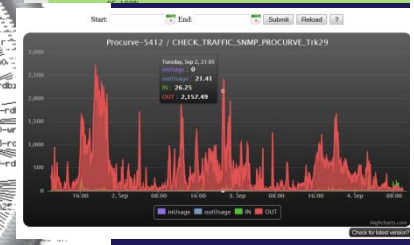


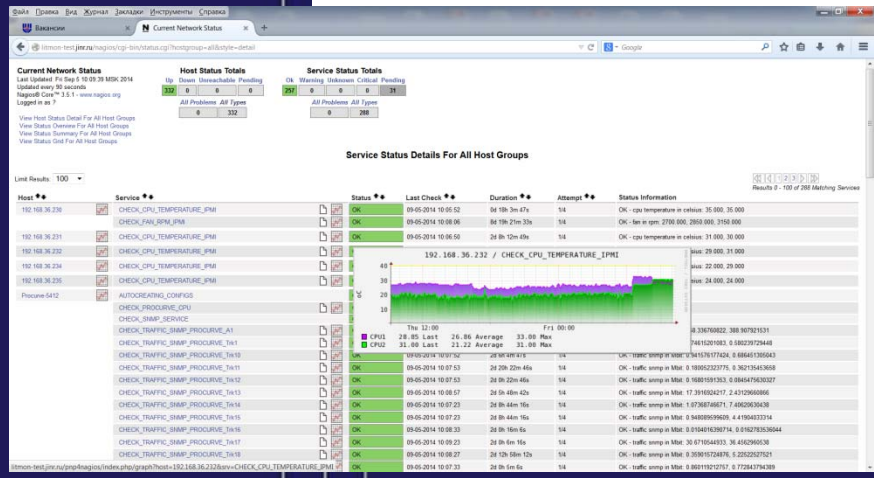
Monitoring

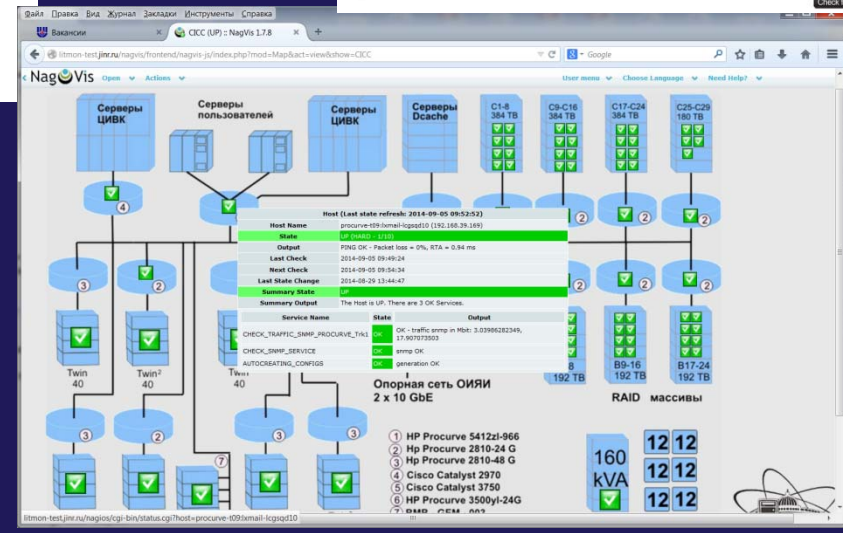
Network monitoring information system - more than 623 network nodes are in round-the-clock monitoring











HybriLIT heterogeneous computing cluster: current state

- **Computing resources:**

CPU Intel Xeon E5-2695v2	**168** cores
GPU K40 & K20	**37248** cores
Intel Xeon Phi 7120P&5110P	**182** cores

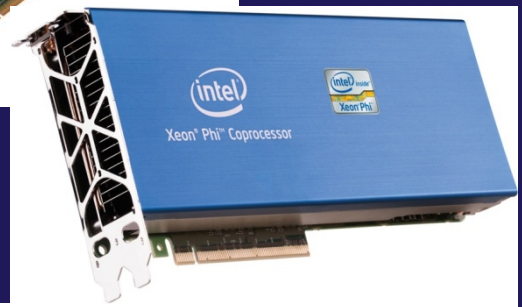
RAM **896** Gb
Disk storage **57** Tbyte
Ethernet
InfiniBand **40** Gb/s

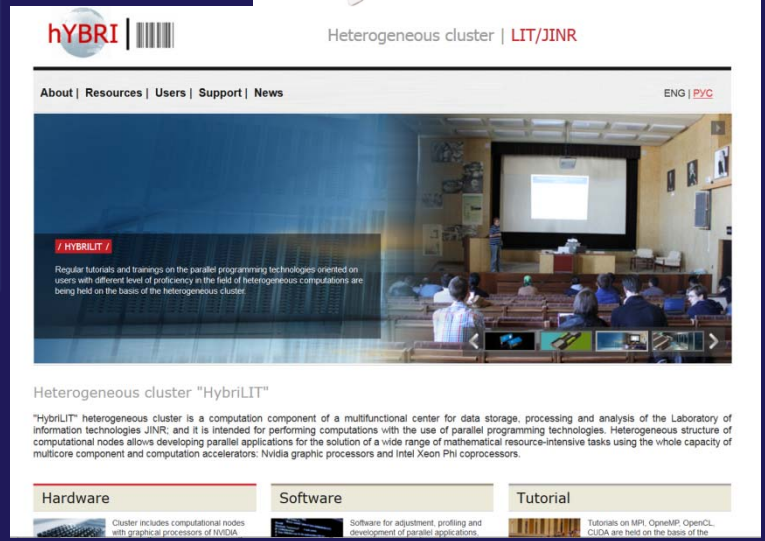
- **Peak performance** for floating point computations

single precision **77** TFLOPS
double precision **29** TFLOPS

Power consumption: 7 kW

Operating system: Scientific Linux 6.5
File systems: EOS and NFS
Batch system: SLURM



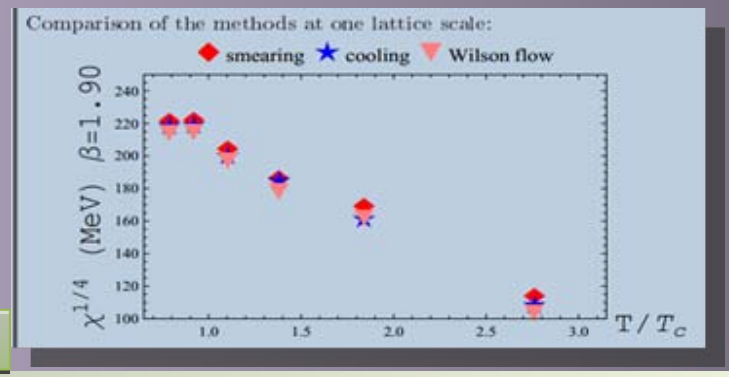


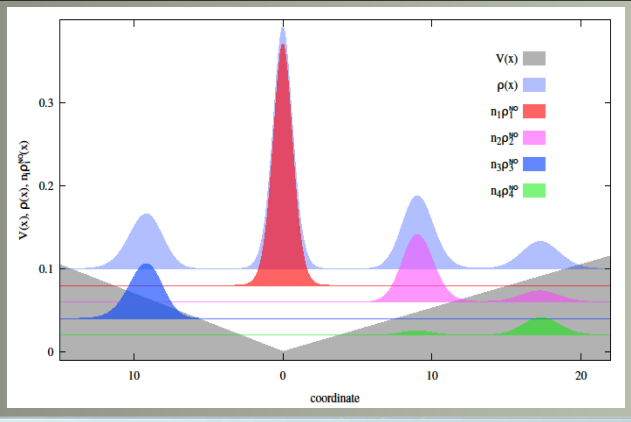
Parallel computing on HybriLIT

Parallel computing for QCD problems:

F. Burger(IP, HU, Berlin,),
M. Müller-Preussker (IP HU, Berlin, Germany),
E.-M. Ilgenfritz (BLTP& VBLHEP, JINR),
A. M. Trunin (BLTP JINR)

http://theor.jinr.ru/~diastp/summer14/program.html#posters





Parallel computing for investigation of Bose-systems:

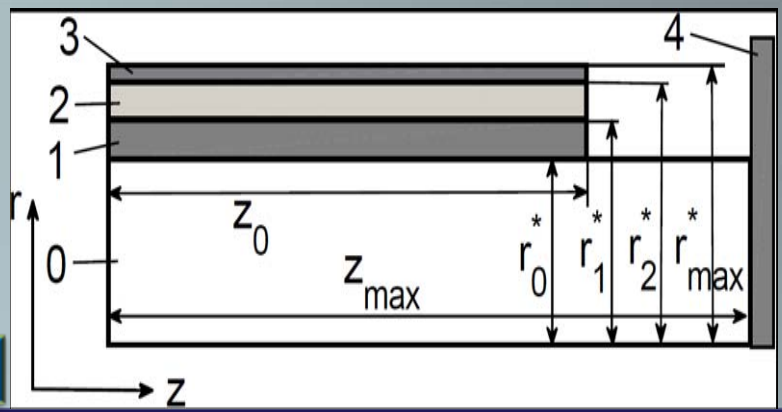
Alexej I. Streltsov ("Many-Body Theory of Bosons" group at CQD, Heidelberg University, Germany),
Oksana I. Streltsova (LIT JINR)

http://MCTDHB.org

Parallel computing for Technical problems:

A. Ayriyan (LIT JINR), J. Busa Jr. (TU of Kŏsice, Slovakia),
E.E. Donets (VBLHEP, JINR),
H. Grigorian (LIT JINR,;Yerevan State University, Armenia),
J. Pribis (LIT JINR; TU of Kŏsice, Slovakia)

arXiv:1408.5853



Training courses on HybriLIT

hYBRI

Parallel programming technologies on hybrid architectures

intel

Intel Xeon Phi

NVIDIA

MPI OpenMP

Cluster Studio

NVIDIA CUDA

OpenCL

7 – 17 July, 2014
Participants From Mongolia, Romania, Russia

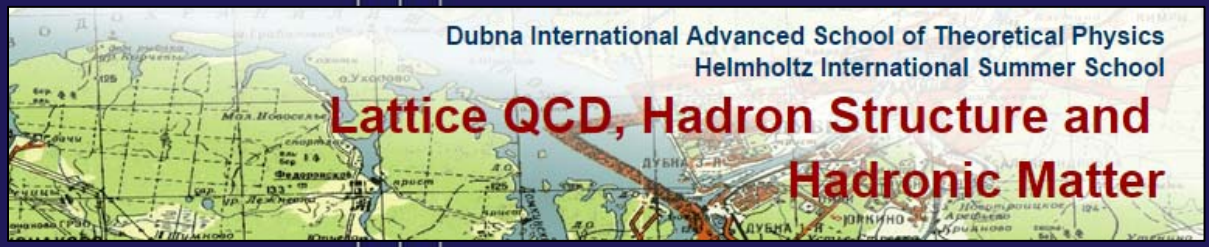
27 August, 2014
Participants from CIS and Russian institutes and companies

International Conference for Young Scientists
«MODERN PROBLEMS OF APPLIED MATHEMATICS & COMPUTER SCIENCE»

MPAMCS 2014

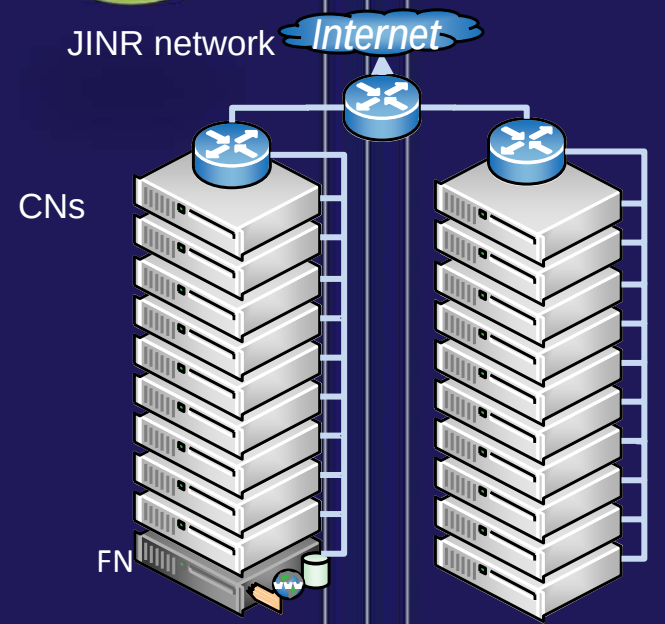
August 25 - 29 2014, Dubna, Russia

1 and 5 September, 2014
Participants from India, Germany, Japan, Ireland, Austria, Ukraine, Russia



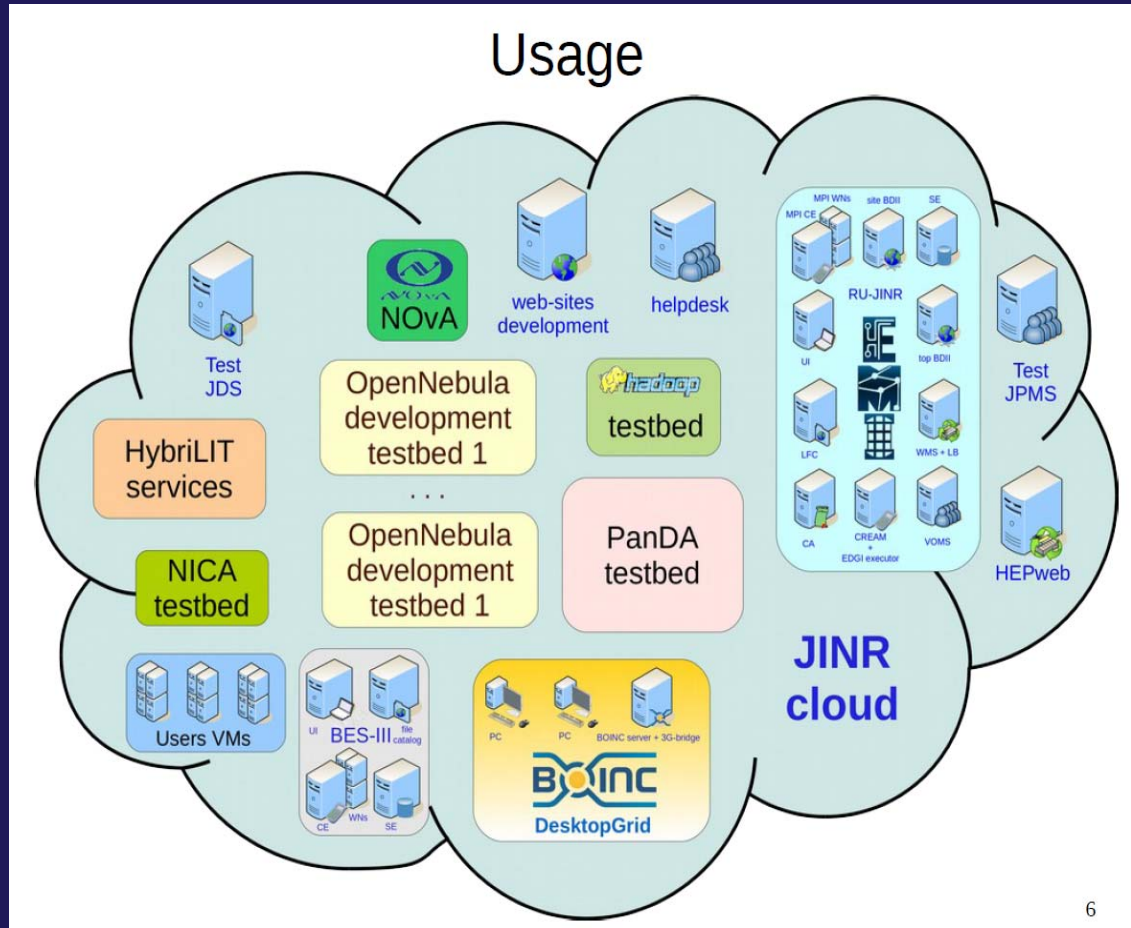
More **100** students and young scientists from Germany, India, Mongolia, Ukraine, Romania, Bulgaria, Moldova , Egypt…

JINR cloud service: current state



FN — front-end node,
CNs — cloud nodes

Cloud characteristics:
- Number of users: 74
- Number of running VMs: 81
- Number of cores: 122
- Occupied by VMs: 134
- Total RAM capacity: 252 GB
- RAM occupied by VMs: 170 GB

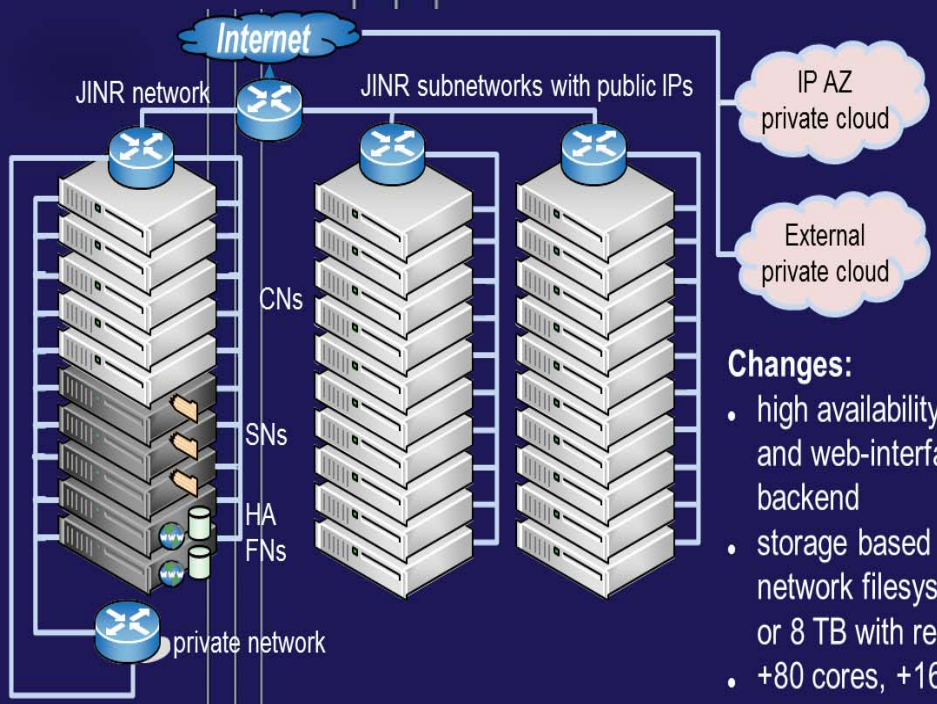


Cloud and heterogeneous cluster development

Advanced cloud infrastructures

–Dynamically reconfigurable computing services

–Large-scale open data repository and access services



HA FNs — highly-available front-end nodes
CNs — cloud nodes
SNs — storage nodes
IP AZ — Institute of physics (Azerbaijan)

Changes:

- high availability for cloud core and web-interfaces as well as DB backend
- storage based on distributed network filesystem (16 TB in total or 8 TB with redundancy=2)
- +80 cores, +160 GB of RAM
- +VMs with private IPs
- connected with external private clouds

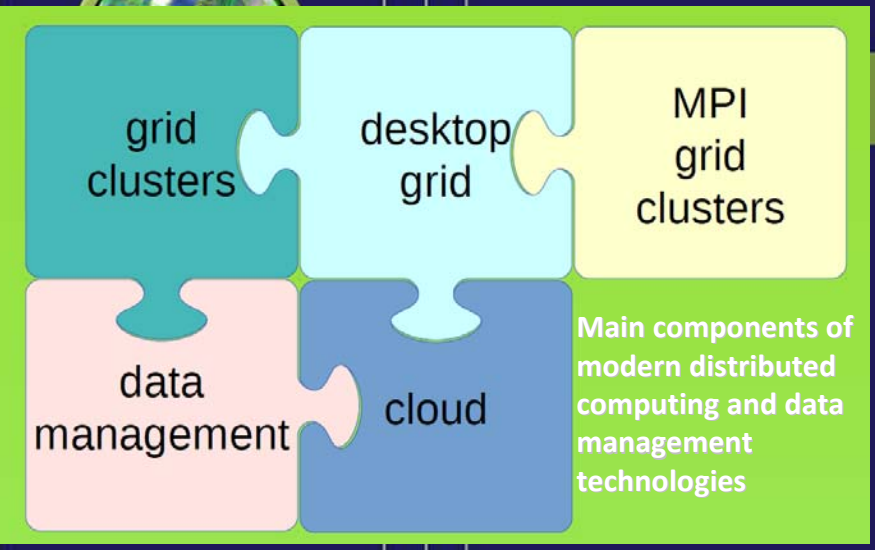
Advanced heterogeneous computing

–User friendly information-computing environment

–New methods and algorithms for parallel hybrid computations

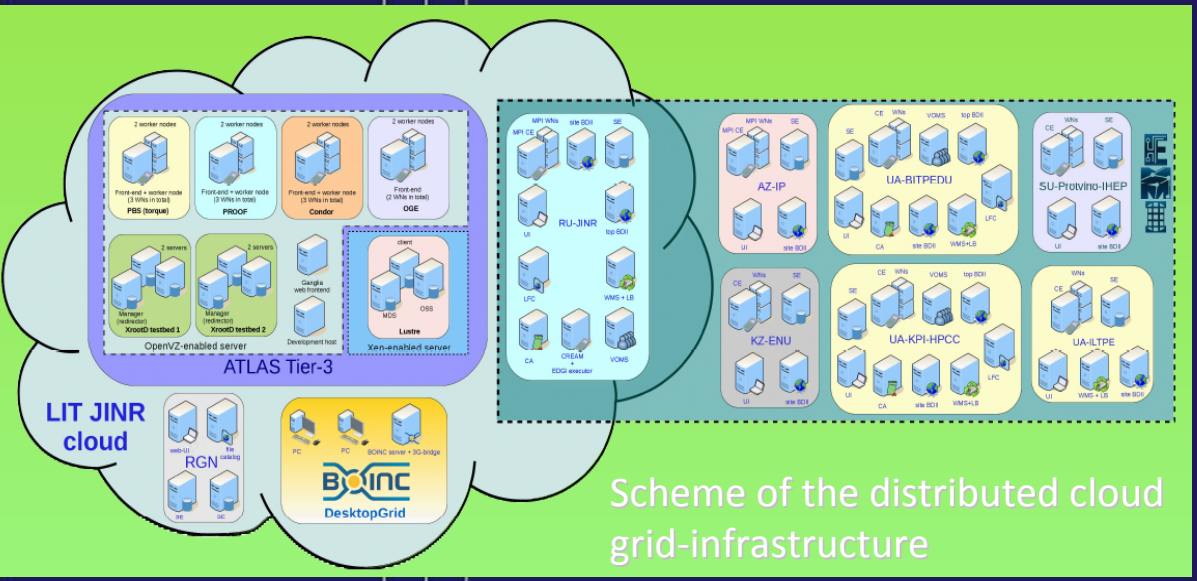
–Infrastructure for tutorials on parallel programming techniques

JINR distributed cloud grid-infrastructure for training and research



Main components of modern distributed computing and data management technologies

There is a demand in special infrastructure what could become a platform for training, research, development, tests and evaluation of modern technologies in distributed computing and data management.
Such infrastructure was set up at LIT integrating the JINR cloud and educational grid infrastructure of the sites located at the following organizations:



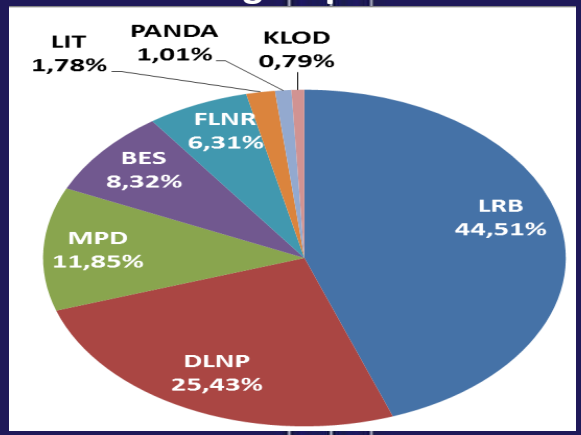
Scheme of the distributed cloud grid-infrastructure

- Institute of High-Energy Physics (Protvino, Moscow region),
- Bogolyubov Institute for Theoretical Physics (Kiev, Ukraine),
- National Technical University of Ukraine "Kyiv Polytechnic Institute" (Kiev, Ukraine),
- L.N. Gumilyov Eurasian National University (Astana, Kazakhstan),
- B.Verkin Institute for Low Temperature Physics and Engineering of the National Academy of Sciences of Ukraine (Kharkov,Ukraine),
- Institute of Physics of Azerbaijan National Academy of Sciences (Baku, Azerbaijan)

JINR Computing Centre for Data Storage, Processing and Analysis

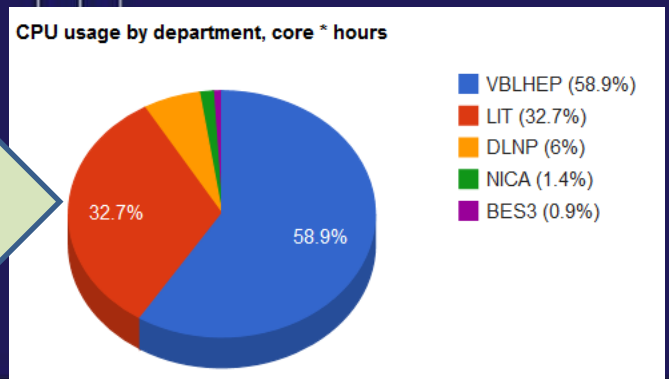
General Purpose Computing Cluster Local users (no grid)

Sharing of the resources according to the processing time among the divisions of the Institute and user groups in 2015.



Cloud Infrastructure

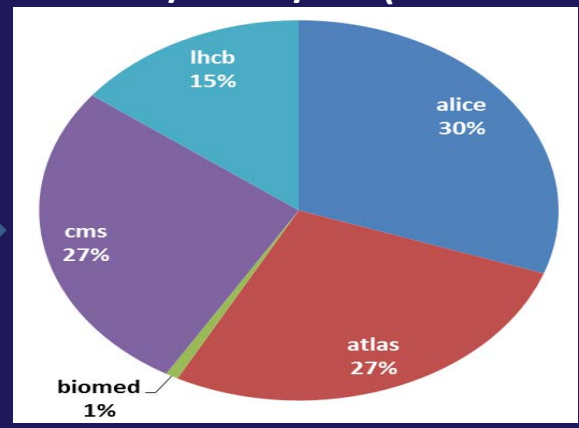
Distribution of cloud resources among the Laboratories and JINR groups in 2015.



Grid-Infrastructure: JINR-LCG2 Tier2 Site JINR-CMS Tier1 Site

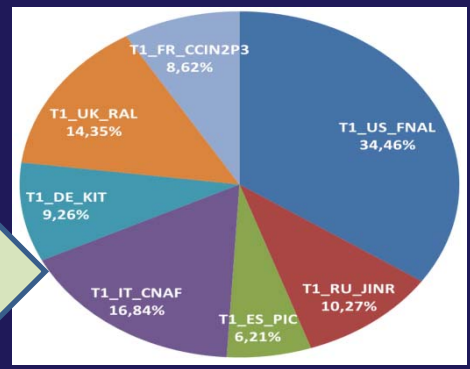
Usage summary of the JINR Tier2 grid-infrastructure by virtual organizations of RDIG/WLCG/EGI (2014-2015)

JINR Tier-2
~ 7 million jobs
~220 million HEPSPEC06-hours



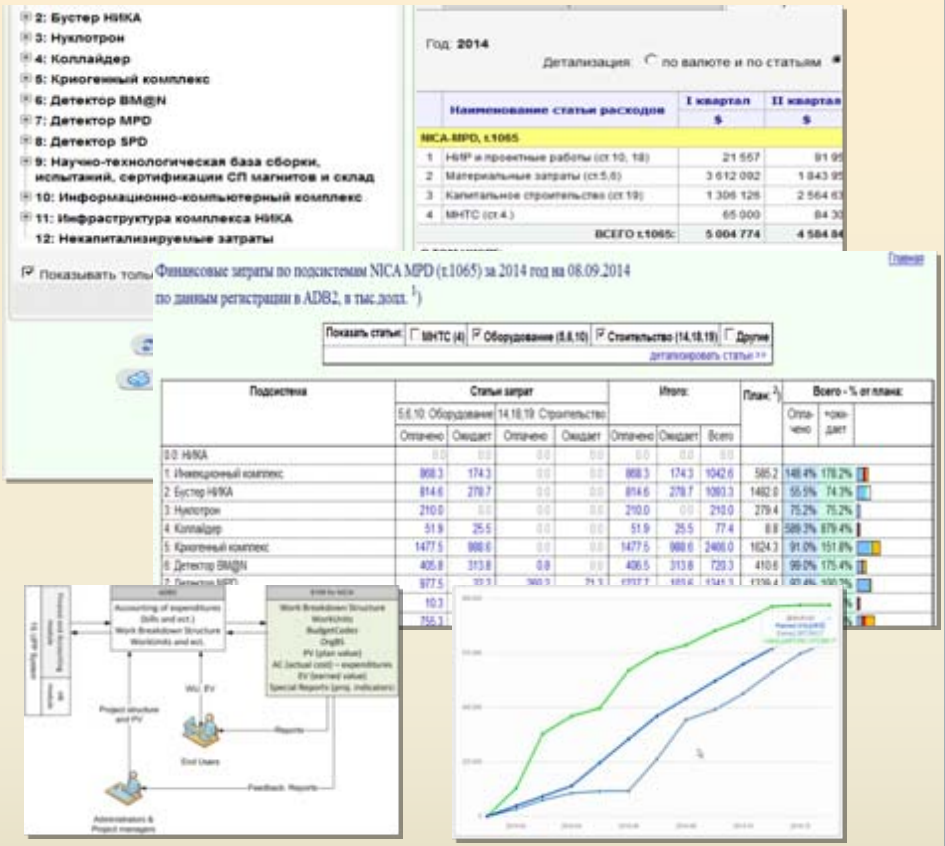
Usage of Tier1 centers by the CMS experiment (last month)

JINR Tier-1 CMS
617 413 jobs



Computing for NICA

Development of management system for NICA project

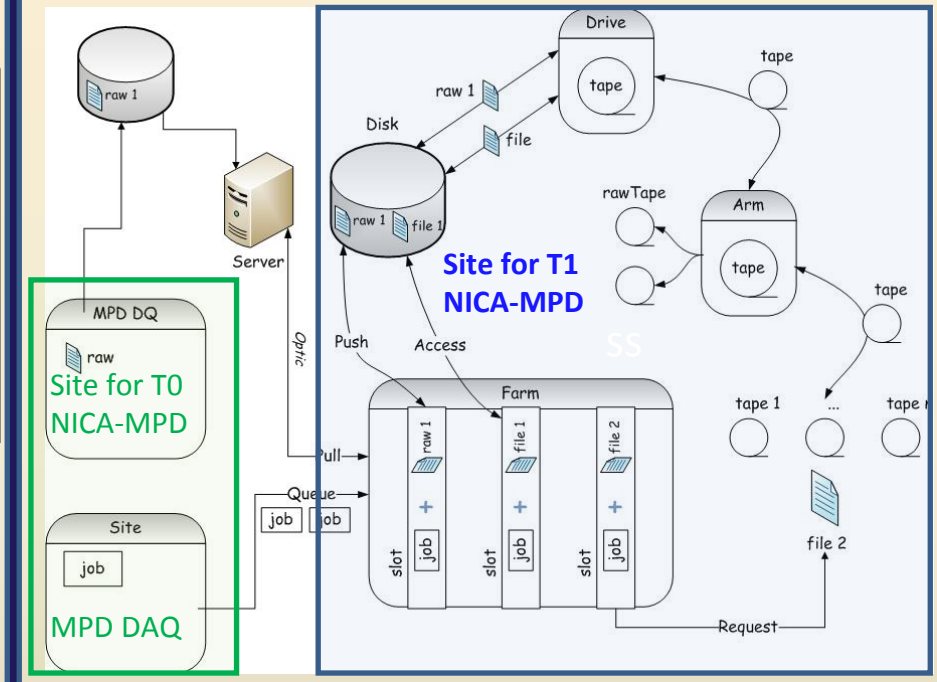


Current status:

- ❑Financial planning and cost control – in production;
- ❑Distributed collection of earned value data – in production;
- ❑Installation of CERN's EVM system at JINR and system integration – finished, in production;
- ❑Development of subsystem for versioning of plans – in progress.

Solution of tasks on processing, storage and security of petabyte data volume of experiments on NICA complex

Aim: get optimal configuration of processors, tape drives, and changers for data processing

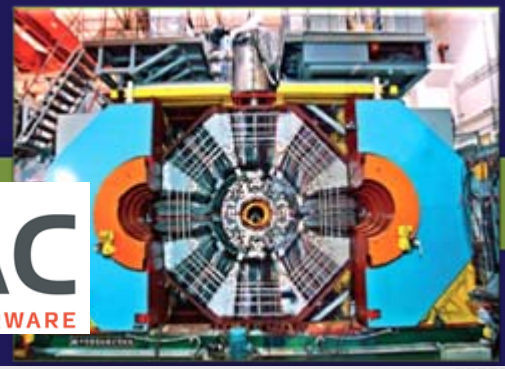


Job & data flow scheme of T0-T1 NICA-MPD

Under study structure composition:

- ✓Tape robot,
- ✓Disk array,
- ✓CPU Cluster.

LIT JINR - China collaboration



LIT team is a key developer of the BES-III distributed computing system

A prototype of BES-III Grid has been built (9 sites including IHEP CAS and JINR). Main developments have been done at IHEP and JINR. The Grid is based on DIRAC interware.

Monitoring

- BES-III grid monitoring system is operational since February 2014.
- Implementation of the new monitoring system based on DIRAC RSS service are in progress

Job management

- Advising on the CE's installation and management
- BES-III jobs can be submitted on JINR cloud service now

Data management

- Installation package for Storage Element was adopted for BES-III Grid
- Solution on dCache-Lustre integration was provided for main data storage in IHEP
- Research on the alternative DB and data management service optimization is in progress

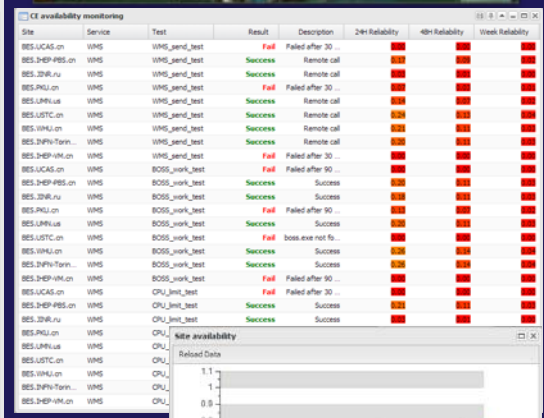
Infrastructure

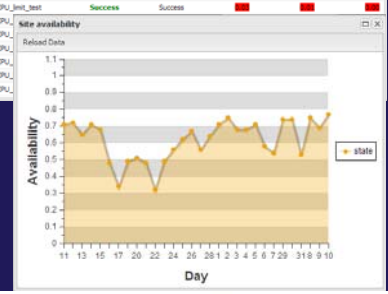
- Creation of the back-up DIRAC services for BES-III grid at JINR is in progress

DIRAC
THE INTERWARE

SE latency monitoring

Source	Destination	Latency(sec)
IHEPD-USER	IHEPD-USER	2.678
IHEPD-USER	JINR-USER	16.316
IHEPD-USER	USTC-USER	15.932
IHEPD-USER	WHU-USER	6.728
JINR-USER	IHEPD-USER	14.322
JINR-USER	JINR-USER	14.24
JINR-USER	USTC-USER	14.827
JINR-USER	WHU-USER	8.516
USTC-USER	IHEPD-USER	3.677
USTC-USER	JINR-USER	17.855
USTC-USER	USTC-USER	2.746
USTC-USER	WHU-USER	624.375
WHU-USER	IHEPD-USER	5.727
WHU-USER	JINR-USER	20.227
WHU-USER	USTC-USER	9.199
WHU-USER	WHU-USER	3.092





Worldwide LHC Computing Grid (WLCG)

The primary goal of the WLCG project is to create a global infrastructure of regional centers for processing, storage and analysis of data of the LHC physical experiments.

The grid-technologies are a basis for constructing this infrastructure.

A protocol between CERN, Russia and JINR on participation in the LCG project was signed in 2003. MoU about participation in the WLCG project was signed in 2007.

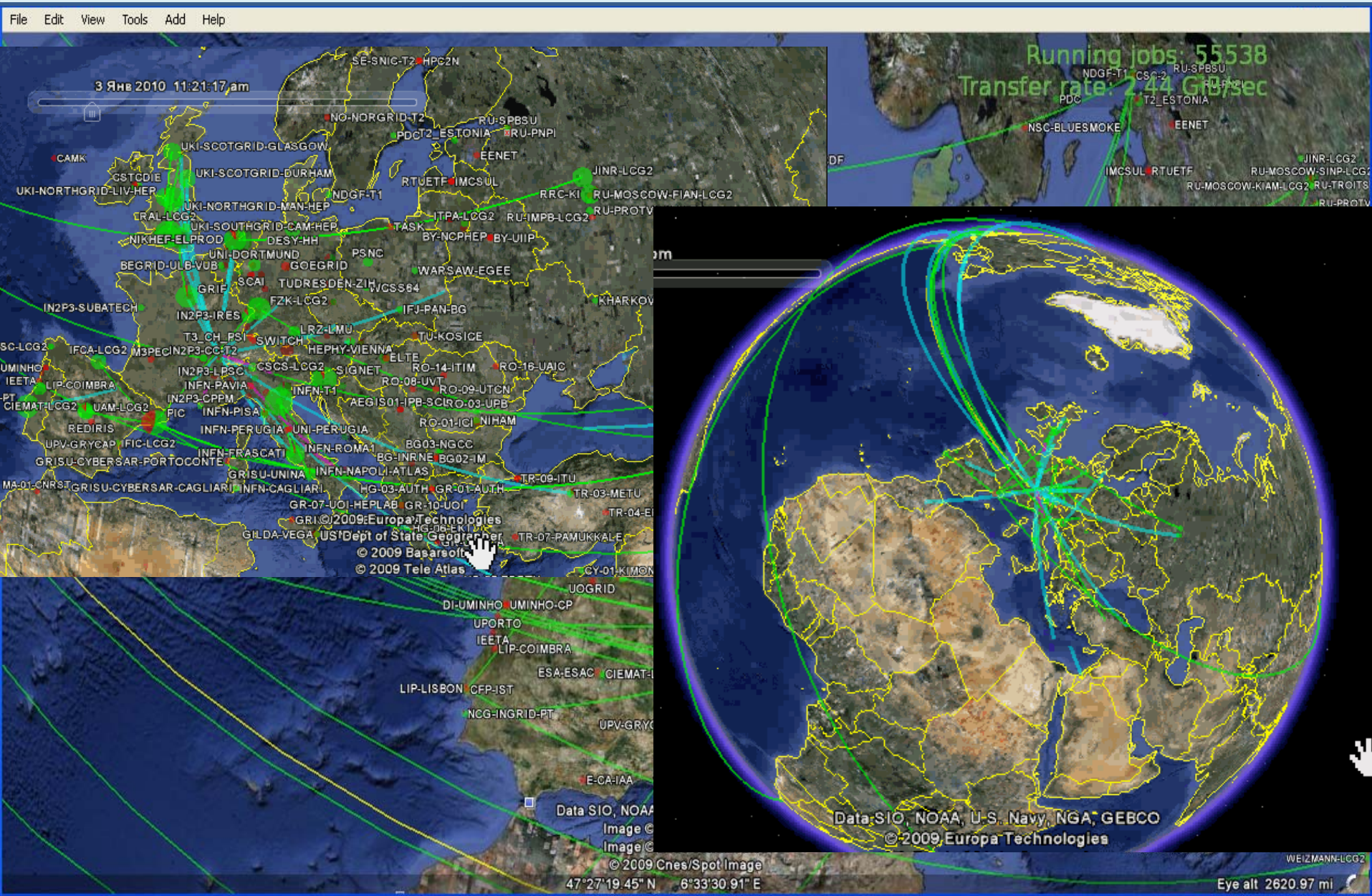
Tasks of the Russian centers and JINR within WLCG :

- Creation of a complex of tests for WLCG software
- Introduction of WLCG services for experiments
- Development of WLCG monitoring systems
- Development of simulation packages for experiments
- Creation of a Tier1 center in Russia

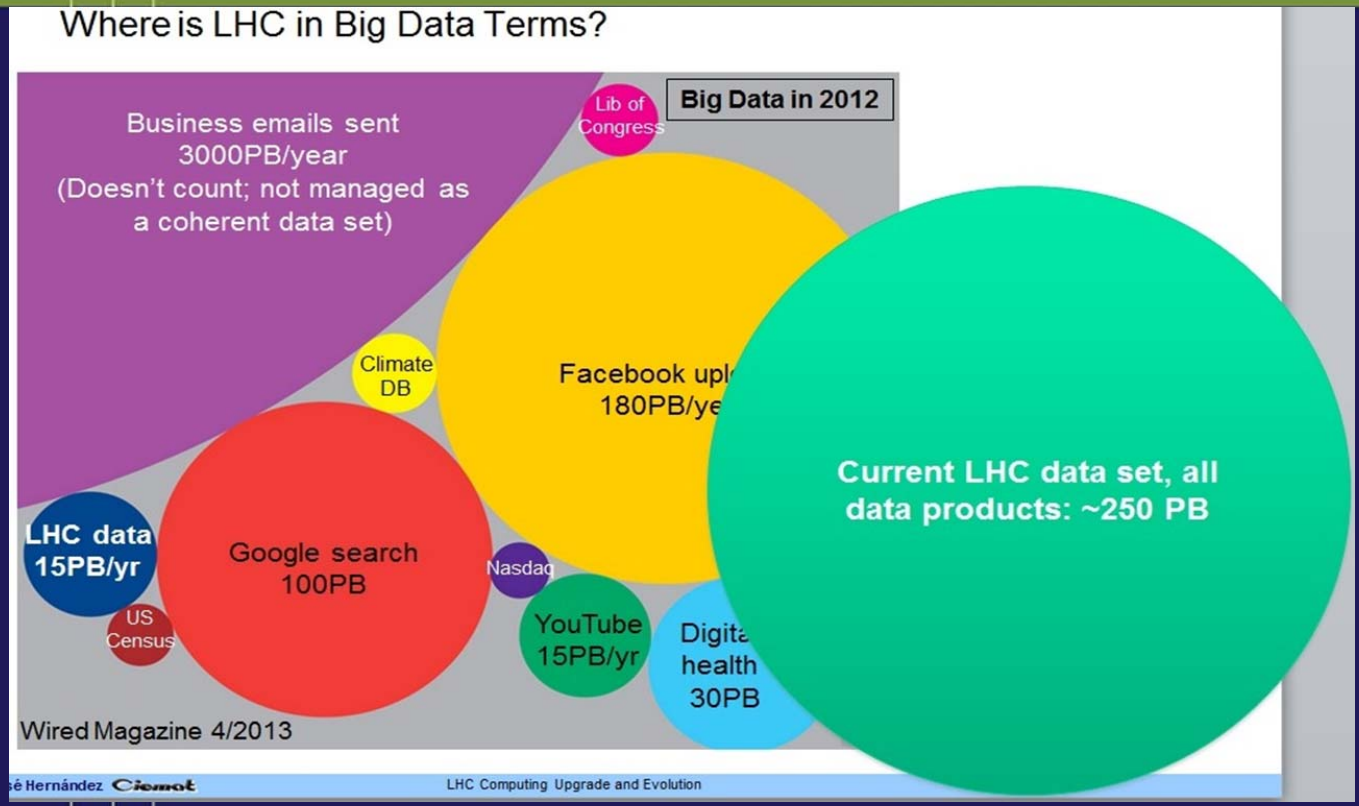
JINR activity at WLCG project

- Participation in development of software for ATLAS, ALICE, CMS
- Development WLCG Dashboard
- Global data transfer monitoring system for WLCG infrastructure
- NOSQL storage
- Integration GRID, Cloud, HPC
- Local and global Monitoring of Tier3 centers
- Development of DDM, AGIS for ATLAS
- GENSER & MCDB

WLCG Google Earth Dashboard



Entering into the era of Big Data

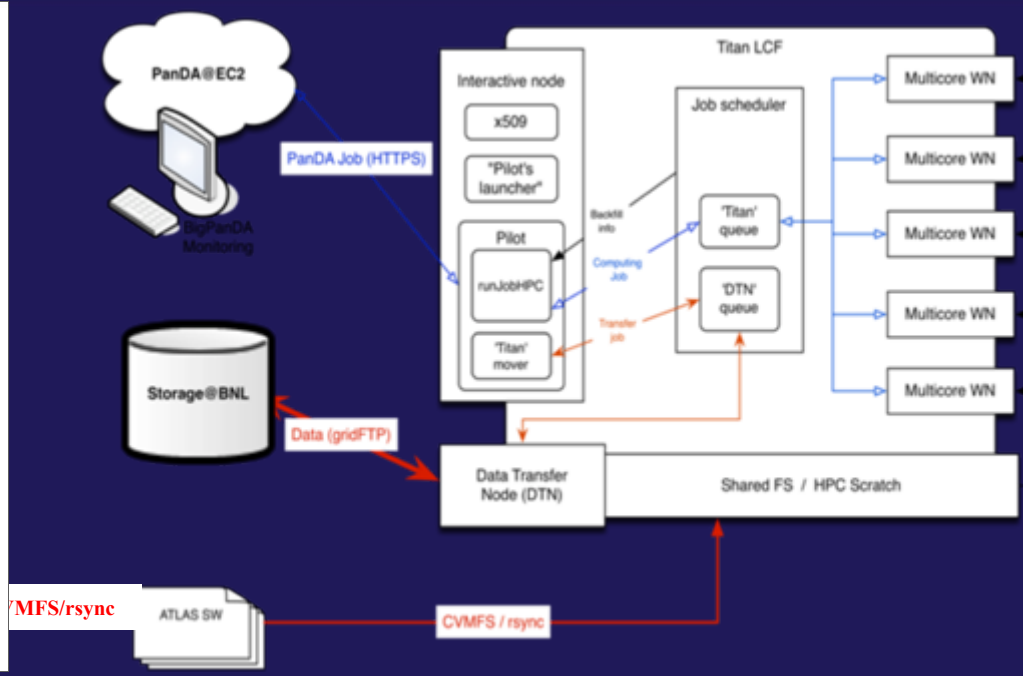


A comparative diagram of processed data evidently shows that the studies underway at CERN are performed under Big Data conditions.

After LHC modernization and start-up in 2015, the data stream will increase 2.5 times thus demanding increase in the resources and optimization of their use.

Evolving PanDA for Advanced Scientific Computing

	Titan System (Cray XK7)		
Peak Performance	27.1 PF 18,688 compute nodes	24.5 PF GPU	2.6 PF CPU
System memory	710 TB total memory		
Interconnect	Gemini High Speed Interconnect	3D Torus	
Storage	Lustre Filesystem	32 PB	
Archive	High-Performance Storage System (HPSS)	29 PB	
I/O Nodes	512 Service and I/O nodes		



ATLAS (BNL, UTA), OLCF, ALICE (CERN,LBNL,UTK), LIT JINR:

- adapt PanDA for OLCF (Titan)
- reuse existing PanDA components and workflow as much as possible.
- PanDA connection layer runs on front-end nodes in user space. There is a predefined host to communicate with CERN from OLCF, connections are initiated from the front-end nodes
- SAGA (a Simple API for Grid Applications) framework as a local batch interface.
- Pilot (payload submission) is running on HPC interactive node and communicating with local batch scheduler to manage jobs on Titan.
- Outputs are transferred to BNL T1 or to local storage

Main objective of the 7-year plan

Creation of a **unified information environment** integrating a number of various technological solutions, concepts, techniques, and software in order to offer **optimal approaches** for solving various types of **scientific and applied** tasks on a global level of the development of advanced information and computation technologies

Unified environment

- Grid
- Supercomputer (heterogeneous)
- Cloud
- Local computing cluster
- ….

Requirements:

- scalability
- interoperability
- adaptability to new technical solutions.
- operates 12 months a year in a 24x7 mode

CICC to MICC

Build up the Multifunctional Information and Computing Complex (MICC)

- fault-tolerant infrastructure with electrical power storage and distribution facilities with expected availability of 99.995%,
- supports and uses a large variety of architectures, platforms, operational systems, network protocols and software products
- provides means for organization of collective development
- supports solution of problems of various complexity and subject matter
- enables management and processing of data of very large volumes and structures (Big Data)
- provides means to organize scientific research processes
- enables training IT infrastructure users

Multifunctional Information&Computing Complex

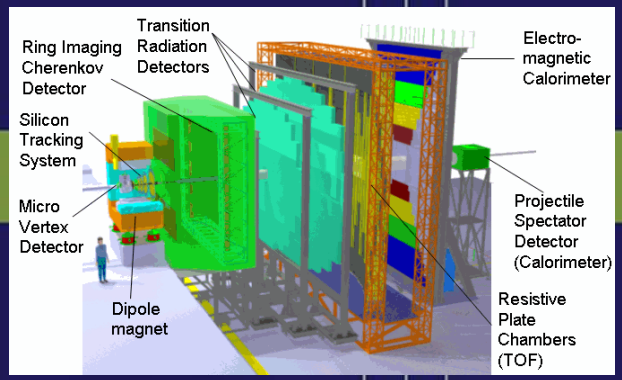
- Engineering infrastructure
- Local network infrastructure and telecommu-nication data links
- Tier1 level grid automated system of data processing of the CMS experiment on the Large Hadron Collider (LHC) , including that as a prototype of the system of data storage and processing of the NICA experiments in a role of the center of Tier0 and Tier1 levels
- Tier-2 level grid-system to support LHC experiments (ATLAS, ALICE, CMS, LHCB), FAIR (PANDA) and other large-scale experiments and projects within the global grid-infrastructure
- High-performance computing system (including parallel computations) beyond the range of heterogeneous and grid systems
- Heterogeneous computer complex for high-efficiency calculations
- Cloud environment

SOFTWARE

Parallel software will be the mainstream:

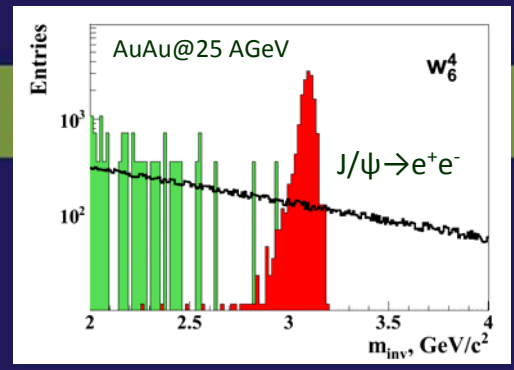
- development and support of the program libraries of general and special purpose;
- creation and support of program libraries and software complexes realized on the parallel programming technologies CUDA, OpenCL, MPI+CUDA, etc.;
- support and development of a specialized service-oriented environment for modeling experimental installations and processes and experimental data processing;
- tools and methods for software development:
  - flexible, platform-independent simulation tools
  - self-adaptive (data-driven) simulation development software

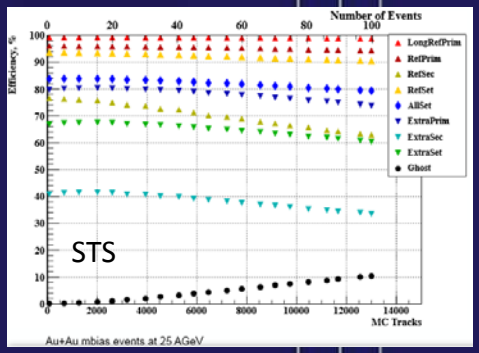
CBM@GSI – Methods, Algorithms & Software for Fast Event Reconstruction

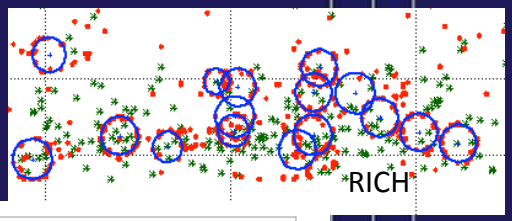


Tasks:

- global track reconstruction;
- event reconstruction in RICH;
- electron identification in TRD;
- clustering in MVD, STS and MUCH;
- participation in FLES (First Level Event Selection);
- development of the Concept of CBM Databases;
- magnetic field calculations;
- beam time data analysis of the RICH and TRD prototypes;
- contribution to the CBMROOT development;
- D0-, vector mesons, $J/\psi \rightarrow e^+e^-$ and $J/\psi \rightarrow \mu^+\mu^-$ reconstruction;

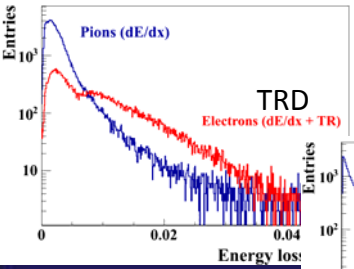


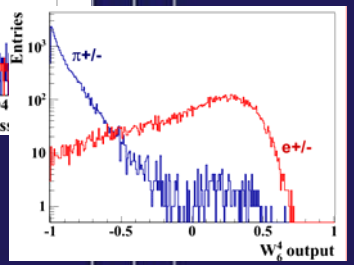




a: $S/Bg_{2\sigma}$, b: Efficiency (%),
c: J/ψ per hour (10 Mhz)

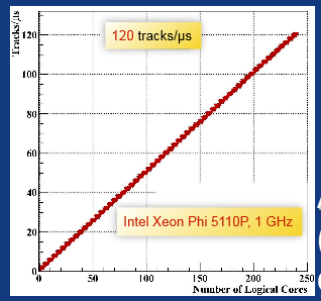
	a	b	c
pC@30GeV	14	22	11
pAu@30GeV	18	22	27
AuAu@10AGeV	0.18	18	64
AuAu@25AGeV	7.5	13.5	5250





Modern parallelization involves multiplicative effects coming from:

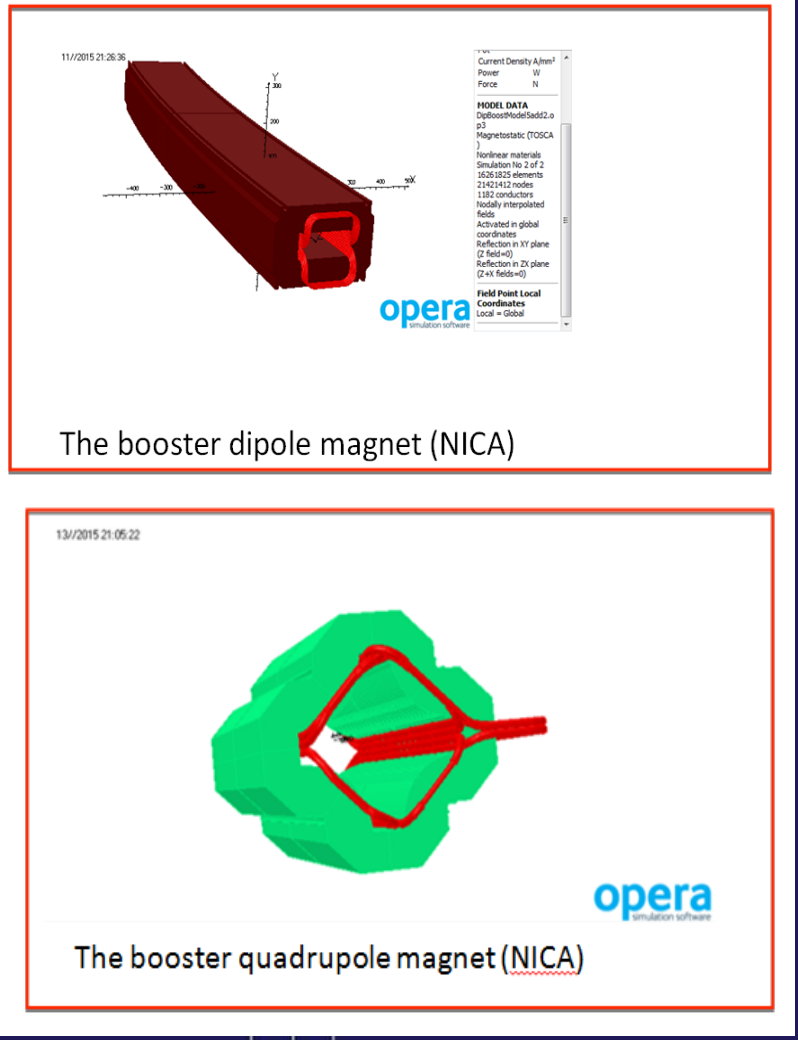
1) Vectorization (SIMD - Single Instruction Multiple Data) factor 2 to 4;
2) Multithreading – factor 4/3 ; 3) υ -Many core processor – factor υ. Total ≈ 4 υ



STS: CA	STS: Kalman Filter	RICH: ring reconstruct.	TRD: track reconstruct.	TRD: el. id. ω(k,n) criterion	KFPar - ticle
164.5	0.5	49.0	1390	0.5	2.5

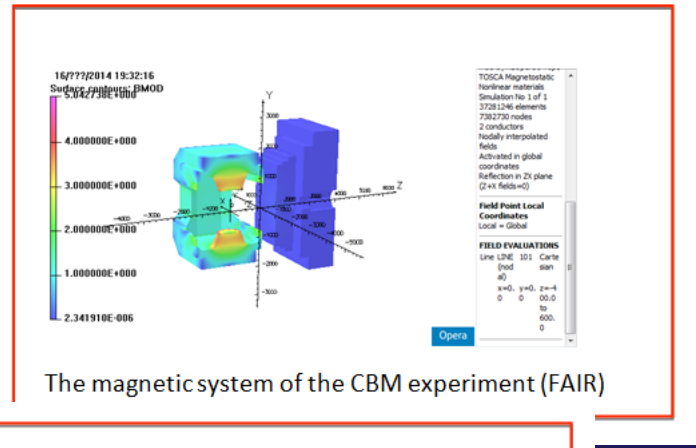
Average time per core (μs/track or μs/ring) of SIMD-algorithms (besides track reconstruction in the TRD) for data processing. Global throughput increases linearly with the number of cores.

The 3D modeling of the magnetic systems

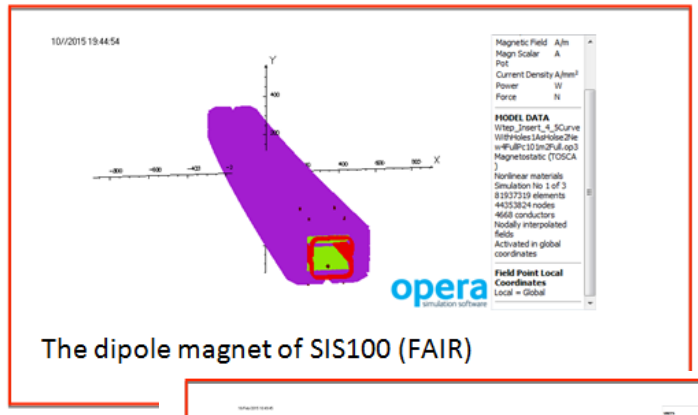


The booster dipole magnet (NICA)

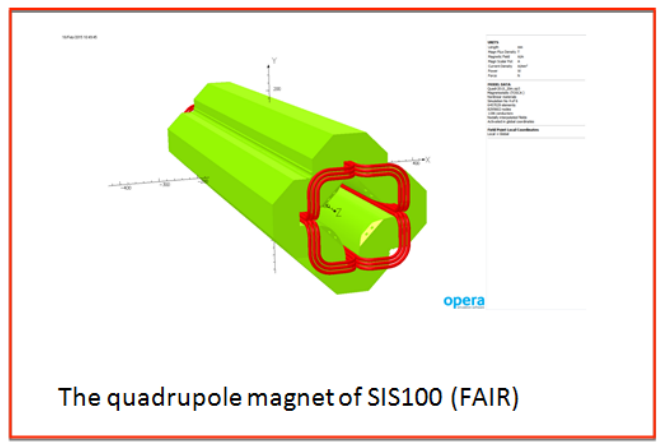
The booster quadrupole magnet (NICA)



The magnetic system of the CBM experiment (FAIR)



The dipole magnet of SIS100 (FAIR)



The quadrupole magnet of SIS100 (FAIR)

HepWeb Overview

http://hepweb.jinr.ru/

Provides: WEB access to computing resources of LIT for Monte Carlo simulations of hadron-hadron, hadron-nucleus, and nucleus-nucleus interactions, by means of most popular generators.
Realization: service - oriented architecture.

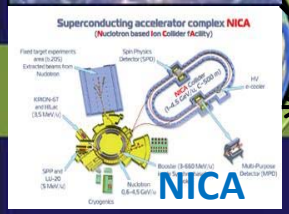
Goals:

- Monte Carlo simulations at the server
- Provide physicists with new calculation/simulation tools
- Mirror site of GENSER of the LHC Computing GRID project
- Provide physicists with informational and mathematical support
- Introduce young physicists into HEP world

Improvement of QGSp in Geant4

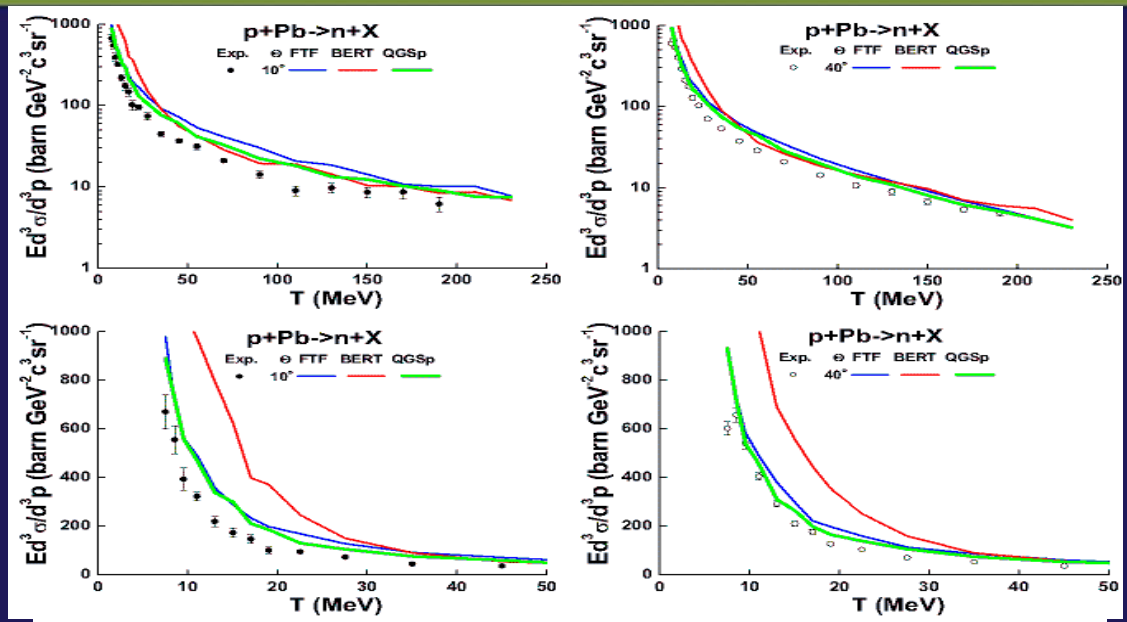
[Author of original code – N.S. Amelin (LIT, JINR)]
Developer – V.V. Uzhinsky (LIT, JINR)

FAIR
Facility for Antiproton and Ion Research in Europe GmbH



Geant 4

Physics List – QGSp_BERT used by ATLAS and CMS



Slow neutron production, ITEP experimental data (1983)

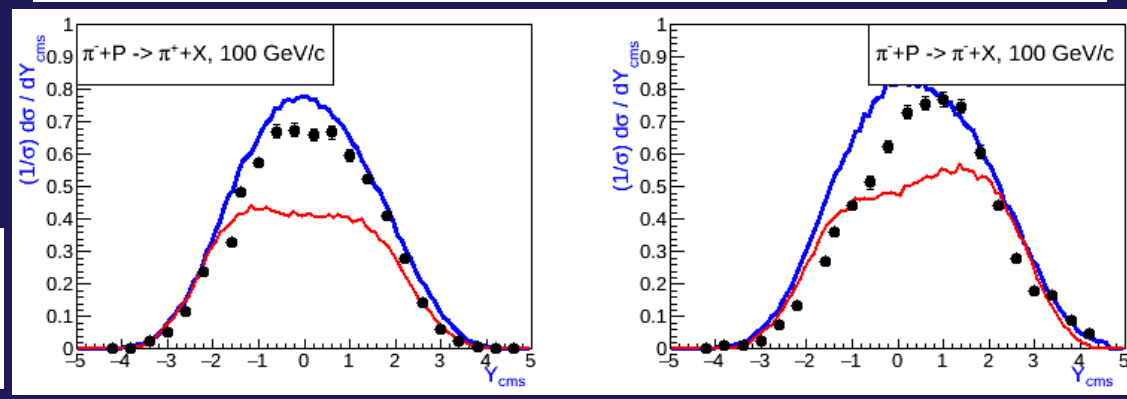
[It is expected this improves shower shape]

Tasks solved (2015):

- Improvement of string fragmentation
- Improvements of processes cross sections
- Inclusion of the Reggeon cascading for correct description of nucleus breakups
- Improvement of parton momenta sampling

To do: fine tuning of the model parameters

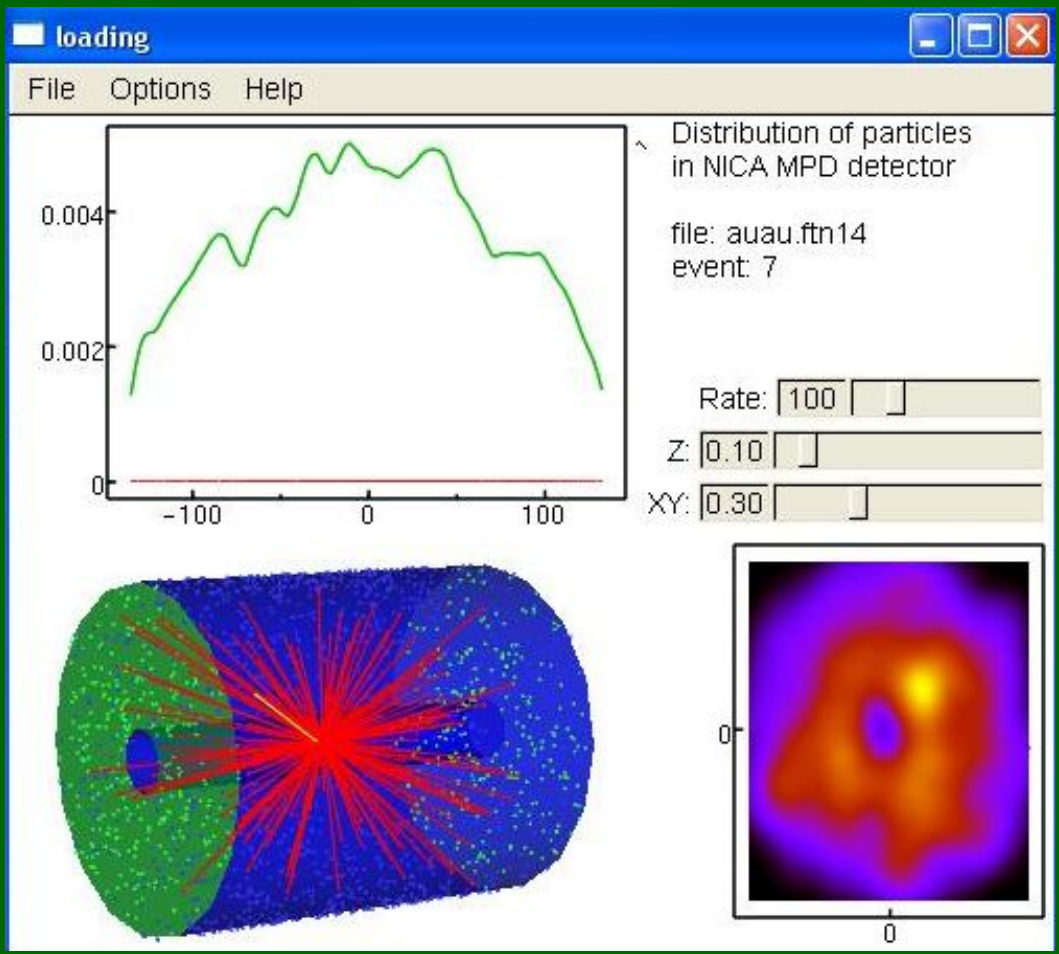
Improved QGSp will be available in G4.10.2.beta (end June 2015)
It is expected that new QGSp will improve calorimeter responses!



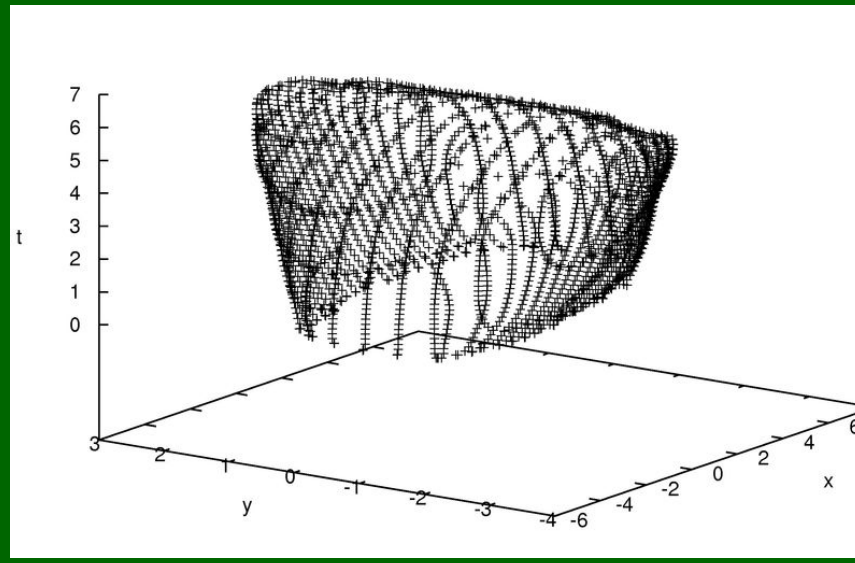
πP interactions at 100 GeV/c

Red lines – old QGSp Blue lines – new QGSp

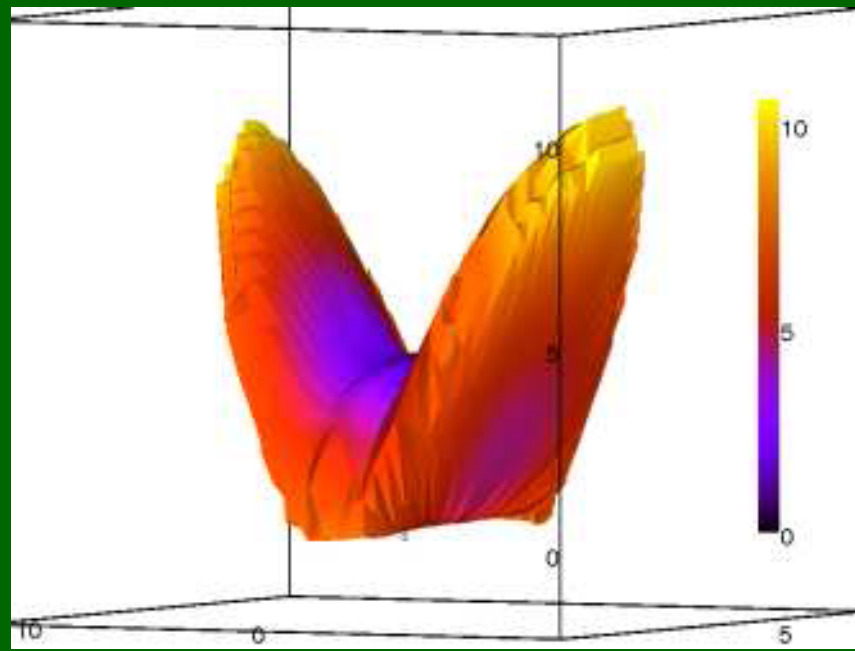
Track visualization in TPC of NICA/MPD
Au + Au at √s = 7 GeV



Visualization of freezeout surface
Au + Au at √s = 7 GeV



Au + Au at √s = 15 GeV



Visualization for Heavy Ion Collision Experiments

G. Musulmanbekov, A. Solovjev (LIT)

Projects in framework Distributed computing

- Worldwide LHC Computing Grid (WLCG)
- EGI-InSPIRE
- RDIG Development
- Project BNL, ANL, UTA "Next Generation Workload Management System for BigData"
- Tier1 Center in Russia (NRC KI, LIT JINR)
- 6 Projects at CERN
- CERN-RFBR project "Global data transfer monitoring system for WLCG infrastructure"
- BMBF grant "Development of the grid-infrastructure and tools to provide joint investigations performed with participation of JINR and German research centers"
- "Development of grid segment for the LHC experiments" with South Africa;
- Development of grid segment at Cairo University and its integration to the JINR GridEdu JINR - FZU AS Czech Republic Project "The grid for the physics experiments"
- NASU-RFBR project "Development and implementation of cloud computing technologies on grid-sites at LIT JINR and BITP for ALICE experiment"
- JINR-Romania cooperation Hulubei-Meshcheryakov programme
- JINR-Moldova cooperation (MD-GRID, RENAM)
- JINR-Mongolia cooperation (Mongol-Grid)
- JINR-China cooperation (BES-III)
- Cooperation with Belarus, Slovakia, Poland, Bulgaria, Kazakhstan, Armenia, Georgia, Azerbaijan…

- On 28 September – 02 October, 2015, Montenegro (Budva), will host the regular JINR XXV Symposium on Nuclear Electronics and Computing - NEC'2015 and students' schools on advanced information technologies
- **http://NEC2015.jinr.ru**

Thank you for your attention!

DISTRIBUTED COMPUTING AND BIG DATA AT JINR

Korenkov V.V.
Laboratory of Information Technologies, JINR, Dubna, Russia
E-mail: korenkov@cv.jinr.ru

The report presents the concept and the evolution of the global computing infrastructure for storage, processing and analysis of experiments at the Large Hadron Collider at CERN. Brief information about the participation of Russia in this process is given. An overview of projects in the field of distributed computing and Big Data, performed at the Laboratory of Information Technologies (LIT JINR) in Russia, CERN, USA, China and JINR member states is presented.

Special attention is paid to the creation of the center of the Tier1 level in Russia for storage and data processing of experiments at the Large Hadron Collider, the development of cloud and hybrid infrastructure, as well as of the computing model of megaproject NICA at JINR. The results and plans for the development of a platform for Big Data management are presented.