Distributed and High Performance Computing for Data Processing in JINR

Sergei Shmatov (MLIT JINR, Dubna)
shmatov@jinr.ru



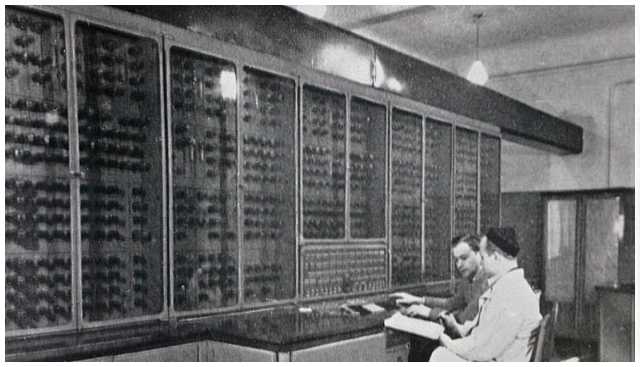


International Seminar "JINR and Latin America: Synergetic partnership for future"
4-6 December, 2025
Havana, Cuba

Meshcheryakov Laboratory of Information Technologies



M.G. Mesheryakov
(17.09.1910 - 24.05.1994)







The Laboratory of Information Technologies (formerly LCTA) at JINR was established in 1966.

Main activities are related to the **development of network, computer, and information infrastructure**, as well as **mathematical and program support for JINR's scientific program**.



N.N. Govorun
(18.03.1930 - 21.07.1989)





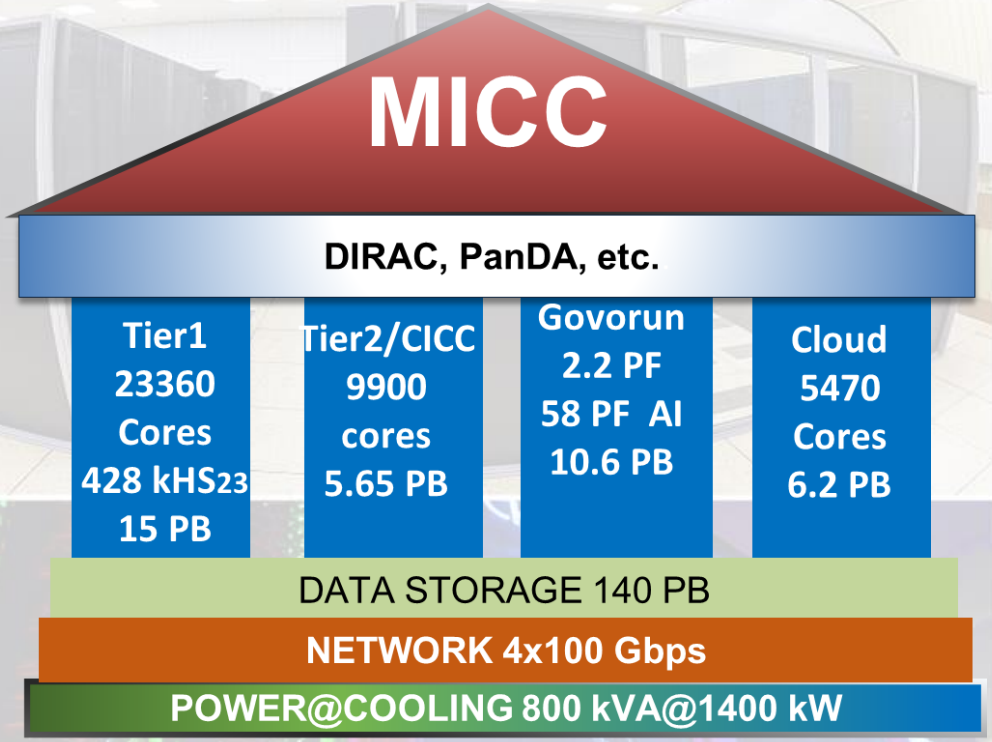


Meshcheryakov Laboratory of Information Technologies

Scientific IT ecosystem:
coordinated development of interconnected IT technologies and computational methods



Scientists Doctors of Science Candidates of Science	**115** 21 62
Specialists (engineers, programmers)	**136**
Management	**18**
Workers	47
Total	316

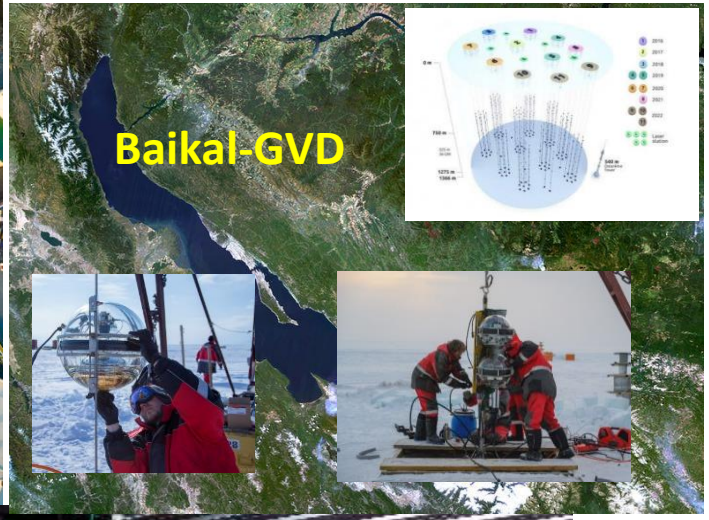


Multifunctional Information and Computing Complex

- **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

JINR Research Program



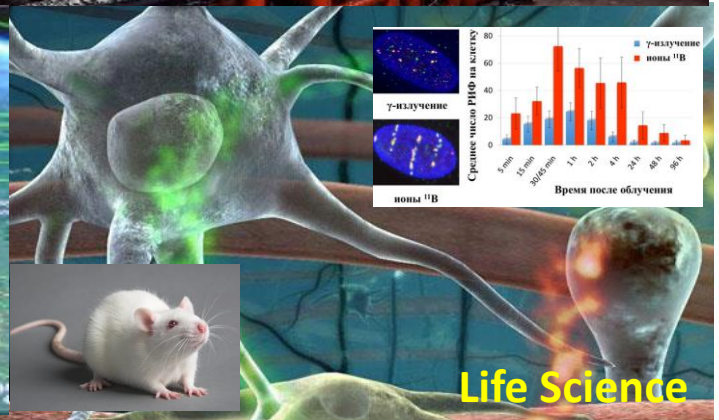












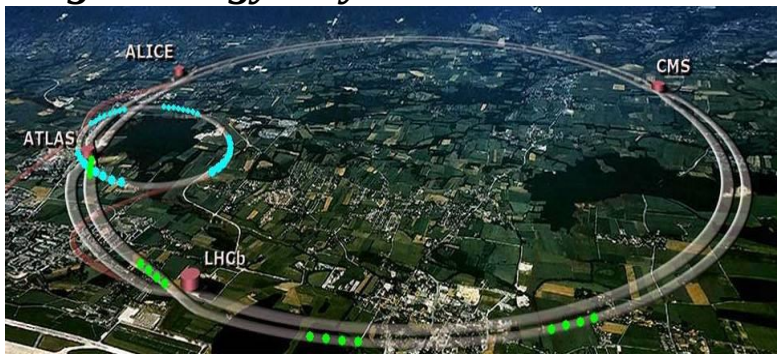


HPC+Big Data+AI for Mega Science Projects

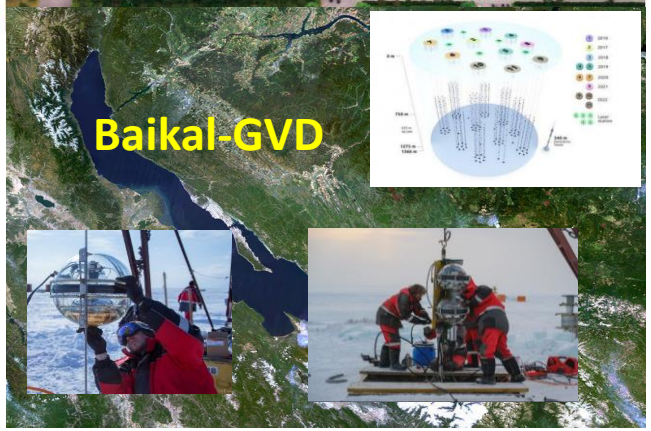




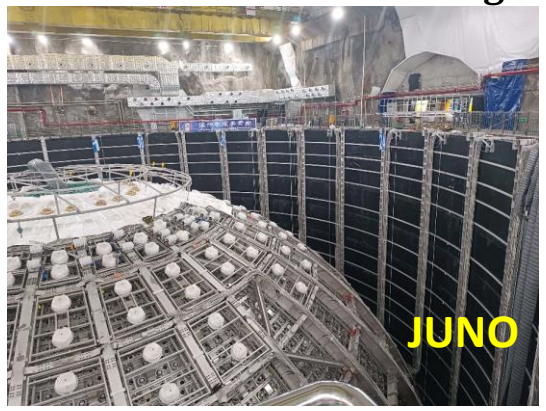
High Energy Physics



CERN Large Hadron Collider > 400 PB/Year

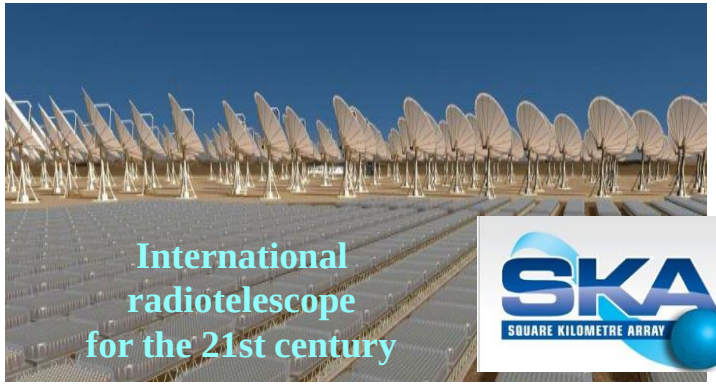


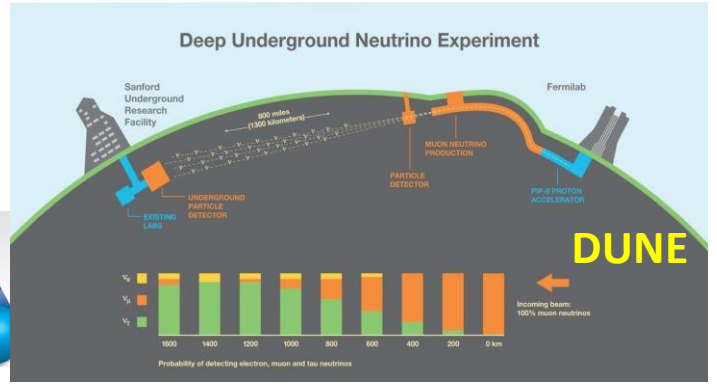


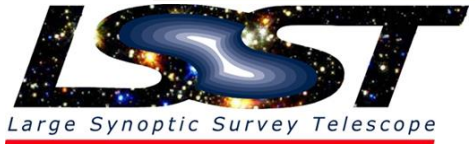


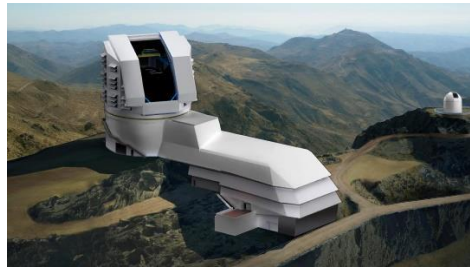


Square Kilometer Array radio telescope > 1 Eb/Year raw data (estimation)



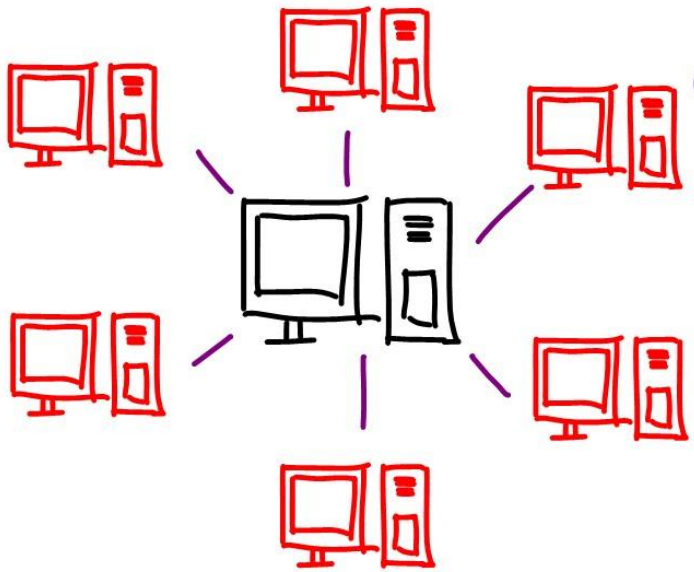




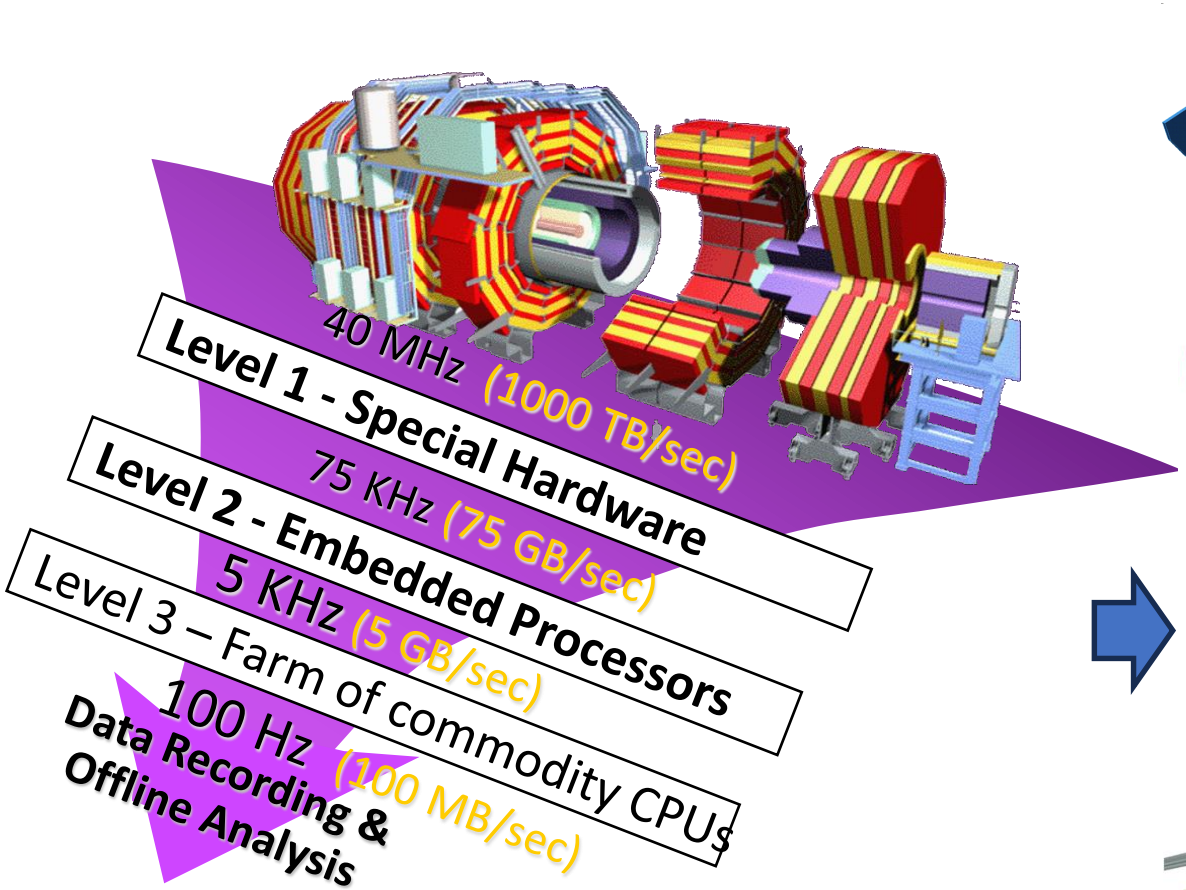


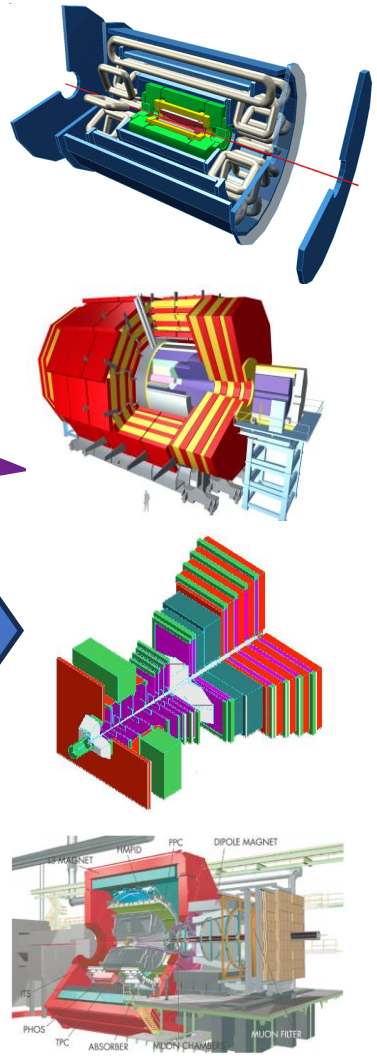
Large Synoptic Survey Telescope > 10 Pb/Year (estimation)

Distributed Computing



Example of Data Flows in HEP Experiments



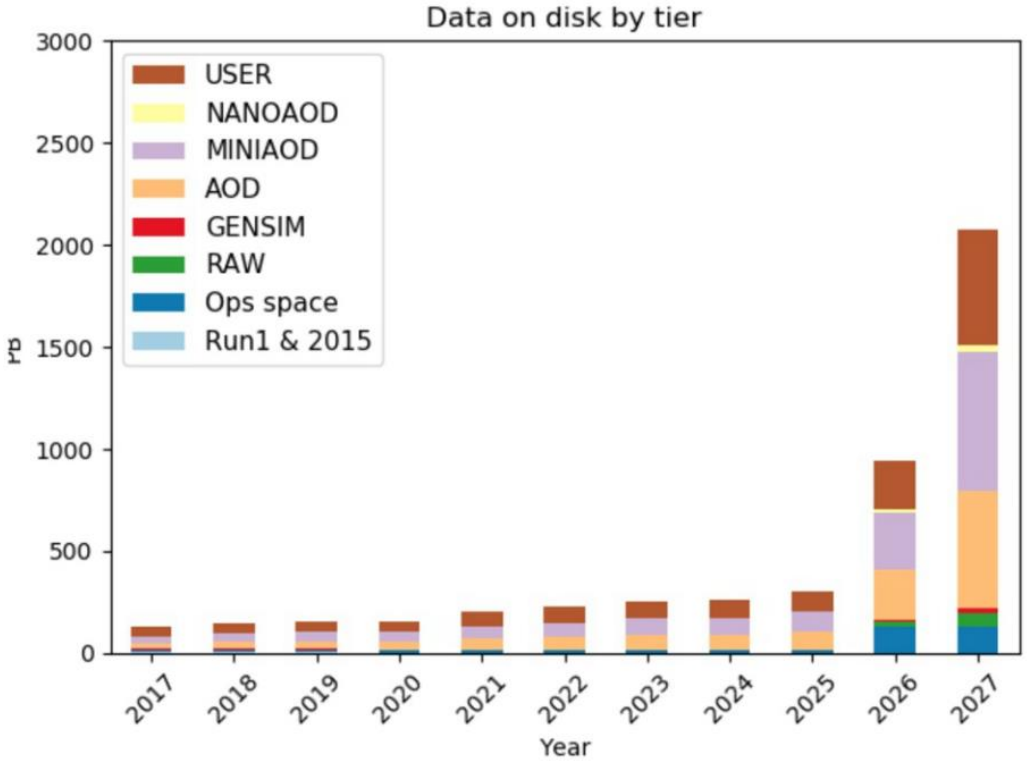


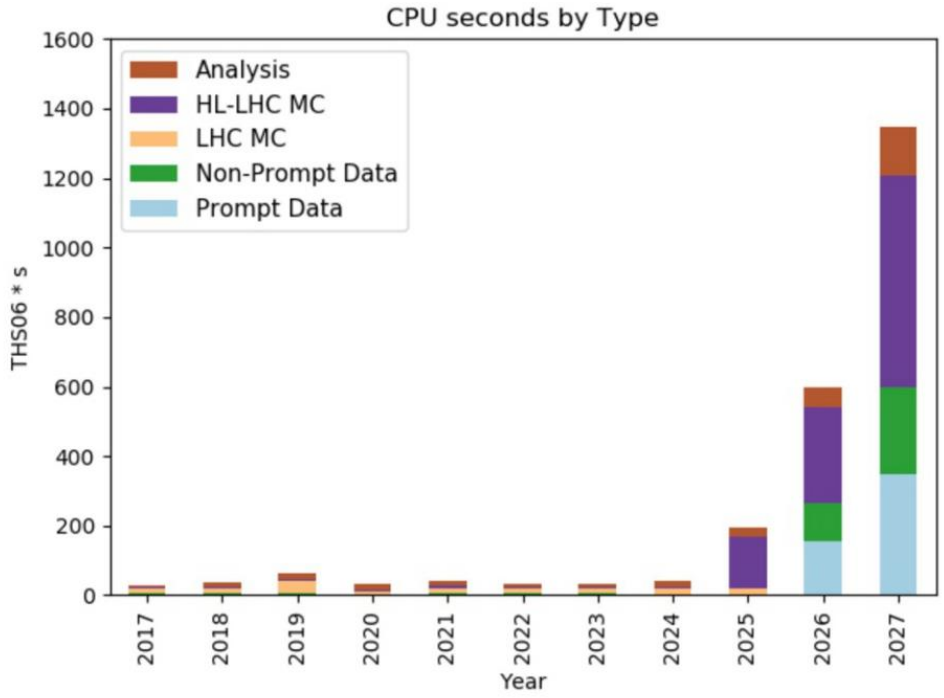
	Level-1 kHz	Event MByte	Storage MByte/s
ATLAS	100	1	100
CMS	100	1	100
LHCb	400	0.1	20
ALICE	1	25	1500

LHC Experiments produce ~400 PB/year (!)

What Do We Expect?

CMS @ LHC Example





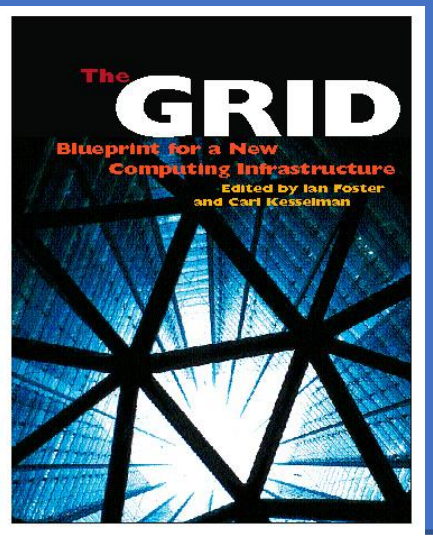
- Facing up to the exabyte (10^{18} bytes) era ⇒ required computing capacity is roughly 10 times higher than today
- an improvement of around a factor 10 in processing capabilities

Grids, clouds, fog, edge, supercomputers…

Grids

- Collaborative environment
- Distributed resources





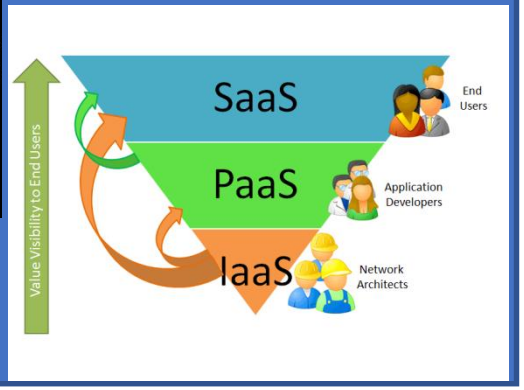
Supercomputers

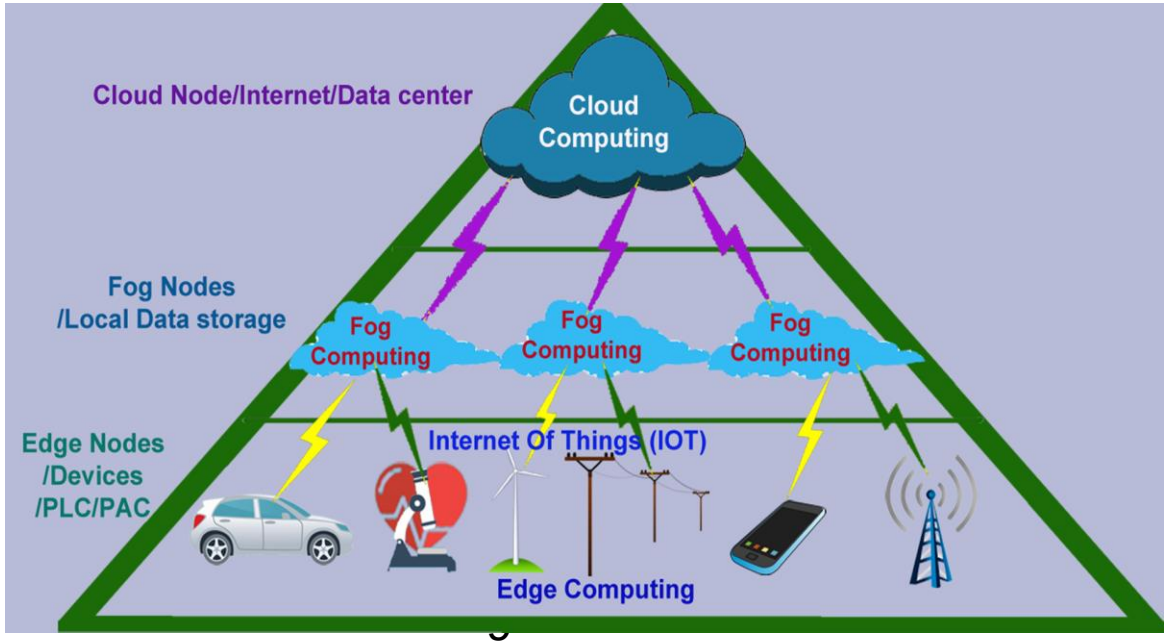




Clouds







GRID concept

«GRID is a system that:

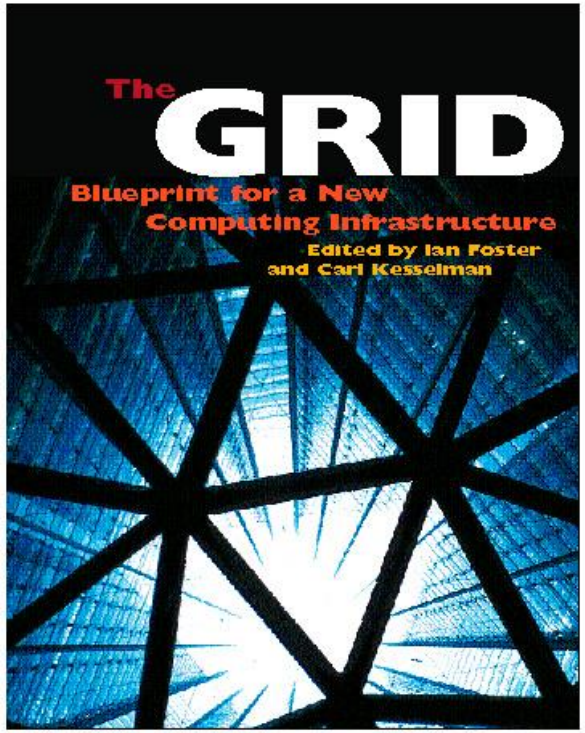
- coordinates resource usage in the absence of centralized management of those resources
- uses standard, open, universal protocols and interfaces
- provides high-quality service»





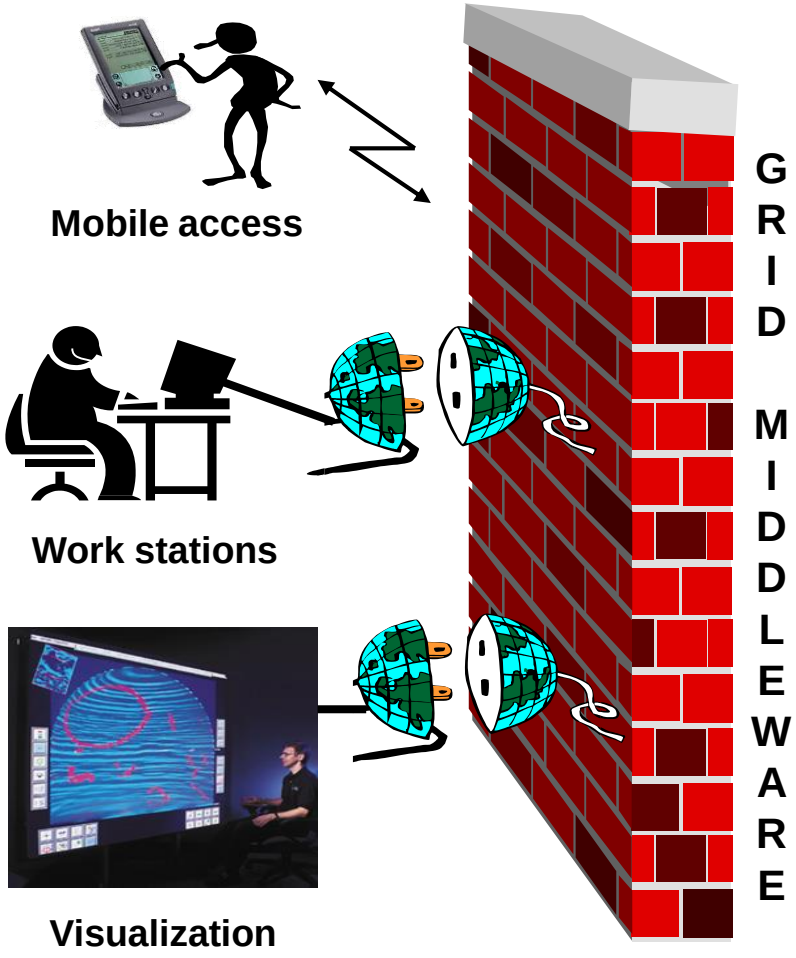
Five Emerging Models of Networked Computing From The Grid:

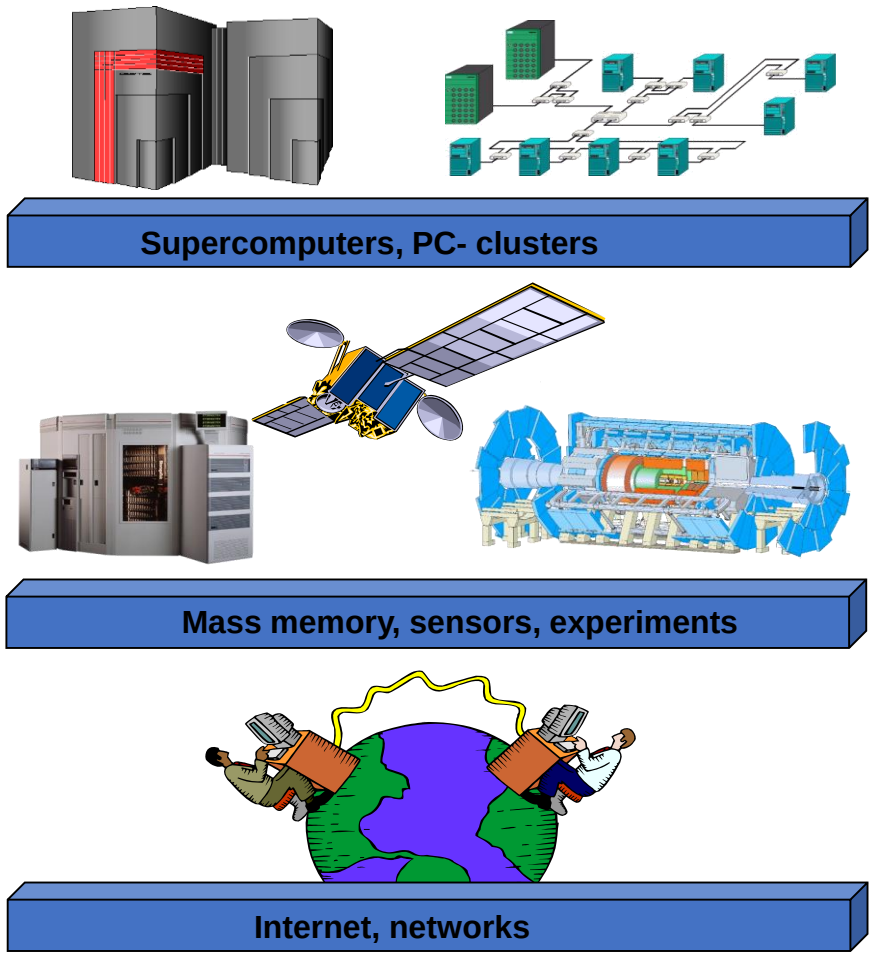
- ❖ Distributed Computing
- ❖ High-Throughput Computing
- ❖ On-Demand Computing
- ❖ Data-Intensive Computing
- ❖ Collaborative Computing



Ian Foster and Carl Kesselman, editors, "The Grid: Blueprint for a New Computing Infrastructure," Morgan Kaufmann, 1999, http://www.mkp.com/grids

GRID is a means of sharing computing power and data storage over the Internet





Worldwide LHC Computing Grid (WLCG)

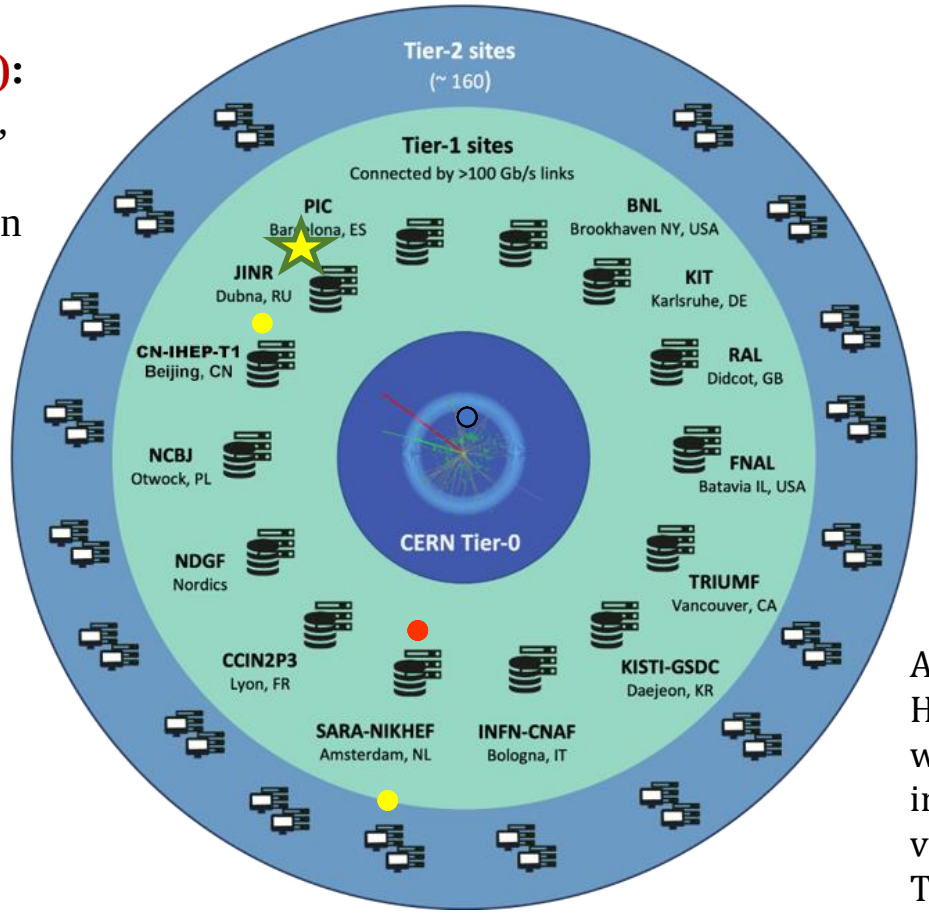
The mission of the **WLCG** is to provide global computing resources for the storage, distribution and analysis ~400 Petabytes of data expected every year of operations from the LHC. **Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists**

WLCG computing enabled physicists to announce the discovery of the Higgs Boson on 4 July 2012.

Tier0 (CERN): data recording, reconstruction and distribution

Tier1: permanent storage, re-processing, analysis

Tier2: Simulation, end-user analysis



42 countries

170 sites

~2 M CPU cores

2 EB of storage

> 2 M jobs/day

100-250 Gb/s links

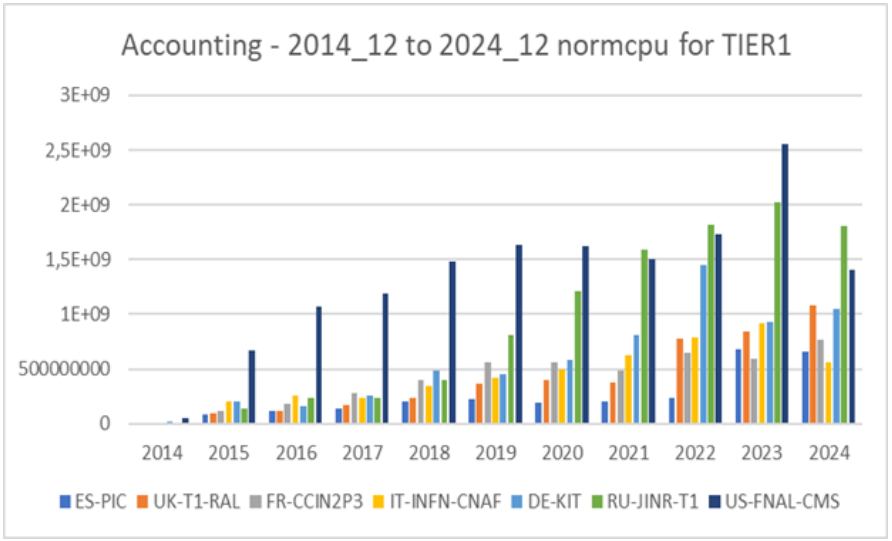


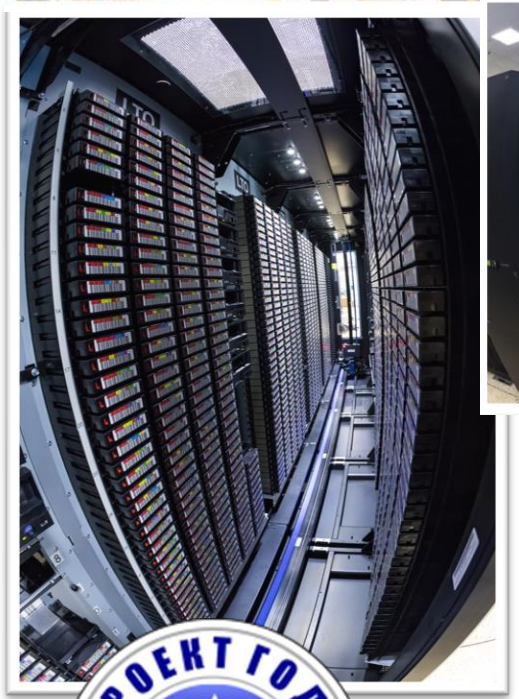
At the Nobel Prize ceremony for the discovery of the Higgs boson, CERN Director Rolf Heuer directly named grid technologies as one of the three pillars of success (along with the LHC accelerator and physics facilities). Without the organization of the grid infrastructure at the LHC, it would be impossible to process and store the colossal volume of data coming from the collider, and therefore, to make scientific discoveries. Today, not a single large project can be implemented without the use of a distributed infrastructure for data processing.

JINR Tier 1 Site for CMS@CERN and NICA Collaborations



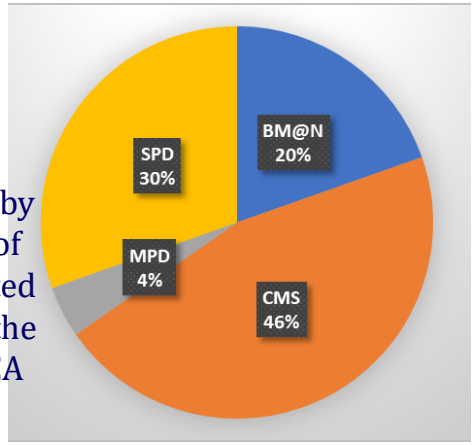
- 23 360 cores
- 428 kHS06
- 15.5 PB disks
- 100 PB tapes
- 100% reliability and availability







Distribution by the number of jobs completed on Tier1 by the CMS and NICA experiments



Since the beginning of 2015, a full-scale WLCG Tier1 site for the CMS experiment has been operating at MLIT JINR.

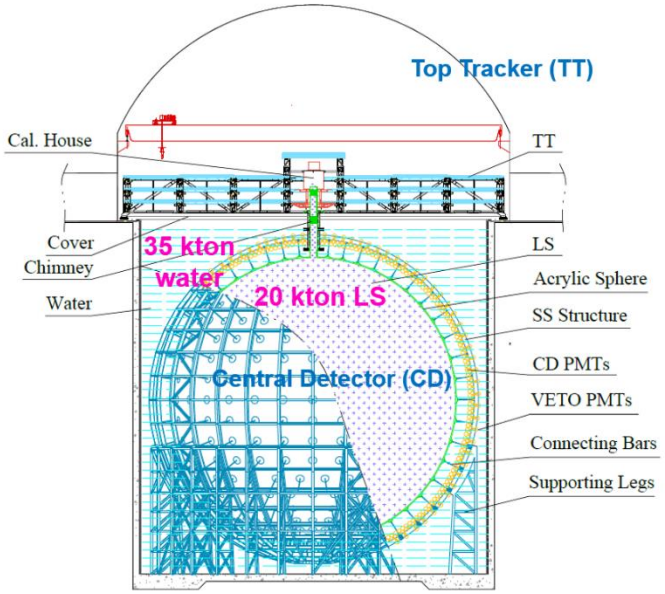
The JINR Tier1 is regularly ranked on top among world Tier1 sites that process data from the CMS experiment at the LHC.

The creation of an accelerator complex at JINR within the NICA megascience project and experimental facilities on it has entailed the extension of the functions of the JINR grid sites with the introduction of their resources in the system for modeling, processing and storing data from the BM@N, MPD and SPD experiments.

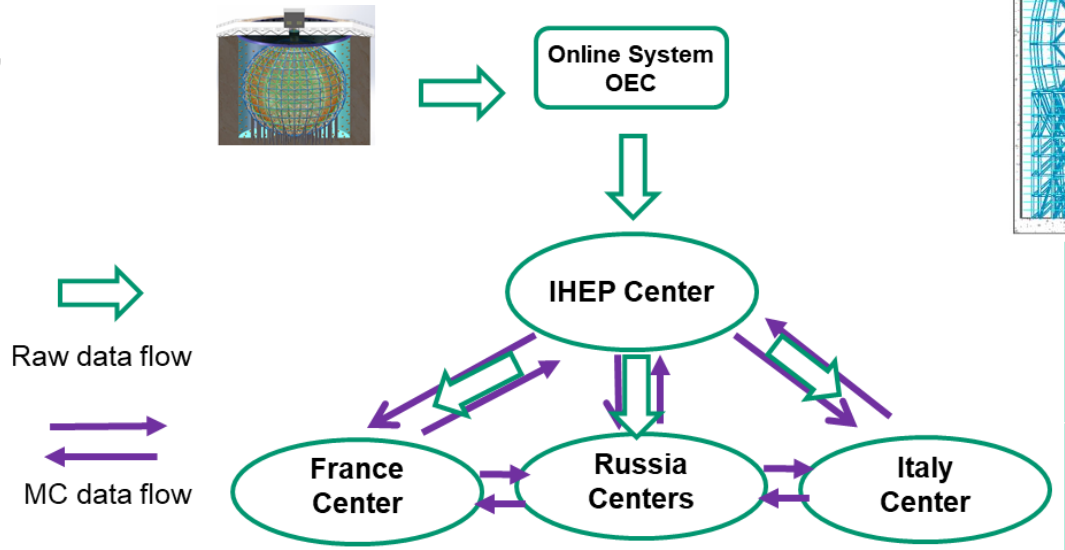
Tier 1 Site for JUNO experiment (Guangzhou, south of China)



JUNO (Jiangmen Underground Neutrino Observatory) is a multipurpose Neutrino Observatory with a rich program in neutrino physics and astrophysics studying neutrinos over a large energy range, including solar, supernova, atmospheric, geo-neutrinos, proton decay, exotic searches.



- Five data centers: IHEP, CC-IN2P3, INFN-CNAF, **JINR**, MSU (~6000 CPU cores, >10PB disk)
- Raw data flows from Online to IHEP which then distributes to other centers
- KUP (Keep up production, 1st reconstruction)* and Calibration run in IHEP
- MC Simulation, PP(Physics Production, 2nd reconstruction) and Analysis run in data centers



	2025	2026	2027	2028	2029	2030
Tape	2.36	5.88	11.75	19.08	26.40	33.73
Disk	2.35	6.15	7.6	7.61	7.61	7.6

JUNO Data Storage Volume Forecast by Storage Type (PB) in Tier-1 Data Centers

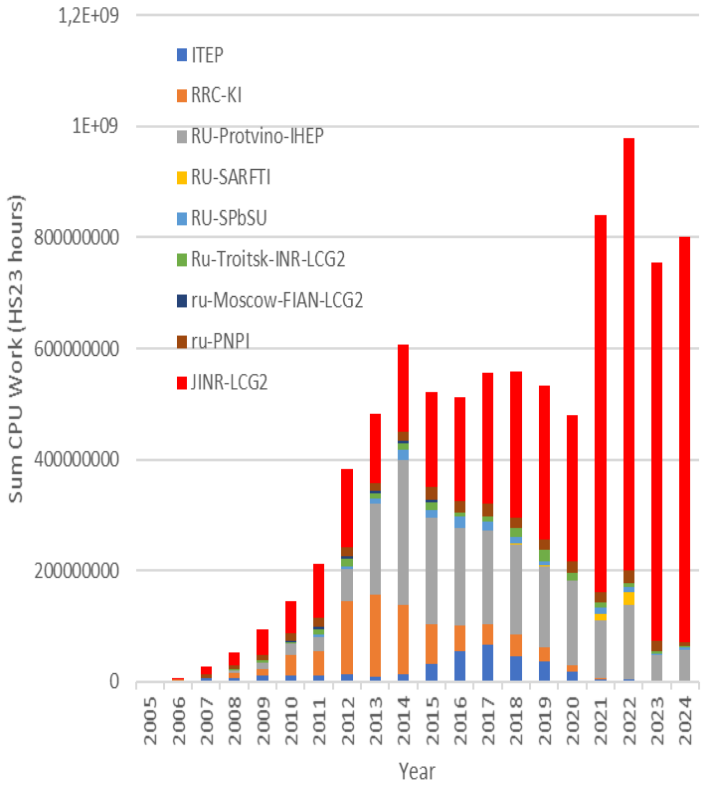
JINR recourses:

- 2000 dedicated cores, with ≈1600 cores available opportunistically
- Storage: ~1.4 PB effective in new dCache + 1 PB effective in EOS.
- Tape quota: ~5 PB in CTA.
- Network: ~3 Gbps

JINR Tier 2 Site for Different VO

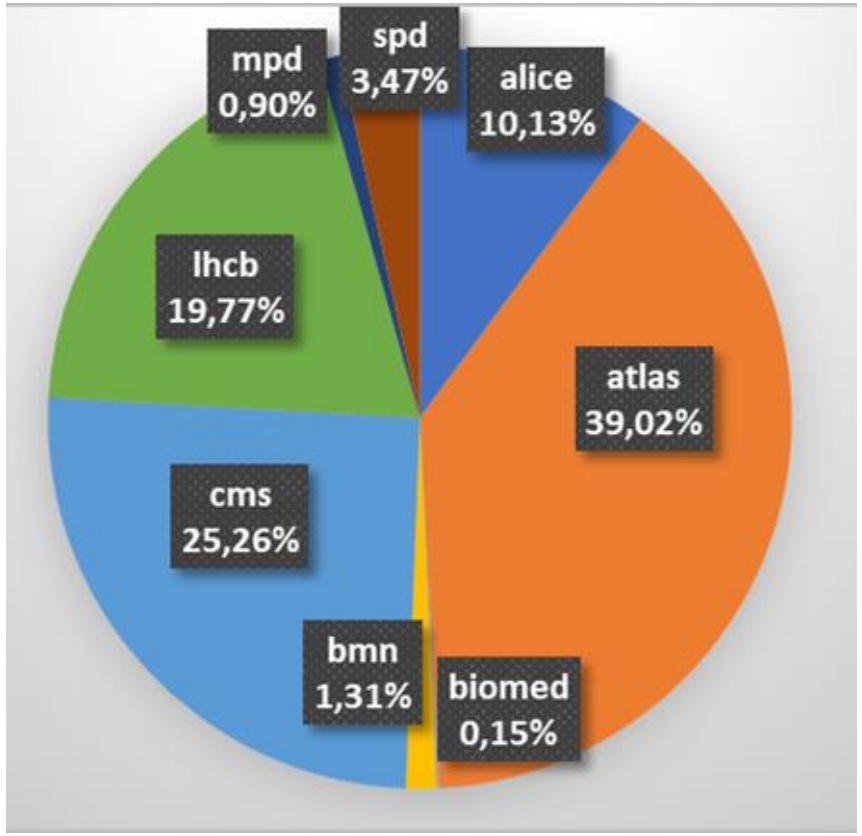
Tier2 at JINR provides computing power and data storage and access systems for the majority of JINR users and user groups, as well as for users of virtual organizations (VOs) of the grid environment (LHC, NICA, FAIR, etc.).

Accounting of RDIG Tier2's and years 2005-2024



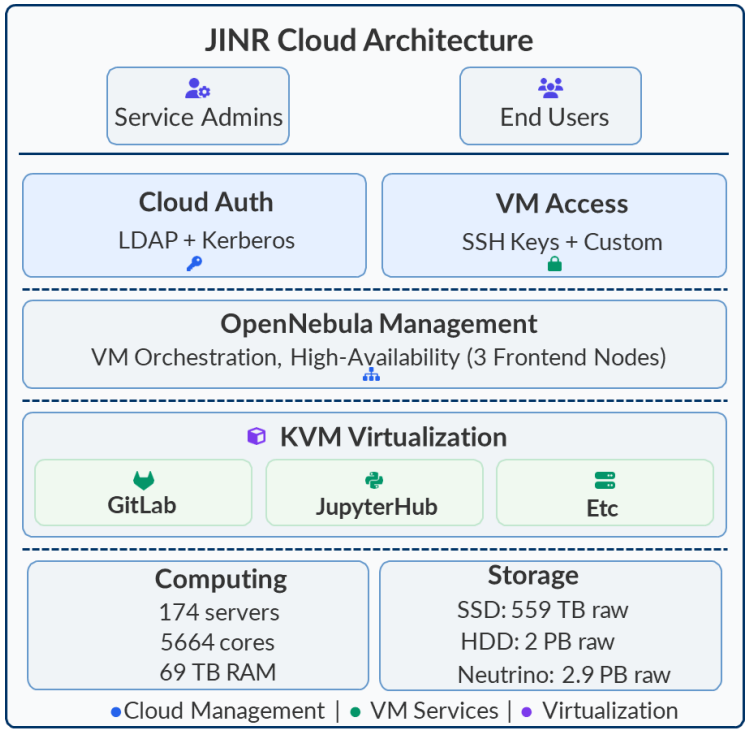


Use of the JINR Tier2 site by virtual organizations within grid projects



Cloud Infrastructure

A universal computing resource that supports individual scientists and provides a variety of common services, ranging from simple web sites to complex multi-user computational system.



Utilization:

- Computational resources for neutrino experiments. Neutrino experiments (JUNO, Baikal-GVD, NOvA) are the major users of the cloud infrastructure.
- VMs for JINR users
- Testbeds for research and development in IT
- COMPASS production system services
- Data Management System of the European Air Pollution Programme (UNECE ICP Vegetation)
- A system for diagnosing diseases of agricultural crops using modern machine learning methods
- Service for data visualization, Gitlab and some others



DIRAC-based distributed information and computing environment (DICE) that integrates the JINR Member State organizations' clouds

“Govorun” Supercomputer

- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 2.2 PFlops DP and 58 Pflops for AI tasks
- GPU component based on NVIDIA H100, V100&A100
- CPU component based on RSC “Tornado” liquid cooling solutions
- The most energy-efficient center in Russia (PUE = 1.06)
- Storage performance >300 GB/s



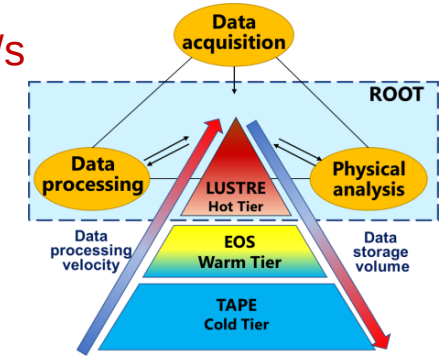
Key projects that use the resources of the SC “Govorun”:

- calculations of lattice quantum chromodynamics,
- computations of the properties of atoms of superheavy elements,
- NICA megaproject,
- studies in the field of radiation biology,
- calculations of the radiation safety of JINR’s facilities.

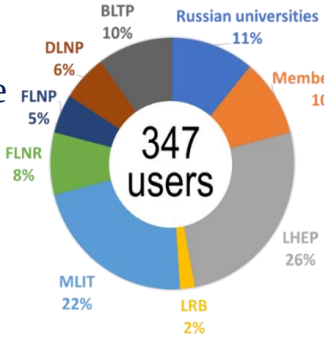




Russian DC Awards 2020 in “The Best IT Solution for Data Centers”



Total number of users : 347 including from the Member States (Armenia, Belarus, Vietnam, Egypt, South Africa)





SC "Govorun" is included into a single supercomputer infrastructure based on the National Research Computer Network of Russia (NIKS).

ML/DL/HPC ecosystem of the HybriLIT platform

The ecosystem is used for machine and deep learning tasks, training for high-performance computations. At the same time, the accumulated tools and libraries can be more widely used for scientific research, including numerical computations; parallel computing on CPUs and GPUs; results visualization; accompanying numerical results with the necessary formulas and explanations.

Component for educational purposes (without GPUs)

For teaching students
https://studhub1.jinr.ru

For conducting workshops within the framework of JINR scientific events
https://studhub2.jinr.ru
https://studhub3.jinr.ru

Component for carrying out resource-intensive computations (with GPUs)

https://jhub1.jinr.ru
https://jhub2.jinr.ru

HPC component for scientific projects (with installed specialized libraries)

BioHlit project
https://cell.jinr.ru
http://mostlit.jinr.ru
http://bio-dashboards.jinr.ru/morris

CVAT cervices
http://159.93.36.88:8080
http://159.93.36.67:8080

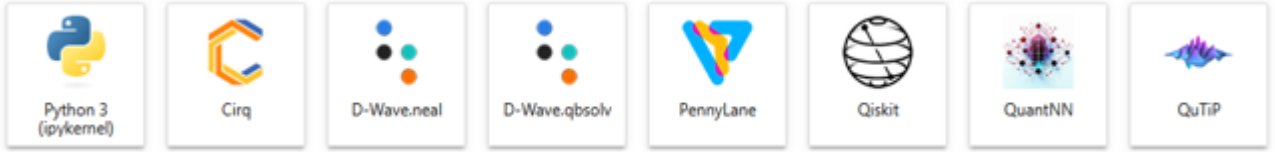
Jupyter Books infrastracture
http://studhub.jinr.ru:8080/jjbook
http://studhub.jinr.ru:8080/books
http://studhub.jinr.ru:8080/itschool2024

A polygon for visualization of brain CT data
hlit-th-ct.jinr.ru

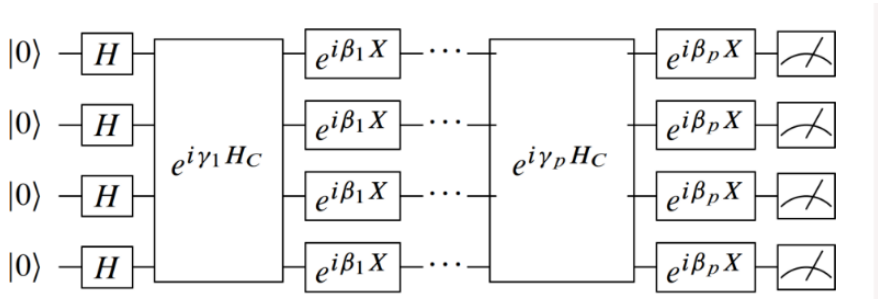
A polygon for quantum computing
https://ampere05.jinr.ru

Quantum Computing Polygon

- the ability to visually develop algorithms, visualize quantum circuits;
- available Python language materials can significantly speed up research.

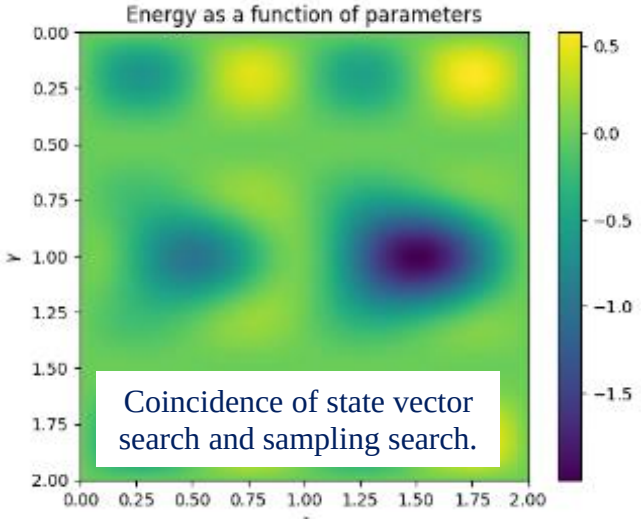


As an example, we present a search for the ground state and its energy in the Ising model with a longitudinal magnetic field using the quantum approximation optimization algorithm (QAOA).



A quantum circuit to the variation ansatz of QAOA

$$|\psi(\gamma,\beta)\rangle = \underbrace{U(\beta_p,B)U(\gamma_p,\mathcal{H})}_{p}\ldots\underbrace{U(\beta_1,B)U(\gamma_1,\mathcal{H})}_{1}H^{\otimes n}|0\rangle^{\otimes n}$$



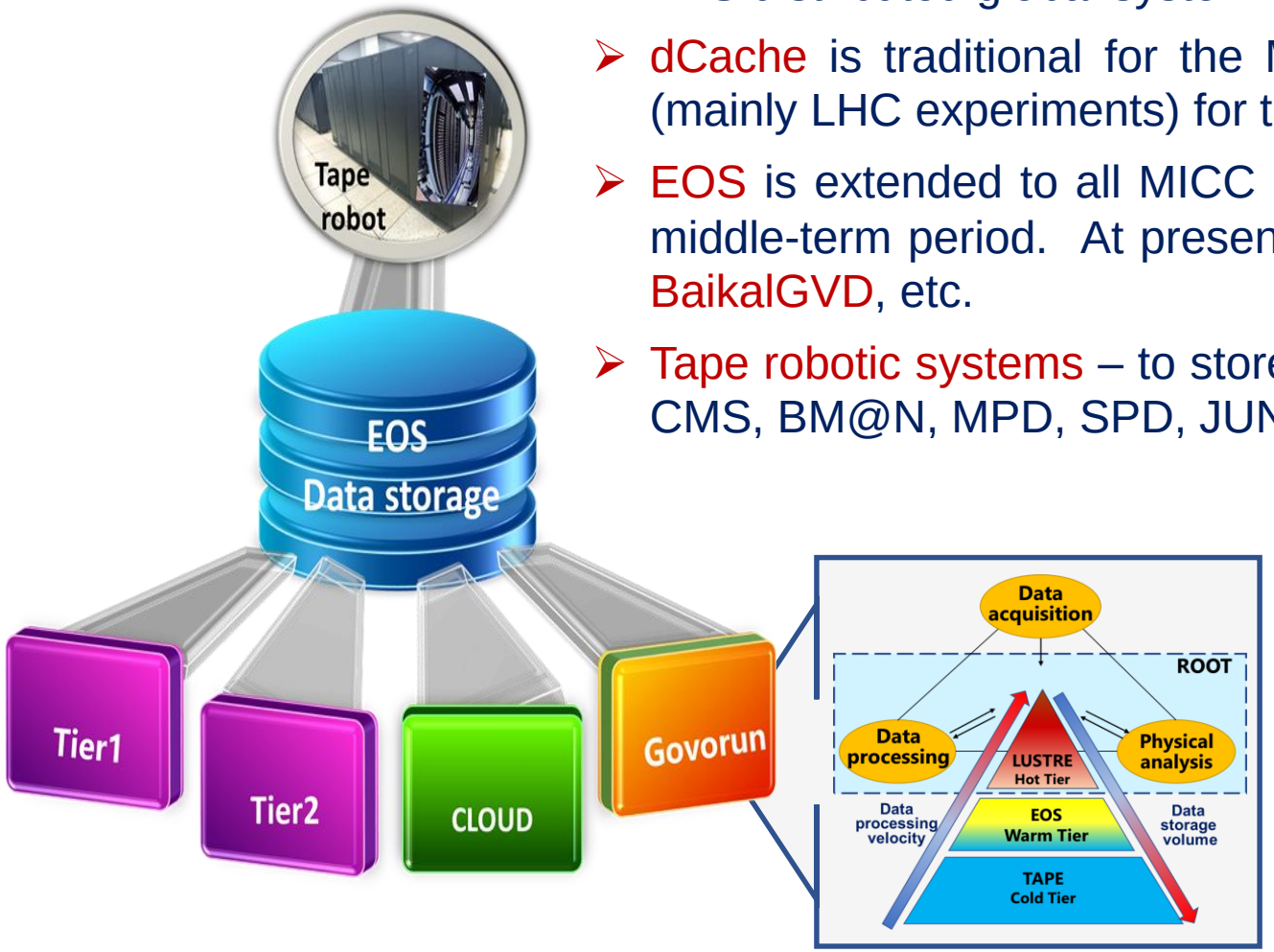
OpenMP for CPU computations and cuStateVec for GPU computations

Ising Model 3x3x3 lattice 27 qub	AMD EPYC 7763, 128 th.	Intel Xeon Platinum 8368Q, 128 threads	NVIDIA A100, cuStateVec
Comp. time	3 h 20 min	3 h 10 min	**14 min 35 sec**

Yu. Palii, A. Bogolubskaya, and D. Yanovich: Quantum Approximation Optimization Algorithm for the Ising Model in an External Magnetic Field // PEPAN, V. 55, N. 3. Pp. 600-602, 2024.

Distributed Multi-layer Data Storage System

- Limited data and short-term storage – to store the OS itself, temporary user files
- AFS distributed global system – to store user home directories and software
- dCache is traditional for the MICC grid sites – to store large amounts of data (mainly LHC experiments) for the middle-term period
- EOS is extended to all MICC resources – to store large amounts of data for the middle-term period. At present, EOS is used for storage by BM@N, MPD, SPD, BaikalGVD, etc.
- Tape robotic systems – to store large amounts of data for the long-term period for CMS, BM@N, MPD, SPD, JUNO – in progress.

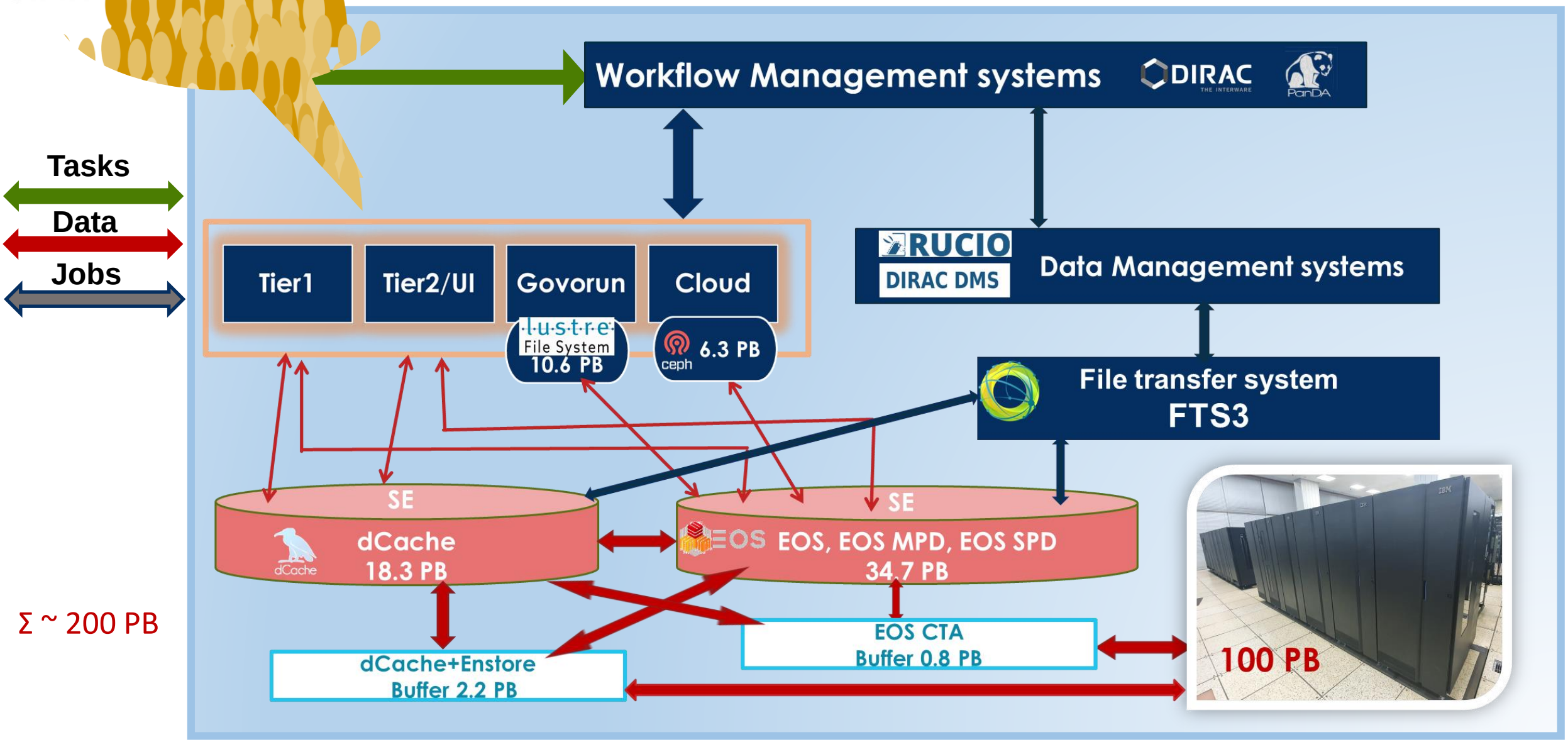


A special hierarchical data processing and storage system with a software-defined architecture was developed and implemented on the “Govorun” supercomputer.

According to the speed of accessing data, there are the following layers:

- hot data (LUSTRE),
- warm data (EOS)
- cold data(TAPE)

Data Flows and Data Management

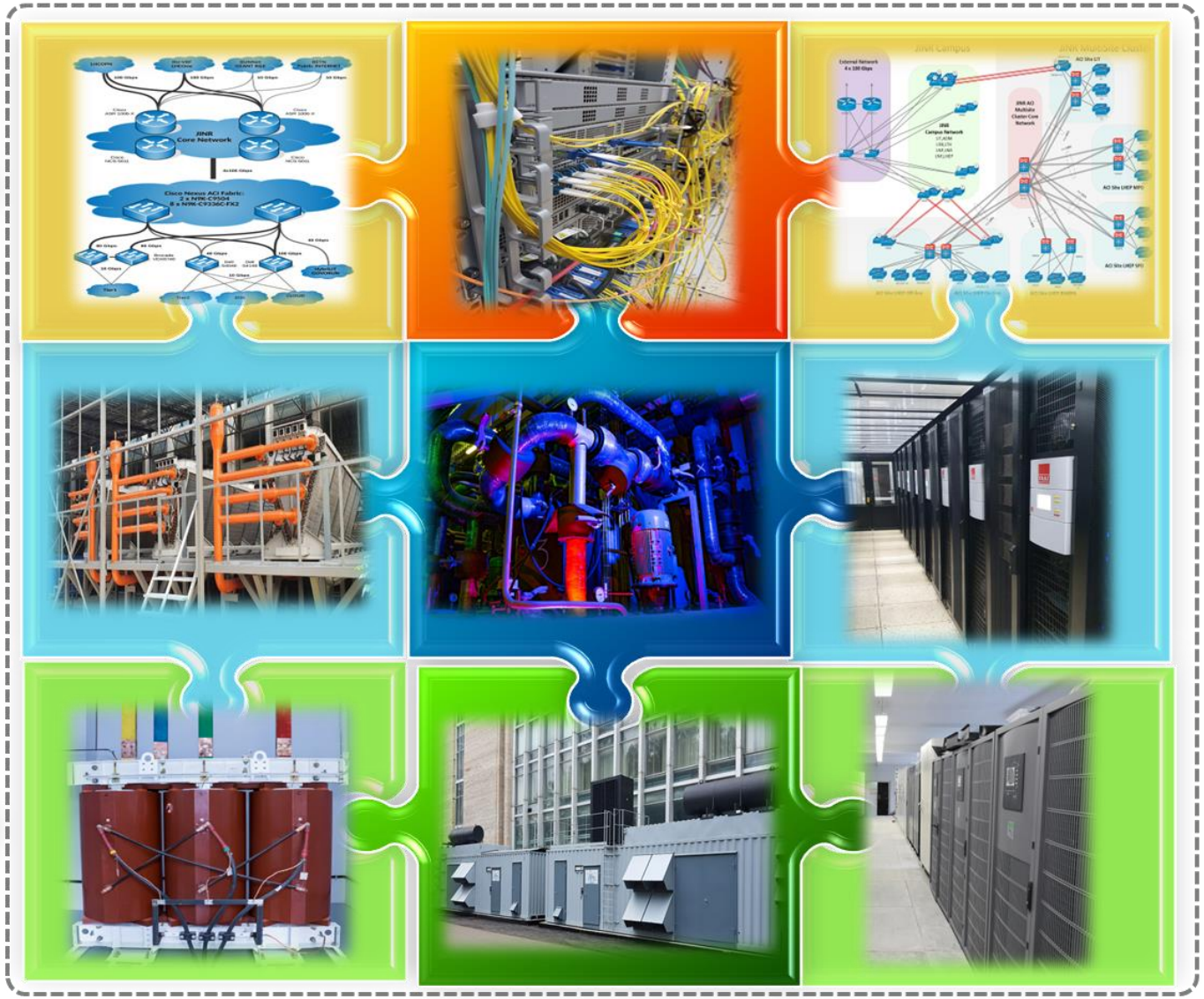


MICC Power @ Cooling @ Network

Network

Cooling

Power

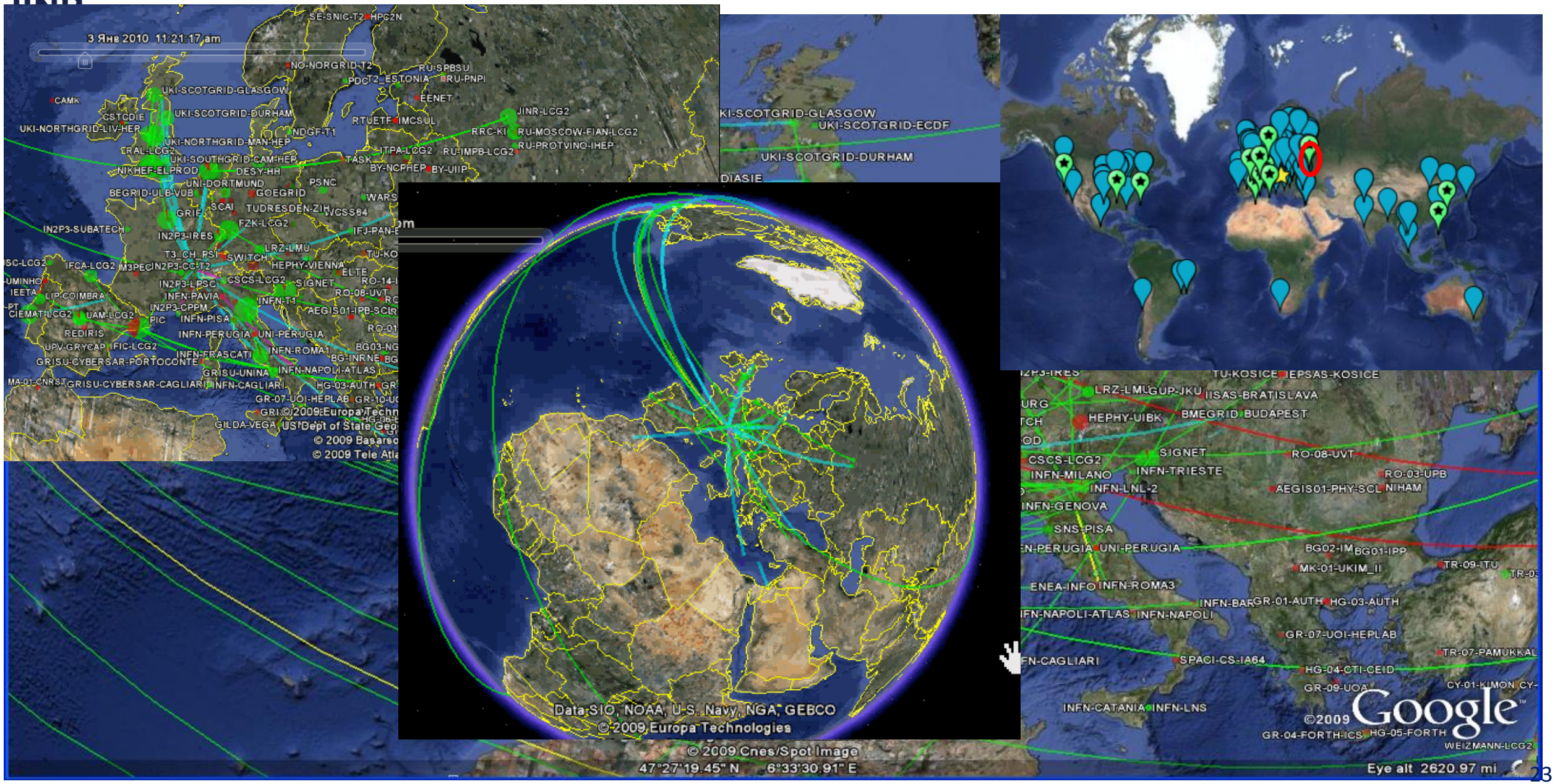


Wide Area Network 4x100 Gbps
Cluster Backbone 4x100 Gbps
Campus Backbone 2x100 Gbps

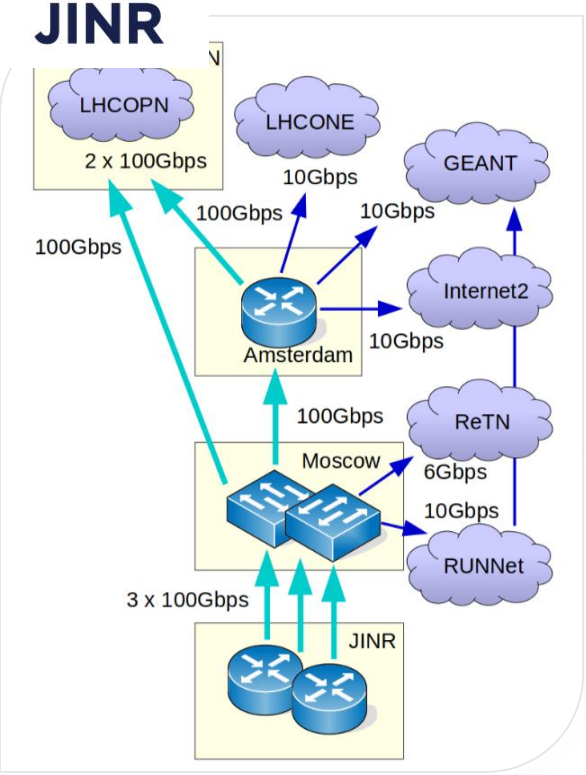
Dry chillers
In-Row systems
Total cooling 1400 kW

Uninterruptible power supplies (UPS) 8x300 kVA
Diesel-generator units (DGU) 2x1500 kVA

The Worldwide LHC Computing Grid (WLCG)

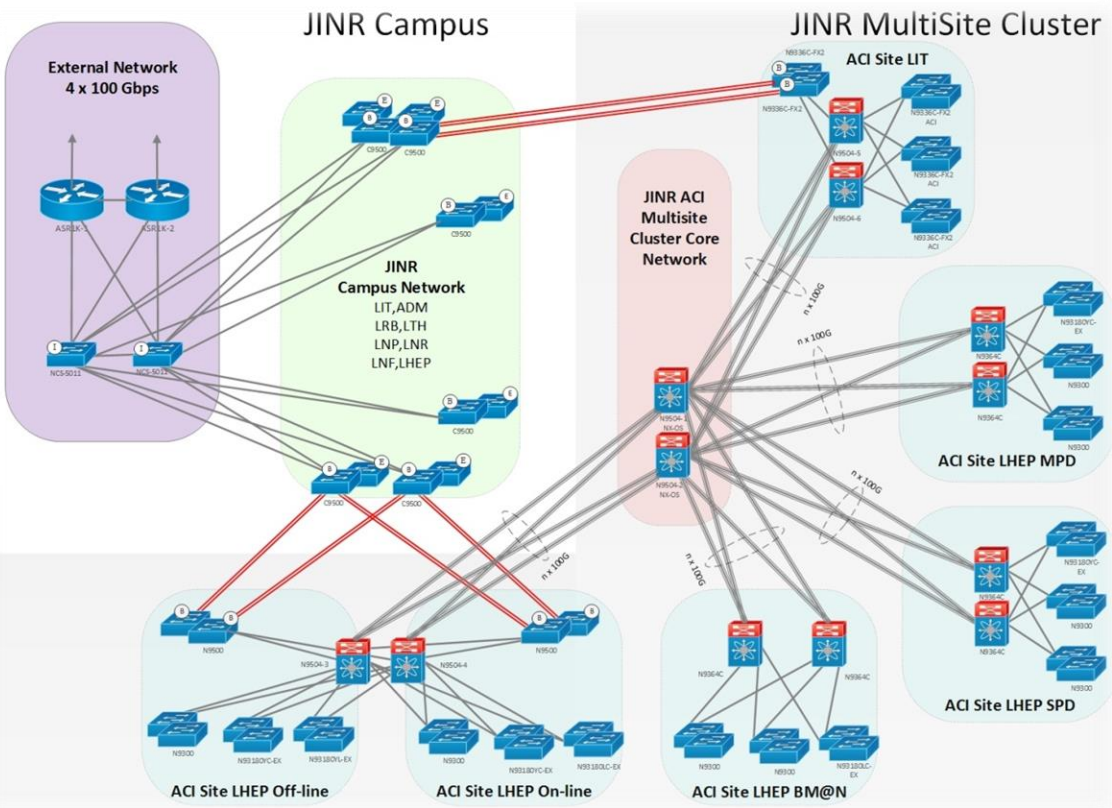


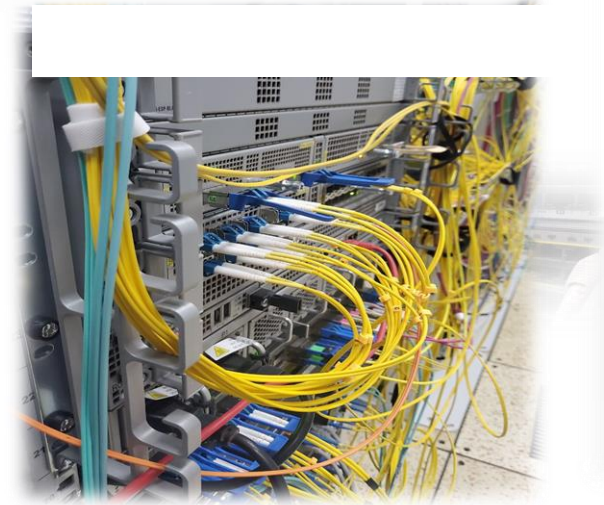
Network Infrastructure



MLIT ensures the reliable and fault-tolerant operation of all components of the network infrastructure:

- JINR-Moscow 4x100 Gbit/s
- JINR-CERN 100 Gbit/s and (?) JINR-Amsterdam 100 Gbit/s
- multi-site cluster network with a bandwidth of 4x100 Gbit/s for the NICA megaproject
- local area network with a bandwidth of 2x100 Gb/s





The JINR LAN comprises:

13395 network elements
23850 IP-addresses
5750 users registered within the network
4785 *.jinr.ru service users
1165 digital library users
920 remote VPN
140 EDUROAM users

Network traffic in 2024

43.45 PB - input
27.28 PB - output

JINR – China Networking (plans)

China-JINR Workshop on Software and Computing for Future HEP Experiments



The purpose of the workshop is to discuss the issues of cooperation between IHEP and JINR in the field of computing and software, including BESIII, JUNO experiments, planned experiments at CEPC and NICA colliders, as well as R&D of new computational technologies in particle physics.

Realization of the MoU

Annex 2: Pledged and planned to be pledged Computing Resources

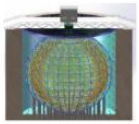
The following tables give the size of pledged and planned to be pledged Computing Resource Levels of the sites participating in the JUNODCI.

Annex 2.1: IHEP Computing Resource Levels

IHEP	Planned to be pledged				
	2023	2024	2025	2026	2027
Tape (PB)	3.0	5.0	5.0	5.0	5.0
Disk (PB)	3.0	5.0	5.0	5.0	5.0
CPU (kHS06)	50	100	100	100	100

Annex 2.2 Raw Data Centres Computing Resource Levels

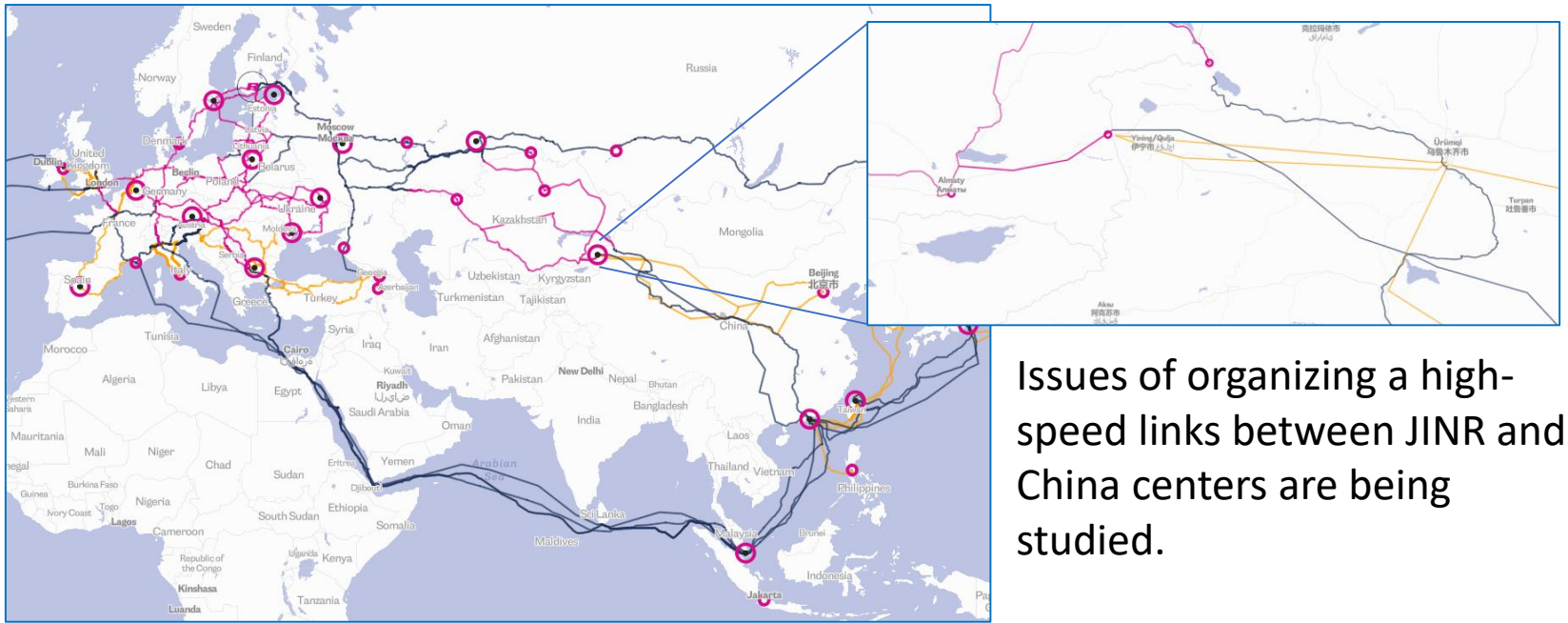
JINR	Planned to be pledged				
	2023	2024	2025	2026	2027
Tape (PB)	5.0	5.0	5.0	5.0	5.0
Disk (PB)	5.0	5.0	5.0	5.0	5.0
CPU (kHS06)	36	36	30	20	10



JUNO

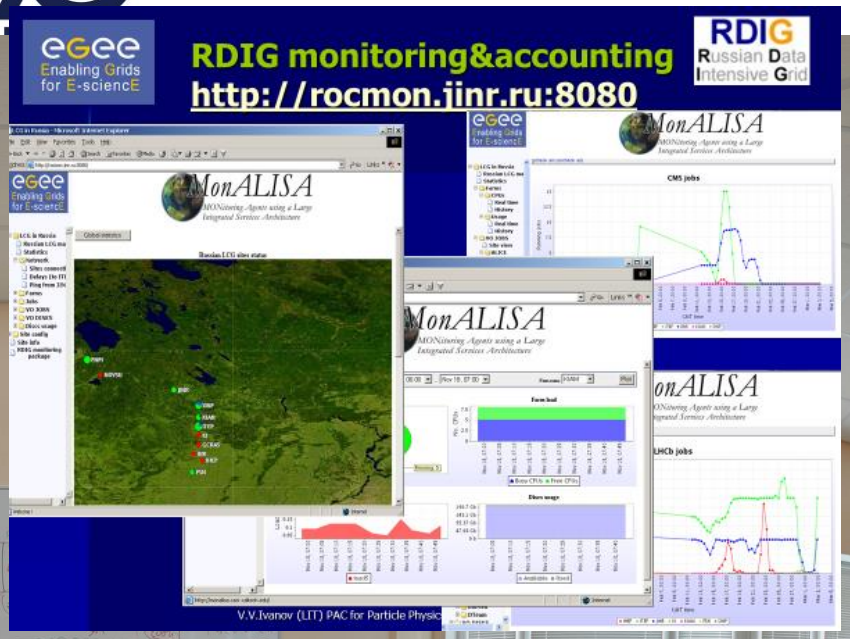
Raw data flow

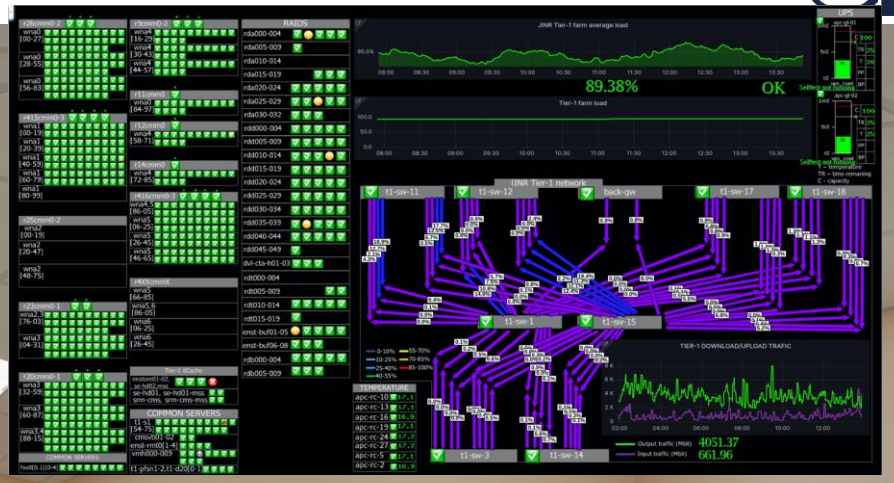
MC data flow



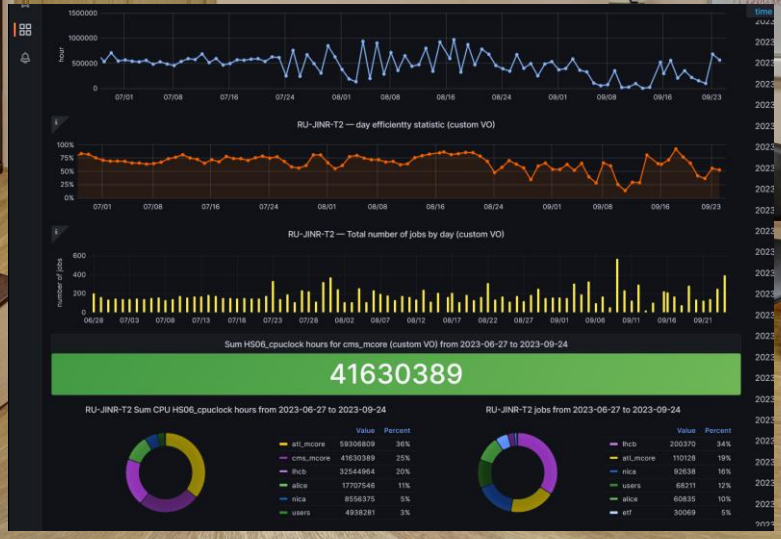
Issues of organizing a high-speed links between JINR and China centers are being studied.

Monitoring

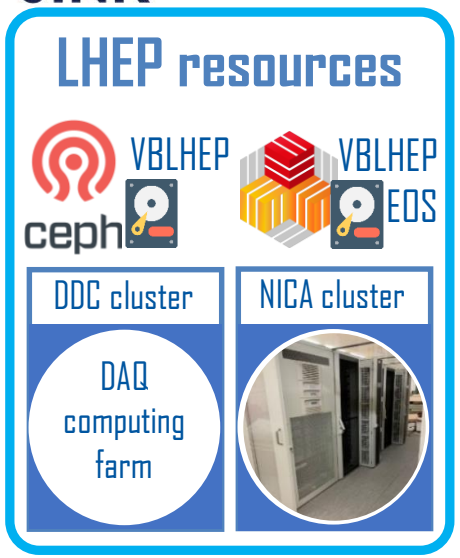

RDIG monitoring&accounting
http://rocmon.jinr.ru:8080
RDIG
Russian Data Intensive Grid
MonALISA
CMS jobs
LHCb jobs
V.V.Ivanov (LIT) PAC for Particle Physic


89.38%
OK
JINR Tier-1 network
TIER-1 DOWNLOAD/UPLOAD TRAFIC
4051.37
661.96

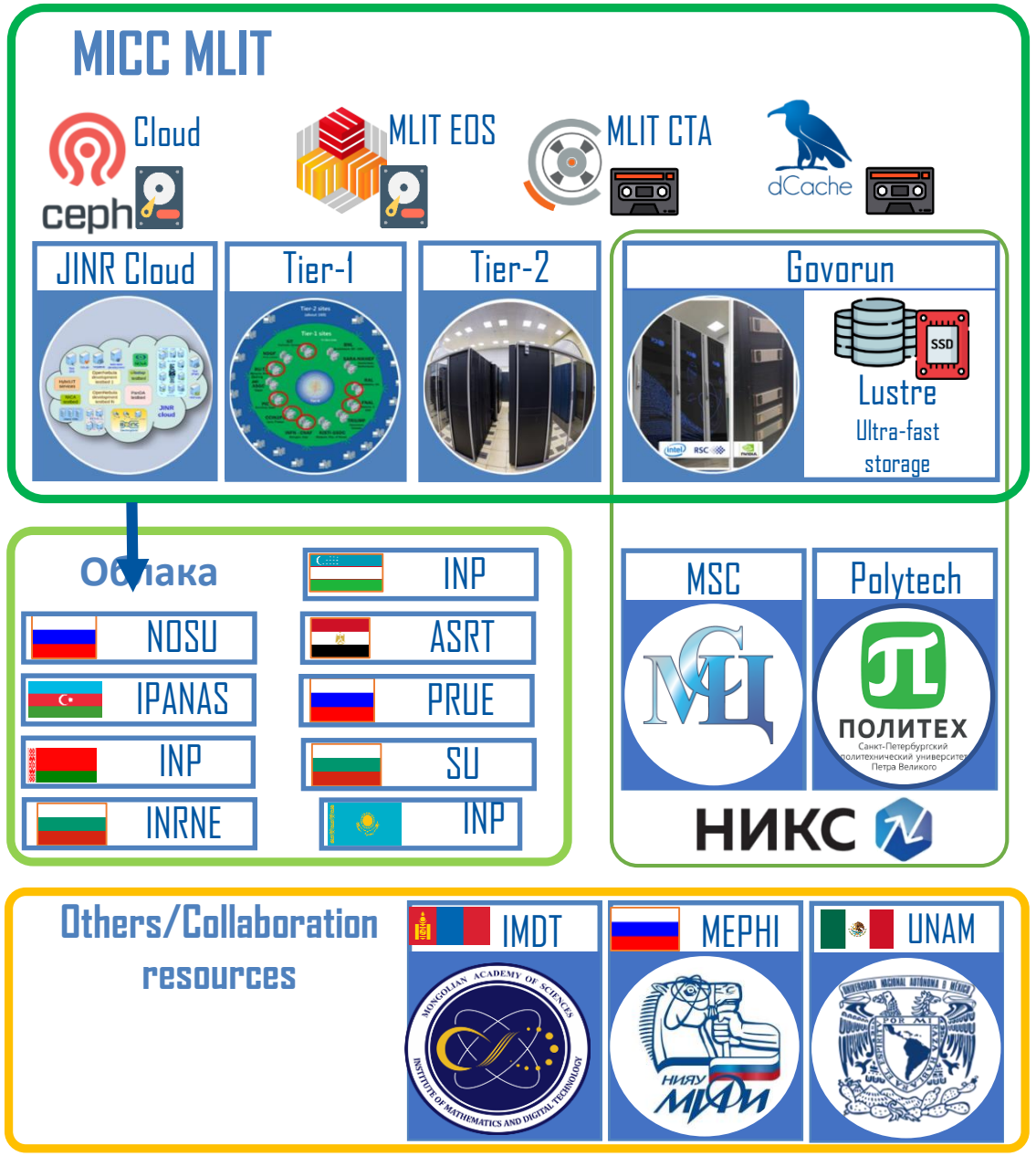

50.6 PB
2.65 PB
11.7 PB
20000
1.99 PB
1.69 PB
1.94 PB
10364
7.51 PB
15680
22.4 PB
4320
5152
60.6 TB
3.84 PB
1.44 PB


RU-JINR-T2 — day efficientty statistic (custom VO)
RU-JINR-T2 — Total number of jobs by day (custom VO)
Sum HS06_cpuclock hours for cms_mcore (custom VO) from 2023-06-27 to 2023-09-24
41630389
RU-JINR-T2 Sum CPU HS06_cpuclock hours from 2023-06-27 to 2023-09-24
RU-JINR-T2 jobs from 2023-06-27 to 2023-09-24

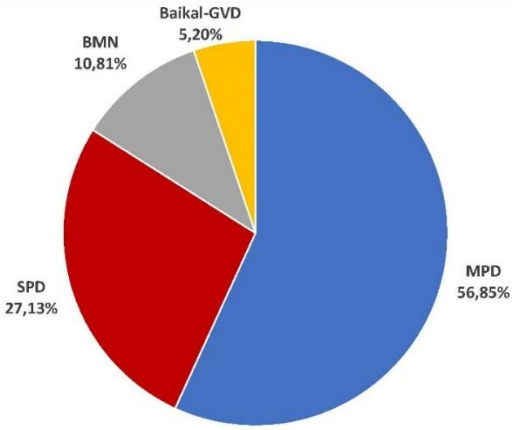
DIRAC-based distributed heterogeneous environment



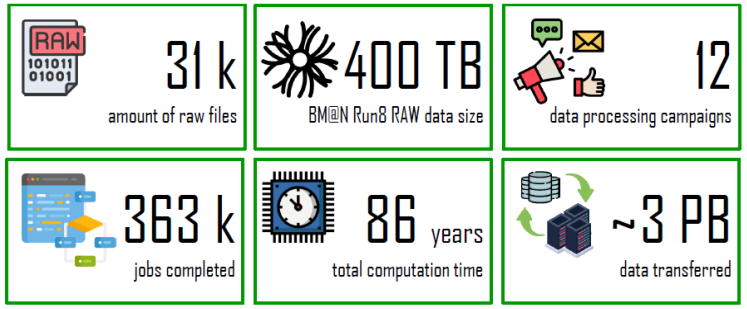
To perform calculations related to the mass generation and processing of data from the NICA mega-science project experiments, a distributed environment based on the DIRAC Interware platform was created.



DIRAC is employed to solve the tasks of collaborations of all three experiments at the NICA accelerator complex, as well as of the Baikal-GVD neutrino telescope.



In 2023, for the first time at JINR, the complete processing of raw data from the 8th run of the BM@N experiment was performed on the distributed heterogeneous computing infrastructure integrated using the DIRAC platform.



Summary statistics of using the DIRAC platform for BM@N Run 8 data processing

Development of the system for training and retraining IT specialists



Training courses, master classes and lectures

MLIT staff and leading scientists from JINR and its Member States

Leading manufacturers of modern computing architectures and software

Parallel programming technologies

OpenMP

MPI

OpenCL

NVIDIA CUDA

Tools for debugging and profiling parallel applications

Intel Cluster Studio

Work with applied software packages

COMSOL MULTIPHYSICS

Wolfram Mathematica

ROOT Data Analysis Framework

MATLAB

GEANT4

Maple

Frameworks and tools for ML/DL tasks

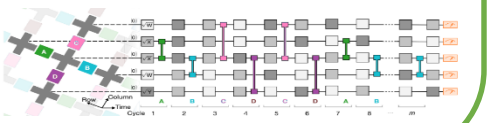
jupyterhub

TensorFlow

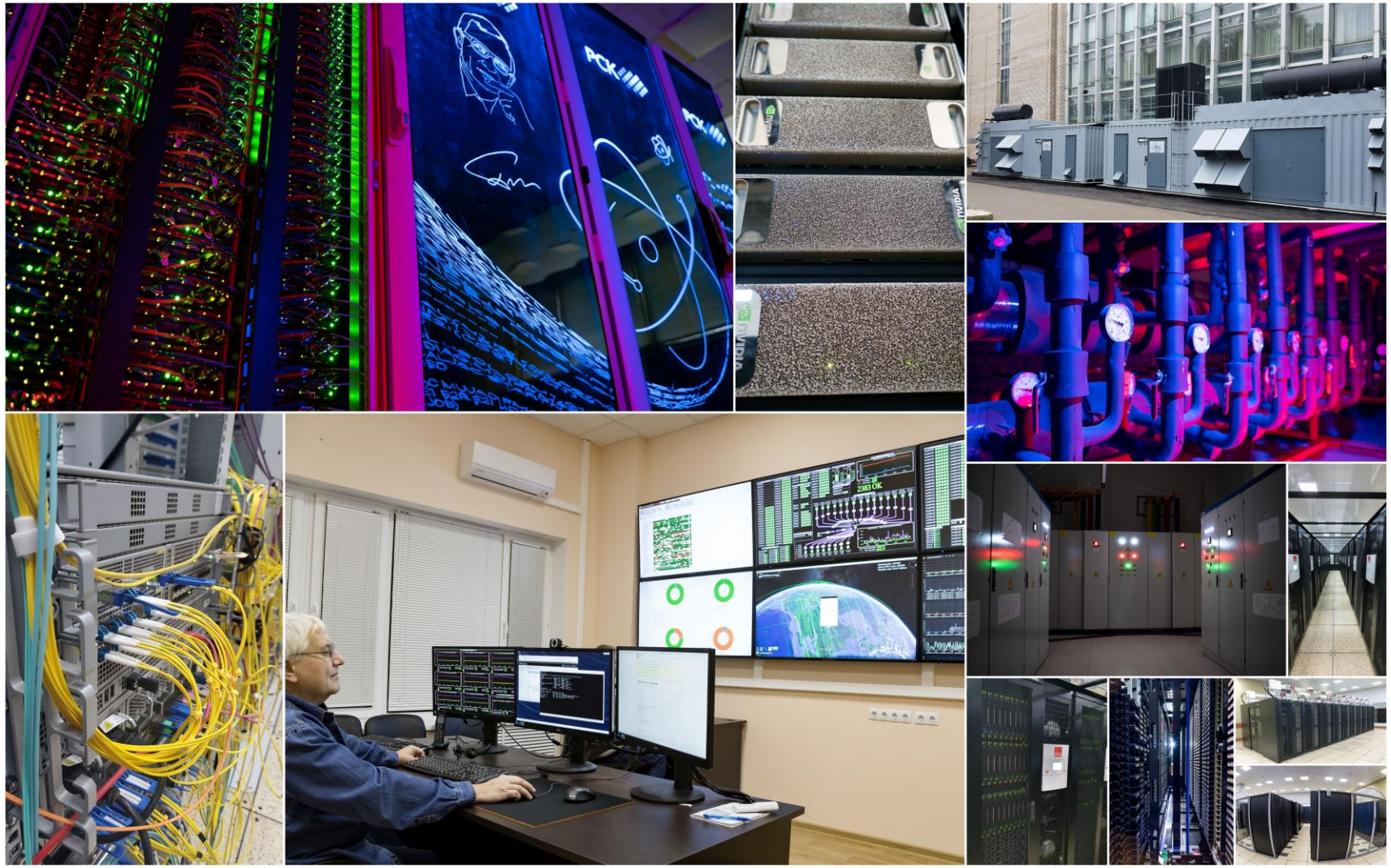
NumPy

scikit learn

Quantum algorithms, quantum programming and quantum control



Thank you for your attention!



https://lit.jinr.ru

t.me/MLIT_news