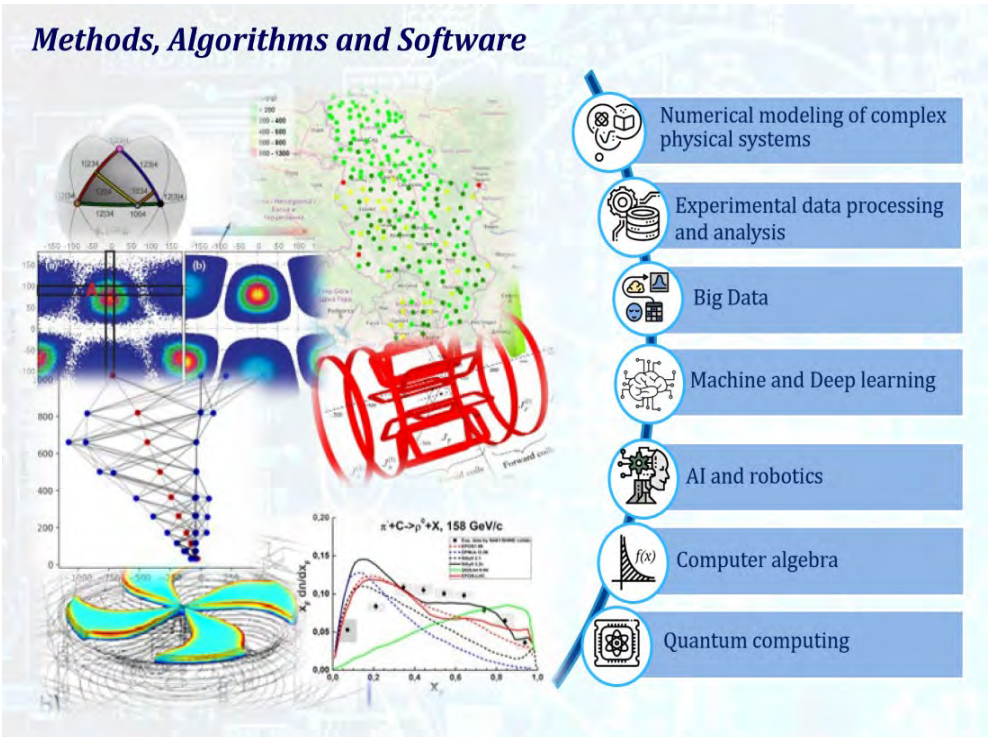
Методы, алгоритмы и программное обеспечение для моделирования физических систем, обработки и анализа экспериментальных данных физики высоких энергий

С.В. Шматов (ЛИТ ОИЯИ)
shmatov@jinr.ru



Рабочее совещание “ОИЯИ-ВШЭ”
ЛФВЭ ОИЯИ, 14 июня 2024

MLIT Projects and Activities

Critical basic scientific infrastructure project

- JINR Multifunctional Information and Computing Complex (MICC)
  - High-throughput computing (Grid Tier1 and Tier2)
  - High-performance computing ("Govorun" supercomputer)
  - Cloud Computing
  - Distributed Data Storage (long-term and middle-term Data lake)
  - Engineering infrastructure (electricity and cooling)
  - Networking (local and worldwide connectivity)

Talks by V. Korenkov, D. Podgainy

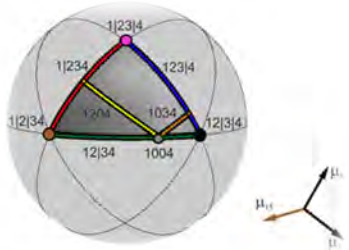
Research projects

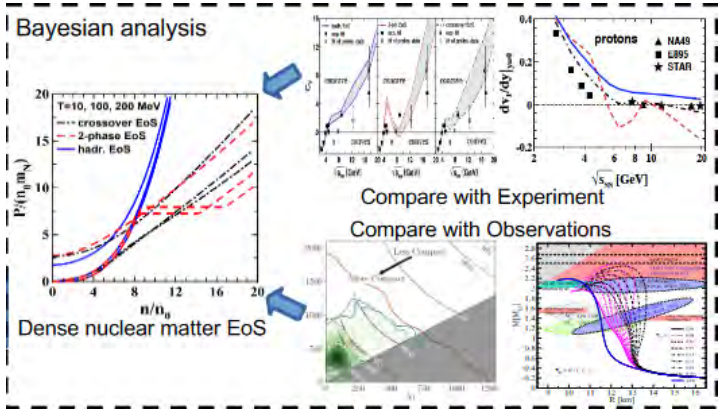
Talk by G. Ososkov

- Mathematical methods, algorithms and software for modeling physical processes and experimental facilities, processing and analyzing experimental data
  - implementation of ML/DL methods in data processing and analysis
- Methods of computational physics for the study of complex systems
  - High-Performance Computing for Complex System Studies
  - Development and application of methods of computational mathematics in quantum information theory

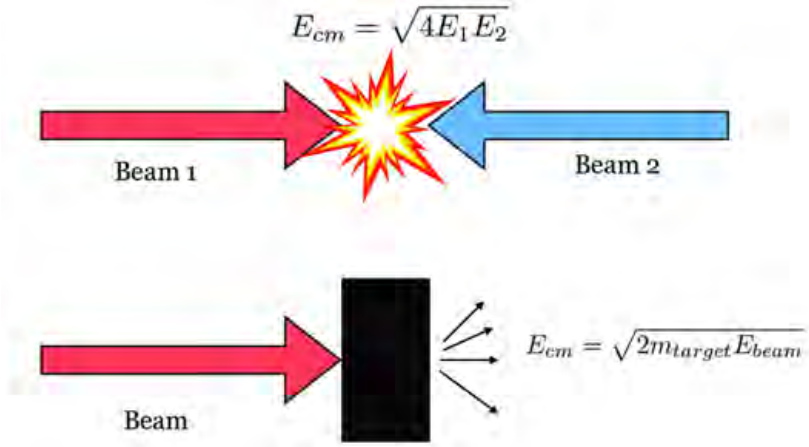
Activities

- Big Data Analytics and Quantum intelligent control of technological processes and physical facilities at JINR
- Digital JINR
- IT-school

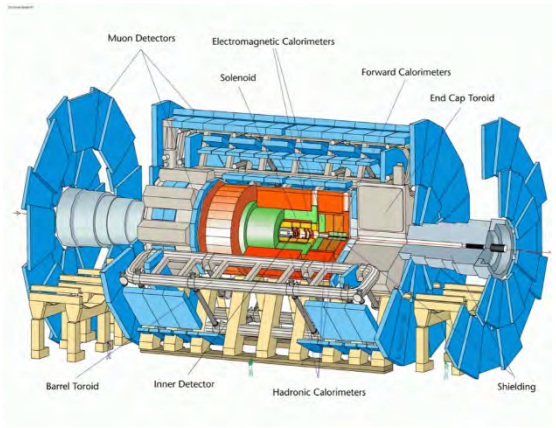




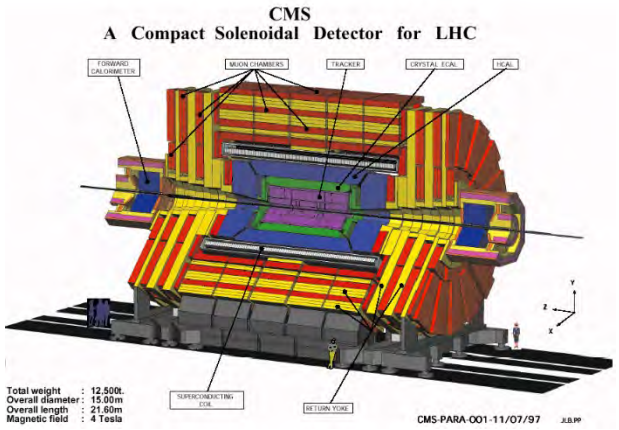
Examples of HEP Experimental Facilities



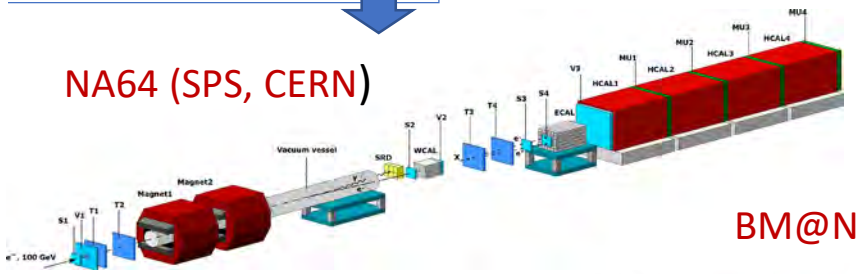
ATLAS (LHC, CERN)



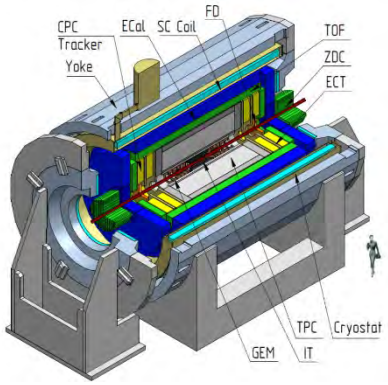
CMS (LHC, CERN)



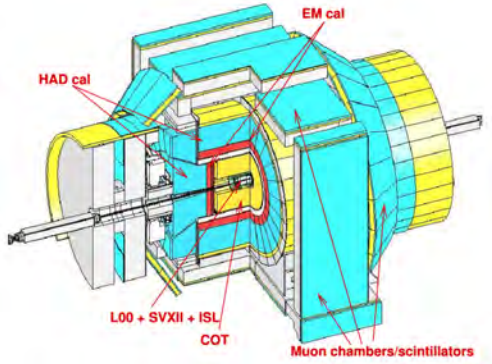
NA64 (SPS, CERN)



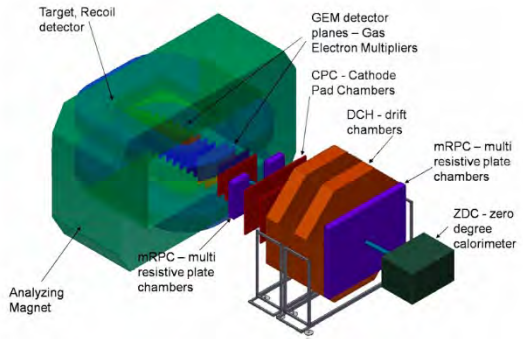
MPD (NICA, JINR)



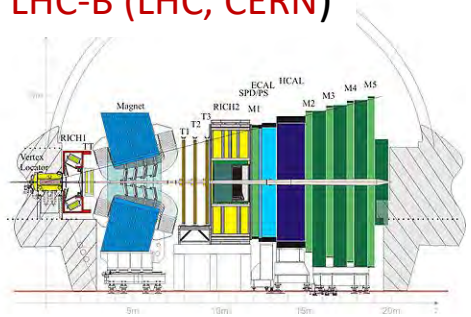
CDF (Tevatron, FLab)



BM@N (NICA, JINR)

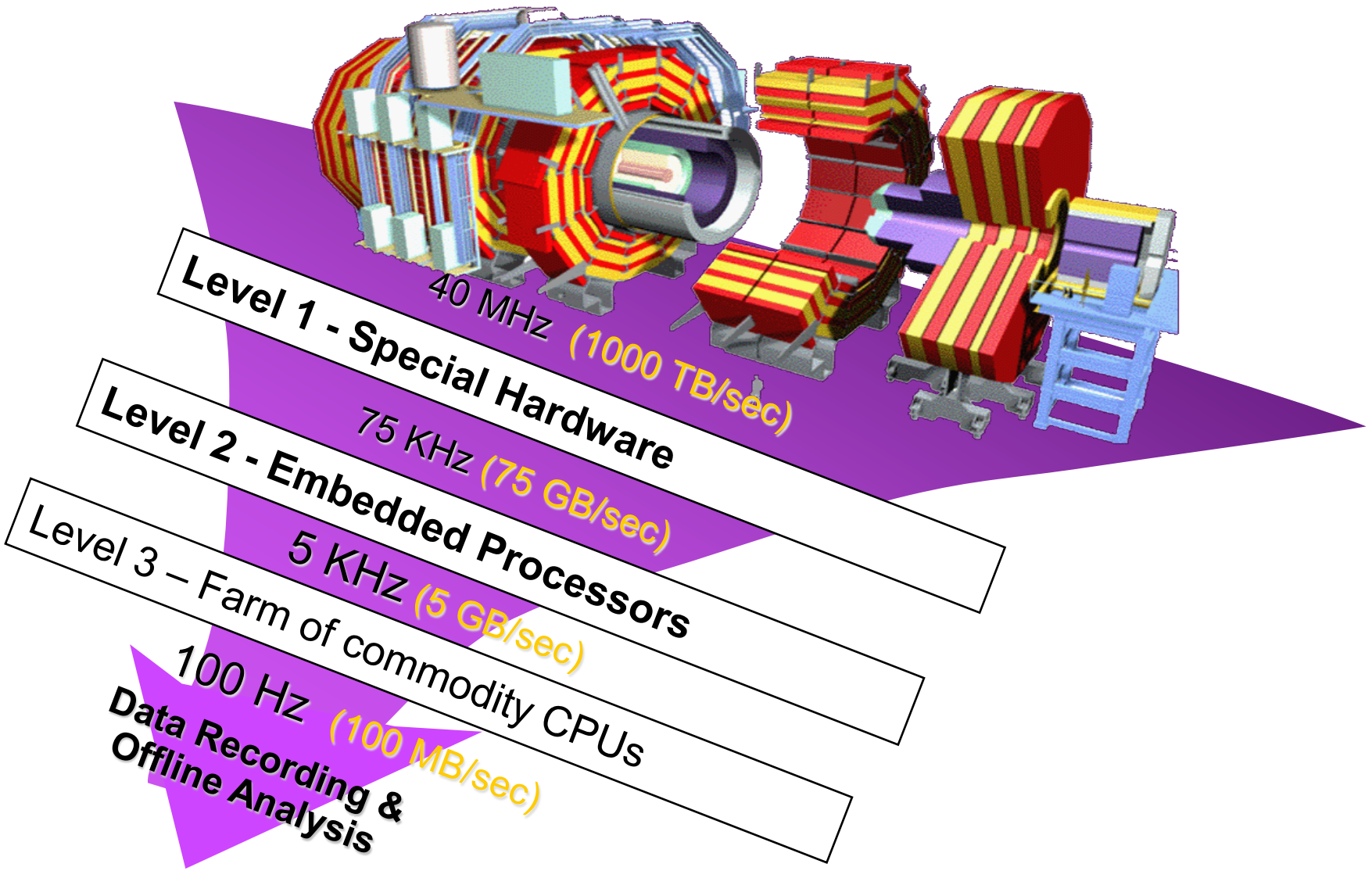


LHC-B (LHC, CERN)

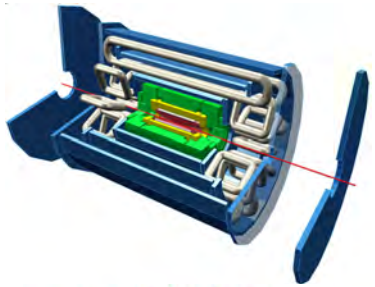


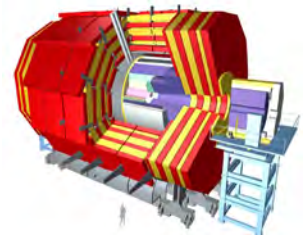
Scale without preservation of proportions

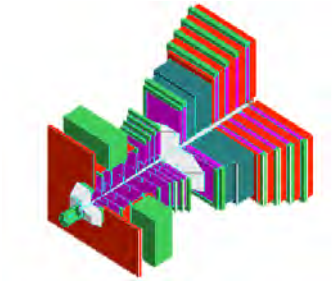
Example of CMS Trigger Rates

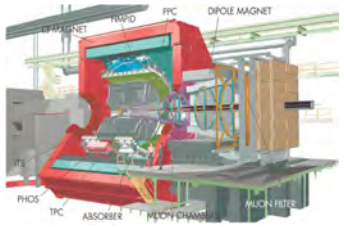


Example of LHC Data Flows









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

~ 3PB/year

Big Data/GRID Computing

BM@N
Baryonic Matter at Nuclotron

NICA MPD

SPD NICA

ARIADNA beamlines

ОНЛАЙН-КЛАСТЕР NICA

МНОГОФУНКЦИОНАЛЬНЫЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫЙ КОМПЛЕКС ОИЯИ

НАЦИОНАЛЬНАЯ
ИССЛЕДОВАТЕЛЬСКАЯ
КОМПЬЮТЕРНАЯ СЕТЬ

- ~ 200 ПБ/год данных
- Скорость передачи данных:
  4×100 Гбит/с

МНОГОФУНКЦИОНАЛЬНЫЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫЙ КОМПЛЕКС ОИЯИ

ОБЛАЧНАЯ ИНФРАСТРУКТУРА

- 5152 вычислительных ядер
- 4 петабайта

ИНФРАСТРУКТУРА РАСПРЕДЕЛЁННЫХ ВЫЧИСЛЕНИЙ

- 30 000 вычислительных ядер
- 20 петабайт

СУПЕРКОМПЬЮТЕР «ГОВОРУН»

- 1,7 петафлопс двойной точности
- 26 петафлопс для задач ИИ
- 8 петабайт

РАСПРЕДЕЛЁННОЕ ХРАНИЛИЩЕ ДАННЫХ

- 130 петабайт для хранения данных

РАСПРЕДЕЛЁННАЯ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНАЯ ПРОГРАММНАЯ СРЕДА – ПЛАТФОРМА DIRAC

КЛАСТЕР NICA/ОИЯИ

NICA – ОИЯИ: РАЗВИТИЕ ГЛОБАЛЬНОЙ СЕТИ ПАРТНЁРСКИХ ВЫЧИСЛИТЕЛЬНЫХ ЦЕНТРОВ

РОССИЯ · АЗЕРБАЙДЖАН · БЕЛАРУСЬ · БОЛГАРИЯ · ГРУЗИЯ · ЕГИПЕТ · УЗБЕКИСТАН · МЕКСИКА

Project for HEP Experiments

The project is aimed at

- the development of mathematical methods and software for modeling physical processes and experimental facilities, processing and analyzing experimental data in the field of elementary particle physics, nuclear physics, neutrino physics, condensed matter, radiobiology, etc.
- development, support and operation of software complexes for the distributed processing and analysis of experimental data as well as information systems to support research at JINR and other world centers.

The Project Structure

Simulation of Physics Processes and Facilities	Reconstruction and Data Analysis	Software Environment for Experiments
Physics event simulation	Particle trajectory reconstruction	Data processing and analysis models
GEANT-simulation of experimental setups	Particle identification	Data models
	Reconstruction of physics processes	Software platforms and systems
	Experimental data analysis	Development and maintenance of DBs
		Event visualization

The main strategy is to use common solutions and methods for different experiments

The priorities are mathematical and computational physics to support the JINR large research infrastructure projects, and first of all the experiments at the NICA accelerator complex and the Baikal-GVD neutrino telescope.

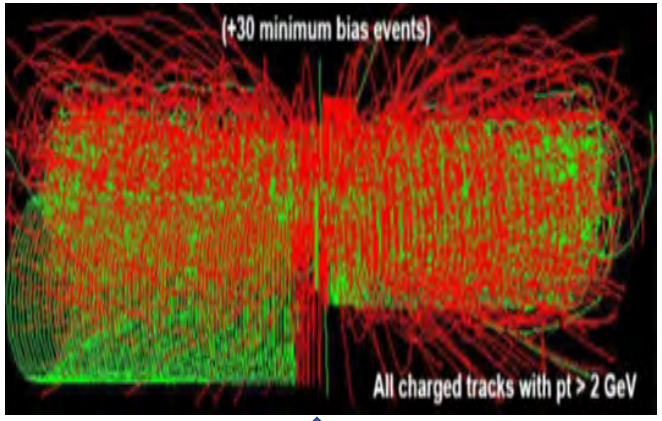
Further cooperation will also be continued with the experiments at the largest world accelerator centers (CERN, BNL, etc.), experiments in the field of neutrino physics and astrophysics, radiobiological research programs.

The possibility of using the developed methods and algorithms within other fundamental science and applied projects is being considered.

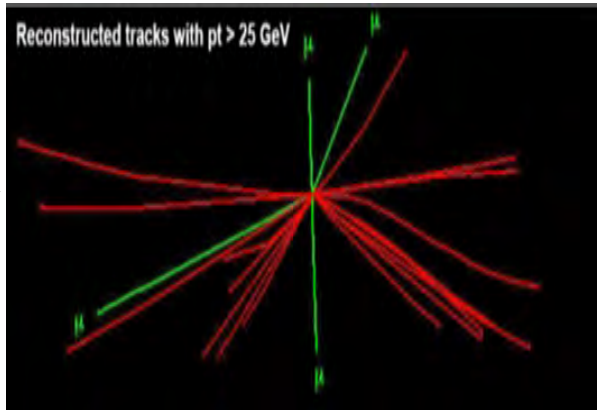
Data Processing and Analysis

RAW Data

Reconstruction

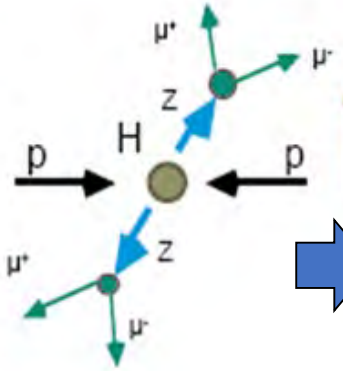


Event Selection

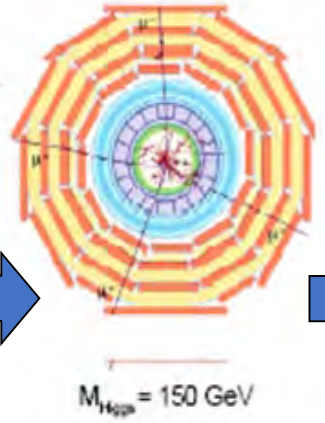


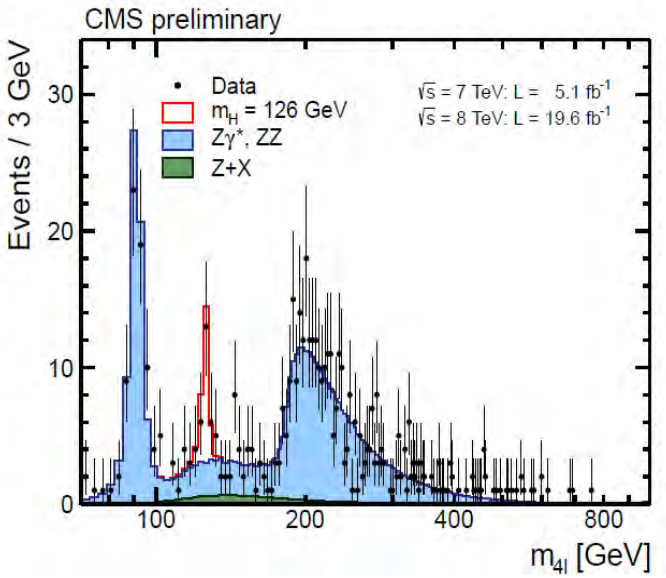
Calibration/Condition/etc Data Bases

Theory/ Monte Carlo Simulation



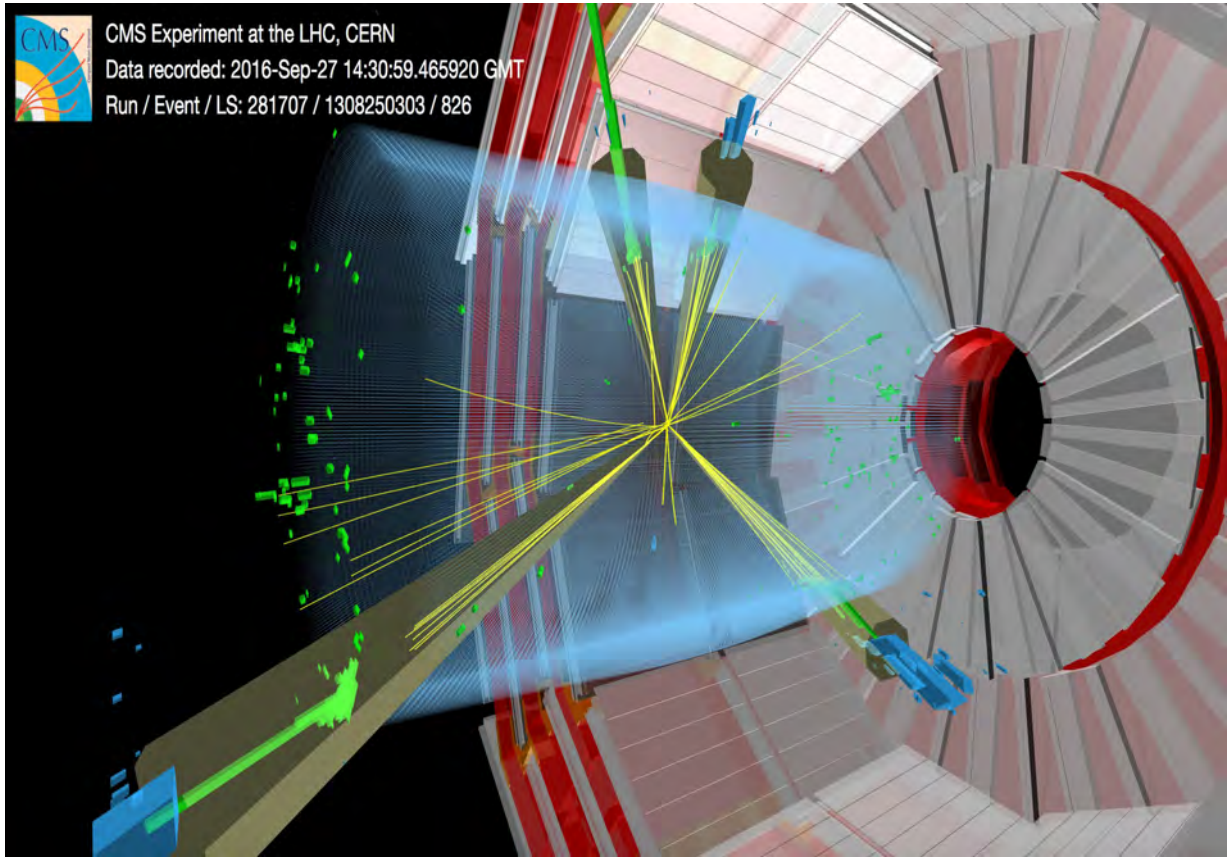
Detector Response Simulation





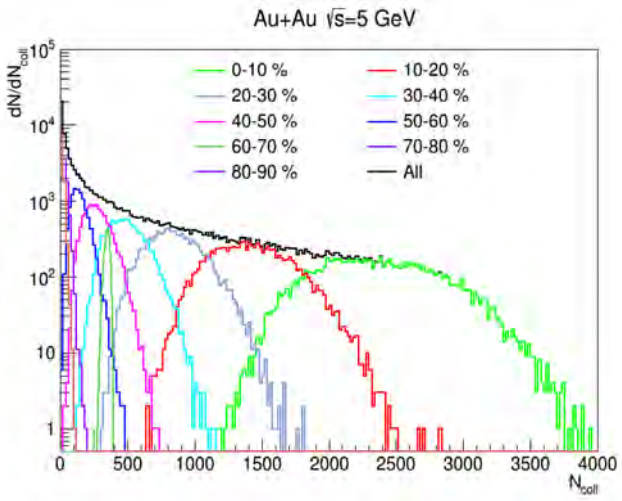
Simulation of physical processes and experimental facilities

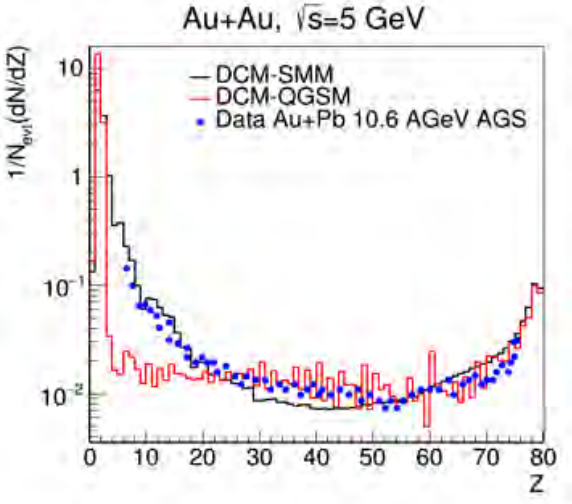
- analytical and numerical calculations of physical processes, software optimization, including tuning and adaptation of physics event generators;
- MC event production, development and support of information systems for event catalogues;
- participation in the creation of computer models of experimental facilities and simulation of elementary particles passing through them based on GEANT4 (and others) and fast simulation of the response of the detectors.



MC Generators for the NICA and LHC Experiments

- Development of the heavy ion collision generators
  - Dubna Cascade Model, Quark-Gluon-String Model, Statistical Multifragmentation Model for the NICA Experiments
  - tuning the HIJING generator with data of NA49 and NA61/SHINE @ CERN, STAR@RHIC (can be used in MPD and SPD experiments)
- Analytical and numerical methods for calculating neutron-proton systems under strong compression at the NICA



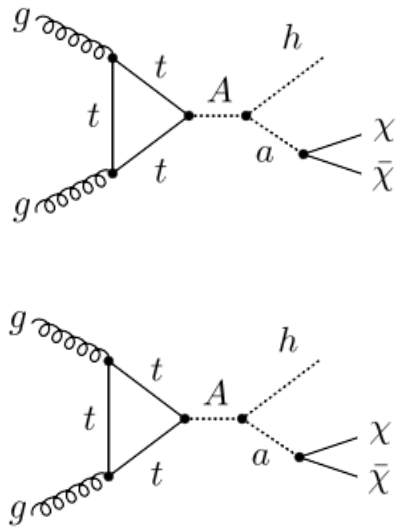


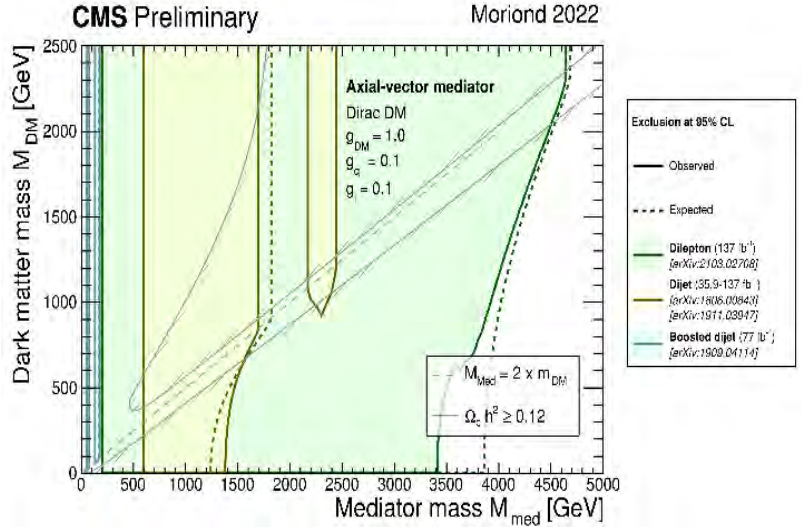
The priority tasks of the JINR Groups in LHC physics program (ATLAS and CMS) include searches for candidates for dark matter particles, tests of predictions of TeV-energy scenarios

- Fine tuning the generators for searches for new physics
  - revision of model parameters for 2HDM+a, 2HDM+s, etc.
  - simulation with Pythia8, QBH, MadGraph5_aMC@NLO + FeynRules (simplified DMM, HDM+a, 2HDM+s, etc.)
  - mass production + Geant4 response

Ex., Dark Matter can be probed with two fermions/two fermions + MET/higgs + MET/Z + MET in the final states

h (→ bbar) + a (→ χχ) = bbar + MET



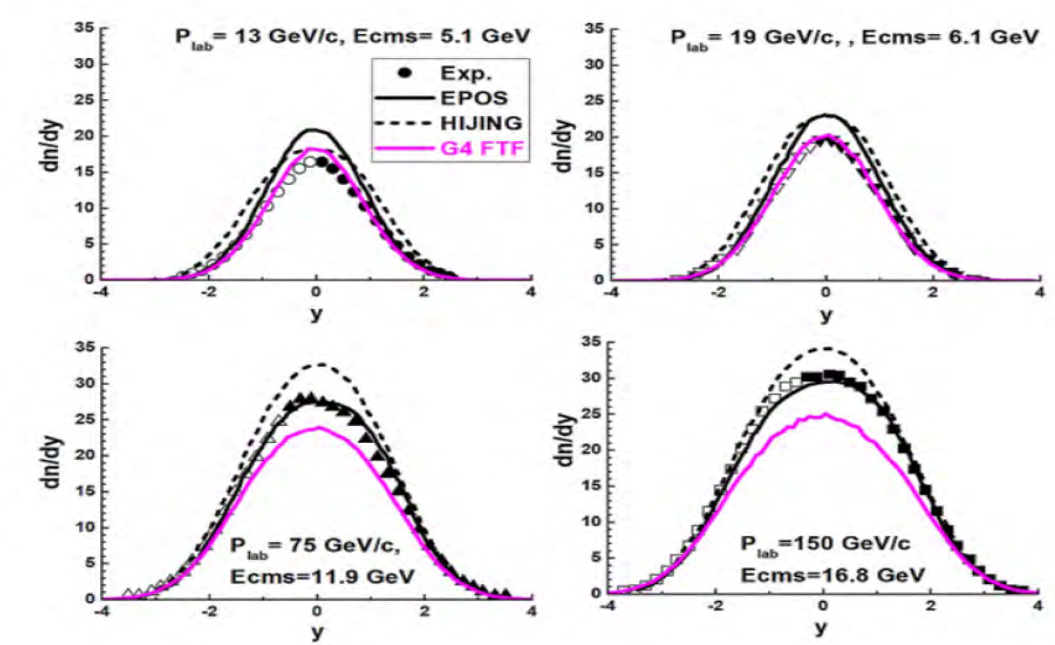


MLIT + BLTP + VBLHEP

Modeling Experimental Facilities for the NICA and LHC

- Development, verification, validation and application of FTF (Fritiof) and QGSM (Quark-Gluon-String-Model) hadronic models

Rapidity distributions of π^- mesons in $^{40}Ar + ^{45}Sc$ interactions (EPJ , C82 (2022)

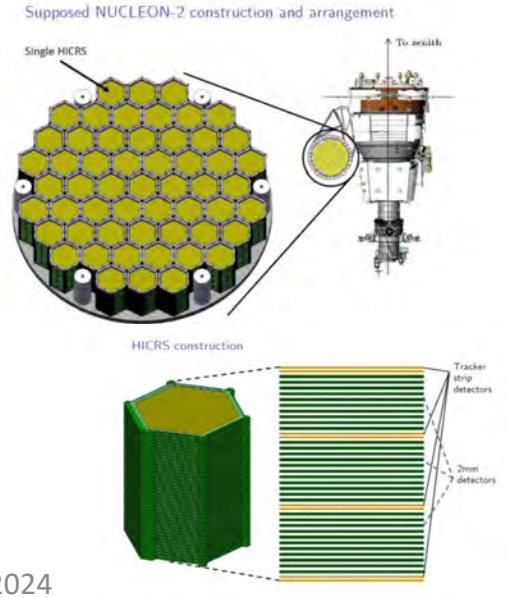


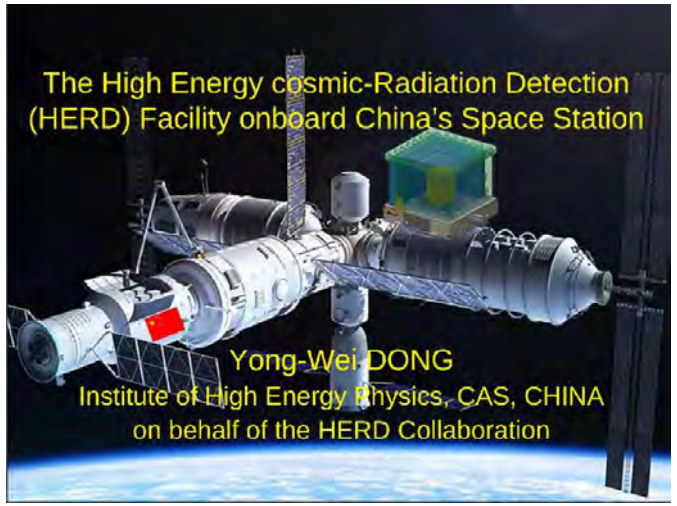
Exp. conclusion: “There is no model (EPOS, UrQMD, HIJING …) able to describe the data!” from NA61/SHINE Collab. on PP, $^{40}Ar + ^{45}Sc$ and $^{7}Be + ^{9}Be$

- Simulation of the HGCal cassettes test stand for the CMS Experiment @ LHC with cosmic muons in Cold room
  - modeling and optimization the test stand geometry
  - scintillator trigger planes optimization
  - performance estimations, detector evaluation algorithms tests

Cold room trigger planes position options

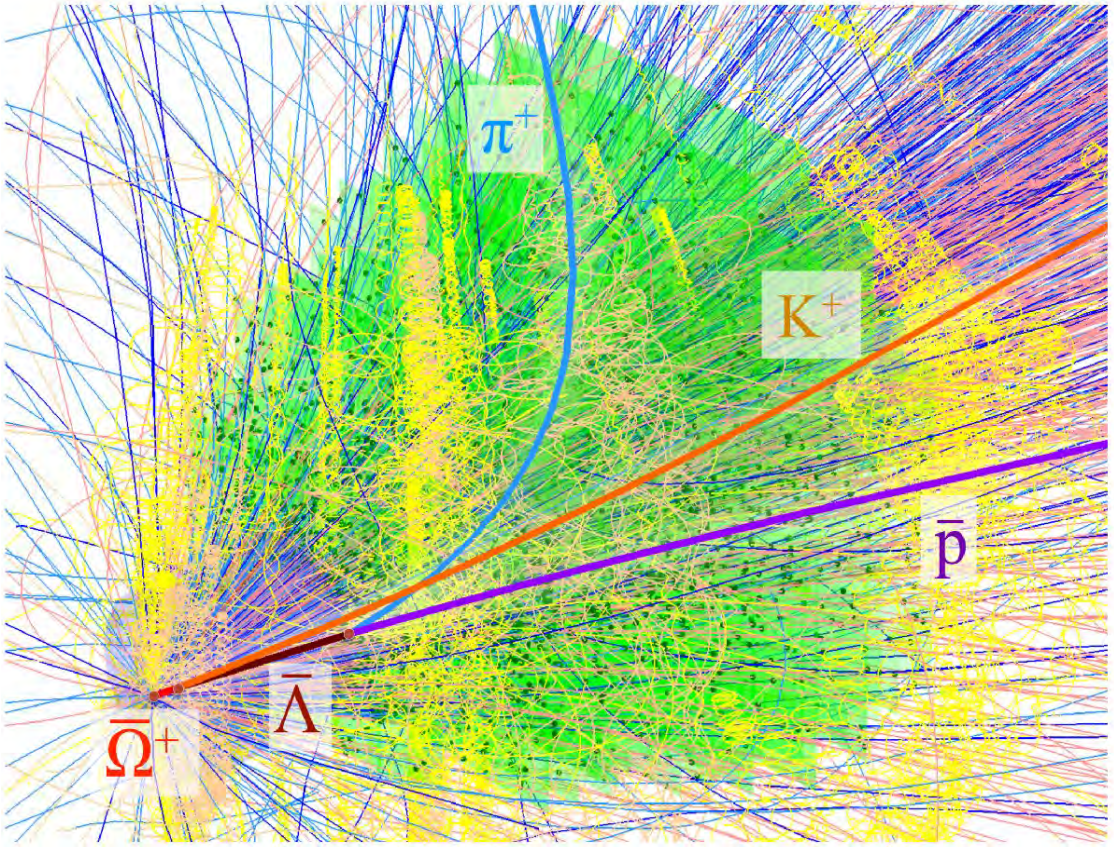
- Simulation and prototype testing for present and future orbital detectors: NUCLEON, NUCLEON-2, HERD





Reconstruction of physical objects and analysis of experimental data

- development of algorithms, including those based on recurrent and convolutional neural networks for machine and deep learning tasks, and creation of corresponding software for the reconstruction of physical objects (tracks, particles, clusters, etc.) and physical processes;
- development of methods and algorithms for data analysis, including statistical analysis;
- adaptation of existing software for specific experiments, reconstruction and analysis of experimental data;
- analysis of Open Data of experiments, in particular, experiments at the LHC;
- conducting a global analysis of data from various experiments (in particular, a combined analysis of data from accelerator and astrophysical experiments in search for candidates for the role of dark matter).



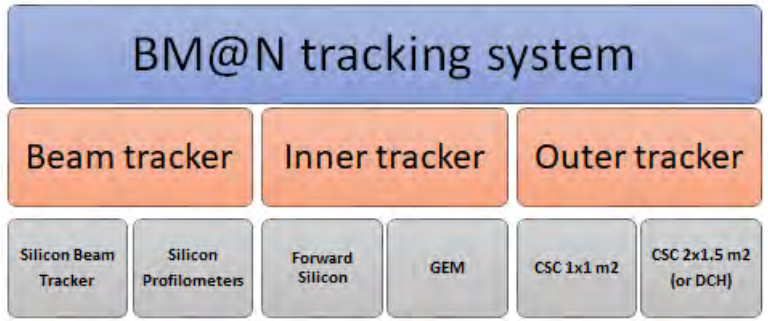
LHC, pp
<μ>~52 @ 13.6 TeV (RUN3, 2023)
~2000 tracks per events

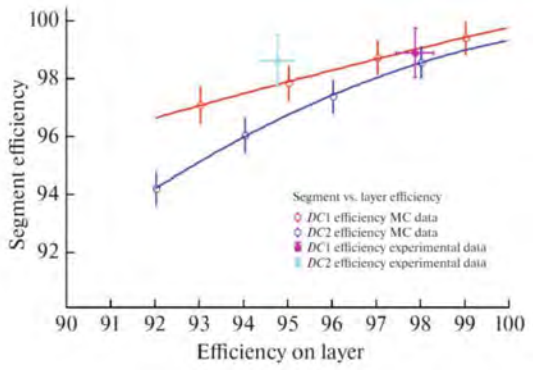
NICA/MPD, AuAu
central collisions @ 11 GeV
~1000 tracks per events

ML methods for AI analysis and processing of HEP data ⇒ G. Ososkov

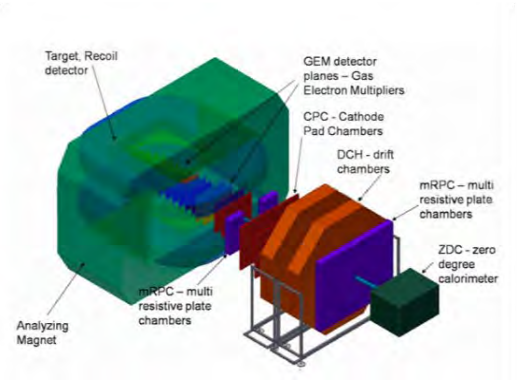
Tracking Algorithms for HEP Experiments

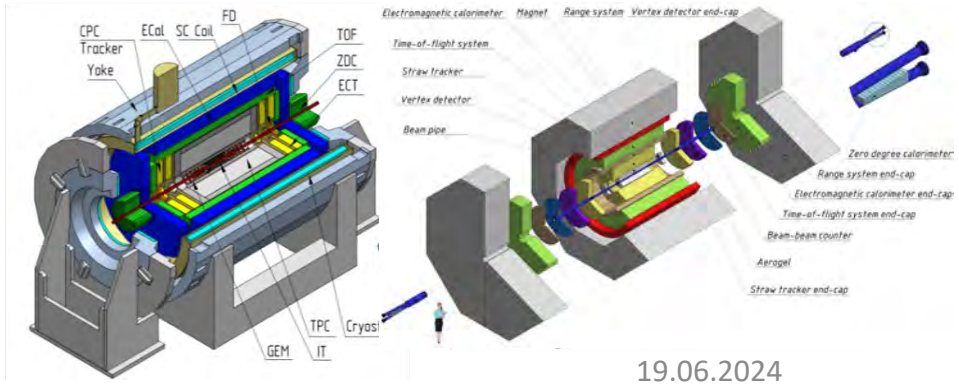
- Mathematical methods and software for processing and analyzing data from the experiments @ NICA
  - software for alignment and calibration of the BM@N STS (silicon chambers) and GEM (gas electron multipliers) track detectors



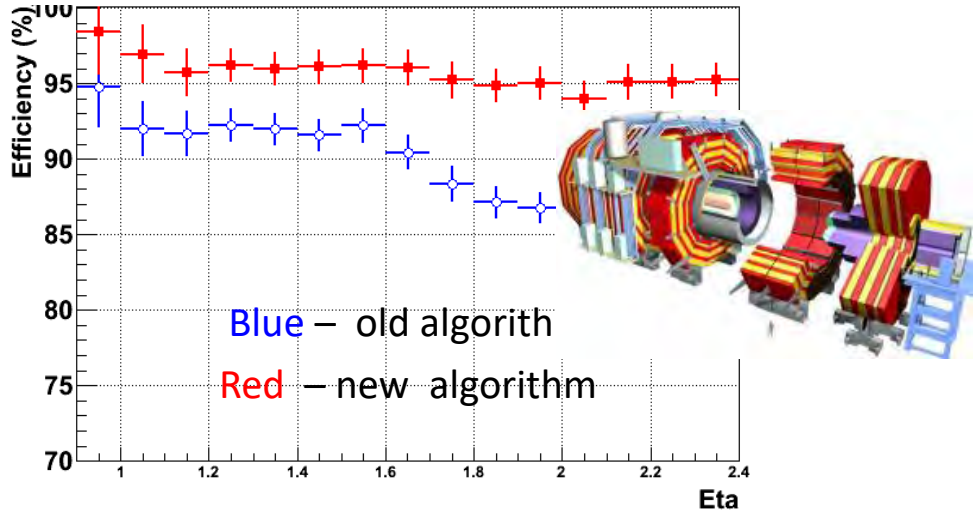


  - development and application of methods and algorithms for processing and analyzing experimental data for the coordinate detectors (MPD/SPD)





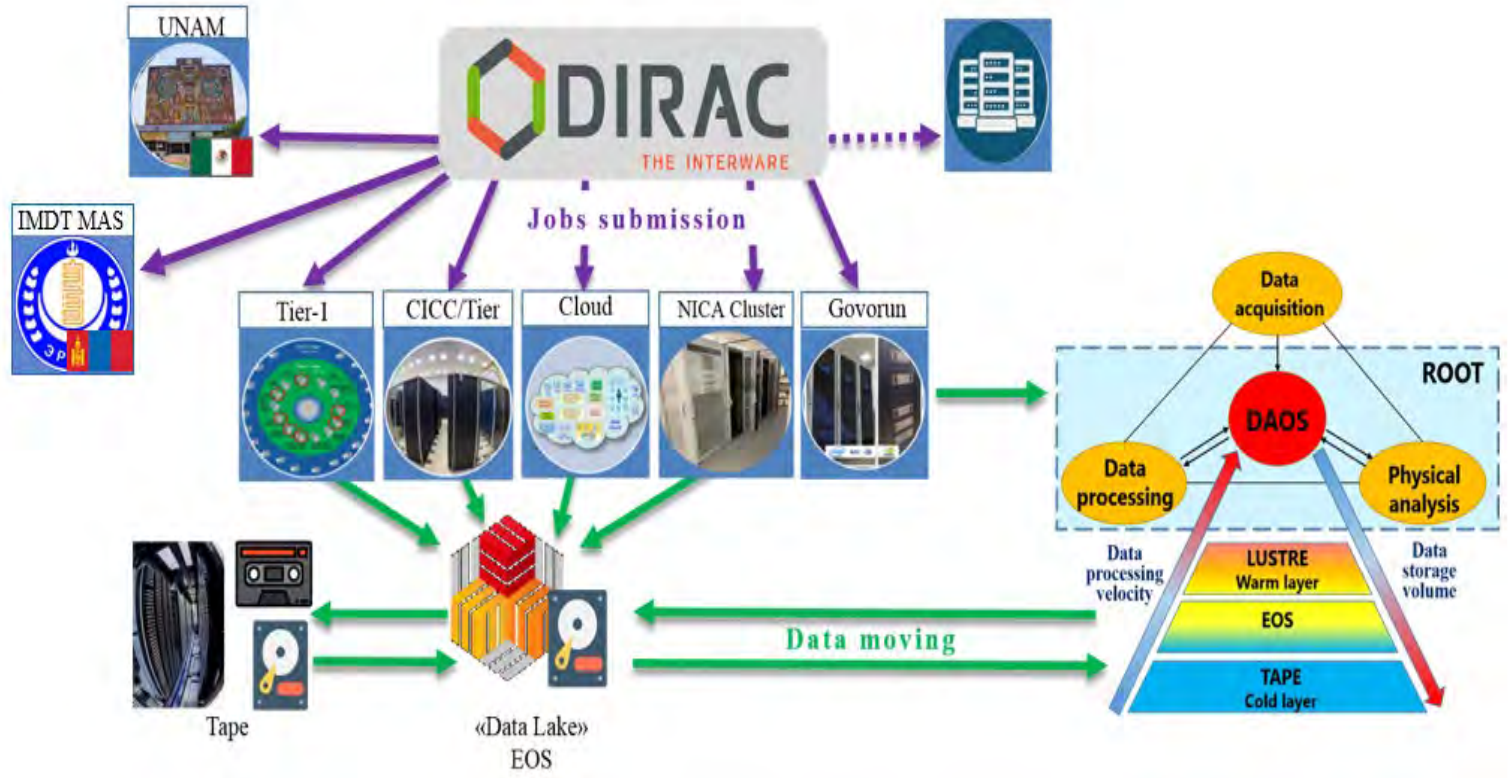
- Mathematical methods and software for muon reconstruction and the estimation of operation parameters of CMS detectors @ LHC
  - reconstruction of the cosmic muon trajectory in the setup for testing active elements of the CMS HGCal, as well as evaluation of the efficiency of HGCal modules;
  - usage of discrete wavelet analysis to recognize the coordinates of close-flying particles from over-lapping signals in the Cathode Strip Chambers (CSC). Evaluation of the operation parameters of CSC detectors and of the rate of background particles for different types of experimental data.



Software for Computing Environment

To ensure uninterrupted operation of MICC and for efficient use of computing resources, core software development is required

- distributed software defined high-performance computing platform for processing and storing data of experiments, combining supercomputer (heterogeneous), grid and cloud technologies for the efficient use of new computing architectures
- multipurpose software and hardware platform of Big Data analytics based on hybrid hardware accelerators (GPU, FPGA, quantum systems); machine learning algorithms; analytics, reporting and visualization tools; support for user interfaces and tasks



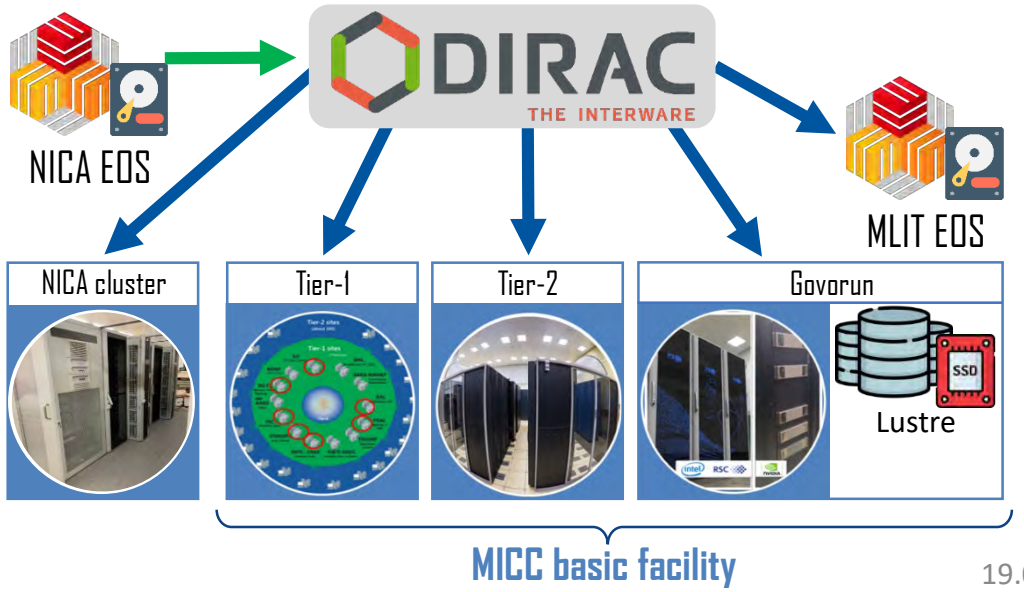
- advanced systems for the protection of cyber infrastructure, computer and user information, public e-services and user authentication

BM@N Run 8 Data

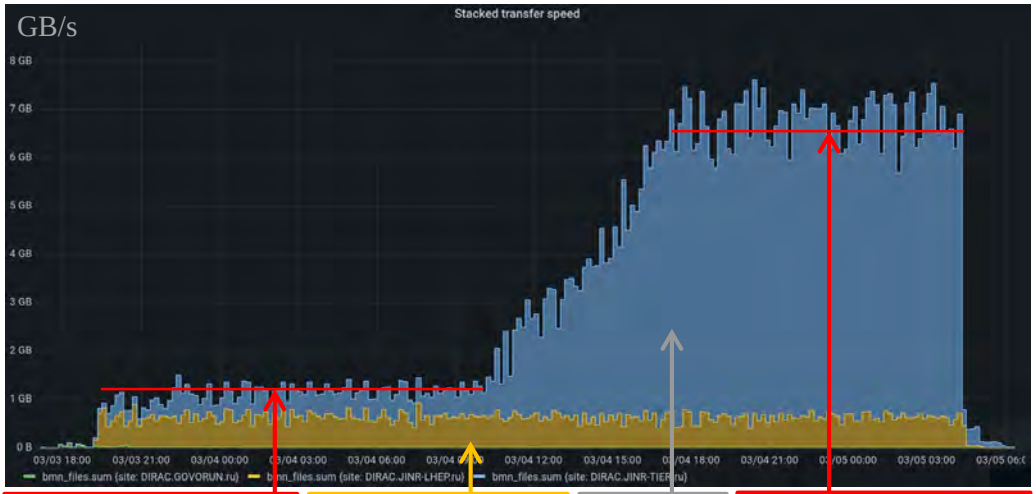
The 8th BM@N physics run was the first time at JINR when the entire computing infrastructure, integrated by DIRAC, was used for the complete reconstruction of raw experimental data. During the session, there were received about 550 million events, which were written in 31,306 files with a total size of more than 430 TB.

The reconstruction process was carried out in two stages:

1. Raw → DIGI (99% processed on Tier1 and the NICA cluster, large files (16 - 250 GB) could only be processed on the “Govorun” SC)
2. DIGI → DST (Tier1, Tier2, NICA cluster and “Govorun” SC)



Raw → DIGI: High disk and network system load



300 jobs
4 MB/s per job

NICA cluster

Tier1

1,580 jobs
4.1 MB/s per job

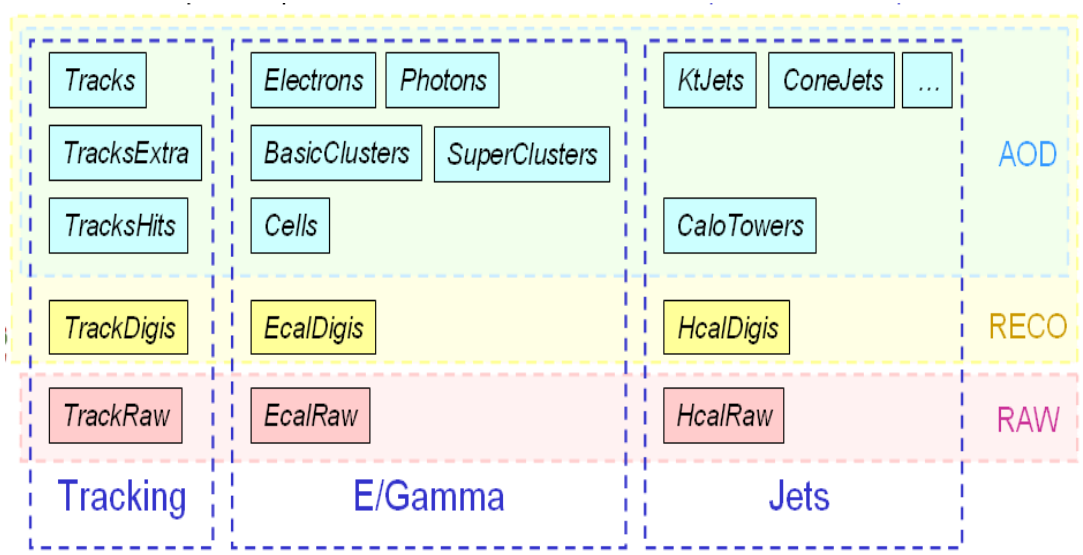
Maximal data transfer speed (Read+Write) with EOS (MLIT) – 7.5 GB/s

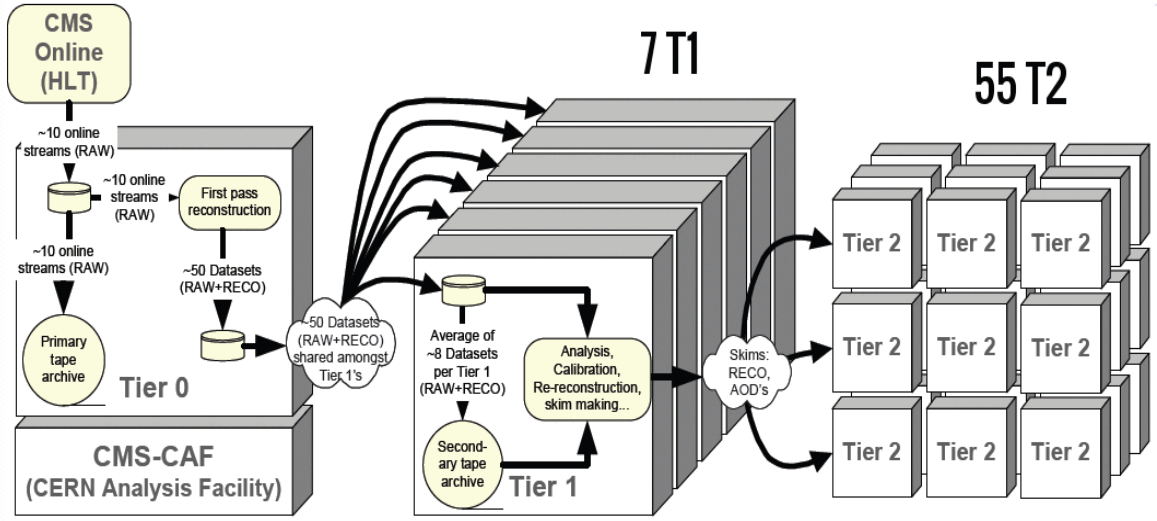
Jobs completed
62612

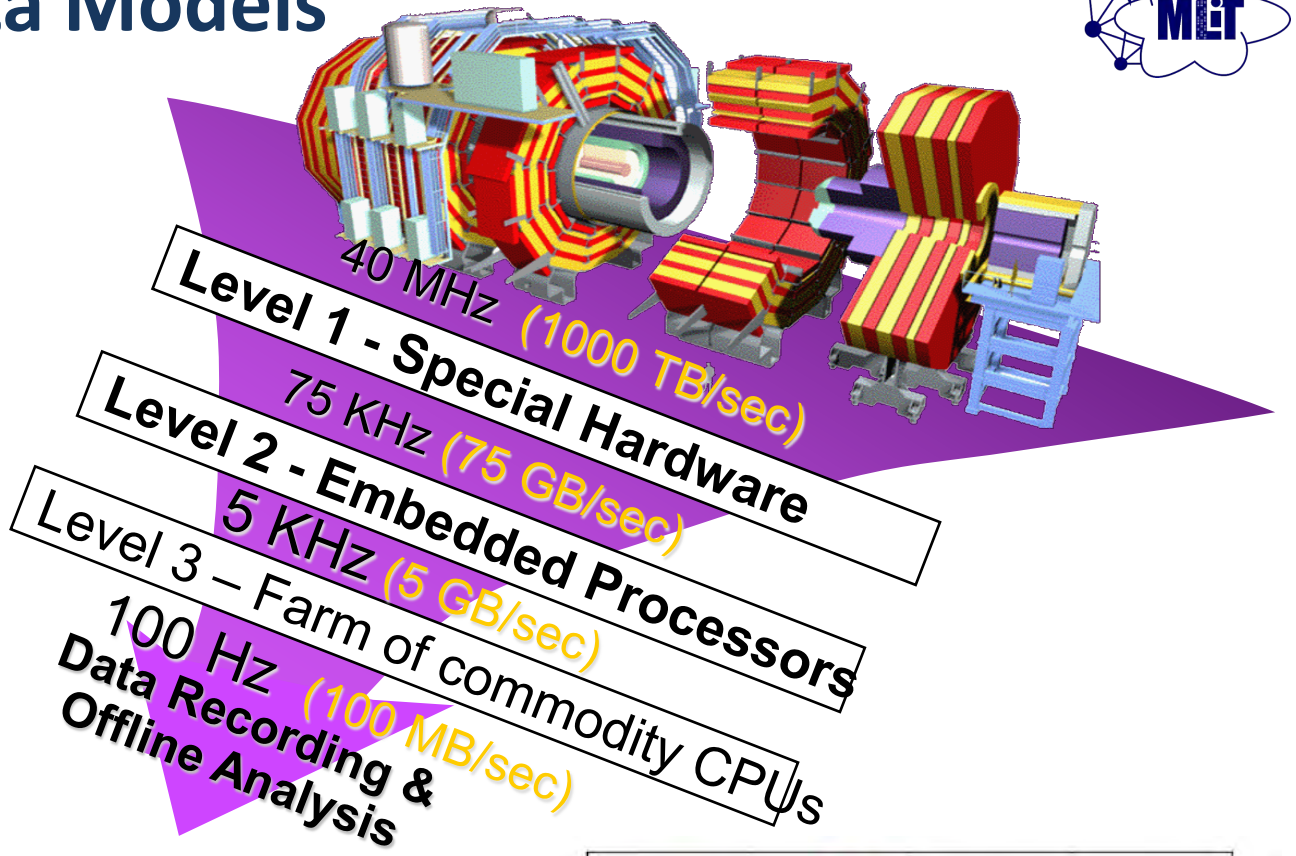
Real time
~48 h

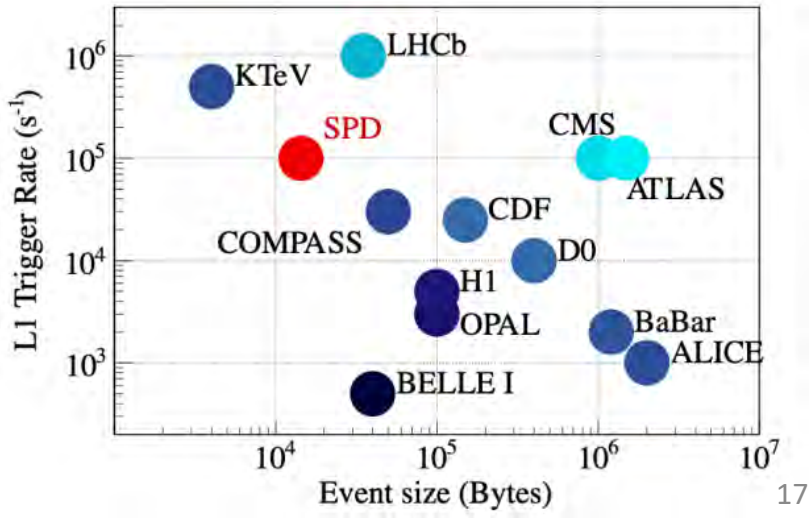
RAW
436 TB
→
DIGI
23 TB
→
DST
53 TB

HEP Data Models





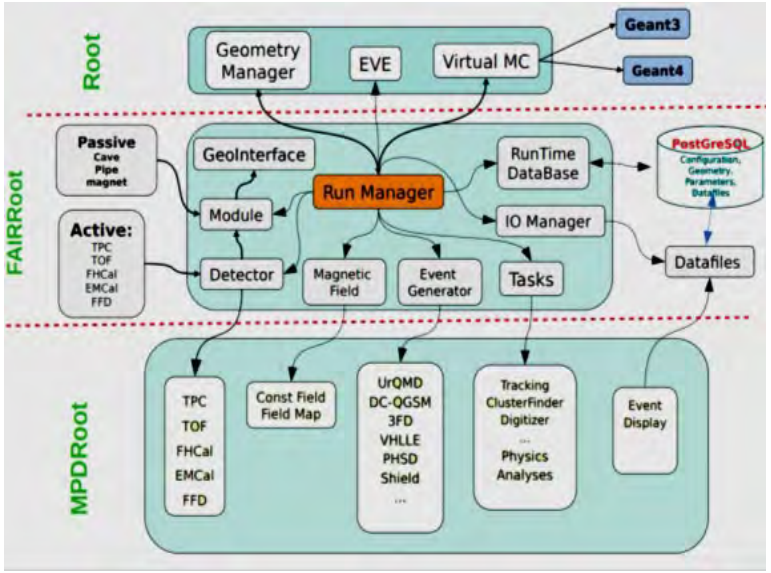




Software Environment for the NICA Experiments

Collaboration SW for MPD

- refactoring of the MPDroot architecture; its adaptation to the latest changes in the FairRoot package
- development and support of the NICAdist system for building modules;
- improving the performance of MPDroot algorithms

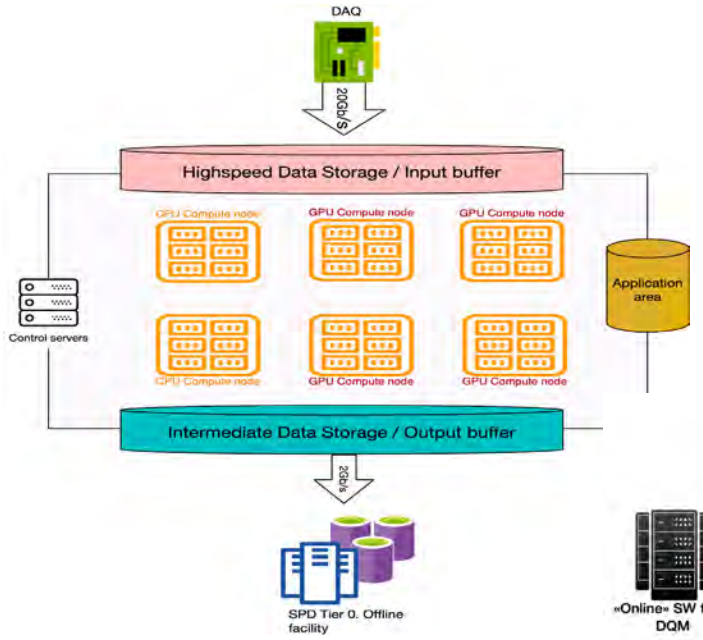


Middleware SW for MPD

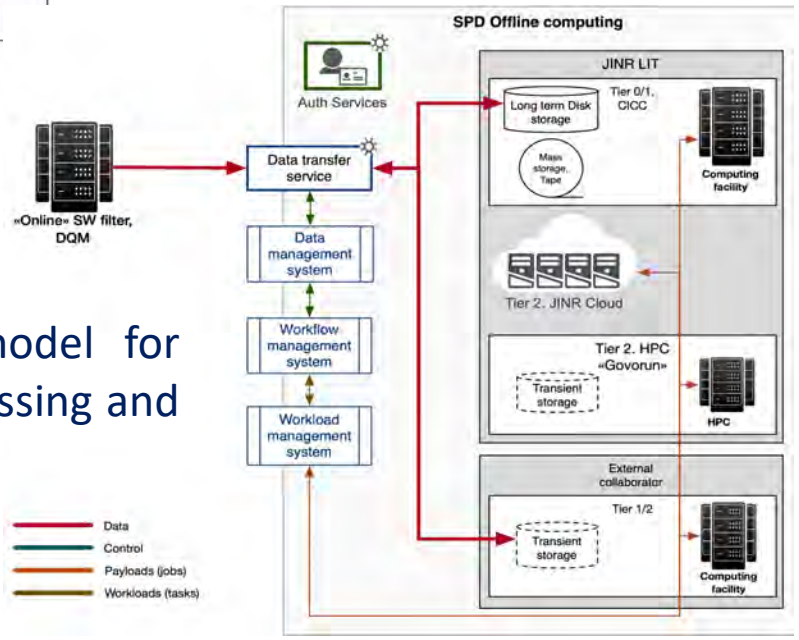
- to run MPD tasks through the DIRAC Interware
- MPD distributed computing model

SPD OnLine Filter for SPD

- design and creation of a software complex for organizing the high-throughput processing of data received from the SPD DAQ



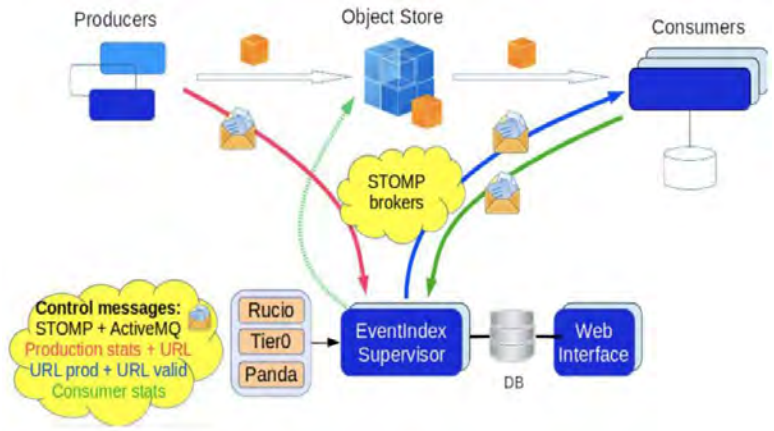
- ✓ data flow ~3M ev./s. (~20 GB/s)
- ✓ no trigger
- ✓ middleware complex for dedicated computing farm
- ✓ methods and approaches to filtering data in real time using artificial intelligence technologies



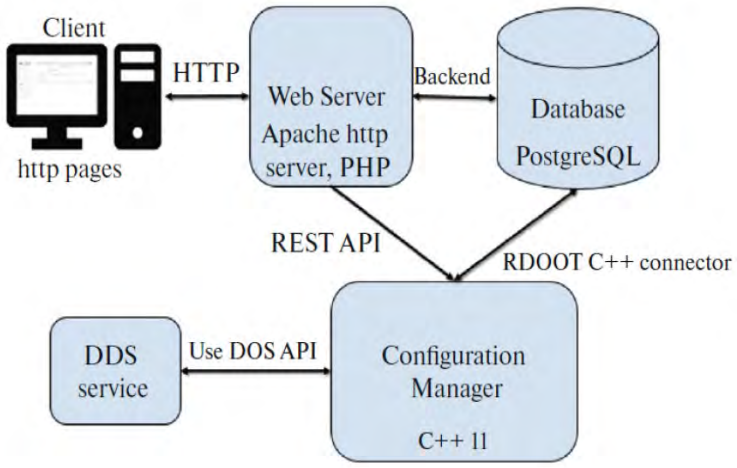
- development of a model for distributed data processing and analyses

Data Bases for HEP Computing

- ATLAS EventPickingService
  - the first version for automated event collection has been created.
  - the service was used for the second stage oa the «γγ→WW» analysis (136 K events)
  - further modernization of the service is ongoing according to the received results

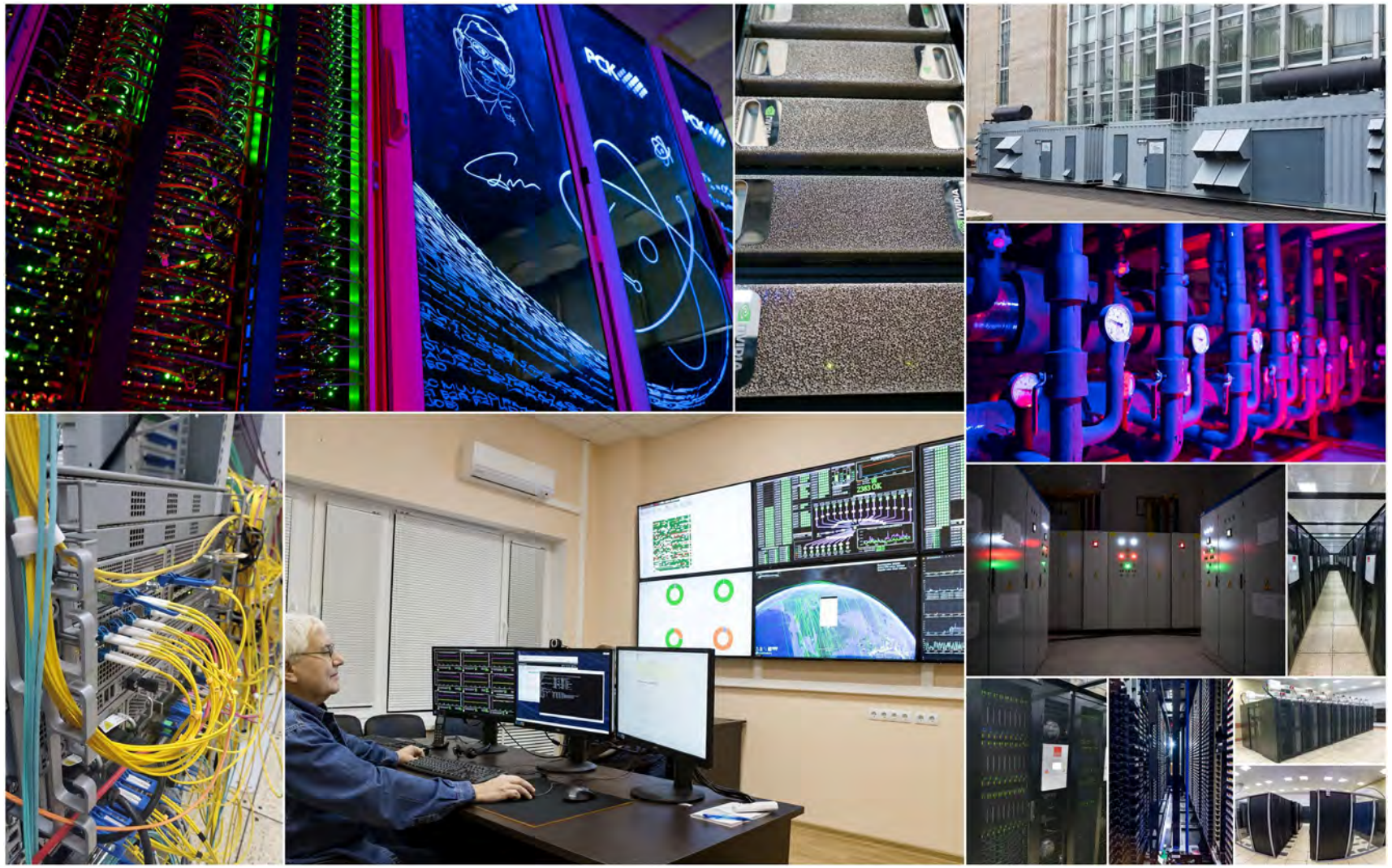


- ATLAS CREST (Condition DB)
  - C++ API for CREST (implemented into Athena software package)
  - COOL2CREST converter (developed, but not yet implemented)
  - both parts require constant improvement for compatibility with the updated CREST server
- Event metadata system for the experiments at NICA (BM@N, MPD and SPD)
  - has been designed to index events of the NICA experiments and store their metadata
  - quick search by required conditions and parameters used in various physics analyses for a set of physics events to use in further event data processing



- Geometry DB
  - The geometry database is the main element of the information system designed to store, process and manage information about the geometric models of detectors. It is implemented into BMNROOT software
- Configuration DB (Configuration Information System, CIS)
  - Configuration database for managing online applications for collecting and processing high energy physics events. It was created for process launching in BM@N environment.

Thank you for your attention!



10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)

- Distributed Computing Systems
- HPC
- Cloud Technologies
- Distributed Storage Sys
- Distributed Computing
  education, industry and
- Computing for MegaSci
- Quantum informatics a
- Big Data, Machine Lear

MMCP

MATHEMATICAL MODELING AND COMPUTATIONAL PHYSICS

- methods, software and program packages for data processing and analysis;
- mathematical methods and tools for modeling complex physical and technical systems, computational biochemistry and bioinformatics;
- methods of computer algebra, quantum computing and quantum information processing;
- hine learning and big data analytics;
- thms for parallel and hybrid
  ions.

MMCP2024 will be held @ Yerevan in October 2024
visit https://mmcp.jinr.ru/ site for details

Modernization and development of MICC

To achieve the main goals of the key JINR projects, a huge amount of experimental data will have to be processed. Very roughly, these are tens of thousands of processor cores and hundreds of petabytes of long-term raw data storage.

- Modernization and development of computing infrastructure
  - distributed computing platform for NICA experiments (Tier-0/1 centers) is assumed to be created (15.6 Pflops CPU, 142 PB Disk, 546 PB Tape for 2030)
  - data storage and computing capacity of the WLCG project should increase annually by 10–20% for CERN experiments (Tier-1/2 grid clusters)
  - creation of a computing infrastructure for the JINR neutrino program
  - expansion of JINR cloud infrastructure and the heterogeneous HybriLIT platform, including the “Govorun” SC
- Modernization of the engineering infrastructure of the MICC JINR, the further development of the JINR network infrastructure (400/800 Gbps), and distributed EOS storage system (data lakes)
- Priorities
  - large research infrastructure projects (NICA, JINR research programme in the field of neutrino physics and astrophysics)
  - science programme in the field of elementary particle physics and new physics beyond the Standard Model (experiments in CERN and in other World research centers, theoretical calculations and modeling)
  - modern radiobiology, astrobiology, neurophysiology, molecular biology and genetics, etc

From JINR 7-y Plan

		2024	2025	2026	2027	2028	2029	2030
LHC Tier1 (CMS)	CPU (Pflops)	1,53	1,69	1,84	2,03	2,22	2,45	2,68
	Disk (PB)	18	20	25	28	31	34	40
	Tape (PB)	46	50	60	70	80	90	100
	Network (Gbps)	200	400	400	600	600	800	800
LHC Tier2 (ATLAS, CMS, ALICE, LHCb) + etc.	CPU (Pflops)	0,73	0,81	0,88	0,96	1,04	1,15	1,27
	Disk (PB)	7,7	8,5	9,2	10	11,	12,80	14
	Network (Gbps)	200	400	400	600	600	800	800
“Govorun” supercomputer	CPU (Pflops)	1,2	2,2	3,2	4,2	5,2	6,2	7,2
	Disk (PB)	8	9	10	11	12	13	14
DataLake	Disk (PB)	60	60	60	80	80	80	100
*NICA Tier 0,1,2	CPU (Pflops)	2,2	2,6	8,6	8,6	15,6	15,6	15,6
	Disk (PB)	17	24	47	75	96	119	142
	Tape (PB)	45	88	170	226	352	444	536
	Network (Gbps)	400	400	400	400	400	400	400
*Baikal-GVD, NOvA, JUNO, DUNE Tier 0,1,2	CPU (Pflops)	0,94	1,02	1,2	1,28	1,36	1,54	1,62
	Disk (PB)	1,9	3,2	3,5	3,8	4,6	4,9	5,2
	Tape (PB)	9	12	15	18	21	24	27
	Network (Gbps)	200	200	200	200	200	200	200

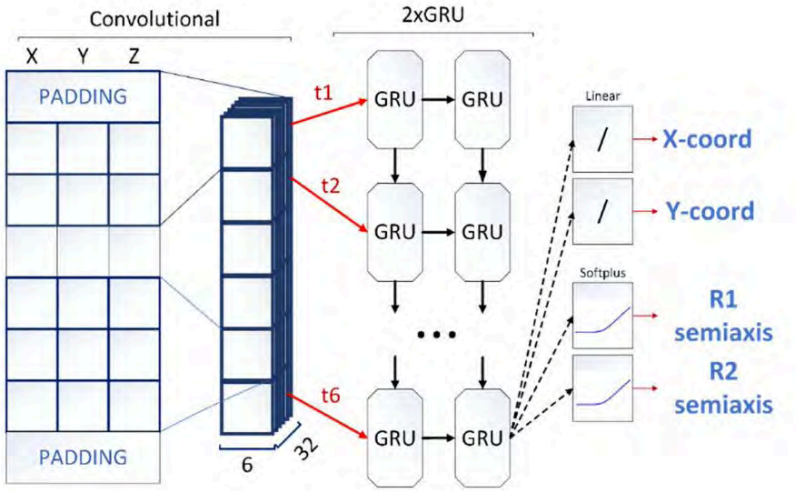
* The financing of computing resources for computing under the NICA project and the JINR neutrino program is carried out within the budgets of the corresponding directions

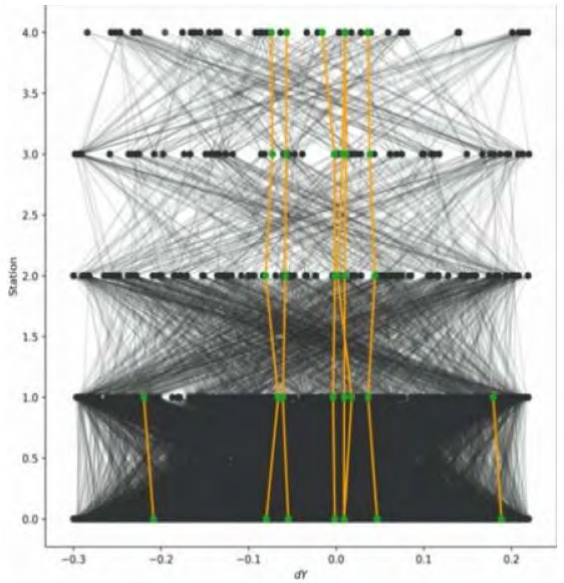
Machine Learning Methods for the Track Reconstruction and PID

- particle identification based on the gradient boosting of decision trees with MPD Time-Projection Chamber (TPC) and Time-of-Flight (TOF)

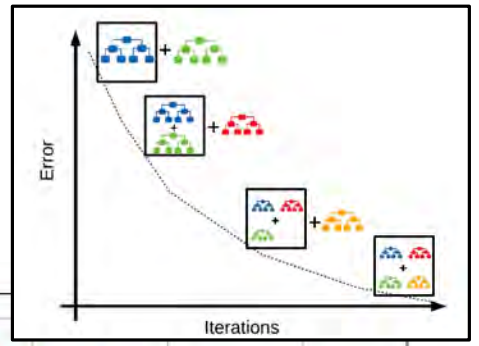
- new approaches to track recognition in SPD strip and pixel detectors based on a recurrent neural network and a graph network (already used for track recognition in the BM@N experiment at JINR and in the BESIII experiment in China)

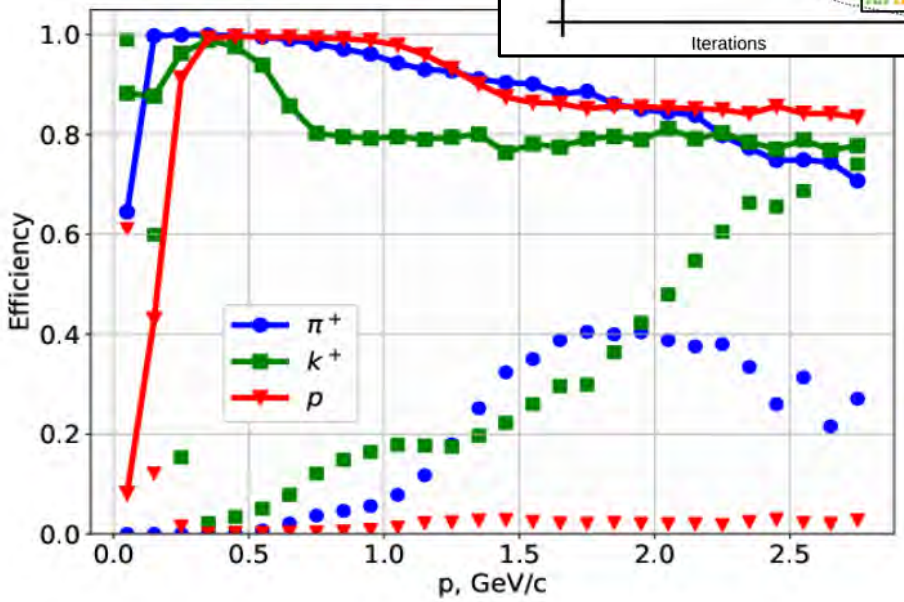
BM@N





MPD





Efficiency and contamination of the identification of positively charged particles in the MPD model data, obtained by the method of decision trees with gradient boosting

Preliminary results: accuracy of about 99% for testing data на (18% of true segments are lost)

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

MATLAB

GEANT4

Maple

Frameworks and tools for ML/DL tasks

jupyterhub

TensorFlow

NumPy

learn

Quantum algorithms, quantum programming and quantum control

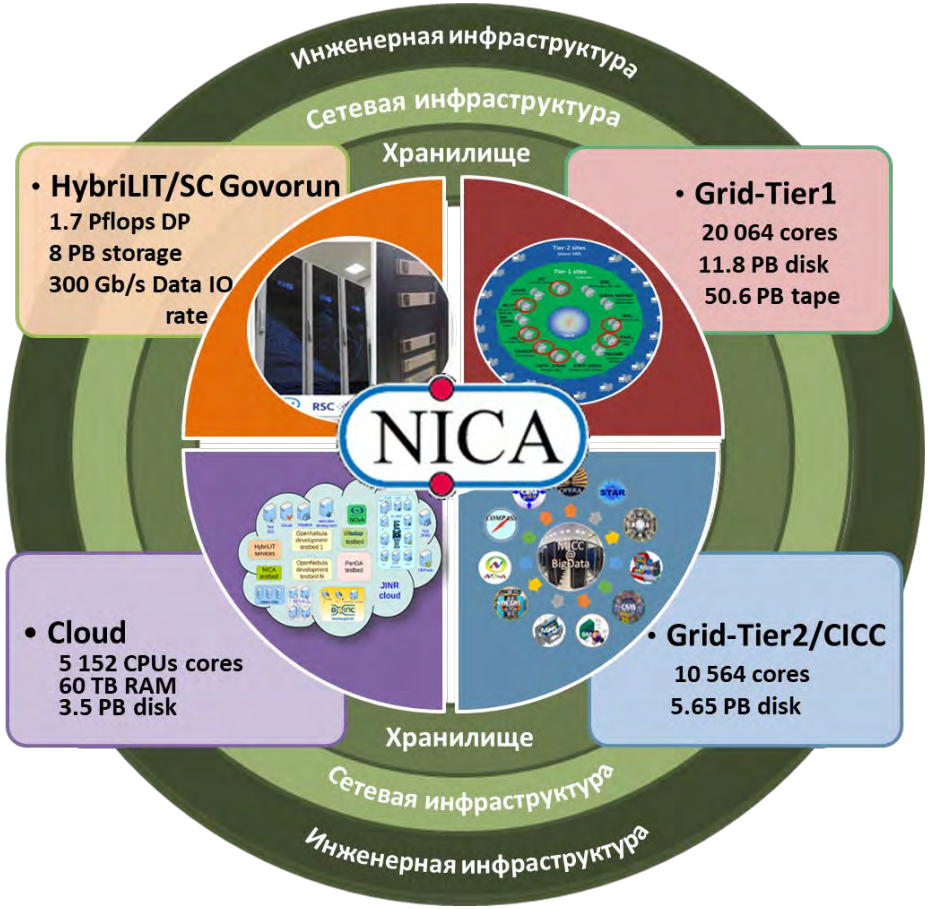
JINR Multifunctional Information and Computing Complex

The Multifunctional Information and Computing Complex (MICC) of MLIT JINR is a key component of the JINR network and information and computing infrastructure, to provide the analysis, processing and storage of data from JINR research programmes.

The MICC uniqueness is ensured by the consolidation of all state-of-the-art information technologies

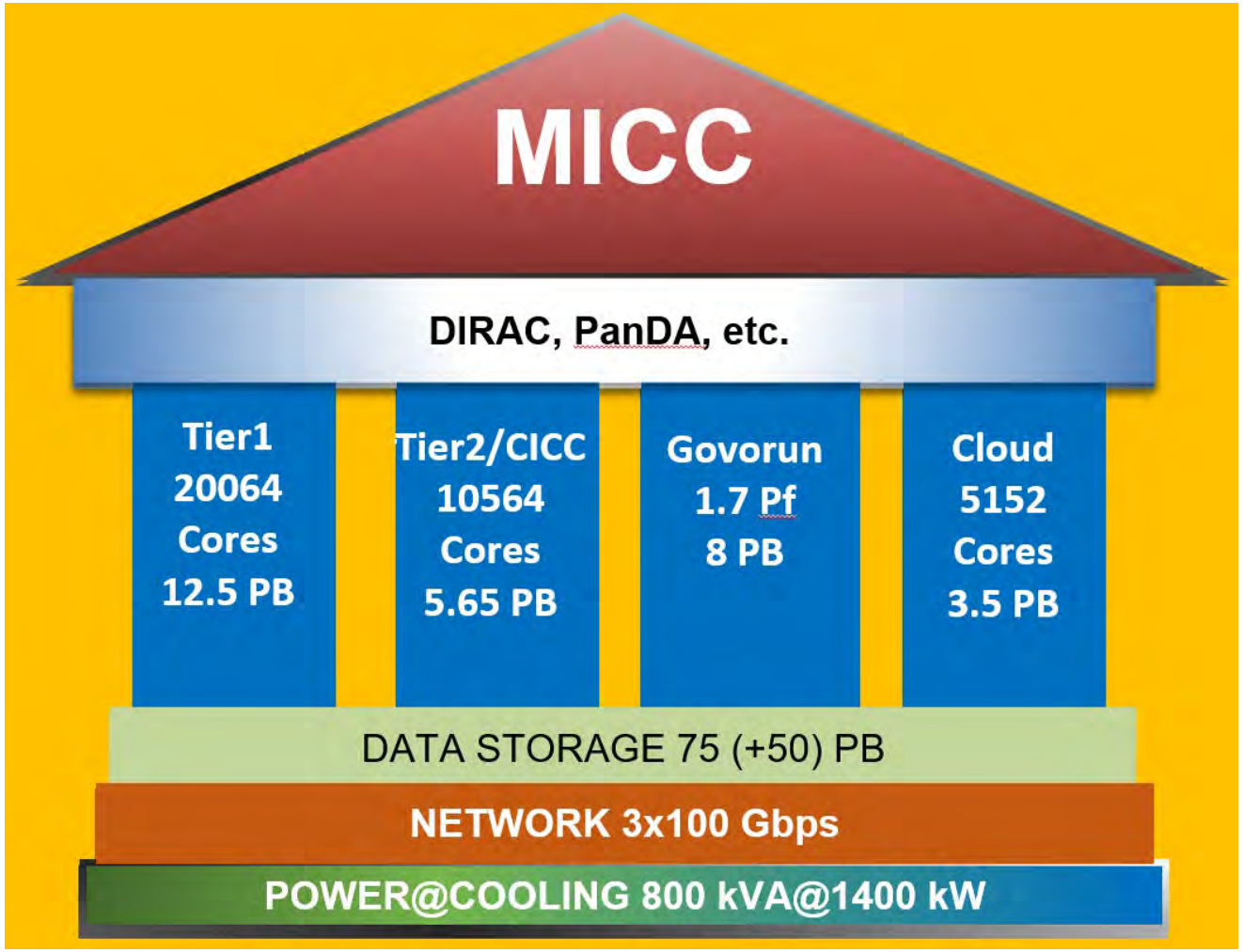
The MICC meets the requirements for a modern highly performant scientific computing complex:

- multi-functionality,
- high performance,
- task-adapted data storage system,
- high reliability and availability,
- information security,
- scalability,
- customized software environment for different user groups,
- high-performance telecommunications and modern local network.



The MICC is on of the five JINR basic scientific infrastructure project.

Multifunctional Information and Computing Complex



4 hardware components

- Tier1 grid site (distributed data processing)
- Tier2/CICC site (distributed data processing)
- Hyperconverged “Govorun” supercomputer
- Cloud infrastructure

Distributed multi-layer data storage system

- Disks
- Robotized tape library

Network

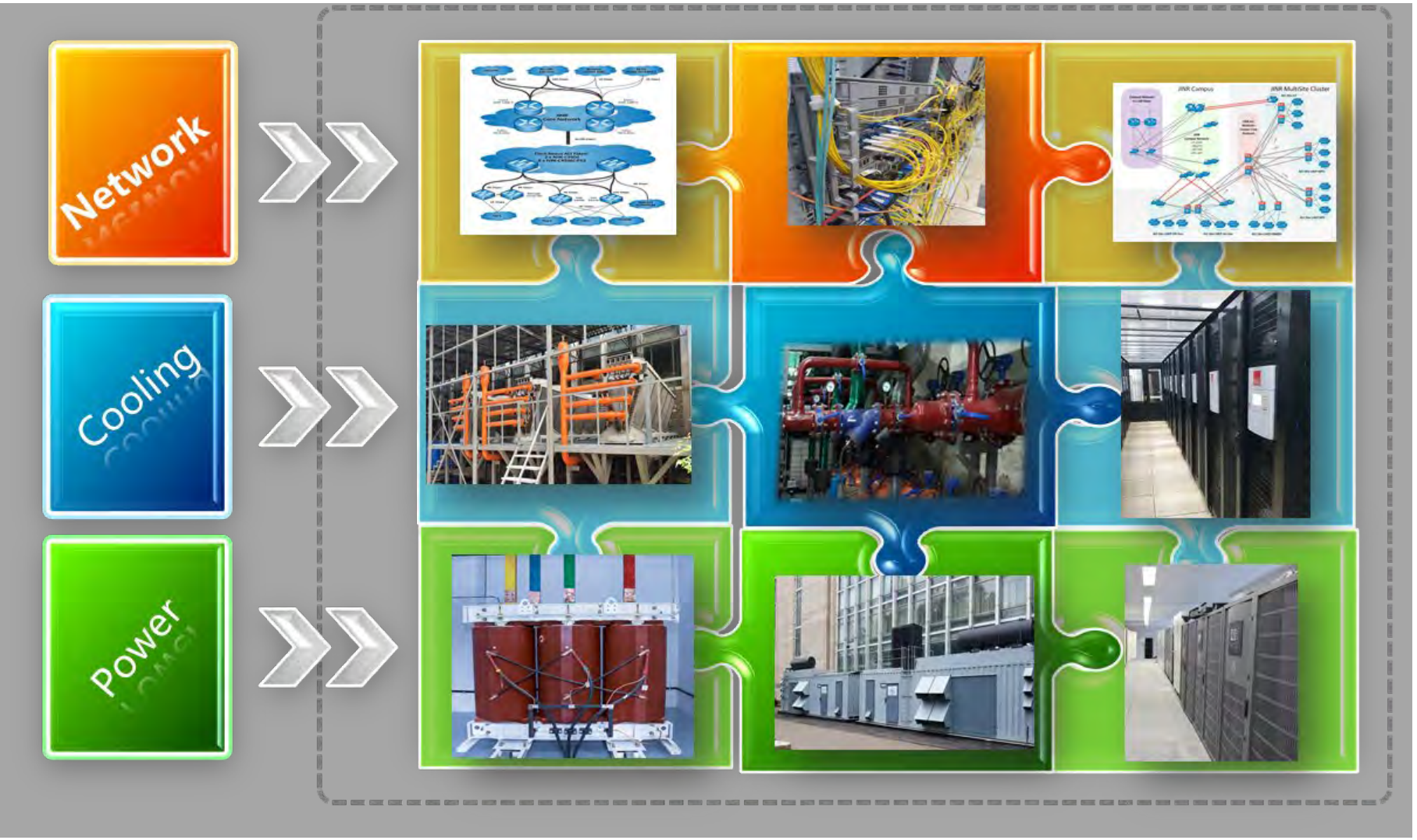
- Wide Area Network
- Local Area Network

Engineering infrastructure

- Power
- Cooling

The main objective of the project is to ensure multifunctionality, scalability, high performance, reliability and availability in 24x7x365 mode for different user groups that carry out scientific studies within the JINR Topical Plan.

Engineering Infrastructure (Power @ Cooling) @ Network



Wide Area Network 3x100 Gbps
Cluster Backbone 4x100 Gbps
Campus Backbone 2x100 Gbps

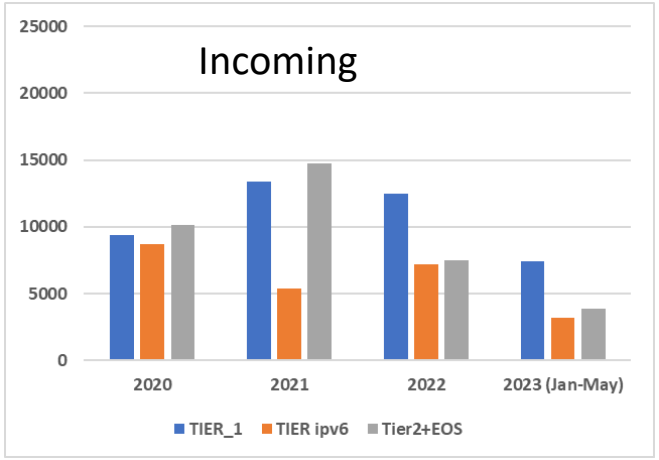
Dry chillers
In-Row systems
Total cooling 1400 kW

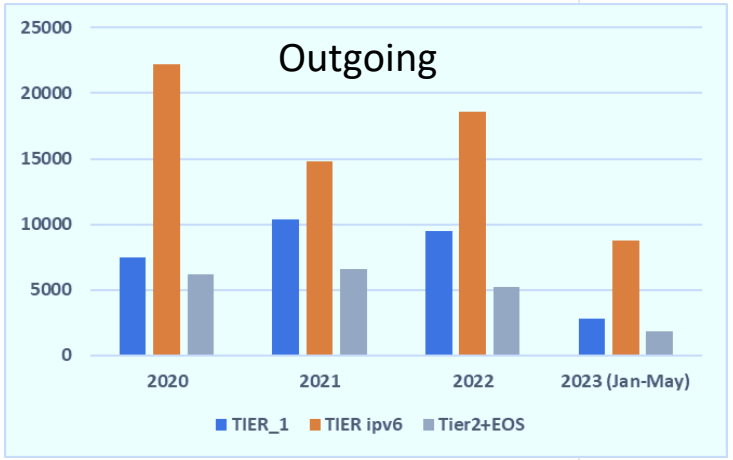
Uninterruptible power supplies (UPS) 8x300 kVA
Diesel-generator units (DGU) 2x1500 kVA
Transformers 2x2500 kVA

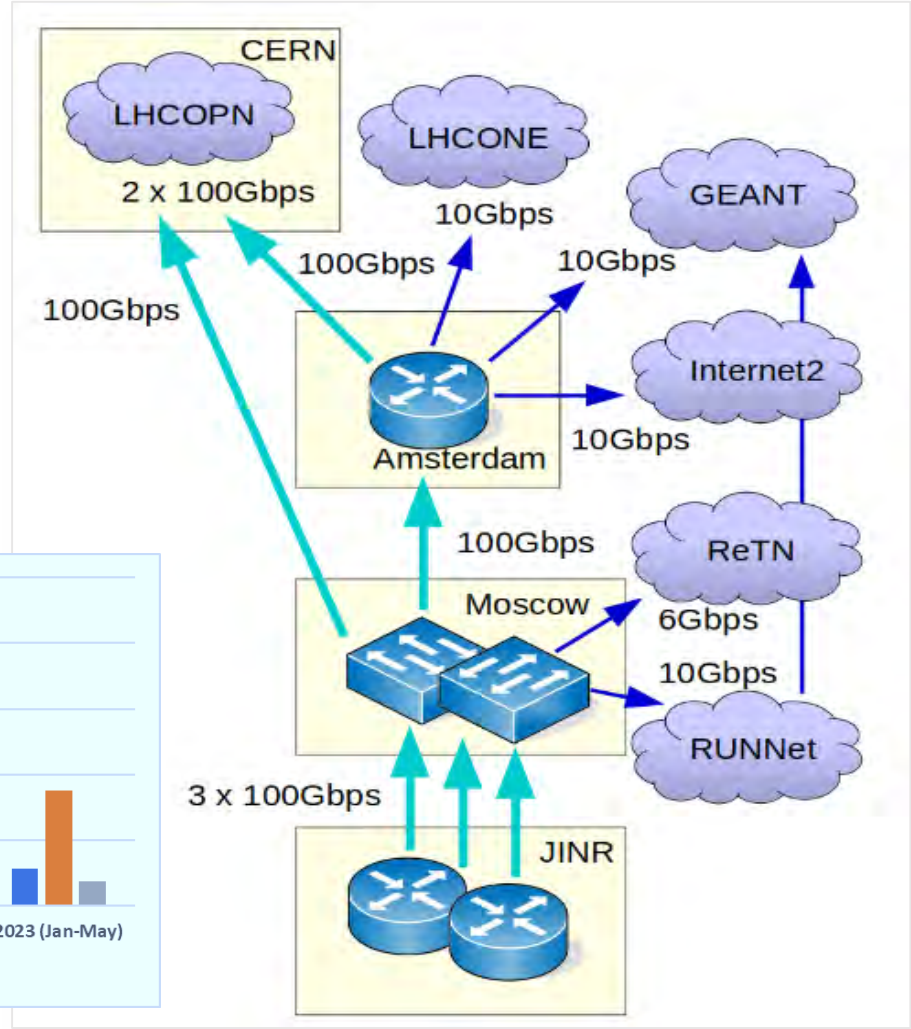
JINR Network Infrastructure

The network infrastructure is a fundamental component of the IT infrastructure of JINR and of the MICC. It provides access to the Internet, computing resources and data storage systems, as well as enables experimental data processing and computing. MLIT ensures the reliable and fault-tolerant operation of all components of the network infrastructure:

- JINR-Moscow 3x100 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:

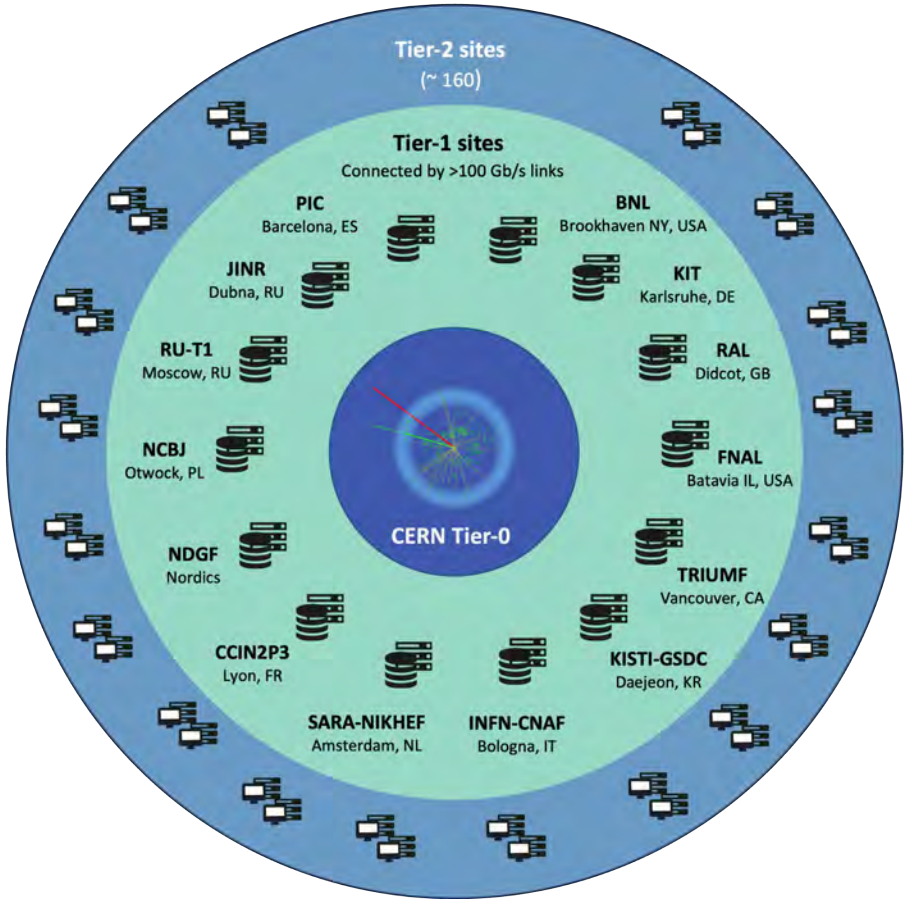
9291 network elements
18044 IP-addresses
6355 users registered within the network
4477 *.jinr.ru service users
1455 digital library users
837 remote VPN
111 EDUROAM users

network traffic in 2022

- 29.56 PB - input
- 34.19 PB - output

The Worldwide LHC Computing Grid

WLCG is an International collaboration to distribute and analyze LHC data. Integrates computer centers worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists



Tier0 (CERN): data recording, reconstruction and distribution

Tier1: permanent storage, re-processing,

Tier2: Simulation, end-user analysis

The mission of the WLCG project is to provide global computing resources to store, distribute and analyze the ~50-70 Petabytes of data expected every year of operations from the Large Hadron Collider.

WLCG computing enabled physicists to announce the discovery of the Higgs Boson.

170 sites

42 countries

> 12k physicists

~1.4 M CPU cores

1.5 EB of storage

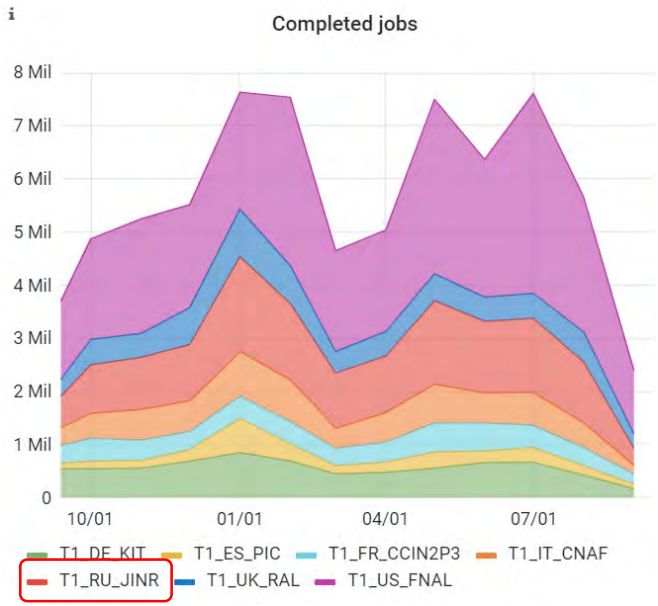
> 2 million jobs/day

100-250 Gb/s links

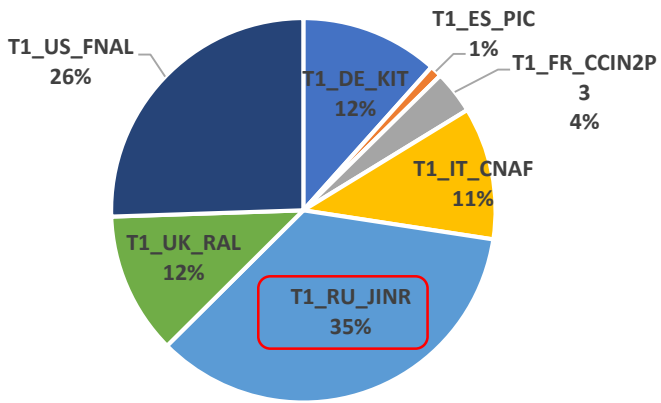
Worldwide LHC Computing Grid - 2019

JINR Tier1 for CMS @LHC and NICA Experiments

Last year



Number of processed events

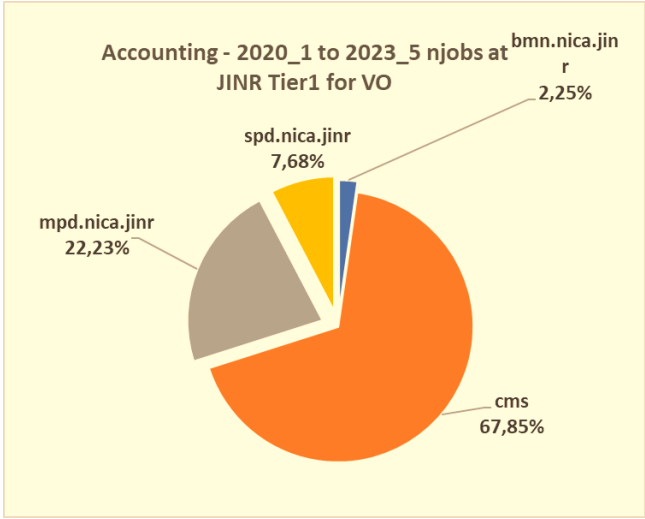


Since the beginning of 2015, a full-scale WLCG Tier1 site for the CMS experiment has been operating at MLIT JINR.

The importance of developing, modernizing and expanding the computing performance and data storage systems of this center is dictated by the research program of the CMS experiment, in which JINR physicists take an active part within the RDMS CMS collaboration.

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

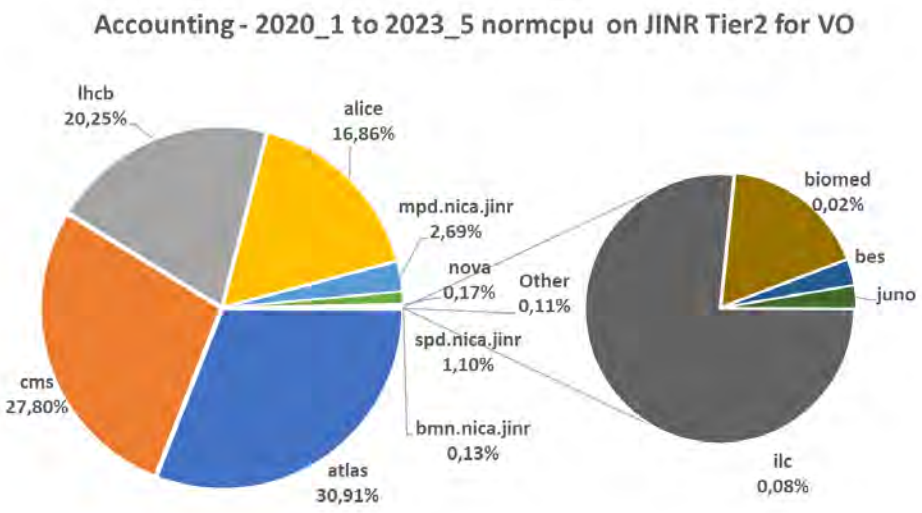
30% of all jobs executed at Tier1 JINR are NICA jobs



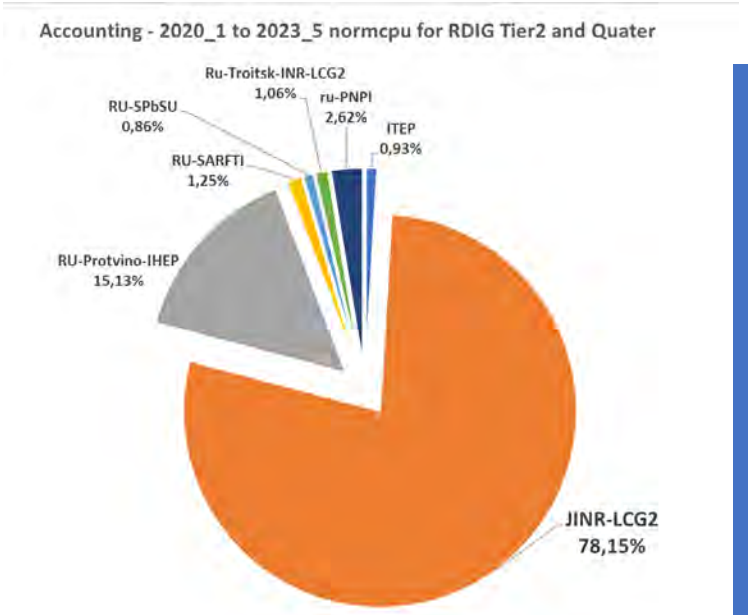
Since 2021, the JINR Tier1 center has demonstrated stable operation not only for CMS (LHC), but also for the NICA experiments.

20096 CPU cores; 360 kHS06
14 PB dCache based disk storage;
50,6 PB Tape storage
100% reliability and availability

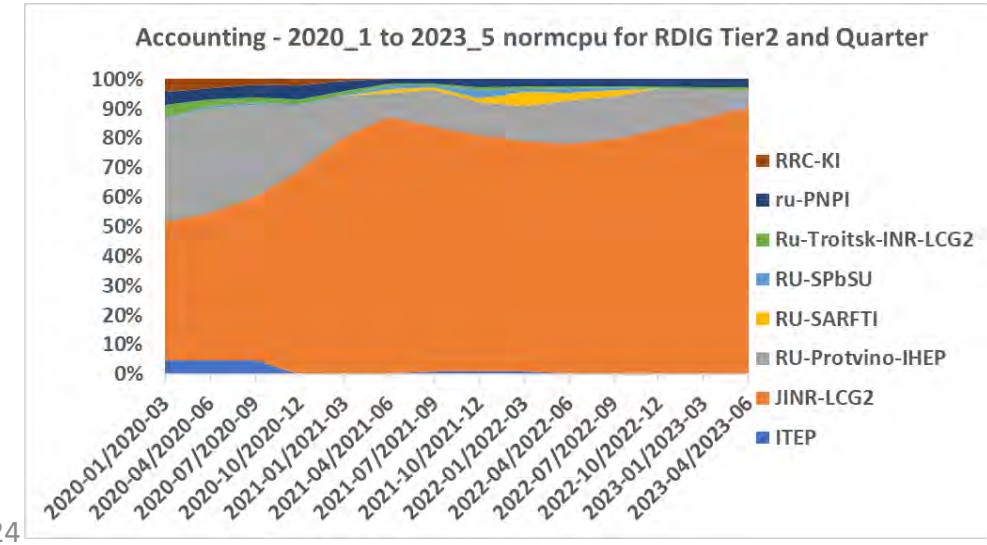
JINR Tier2 in WLCG & RDIG



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, etc.).

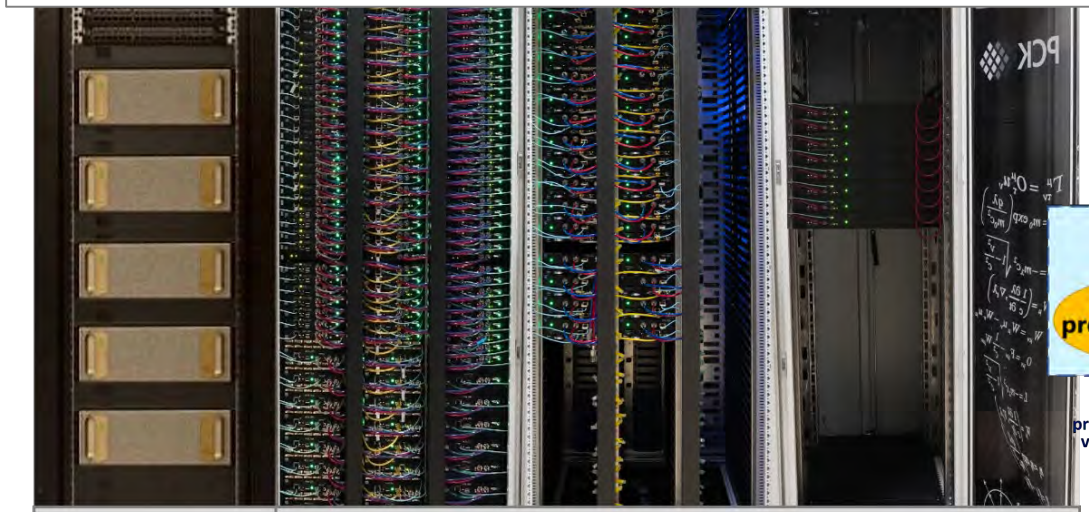


JINR Tier2 is the most productive in the Russian Data Intensive Grid (RDIG) Federation. Almost 80% of the total CPU time in the RDIG is provided by JINR Tier2 site.

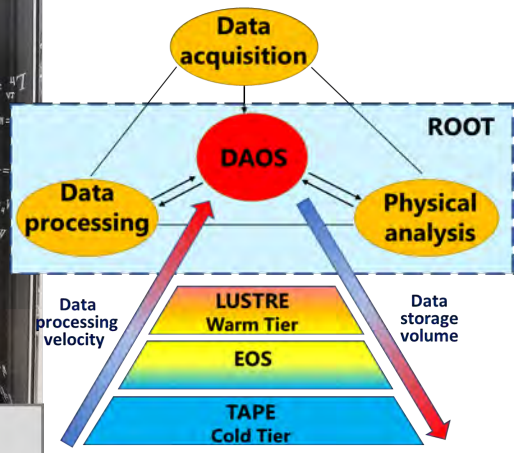


“Govorun” Supercomputer

- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 1.7 PFlops DP
- Total capacity of Hierarchical Storage: 8.6 PB
- GPU component based on NVIDIA
- CPU component based on RSC “Tornado” liquid cooling solutions
- The most energy-efficient center in Russia (PUE = 1.06)
- Data IO rate: 300 Gb/s



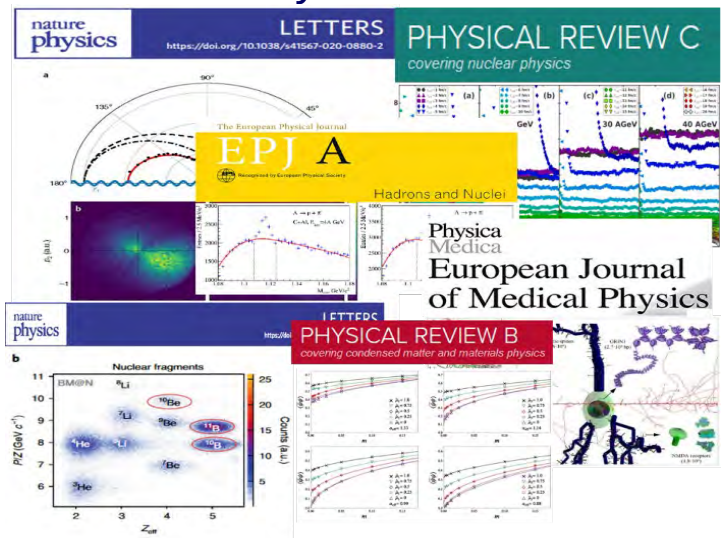
GPU-accelerator | Hyperconverged CPU and Distributed Storage Nodes



The resources of the “Govorun” supercomputer are used by scientific groups from all the Laboratories of the Institute for solving a wide range of tasks in the field of theoretical physics, as well as for physics modeling and experimental data processing.

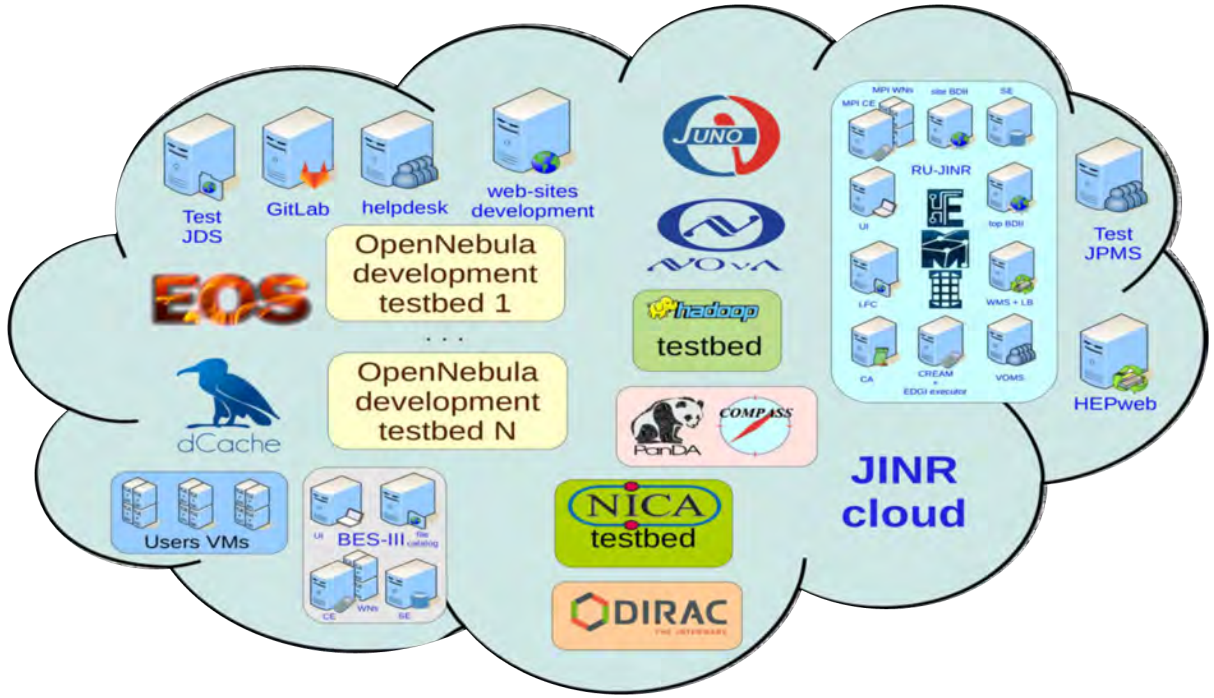
Key projects that use the resources of the SC “Govorun”:

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



Cloud Infrastructure

- Cloud Platform - OpenNebula
- Virtualization - KVM
- Storage (Local disks, Ceph)
- Total Resources
  ~ 5,000 CPU cores; 60 TB RAM; 3.1 PB of raw ceph-based storage

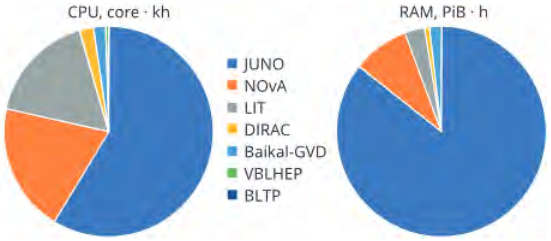


DIRAC-based distributed information and computing environment that integrates the JINR Member State organizations' clouds

- VMs for JINR users
- Computational resources for neutrino experiments
- Testbeds for research and development in IT
- COMPASS production system services
- Data management system of the UNECE ICP Vegetation
- Scientific and engineering computing
- Service for data visualization
- Gitlab and some others

Cloud Infrastructure

- Computational resources for neutrino experiments:
- VMs for JINR users
- Testbeds for research and development in IT
- COMPASS production system services
- Data management system of the UNECE ICP Vegetation
- Service for data visualization, Gitlab and some others

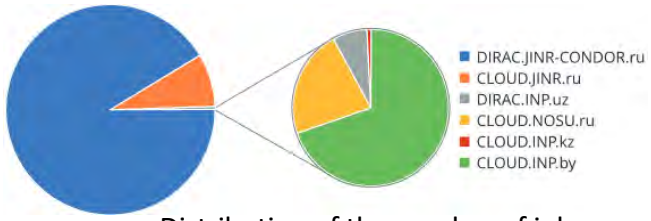


Usage of cloud computing by experiments and JINR subdivisions

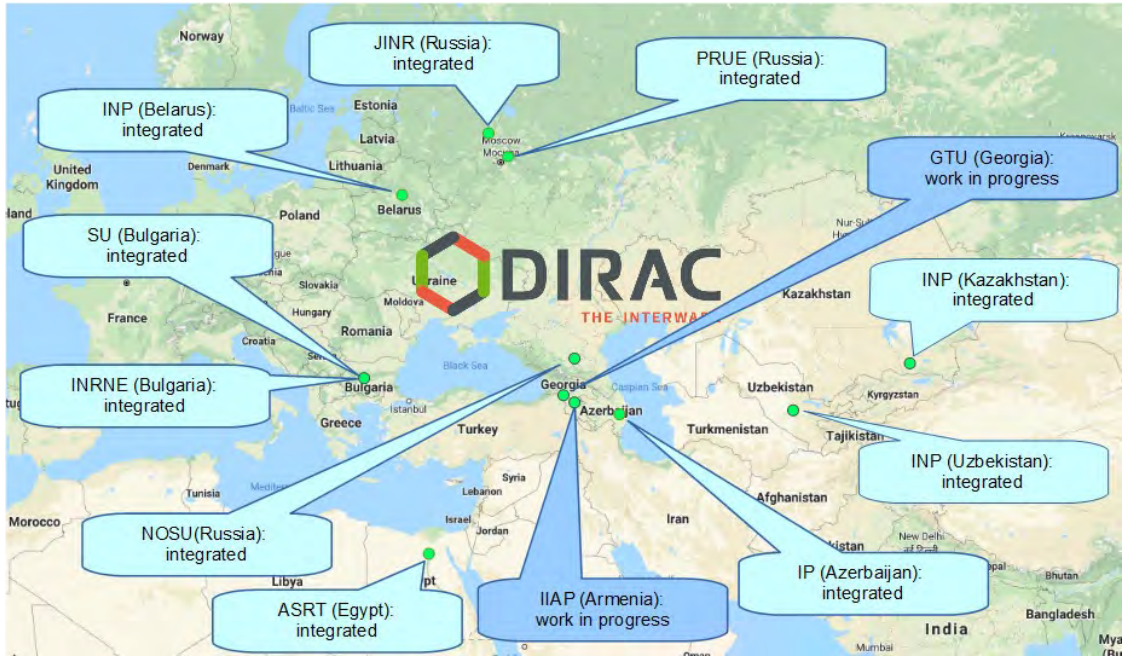
The Baikal-GVD, NOvA and JUNO experiments are the major users of the cloud infrastructure.

Most of the jobs in the JINR DICE are performed on the neutrino computing platform (DIRAC.JINR-CONDOR.ru).

The main consumer of the JINR DICE resources is the Baikal-GVD experiment (96%).

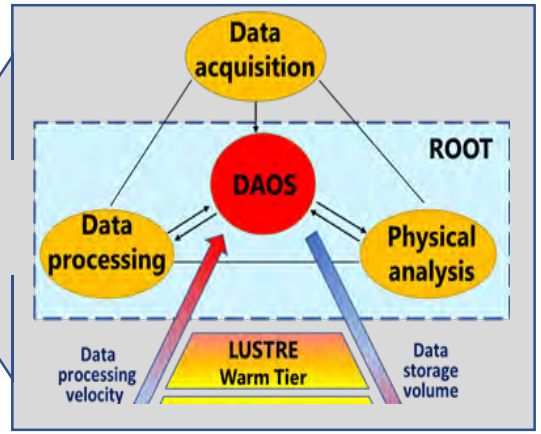


Distribution of the number of jobs completed in the JINR DICE by participants



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

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. At present, 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:

- ✓ very hot data (DAOS (Distributed Asynchronous Object Storage)),
- ✓ the most demanded data (fastest access),
- ✓ hot data,
- ✓ warm data (LUSTRE).

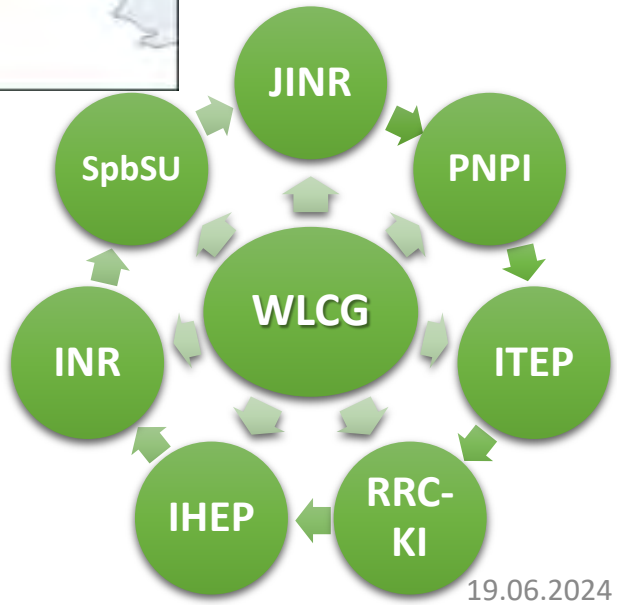
From RDIG to RDIG-M

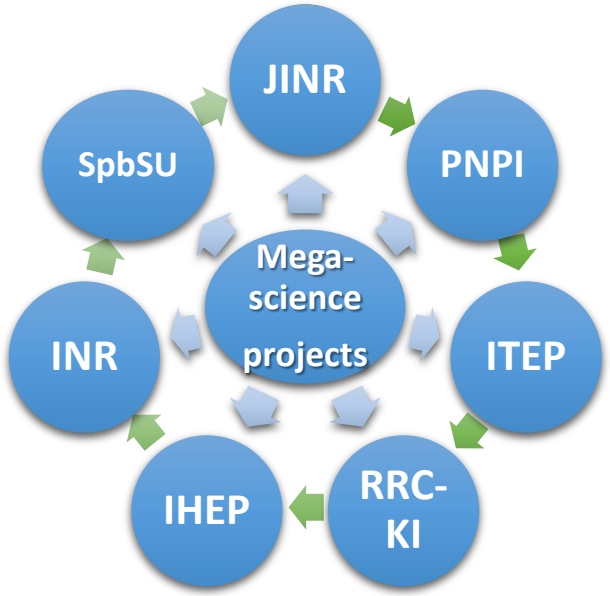
The Russian consortium RDIG (Russian Data Intensive GRID) was set up in September 2003 as a national federation in the EGEE project.

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

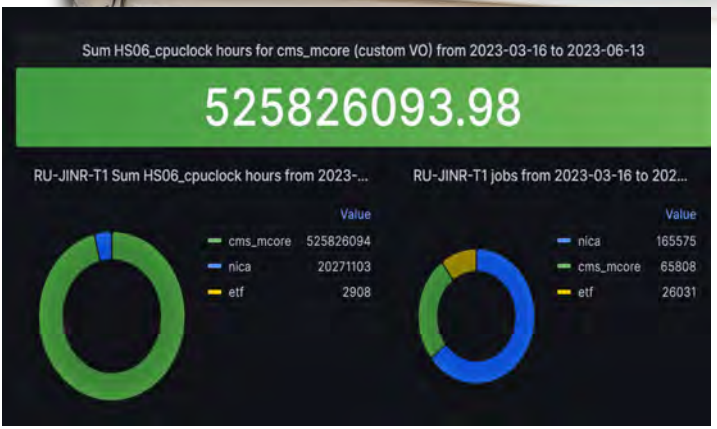
Consortium RDIG-M – Russian Data Intensive GRID for Megascience projects

RDIG
Russian Data Intensive Grid





MICC Monitoring @Accounting



- ❖ 3 monitoring servers
- ❖ About 1800 nodes
- ❖ About 16000 service checks

The successful functioning of the computing complex is ensured by the system that monitors all MICC components. We must

- ➢ expand the monitoring system by integrating local monitoring systems for power supply systems into it (diesel generators, power distribution units, transformers and uninterruptible power supplies);
- ➢ organize the monitoring of the cooling system (cooling towers, pumps, hot and cold water circuits, heat exchangers, chillers);
- ➢ create an engineering infrastructure control center (special information panels for visualizing all statuses of the MICC engineering infrastructure in a single access point);
- ➢ account each user job on each MICC component.

It is required to develop intelligent systems that will enable to detect anomalies in time series on the basis of training samples, which will result in the need to create a special analytical system within the monitoring system to automate the process.

Using JINR computing resources for your research FAQ

What is the recipe of getting access and use JINR's computing resources?

1. Find a theme or project appropriate for your research in the JINR Topical Plan (available online at: http://www.jinr.ru/staff/science_ptp/)
2. Start collaborating in the chosen theme or project with JINR researchers;
3. Conclude a formal agreement with JINR about your institution/research center/university status;
4. Receive one of the statuses of JINR personnel: staff, associated staff, visitor, user(under development);
5. Get access to needed resources for your research.

Can JINR resources be used for personal or commercial purposes?

- NO

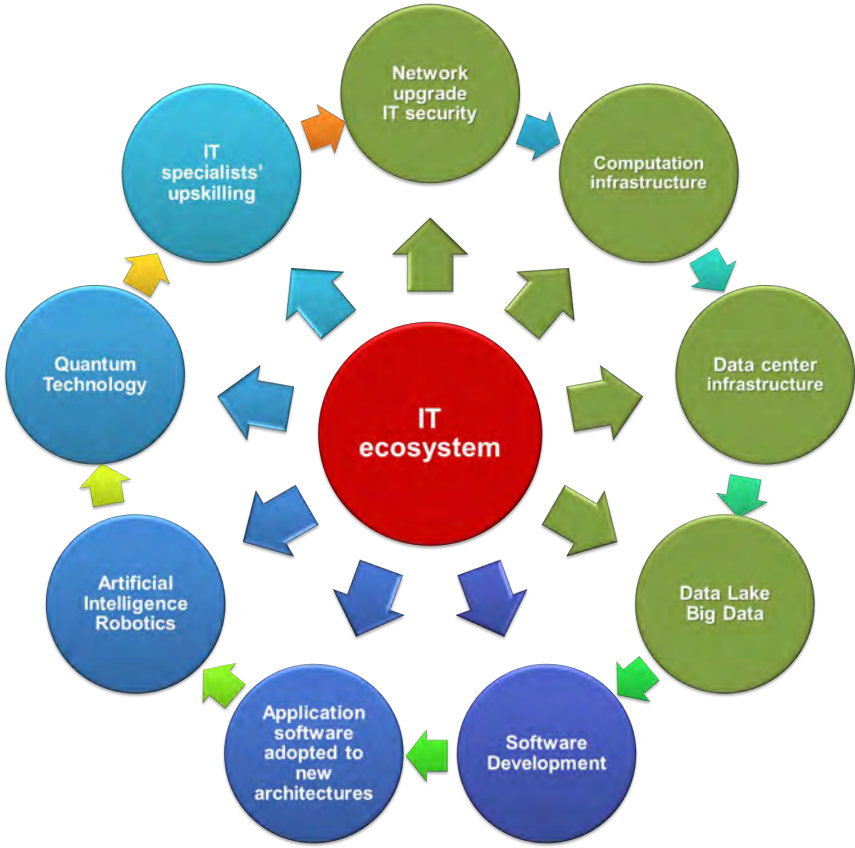
Strategy for Information Technology and Scientific Computing at JINR

Scientists 25 ScD 60 PhD	96
Specialists (engineers, software engineers)	150
Managers	19
Technical workers	50
TOTAL	**315**

For more than 50 years from the date of its formation, the MLIT has come a long way along which it is possible to trace the development of computational and information technologies.

The concept of the development of information technologies and scientific computing aimed at solving the strategic tasks of JINR, in the JINR Seven-Year Plan for 2024-2030, provides for the creation of a Scientific IT Ecosystem.

The IT ecosystem implies introduction and development of a whole range of advanced IT solutions, integrated into a single computing environment that combines a variety of technological concepts and methods.



Strategic long-term plan

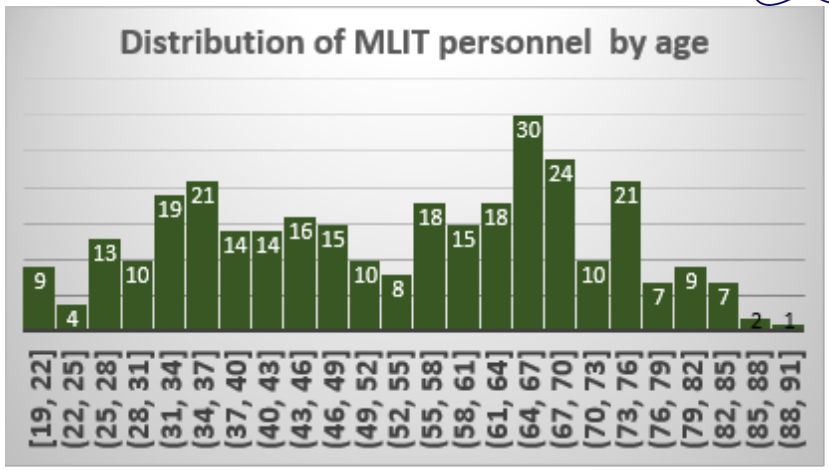
- Dynamically evolving IT-platform, which responds to the rapidly developing IT-world.
- The promising directions of modern information technologies are Artificial Intelligence, Quantum Technologies and Big Data Analytics.
- The development of the scientific IT-ecosystem will depend on novel technologies for acquiring, analysing and sharing data.
- This system must be very flexible and open to new computing methods such as quantum, cognitive calculations, machine learning methods and data mining, as well as to any developments of new algorithmic bases.

Personnel, Scientific and Educational Activities

To implement the research program, MLIT is pursuing policy, which is balancing between hiring a professional and involving young people. Targeted training and human capacity-building as part of research efforts include

- Increase numbers of younger staff (<30)
- Increase number of employees with PhD/Doctoral degrees
- Increase number of publications
- Participation in conferences and schools
- Scientific and Educational Activities (IT-School)
  - retraining and advanced training of personnel
  - organization of trainings, master classes, lectures for young scientists and students
  - involvement of students, PhDs, and young specialists for preparation of qualification works
  - communications with leading educational centers from Russia and from abroad

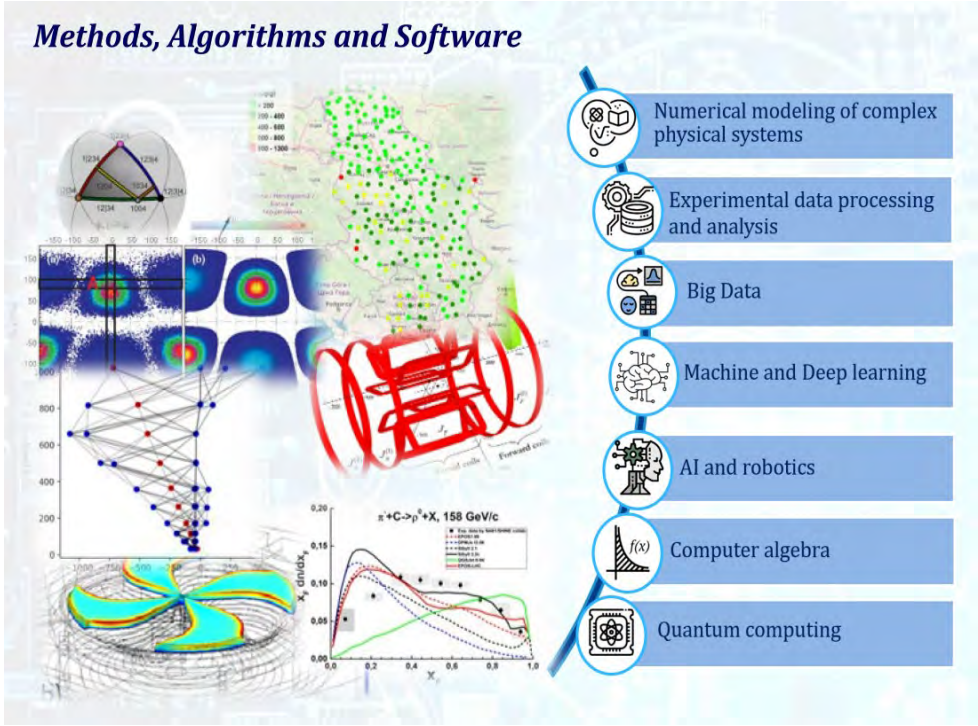
Scientists 25 ScD 60 PhD	96
Specialists (engineers, software engineers)	150
Managers	19
Technical workers	50
TOTAL	**315**



Mathematical, algorithmic and software support for experimental and theoretical researches

The critical challenge is the provision of mathematical, algorithmic and software support for experimental and theoretical research performed at JINR (particle physics experiments, including the NICA megaproject, neutrino experiments, LHC experiments, etc; Factory of Superheavy Elements; irradiation facilities for research in the field of materials science and radiation biology; etc)

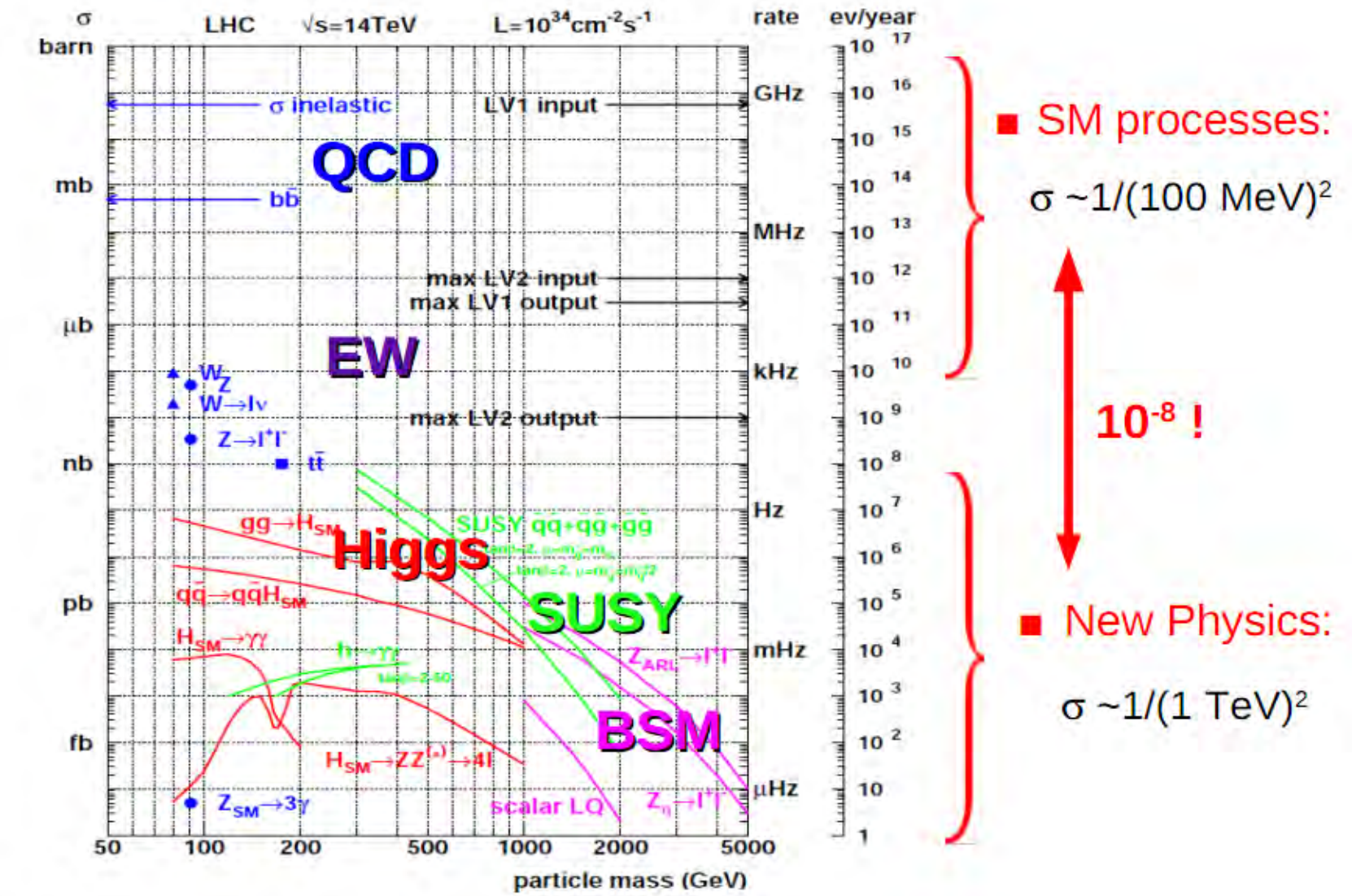
- Simulation of physics processes and experimental facilities
- Development of new-generation algorithms and software systems for data processing and analysis
- New data processing and analysis algorithms for solving applied problems
  - deep and machine learning, including artificial intelligence and artificial intelligence and cognitive intelligent robotics
  - systems of quantum intelligent control, the development of methods of computer algebra and quantum computing
  - Big Data methods



Active participation in the organization and implementation of the processing of experimental data and the analysis of physics information within the priority fields of JINR research, including open data from world experiments (Open Data)

- distributed computing
- comprehensive MDL data analysis with SC «Govorun» (ex. Colliders & Astrophysics)

Physics Processes at LHC



- SM processes:

$\sigma \sim 1/(100\ \text{MeV})^2$

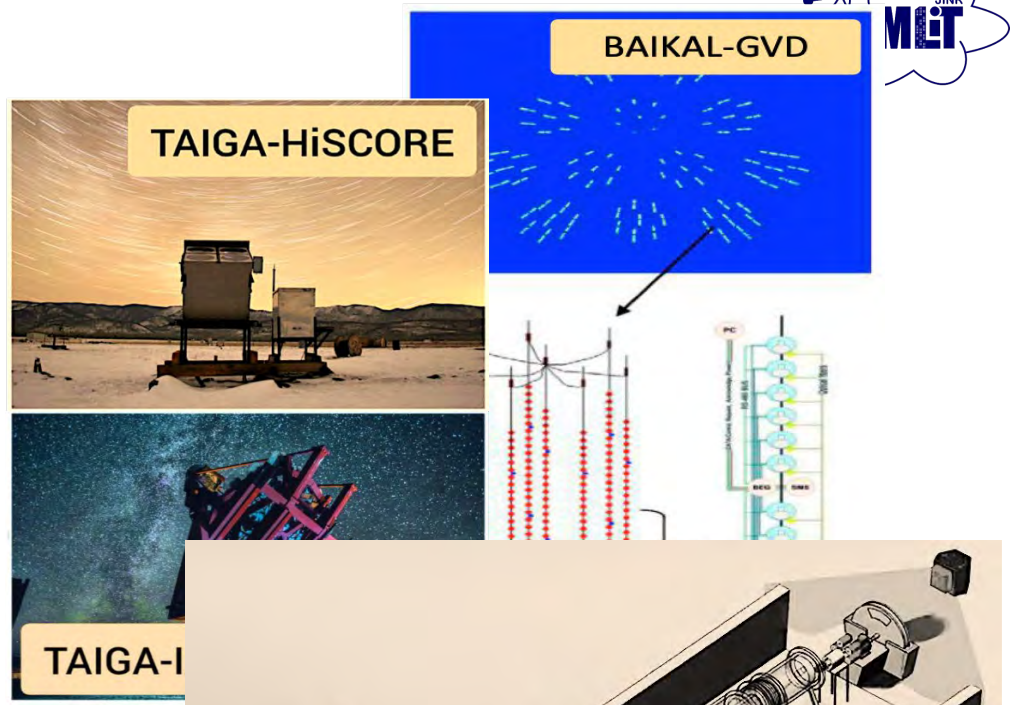
10^{-8} !

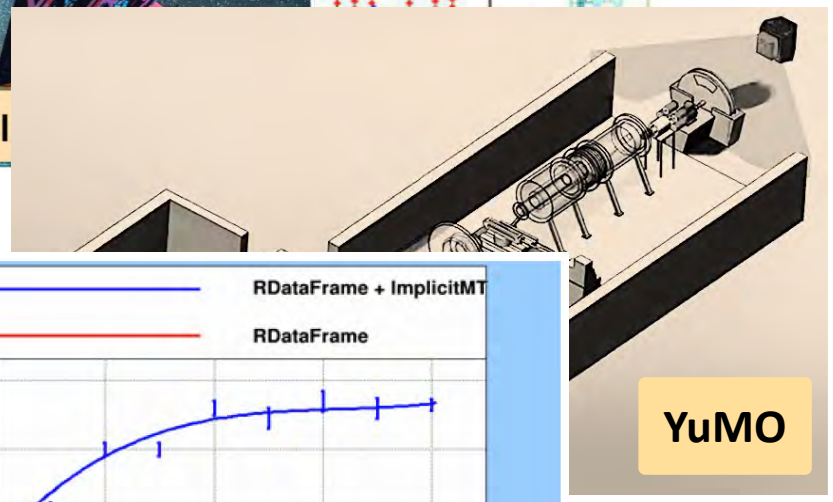
- New Physics:

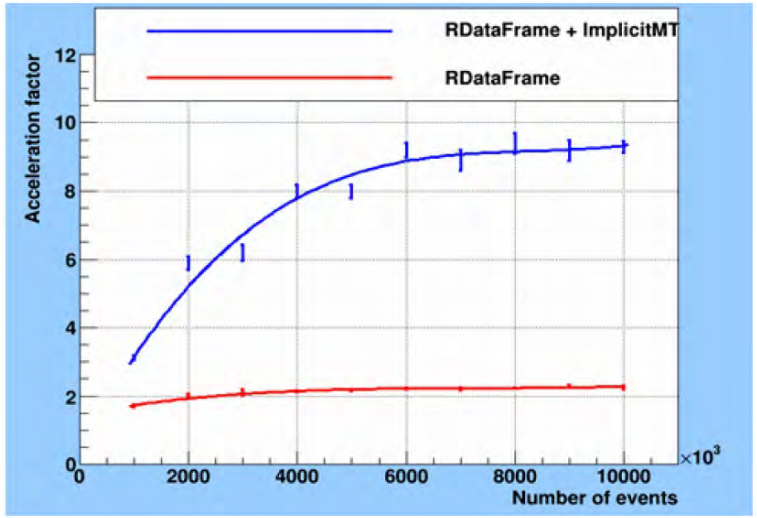
$\sigma \sim 1/(1\ \text{TeV})^2$

… and other fields

- reconstruction of cosmic rays at the TAIGA gamma-ray observatory
  - several atmospheric Cherenkov telescopes (IACT) with a network of TAIGA-HiSCORE wide-aperture Cherenkov optical detectors
- data processing of the Baikal Neutrino Telescope
  - estimate of the positions of the underwater detector components
  - substantial acceleration of the main code of the Cascade Recovery Collaboration, which is important for tracking transient astrophysical neutrino sources
- software complex for data processing of the YuMO small-angle neutron scattering spectrometer at IBR-2
  - data fitting (increase the speed of the data processing with a factor up to 5.5)
  - parallel ROOT (PROOT) tools, also ca be used for experimental data analysis the Baikal-GVD
- processing and analysis of neutron noise of the IBR-2M reactor
- track reconstruction in the proton digital calorimeter for proton therapy



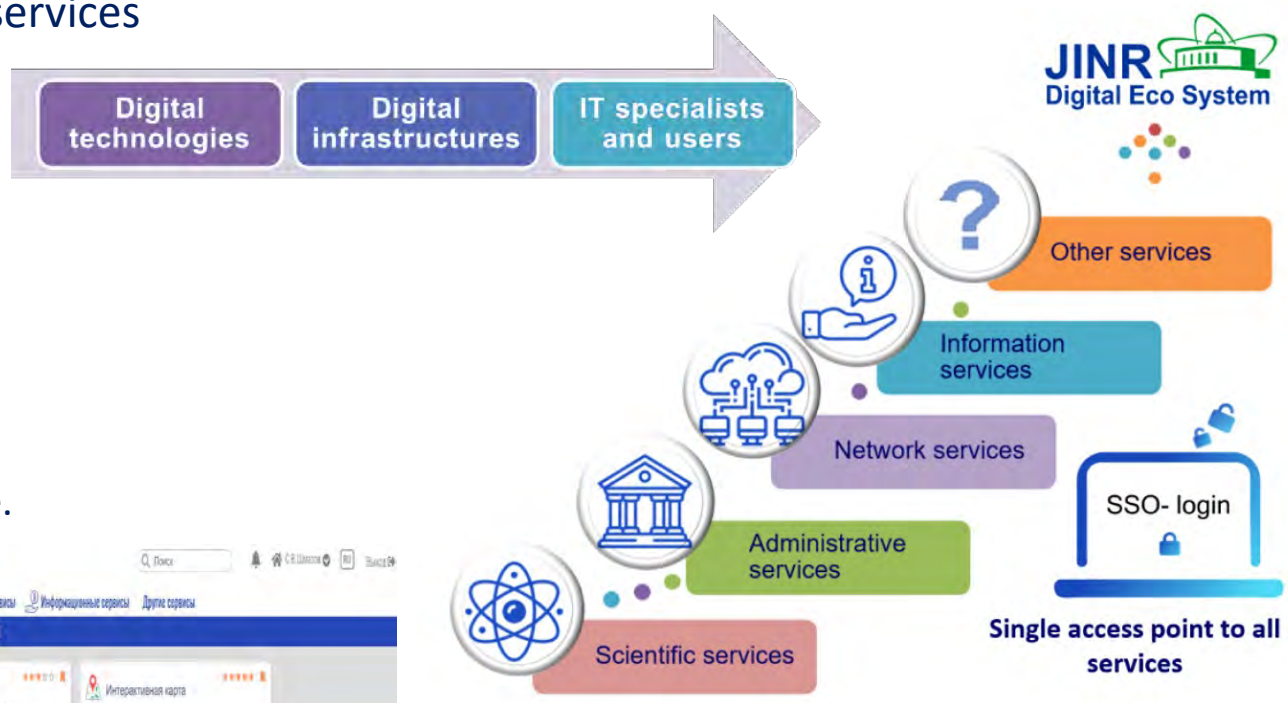


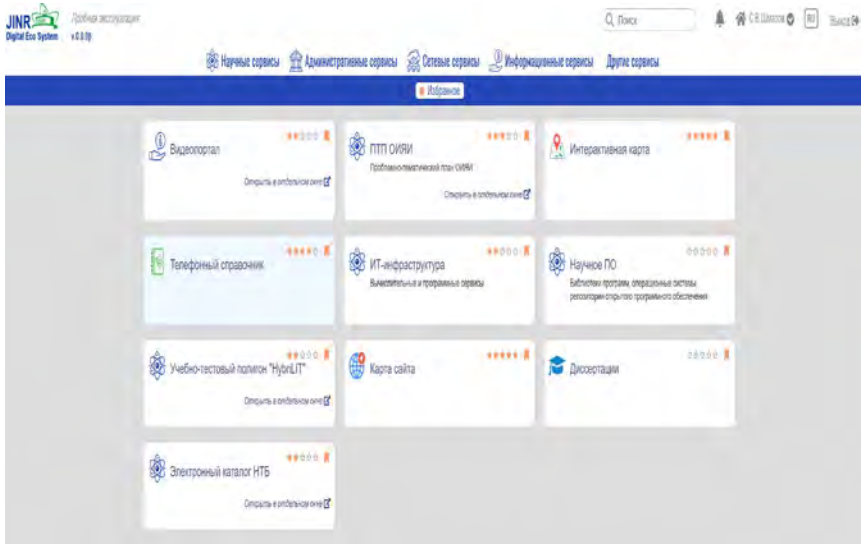


JINR Digital EcoSystem

Creation and development of the JINR-wide digital platform “JINR Digital Ecosystem” are uppermost in the priorities of JINR Seven-Year Plan. This platform integrates existing and future services

- to support
  - scientific, administrative and social activities,
  - maintenance of the engineering and IT infrastructures
- to provide
  - reliable and secure access to various types of data
- to enable
  - a comprehensive analysis of information
- using
  - modern Big Data technologies and artificial intelligence.





- User registration (Single Sign-On, SSO)
- Personal account of a JINR employee
- Notifications in a personal account
- Responsive interface, customizable by the user
- Easy access, convenient navigation and search for information on a large-scale network of a wide variety of JINR services

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

MATLAB

GEANT4

Maple

Frameworks and tools for ML/DL tasks

jupyterhub

TensorFlow

NumPy

learn

Quantum algorithms, quantum programming and quantum control

10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)

- Distributed Computing Systems
- HPC
- Cloud Technologies
- Distributed Storage Systems
- Distributed Computing and HPC Applications in science, education, industry and business, open data.
- Computing for MegaScience Projects
- Quantum informatics and computing
- Big Data, Machine Learning and Artificial Intelligence

MMCP

MATHEMATICAL MODELING AND COMPUTATIONAL PHYSICS

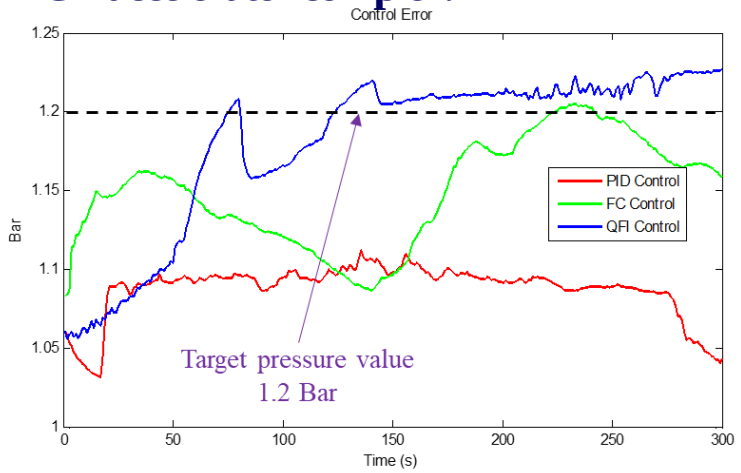
- methods, software and program packages for data processing and analysis;
- mathematical methods and tools for modeling complex physical and technical systems, computational biochemistry and bioinformatics;
- methods of computer algebra, quantum computing and quantum information processing;
- machine learning and big data analytics;
- algorithms for parallel and hybrid calculations.

JINR School of Information Technology

Quantum Intelligent Control of Technological Processes and Physical Facilities at JINR

1. Development of built-in quantum self-organizing controllers for systems of the intelligent control of technological processes, devices and facilities at JINR (including for unforeseen and unpredictable situations).
2. Development of an artificial intelligence platform based on quantum optimization for the tasks of intelligent cognitive robotics and quantum intelligent control in JINR projects.
3. Development of quantum software engineering methods for quantum deep learning based on quantum algorithms, quantum programming, quantum genetic algorithms and quantum soft computing.

A software and hardware platform has been developed on the basis of quantum fuzzy controllers embedded into the control loop to the control of the pressure and flow of liquid nitrogen of the superconducting magnets of the cryogenic system of the NICA accelerator complex.

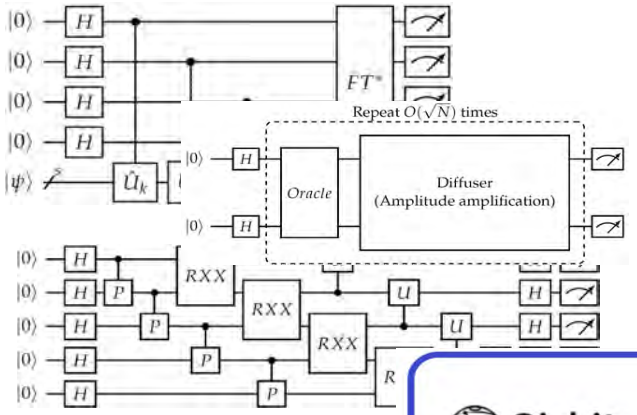


The quantum controller demonstrated the highest speed in achieving the target value, low overshoot and accuracy of achieving the control goal compared to other types of controllers.

Butenko A.V., Zrelov P.V., Korenkov V.V., Kostromin S. A., Nikiforov D.N., Reshetnikov A.G., Semashko S.V., Trubnikov G.V., Ulyanov S.V. Intelligent system for remote control of liquid nitrogen pressure and flow in the cryogenic system of superconducting magnets: hardware and software platform // PEPAN Letters (accepted for publication)

Quantum Computing and Quantum Algorithms

Software quantum simulators for computing on computers of a classical architecture using CPUs and GPUs is of particular interest for solving a number of problems in condensed matter, high-energy physics, quantum chemistry, AI, etc.



Quantum algorithms

Quantum simulators

Qiskit

QuEST

qsim

CUQUANTUM

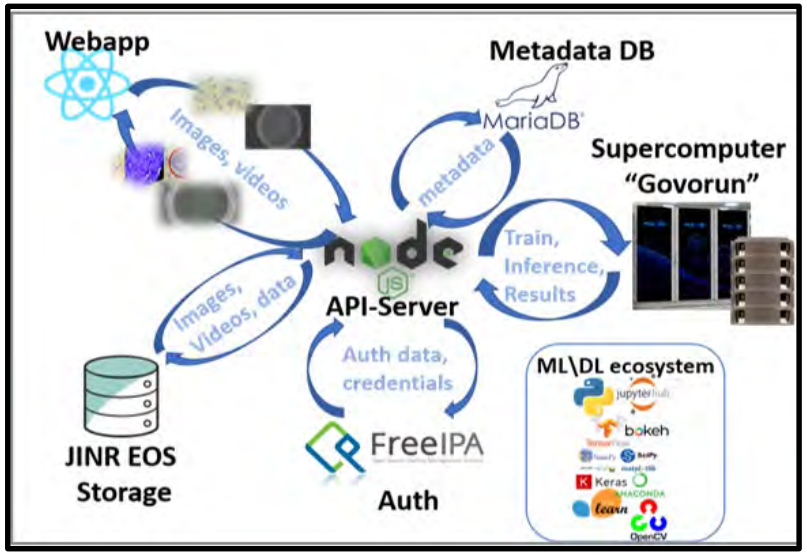
SC “Govorun”

TASKS

- Form a list of QAs required to solve tasks within the studied physical models
- Select the type of quantum simulator to simulate a classical architecture on computers
- Define resources for the selected quantum-limiting capabilities of available computing simulators (number of qubits and computation time)
- Search for exact solutions to urgent problems of quantum chemistry and study the chemical properties of heavy elements

Radiobiology and Life Science

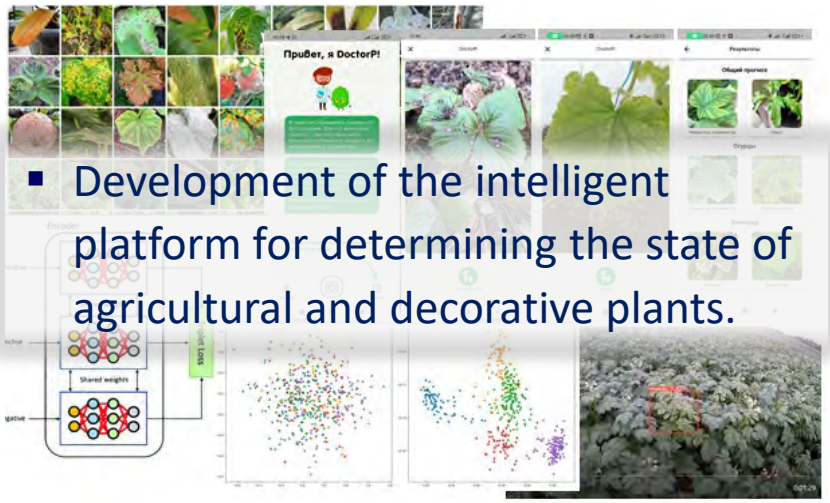
Application of accumulated experience to various classes of problems!



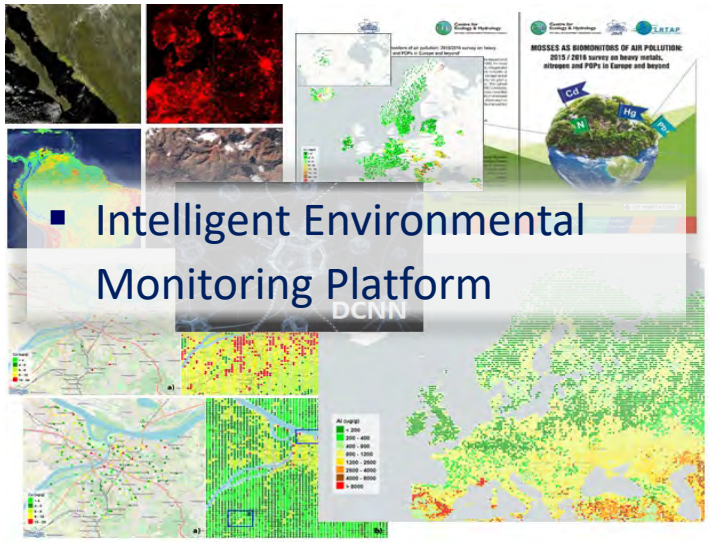
The ML/DL/HPC ecosystem

- top of ML/DL technologies
- modern IT solutions for data storage, processing and visualization
- statistical analysis

- Information System (IS) for Radiation Biology Tasks
  - to store, quickly access and process data from experiments at LRB
  - using a stack of neural network and classical algorithms of computer vision, providing a wide range of possibilities for automating routine tasks.



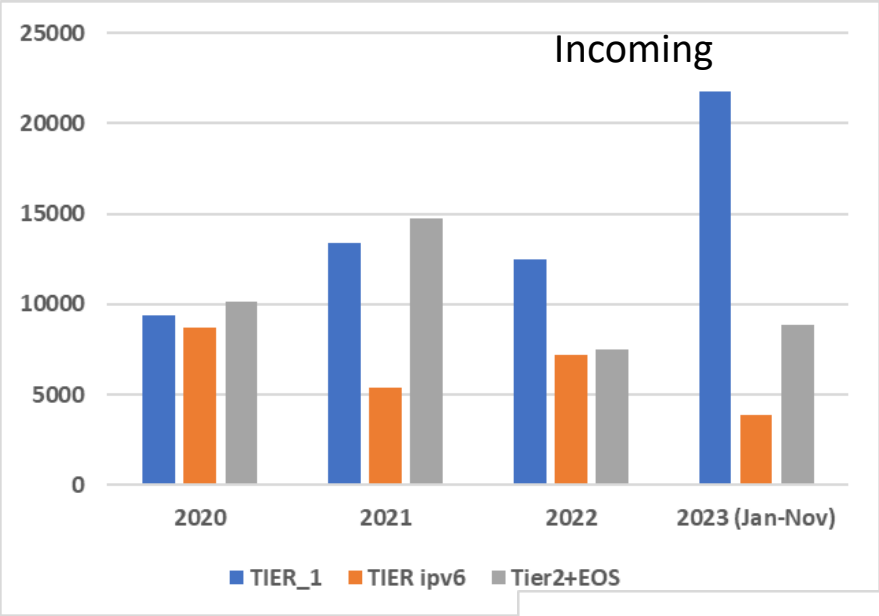
- Development of the intelligent platform for determining the state of agricultural and decorative plants.

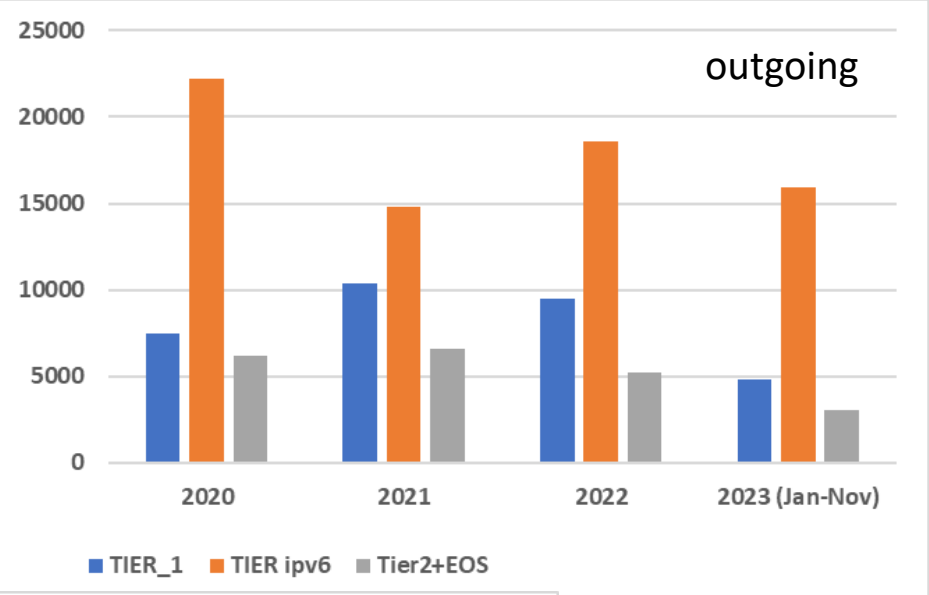


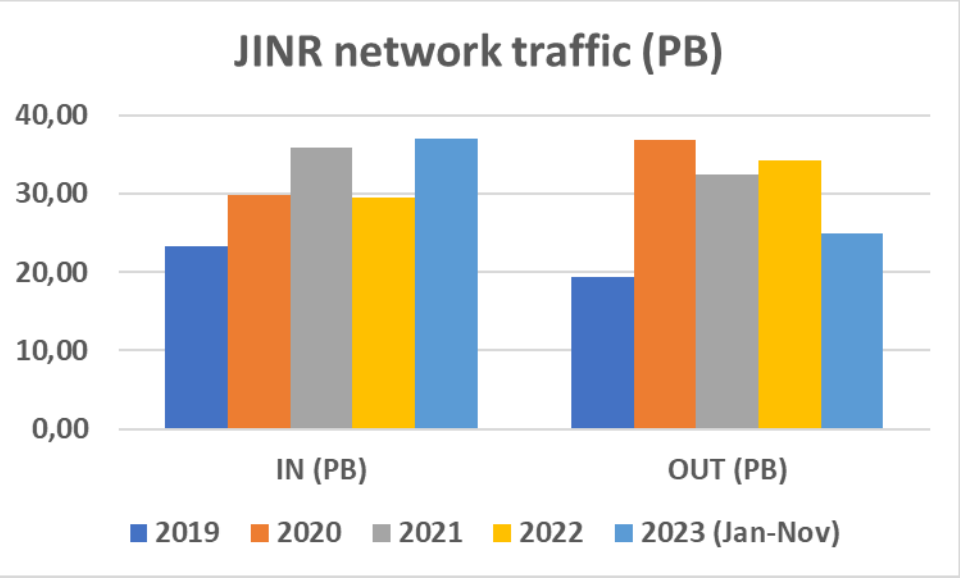
- Intelligent Environmental Monitoring Platform

Networking @ Traffic

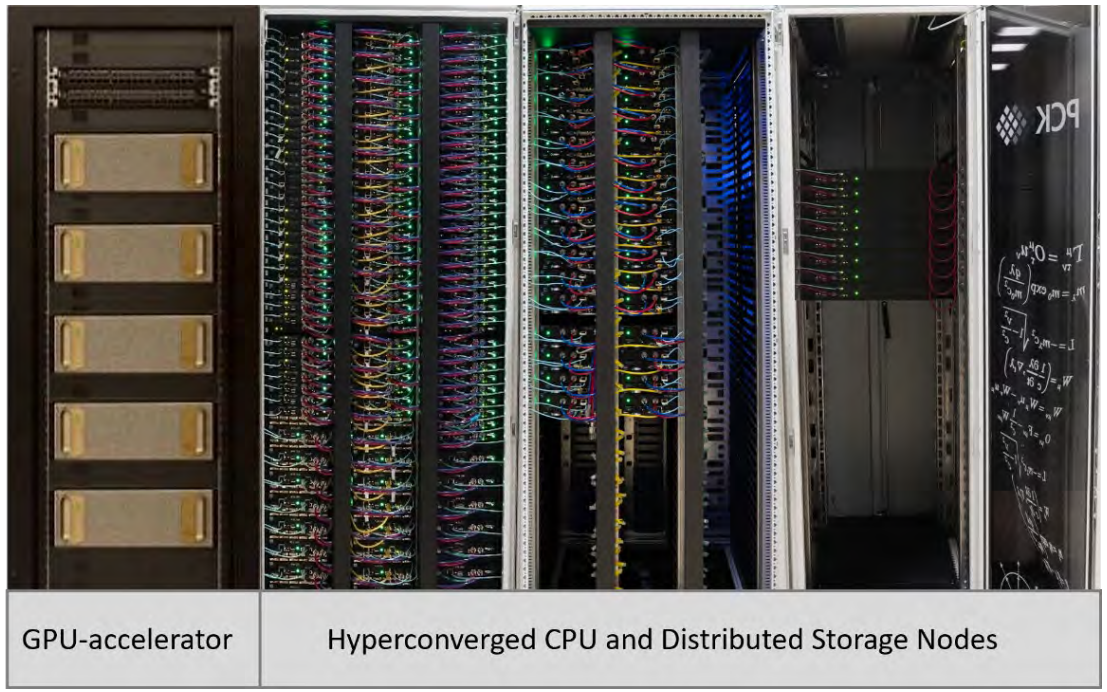
Distribution of the incoming and outgoing traffics by the JINR MICC in 2020-2023 (TB)







"Govorun" supercomputer modernization in 2022 - 2023



GPU-accelerator | Hyperconverged CPU and Distributed Storage Nodes

\+ 40 NVIDIA A100 GPU accelerators

Performance: + 600 Tflops DP

+32 hyperconverged compute nodes

+2 432 new computational cores

Performance: +239 Tflops DP

"New cores"/"old cores" performance increase by more than 1.5 times

+8 distributed storage nodes

Lustre, EOS increase: +8 PB

DAOS increase: +1.6 PB

+0.4 PB for MPD mass production storages integrated into the DIRAC File Catalog

+1 PB for the MPD EOS storage

\+

5 servers with 8 NVidia A100 GPUs in each

\+

Computation field: +32 hyperconverged compute nodes

\+

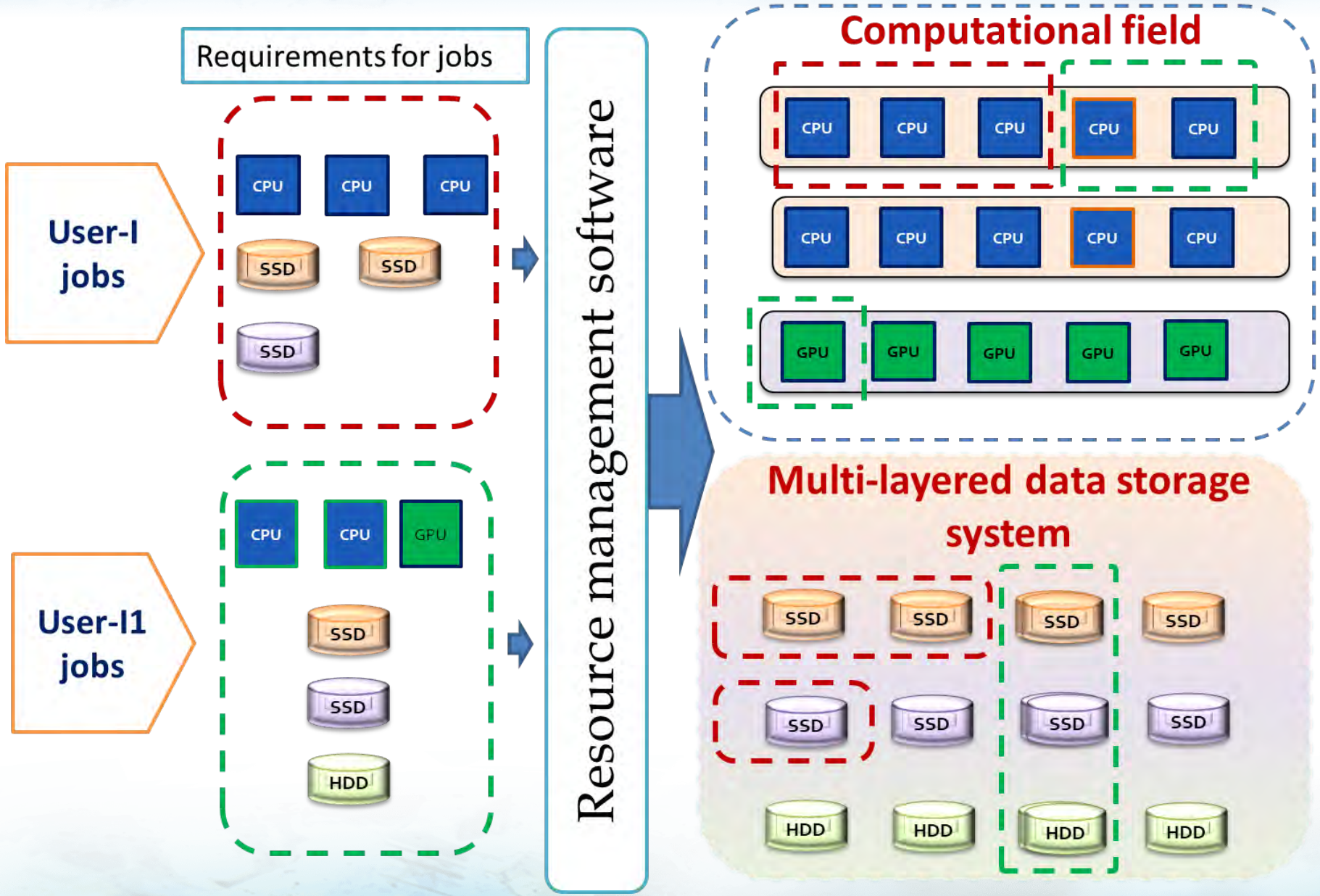
Hierarchical Storage: +8 distributed storage nodes

"Govorun" SC total peak performance: 1.7 PFlops DP

Total capacity of Hierarchical Storage: 8.6 PB

Data IO rate: 300 Gb/s

Orchestration and hyperconvergence on the SC “Govorun”



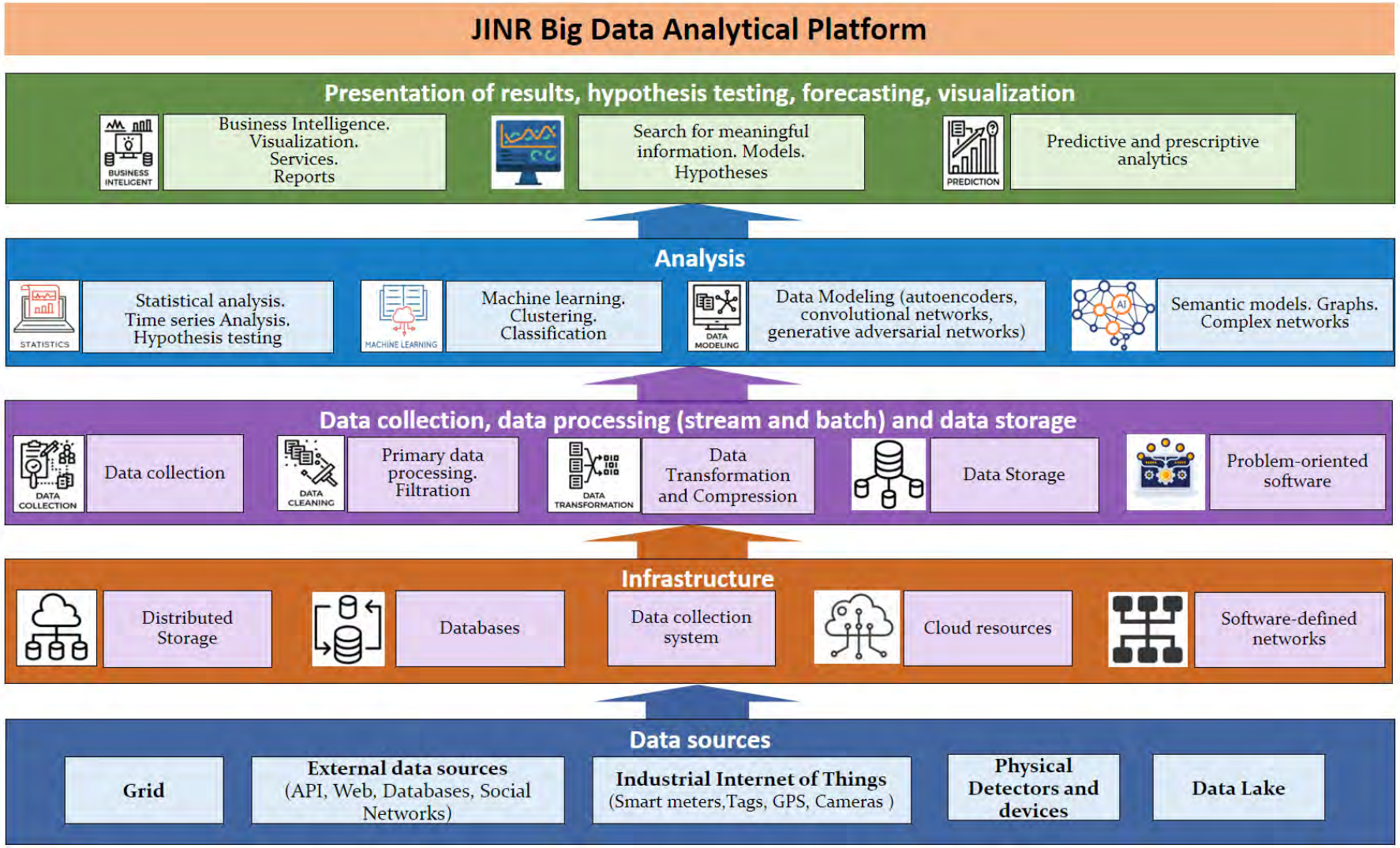
The SC “Govorun” has unique properties for the flexibility of customizing the user’s job.

For his job the user can allocate the required number and type of computing nodes and the required volume and type of data storage systems.

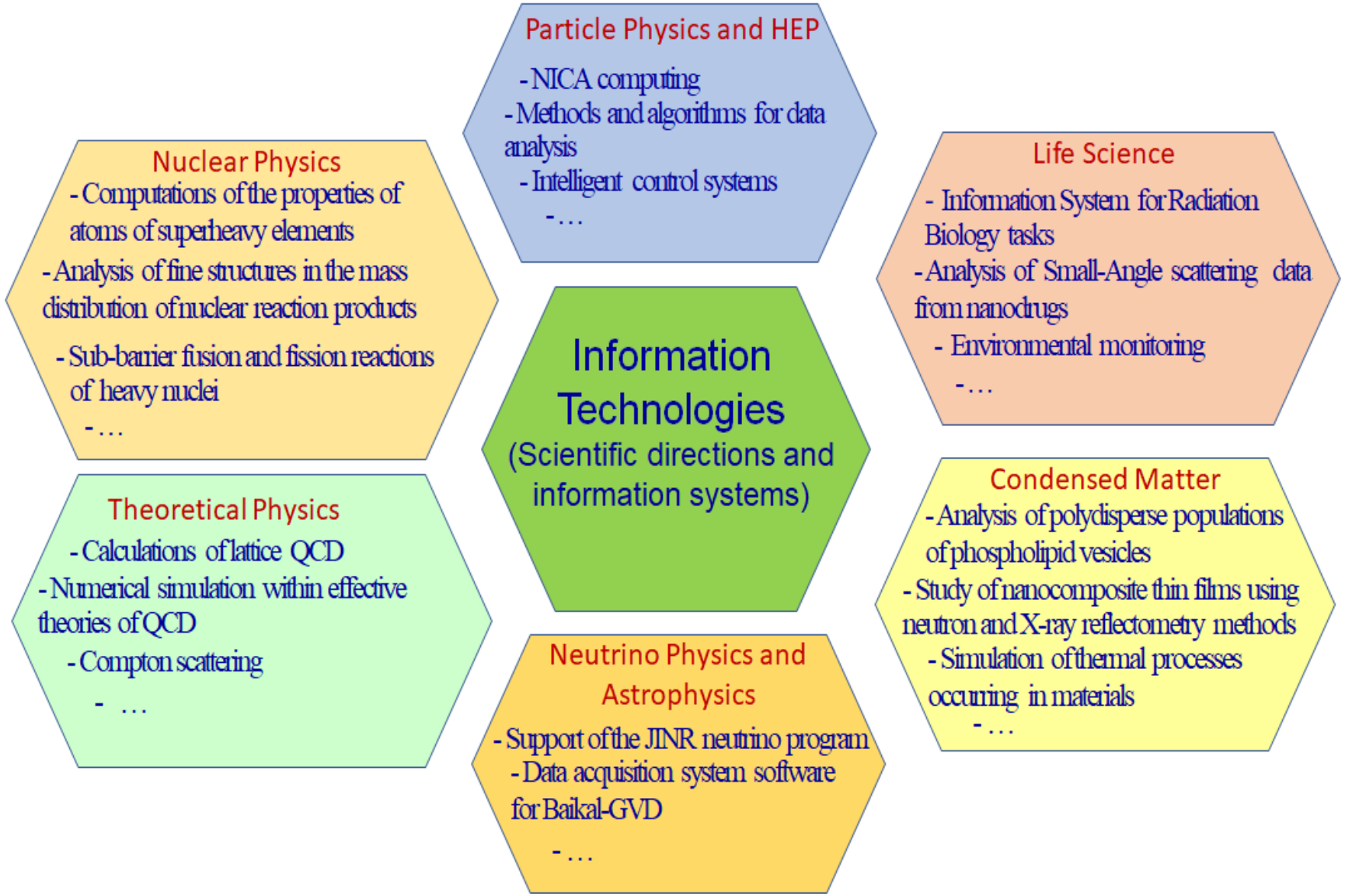
This property enables the effective solution of different tasks, which makes the SC “Govorun” a unique tool for research underway at JINR.

Methods of Artificial Intelligence and Big Data Analytics

- Bringing best of Big Data approaches to JINR practices
- Providing the Big Data infrastructure for users



Cooperation with All JINR Laboratories



Collaboration

Collaborations with 19 research centers and universities from 9 countries (4 members, 1 associate, 4 non-members), Collaborations @ NICA, CERN, etc					
	AANL, Yerevan		GSU, Gomel INP BSU, Minsk		Oxford Un., GB
	CEA, Saclay		INFN, Genova		UTA, Arlington
	INP AS RUz., Tashkent		CERN, Geneva		
	NOSU, Vladikavkaz; NRC KI PNPI; ITEP; SINP MSU; NNRU “MEPhI”; LPI RAS; INR RAS, IHEP; SSU, Samara; SPbSU				
Member States of the Collaborations @ NICA (BM@N, MPD, SPD)		Member States of the Collaborations @ CERN (ALICE, ATLAS, CMS, AMBER, GEANT4, etc)		Member States of the BAIKAL-GVD anf TAIGA projects	

International Large-scale projects

Russian research institutes and universities actively participate in international large-scale projects:

- LHC, CERN (experiments: ATLAS, ALICE, LHCb, CMS)
- XFEL, DESY (European free electron laser)
- ESRF, France (European synchrotron center)
- FAIR, GSI, Germany (CBM, PANDA experiments)
- ITER, France …

International large-scale projects are being prepared in Russia:

- **NICA**, JINR, Dubna (proton and heavy ion collider)
- **PIK**, PNPI, Gatchina (high-flow reactor complex)
- **SKIF**, INP SB RAS Novosibirsk (Siberian ring photon source)
- **Super S-Tau Fabric**, Sarov (electron-positron collider)
- **Нейтринная программа (Байкал, JUNO, NOVA, DUNE …)**
- **синхротронно-нейтронная программа, науки о жизни**

Support and development of the software environment for JINR research

- Development of the software environment for processing and analyzing data from the NICA experiments
- Creation, implementation and development of an information and computing complex for processing, analyzing and storing data for the SPD experiment
- Creation of specialized databases and information systems for the Collaborations @ NICA and LHC (ATLAS, BM@N, MPD)
- Development and creation of an information and computing system for
  - automating the processing of data from radiobiological studies
  - intelligent determining the state of agricultural and decorative plants
  - monitoring and predicting the state of the environment