Information and Computing Infrastructure of JINR
(05-6-1118-2014/2023),
including the Project
Multifunctional Information and Computing Complex
of MLIT JINR

(Written report)

Leader: V.V. Korenkov – korenkov@jinr.ru

Deputy: T.A. Strizh – strizh@jinr.ru

April, 2023

CONTENTS

Introduction

In 2020-2023, within the “Networks, Computing, Computational Physics” direction, the Meshcheryakov Laboratory of Information Technologies (MLIT) carried out research on the theme of the first priority: 05-6-1118-2014/2023 “Information and Computing Infrastructure of JINR”, the purpose of which is to develop the network, information and computing infrastructure of JINR for the research and production activity of the Institute and its Member States on the basis of state-of-the-art information technologies in accordance with the Seven-Year Plan for the development of JINR. The major direction within the theme is the development of the JINR MLIT **Multifunctional Information and Computing Complex (MICC)** presented as **the Project**. The JINR computer infrastructure encompasses the IT ecosystem for the NICA project (BM@N, MPD, SPD), which includes all MICC computing components and storage systems owing to grid technologies; the Tier1 grid site for the CMS experiment at the LHC; Tier2/CICC that provides support for the experiments at NICA (BM@N, MPD, SPD), LHC (ATLAS, ALICE, CMS) and other large-scale experiments, as well as support for users of JINR’s Laboratories and Member States; the integrated cloud environment of the Member States to support users and experiments (NICA, ALICE, BESIII, NOvA, Baikal-GVD, Daya Bay, JUNO, etc.); the HybriLIT platform with the “Govorun” supercomputer as the major resource for high-performance computing.

The MICC project is aimed at:

- Development of the network, information and computing infrastructure to ensure the implementation of the Seven-Year Plan for the development of JINR with the necessary means of modern information technologies. Creation of a single space for resources existing at JINR: computing, information and data storage ones.
- Development of the external and local network infrastructures that provide the possibility of data exchange between the Institute’s subdivisions, the JINR Member States and international organizations cooperating with JINR; creation of a network infrastructure for receiving and transmitting data between the BM@N, MPD, SPD facilities and the on/offline clusters of the NICA megaproject; support and development of common network services, such as electronic mail (e-Mail), domain name management (DNS), data caching (Proxy), resource management (IPDB), monitoring (NMIS), single authorization service (SSO), information security system.
- Modernization and development of the MICC engineering infrastructure, including power supply and uninterruptible power supply systems, air conditioning and ventilation systems, the fire safety complex in accordance with the growth of computing power and data storage volumes.
- Creation of an offline cluster based on the MICC within the development of computing for the NICA megaproject, of a unified information and computing platform (environment) for the implementation of the JINR neutrino program.
- Extension of computing resources and data storage systems:
  - MICC Tier1, Tier2/CICC grid components in accordance with the JINR Seven-Year Plan, which will provide the required level of resources for all LHC collaborations at Tier1 and Tier2 at JINR.
  - MICC cloud component in order to expand the range of services provided to users. Creation of an integrated cloud environment with the clouds of the JINR Member States.

- “Govorun” supercomputer to meet the needs of users from JINR and its Member States with computing resources for solving tasks related to high-performance computing (HPC). Provision of users with modern IT solutions and services in the field of HPC.
- Creation of the JINR Data Lake on the basis of MICC storage systems.

- Creation of an information-analytical intelligent monitoring system that allows aggregating information from different levels of a computing center: engineering infrastructure, networks, compute nodes, task launch systems, data storage elements, grid services, etc., which will ensure a high level of the MICC reliability.

The other key objectives of the theme are:

— **Information and software support of the JINR research and production activity**

One of the important tasks of the theme is the maintenance and further development of services related to the corporate information system and designed to simplify and optimize the scientific and administrative activities of the Institute. These services include: the 1C system; the NICA project management system (APT EVM); the information search system (ISS); the system of the electronic signing, storage and search of documents of the JINR main office administration as well as documents of MES&CC (Management of Economic Services and Capital Construction) and PLS (Procurement and Logistics Service); the information system of scientific event management (Indico); the information system for storing and managing data on the results of the research activity of JINR staff members (Personal Information System (PIN)); the management accounting system (ADB2); the JINR electronic document management system (EDMS JINR), etc.

— **Development of the system for training and retraining IT specialists on the basis of the JINR MICC and its educational components**

This section of the theme is aimed at creating a special training polygon based on the MICC for conducting regular courses on modern IT technologies, including

- ✓ special courses and tutorials on novel HPC technologies, technologies and tools for solving applied tasks on the basis of machine and deep learning methods for the Institute’s staff members, students and young scientists from the JINR Member States within practices organized by the UC, as well as within conferences and schools organized by JINR.
- ✓ special courses and tutorials in the JINR Member States in accordance with international cooperation programs.
- ✓ specialized courses on training IT specialists to solve tasks related to data processing and analysis for megascience experiments, including the NICA project.
- ✓ IT schools for young scientists and specialists.
- ✓ creation of a laboratory of intelligent robotics for the development of cognitive control systems; development of a laboratory workshop on robotics.

MICC project

The main information and computing resources of JINR are concentrated in the MICC [1], which can be considered as JINR’s unique basic facility that plays a defining role in scientific research, entailing modern computing power and storage systems. The MICC uniqueness is ensured by the consolidation of all state-of-the-art information technologies from the network infrastructure with a bandwidth to 4x100 Gb/s and the distributed data processing and storage system based on grid technologies and cloud computing to the hyperconverged computing infrastructure with liquid cooling for supercomputer applications. Multifunctionality, high reliability and availability in 24x7x365 mode, scalability and high performance, a reliable data storage system, information security and an advanced software environment are the main requirements that the MICC meets.

The MICC development encompassed the continuation of the extensive modernization of the MICC cooling and power supply systems, which started in 2019, the modernization and development of the MICC computing resources and data storage systems, the development of the IT infrastructure of the NICA megascience project, the expansion of the performance of the grid components, i.e., Tier1 and Tier2, the extension of the cloud component and creation of an integrated cloud environment for JINR experiments, the enlargement of the HybriLIT heterogeneous computing platform, including the “Govorun” supercomputer.

The active use of the MICC resources for JINR’s research and applied tasks was underway. Thanks to grid technologies (DIRAC Interware), which brought together the dedicated computing resources of all MICC components, simulation campaigns for the MPD experiment of the NICA complex were successfully held. The Tier1 grid site for the CMS experiment at the LHC continued to be a leader among similar world sites. Tier2/CICC provided data processing for the experiments at the LHC, NICA, FAIR and other large-scale experiments, as well as support for users from the JINR Laboratories and Member States. The cloud environment of JINR and its Member States was mainly used for computing within the JINR neutrino program. The HybriLIT platform, which includes the basic resource for high-performance computing, the “Govorun” supercomputer, and the education and testing polygon, was actively used by registered users.

One can find below more detailed information on the results obtained during the implementation of the MICC project in the 2020-2023 time period.

Network infrastructure

During the reporting period, a significant modernization of the JINR network infrastructure, which is a fundamental component of the IT infrastructure of JINR and of the MICC, was carried out. This is an intricate complex of multifunctional network equipment and specialized software. It is the foundation for the JINR information and computing infrastructure, which was created and is constantly developing, providing access to the Internet, the computing resources and the data storage systems both within JINR and in external scientific organizations cooperating with JINR. The JINR network infrastructure consists of the following functional components: the external optical telecommunication data transmission channel JINR – Moscow, the backbone of the JINR local area network, the local area networks of the Institute’s subdivisions.

A project to increase the capacity of the JINR – Moscow backup channel with a bandwidth from 100 Gb/s to 4x100 Gb/s was implemented, the capacity of the local area network was increased to 2x100 Gb/s, and a distributed computing cluster network was built between the DLNP and VBLHEP sites with a capacity of 400 Gb/s, with double redundancy to improve the reliability of the optical backbone.

At present, the JINR external distributed network (Fig. 1) is represented by the 100 Gb/s JINR – CERN direct channel and its 100 Gb/s backup channel, passing through Moscow and Amsterdam, which ensure the operation of the LHCOPN network for connection between Tier0 (CERN) and Tier1 (JINR) and of the LHCONE external overlay network allocated to the JINR Tier2 center, for communication with the RUHEP collaboration and the networks of the National Research Computer Network of Russia and RetN using RU-VRF technology [2-3]. IPv6 routing was implemented for the Tier1 and Tier2 sites. In 2021, the Nortel DWDM (Dense Wave Division

Multiplexing) equipment on the RSCC fiber-optic links route (Dubna, Radishchevo, Moscow) was replaced with the Infinera new equipment, which made it possible to broaden the bandwidth of all Dubna – Moscow channels.

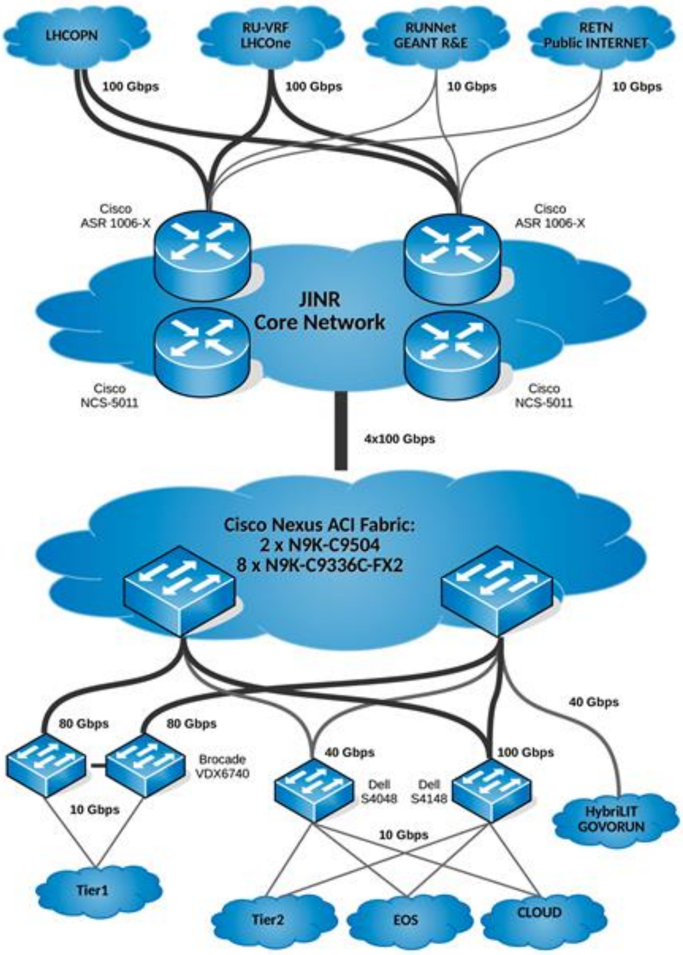

Fig. 1. Structure of the JINR external network

The distribution of the incoming and outgoing traffics by the JINR subdivisions in 2020-2022 (exceeding 25 TB by the incoming traffic) is shown in Table 1.

Table 1.

Subdivision	2020		2021		2022	
	Ingoing (TB)	Outgoing (TB)	Ingoing (TB)	Outgoing (TB)	Ingoing (TB)	Outgoing (TB)
Laboratory of Nuclear Problems	425.84	271.81	941.9	158.27	229.59	113.38
Laboratory of High Energy Physics	208.03	130.76	280.18	160.98	430.2	209.23
Laboratory of Neutron Physics	97.7	130.15	106.46	69.34	137.05	42.05
Laboratory of Information Technologies	93.26	87.99	325.56	449.63	330.76	204.99
Hotel&Restaurant Complex	87.04	27.71	171.74	33.52	413.24	60.25
Dubna University	86.97	51.16	129.73	41.08	139.03	39.38
Laboratory of Nuclear Reactions	81.98	70.52	108.59	48.07	137.7	32.69
Remote Access Node	63.91	10.89	77.75	10.44	84.88	11.84

JINR Directorate	56.09	90.55	76.42	76.51	27.63	2.94
University Centre	31.4	10.59	29.83	6.35	84.88	11.84
Laboratory of Theoretical Physics	24.39	27.85	26.22	21.64	35.57	15.96

The overall incoming traffic of JINR, including the general-purpose servers, Tier1, Tier2, the computing complex, the "Govorun" supercomputer and cloud computing, amounted to 29.91 PB in 2020, 33.23 PB in 2021, 29.56 in 2022, while the overall outgoing traffic reached 36.94 PB in 2020, 35.86 PB in 2021, 34.19 PB in 2022. The traffic with the scientific and educational networks is overwhelming.

The local area network (LAN) is based on the JINR backbone network with a bandwidth of 2x100 Gb/s and the distributed multi-node cluster network between the DLNP and VBLHEP sites (4x100 Gb/s).

The MICC internal network has a Tier1 segment built at the Brocade factory with a bandwidth of 80 Gb/s. The network segments of the EOS data storage system, Tier2, cloud computing and the "Govorun" supercomputer are built on the Dell and Cisco hardware. 10 Gb/s and 100 Gb/s ports are used to connect server components to the switches of the MICC network core built on Cisco Nexus 9504 and Nexus 9336C switches with an N x 100 Gb/s port bandwidth.

The internal network of the "Govorun" supercomputer consists of three main parts, namely, a communication and transport network, a control and monitoring network, a task management network. The communication and transport network uses Intel OmniPath 100 Gb/s technology. The network is built on a "thick tree" topology based on 48-port Intel OmniPath Edge 100 Series switches with full liquid cooling. The control and monitoring network enables the unification of all compute nodes and the control node into a single Fast Ethernet network. This network is built using Fast Ethernet HP 2530-48 switches. The task management network connects all compute nodes and the control node into a single Gigabit Ethernet network. The network is built using HPE Aruba 2530 48G switches.

The work on the development and improvement of the network components of the JINR IT infrastructure was in progress. The Cisco ACI factory based on the Cisco Nexus 9504 and Cisco Nexus C9336C-FX2 equipment, allowing one to connect the MICC components at speeds of up to 100 Gb/s and more, was put into operation. The EOS distributed storage network and the cloud computing network were connected to the RU-VRF/LHCONE external network. The commissioning of the 4^{th} module in the MICC hall and its equipping with computing resources were performed together with setting up and connecting central and rack switches.

The central network virtual cluster of the JINR network service (NOC – Network Operation Center), which is built on top of the Proxmox VE (Virtual Environment) open-source software under the GNU license, was modernized. This approach made it possible to use the NOC central cluster in 24x7 mode. The architectural solution ensures the organization of non-stop operation during updates of both software and hardware components of the network cluster. It is noteworthy that the virtual machines operating in the central cluster serve all essential elements of the JINR network. The NOC cluster ensures the operation of the NOC and JINR services such as DNS (Domain Name System), DHCP (Dynamic Host Configuration Protocol), SMTP (Simple Mail Transfer Protocol), SNMP (Simple Network Management Protocol), SSO (Single Sign-On), registration (user), authorization (devices), authentication (users), switching, routing, security, video conference communication, nmis monitoring, sshgate, centralized logging, sip telephony, etc.

The NOC regularly updates software on 15-20 servers (webmail.jinr.ru, indico.jinr.ru, mail.jinr.ru, aillist.jinr.ru, mx1.jinr.ru, mx2.jinr.ru, auth-1.jinr.ru (login.jinr.ru), auth-2.jinr.ru, etc.), which keeps the systems up to date.

The work on the improvement of the mail.jinr.ru service was systematically performed: new software (Proxmox Mail Gateway) was introduced, it significantly reduced the number of spams by training the system in spam filtering mechanisms; a new hypervisor was prepared and put into operation, scripts were developed, a new adm-mail.jinr.ru server was created for the “cold” copy of mail.jinr.ru. Support for the mailing services (mailist.jinr.ru), “Personal Account”, News, VPN, Edurom, Elibs, IPDB was carried out.

In 2022, the jinr.int zone was registered, and the mail.jinr.ru mail server was adapted to work with the jinr.int zone.

Continuous availability monitoring and the logging of logins to network sessions were added to the SSO system. The ability to edit LDAP entries for working in external services and the ability to register non-JINR staff members were added. The coverage of the eduroam WiFi network on the JINR territory was extended.

The functionality of the system for network traffic analysis was expanded with the help of new written scripts that can identify infected and hacked user computers. Since March 2022, an enhanced network protection regime has been provided. More than 80 incidents related to the hacking of the JINR network resources, copyright infringement, etc. were identified and processed. As part of cooperation with third-party scientific organizations, VPN access to the network was provided for more than 110 users of the computing resources. A system for monitoring and tracking the status of over 770 network elements was ensured. The sshgate remote access service was put into operation. A mandatory check for vulnerabilities is performed for websites opened for access from the outside.

The JINR LAN comprises 9,291 network elements and 18,044 IP addresses, 6,355 network users, 4,477 users of mail.jinr.ru, 1,455 users of electronic libraries, 837 users of the remote access service and 111 users of the EDUROAM service.

Engineering infrastructure

The work on the improvement of the MICC engineering infrastructure, designed to ensure the reliable, uninterrupted and fault-tolerant operation of the information and computing system and the network infrastructure, was in progress. Using the integrated approach to building the MICC engineering infrastructure allowed one to elaborate algorithms of the equipment operation and interaction of separate systems both in a normal operation mode and in emergencies, which ensured the uninterrupted performance regardless of external factors. The uninterruptible power supply system provides guaranteed power to connected consumers, the automatic launch of diesel-generator units (DGUs), the automatic load transfer from the main external power supply network to DGUs and vice versa, and allows one to send messages to the dispatcher post in the case of an emergency with DGUs.

The MICC computing facilities are hosted in one computing hall of 800 m^2 of floor-space at the 2^{nd} floor of the MLIT building. It currently consists of eight separate IT equipment modules (Fig. 2) with 2 MW power:

- Module 1 and Module 2: 22.55 m^2 of floor-space each, 33 server racks and 20 kW per rack;
- Module Tier1: 29.33 m^2 of floor-space, 16 server racks and 35 kW per rack;

- Tape library space: 13 m^2 of floor-space, two robotic tape libraries IBM TS3500 and IBM TS4500 with a total capacity of 50.6 PB;
- “Govorun” supercomputer: 1.97 m^2 of floor-space, 4 racks and 100 kW per rack;
- Module that hosts critical services of the standard business computing type (administrative systems and databases, etc.);
- Module 4: 36.12 m^2 of floor-space, 20 server racks and 35 kW per rack;
- Network equipment module that hosts the main network services for the MICC, JINR local and wide area networks.

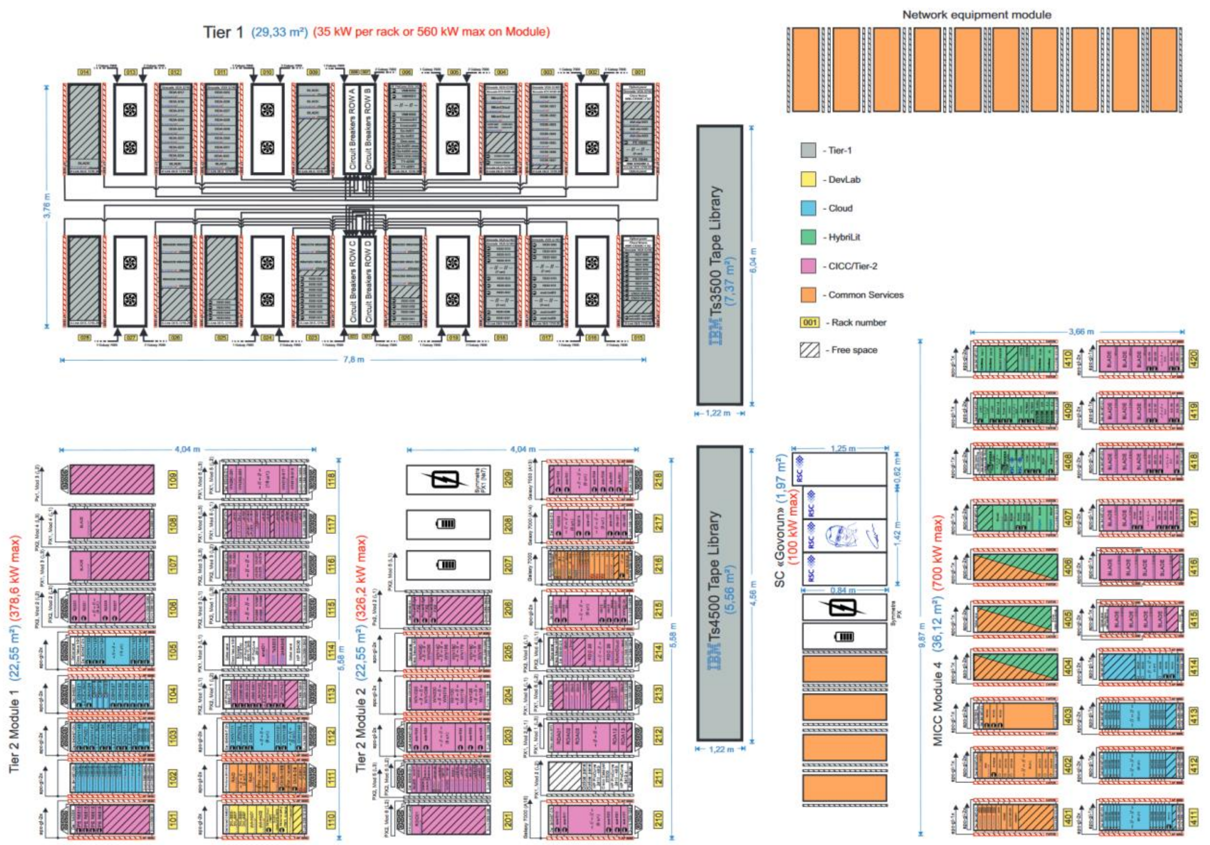

Fig. 2. Location of equipment in the MICC server hall

All racks are uninterruptible power supply (UPS) backed up with an autonomy of 10–15 minutes. Racks are equipped with intelligent (switched and metered) power distribution units, which enable the fine-grained monitoring of power consumption. There are two diesel-generator backups in operation to provide the computer center with electricity in the case of disconnection from external power supply networks.

The MICC existing climate control system is a complex of the interconnected equipment with different air and liquid cooling schemes, with the help of which the corresponding temperature regime ensuring the MICC functioning in 24x365 mode is created. Module Tier1 and Module 4 are air-cooled with in-row racks arranged between server racks. Modules 1 and 2 are air-cooled, and the cold air is blown through large ducts underneath the false floor, where it diffuses into cold corridors through perforated floor tiles. The “Govorun” supercomputer is fully “hot” water-cooled, which allows for a power density of 100 kW per rack and PUE = 1.06. According to the type of heat removal, the MICC climate control system refers to the mixed type that combines systems with the evaporation of a coolant and systems with an intermediate coolant.

All technological equipment that provides both the guaranteed power supply to the MICC and the cooling system is located at the first and basement floors of the building. Chillers, dry coolers and diesel generators are located on the territory adjacent to the MLIT building.

The DCIM (Data Center Infrastructure Management) system was put into operation for controlling and accounting the equipment of the MICC hall. This software allows one to visualize and control the MICC physical infrastructure on the basis of data on the equipment and its location entered in the DCIM database, to provide management and monitoring services. Figure 3 illustrates examples of equipment visualization.

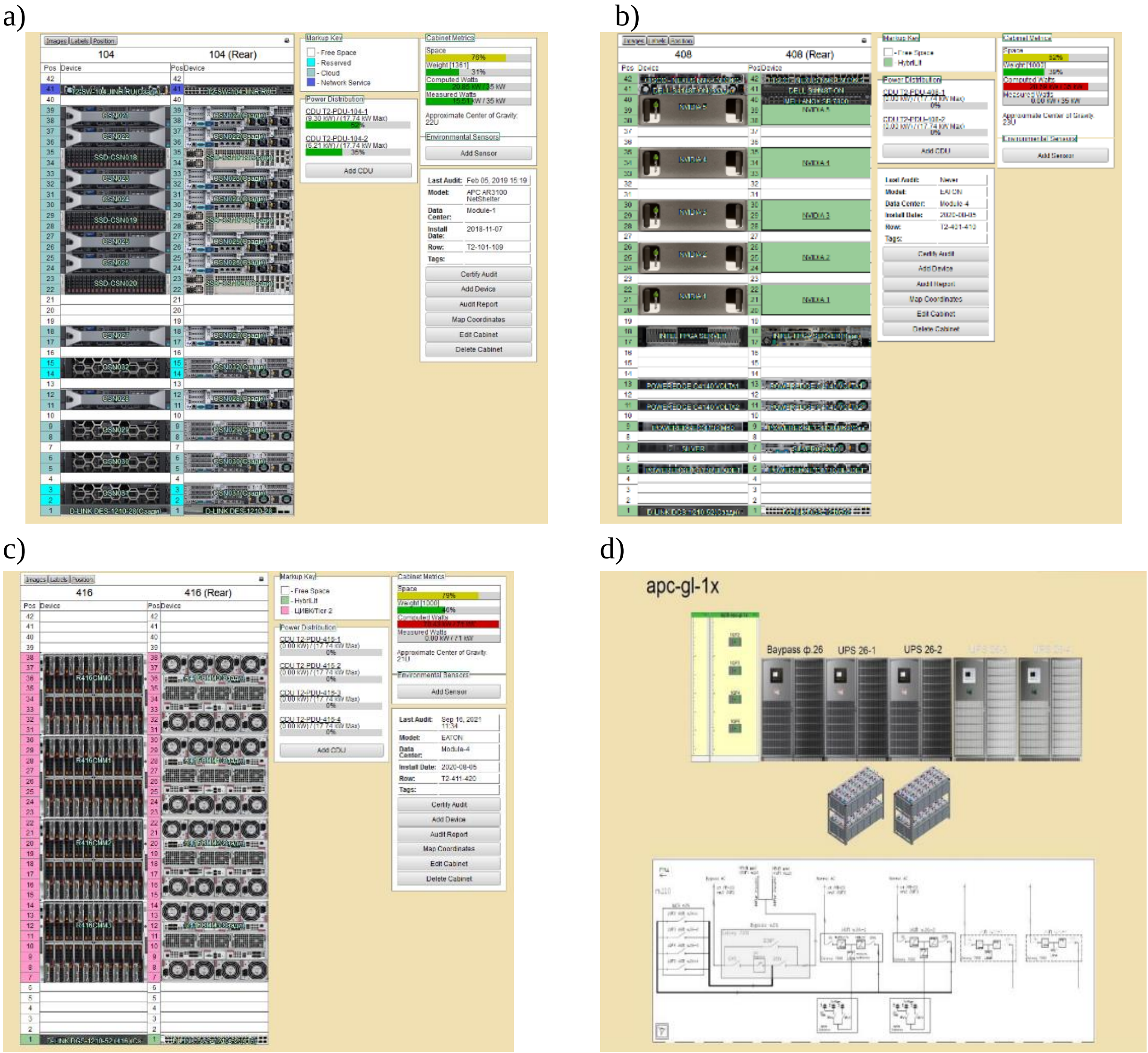

Fig. 3. Examples of the equipment visualization system in racks 104 (a), 403 (b) and 416 (c) of the MICC hall and of the uninterruptible power supply system (d)

JINR grid environment (Tier1 and Tier2 sites)

The JINR grid infrastructure is represented by the Tier1 center for the CMS experiment at the LHC and the Tier2 center for processing data from the NICA, LHC, BES, BIOMED, NOvA, ILC experiments and others [4]. Both JINR grid sites ensure on average 100% availability and reliability of services.

For many years, the Tier1 resource center was used only to perform jobs of the CMS experiment at the LHC within the participation of JINR and its Member States in this experiment. Since 2021, the introduction of the DIRAC platform has made it possible to utilize the allocated resources of this center for simulation jobs of the MPD experiment of the NICA project. The amount of resources is regulated by requests from the MPD collaboration.

During the reporting period, the Tier1 data processing system was increased to 20,096 cores, providing a performance of 323,820.54 HEP-SPEC06. The following software and compilers are used: CentOS Scientific Linux version 7.9, gcc (GCC) 4.8.5, gcc-11.2.1, gcc-c++-11.2.1, gcc-gfortran-11.2.1, C++ (g++ (GCC) 4.8.5, GNU Fortran (GCC) 4.8.5, DCACHE-6.2 for data storage, Enstore 6.3 for tape libraries and FTS. The total usable capacity of disk servers was increased to 14 PB. In 2020, a new tape library, IBM TS4500, with a total volume of 40 PB was commissioned. To date, the data long-term storage system consists of the IBM TS3500 and IBM TS4500 libraries with a total volume of 50.6 PB and is focused on servicing the NICA and CMS experiments. Software for NICA is installed in CVMFS using GitLab by users/software developers themselves.

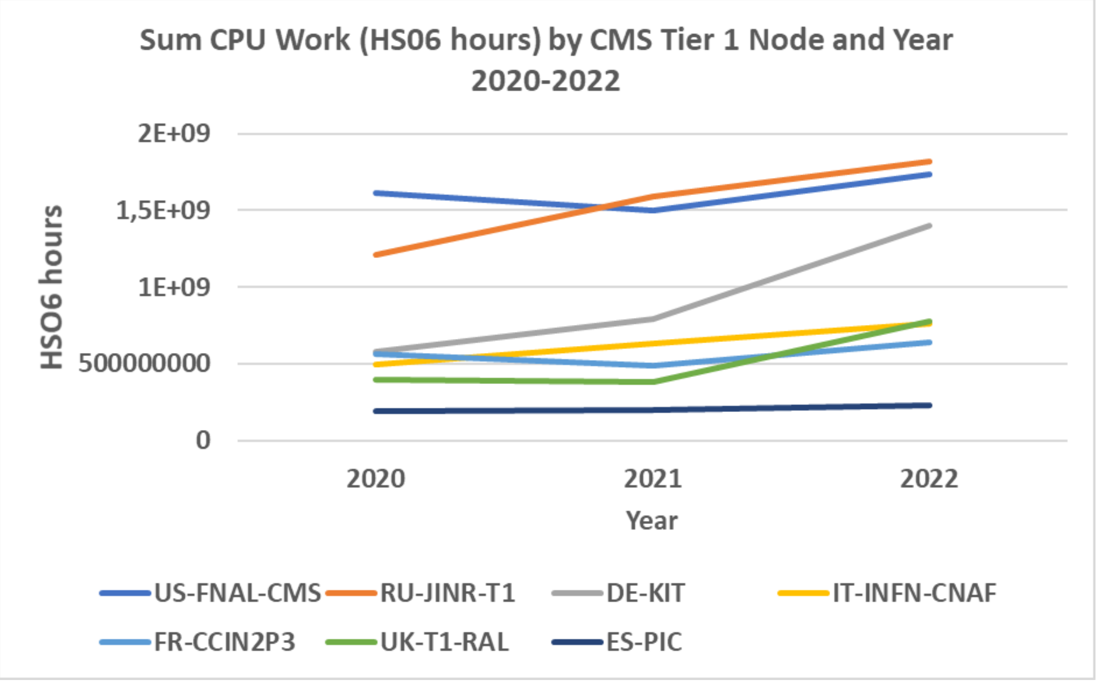

Fig. 4. Distribution by the normalized CPU time in HEP-SPEC06 hours for the world Tier1 centers of CMS experimental data processing in 2020-2022

In 2020, at Tier1 there happened a transition related, on the one hand, with the end of support of the software used for computing elements, i.e., CREAM-CE, the Torque batch processing system and the Maui scheduler, and, on the other hand, with the fact that the previous software and systems could not cope with the increased load and a large number of computing machines. To replace CREAM-CE, ARC-CE (Advanced Resource Connector-Computing Element) was chosen; it is widely used in the WLCG (Worldwide LHC Computing Grid). SLURM, an open source, highly scalable, fault-tolerant cluster manager and job scheduler for large clusters, was selected as a resource manager. It enables flexible planning with priorities, a fair distribution of resources between different users and the optimization of computing resource utilization. SLURM is also used on the "Govorun" supercomputer.

The CMS Tier1 center at JINR has demonstrated stable work throughout the reporting period and since 2021, in terms of performance, Tier1 has been ranked first among Tier1 world centers for the CMS experiment (Fig. 4). In 2020-2022, more than 890 million events were processed, which accounts for 19.6% of the total number of processed events (Fig. 5a) and 19.1% of the total CPU load of all Tier1 centers for the CMS experiment (Fig. 5b).

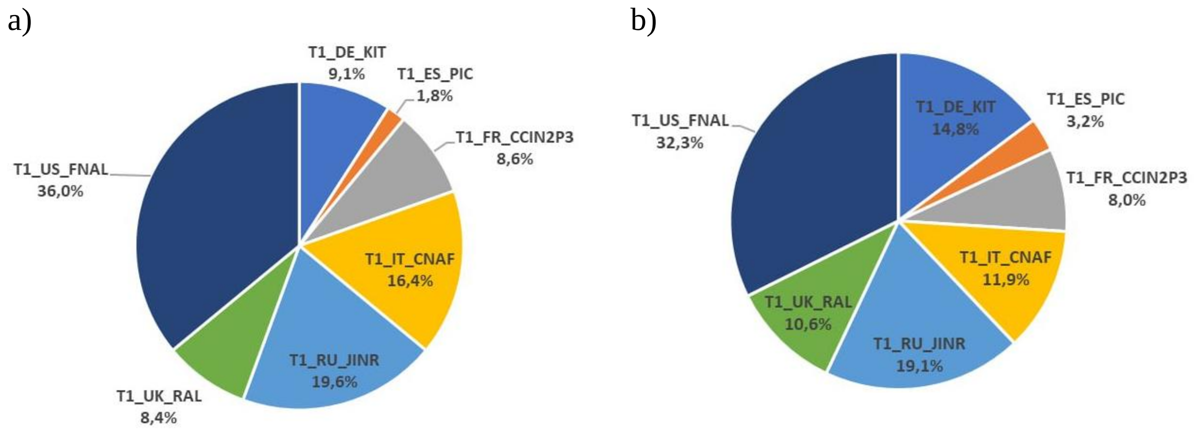

Fig.5. Contribution of the world Tier1 centers to CMS experimental data processing in 2020-2022: a) number of processed events; b) distribution by the normalized CPU time in HEP-SPEC06 hours

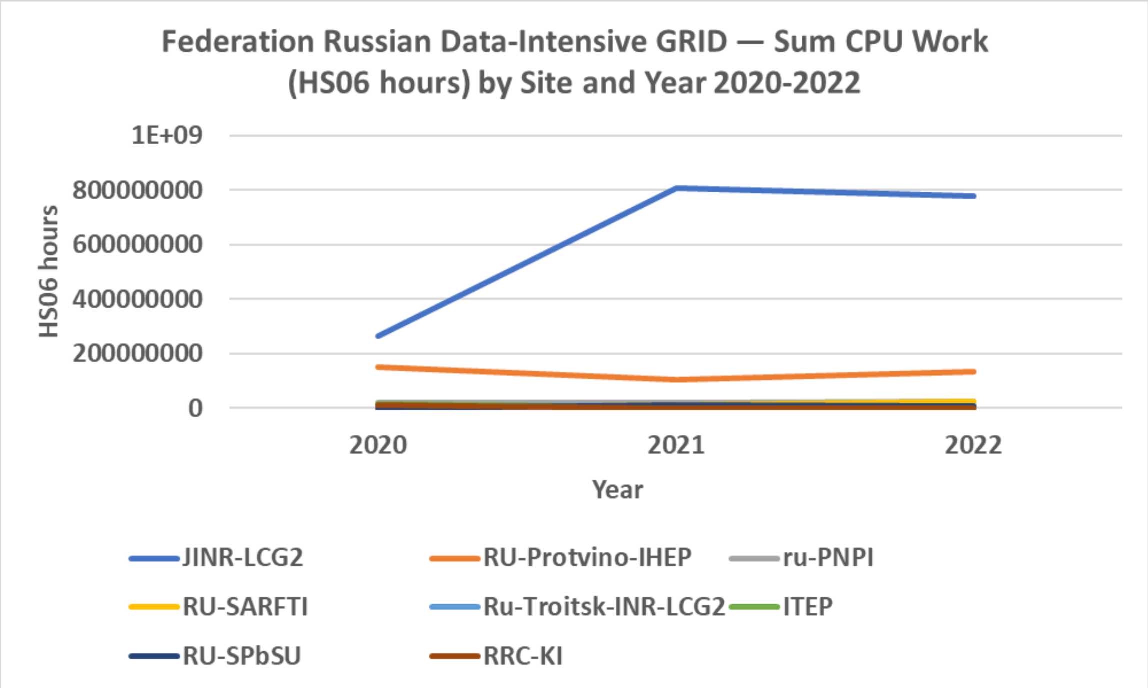

Fig. 6. Distribution by the normalized CPU time in HEP-SPEC06 hours for the RDIG grid sites

The computing resources of the Tier2 center were extended to 9,244 cores, which currently provides a performance of 149,938.7 HEP-SPEC06. The total usable capacity of disk servers is 4,763 TB for ATLAS, CMS, ALICE and 140 TB for other virtual organizations. The JINR Tier2 output is the highest in the RDIG (Russian Data Intensive Grid) Russian Consortium. Figure 6 illustrates the distribution by the normalized CPU time in HEP-SPEC06 hours for the RDIG grid sites. More than 75% of the total CPU time in the RDIG is used for computing on our site.

The MICC allows users to perform calculations outside the grid environment. This is necessary for some experiments and local users of the JINR Laboratories. JINR and grid users have access to all computing power via a unified batch processing system.

In the reporting period, the EOS-based data storage system was expanded. At present, 17 PB of disk space is available for EOS users. EOS is visible as a local file system on the MICC working nodes and allows authorized users (by the kerberos5 protocol) to read and write data. Figure 7 demonstrates the statistics on the use of the EOS system in 2022.

The stable and efficient operation of Tier1, Tier2, storage systems and the required level of cluster cybersecurity were ensured by the systematic updating of the firmware of the server components, the version of the operating system kernel and the firmware of the service modules of IDRAC/IPMI servers.

To ensure the guaranteed and stable functioning of the infrastructure under constant load conditions, centralized and timely software maintenance and the rapid introduction of new compute nodes are required. As a solution to this problem, the Lifecycle Management Service (LMS) was created; its purpose is to automate the process related to software maintenance and the commissioning of new computing resources [5].

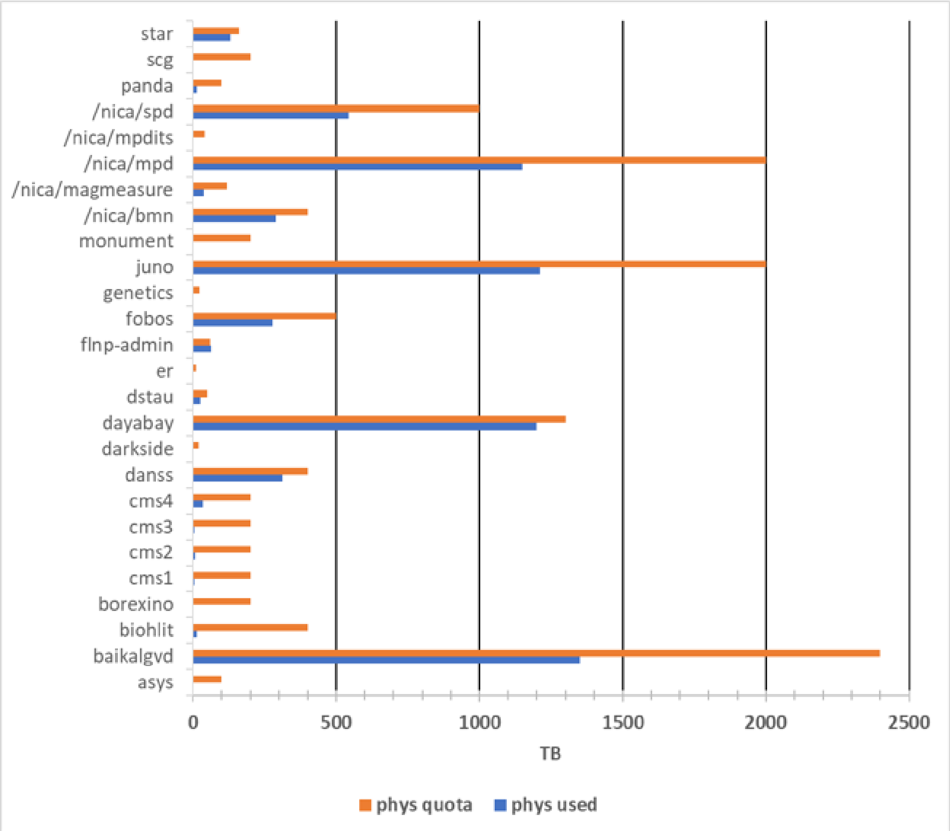

Fig. 7. Statistics on the use of the EOS system by user groups and experiments in 2022

Cloud environment

In 2020-2023, the work on expanding the JINR cloud environment and combining the computing power of the organizations of the Institute's Member States into a unified information and computing environment was in progress [6].

The resources of the cloud infrastructure were enlarged due to the contributions of the NOvA experiment (480 CPU cores, 2.88 TB of RAM, 1.728 PB of disk space for ceph-based storage) and the commissioning of 2,880 CPU cores with 46.08 TB of RAM purchased by the JUNO experiment.

Besides the general-purpose ceph-based storage with a total raw capacity of 1.1 PB, two new storage elements were deployed in 2021: the first is devoted to NOvA experiment needs only, and the second is an SSD-based ceph storage for a set of production services and users with high demands in terms of disk I/O. The main parameters of all these cloud storage systems are listed in Table 2.

Table 2. Characteristics of the cloud storage elements.

	Disk type	Consumers	Ceph version	Raw capacity, PiB	Replication	Connectivity
Regular cloud storage	HDD	All	14.2.21	1.1	3x	2x10GBase-T
NOvA storage	HDD	NOvA	15.2.11	1.5	3x	2x10GBase-T
Fast cloud storage	SSD	High disk I/O demands	15.2.13	0.419	3x	4x10GBase-T + 2x100Gbps –

The total amount of the resources located in the JINR cloud infrastructure is currently 5,150 CPU cores, 60 TB of RAM and 3.41 PB of raw disk space in the ceph storage.

A fairly wide set of software was used for JINR cloud servers and for the monitoring of some of its services (Nagios, InfluxDB time series database (TSDB), Prometheus TSDB, etc.). node_exporters were deployed on all cloud servers to provide Prometheus with servers state data. Alerting is implemented at the Prometheus level. Grafana is used for data visualization.

The JINR cloud is managed in accordance with the “Infrastructure-as-a-Code” (IaC) approach, where host provisioning and management are performed via configuration files. The Foreman and Puppet software are used for this purpose.

Figure 8 provides information on the consumption of the cloud infrastructure resources in 2022: the main users are neutrino experiments and MLIT.

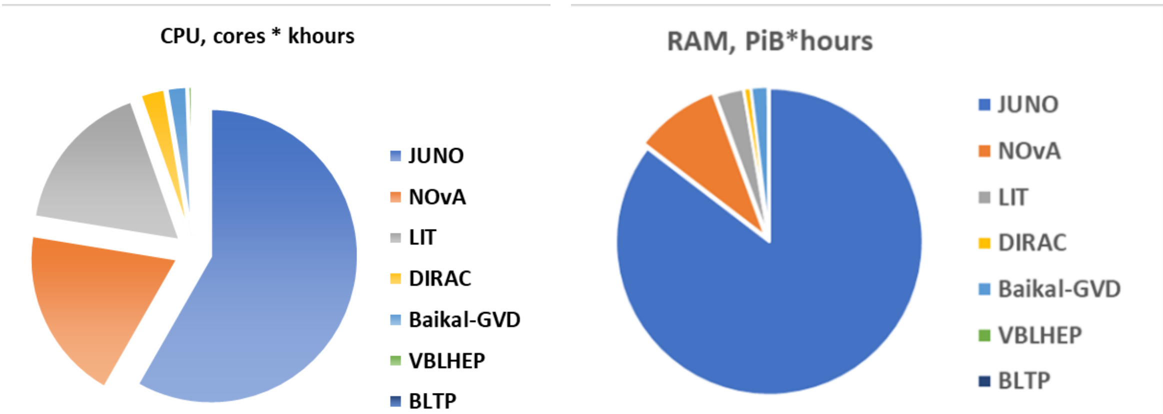

Fig. 8. Use of cloud computing by the experiments and the JINR subdivisions in 2022

The work on the development and support of the computing platform for neutrino experiments was underway. Due to the termination of support for the GSI authentication and the transition of the DUNE and NOvA experiments to an authentication system using tokens, the HTCondor cluster of the cloud neutrino platform was updated to version 9.0, in which support for authentication by JSON web tokens (i.e., JWT) was implemented. For the DUNE experiment, the StashCache caching data storage with a total volume of 1 TB was deployed and connected to the Open Science Data Federation (OSDF).

At the request of the JUNO experiment, an exporter of data both on the current load in the batch cluster of the neutrino platform to use this data by the experiment in order to optimize the distribution of computing jobs in the JUNO global grid infrastructure, and on the EOS storage used by the experiment and the load of JINR network channels was developed.

In the jupyter.jinr.ru interactive computing service, at the request of users, instead of a set of different images with different software, one universal image based on Datascience Notebook from the Jupyter Docker Stacks set, including all the basic software and modified environment necessary for the correct operation of the ROOT software, was prepared for data analysis.

A server with the NVIDIA A100 Ampere 40 GB graphics card was purchased and put into operation for neutrino computing platform users involved in the development of machine learning algorithms, as well as using the corresponding application software.

In the JINR cloud storage, an additional local replica of modeled and real data sets in the Near and Far Detectors of the NOvA experiment was created to perform the oscillation analysis of the experiment completely independently of FermiLab infrastructure tools.

About 60 TB of data from the Borexino experiment was copied from the CNAF data center (Italy) to the EOS storage at JINR. A fairly new approach together with the IAM (Identity & Access Management) service was used to authenticate and authorize the user under whom the data was copied.

The JINR cloud is part of the distributed information and computing environment (DICE) based on the resources of JINR and its Member States' organizations [7]. The amount of JINR cloud resources contributed to the DICE varies depending on its load.

In 2020, cloud infrastructures at North Ossetian State University, Sophia University "St. Kliment Ohridski" and the Institute for Nuclear Research and Nuclear Energy of the BAS were deployed and connected to the DICE. In 2021, the cloud of the Egyptian National STI Network of the Academy of Scientific Research and Technology and in 2022, the cloud of the Institute of Nuclear Physics of the Academy of Sciences of Uzbekistan were put into operation and integrated into the DICE. The server and network hardware for the cloud infrastructure of the Institute of Nuclear Physics (Alma-Ata, Kazakhstan) was purchased and supplied. The work on putting the hardware into operation and expanding the cloud capacity of this organization is underway. At present, Plekhanov Russian University of Economics (Russia), the North Ossetian State University (Russia), the Institute of Nuclear Physics (Kazakhstan), the Institute of Physics of the National Academy of Sciences of Azerbaijan, the Institute for Nuclear Research and Nuclear Energy (Bulgaria), Sophia University "St. Kliment Ohridski", the Scientific Research Institute for Nuclear Problems of the Belarusian State University, the Institute of Nuclear Physics (Uzbekistan), the Egyptian National STI Network of the Academy of Scientific Research and Technology (Egypt) were fully integrated into the JINR DICE, and the Georgian Technical University is in the progress of integration (Fig. 9).

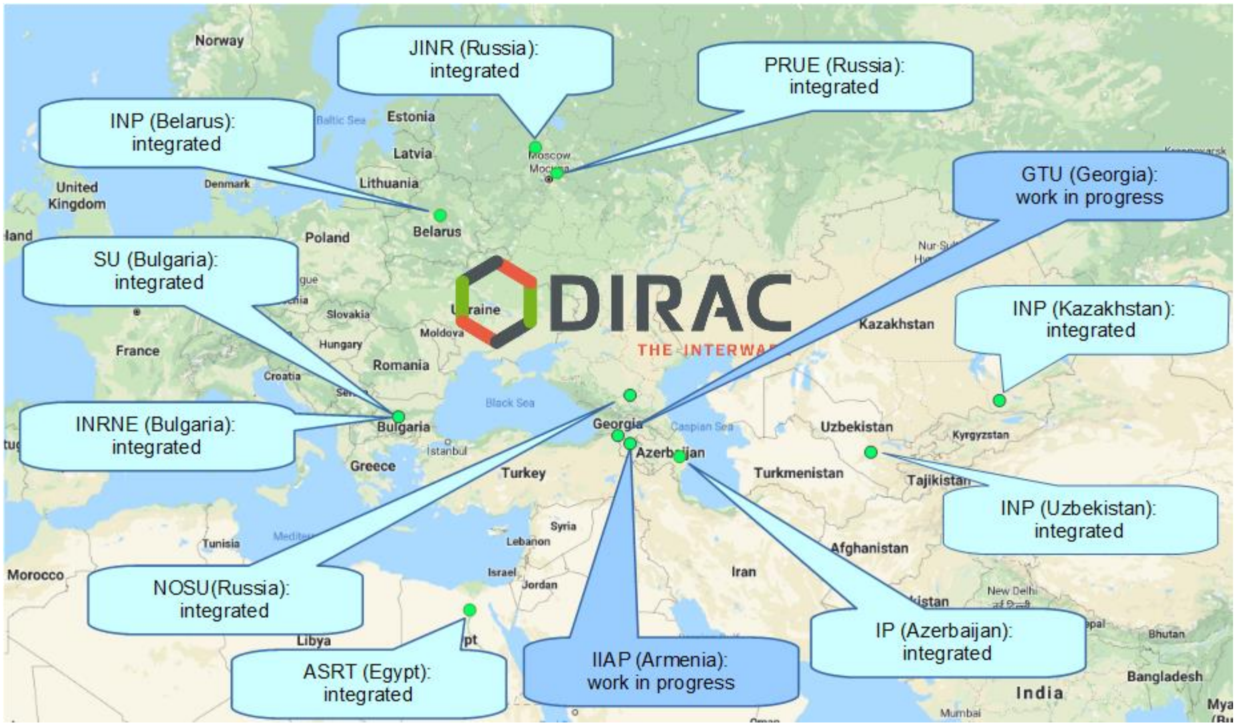

Fig. 9. Participants of the distributed information and computing environment (JINR DICE)

The http:/dice.jinr.ru web portal with information about the participants of the distributed information and computing environment (JINR DICE), which combines the JINR cloud and the cloud infrastructures of the JINR Member States, and about conducted training events [8] on work at the JINR DICE was put into operation. In 2020, the Baikal-GVD experiment joined the utilization of the DICE computing power and now it is the main consumer of the JINR DICE resources (96%, Fig. 10).

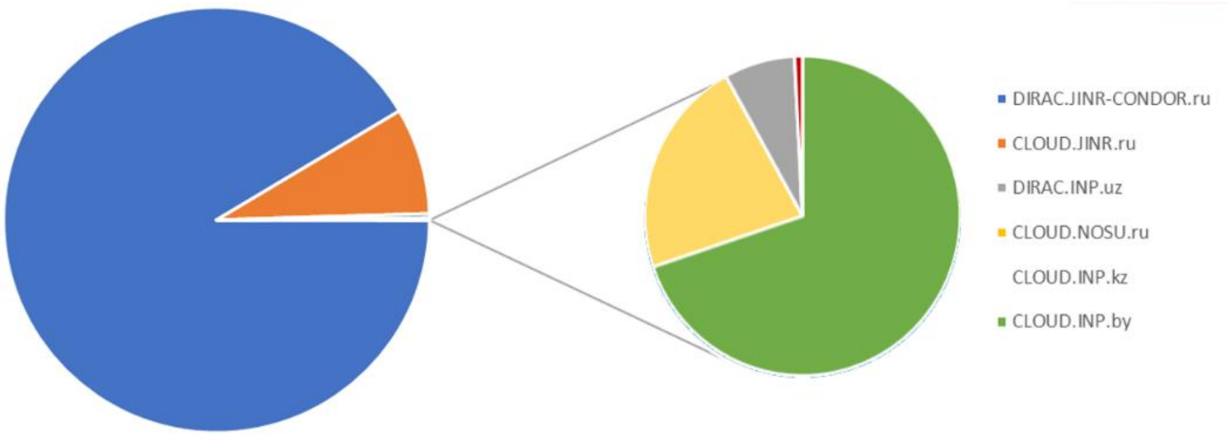

Fig.10. Distribution of the number of jobs completed in the JINR DICE by participants in 2022

In 2020-2021, idle resources of the JINR DICE were involved in research on the SARS-CoV-2 virus within the Folding@Home platform.

The service that provides access to the MICC resources for carrying out a wide range of scientific calculations via a problem-oriented web interface [9], which ensures extended capabilities for launching jobs and notifying the user about the job status, was developed and improved in the JINR cloud.

In addition, a number of auxiliary cloud services for different JINR scientific experiments were deployed on top of the JINR cloud: several documentation storage systems for experiments and user groups (ad-docs.jinr.ru for the VBLHEP Accelerator Department, spd-docs.jinr.ru and bmn-docs.jinr.ru for the SPD and BMN experiments, respectively, baikal-docs.jinr.ru for the Baikal-GVD experiment, as well as neutrino-docdb.jinr.ru for joint use by the participants of the JINR neutrino program); the first version of the ad-operations.jinr.ru electronic journaling system was realized and implemented for the VBLHEP Accelerator Department.

Heterogeneous infrastructure

Resource-intensive massively parallel computing and work with Big Data are provided by the JINR MICC heterogeneous infrastructure represented by the HybriLIT platform, which involves the education and testing polygon, the ML/DL/HPC ecosystem and the "Govorun" supercomputer, driven by a common software and information environment. The "Govorun" supercomputer is the key computing component of the HybriLIT platform and has an innovative hyperconverged software-defined architecture with unique properties for customization flexibility for the user's task [10]. The "Govorun" supercomputer comprises a GPU component, a CPU component and a hierarchical data processing and storage system with a read/write speed of 300 Gb/s.

In 2020, the development and implementation of an ecosystem for machine/deep learning and high-performance computing (ML/DL/HPC ecosystem) into this environment were completed; the ecosystem is actively used to create algorithms based on neural network approaches for solving applied tasks [11].

Since 2021, an information and computing system (ICS) has been developed on the platform at a very rapid rate to solve tasks related to the calculations of electron shells of superheavy elements. The ICS encompasses the computing resources of the "Govorun" supercomputer and a set of IT solutions and software required for modeling electron shells. The created system enables the solution of tasks of different types, with distinct requirements for the amount of computing resources, the amount of data and the requested speed of access to them. The ICS itself is based on the on-demand computing system created on the "Govorun" supercomputer, which contains 288 physical cores (576 logical cores) and a 7 TB file storage managed by the NFS file system. Intensive computing using the AMS and DIRAC software was carried out on this system to calculate the electronic properties of superheavy elements. For the development of quantum algorithms implemented on quantum computing simulators, a number of quantum simulators (QuEST, Qiskit, CuQuantum), capable of working on different computing architectures, were implemented in the ICS.

Another area of research, in which the resources of the "Govorun" supercomputer are involved, is the creation of a unified scalable research supercomputer infrastructure based on the National Research Computer Network of Russia (NIKS) at the end of 2021. At present, in addition to the "Govorun" supercomputer, this infrastructure combines the supercomputers of the Interdepartmental Supercomputer Center of the Russian Academy of Sciences and Peter the Great St. Petersburg Polytechnic University. The created infrastructure allows the participants to enlarge their local computing power, to provide access to the means for storing and processing large data volumes, to distributed data storages (data hubs), as well as to utilize each other's capacities in the case of peak loads. Such an infrastructure is in demand primarily for the tasks of the NICA megascience project. In 2022, the first joint experiment on the use of the unified supercomputer infrastructure for the tasks of the NICA megascience project was successfully completed. In total, 3,000 Monte Carlo data generation and event reconstruction tasks were launched for the MPD experiment. As a result, about 3 million events were generated and reconstructed. The obtained data was transferred to JINR for further processing and physics analysis.

Based on the rapid development of IT technologies and user requests, the supercomputer was modernized in November 2019, i.e., the transition to new Intel® Xeon® Scalable gen 2 (Intel® Xeon® Platinum 8268 models) processors and novel Intel® SSD DC P4511 high-speed solid-state disks with the NVMe interface and a capacity of 2 TB was carried out. As a result of the modernization, the performance of the CPU component increased three times, and the total peak performance of the supercomputer reached 860 TFlops for double-precision operations and 1.7 PFlops for single-precision operations, which in turn allowed the CPU component of the "Govorun" supercomputer to take 10th place in the TOP50 list of the most powerful supercomputers in Russia and the CIS.

The CPU component of the supercomputer is implemented on the high-density architecture "RSK Tornado" with direct liquid cooling, which ensures a high density of compute nodes, i.e., 150 nodes per computing rack, and high-energy efficiency about 10 GFlops/WB. The average annual PUE indicator of the system, reflecting the level of energy efficiency, is less than 1.06.

In 2020, to increase the efficiency of solving user tasks, as well as to expand the efficiency of the utilization of both the computing resources and data storage resources on the "Govorun" supercomputer, an approach to the management of computing and data storage resources, namely, resource orchestration, was elaborated and implemented. This notion means the software disjunction of a compute node, i.e., the separation of compute nodes and data storage elements (SSDs) with their subsequent integration in accordance with the requirements of user tasks. Thus, the computing elements (CPU cores and graphics accelerators) and data storage elements (SSDs) form independent fields. Due to orchestration, the user can allocate for his task the required number and type of compute nodes (including the required number of graphics accelerators), the required volume and type of data storage systems. After the task is completed, the compute nodes and storage elements are returned to their corresponding fields and are ready for the next use. This feature allows one to effectively solve user tasks of different types, to enhance the confidentiality level of working with data and avoid system errors that occur when crossing the resources for different user tasks.

To work with Big Data, including for the NICA megaproject, a hierarchical data processing and storage system with a software-defined architecture was developed and implemented on the "Govorun" supercomputer [12]. According to the speed of accessing data, the system is divided into layers that are available for the user's choice. The fastest layer of the hierarchical system is based on the latest DAOS (Distributed Asynchronous Object Storage) technology. DAOS was deployed on eight nodes of the "Govorun" supercomputer and demonstrated a high read/write speed, ranking 16th in the "10 node challenge" nomination in the 2021 edition of the IO500 list (https://io500.org/list/isc21/ten). Great prospects for the use of this technology are associated with its application to the NICA project at all stages of its work, from experimental data acquisition to final physics analysis (Fig. 11).

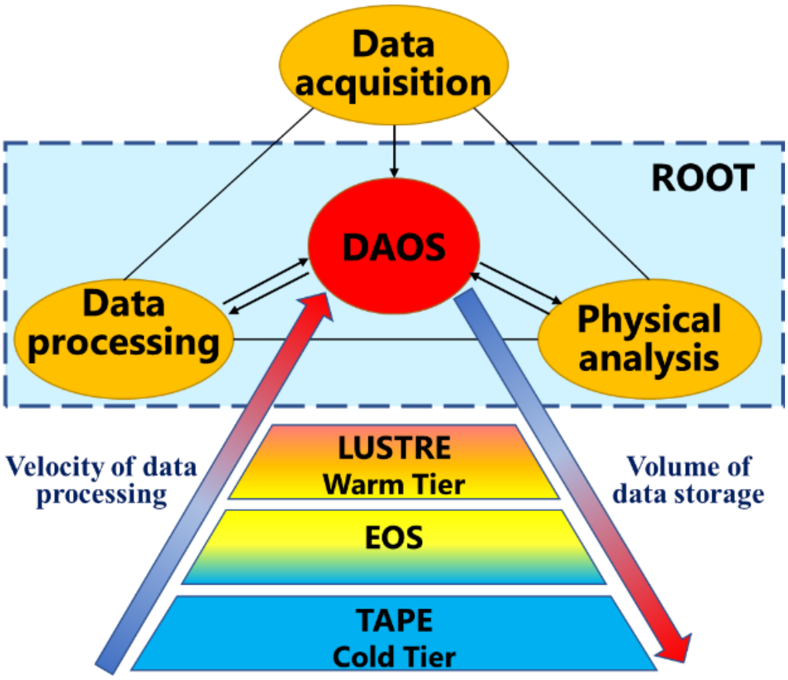

Fig. 11. Hierarchical data processing and storage system of the "Govorun" supercomputer

The use of DAOS for the NICA project will enable the storage and read of multidimensional data structures of TB size in a single address space. Moreover, DAOS technology looks promising in applications to other types of tasks related to Big Data. These are primarily ML/DL tasks and quantum computing [13, 14].

The operation of the first stage of the "Govorun" supercomputer made it possible to perform a number of complex resource-intensive calculations in the field of lattice quantum chromodynamics, to qualitatively increase the efficiency of modeling the dynamics of collisions of relativistic heavy ions, to speed up the process of event generation and reconstruction for conducting experiments within the NICA megaproject implementation, to carry out computations of the radiation safety of JINR experimental facilities and enhance the efficiency of solving applied tasks [15].

In November 2022, the next stage of the "Govorun" supercomputer modernization, which is associated with the expansion of the CPU component, implemented as part of a hyperconvergent approach to building a computing complex, underlying the "Govorun" supercomputer, took place. As a result of the modernization, the CPU component was extended to 32 hyperconverged compute nodes. Each node contains two Intel Xeon Platinum 8368Q processors (frequency 2.6 GHz, 38 cores, cache 57 MB, TDP 270 W), eight DDR4 RAM modules (256 GB), eight Intel Optane DC Persistent Memory modules (2 TB), four EDSFF E1.S NVMe SSDs (16 TB) and an M.2 NVMe SSD with a capacity of 128 GB. In addition, each node is equipped with two 100 Gb/s Intel Omni-Path adapters (Fig. 12).

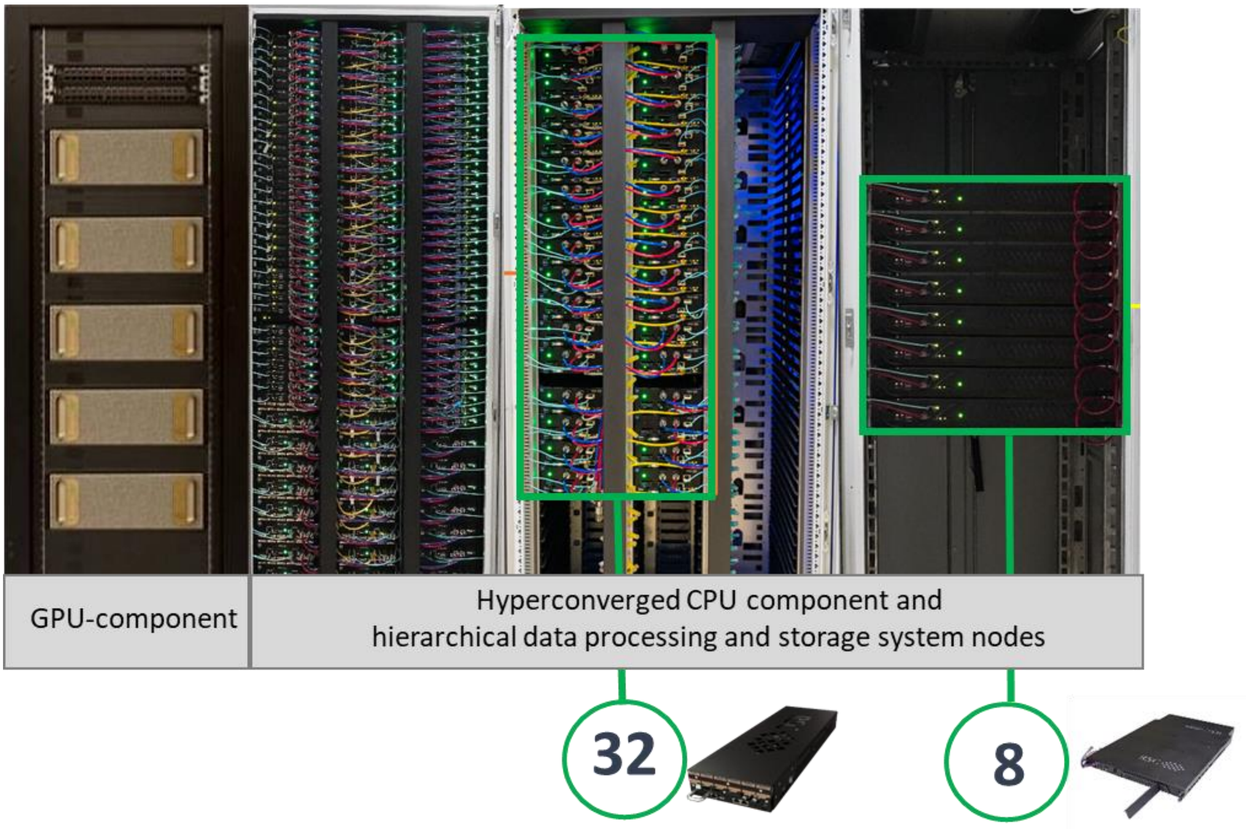

Fig. 12. Location of the new nodes of the "Govorun" supercomputer

As a result of the modernization of the CPU component, the performance of the "Govorun" supercomputer increased by 239 TFlops, and the volume of the hierarchical data processing and storage system of the "Govorun" supercomputer was enlarged by 1.6 PB for the "very hot data" DAOS layer and by 8 PB for the "warm data" layer. Consequently, the performance of the "Govorun" supercomputer enhanced by 23.5% and reached 1.1 PFlops, and the total capacity of the hierarchical storage increased to 8.6 PB.

The hyperconvergence of the new compute nodes already enabled their use for data mass generation and reconstruction tasks within the NICA MPD experiment. It is noteworthy that for a number of MPD tasks, there was a need for a large amount of RAM, which is satisfied by the new nodes. The computing resources and the hierarchical data processing and storage system of the "Govorun" supercomputer were integrated into a DIRAC-based distributed heterogeneous environment that includes the resources of JINR and its Member States.

The above modernization of the "Govorun" supercomputer will make it possible to speed up studies in the field of lattice quantum chromodynamics, to qualitatively increase the efficiency of modeling the dynamics of relativistic heavy-ion collisions, to carry out calculations of the radiation safety of JINR experimental facilities and enhance the efficiency of solving applied tasks. The modernized "Govorun" supercomputer makes it possible not only to perform computing, but also to use the supercomputer as a research polygon for developing software-hardware and IT solutions for JINR tasks.

The results obtained using the resources of the "Govorun" supercomputer from the moment it was put into operation in July 2018 to April 2023 are reflected in more than 240 user publications, two of them in the Nature Physics journal.

Thus, the operation experience of the "Govorun" supercomputer has shown the relevance and effectiveness of using both novel hyperconverged computing architectures and the hierarchical data processing and storage system being part of it. At present, the resources of the "Govorun" spercomputer are used by scientific groups from all the Laboratories of the Institute within 25 themes of the JINR Topical Plan. The number of users of the "Govorun" supercomputer is 323 people, of which 262 are JINR staff members, and 61 are from the Member States. Access to the supercomputer resources is provided only to those users who are directly involved in the implementation of the JINR Topical Plan.

Integration of computing resources

A significant feature of the created infrastructure is the integration of distributed computing resources [16]. For five years now, the integration of heterogeneous distributed computing resources on the basis of the DIRAC Interware platform has been functioning at JINR. The DIRAC Interware is a product for combining heterogeneous computing and storage resources into a unified platform, based on the use of standard data access protocols (xRootD, GridFTP, etc.) and pilot jobs.

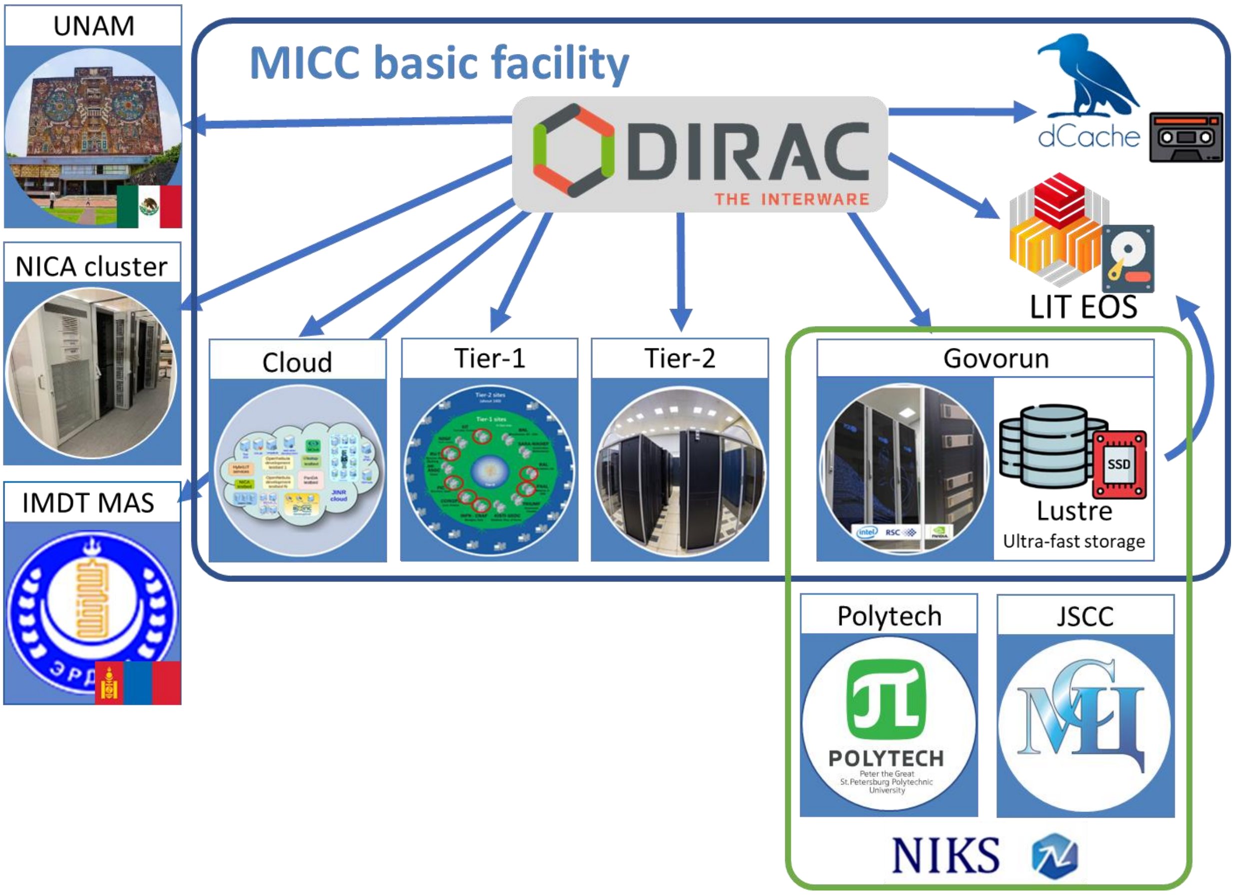

Fig. 13. Heterogeneous distributed environment based on the DIRAC platform

In August 2019, the first set of Monte-Carlo data simulation jobs for the MPD experiment was sent to the resources of the Tier1 and Tier2 grid clusters via DIRAC. Then the "Govorun" supercomputer and the JINR DICE were integrated into the distributed computing platform (DCP). In the summer of 2020, the NICA cluster and the cluster of the National Autonomous University of Mexico (UNAM) were added. It is noteworthy that the UNAM cluster became the first computing resource located outside Europe or Asia and included in the DIRAC infrastructure at

JINR. dCache, which manages disk and tape storages, EOS and the hierarchical data processing and storage system of the "Govorun" supercomputer were integrated as storage systems [17]. In 2022, the computing cluster of the Institute of Mathematics and Digital Technologies of the Mongolian Academy of Sciences (IMDT MAS) and NIKS (National Research Computer Network, the Russia's largest research and education network) were integrated into the heterogeneous distributed environment based on the DIRAC platform (Fig. 13).

Thanks to the integration via DIRAC, it became possible to use the largest amount of computing resources for centralized data generation by the Monte Carlo method for the MPD experiment [18]. The experience of using different computing resources of JINR and other MPD collaboration institutes has shown that at present, the use of the "Govorun" supercomputer resources is the most efficient (Fig. 14).

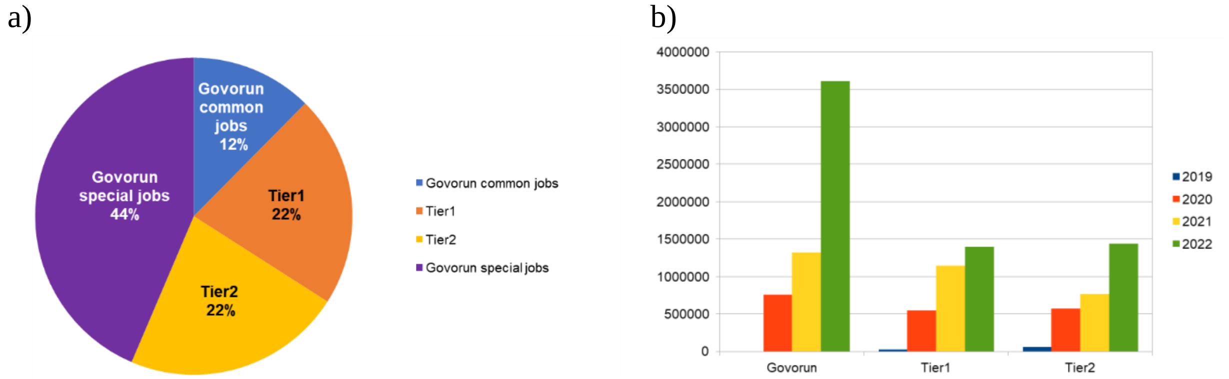

Fig. 14. DIRAC-based distributed heterogeneous environment for MPD tasks: a) share of use of different MICC components in 2022; b) increase in the share of the MICC computing resources in normalized CPU HEP-SPEC06 days

Figure 14a shows the diagram of the use of different MICC components in the DIRAC-based distributed heterogeneous environment for MPD jobs. At the same time, the share of those jobs that can only be calculated on the "Govorun" supercomputer was about 44% in 2022, and the total contribution of the "Govorun" supercomputer resources to event mass generation and reconstruction for the MPD experiment was about 56%. Figure 14b illustrates the increase in the share of the MICC computing resources in the heterogeneous environment for MPD jobs. The sharp increase in the share of the "Govorun" supercomputer is related to the modernization carried out in 2022; thus, over 50 million events were generated on the new nodes.

To simplify the access of MPD collaboration members to the computing resources, a special application was developed; it enables the description of the data generation job in physical terms, i.e., collision energy, type of the generator used, target and beam material [19]. The developed application was integrated into the web interface of the DIRAC platform at JINR and can be accessed by all MPD users.

To date, thanks to resource integration using DIRAC, 2.1 million jobs have been completed on the distributed platform. The number of calculations performed is estimated at 13.5 million HEP-SPEC06 days, which is equivalent to 2,000 years of computing on one core of the central processor. 90% of computing were performed on the "Govorun" supercomputer, Tier1 and Tier2, 6% was carried out on the JINR cloud infrastructure, 3% was done on the NICA cluster.

The major user of the distributed platform is the MPD experiment, whose share accounts for 90%. Using DIRAC, a program of mass data simulation runs within the MPD experiment is performed. More than 1.283 billion events were successfully modeled with the help of UrQMD, GSM, 3 Fluid Dynamics, vHLLE_UrQMD and other generators; 440 million events were subsequently reconstructed. The total amount of data received exceeds 1.3 PB. 5% of computing was performed by the Baikal-GVD experiment, 3% was carried by Folding@Home, 2% was done by BM@N.

Monitoring system

The developed integrated monitoring system of the MICC allows one to receive information from different components of the computing complex: the engineering infrastructure, networks, compute nodes, task launch systems, data storage elements and grid services, which guarantees a high level of reliability of the MICC (Fig. 15). In 2020, the cloud infrastructure was connected to the common monitoring system. Within the development of the resource monitoring system for the Tier1 and Tier2 grid sites, a new accounting system was created at JINR in 2021 [20]; it enabled to significantly expand the functionality of the original system, as well as to reduce the time of obtaining statistical data due to the creation of automatic data processing by the visualization system; it also allows collecting statistical data on the use of resources by user jobs (for any time interval), namely, the astronomical job execution time and CPU time in HEP-SPEC06 hours, the number of jobs and efficiency of using the computing cluster.

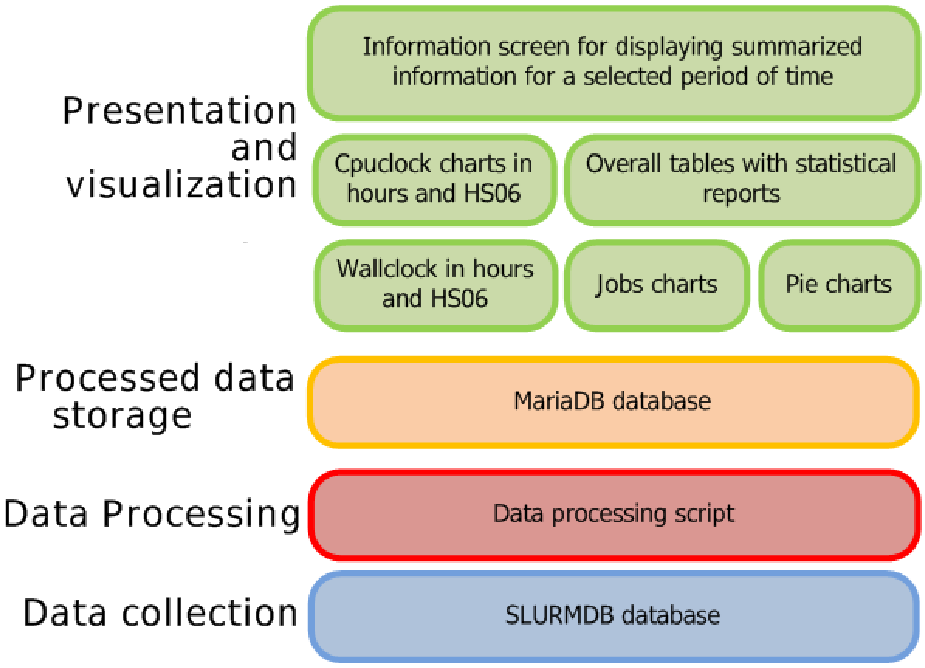

Fig. 15. Scheme of operation of the MICC monitoring system

The implemented approach provides the statistical data display directly from SLURM (Simple Linux Utility for Resource Management) and allows accounting for the resources and their use both within the distributed data processing system and locally. The visualization system yields a powerful tool for analyzing and compiling different reports and presentations. It is also noteworthy that the accounting system was integrated into the general MICC monitoring system, i.e., LITMon [21]. This made it possible to organize a single entry point and combine disparate accountings into a unified structure.

Applied research on the MICC

The studies, carried out in collaboration with FLNP within the international program “UNECE International Cooperative Program (ICP) Vegetation” for monitoring and predicting air pollution processes in Europe and Asia, were in progress [22]. A mobile application, which allows automatically filling in information about places of moss samples in accordance with the standards of UNECE ICP Vegetation, was elaborated. The application is integrated with the system of managing data of the ICP Vegetation program (DMS), developed at MLIT on the basis of the cloud platform. The DMS is intended to provide its participants with a modern unified system of collecting, analyzing and processing biological monitoring data. A combination of satellite data, biomonitoring measurements, and different machine and deep learning technologies was used to predict potentially toxic elements. The possibility of using Earth remote sensing data together with machine learning methods for predicting air pollution by heavy metals was investigated [23]. The average accuracy of the models exceeded 89%. Models of pollution by aluminum, iron and antimony in the central region of Russia were constructed.

In 2020, an information system (IS) for the tasks of radiation biology was developed [24] within a joint project of MLIT and LRB using the ML/DL/HPC ecosystem of the HybriLIT platform. The IS is aimed to store experimental data and analyze changes in the central nervous system of mammals on the basis of molecular, pathomorphological and behavioral changes in the mammalian brain when exposed to ionizing radiation and other factors (Fig. 16).

Algorithms for experimental data processing based on machine and deep learning were implemented into the developed system. The IS comprises reliable modern means of authentication and hierarchical delimitation of access to data, a data storage system, as well as components for convenient work and the visualization of data analysis results.

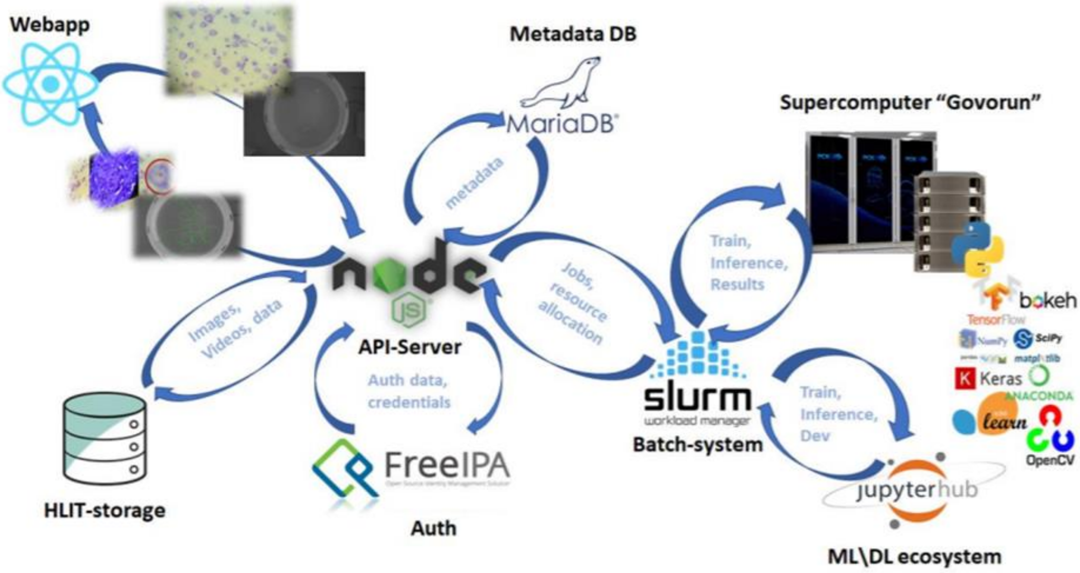

Fig. 16. Architecture of the information system for radiation biology tasks

Within the ML/DL/HPC ecosystem, using the example of solving a specific problem to investigate the dynamics of magnetization in a Josephson ϕ0-junction, a methodology for developing software modules on top of JupyterHub [25], which enable not only to carry out calculations, but also to visualize the results of the study and accompany them with the necessary formulas and explanations, was presented (Fig. 17). A parallel implementation of the algorithm for performing computing for different values of the model parameters based on the Joblib Python library and modules with the integration of the Matlab code into Jupyter Notebook, which make it possible to effectively carry out applied computations for image analysis, were developed.

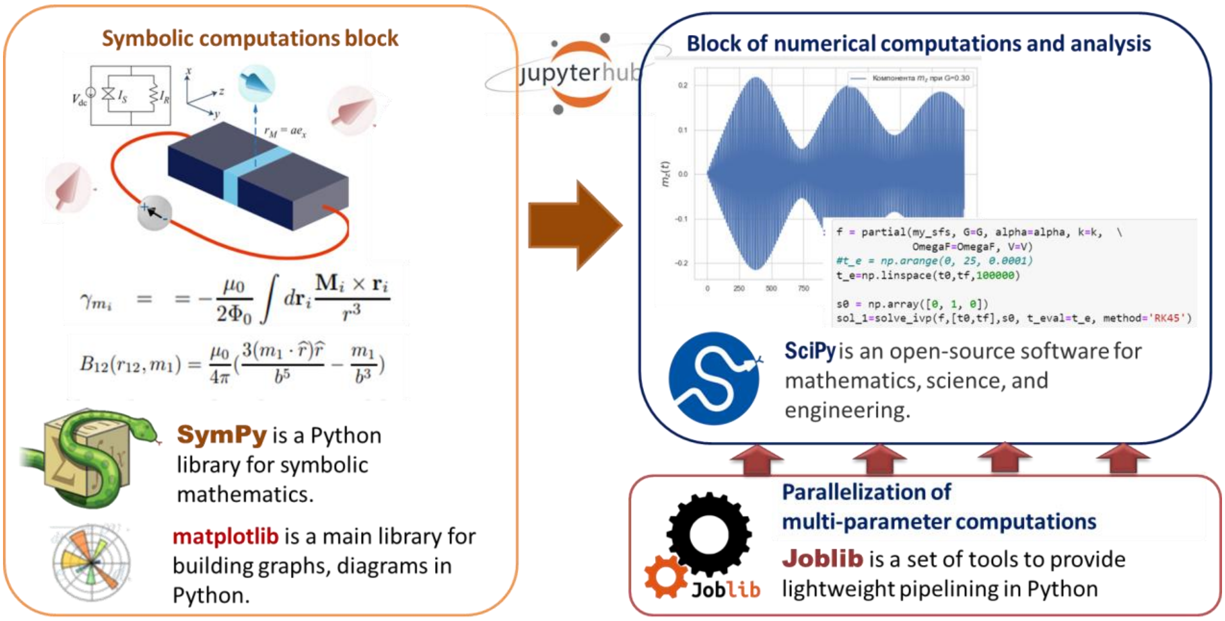

Fig. 17. Scheme of software modules for studying systems with Josephson junctions

A software complex that enables the simulation of a distributed computing system for acquiring, storing and processing data from the BM@N experiment of the NICA project under different scenarios for launching jobs for the next run was developed [26] The purpose of the modeling was the optimal distribution of the flows of primary data processing for the BM@N experiment to the compute nodes in order to minimize hardware downtime during job execution. Based on the simulation results, one can predict the load of the compute nodes and telecommunication channels.

A service for planning and accounting for excursions to JINR, i.e., https://jinrex.jinr.ru, was developed together with the UC [27]. The main functions of the service are as follows: saving information on excursions; excursion coordination; demonstration of the workload of visiting points; collecting analytics and demonstration of statistics.

Information and software support of the JINR research and production activity

During the reporting period, the work on the maintenance and development of the JINRLIB program library was in progress. The library was replenished with the following programs developed by MLIT specialists: EORP 2020, a program for computing closed equilibrium orbits (Authors: I.V. Amirkhanov, I.N. Kiyan); Split, a parallel implementation of the numerical solution of a system of algebraic equations with a tridiagonal matrix using the partition algorithm and MPI technology (Author: A.V. Volokhova); SIR-model, the simplest epidemic process model (Author: V.S. Rikhvitsky); RK4-MPI, a parallel implementation of the numeral solution of the Cauchy problem by the fourth-order Runge-Kutta method using MPI technology (Author: A.V. Volokhova); INQSIM, a program for converting PI-type fully symmetric quadrature rules on 2-,..., 6-simplexes from compact to expanded forms (Authors: G. Chuluunbaatar, O. Chuluunbaatar, A.A. Gusev, S.I. Vinitsky); FITTER_WEB, a program for fitting experimental data obtained on a small-angle neutron scattering spectrometer, implemented as a web application (Authors: A.G.

Soloviev, T.M. Solovjeva). At present, the JINRLIB library has over 65 software packages, most of which solve the tasks of experimental data processing automation and computational mathematics.

In 2020-2023, a number of works on the development and current maintenance of the “Dubna” electronic document management system (EDMS) were completed. In particular, a module for maintaining the procurement plan was worked out, the ability to sign invoices for payment using an enhanced electronic signature was implemented, a subsystem for the automated formation of supply contracts on the basis of standard forms was developed, a module for monitoring, electronic archive storage and search for supply contracts was elaborated. “Advance Reports” was developed and put into operation; it is designed for electronic recording in the JINR Accounting Department to receive prepayment and report on business trips and household expenses. The system enables the preliminary loading of the required data to accelerate the work of the Accounting Department and generates a number of financial reports. New documents “Instruction” and “Request” with the possibility of automatic control of deadlines, the document “Strabag (KS-2, KS-3, invoice)”, “SEDNPSS archive” for storing the technical documentation of the Scientific and Engineering Department of Nuclotron Power Supply Systems of VBLHEP were introduced. For the further development and optimization of electronic document management, the EDMS software was configured for the process of coordinating technical specifications and design assignments for the development of design and/or working documentation for all capital investment objects of the Institute, for the document management of accounts and acts of work performed under budget items 18 “Design works” and 19 “Construction of buildings and technological systems”, for the document management of accounts and acts of work performed KS-2, KS-3 under budget item 14, paragraphs a, b, etc. A number of works on the adaptation of the “Dubna” EDMS to accommodate the changes in the organization of procurement procedures at JINR were carried out. The protection of the EDMS and other systems from hacker attacks was significantly enhanced through the use of pro-active methods for diagnosing attacks and protecting against them.

The work on the current maintenance and development of the APT EVM project management system for NICA was in progress. Specifically, the integration of Cost Book data with the procurement plan in the “Dubna” EDMS was implemented. The ongoing maintenance and development of the following information systems upon user requests: HR LHEP, ADB2, PIN, ISS, INDICO, Document Base and Electronic Photo Archive, were performed.

A new version of the CERN DB information system for registering business trips at CERN, managing accommodation and accounting financial expenses was developed and put into operation.

The work on enhancing and applying the WALT (Web Application Lego Toolkit) platform, which is a template-oriented platform designed for developing web applications of different complexity, was in progress [28]. In contrast to many other platforms that are “magic black boxes”, the WALT underlying idea is to provide transparent, extensible and modifiable tools to solve specific tasks arising in the development of web applications. The WALT platform is used to develop JINR’s corporate web applications such as ADB2, PIN, NICA EVM, EDMS, etc. (Fig. 18).

ADB2	JINR's management accounting
PIN	Staff information
EDMS "Documents DB"	Electronic storage for administrative activity documents
NICA EVM	Project structure, workplans, expenses, Costbook, reports, etc.
EDMS "Dubna"	JINR electronic document management system
HR JINR	Staff administrative information
MAP JINR	Basic map of JINR's sites
Gateway	Universal gateway for data exchange between various systems
ISSC	Scientific attestation support system
CERN DB	JINR's staff at CERN: trips, accommodation, reports, etc.
EDMS "Advance reports"	Data preparation for business trip accounting reports
Checkpoint lists	Lists for access to JINR's sites in a limited access mode
DES	JINR Digital EcoSystem that integrates existing and future services

Fig. 18. JINR's corporate web applications developed using WALT

In 2020, a personal account with the possibility of online payment for tenants of the Institute's housing stock was developed and put into operation. Together with the electronic document management system, a system for processing invoices in an electronic form, for the signing of which an internal certification center was implemented, was created. In the personnel system, a subsystem for accounting electronic employment record books was created, and the system for special job assessment was completely revised.

The Institute's management reporting was improved; as part of the development of project management, a corresponding module, which allows one to track the work progress, to draw a Gantt chart and assign financial resources, was elaborated in the 1C program.

The JINR Information System for Scientific Certification (ISSC) (https://dissertations.jinr.ru/) and the "Visit Centre" portal (https://visitcentre.jinr.ru/) were maintained. The modernization and support of the web site of the PEPAN and PEPAN Letters journals (http://pepan.jinr.ru) continued. The traditional development, creation and maintenance of web sites of conferences, symposia at the request of the Laboratories and other JINR subdivisions were in progress.

The maintenance of the following servers and systems of general use was conducted: the infrastructure of site hosting (www.jinr.ru, flnph.jinr.ru, flerovlab.jinr.ru, micc.jinr.ru, mpdroot.jinr.ru, etc.), the infrastructure of administrative servers (pin.jinr.ru, adb2.jinr.ru, sed.jinr.ru, etc.), the automated project management system (pm.jinr.ru) and the cloud storage service for the JINR staff (disk.jinr.ru).

A prototype of the MLIT Information and Analytical System for Maintaining Licenses (IAL, http://soft-lit.jinr.ru) was developed and put into trial operation [29]. The IAL main task is the automation of the management, purchase, maintenance and use of licensed software products (LSP), asked by the need to both plan and optimize the purchase of licensed software and the necessity of controlling the compliance with the rules of the licensing policy. The system database contains all the information on the purchased licenses, on the basis of which the user's personal account (PA) is formed. The PA stores data about licenses owned by the user. The Administrator/Auditor PA is designed for LSP management and monitoring. Login to the IAL is made through the JINR SSO user authentication system. The web interface is implemented in the WALT development environment using the LegoToolkit web application.

In 2022, the work on the creation of the Digital JINR platform started (Fig. 19); its main purpose is to provide a unified environment for the creation and development of digital services, their integration with each other and the analysis of information on all aspects of JINR's activities. The Digital EcoSystem will encompass a wide range of services, from resources for users of basic facilities to handling business trips, vouchers, ordering certificates, etc. The major groups of services are administrative (the area of responsibility of DDSD JINR) and scientific. At present, a prototype of the Digital EcoSystem single access point (https://eco.jinr.ru), a number of network services, a telephone directory and some others were implemented. Access to the system is based on the JINR Single Sign-On (SSO) authentication service.

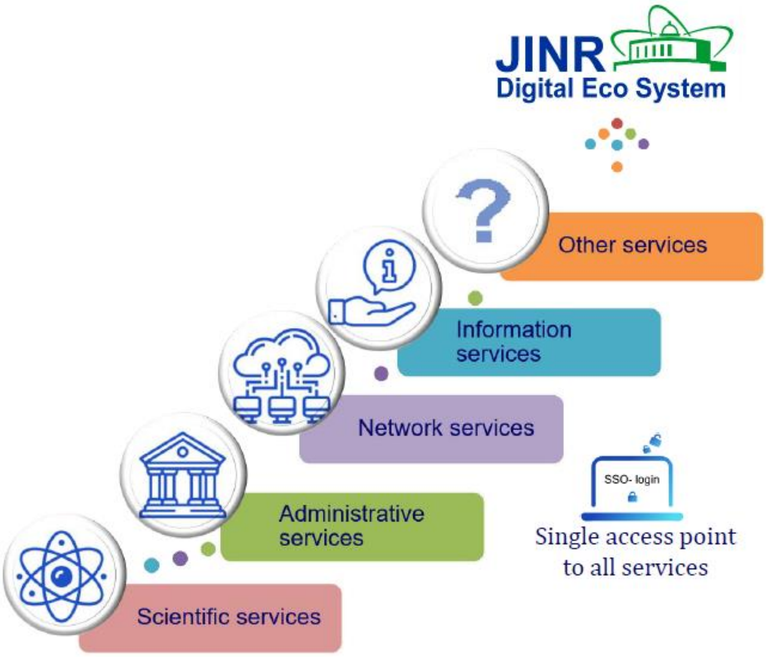

Fig. 19. JINR Digital EcoSystem

Educational program on the education and testing polygon

An important aspect of the activity that involves the resources of the HybriLIT platform is the educational direction related to conducting both training courses for JINR staff members and practical classes for students of Dubna State University, Tver State University and other universities. In 2020, tutorials and master classes were remotely held for students from the Czech Republic and Armenia.

In the reporting period, practical classes on "Architecture of Computer Systems", "High-performance Computing Technologies", "Modern Methods for Analyzing Complex Systems", "Machine Learning and Data Mining", "Languages and Technologies for Data Analysis", "Mathematical Apparatus and Tools for Data Analysis" were held using the resources of the education and testing polygon and the ML/DL/HPC ecosystem, which enabled students to master state-of-the-art technologies for developing parallel algorithms on novel computing hybrid architectures and tools (libraries and frameworks) for machine and deep learning tasks.

The platform resources were also actively used to train IT specialists within the International School of Information Technologies "Data Science" [30], whose students are engaged in real scientific projects of JINR. The results of graduates are presented in a collection of scientific and project activity reports [31].

Since 2019, JINR, together with North Ossetian State University, has been holding the regular IT School for Young Scientists “Distributed Computing and Data Science”. The IT School gathers more than 60 students and teachers from universities of the South of Russia (North Ossetia, Kabardino-Balkaria, Chechnya). Within the School, JINR scientists deliver lectures on JINR’s scientific projects, information technologies and solutions for scientific tasks being developed at JINR; master classes and training courses on distributed computing, virtualization and cloud technologies, machine and deep learning algorithms for analyzing data of a complex structure are conducted.

The resources of the platform were also actively used during the first JINR Autumn School of Information Technologies, which was held on 14-19 November 2022 at MLIT. School participants attended training courses on topical issues in the field of distributed and high-performance computing, machine learning and artificial intelligence, mathematical modeling, modern methods and technologies for data processing and analysis. During the second stage of the School, i.e., the JINR Spring School of Information Technologies (May 2023), students will present the results of joint work with JINR specialists on selected topics of graduation theses.

In 2020-2023, four PhD theses and more than 80 master’s and bachelor’s theses were prepared using the resources of the HybriLIT platform.

Conferences

On 5-9 July 2021, the **9th International Conference “Distributed Computing and Grid Technologies in Science and Education” (GRID’2021)** was held in the Meshcheryakov Laboratory of Information Technologies in a mixed format. The GRID conference, which takes place every two years, was dedicated to the 65th anniversary of JINR and the 55th anniversary of the foundation of the Laboratory of Computing Techniques and Automation (LCTA, now MLIT).

The conference traditionally attracted a large community of Russian and foreign experts ready to discuss emerging challenges and prospects related to the use and development of distributed grid technologies, heterogeneous and cloud computing in different fields of science, education, industry and business. In 2021, the list of conference topics was complemented with quantum computing.

More than 270 scientists (103 – in person, over 170 – remotely) from research centers of Armenia, Belarus, Bulgaria, China, the Czech Republic, Egypt, France, Germany, Georgia, Iran, Italy, Moldova, New Zealand, Poland, Romania, Slovakia, Sweden and Switzerland took part in the conference. Russia was represented by participants from 28 universities and research centers. The conference was organized in ten sessions, where issues associated with the development of distributed computing technologies, cloud technologies, heterogeneous computing, volunteer computing and Big Data analytics, machine learning and quantum information processing were discussed. 23 plenary and 140 sessional talks were delivered.

Round tables organized within the conference were dedicated to the use of IT in education; to the Russian segment of the WLCG (Worldwide LHC Computing Grid), i.e., RDIG; to supercomputer technologies. At one of the round tables Intel presented its new product, the DAOS high-speed file system.

A number of plenary talks at the conference were made by representatives of the IT industry, who were sponsors of the conference. Among them were IBS Platformix, IT Cost, Niagara Computers, Dell EMC, RSC Group, Intel, Softline.

The presentations of the talks and photos are available at https://indico.jinr.ru/event/1086/. Selected proceedings of the conference were published in CEUR Workshop Proceedings (CEUR-WS.org) – https://ceur-ws.org/Vol-3041/.

On 29-31 March 2022, the Meshcheryakov Laboratory of Information Technologies hosted the international scientific conference **"Parallel Computational Technologies (PCT) 2022"**, the sixteenth in a series of annual conferences dedicated to the development and application of parallel computing technologies and machine learning in versatile areas of science and technology. The conference is organized by the Ministry of Science and Higher Education of the Russian Federation and the Supercomputer Consortium of Russian Universities.

More than 110 scientists from Belarus, Brazil, Egypt, Mongolia, Romania, Slovakia took part in the conference. Russia was represented by participants from 40 universities, research centers, IT and industrial companies. The conference was organized in nine sessions, where issues associated with the application of cloud, supercomputer and neural network technologies in science and technology, including applications, hardware and software, specific models, languages, libraries and packages, were discussed. 7 plenary, 38 sessional and 10 poster talks were delivered.

On working days of the conference, there was organized a supercomputer exhibition, where RSC Group and Karma Group presented their latest developments in the field of high-performance computing.

The presentations of the talks and photos are available at the conference website http://agora.guru.ru/pavt2022/. Selected proceedings of the conference were published in Springer's Communications in Computer and Information Science series and in the Computational Mathematics and Software Engineering series of the Bulletin of South Ural State University.

Publications

In 2020-2023, MLIT specialists co-authored 223 publications of the CMS collaboration, 34 publications of the Baikal-GVD collaboration and other collaborations, including NOvA, DUNE, PANDA, etc. More than 70 articles were published in refereed journals. Selected papers are presented below.

[1] *Korenkov V.V.* The JINR Multifunctional Information and Computing Complex// Nuclear Physics, V. 83, No. 6, pp. 534–538, 2020, in Russian; *V. Korenkov* // 2020 International Scientific and Technical Conference Modern Computer Network Technologies (MoNeTeC), Moscow, Russia, 2020 pp. 1-4, doi: 10.1109/MoNeTeC49726.2020.9258311

[2] *Baginyan A., et al.* JINR Network Infrastructure for Megascience Projects // 2020 International Scientific and Technical Conference Modern Computer Network Technologies (MoNeTeC), Moscow, Russia, 2020, pp. 1-5., doi: 10.1109/MoNeTeC49726.2020.9258004

[3] *Velikhov V.E., et al.* // 2022 Intern. Sci. and Techn. Conf. "Modern Network Technologies" (MoNeTec), Moscow, 2022. P. 1–4.

[4] *Baginyan A., et al.* Current Status of the MICC: an Overview // CEUR Workshop Proc. 2021. V. 3041. P. 1–6.

[5] *Baranov A., et al.* Lifecycle Management Service for the compute nodes of Tier1, Tier2 sites (JINR)// CEUR Workshop Proc. 2021. V. 3041. P. 429–433, 2021

[6] *Kutovskiy N., Pelevanyuk I., Zaborov D.* Using Distributed Clouds for Scientific Computing// CEUR Workshop Proc. 2021. V. 3041. P. 196–201, 2021.
[7] *Balashov N., et al.* JINR Distributed Information and Computing Environment: Participants, Features and Challenges // CEUR Workshop Proc. 2021. V. 3041. P. 280–284, 2021.
[8] *Balashov N., et al.* Creating a Unified Educational Environment for Training IT Specialists of Organizations of the JINR Member States in the Field of Cloud Technologies // Modern Information Technology and IT Education, V. 1201, pp. 149-162, 2020.
[9] *Balashov N.A., Kutovskiy N.A., Sokolov I.V.*, Problem-oriented interface for MICC // AIP Conference Proceedings, V. 2377, 040002, 2021, https://doi.org/10.1063/5.0064008, 2021
[10] *Belyakov D, et al.* "Govorun" Supercomputer for JINR Tasks // CEUR Workshop Proc., vol. 2772, pp. 1–12, 2020
[11] *Zuev M.I. et al.* ML/DL/HPC Ecosystem of the HybriLIT Heterogeneous Platform (MLIT JINR): New Opportunities for Applied Research // PoS (DLCP2022). 2022. V.429. P. 027.
[12] *Kokorev A., Belyakov D., Lyubimova M.*, Data storage systems of the "HybriLIT" heterogeneous computing platform for scientific research carried out in JINR: filesystems and RAIDs performance research // CEUR Workshop Proc., vol. 3041, pp. 296-303, 2021
[13] *Moshkin A.A., et al.* Approaches, services, and monitoring in a distributed heterogeneous computing environment for the MPD experiment // Russian Supercomputing Days: Proc. of the Int. Conf. 2021. P. 4–11.
[14] *Kudryavtsev A.O., Podgainy D.V., Moskovsky A.A.* // ISC High Performance 2021. May 29 – June 2, 2021. Humburg. Germany (https://www.isc-hpc.com/); *Moskovsky A.A., Brekhov A.T., Podgainy D.V., Kudryavtsev A.O.* // Sixth International Parallel Data Systems Workshop. November 15. 2021 (https://sc21.supercomputing.org/session/?sess=sess332); *Vaľa M., Podgainy D., Lavrenko P., Brekhov A.* High Energy Physics experiment data processing with DAOS in multi-tier storage environment based on RSC Storage on-Demand // The 5th annual DAOS User Group meetingю. November 19. 2021 (https://daosio.atlassian.net/wiki/spaces/DC/pages/11015454821/DUG21)
[15] *Podgainy D., et al.* IT Solutions for JINR Tasks on the "GOVORUN" Supercomputer // CEUR Workshop Proc. 2021. V. 3041. P. 612–618.
[16] *Korenkov V.V., Pelevanyuk I.S., Tsaregorodtsev A.Yu.*, Integration of the JINR Hybrid Computing Resources with the DIRAC Interware for Data Intensive Applications // Communications in Computer and Information Science, V. 1223, pp. 31-46, 2020
[17] *Pelevanyuk I.*, Performance evaluation of computing resources with DIRAC interware // AIP Conference Proceedings, V. 2377, N. 1, 040006, 2021
[18] *Kutovskiy N., et al.* Integration of distributed heterogeneous computing resources for the MPD experiment with DIRAC Interware, PEPAN, V. 52, N. 4, pp. 999-1005, 2021
[19] *Moshkin A., Pelevanyuk I., Rogachevskiy O.* Design and Development of Application Software for the MPD Distributed Computing Infrastructure// CEUR Workshop Proc. 2021. V. 3041. P. 321–325, 2021.
[20] *Kashunin I., Mitsyn V., Strizh T.* JINR WLCG Tier1 & Tier2/CICC accounting system // CEUR Workshop Proc. 2021. V. 3041. P. 285–290, 2021.
[21] *Kashunin I.A., Mitsyn V.V., Strizh T.A.* Operating Principles of the Accounting System for JINR's Grid Sites // Phys. Part. Nucl. Letters., V. 19, No. 6, pp. 820–825, 2022.
[22] *Uzhinskiy A.V.*, Intelligent platform for environmental monitoring // Open Systems, V. 2, pp. 21-23, 2021, in Russian
[23] *Uzhinskiy A., Vergel K.* Central Russia heavy metal contamination model based on satellite imagery and machine learning // Computer Optics, Vol. 47, No. 1, pp. 137-151, 2023.
[24] *Butenko Yu., Marov D., Nechaevskiy A., Podgainy D.*, Development of a Service for Conducting Radiobiological Studies on the HybriLIT Platform // CEUR Workshop Proc., V. 2743. P. 26-33, 2020

[25] *Butenko Yu., Nechaevskiy A., Rahmonv I., Streltsova O.,*Virtual Research Environment NanoHLIT for Hybrid Nanostructures Research on the HybriLIT Platform // CEUR Workshop Proc., V. 3041. P. 407-412, 2021
[26] *Priakhina D., Trofimov V., Ososkov G. and Gertsenberger K.* Data center simulation for the BM@N experiment of the NICA project // AIP Conference Proceedings, V. 2377, 040007, 2021
[27] *Ilyina A.V., Pelevanyuk I.S.,* System for planning and logging excursions at JINR, Phys. Part. Nucl. Letters., 2023 (in press)
[28] *Korenkov V., Kuniaev S., Semashko S., Sokolov I.* Walt platform for web application development // CEUR Workshop Proc. 2021. V. 3041. P. 387–392.
[29] *Alexandrov E., et al.* Development of an information-analytical system for the support and maintenance of licenses at MLIT JINR // Proceedings of the ISA RAS, 2023 (in press)
[30] *Korenkov V.V., et al.* Concept and implementation of the educational program of the International School of Information Technologies “Data Science” at Dubna State University // System Analysis in Science and Education: network scientific edition, No. 3, pp. 1–7, 2020, in Russian
[31] Collection of scientific and project activity reports of graduates of the International School of Information Technologies “Data Science”. Issues 1-3, Editors: Korenkov V.V. et al. Publisher: JINR, 2020-2022