DISTRIBUTED INFORMATION AND COMPUTING INFRASTRUCTURE OF THE JINR MEMBER STATE ORGANIZATIONS

N.A. Balashov[1], A.N. Makhalkin[1], Ye.M. Mazhitova[1,2], N.A. Kutovskiy[1], I.S. Pelevanyk[1], R.N. Semenov[1,3]

emazhitova@jinr.ru

[1] Meshcheryakov Laboratory of Information Technologies, JINR

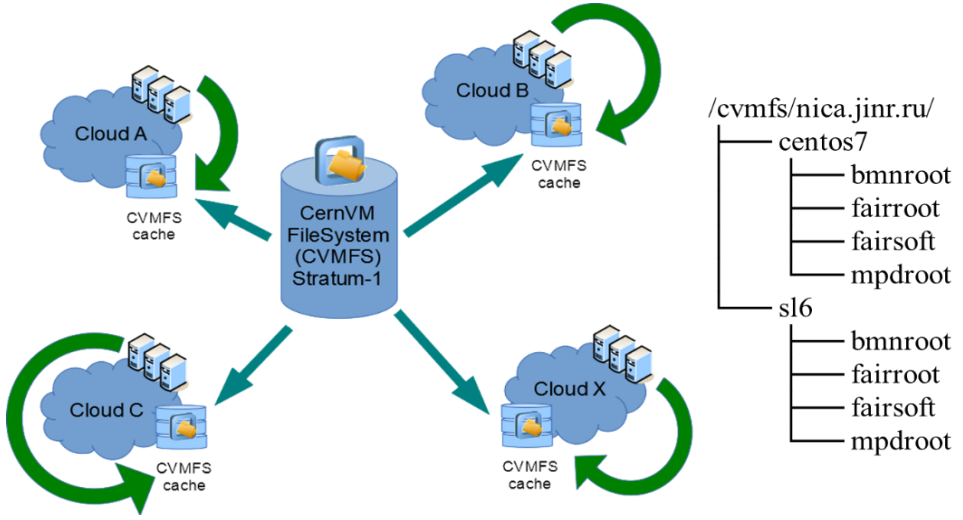
[2] Institute of Nuclear Physics, Almaty, Kazakhstan

[3] Plekhanov Russian University of Economics, Moscow, Russia

The integration of computing power of the Joint Institute for Nuclear Research (JINR) Member State organizations into a unified distributed information and computing environment (DICE) is an important and topical task, the solution of which would significantly reduce the time of research with obtaining significant scientific results. This paper describes the motivation of creating the distributed cloud environment, deployed on the basis of the resources of the JINR Laboratory of Information Technologies and some JINR Member State organizations, and the approach it is based on, as well as outlines plans for using and developing the created infrastructure.

SOFTWARE DISTRIBUTION MODEL

Software distribution in the JINR DICE is done via CernVM File System (CVMFS). CVMFS is a web-based, global, and versioning file system optimized for software distribution. The file system content is installed on a central web server from where it can be mirrored and cached by other web servers and web proxies. File system clients download data and meta-data on demand and cache them locally. Data integrity and authenticity is ensured by cryptographic hashes and digital signatures.

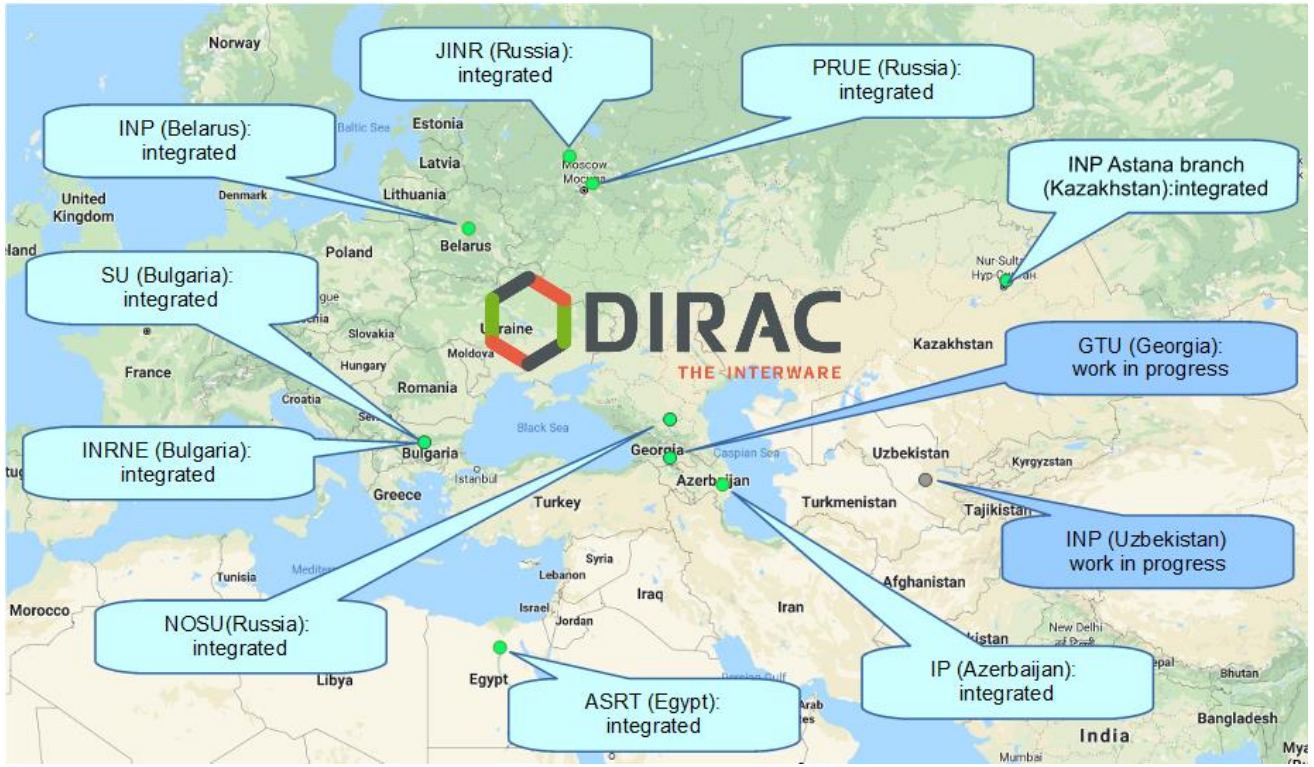


PARTICIPANTS OF THE INFRASTRUCTURE

The cloud infrastructures of JINR and its Member State organizations are based on the open source solution OpenNebula. The JINR cloud is the core of this infrastructure. It hosts DIRAC services, which manage computational tasks and data over the resources of JINR and its Member State organizations. DIRAC provides all the necessary components to build ad-hoc grid infrastructures that interconnect computing resources of different types (*computational grids, clouds or clusters*), enabling interoperability and simplifying interfaces.

At the moment, the integration of clouds of organizations from the JINR Member States is at different stages, in particular:

1. Plekhanov Russian Economic University (Moscow, Russia, integrated);
2. Astana branch of the Institute of Nuclear Physics (Astana, Kazakhstan, integrated);
3. Georgian Technical University (Tbilisi, Georgia, in the process of integration);
4. Institute of Nuclear Physics (Tashkent, Uzbekistan, in the process of integration);
5. Institute of Physics of the National Academy of Sciences of Azerbaijan (Baku, Azerbaijan, integrated);
6. Academy of Scientific Research & Technology - Egyptian National STI Network (Egypt, Africa, integrated);
7. North Ossetian State University (Vladikavkaz, Russia, integrated);
8. Institute for Nuclear Research and Nuclear Energy (Sofia, Bulgaria, integrated);
9. St. Sophia University «St. Kliment Ohridski» (Sofia, Bulgaria, integrated);
10. Scientific Research Institute of Nuclear Problems of the Belarusian State University (Minsk, Belarus, integrated);



RESEARCH ON SARS-COV-2

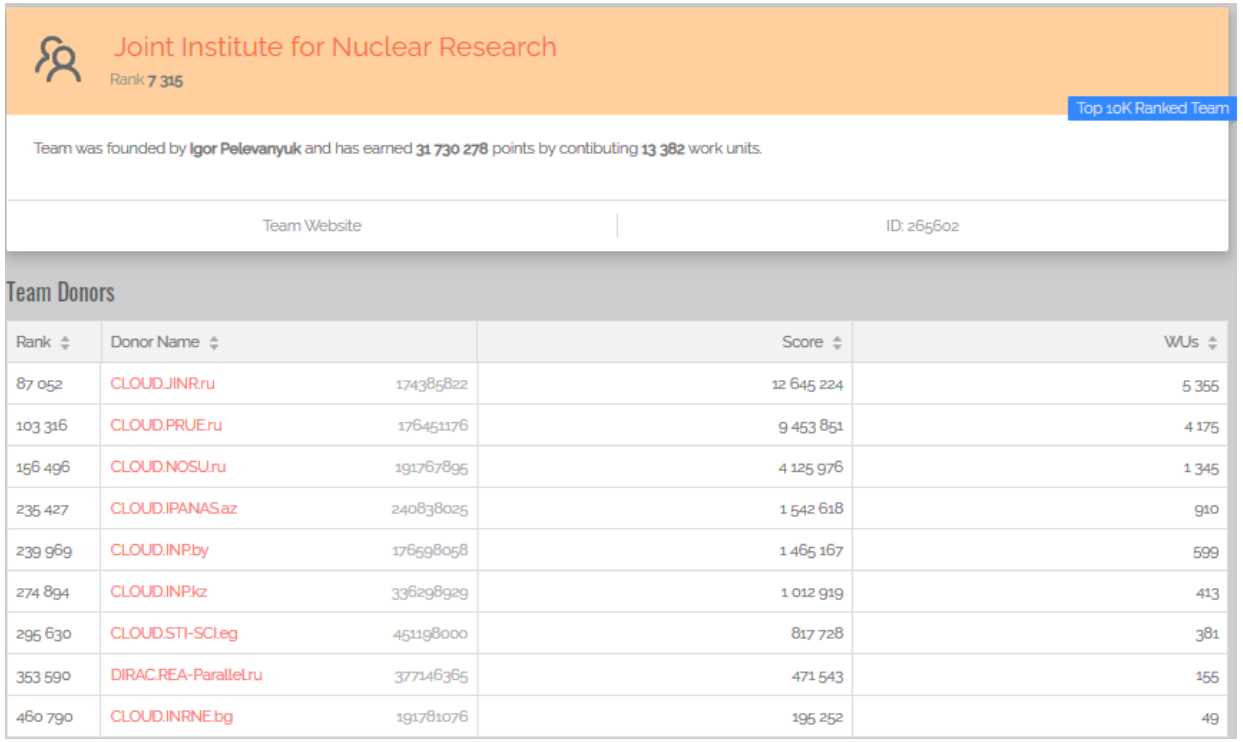
Idle resources of the JINR DICE infrastructure are used to study the SARS-CoV-2 virus with the help of the Folding@home (F@h) project. It is a distributed computing project aimed to help scientists develop new therapeutics for various diseases by simulating protein dynamics. This includes protein folding and protein movements, and is reliant on simulations run on volunteers' personal computers.

In March 2020, F@h launched a program to assist researchers around the world who are working on finding a cure and learning more about the coronavirus pandemic. The initial wave of projects simulate potentially druggable protein targets from the SARS-CoV-2 virus, and the related SARS-CoV virus, about which there is significantly more data available.

DIRAC running at JINR was tuned to perform Folding@home jobs within a week after the initial idea. No additional effort was required from the administrators of the already integrated clouds. Since small chunks of idle resources can be utilized by such payload, F@h jobs allow to enhance the overall efficiency of the shared resources.

F@h discover new antiviral candidates in animal testing and publish four papers so far:

1) SARS-CoV-2 Simulations Go Exascale to Capture Spike Opening and Reveal Cryptic Pockets Across the Proteome,
2) SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential,
3) The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA,
4) COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning.



Joint Institute for Nuclear Research

Rank 7 315

Top 10K Ranked Team

Team was founded by Igor Pelevanyuk and has earned 31 730 278 points by contibuting 13 382 work units.

Team Website | ID: 265602

Team Donors

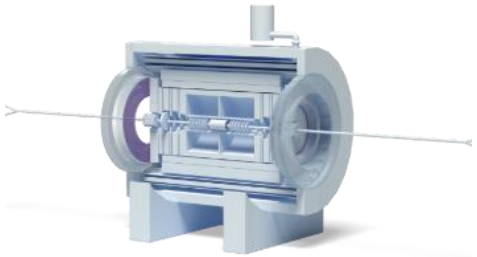
Rank	Donor Name		Score	WUs
87 052	CLOUD.JINR.ru	174385822	12 645 224	5 355
103 316	CLOUD.PRUE.ru	176451176	9 453 851	4 175
156 496	CLOUD.NOSU.ru	191767895	4 125 976	1 345
235 427	CLOUD.IPANAS.az	240838025	1 542 618	910
239 969	CLOUD.INP.by	176598058	1 465 167	599
274 894	CLOUD.INP.kz	336298929	1 012 919	413
295 630	CLOUD.STI-SCI.eg	451198000	817 728	381
353 590	DIRAC.REA-Parallel.ru	377146365	471 543	155
460 790	CLOUD.INRNE.bg	191781076	195 252	49

https://stats.foldingathome.org/team/265602

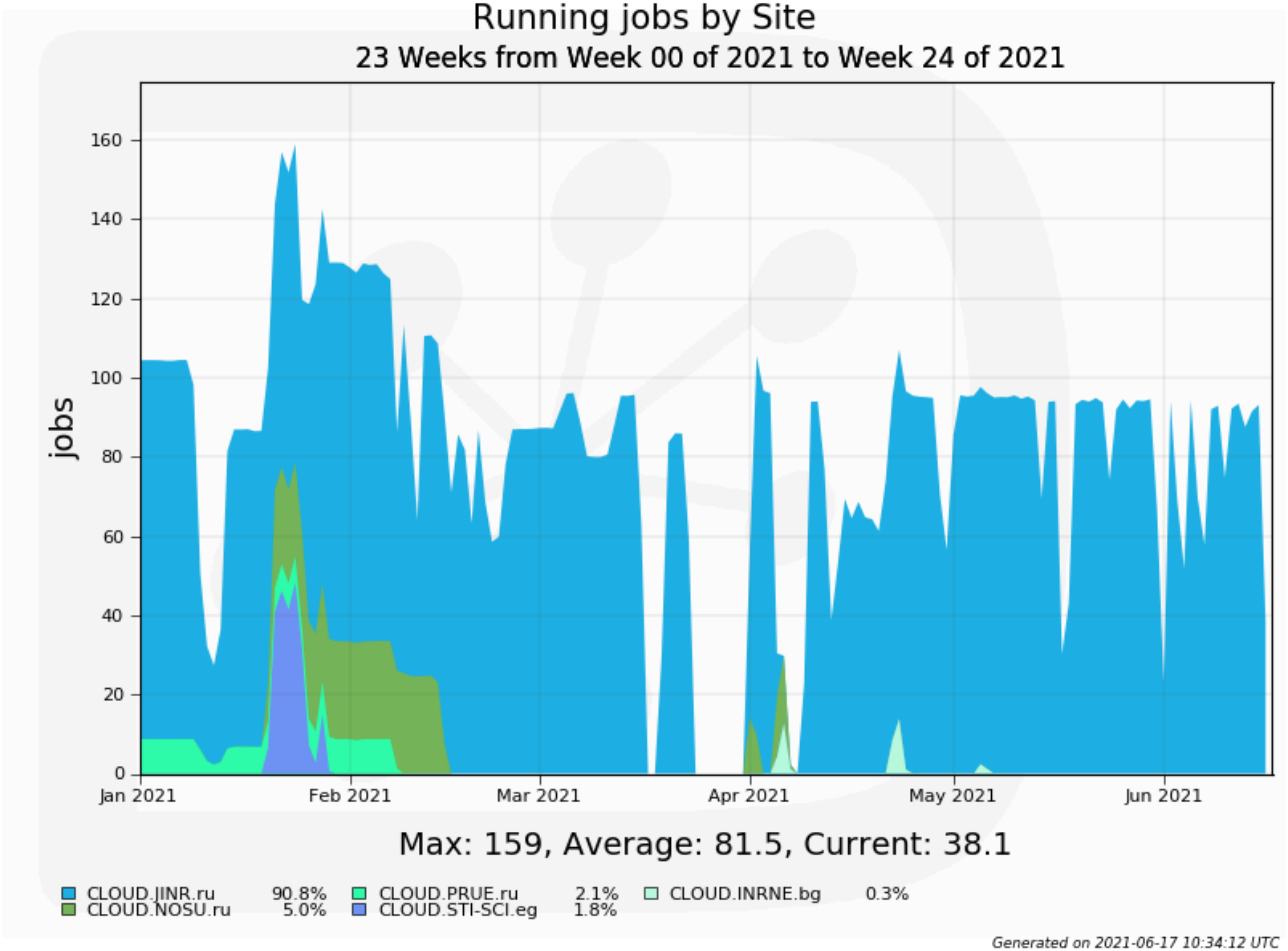
COMMUNICATION WITH THE EXPERIMENTS

The JINR DICE resources are configured to support the following virtual organizations (VOs) representing a scientific experiment and/or collaboration: BM@N, MPD, and Baikal-GVD. More than 15,000 Monte-Carlo simulation jobs for the BM@N VO were successfully completed as well as about 9000 jobs for Baikal-GVD VO. Testing jobs for MPD were performed on the distributed cloud infrastructure.

BAIKAL-GVD



BM@N
Baryonic Matter at Nuclotron



The statistics on the usage of the JINR DICE resources within the F@h project and by other users, including the experiments mentioned above.

CONCLUSION

Combining the resources of organizations from the JINR Member States and the efforts of its colleagues made it possible to build a distributed information and computing infrastructure and successfully use its resources for various scientific projects and experiments. The DIRAC Interware plays a function of a glue layer that transforms the computing power distributed across the world into a single environment, providing users with a single entry point for jobs and data management. Software distribution is implemented with the help of CVMFS tools and services. The JINR DICE resources are used not only by users from scientific experiments, in which JINR participates, but also the idle part of the computing power is utilized to study the SARS-CoV-2 virus.

PLANS

- Update to the latest stable OpenNebula version.
- Expand collaboration with experiments.
- Provide scientists from the JINR Member States with access to the computing power of the JINR supercomputer named after N.N. Govorun.
- Complete the integration with organizations that are in the process of integration at the moment.
- Start integration with new participants.

LATEST PUBLICATIONS

- Nikita Balashov et al. Creating a Unified Educational Environment for Training IT Specialists of Organizations of the JINR Member States in the Field of Cloud Technologies // International Conference on Modern Information Technology and IT Education SITITO 2018: Modern Information Technology and IT Education, ISBN: 978-3-030-46895-8, vol. 1201, 2020, pp. 149-162.
- Nikita Balashov et al. Present Status and Main Directions of the JINR Cloud Development // The 27th Symposium on Nuclear Electronics and Computing (NEC'2019), CEUR Workshop Proceedings, ISSN: 1613-0073, vol. 2507, 2019, pp. 185-189.
- Nikita Balashov et al. Cloud Integration Within the DIRAC Interware // The 27th Symposium on Nuclear Electronics and Computing (NEC'2019), CEUR Workshop Proceedings, ISSN: 1613-0073, vol. 2507, 2019, pp. 256-260.
- Vladimir Korenkov et al. The JINR distributed computing environment // 23rd International Conference on Computing in High Energy and Nuclear Physics (CHEP 2018), EPJ Web of Conferences, ISSN:2100-014X, vol. 214, 2019, P. 1-8.
- Ye. Mazhitova et al. Cloud infrastructure of INP'S Astana branch - PE "NULITS" and its integration with distributed JINR cloud infrastructure // «The XXII International Scientific Conference of Young Scientists and Specialists (AYSS-2018)», EPJ Web Conf. Vol. 201, 2019, pp. 1-5. doi:https://doi.org/10.1051/epjconf/201920105003.