Distributed information and computing infrastructure of JINR member states' organizations

Cite as: AIP Conference Proceedings **2377**, 040001 (2021); https://doi.org/10.1063/5.0063809
Published Online: 24 September 2021

Nikita A. Balashov, Nikolay A. Kutovskiy, Alexander N. Makhalkin, et al.

Distributed Information and Computing Infrastructure of JINR Member States' Organizations

Nikita A. Balashov[1], Nikolay A. Kutovskiy[1], Alexander N. Makhalkin[1], Yelena Mazhitova[1,2, a)], Igor S. Pelevanyuk[1], and Roman N. Semenov[1,3]

[1]*Laboratory of Information Technologies, Joint Institute for Nuclear Research, 6 Joliot-Curie, 141980, Dubna, Moscow region, Russia*
[2]*Institute of Nuclear Physics, 1 Ibragimova street, 050032, Almaty, Kazakhstan*
[3]*Plekhanov Russian University of Economics, 36 Stremyanny per., 117997, Moscow, Russia*

[a)]*Corresponding author: emazhitova@jinr.ru*

Abstract. For a significant reduction of time spent on research to obtain meaningful results in scientific fields, the computing resources of the Joint Institute for Nuclear Research (JINR) and some organizations of its Member States were integrated into a distributed information and computing environment (DICE). The technical possibility for running tasks in the environment was implemented for users of the BM@N, MPD and Baikal-GVD collaborations. The resources not occupied by computational tasks within the main scientific fields of JINR are used to conduct studies on the SARS-CoV-2 virus that causes the COVID-19 disease. In addition, the paper gives a description of the DICE technical implementation, lists the participating organizations and provides some statistics on the use of their resources.

DISTRIBUTED CLOUD INFRASTRUCTURE

JINR participates in a large number of research projects, in many of which computing infrastructures are an important tool for obtaining relevant scientific results. In this connection, the pooling of computational resources of JINR and organizations from its Member States is a crucial task, the solution of which will significantly accelerate scientific studies.

The cloud infrastructures of JINR and its Member States' organizations are based on the open source solution OpenNebula [1]. The JINR cloud is the core of this infrastructure [2]. It hosts DIRAC services, which manage computational tasks and data over the resources of JINR and its Member States' organizations. DIRAC provides all the necessary components to build ad-hoc grid infrastructures that interconnect computing resources of different types, enabling interoperability and simplifying interfaces [3].

Software distribution model

Software distribution on the remote resources was performed using the CernVM File System. CernVM-FS is a web-based, global, and versioning file system optimized for software distribution [4]. The content of the file system 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 metadata on demand and cache them locally. Data integrity and authenticity is ensured by cryptographic hashes and digital signatures. Using CernVM-FS, the optimization of methods for delivering the software of the experiments to the DICE resources was carried out in order to increase the speed of launch and operation of application software (see Fig. 1).

Proceedings of the 24th International Scientific Conference of Young Scientists and Specialists (AYSS-2020)
AIP Conf. Proc. 2377, 040001-1–040001-5; https://doi.org/10.1063/5.0063809
Published by AIP Publishing. 978-0-7354-4132-3/$30.00

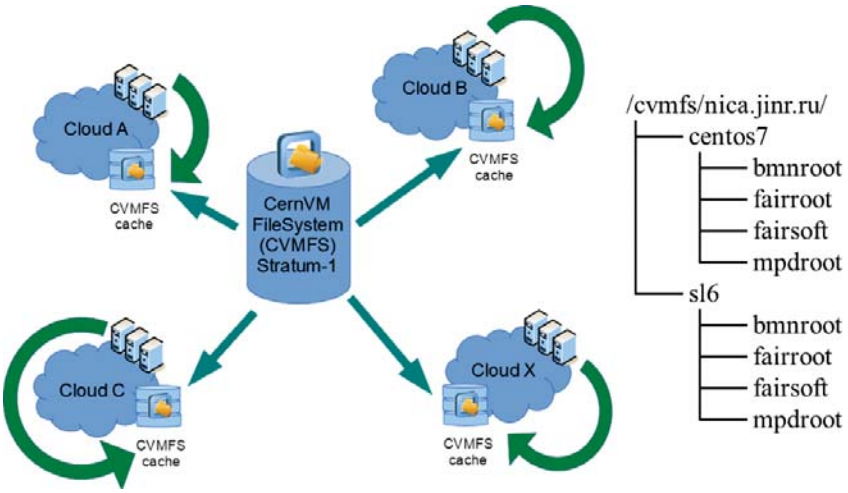

FIGURE 1. Software distribution model of the experiments

Participants of the infrastructure

The JINR cloud is the core of the DICE, and other organizations are in the process of integrating with the JINR cloud. Some information about the DICE clouds of organizations from the JINR Member States and their resources is given in Table 1.

TABLE 1. Stages of integration and resources of the participants of the infrastructure

№	Name of the organization	City	Country	Status	Number of CPU cores, pcs.	Total RAM, GB	Total disk space, TB
1	Joint Institute for Nuclear Research	Dubna	Russia	Core of the DICE	4,532	57,856	3,072
2	Plekhanov Russian University of Economics	Moscow	Russia	Integrated	132	608	51.1
3	Astana branch of the Institute of Nuclear Physics	Nur-Sultan	Kazakhstan	Integrated	84	840	6.8
4	Georgian Technical University	Tbilisi	Georgia	In the process of deployment	50	308	20
5	Institute of Nuclear Physics	Tashkent	Uzbekistan	In the process of deployment	-	-	-
6	Institute of Physics of the National Academy of Sciences of Azerbaijan	Baku	Azerbaijan	Integrated	16	96	56
7	North Ossetian State University	Vladikavkaz	Russia	Integrated	84	672	17.2
8	Academy of Scientific Research & Technology - Egyptian National STI Network	Egypt	Africa	In the process of deployment	98	704	13.8
9	Institute for Nuclear Research and Nuclear Energy	Sofia	Bulgaria	Integrated	8	64	4
10	St. Sophia University “St. Kliment Ohridski”	Sofia	Bulgaria	Integrated	48	250	4.7
11	Research Institute for Nuclear Problems of Belarusian State University	Minsk	Belarus	Integrated	132	290	127

Figure 2 shows the location of the participants in the distributed cloud infrastructure on the map.

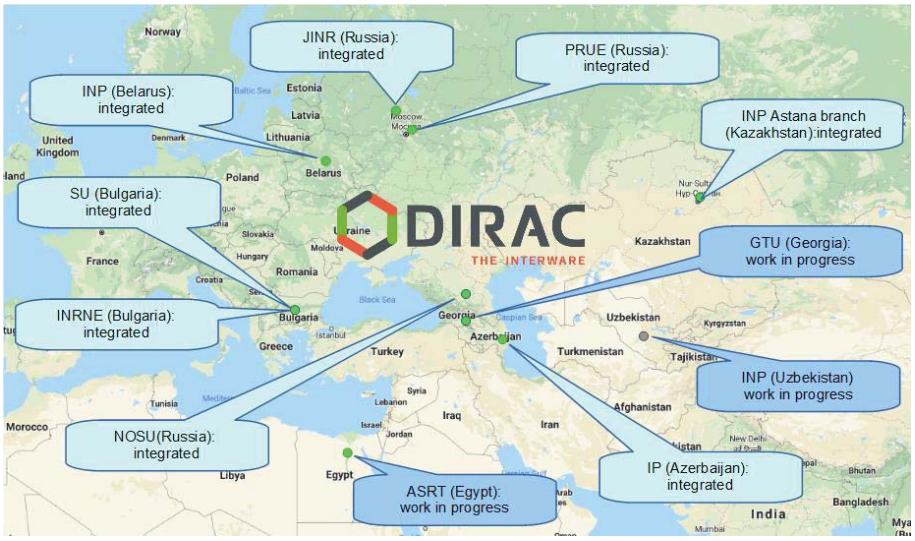

FIGURE 2. Map with the participants of the distributed cloud infrastructure

One of the uppermost tasks is to provide scientists from the JINR Member States with access to the computing power of the JINR supercomputer named after N.N. Govorun. The integration of its hardware and the cloud resources of the participating countries significantly expand the total computing capacity of the entire DICE.

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 [5], MPD [6], and Baikal-GVD [7]. More than 15,000 Monte-Carlo simulation jobs for the BM@N VO were successfully completed. Testing jobs for MPD were performed on the distributed cloud infrastructure. Users of the Baikal-GVD VO use the JINR DICE resources very intensively.

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 [8]. 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.

The project utilizes central processing units (CPUs), graphics processing units (GPUs), and other devices. Due to the increased interest in the project as a result of the COVID-19 pandemic, the system reached a speed of approximately 1.22 exaflops by the end of March 2020 and 2.43 exaflops by 12 April 2020, which made it the world's first exaflop computing system.

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. If necessary, stopping F@h is a matter of several commands.

The statistics on the usage of the JINR DICE resources within the F@h project and by other users, including the experiments mentioned above, are shown in Figures 3-6.

Team: Joint Institute for Nuclear Research

Date of last work unit	2020-10-21 06:04:11
Active CPUs within 50 days	155
Team Id	265602
Grand Score	27,693,728
Work Unit Count	12,018
Team Ranking	6756 of 255314
Homepage	http://www.jinr.ru/main-en/

Team members

Rank	Name	Credit	WUs
68	416 CLOUD.JINR.ru	12,645,224	5,355
81	988 CLOUD.PRUE.ru	9,453,851	4,175
200	259 CLOUD.IPANAS.az	1,542,618	910
204	504 CLOUD.INP.by	1,465,167	599
230	877 CLOUD.NOSU.ru	1,083,663	395
237	027 CLOUD.INP.kz	1,012,919	413
311	109 DIRAC.REA-Parallel.ru	471,543	155
N/A	CLOUD.INRNE.bg	18,743	16

FIGURE 3. Statistics of using the JINR DICE resources from the F@h site

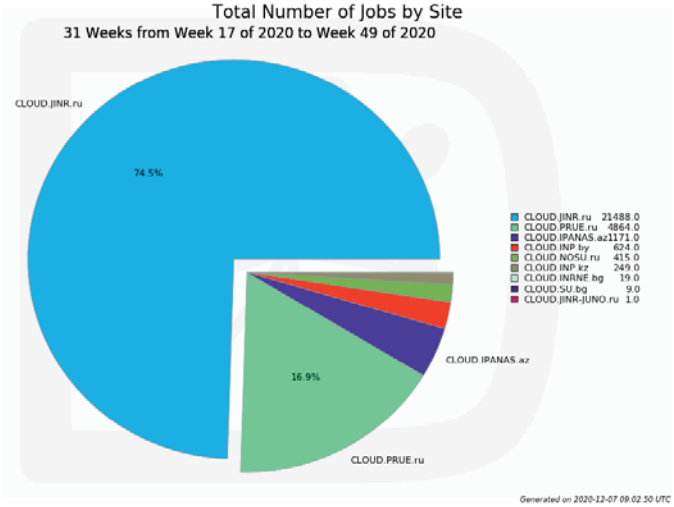

FIGURE 4. Pie chart with the number of completed tasks on different clouds

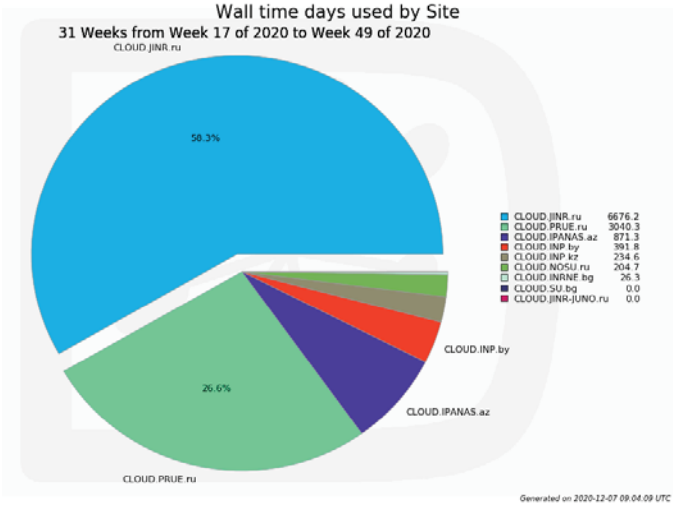

FIGURE 5. Pie chart with the wall-time days on different clouds

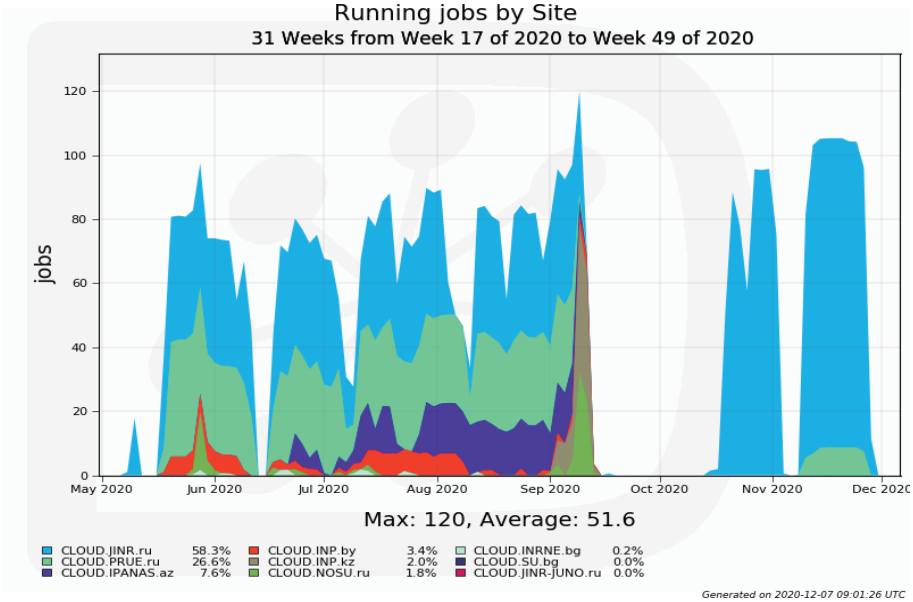

FIGURE 6. Statistics of simultaneously DIRAC F@h+Baikal running jobs

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 CernVM-FS 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.

REFERENCES

1. R. Kumar et al., "OpenNebula: Open Source IaaS Cloud Computing Software Platforms", in *National Conference on Computational and Mathematical Sciences (COMPUTATIA-IV)*, 1-5 (2014).
2. N.A. Balashov et al., Phys. Part. Nuclei Lett. 13, 672–675 (2016).
3. N.A. Balashov et al., CEUR Workshop Proc., 2507, 256-260 (2019).
4. J. Blomer et al., EPJ Web of Conferences, 214, 09007 (2019).
5. K. Gertsenberger et al., EPJ Web of Conferences, 226, 03008 (2020).
6. V. Golovatyuk et al., Nucl. Phys. A 982, 963-966 (2019).
7. A.D. Avrorin et al., Bull. Russ. Acad. Sci. Phys. 83, 921–922 (2019).
8. M. I. Zimmerman et al., bioRxiv:2020.06.27.175430 (2020).