The INP Almaty cloud and it's integration into a JINR DICE infrastructure

Nikita Balashov[1], Nassurlla Burtebayev[2], Vladimir Korenkov[1], Nikolay Kutovskiy[1], Alexander Makhalkin[1], Yelena Mazhitova[1,2], Igor Pelevanyuk[1], Ilyas Satyshev[1,2], Roman Semenov

[1] Meshcheryakov Laboratory of Information Technologies, Joint Institute for Nuclear

[2] Institute of Nuclear Physics

Relevance

Nowadays, the successful implementation of a significant part of scientific projects involves the use of a distributed information and computing environment (DICE) for storing, processing and analyzing data.

The JINR DICE initiative is dedicated to the creation, support and development of such an environment by combining the resources of educational and research organizations of the JINR Member States.

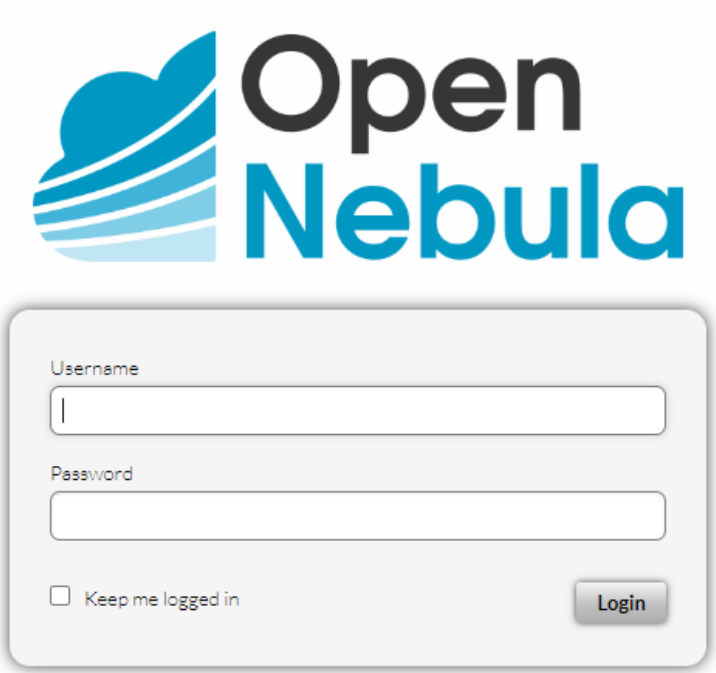
One of such organizations is the Institute of Nuclear Physics in Almaty (Kazakhstan).

OpenNebula

To create a cloud infrastructure an open source solution OpenNebula was chosen.

It is a cloud computing platform for managing heterogeneous distributed data center infrastructures. This platform manages the virtual infrastructure of the data center to create private, public and hybrid implementations of infrastructure as a service.

OpenNebula allows to deploy an IaaS model that provides the ability to use cloud infrastructure to independently manage processing, storage, networks and other fundamental computing resources, as well as provide a hybrid scheme by combining the resources of a local data center and external cloud providers.



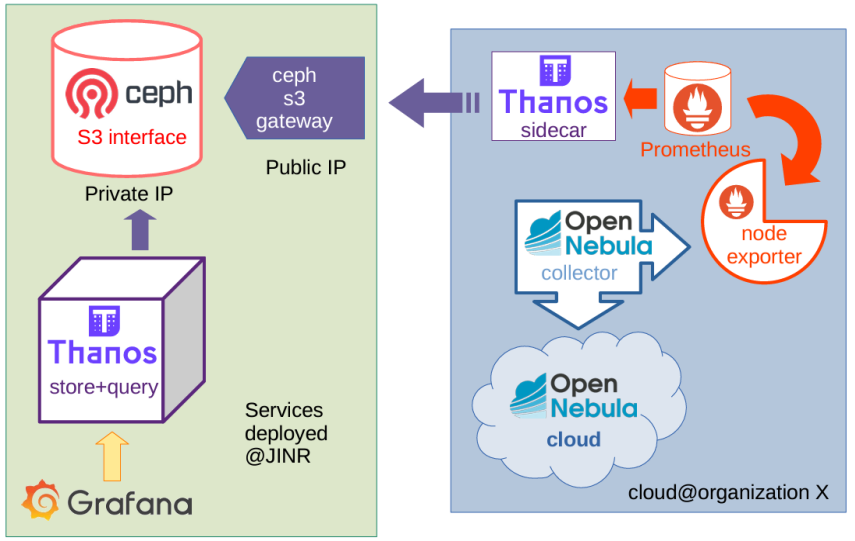
Monitoring system

One of the necessary elements of software testing and debugging is a monitoring system.

Several years ago, a custom collector for collecting cloud metrics based on OpenNebula was developed at JINR.

It can store the collected data in the Prometheus time series database via the node_exporter component. Using the Thanos sidecar, the collected metrics are sent to the ceph object storage. Grafana is used for data visualization. It requests data in the object storage via Thanos Querier and Thanos Store.

This collector was installed in the INP cloud infrastructure.



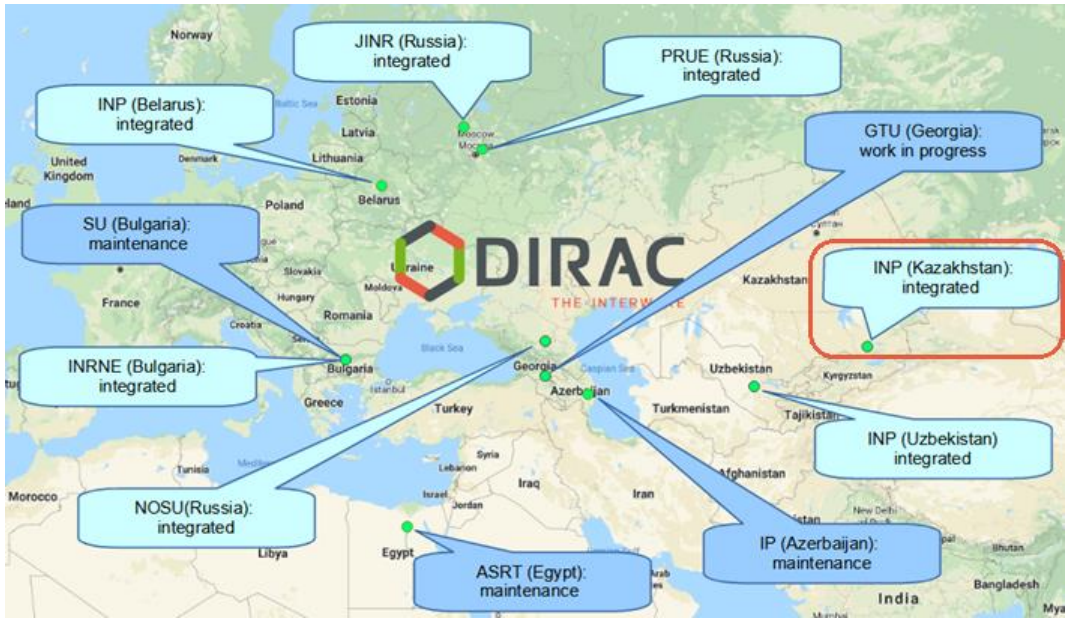
Distributed information and computing environment

The next step was the integration of the INP cloud into the JINR DICE.

In particular the JINR distributed informational and computing environment was create to combine computational power for solving common scientific tasks, to distribute peak loads across participants as well as to disseminate knowledge and practical skills on cloud and grid technologies among users and admins of the JINR Member State research and educational organizations. Currently JINR DICE based on the computational and storage facilities of the participants shown on the map. And INP is one of this organizations.

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 plays the role of a connecting layer that transforms computing resources (*computational grids, clouds or clusters*) distributed around the world into a unified environment, providing users with a single entry point for working and managing data.

Thanks to the integration of the INP cloud infrastructure into the JINR DICE, users can access much more infrastructure resources than when using only the local cloud.



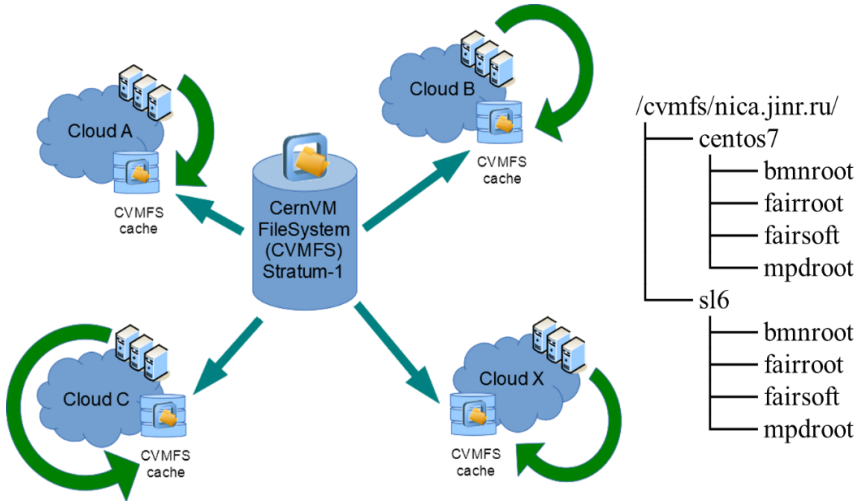
Detailed information about the JINR DICE, its participants and use can be found on the website https://dice.jinr.ru

Software distribution model

The software is distributed to remote resources using the CernVM file system.

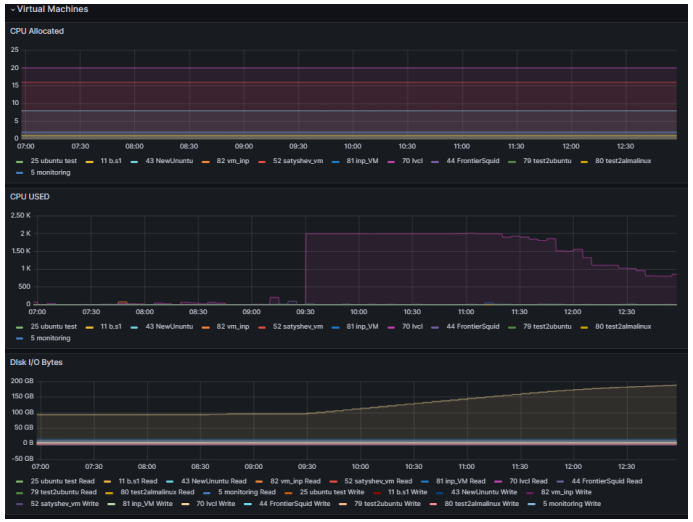
It is a web–oriented global file system with version control optimized for software distribution. The contents of the file system are installed on a central web server, from where they can be mirrored and cached by other web servers and web proxies. File system clients load data and metadata on request and cache it locally. Data integrity and authenticity are ensured by cryptographic hashes and digital signatures.

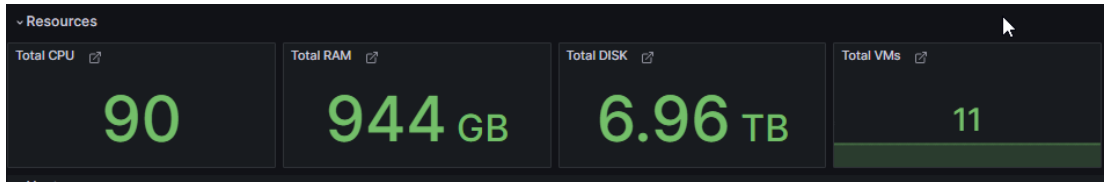
For more stable operation of CernVM-FS, the Frontier-squid software package, which is a corrected version of the standard squid HTTP-proxy caching software, is installed. The frontier-squid package contains default settings and bug fixes that work well with applications used on the network.

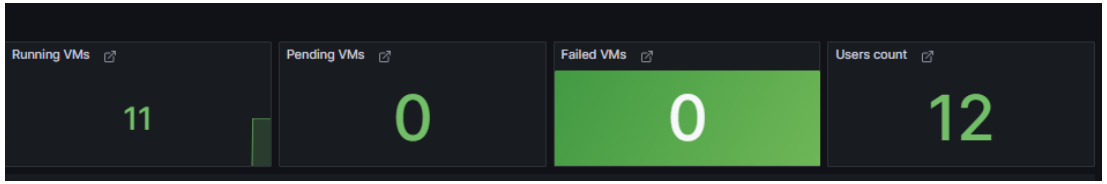


The INP's cloud infrastructure

The INP's cloud infrastructure has the following hardware resources: the number of central processor cores is 90, the total amount of random access memory is 944 GB, the total ceph-based storage capacity is 6.96 TB.







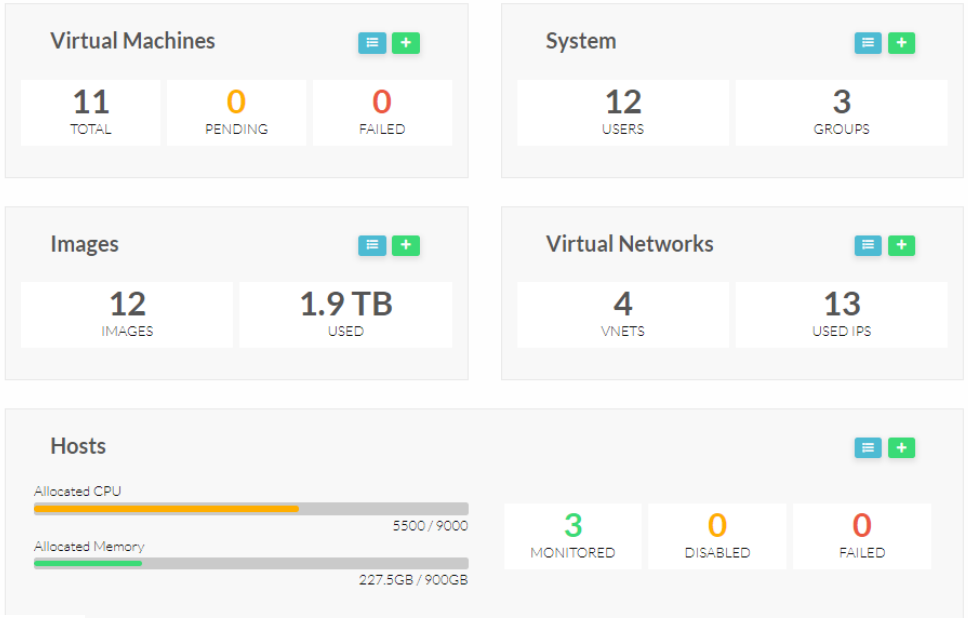
The INP's cloud infrastructure

Infrastructure administrators tested the OpenNebula software by downloading operating system images, creating virtual machine templates, creating virtual machines, running and working in them.

Cloud users search was conducted, the hardware and software they needed were identified.

There was organized and conducted a set of training events for specialists working in the cloud infrastructure based on OpenNebula, after which users were invited to create their own virtual machines and perform work tasks on them. Users coped with the task. The approbation was successful.

Nowadays 12 users are registered, 11 virtual machines are successfully launched, 55 CPU cores, 227.5 GB of RAM and 1.9 TB of disk storage are used.



Full name	Direction	Programs	Data volumes per task (input/output)	Tasks can be launched in batch mode	CPU cores (90)	RAM (944 GB)	Disk (6.96 TB)
Asel Mukhamedzhanova	SPD	Garfield++	~ 10 MB ~ 20 GB	yes	40 **(20)**	**100 GB**	5TB **(1TB)**
Nurzhan Erezhepov	Tasks in the field of cosmic ray physics and high and ultra-high energy physics	Corsika Geant4 ROOT	1. input 1-5 kb, output 100 MB up to 300-500 GB 2. input 1-5 MB, output 100 MB up to 10 GB 3. depends on data received in 1.1 and 2.1	1 yes 2 no 3 partially	40 **(20)**	**64 GB**	4 TB **(1TB)**
Igor Lebedev	MPD, B@MN	FORTRAN GEANT	~100 MB and more	It is possible, but needs to be checked	**4**	**16 GB**	**1 TB**
Ilyas Satyshev	OLVE-HERO	Geant4 ROOT	100 MB - 100 GB	It is possible, but needs to be checked	**16**	**100 GB**	**1 TB**
Pavel Krasovitsky	theoretical physics	FORTRAN	-	Most likely not	**10**	**50 GB**	**0,5 TB**
Meirzhan Kakenov	Application of the dibaryon model to study the two-proton decay of 6Be. Determination of the spectrum of excited states for light nuclei with A=6: 6Li, 6He and 6Be.	C/C++ FORTRAN	-	-	**10**	**50 GB**	**0,5 TB**
Bekarystan Sattar	IT specialist at INP						

The INP's cloud infrastructure

Currently, the cloud infrastructure of the INP is used by representatives of the following organizations:

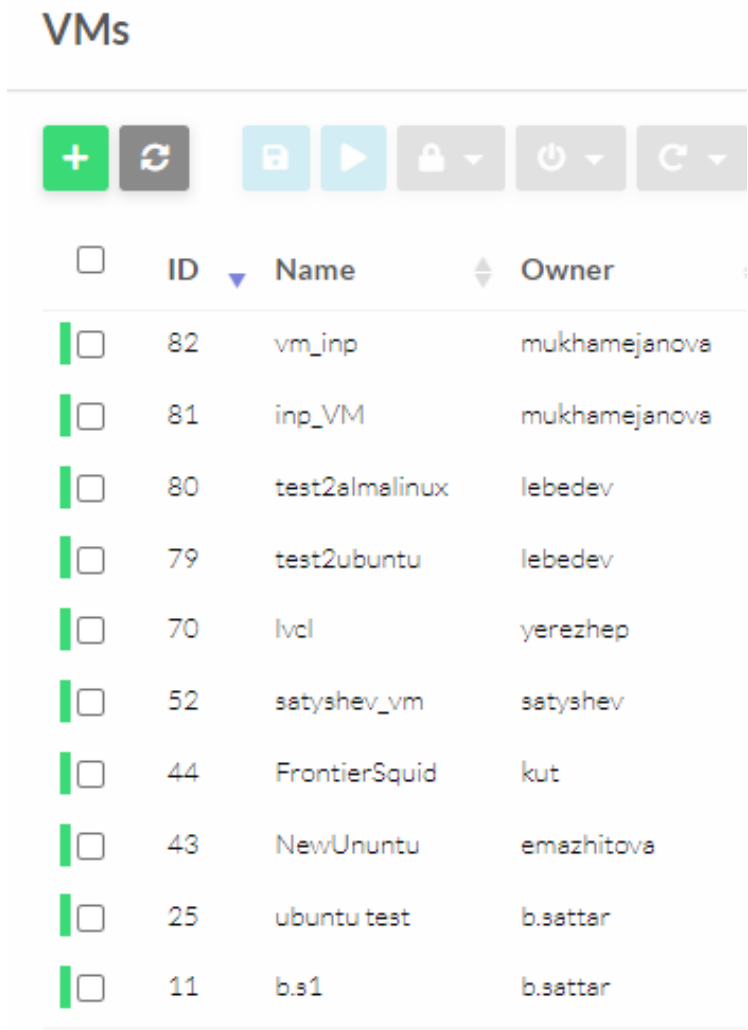
- Institute of Nuclear Physics (Almaty, Kazakhstan),
- Joint Institute for Nuclear Research (Dubna, Russia),
- Al-Farabi Kazakh National University (Almaty, Kazakhstan),
- L.N. Gumilev Eurasian National University (Astana, Kazakhstan),
- Institute of Physics and Technology (Almaty, Kazakhstan).

The following tasks are being solved:

- SPD (Spin Physics Detector), MPD (Multi-Purpose Detector), BM@N (Baryonic Matter at Nuclotron) experiments of NICA megascience project,
- OLVE-HERO - High Energy Cosmic Ray Observatory studies cosmic radiation in the «knee» region 1012 - 1016 eV,
- Tasks in the field of cosmic ray physics and high and ultra-high energy physics in particular CORSIKA. (COsmicRaySImulationsforKAscade - is a Monte Carlo tool designed to analyze the properties and evolution of widespread air showers).

Moreover, virtual machines are deployed in the cloud infrastructure to test the operation of various software and the operation of the cloud infrastructure in general.

The INP's cloud infrastructure available to users at https://cloud.inp.kz/



ID	Name	Owner
82	vm_inp	mukhamejanova
81	inp_VM	mukhamejanova
80	test2almalinux	lebedev
79	test2ubuntu	lebedev
70	lvcl	yerezhep
52	satyshev_vm	satyshev
44	FrontierSquid	kut
43	NewUnuntu	emazhitova
25	ubuntu test	b.sattar
11	b.s1	b.sattar

Conclusion

Based on the INP's resources:

- A cloud infrastructure was installed,
- The necessary services were deployed.

The INP cloud was integrated into the JINR DICE.

A group of pioneer scientists was selected who start using the INP's cloud resources to conduct research in their areas.

Presently, the work is focused on maintaining software, running services, failure cloud recovery, as well as user support, which includes conducting training events on working on the OpenNebula portal, assistance in installing the necessary software products (including graphical interfaces) to solve their problems on virtual machines.

«On 11 November 2023, within the guest CP JINR session, a new cloud computing cluster of the Joint Institute for Nuclear Research was launched».
https://www.jinr.ru/posts/cloud-cluster-launched-and-jinr-information-centre-opened-at-inp-in-almaty/

References

- Balashov N., Burtebayev N., Korenkov V., Kutovskiy N., Mazhitova Ye., Satyshev I., Semenov R. CERN-JINR-INP-KazNU Data Center: Current Status and Plans // Physics of Particles and Nuclei Letters. – 2022. – Vol. 19, No. 5. – P. 547–549. Q3
- DIRAC web portal.. URL: https://dirac.readthedocs.io/en/latest/ (date of access 25.10.2024).
- OpenNebula web portal. URL: https://opennebula.io/ (date of access 25.10.2024).
- Milojicic D., Llorente I.M., Montero R.S. “OpenNebula: A cloud management tool”. IEEE Internet Computing, Vol.15, March 2011.
- Prometheus web portal. URL: https://prometheus.io (date of access 25.10.2024).
- Prometheus node_exporter github page. URL: https://github.com/prometheus/node_exporter (date of access 25.10.2024).
- Thanos sidecar github page. URL: https://github.com/thanos-io/thanos (date of access 25.10.2024).
- Grafana web portal. URL: https://grafana.com/grafana/ (date of access 25.10.2024).
- Balashov N.A., Kuprikov I.S., Kutovskiy N.A., Makhalkin A.N., Mazhitova Ye., Pelevanyuk I.S., Semenov R.N. JINR distributed information and computing environment: participants, features and challenges // CEUR Workshop Proceedings, ISSN:1613-0073, 2021. – Vol. 3041. – P. 280-284.
- Balashov N.A., Kutovskiy N.A., Makhalkin A.N., Mazhitova Ye., Pelevanyuk I.S., Semenov R.N. Distributed information and computing infrastructure of JINR member states’ organizations // AIP Conference Proceedings, ISBN: 978-0-7354-4132-3, 2021. – Vol. 2377. – 040001.
- CernVM File System web portal. URL: https://cernvm.cern.ch/fs/ (date of access 25.10.2024).
- Frontier-squid cache server installing page. URL: https://twiki.cern.ch/twiki/bin/view/Frontier/InstallSquid (date of access 25.10.2024).

ACKNOWLEDGEMENT

The work was supported by the Program # BR10965191 of the Ministry of Education and Science of the Republic of Kazakhstan.

Thank you for attention!

Q&A