INTERNATIONAL
INTERGOVERNMENTAL
ORGANIZATION

JOINT INSTITUTE
FOR NUCLEAR RESEARCH

JINR and its Member States Cloud Infrastructures Status

N. Balashov[1], I. Hristov[2], R. Hristova[2], N. Kutovskiy[1], A. Makhalkin[1], Ye. Mazhitova[1,3], I. Pelevanyuk[1], R. Semenov[1]

1 Meshcheryakov Laboratory of Information Technologies, JINR

2 Sofia University, Bulgaria

3 Institute of Nuclear Physics, Almaty, Kazakhstan

The 11th International Conference

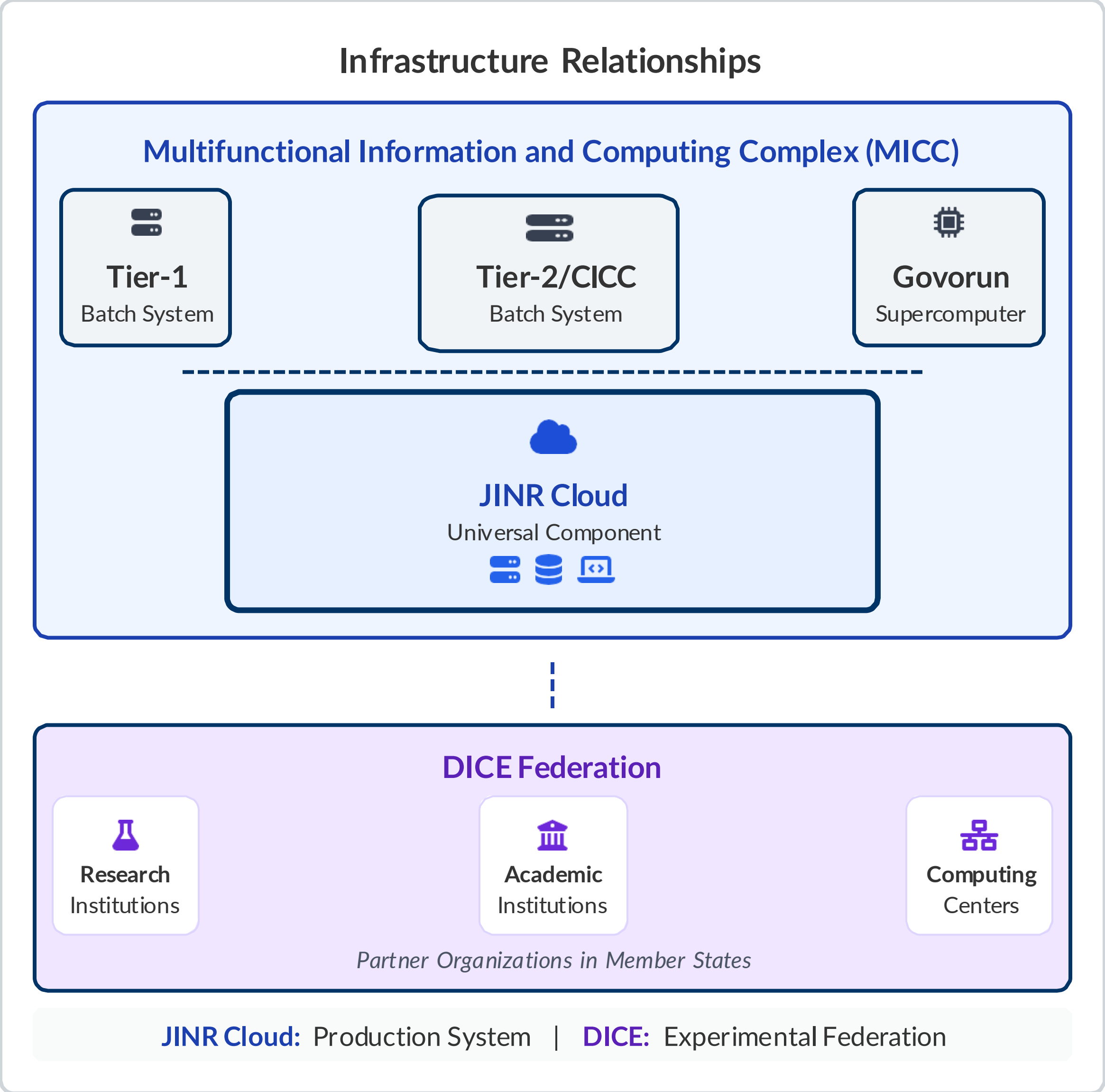
"Distributed Computing and Grid technologies in Science and Education"

(GRID'2025)

July 7-11, 2025, JINR, Dubna, Russia

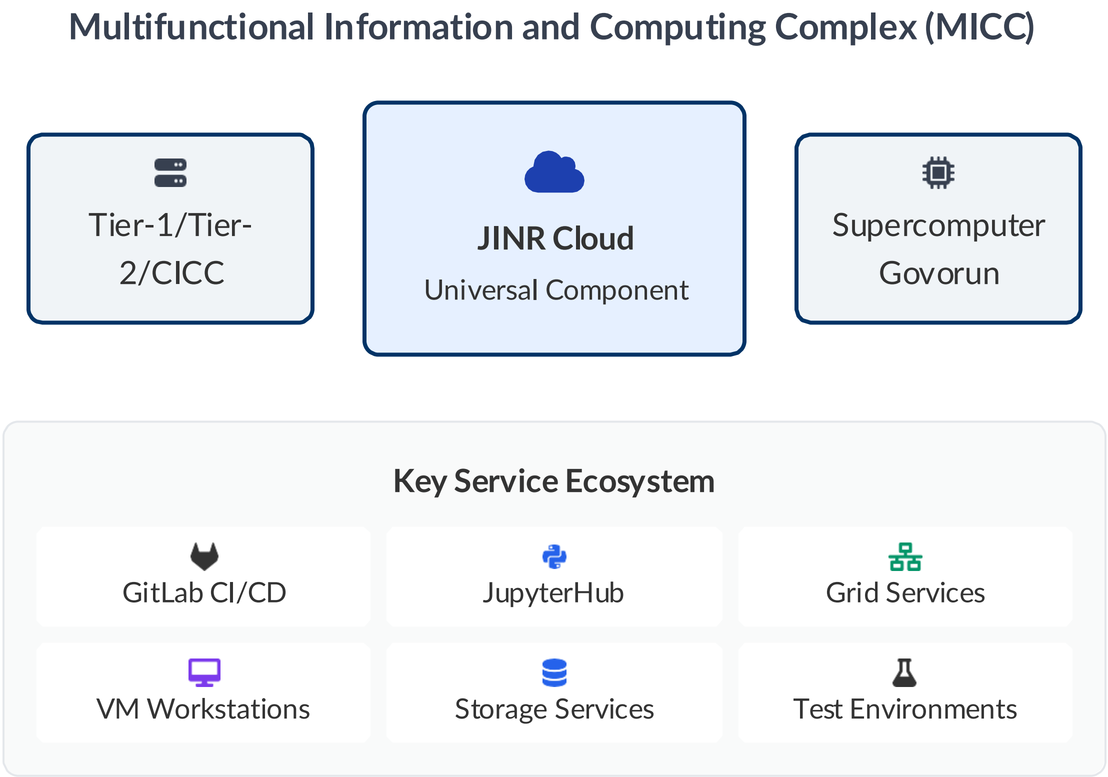
Overview & Agenda

- **JINR Cloud:** Production system within MICC
- **DICE:** Experimental partnership cloud environment
- **Operational Insights:** Usage patterns, service management, monitoring and support systems



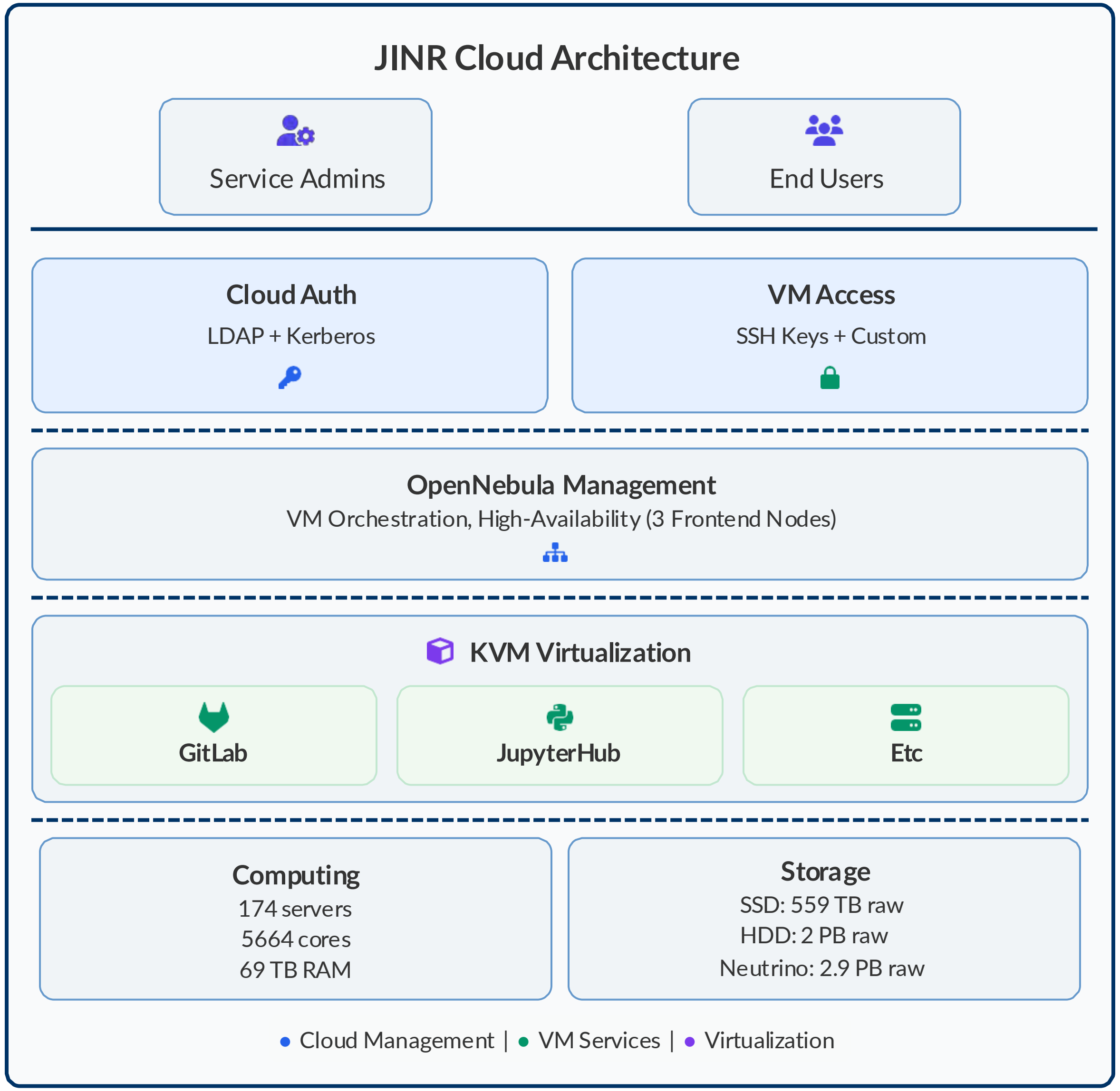
The Role of JINR Cloud in MICC

- JINR Cloud serves as the universal, flexible component within the Multifunctional Information and Computing Complex (MICC)
- Enables both service-oriented infrastructure and non-standard/scalable use-cases
- Complements static batch systems and the supercomputer with dynamic virtualized resources
- Simplifies service administration - cloud team manages hardware/virtualization, service admins manage VMs
- Offers on-demand scalability for services with unpredictable resource requirements



JINR Cloud Architecture

- **Platform:** OpenNebula (v6.10 CE), KVM virtualization
- **Cloud Authentication:** JINR central user database (LDAP+Kerberos)
- **VM Authentication:** SSH keys by default, users can configure additional methods
- **Management:** Cloud team handles hardware & virtualization; service admins manage VMs
- **Hardware for VMs:** 176 servers, 5664 cores (+504 since Grid2023), 69TB RAM
- **Server specs:** 20-192 non-HT CPU cores per server, 5-16GB RAM per core
- **Storage back-end:** Ceph block-device for KVM VM images, 36 servers
- **API:** provides XML-RPC for managing VMs programmatically



Ceph Storage

- **Access protocols:** RBD for VM disks, CephFS for experimental data, S3 for backups and artifacts storage
- **Data protection:** Triple replication (3x) across all storage clusters for high availability and data security
- **SSD Cluster:** Optimized for high I/O workloads, hosts critical services (Prometheus DB, COMPASS, GitLab, etc)
- **HDD Cluster:** Multi-protocol storage with balance between capacity and performance for general VM workloads
- **Neutrino Cluster:** Dedicated to experimental data with enhanced security measures and isolation
- **Availability scope:** SSD and HDD clusters serve all JINR infrastructure, Neutrino cluster exclusive to Neutrino platform

Ceph Storage Infrastructure

All-Flash SSD Storage
Ceph v18.2.7

Raw Capacity: 559 TiB
100 Gbps Connectivity

VM Disks (high I/O)
257 TiB used / 302 TiB free

HDD-based Storage
Ceph v18.2.7

Raw Capacity: 2 PiB
2x10GBase-T

VM Disks (mild/low I/O), CephFS, S3
0.6 PiB used / 1.7 PiB free

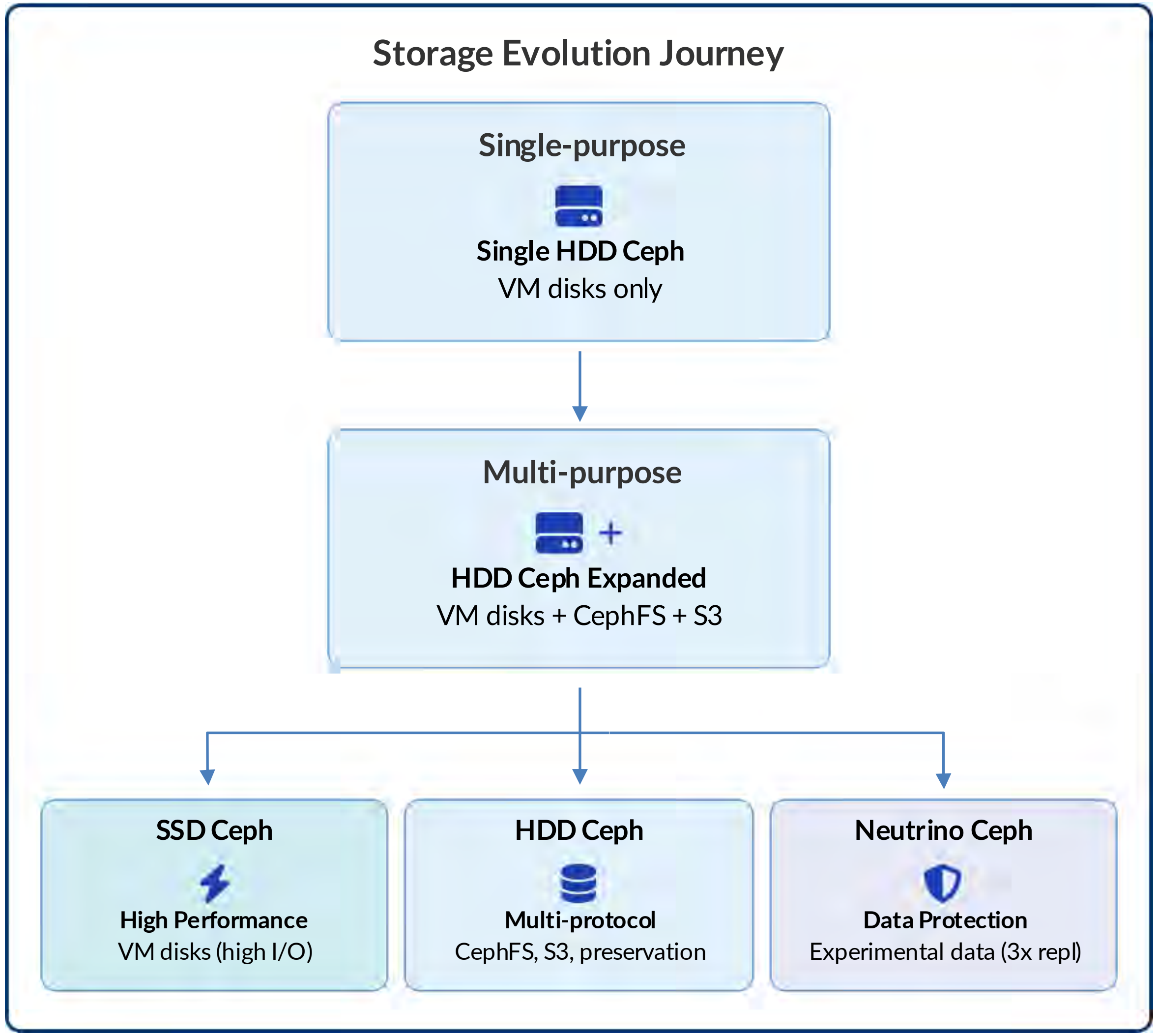
Neutrino Storage
Ceph v18.2.4

Raw Capacity: 2.9 PiB
2x10GBase-T

Experimental Data with Enhanced Protection
1.96 PiB used / 0.94 PiB free

Ceph Storage Evolution

- **Single-purpose storage** (HDD for VMs) → **Multi-purpose storage** (HDD for CephFS/S3/VMs) → **Specialized storage clusters**
- **Current state (2025):**
  - **SSD Ceph:** VM disks (high I/O workloads)
  - **HDD Ceph:** Shifts towards preservation storage, CephFS, S3
  - **Neutrino Ceph:** Protected experimental data
- **Benefits of specialization:**
  - Optimized performance for specific workloads
  - Better resource allocation based on needs
  - Enhanced data protection for sensitive information



JINR Cloud Usage Patterns

- **Service infrastructure:** Cloud team manages hardware, service admins manage VMs
- **Scalable services:** GitLab CI/CD runners, JupyterHub, grid control nodes
- **Interactive computing:** Personal VMs for researchers via SSH
- **Testing environments:** Flexible resources for system validation
- **Non-standard use-cases:** Support for unpredictable demand and varied workloads
- **Resource distribution:** Dynamic allocation based on priority and demand

Web Services

GitLab, document systems, web portals, monitoring

Data Analysis

JupyterHub, interactive computing, data visualization

Grid Services

Control nodes, batch system, CEs

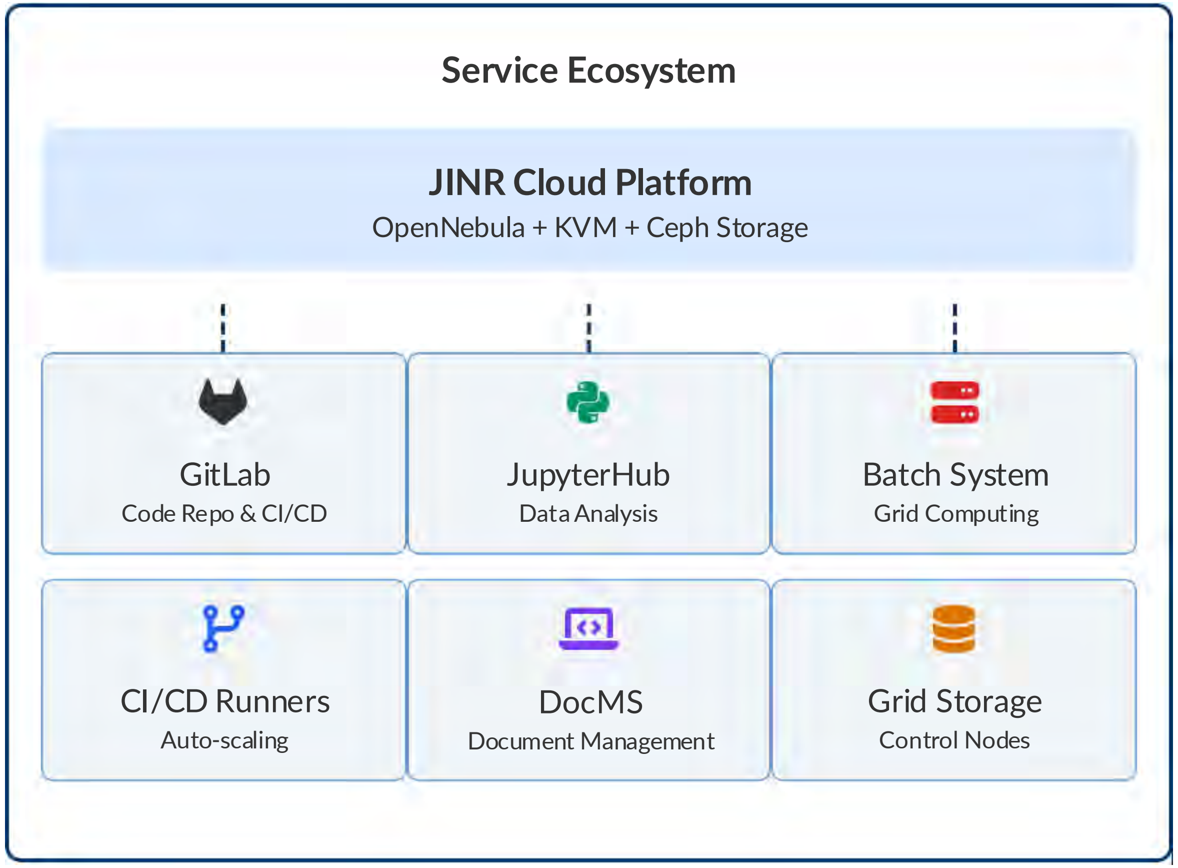
Development

Testing environments, CI/CD pipelines, integration

Service Admins | Researchers | Students

Services Managed by Cloud Team

- **GitLab:** Version control and repository management with dynamic CI/CD runners
- **JupyterHub:** Interactive data analysis platform for researchers
- **HTCondor batch system:** Part of neutrino platform for distributed computing
- **Compute Elements (CEs):** Grid computing interfaces for job submission and management
- **DocMS:** Document Management System for organizational content

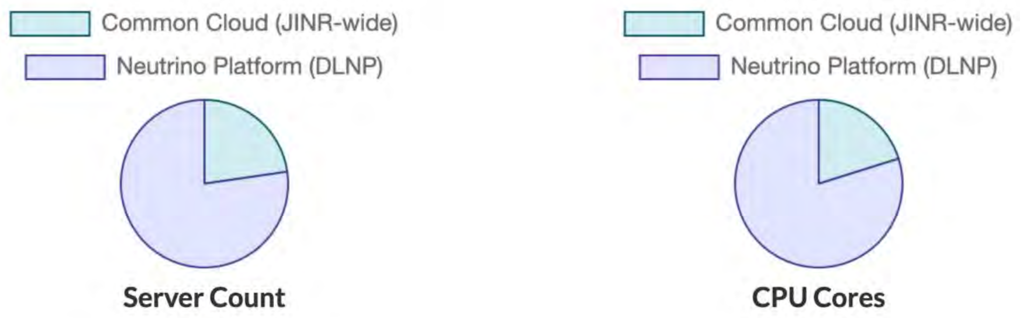


JINR Cloud Resource Distribution

- **Total Infrastructure:** 176 servers, 5,664 CPU cores, 854 virtual machines
- **User Base:** 103 active users including service accounts
- **Common Resources:** 40 servers with 1,136 CPU cores serving institute-wide needs
- **Neutrino Platform (DLNP):** 136 servers with ~4,500 CPU cores dedicated to neutrino experiments
- **Storage Specialization:** General purpose (HDD+SSD) JINR-wide, dedicated storage (n-ceph, 2.9 PiB) for neutrino experiments
- **NICA Projects:** BM@N (16 VMs, 96 cores) and SPD (46 VMs, 180 cores) - distinct research-specific resource groupings within the common resources

Server & Core Distribution

Common Cloud	Neutrino Platform
23% of total servers	77% of total servers

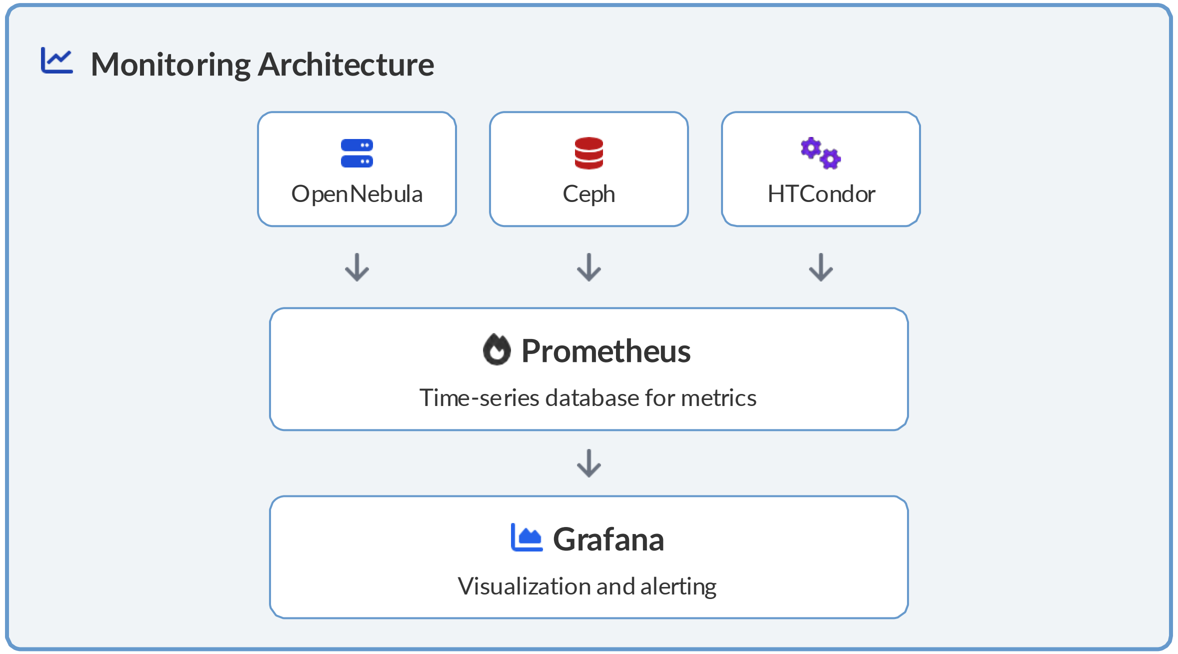


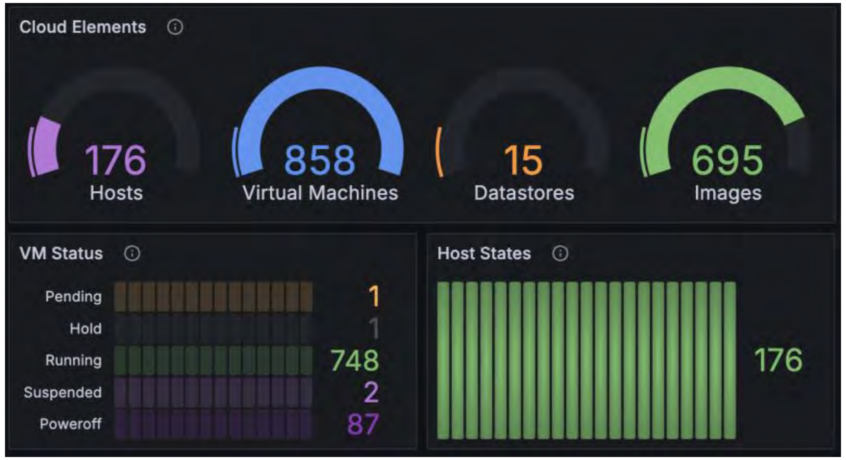
Neutrino Platform Allocation

Experiment	CPU Cores	RAM (TB)	Storage (PB)
Baikal-GVD	484	7.8	0.6
JUNO Grid	2,232	37.0	0.1
NOvA	720	4.0	2.1
Neutrino Common	944	17.5	0.0
TAIGA	0	0.0	0.1
Total	**4,380**	**66.3**	**2.9**

Monitoring Infrastructure

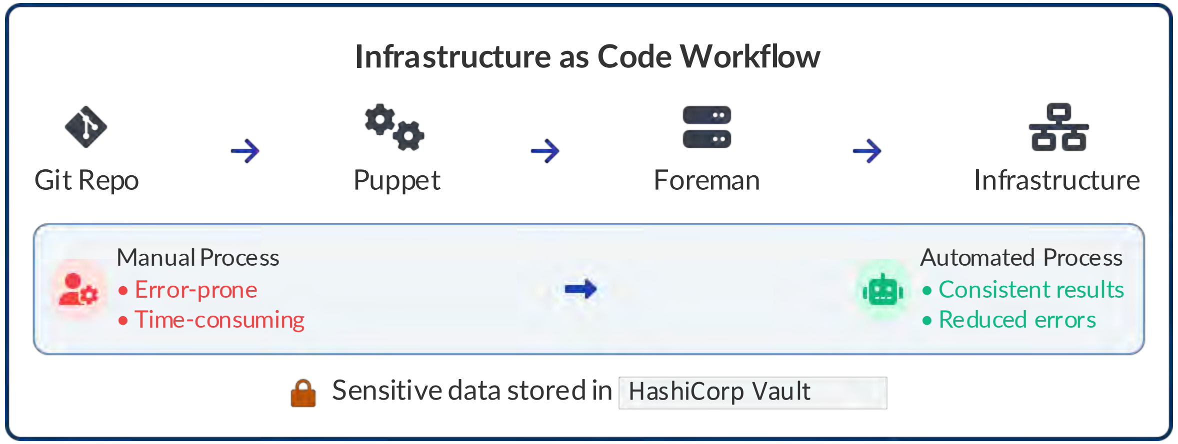
- **Core monitoring stack:** Prometheus for data storage and Grafana for visualization
- **OpenNebula metrics:** Migrated from custom collector to native OpenNebula metrics (available since CE 6.10+)
- Working on adapting custom Grafana panels to the new metrics source
- **Ceph monitoring:** Using built-in Ceph Prometheus module with exporters and official Grafana panels
- **Custom exporters:** HTCondor exporter and other service-specific data collectors
- **Real-time visibility:** Performance, availability, and capacity across all infrastructure components

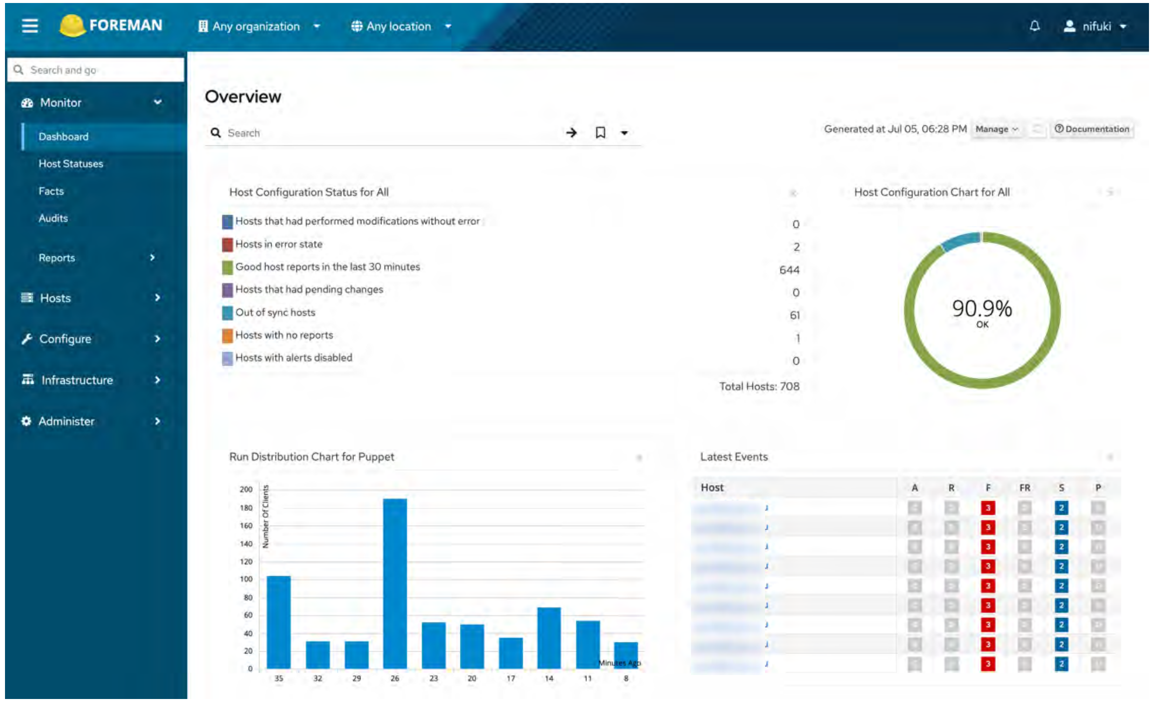




Infrastructure Management

- **Infrastructure as Code (IaaC):** Automated provisioning and configuration management
- **Minimize manual intervention:** Reduce human error and increase consistency through automated workflows
- **Foreman:** Server provisioning and configuration management
- **Puppet:** Configuration management using Profile + Role model for both physical and virtual machines
- **Git integration:** Version controlled infrastructure with Puppet manifests managed via git
- **Security management:** Sensitive information stored in HashiCorp Vault



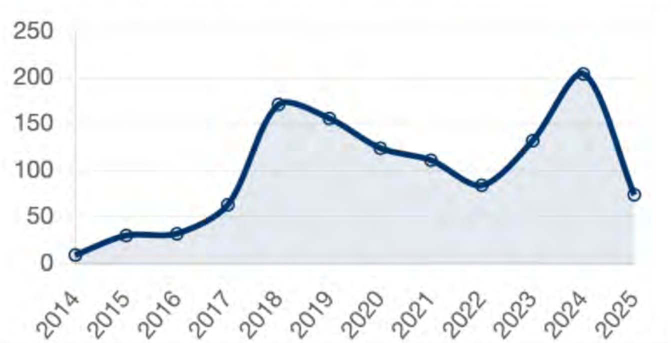


iTop: Support System & Inventory

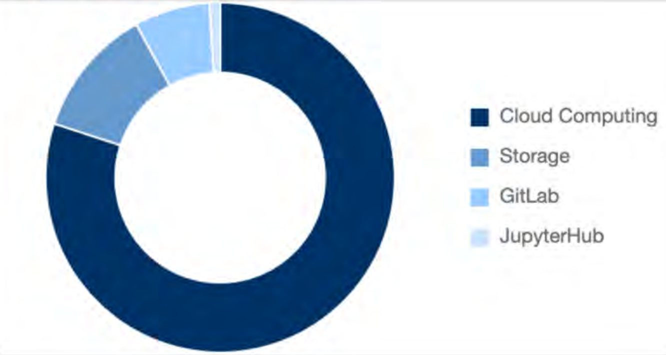
- **Dual purpose system:** Integrated solution for both hardware inventory management and user support helpdesk
- **Hardware tracking:** Complete lifecycle management of servers and stock of spare parts
- **Ticket system:** Request tracking, statistics, and resolution workflows for cloud and storage support
- **Automation interface:** REST API for programmatic integration with server testing system
- **Communication challenges:** Users sometimes bypass helpdesk via direct email/messaging, reducing ticket visibility and hindering prioritization

User Communication Channels
Official tickets | Direct emails | Messengers

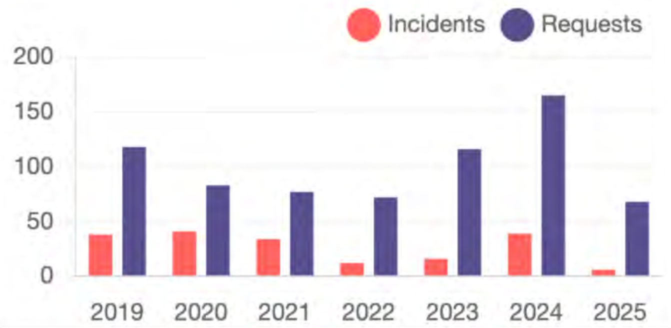
Total Tickets by Year (2014-2025)



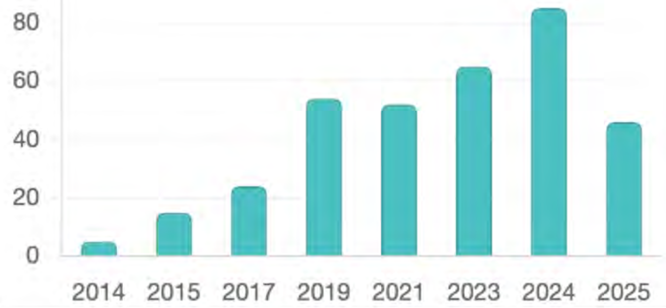
Service Distribution



Incidents vs Requests



User Dynamics by Year



Key Insights
- Peak activity in 2024 (204 tickets)
- Requests dominate (68-74%)
- 85 unique users in 2024 (17x growth)
- Cloud computing: 80% of all tickets

T-Cloud and Student Research Projects

- **T-Cloud purpose:** Dedicated environment for both testing and training activities
- **Training capabilities:** User onboarding, administrator practice, classroom environment for workshops
- **Testing platform:** Safe environment for software testing, integration testing, and infrastructure validation
- **Student engagement:** Active opportunities for students to work on real infrastructure challenges
- **Knowledge transfer:** Bridge between academic research and production infrastructure needs

T-Cloud Specifications

Hardware:
- 32 x Supermicro servers
- 384 CPU cores total
- 768 GB RAM total

Software:
- OpenNebula 6.0.0.2
- Ceph v17.2.5
- ~400 TB raw storage

t-cloud.jinr.ru

2025 Student Research Projects

Hardware Failure Prediction System

Master Thesis Project #1

Technology Stack:
- SMART data collection
- OpenSearch for storage
- Statistical analysis
- Predictive visualization

Benefits:
- Predict disk failures before they occur
- Reduce unplanned downtime
- Optimize hardware replacement cycles

Automated Server Stress Testing Suite

Master Thesis Project #2

Technology Stack:
- Stress-ng for component testing
- HASS methodology implementation
- Prometheus monitoring integration
- iTop inventory integration

Benefits:
- Early detection of hardware defects
- Thermal stress pattern identification
- Automated burn-in testing
- Integrated hardware lifecycle

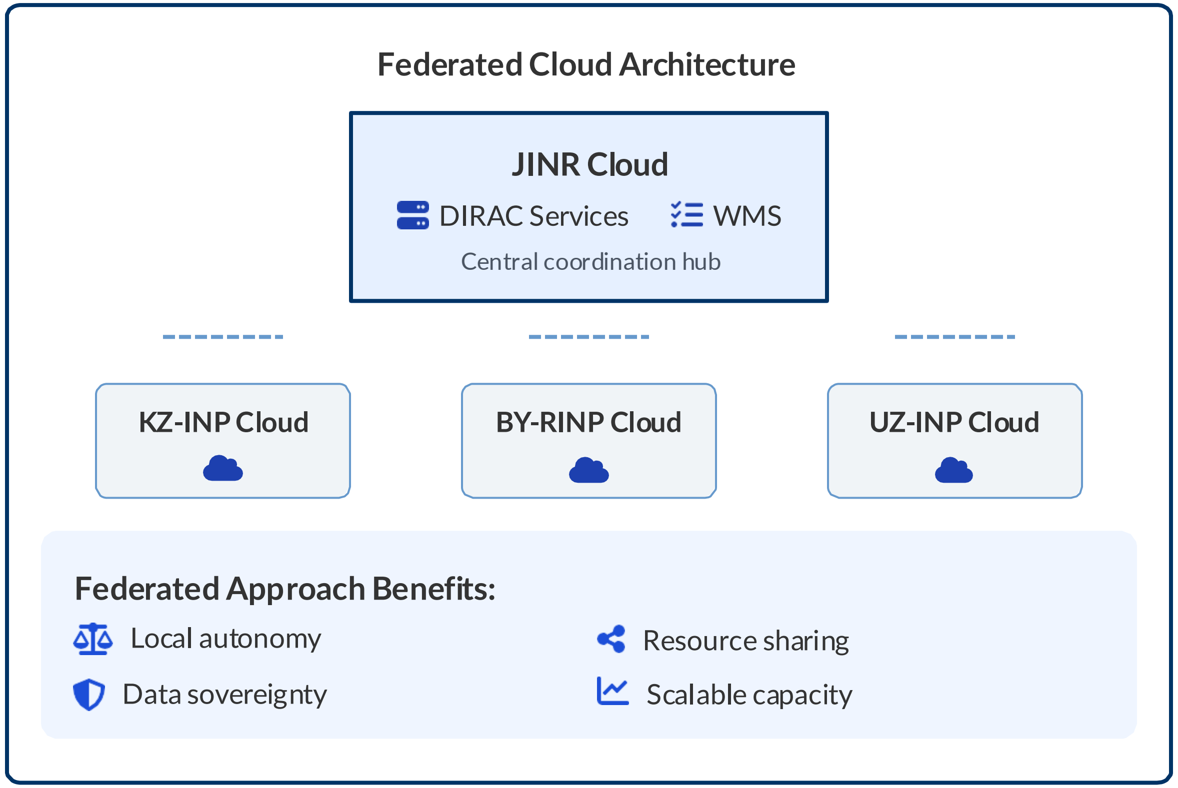
DICE

Distributed Information and Computational Environment

> Joining small institutional clouds into a federated ecosystem.
> Enabling shared scientific computing across JINR Member States.

DICE Mission and Structure

- **Federation mission:** Connect clouds from JINR Member States into a unified computational environment
- **Resource sovereignty:** Organizations maintain control of their resources while sharing excess capacity
- **Technology standardization:** OpenNebula-based clouds at partner sites for unified management
- **Workload distribution:** DIRAC WMS as the central job management system
- **Knowledge transfer:** Disseminating cloud and grid technologies across Member State organizations
- **Peak load handling:** Distributed resources absorb computational spikes across the federation



DICE Participants & Resource Overview

Resource Statistics

Total Organizations:	10
Total CPU Cores:	762
Total RAM:	4722 GB
Total Storage:	287 TB

Status Legend

- Active
- Pending

Participant Countries

- **Active:** Kazakhstan, Belarus, Uzbekistan, Russia, Bulgaria
- **Pending:** Azerbaijan, Bulgaria, Egypt, Georgia

Organization	Country	Status	Cores	RAM (GB)	Storage (TB)
Institute of Nuclear Physics	KZ	Active	84	840	6.8 (SSD)
Research Institute of Nuclear Problems	BY	Active	132	290	127
Institute of Nuclear Physics	UZ	Active	98	890	6.6 (SSD)
Total			**314**	**2020**	**140.4**

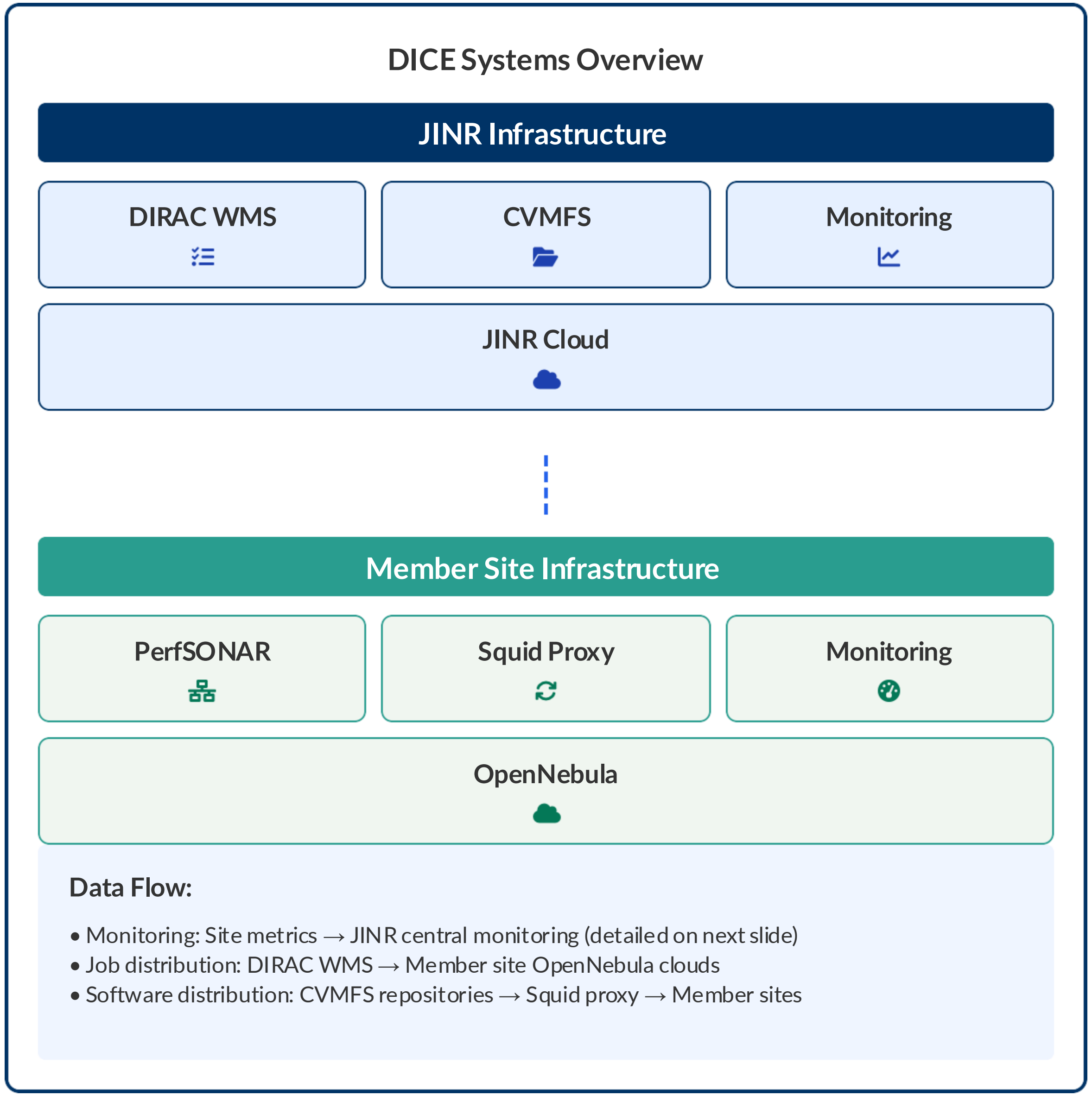
Main Components of DICE

At JINR:

- **DIRAC WMS:** Central workload management system for job distribution
- **CVMFS repositories:** Software distribution across federation sites
- **Monitoring components:** Aggregated metrics collection, S3 gateway, Thanos store+query

At each site:

- **OpenNebula cloud:** Local virtualization platform
- **Squid proxy:** Caching CVMFS traffic for efficient software distribution
- **Monitoring components:** Collectors, exporters, local storage, Thanos sidecar
- **PerfSONAR:** Network performance monitoring

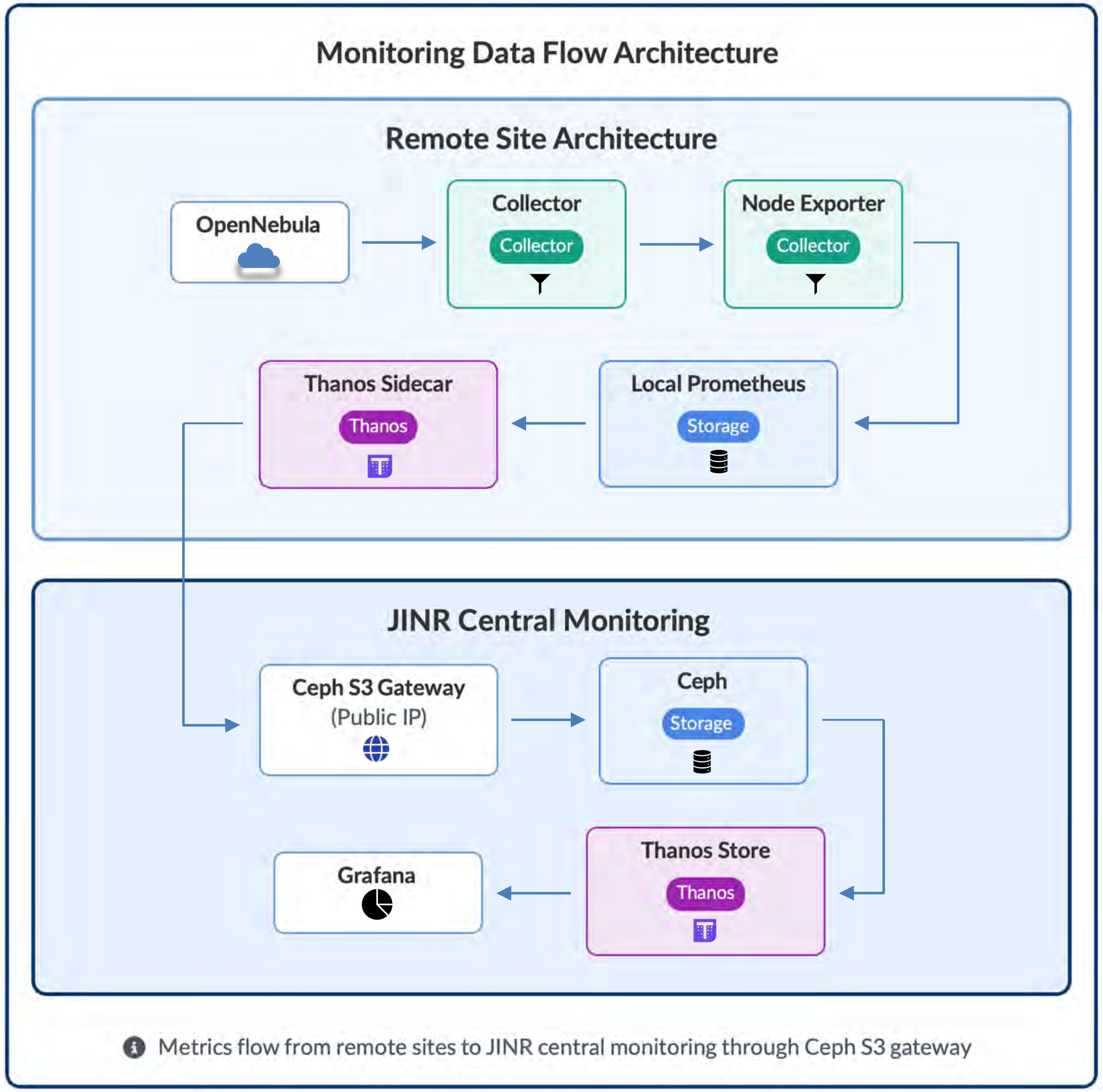


DICE Remote Site Monitoring

- **Remote site monitoring:** Collecting and aggregating metrics from member cloud infrastructures despite NAT and network constraints
- **Local metrics collection:** Each site collects OpenNebula and hardware metrics through node_exporter and local Prometheus
- **Thanos architecture:** Long-term metrics storage with Thanos sidecar and store components
- **S3 gateway bridge:** Public IP endpoint allowing remote sites to push metrics through network restrictions
- **Centralized visualization:** Unified Grafana dashboards at JINR for monitoring all federation resources

Key Benefits:

- NAT-compatible architecture
- Long-term metrics storage
- Unified visualization
- Scalable to all member sites



Use Case Highlight

Classical Three-body Gravitational Problem

- C-code developed by **Ivan Hristov** and **Radoslava Hristova (Sofia University)** successfully ported to run on DICE resources
- **Problem type 1:** Computing decay times of the three-body system in considered sections
- **Problem type 2:** Searching for periodic orbits with Newton's method in considered sections
- Ideal for DICE's distributed approach: minimal data transfer requirements

Monte-Carlo simulation on DICE

- Wide-spread in physics
- Often optimal for limited-bandwidth resources: small input data, computation-intensive workloads

Three-Body Problem Visualization

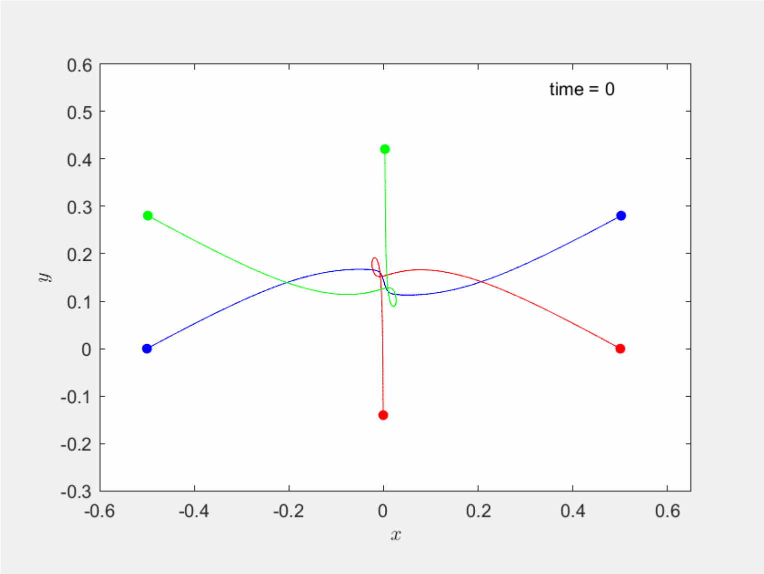


Illustration of a three-body system with gravitational interactions

Dice Information Portal

Web: dice.jinr.ru | **Email:** dice@jinr.ru

Conclusions and Plans

- JINR Cloud is growing steadily to meet the demand
- Continue our efforts on development of DICE and porting user workloads
- Rebuild custom Grafana panels to the new metrics source
- Commission a web-interface for managing Ceph storages
- Migrate from Puppet to Open Vox
- Consider sharing Foreman/Puppet with other MICC teams

Thanks!

Web: cloud-info.jinr.ru | **Email:** coud-info@jinr.ru