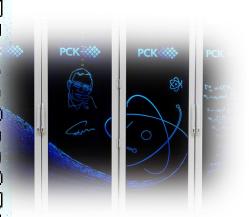


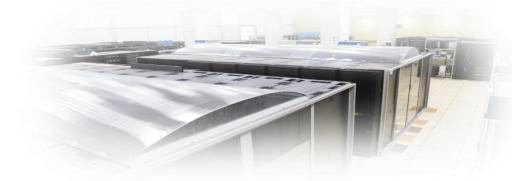


# Distributed Computing at JINR: yesterday, today, tomorrow

#### **Vladimir Korenkov**

**Meshcheryakov Laboratory of Information Technologies** 





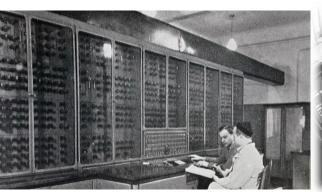
The International Conference GRID-2025 Dubna, JINR, 7 July 2025

#### **Meshcheryakov Laboratory of Information Technologies**





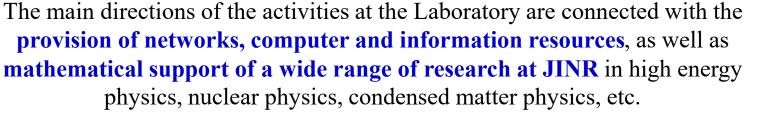
**M.G. Meshcheryakov** (17.09.1910 - 24.05.1994)







The Laboratory of Computing Techniques and Automation (now MLIT) was founded in August 1966.





**N.N. Govorun** (18.03.1930 - 21.07.1989)







# Distributed Computing at JINR: first steps



### Condor

# **NIKHEF 1** 15 Sparc stations **FWI-UVA 1** 15 Sparc **CERN SMC** 30 Sparc stations **CS Madison UW** 250 workstations DEC, Sparc, HP, SGI, R6000

### **PC-cluster**



**CCIC JINR** 

150 CPU

**18TB RAID-5** 

39TB Certon100

8 - Interactive & UI

32 - Common PC-farm

**MYRINET** (Parallel)

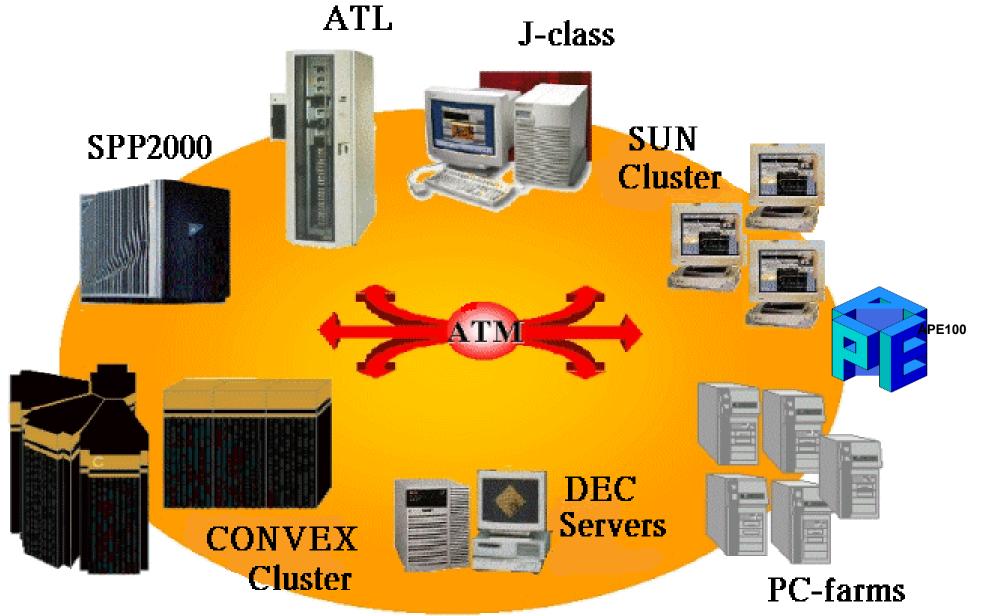
52 - LCG

61 - servers









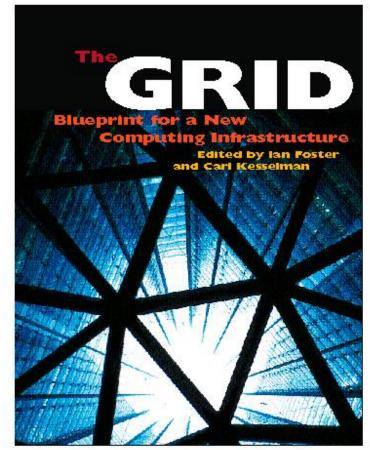
### What is GRID



# Five Emerging Models of Networked Computing From The Grid:

- Distributed Computing
- High-Throughput Computing
- **❖**On-Demand Computing
- Data-Intensive Computing
- Collaborative Computing

Ian Foster and Carl Kesselman, editors, "The Grid: Blueprint for a New Computing Infrastructure," Morgan Kaufmann, 1999, http://www.mkp.com/grids

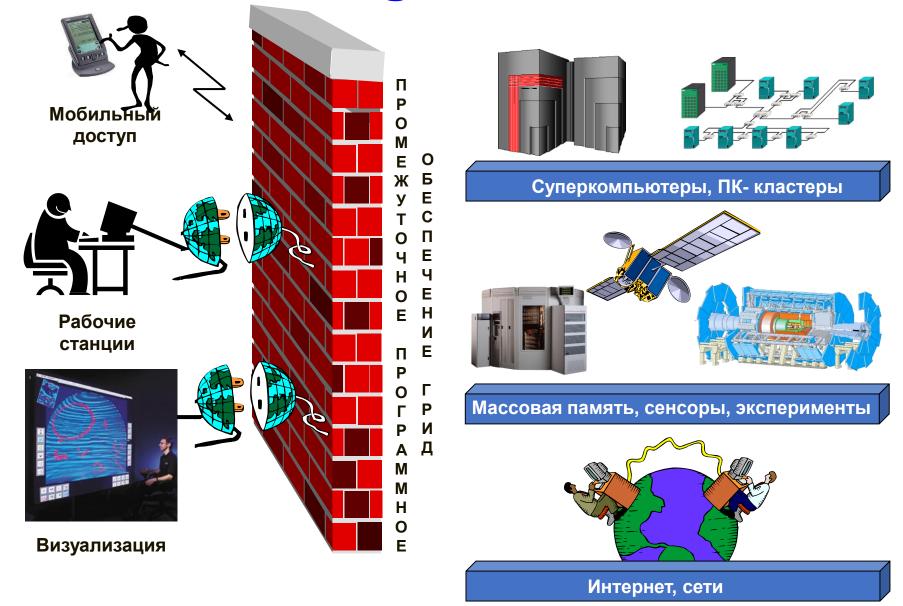


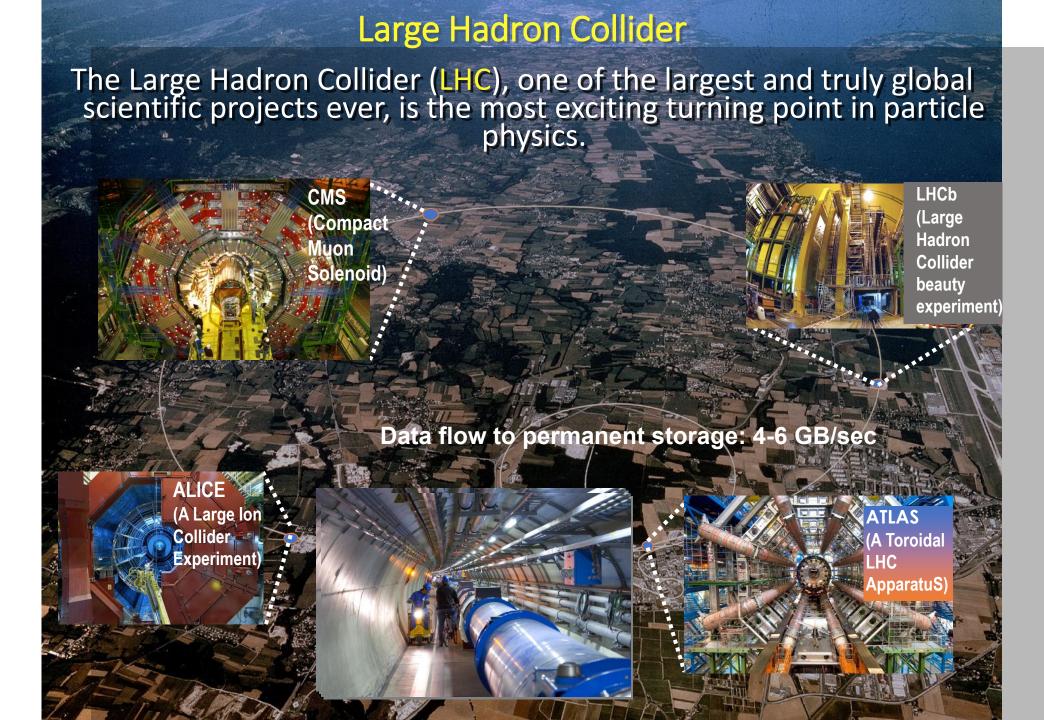




# GRID- is a means for sharing computing power and data storage via the Internet



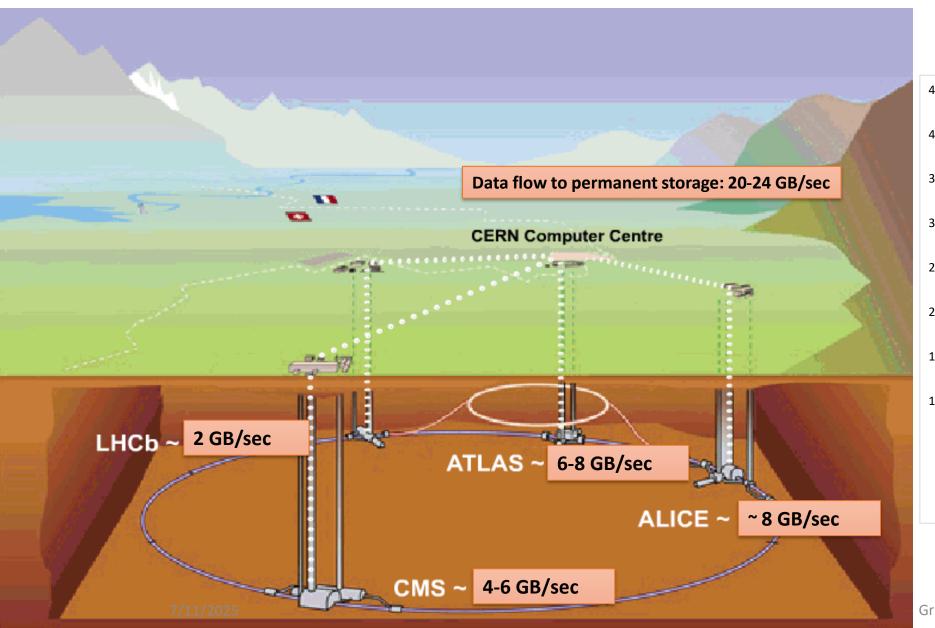


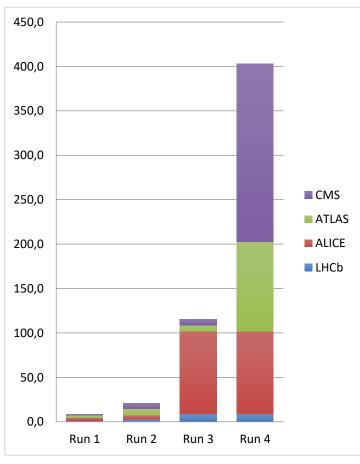




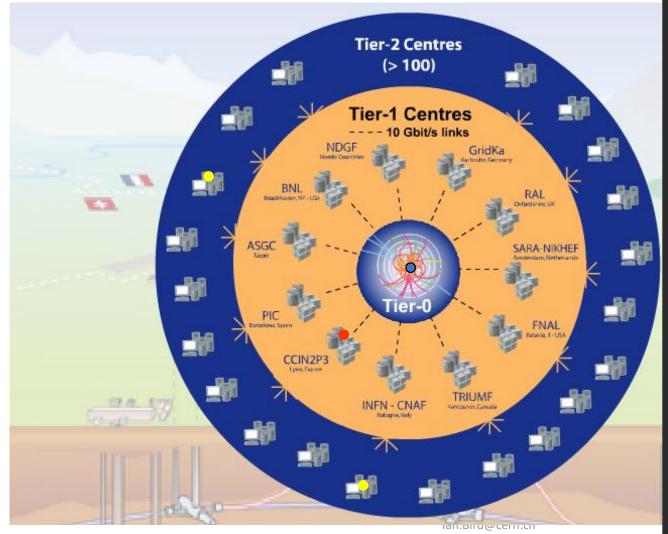
# **Data Collection and Archiving at CERN**







#### Tier Structure of GRID Distributed Computing: Tier-0/Tier-1/Tier-2



#### Tier-0 (CERN):

- accepts data from the CMS
   Online Data Acquisition
   and Trigger System
- archives RAW data
- the first pass of reconstruction and performs Prompt Calibration
- data distribution to Tier-1

#### Tier-1 (11 centers):

- receives a data from the Tier-0
- data processing (rereconstruction, skimming, calibration etc)
- distributes data and MC to the other Tier-1 and Tier-2
- secure storage and redistribution for data and MC

#### **Tier-2 (>200 centers):**

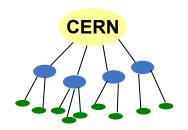
- simulation
- user physics analysis



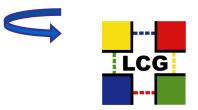
## **Some history**



- 1999 Monarc Project
  - Early discussions on how to organise distributed computing for LHC
- 2001-2003 EU DataGrid project
  - middleware & testbed for an operational grid
- 2003 RDIG in Russia (Tier2)
- 2002-2005 LHC Computing Grid LCG
- 2004-2006 EU EGEE project phase 1
- 2005-2006 PanDA and DIRAC
- 2006-2008 EU EGEE-II
- 2008-2010 EU EGEE-III
- 2010-2012 EGI-InSPIRE
- 2010 GRID-Cloud
- 2012 discovery of the Higgs Boson
- 2013 Tier1 in Russia (JINR and KI)
- 2013 GRID-Supercomputer TITAN



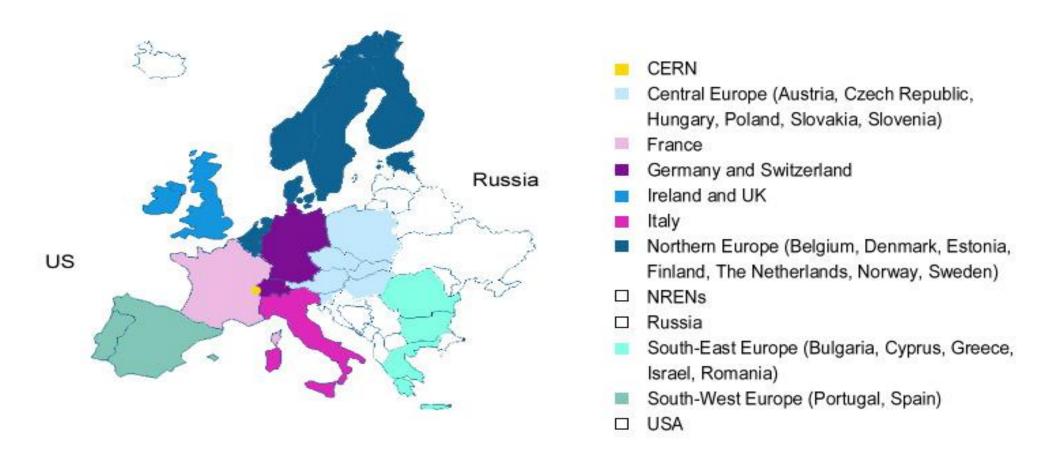






## **EGEE - Enabling Grids for E-sciencE**





The EGEE project - Enabling Grids for E-sciencE aims to create international Grid-based workforces. The project is being implemented by a consortium of **70 institutes in 27 countries**, organized into regional grids.

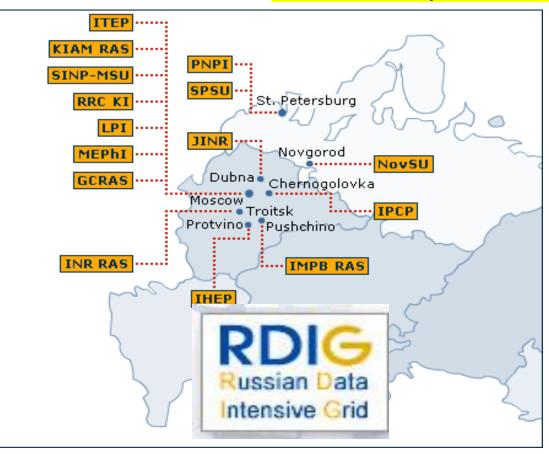
# JINR in the Russian Data Intensive Grid infrastructure (RDIG)



The Russian consortium RDIG (Russian Data Intensive Grid), was set up in September 2003 as a national federation in the EGEE project.

Now the RDIG infrastructure comprises 17 Resource Centers with

> 5000 CPU (12000 kSI2K) and > 3200 TB of disc storage.



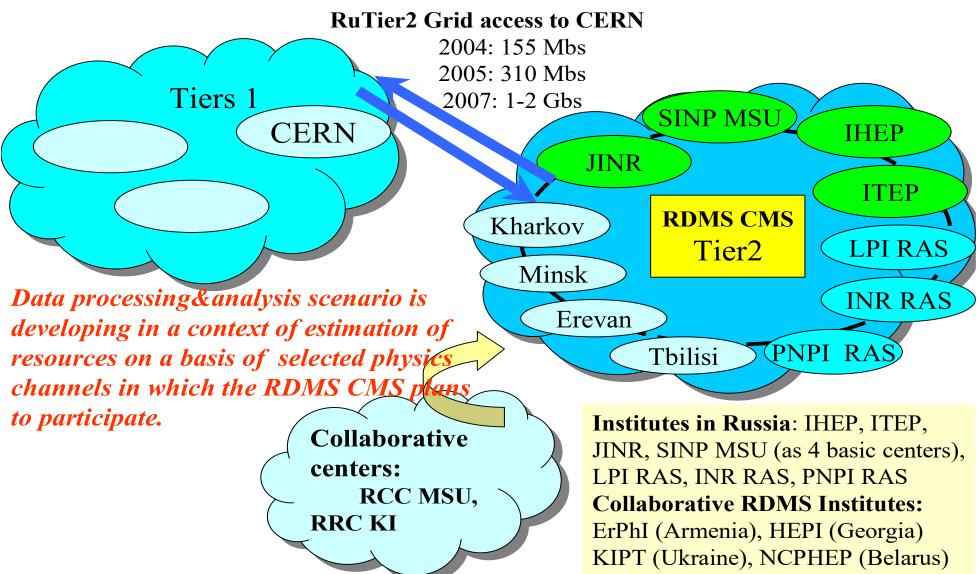
**RDIG Resource** 

**Centres:** 

- ITEP
- JINR-LCG2
- Kharkov-KIPT
- RRC-KI
- RU-Moscow-KIAM
- RU-Phys-SPbSU
- RU-Protvino-IHEP
- RU-SPbSU
- Ru-Troitsk-INR
- ru-IMPB-LCG2
- ru-Moscow-FIAN
- ru-Moscow-GCRAS
- ru-Moscow-MEPHI
- ru-PNPI-LCG2
- ru-Moscow-SINP
- BY-NCPHEP

# **Development of RDMS CMS computing model**





**RDMS CMS computing centers** 



# RDIG monitoring&accounting http://rocmon.jinr.ru:8080



- Monitoring allows to keep an eye on parameters of Grid sites' operation in real time
- Accounting resources utilization on Grid sites by virtual organizations and single users

#### **Monitored values**

**CPUs** - total /working / down/ free / busy

Jobs - running / waiting

Storage space - used / available

Network - Available bandwidth

**Accounting values** 

Number of submitted jobs

#### **Used CPU time**

Totally sum in seconds

Normalized (with WNs productivity)

Average time per job

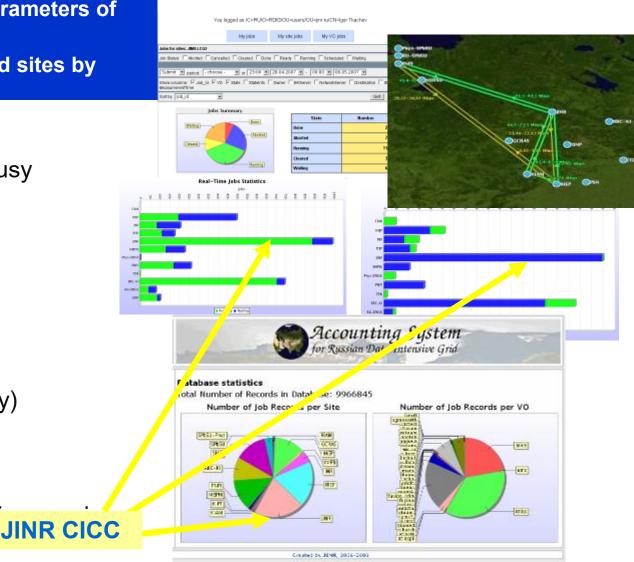
#### Waiting time

Totally sum in seconds

Average ratio waiting/used CPU

#### Physical memory

Average per job

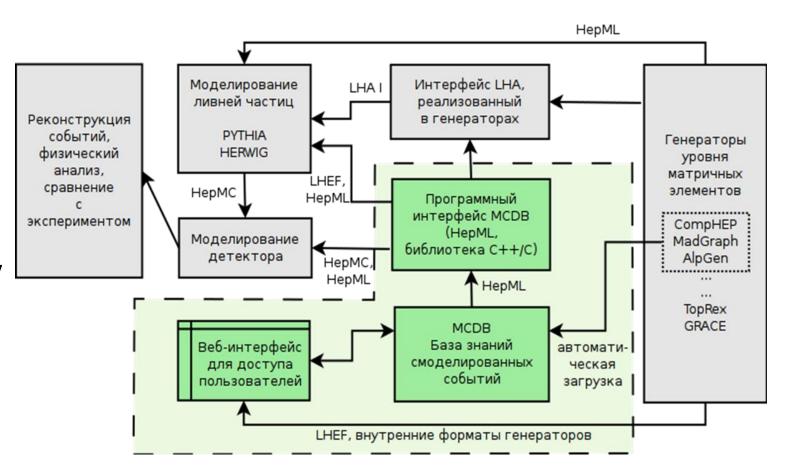




### **WLCG MCDB - project**



- MCDB access libraries have been integrated to CMSSW software
- Several improvements have been made in MCDB software
- New improved XML schema has been developed for High Energy Physics Markup Language (HepML)
- The program libraries have been developed to work with new HepML schema

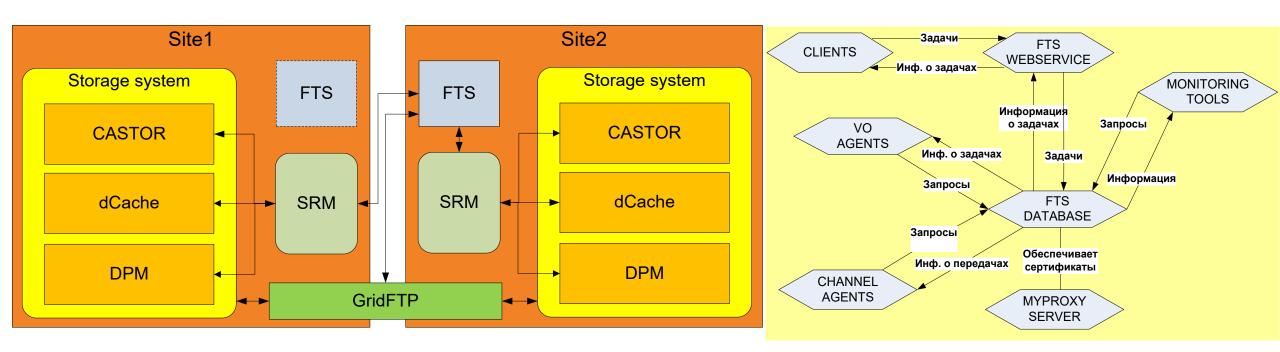




# FTS (FILE TRANSFER SERVICE) MONITORING FOR WORLDWIDE LHC COMPUTING GRID PROJECT



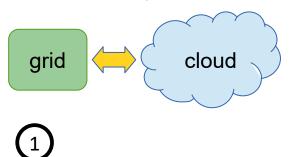
A monitoring system is developed which provides a convenient and reliable tool for receiving detailed information about the FTS current state and the analysis of errors on data transfer channels, maintaining FTS functionality and optimization of the technical support process. The system could seriously improve the FTS reliability and performance.



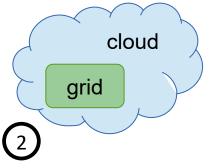
## **Clouds and grids**

MLIT

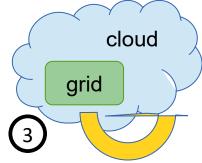
- Both ones share two main attributes:
  - they provide access to remote computing resources; and
  - they provide a service.
- But they are based on different paradigms:
  - clouds are being based on virtualization of resources,
  - grids being based on the sharing of resources across boundaries.
- Modern trend is a synthesis of these two technologies:



Cloud resources suppliments grid ones (e.g. during peak load to provide required QoS)



Services of grid site are deployed on cloud VMs (to increase hardware utilization efficiency and simplify admins' work)

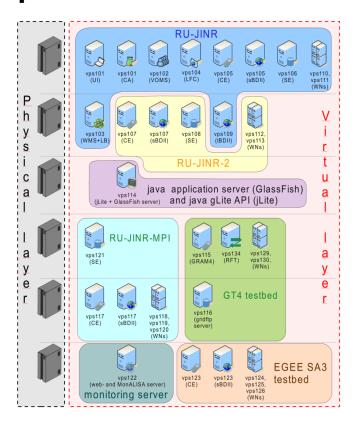


Synergy of 1st and 2nd approach

Grid training and education - distributed training

infrastructure

gLite user trainings for students of Dubna University and University Centre of JINR, grid site administrators trainings for JINR member-states, testbed for grid developers, testbed for middleware evaluation, GILDA cooperation



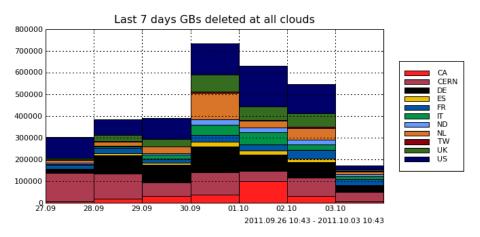


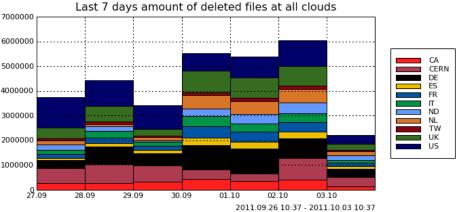


### **DDM DQ2 Deletion service**



- The ATLAS Distributed Data Management project DQ2 is responsible for the replication, access and bookkeeping of ATLAS data across more than 120 distributed grid sites. It also enforces data management policies decided on by the collaboration and defined in the ATLAS computing model.
- The DQ2 Deletion Service is one of the most important DDM services. This distributed service interacts with 3rd party grid middleware and the DQ2 catalogues to serve data deletion requests on the grid. Furthermore, it also takes care of retry strategies, check-pointing transactions, load management and fault tolerance.
- Current version of Deletion Service was developed (and maintained) by JINR LIT specialists and is used by ATLAS Distributed Computing

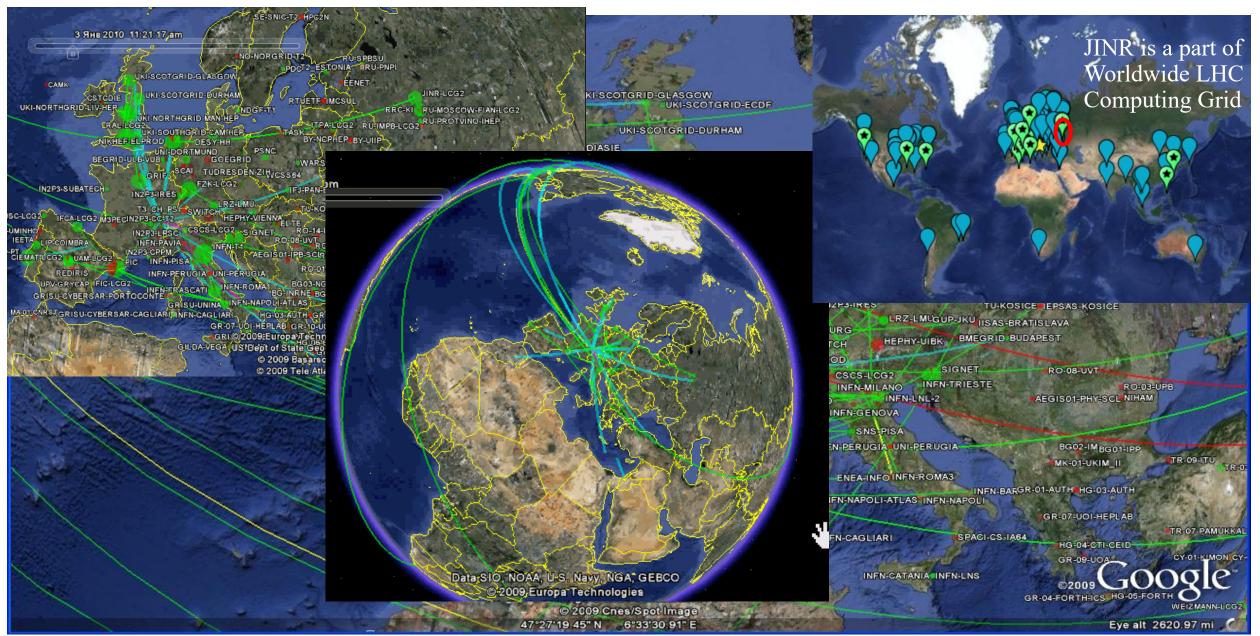




Deletion Service serves more than 120 sites. In usual operation it deletes 2-2,5M of files per day, which correspond to 400 - 500 TB per day. During the deletion campaigns when deletion was carried out on most sites, deletion rate achieved is more than 6M of files per day, reaching up to 300k files per hour.

## The Worldwide LHC Computing Grid (WLCG)

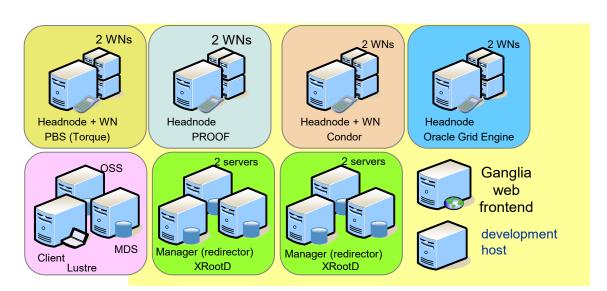


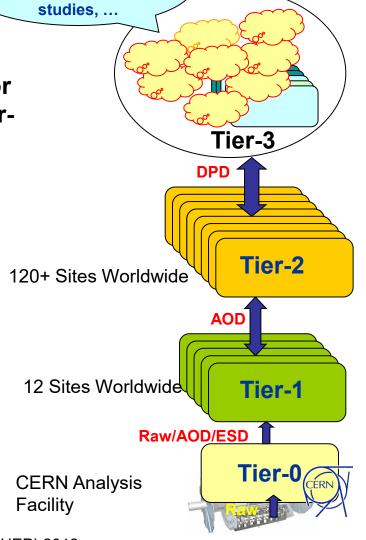


## Tier3 sites monitoring project



- Traditional LHC Distributed Computing Tier-0 (CERN) → Tier-1 → Tier-2
- Additional → Tier-3
- Needs → a global view of the LHC computing activities
- LIT participates in the development of a software suite for Tier-3 sites global monitoring and ATLAS off-Grid sites (Tier-3) monitoring
- A virtual testbed has been created at JINR which allows simulation of various Tier3 clusters and solutions for data storage



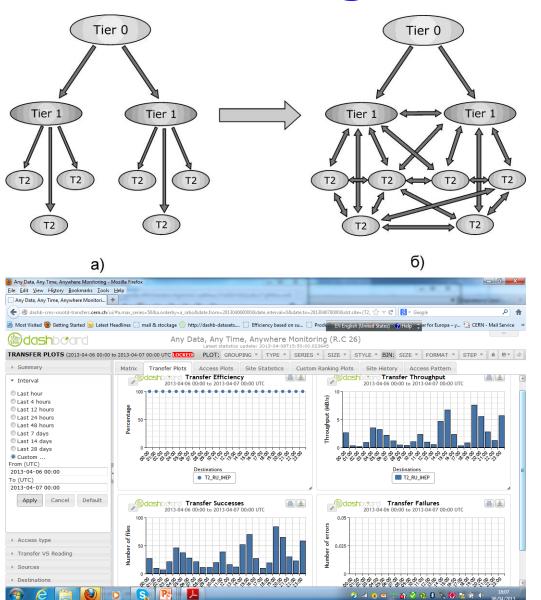


nteractive analysis plots, fits, toy MC,

Poster reports at the Conference "Computing in High Energy and Nuclear Physics" (CHEP) 2012 21-25 May 2012 New York City, NY, USA

# **Monitoring of the XRootD federations**





- The computing models of the LHC experiments are gradually moving from hierarchical data models with centrally managed data pre-placement towards federated storage which provides seamless access to data files independently of their location and dramatically improved recovery due to fail-over mechanisms.
- Construction of the data federations and understanding the impact of the new approach to data management on user analysis requires complete and detailed monitoring.
- In the WLCG context, there are several federations currently based on the XRootD technology. Monitoring applications for ATLAS XRootD federation (FAX Dashboard) and CMS XRootD federation (AAA Dashboard) were developed in collaboration of CERN IT and JINR LIT



## **WLCG Transfer Dashboard**



- Along with data processing, data distribution is the key computing activity on the WLCG infrastructure.
- Monitoring of the data distribution is a challenging task because of the high scale of the activity and the heterogeneity of the infrastructure (various storage implementations and transport protocols used)
- The WLCG Transfer Dashboard developed during last two years, provides cross-experiment and cross-technology view of data transfers performed by the LHC experiments on the WLCG infrastructure
- JINR LIT actively participated in the development,
- Namely
  - development of the File Transfer Service (FTS) publisher to ActiveMQ,
  - monitoring of the state of the FTS queues,
  - integration of the traffic of the ALICE experiment which in difference with other LHC experiments does not use File Transfer Service (FTS)

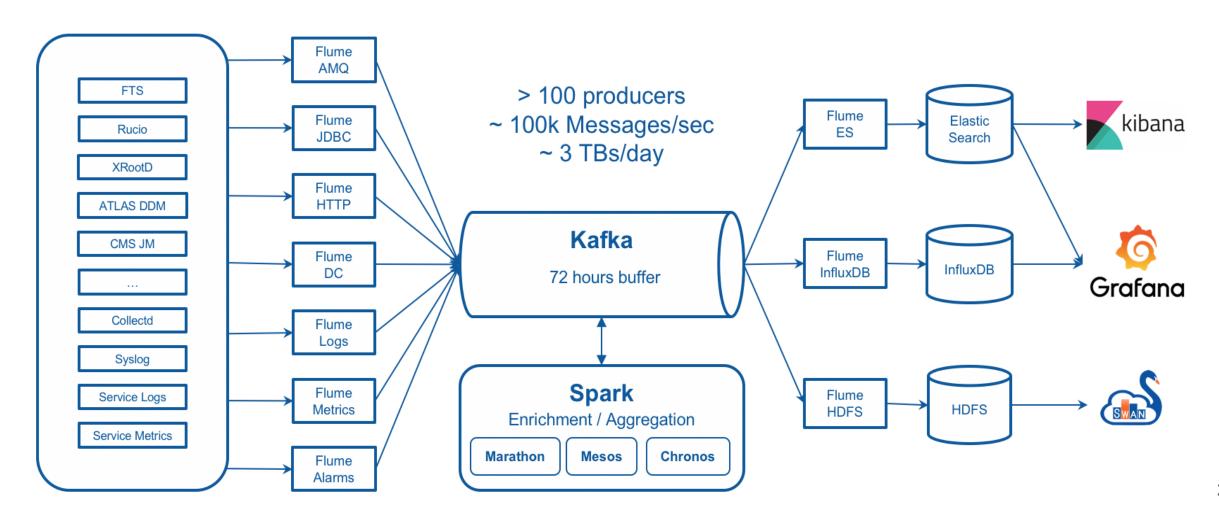


WLCG home page contains the plot of the WLCG Transfer Dashboard

# Organization of monitoring using Big Data technologies



Sources > Transport > (Processing) > Storage > Access



## Implementation of file transfer monitoring in ATLAS



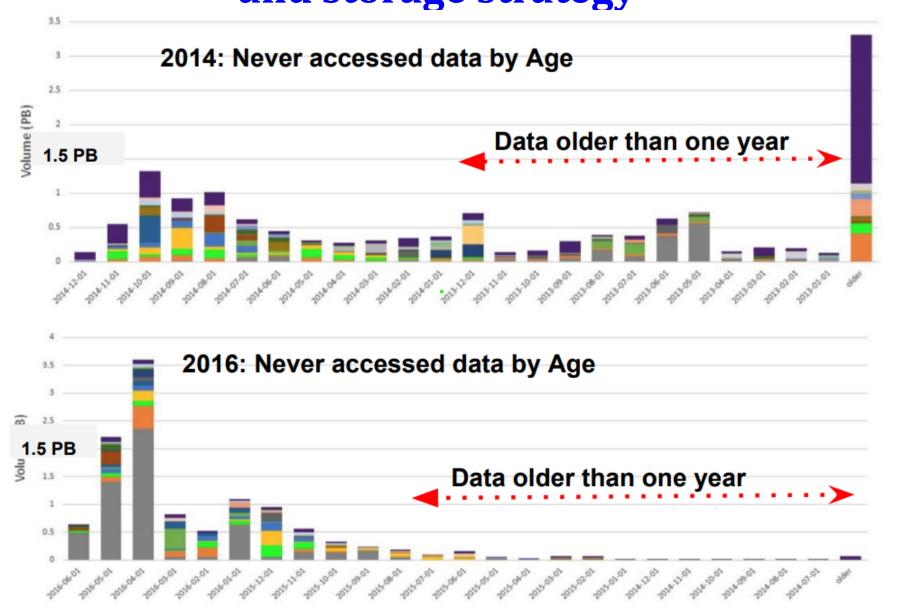


- Scaling: millions of transmissions per day
- Monitoring of both individual transmissions and global processes – close to real time
- Statistical slices by various parameters

Statistic of data transfer monitoring in ATLAS (project MONIT). Sept-Okt 2024

# Analysis of data demand and change of data replication and storage strategy





## Remote access to ATLAS and CMS

MET

- System of remote access in real time (SRART) for monitoring and quality assessment of data from the ATLAS at JINR
  - One of the most significant results of the team TDAQ ATLAS at LIT during the last few years was the participation in the development of the project TDAQ ATLAS at CERN.
- JINR CMS Remote Operation Centre
  - Monitoring of detector systems
  - Data Monitoring / Express Analysis
  - Shift Operations (except for run control)
  - Communications of JINR shifter with personal at CMS Control Room (SX5) and CMS Meyrin centre
  - Communications between JINR experts and CMS shifters
  - Coordination of data processing and data management
  - Training and Information





### Extending PanDA to Oak Ridge Leadership Computing Facilities



Multicore WN

Multicore WN

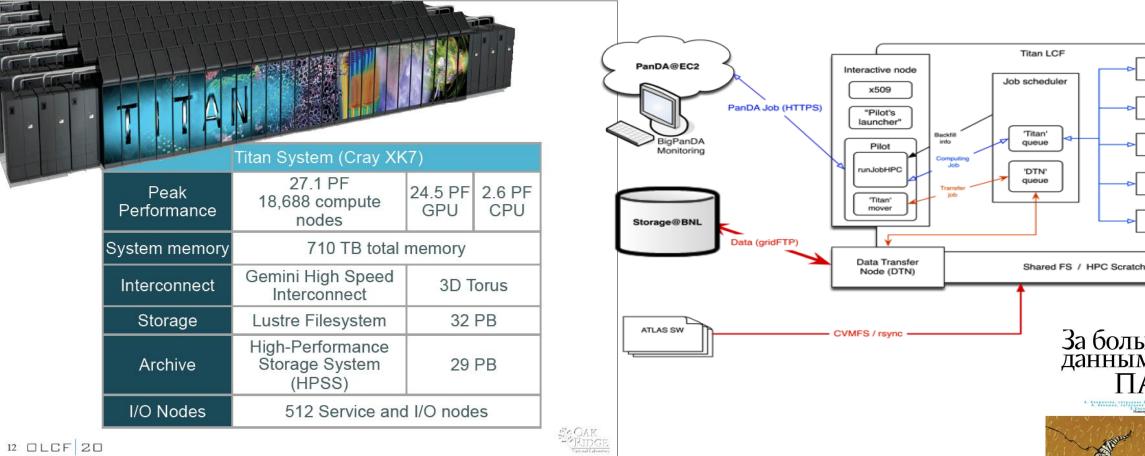
Multicore WN

Multicore WN

Multicore WN

За большими

данными следит



PanDA development in the direction of integration of various distributed and parallel computing systems (grid, cloud, clusters, data centers, supercomputers) with the aim of creating a universal platform for largescale big data management projects.



### **New COMPASS Production System**



**COMPASS Grid Production System** was developed and provides automated task processing from definition till archiving

#### **Key features:**

Production management via Web UI, allows one to define a task, send, follow and manage it during lifecycle.

Via **PanDA** job execution layer jobs are being sent to any available type of computing resource: **Condor, LSF, PBS, etc.** 

**Computing sites:** CERN Tier-1, JINR Tier-2

**Storage:** EOS and CASTOR at CERN

## All management services deployed at JINR Cloud Service

In production since August, statistics:

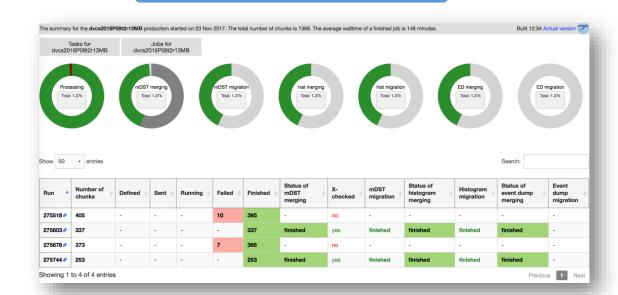
- •~1 000 000 chunks of raw data processed,
- •~30 billions of events,
- •~200 TB of data produced,
- •~3 000 000 jobs processed:
- reconstruction, ddd filtering,
- merging of mDST, histograms
- and event dumps.

Tasks requests layer: Web UI

Task definition layer: ProdSys

Job definition layer: ProdSys

Job execution layer: PanDA





# **GridNNN infrastructure**



- Grid support for Russian national nanotechnology network
  - To provide for science and industry an effective access to the distributed computational, informational and networking facilities
  - Expecting breakthrough in nanotechnologies
  - Supported by the special federal program
- Main points
  - based on a network of supercomputers (about 15-30)
  - has two grid operations centers (main and backup)
  - is a set of grid services with unified interface
  - partially based on Globus Toolkit 4

10 resource centers in different regions of Russia

 RRC KI, «Chebyshev» (MSU), IPCP RAS, CC FEB RAS, ICMM RAS, JINR, SINP MSU, PNPI, KNC RAS, SPbSU











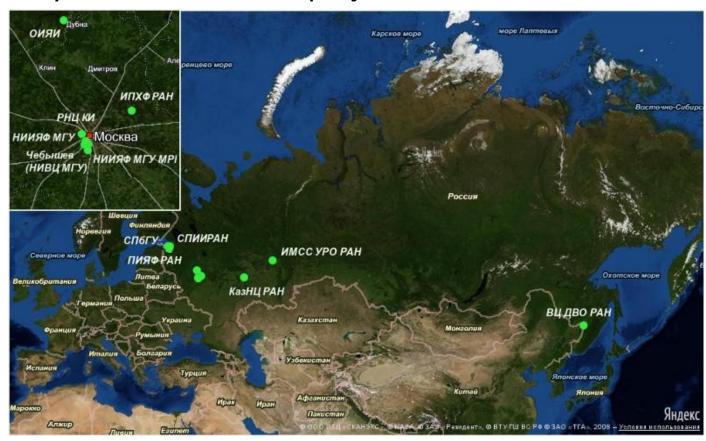




## **Russian Grid Network**



- Goal to make a computational base for hi-tech industry and science
- Using the network of supercomputers and original software created within recently finished GridNNN project
- Some statistics
  - 19 resource centers
  - 10 virtual organizations
  - 70 users
  - more than 500'000 tasks processed



### Tier1 center

March 2011 - Proposal to create the LCG Tier1 center in Russia (official letter by Minister of Science and Education of Russia A. Fursenko has been sent to CERN DG R. Heuer):

NRC KI for ALICE, ATLAS, and LHC-B LIT JINR (Dubna) for the CMS experiment

<u>The Federal Target Programme Project</u>: «Creation of the automated system of data processing for experiments at the LHC of Tier-1 level and maintenance of Grid services for a distributed analysis of these data»

**Duration: 2011 – 2013** 

September 2012 – Proposal was reviewed by WLCG OB and JINR and NRC KI Tier1 sites were accepted as a new "Associate Tier1"

Full resources - in 2014 to meet the start of next working LHC session.



#### ORGANISATION EUROPEENNE POUR LA RECHERCHE NUCLEAIRE EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH



Laboratoire Européen pour la Physique des Particules European Laboratory for Particle Physics

Mail address: Dr. Ian Bird CERN, IT Department CH-1211 GENEVE 23 Switzerland Tel: +41 22 767 5888

E-mail: lan.Bird@cern.ch

Prof. Mikhail Kovalchuk Director of National Research Centre "Kurchatov Institute" 1, Akademia Kurchatova pl., Moscow 123182, Russia

Prof. Victor Matveev Director of Joint Institute for Nuclear Research Joliot-Curie 6 141980 Dubna, Moscow Region, Russia

Votre référence/Your reference: Notre référence/Our reference:

Subject: Acceptance of the proposal to build Tier 1 centres in Russia

Geneva, October 12, 2012

Dear Directors,

As you know, the proposals from the National Research Centre — "Kurchatov Institute" and the Joint Institute for Nuclear Research, Dubna, to build Tier 1 centres for LHC data analysis were discussed in the recent WLCG Overview Board held on September 28. I am very happy to report that the proposals were well received by the members of the board, and that the decision was made to accept the Russian sites as a new "Associate Tier 1". This decision will be noted in the formal minutes of the meeting.

The next step is now to proceed to signing the WLCG Memorandum of Understanding. The WLCG project office will assist in drafting this MoU, which should be signed by the relevant funding agencies for the two Russian Institutes, or their designated agents.

I am at your disposal for any assistance or to provide further details of the process.

Yours Sincerely,

r. lan Bird

LHC Computing Grid Project Leader

IT Department

CERN



# **CMS Tier-1 at JINR**

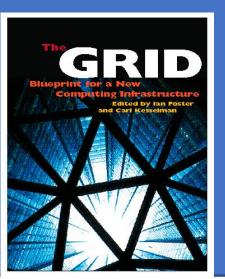
	2012	2013	2014
CPU (HEPSpec06) Number of core	14400 1200	28800 2400	57600 4800
Disk (Terabytes)	720	3500	4500
Tape (Terabytes)	72	5700	8000
Link CERN-JINR	4	10	40



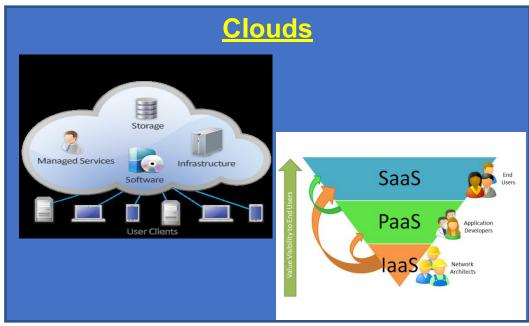
## Grids, clouds, fog, edge, supercomputers...

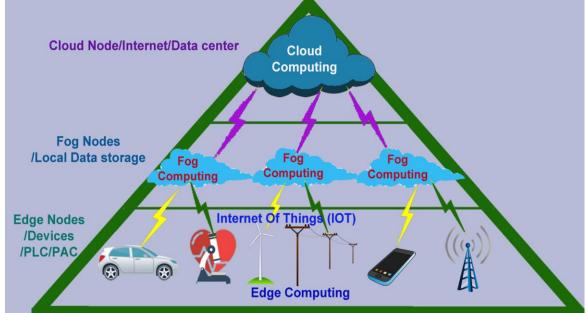












## **HPC+Big Data+AI**





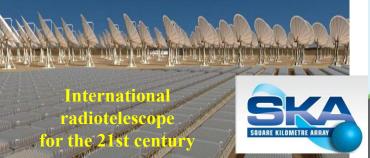


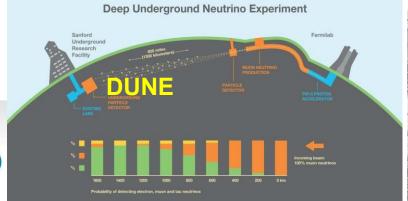


CERN Large Hadron Collider > 400 PB/Year



Square Kilometer Array radio telescope > 1 Eb/Year raw data (estimation)









#### Worldwide LHC Computing Grid (WLCG)

The mission of the WLCG is to provide global computing resources for the storage, distribution and analysis ~50-70 Petabytes of data expected every year of operations from the LHC. Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists

Tier0 (CERN):

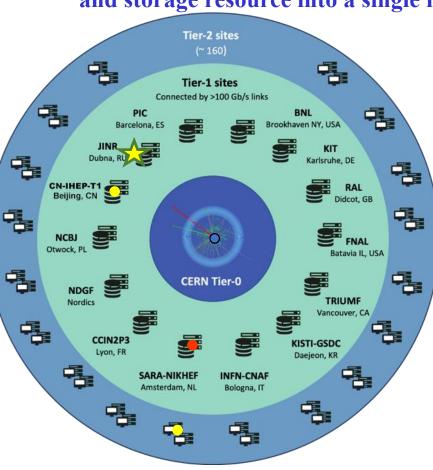
data recording, reconstruction and distribution

#### Tier1:

permanent storage, re-processing, analysis

#### Tier2:

Simulation, end-user analysis



The mission of the WLCG project is to provide global computing resources to store, distribute and analyze the ~400 PB of data expected every year of operations from the Large Hadron Collider.

WLCG computing enabled physicists to announce the discovery of the Higgs Boson on 4 July 2012.

42 countries
170 sites
~2 M CPU cores
2 EB of storage
> 2 M jobs/day
100-250 Gb/s links



At the Nobel Prize ceremony for the discovery of the Higgs boson, CERN Director Rolf Heuer directly named grid technologies as one of the three pillars of success (along with the LHC accelerator and physics facilities). Without the organization of the grid infrastructure at the LHC, it would be impossible to process and store the colossal volume of data coming from the collider, and therefore, to make scientific discoveries. Today, not a single large project can be implemented without the use of a distributed infrastructure for data processing.

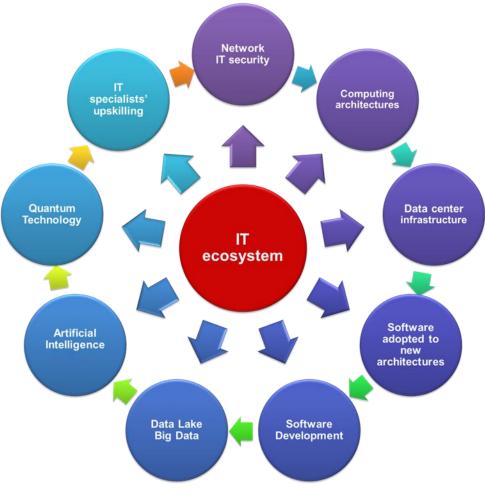
11.07.2025

#### **Meshcheryakov Laboratory of Information Technologies**

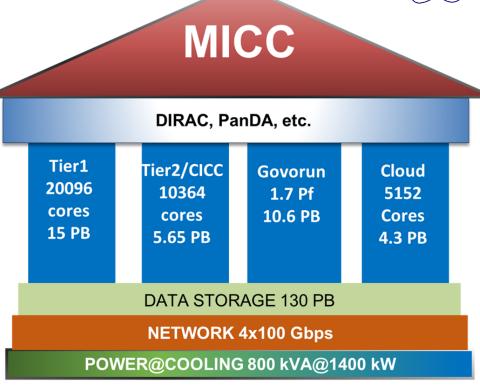
# MET

#### **Scientific IT ecosystem:**

coordinated development of interconnected IT technologies and computational methods



Scientists	115
Doctors of Science	21
Candidates of Science	62
Specialists	136
(engineers,	
programmers)	
Management	18
Workers	47
Total	316

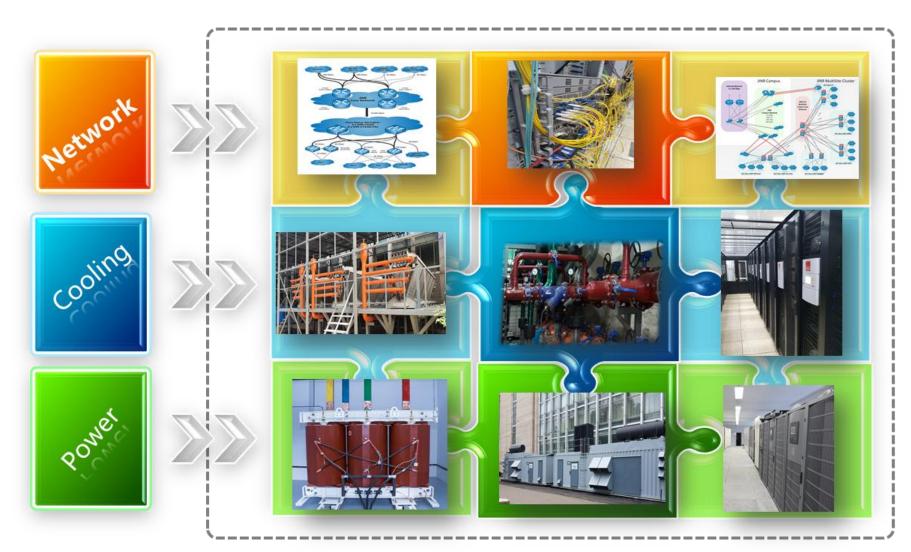


**Multifunctional Information and Computing Complex** 

- **Provide IT services** necessary for the fulfillment of the JINR Topical Plan on Research and International Cooperation in an efficient and effective manner
- Building world-class competence in IT and computational physics
- 24/7 support of computing infrastructure and services such availability is called nonstop service

## MICC Power @ Cooling @ Network





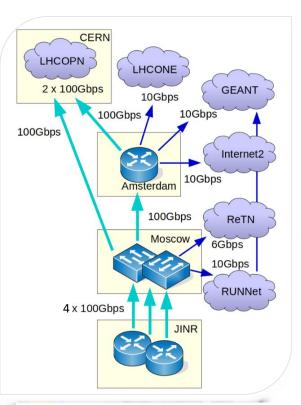
Wide Area Network 4x100 Gbps Cluster Backbone 4x100 Gbps Campus Backbone 2x100 Gbps

Dry chillers In-Row systems Total cooling 1400 kW

Uninterruptible power supplies (UPS) 8x300 kVA
Diesel-generator units (DGU)
2x1500 kVA

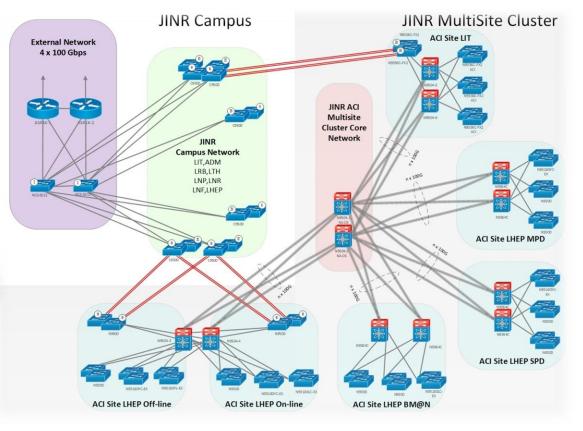
#### **Network Infrastructure**





MLIT ensures the reliable and fault-tolerant operation of all components of the network infrastructure:

- JINR-Moscow 4x100 Gbit/s
- JINR-CERN 100 Gbit/s and JINR-Amsterdam 100 Gbit/s
- multi-site cluster network with a bandwidth of 4x100 Gbit/s for the NICA megaproject
- local area network with a bandwidth of 2x100 Gb/s



#### The JINR LAN comprises:

14017 network elements

23276 IP-addresses ipv4

1491 IP-addresses ipv6

5934 users registered within

the network

4937 \*.jinr.ru service users

1153 digital library users

967 remote VPN

157 EDUROAM users

#### Network traffic in 2024

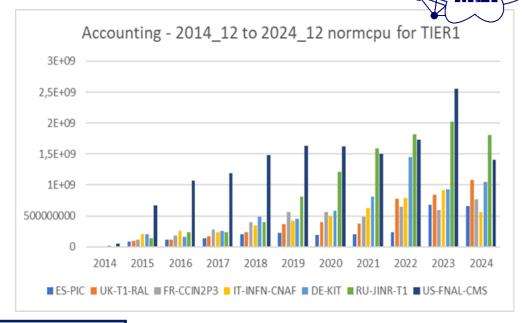
42.53 PB - input

20.62 PB - output



## Tier1 for CMS at JINR









- 20096 cores
- 360 kHS23
- 15.5 PB disks
- 100 PB tape
- 100% reliability and availability

Global CIO

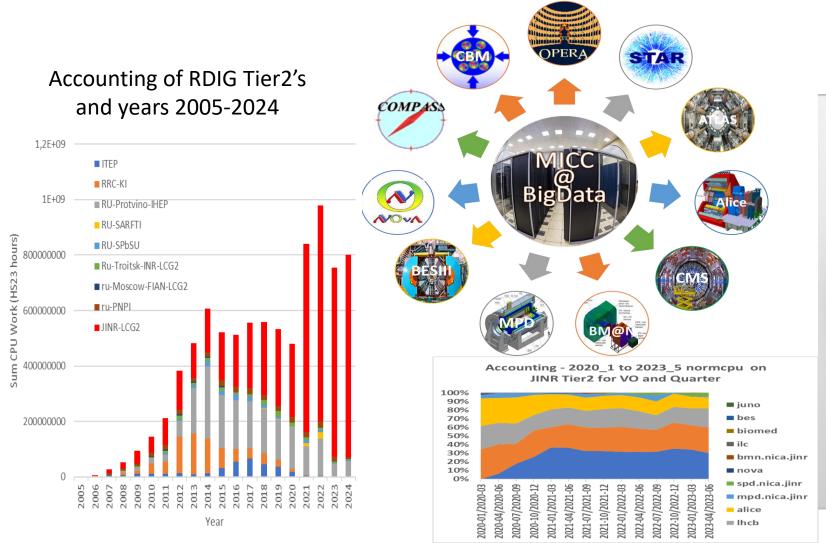
#### **Tier1 CMS 2024**

RU-JINR-T1	1,888,913,532	<b>24</b> %
US-FNAL-CMS	1,665,321,019	<b>21</b> %
UK-T1-RAL	1,132,299,978	14 %
DE-KIT	1,097,134,332	13 %
FR-CCIN2P3	807,778,212	10 %
ES-PIC	671,291,604	8 %
IT-INFN-CNAF	618,136,609	7 %

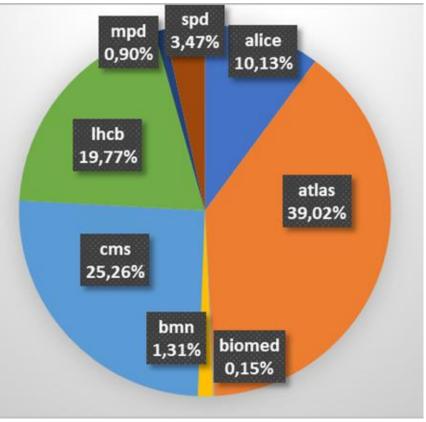
## Tier2 at JINR



Tier2 at JINR provides computing power and data storage and access systems for the majority of JINR users and user groups, as well as for users of virtual organizations (VOs) of the grid environment (LHC, NICA, FAIR, etc.).



Use of the JINR Tier2 site by virtual organizations within grid projects



#### **Cloud Infrastructure**

MEI

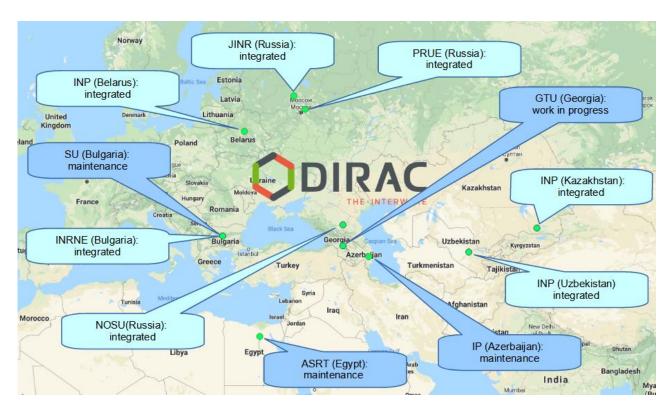
- Cloud Platform

   OpenNebula,
- Visualization KVM
- Computational resources for neutrino experiments

Neutrino experiments (JUNO, Baikal-GVD, NOvA) are the major users of the cloud infrastructure.

- VMs for JINR users
- Testbeds for research and development in IT
- COMPASS production system services
- Data Management System of the European Air Pollution Programme (UNECE ICP Vegetation)
- A system for diagnosing diseases of agricultural crops using modern machine learning methods
- Service for data visualization, Gitlab and some others

- Storage (Local disks, Ceph)
- Resources: ~ 5,152 CPU core; 6.3 PB ceph-storage

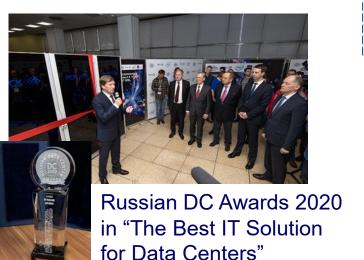


DIRAC-based distributed information and computing environment (DICE) that integrates the JINR Member State organizations' clouds

#### "Govorun" Supercomputer

MiT

- Hyper-converged software-defined system
- Hierarchical data processing and storage system (1,6-10-30-100 PB)
- Scalable solution Storage-on-demand
- Total peak performance: 2.2 PFlops DP and 26 Pflops for AI tasks
- GPU component based on NVIDIA Tesla V100&A100 (H100)
- CPU component based on RSC "Tornado" liquid cooling solutions
- The most energy-efficient center in Russia (PUE = 1.06)
- Storage performance >300 GB/s







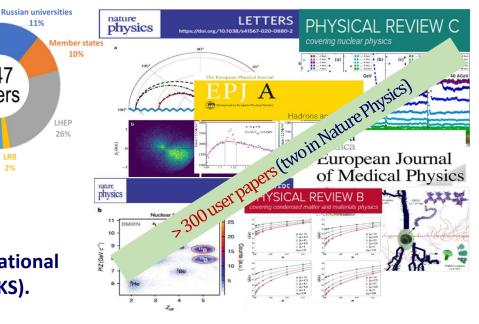
347

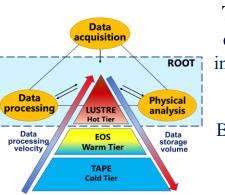
users

SC "Govorun" is included into a single supercomputer infrastructure based on the National Research Computer Network of Russia (NIKS).

## Key projects that use the resources of the SC "Govorun":

- NICA megaproject,
- calculations of lattice quantum chromodynamics,
- computations of the properties of atoms of superheavy elements,
- studies in the field of radiation biology,
- calculations of the radiation safety of JINR's facilities.



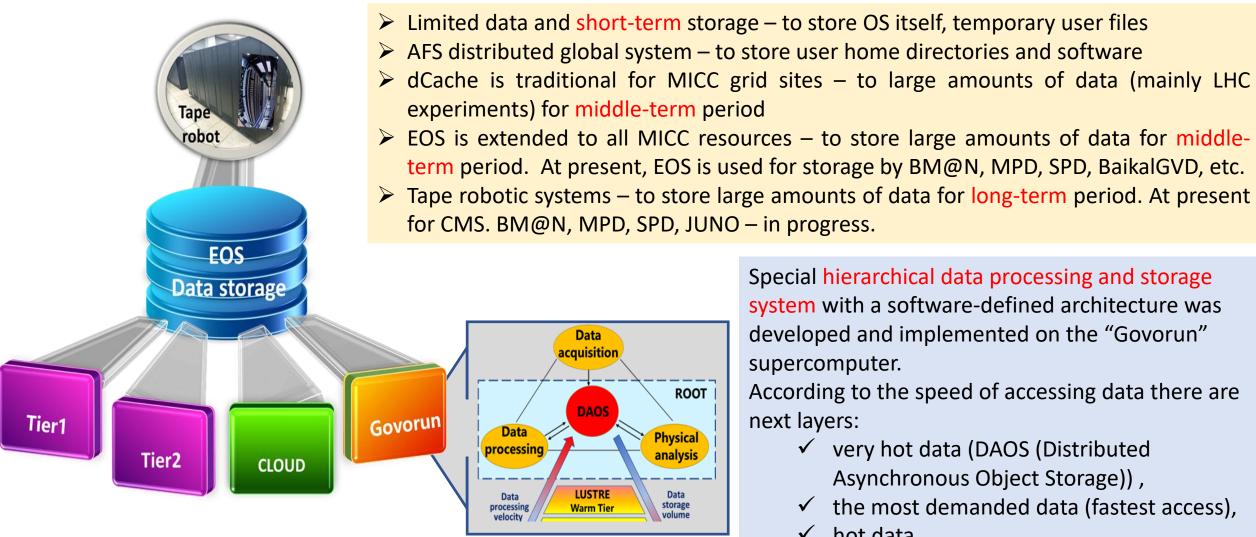


**Monitoring** 



## Distributed Multilayered Data Storage System





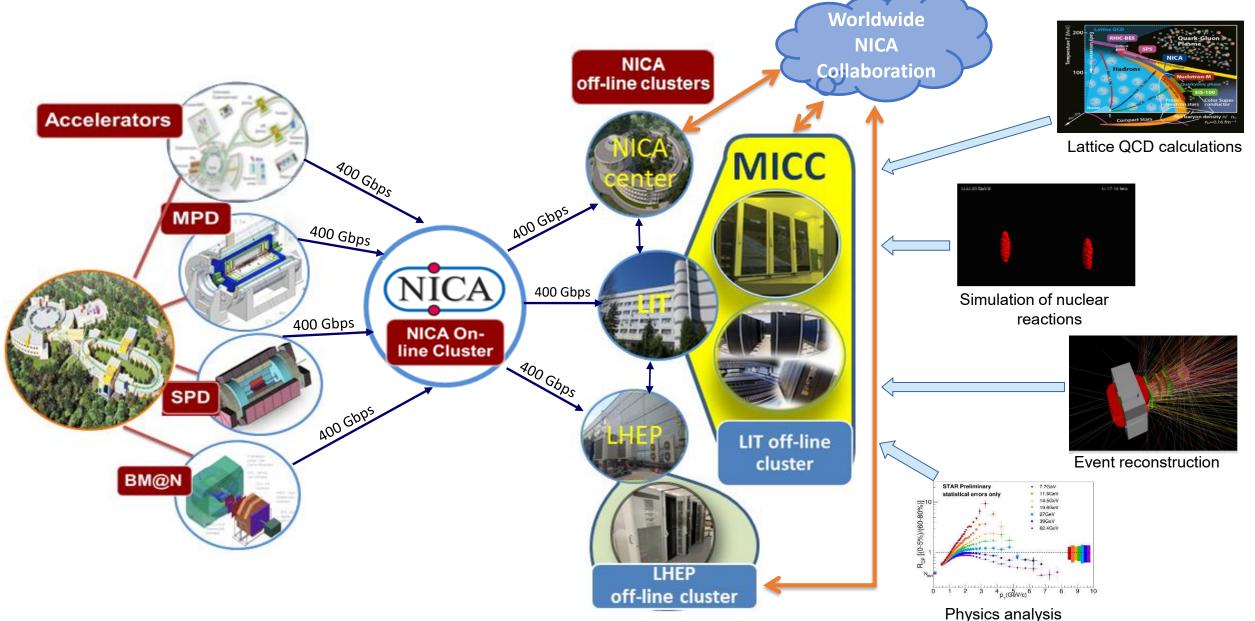
Special hierarchical data processing and storage system with a software-defined architecture was developed and implemented on the "Govorun" supercomputer.

According to the speed of accessing data there are next layers:

- ✓ very hot data (DAOS (Distributed) Asynchronous Object Storage)),
- the most demanded data (fastest access),
- hot data
- ✓ warm data (LUSTRE).

## **NICA Computing Concept & Challenges**



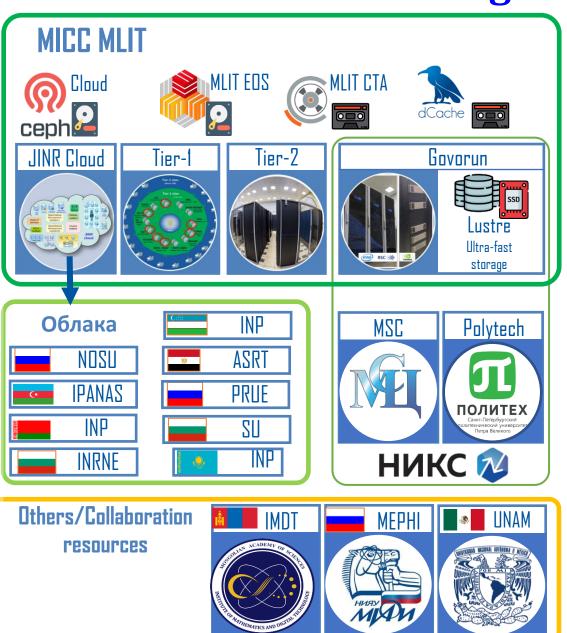


## DIRAC-based distributed heterogeneous environment

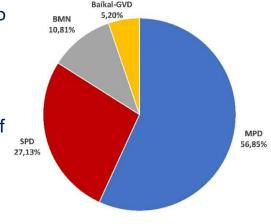




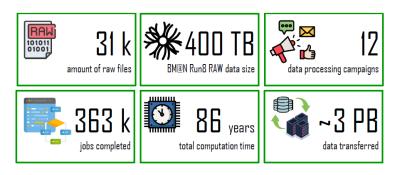
To perform calculations related to the mass generation and processing of data from the NICA mega-science project experiments, a distributed environment based on the DIRAC Interware platform was created.



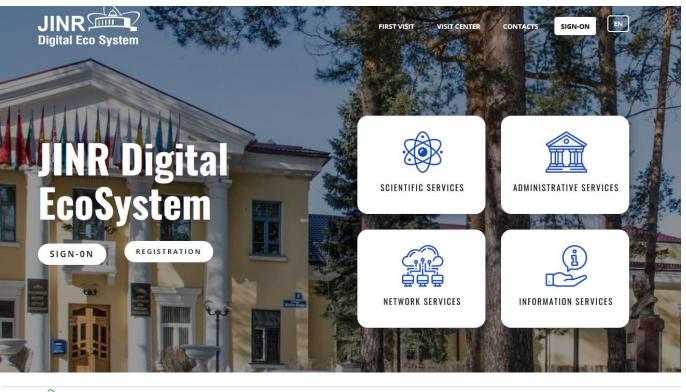
DIRAC is employed to solve the tasks of collaborations of all three experiments at the NICA accelerator complex, as well as of the Baikal-GVD neutrino telescope.

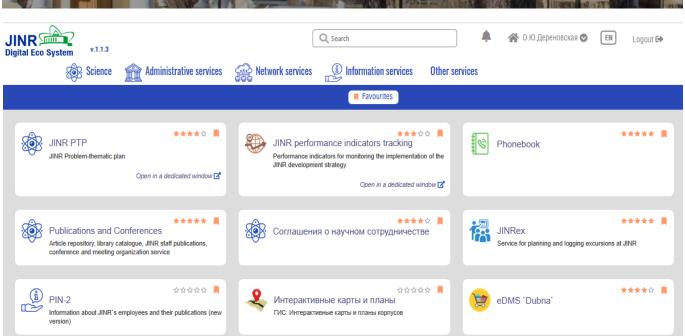


In 2023, for the first time at JINR, the complete processing of raw data from the 8th run of the BM@N experiment was performed on the distributed heterogeneous computing infrastructure integrated using the DIRAC platform.



Summary statistics of using the DIRAC platform for BM@N Run 8 data processing





## JINR Digital Ecosystem

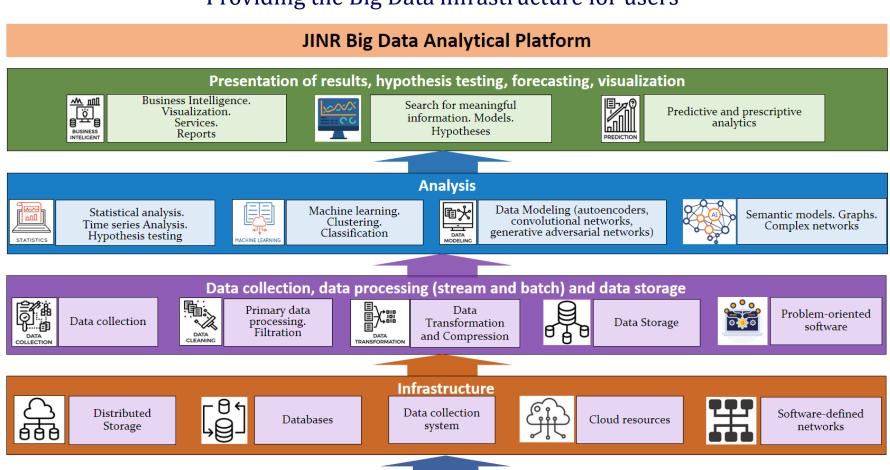
a single window into the JINR digital environment

- integration of existing and prospective services for supporting scientific, administrative and social activities, as well as maintenance of the Institute's engineering and IT infrastructure.
- Integrated Personal Account of a JINR Employee
- Notifications in the Personal Account
- Some resources are available to unregistered users
- Relevance. Information is updated promptly and regularly by service owners
- Convenient interface for service administrators
- Supports bilingualism: Russian and English
- Mobile version of the system

## **Methods of Artificial Intelligence and Big Data Analytics**



- Bringing best of Big Data approaches to JINR practices
- Providing the Big Data infrastructure for users



External data sources (API, Web, Databases, Social Networks)

Grid

Data sources

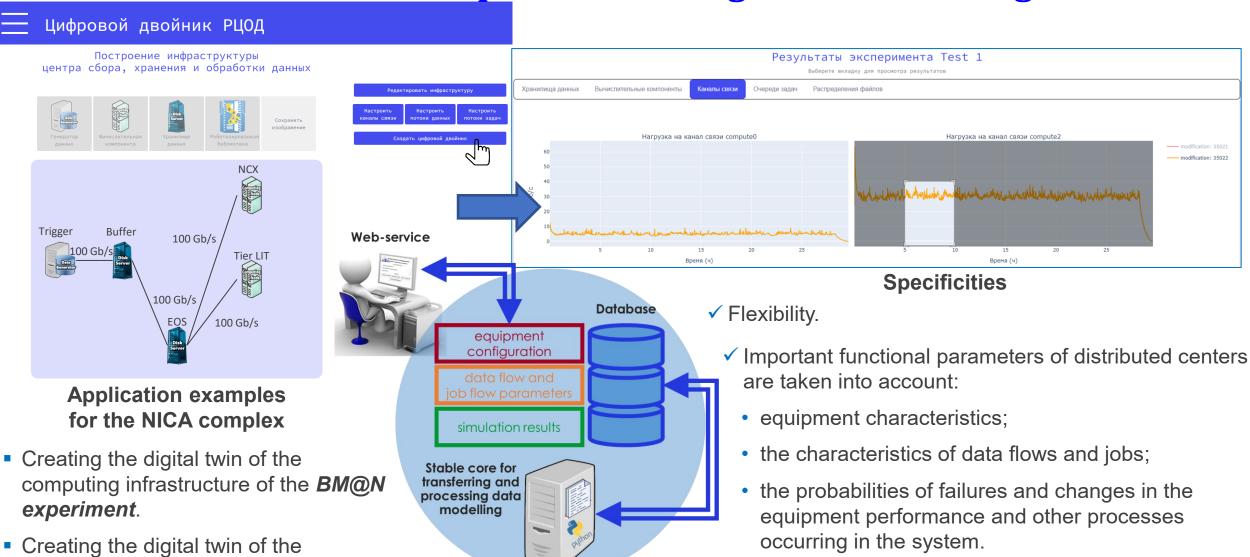
Industrial Internet of Things
(Smart meters, Tags, GPS, Cameras)

Physical Detectors and devices

Data Lake

#### **Software Complex for Creating Digital Twins**

of Distributed Data Acquisition, Storage and Processing Center \$\sqrt{\text{S}}



The results of the digital twin's work differ from the results of the

existing distributed center by no more than 20%.

computing system of the online data

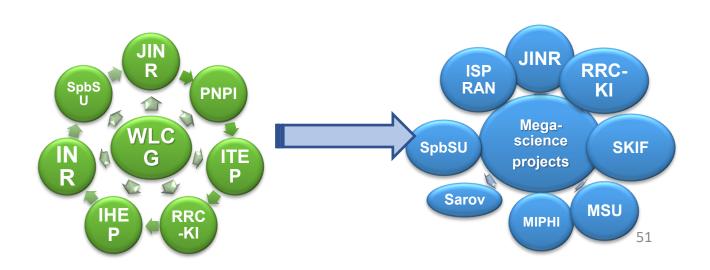
filter of the **SPD experiment**.

# **Creation of a consortium for IT support of large-scale research infrastructure**



- In Russia, a program of large-scale scientific projects is being implemented, the
  most important part of which is the development of distributed heterogeneous
  computer systems (including systems with parallelism) for processing, storing,
  analyzing experimental data, development and implementation of effective
  methods, algorithms and software for modeling physical systems, mathematical
  processing and analysis of experimental data, development of methods of
  machine learning, artificial intelligence, quantum computing.
- To solve this large-scale task, it is necessary to develop a distributed computer infrastructure that unites key scientific and educational institutes participating in mega-science projects - RDIG-M. The consortium created on the basis of JINR, NRC Kurchatov Institute, ISP RAS should become the core for IT support of the research infrastructure of the "mega-science" class.







Involving young specialists in solving tasks that face JINR using high-performance and distributed computing, data analysis methods and algorithms, state-of-the-art information technologies

Stage 1

#### Autumn School

acquaintance with the directions of JINR scientific research

tage 2

## Spring School

presentation the results of joint work with the Institute's specialists







#### MSU branch in Dubna







Training of specialists to work in the field of theoretical and experimental high energy physics, relativistic nuclear physics on the basis of the Joint Institute for Nuclear Research (JINR, Dubna), as well as for applied research and developments that are being carried out in medicine, biology and other fields using nuclear physical methods

and information technologies.

Fields of research in 2024

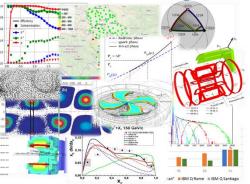
#### 03.04.02 "Physics":

- Master's program "Physics of Elementary Particles"
- Master's program "Fundamental and Applied nuclear physics"

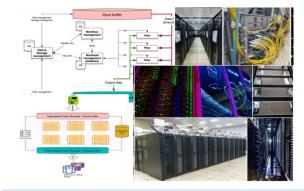
#### **Master's program**

Methods and technologies for data processing in heterogeneous computing environments

**Future:** 



SOUND TO SERVICE STORY OF THE PROPERTY OF THE



Mathematical modeling, numerical methods and software packages

Deep machine learning and big data analytics

Computing (software and models) for megascience class projects

# Trends in the development of distributed computing for large-scale scientific projects



Development of computer architectures and their integration:

- Peta-ExaFlop scale supercomputers, graphic, quantum, photonic processors as specialized computers around a universal system
- Peta-ExaByte scale distributed hierarchical information storage systems working with common metadata for secure efficient access and reliable storage of information
- cloud-based data centers that provide efficient services to users and support the operation of fog and edge computing systems
- intelligent task management systems in a distributed heterogeneous environment that includes various computer architectures (supercomputers, grid, clouds, clusters, servers, volunteer computing systems)
- development of methods, algorithms, platforms, systems of Big Data analytics, intelligent data analysis, ML/DL, artificial intelligence, digital services
- creation of digital twins of experimental setups, distributed computer systems and other complex objects
- training of highly qualified specialists in the field of distributed computing, storage, processing, and analysis of data for scientific projects





GRID'2004 – 204 participants from 8 countries;

GRID'2006 – 173 participants from 8 countries;

GRID'2008 – 211 participants from 20 countries;

GRID'2010 – 198 participants from 21 countries;

GRID'2012 – 243 participants from 22 countries

GRID'2014 – 190 participants from 12 countries;

GRID'2016 – 280 participants from 18 countries;

GRID'2018 – 260 participants from 16 countries;

GRID'2021 – 239 participants from 19 countries;

GRID'2023 – 280 participants from 17 countries;

**GRID'2025 – 298 participants from 14 countries** 37 plenary and over 135 sessional talks













Grid'2014









JOINT INSTITUTE FOR NUCLEAR RESEARCH LABORATORY OF INFORMATION TECHNOLOGIES 29 June - 2 July 2004 Dubna, Russia Distributed Computing and Grid-technologies in Science and Education Topics:
- creation and exploitation of Grid-infrastructures in various fields of ciscince and education;
of ciscince and educatio 





































## **Distributed Computing Team of MLIT**

- Tikhonenko Elena, Uzhinskiy Aleksandr, Kutovskiy Nikolay, Oleynik Danila,
   Petrosyan Artem, Belov Sergey, Priakhina Daria, Pelevanyuk Igor
- Strizh Tatiana, Mitsyn Valery, Trofimov Vladimir, Dolbilov Andrey, Astakhov Nikolay, Baginyan Andrey, Balandin Anton, Gorbunov Iliya, Gromova Natalia, Poze Viktor, Dmitrienko Pavel, Tkachev Igor, Sidorova Irina, Mitsyn Sergey, Kadochnikov Ivan, Nechaevskiy Andrey, Semenov Roman, Balashov Nikita, Zhiltsov Viktor, Golunov Aleksei, Kashunin Ivan, Mazhitova Elena, Voytishin Nikolay, Gavrish Andrey, Vorontsov Aleksey
- Shmatov Sergey, Podgainy Dmitry, Streltsova Oksana, Moybenko Aleksandr, Belyakov Dmitry, Kokorev Aleksandr









https://lit.jinr.ru



t.me/MLIT\_news