JINR computing infrastructure

Korenkov Vladimir

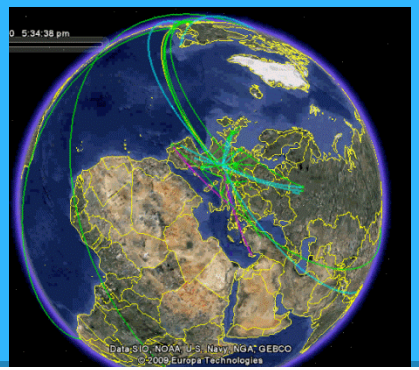
Director LIT JINR

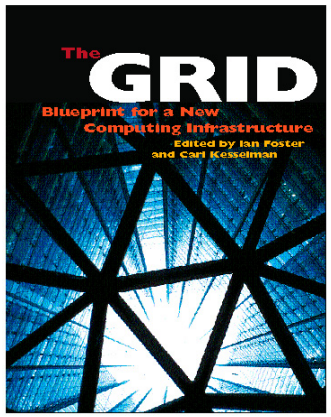
NEC-2017, Budva, Montenegro
25 September 2017

Computing in High Energy Physics and Nuclear Physics

Grids

- Collaborative environment
- Distributed resources



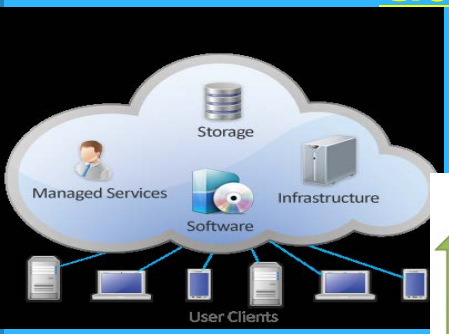


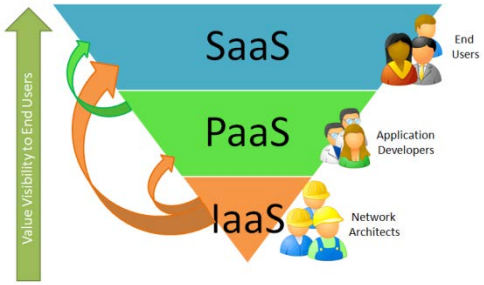
Supercomputers

	Titan System (Cray XK7)		
Peak Performance	27.1 PF 18,688 compute nodes	24.5 PF GPU	2.6 PF CPU
System memory	710 TB total memory		
Interconnect	Gemini High Speed Interconnect	3D Torus	
Storage	Lustre Filesystem	32 PB	
Archive	High-Performance Storage System (HPSS)	29 PB	
I/O Nodes	512 Service and I/O nodes		

12 OLCF | 2D

Clouds



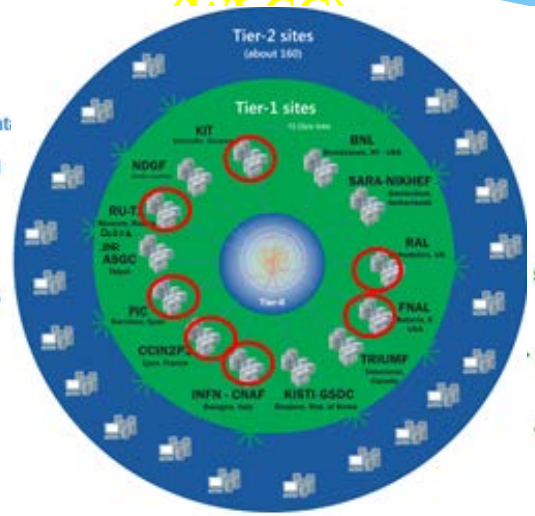


Big Data

- *Volume*
- *Velocity*
- *Variety*

Computing in High Energy Physics and Nuclear Physics

The Worldwide LHC Computing Grid (WLCG)



Tier-0 (CERN): data recording, reconstruction and distribution

Tier-1: permanent storage, re-processing, analysis

Tier-2: Simulation, end-user analysis

nearly 170 sites, 40 countries

~350'000 cores

500 PB of storage

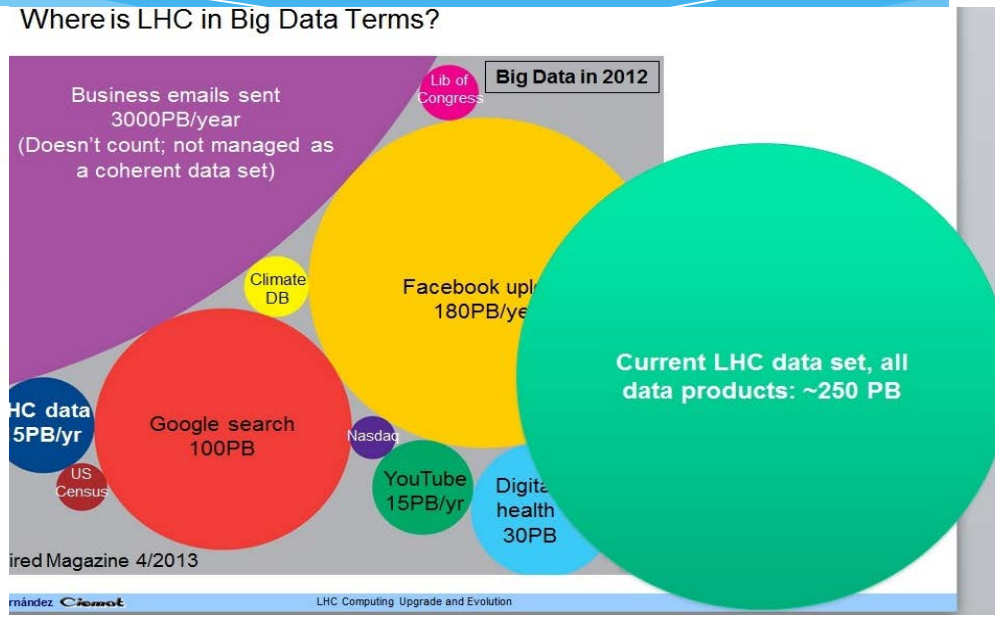
2 million jobs/day

10-100 Gb links

WLCG:
An International collaboration to distribute and analyse LHC data

Integrates computer centres worldwide that provide computing and storage resource into a single infrastructure accessible by all LHC physicists

Entry into the Big Data era



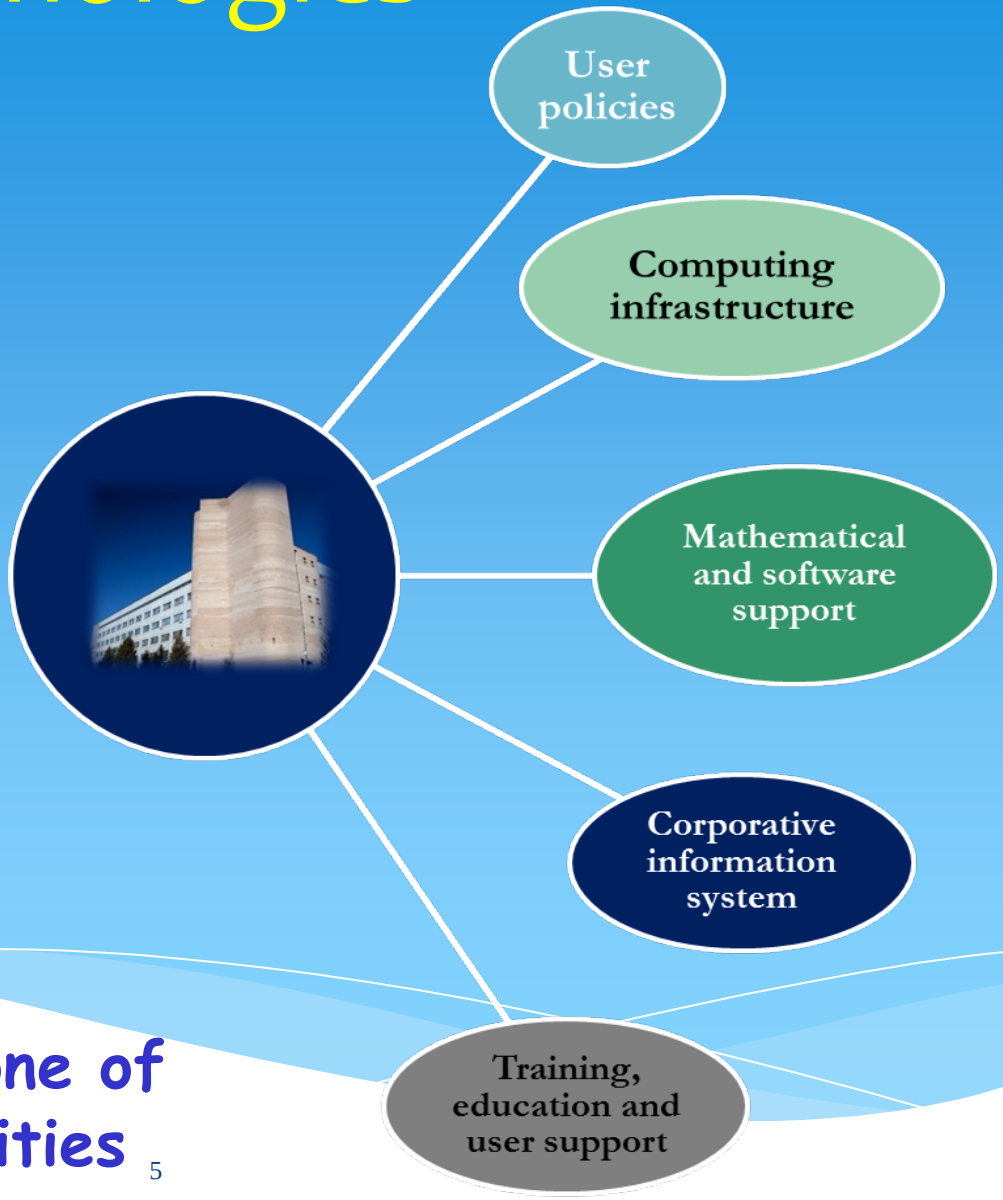
On a festivity dedicated to receiving the Nobel Prize for discovery of Higgs boson, CERN Director professor Rolf Dieter Heuer directly called the grid-technologies one of three pillars of success (alongside with the LHC accelerator and physical installations)

Nowadays, every large-scale project will fail without using a distributed infrastructure and Big Data Analytics for data processing

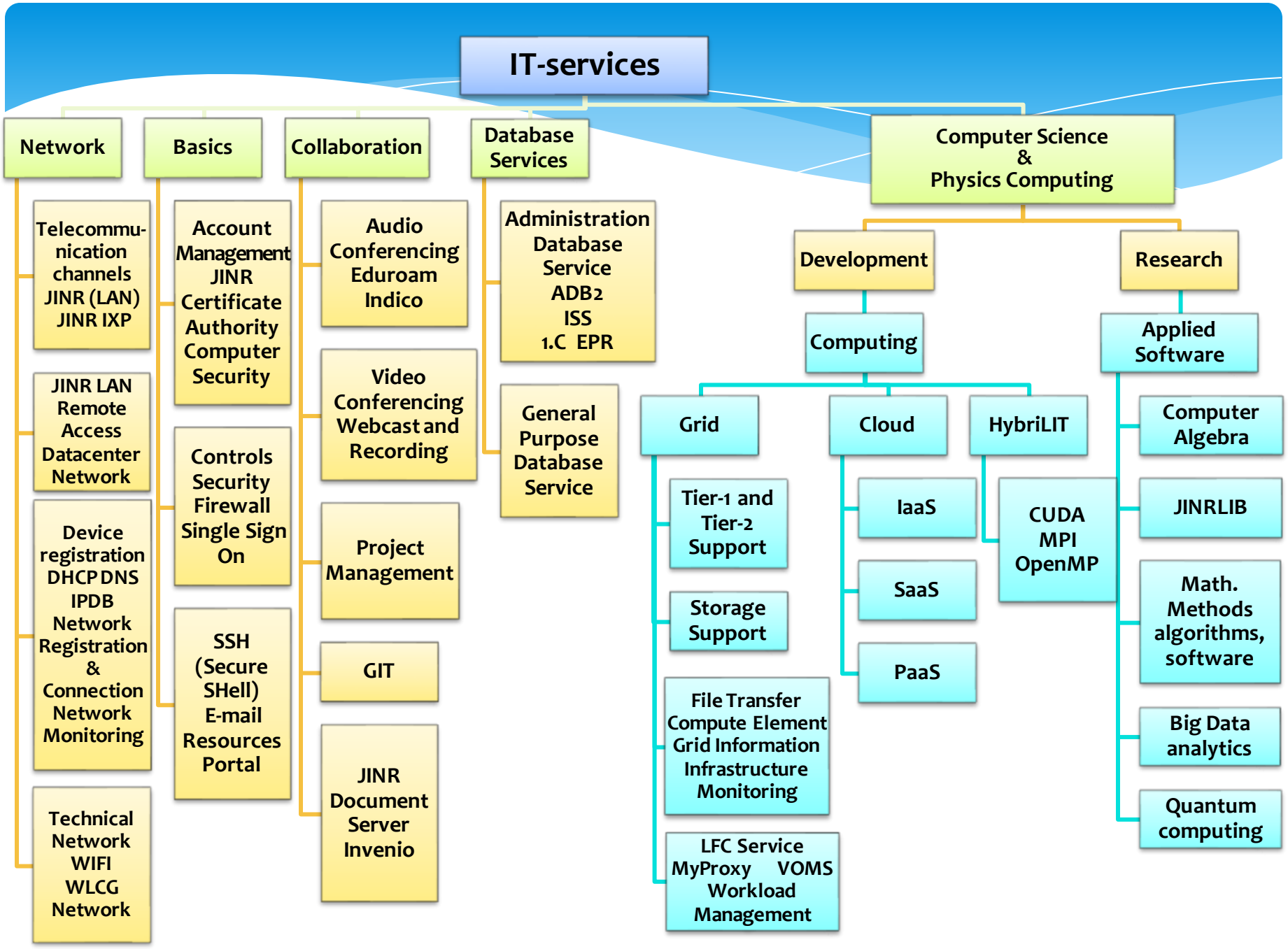
50 Years – LIT JINR

IT technologies

- The computer infrastructure development program.
- Mathematical support of studies conducted at JINR.
- Development of the multifunctional information and computing complex.



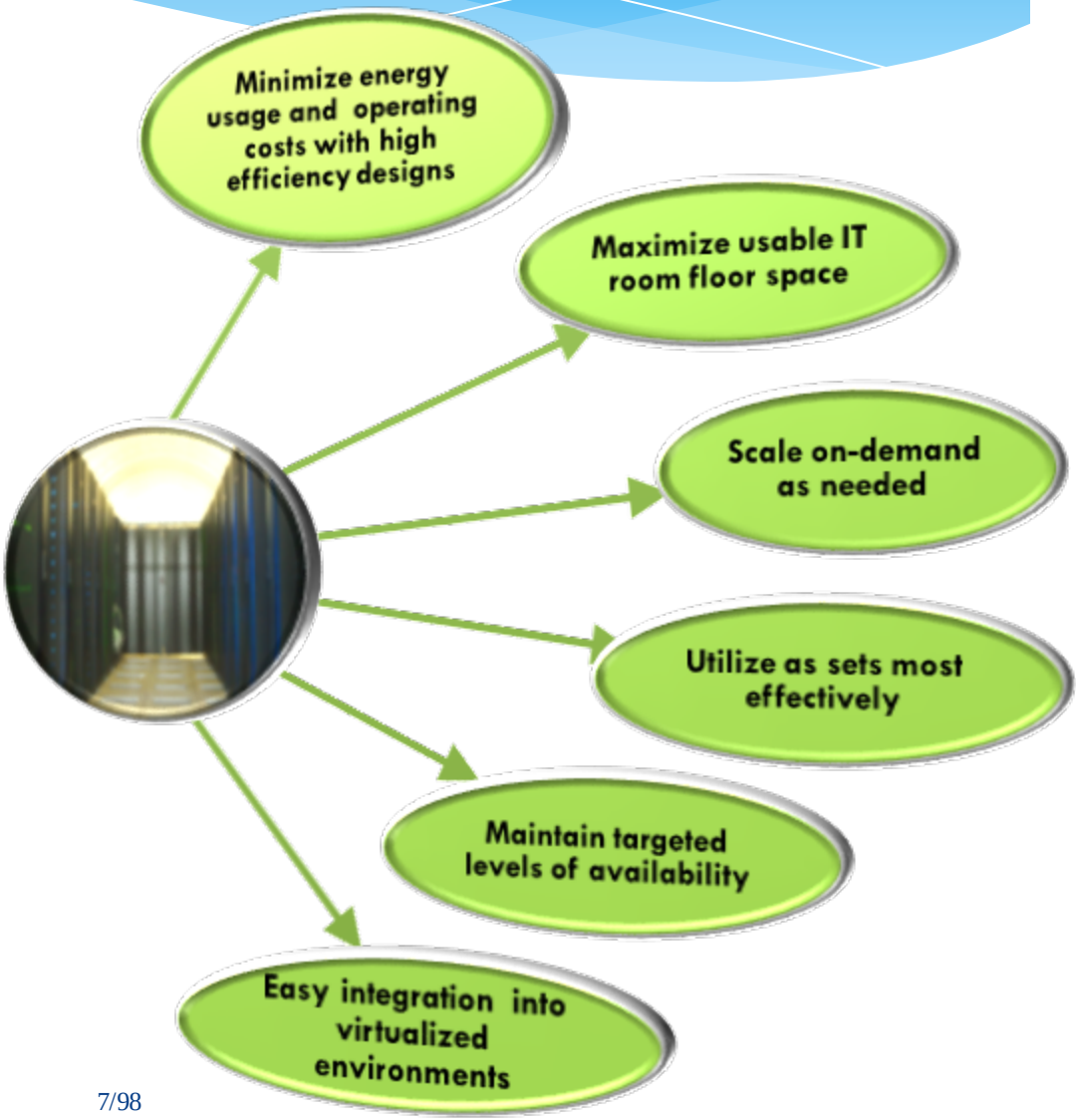
IT-infrastructure is one of the JINR basic facilities


IT-services
Network
Basics
Collaboration
Database Services
Computer Science & Physics Computing
Telecommu-nication channels JINR (LAN) JINR IXP
JINR LAN Remote Access Datacenter Network
Device registration DHCP DNS IPDB Network Registration & Connection Network Monitoring
Technical Network WIFI WLCG Network
Account Management JINR Certificate Authority Computer Security
Controls Security Firewall Single Sign On
SSH (Secure SHell) E-mail Resources Portal
Audio Conferencing Eduroam Indico
Video Conferencing Webcast and Recording
Project Management
GIT
JINR Document Server Invenio
Administration Database Service ADB2 ISS 1.C EPR
General Purpose Database Service
Development
Research
Computing
Applied Software
Grid
Cloud
HybriLIT
Tier-1 and Tier-2 Support
Storage Support
File Transfer Compute Element Grid Information Infrastructure Monitoring
LFC Service MyProxy VOMS Workload Management
IaaS
SaaS
PaaS
CUDA MPI OpenMP
Computer Algebra
JINRLIB
Math. Methods algorithms, software
Big Data analytics
Quantum computing

Project MICC (Multifunctional information and computing complex)

Build up the Multifunctional Information and Computing Complex (MICC)

- fault-tolerant infrastructure with electrical power storage and distribution facilities with expected availability of 99.995%,
- supports and uses a large variety of architectures, platforms, operating systems, network protocols and software products
- provides means for organization of collective development
- supports solution of problems of various complexity and subject matter
- enables management and processing of data of very large volumes and structures (Big Data)



IT-infrastructure of JINR

Project MICC: main components

- LAN: 10 Gbps
- WAN: 100 Gbps + 2x10 Gbps
- Tier-1: 4160 core, 5,8 PB disk, 11 PB tape
- CICC/Tier-2: 3500 core, 2 PB disk
- HybriLIT: 252 CPU, 77184 GPU cores, 182 PHI-cores, 2.4 TB RAM, 57.6 TB HDD, 142 Tflops
- Cloud: 700 CPU, 1 TB RAM

JINR grid sites of WLCG/EGI: Tier-1 for CMS
Tier-2 for ALICE, ATLAS, CMS, STAR, LHCb, BES, biomed, fermilab

Cloud infrastructure

Heterogeneous(CPU + GPU) computing cluster HybriLIT

Off-line cluster and storage system for BM@N, MPD, SPD Storage and computing facilities for local users

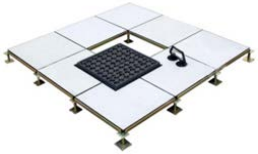
Network infrastructure

Engineering infrastructure

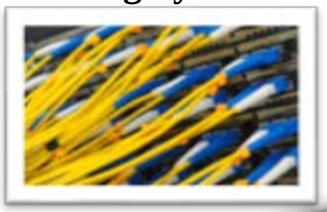
Engineering infrastructure

MICC engineering infrastructure

Raised Flooring System

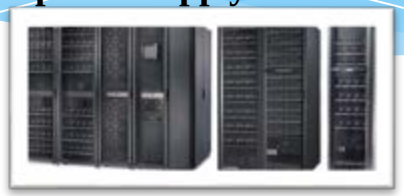


Fibre Optic & Data Structured Cabling System

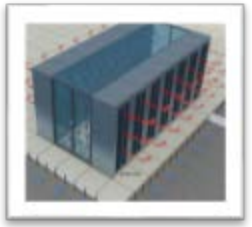


Diesel generator set

Uninterruptible power supply



High Density Heat Containment System



Computer Room Air Conditioner

Water Detection System

MICC Monitoring System

Biometric Access System

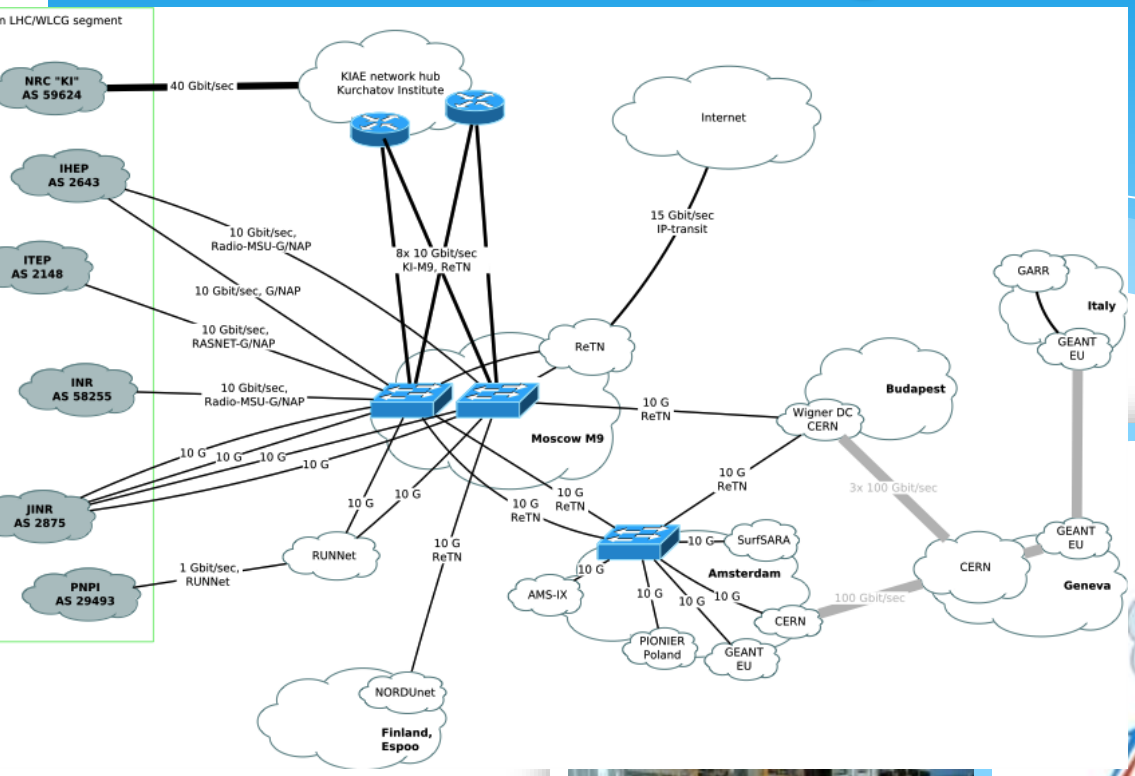
Fire Suppression System

Surveillance System

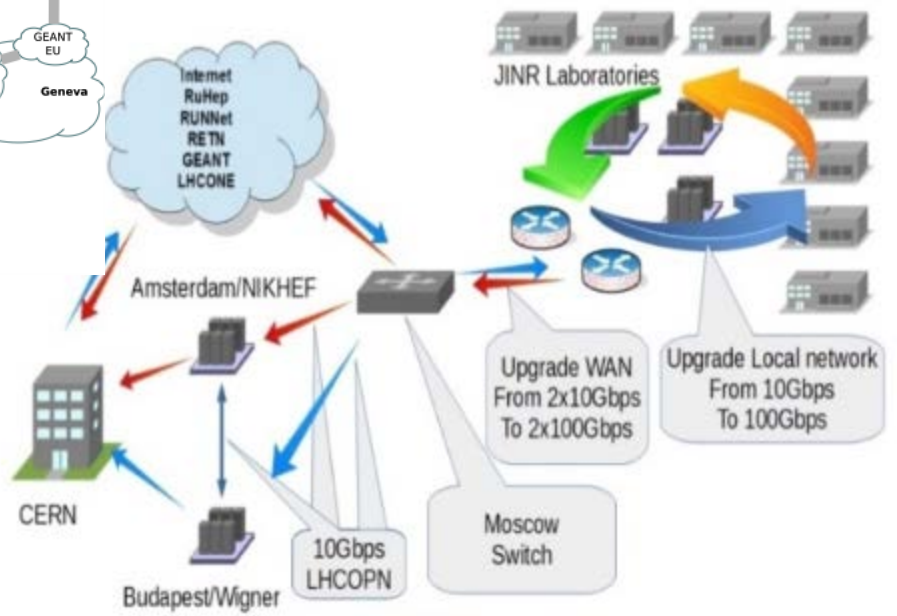
VESDA (Very Early Smoke Detection Apparatus)

Networking infrastructure

One of the most important components of JINR and MICC providing access to resources and the possibility to work with the big data is a network infrastructure.



JINR Local Area Network
Comprises 7762 computers & nodes
Users – 4320, IP – 13616, Remote VPN users – 376, E-library- 1520, mail.jinr.ru-2397, AFS -367, VOIP – 127, LAN high-speed transport -10 Gb/s)



Development plans for the network infrastructure

Increasing the channel capacity of the external JINR data link: 2 x 100 Gbps

Modernization of optical backbone of the local area network of JINR: 100 Gbps

Development of network services:

- Implement IPv6
- The use of new data transfer protocols
- Improved email service
- Wi-Fi authorization service
- Project “Personal office”

Local network of the NICA project: The projected capacity is stated as a data transmission channel with a throughput of 2 x 100 GbE.

Tier-1 JINR participation in Worldwide LHC Computing Grid (WLCG)

Grid Tier-1 is one of the 7 centres in the world intended for large-scale processing of experimental and event-modeling data coming from the centres of Tier-0 (CERN), as well as Tier-1 and Tier-2 of the WLCG for the CMS experiment. This component is considered as a prototype of the system of processing and storage of the experimental data from the mega-project NICA as a centre of Tier-0 and Tier-1.

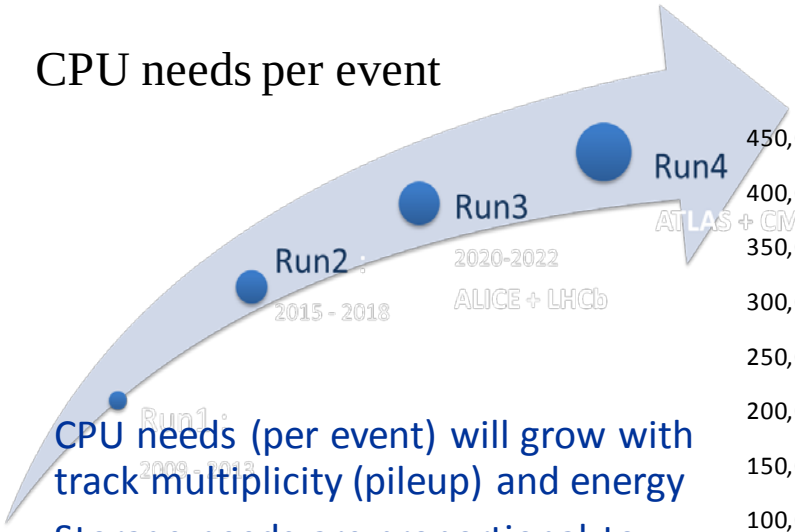
Current configuration Tier1 JINR 4160 core, 5,8 PB disk (1.1 PB MSS), 11 PB tape, 100% R/A, 70 TB per day

Development plans for the Tier-1 centre

Planned growth of Tier-1resources in years in absolute values with percentage growth over previous year.

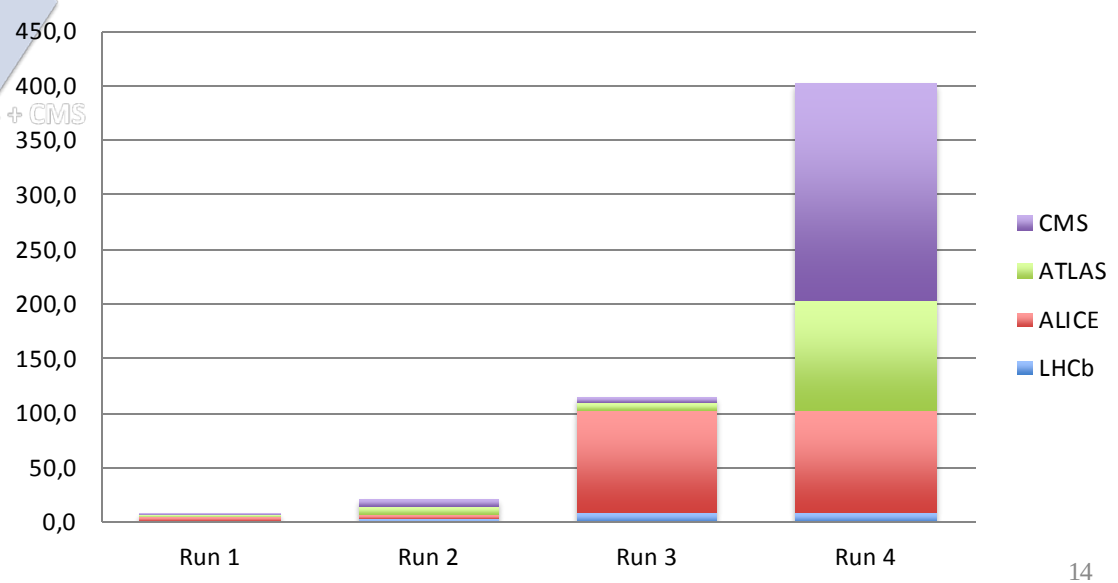
	2017	2018	2019
Processor capacity of the core/kHS06	4200/67,2	5200/83,2	10000/160
Disc storage (TB)	5070	6100	8000
Tape storage (TB)	11000	20000	20000

CPU needs per event



- CPU needs (per event) will grow with track multiplicity (pileup) and energy
- Storage needs are proportional to accumulated luminosity
- Grid resources are limited by funding and fully utilized

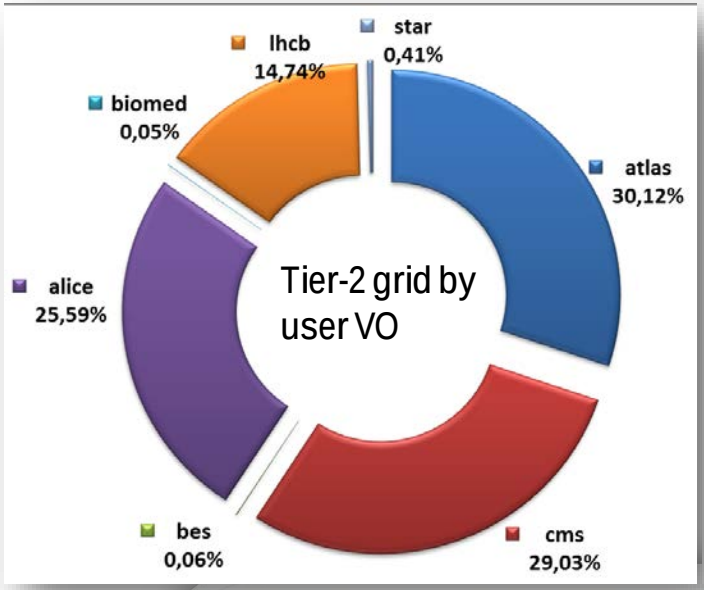
LHC Upgrade 2019-2021. Computing Needs

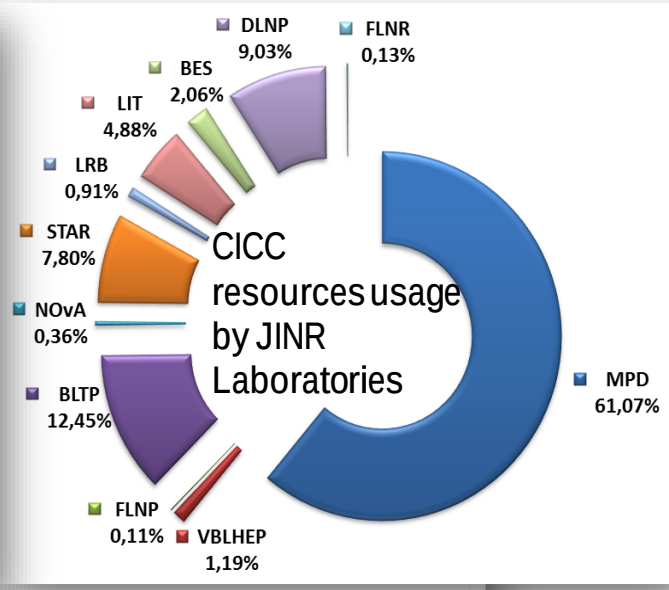


Tier2 & CICC

Experiments

BM@N,
MPD,
CMS,
ATLAS,
ALICE,
LHCb,
COMPASS,
DIRAC,
PANDA,
CBM,
STAR,
NOvA,
BESIII,
OPERA
NEMO
Mu2e,
NUCLON,
HONE,
FUSION
BIOMED





Tier-1 and Tier-2:
JINR in Worldwide LHC Computing Grid

The CICC/Tier-2 development plans

In the future we suppose to use and increase the CICC resources for

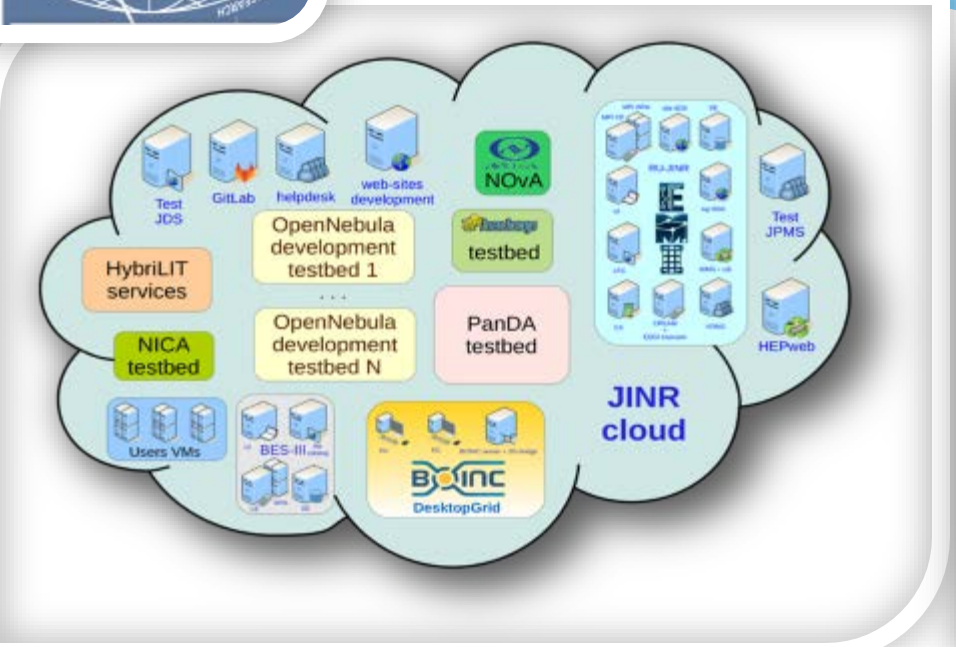
- Organization and support of the Tier-2 level sites for four LHC experiments;
- Provision of countable resources as well as data storage and data access for third-party collaborations , local user groups and individual JINR users.

One of the main directions is the creation of the infrastructure to support work of collaboration MPD NICA and the whole NICA project at all stages of its work.
It was installed hardware-software DevLab for testing new hardware solutions and software systems for designing and creating a data processing complex for NICA.

The planned growth of the CICC/Tier-2 resources over years

	2017	2018	2019
Comp.cores / kHS06	3700/59,2	4700/75,2	6000/96,0
Disk (TB)	2970	3400	5000

JINR Cloud



A spectrum of the tasks solved with the JINR:

- PanDA testbed was deployed for PanDA software validation and extensions development for ATLAS and COMPASS experiments;
- DIRAC-based testbed for BESIII experiment;
- a set of VMs of NOvA experiment users for analysis and software development;
- NICA testbed for grid middleware evaluation for NICA computing model development;
- Spark instance for Machine Learning and BigData analysis.

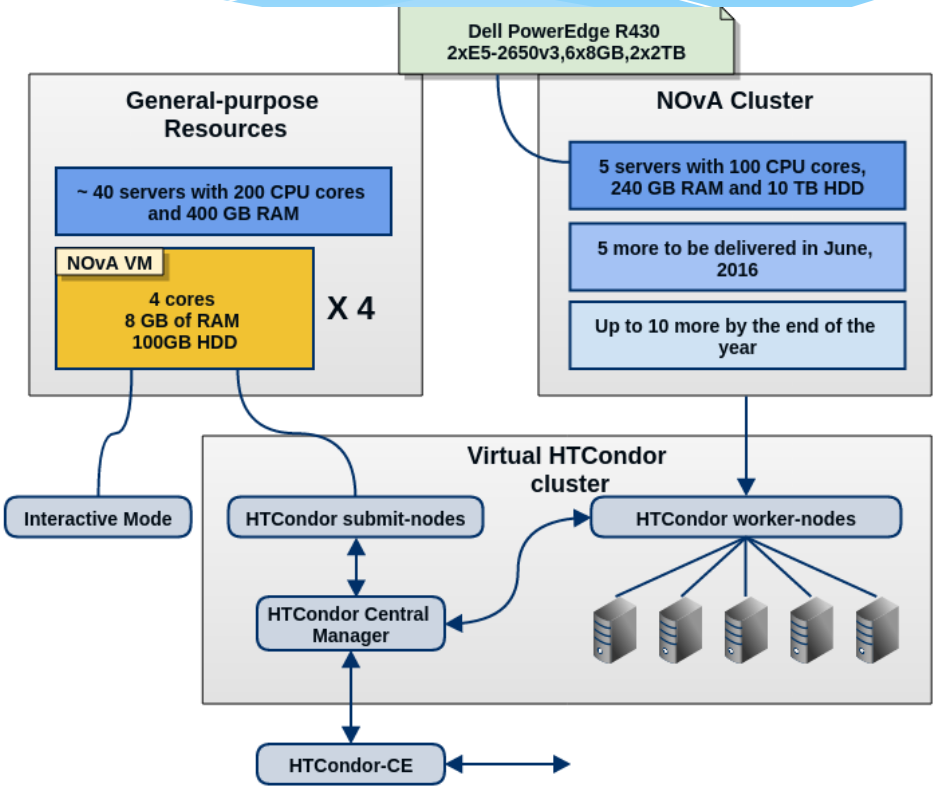
Geographical locations of the partner cloud infrastructures from JINR Member States which provide part of their computational resources being integrated into the JINR cloud

Computing Support for Neutrino Projects

NOvA (Fermilab, USA) is the first neutrino experiment actively using JINR Cloud:

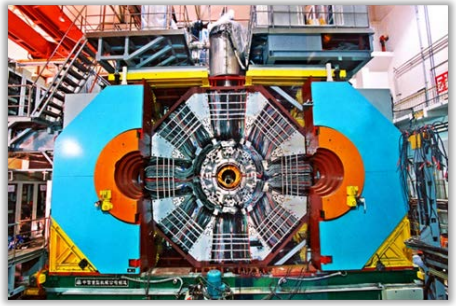
- ✓ 4 VMs for interactive/batch processing used by local JINR NOvA team
- ✓ Virtual batch-cluster based on HTCondor and connected to OSG
- ✓ 100 CPU, 240 GB RAM and 10 TB HDD already available
- ✓ Up to 400 CPU, 1 TB RAM and 80 TB HDD by the end of the year
- ✓ Computing support team was formed including physicists and IT specialists

These resources may also be used by other future experiments at Fermilab, such as DUNE and mu2e.



Reactor neutrino experiments Daya Bay and JUNO also showed its interest in using JINR cloud resources. At the moment the experiments' tasks and required computing capacities are being discussed.

BES-III Distributed Computing



GRID.JINR.ru
100
30 TB
CLOUD.JINR.ru
5

GRID.INFN-Torino.it
200
30 TB
CLOUD.TORINO.it
101

GRID.INFN-ReCas.it
50
30 TB

CLOUD.IHEP-OPENSTACK.cn
96
66 TB

CLOUD.IHEP-OPENNEBULA.cn
178
126 TB

CLUSTER.WHU.c
120
39 TB

CLUSTER.UMN.u
768
50 TB

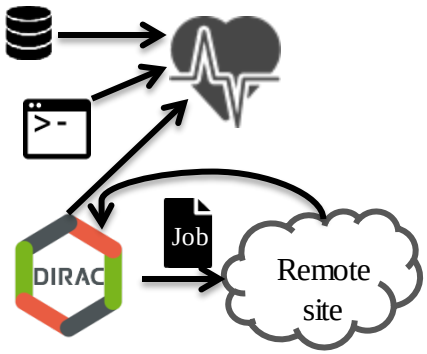
CLUSTER.USTC.c
200
24 TB

What have been done in computing:

- Grid monitoring system developed from scratch
- JINR cloud was integrated in BES-III infrastructure
- 6 % of all jobs was done in JINR during the past year

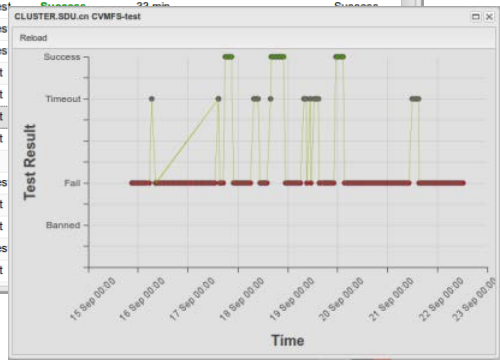
Planning to continue participate in BES-III experiment by:

- Improving monitoring
- Research on clouds in grid
- Providing storage and CPU cores



CE availability monitoring

Site	Test	Result	Received ago	Description
CLUSTER.SDU.cn	WMS-test	Success	43 min	Remote call
GRID.INFN-Torino.it	WMS-test	Success	8 min	Remote call
CLOUD.IHEP-OPENSTACK.cn	WMS-test	Success	86 min	Remote call
GRID.INFN-ReCas.it	WMS-test	Success	40 min	Remote call
CLOUD.CNIC.cn	WMS-test	Success	44 min	Remote call
CLUSTER.UMN.us	CVMFS-test	Success	37 min	Success
GRID.INFN-Torino.it	CVMFS-test	Success	8 min	Success
CLOUD.IHEP-OPENSTACK.cn	CVMFS-test	Success	30 min	Success
CLOUD.IHEP-OPENNEBULA.cn	CVMFS-test			
GRID.INFN-ReCas.it	CVMFS-tes			
CLOUD.CNIC.cn	CVMFS-tes			
CLUSTER.UMN.us	BOSS-test			
GRID.INFN-Torino.it	BOSS-test			
GRID.INFN-ReCas.it	BOSS-test			
CLOUD.CNIC.cn	BOSS-test			
CLOUD.AWS.cn	WMS-test			
CLOUD.AWS.cn	CVMFS-tes			
BOINC.IHEP.cn	BOSS-test			
CLOUD.AWS.cn	BOSS-test			
CLUSTER.SDU.cn	CVMFS-tes			
CLUSTER.SDU.cn	BOSS-test			



Heterogeneous computing component

In total, the cluster includes:

— 252 CPU-cores,
— 77184 CUDA-cores;
— 182 PHI-cores;
— 1920 GB RAM;
— 60 TB HDD.

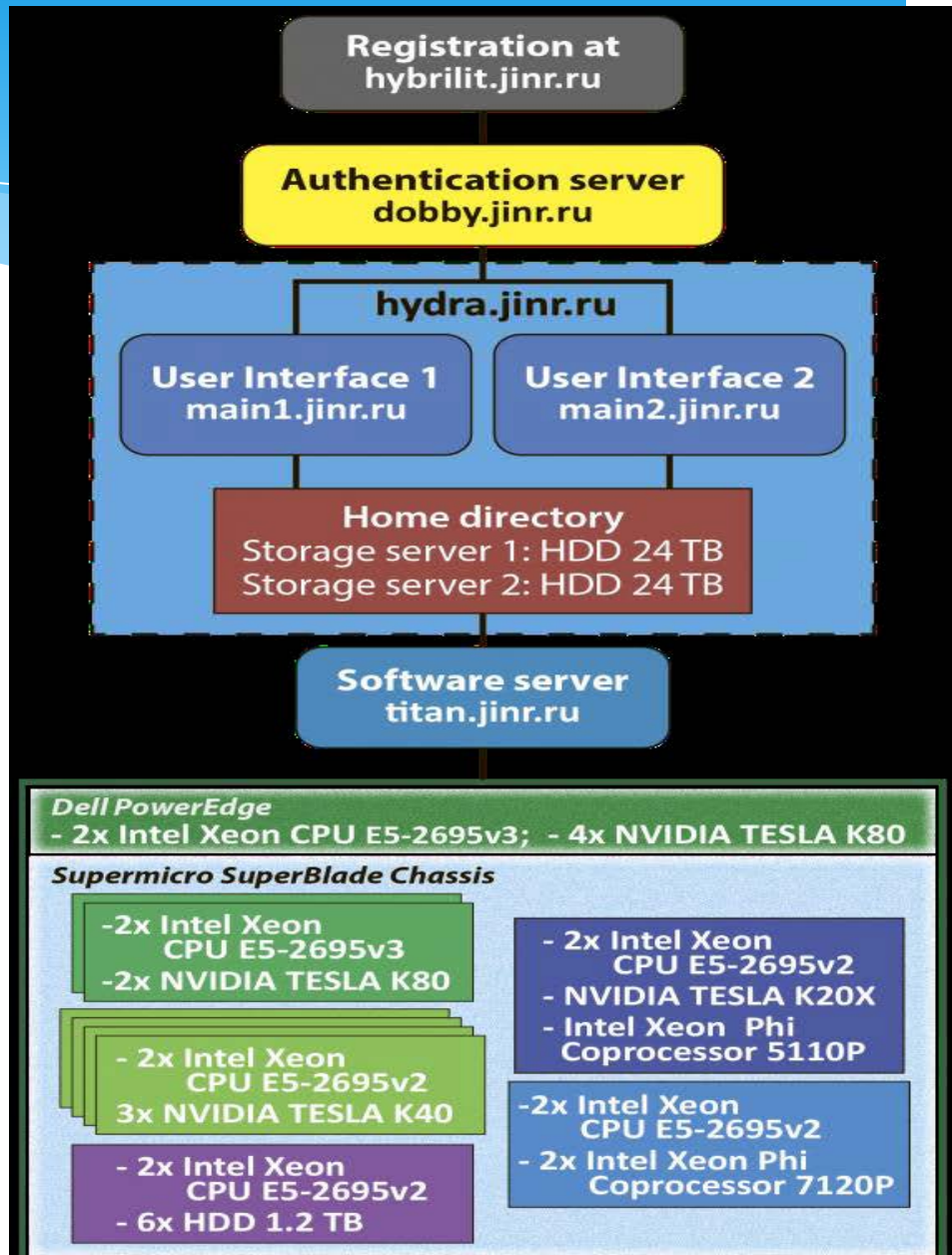
Total performance:

142 TFlops for operations with single precision;

50 TFlops for operations with double precision.

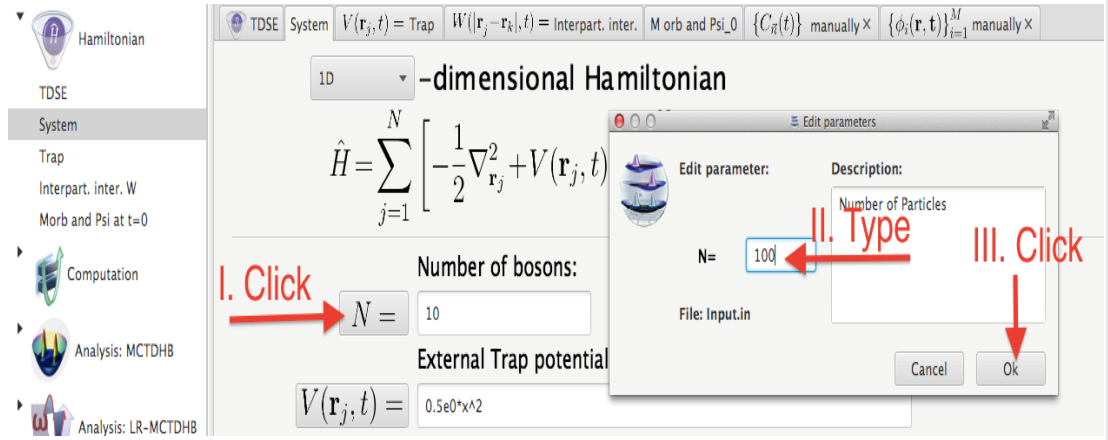
At present, the cluster is used by:

120 persons, including **26** from JINR Member States and **19** – Russia Universities.



In 2015 year was presented MCTDHB-Lab package – solver of the many-boson Time-Dependent Schrödinger Equation. This solver can be used for theoretical investigations of the highly-non-equilibrium quantum dynamics realized in trapped systems of ultra-cold atoms and molecules.

MCTDHB-Lab is available as a FREE-for-download, cross-platform software with a mouse-click interface



Center for Quantum Dynamics

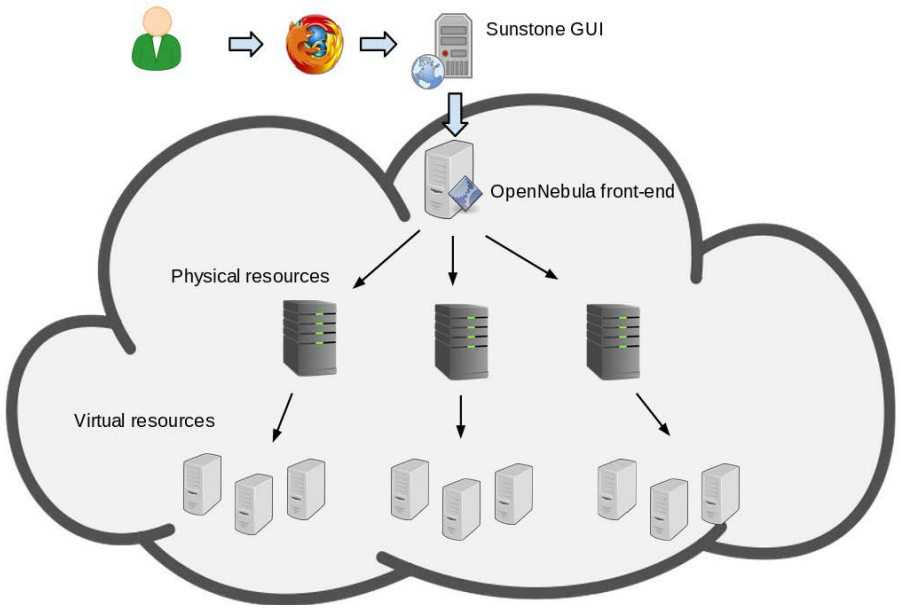
Many-body theory of bosons group Heidelberg, Germany

Heterogeneous Computations **HybriLIT** - team

Cloud and heterogeneous cluster

Advanced cloud infrastructures

- Dynamically reconfigurable computing services
- Large-scale open data repository and access services



	2017	2018	2019
Cores	1400	1800	2200
RAM/GB	6160	8080	10000
Disk serves/TB	576	768	960

Advanced heterogeneous computing

- User friendly information-computing environment
- New methods and algorithms for parallel hybrid computations
- Infrastructure for tutorials on parallel programming techniques

- **The annual increase in computing resources - 500 Tflops and in disk storages - 20TB**
- **Systematic research on the possibilities of new computing architectures appearing in the field of high-performance computing (HPC) include them into the cluster configuration in the future.**

Monitoring and Control room for MICC

The monitoring system of the JINR Computing Complex has been developed and put into exploitation. System allows one, in a real time mode, to observe the whole computing complex state and send the system alerts to users via e-mail, sms, etc.

690 elements are under observation
3497 checks in real time

Worldwide LHC Computing Grid Project (WLCG)

The primary goal of the WLCG project is to create a global infrastructure of regional centers for processing, storage and analysis of data of the LHC physical experiments.
The grid-technologies are a basis for constructing this infrastructure.

A protocol between CERN, Russia and JINR on participation in the LCG project was signed in 2003. MoU about participation in the WLCG project was signed in 2007.

Tasks of the Russian centers and JINR within WLCG :

- **Data management for WLCG e**
- **Introduction of WLCG services for experiments**
- **Development of WLCG monitoring systems**
- **Development of simulation packages for experiments**
- **Creation of a Tier1 center in Russia**

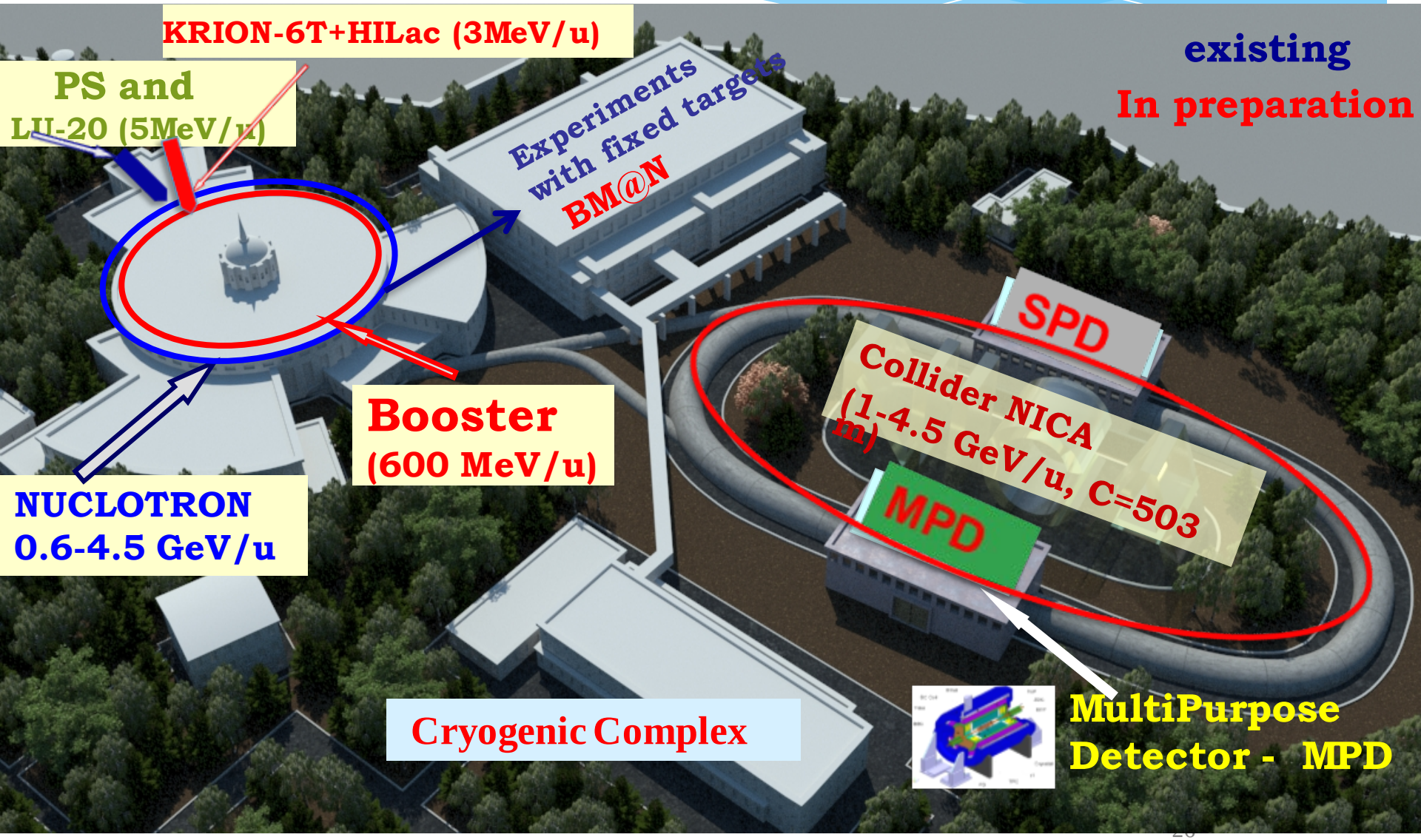
JINR activity at WLCG project

- **Participation in software development for ATLAS, ALICE, CMS**
- **Development of WLCG Dashboard (WLCG Google Earth monitoring, Global data transfer monitoring, Local and global Monitoring of Tier3 centers, xROOTd Monitoring,…)**
- **NoSQL data storage integration (Hadoop, ElasticSearch,…)**
- **Integration of Grid, Clouds and HPC**
- **Development of DDM (Deletion service) and AGIS for ATLAS**
- **BigPanda development (TITAN, IT4, COMPASS,…)**
- **GENSER & MCDB**
- **Creation of the Tier1 center in Russia**

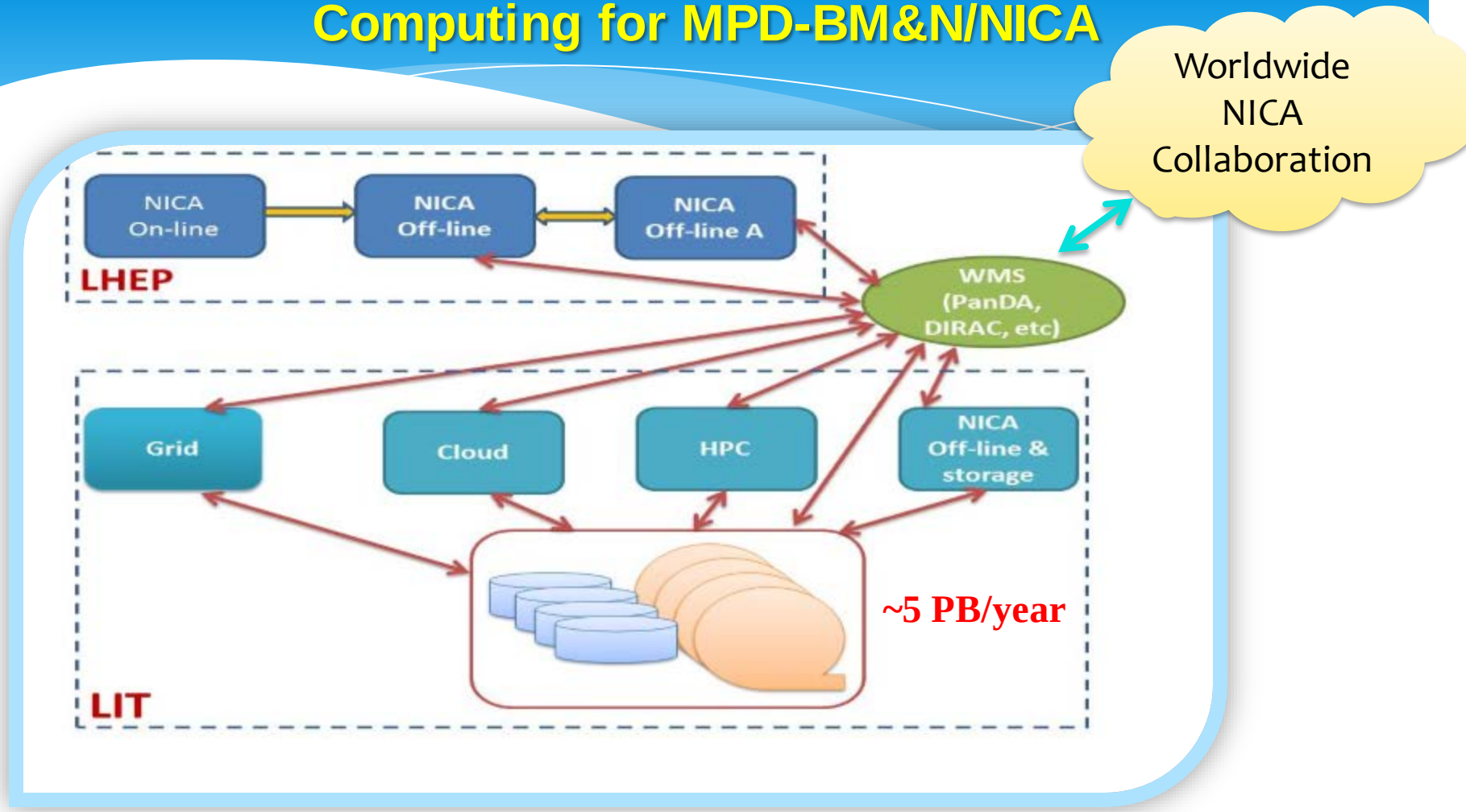
NICA Complex: *New era in the hot dense matter science*

Collider basic parameters:

$\sqrt{S_{NN}}$ = 4-11 GeV; *beams: from* p to Au; L~10^{27} cm^{-2} c^{-1} (Au), ~10^{32} cm^{-2} c^{-1} (p)



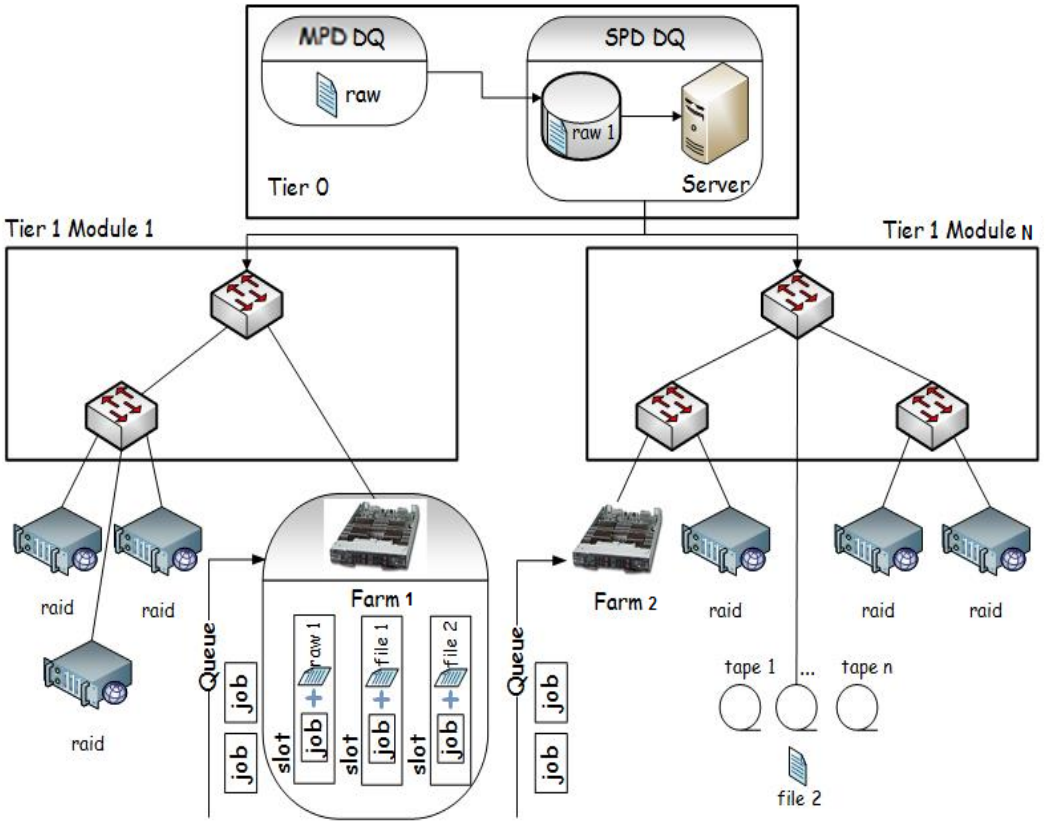
Computing for MPD-BM&N/NICA



Integration of various resources from NICA VBLHEP and MICC LIT is for the development of computer infrastructure of the NICA.

Requirements: reliability, scalability, interoperability, adaptability to new technical solutions, year-round operation in 24x7 mode

Creation of the computing off-line complex for simulation, processing, analysis and storing data of the NICA project



SyMSim Simulation System > NICA > 3 models

1. A data model – description of the data generation processes, their volumes and storage conditions.
2. Data processing model - rules for the use of resources such as CPU, memory and I/O between concurrently running tasks .
3. Model of communications of data traffic for different protocols in local and global networks.

The goal of the planned work:

– creating a MICC component to store and process physical data taking specific parameters of the experiments into account. For this, it is necessary to properly describe and predict the performance and limitations of the developed system of processing and storing the NICA data.

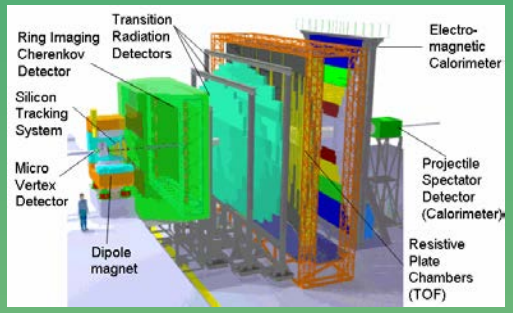
In frames of the work conducted, an answer to the question should be received what kind of architecture of the system is preferable from a viewpoint of a reasonable balance of time, financial and technological costs.

J/ψ→e⁺e⁻ decay selection criteria for Au+Au at 10 AGeV in the CBM experiment

The study of charmonium production is one of the key objectives of the CBM experiment.

Main difficulty of J/ψ→e+e- :

- extremely low yield of *J/ψ* mesons
- low probability (about 6%) of decay
- intense hadron background



The investigation goal is fast and efficient selection of the signal events for J/ψ→e+e- decays reconstruction in the real time experiment

J/ψ reconstruction technique

These criteria are applied successively to a sample formed as a result of applying the preceding criterion.

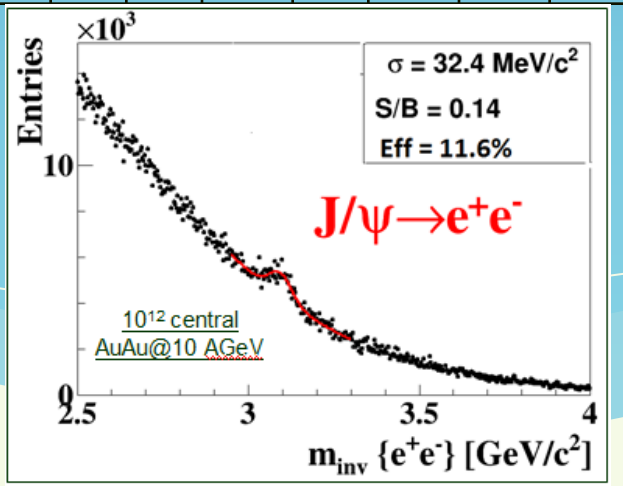
1. Transverse momentum $p_t > 1$ GeV/c
2. Number of hits in the TRD-track > 2
3. Electron identification in TRD with ANN
4. Limit on the track deviation in the TRD
5. Electron identification in TOF
6. The presence of the Cherenkov ring in the RICH

J/ψ reconstruction technique allows to almost completely suppress the background and accumulate considerable J/ψ meson statistics upon its decay into the dielectron channel using the CBM setup.

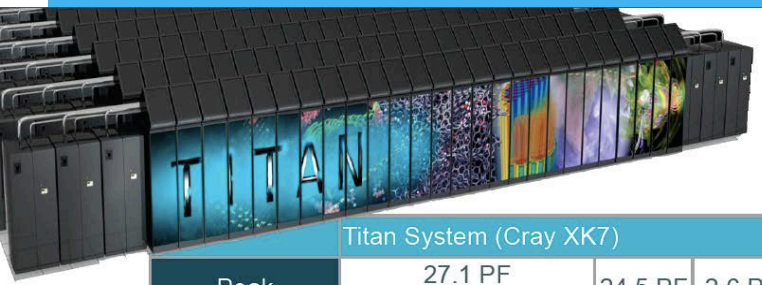
Results

The loss of signal events and the corresponding suppression background in each stage independently

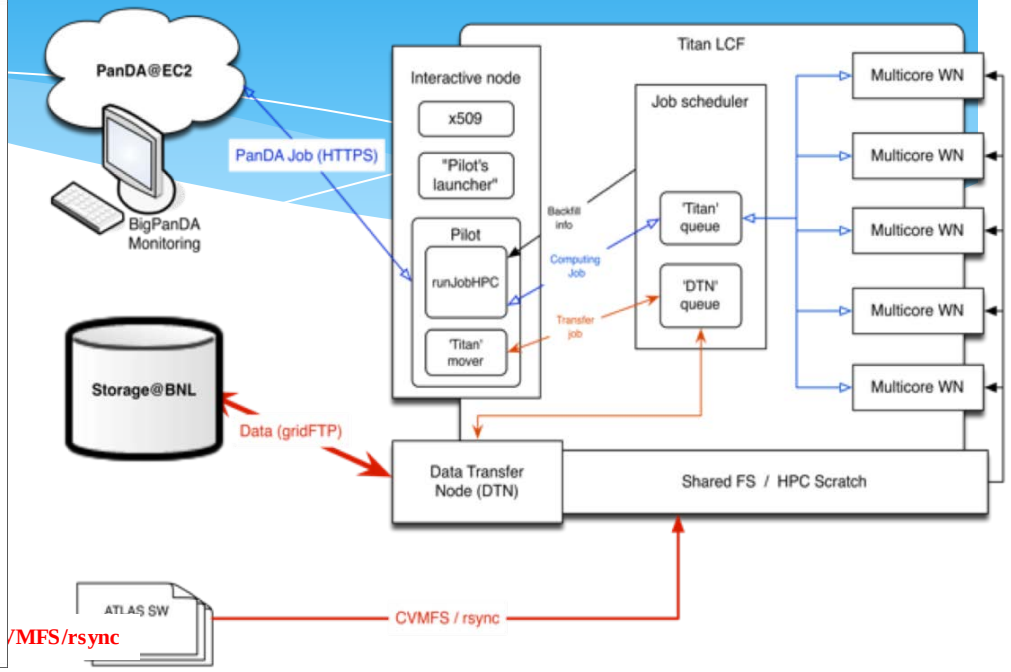
	1	2	3	4	5	6	Total
signal loss (%)	16	18	28.5	7.5	1.1	1.6	56
Bg supp(%)	92	26	99.5	16	54	61.5	99.9



Evolving PanDA for Advanced Scientific Computing



	Titan System (Cray XK7)		
Peak Performance	27.1 PF 18,688 compute nodes	24.5 PF GPU	2.6 PF CPU
System memory	710 TB total memory		
Interconnect	Gemini High Speed Interconnect	3D Torus	
Storage	Lustre Filesystem	32 PB	
Archive	High-Performance Storage System (HPSS)	29 PB	
I/O Nodes	512 Service and I/O nodes		



ATLAS (BNL, UTA), OLCF, ALICE (CERN,LBNL,UTK), LIT JINR:

- adapt PanDA for OLCF (Titan)
- reuse existing PanDA components and workflow as much as possible.
- PanDA connection layer runs on front-end nodes in user space. There is a predefined host to communicate with CERN from OLCF, connections are initiated from the front-end nodes
- SAGA (a Simple API for Grid Applications) framework as a local batch interface.
- Pilot (payload submission) is running on HPC interactive node and communicating with local batch scheduler to manage jobs on Titan.
- Outputs are transferred to BNL T1 or to local storage

BigPanDA. Generalizing PanDA. Beyond Grid and ATLAS

COMPASS

Google Cloud Platform

Alpha Magnetic Spectrometer AMS-02

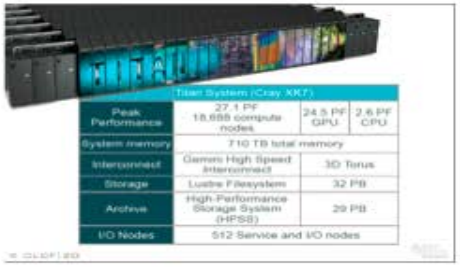
ALICE

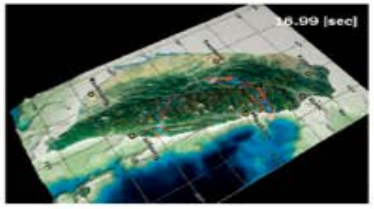
ATLAS

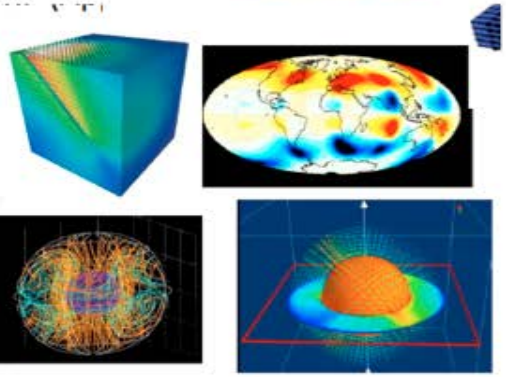
NICA

nectar

amazon web services | EC2







Functional-technological core of the decision support system based on the Big Data approach

Sources of information

- Social Networks
- Official information
- Websites
- Forums
- Information provided by the customer

data collection and storage

- Automatic data extraction (Flume, Elasticsearch)
- subsystem for manual input
- Data storing (HDFS, lizardfs, dCache, DPM)
- Data preprocessing (Avro, Sqoop)
- NoSQL DB (MongoDB, Hbase, Neo4j, Cassandra)
- SQL DB (PostgreSQL,MySQL)

data processing

- Clustering data (MapReduce, Hive, Pig)
- Dispatching system of tasks (MapReduce, Hive, Pig)
- Execution of packages of tasks (Panda)
- Intelligent analysis (Data Mining, Mahaut)

data analysis

- Visualization (Table au, Microstrategy)
- Hypothesis testing
- Interpretation of results

Identifying trends and forecasting

- Interpretation of results
- Modeling
- Estimation of applicability

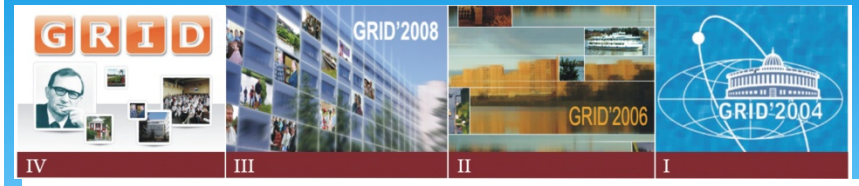
Solutions

- Templates
- Thesaurus
- Reference books
- Criteria
- Library of application interfaces

Projects in framework Distributed computing

- JINR-CERN cooperation (Worldwide LHC Computing Grid (WLCG), CMS, ATLAS, ALICE, COMPASS, NA62…)
- RDIG Development
- BigPANDA cooperation “Next Generation Workload Management System for BigData”
- Tier1 Center in Russia (NRC KI, LIT JINR)
- BMBF grant “Development of the grid-infrastructure and tools to provide joint investigations performed with participation of JINR and German research centers”
- “Development of grid segment for the LHC experiments” with South Africa;
- Development of grid segment at Cairo University and its integration to the JINR GridEdu
- JINR - FZU AS Czech Republic Project “The grid for the physics experiments”
- JINR-BITP(Kiev) project “Development and implementation of cloud computing technologies on grid-sites at LIT JINR and BITP for ALICE experiment”
- JINR-Romania cooperation Hulubei - Meshcheryakov programme
- JINR- Mongolia cooperation (MongolGRID)
- JINR-Bulgaria cooperation in cloud computing (INRNE, SU)
- JINR- Kazakhstan cooperation in cloud computing (NU, ENU)
- JINR-Belarus cooperation in grid and cloud computing (BU, BTU)
- JINR- Azerbaijan cooperation in cloud computing (IP)
- JINR-China cooperation (BES-III)
- JINR-DESY cooperation (dCache)
- JINR- FNAL cooperation in computing (NOVA)
- JINR-BNL cooperation in computing (STAR)
- Cooperation with Armenia, Georgia, Moldova, Poland, Slovakia, …

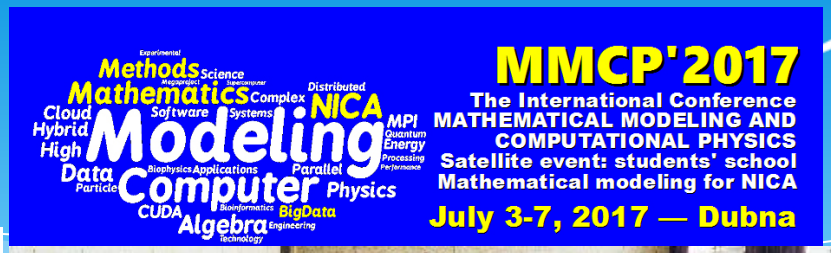
LIT traditional conferences



Distributed Computing and Grid-technologies in Science and Education

Mathematics. Computing. Education

DIGITAL LIBRARIES:
ADVANCED METHODS AND TECHNOLOGIES,
DIGITAL COLLECTIONS



IFIN-HH

ÚEF SAV

LIT schools

JINR / CERN

GRID AND ADVANCED INFORMATION SYSTEMS

History of Symposium

I	1963	Budapest
II	1964	Dubna
III	1965	Dresden
IV	1966	Prague
V	1968	Alushta
VI	1971	Warsaw
VII	1973	Budapest
VIII	1975	Dubna
IX	1977	Varna
X	1980	Dresden
XI	1983	Bratislava
XII	1985	Dubna
XIII	1988	Varna
XIV	1990	Warsaw
XV	1992	Warsaw
XVI	1994	Varna
XVII	1997	Varna
XVIII	2001	Varna
XIX	2003	Varna
XX	2005	Varna
XXI	2007	Varna
XXII	2009	Varna
XXIII	2011	Varna
XXIV	2013	Varna
XXV	2015	Budva
XXVI	2017	Budva

25-29 September 2017, Montenegro, Budva, Becici.