



Методы и технологии обработки данных в гетерогенных вычислительных средах

Кореньков Владимир Васильевич

Научный руководитель Лаборатории информационных технологий
имени М.Г. Мещерякова ОИЯИ

Зав. кафедрой «Распределенные информационно-вычислительные системы»
Государственного университета «Дубна»

Осенняя студенческая ИТ-школа,
Дубна, ЛИТ ОИЯИ, 7 октября 2024 г.

Joint Institute for Nuclear Research

International intergovernmental organization

Established on 26 March 1956 and registered with the United Nations

Located in Dubna, Russian Federation



JINR IS
A MEMBER
OF
39
COLLABORATIONS
AT SCIENTIFIC
CENTRES
AROUND THE WORLD

>900
PARTNER
NETWORK
ORGANIZATIONS



230M\$ budget

JINR in Figures:



16 Member States



1500 scientific publications
per year



5260 staff members



over **70** international
conferences
and workshops
per year



1200 researchers



1000 Doctors and
Candidates of Sciences



800 partner universities,
educational and
research centres
in more than
70 countries



2000 engineers and
technicians

Unique Park of Basic Facilities:

- World's Top Pulsed Neutron Source
- Heavy Ion Accelerators in a Wide Energy Range
- Megascience Project: Superconducting Collider NICA

7 JINR Laboratories,
each being comparable
with a large research institute
in the scale of investigations
performed



Frank Laboratory
of Neutron Physics



flnph.jinr.ru



Veksler and Baldin Laboratory
of High Energy Physics



lhph.jinr.ru



Flerov Laboratory
of Nuclear Reactions



flerovlab.jinr.ru



Dzhelepov Laboratory
of Nuclear Problems



dlnp.jinr.ru



Meshcheryakov Laboratory
of Information Technologies



lit.jinr.ru



Bogoliubov Laboratory
of Theoretical Physics



theor.jinr.ru



Laboratory
of Radiation Biology



lrb.jinr.ru

Implementation of the JINR Development Program



NICA complex



Baikal-GVD



IBR-2M



SHEF



Nuclotron



DC-280 SHE-factory



U-400 Heavy and super-heavy nuclei



U-400M Light exotic nuclei



ACCULINNA-2 Fragment separator



NanoLab



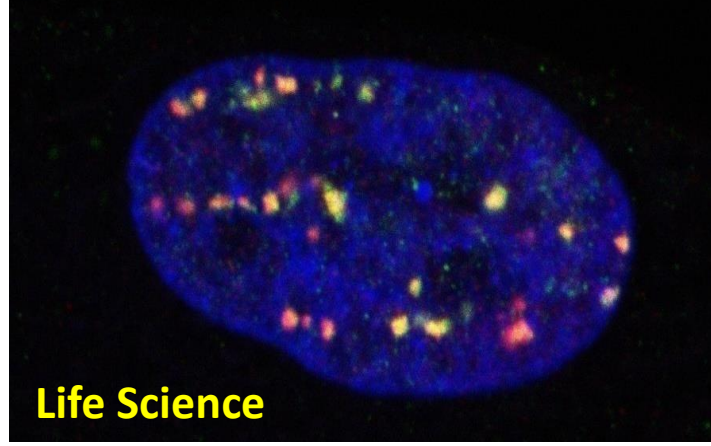
DRIBS-III



IT & CC



Life Science



Meshcheryakov Laboratory of Information Technologies



M.G. Meshcheryakov

(17.09.1910 - 24.05.1994)



N.N. Govorun

(18.03.1930 - 21.07.1989)

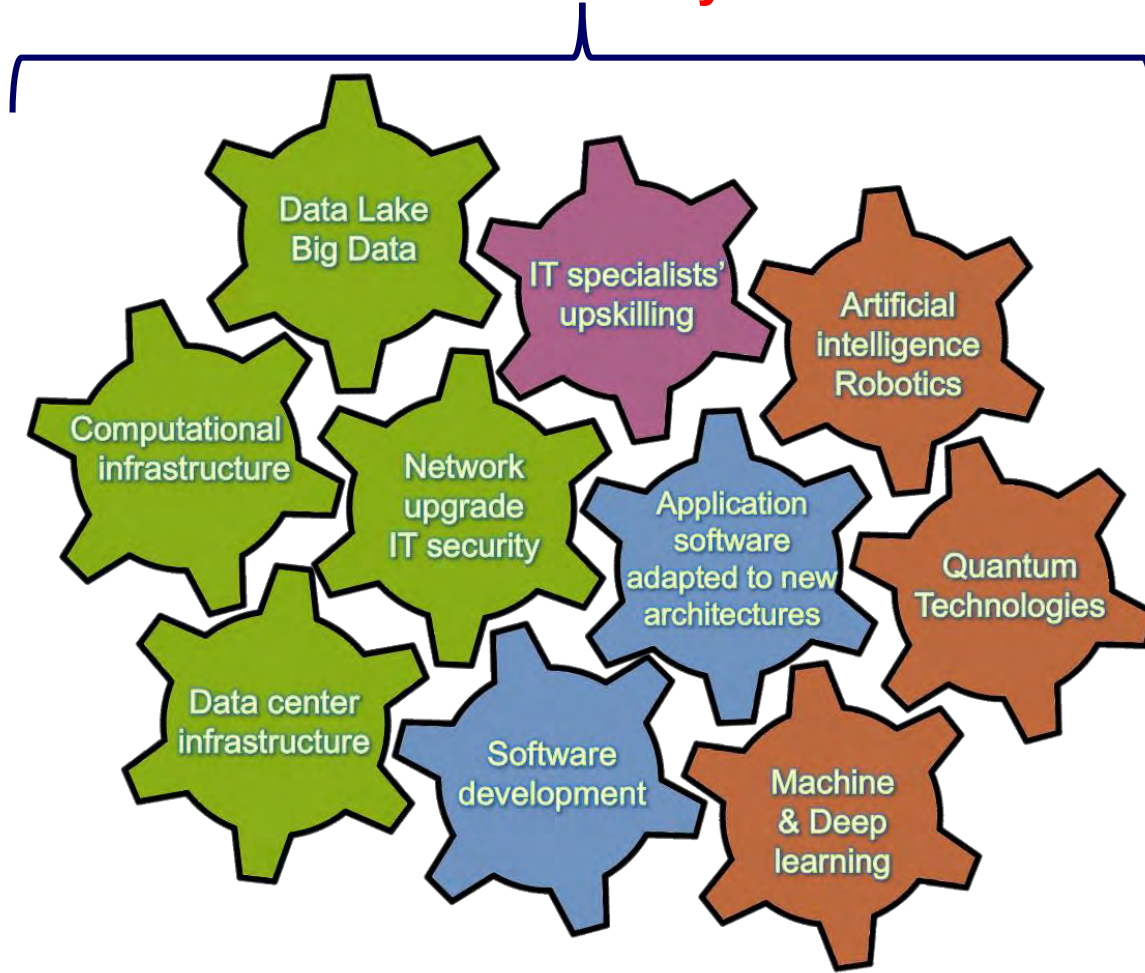
Meshcheryakov Laboratory of Information Technologies of the Joint Institute for Nuclear Research in Dubna was founded in August 1966. The main directions of the activities at the Laboratory are connected with the provision of networks, computer and information resources, as well as mathematical support of a wide range of research at JINR.



Strategy for Information Technology and Scientific Computing at JINR



Scientific IT ecosystem:



Coordinated development of interconnected IT technologies and computational methods

It will be a steady implementation/upgrades of

- Networking (**Tb/s** range),
- Computing infrastructure within the **Multifunctional Information & Computing Complex (MICC)** and
- “Govorun” Supercomputer,
- Data center infrastructure,
- Data Lake & long-term storage for all experiments.

The development of new data processing and analysis algorithms based on

- ML/DL,
- Artificial intelligence,
- Big Data
- Quantum technologies.

A variety of means will be used for IT specialists' upskilling.

Грид технологии – путь к успеху



На торжестве по поводу получения Нобелевской премии за открытие бозона Хиггса директор ЦЕРНа Рольф Хойер прямо назвал **грид-технологии одним из трех столпов успеха** (наряду с ускорителем LHC и физическими установками).

Без организации грид-инфраструктуры на LHC было бы невозможно обрабатывать и хранить колоссальный объем данных, поступающих с коллайдера, а значит, совершать научные открытия.

Сегодня уже ни один крупный проект не осуществим без использования распределенной инфраструктуры для обработки данных.



Эволюция ИТ

- менялись концепции
- круг и сложность решаемых задач
- возникал новый технологический набор
- углублялась специализация разработчиков
- сокращалось время ввода новых продуктов и сервисов

первые цифровые платформы
поддержки, работающих в
реальном времени

Первое поколение (1960-е годы) — мейнфреймы

Второе поколение (1970-е годы) — универсальные ЭВМ

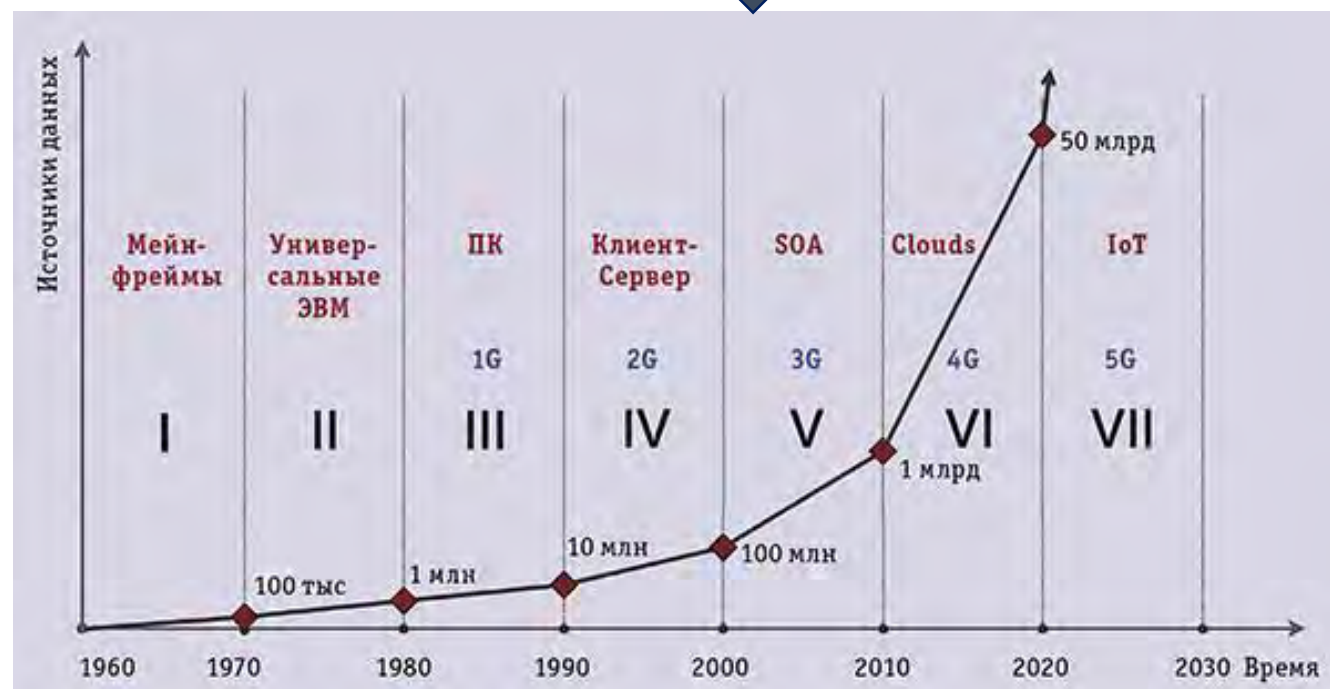
Третье поколение (1980-е годы) — персональные компьютеры

Четвертое поколение (1990-е годы) — клиент-сервер

Пятое поколение (2000-е годы) — сервисная архитектура

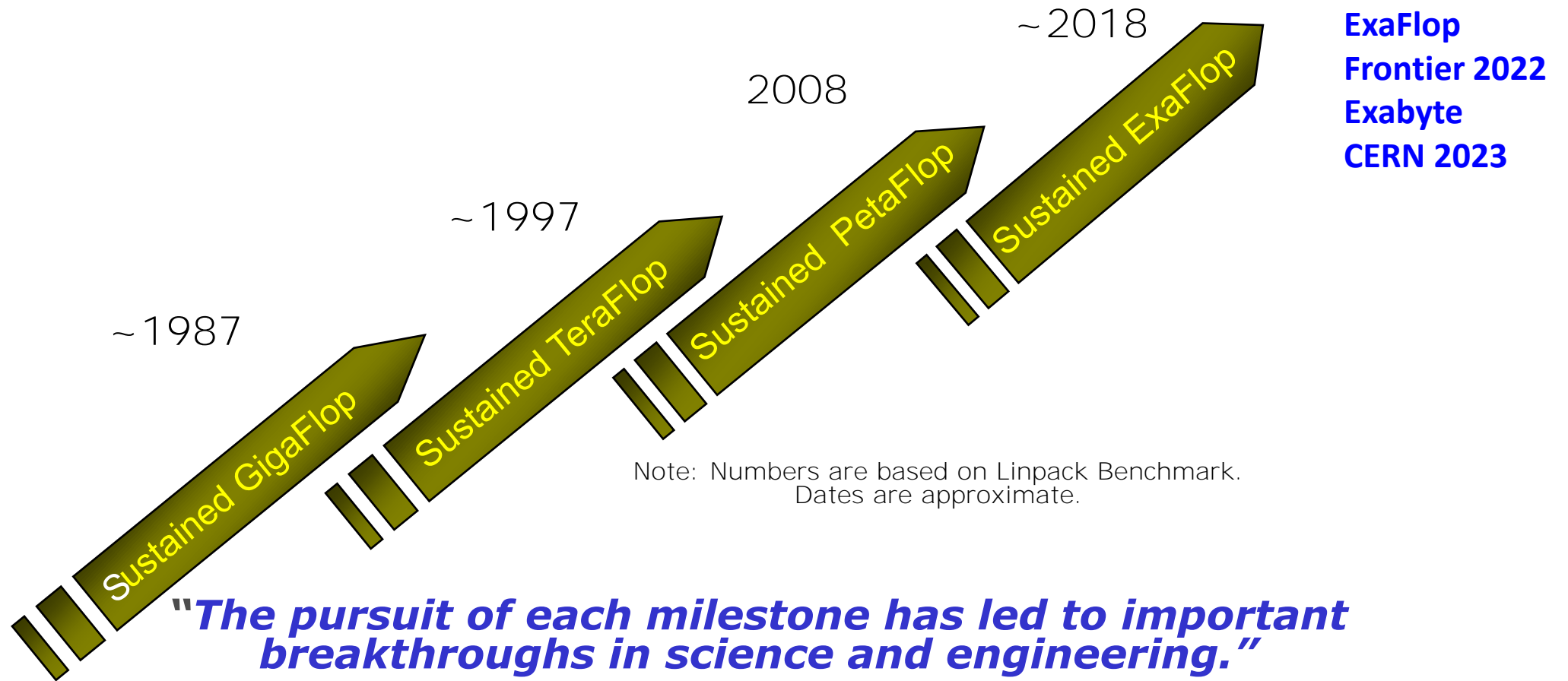
Шестое поколение (2010-е годы) — облака

Седьмое поколение (2020-е годы) — IoT, искусственный интеллект, квантовые вычисления



Reach Exascale by 2018

From GigFlops to ExaFlops



Note: Numbers are based on Linpack Benchmark.
Dates are approximate.

"The pursuit of each milestone has led to important breakthroughs in science and engineering."

Source: IDC "In Pursuit of Petascale Computing: Initiatives Around the World," 2007

Концепция Грид

«Грид - это система, которая:

- координирует использование ресурсов при отсутствии централизованного управления этими ресурсами
- использует стандартные, открытые, универсальные протоколы и интерфейсы.
- обеспечивает высококачественное обслуживание»

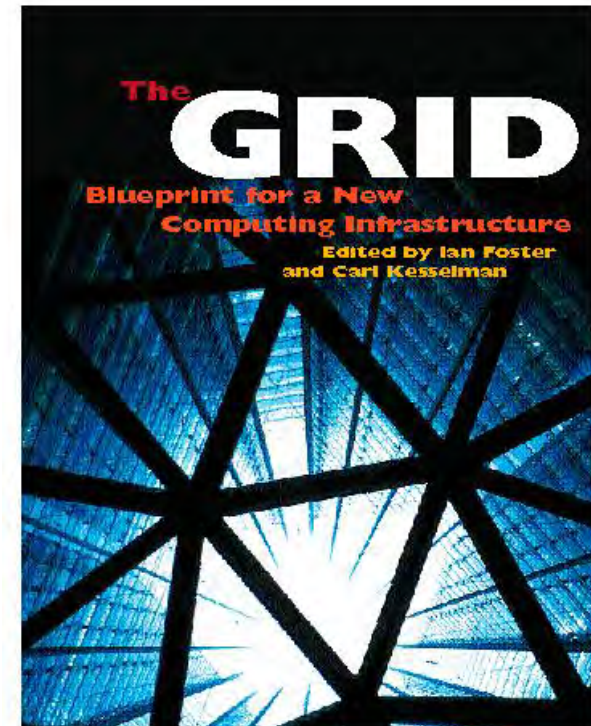
(Ian Foster: "What is the grid? ", 2002 г.)

Модели грид:

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

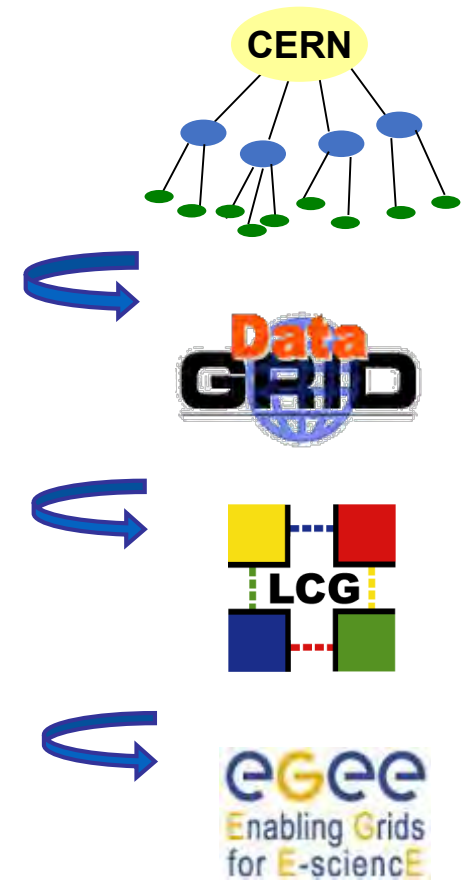
Междисциплинарный характер грид: развиваемые технологии применяются в физике высоких энергий, космофизике, микробиологии, экологии, метеорологии, различных инженерных и бизнес приложениях.

Виртуальные организации (VO)



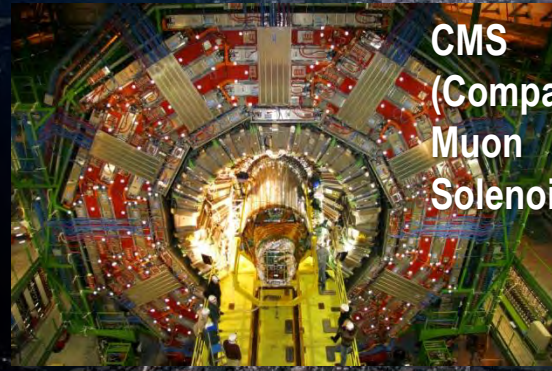
Some history

- **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**



Large Hadron Collider

The Large Hadron Collider (**LHC**), one of the largest and truly global scientific projects ever, is the most exciting turning point in particle physics.



CMS
(Compact Muon Solenoid)



LHCb
(Large Hadron Collider beauty experiment)

Data flow to permanent storage: 4-6 GB/sec

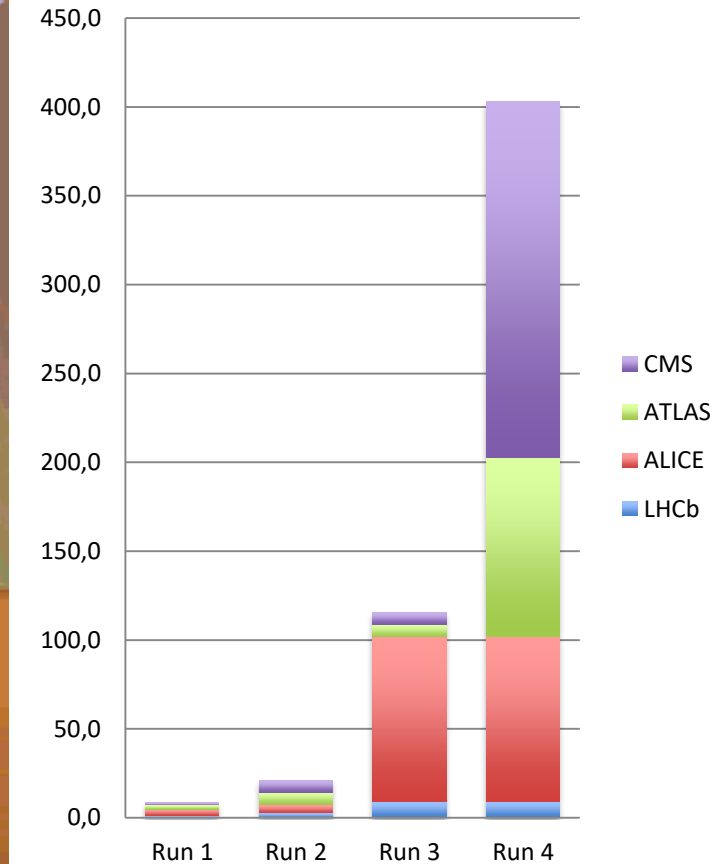
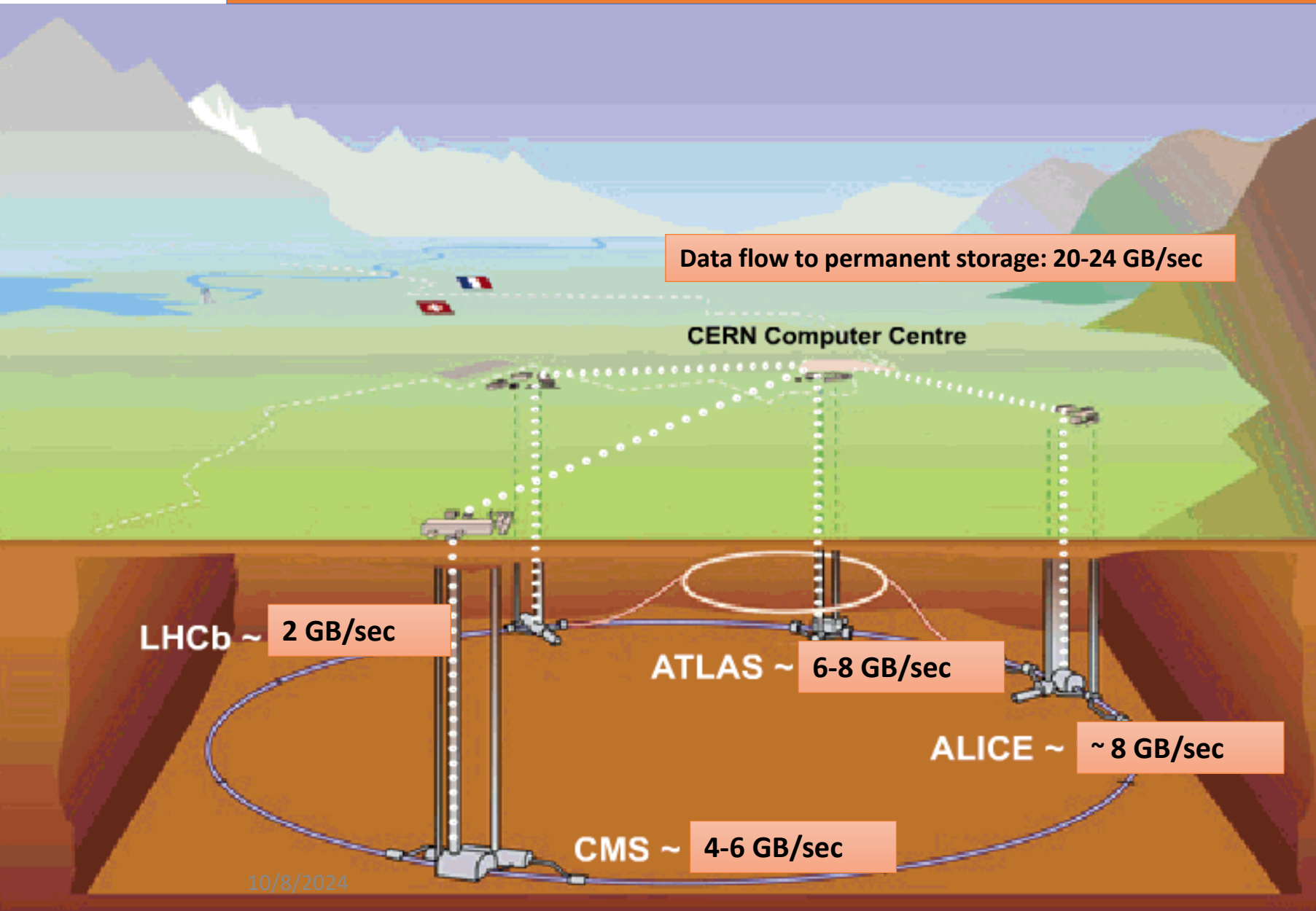


ALICE
(A Large Ion Collider Experiment)

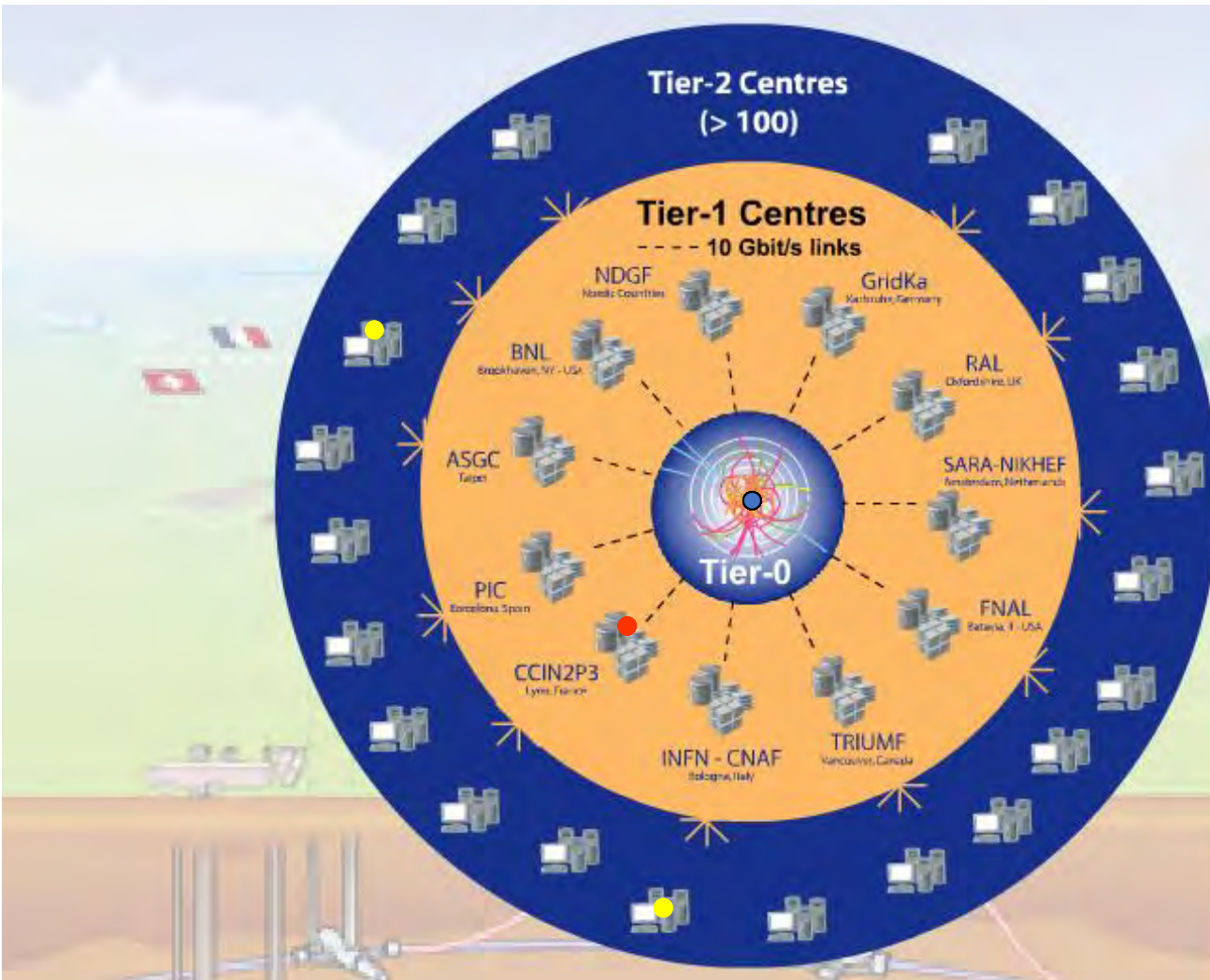


ATLAS
(A Toroidal LHC ApparatuS)

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

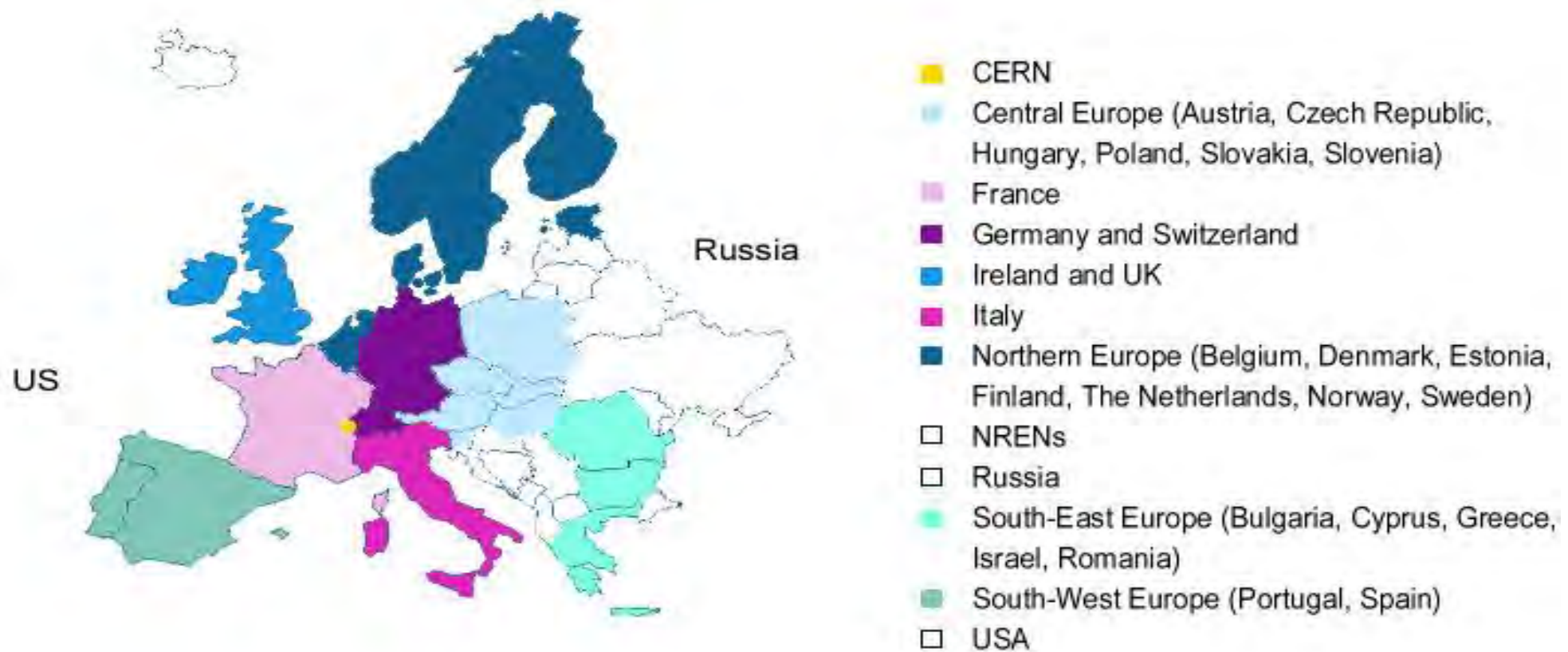
Tier-1 (11 centers):

- receives a data from the Tier-0
- data processing (re-reconstruction, 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

Проект EGEE - Enabling Grids for E-science



Проект EGEE - Enabling Grids for E-science направлен на создание международной инфраструктуры, основанной на технологиях грид. Проект выполняется консорциумом из **70 институтов в 27 странах**, объединенных в региональные гриды.



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. In 2010 the RDIG infrastructure comprises **12 Resource Centers with > 3000 CPU and > 5000 TB of disc storage.**



RDIG Resource Centres:

- ITEP
- JINR-LCG2 (Dubna)
- RRC-KI
- RU-Moscow-KIAM
- RU-Phys-SPbSU
- RU-Protvino-IHEP
- RU-SPbSU
- Ru-Troitsk-INR
- ru-IMPB-LCG2
- ru-Moscow-FIAN
- ru-Moscow-MEPHI
- ru-PNPI-LCG2 (Gatchina)
- ru-Moscow-SINP
- **Kharkov-KIPT (UA)**
- **BY-NCPHEP (Minsk)**
- **UA-KNU**
- **UA-BITP**



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

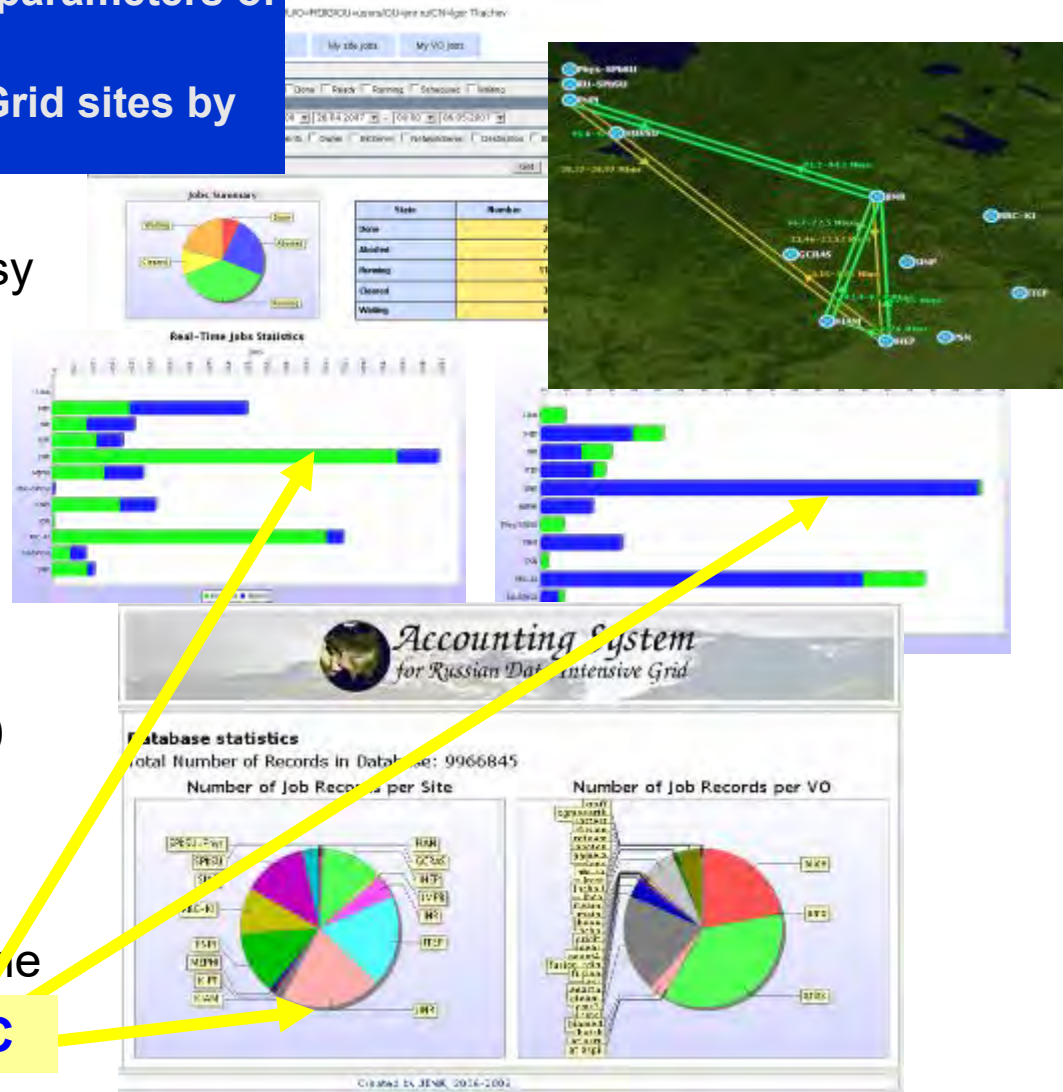
Waiting time

Totally sum in seconds

Average ratio waiting/used CPU time per job

Physical memory

Average per job



JINR CICC

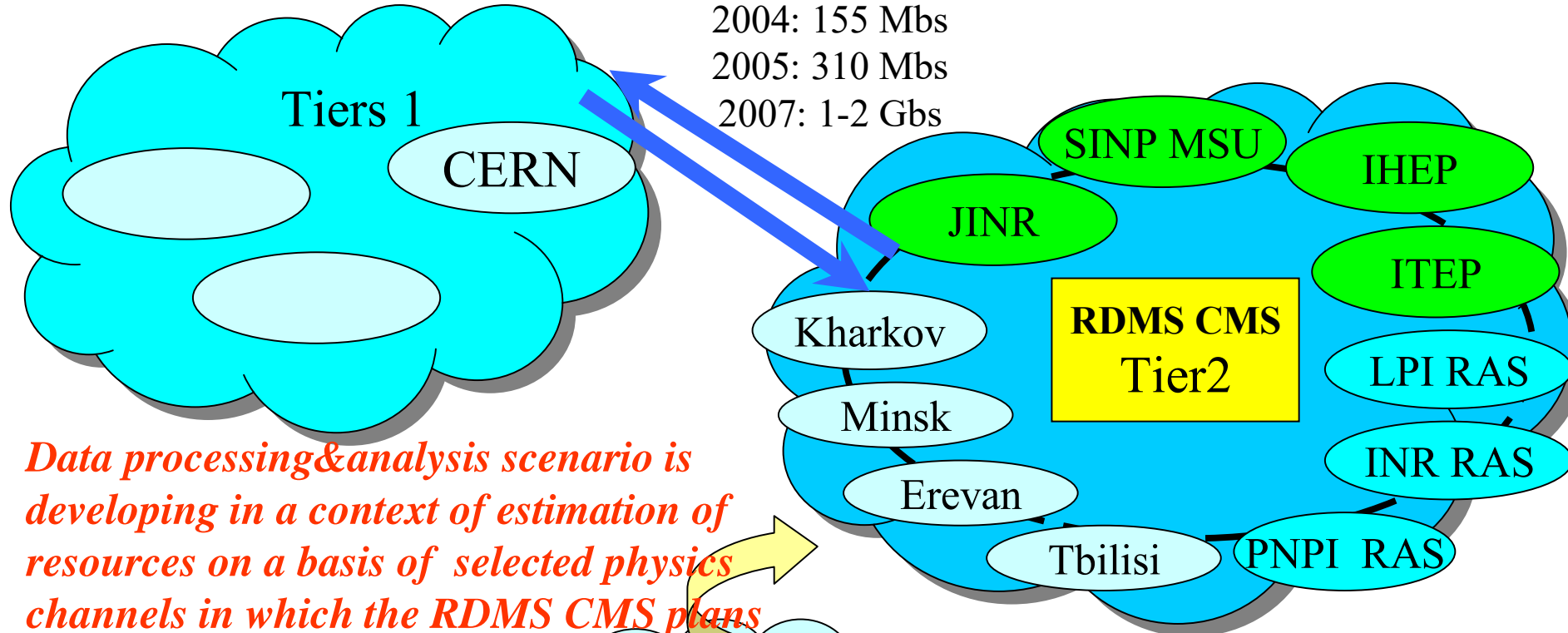
Development of RDMS CMS computing model

RuTier2 Grid access to CERN

2004: 155 Mbs

2005: 310 Mbs

2007: 1-2 Gbs



Data processing&analysis scenario is developing in a context of estimation of resources on a basis of selected physics channels in which the RDMS CMS plans to participate.

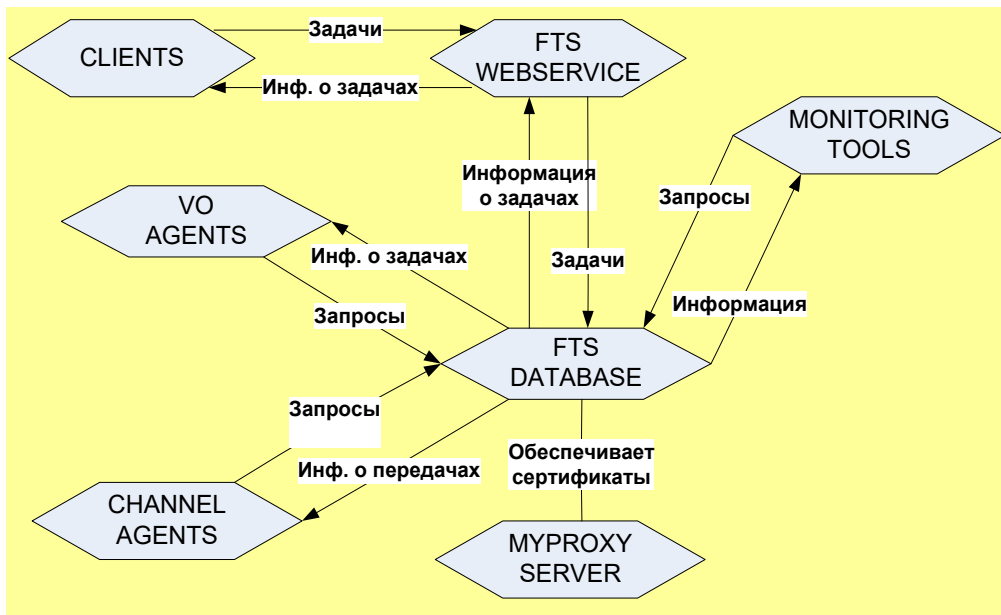
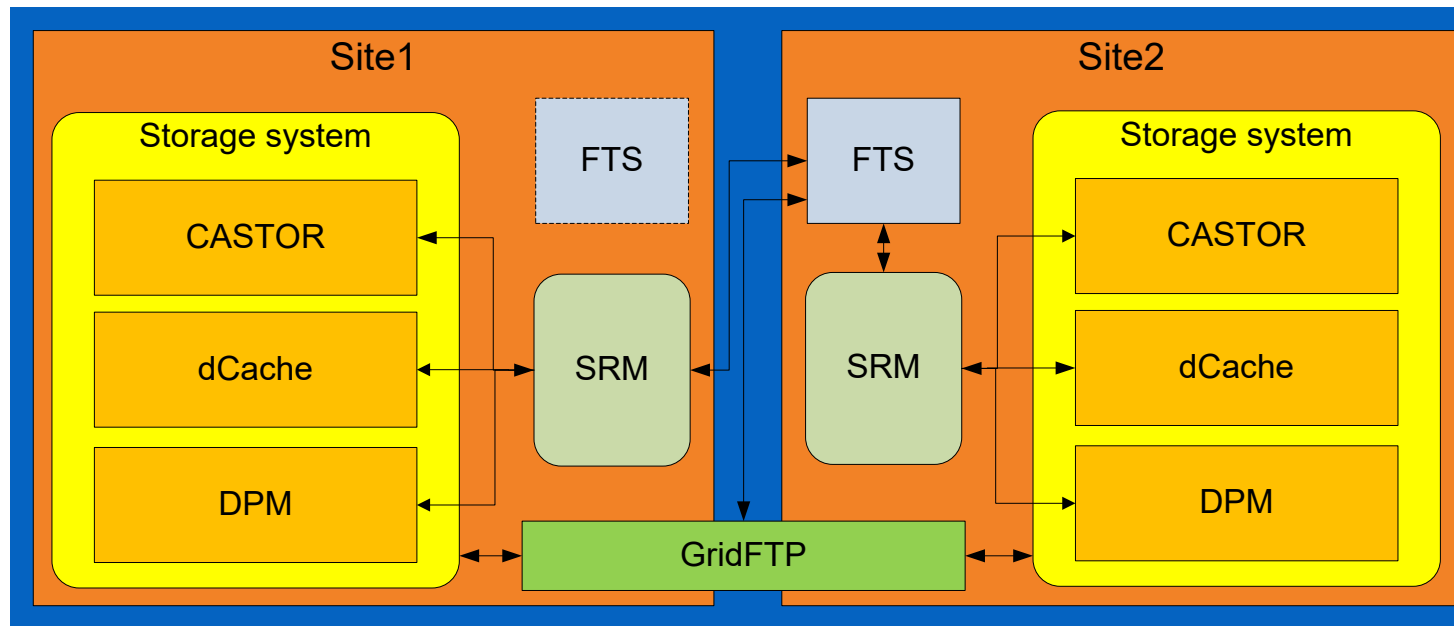
Collaborative centers:
RCC MSU,
RRC KI

Institutes in Russia: IHEP, ITEP, JINR, SINP MSU (as 4 basic centers), LPI RAS, INR RAS, PNPI RAS
Collaborative RDMS Institutes: ErPhI (Armenia), HEPI (Georgia) KIPT (Ukraine), NCPHEP (Belarus)

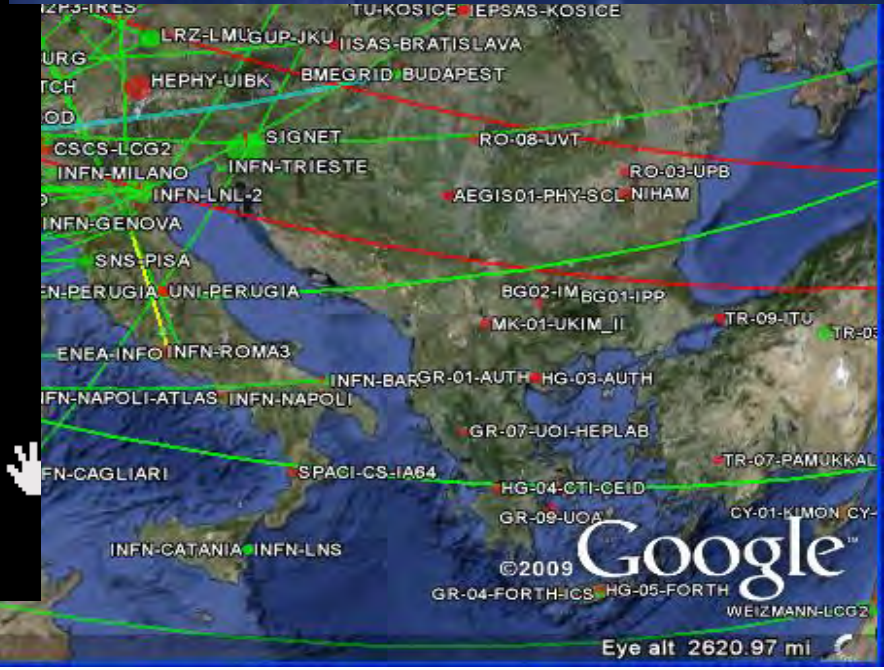
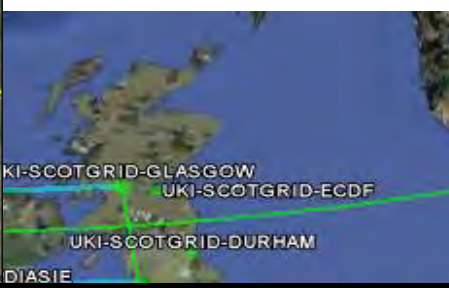
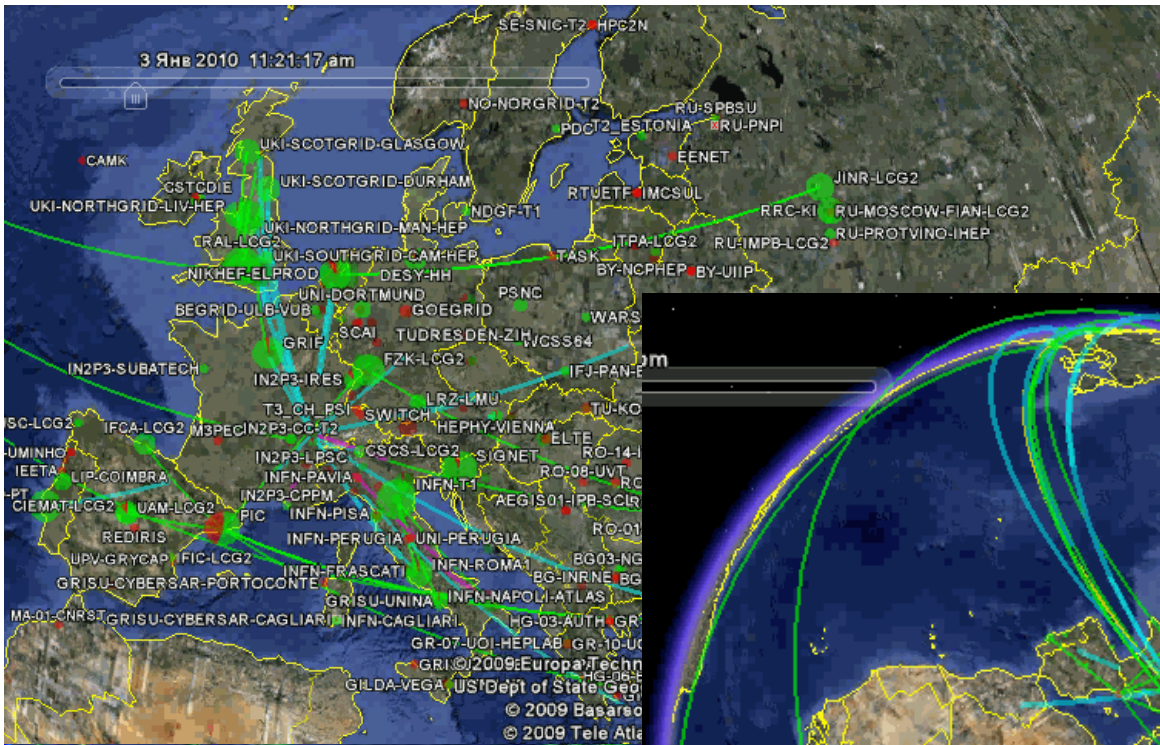
RDMS CMS computing centers



File Transfere System Monitoring and Testing



The Worldwide LHC Computing Grid (WLCG)



© 2009 Cnes/Spot Image
47°27'19.45" N 6°33'30.91" E



Participation in WLCG MCDB Development

- 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

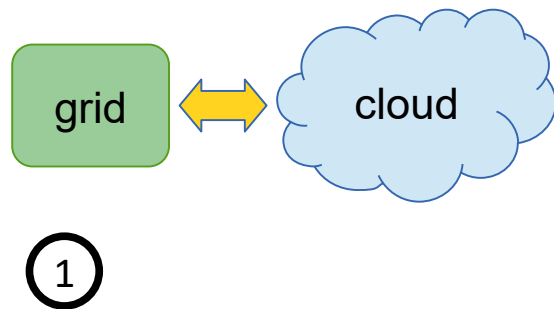
A screenshot of a web browser window displaying the Monte-Carlo Data Base website. The browser title is "Monte-Carlo Data Base - Windows Internet Explorer provided by CERN". The address bar shows "http://mcdb.cern.ch/". The website header features the text "MCDB - MonteCarlo Database". Below the header, there is a search bar on the left and a main content area on the right. The search bar contains the text "Search this site" and a "Go" button. The main content area displays search results for "P,P->WBBJ->L,MU,B,B,J PROCESS FROM COMPHEP". The results show a list of articles with details such as author names, publication dates, and IDs. On the right side of the page, there are several navigation and utility links, including "Login to MCDB", "Help and support", "HepML", and "MCDB software". The footer of the page includes statistics and copyright information: "0002414 times visited since October Statistics of visits to MCDB © 2005-2007 Monte Carlo Generators group, LCG".

<http://mcdb.cern.ch/>

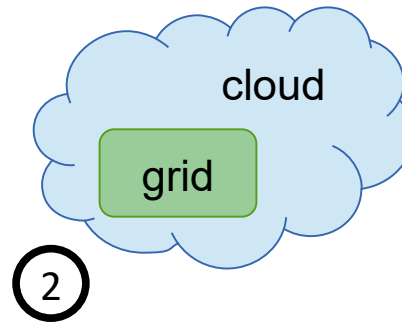


Clouds and grids

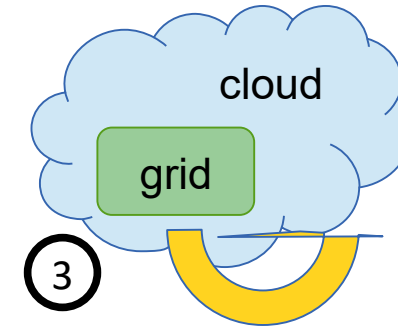
- Both ones **share** two main attributes:
 - they provide access to remote computing resources; and
 - they provide a service.
- But they are based on different paradigms:
 - clouds are being based on **virtualization of resources**,
 - grids being based on the **sharing of resources across boundaries**.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



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

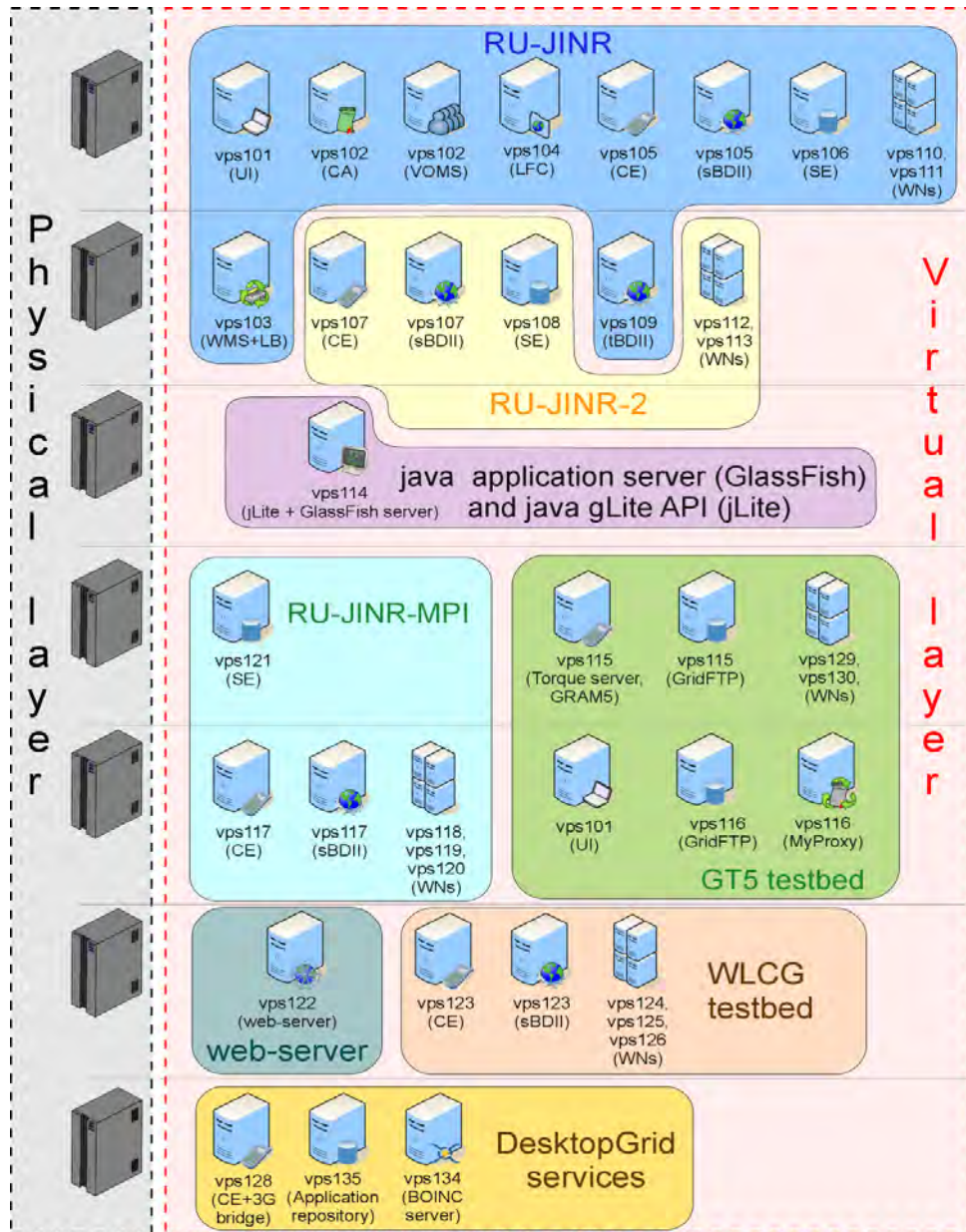


Services of grid site are
deployed on cloud VMs
(to increase hardware
utilization efficiency and
simplify admins' work)



Synergy of 1st and 2nd
approach

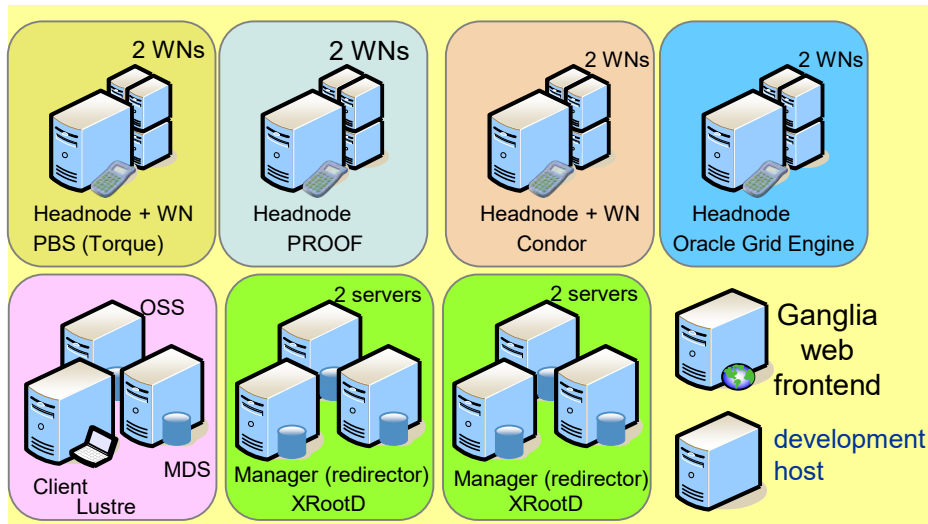
T-infrastructure implementation



- All services are deployed on VMs (OpenVZ)
- Main parts:
 - three grid sites on gLite middleware,
 - GT5 testbed,
 - desktop grid testbed based on BOINC,
 - testbed for WLCG activities.
- Running since 2006

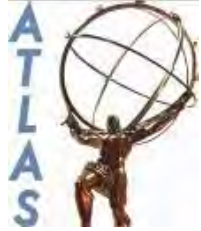
Система мониторинга Tier3-центров для анализа данных экспериментов БАК

Для исследования центров уровня Tier3 в ОИЯИ была разработана архитектура тестовой инфраструктуры, которая позволяет создавать прототипы различных конфигураций центров уровня Tier3. С применением технологий виртуализации такая инфраструктура была реализована на базе виртуальных кластеров, что позволило разработать для каждого варианта конфигурации документацию, настроить или разработать средства локального мониторинга, выработать полные рекомендации по системе сбора информации для глобального мониторинга центров уровня Tier3.



Реализация системы мониторинга Tier3 – центров (не-грид) имеет огромное значение для координации работ в рамках виртуальной организации, так как обеспечивается глобальный взгляд на вклад Tier3 сайтов в вычислительный процесс.





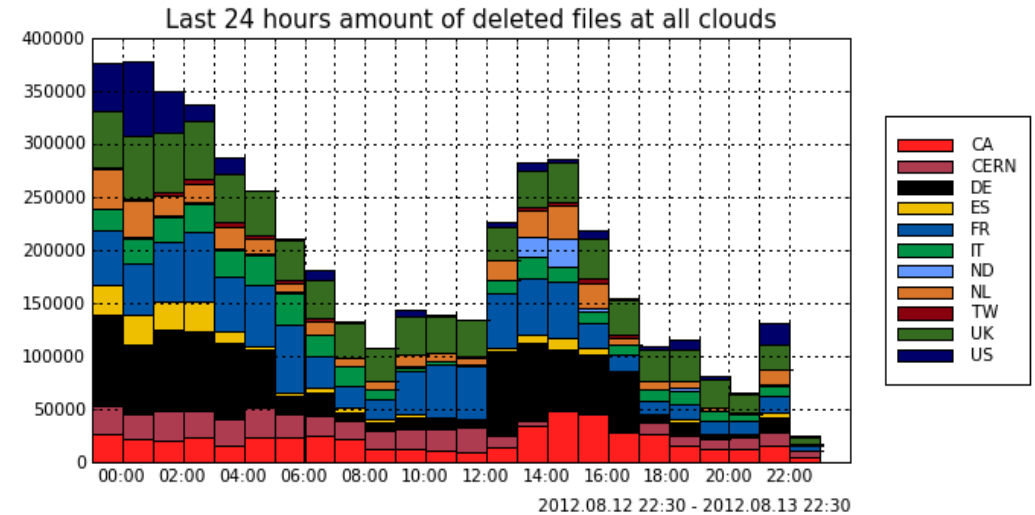
Архитектура системы распределенного хранения данных эксперимента ATLAS

Разработка новой архитектуры сервиса удаления данных для обеспечения целостности распределенного хранения информации эксперимента ATLAS.

- обслуживает запросы на удаление,
- организует балансировку нагрузки,
- обеспечивает масштабируемость и отказоустойчивость системы,
- корректную обработку исключений, возникающих в процессе работы,
- стратегию повтора операций в случае возникновения отказов.

Разработаны:

- новый интерфейс между компонентами сервиса
- создана новая схема базы данных,
- перестроено ядра сервиса,
- созданы интерфейсы с системами массового хранения,
- модернизирована система мониторинга работы сервиса.



Созданный сервис обеспечивает целостность хранения информации в географически распределенной среде. Данные эксперимента ATLAS распределены более, чем на 130 GRID-сайтах с общим объемом дискового пространства более 120 петабайт, в котором хранятся сотни миллионов файлов. Недельный объем удаляемых данных составляет 2 Пб (20 000 000 файлов).

Глобальная система мониторинга передачи данных в инфраструктуре WLCG

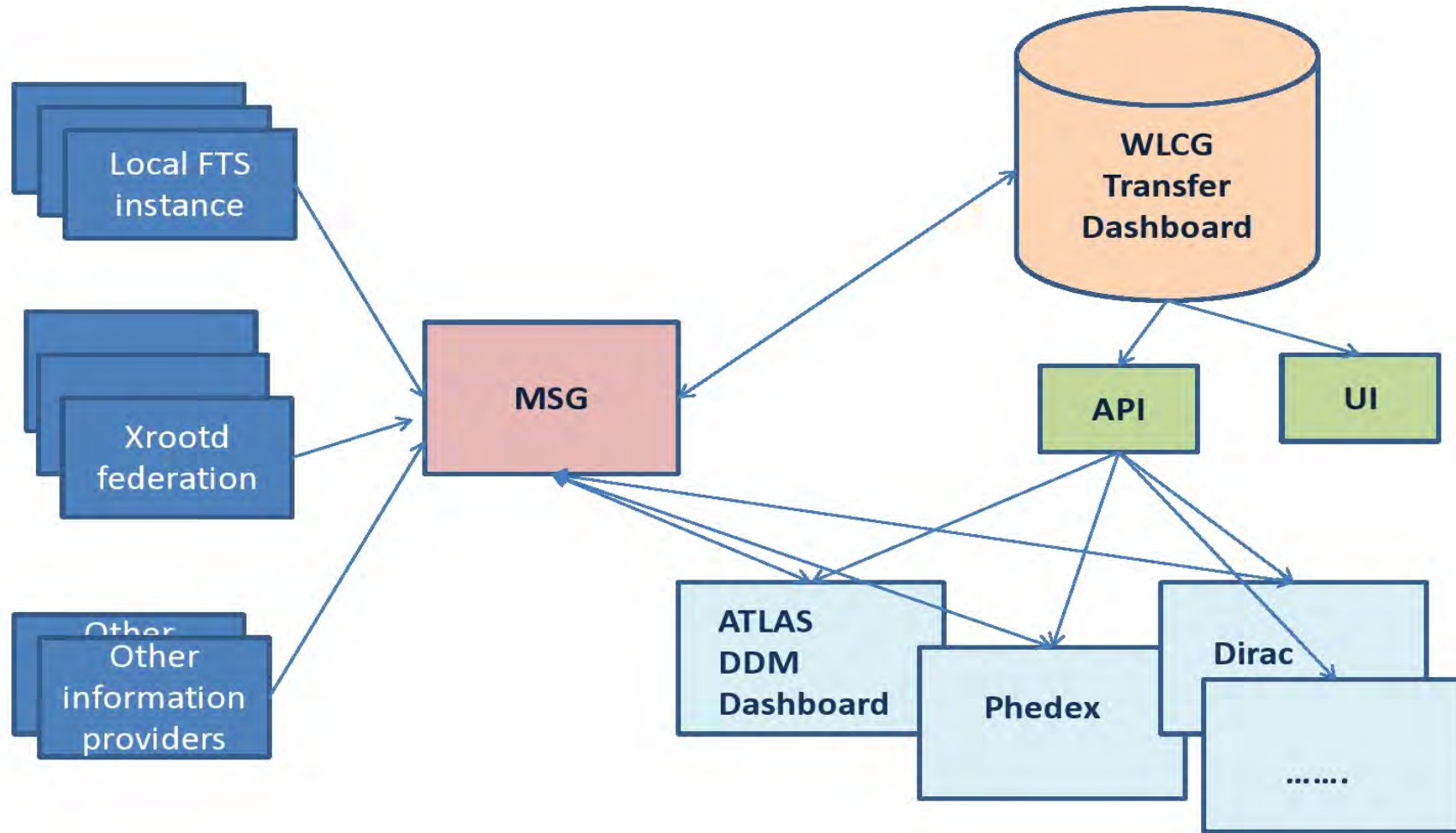
Суть проекта состоит в создании универсальной системы мониторинга, способной собирать информацию :

- о каждой передаче данных (Петабайта в день),
- независимо от метода осуществления (несколько протоколов и сервисов передачи файлов, FTS, xROOTd),
- уровень ресурсного центра (Tier-0, Tier-1, Tier-2),
- принадлежности данных определенной виртуальной организации;
- передавать с высокой степенью надежности собранную информацию в центральное хранилище (ORACLE, HADOOP);
- обрабатывать собранные данные для предоставления различным потребителям и программные интерфейсы для получения данных.



Система позволяет полностью удовлетворить потребности в информации различных типов пользователей и администраторов инфраструктуры WLCG.

Архитектура глобальной системы мониторинга передачи данных в инфраструктуре WLCG



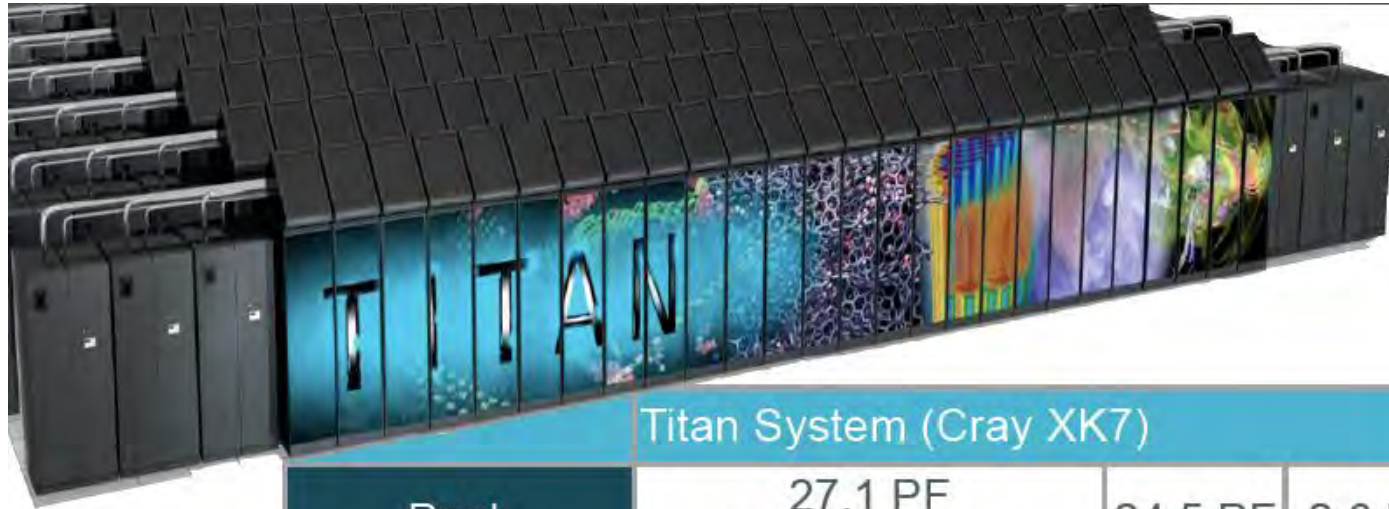
JINR CMS Remote Operation Centre

Founded in 2009 as a part of GRID-based JINR Tier-2

- ❑ 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

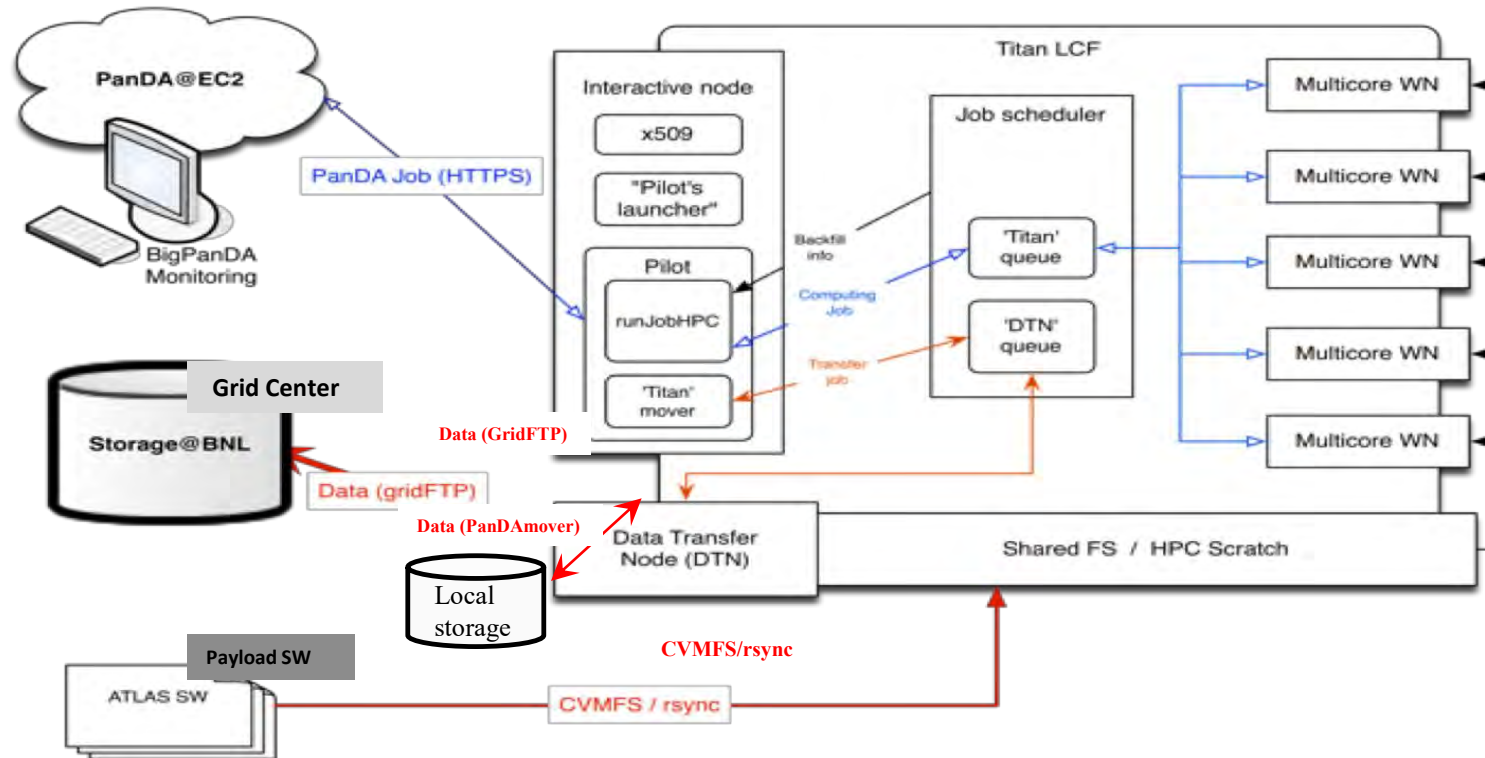


Leadership Computing Facilities. Titan



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		

Extending PanDA to Oak Ridge Leadership Computing Facilities



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

Интеграция технологий распределенных вычислений для управления большими данными

RanDa, как платформа, обеспечивающая прозрачность процесса хранения, обработки и управления данными для приложений с большими потоками данных и массивными вычислениями.

Развитие RanDA в направлении интеграции различных систем распределенных и параллельных вычислений (грид, cloud, кластеры, ЦОД, суперкомпьютеры) с целью создания универсальной платформы для крупных проектов управления большими данными в науке, государственном управлении, медицине, высокотехнологической промышленности, бизнесе.

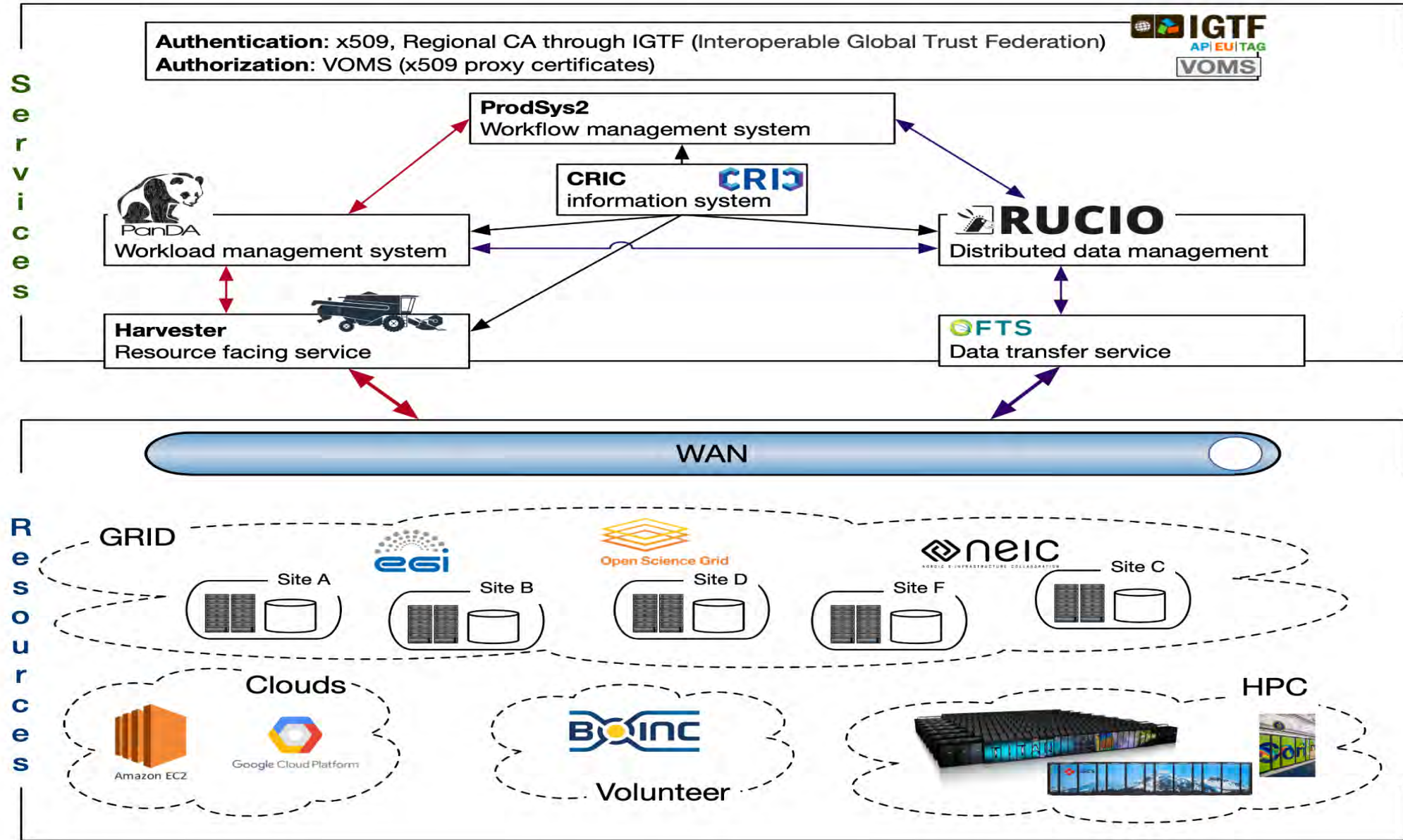
За большими данными следит ПАНДА

Текст
А. Каменцов, сотрудник Брунновской национальной лаборатории,
А. Ваньшин, сотрудник Яготинской национальной лаборатории,
В. Кореньков, ОИЯИ, г. Дубна
Иллюстрация Владимир Камаса

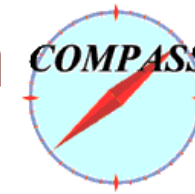


Броский и звучный термин «Big Data» прочно занял свое место в лексиконе не только ученых и компьютерных профессионалов, но и политиков разного ранга.

ATLAS computing



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



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

The Federal Target Programme Project: «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 : Ian.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,

Dr. Ian Bird
LHC Computing Grid Project Leader
IT Department
CERN

Cc: Prof. Sergio Bertolucci, Dr. Viacheslav Ilyin

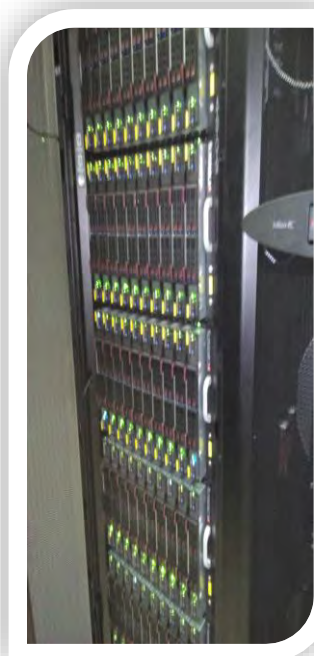
Создание CMS Tier-1 в ОИЯИ

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

- Инженерная инфраструктура (система бесперебойного электропитания, климат-контроля);
- Высокоскоростная надежная сетевая инфраструктура с выделенным резервируемым каналом в ЦЕРН (LHCOPN);
- Вычислительная система и система хранения на базе дисковых массивов и ленточных библиотек большой емкости;
- Надежность и доступность 100%



Tier1 center JINR for the CMS experiment



Current configuration

6000 cores

6 PB disks

11 PB tapes

100% R/A

70 TB per day

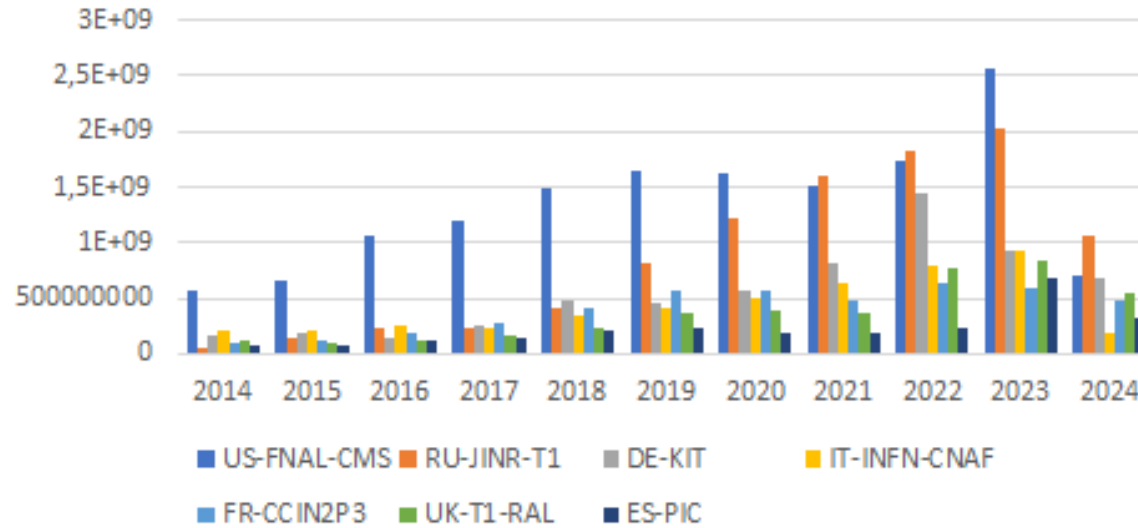
Grid Tier-1 is one of the 7 centres in the world intended for large-scale processing of experimental and event-modeling data coming from the centres of Tier-0 (CERN), as well as Tier-1 and Tier-2 of the Worldwide Computing Grid WLCG for the CMS experiment

JINR Tier1 for CMS

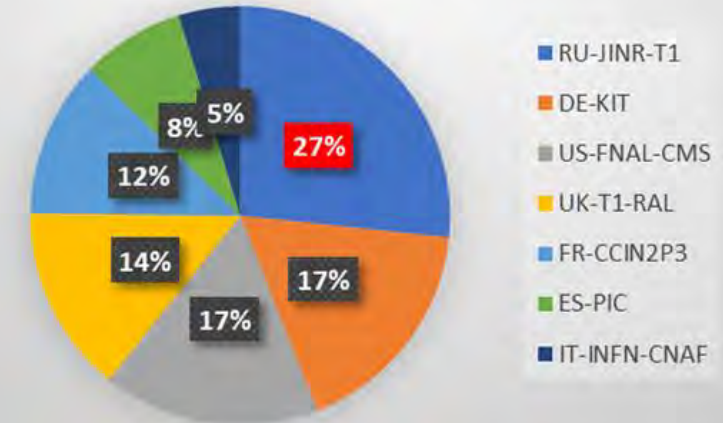
- 20064 cores
- 360 kHS23
- 15 PB disks
- 103 PB tapes
- 100% reliability and availability



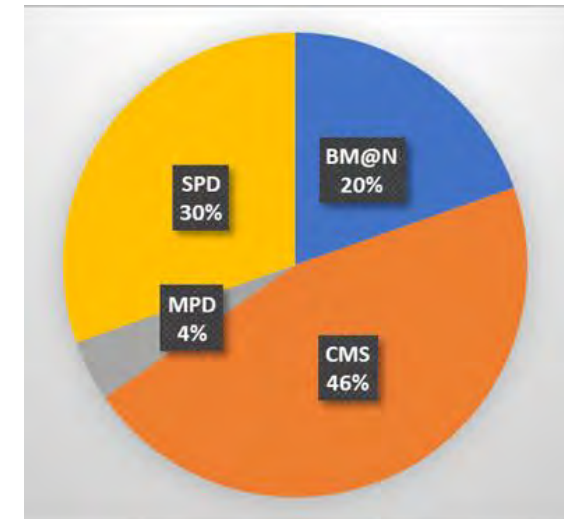
Accounting - 2014_1 to 2024_8 normcpu for CMS TIER1 and Year



Accounting - 2024_1 to 2024_8 normcpu for CMS TIER1 and DATE



Tier1 CMS	2023 July - 2024 Sep	%
RU-JINR-T1	2,329,354,057	25,43
US-FNAL-CMS	1,995,967,865	21,61
DE-KIT	1,318,968,102	14,19
UK-T1-RAL	1,206,928,818	13,00
FR-CCIN2P3	873,169,312	9,04
IT-INFN-CNAF	795,249,12	8,61
ES-PIC	709,757,725	7,68



Tier2 at JINR

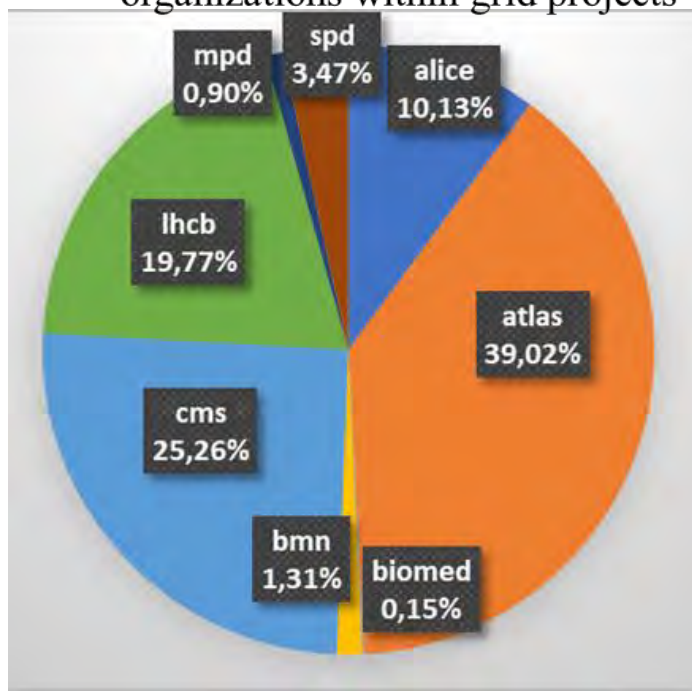
Использование Tier2 сайта ОИЯИ (JINR-LCG2) виртуальными организациями в рамках грид-проектов



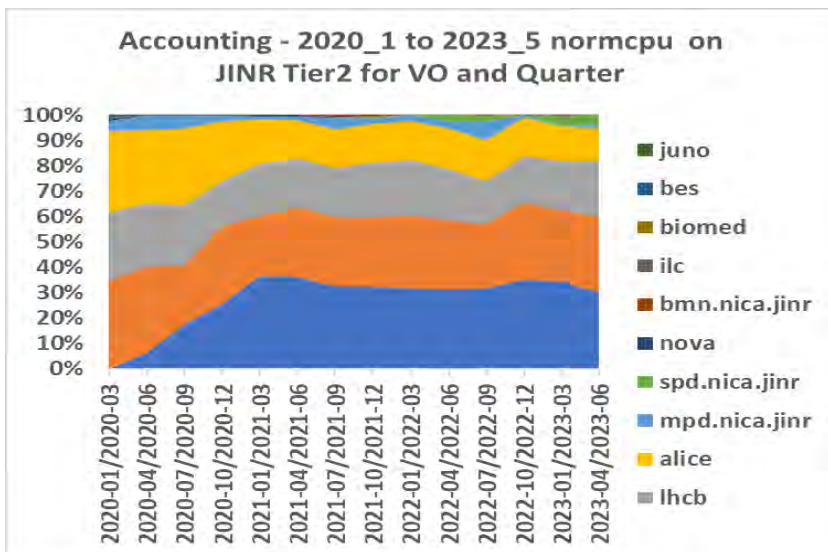
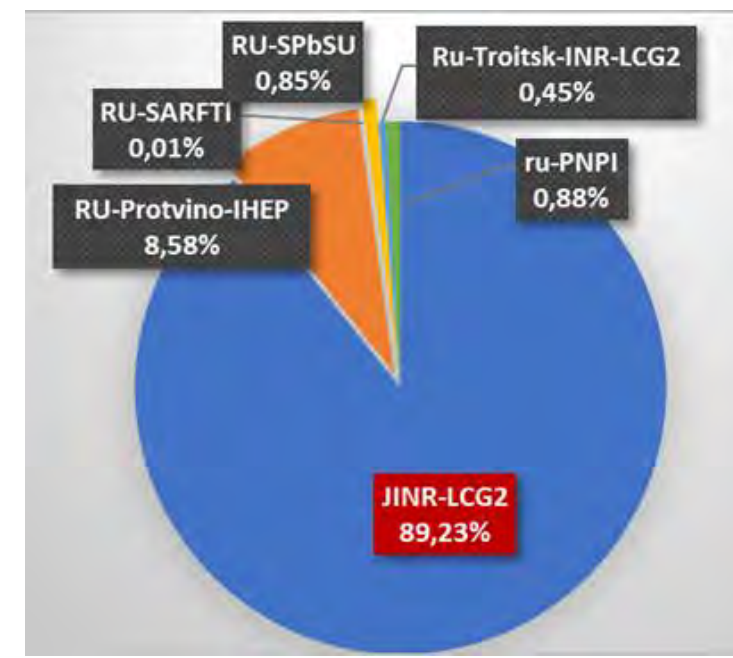
Tier2 at JINR provides computing power and data storage and access systems for the majority of JINR users and user groups, as well as for users of virtual organizations (VOs) of the grid environment (LHC, NICA, FAIR, etc.).

The JINR Tier2 output is the highest (89.58%) in the Russian Data Intensive Grid (RDIG) Federation.

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



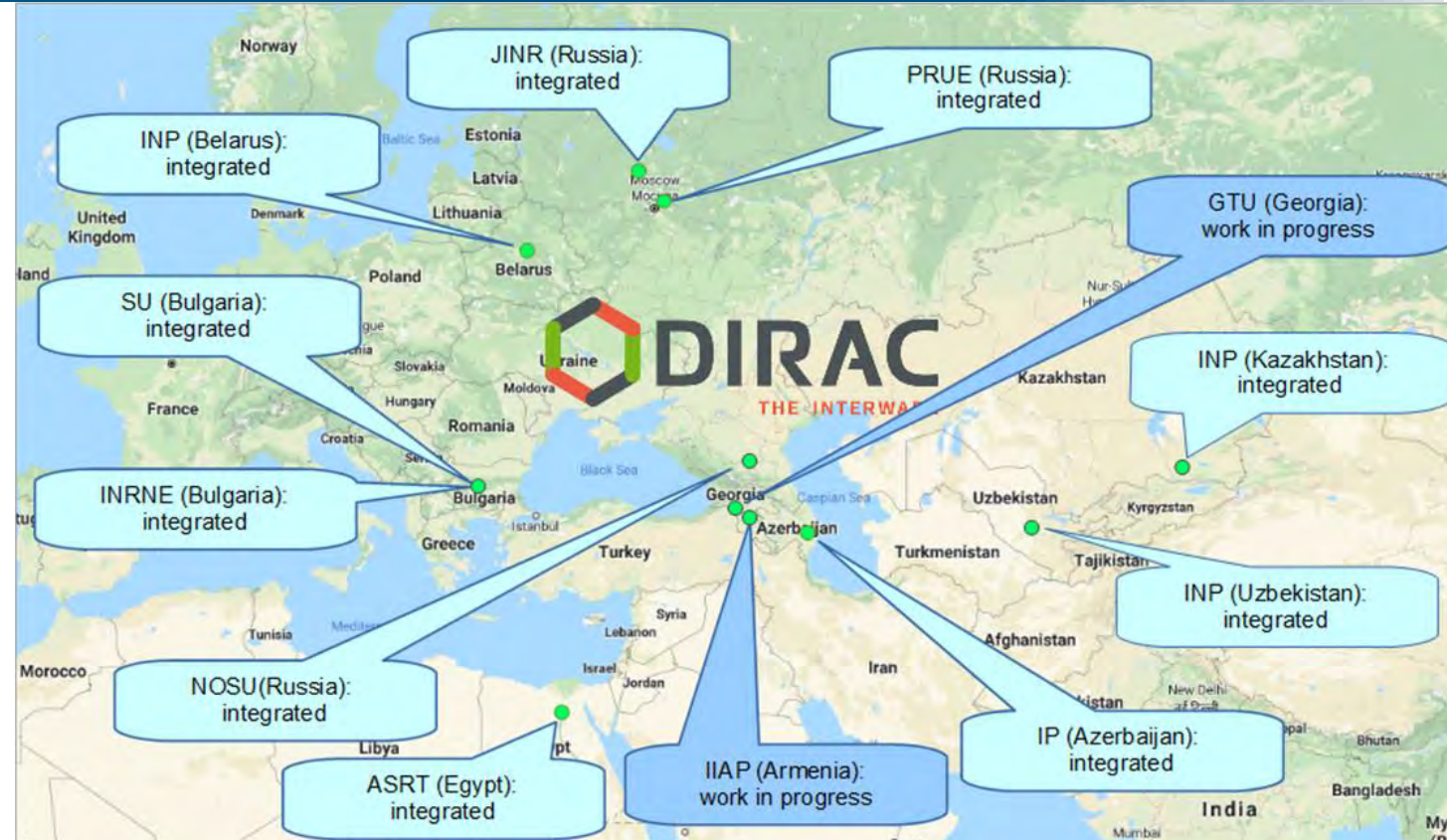
Distribution of RDIG grid sites by the CPU time for data processed



Cloud Infrastructure



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



- Cloud Platform - OpenNebula
- Virtualization - KVM
- Storage (Local disks, Ceph)
- Total Resources
~ **5,152** CPU cores; 80 TB RAM;
3.5 PB of raw ceph-based storage

- Computational resources for neutrino experiments
- Testbeds for research and development in IT
- COMPASS production system services
- Data management system of the UNECE ICP Vegetation
- Scientific and engineering computing
- Service for data visualization
- VMs for JINR users



Суперкомпьютер имени Н.Н. Говоруна



40
узлов

21
узел

5
узлов

CPU-компонента
Intel Xeon Scalable

CPU-компонента
Intel Xeon Phi

GPU-компонента
GPU DGX-1 Volta
(NVIDIA Tesla V100)

Пиковая производительность:
Двойная точность: **500 Tflops**
Одинарная точность: **1000 Tflops**



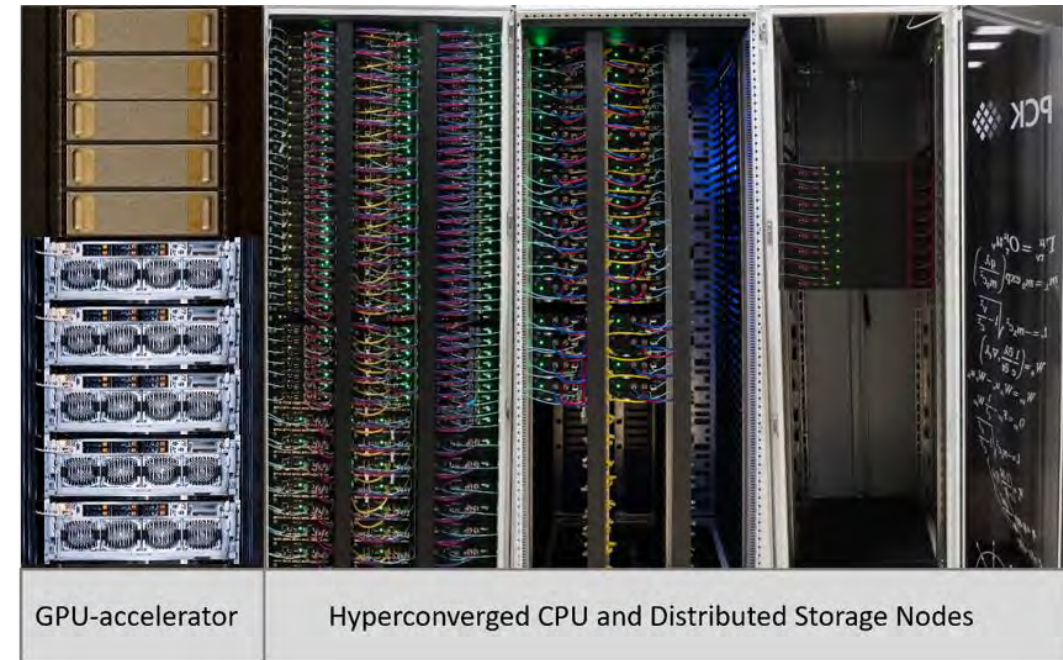
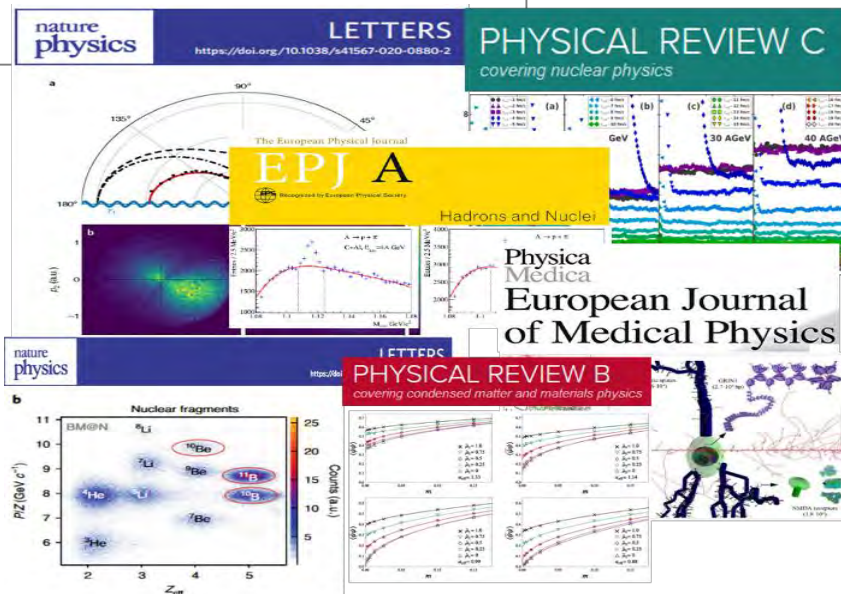
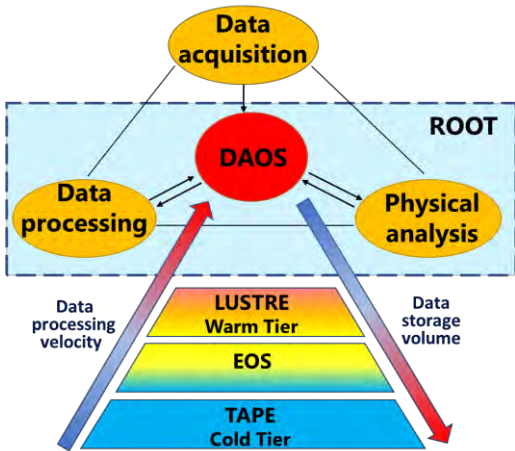
“Govorun” Supercomputer



- Hyper-converged software-defined system
- Hierarchical data processing and storage system
- Scalable solution Storage-on-demand
- Total peak performance: 1.7 PFlops DP
- 26 PFlops for half-precision computations
- Total capacity of the hierarchical storage: 8.6 PB
- Data IO rate: 300 Gb/s
- GPU component based on NVIDIA Tesla V100&A100
- CPU component based on RSC “Tornado” liquid cooling solutions
- The most energy-efficient center in Russia (PUE = 1.06)

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.



> 300 user papers (two in Nature Physics)

Unified Scalable Supercomputer Research Infrastructure

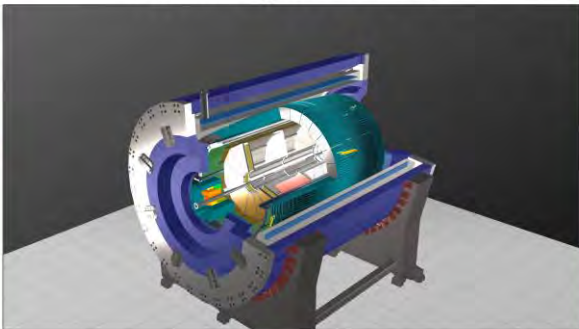


Based on the integration of the supercomputers of JINR, of the Interdepartmental Supercomputer Center of the Russian Academy of Sciences and of Peter the Great St. Petersburg Polytechnic University, a **unified scalable supercomputer research infrastructure** based on the National Research Computer Network of Russia (NIKS) was created. Such an infrastructure is in demand for the tasks of the NICA megaproject.

ДААННЫЕ



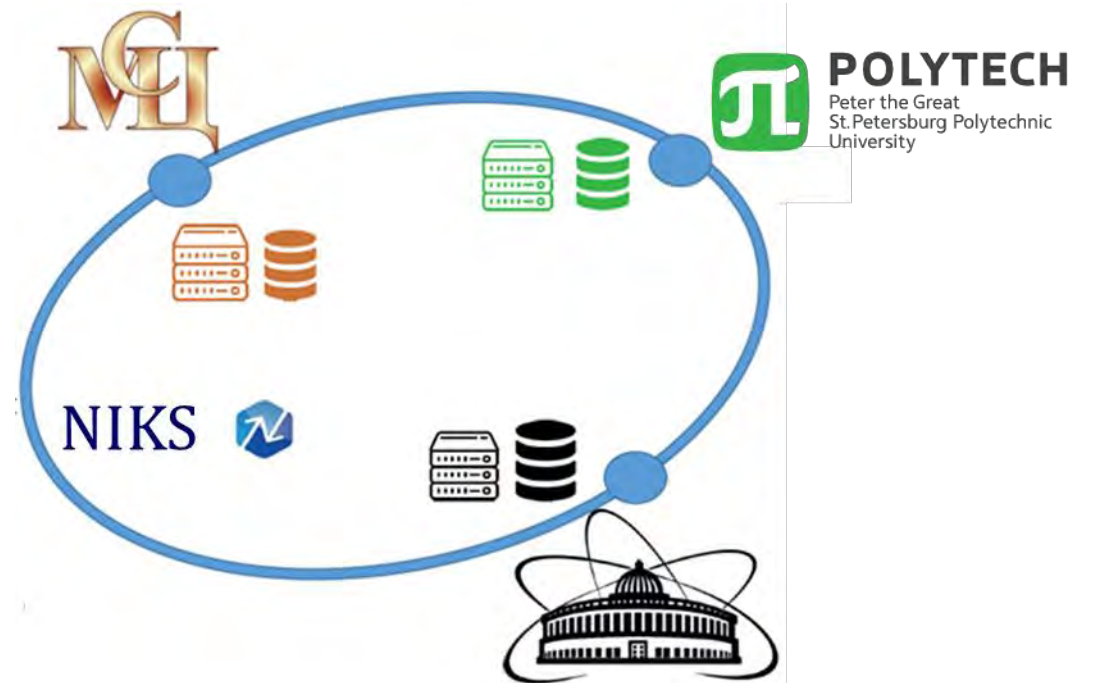
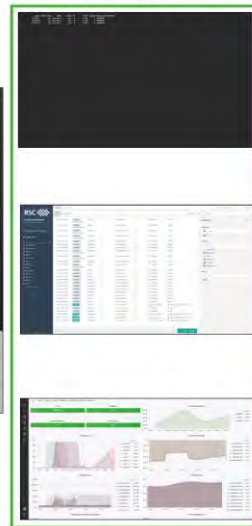
Центр управления виртуальным экспериментом Multi-Purpose Detector



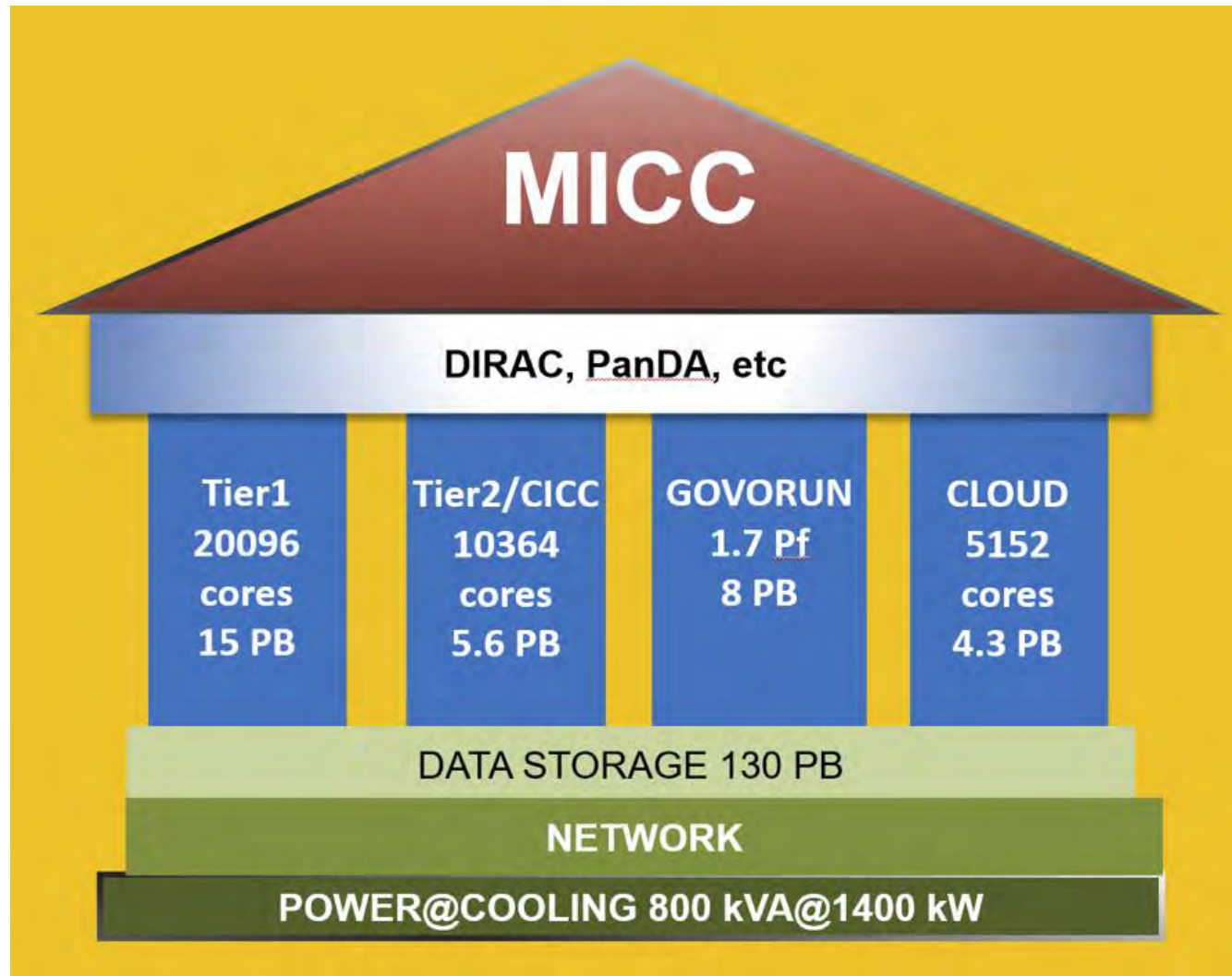
00:00:44:19



ЗАДАЧИ



Multifunctional Information and Computing Complex (MICC)



4 advanced software and hardware components

- Tier1 grid site
- Tier2 grid site
- hyperconverged “Govorun” supercomputer
- cloud infrastructure

Distributed multi-layer data storage system

- Disks
- Robotized tape library

Engineering infrastructure

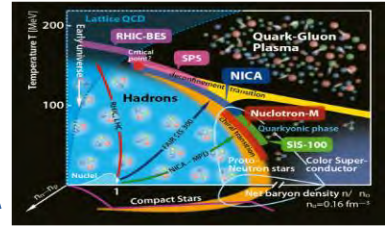
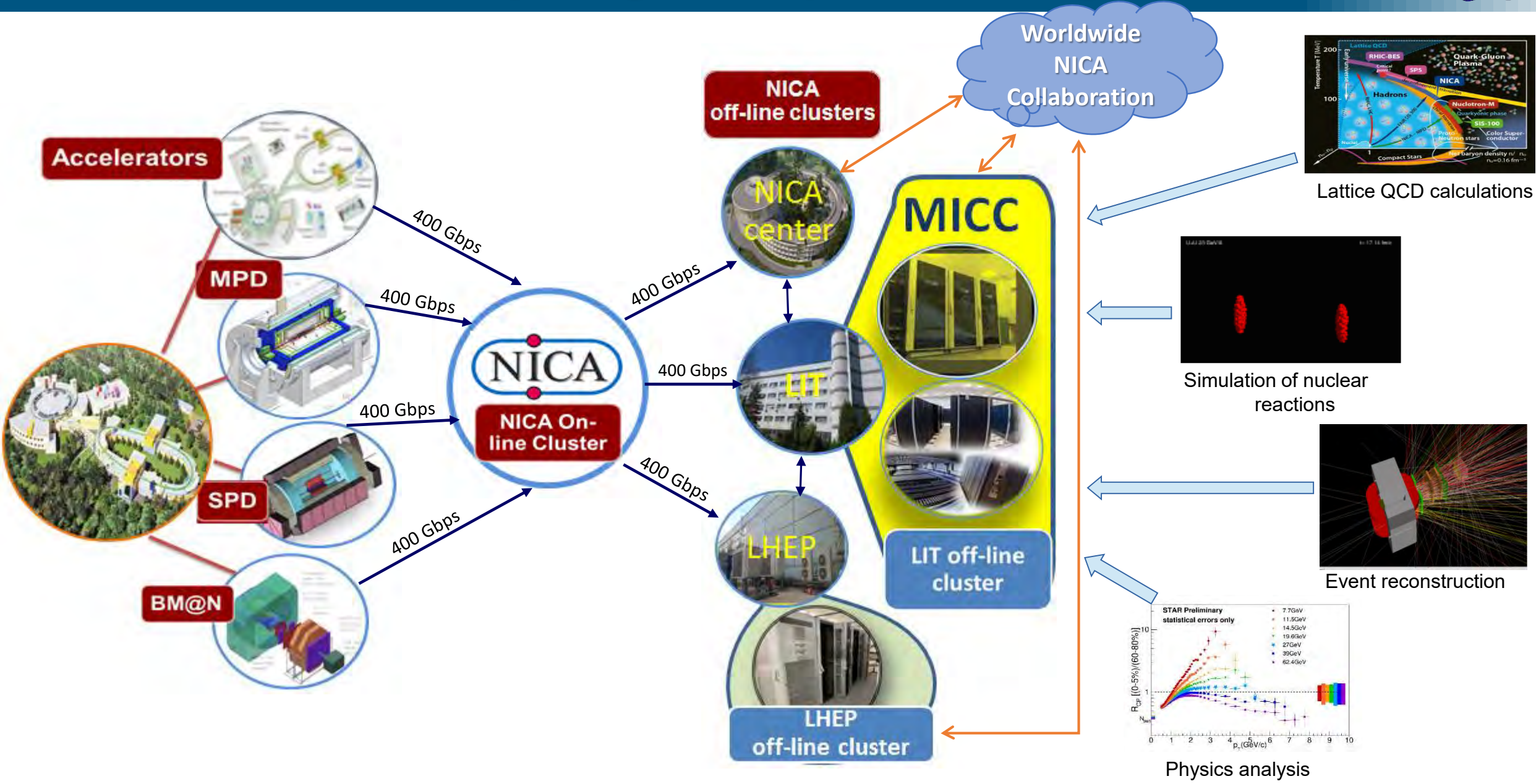
- Power
- Cooling

Network

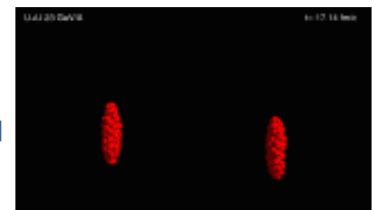
- Wide Area Network
- Local Area Network

The main objective of the project is to ensure multifunctionality, scalability, high performance, reliability and availability in 24x7x365 mode for different user groups that carry out scientific studies within the JINR Topical Plan

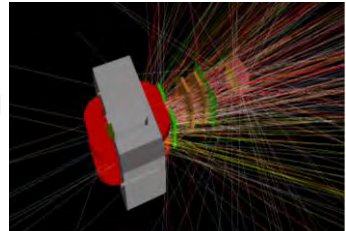
NICA Computing Concept & Challenges



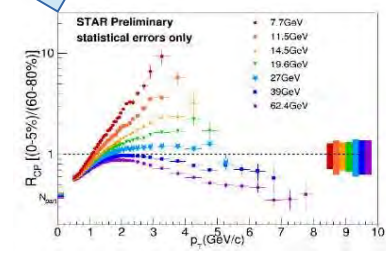
Lattice QCD calculations



Simulation of nuclear reactions

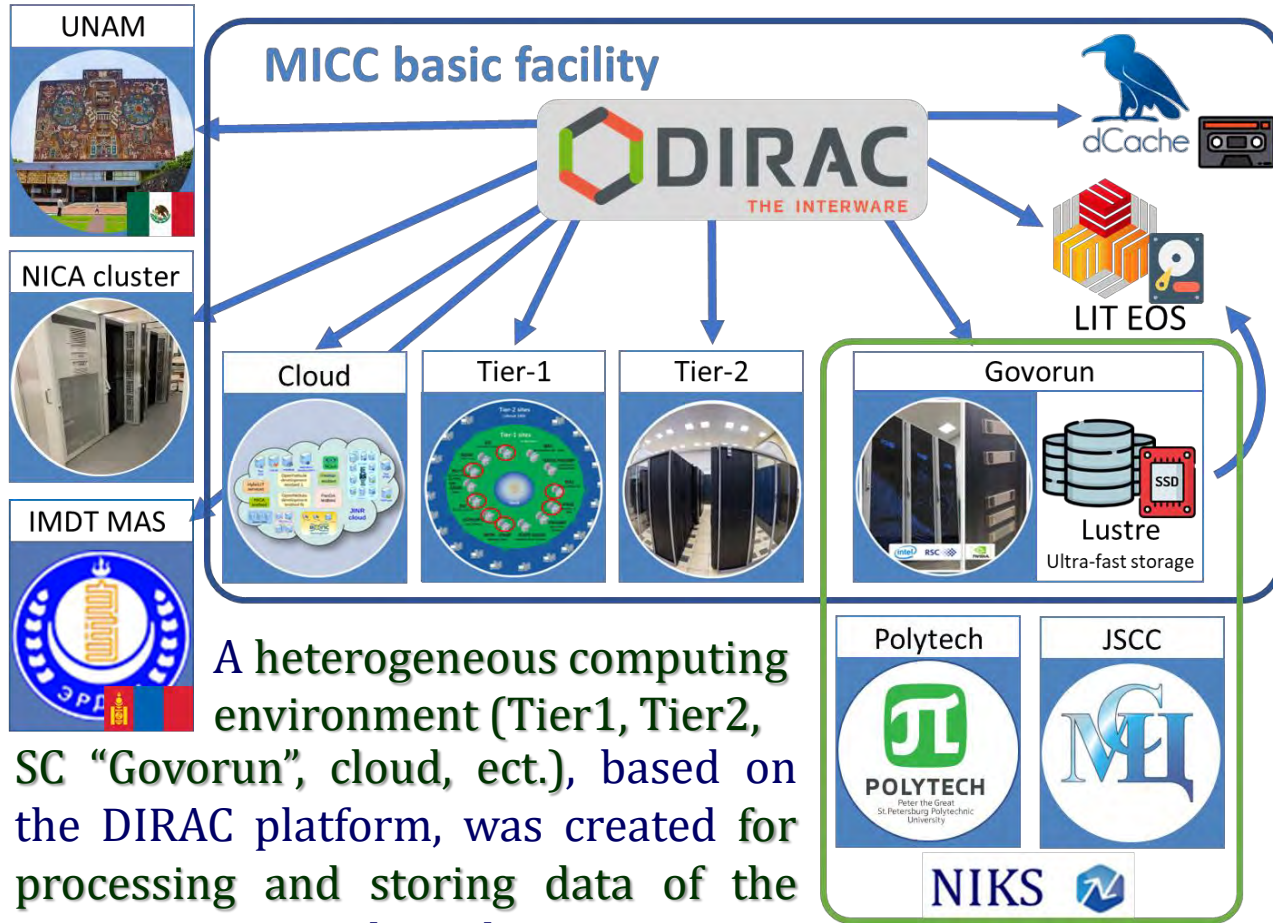


Event reconstruction



Physics analysis

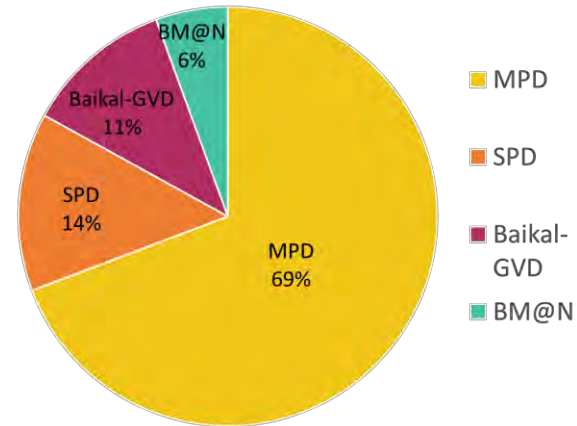
DIRAC-based distributed heterogeneous environment



A heterogeneous computing environment (Tier1, Tier2, SC "Govorun", cloud, ect.), based on the DIRAC platform, was created for processing and storing data of the experiments conducted at JINR.

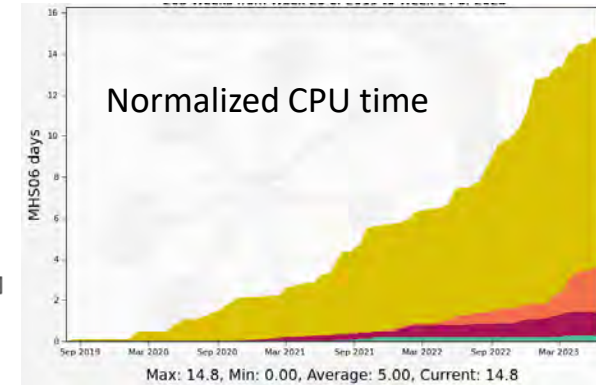
The distributed infrastructure is used by the MPD, Baikal-GVD, BM@N, SPD.

Use of DIRAC platform by experiments in 2019-2022

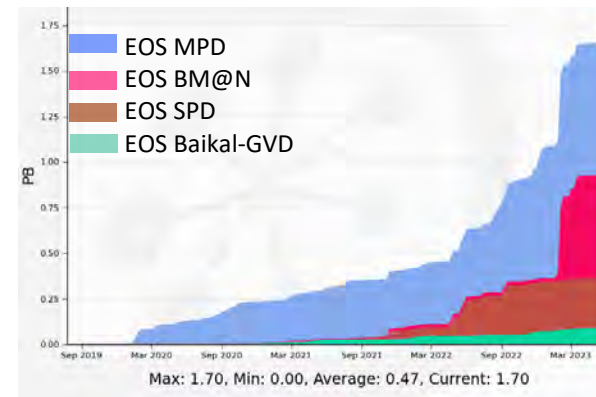


Total Number of executed jobs

The major user of the distributed platform is the MPD experiment



Data processed by experiments

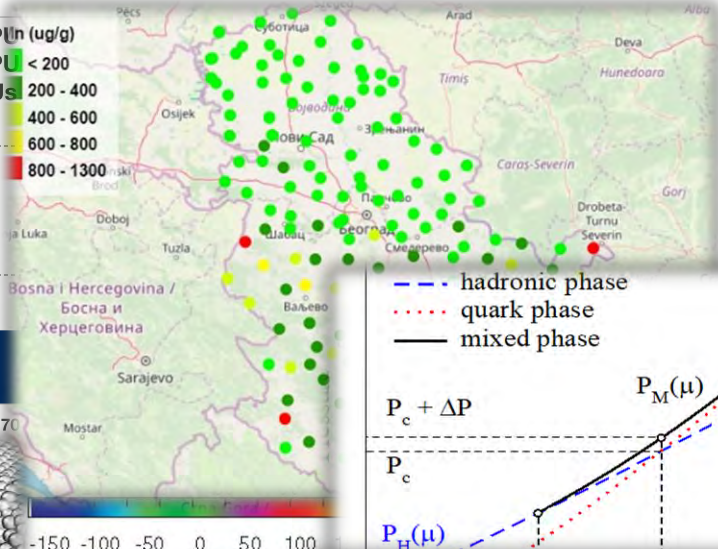
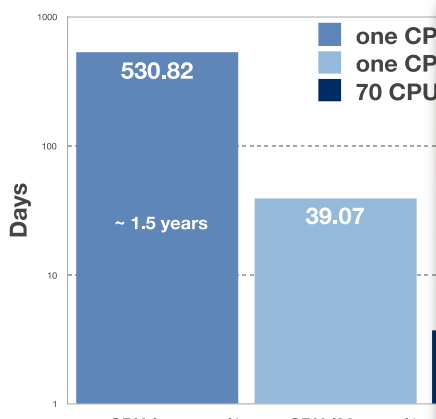


Summary statistics of using the DIRAC platform for MPD tasks in 2019-2022

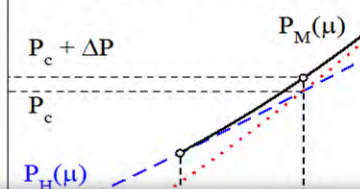
<p>31 data generation campaigns</p>	<p>1.283 B events generated</p>	<p>440M events reconstructed</p>	<p>1.5M jobs completed</p>	<p>1557 years total computation time</p>	<p>1.3 PB MPD data produced</p>
-------------------------------------	---------------------------------	----------------------------------	----------------------------	------------------------------------------	---------------------------------

Methods, Algorithms and Software

Govorun Supercomputer



- - - hadronic phase
 ····· quark phase
 ——— mixed phase



Numerical modeling of complex physical systems



Experimental data processing and analysis



Big Data



Machine and Deep learning



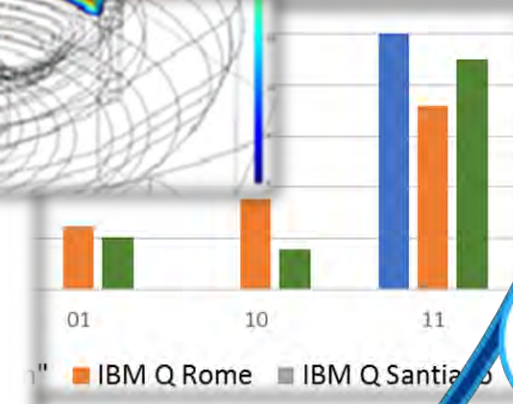
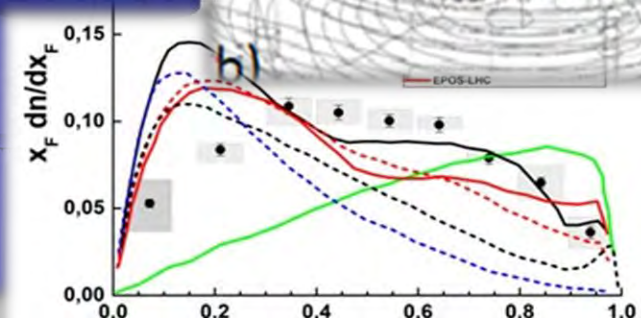
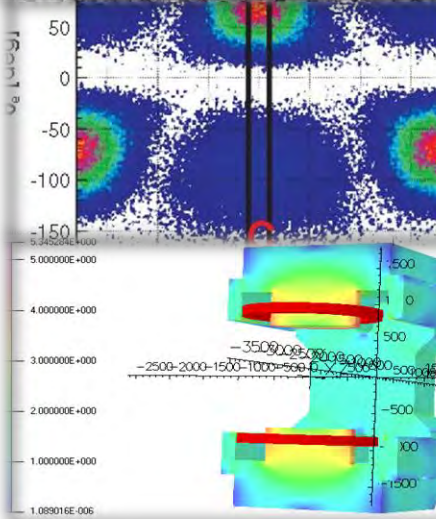
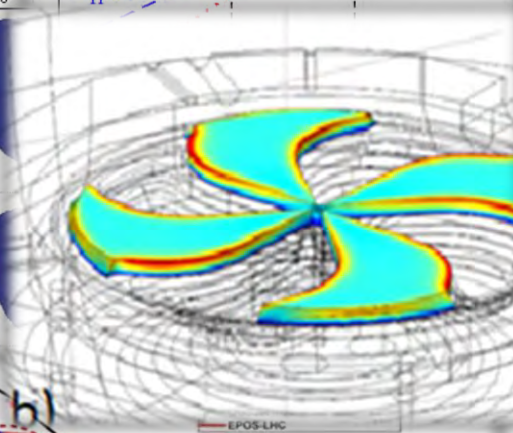
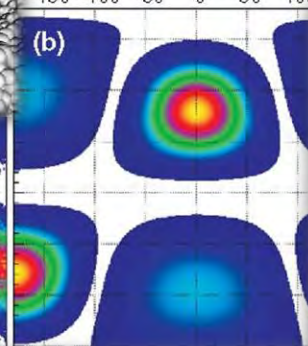
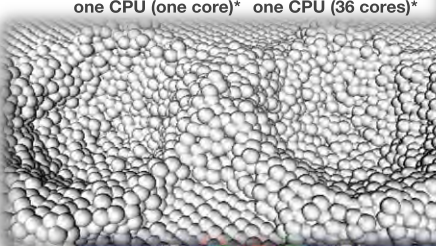
AI and robotics



Computer algebra



Quantum computing



Activity: Digital ecosystem (Digital JINR)



The digital platform “**JINR Digital EcoSystem**” integrates existing and future services

to support

scientific,
administrative and social activities,
maintenance of the engineering and IT infrastructures

to provide

reliable and secure access to various types of data

to enable

a comprehensive analysis of information

using

modern Big Data technologies and artificial intelligence.

JINR
Digital Eco System



Other services



Information services



Network services



Administrative services



Scientific services



Single access point to all services

Digital technologies

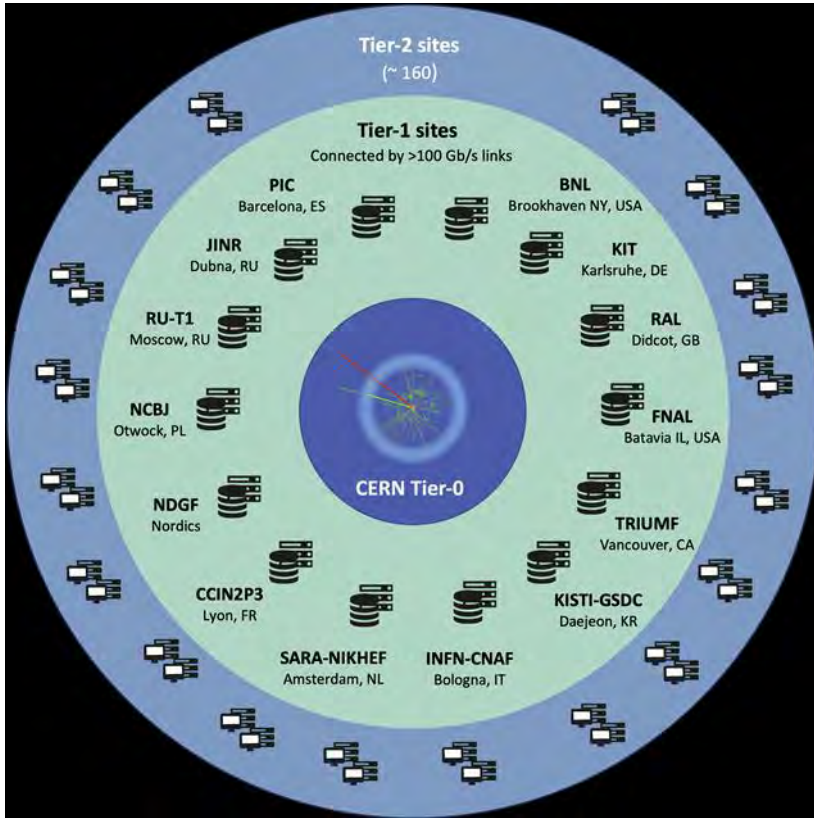
Digital infrastructures

IT specialists and users

The Worldwide LHC Computing Grid



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



The mission of the WLCG project is to provide global computing resources to store, distribute and analyze the **~250-300 Petabytes** of data expected every year of operations from the Large Hadron Collider.

WLCG computing enabled physicists to announce the discovery of the Higgs Boson.

170 sites

42 countries

> 12k physicists

~1.6 M CPU cores

~2 EB of storage (1 EB - CERN)

> 2.5 million jobs/day

100-400 Gb/s links



Tier0 (CERN):
data recording,
reconstruction
and distribution

Tier1:
permanent
storage,
re-processing,
analysis

Tier2:
Simulation,
end-user
analysis

Проекты класса мегасайенс



Российские исследовательские институты и университеты активно участвуют в международных мегапроектах: LHC, CERN (эксперименты: ATLAS, ALICE, LHCb, CMS), XFEL, DESY (европейский лазер на свободных электронах), ESRF, France (европейский синхротронный центр), FAIR, GSI, Germany (эксперименты CBM, PANDA), ITER, France ...

В России идет подготовка проектов класса мегасайенс:

НИКА, ОИЯИ, Дубна (Коллайдер протонов и тяжелых ионов), ПИК, НИЦ ПИЯФ, Гатчина (Высокопоточный реакторный комплекс), СКИФ, ИЯФ СО РАН Новосибирск (Сибирский кольцевой источник фотонов) СИЛА, НИЦ КИ ИФВЭ, Протвино, Супер С-тау фабрика, (Электрон-позитронный коллайдер), Саров Реализуется нейтринная программа: проекты в России (Байкал,), Китае (JUNO), США (NOVA, DUNE), синхротронно-нейтронная программа

Уникальные экспериментальные установки в ОИЯИ



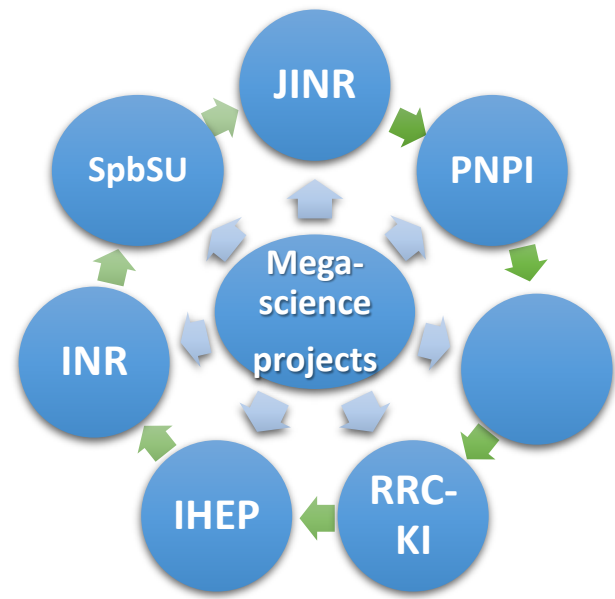
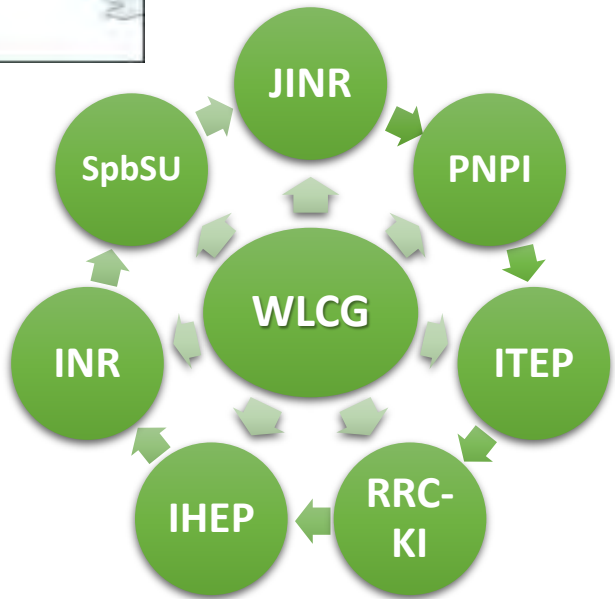
Институт ядерной физики
имени Г. И. Будкера СО РАН



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

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

Consortium RDIG-M – Russian Data Intensive GRID for Megascience projects



Создание консорциума для IT-обеспечения исследовательской инфраструктуры класса «мегасайенс»

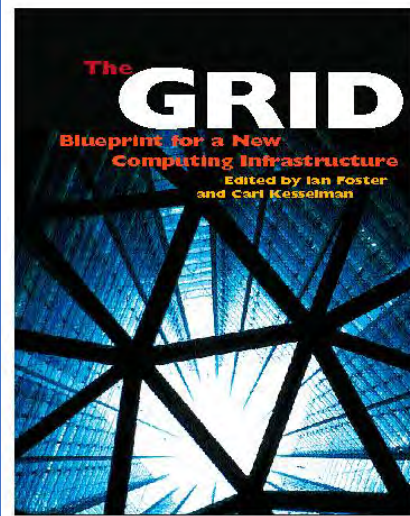


- Консорциум Российский GRID для интенсивных операций с данными (РДИГ) был создан в 2003 году для активного участия в распределенной обработке данных экспериментов на Большом адронной коллайдере LHC в рамках научной коллаборации LHC WLCG (Worldwide LHC Computing Grid). Созданная инфраструктура RDIG имеет огромное значение для эффективного участия ученых России в научной программе экспериментов на LHC.
- В России реализуется программа масштабных научных проектов, важнейшей частью которых является развитие распределенных гетерогенных компьютерных систем (включая системы с экстремально параллелизмом) для обработки, хранения, анализа экспериментальных данных, разработка и внедрение эффективных методов, алгоритмов и программного обеспечения для моделирования физических систем, математической обработки и анализа экспериментальных данных, развитие методов машинного обучения, искусственного интеллекта, квантовых вычислений.
- Для решения этой масштабной задачи необходимо развивать распределенную компьютерную инфраструктуру, объединяющую ключевые научные и образовательные институты, участвующие в проектах мегасайенс - РДИГ-М. Созданный консорциум на базе ОИЯИ, НИЦ Курчатовский институт, ИСП РАН должен стать ядром для IT-обеспечения исследовательской инфраструктуры класса «мегасайенс».

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

Grids

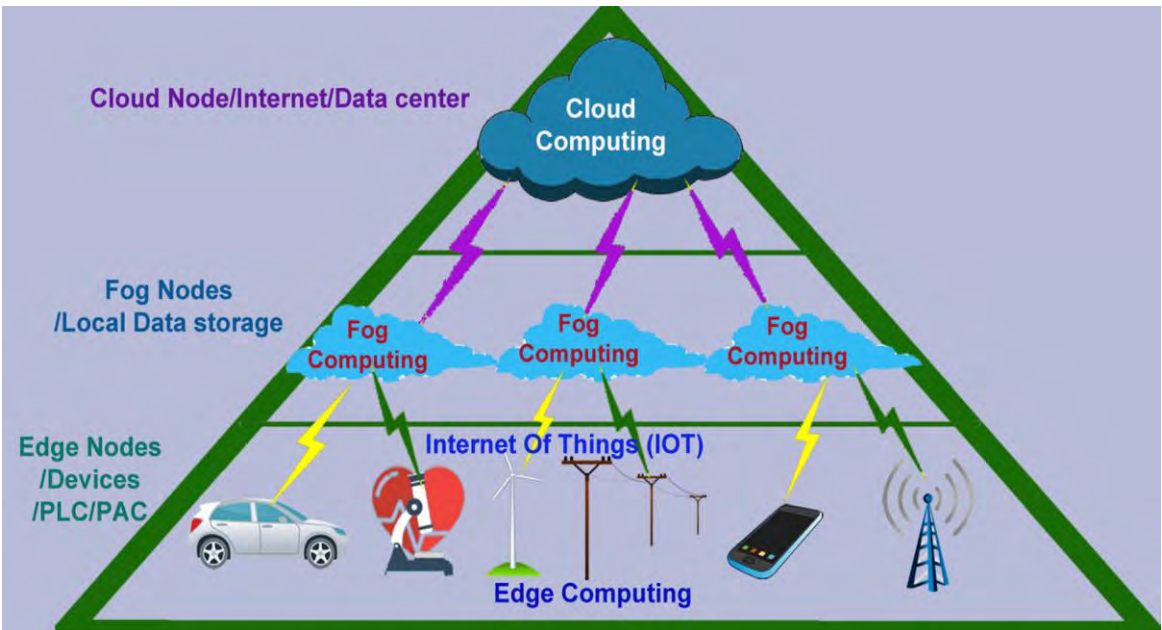
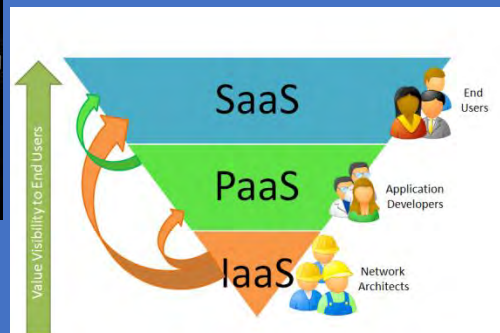
- Collaborative environment
- Distributed resources



Supercomputers



Clouds



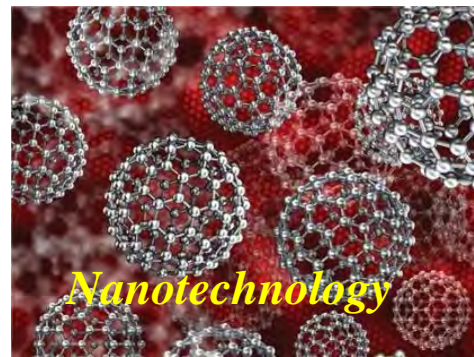
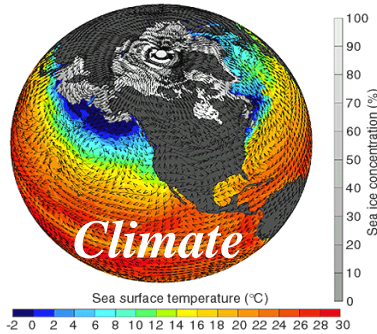
HPC+Big Data+Artificial intelligence



High Energy Physics



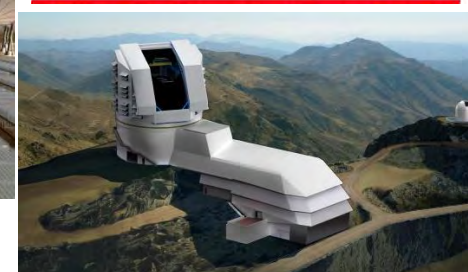
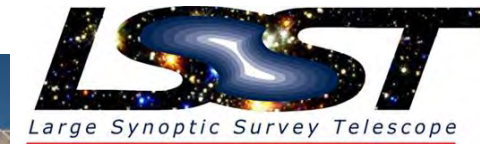
CERN Large Hadron Collider > 600 Pb/Year



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



International radiotelescope for the 21st century



Large Synoptic Survey Telescope > 10 Pb/Year (estimation)



... et cetera

TOP500 List – June 2024

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1.	Frontier - HPE Cray DOE/SC/Oak Ridge National Laboratory US	8,699,904	1,206.00	1,714.81	22,786
2	Aurora - HPE Cray DOE/SC/Argonne National Laboratory US	9,264,128	1,012.00	1,980.01	38,698
3	Eagle - Microsoft, NVIDIA H100, NVIDIA US	2,073,600	561.20	846.84	
4	Fugaku -RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
5	LUMI - HPE Cray Finland	2,752,704	379.70	531.51	7,107
6	Alps - HPE Swiss National Supercomputing Centre (CSCS) Swiss	1,305,600	270.00	353.75	5,194
7	Leonardo – Bull Sequana Italy	1,824,768	241.20	306.31	7,494
8	MareNostrum 5 ACC – Bull EuroHPC/BSC Spain	663,040	175.30	249.44	4,159
9	Summit - IBM DOE/SC/Oak Ridge National Laboratory US	2,414,592	148.60	200.79	10,096
10	Eos NVIDIA DGX NVIDIA Corporation US	485,888	121.40	188.65	
42	Chervonenkis –NVIDIA A100 80GB Yandex Russia	193,440	21.53	29.42	

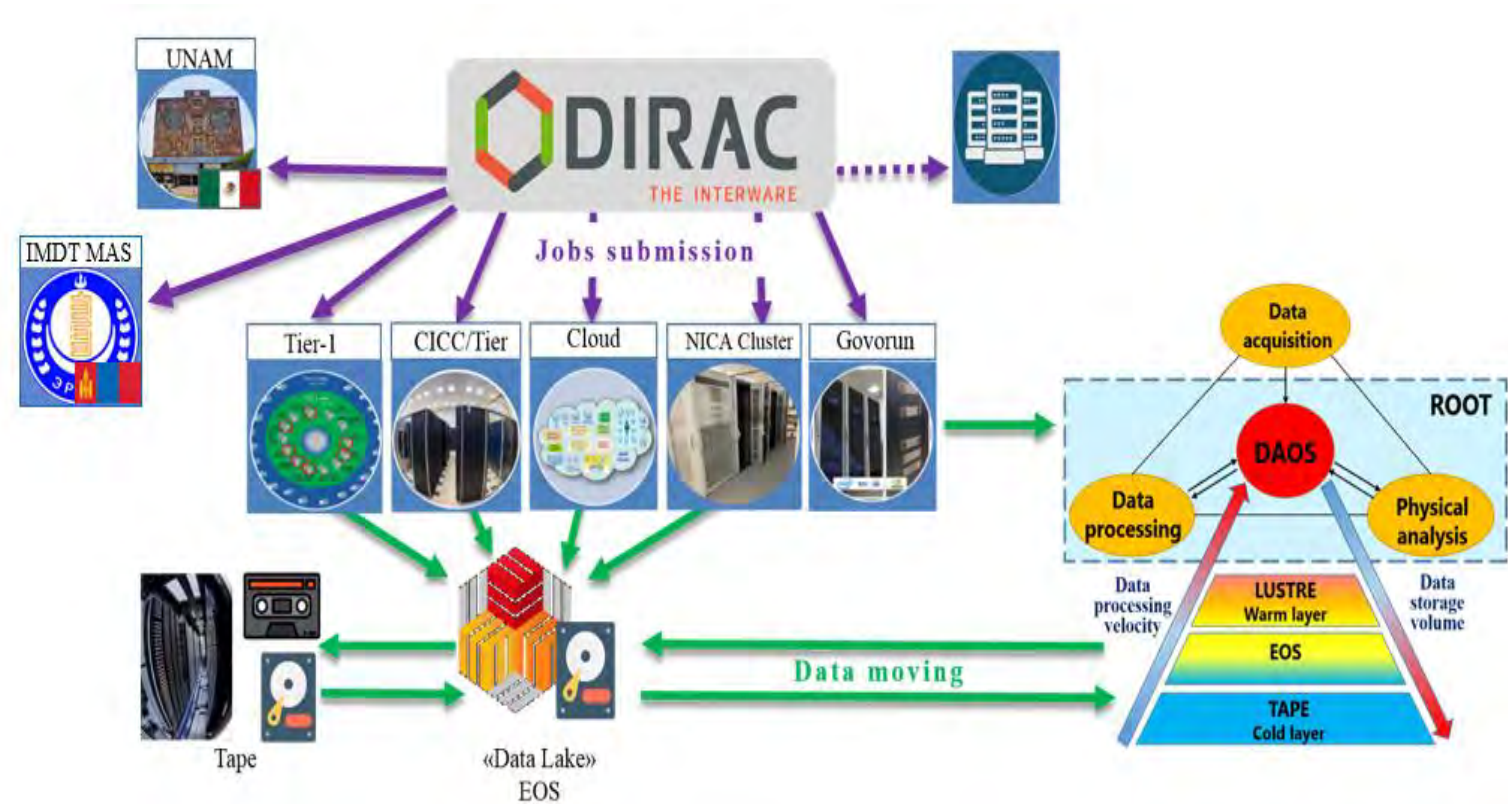


Software for Computing Environment



To ensure uninterrupted operation of MICC and for efficient use of computing resources, core software development is required

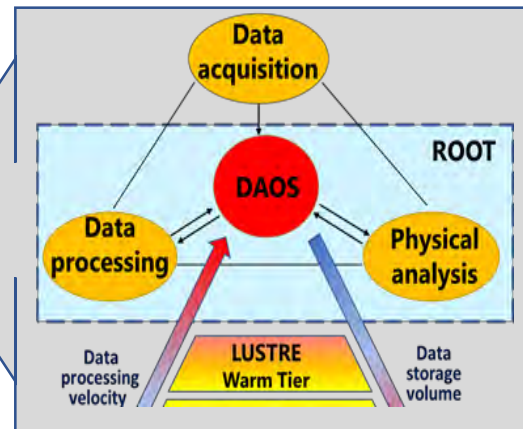
- distributed software defined high-performance computing platform for processing and storing data of experiments, combining supercomputer (heterogeneous), grid and cloud technologies for the efficient use of new computing architectures
- multipurpose software and hardware platform of Big Data analytics based on hybrid hardware accelerators (GPU, FPGA, quantum systems); machine learning algorithms; analytics, reporting and visualization tools; support for user interfaces and tasks
- advanced systems for the protection of cyber infrastructure, computer and user information, public e-services and user authentication



Distributed Multilayered Data Storage System



- Limited data and **short-term** storage – to store OS itself, temporary user files
- AFS distributed global system – to store user home directories and software
- dCache is traditional for MICC grid sites – to large amounts of data (mainly LHC experiments) for **middle-term** period
- EOS is extended to all MICC resources – to store large amounts of data for **middle-term** period. At present, EOS is used for storage by BM@N, MPD, SPD, BaikalGVD, etc.
- Tape robotic systems – to store large amounts of data for **long-term** period. At present for CMS. BM@N, MPD, SPD, JUNO – in progress.



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).

Distributed Storage and scientific data management

EOS Open Storage - Disk Storage at CERN

dCache - Distributed Storage for scientific data

CTA - the CERN Tape Archive

CEPH - Object & Application Storage

CERNBox - Sync & Share platform for collaboration

Filesystems - AFS, DFS, Samba, NFS3/4

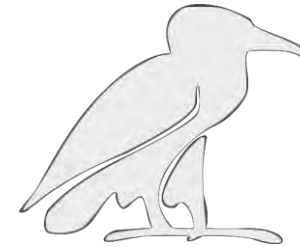
CernVM FS - CernVM File System

FTS & XRootd- File Transfer Service & File Access

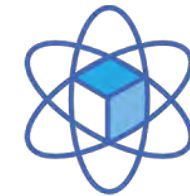
Rucio - Scientific Data Management

DAOS - high-performance storage system

Lustre - open-source, parallel file system



ceph



CERNBox



CERN
Tape Archive

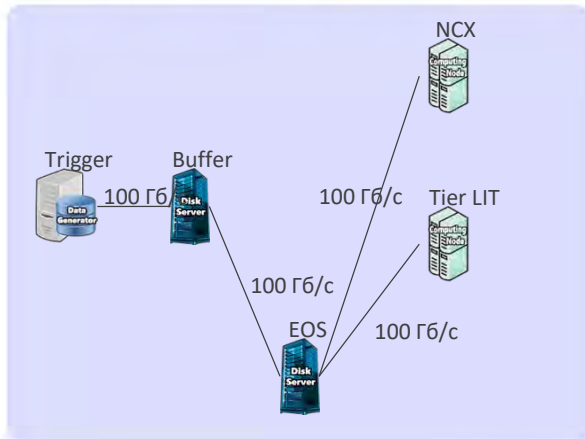
DAOS

l.u.s.t.r.e.[®]

Программный комплекс для создания цифровых двойников распределенных центров сбора, хранения и обработки данных

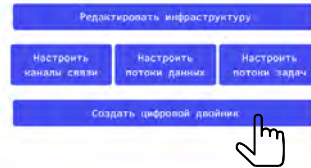
Цифровой двойник РЦОД

Построение инфраструктуры центра сбора, хранения и обработки данных



Примеры применения для экспериментов комплекса NICA

- Построение цифрового двойника вычислительной инфраструктуры эксперимента VM@N.
- Построение цифрового двойника вычислительной системы онлайн-фильтра данных эксперимента SPD.



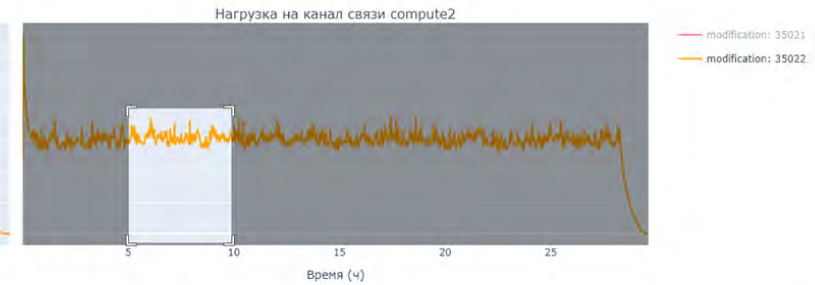
Веб-сервис



Результаты эксперимента Test 1

Выберите вкладку для просмотра результатов

Хранилища данных | Вычислительные компоненты | **Каналы связи** | Очереди задач | Распределения файлов



Особенности

✓ Универсальность.

✓ Учитываются важные функциональные параметры распределенных центров:

- характеристики оборудования;
- характеристики потоков данных и задач;
- вероятности сбоев, отказов и изменений в производительности оборудования и других процессов, происходящих в системе.

✓ Результаты работы отличаются от результатов работы существующего распределенного центра не более, чем на 20%.

Тенденции развития распределенных вычислений для масштабных научных проектов

Развитие компьютерных архитектур и их интеграция:

- суперкомпьютеры масштаба ExaFlop-ZetaFlop, графические, квантовые, фотонные процессоры как специализированные вычислители вокруг универсальной системы
- распределенные иерархические системы хранения информации масштаба ExaByte-ZetaByte, работающие с общими метаданными для защищенного эффективного доступа и надежного хранения информации
- центры обработки данных на базе облачных технологий, обеспечивающие эффективные сервисы для пользователей и поддержку функционирования систем туманных и граничных вычислений
- интеллектуальные системы управления заданиями в распределенной гетерогенной среде, включающей различные компьютерные архитектуры (суперкомпьютеры, грид, облака, кластеры, серверы, системы добровольных вычислений)
- развитие методов, алгоритмов, платформ, систем аналитики Больших данных, интеллектуального анализа данных, ML/DL, искусственного интеллекта, цифровых сервисов
- создание цифровых двойников экспериментальных установок, распределенных компьютерных комплексов и других сложных объектов
- подготовка высококвалифицированных специалистов в области распределенных вычислений, хранения, обработки, анализа данных для научных проектов

Development of the system for training and retraining IT specialists

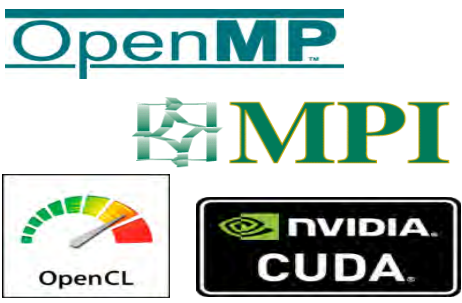


Training courses, master classes and lectures

MLIT staff and leading scientists from JINR and its Member States

Leading manufacturers of modern computing architectures and software

Parallel programming technologies



Tools for debugging and profiling parallel applications



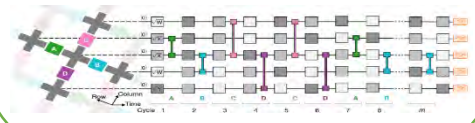
Work with applied software packages



Frameworks and tools for ML/DL tasks



Quantum algorithms, quantum programming and quantum control





Общие отцы-основатели ВМК МГУ и ЛВТА ОИЯИ (ныне ЛИТ им. М.Г. Мещерякова) внесли огромный вклад в развитие программного обеспечения для БЭСМ-6, численных методов, вычислительной физики и компьютеринга. И сегодня продолжает развиваться сотрудничество между МГУ и ЛИТ в области:

- математических методов и математического моделирования,
- грид-технологий в России,
- аналитики Больших данных,
- суперкомпьютерных технологий и методов параллельных вычислений,
- организации и проведения конференций и школ.



Академик Самарский Александр Андреевич



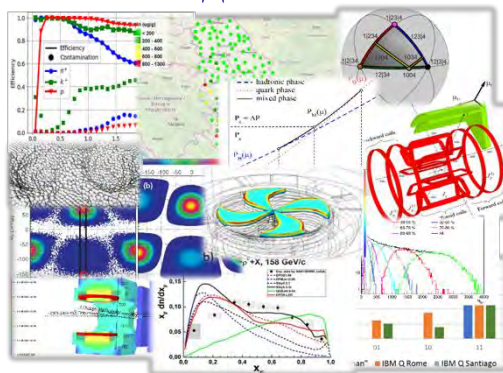
Академик Тихонов Андрей Николаевич

Член-корр. Говорун Николай Николаевич

Потребность в подготовке высококвалифицированных кадров в области математического моделирования и обработки данных проектов класса мегасайнс с применением методов аналитики Больших данных и искусственного интеллекта привела к идее создания на базе филиала МГУ в Дубне направления подготовки «Прикладная математика и информатика»

Магистерской программы

«Методы и технологии обработки данных в гетерогенных вычислительных средах»



Математическое моделирование, численные методы и комплексы программ



Глубокое машинное обучение и аналитика больших данных



Компьютинг (программные средства и модели) для проектов класса мегасайнс

Магистерская программа подготовлена, прошла экспертизу и принята Ученым Советом МГУ

10

GRID2023
3-7 July 2023

10th International Conference
“Distributed Computing and Grid Technologies in
Science and Education”



More than **275** participants

In person - 216

Remotely - 60

30 Plenary reports

135 Sessional reports

17 Countries: Azerbaijan, Armenia, Belarus, Bulgaria, the Czech Republic, Egypt, Germany, Georgia, Iran, Kazakhstan, Mexico, Moldova, Mongolia, Serbia, CERN and Uzbekistan. **Russia** was represented by participants **from 41 universities and research centers.**

Conference Topics:

1. Distributed Computing Systems
2. HPC
3. Distributed Computing and HPC Application
4. Cloud Technologies
5. Computing for MegaScience Projects
6. Quantum Informatics and Computing
7. Big Data, M/D Learning, Artificial Intelligence
8. Student session

Workshop “Computing for radiobiology and medicine”

Workshop “Modern approaches to the modeling of research reactors, creation of the “digital twins” of complex systems”

Round table “RDIG-M - Russian distributed infrastructure for large-scale scientific projects in Russia”

Round table on IT technologies in education



Joint Institute for Nuclear Research
Meshcheryakov Laboratory of Information Technologies

10

GRID2023

3-7 July 2023



10th International Conference
"Distributed Computing and Grid Technologies in
Science and Education"

Social events

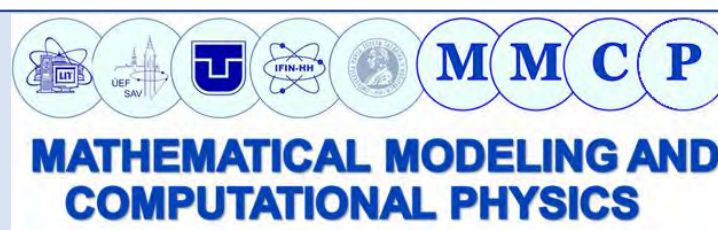


500 shashlik skewers



10th International Conference "Distributed Computing and Grid Technologies in Science and Education" (GRID'2023)

- Distributed Computing Systems
- HPC
- Cloud Technologies
- Distributed Storage Systems
- Distributed Computing in education, industry and science
- Computing for MegaScience
- Quantum informatics and quantum computing
- Big Data, Machine Learning



- ❑ methods, software and program packages for data processing and analysis;
- ❑ mathematical methods and tools for modeling complex physical and technical systems, computational biochemistry and bioinformatics;
- ❑ methods of computer algebra, quantum computing and quantum information processing;
- ❑ machine learning and big data analytics; algorithms for parallel and hybrid computations.

MMCP2024 will be held @ Yerevan 21-25 October 2024
visit <https://mmcp.jinr.ru/> site for details

