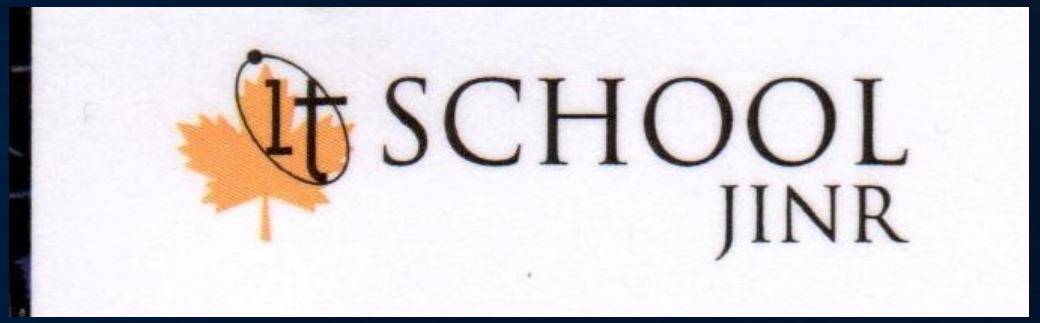


Статус и перспективы Многофункционального информационно-вычислительного комплекса ОИЯИ

Т.А. Стриж
Зам. научного руководителя ЛИТ ОИЯИ

EVOLUTION OF SCIENCE PARADIGMS

1st paradigm: Empirical science

- Descriptive

2nd paradigm: Theoretical science

- Models

3rd paradigm: Computational science

- Simulations

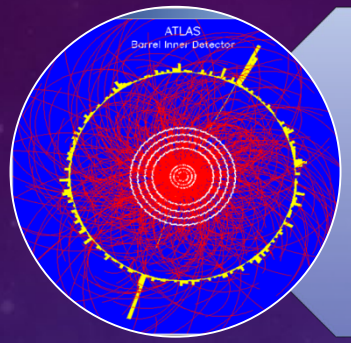
4th paradigm:

Data-intensive exploration (e-Science)

- Unifies the rest to explore large data

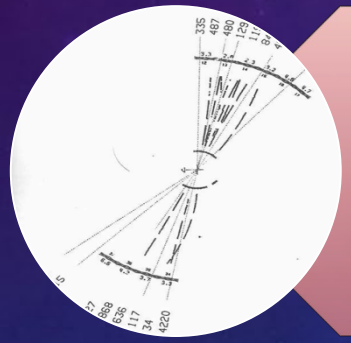
after Jim Gray

История: от малых данных к большим данным (пример физики элементарных частиц)



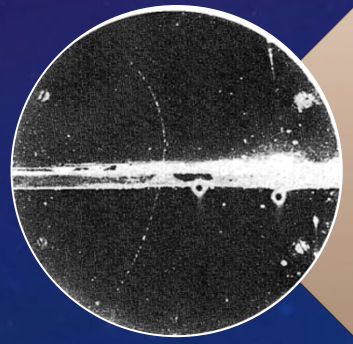
Открытие сегодня

- в основном инклюзивные измерения
- ~3000 ученых в ~150 странах
- сотни Linux серверов, суперкомпьютеры, Grid, Clouds и т.д.



Открытие 70-х

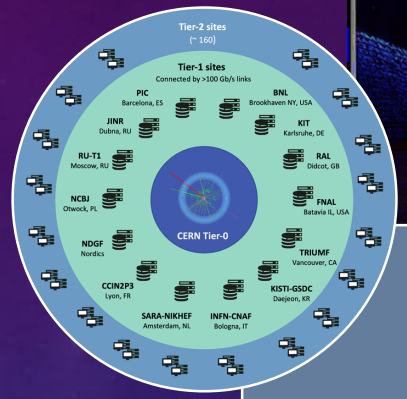
- более инклюзивные измерения
- ~200 ученых в ~10-ти странах
- мейнфреймы



Открытие 1930-х

- **<u>единичные измерения</u>**
- ~2 ученых в 1-ой стране
- ручка и бумага

История ЭВМ в ЛИТ: от средних ЭВМ к GRID, кластерам, облакам и суперкомпьютеру «Говорун»



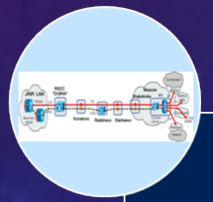
Грид = кластеры + облака +суперкомпьютеры

ПК Фермы, рабочие станции

Мэйнфреймы ЕС 1060, VAX, CONVEX

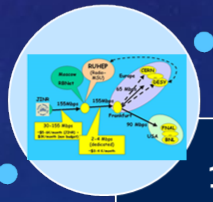
Телекоммуникационные канала и локальная сеть

3x100 Гбит/с

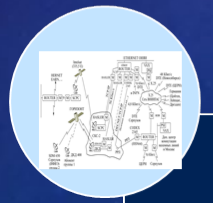


20 Гбит/с

1 Гбит/с

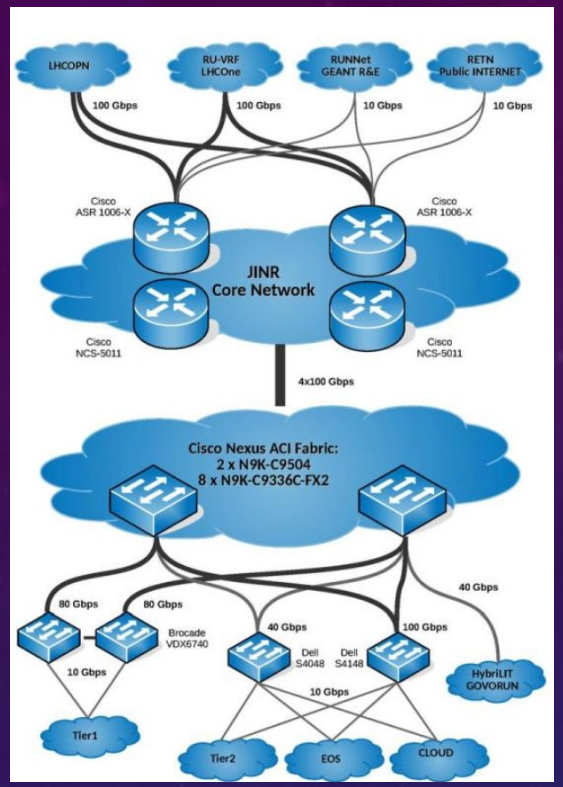


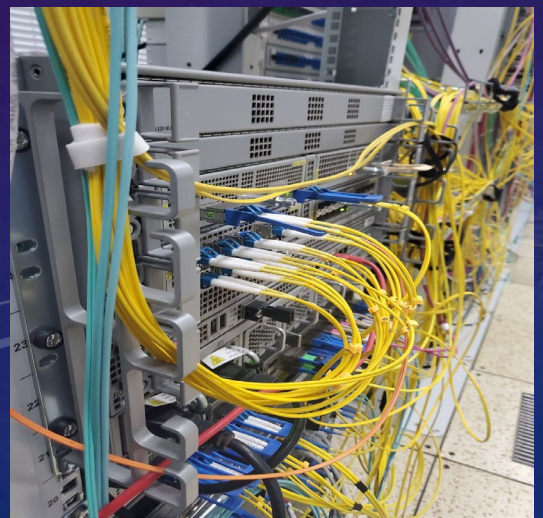
155 Мбит/с



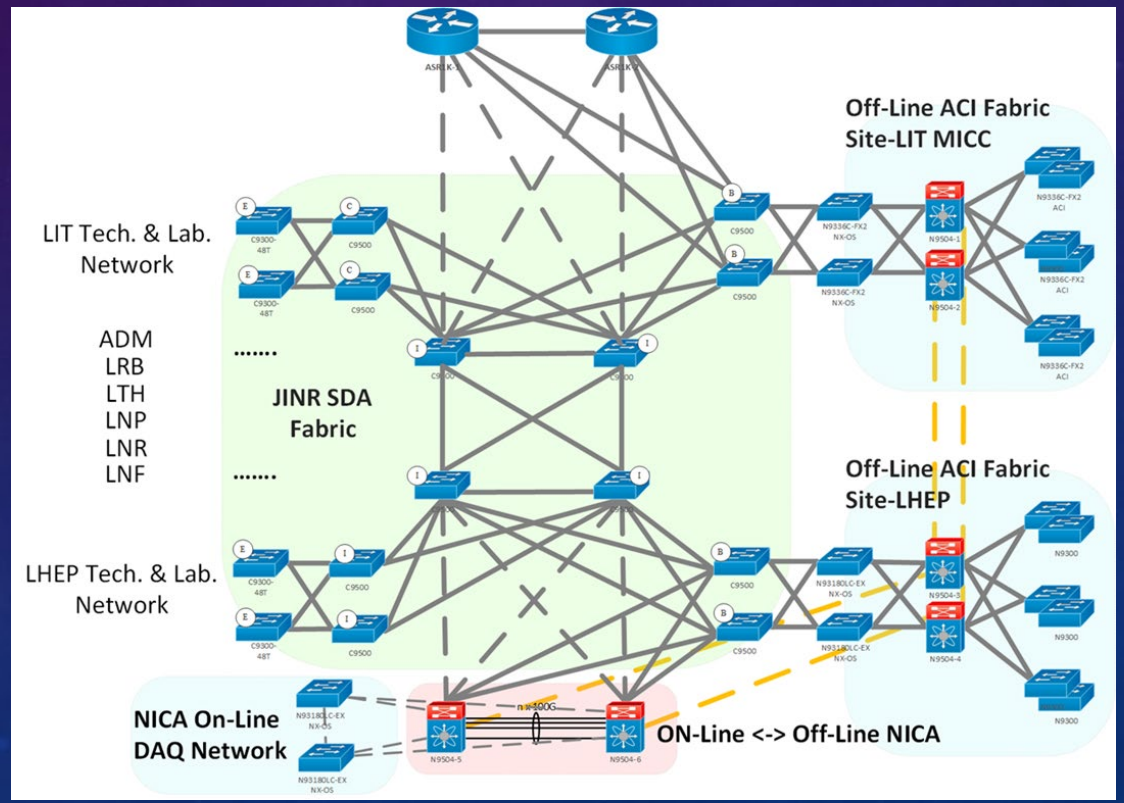
128 Кбит/с

Сетевая инфраструктура





- ОИЯИ- Москва 4x100 Gbit/s
- ОИЯИ - ЦЕРН - 100 Gbit/s и ОИЯИ - Амстердам 100 Gbit/s для сетей LHCOPN, LHCONE, GEANT
- Прямые каналы связи до 100 Gbit/s для связи с RUHEP центрами и сетями Runnet, ReTN
- Мультикластерная сеть 4x100 Gbit/s между ЛФВЭ и ЛИТ

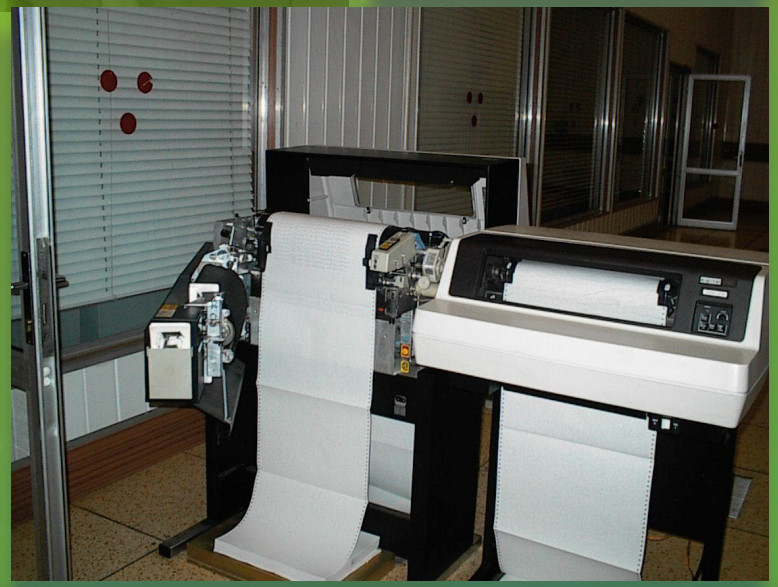


Локальная сеть ОИЯИ:

Пользователей - 5758
из них сотрудников ОИЯИ - 5570
не сотрудников - 188
Сетевых элементов - 13395
IP адресов ipv4 - 22430
IP адресов ipv6 - 1421
Удалённый доступ - 920
Электронные библиотеки - 1163
EDUROAM - 140
Email адреса @jinr.int - 4786
Сетевой трафик в 2023 году

- 41,45 PB - входящий
- 27,28 PB - исходящий

Мэйнфрейм



GRID в ОИЯИ

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
- 2002-2005 – LHC Computing Grid – LCG
  - deploying the results of DataGrid to provide a production facility for LHC experiments
- 2004-2006 – EU EGEE project phase 1
  - starts from the LCG grid
  - shared production infrastructure
  - expanding to other communities and sciences
- 2006-2008 – EU EGEE-II
  - Building on phase 1
  - Expanding applications and communities …

LHC Computing Grid Project (LCG)

The protocol between CERN, Russia and JINR on a participation in LCG Project has been approved in 2003.

The tasks of the Russian institutes in the LCG:

- ✓ **LCG software testing;**
- ✓ **evaluation of new Grid technologies (e.g. Globus toolkit 3) in a context of using in the LCG;**
- ✓ **event generators repository, data base of physical events: support and development;**

Структура комплекса

130 CPU

18TB RAID-5

ATL~ 5 (15) TB

- 6 – Interactive
- 18 – Common PC-farm
- 30 – LHC
- 14 – MYRINET (Parallel)
- 20 – LCG
- 20 – File servers
- 8 – LCG-user interface

The Russian consortium RDIG

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

IHEP - Institute of High Energy Physics (Protvino),
IMPB RAS - Institute of Mathematical Problems in Biology (Pushchino),
ITEP - Institute of Theoretical and Experimental Physics
JINR - Joint Institute for Nuclear Research (Dubna),
KIAM RAS - Keldysh Institute of Applied Mathematics
PNPI - Petersburg Nuclear Physics Institute (Gatchina),
RRC KI - Russian Research Center "Kurchatov Institute"
SINP-MSU - Skobeltsyn Institute of Nuclear Physics, MSU,

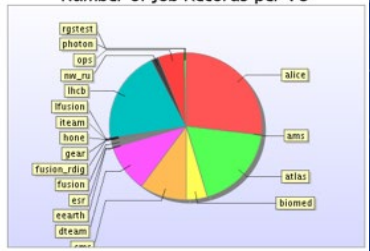
2006

RDIG accounting

Total number of records in Database of RDIG accounting System – 1 384 800

Number of Job Records per Site

Number of Job Records per VO



http://www.egee-rdig.ru

PAH 06.12.2006

Первые шаги LHC – PC фермы

CCIC JINR

130 CPU

17TB RAID-5

10 – Interactive & UI
32 – Common PC-farm
30 – LHC
14 – MYRINET (Parallel)
20 – LCG
24 – servers

3. Creation of a distributed high-performance computing infrastructure and mass storage resources

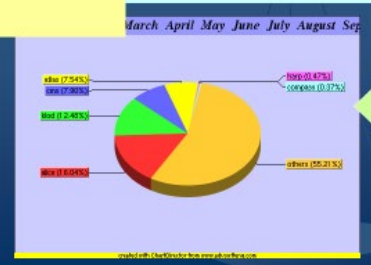
- Development of the JINR CICC as a core of the distributed infrastructure.
- Development of the hard- and software multipurpose infrastructure of the JINR CICC according to the requirements of collaborations and users of JINR and its member states as tabulated:

Year	2005	2006	2007	2010	2015
CPU(kSI2000)	100	660	1000	4000	10000
Disk Space(TB)	50	200	400	800	4000
Tape(TB)	1.5	50	450	1000	6000

Total 501 users
LIT - 171
DLNP - 104
LPP - 53
VBLHE - 44
FLNR - 28
NOJINR - 29
BLTP- 14
FLNP –12
Adm. - 9

Total 17 experiments
ATLAS -44
CMS - 24
ALICE -24
HARP -9
COMPASS – 7
DIRAC -6
DO -3
NEMO -6
OPERA -3

Special groups for ATLAS, CMS, ALICE, LHCb, HARP, COMPASS, DIRAC, D0, NEMO, OPERA, HERMES, H1, NA48, HERA-B, IREN, STAR, KLOD

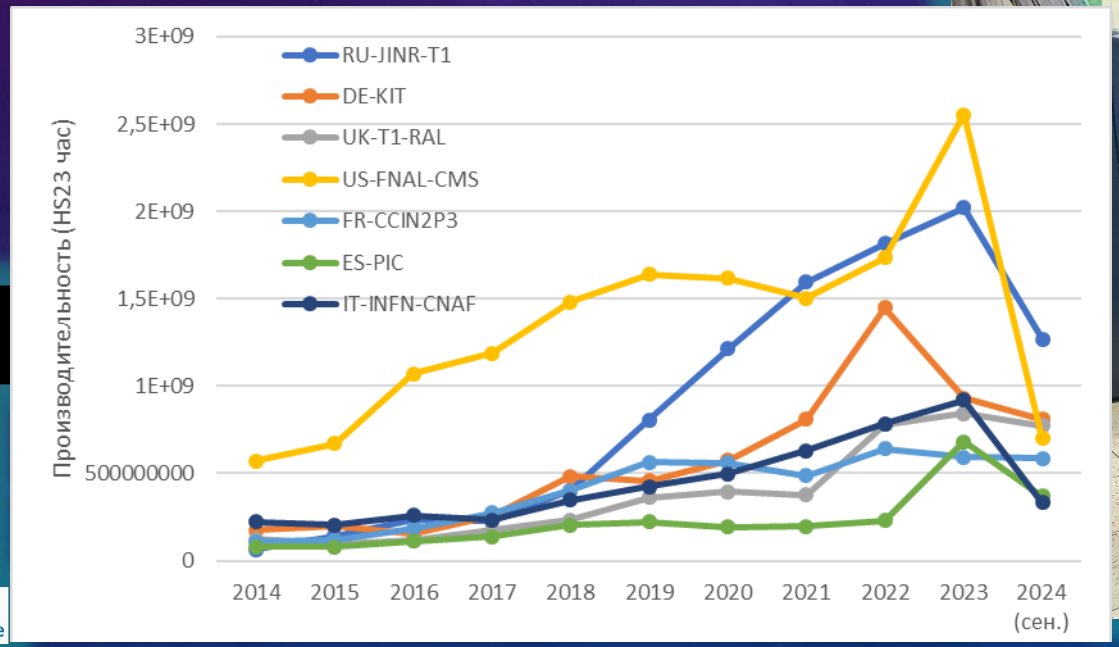


Group statistics (9 months 2005)

Tier2 - RDIG

eGee
Enabling Grids for E-sciencE

2012 –Tier1 для CMS - прототип
1200 ядер, 720 ТВ диски, 72 ТВ ленты



Data GRID

LCG

WLCG
Worldwide LHC Computing Grid

eGee
Enabling Grids for E-sciencE

egi European Grid Infrastructure

THE WORLDWIDE LHC COMPUTING GRID (WLCG)

WLCG – Международное сотрудничество по распространению и анализу данных БАК. Объединяет компьютерные центры по всему миру, предоставляющие вычислительные ресурсы и ресурсы хранения данных, в единую инфраструктуру, доступную всем физикам БАК.

Миссия проекта WLCG — предоставить глобальные вычислительные ресурсы для хранения, распространения и анализа примерно **50-70 петабайт** данных, ожидаемых каждый год от Большого адронного коллайдера.

Компьютерные технологии WLCG позволили физикам объявить об открытии бозона Хиггса (Нобелевская премия 2013 года).

42 страны
170 вычислительных центров
> 1,4 миллиона ядер
2 эксабайта хранилища
> 2 миллионов задач/день
Каналы связи **100-250 Gb/s**

Tier0 (CERN):
запись данных, реконструкция и распределение данных

Tier1:
постоянное хранение, повторная обработка

Tier2:
моделирование, физический анализ

Основные функции грид сайтов

Tier1

- Получение необработанных (RAW) экспериментальных данных от Tier0 в объеме, определенном соглашением WLCG
- Архивирование и ответственное хранение полученных экспериментальных данных.
- Последовательная и непрерывная обработка данных
- Дополнительная обработка (скимминг) данных RAW, RECO (RECOnstructed) и AOD (данные объекта анализа).
- Повторная обработка данных с использованием нового программного обеспечения или новых констант калибровки и юстировки установки CMS.
- Обеспечение доступности наборов данных AOD
- Передача наборов данных RECO и AOD на другие сайты уровней 1/2/3 для их дублированного хранения (репликации) и физического анализа.
- Проведение производственной переработки с использованием нового программного обеспечения и новых калибровочных и юстировочных констант частей установки CMS, защищенное хранение моделируемых событий.
- Получение смоделированных данных и анализ данных, записанных в ходе эксперимента CMS.
- Производство смоделированных данных и их анализ для экспериментов NICA (MPD, BM@N, SPD)

Tier2

- Производство смоделированных данных и анализ данных для всех виртуальных организаций, зарегистрированных в РДИГ и всех экспериментов с участием ОИЯИ, использующих грид.
- Производство смоделированных данных и их анализ для экспериментов NICA (MPD, BM@N, SPD)

Инфраструктура и сервисы

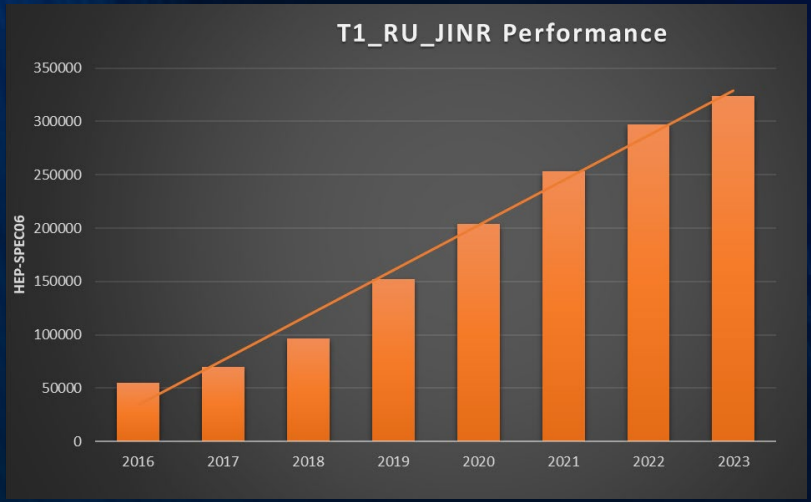
Инфраструктура и сервисы Tier1 (JINR-T1) и Tier2 (JINR-LCG2) обеспечивают работу:

- вычислительного сервиса,
- сервиса хранения данных,
- сервиса доступа к домашним каталогам пользователей,
- сервиса доступа к версиям пользовательского ПО,
- сервисов поддержки грид,
- сервиса передачи данных,
- сервиса управления распределенными вычислительными системами,
- информационных сервисов (мониторинг, информационные сайты).

Общие сервисы для большинства компонент МИВК:

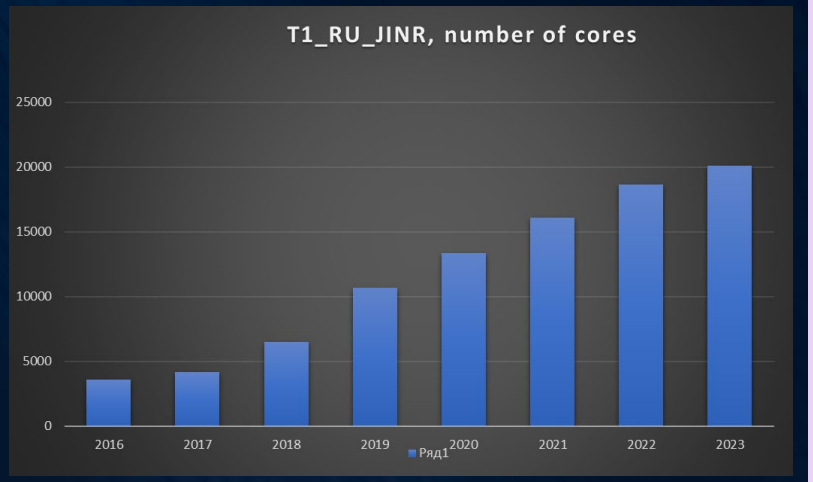
- kerberos, VOMS — аутентификация и авторизация доступа;
- AFS — домашние каталоги пользователей, установка и доступ к пользовательскому и групповому программному обеспечению, доступному по всему миру, как локальная файловая система с доступом POSIX;
- Серверы CVMFS (CernVM-File System) (stratum0/1) — установка и хранение программного обеспечения для совместной работы со многими версиями программного обеспечения, доступными по всему миру, например, локальная файловая система с доступом POSIX.
- Клиенты CVMFS и кеширование — доступ к программному обеспечению для совместной работы (только для чтения), используемому для доступа к локальным CVMFS и глобальным репозиториям со всего мира;
- EOS — хранение и доступ к большим объемем данных, доступных на интерактивных и вычислительных машинах, таких как локальная FS с доступом POSIX, доступ по всему миру через протоколы xroot и http;
- GIT — сервис для сборки и тестирования программного обеспечения для совместной работы с последующей установкой в CVMFS.

Инфраструктура и сервисы(Tier1 2023)



Вычислительный ресурс (CE)

323820.54 HEP-SPEC06, 20096 ядер
Среднее HEP-SPEC06 на ядрое = 16.11
468 машин
CMS (пилоты по 16 ядер):
Max: 20096 ядер
NICA (через DIRAC)
Max: 4000 ядер



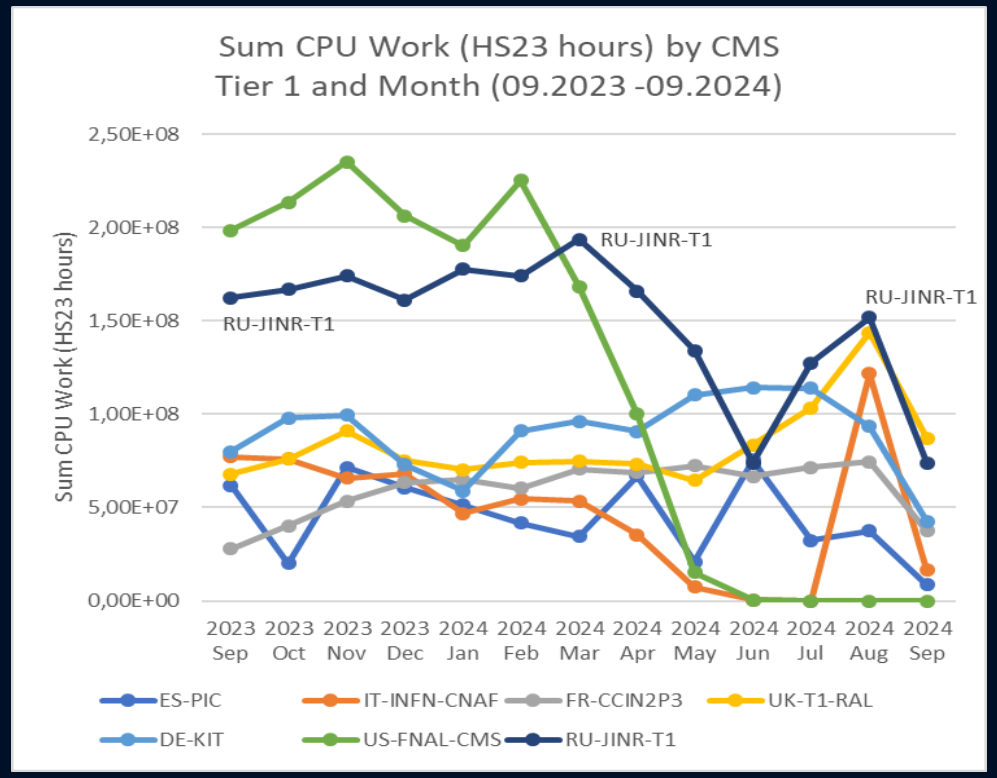
Системы хранения (SE)

dCache: SE disks: 11763.44 PB
CMS @ dcache mss Total: 2642.24 TB
Tapes@Enstore: 35562,00 TB
Ленточные роботы: 51.5PB, IBM TS3500(11.5PB) + IBM T4500(90PB)
EOS: 21829.01 TB
CVMFS
2 squid servers cache CVMFS

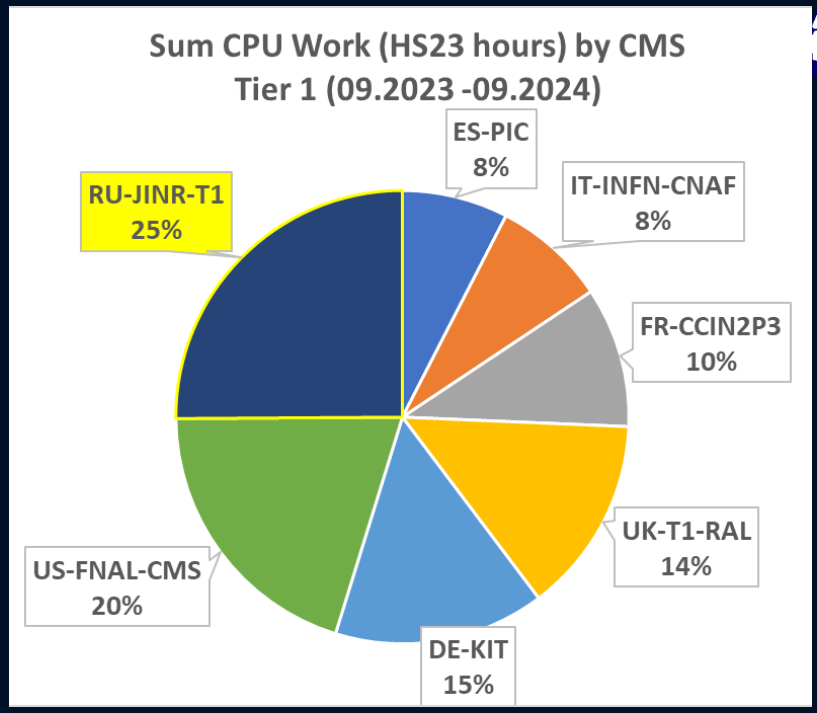
Программное беспечение:

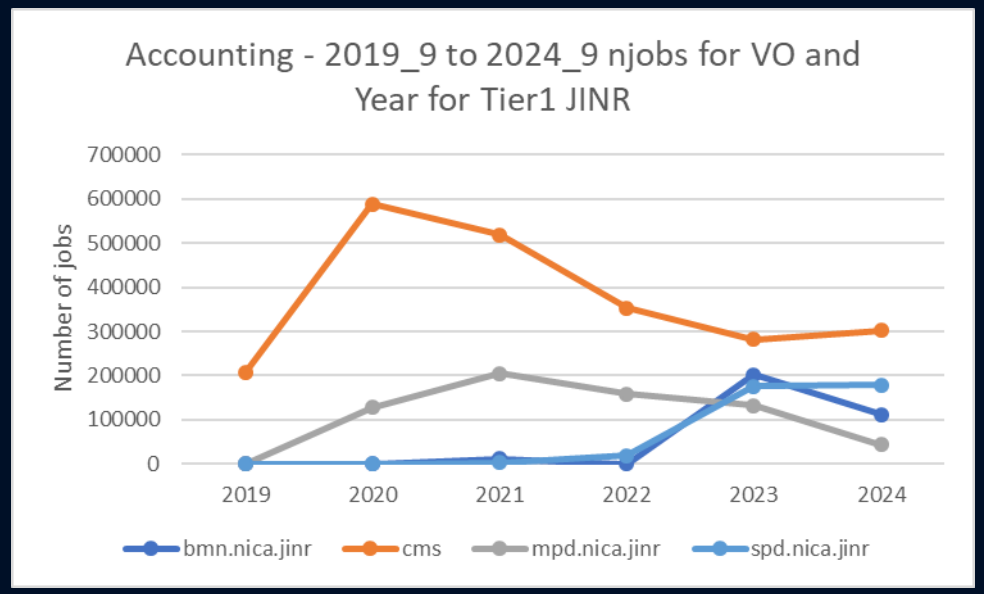
OC: Scientific Linux версия 7.9.
EOS 5.1.23
dCache 8.2,
Enstore 6.3.
Slurm 20.11.
grid UMD4 + EPEL (текущая версия)
ARC-CE
FairSoft
FairRoot
MPDroot

JINR Tier1

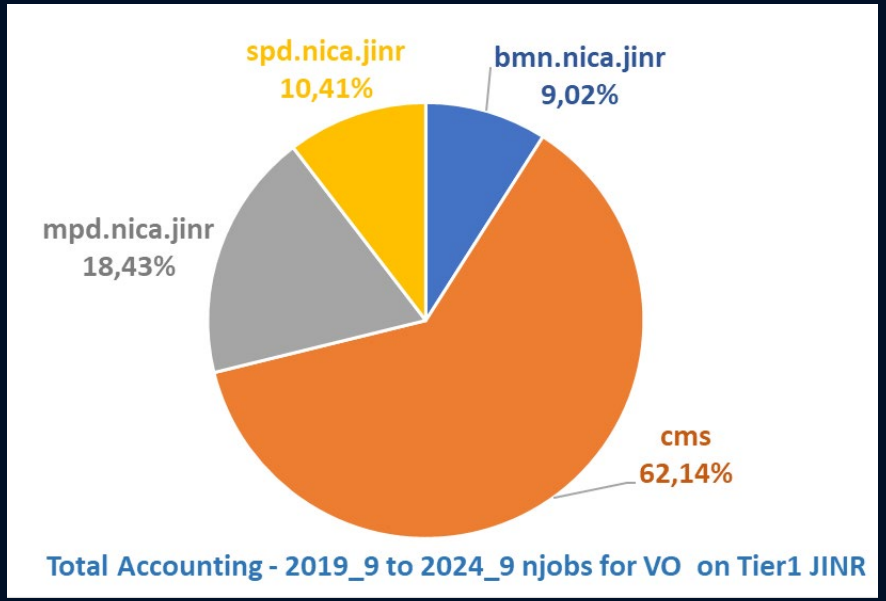


Наш Tier 1 регулярно занимает лидирующее место среди Tier1, обрабатывающих данные эксперимента CMS на БАК.

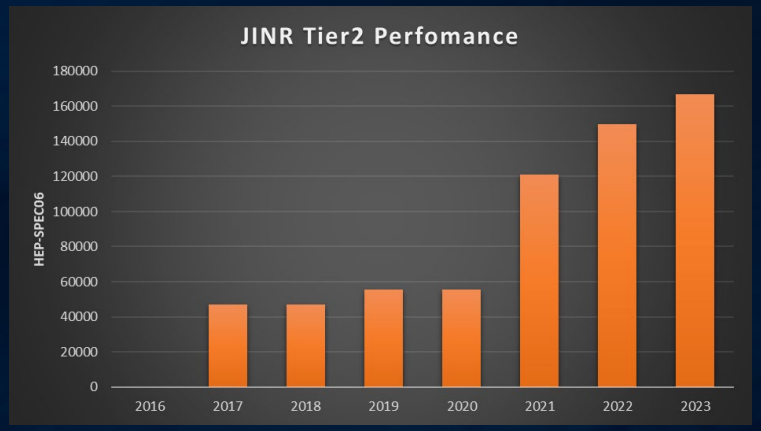


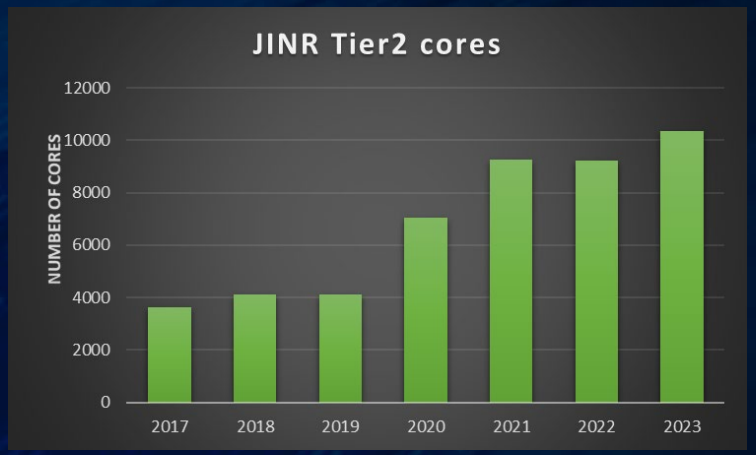


С 2019 года Tier1 ОИЯИ демонстрирует стабильную работу не только для CMS (БАК), но и для экспериментов NICA.



Инфраструктура и сервисы (Tier2 2023)





Вычислительные ресурсы(CE):

Интерактивный кластер: lxpub [01-05] .jinr.ru
Интерфейс пользователей lxui [01-04] .jinr.ru (шлюз для внешнего соединения)
Вычислительный кластер.

485 машин
10356 ядер
166788.4 HEP-SPEC06
16.11 HEP-SPEC06 среднее на ядро

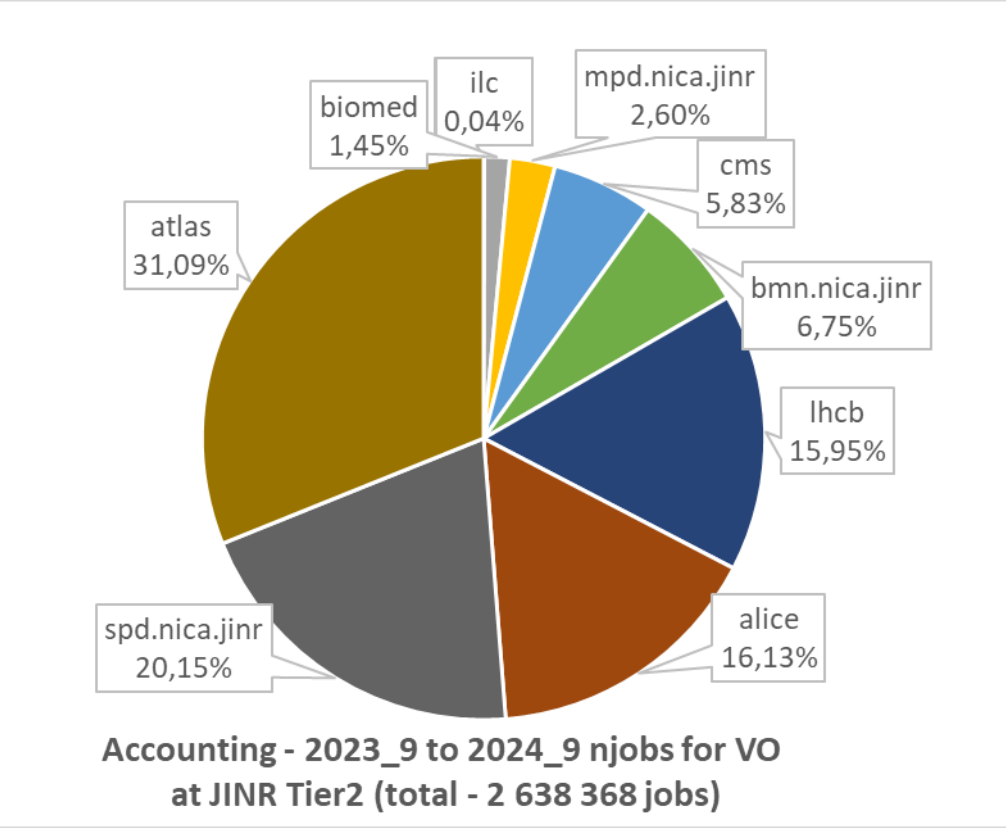
Система хранения(SE)

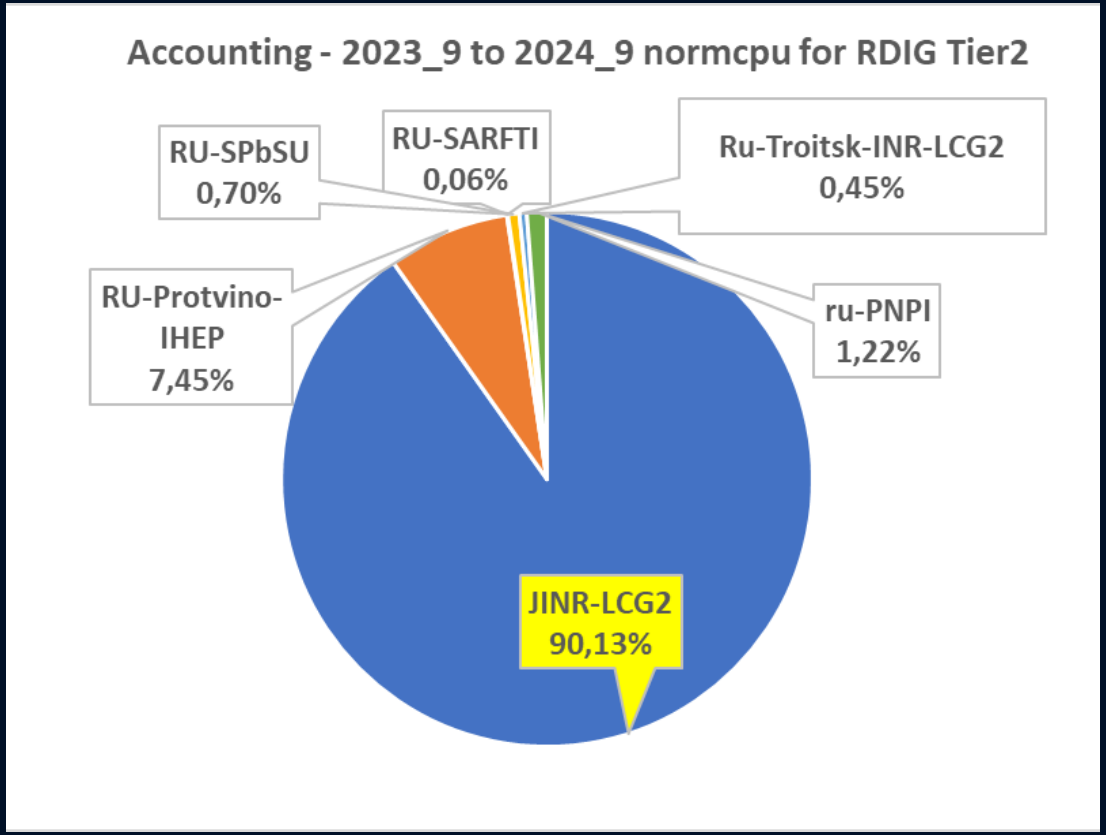
EOS=21829.04 TB
ALICE @ EOS 1653.24 TB
AFS: ~12.5TB (пользовательские директории)
CVMFS: 3 машины: 1 stratum0, 2 stratum
2 squid servers cache CVMFS (VOs: NICA (MPD, B@MN, SPD), dstau, jjnano, juno, baikalgvd).
dCache : SE disks = 3753,69 TB
for CMS: 1903.2695 TB
for ATLAS: 1850.4248 TB
Local & EGI @ dcache2 Total: 256.91 TB

Программное обеспечение:

OC: Scientific Linux release 7.9.
EOS 5.1.23
dCache 8.2
BATCH: Slurm 20.11 адаптированный к kerberos и AFS
grid UMD4 + EPEL (текущие версии)
ARC-CE
FairSoft
FairRoot
MPDroot

JINR Tier2



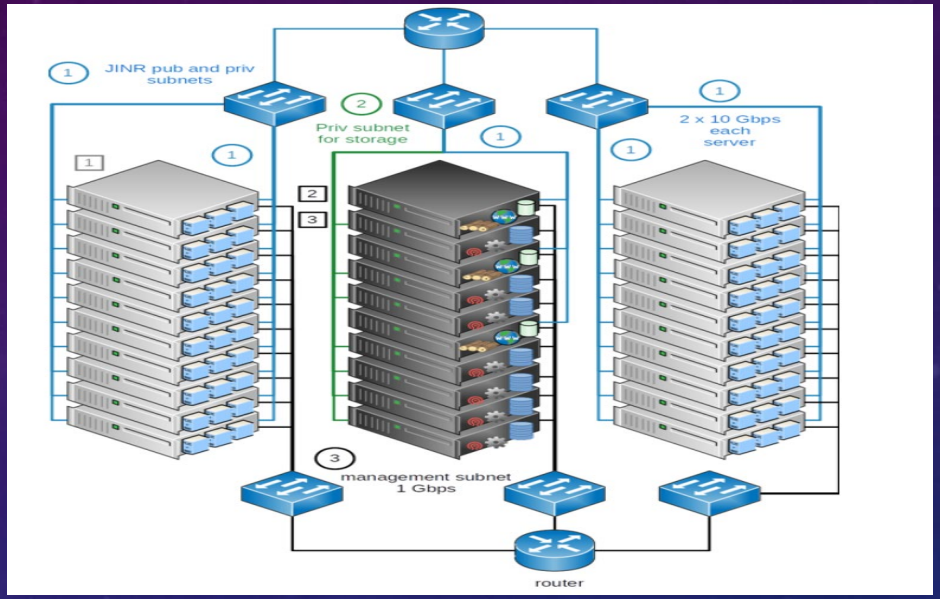


Tier2 в ОИЯИ предоставляет вычислительные мощности, системы хранения данных и доступа к ним для большинства пользователей ОИЯИ и групп пользователей, а также для пользователей виртуальных организаций (VO) грид-среды (LHC, NICA и др.).

Tier2 ОИЯИ является самым производительным в Российском грид для интенсивных операций с данными (RDIG).
Более 80% общего процессорного времени в RDIG используется для вычислений на нашем сайте.

Облачные вычисления

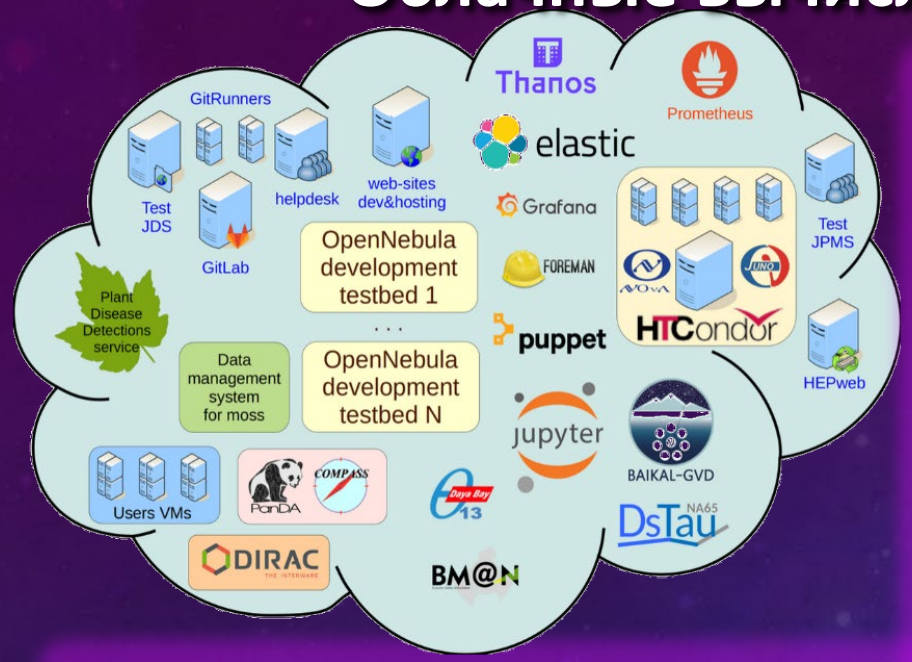
Облачные инфраструктуры ОИЯИ и стран участниц основаны на решении с открытым исходным кодом OpenNebula. Облако ОИЯИ является ядром этой инфраструктуры. На нем размещаются службы DIRAC, которые управляют вычислительными задачами и данными с использованием ресурсов ОИЯИ и стран участниц.

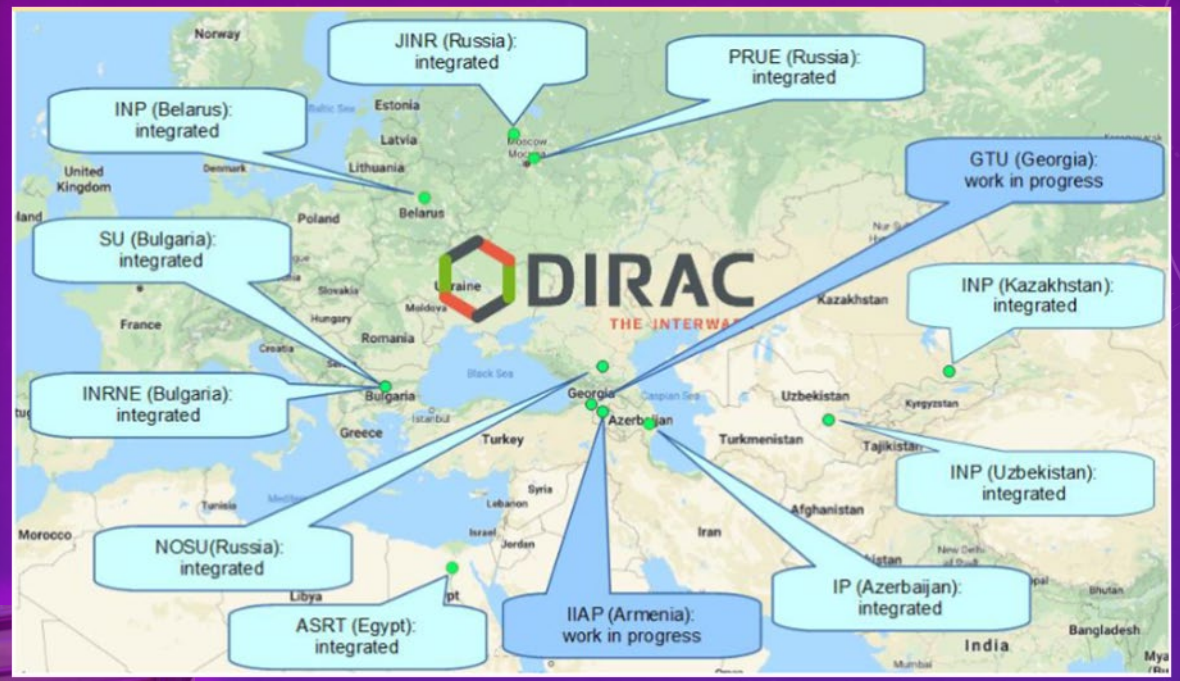


- Вычислительные ресурсы для нейтринных экспериментов:
- Виртуальные машины для пользователей ОИЯИ
- Испытательные стенды для исследований и разработок в области ИТ
- Сервисы системы обработки данных эксперимента COMPASS
- Система управления данными программы по контролю за загрязнением воздуха Европы (UNECE ICP Vegetation)
- Система диагностирования болезней агрокультур посредством использования современных методов машинного обучения
- Сервис для визуализации данных, Gitlab и некоторые другие.
- Распределенная информационно-вычислительная среда на базе DIRAC (DICE), которая интегрирует облака организаций стран-членов ОИЯИ и т.д.

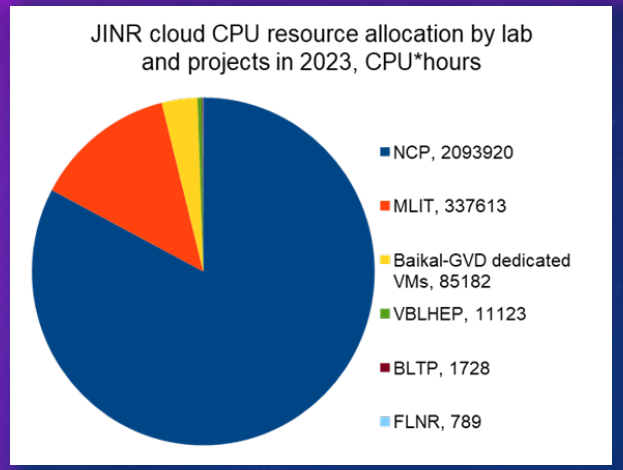
– Облачная платформа: OpenNebula (v5.12.0.4 CE)
– Виртуализация: KVM
– Серверное хранилище для образов виртуальных машин KVM: блочное устройство ceph
– Пользовательские интерфейсы: веб-интерфейс и интерфейс командной строки
– Аутентификация в облачном веб-интерфейсе: центральный пользователь ОИЯИ база данных (LDAP+Kerberos)
● Аппаратное обеспечение: 174 сервера для ВМ: >5000 ядер ЦП, ОЗУ на каждое ядро ЦП: 5,3–16 ГБ
– 24 сервера для хранилищ Ceph 3 ПБ
● URL веб-интерфейса: http://cloud.jinr.ru.

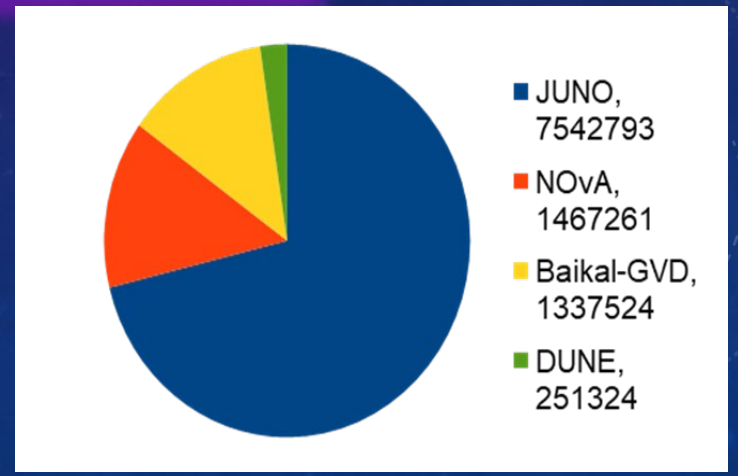
Облачные вычисления





Распределенная информационно-вычислительная среда (DICE) на базе DIRAC, объединяющая облака организаций государств-участников ОИЯИ.

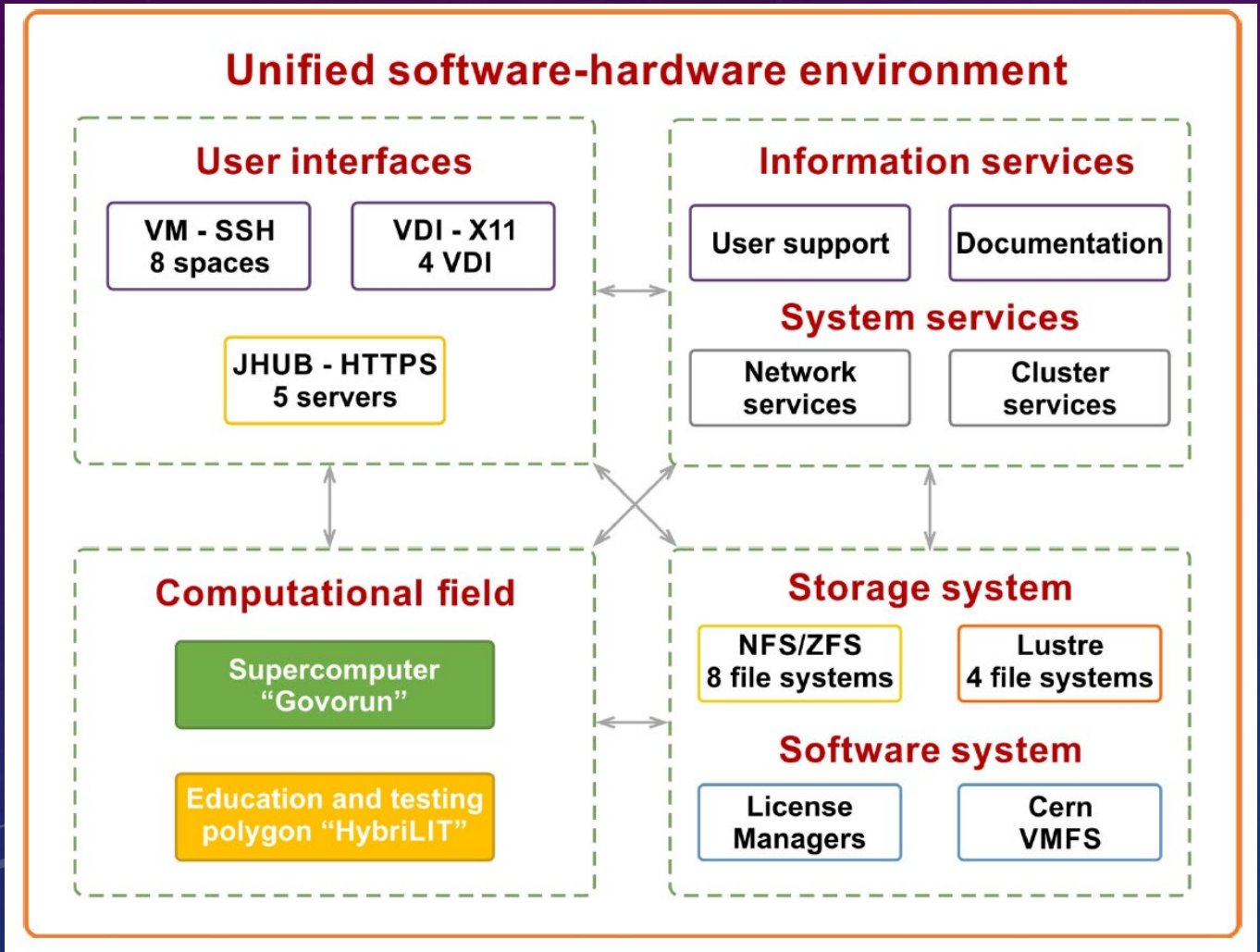




Нейтринные эксперименты являются основными пользователями облачной инфраструктуры.

Гетерогенная платформа «HybriLIT»

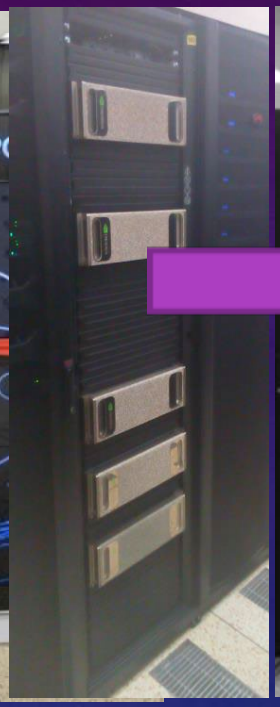
Гетерогенная платформа состоит из Суперкомпьютера «Говорун» и учебно-тестового полигона «HybriLIT».



Учебно-тестовый полигон имеет гетерогенную структуру вычислительных узлов и позволяет разрабатывать параллельные приложения для проведения расчетов на различных вычислительных архитектурах, таких как многоядерные процессоры, сопроцессоры Intel Xeon Phi и линейки графических процессоров NVIDIA (Testla K20, K40, K80), а также проводить учебные курсы по технологиях параллельного программирования, позволяющим студентам осваивать работу на новейших вычислительных архитектурах.

Единая двухуровневая программно-информационная среда для Учебно-тестового полигона и суперкомпьютера "Говорун" позволяет пользователям протестировать и отладить свое приложение до отправки его на «Говорун».

Развитие гетерогенной платформы HybriLIT



2014

Кластер HybriLIT:
140 TFlops – одинарная точность
50 TFlops - двойная точность

2018

Суперкомпьютер «Говорун»
1 PFlops – одинарная точность
500 TFlops - двойная точность
9-ый в рейтинге IO500 (июль 2018)
#18 в Top50

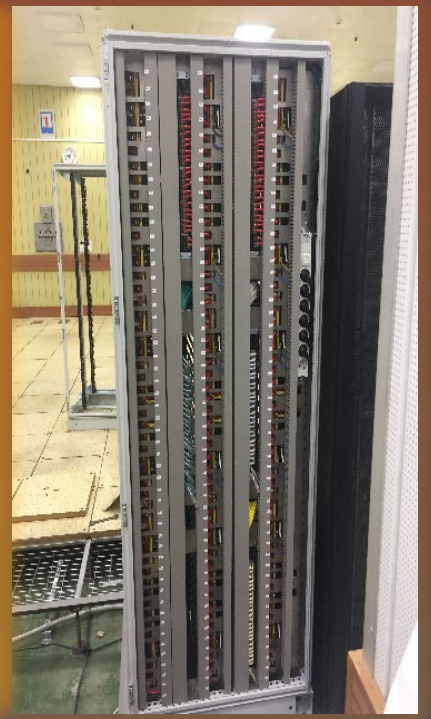
2019

Суперкомпьютер «Говорун»
1,7 PFlops – одинарная точность
860 TFlops - двойная точность
288 ТВ ССХД I/O > 300 Gb/s
17-ый в рейтинге IO500 (июль 2020)
#10 в Top50

2020

RUSSIAN DATA CENTER AWARDS 2020
«Лучшее ИТ-решение для ЦОДа»

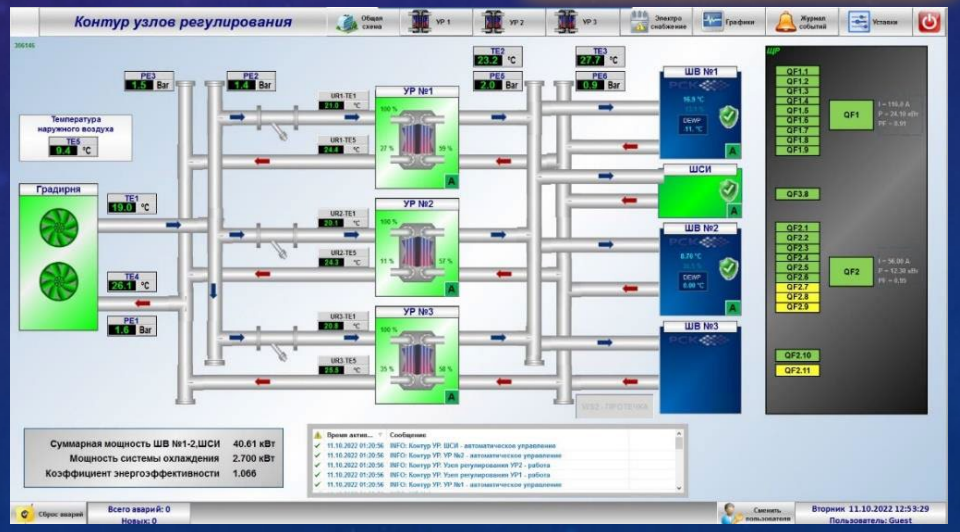
Суперкомпьютер «Говорун». Охлаждение горячей водой



В суперкомпьютер поступает вода, охлажденная до температуры 45 градусов. Пройдя весь контур суперкомпьютера, нагретая до 50 градусов вода возвращается в теплообменник, где охлаждается, передавая тепловую энергию в гидравлический контур сухой градирни.

PUE ~ 1,06

Power usage effectiveness



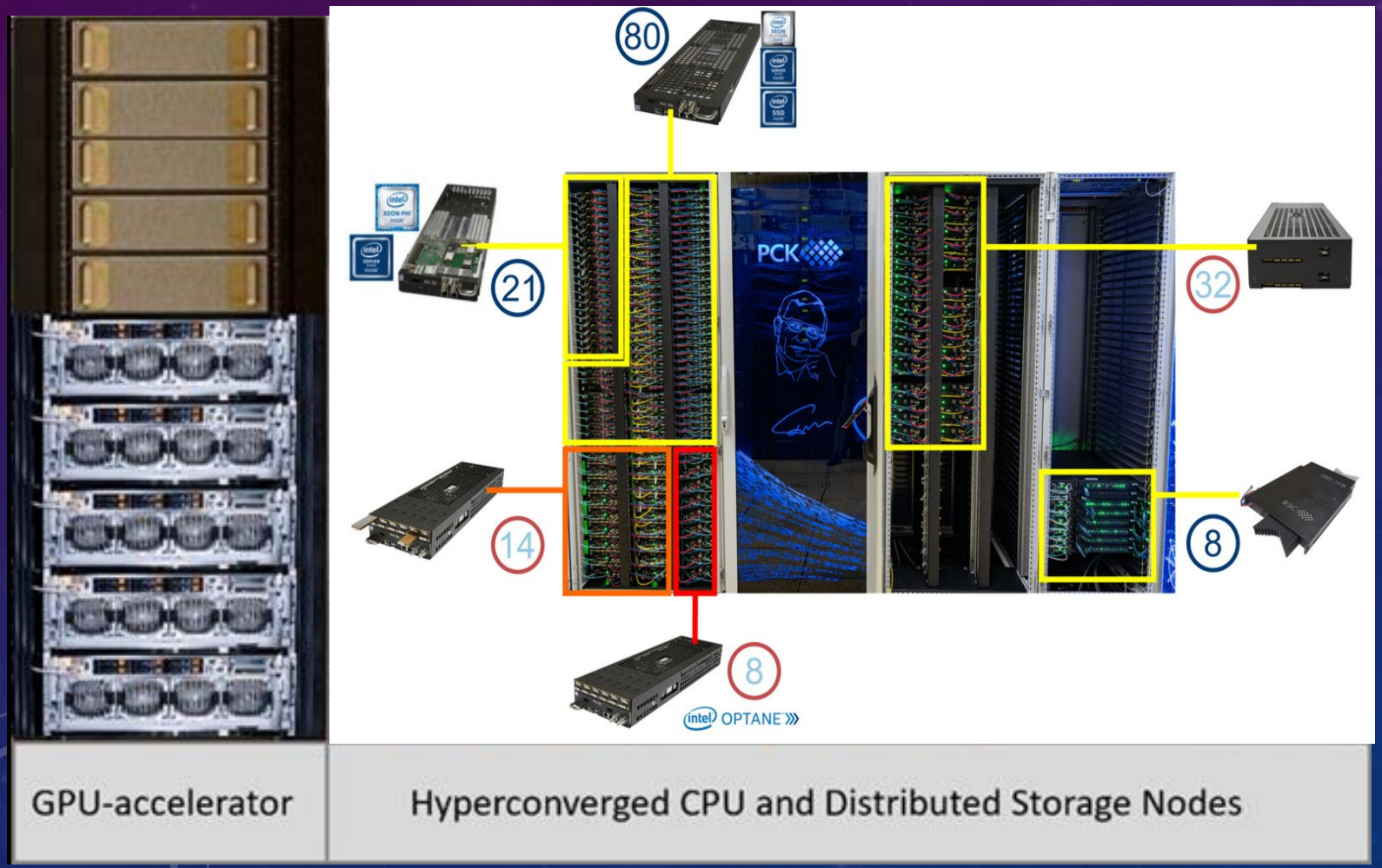
Высокая доступность, отказоустойчивость и простота использования вычислительных систем, обеспечиваются благодаря передовой системе управления и мониторинга на базе ПО «РСК БазИС».

Система охлаждения имеет плавную регулировку производительности, которая позволяет увеличивать или уменьшать мощность системы охлаждения в соответствии с реальной загрузкой. Это позволяет значительно снизить потребление электроэнергии при частичной загрузке.

Суперкомпьютер «Говорун» сегодня

Двухкомпонентная система:

CPU-компонента, базирующаяся на новейших архитектурах Intel (процессоров Intel Xeon Phi и Intel Skylake);
GPU-компонента, базирующаяся на узлах Niagara R4206SG Ampere A100 и NVIDIA DGX-1 Volta V100.



- Гиперконвергентная программно-конфигурируемая система
- Иерархическая система обработки и хранения данных
- Масштабируемое решение
- Общая пиковая производительность: 1,7 Пфлопс DP
- Общая емкость иерархического хранилища: 8,6 ПБ
- GPU компонента на базе Niagara Ampere A100 и NVIDIA DGX-1 Volta V100
- CPU компонента на базе решений жидкостного охлаждения РСК «Торнадо»
- Самый энергоэффективный центр России (PUE=1,06)
- Скорость ввода-вывода данных: 300 Гбит/с

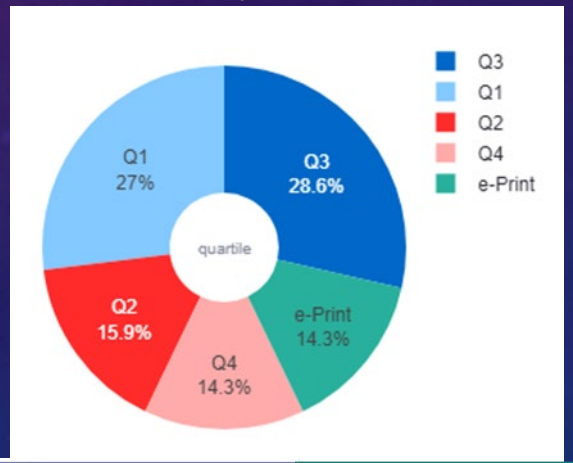
Суперкомпьютер «Говорун»

Ресурсы суперкомпьютера “Говорун” используются научными группами из всех лабораторий института для решения широкого круга задач в области теоретической физики, а также для физического моделирования и обработки экспериментальных данных.

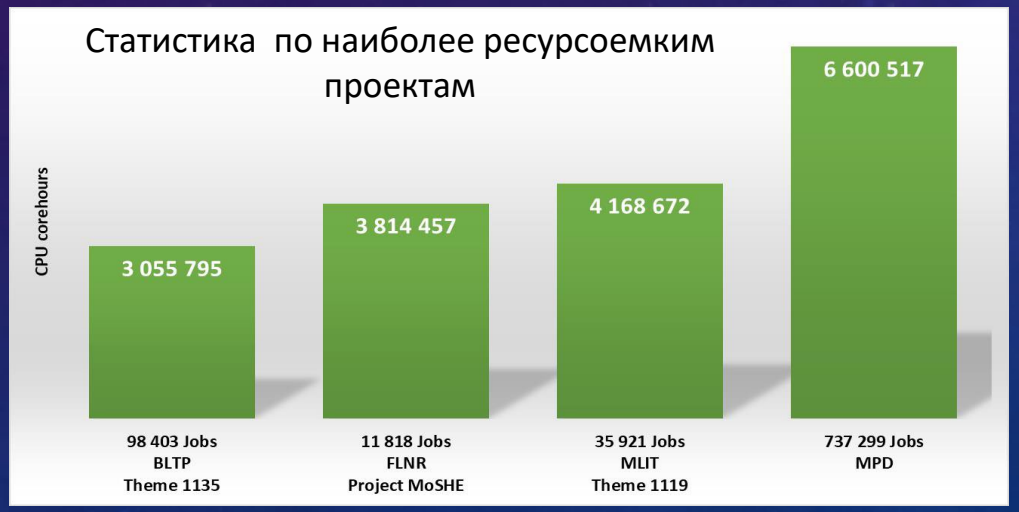
На СК «Говорун» развернут полигон для квантовых вычислений (симуляторы квантовых вычислений)

Ключевые проекты, в которых используются ресурсы СК “Говорун”:

- ❑ мегапроект NICA,
- ❑ расчеты решеточной квантовой хромодинамики,
- ❑ расчеты свойств атомов сверхтяжелых элементов,
- ❑ исследования в области радиационной биологии,
- ❑ расчеты радиационной безопасности объектов ОИЯИ.



В течение 2023 года все группы пользователей СК “Говорун” выполнили 640 861 задание на CPU и 7 808 заданий на GPU компонентах. Средняя загрузка CPU составила **96,4%,** в то время как загрузка GPU - **91,2%.**



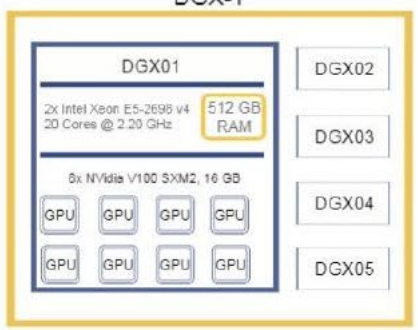
За 2023 год пользователи HybriLIT опубликовали 65 статей в различных областях:

- физика элементарных частиц и атомного ядра,
- физика высоких энергий,
- биофизика и химия,
- нейросетевой подход, методы и алгоритмы машинного
- и глубокого обучения и др.

GPU КОМПОНЕНТА

2017

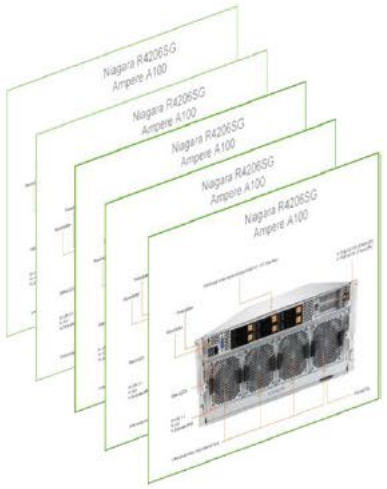
DGX-1



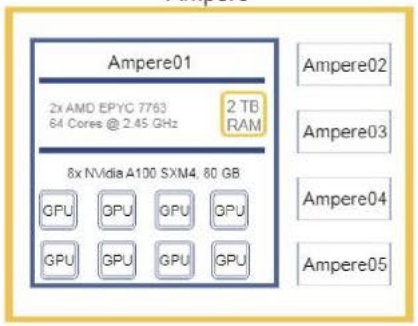
40 NVIDIA V100

+

2023



Ampere

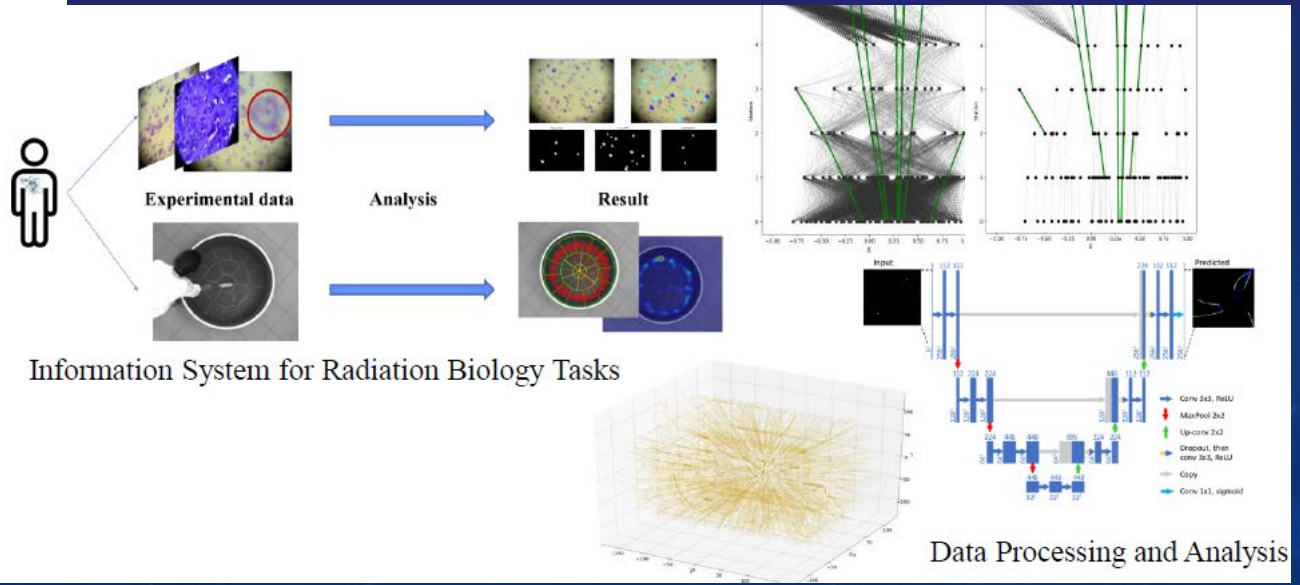


40 NVIDIA A100

Пиковая производительность GPU-компоненты:
900 Tflops для вычислений с двойной точностью
26 Pflops для вычислений с половинной точностью

GPU-компонента дает пользователям суперкомпьютера возможность использовать алгоритмы машинного обучения и глубокого обучения для решения прикладных задач с помощью нейросетевого подхода:

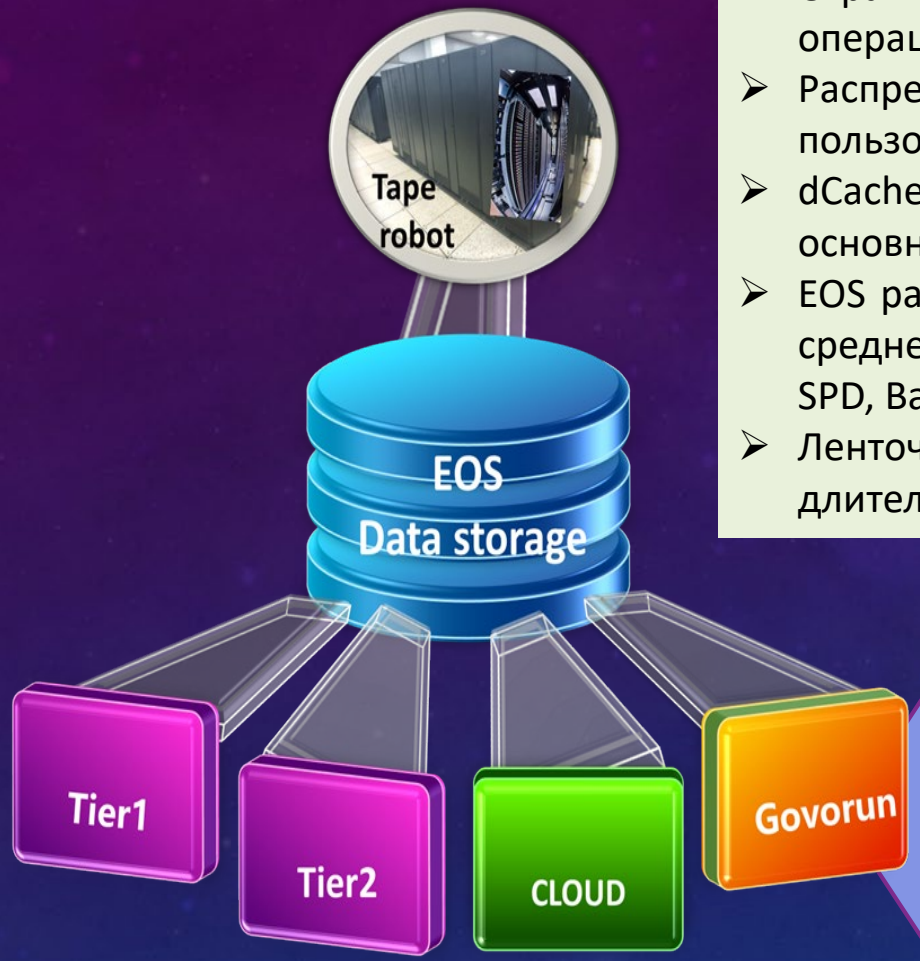
- обработка данных экспериментов в ЛРБ в составе Информационной системы для задач радиационной биологии;
- обработка и анализ экспериментальных данных на NICA и т.д.



Information System for Radiation Biology Tasks

Data Processing and Analysis

МНОГОУРОВНЕВАЯ СИСТЕМА ХРАНЕНИЯ ДАННЫХ



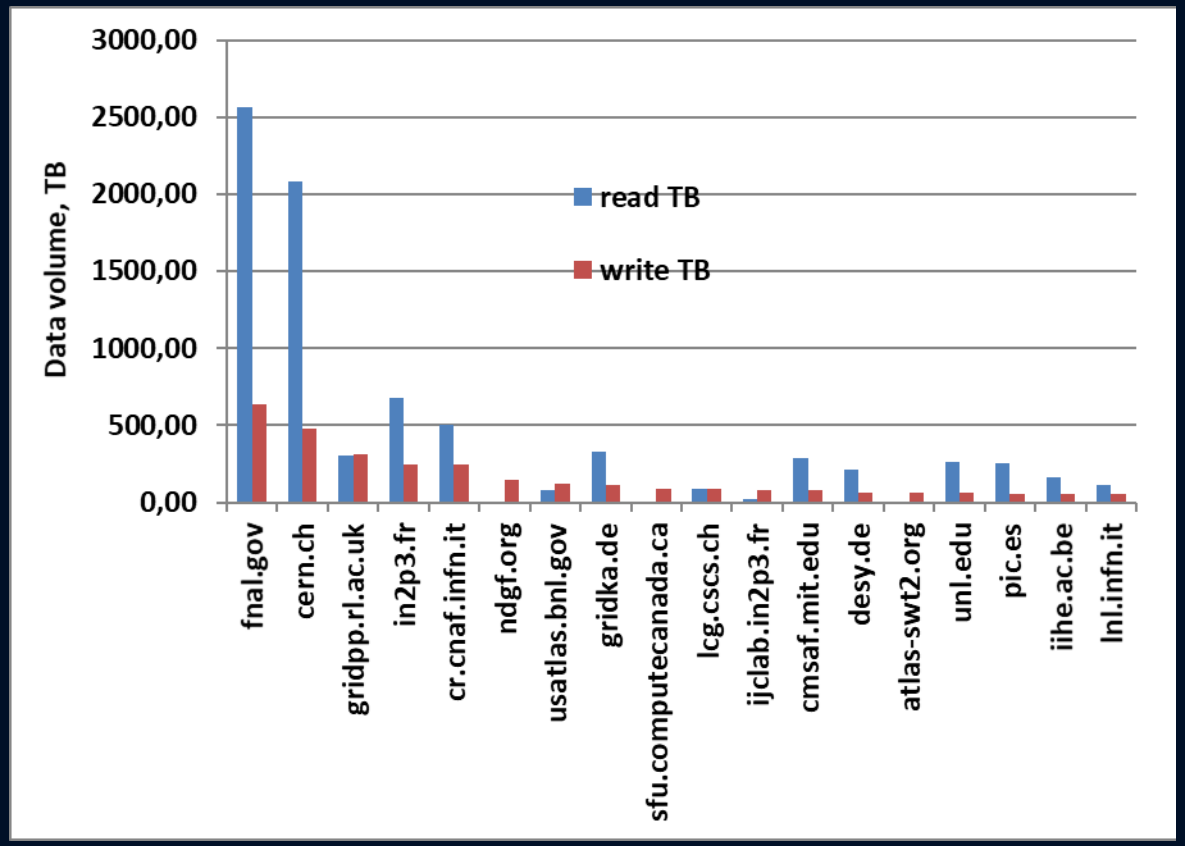
- Ограниченное количество данных и краткосрочное хранилище - для хранения самой операционной системы, временных пользовательских файлов
- Распределенная глобальная система AFS – для хранения домашних каталогов пользователей и программного обеспечения
- dCache является традиционным для Grid – для хранения больших объемов данных (в основном для экспериментов на LHC) на среднесрочный период
- EOS распространяется на все ресурсы МИВК – для хранения больших объемов данных на среднесрочный период. В настоящее время EOS используется для хранения BM@N, MPD, SPD, BaikalGVD и др.
- Ленточные роботизированные системы – для хранения больших объемов данных на длительный период. В настоящее время - для CMS. BM@N, MPD, SPD, JUNO – в разработке.

На суперкомпьютере “Говорун” была разработана и внедрена специальная иерархическая система обработки и хранения данных с программно-определяемой архитектурой.
По скорости доступа к данным различают следующие уровни:

- горячие данные(LUSTRE),
- теплые данные (EOS)
- холодные данные (TAPE)

Система долговременного хранения

Статистика по обмену данными с начала 2023



Системы хранение и данные.

TS3500 в режиме ожидания, в данный момент подключен к CTA

TS4500 работает на CMS, половина емкости зарезервирована для NICA.

TS3500 12 ПБ свободно

TS4500 всего 90 ПБ

Объем записанных и считанных в 2023 году данных в PB

	T2	T1	T1mss
write	3.0	4.3	2.0
read	8.2	19.8	1.1

Развитие на ближайшие год-два

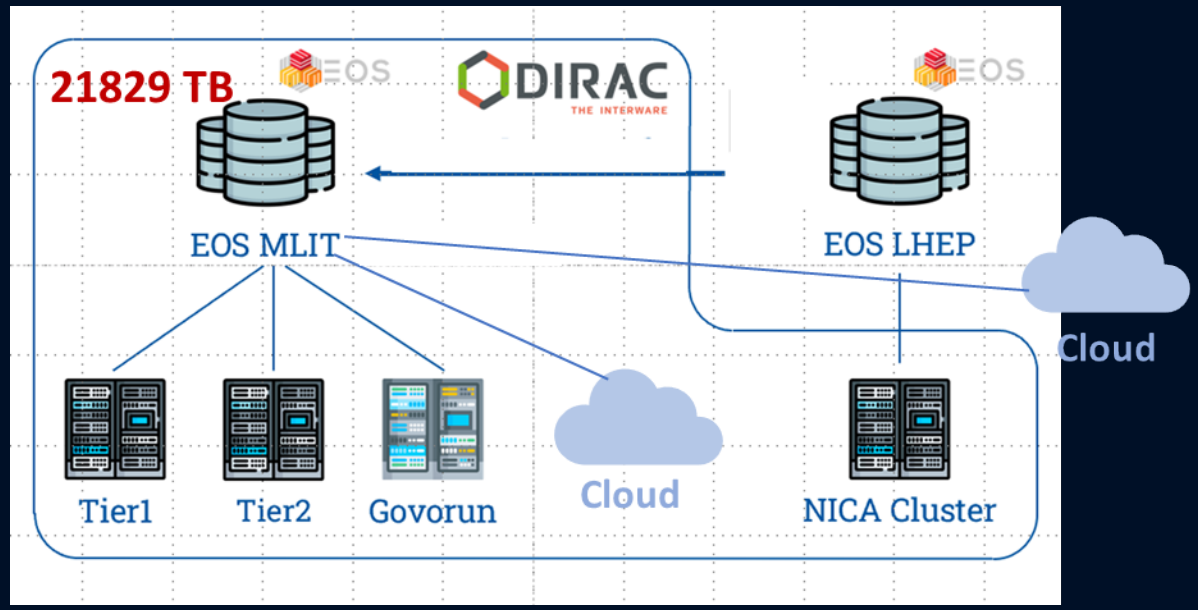
TS3500 12 PB, 12 LTO6 используются в качестве испытательной площадки для установки EOSCTA. Будет хранилищем для экспериментов, не связанных с WLCG.

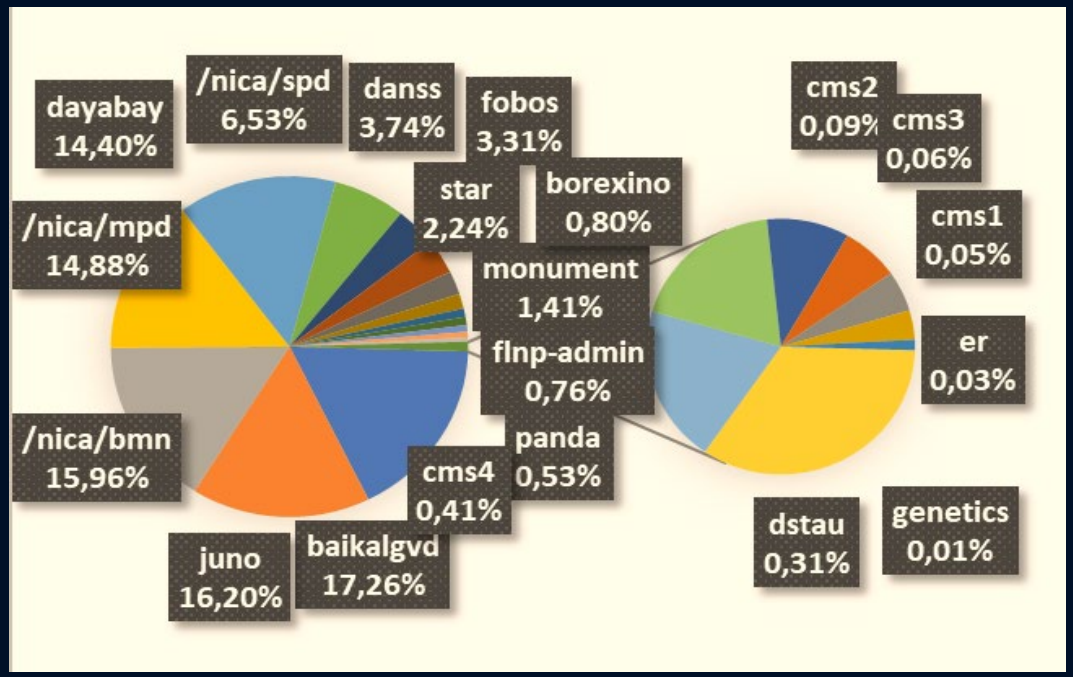
TS4500 90 PB 12 драйвов 3592-60F Jaguar будет разделен на несколько логических библиотек

Часть - под управлением Enstore для CMS

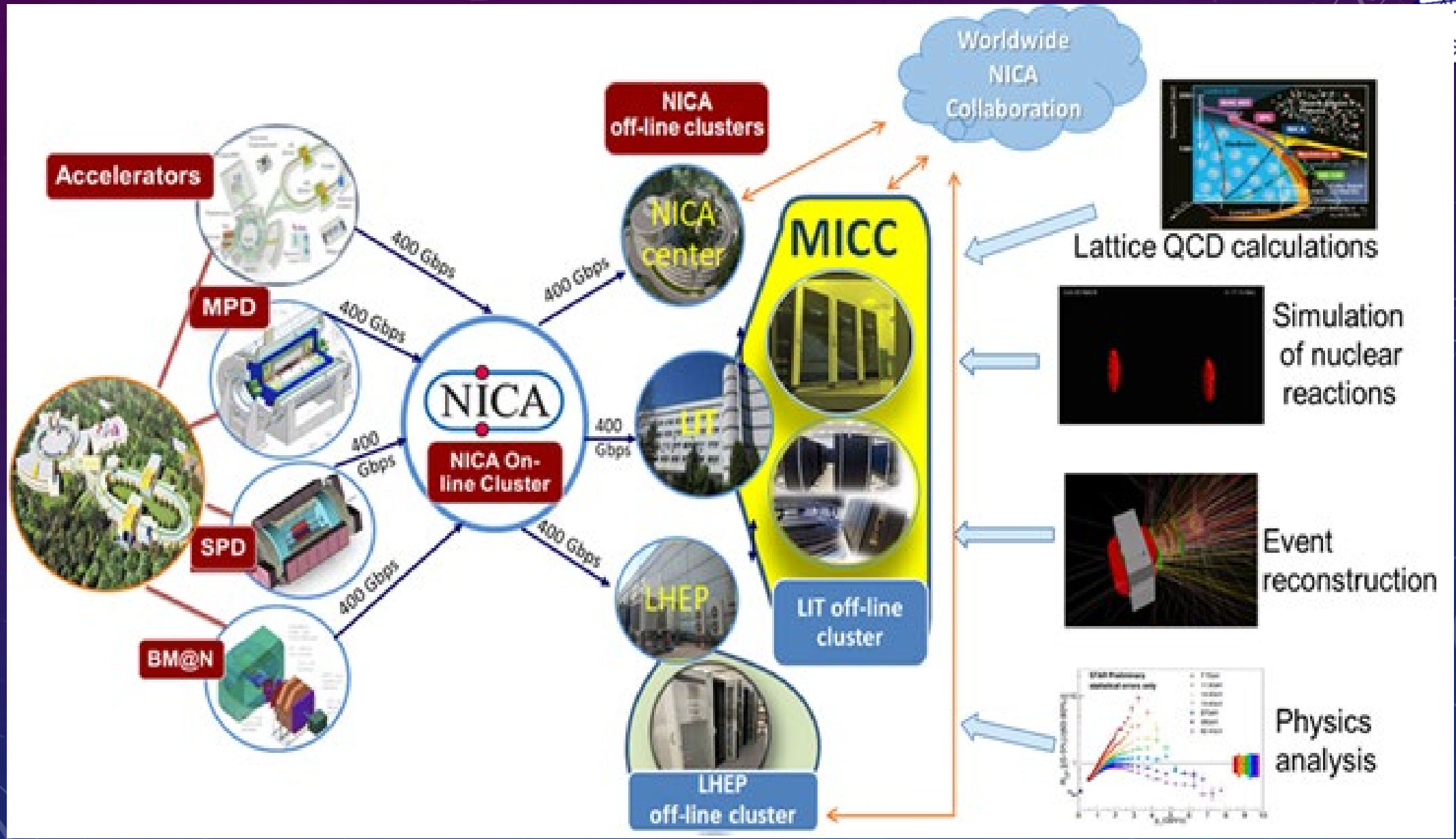
Часть - под EOSCTA для NICA

Система среднесрочного хранения





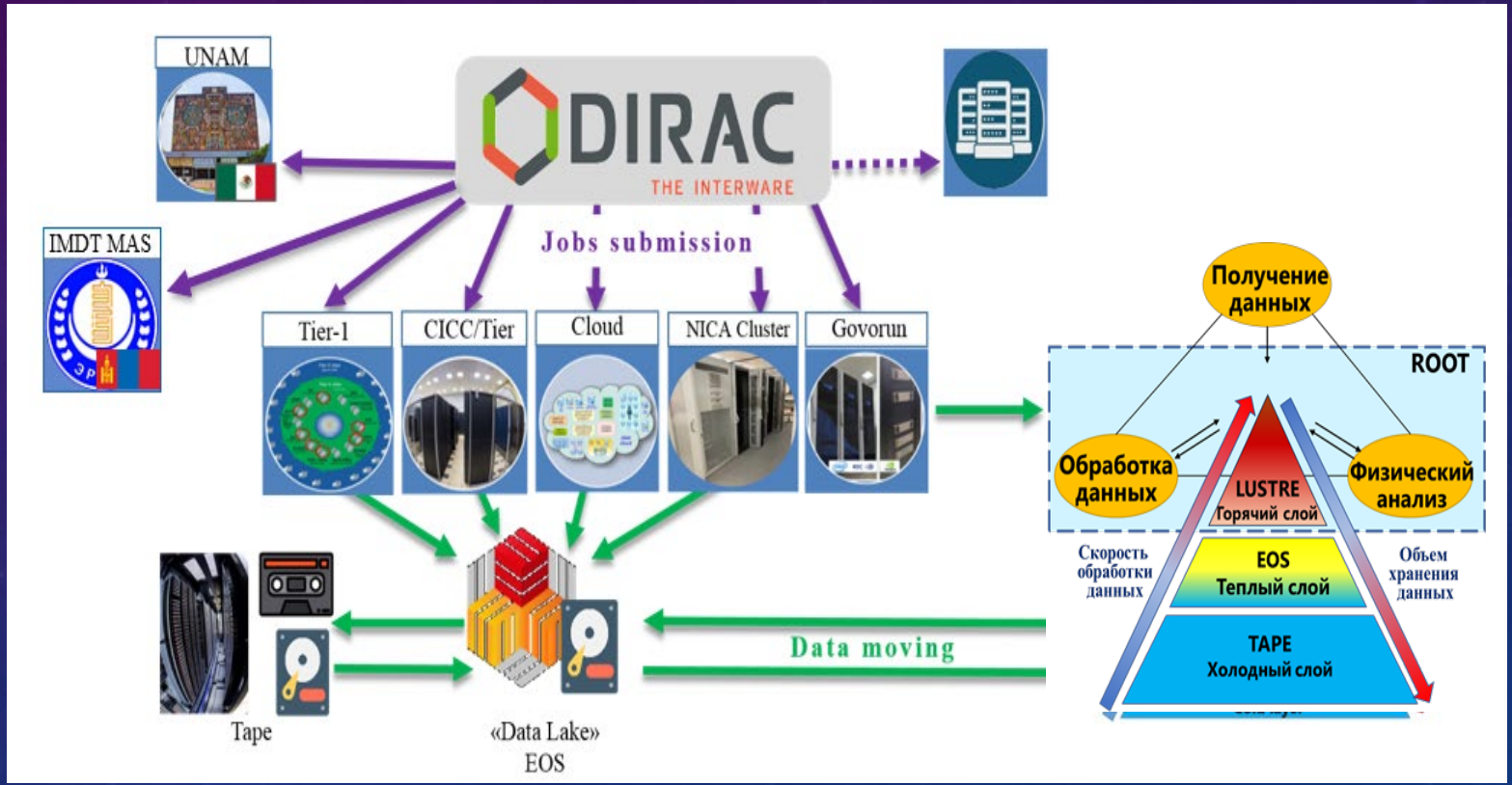
EOS является системой хранения очень больших объемов данных.
Оптимален по соотношению стоимость/объем хранения.
Удобна для пользователей почти как локальная файловая система.
Поддерживает множество протоколов доступа: POSIX при установке на пользовательском компьютере; xroot и http для быстрого удаленного доступа.
Высокая надежность хранения данных за счет дублирования на разных серверах, хранения на разных серверах в формате вертикального RAID с контрольными суммами.
Высокая скорость доступа к данным за счет параллельного копирования с множества серверов.
Защита данных с помощью расширенного списка доступа для групп и отдельных пользователей.


Worldwide
NICA
Collaboration
NICA
off-line clusters
Accelerators
MPD
SPD
BM@N
400 Gbps
400 Gbps
400 Gbps
400 Gbps
400 Gbps
400 Gbps
400 Gbps
NICA
NICA On-
line Cluster
NICA
center
LIT
LHEP
MICC
LIT off-line
cluster
LHEP
off-line cluster
Lattice QCD calculations
Simulation
of nuclear
reactions
Event
reconstruction
Physics
analysis

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ

Для обеспечения бесперебойной работы МИВК и эффективного использования вычислительных ресурсов необходима разработка базового программного обеспечения.

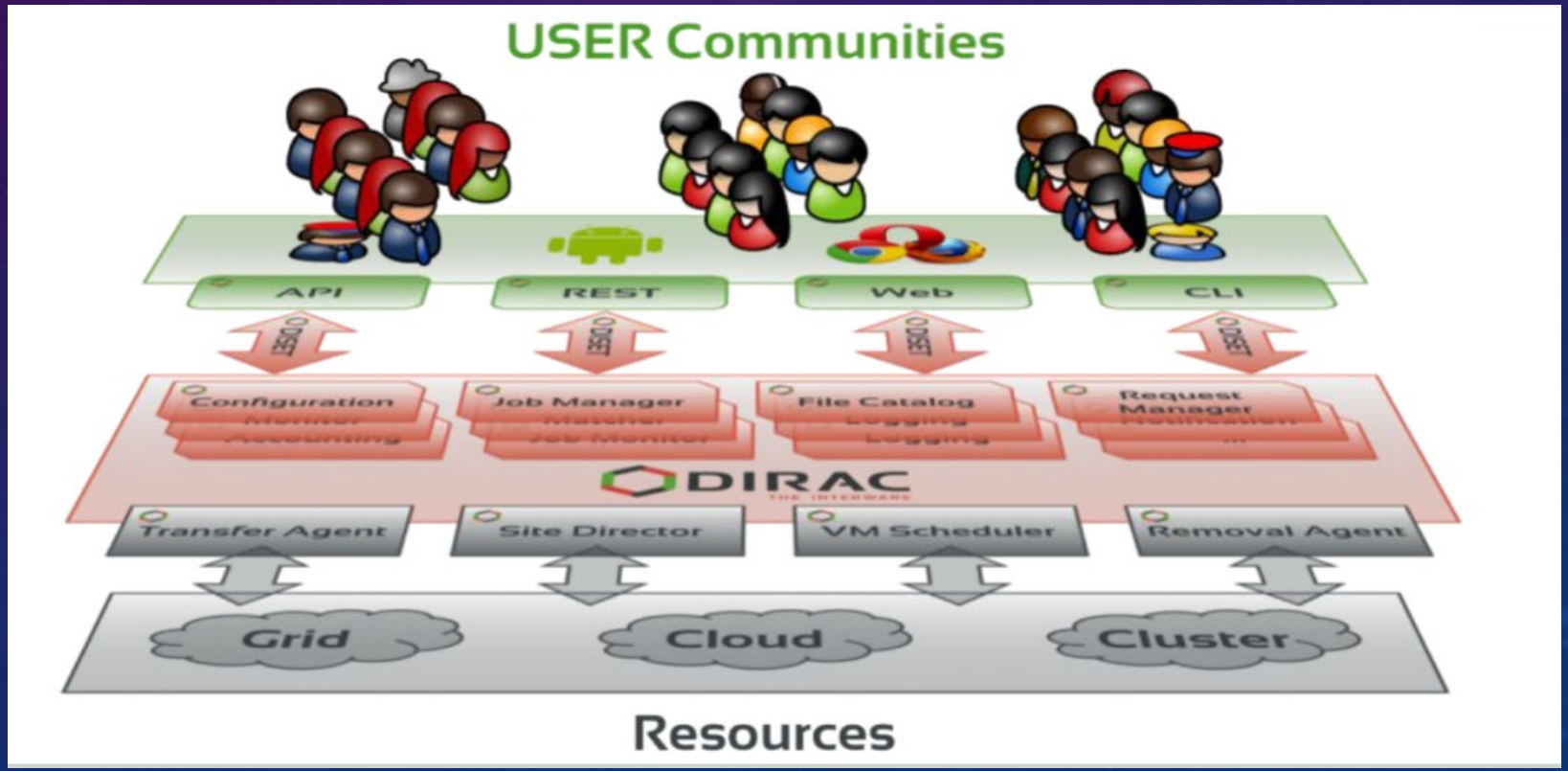
- распределенная программно-определяемая высокопроизводительная вычислительная платформа для обработки и хранения данных экспериментов, объединяющая суперкомпьютерные (гетерогенные), грид- и облачные технологии для эффективного использования новых вычислительных архитектур
- многофункциональная программно-аппаратная платформа аналитики больших данных на базе гибридных аппаратных ускорителей (GPU, FPGA, квантовые системы); алгоритмы машинного обучения; инструменты аналитики, отчетности и визуализации; поддержка пользовательских интерфейсов и задач



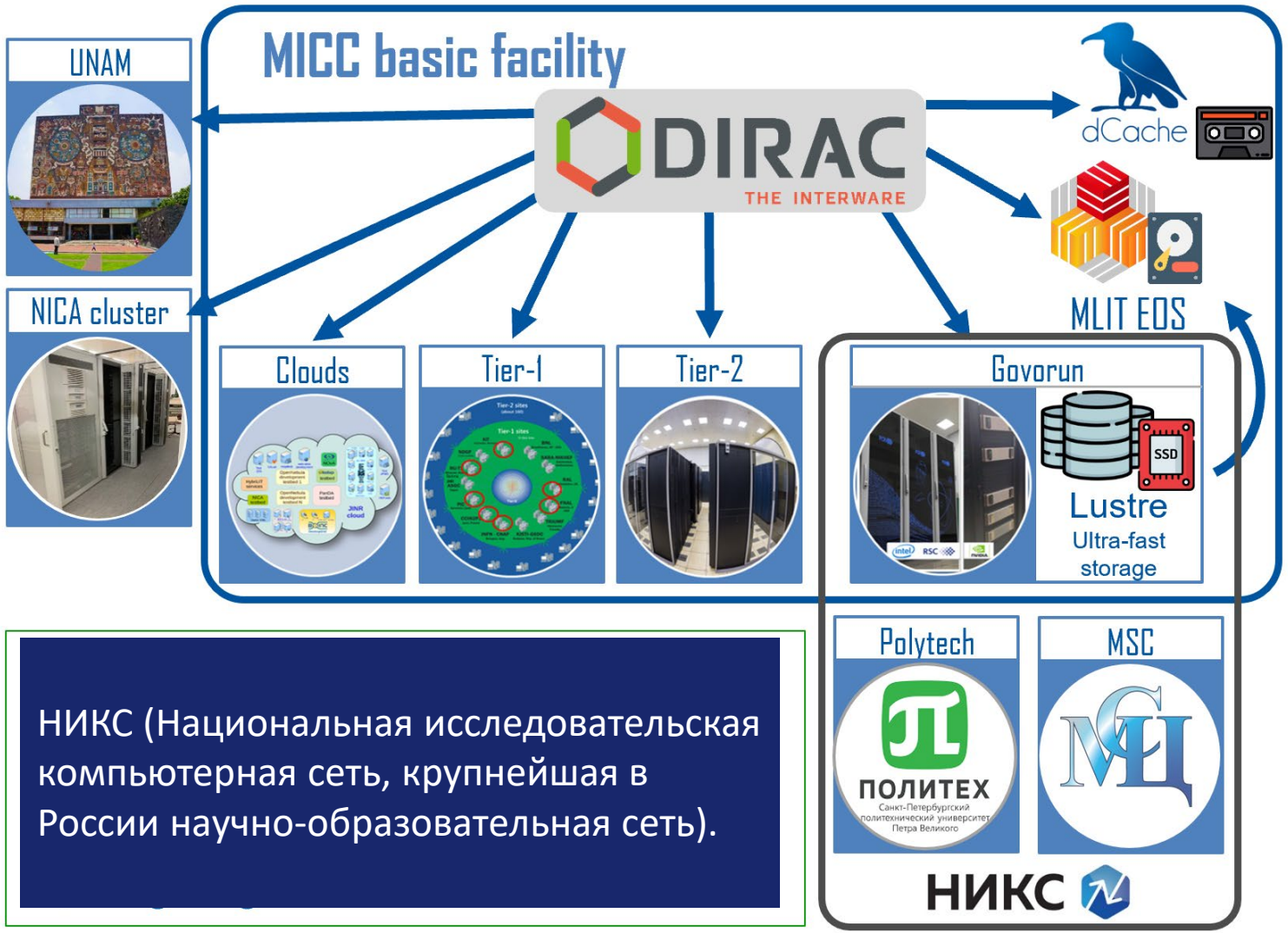
- передовые системы защиты киберинфраструктуры, компьютерной и пользовательской информации, публичных электронных услуг и аутентификации пользователей

Распределенная гетерогенная среда на основе DIRAC

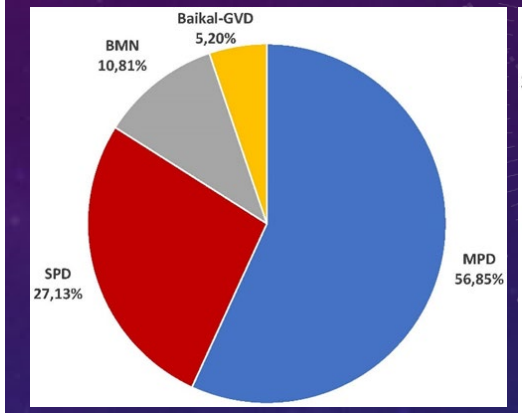
На текущий момент **DIRAC** Interware (Distributed Infrastructure with Remote Agent Control) – единственная система, которая интегрирует все компоненты МИВК. **DIRAC** Interware — это программная среда для распределенных вычислений, обеспечивающая полное решение для одного (или более) сообщества пользователей, требующего доступа к распределенным ресурсам. **DIRAC** создает промежуточный слой между пользователями и ресурсами, предлагая общий интерфейс для ряда гетерогенных поставщиков, интегрируя их бесшовно, обеспечивая интероперабельность, одновременно с оптимизированным, прозрачным и надежным использованием ресурсов.



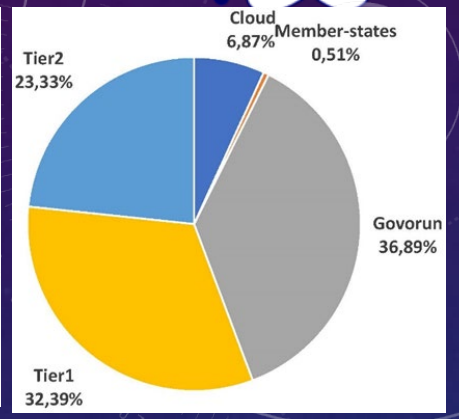
Распределенная гетерогенная среда на основе DIRAC



НИКС (Национальная исследовательская компьютерная сеть, крупнейшая в России научно-образовательная сеть).



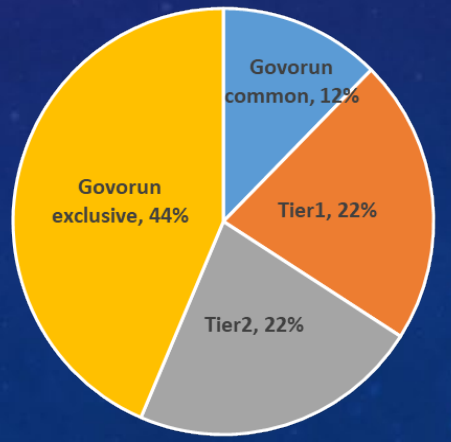
Использование платформы DIRAC экспериментами в 2023 г.

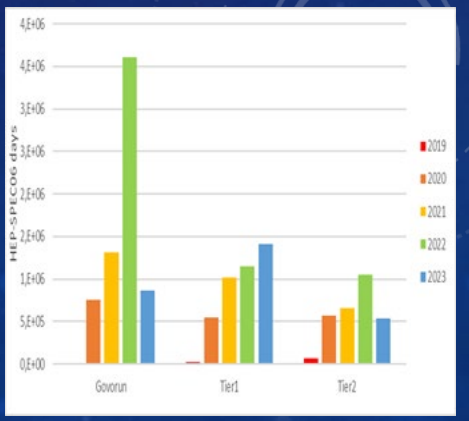


Доля использования компонентов МИВК в DIRAC в 2023 г

Основным пользователем платформы является эксперимент MPD

MPD Monte-Carlo

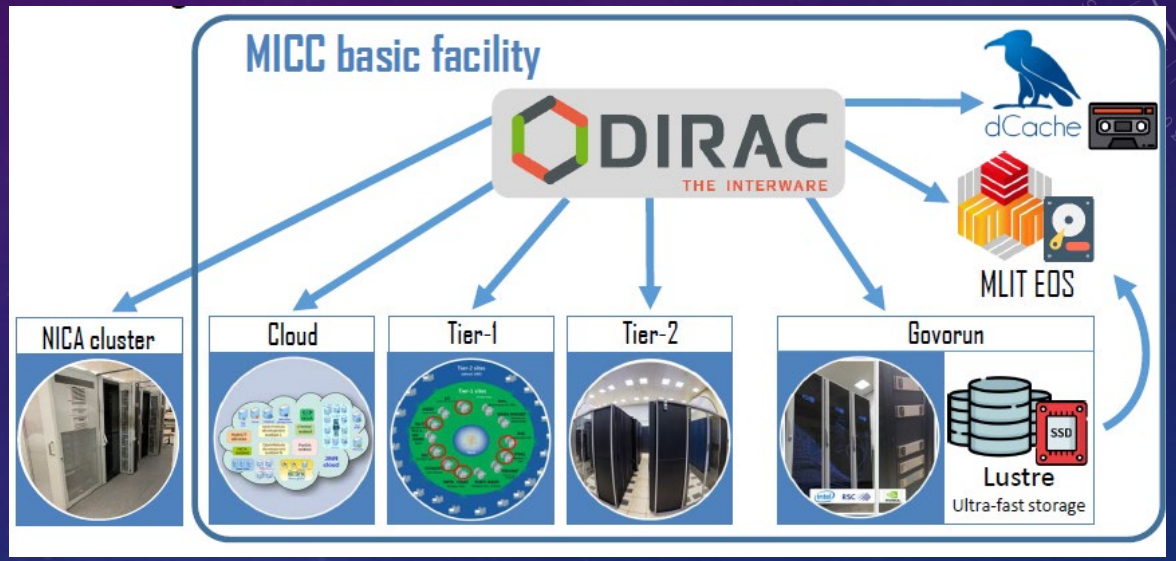


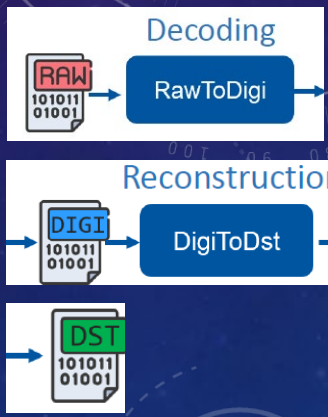


Распределенная гетерогенная среда на основе DIRAC

В 2023 году впервые в ОИЯИ на распределённой гетерогенной вычислительной инфраструктуре, объединённой с помощью платформы **DIRAC,** реализована полная обработка сырых данных 8-го сеанса эксперимента **BM@N.** В ходе сеанса было собрано **430 ТБ** данных в виде **~30000** файлов. В течение года, после внесения изменений в пакет BmnRoot, который используется для реконструкции данных, эта процедура применялась несколько раз для переобработки исходных данных. Всего за год были выполнены 5 больших и 7 малых сеансов по реконструкции/генерации данных.

BM@N Run 8 data processing

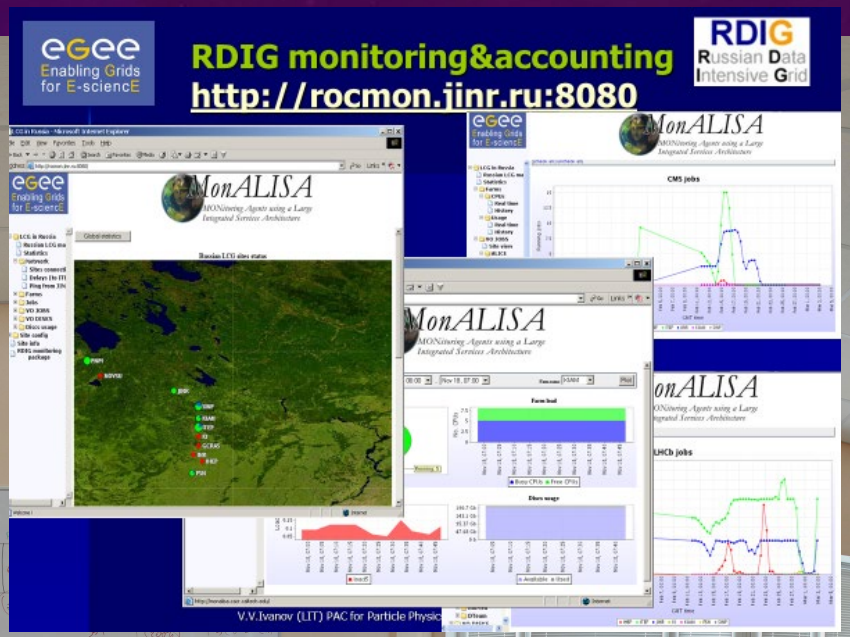


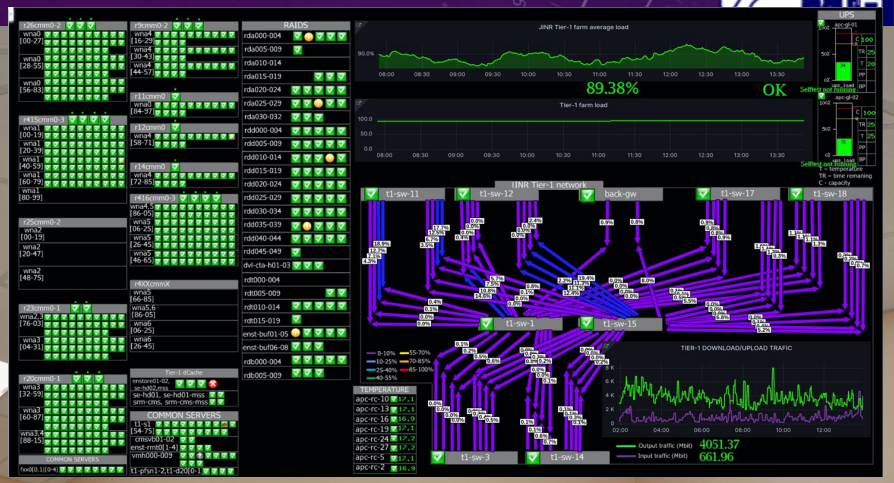


31 k	400 TB	12	363 k	86 years	~3 PB
amount of raw files	BM@N Run8 RAW data size	data processing campaigns	jobs completed	total computation time	data transferred

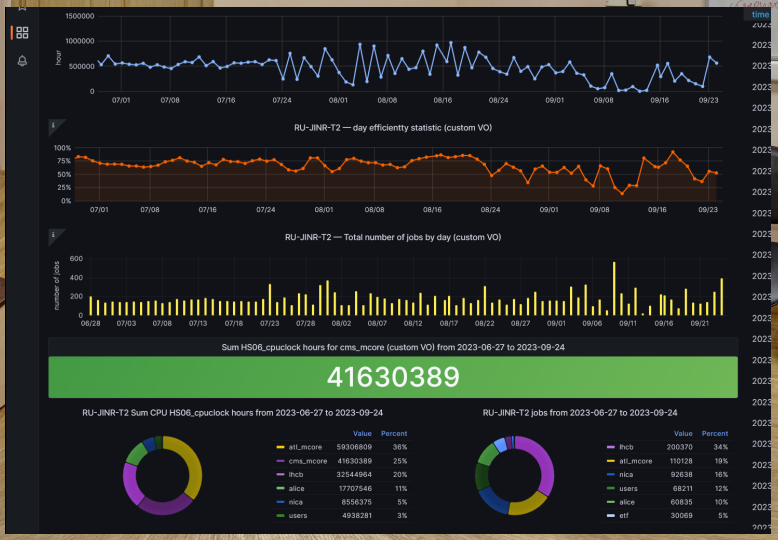
Сводная статистика использования платформы DIRAC для обработки данных BM@N Run 8

МОНИТОРИНГ

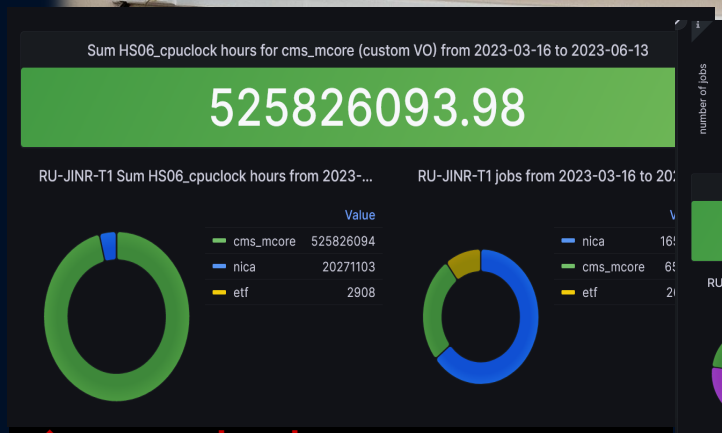

RDIG monitoring&accounting
http://rocmon.jinr.ru:8080
RDIG
Russian Data Intensive Grid
MonALISA
V.V.Ivanov (LIT) PAC for Particle Physic

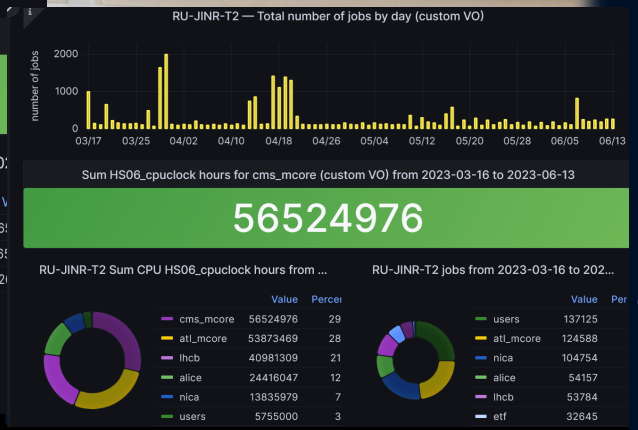

JINR Tier-1 farm average load
89.38%
OK
Tier-1 farm load
JINR Tier-1 network
TIER-1 DOWNLOAD/UPLOAD TRAFIC
4051.37
661.96


General / start_Dashboard
50.6 PB
2.65 PB
11.7 PB
20000
1.99 PB
1.69 PB
1.94 PB
10364
7.51 PB
15680
22.4 PB
4320
5152
60.6 TB
3.84 PB
1.44 PB


RU-JINR-T2 — day efficienty statistic (custom VO)
RU-JINR-T2 — Total number of jobs by day (custom VO)
Sum HS06_cpuclock hours for cms_mcore (custom VO) from 2023-06-27 to 2023-09-24
41630389
RU-JINR-T2 Sum CPU HS06_cpuclock hours from 2023-06-27 to 2023-09-24
RU-JINR-T2 jobs from 2023-06-27 to 2023-09-24

МИВК мониторинг и аккаунтинг





- 3 monitoring servers
- About 16000 service checks
- About 1800 nodes

Успешное функционирование вычислительного комплекса обеспечивается системой, которая контролирует все компоненты МИВК.

Необходимо:

- расширить систему мониторинга, интегрировав в нее локальные системы мониторинга систем электроснабжения (дизель-генераторы, блоки распределения электроэнергии, трансформаторы и источники бесперебойного питания).;
- организовать мониторинг системы охлаждения (градирни, насосы, контуры горячей и холодной воды, теплообменники, чиллеры).;
- создать центр управления инженерной инфраструктурой (специальные информационные панели для визуализации всех статусов инженерной инфраструктуры МИВК в единой точке доступа);
- учитывать каждое пользовательское задание в каждом компоненте МИВК.

Требуется разработать интеллектуальные системы, которые позволят обнаруживать аномалии, что приведет к необходимости создания специальной аналитической системы в рамках системы мониторинга для автоматизации процесса.

Семилетний план развития МИВК

1. Семилеткой предусмотрено создание на базе ЛИТ центра долгосрочного хранения данных на ресурсах МИВК.
2. Процесс моделирования, обработки и анализа экспериментальных данных, полученных с детекторов BM@N, MPD и SPD, будет реализован в распределенной вычислительной среде на базе МИВК и вычислительных центров ЛФВЭ и стран-участниц коллабораций.
3. Региональный центр обработки данных, предназначенный для производства, хранения и обработки для эксперимента JUNO. Ожидается, что этот центр обработки данных станет одним из трех европейских центров обработки данных JUNO. Ресурсы, необходимые для обработки и хранения данных JUNO, были одобрены сторонами в рамках «Меморандума о взаимопонимании по сотрудничеству в развертывании и эксплуатации вычислительной сети JUNO», подписанного между ИФВЭ и ОИЯИ 1 сентября 2022 г.
4. Продолжение работы в качестве Tier1 и Tier2 для LHC (HL-LHC).
5. Расширение инфраструктуры облачных вычислений.
6. Дальнейшее развитие, наращивание производительности и возможностей суперкомпьютера «Говорун».

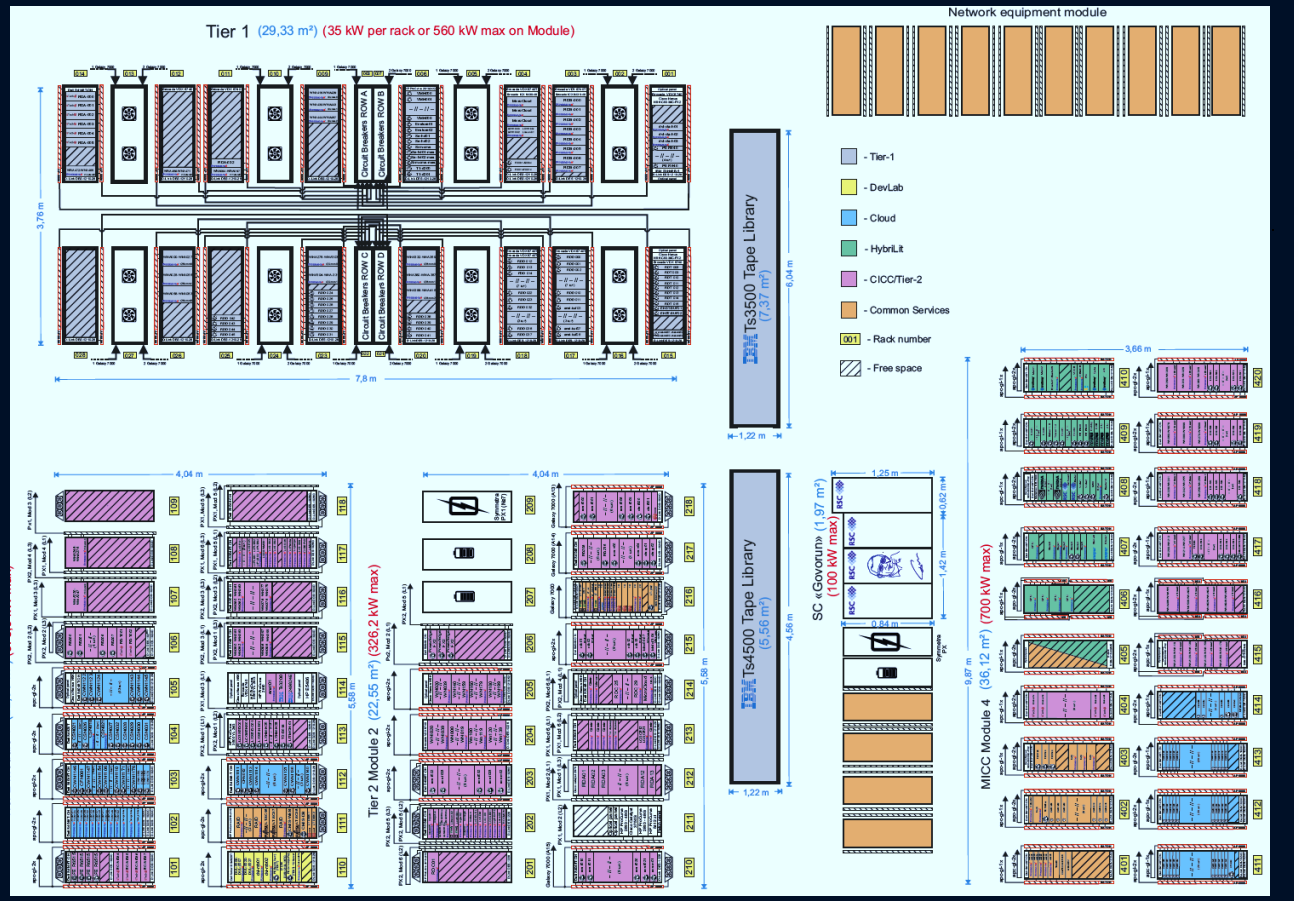
Информационно-вычислительный блок комплекса NICA в ОИЯИ включает в себя:
1. онлайн-кластер NICA,
2. автономный кластер NICA в ЛФВЭ,
3. все компоненты МИВК (Tier0, Tier1, Tier2, суперкомпьютер «Говорун», облачные вычисления),
4. многоуровневую систему хранения данных,
5. распределенную вычислительную сеть.

NICA Tier 0,1,2	2024	2025	2026	2027	2028	2029	2030
CPU (PFlops)	2.2	2.6	8.6	8.6	15.6	15.6	15.6
DISK (PB)	17	24	47	75	96	119	142
TAPE (PB)	45	88	170	226	352	444	536
NETWORK (Gbps)	400	400	800	800	800	1000	1000

Развитие серверных залов МИВК

Сегодня (1000 кВт)

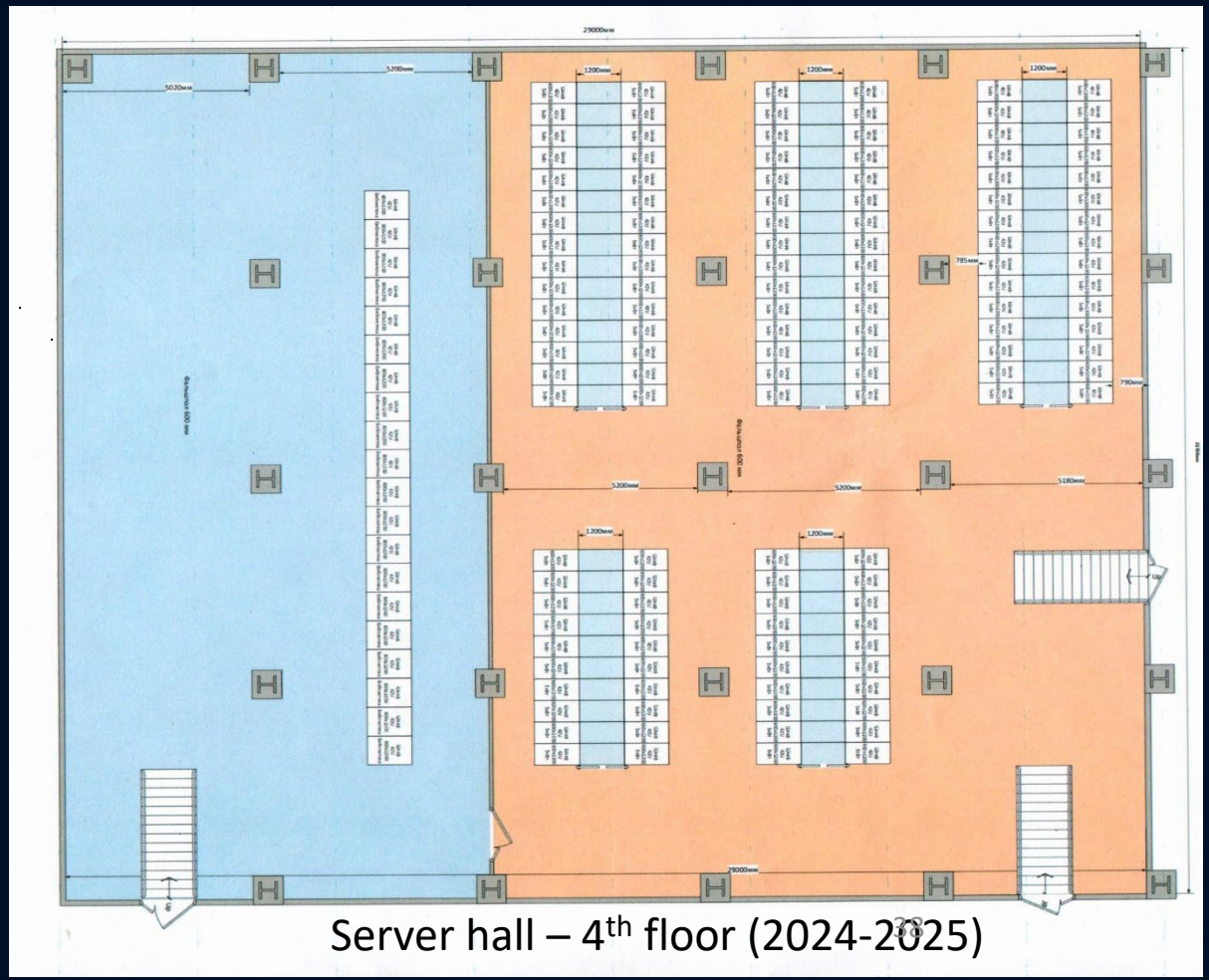
- 69 стоек для серверов
- 4 стойки для СК «Говорун»
- 10 стоек для сетевого оборудования
- 4 стойки для административных сервмсов
- 2 роботизированные ленточные библиотеки



Server hall – 2nd floor (2023)

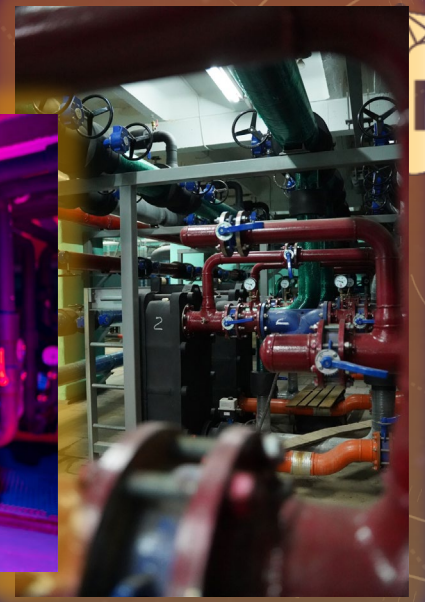
Планируем – новый серверный зал МИВК (600 кВт)

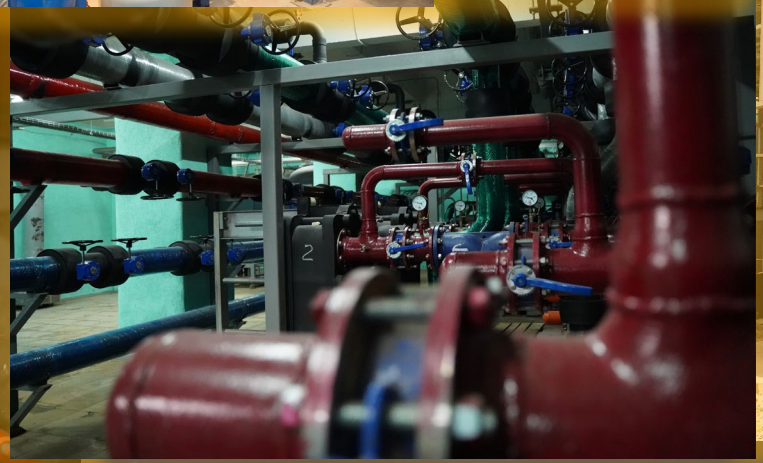
- зона роботизированных ленточных библиотек
- 130 стоек для серверов



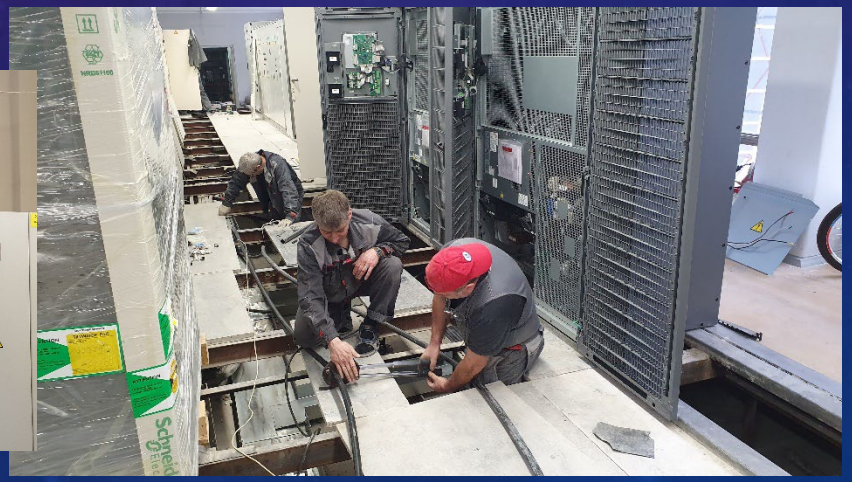
Server hall – 4th floor (2024-2025)

Охлаждение





ЭНЕРГООБЕСПЕЧЕНИЕ



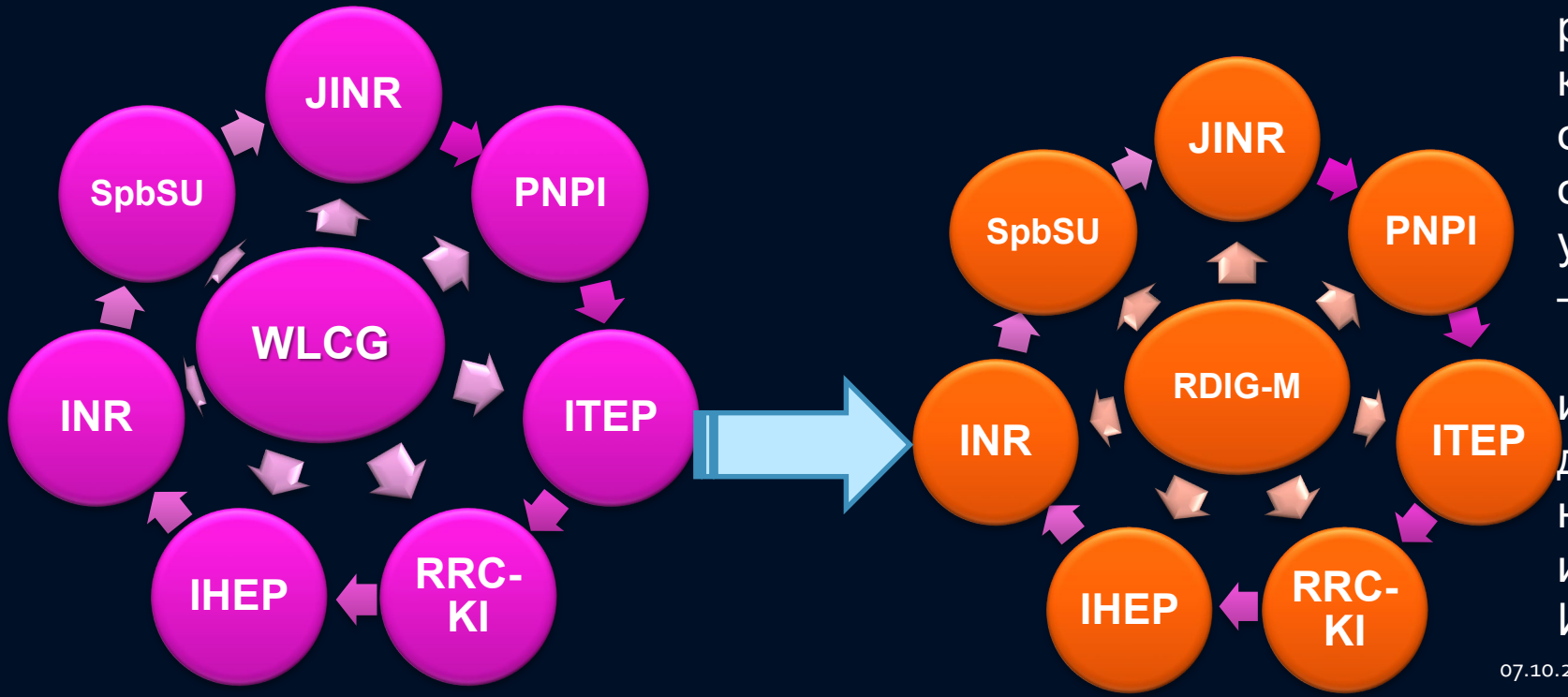
От РДИГ к РДИГ-М

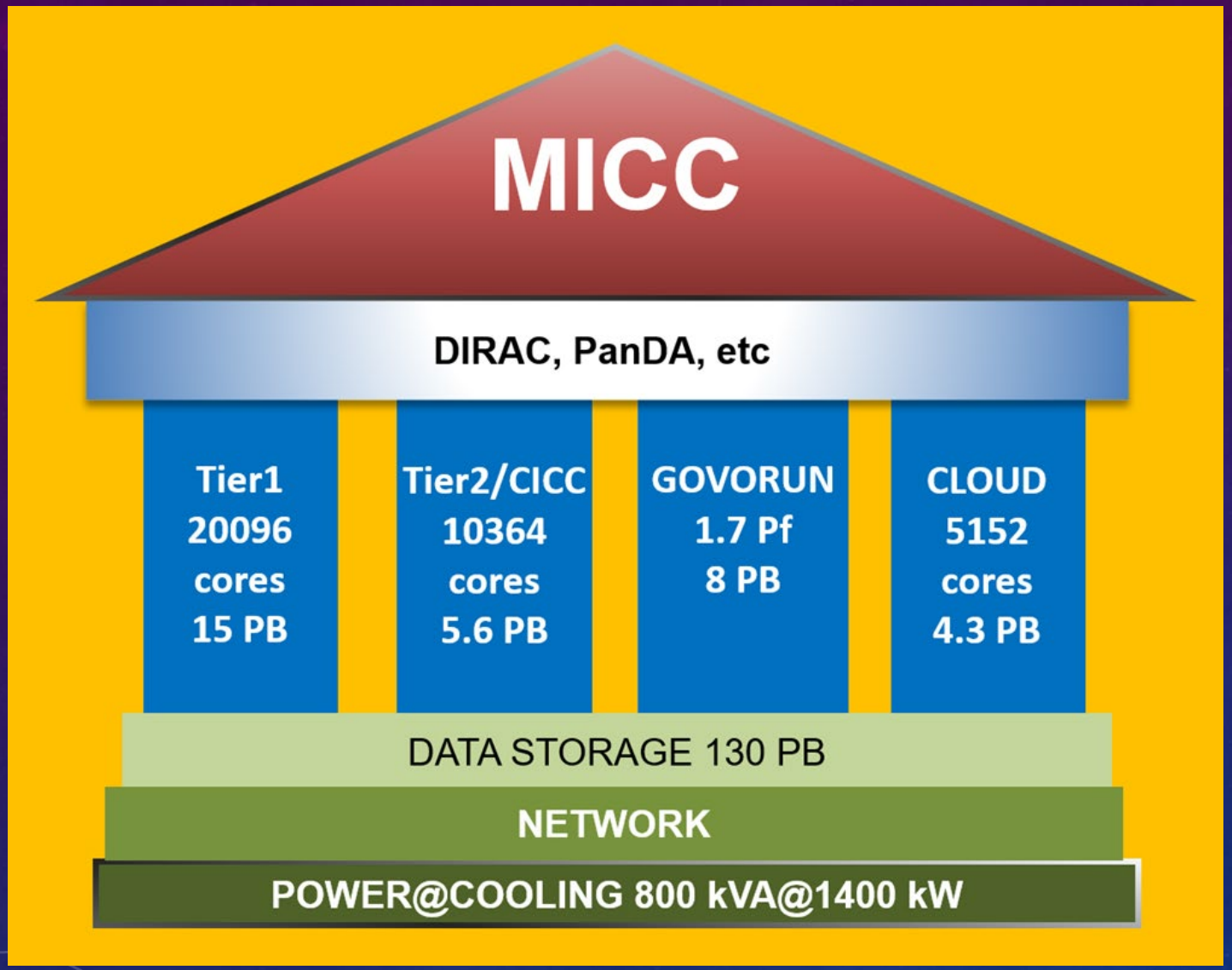
Российский консорциум РДИГ (Российский Data Intensiv GRID) был создан в сентябре 2003 года как национальная федерация проекта EGEE.

Протокол между ЦЕРН, Россией и ОИЯИ об участии в проекте LCG был подписан в 2003 году. Меморандум о взаимопонимании об участии в проекте WLCG был подписан в 2007 году.

В России реализуется программа масштабных научных проектов. Для решения этой задачи необходимо развитие распределенной компьютерной инфраструктуры, объединяющей ключевые научные и образовательные учреждения, участвующие в меганаучных проектах, – РДИГ-М.

Ядром ИТ-поддержки такой исследовательской инфраструктуры должен стать созданный в 2024 году на базе ОИЯИ, НИЦ «Курчатовский институт», ИСП РАН консорциум по ИТ-поддержке мегасайенс-проектов.





Основная цель проекта — обеспечить

- многофункциональность,
- масштабируемость,
- высокую производительность,
- надежность и доступность в режиме 24x7x365

для различных групп пользователей, выполняющих научные исследования в рамках Тематического плана ОИЯИ.

EN

МИВК

МНОГОФУНКЦИОНАЛЬНЫЙ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫЙ КОМПЛЕКС ОИЯИ

Облачная инфраструктура

Вам нужно больше компьютеров для исследований?

Создайте их в нашем облачном веб интерфейсе. Выберите необходимое Вам количество ядер, ОЗУ и операционную систему для своих целей.

Гетерогенная платформа

Нужны параллельные преимущества современных графических ускорителей?

Используйте 1000 ядер в один момент, чтобы получить результаты так быстро, как это возможно.

Грид-инфраструктура

Нужен анализ данных экспериментов БАК?

Получите доступ к нашему грид кластеру для выполнения анализа.

ЦИВК

Нужны ресурсы для длительных вычислений?

ЦИВК - это набор серверов, которые вы можете загружать своими задачами. Чтобы использовать параллельные функции фермы, используйте MPI задачи.

https://micc.jinr.ru/

Спасибо за внимание