

Интеграция вычислительных ресурсов



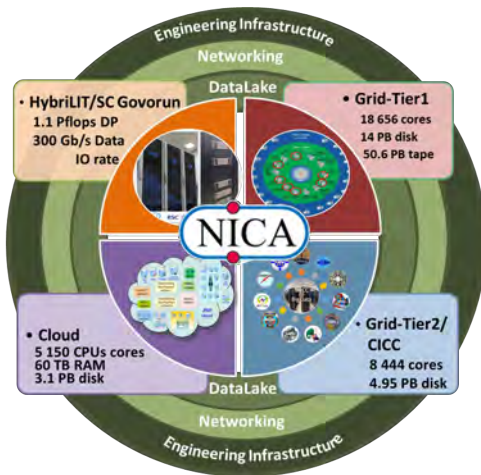
Научный сотрудник

Игорь Пелеванюк

Лаборатория
информационных технологий
им. М.Г. Мещерякова

Вычислительные ресурсы

MICC



NICA cluster



JINR Member States



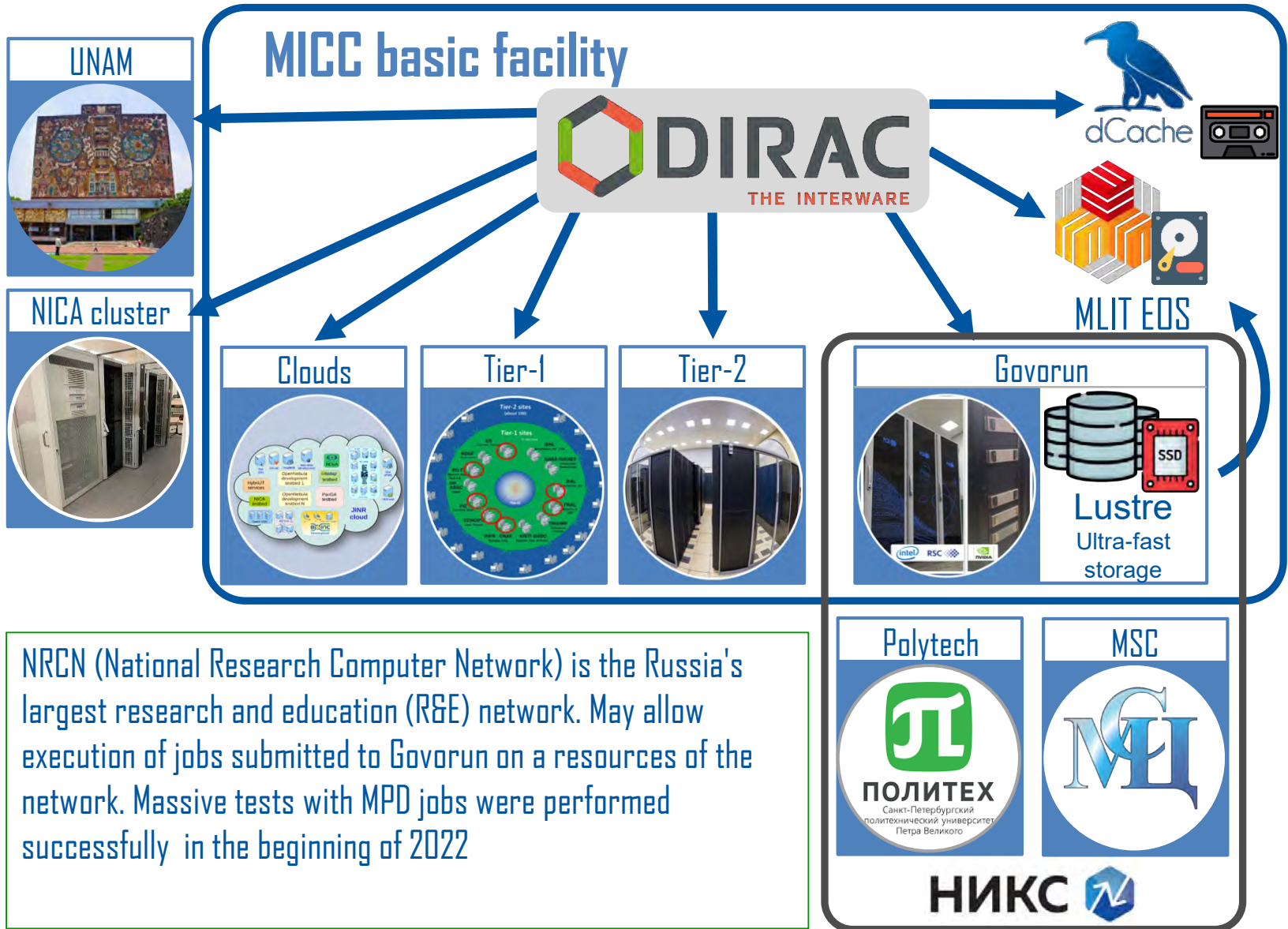
Others...

Зачем интегрировать ресурсы

Большое количество однотипных вычислительных задач - сотни тысяч задач, сотни CPU-лет в год.

1. Запуск таких объёмов на любом из доступных ресурсов приведёт к одному из двух исходов:
 - Задачи будут считаться слишком долго
 - Вычислительный ресурс будет недоступен для других пользователей
2. Ручной запуск на разных ресурсах превратится в большую проблему для ответственного:
 - Придётся эмпирически делить нагрузку между ресурсами
 - Входные данные придётся вручную распределять по разным ресурсам. Сбор результатов так же потребует дополнительных усилий.
 - Вопросы использования новых появившихся ресурсов – это на 100% проблема конечного пользователя. Ему придётся быть в курсе изменений на всех используемых ресурсах.
 - В случае появления проблем на ресурсах или обнаружения ошибок в запускаемых задачах пользователю потребуется перезапускать или перенаправлять нагрузку в «ручном режиме»

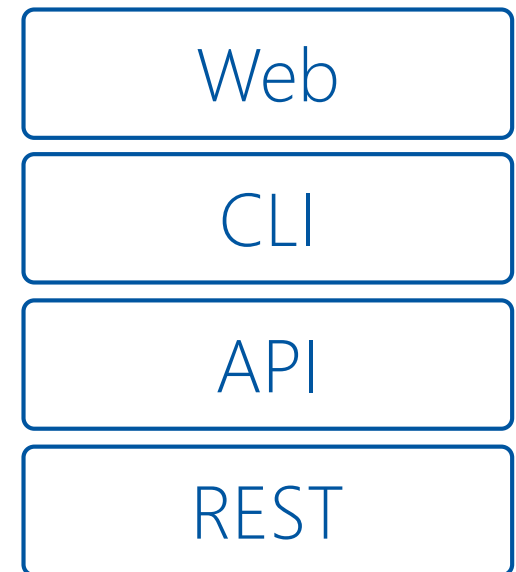
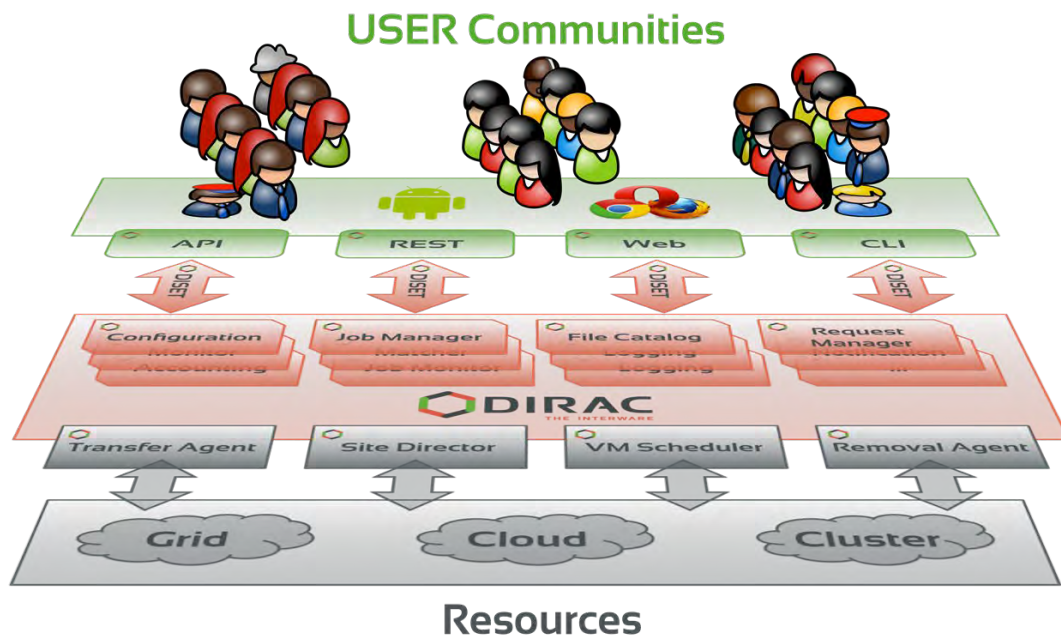
Что было сделано



NRCN (National Research Computer Network) is the Russia's largest research and education (R&E) network. May allow execution of jobs submitted to Govorun on a resources of the network. Massive tests with MPD jobs were performed successfully in the beginning of 2022

What is DIRAC?

DIRAC provides all the necessary components to build ad-hoc grid infrastructures **interconnecting** computing resources of different types, allowing **interoperability** and simplifying **interfaces**. This allows to speak about the DIRAC *interware*.



Birth of DIRAC

- LHC experiments, all developed their own middleware
 - PanDA, AliEn, glideIn WMS, PhEDEx, ...
- DIRAC is developed originally for the LHCb experiment
- The experience collected with a production grid system of a large HEP experiment is very valuable
 - Several new experiments expressed interest in using this software relying on its proven in practice utility
- In 2009 the core DIRAC development team decided to generalize the software to make it suitable for any user community.
- The results of this work allow to offer DIRAC as a general purpose distributed computing framework

Why DIRAC?

1. Single system for all aspects of computing



User Interface

API

Central configuration

Workload management

Data management

Integration tools

File Catalog

Workflow management

Metadata management

Accounting

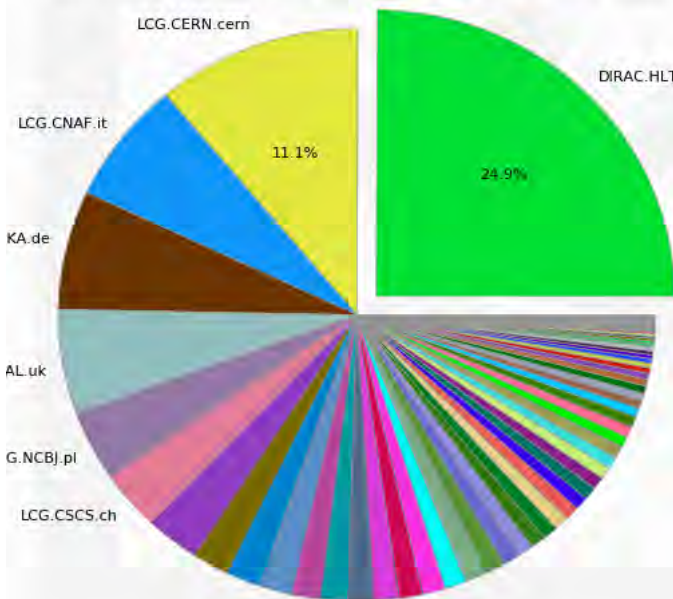
Management

Why DIRAC?

2. Good performance

Wall time days used by Site

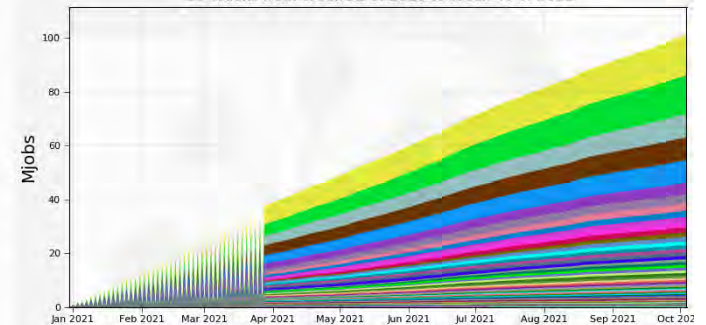
39 Weeks from Week 00 of 2021 to Week 40 of 2021



DIRAC.HLTFarm.lhcb	8735245.1
LCG.CERN.cern	3895823.4
LCG.CNAF.it	2421484.2
LCG.GRIDKA.de	2353391.6
LCG.RAL.uk	2061729.7
LCG.NCBJ.pl	1390750.6
LCG.CSCS.ch	1172944.7
LCG.IN2P3.fr	1042061.3
LCG.Manchester.uk	689467.2
LCG.RRCKI.ru	663460.3
LCG.UKI-LT2-QMUL.uk	612525.1
LCG.MIT.us	549385.2
LCG.UKI-LT2-IC-HEP.uk	515392.8
LCG.LAL.fr	490277.7
LCG.NIKHEF.nl	474998.9
DIRAC.Client.fr	446574.6
DIRAC.Client.ch	435127.7
DIRAC.Client.de	408990.6
LCG.CPPM.fr	407705.7
LCG.NIPNE-07.ro	380158.5
LCG.Lancaster.uk	332407.3
LCG.CBPF.br	291745.9
DIRAC.Client.it	286298.3
LCG.Liverpool.uk	284122.6
LCG.JINR.ru	283145.2
LCG.LAPP.fr	273415.1
LCG.SARA.nl	255426.3
DIRAC.Client.nl	248673.9
LCG.RAL-HEP.uk	244129.9
LCG.Beijing.cn	241335.1
LCG.DURHAM.uk	235125.9
LCG.UKI-LT2-RHUL.uk	233936.2
LCG.PIC.es	214956.9
DIRAC.UZH.ch	212277.2
LCG.LPNHE.fr	211377.3
LCG.BEER.cern	178553.2
LCG.CINECA.it	152835.0
LCG.IHEP.ru	150573.2
LCG.Bristol.uk	149147.8
LCG.LPC.fr	131954.5
VAC.Glasgow.uk	121261.9
LCG.Oxford.uk	106505.4
...	plus 49 more

Generated on 2021-10-05 16:03:06 UTC

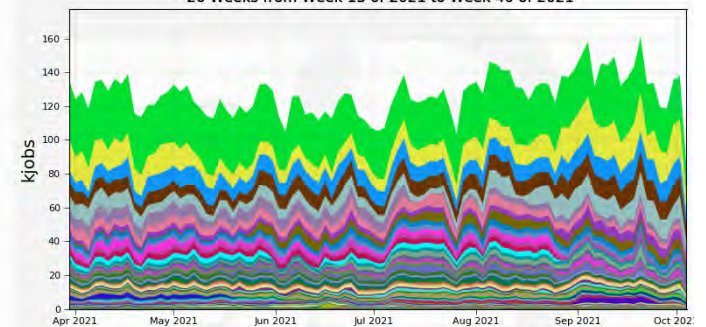
Cumulative Jobs by Site
39 Weeks from Week 52 of 2020 to Week 40 of 2021



Max: 101, Min: 0.00, Average: 41.1, Current: 101

LCG.CERN.cern	15.2	LCG.NIKHEF.nl	2.1	LCG.SARA.nl	1.2	LCG.LAPP.fr	0.7
DIRAC.HLTFarm.lhcb	14.3	DIRAC.Client.ch	1.8	LCG.CPPM.fr	1.1	LCG.RAL-HEP.uk	0.7
LCG.RAL.uk	8.7	DIRAC.Client.fr	1.8	DIRAC.Client.nl	1.0	LCG.CBPF.br	0.6
LCG.GRIDKA.de	8.5	LCG.Manchester.uk	1.7	LCG.PIC.es	1.0	LCG.UKI-LT2-RHUL.uk	0.6
LCG.CNAF.it	8.4	LCG.UKI-LT2-QMUL.uk	1.6	LCG.Lancaster.uk	1.0	LCG.Liverpool.uk	0.6
LCG.IN2P3.fr	4.3	DIRAC.Client.de	1.4	LCG.Beijing.cn	0.9	LCG.DURHAM.uk	0.6
LCG.NCBJ.pl	3.4	LCG.UKI-LT2-IC-HEP.uk	1.4	DIRAC.Client.it	0.9	DIRAC.UZH.ch	0.6
LCG.CSCS.ch	2.8	LCG.MIT.us	1.3	LCG.NIPNE-07.ro	0.9	LCG.LPNHE.fr	0.5
LCG.RRCKI.ru	2.3	LCG.LAL.fr	1.3	LCG.JINR.ru	0.8	...	plus 56 more

Running jobs by Site
26 Weeks from Week 13 of 2021 to Week 40 of 2021



Max: 161, Min: 81.9, Average: 127, Current: 81.9

DIRAC.HLTFarm.lhcb	24.2%	LCG.RRCKI.ru	2.0%	LCG.NIKHEF.nl	1.2%
LCG.CERN.cern	9.9%	LCG.UKI-LT2-QMUL.uk	1.8%	LCG.NIPNE-07.ro	1.2%
LCG.CNAF.it	7.3%	DIRAC.Client.ch	1.8%	LCG.Lancaster.uk	1.2%
LCG.GRIDKA.de	6.5%	LCG.MIT.us	1.6%	DIRAC.Client.it	1.1%
LCG.RAL.uk	5.9%	DIRAC.Client.fr	1.6%	LCG.CBPF.br	0.9%
LCG.NCBJ.pl	3.8%	LCG.Client.de	1.4%	DIRAC.Client.nl	0.9%
LCG.CSCS.ch	3.7%	LCG.UKI-LT2-IC-HEP.uk	1.4%	LCG.Liverpool.uk	0.9%
LCG.IN2P3.fr	2.6%	LCG.CPPM.fr	1.4%	LCG.LAPP.fr	0.8%
LCG.Manchester.uk	2.5%	LCG.LAL.fr	1.3%	...	plus 56 more

Generated on 2021-10-05 16:13:38 UTC

Why DIRAC?

3. Active users and developers community



GridPP
UK Computing for Particle Physics



- ▶ Dedicated installations
 - ▶ LHCb, Belle II, CTA
- ▶ Multi-community services
 - ▶ ILC, CALICE
 - ▶ IHEP: BES III, Juno, CEPC
 - ▶ FG-DIRAC
 - ▶ GridPP
 - ▶ DIRAC4EGI
 - ▶ PNNL

▶ **DIRAC@JINR**

▶ DIRAC@CNAF

- ▶ Several DIRAC evaluations are ongoing
 - ▶ Auger, ELI, NICA, Virgo, LSST, ...



DIRAC standard job workflow

1. Initial configuration
2. Input data download
3. Processing
4. Output data upload
5. Finalization

Job example

```
dirac-configure config.cfg
```

```
dirac-dms-get-file /bmn/raw/AuAu/data1002.raw
```

```
root -l -q -b reco.MC("./data1002.raw")
```

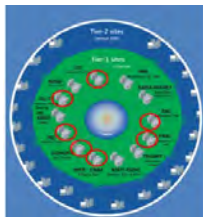
```
dirac-dms-put-file /bmn/reco/AuAu/data1002.root \\  
                                                           data1002.root \\  
                                                           JINR-EOS-BMN
```

```
rm -f data1002.raw data1002.root
```

Workload management



Submit thousand of jobs to DIRAC Job Queue



Tier-1



CICC/Tier-2



Clouds



Govorun



NICA Cluster

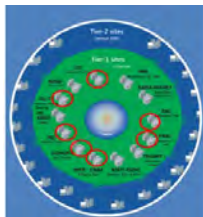


UNAM

Workload management



Submit thousand of jobs to DIRAC Job Queue



Tier-1



CICC/Tier-2



Clouds



Govorun



NICA Cluster

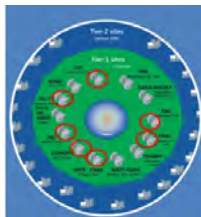


UNAM

Workload management



Submit thousand of jobs to DIRAC Job Queue



Tier-1



CICC/Tier-2



Clouds



Govorun

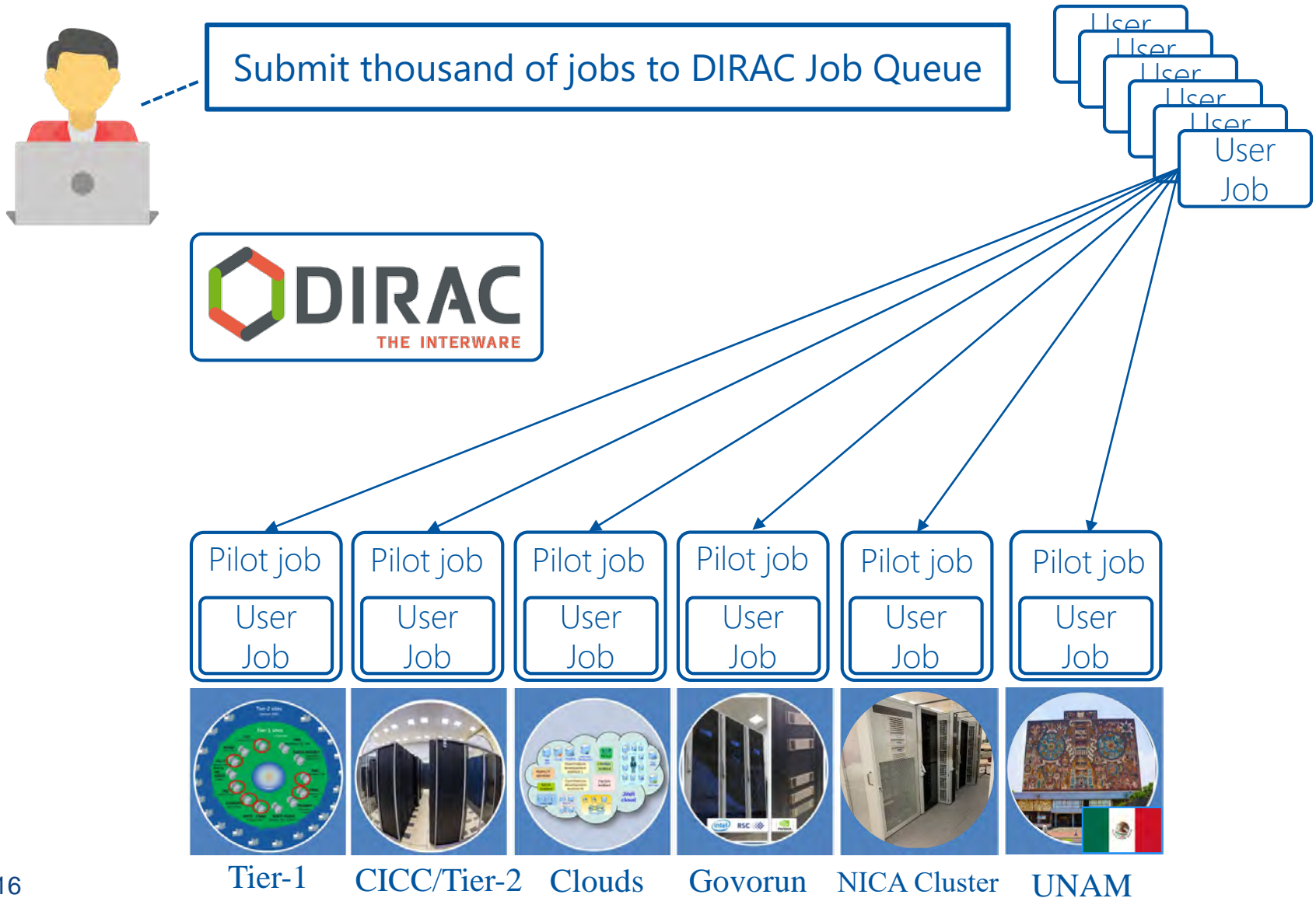


NICA Cluster



UNAM

Workload management



History of DIRAC at JINR

2013 – Development of monitoring system for BES-III installation. First tries to setup and configure DIRAC infrastructure.

2017 – DIRAC Interware installed; basic configuration done. Used for educational purposes. **dCache** storage integrated, **Tier2** integrated.

2018 – **HybriLIT** integrated. **JINR cloud** integrated using OCCl protocol. Tests of full cycle of Monte-Carlo for **BM@N** were performed.

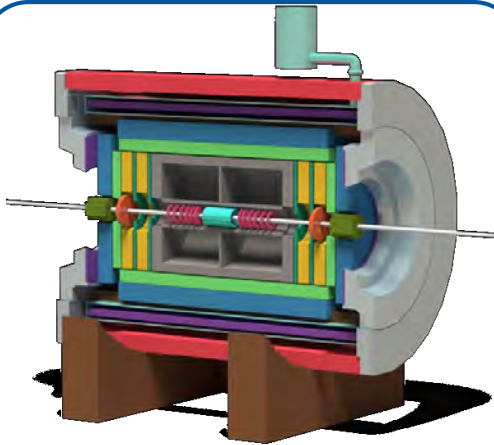
2019 – **Clouds** of JINR Member-States integrated by module developed in JINR. **MPD** starts using DIRAC for massive Monte-Carlo production. **Tier1**, **Govoron** and **EOS** integrated in DIRAC.

2020 – **Folding@Home** jobs submitted to clouds via DIRAC. **Baikal-GVD** jobs submitted to JINR and PRUE clouds.

2021 – First tests for **SPD** Monte-Carlo successfully done. First million jobs done!

2022– Total walltime exceeds 1000 years. DIRAC in JINR updated to use Python 3.

What do we use DIRAC for?



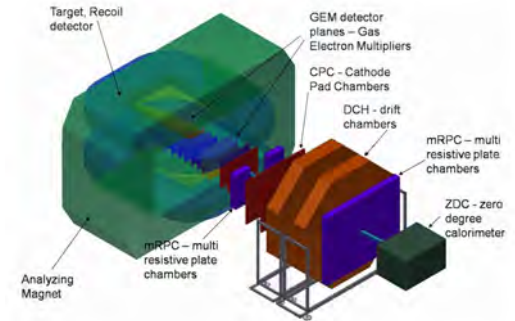
MPD@NICA

Monte-Carlo – Real
Analysis – Maybe



Baikal-GVD

Monte-Carlo – Real



BM@N

Monte-Carlo – Real
Reconstruction – Real



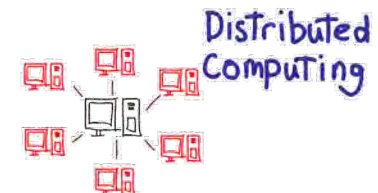
SPD@NICA

Monte-Carlo – Real



FOLDING
@HOME

Folding@HOME

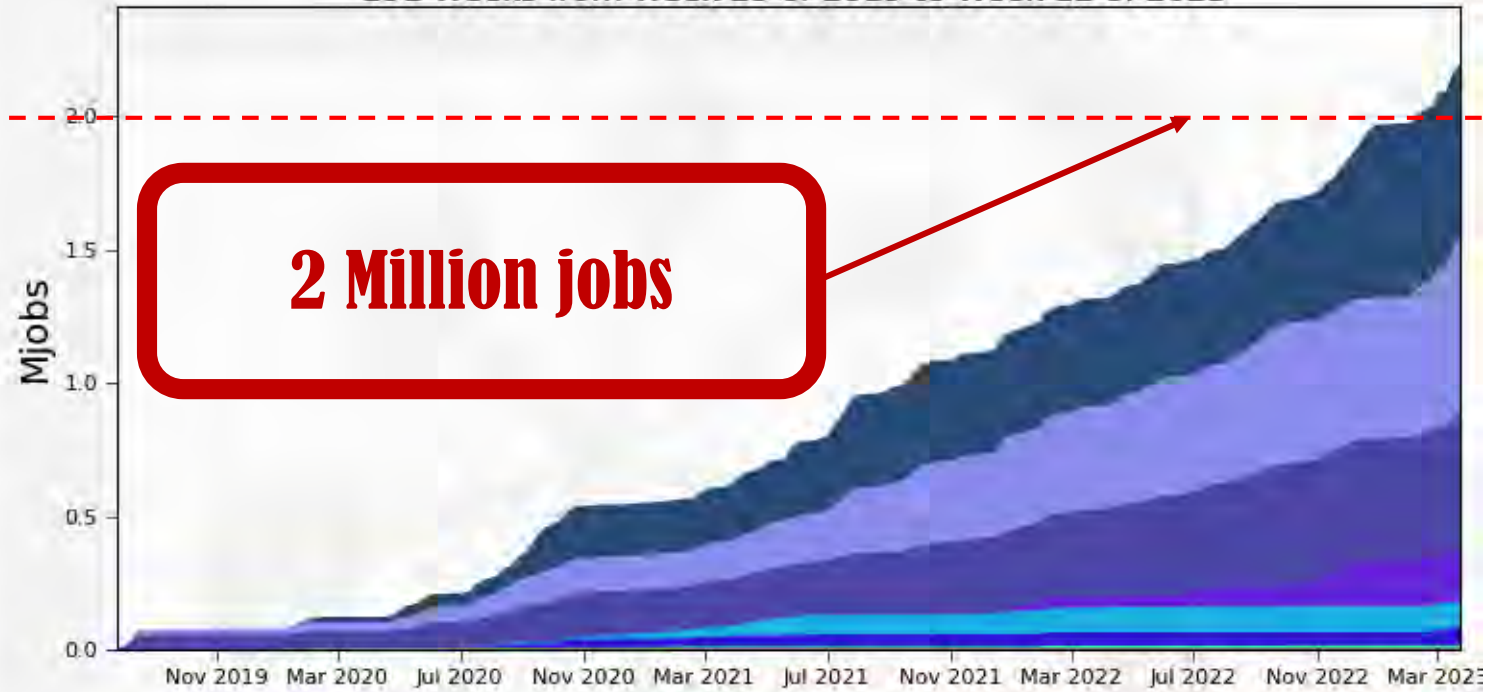


Teaching

Statistics: jobs done

Cumulative Jobs by Site

191 Weeks from Week 29 of 2019 to Week 12 of 2023



Max: 2.19, Min: 0.00, Average: 0.85, Current: 2.19

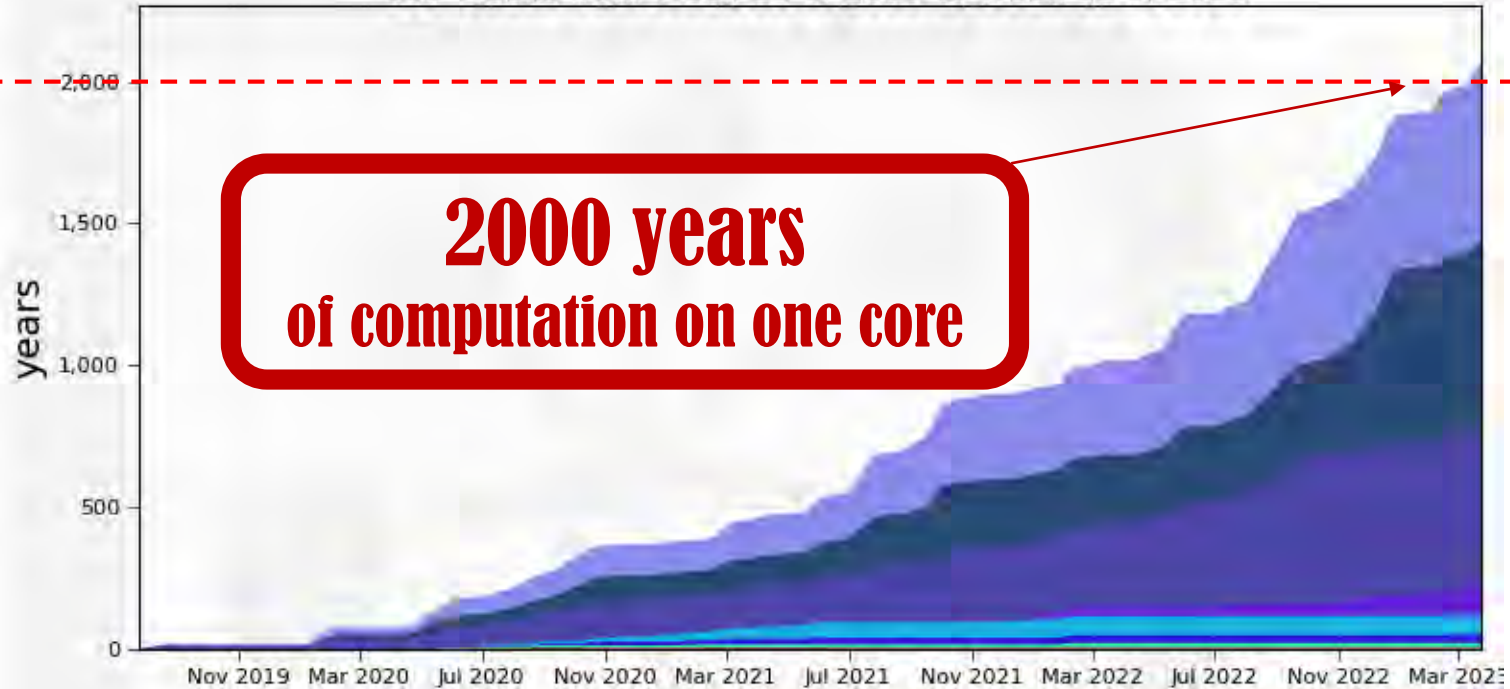
DIRAC.GOVORUN.ru	0.7	DIRAC.UNAM.mx	0.0	CLOUD.NU.kz	0.0
DIRAC.JINR-TIER.ru	0.7	CLOUD.IPANAS.az	0.0	CLOUD.INRNE.bg	0.0
DIRAC.JINR-CREAM.ru	0.5	DIRAC.INP.uz	0.0	DIRAC.SSH.ru	0.0
DIRAC.JINR-CONDOR.ru	0.2	CLOUD.INP.by	0.0	DIRAC.IMDT.mn	0.0
CLOUD.JINR.ru	0.1	CLOUD.STI-SCI.eg	0.0	CLOUD.JINR-JUNO.ru	0.0
DIRAC.JINR-LHEP.ru	0.1	DIRAC.JINR-SANC.ru	0.0	DIRAC.TEST.ru	0.0
CLOUD.PRUE.ru	0.0	CLOUD.INP.kz	0.0	DIRAC.UPJS.sk	0.0
DIRAC.NIKS-JSCC.ru	0.0	CLOUD.GRENA.ge	0.0		
CLOUD.NOSU.ru	0.0	DIRAC.REA.ru	0.0		

Generated on 2023-03-27 22:35:12 UTC

Statistics: walltime

Cumulative wall time by Site

191 Weeks from Week 29 of 2019 to Week 12 of 2023



Max: 2,062, Min: 0.01, Average: 696, Current: 2,062

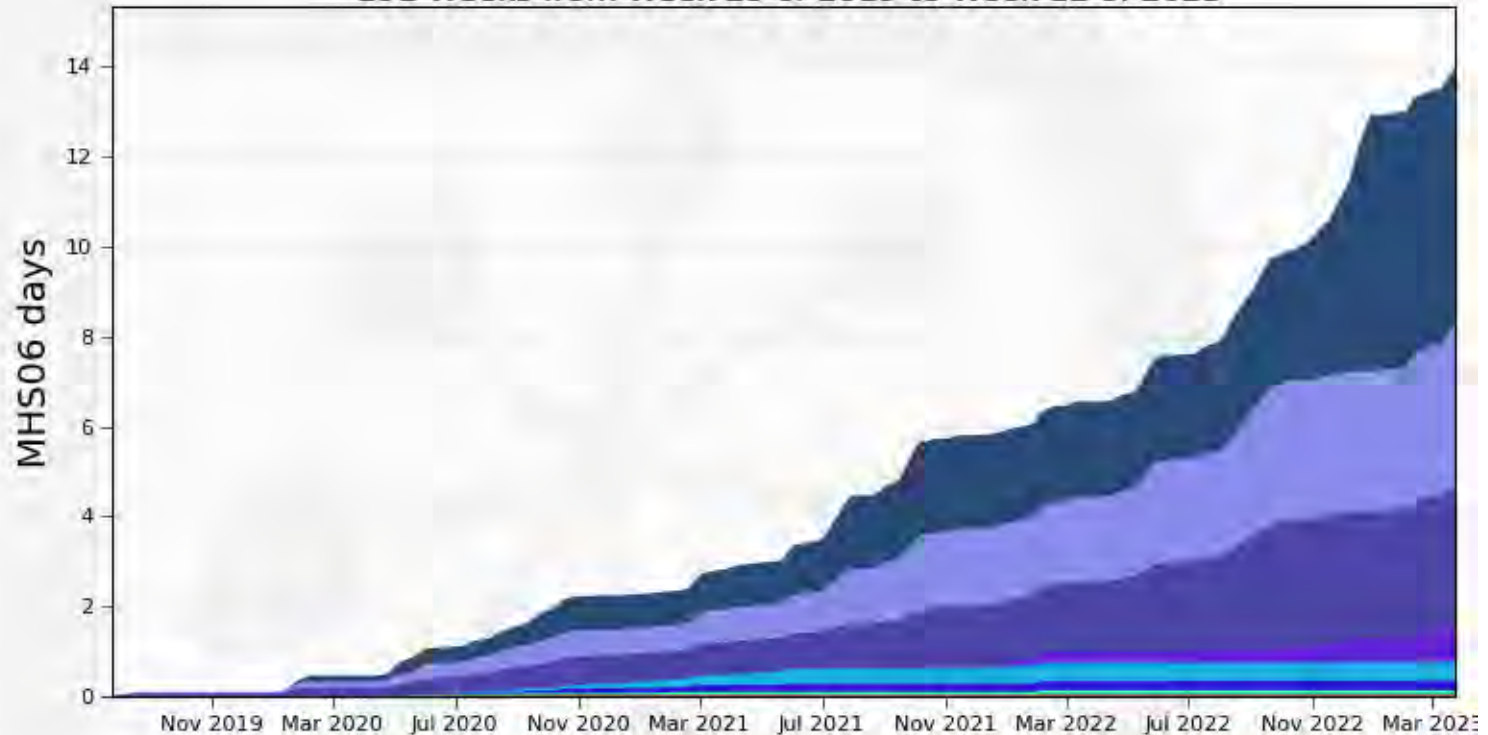
DIRAC.JINR-TIER.ru	629.2	CLOUD.IPANAS.az	2.0	CLOUD.GRENA.ge	0.0
DIRAC.GOVORUN.ru	624.4	DIRAC.UNAM.mx	1.3	DIRAC.JINR-SANC.ru	0.0
DIRAC.JINR-CREAM.ru	585.1	CLOUD.INP.by	1.0	DIRAC.IMDT.mn	0.0
DIRAC.JINR-CONDOR.ru	103.7	CLOUD.STI-SCI.eg	0.7	DIRAC.UPJS.sk	0.0
CLOUD.JINR.ru	68.5	CLOUD.INP.kz	0.2	DIRAC.SSH.ru	0.0
DIRAC.JINR-LHEP.ru	29.2	CLOUD.INRNE.bg	0.2	DIRAC.TEST.ru	0.0
CLOUD.PRUE.ru	9.2	DIRAC.REA.ru	0.2	CLOUD.JINR-JUNO.ru	0.0
DIRAC.NIKS-JSCC.ru	4.9	DIRAC.INP.uz	0.0		
CLOUD.NOSU.ru	2.3	CLOUD.NU.kz	0.0		

Generated on 2023-03-27 22:37:29 UTC

Statistics: normalized time

Normalized CPU used by Site

191 Weeks from Week 29 of 2019 to Week 12 of 2023



Max: 13.9, Min: 0.00, Average: 4.49, Current: 13.9

DIRAC.GOVORUN.ru	5.7	DIRAC.NIKS-JSCC.ru	0.0	CLOUD.INP.kz	0.0	DIRAC.IMDT.mn	0.0
DIRAC.JINR-TIER.ru	3.6	CLOUD.NOSU.ru	0.0	DIRAC.INP.uz	0.0	DIRAC.SSH.ru	0.0
DIRAC.JINR-CREAM.ru	3.1	CLOUD.IPANAS.az	0.0	CLOUD.NU.kz	0.0	DIRAC.TEST.ru	0.0
DIRAC.JINR-CONDOR.ru	0.8	DIRAC.UNAM.mx	0.0	CLOUD.GRENA.ge	0.0	CLOUD.JINR-JUNO.ru	0.0
CLOUD.JINR.ru	0.4	CLOUD.INP.by	0.0	DIRAC.JINR-SANC.ru	0.0		
DIRAC.JINR-LHEP.ru	0.2	CLOUD.STI-SCL.eg	0.0	DIRAC.REA.ru	0.0		
CLOUD.PRUE.ru	0.1	CLOUD.INRNE.bg	0.0	DIRAC.UPJS.sk	0.0		

Generated on 2023-03-27 22:38:43 UTC

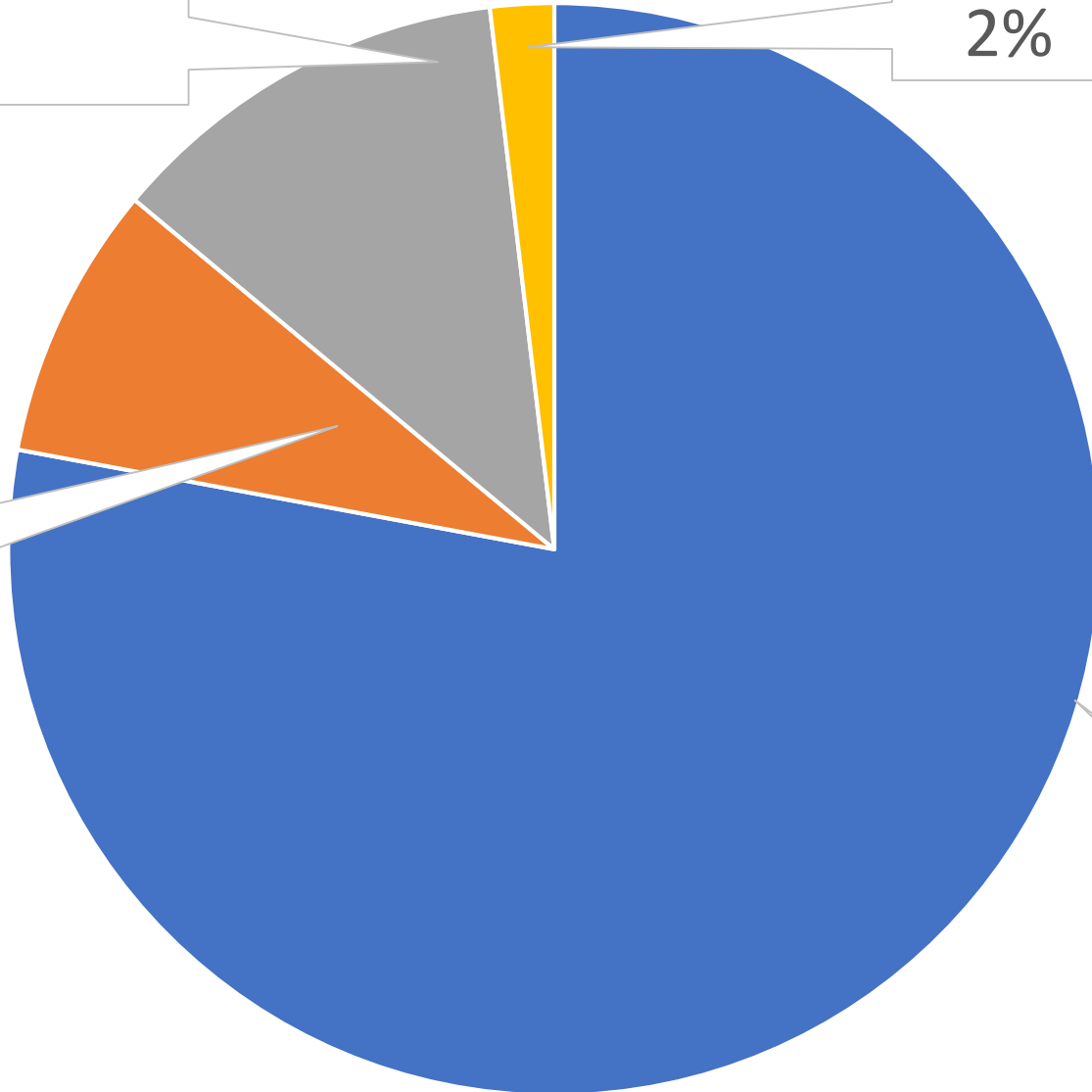
Ratio: between experiments

SPD
12%

BM@N
2%

Baikal-GVD
8%

MPD
78%



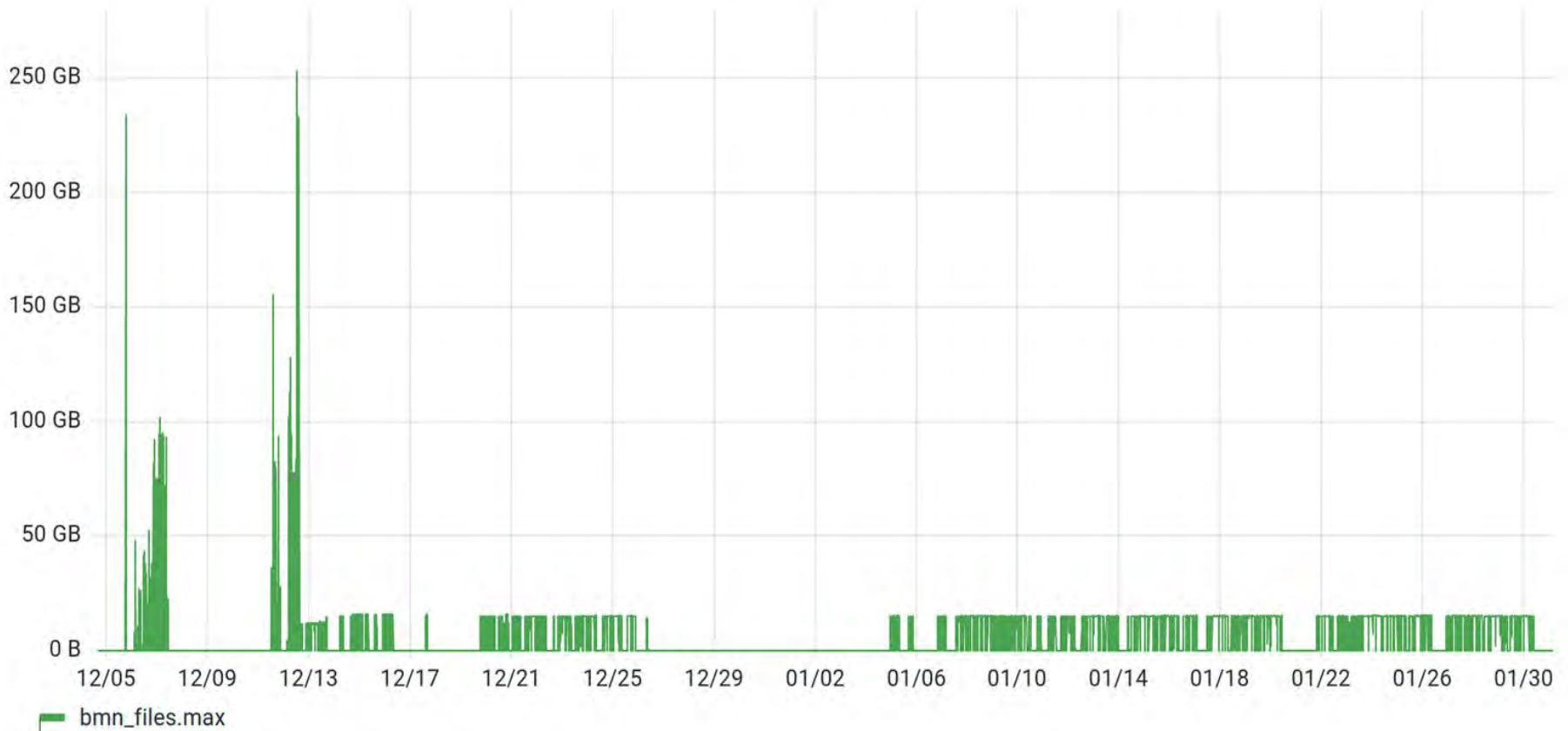
Example:

BM@N run8 raw data processing (with Konstantin Gertsenberger)

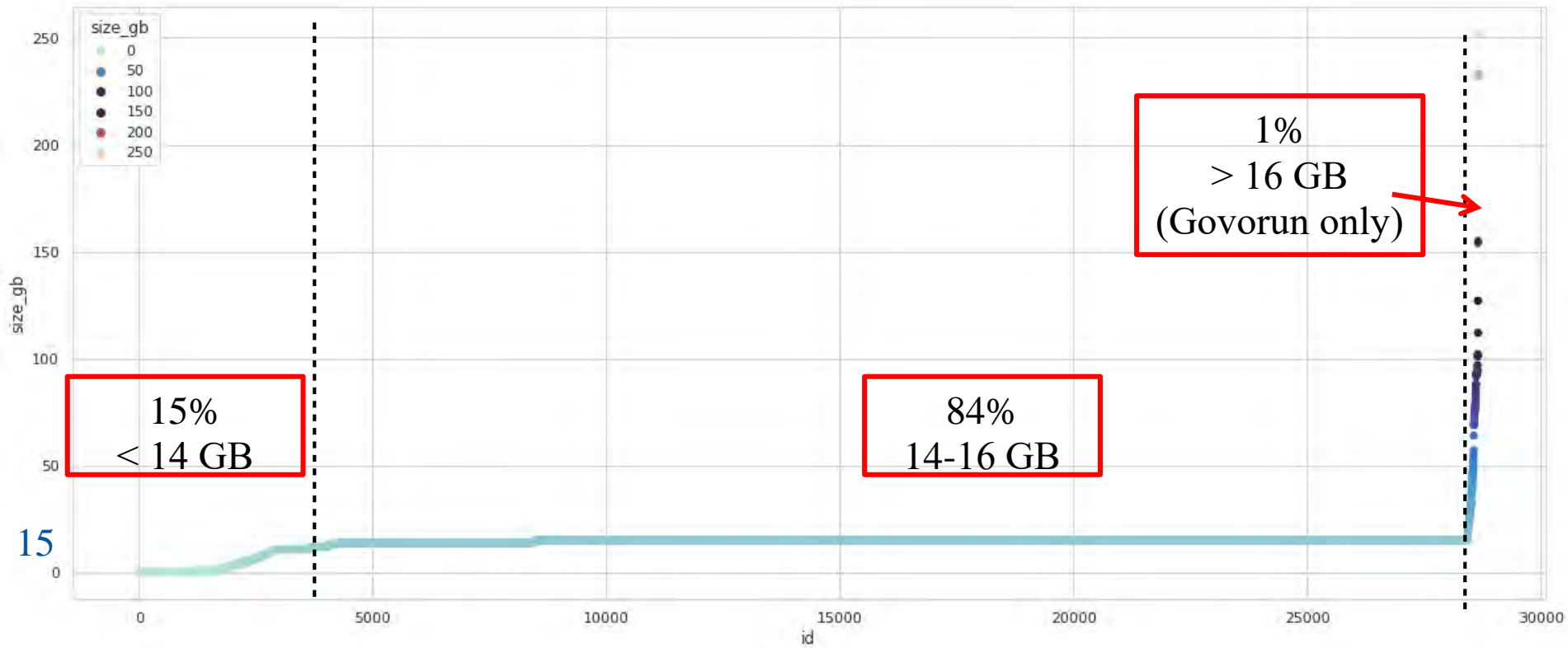
Total files ~ **30000**

Total raw size ~ **400TB**

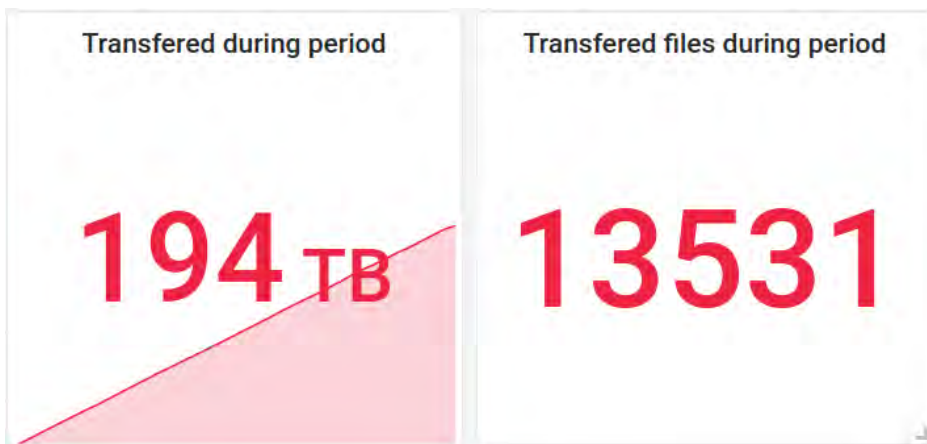
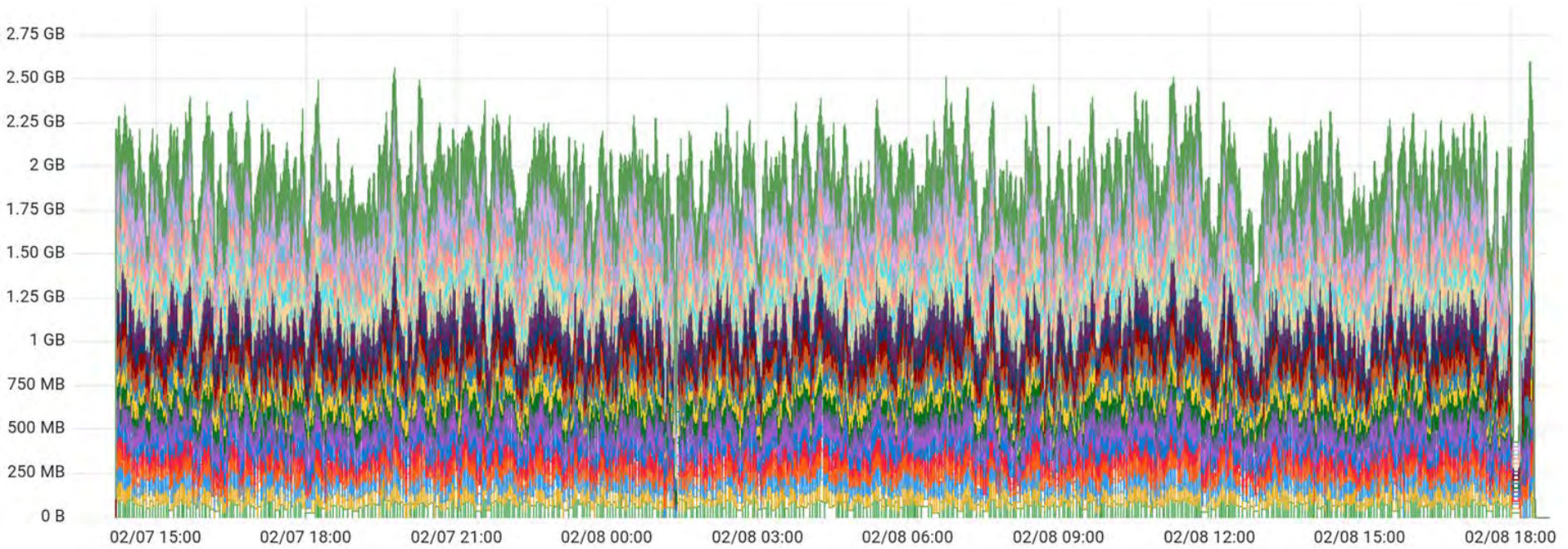
File size



BM@N run8 raw data size

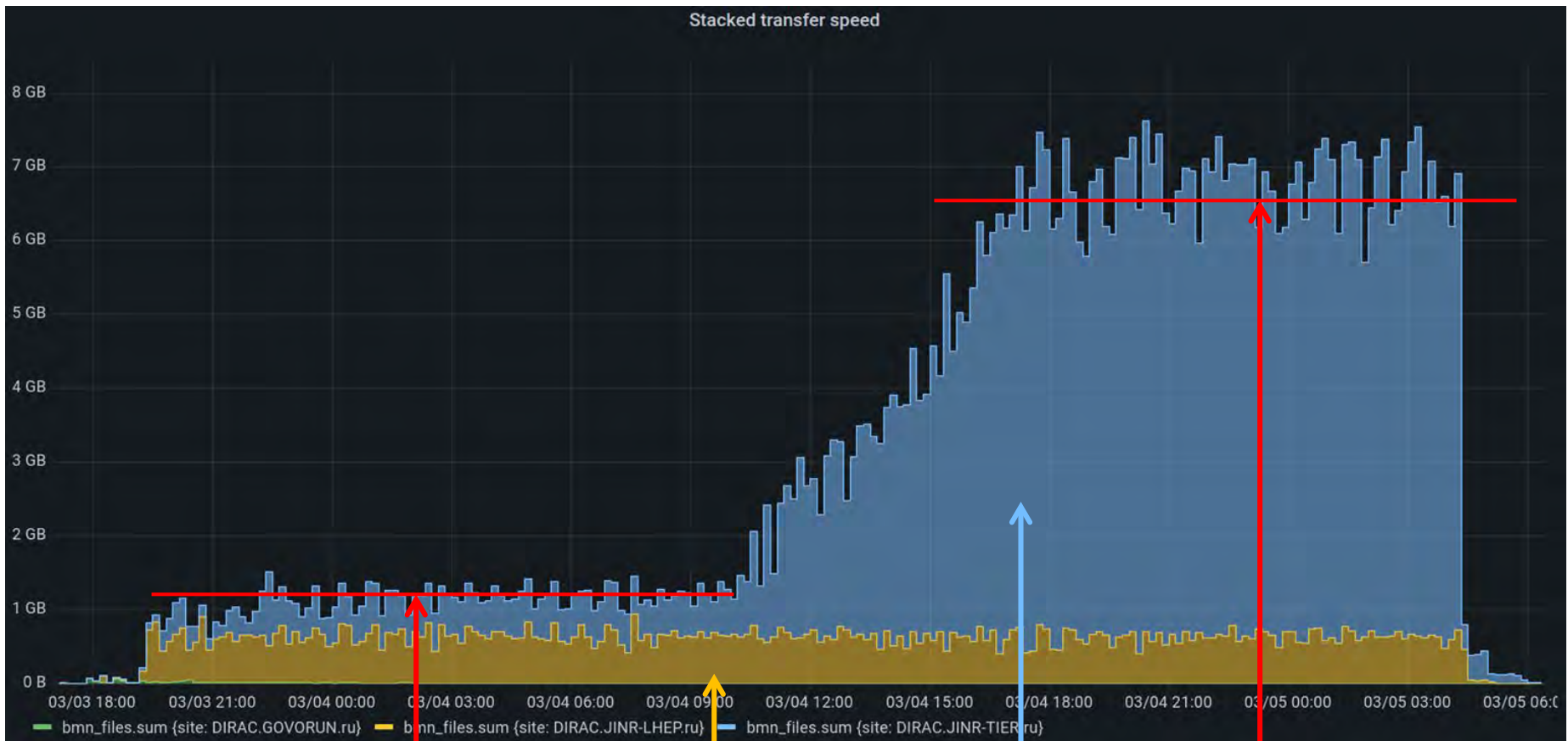


EOS LHEP -> EOS LIT



Average transfer speed on 20 streams
1.92 GB/s

BM@N Digi->Dst



300 jobs running
4 MB/s per job

NICA Cluster

Tier1

1580 jobs running
4.1 MB/s per job

List of participants

DIRAC: Igor Pelevanyk, Andrey Tsaregorodtzev

Baikal-GVD: Dmitry Zaborov

BM@N: Konstantin Gertsenberger

MPD: Oleg Rogachevskiy, Andrey Moshkin

SPD: Alexey Zhemchugov, Katherin Shtejer

Responsible for resources:

Govorun: Dmitry Podgainy, Dmitry Belyakov, Aleksandr Kokorev, Maxim Zuev,

NICA cluster: Boris Schinov, Ivan Slepov

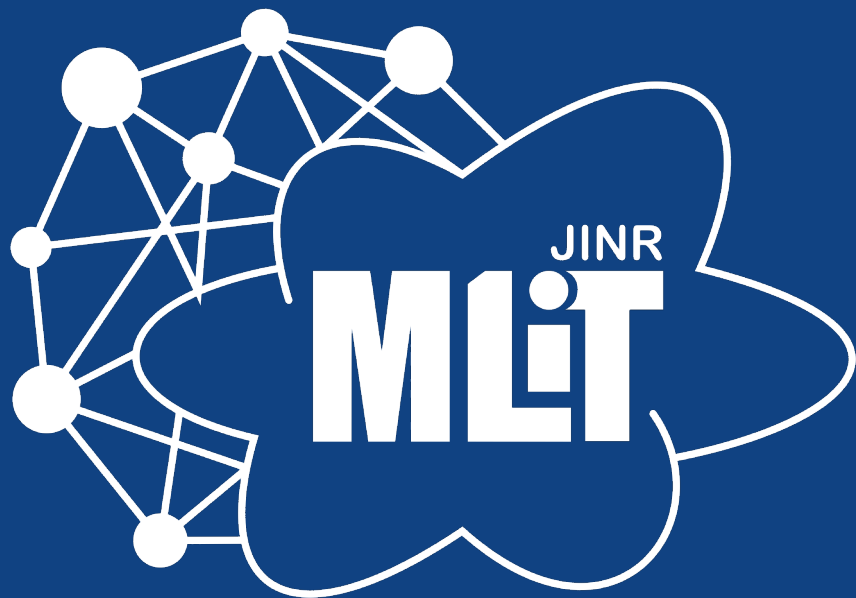
Tier-1, Tier-2, EOS: Valery Mitsyn

Cloud: Nikolay Kutovskiy, Nikita Balashov

dCache: Vladimir Trofimov

Detailed articles

1. Gergel, V., V. Korenkov, I. Pelevanyuk, M. Sapunov, A. Tsaregorodtsev, and P. Zrellov. 2017. **Hybrid Distributed Computing Service Based on the DIRAC Interware**.
2. Korenkov, V., Pelevanyuk, I. & Tsaregorodtsev, A. 2019, "**Dirac system as a mediator between hybrid resources and data intensive domains**", CEUR Workshop Proceedings, pp. 73.
3. Balashov, N.A., Kuchumov, R.I., Kutovskiy, N.A., Pelevanyuk, I.S., Petrunin, V.N. & Tsaregorodtsev, A.Y. 2019, "**Cloud integration within the DIRAC Interware**", CEUR Workshop Proceedings, pp. 256.
4. Korenkov, V., Pelevanyuk, I. & Tsaregorodtsev, A. 2020, **Integration of the JINR hybrid computing resources with the DIRAC interware for data intensive applications**.
5. Kutovskiy, N., Mitsyn, V., Moshkin, A., Pelevanyuk, I., Podgayny, D., Rogachevsky, O., Shchinov, B., Trofimov, V. & Tsaregorodtsev, A. 2021, "**Integration of Distributed Heterogeneous Computing Resources for the MPD Experiment with DIRAC Interware**", Physics of Particles and Nuclei, vol. 52, no. 4, pp. 835-841.
6. Pelevanyuk, I., "**Performance evaluation of computing resources with DIRAC interware**", AIP Conference Proceedings 2377, 040006 (2021)



Computing Resources

JINR



Tier-1
920 slots

CICC/Tier-2
1000 slots

Cloud
80 slots

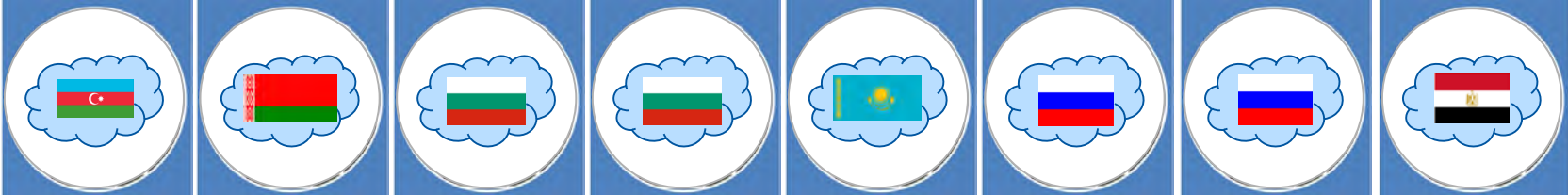
Govorun
184-3000
slots

NICA Cluster
250 slots

UNAM
100 slots

MPD
collaboration

Clouds



IPANAS
16 slots

INP
132 slots

SU
48 slots

INRNE
20 slots

INP
50 slots

REA Plehanova
132 slots

NOSU
84 slots

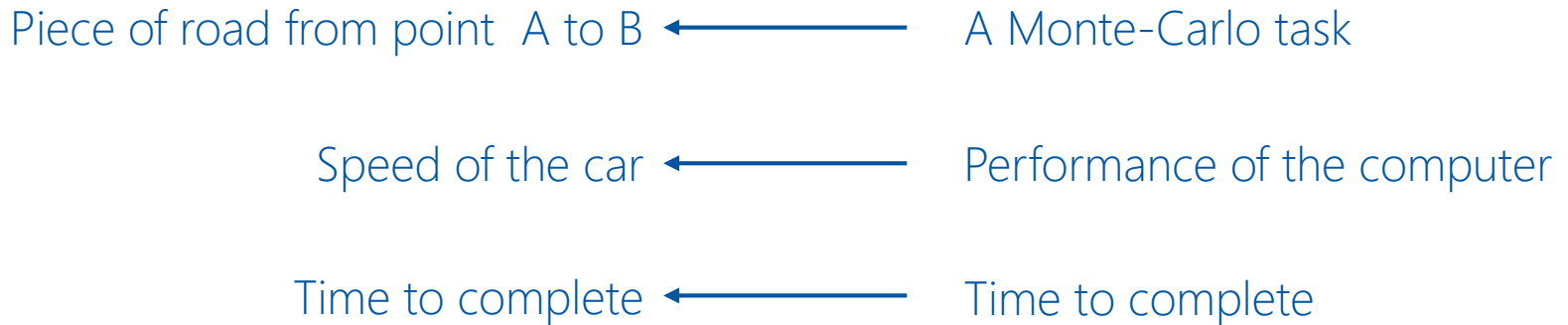
STI-SCI
98 slots

Total amount of cores exceeds 5000

Individual CPU core performance study

- Centralized job management gives possibility for centralized and unified performance study of different computing resources.
- Before running user jobs DIRAC Pilots execute benchmark for CPU core they are running on.
- Benchmark is DiracBenchmark2012 or DB12. It evaluate just CPU core performance. Disk I/O, RAM speed, Network, CPU caches and other highly important aspects of performance are **neglected by DB12**.

DB12 benchmark study

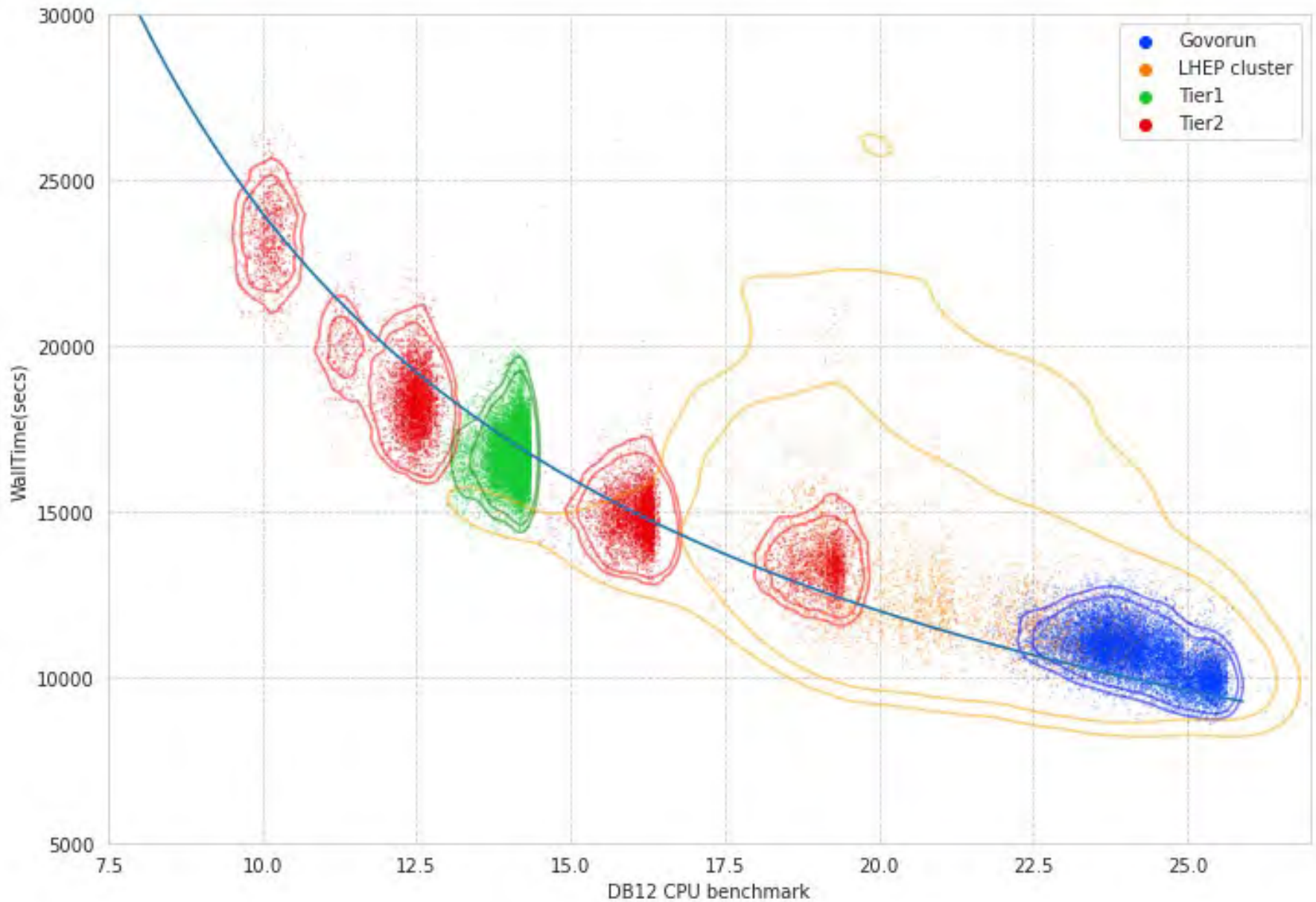


$$Time = \frac{Amount\ of\ work}{Speed\ of\ computer}$$

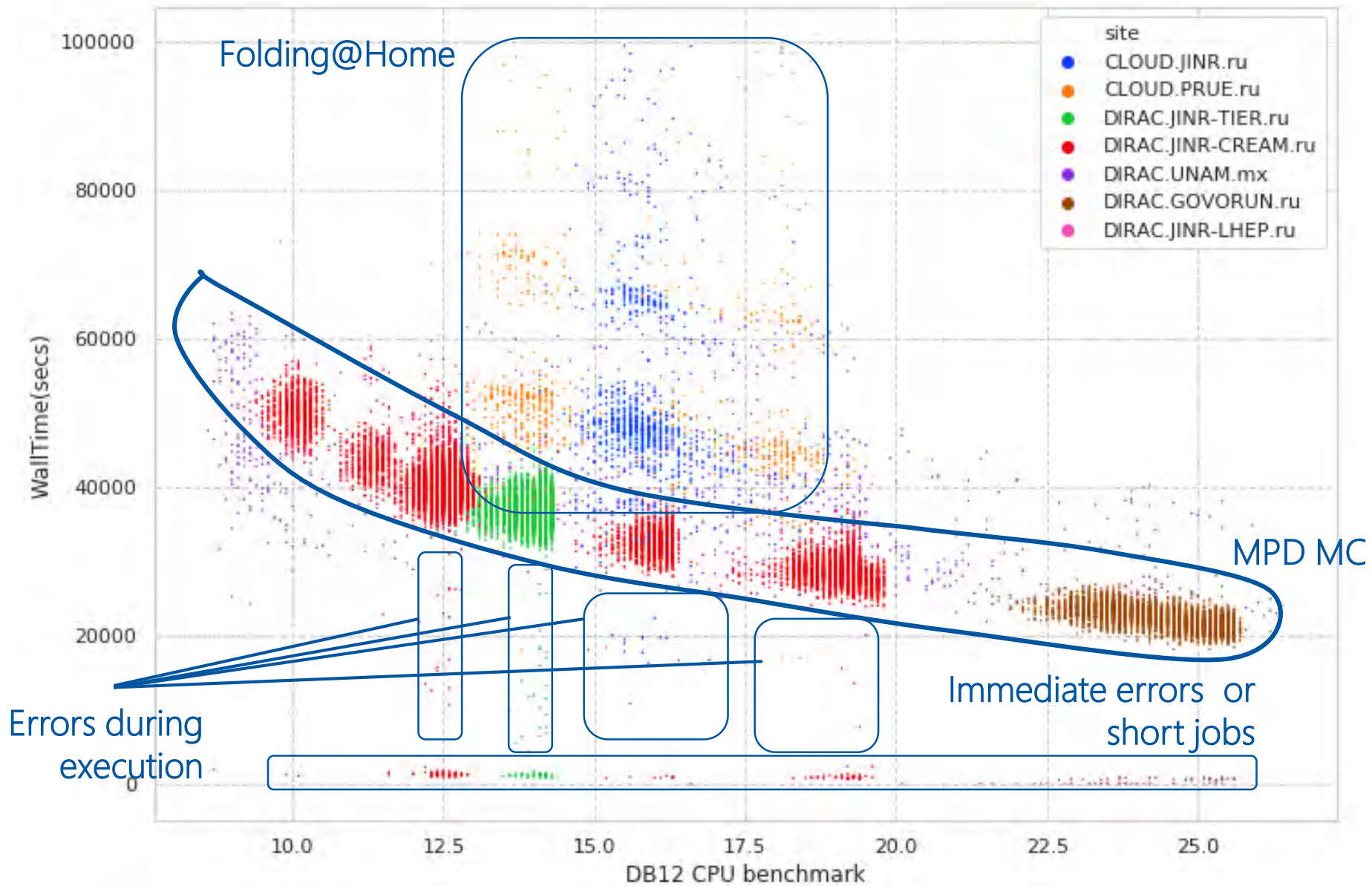
DB12 gives results like: 10(old slow core), 17 (standard server core), 27 (high performance core)

What if we build a plot, where X is DB12 result, Y is time in seconds. Then, every point on the plot represent one job. It would be mostly useless if all jobs were unique and different. But, in the real life there are usually many similar jobs.

Performance analysis



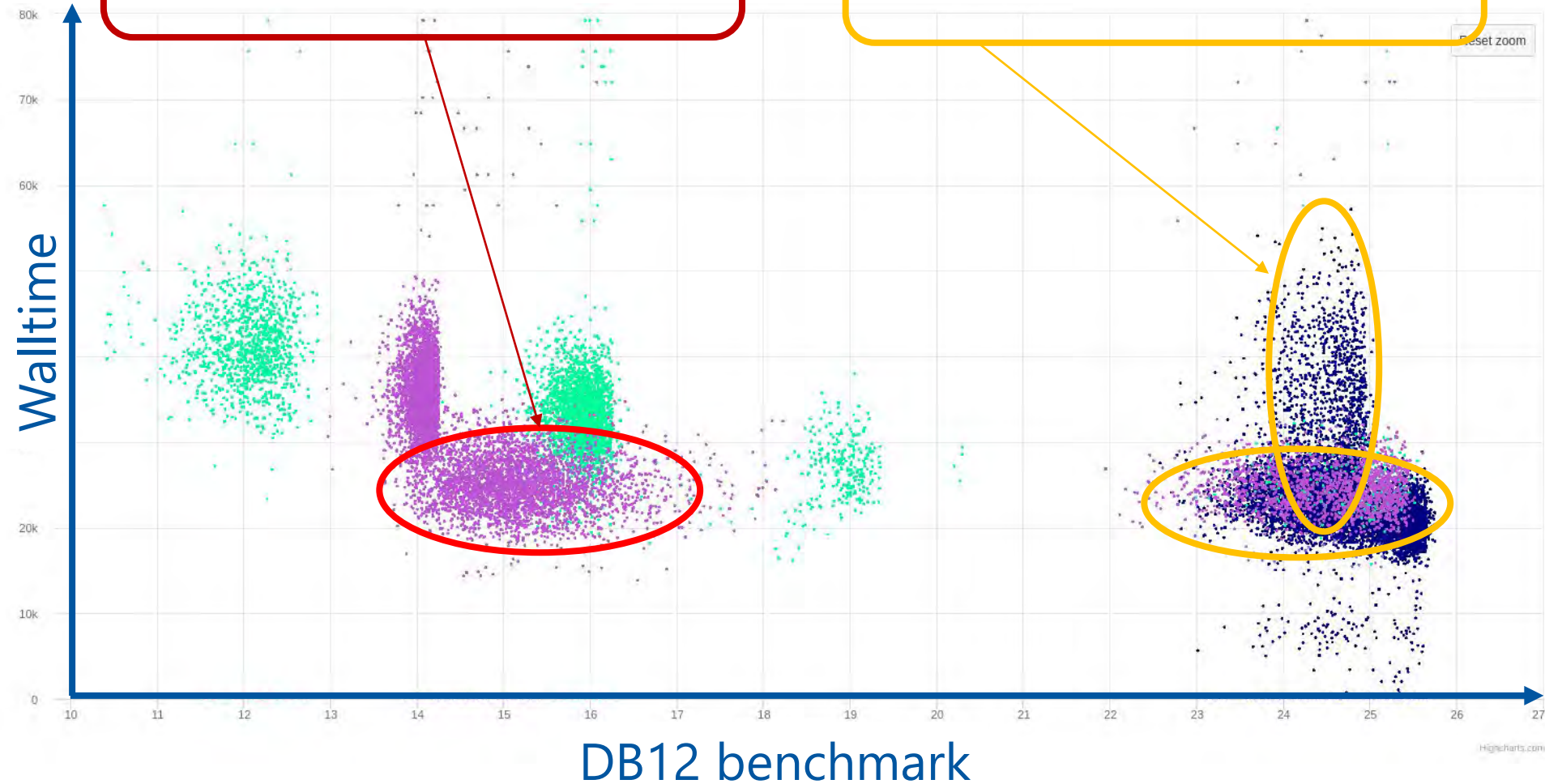
Performance analysis



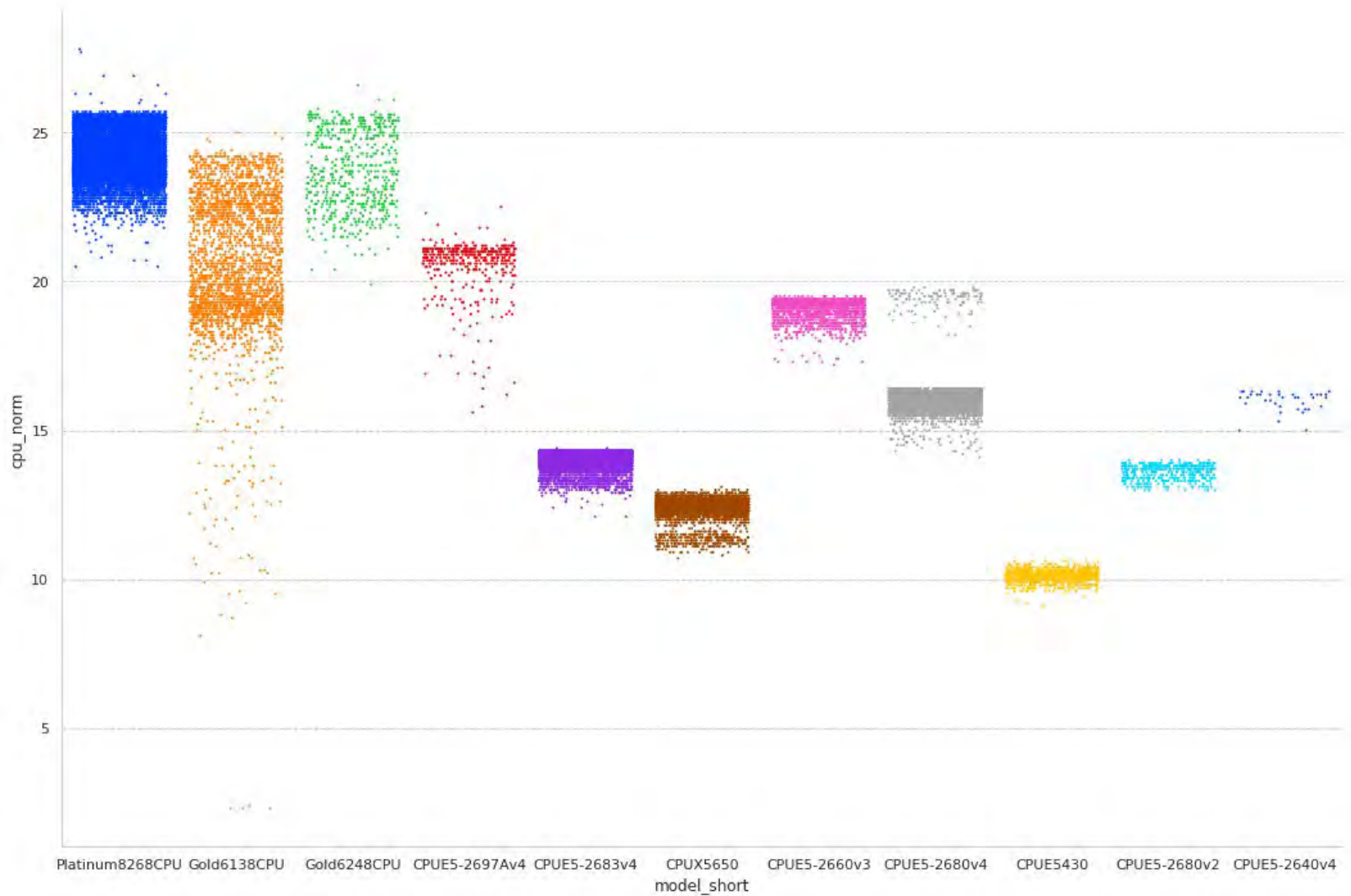
Discoveries

Wrong AMD processors estimation

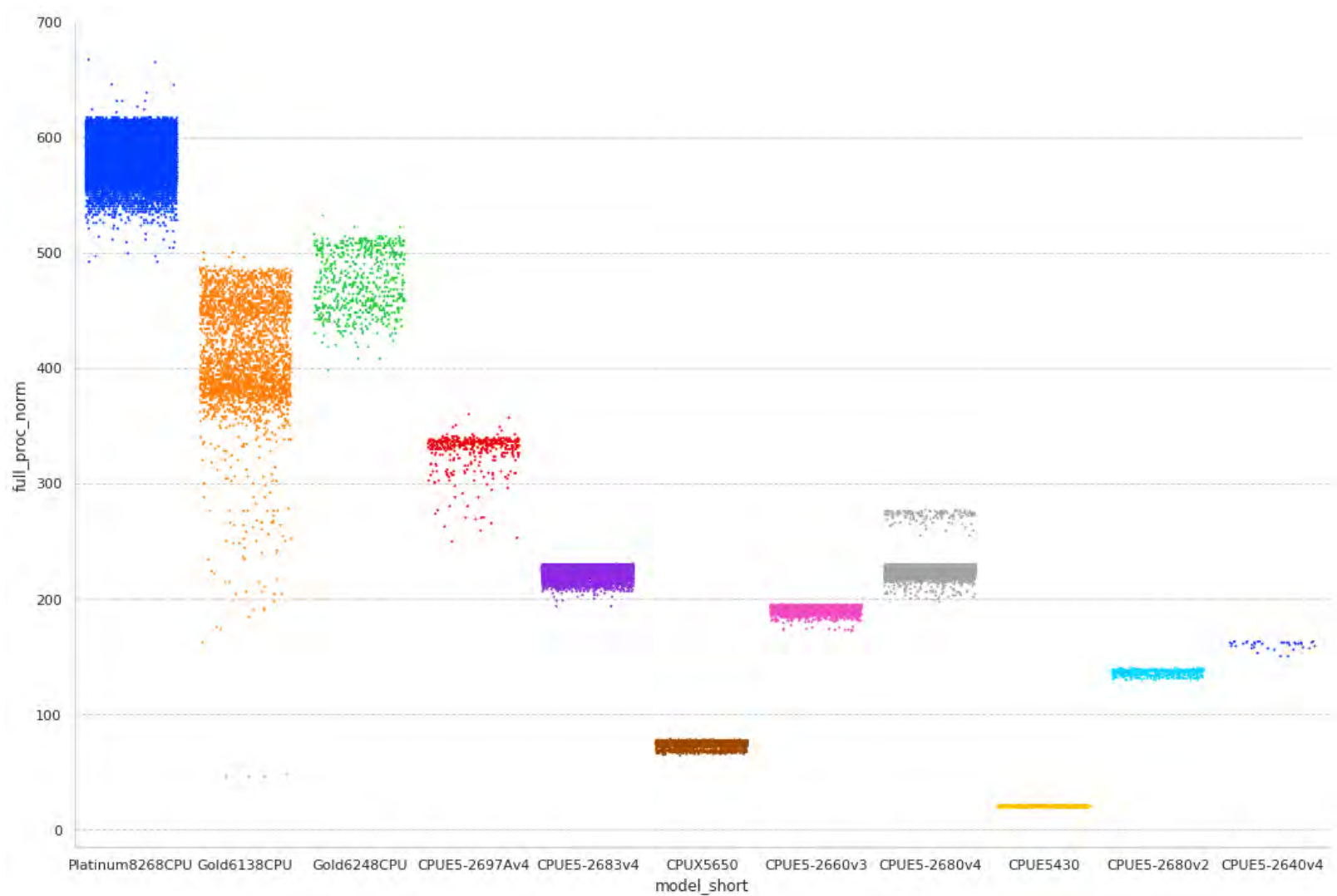
Occasional speed loss on high ram Govorun nodes



CPU core performance

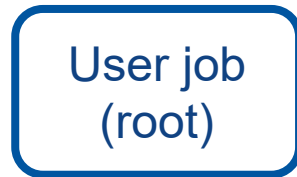


Total CPU performance

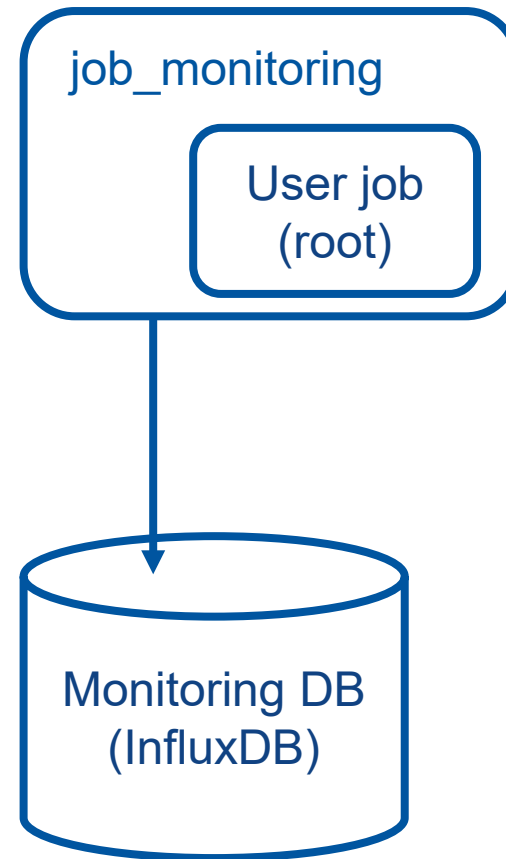


User job monitoring

```
$ root macro.c(input)
```



```
$ job_monitoring root macro.c(input)
```



User job monitoring

GenToDst job on Govorun

