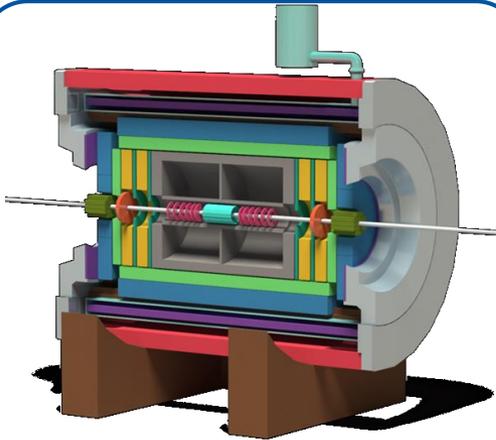


DUBNA

DIRAC@JINR report

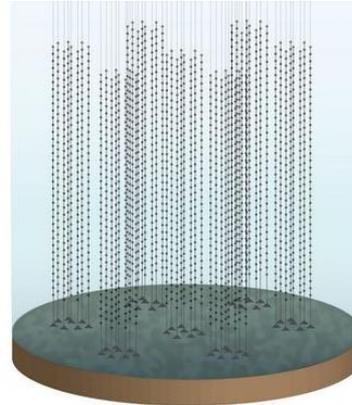
Speaker: Igor Pelevanyuk
Joint Institute for Nuclear Research

What do we use DIRAC for?



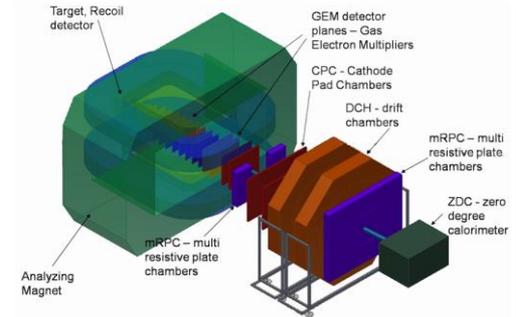
MPD@NICA

Monte-Carlo – Real
Analysis – Maybe



Baikal-GVD

Monte-Carlo – Real



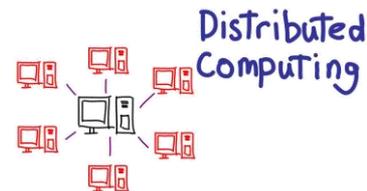
BM@N

Monte-Carlo – Tests
Reconstruction – Tests



**FOLDING
@HOME**

Folding@HOME

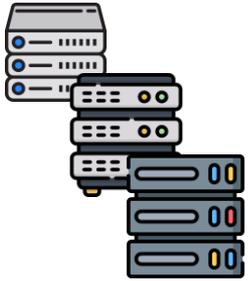


Distributed
Computing

Teaching

DIRAC features we use

Integration



ARC CE
Slurm
OpenNebula
SunGridEngine
EOS

FileCatalog

Not so much MetaData,
but it will change

Multi-VO

Cant imagine life
without it

VMDIRAC



Baikal-GVD and
Folding@HOME
rely solely on clouds

WebApp



Mostly JobMonitor
and Accounting
(and Configuration)

DIRAC features we do not use

RSS

Once upon a time it did not worked out, may be we should try again.

Interested

Request MS

We have just one SE really working. Will try it when there will be two of them.

Interested in future

Transformation S

MC transformation tested.
Going to try transformation initiated by meta data.

Interested

Any notable operations incident in the last year?

No major incidents from DIRAC

Do you have a DIRAC extension? Why?

We do not use extensions,

Cause I do not know how it may be useful for us

What is your biggest frustration with DIRAC?

To understand some nasty errors it is necessary to add `logger.debug()` in DIRAC code

You can magically add one feature to DIRAC, what is it?

We do not use all the features right now, can't imagine something big enough

To support your "Grid", do you have to use other systems than DIRAC?

- Ansible – restart all services/agents
- InfluxDB + Telegraf + Chronograf
for monitoring and some special use-cases like users' job monitoring, "zombie" VirtualMachines checks
- DIRAC CA – for educational purposes
and Folding@HOME

Computing resources

JINR



Tier-1
700 slots
Quota increased

CICC/Tier-2
600 slots

Clouds
100 slots

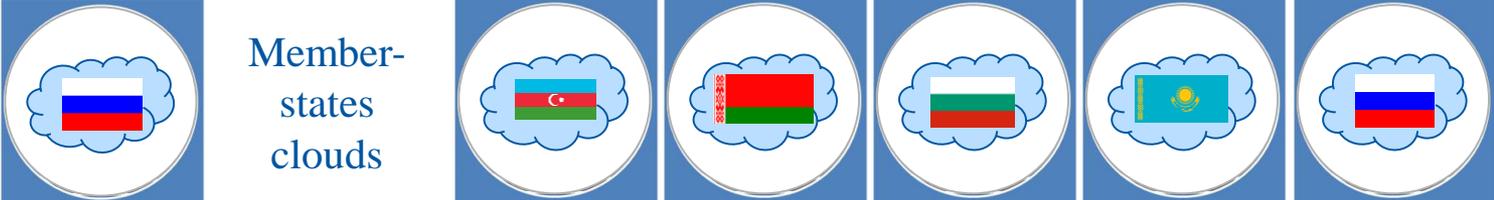
Govorun
184-800 slots

NICA Cluster
250 slots
(single user quota)

UNAM
100 slots

MPD collaboration

Member-states clouds



Plekhanov REA
40 slots

IPANAS
20 slots

INP
10 slots

INRNE
3 slots

INP
50 slots

NOSU
60 slots

Quotas in different resources may be increased in case of successful and effective usage.

MPD+DIRAC

- MPD is the first experiment that really used DIRAC.
- Only Monte-Carlo workload.
- >600k jobs successfully done
- ~200TB data written to EOS disks(all registered in DIRAC FileCatalog)
- Some resources are not presented in accounting:
 - JINR Cloud and other clouds were not actively used up to now.
 - Mexican cluster: 1000 jobs completed as an experiment. Network is weak point. The use of local storage will solve the issue.
 - dCache Tapes access over DIRAC is successfully tested. Mostly needed for RAW data from detector.

Clouds studying SARS-CoV-2

In March 2020, Folding@home launched a program to assist researchers around the world who are working on finding a cure and learning more about the coronavirus pandemic. The initial wave of projects simulate potentially druggable protein targets from SARS-CoV-2 virus, and the related SARS-CoV virus, about which there is significantly more data available.

Team members

Rank	Name	Credit	WUs
79 896	CLOUD.JINR.ru	12,645,224	5,355
94 530	CLOUD.PRUE.ru	9,453,851	4,175
144 494	CLOUD.NOSU.ru	4,125,976	1,345
219 931	CLOUD.IPANAS.az	1,542,618	910
224 325	CLOUD.INP.by	1,465,167	599
257 967	CLOUD.INP.kz	1,012,919	413
278 111	CLOUD.STI-SCL.eg	817,728	381
334 763	DIRAC.REA-Parallel.ru	471,543	155
N/A	CLOUD.INRNE.bg	80,796	30

It was decided to use idle cloud resources to participate in F@H.

Our priorities:

1. Do not interfere with other VMs
2. Control of the usage
3. Accounting
4. Only COVID-19 jobs accepted

Folding@HOME+DIRAC

- Clouds were underloaded.
- It took just one week to do technical work to submit F@H jobs to clouds through DIRAC.
- It became a good test for VMDIRAC and cloud performance
- Only Monte-Carlo workload.

- Issue: we have noticed that sometimes VMs are not deleted by VMDIRAC. Debug led to almost nothing. So, simple monitoring system and manual VMs removal script were introduced.

Baikal-GVD + DIRAC

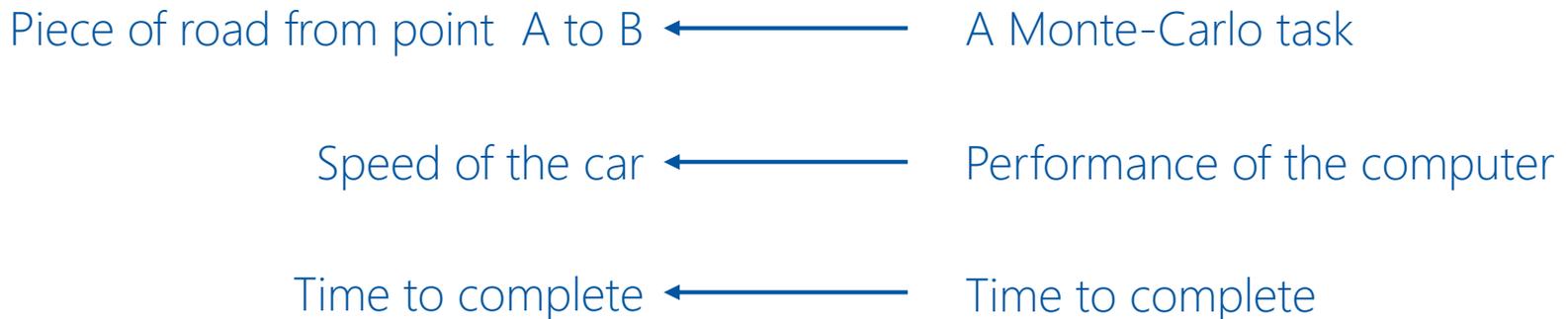
- Started to work with DIRAC after F@H proof of concept.
- Basic scripts and examples were presented and after that Baikal-GVD workload went to clouds instead of F@H.
- Only Monte-Carlo workload.
- Issue: typical job requires to download relatively big file. Current VMDIRAC version use old pilot version and does not allow to switch off watchdog, which kills jobs during initial data download and result upload.
- Solution: fake load script was created. It occupies CPU for 20% during data transfer. Requires user to manually run it in the job script.

BM@N+DIRAC

- Tests for Monte-Carlo and data reconstruction
- Resources successfully tested so far
 - Govorun, Tier1, Tier2, NICA cluster
 - JINR EOS in MLIT
- VOMS bmn.jinr.ru created and working.
- Waiting for real Monte-Carlo work to be done.

- Issue: nasty problems with big files upload (>100GB). Solved by careful gfal2 configuration. (Topic on the forum)

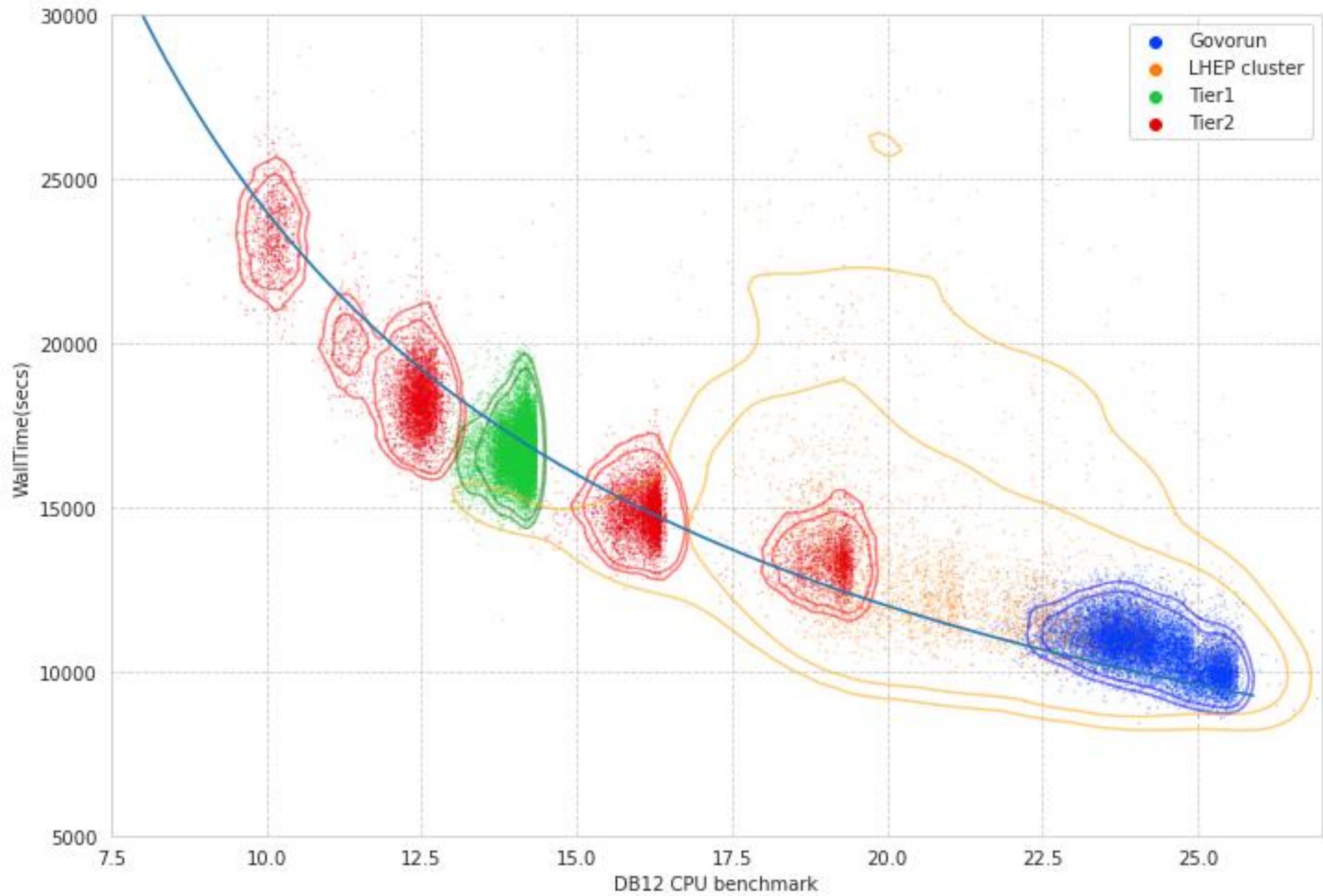
DB12 benchmark study



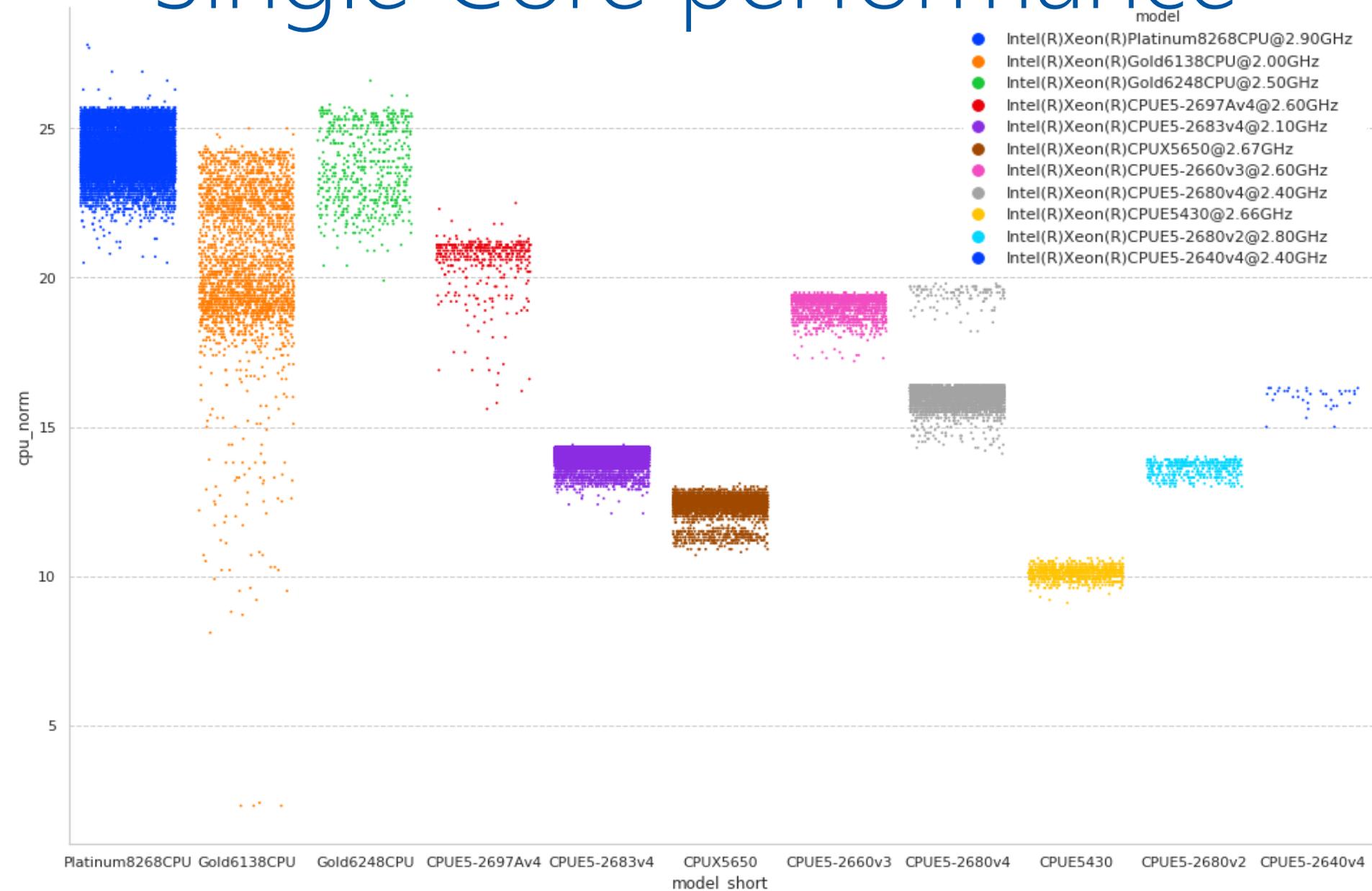
$$Time = \frac{Amount\ of\ work}{Speed\ of\ computer}$$

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

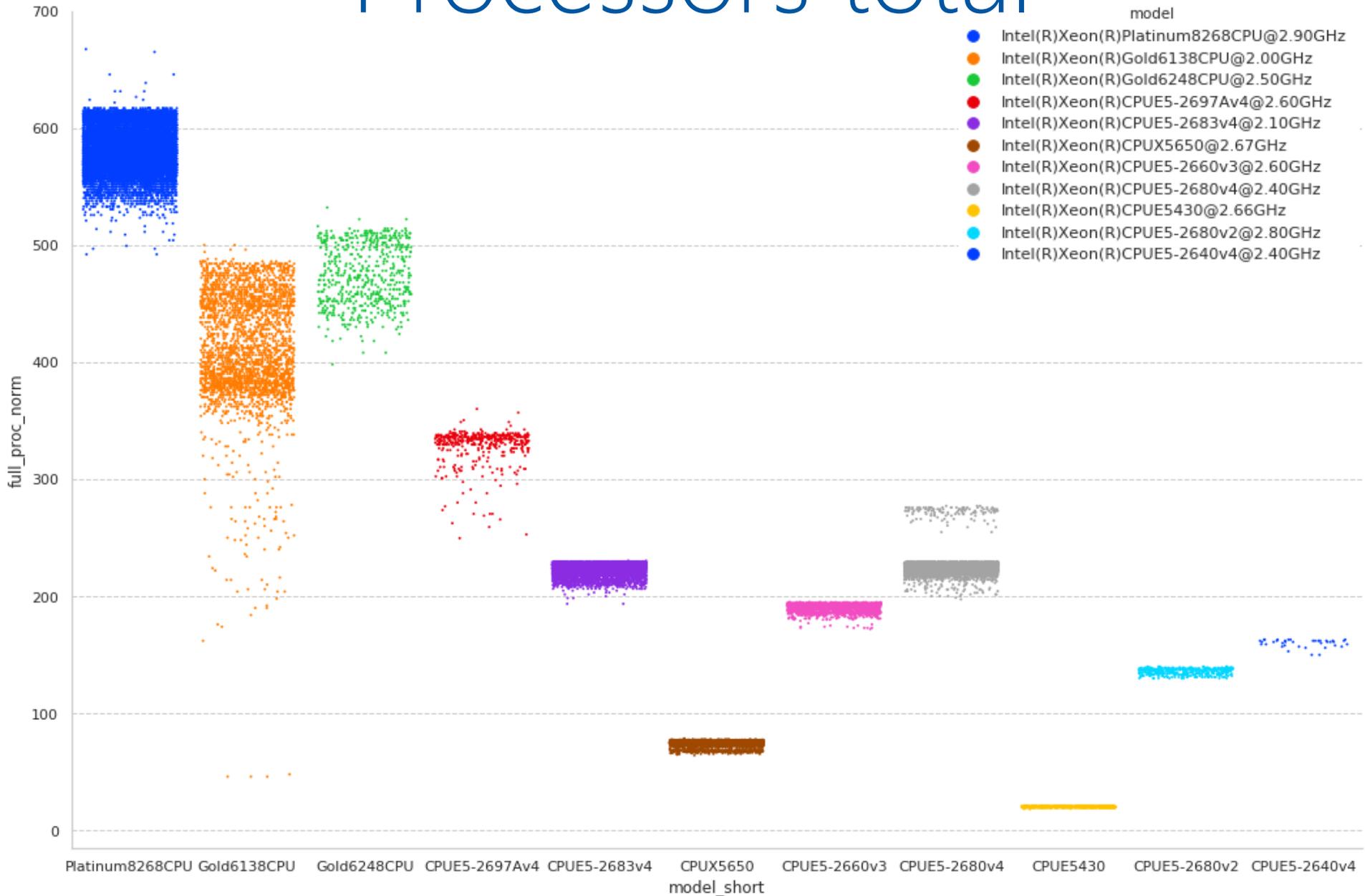
What if we build a plot, where X is DB12 result, Y is time in seconds. Then, every point on the plot represent one job. It would be mostly useless if all jobs were unique and different. But, in the real life there are usually many similar jobs.



Single Core performance

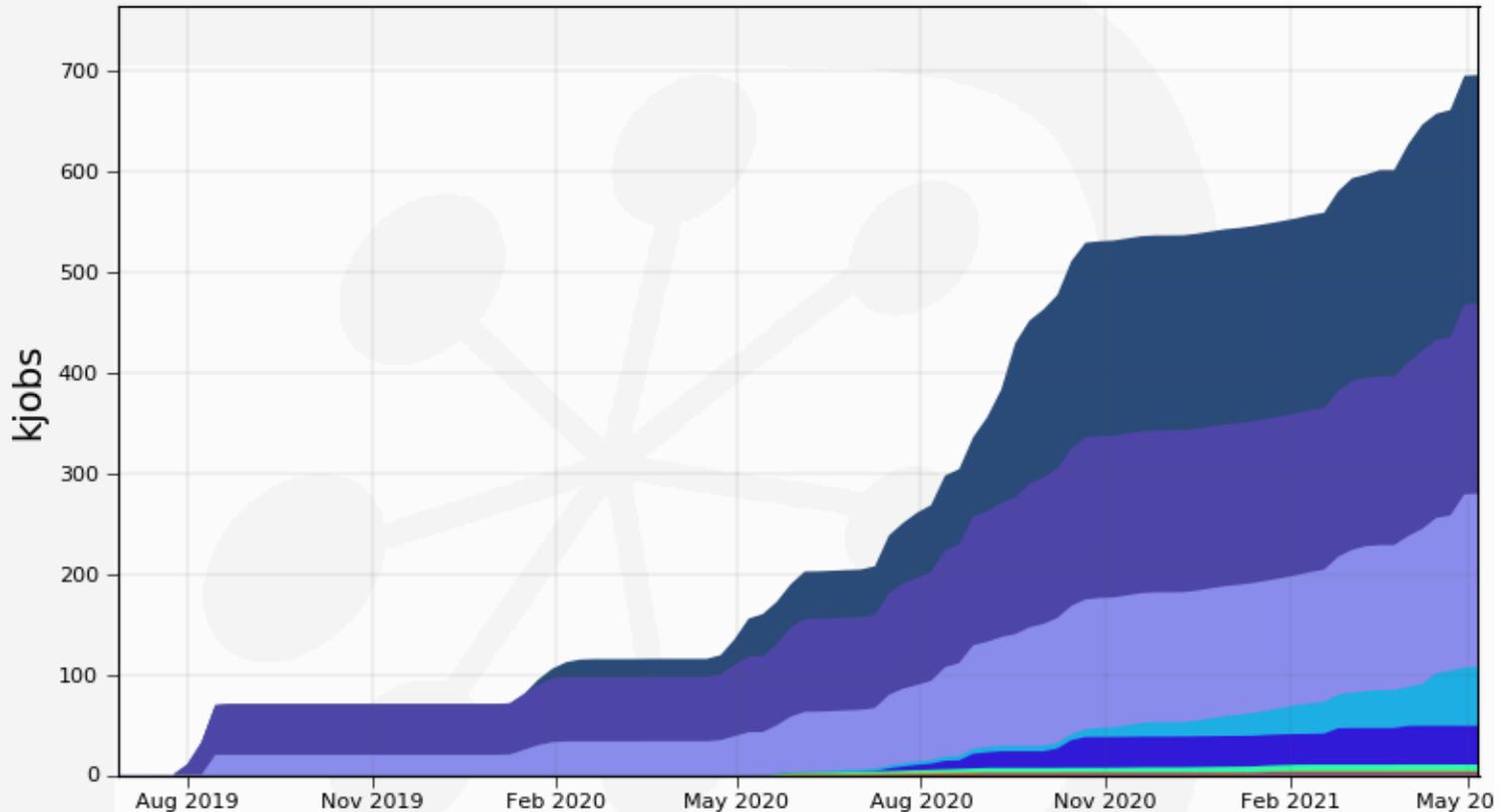


Processors total



DIRAC Jobs Done

Cumulative Jobs by Site
97 Weeks from Week 25 of 2019 to Week 18 of 2021



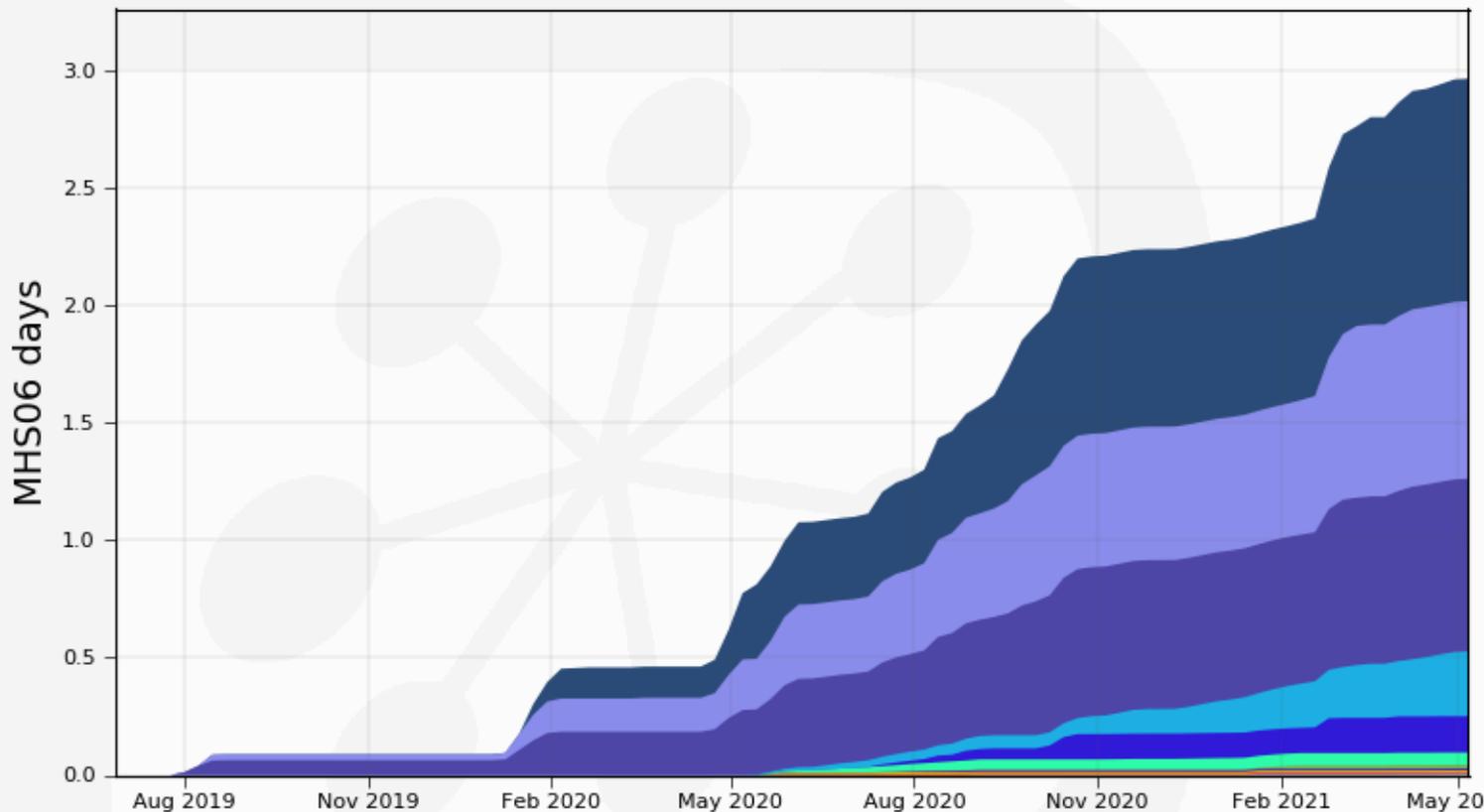
Max: 695, Min: 0.29, Average: 275, Current: 695

DIRAC.GOVORUN.ru	226.6	DIRAC.JINR-LHEP.ru	38.4	CLOUD.IPANAS.az	0.9	CLOUD.INP.kz	0.1
DIRAC.JINR-CREAM.ru	189.1	CLOUD.PRUE.ru	5.9	CLOUD.INP.by	0.6	CLOUD.INRNE.bg	0.1
DIRAC.JINR-TIER.ru	172.2	CLOUD.NOSU.ru	1.2	CLOUD.STI-SCI.eg	0.4		
CLOUD.JINR.ru	58.4	DIRAC.UNAM.mx	1.0	DIRAC.REA.ru	0.2		

Generated on 2021-05-08 19:46:28 UTC

DIRAC Normalized CPU

Normalized CPU used by Site
97 Weeks from Week 25 of 2019 to Week 18 of 2021



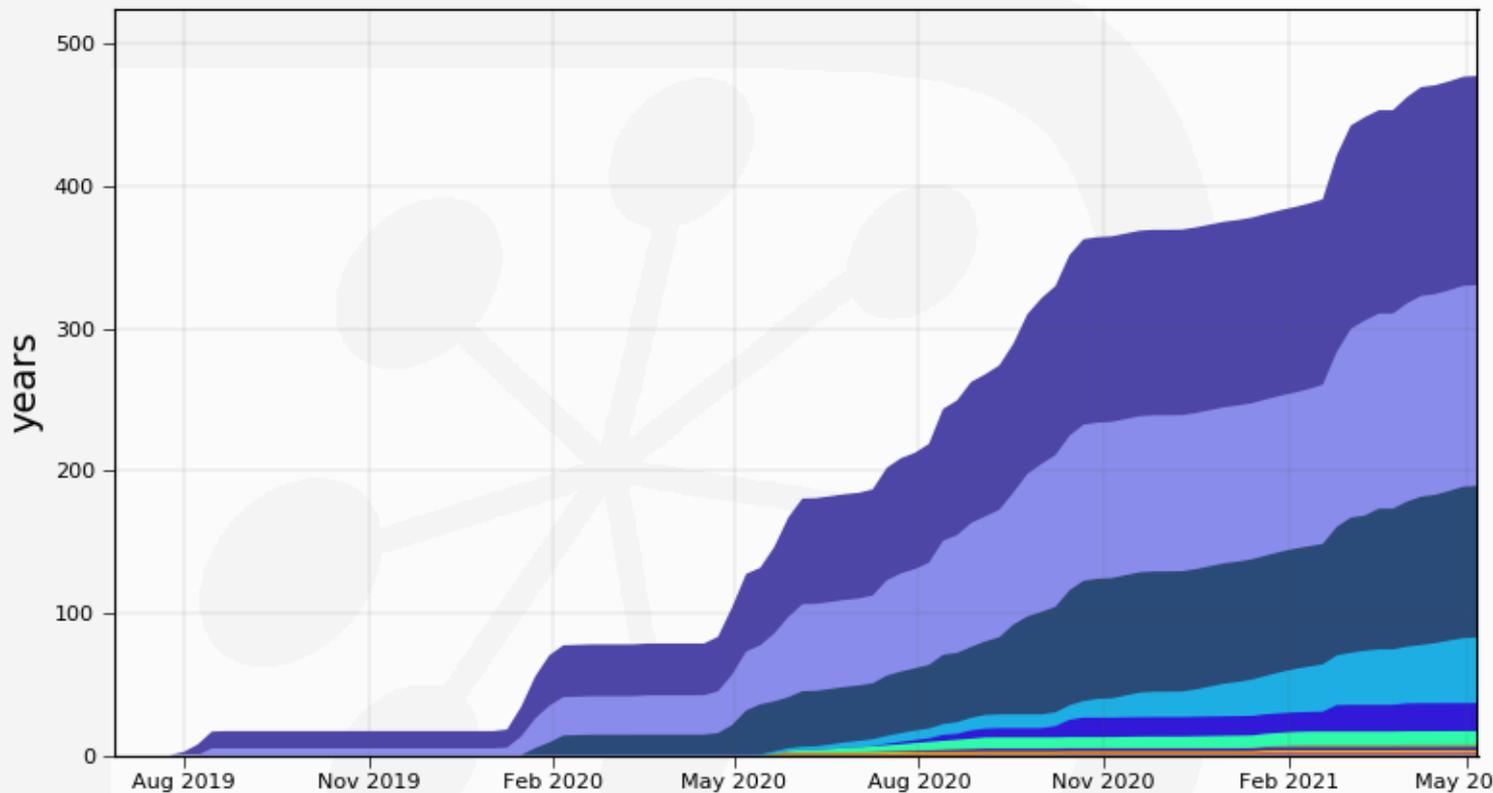
Max: 2.96, Min: 0.00, Average: 1.16, Current: 2.96

DIRAC.GOVORUN.ru	0.9	DIRAC.JINR-LHEP.ru	0.2	DIRAC.UNAM.mx	0.0	CLOUD.INP.kz	0.0
DIRAC.JINR-TIER.ru	0.8	CLOUD.PRUE.ru	0.1	CLOUD.INP.by	0.0	DIRAC.REA.ru	0.0
DIRAC.JINR-CREAM.ru	0.7	CLOUD.NOSU.ru	0.0	CLOUD.STI-SCI.eg	0.0		
CLOUD.JINR.ru	0.3	CLOUD.IPANAS.az	0.0	CLOUD.INRNE.bg	0.0		

Generated on 2021-05-08 19:45:48 UTC

DIRAC Waltime CPU

Cumulative wall time by Site
97 Weeks from Week 25 of 2019 to Week 18 of 2021



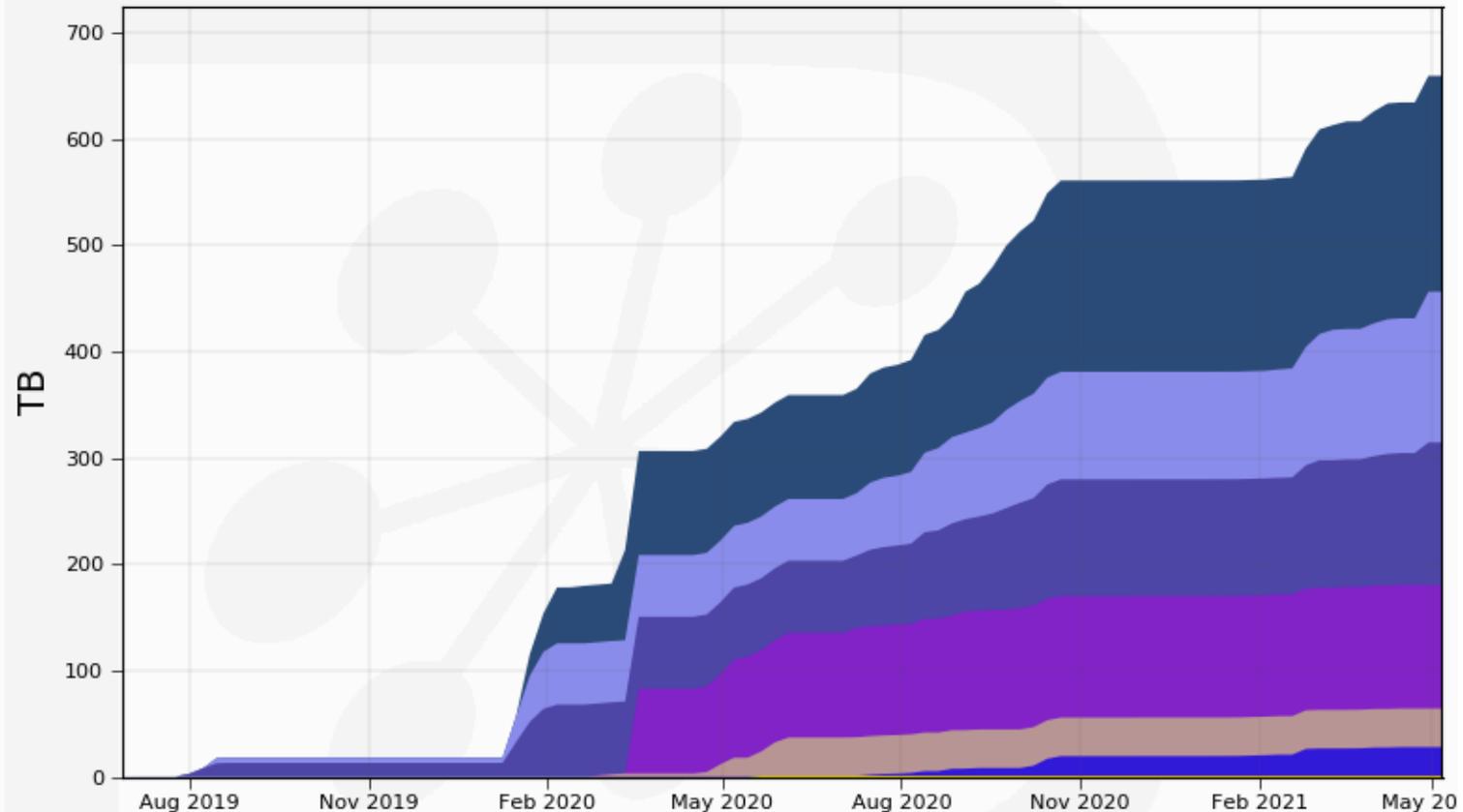
Max: 478, Min: 0.00, Average: 192, Current: 478

DIRAC.JINR-CREAM.ru	147.0	CLOUD.PRUE.ru	9.2	CLOUD.STI-SCL.eg	0.7
DIRAC.JINR-TIER.ru	141.3	CLOUD.NOSU.ru	2.1	DIRAC.REA.ru	0.2
DIRAC.GOVORUN.ru	106.5	CLOUD.IPANAS.az	2.0	CLOUD.INRNE.bg	0.2
CLOUD.JINR.ru	46.2	DIRAC.UNAM.mx	1.3	CLOUD.INP.kz	0.1
DIRAC.JINR-LHEP.ru	19.7	CLOUD.INP.by	1.0		

Generated on 2021-05-08 19:46:03 UTC

DIRAC data transfer

Transferred data by ExecutionSite
97 Weeks from Week 25 of 2019 to Week 18 of 2021



Max: 659, Min: 0.03, Average: 314, Current: 659

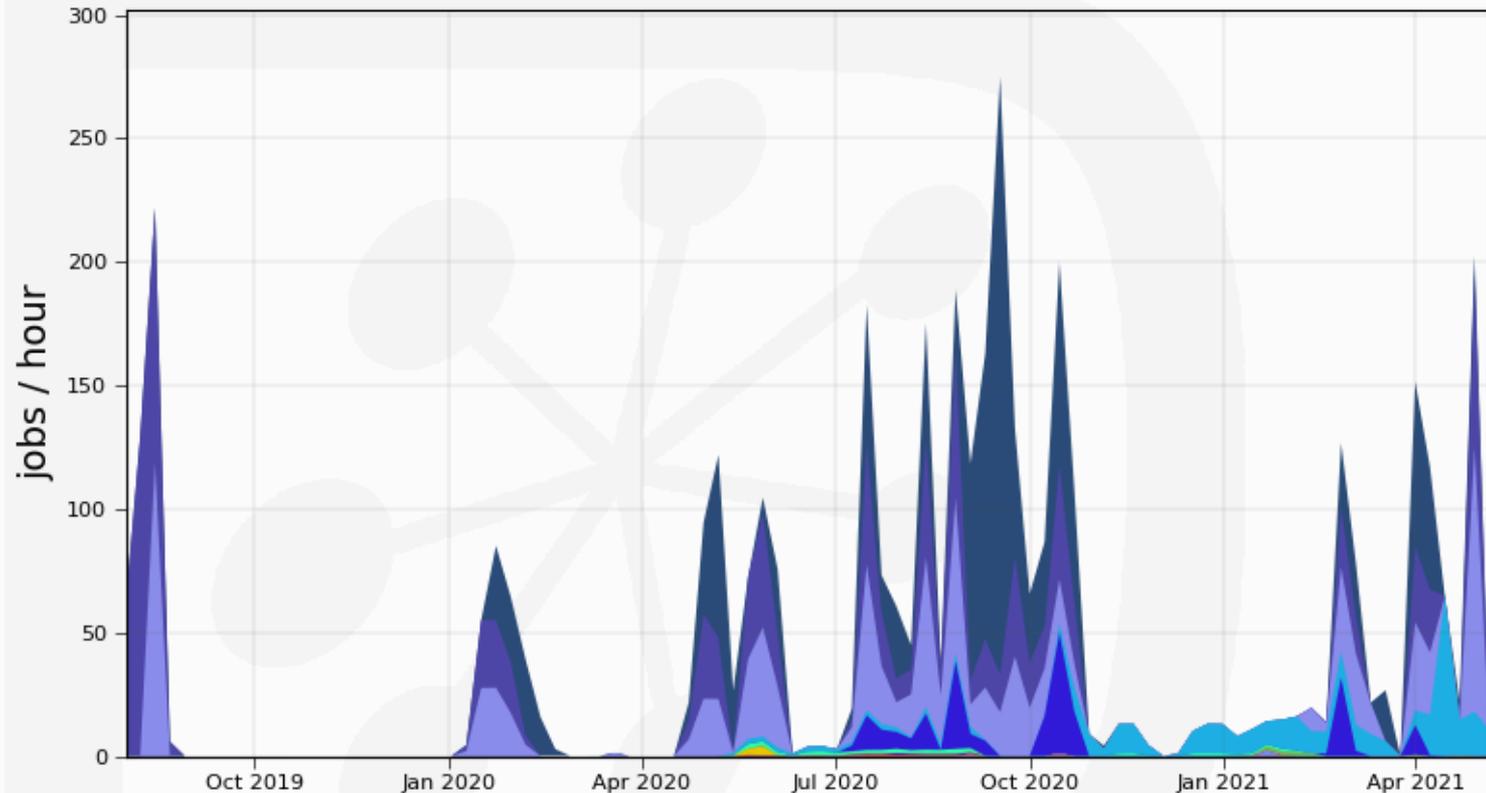
DIRAC.GOVORUN.ru	203.1	DIRAC.Client.local	116.3	DIRAC.Client.mx	0.6
DIRAC.JINR-TIER.ru	141.8	DIRAC.Client.ru	36.2	CLOUD.JINR.ru	0.0
DIRAC.JINR-CREAM.ru	134.1	DIRAC.JINR-LHEP.ru	27.3	DIRAC.UNAM.mx	0.0

Generated on 2021-05-08 20:42:53 UTC

DIRAC data transfer

Jobs by Site

92 Weeks from Week 30 of 2019 to Week 18 of 2021



Max: 275, Average: 44.5, Current: 10.7

DIRAC.GOVORUN.ru	32.6%	CLOUD.PRUE.ru	0.8%	CLOUD.STI-SCI.eg	0.1%
DIRAC.JINR-CREAM.ru	27.1%	CLOUD.NOSU.ru	0.2%	DIRAC.REA.ru	0.0%
DIRAC.JINR-TIER.ru	24.8%	DIRAC.UNAM.mx	0.1%	CLOUD.INP.kz	0.0%
CLOUD.JINR.ru	8.5%	CLOUD.IPANAS.az	0.1%	CLOUD.INRNE.bg	0.0%
DIRAC.JINR-LHEP.ru	5.5%	CLOUD.INP.by	0.1%		

Acknowledgments

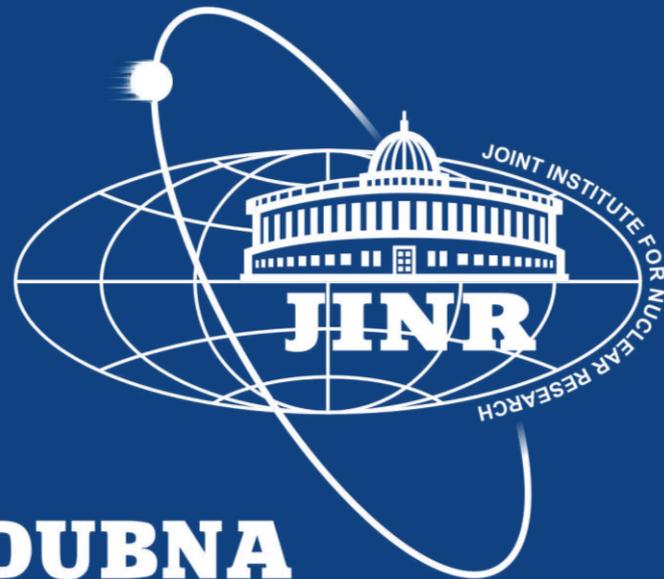
- JINR installation would not be possible without Andrei help and support during all these years.
- Thanks to all of you who helped with DIRAC, especially: Federico, Daniela, Christophe, Xiaomei, Andre, Marko and Christopher

How would you rate the communication?

- Communication is a “killer feature” of DIRAC.
 - 10 conversations on DIRAC Forum
 - BiLD-dev meetings are super useful for feeling the heartbeat of DIRAC.
- Thanks for DIRAC Communities roundtable, it makes me confident that I am not alone with DIRAC.

Conclusions

- In JINR DIRAC allowed using all major computing resources. Biggest computing resource right now is just around 40% of peak united performance.
- Integration of resources requires some effort.
- When the system is operational, and users submit jobs, we may get intelligence about the performance, structure, components of computing resources “almost” “for free”.
- Availability of system integrated with DIRAC allowed to start to participate in Folding@HOME in one week for all cloud resources. Later it allowed running Baikal-GVD jobs.



DUBNA