

Belle II

Dr. Silvio Pardi

WLCG Workshop - Lancaster

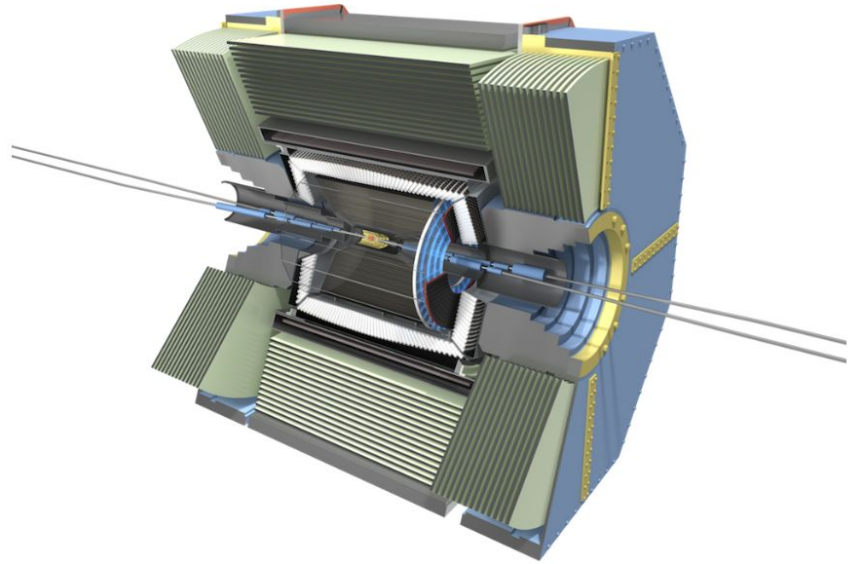
7 November 2022

Belle II Collaboration

26 Countries/regions

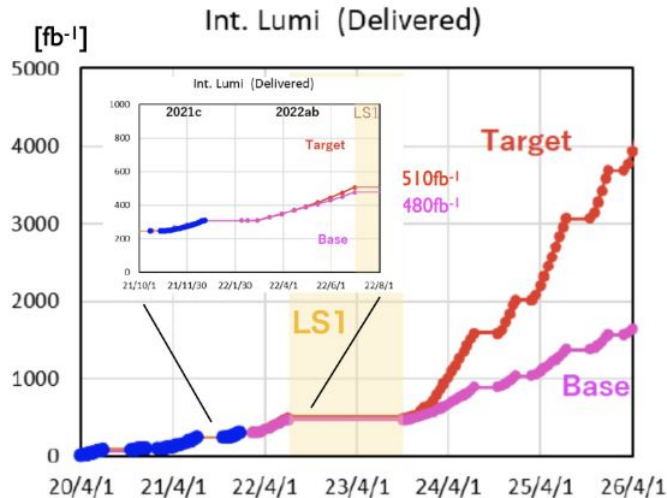
123 Institutes

1.075 Researches



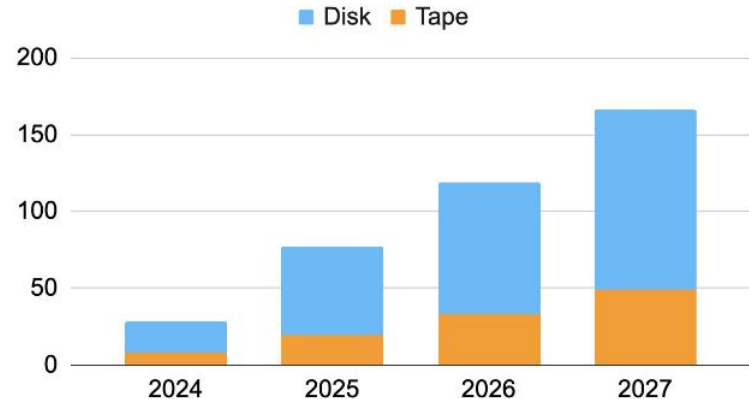
Belle II Numbers

- Integrated luminosity expected by the end of the experiment: 50 ab^{-1}
- Estimated size of the dataset collected by the experiment is $\sim \mathbf{O(10) \text{ PB/year}}$.



- Data must be distributed and analyzed by > 1000 collaborators around the world.

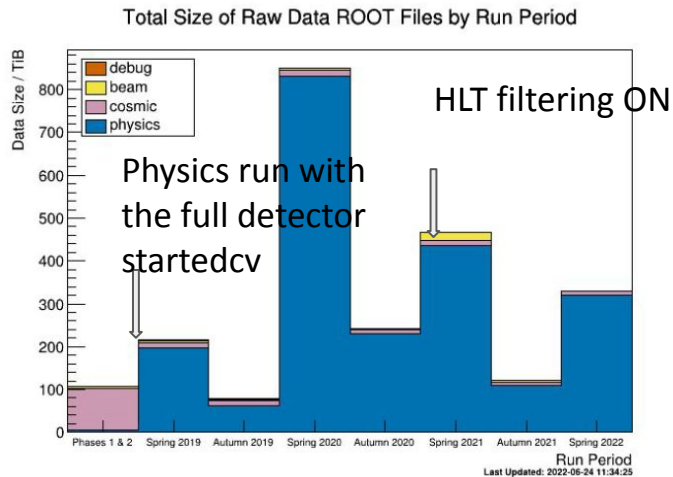
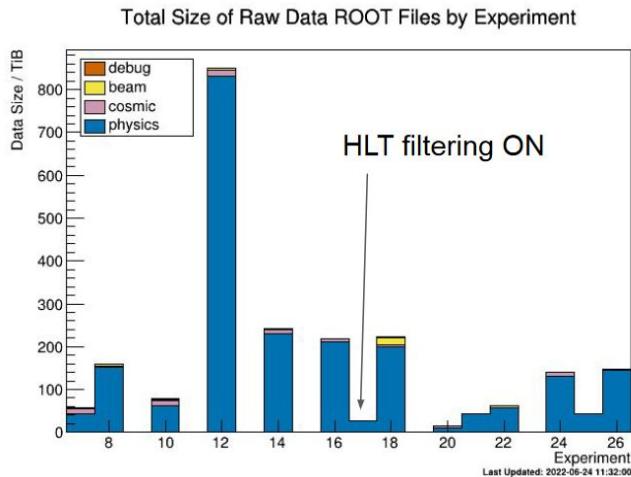
Space Occupancy (PB)



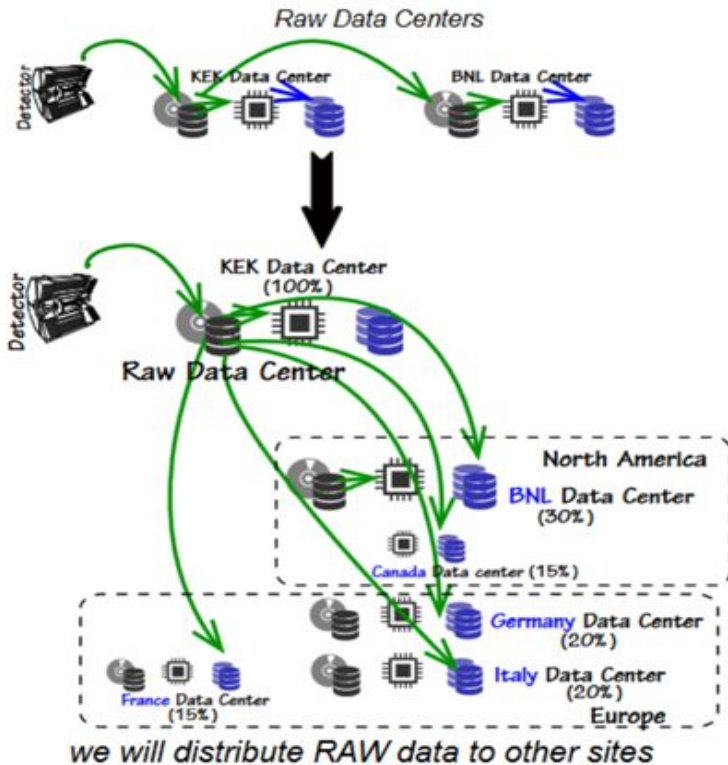
- Not as large when compared to HL-LHC scales, but corresponds to 10^{12} events, representing a significant data management challenge.

Belle II Status and Plans

- More than 2PB of RAW Data Collected so far, since 2019
- Currently we are in Long Shutdown for upgrade
- Data taking will start again in the last quarter of 2023



RAW Data distribution



We have gradually implemented the full RAW Data distribution schema, starting to distribute them since 2021 JFY according with the following table

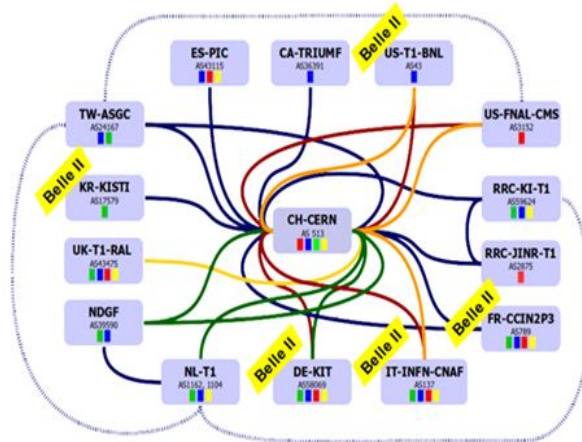
SITE	2019-2020	2021-2024
BNL - USA	100%	30%
CNAF - Italy	0%	20%
DESY - Germany	0%	10%
KIT - Germany	0%	10%
IN2P3CC - France	0%	15%
UVIC - Canada	0%	15%

Belle II Network

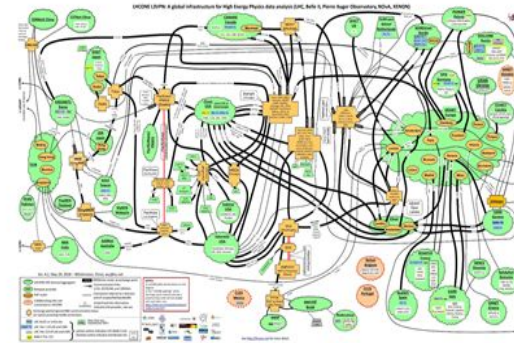
100G Global Ring
runned by SINET



LHCOPN Optical
infrastructure that can
be used without
jeopardizing resources



LHCONE L3 VPN
Connecting all the major
Data Centres



DIRAC Framework and Grid services

Production Infrastructure

11 DIRAC servers + 4 DB servers + 2 Web servers (KEK)

Test Infrastructure at BNL

Certification: validation of new BelleDIRAC releases.

Migration Infrastructure: test of base DIRAC upgrades.

Other Grid Services

FTS - File transfers

AMGA - Metadata Catalog

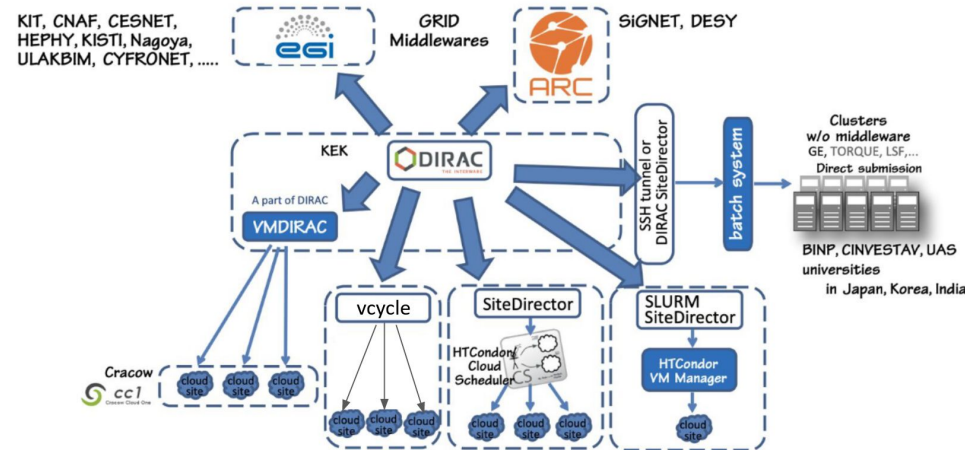
VOMS - Authorization

CVMFS - Software (basf2) and DIRAC + BelleDIRAC distribution

GGUS - Issue tracking

GOcdb - Downtime Information from sites

VCYCLE/CloudScheduler/TARDIS - For Cloud.



The Data Management System : RUCIO

Rucio is a highly-scalable, policy-driven data management system.

Originally built for ATLAS, Rucio has been interfaced, initially with BelleDIRAC, then DIRAC and is now responsible of the Data Management part for Belle II:

- As Distributed Data Management System
- As File Catalog
- Rucio Client

Gradually enabled more and more features from Rucio.

In evaluation the usage of Rucio as metadata service (see also presentation at the Rucio workshop <https://indico.cern.ch/event/1185600/contributions/5120132/>)

Distributed Computing Infrastructure as of 2022

Storage Elements (SEs)

- 29 storages
- 5 tape systems

Storage	Space (PB)
Disk	15.5
Tape	12.4

Computing elements (CEs)

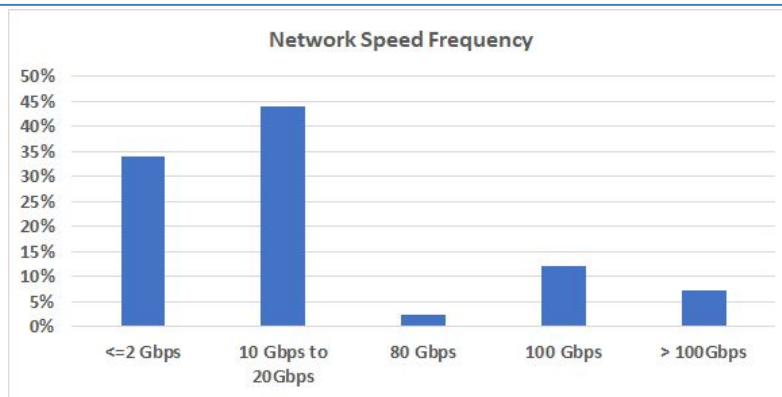
- 56 sites registered in DIRAC
 - 30 sites Providing Pledged CPUs
 - 16 Sites Pledged+Opportunistic
 - 10 Sites Opportunistic Only

CPU	kHS06	Job slots
Pledged CPU	466	32 kJS
Opportunistic CPU (Maximum)	385	32 kJS
TOTAL	852	64 kJS

Network Overview

Network	#Sites
LHCONE	48%
GeneralIP	52%

More than 80% of kHS06
Running on LHCONE
More than 90% of Storage on
LHCONE



IPv6 deployment	#Sites
Storage Dual Stack	38%
WorkerNode Dual Stack	13%

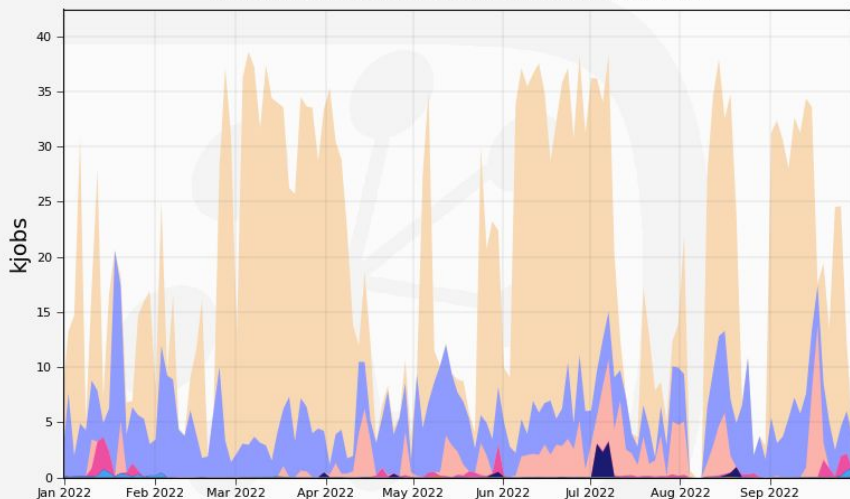
11.3 PB reachable via
IPv6 over of 15.5 PB

Highlight

- KEK 80Gbps on LHCONE
- BNL going to 300TB will increase up to 800 Gbps and 1.2Tbps
- CNAF, and KIT 200Gbps

Belle II Status

Running jobs by JobType
38 Weeks from Week 52 of 2021 to Week 39 of 2022



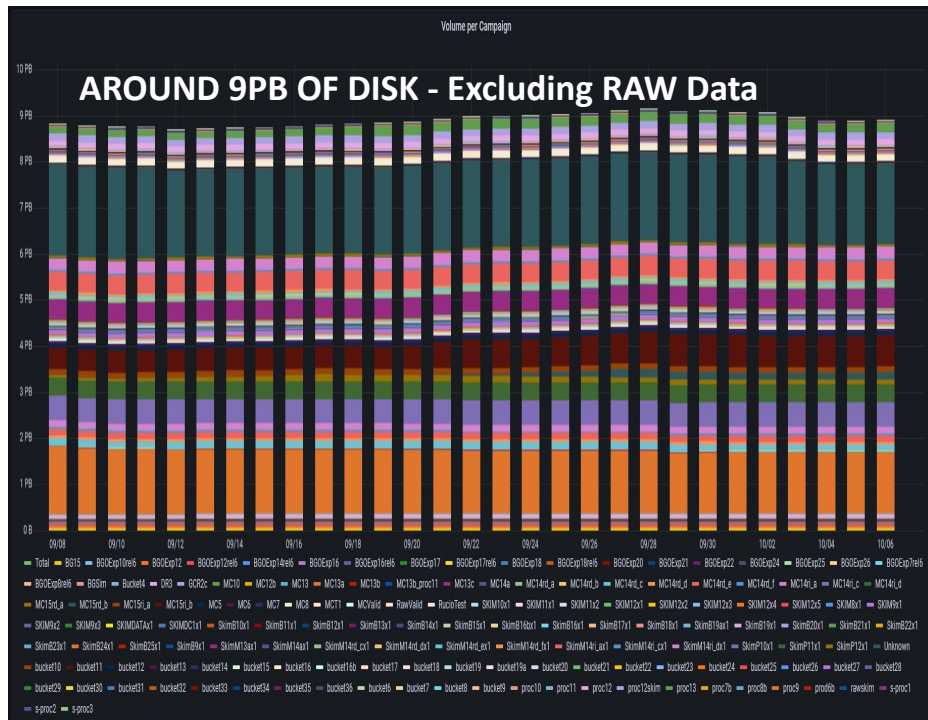
Max: 38.6, Min: 0.00, Average: 20.2, Current: 4.91

MCProduction	70.0%	MCSkim	1.0%	Merge	0.1%	Test	0.0%
User	22.0%	MCProductionBGx0	0.4%	DataMerge	0.0%	unknown	0.0%
RawProcessing	6.3%	DataSkim	0.2%	UserScout	0.0%		

Generated on 2022-10-07 07:14:19 UTC

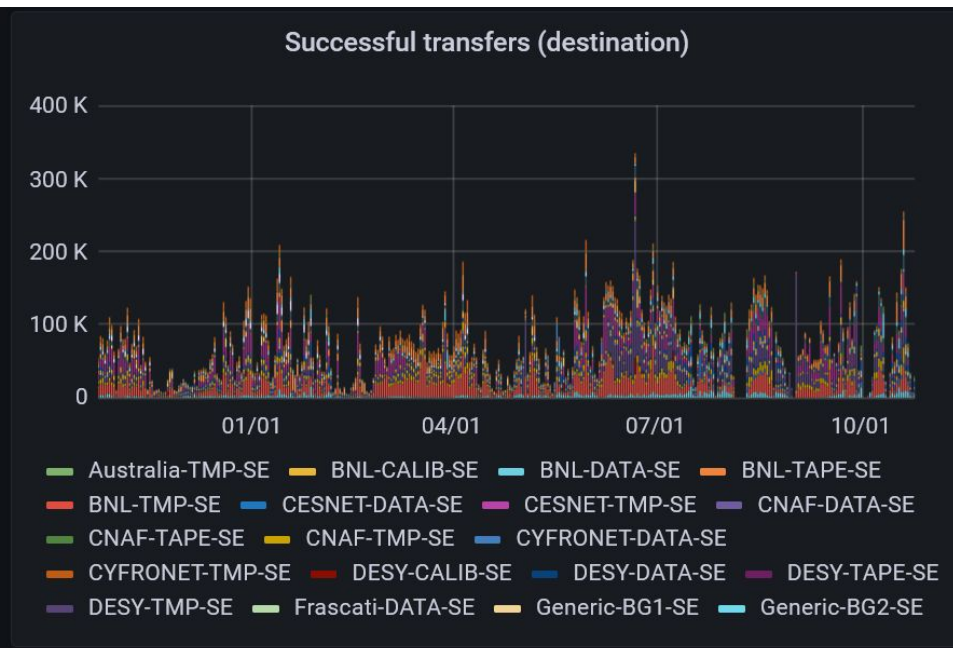
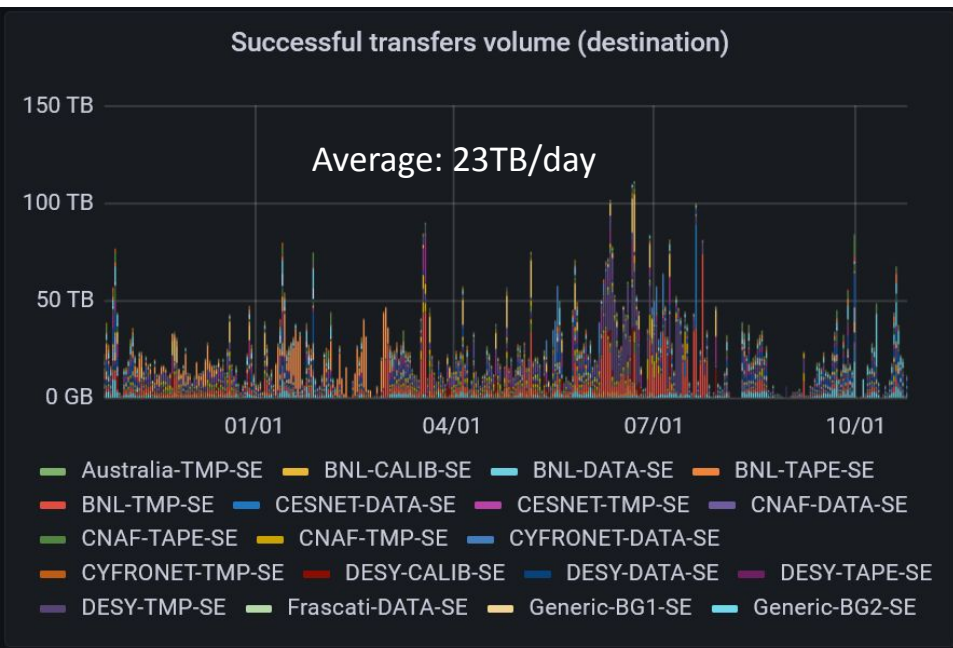
More that 38kJobs Running concurrently over the distributed infrastructure. MC dominant.

AROUND 9PB OF DISK - Excluding RAW Data



Statistics from Rucio Monitoring in the last 12 months

Third Party Copy transfers.



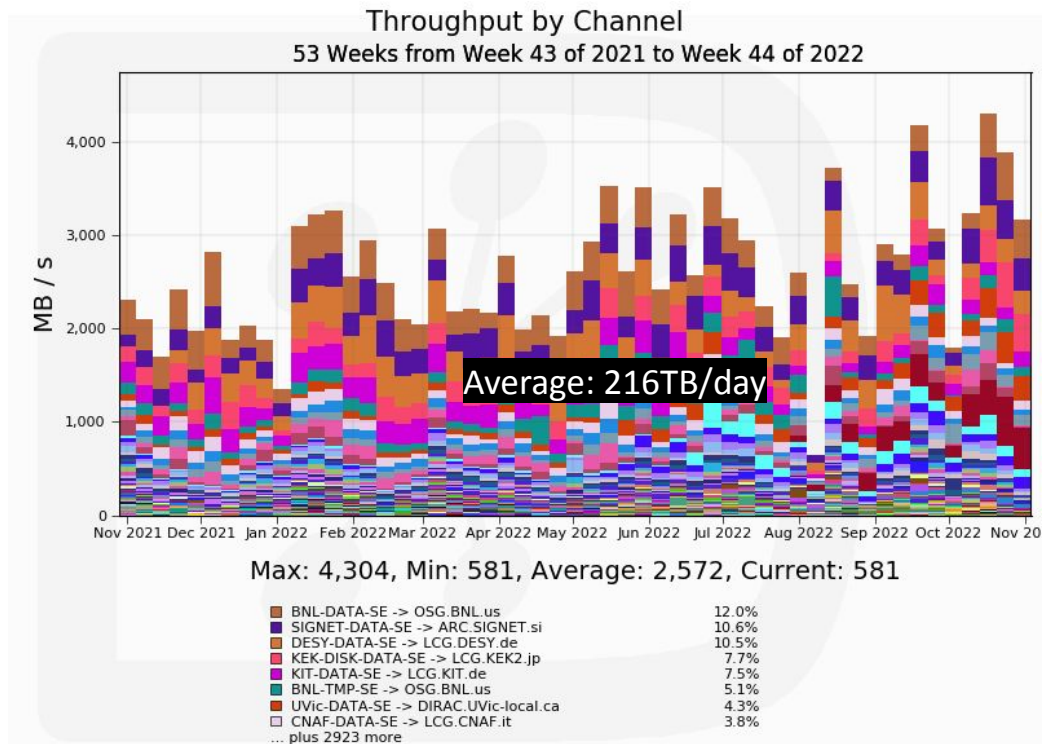
Data Analysis and Data production Throughput

From a global statistics we have roughly

>95% on LAN

<5% over WAN

Around 10TB/Day of global WAN Traffic



Generated on 2022-11-04 10:38:13 UTC

Expected Traffic in near term

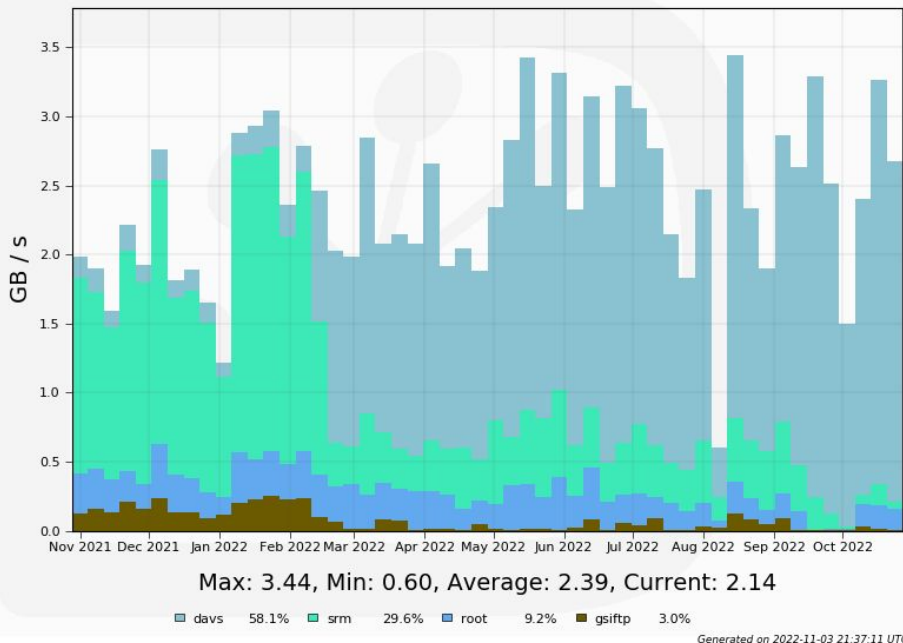
From the latest resource estimation data sheet.

(Values for 2023 bit biased by duration of LS1, RAW data production may restart latest quarter of 2023)

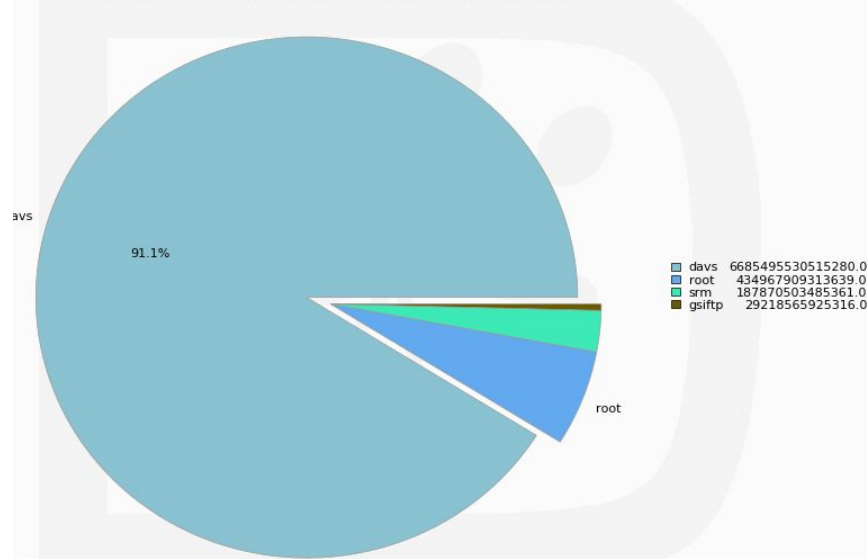
	2022	2023	2024
Estimated Mean	39 TB/day	52TB/day	67TB/day
Estimated Peak	190TB/day	260TB/day	339TB/day
Measured Mean	>33TB/day		
Measured Peak	>110 TB/day		

Migration to DAVS for data access

Throughput by Protocol
52 Weeks from Week 43 of 2021 to Week 43 of 2022



Total data transferred by Protocol
13 Weeks from Week 39 of 2022 to Week 52 of 2022

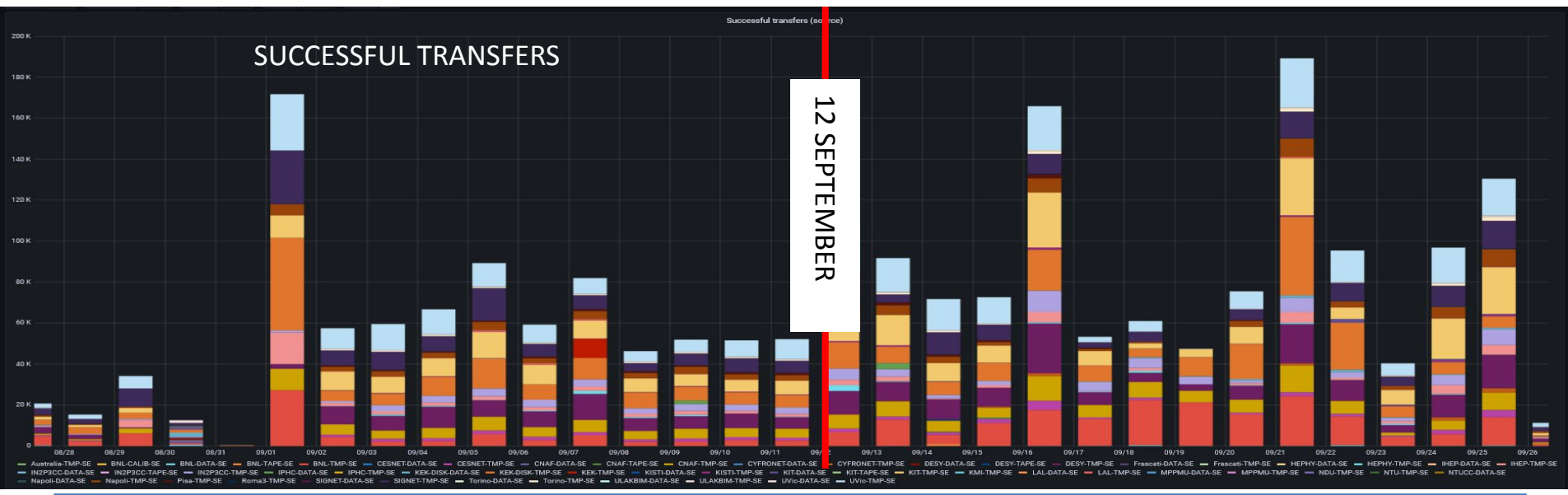


DAVS vs (SRM+gsiftp)/gridftp/root
in the last 3 months

Generated on 2022-11-03 21:31:40 UTC

Migration to DAVS for data transfer

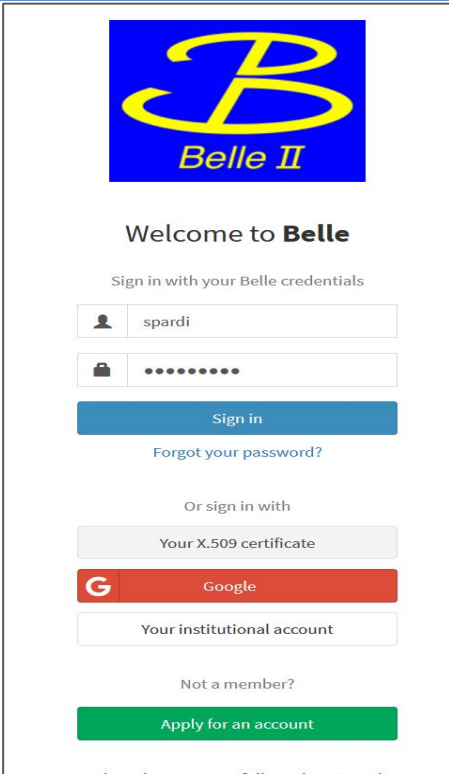
Ongoing activity for the extensive usage of davs protocol for data access and Third-Party-Copy (TPC) in substitution of srm+gsiftp.



Token Based Authentication

Following WLCG and OSG agenda, Belle II is working to supports token based authentication in substitution of the Grid Security Infrastructure (GSI)

- Indigo IAM service in place at CNAF for early tests
- Pre-production and Development IAM services in place at KEK.
- Token Based Authentication ongoing vs a selected set of Computing Elements and Storage Elements without DIRAC
- Tests the full workflow with DIRAC after the upgrading to the future versions



The image shows a Belle II login interface. At the top is the Belle II logo. Below it, the text "Welcome to Belle" is displayed. The user is prompted to "Sign in with your Belle credentials". There are two input fields: one for the username "spardi" and one for the password, represented by dots. A blue "Sign in" button is below the password field. A link "Forgot your password?" is positioned below the "Sign in" button. Below this, the text "Or sign in with" is shown. There are three options for signing in: "Your X.509 certificate" (grey button), "Google" (red button with the Google logo), and "Your institutional account" (white button). At the bottom, there is a link "Not a member?" and a green button "Apply for an account".

Token Testbed


Resources tested with CNAF IAM Service



- HTCondor-CE: CNAF, BNL, DESY, Napoli, IN2P3CC, KIT, Roma3
 - Test: condor submission
- Storage Elements: CNAF (STORM), IN2P3CC (dCache)
 - Test: full set of ls, mkdir, copy, delete with both null and production role implemented via optional group









Resources in testing at KEK

- FTS Server
- KEK storage server based on STORM
- KEK cluster under ARC-CE

COMPUTING ELEMENT - CONDOR PING TEST WITH TOKEN AUTHENTICATION

 New filter



 host_group:condor_ce_token 









 ACKNOWLEDGE  SET DOWNTIME  CHECK    Rows per page 30 1-6 of 6  

<input type="checkbox"/>	s	Status ↑	Resource	Parent	N	A	G	Duration	Tries	Last check	Information	State
<input type="checkbox"/>		UP	h cccondorce03.in2p3.fr					1h 18m	1/3 (H)	14m 10s	OK - condor_ping to cccondorce03.in...	
<input type="checkbox"/>		UP	h bgk01.sdcc.bnl.gov					1h 18m	1/3 (H)	14m 10s	OK - condor_ping to bgk01.sdcc.bnl...	
<input type="checkbox"/>		UP	h pps-token-htcondor-ce.gridka.de					1h 18m	1/3 (H)	1m 5s	OK - condor_ping to pps-token-htcon...	
<input type="checkbox"/>		UP	h condor-02.roma3.infn.it					1h 23m	1/3 (H)	14m 10s	OK - condor_ping to condor-02.roma...	
<input type="checkbox"/>		UP	h ce07-htc.cr.cnaf.infn.it					1h 46m	1/3 (H)	14m 10s	OK - condor_ping to ce07-htc.cr.cnaf...	
<input type="checkbox"/>		UP	h htc-belle-ce02.na.infn.it					5d 20h	1/3 (H)	14m 10s	OK - condor_ping to htc-belle-ce02.n...	

STORAGE ELEMENT - LS VIA DAVS WITH TOKEN AUTHENTICATION

 New filter

 host_group:storage_token 

 ACKNOWLEDGE  SET DOWNTIME  CHECK    Rows per page 30 1-2 of 2  

<input type="checkbox"/>	s	Status ↑	Resource	Parent	N	A	G	Duration	Tries	Last check	Information	State
<input type="checkbox"/>		UP	h ccdcacli303.in2p3.fr					5m 29s	1/3 (H)	29s	monitor\nTPC\ngfal-ls davs://ccdcacli303.in...	
<input type="checkbox"/>		UP	h xfer-archive.cr.cnaf.infn.it					11m 14s	1/3 (H)	3m 54s	e0003\nmonitor\nTPC\nDC\nddm_test\ngfal...	

Other ongoing activities

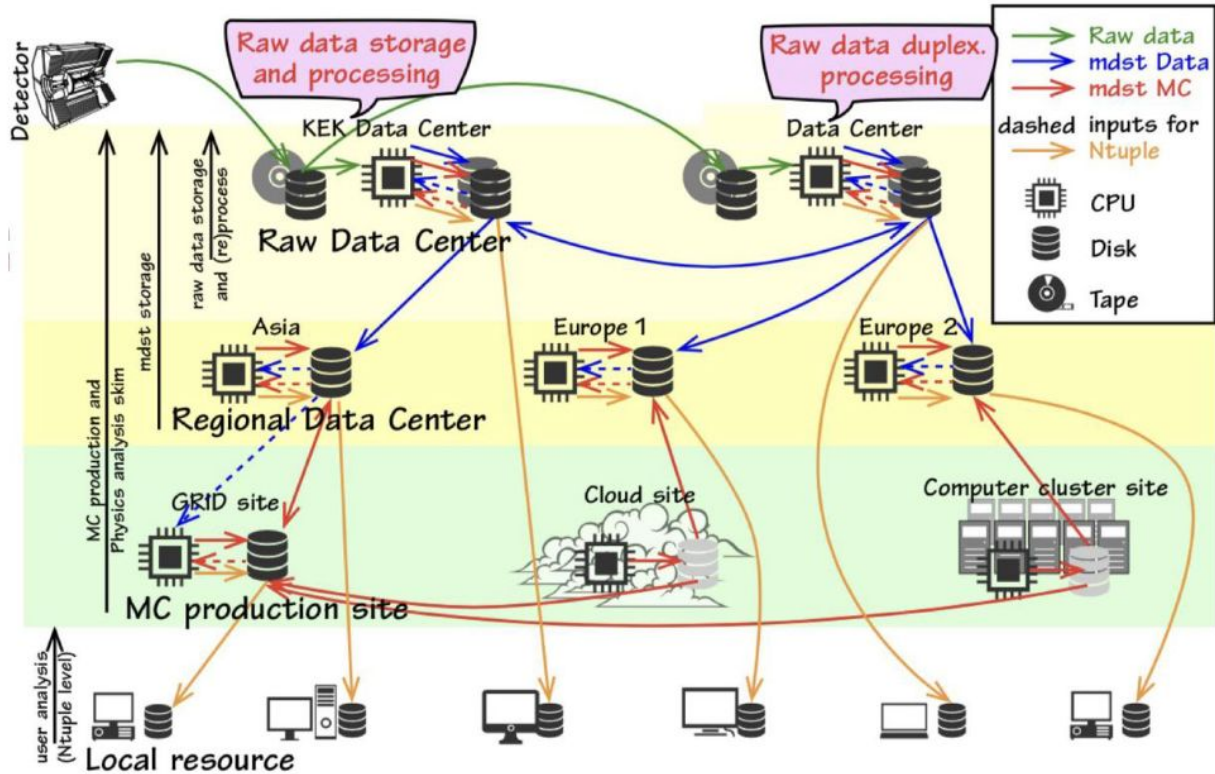
- Enabling Multicore jobs
- Integration of additional Rucio features into our workflow: Metadata in Rucio, data popularity, user quota.
- Review of the scalability in the user analysis towards a x10 luminosity scenario in 2026.
- Improving automatisisation of production activities
- Migration to DIRAC 7.3 and to 8.0

Conclusions

- Belle II is a large International collaboration.
- Data Processing and analysis is done over a distributed computing infrastructure.
- 56 sites providing Computing and Storage resources, 6 of them are Raw Data Centers
- Access to LHCONE for the largest sites.
- Continuous update of the computing infrastructure
- 2 PB of data has been collected so far, at the maximum luminosity we expect to collect 10PB/year.

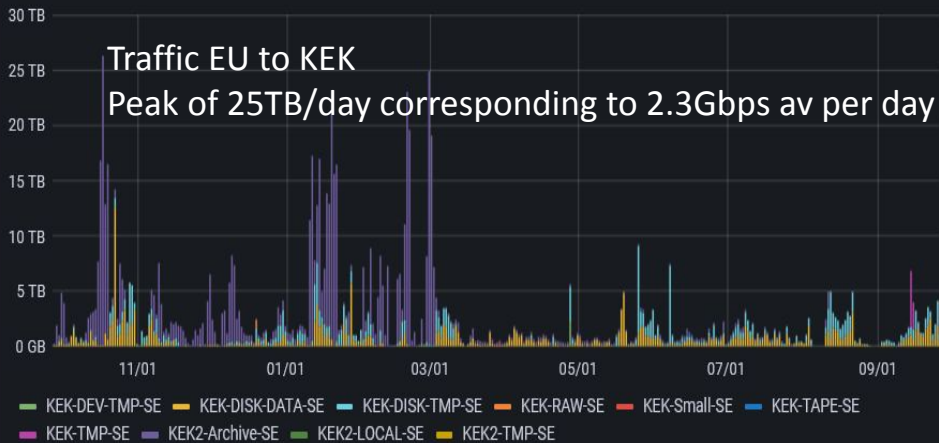
Backup

The Belle II distributed computing model



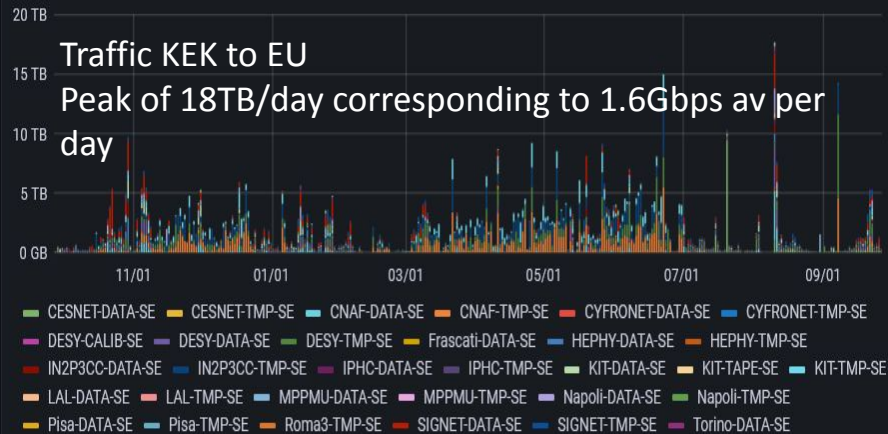
Successful transfers volume (destination)

Traffic EU to KEK
Peak of 25TB/day corresponding to 2.3Gbps av per day

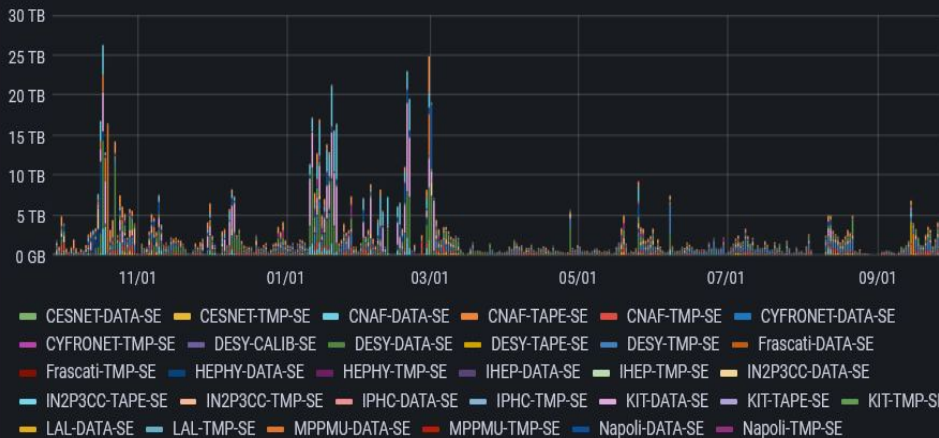


Successful transfers volume (destination)

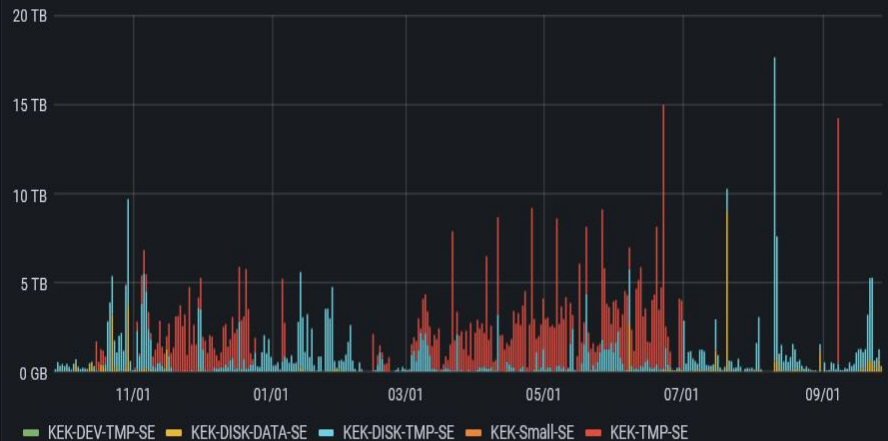
Traffic KEK to EU
Peak of 18TB/day corresponding to 1.6Gbps av per day



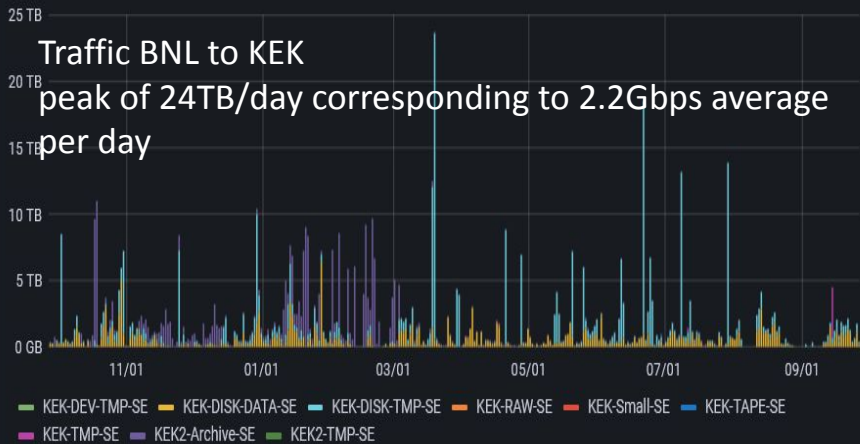
Successful transfers volume (source)



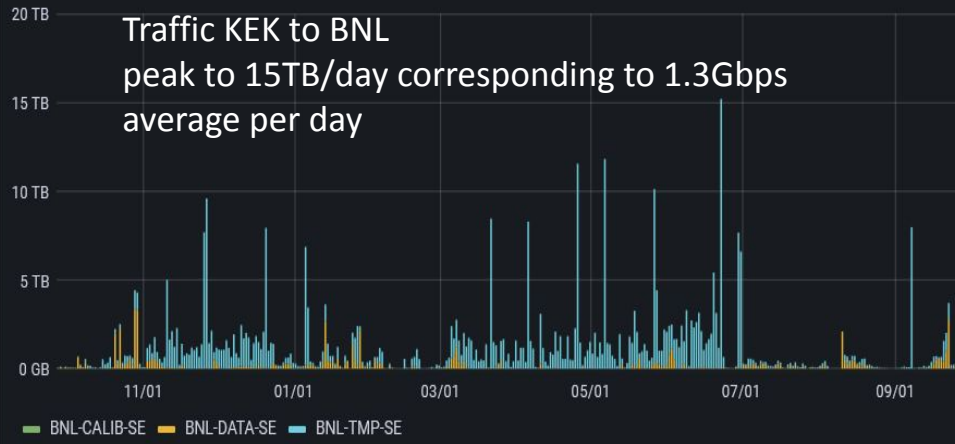
Successful transfers volume (source)



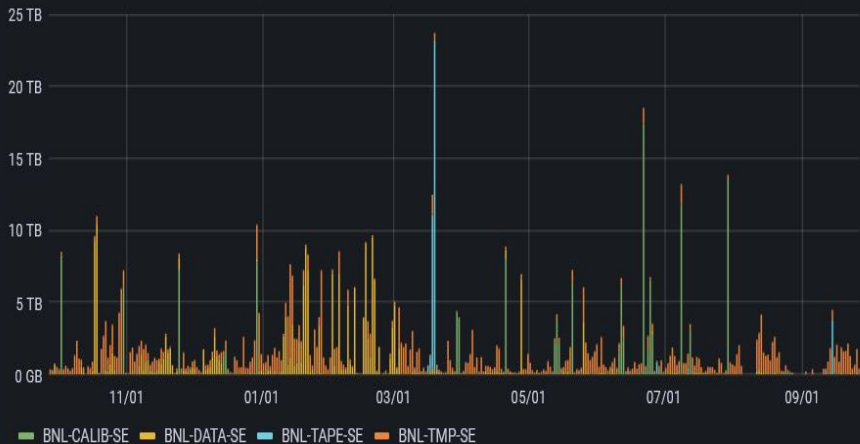
Successful transfers volume (destination)



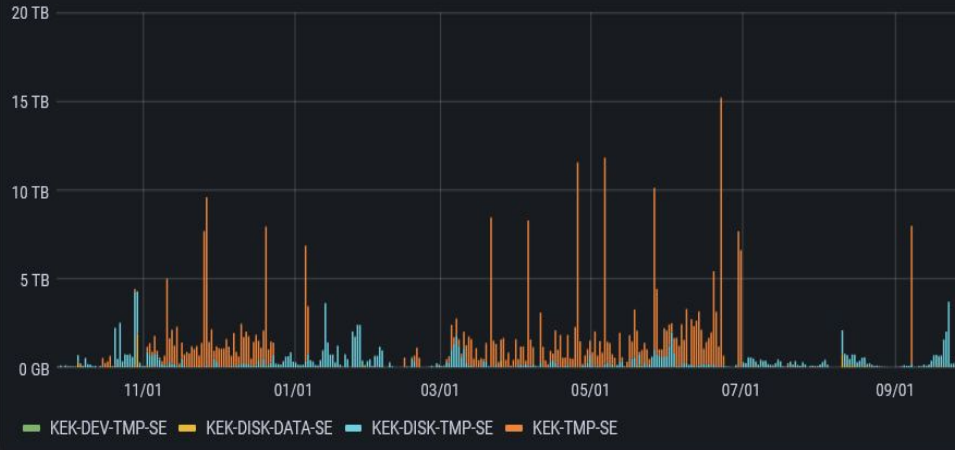
Successful transfers volume (destination)



Successful transfers volume (source)



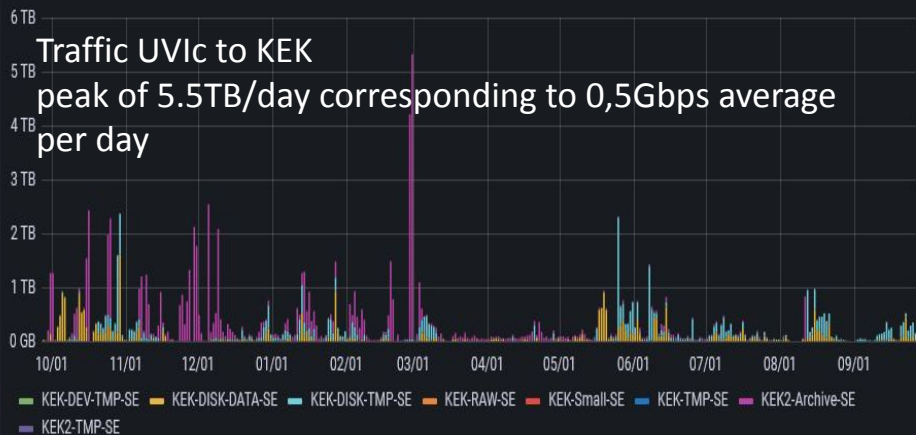
Successful transfers volume (source)



Successful transfers volume (destination)

Traffic UVic to KEK

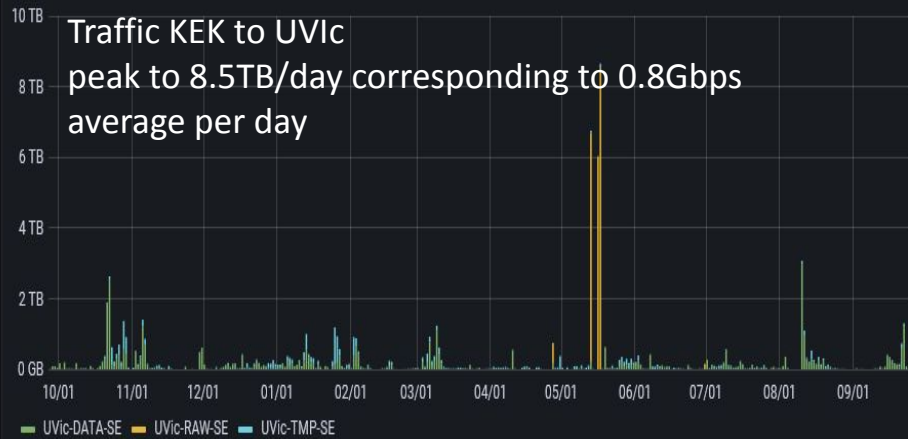
peak of 5.5TB/day corresponding to 0,5Gbps average per day



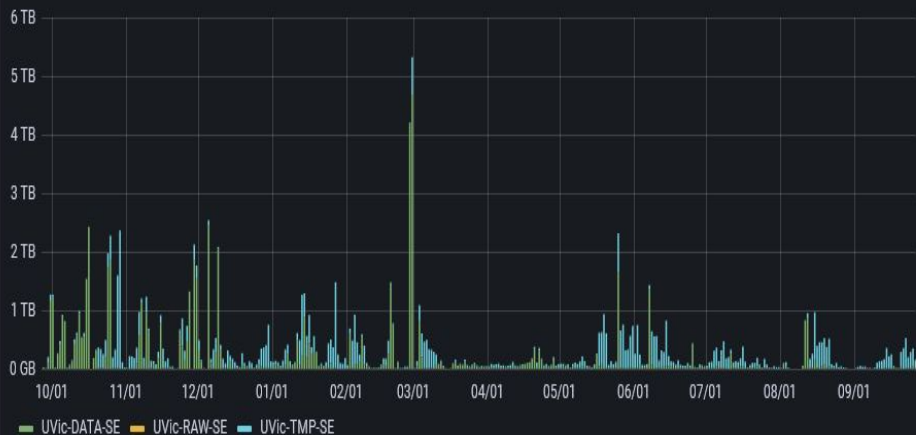
Successful transfers volume (destination)

Traffic KEK to UVic

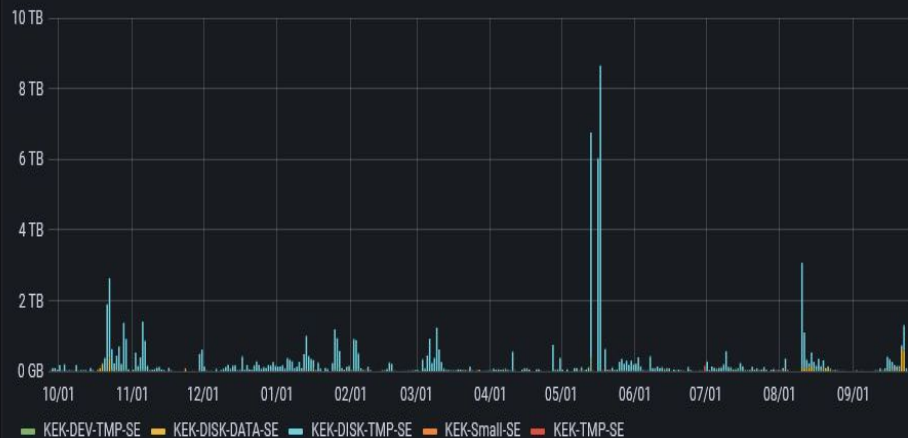
peak to 8.5TB/day corresponding to 0.8Gbps average per day



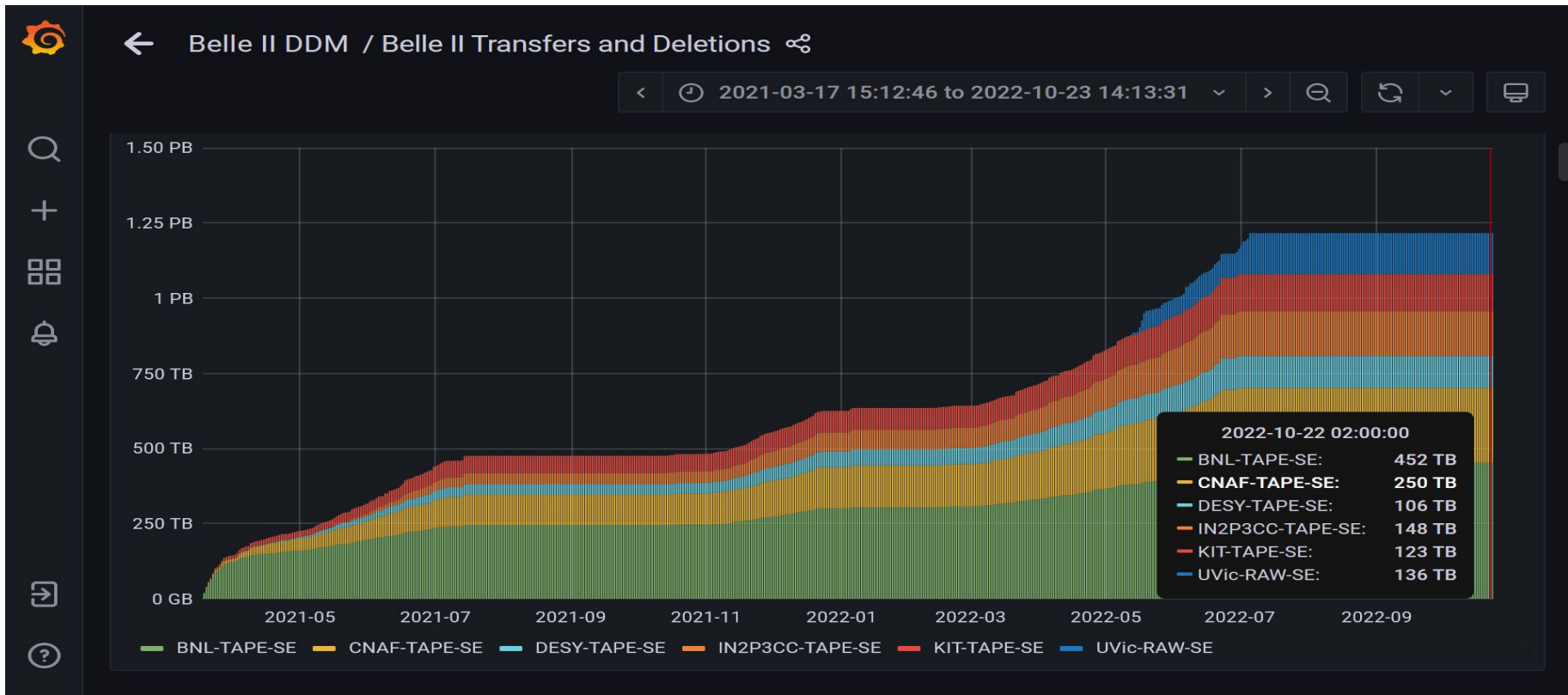
Successful transfers volume (source)



Successful transfers volume (source)



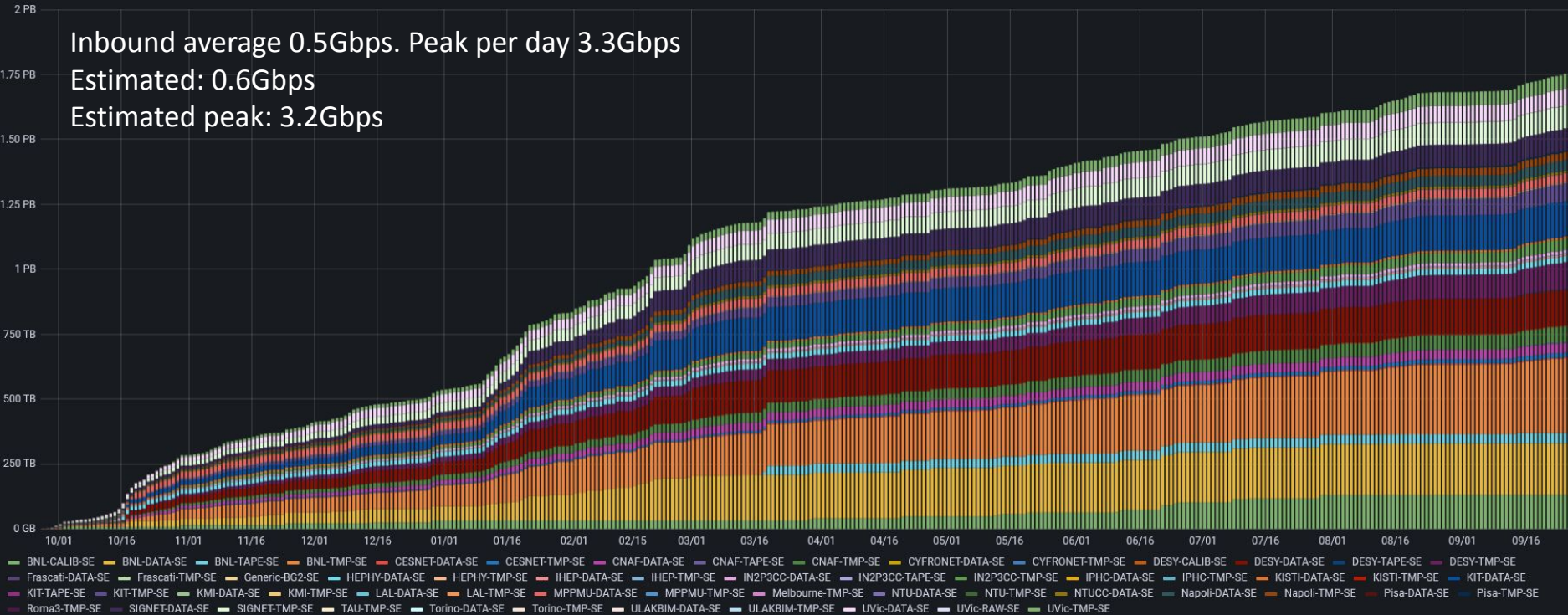
Raw Data Distribution with Rucio



Global traffic to KEK in the last 12 month

Successful transfers volume per source (aggregation)

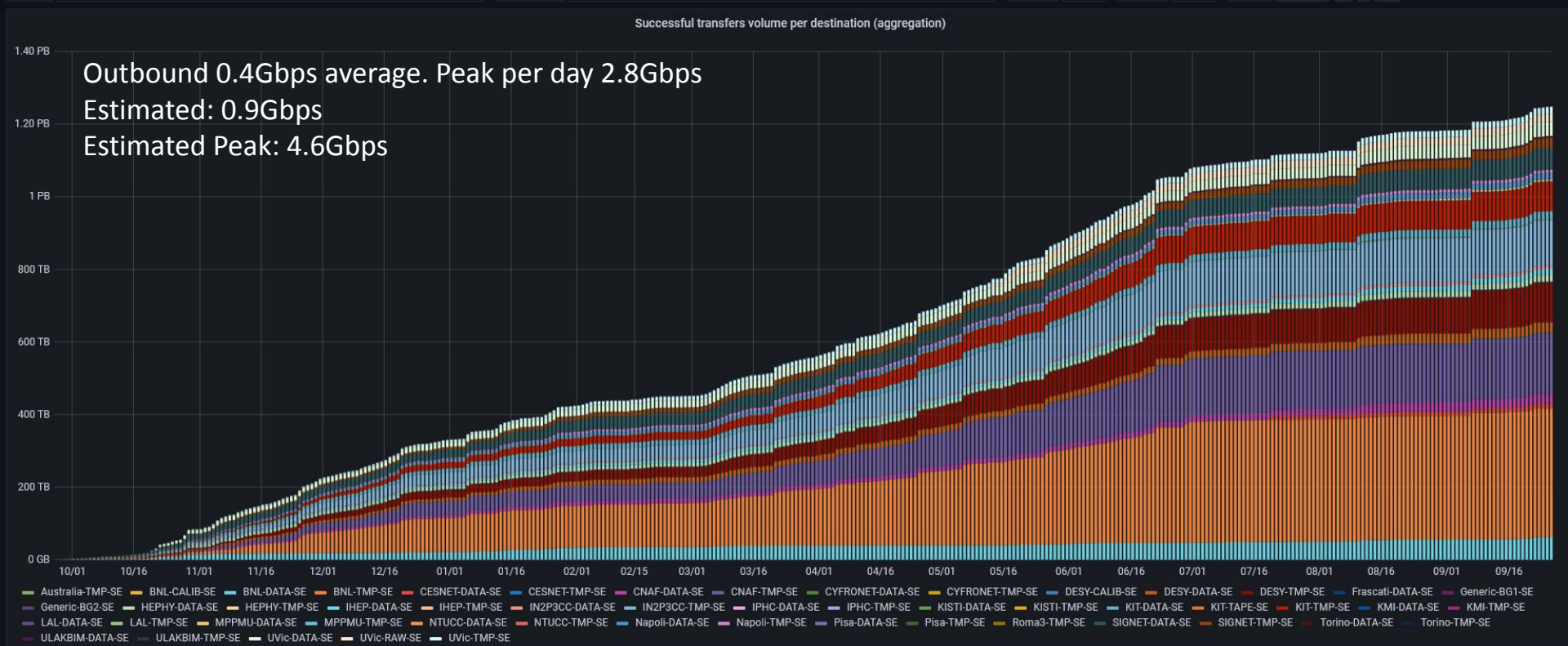
Inbound average 0.5Gbps. Peak per day 3.3Gbps
 Estimated: 0.6Gbps
 Estimated peak: 3.2Gbps



Global traffic from KEK in the last 12 month

Successful transfers volume per destination (aggregation)

Outbound 0.4Gbps average. Peak per day 2.8Gbps
 Estimated: 0.9Gbps
 Estimated Peak: 4.6Gbps

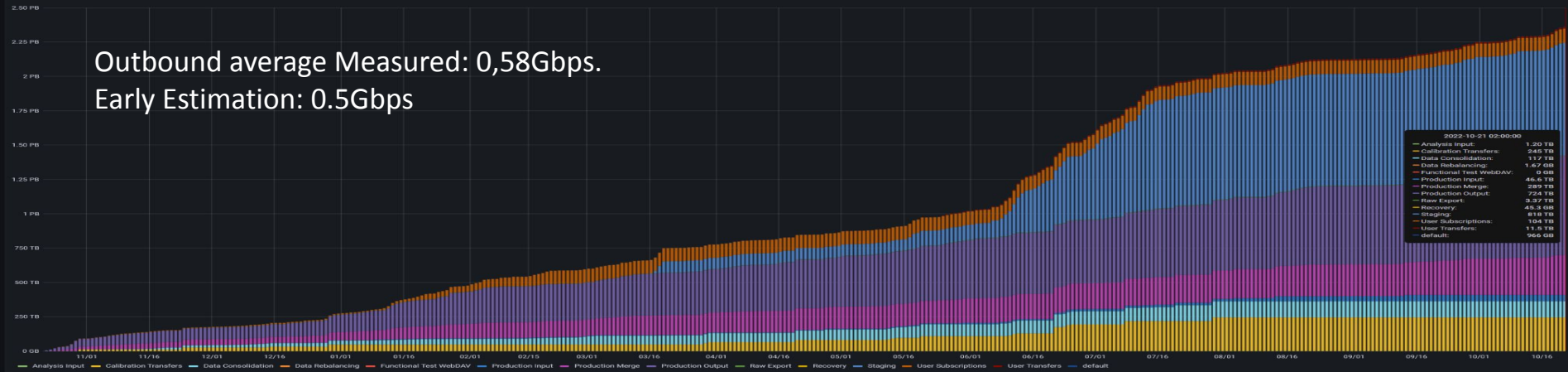


Outbound traffic BNL in the last 12 month



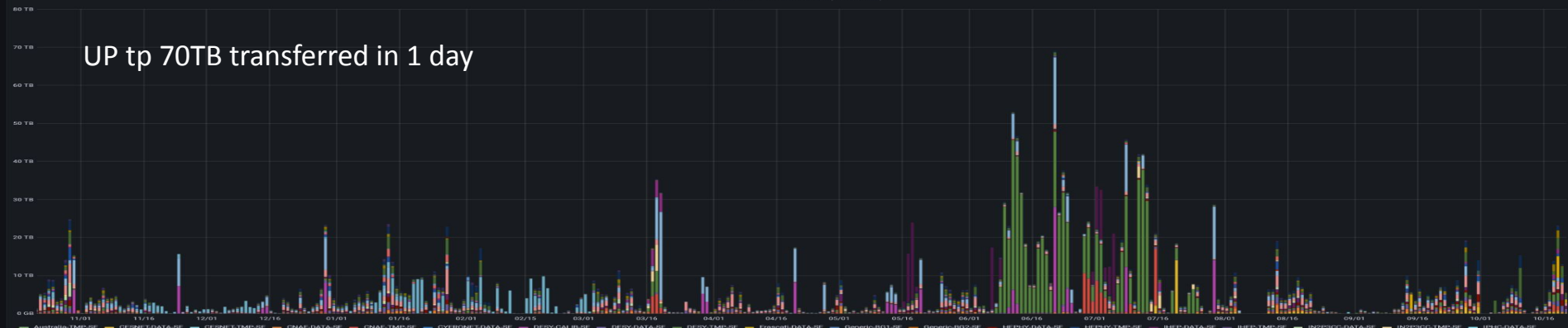
Outbound average Measured: 0,58Gbps.
Early Estimation: 0.5Gbps

Successful transfers volume per activity (aggregation)



Successful transfers volume (destination)

UP to 70TB transferred in 1 day

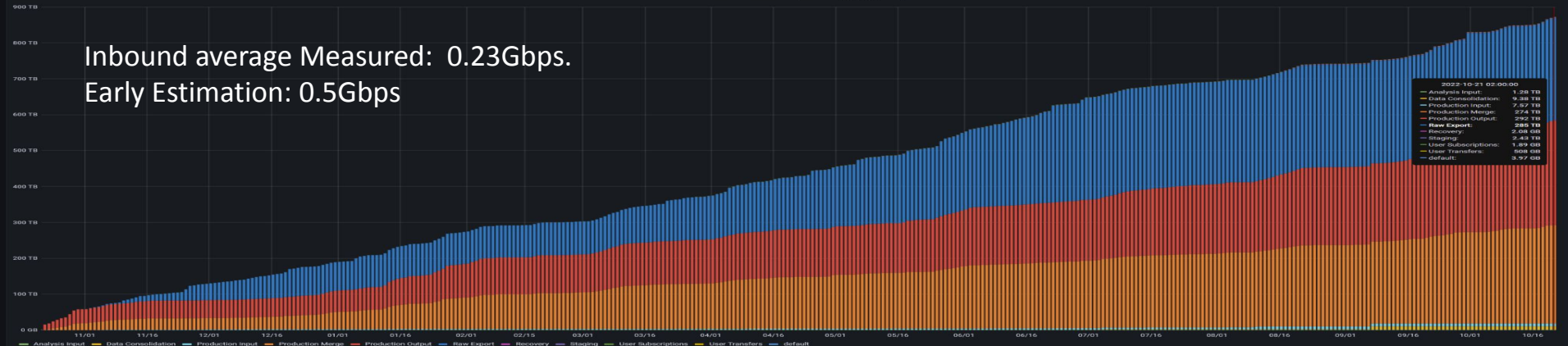




Inbound traffic BNL in the last 12 month

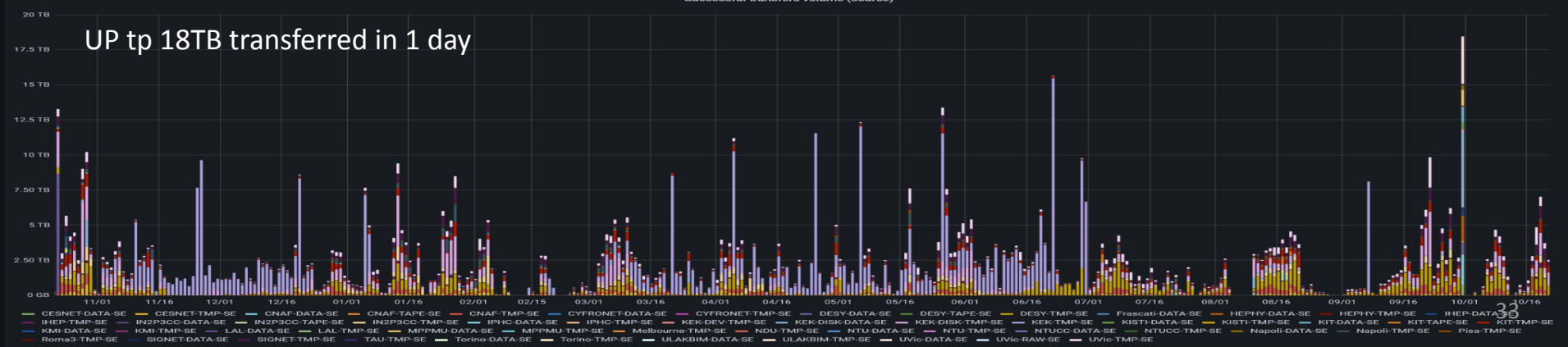
Inbound average Measured: 0.23Gbps.
Early Estimation: 0.5Gbps

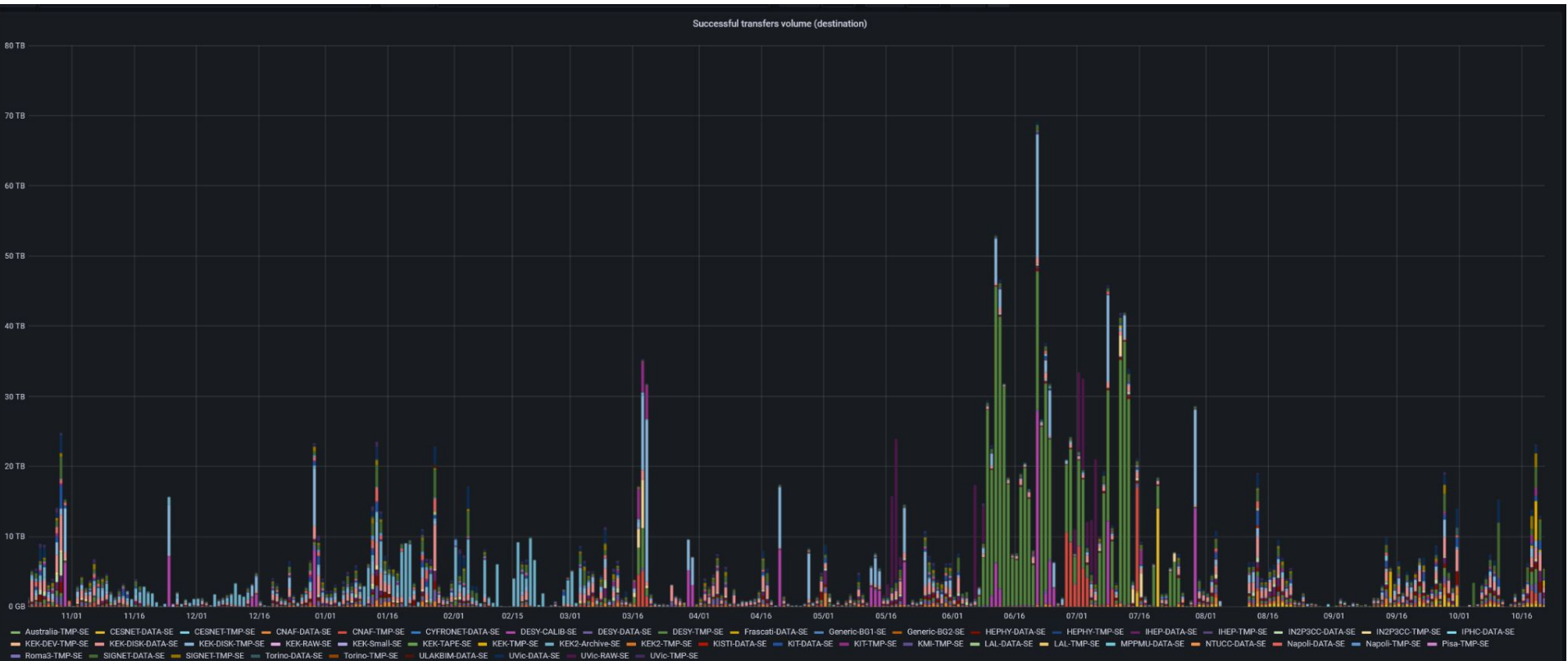
Successful transfers volume per activity (aggregation)



UP to 18TB transferred in 1 day

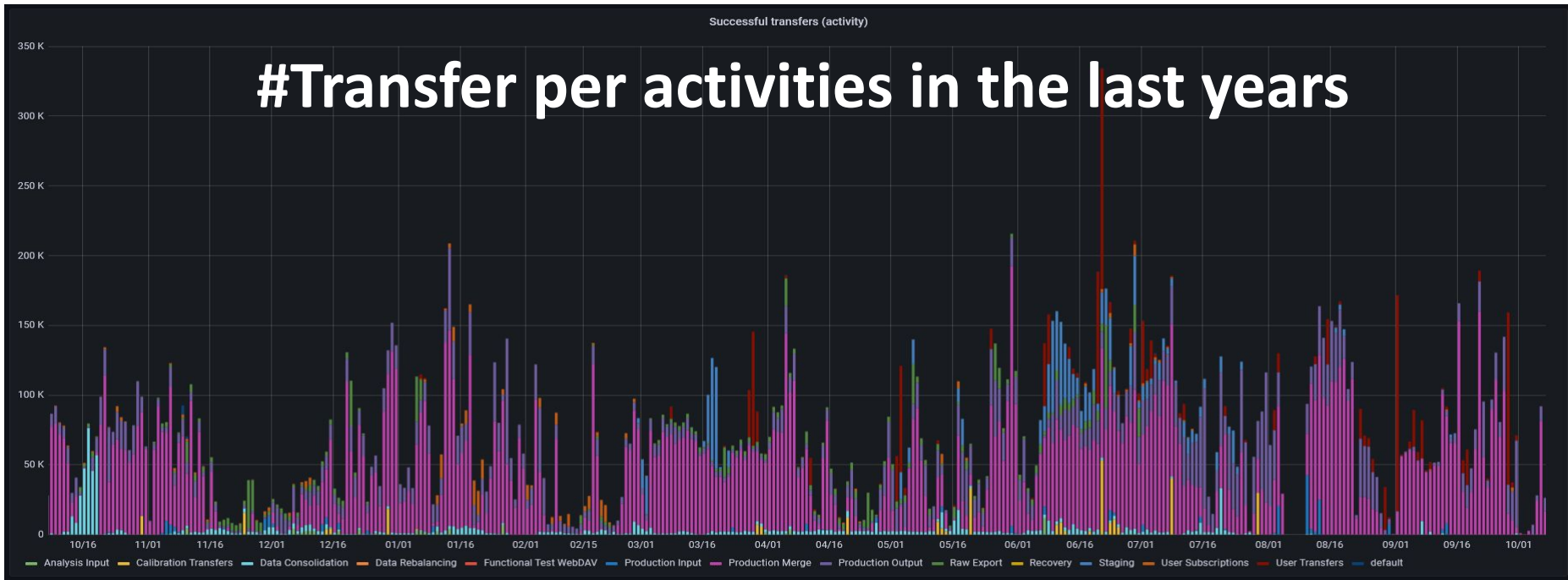
Successful transfers volume (source)





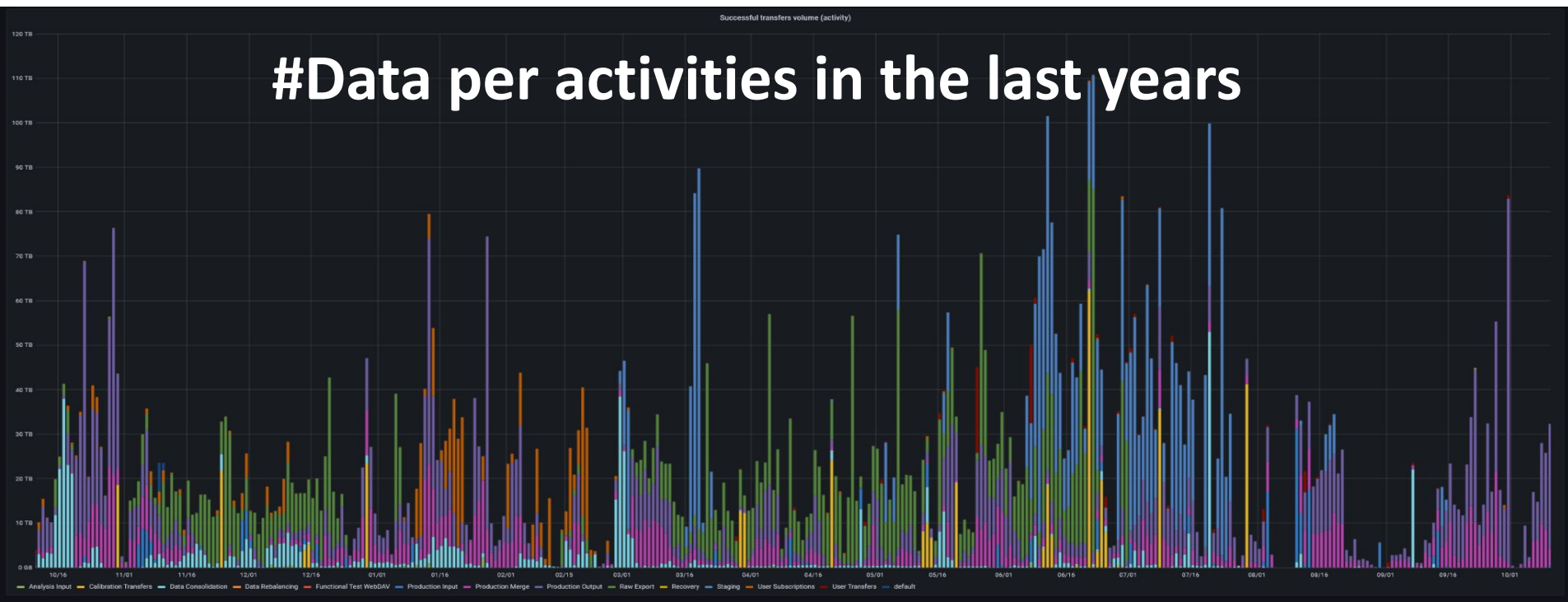
Rucio monitoring system

#Transfer per activities in the last years

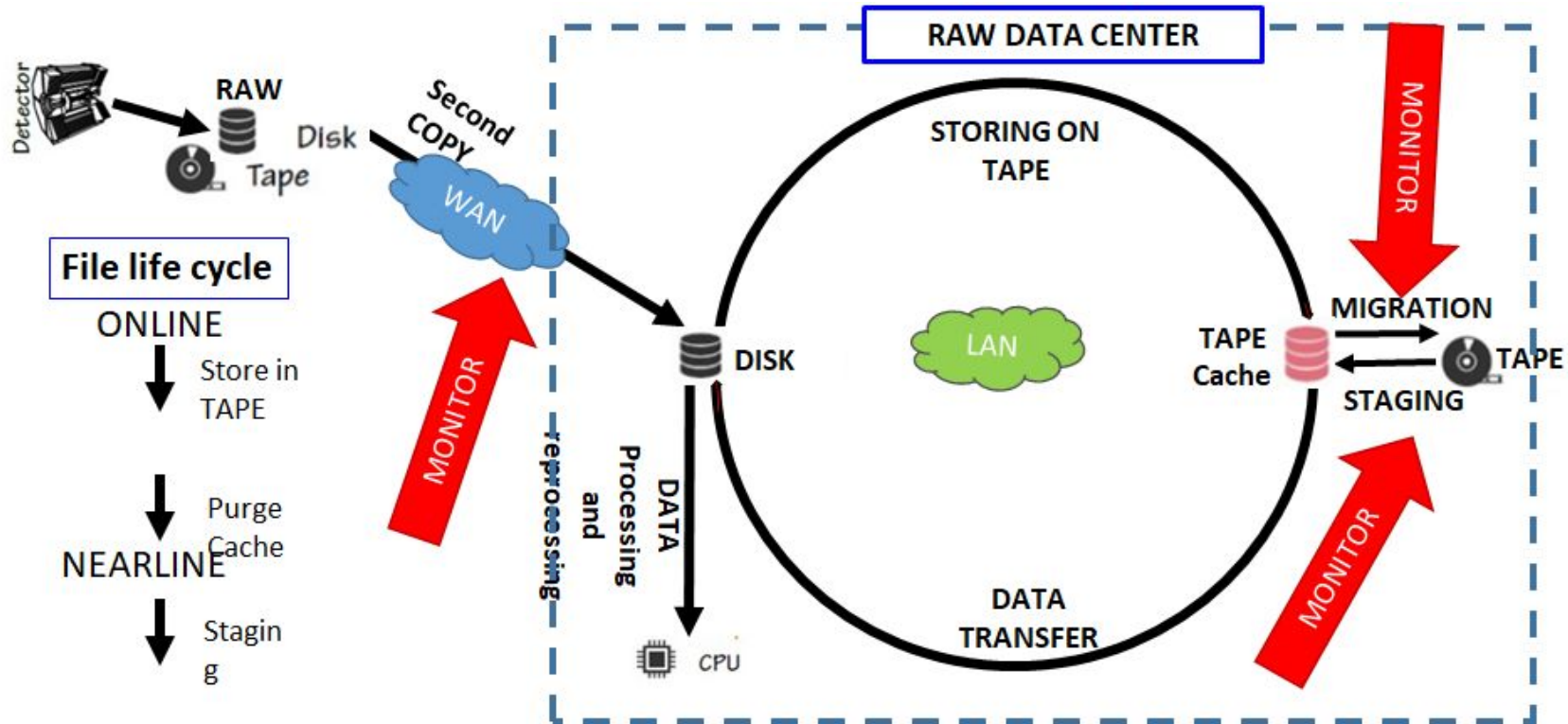


<https://monitoring.sdcc.bnl.gov/pub/grafana/d/belle2xfers/belle-ii-transfers-and-deletions?orgId=1>

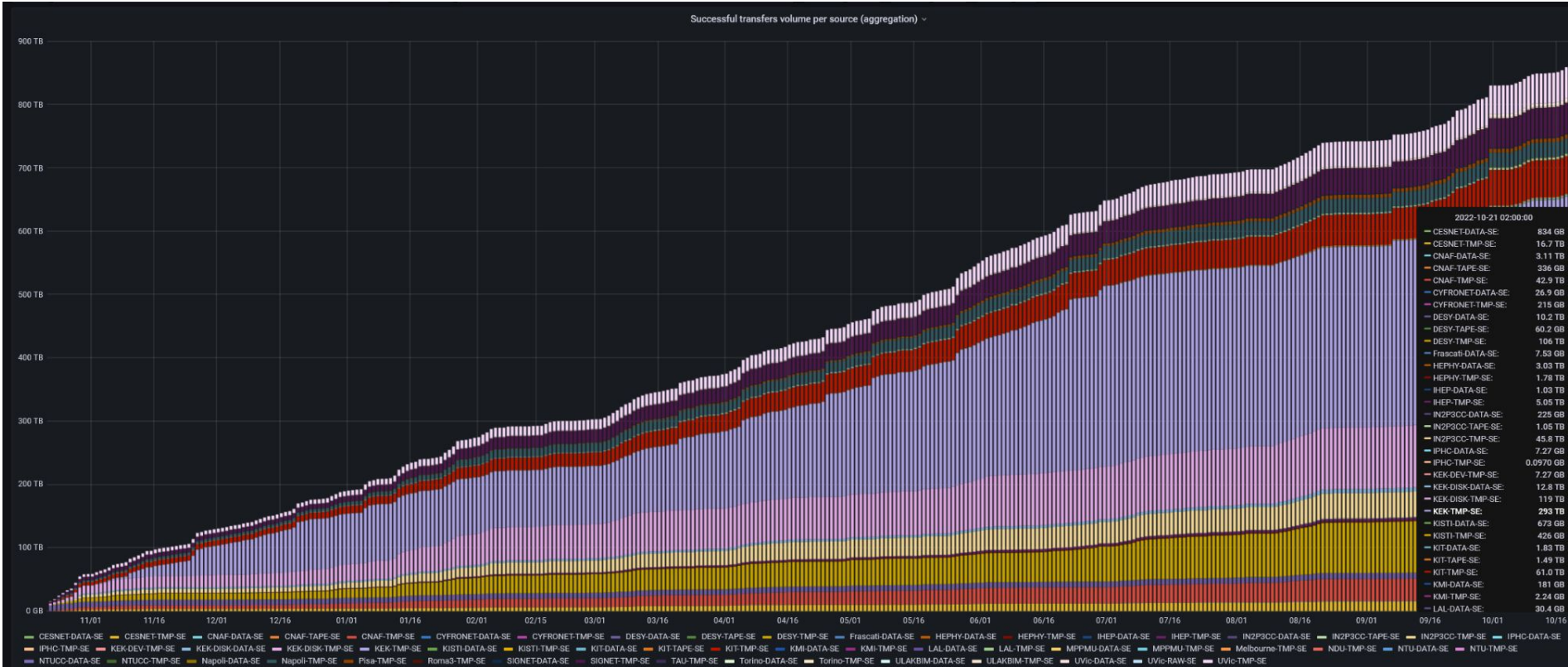
Rucio monitoring system



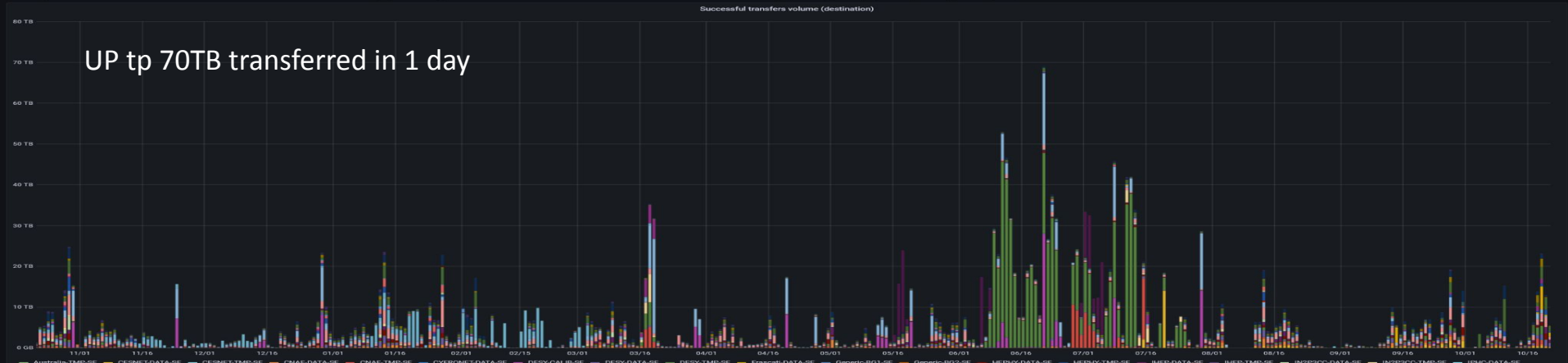
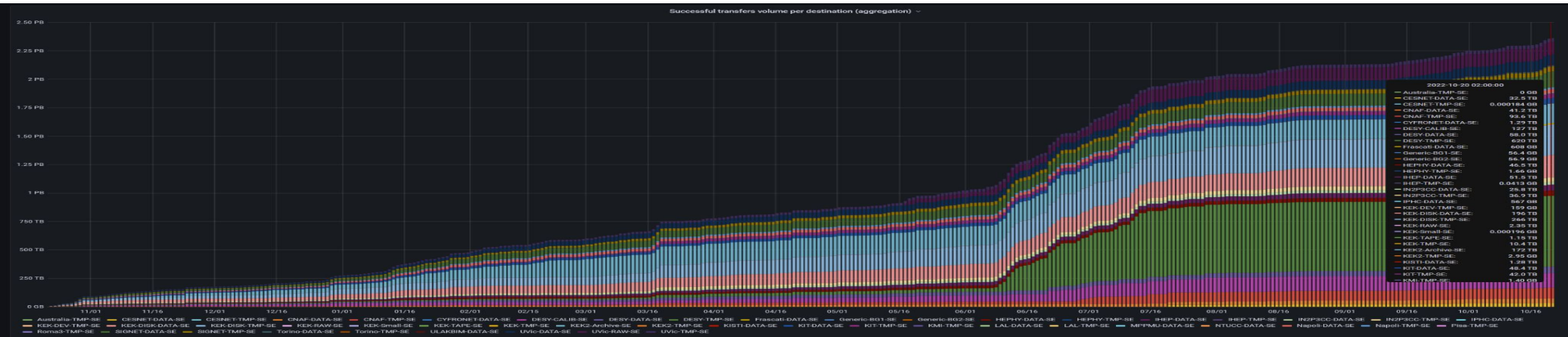
Raw Data Cycle



Inbound traffic BNL in the last 12 month

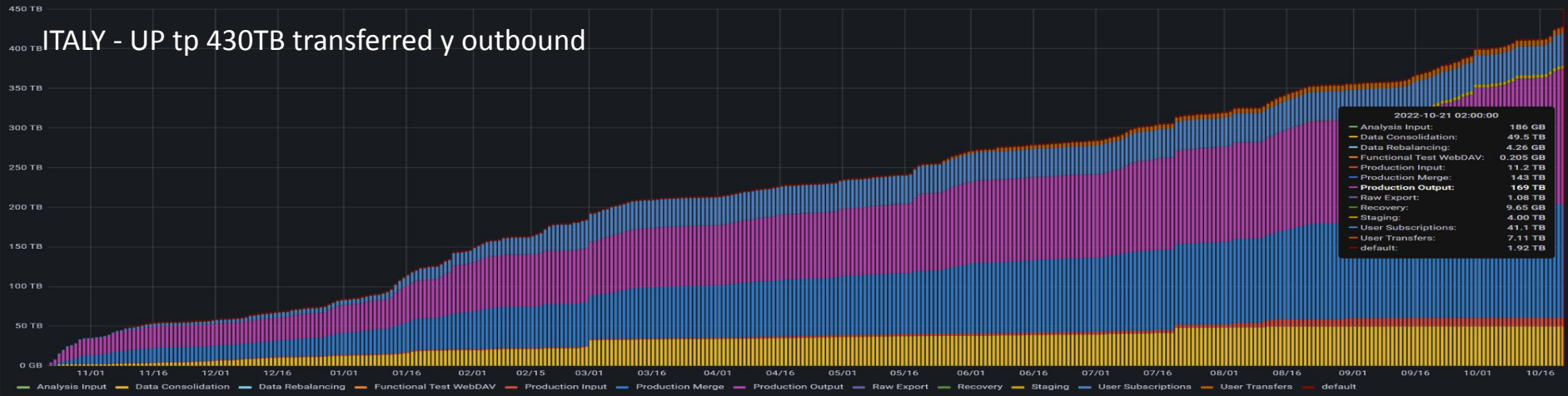


Outbound traffic BNL in the last 12 month



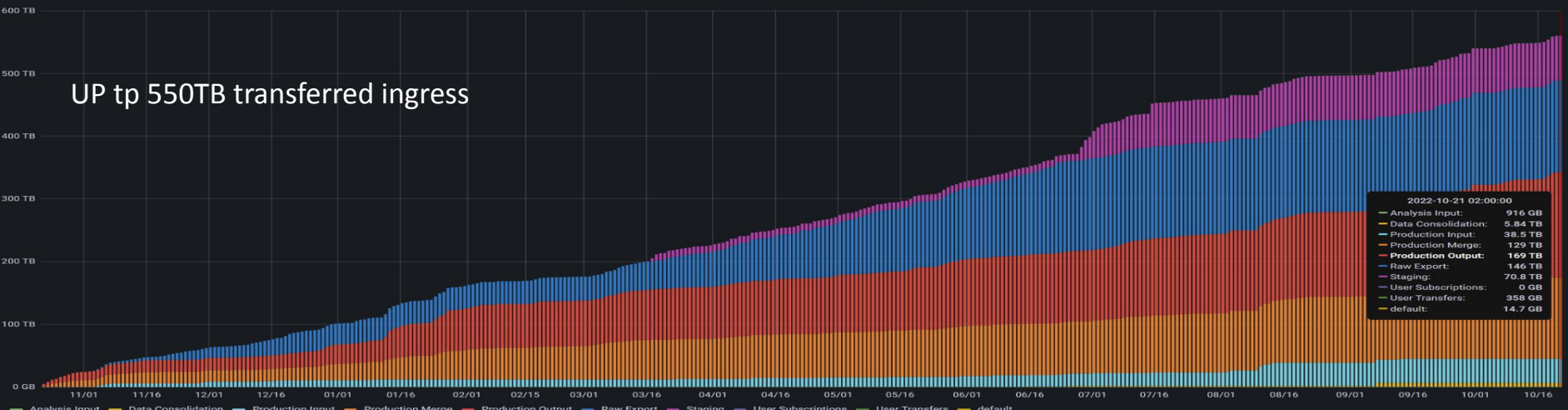
Successful transfers volume per activity (aggregation) -

ITALY - UP tp 430TB transferred y outbound



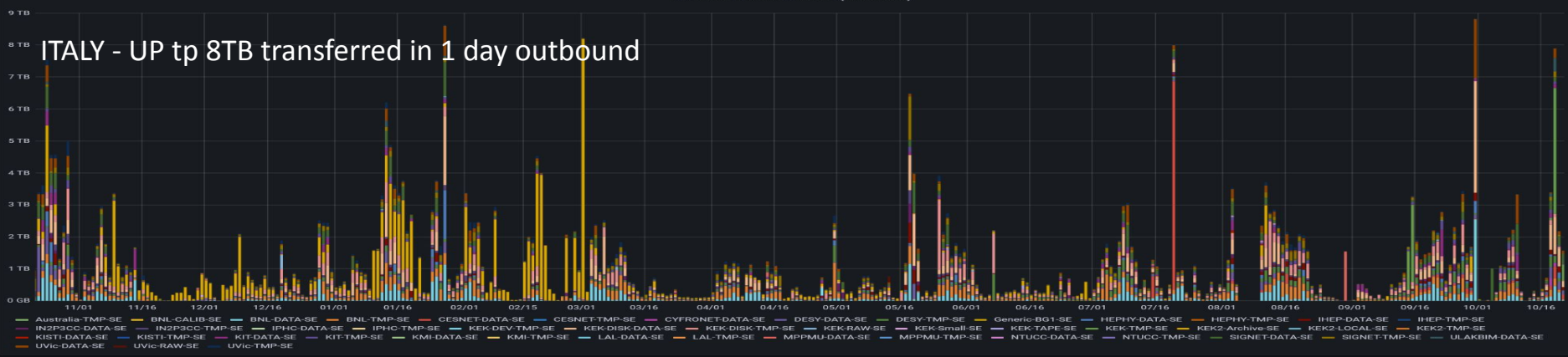
Successful transfers volume per activity (aggregation) -

UP tp 550TB transferred ingress



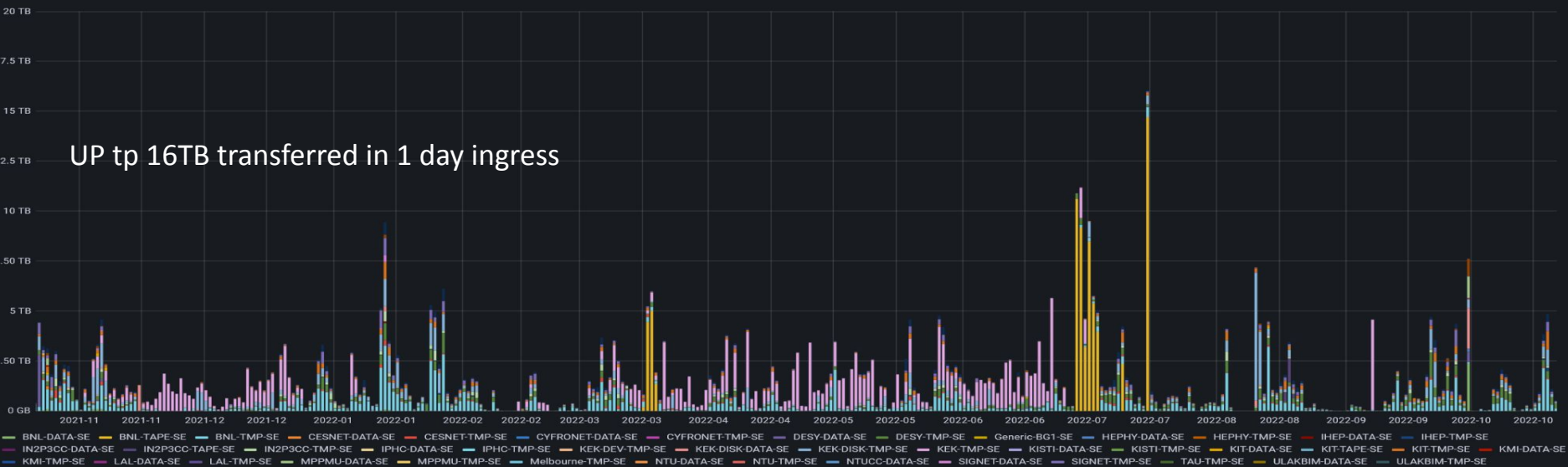
Successful transfers volume (destination)

ITALY - UP tp 8TB transferred in 1 day outbound



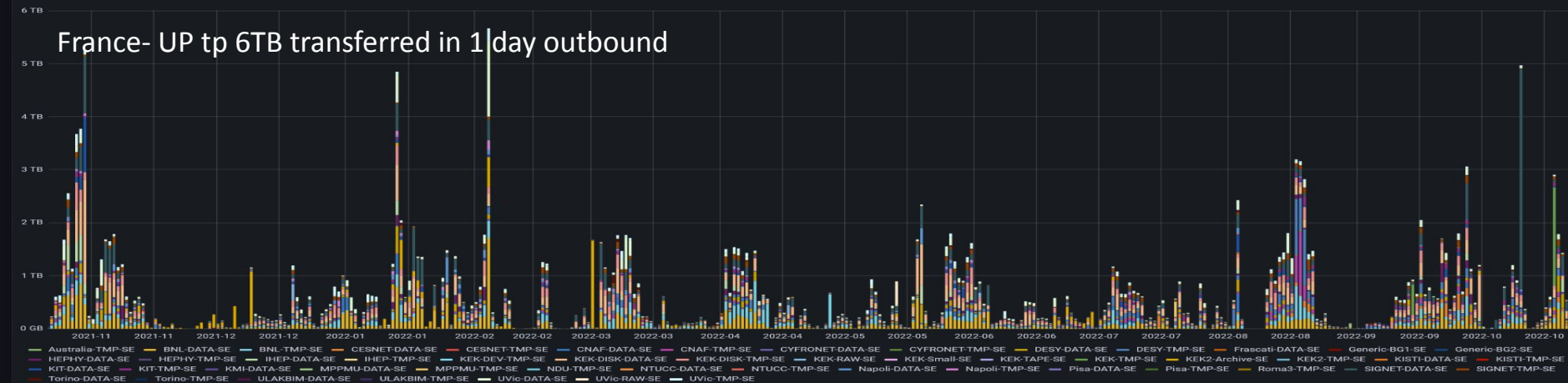
Successful transfers volume (source)

UP tp 16TB transferred in 1 day ingress



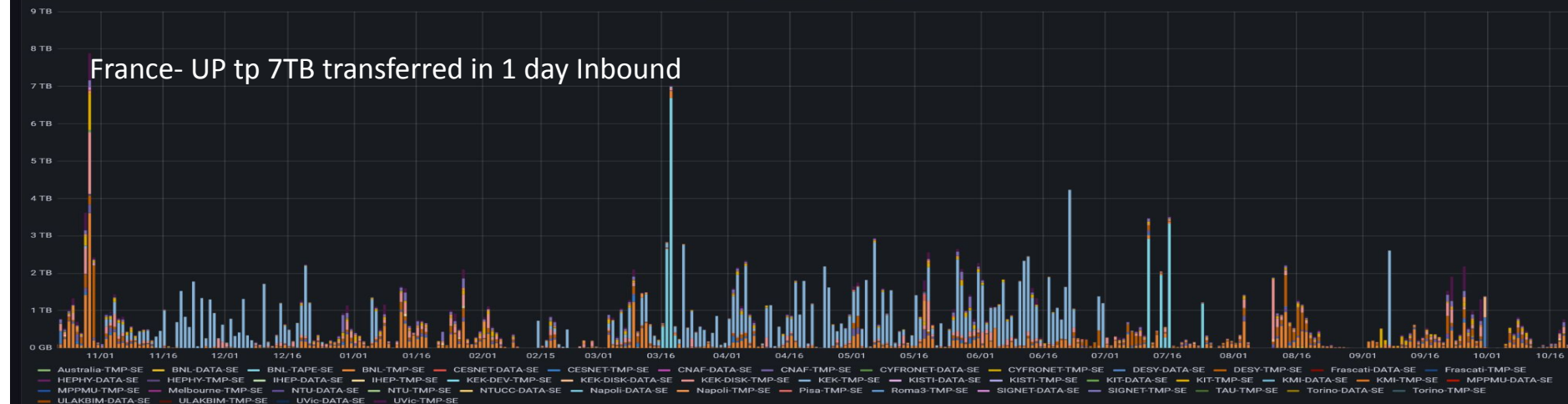
Successful transfers volume (destination)

France- UP tp 6TB transferred in 1 day outbound

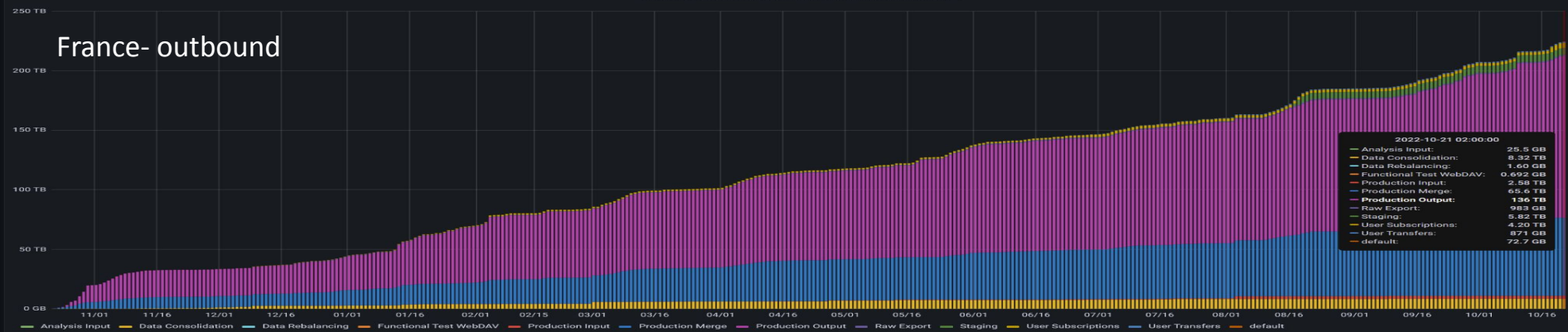


Successful transfers volume (source)

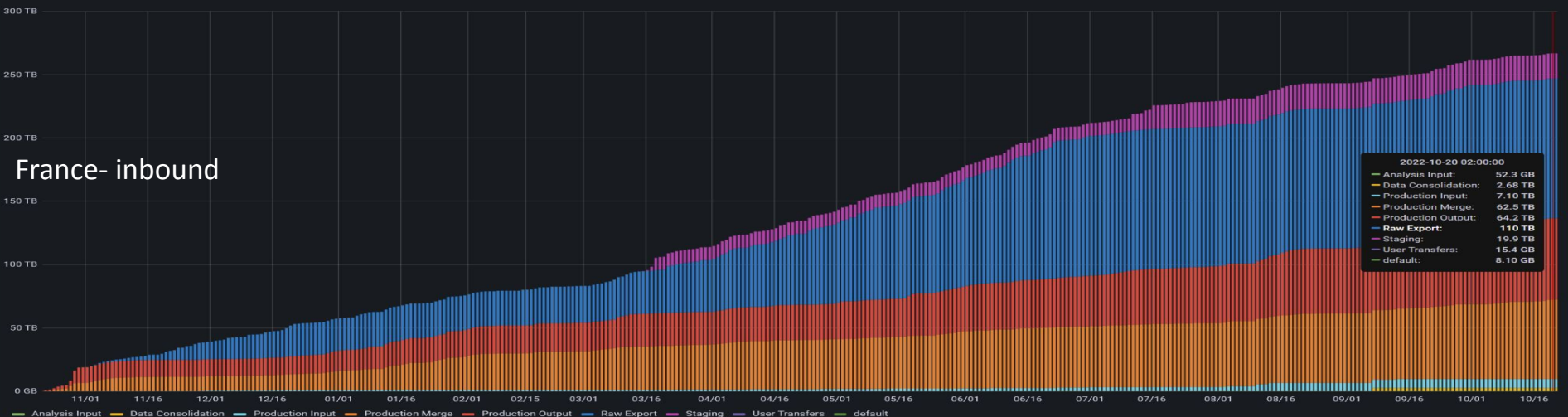
France- UP tp 7TB transferred in 1 day Inbound



France- outbound

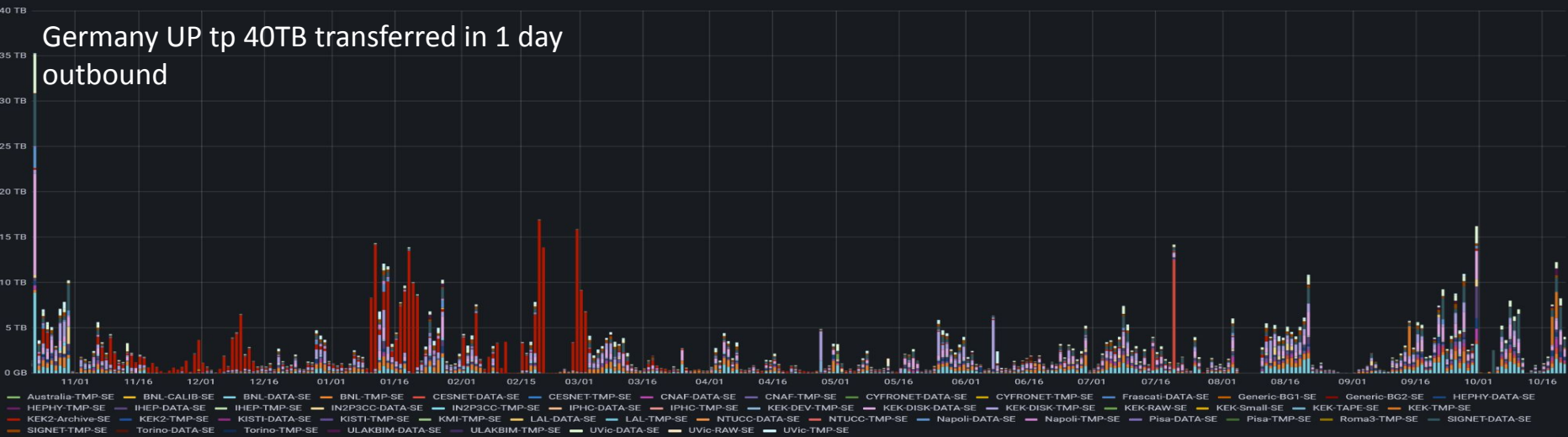


France- inbound



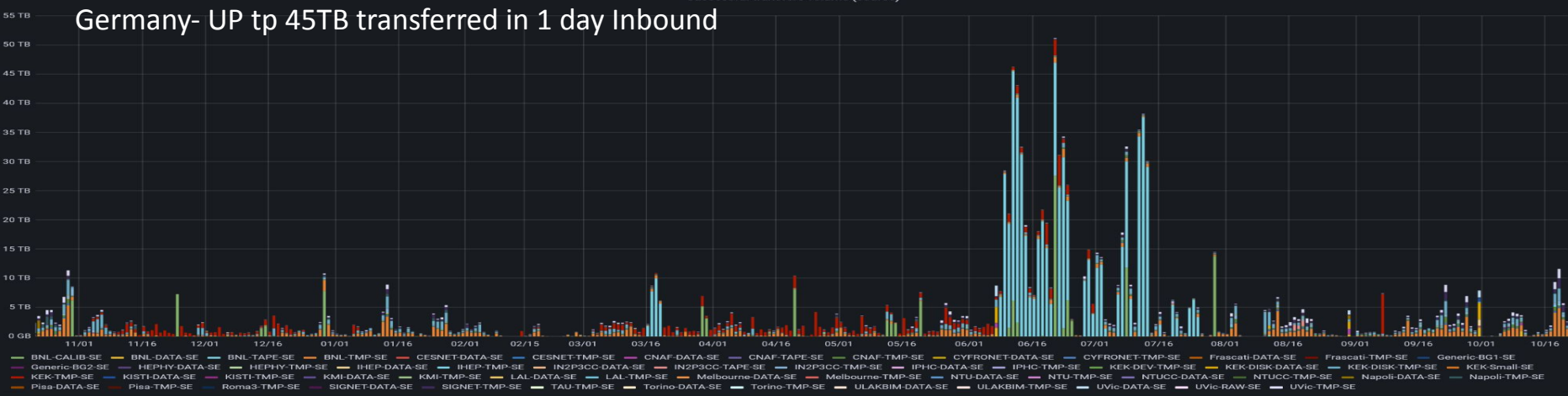
Successful transfers volume (destination)

Germany UP tp 40TB transferred in 1 day
outbound

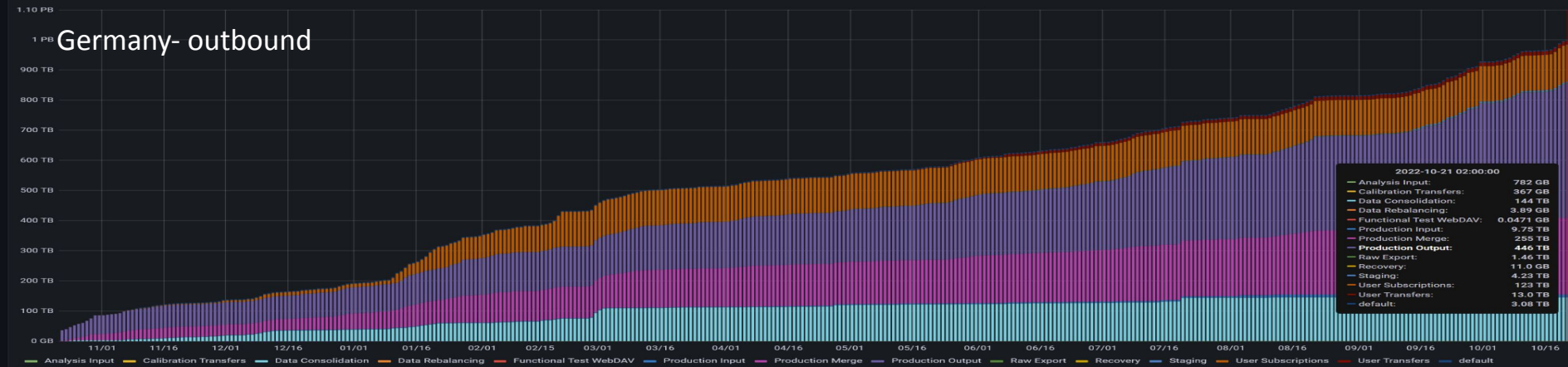


Successful transfers volume (source)

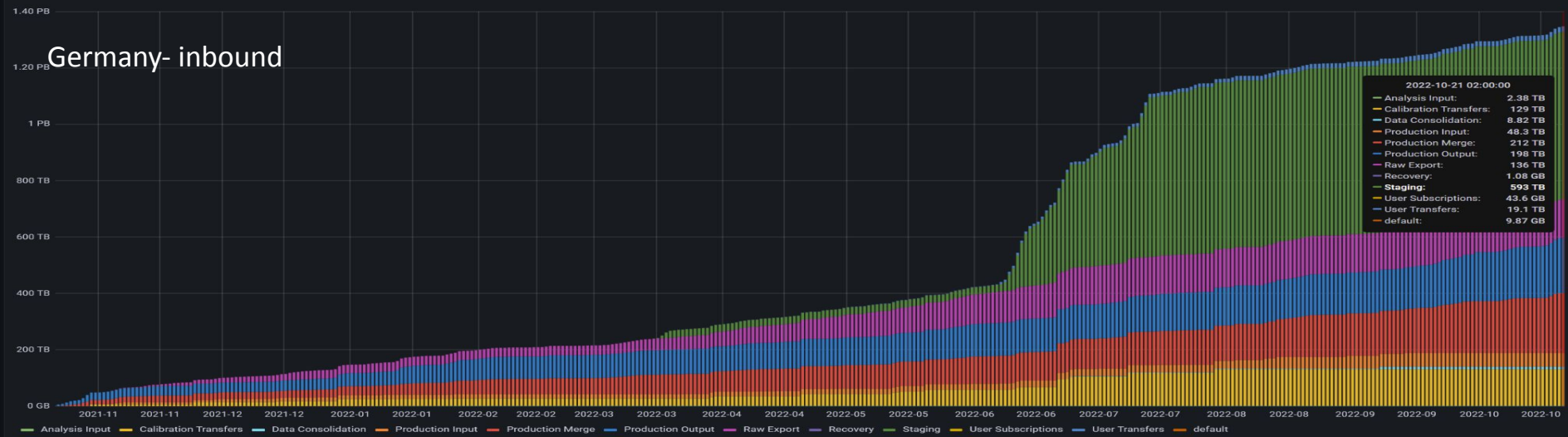
Germany- UP tp 45TB transferred in 1 day Inbound



Germany- outbound



Germany- inbound



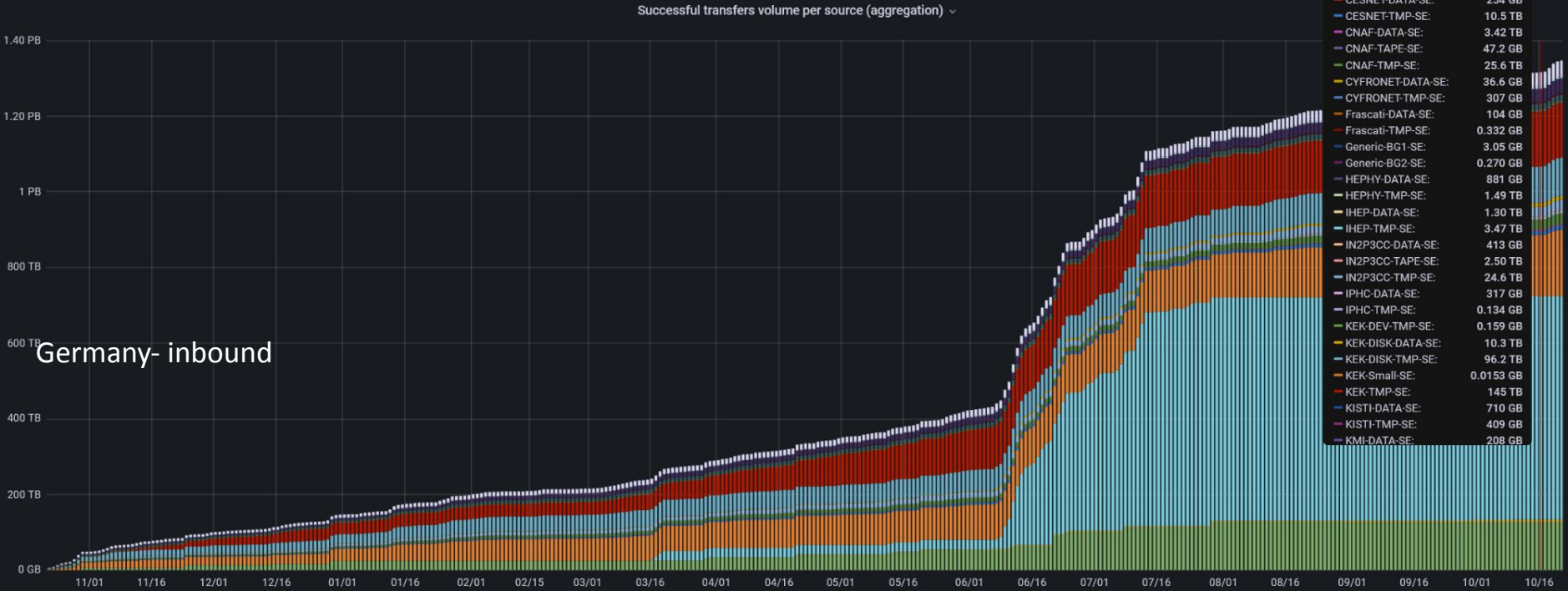


Source Australia-TMP-SE + BNL-CALIB-SE + BNL-DATA-SE + BNL-TAPE-SE + BN...

Destination DESY-CALIB-SE + DESY-DATA-SE + DESY-TAPE-SE + DESY-TAPE-TEST + ...

Activity All Binning 1d Filter

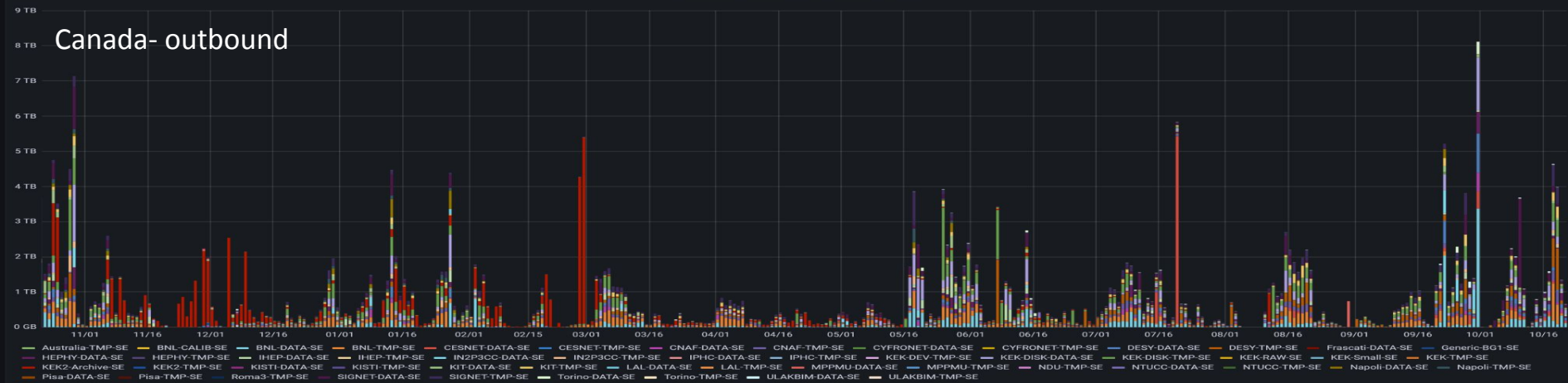
BNL-CALIB-SE:	127 TB
BNL-DATA-SE:	6.43 TB
BNL-TAPE-SE:	590 TB
BNL-TMP-SE:	161 TB
CESNET-DATA-SE:	254 GB
CESNET-TMP-SE:	10.5 TB
CNAF-DATA-SE:	3.42 TB
CNAF-TAPE-SE:	47.2 GB
CNAF-TMP-SE:	25.6 TB
CYFRONET-DATA-SE:	36.6 GB
CYFRONET-TMP-SE:	307 GB
Frascati-DATA-SE:	104 GB
Frascati-TMP-SE:	0.332 GB
Generic-BG1-SE:	3.05 GB
Generic-BG2-SE:	0.270 GB
HEPHY-DATA-SE:	881 GB
HEPHY-TMP-SE:	1.49 TB
IHEP-DATA-SE:	1.30 TB
IHEP-TMP-SE:	3.47 TB
IN2P3CC-DATA-SE:	413 GB
IN2P3CC-TAPE-SE:	2.50 TB
IN2P3CC-TMP-SE:	24.6 TB
IPHC-DATA-SE:	317 GB
IPHC-TMP-SE:	0.134 GB
KEK-DEV-TMP-SE:	0.159 GB
KEK-DISK-DATA-SE:	10.3 TB
KEK-DISK-TMP-SE:	96.2 TB
KEK-Small-SE:	0.0153 GB
KEK-TMP-SE:	145 TB
KISTI-DATA-SE:	710 GB
KISTI-TMP-SE:	409 GB
KMI-DATA-SE:	208 GB



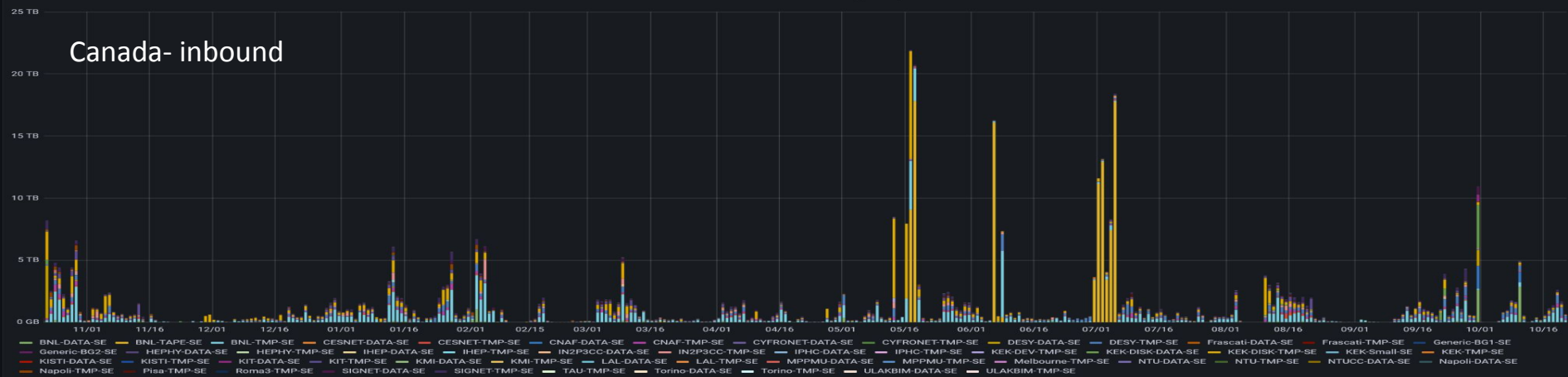
Germany-inbound

- BNL-CALIB-SE
- BNL-DATA-SE
- BNL-TAPE-SE
- BNL-TMP-SE
- CESNET-DATA-SE
- CESNET-TMP-SE
- CNAF-DATA-SE
- CNAF-TAPE-SE
- CNAF-TMP-SE
- CYFRONET-DATA-SE
- CYFRONET-TMP-SE
- Frascati-DATA-SE
- Frascati-TMP-SE
- Generic-BG1-SE
- Generic-BG2-SE
- HEPHY-DATA-SE
- HEPHY-TMP-SE
- IHEP-DATA-SE
- IHEP-TMP-SE
- IN2P3CC-DATA-SE
- IN2P3CC-TAPE-SE
- IN2P3CC-TMP-SE
- IPHC-DATA-SE
- IPHC-TMP-SE
- KEK-DEV-TMP-SE
- KEK-DISK-DATA-SE
- KEK-DISK-TMP-SE
- KEK-Small-SE
- KEK-TMP-SE
- KISTI-DATA-SE
- KISTI-TMP-SE
- KMI-DATA-SE
- KMI-TMP-SE
- LAL-DATA-SE
- LAL-TMP-SE
- Melbourne-DATA-SE
- Melbourne-TMP-SE
- NTU-DATA-SE
- NTU-TMP-SE
- NTUCC-DATA-SE
- NTUCC-TMP-SE
- Napoli-DATA-SE
- Napoli-TMP-SE
- Pisa-DATA-SE
- Pisa-TMP-SE
- Roma3-TMP-SE
- SIGNET-DATA-SE
- SIGNET-TMP-SE
- TAU-TMP-SE
- Torino-DATA-SE
- Torino-TMP-SE
- ULAKBIM-DATA-SE
- ULAKBIM-TMP-SE
- UVic-DATA-SE
- UVic-RAW-SE
- UVic-TMP-SE

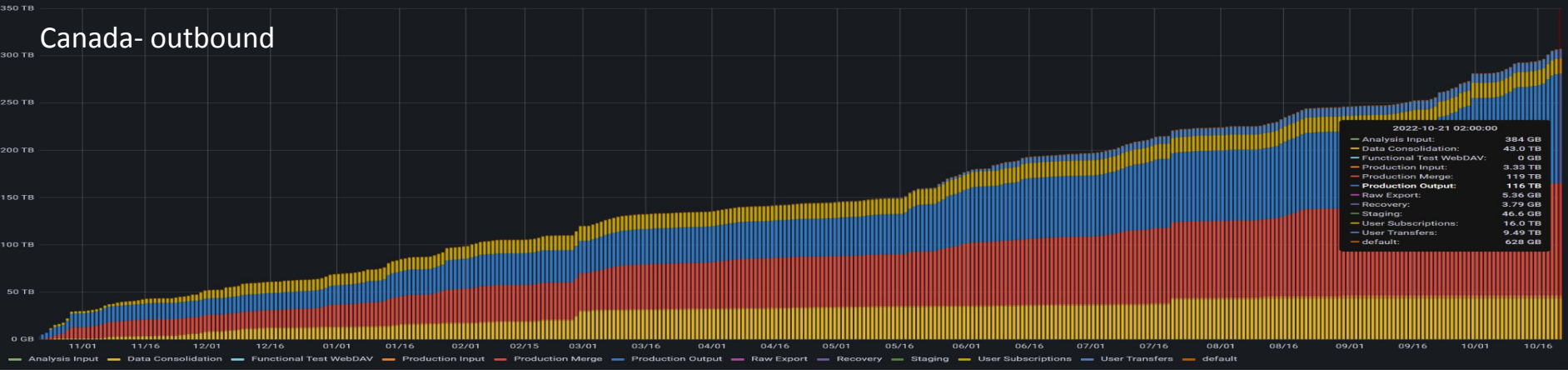
Canada- outbound



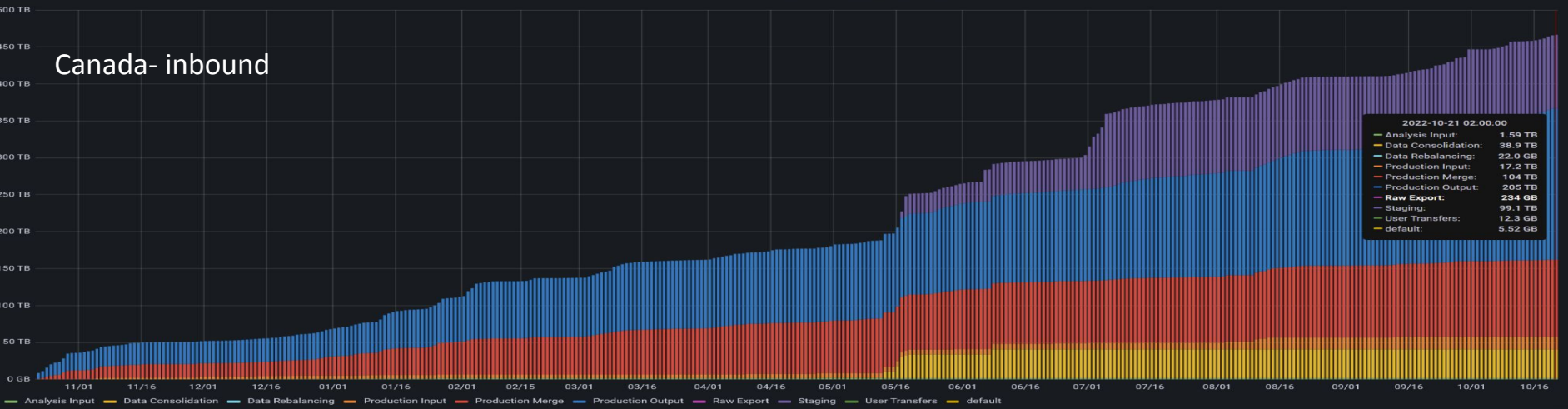
Canada- inbound



Canada- outbound



Canada- inbound



DPM Transition

Current information from DPM Sites.

- CESNET migrated the SE to a dCache system.
- Frascati-SE : in migration to dCache
- Napoli-SE: in migration to dCache
- IPHC-SE : Plan to move to EOS.
- KIRSTI: Plan to migrate to dCache
- IHEP: Plan to migrate to dCache or EOS
- NTU: Plan to migrate to dCache
- CYFRONET: To be check
- LAL: To be Check
-

DynaFed long term support

DynaFed seems that will have the same of roadmap of DPM.

DynaFed is not in the list of the technologies to test within the WLCG JWT Compliance test.

However

- As of today, DynaFed looks to be a good solution to use S3 storages thanks to the UVic expertise.
- DynaFed is used in BONIC (volunteer computing)
- Other work on DynaFed (see “IRIS DynaFed: IAM-Integrated Echo Storage” <https://indico.cern.ch/event/970568/contributions/4193736/attachments/2180300/3682748/IRIS%20DynaFed%20-%20IAM-Integrated%20Echo%20Storage.pdf>)
- Investigation is needed