

Stato del Computing in Belle II

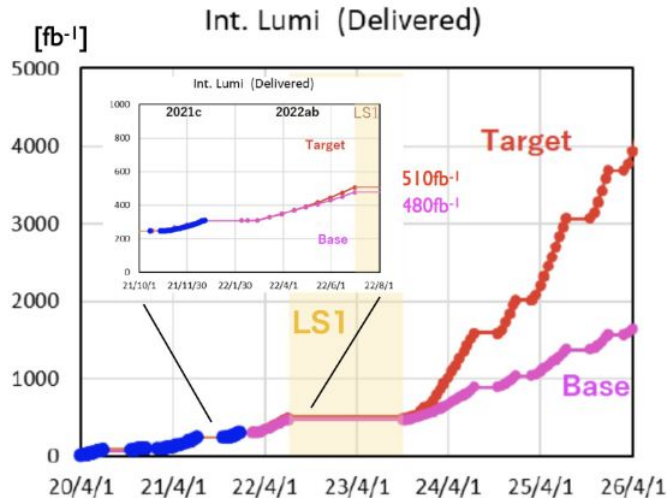
Dr. Silvio Pardi

Phone

25 Novembre 2022

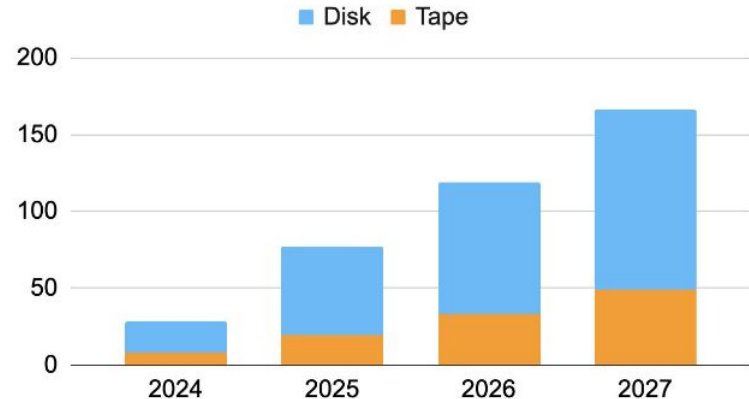
Belle II Numbers

- 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.

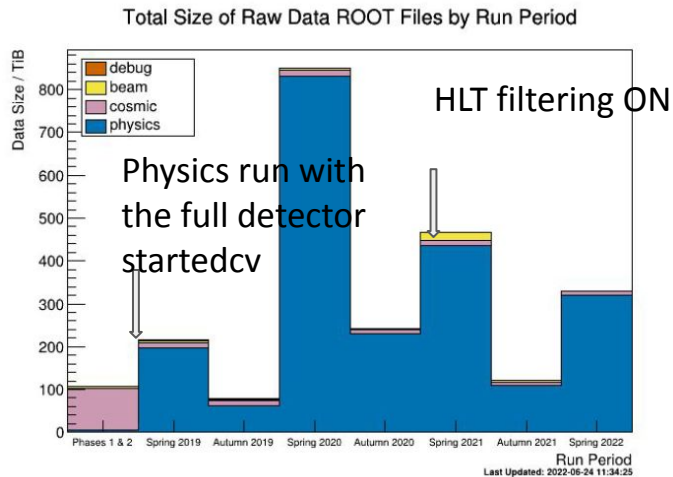
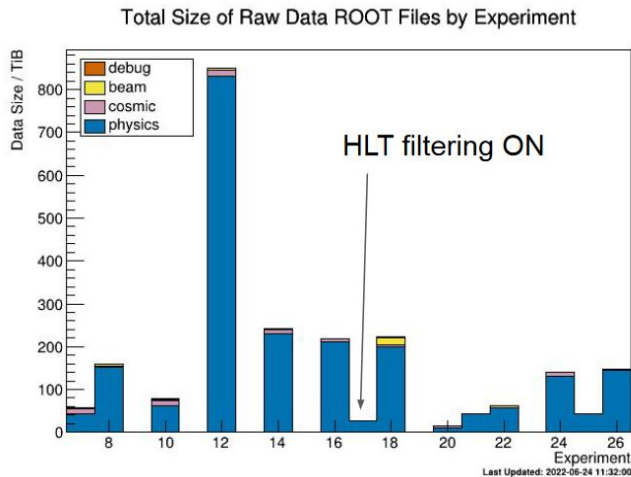
Space Occupancy (PB)



- 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



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

Siti Italiani

	CPU Pledge (kHS06)	CPU Opport. (kHS06)	Storage (TB)	Tape (TB)
CNAF	27		820	650
Cosenza	1			
Napoli	13	10	390 (+200)	
Pisa	8	10	200	
Torino	6	24	350	
Frascati		0,5	11	
LNL		1		
Roma3		2	2	
TOTALE	55 kHS06	47,5 kHS06	1.973 TB	650 TB

Richiesti per il 2023 ulteriori 200 TB da installare presso il CNAF

From RUCIO Workshop

Size of the DB

- Very different range if one compares the number of rows of DID table. To be fair, some collaborations (*) are in data taking mode, whereas others not yet :
 - ATLAS* : 1.3B
 - Belle II* : 104M
 - CMS* : 92M
 - Dune : 3.3M
 - SKAO : 266k
- Table partitioning :
 - ATLAS : All “active” tables partitioned by scope, archived tables partitioned by time
 - CMS : History table + bad_replicas tables partitioned
 - Belle II : History tables recently partitioned

L'infrastruttura di calcolo di Belle II sta crescendo molto.

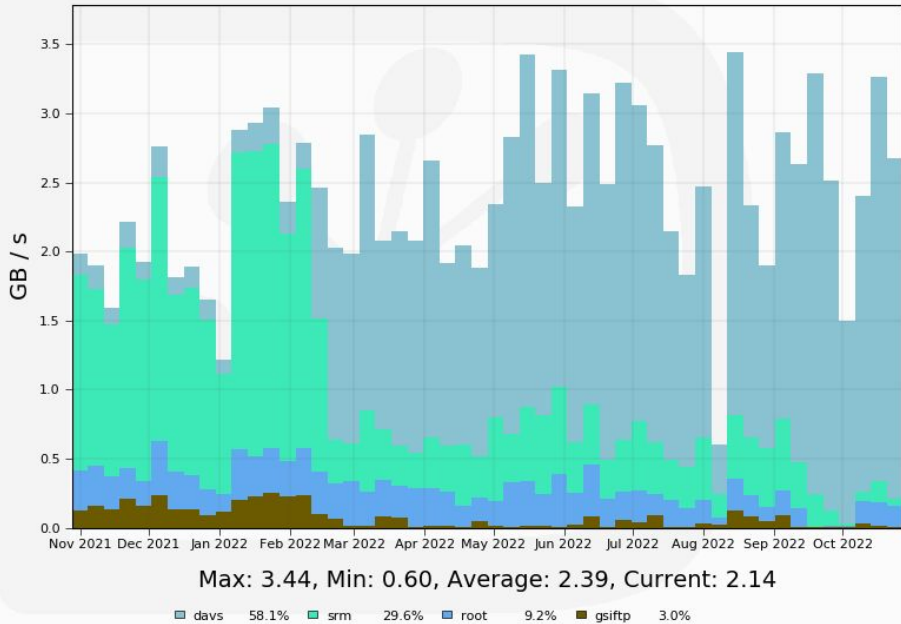
Al Rucio Workshop 2022 sono stati confrontati i DB delle maggiori installazioni di Rucio.

Belle II risulta essere il secondo dopo ATLAS.

Migration to DAVS: Data access

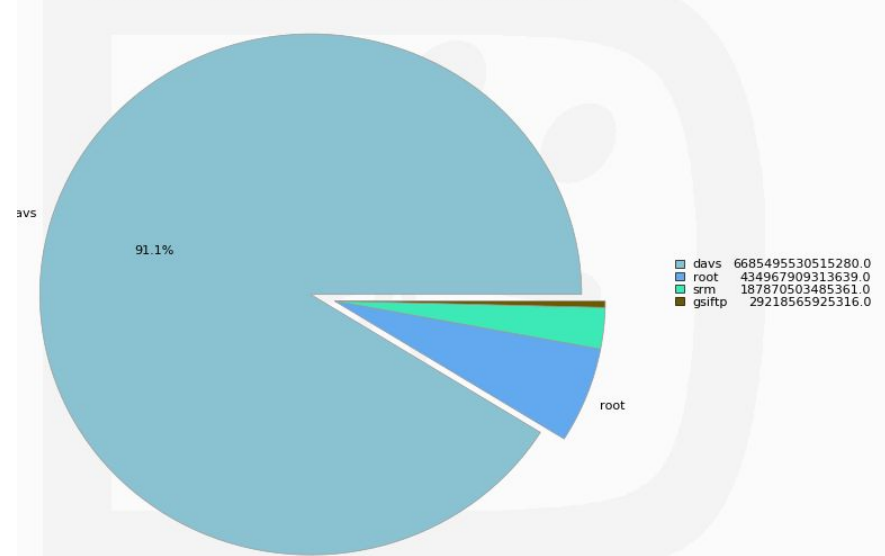
Throughput by Protocol

52 Weeks from Week 43 of 2021 to Week 43 of 2022



Total data transferred by Protocol

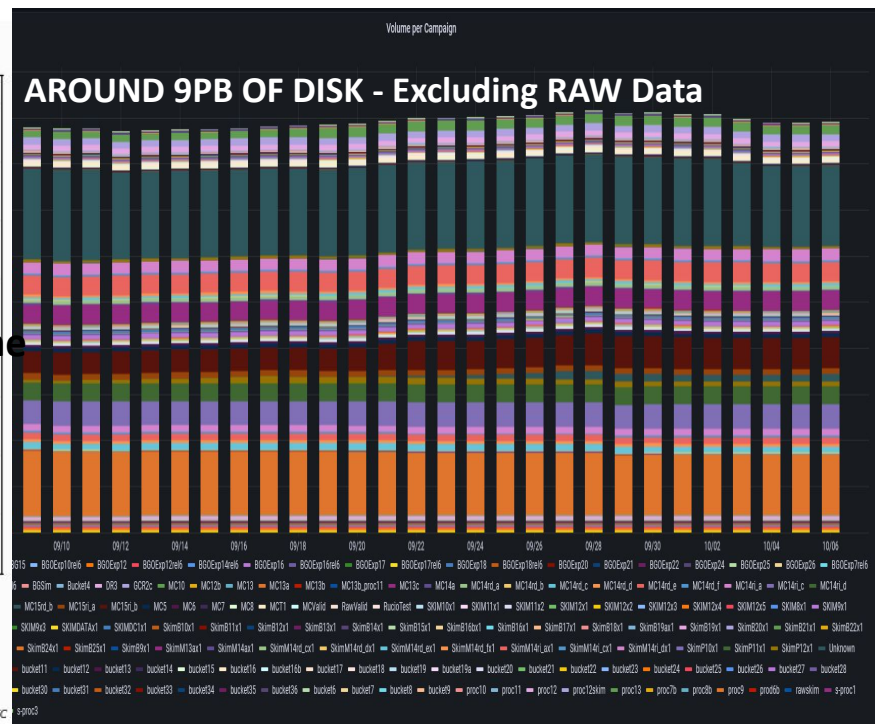
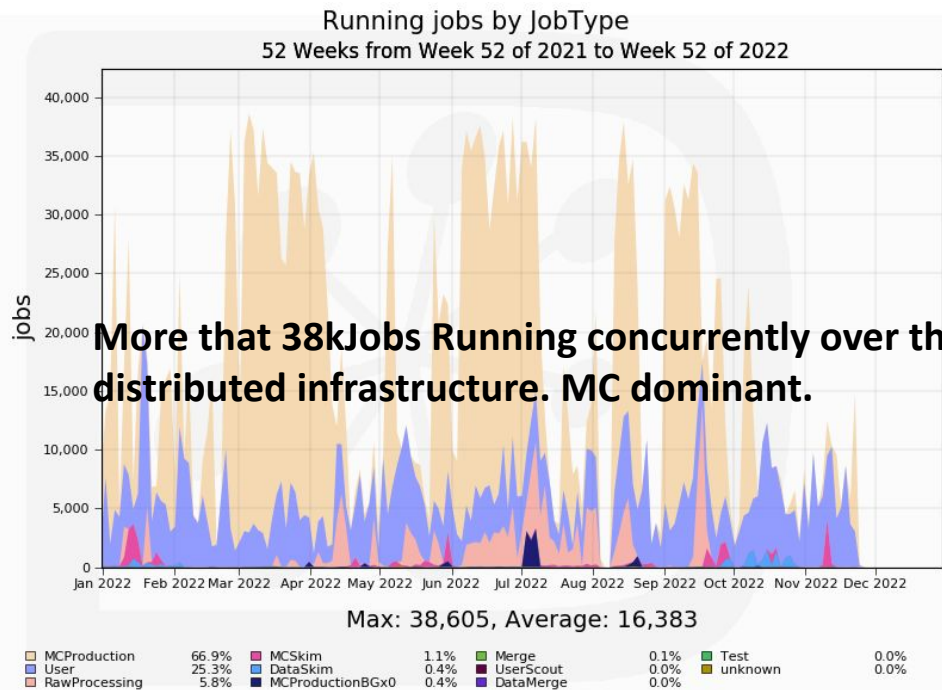
13 Weeks from Week 39 of 2022 to Week 52 of 2022



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

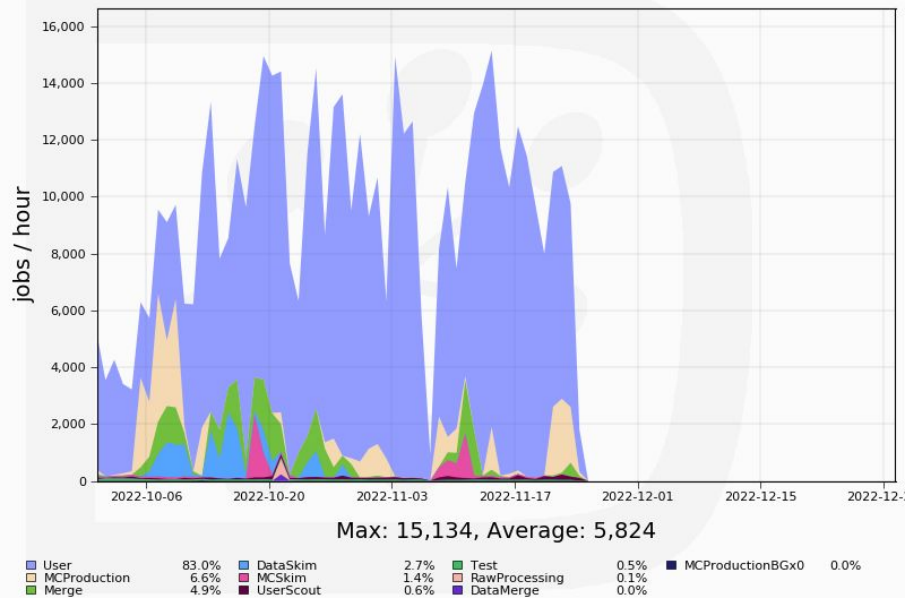
Generated on 2022-11-03 21:31:40 UTC

Belle II Status



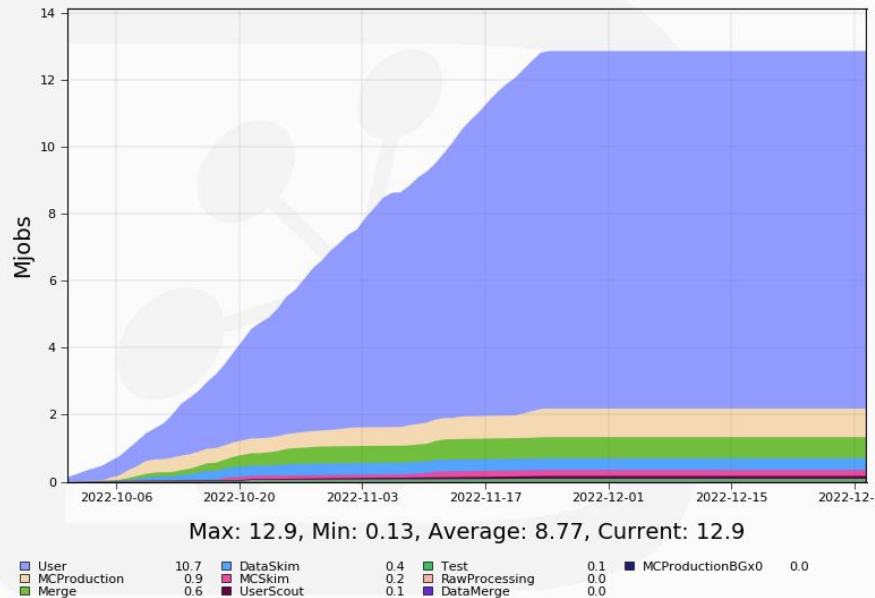
Ultimo Quadrimestre oltre 12.9MJobs eseguiti

Jobs by JobType
13 Weeks from Week 39 of 2022 to Week 52 of 2022



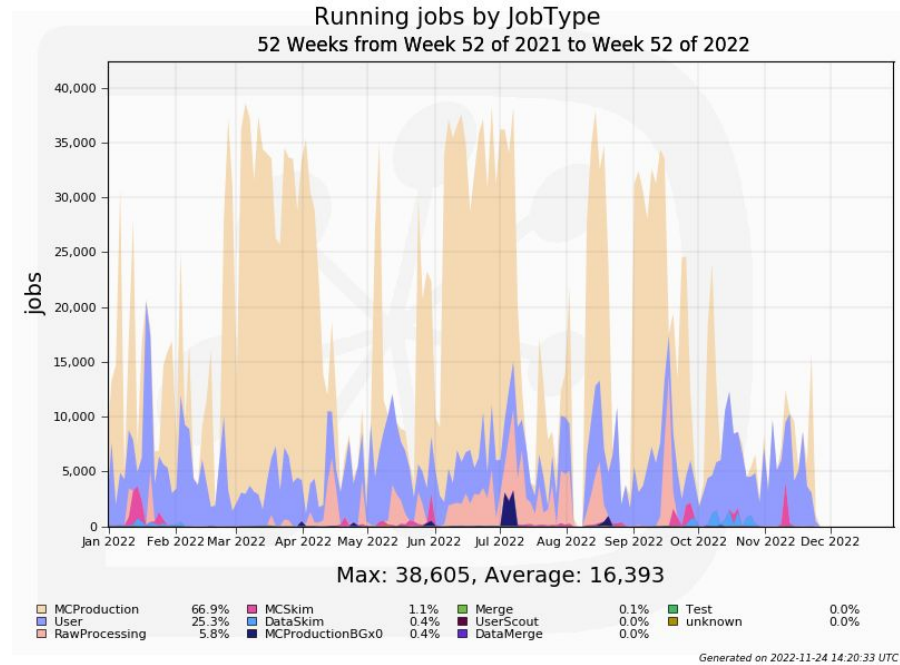
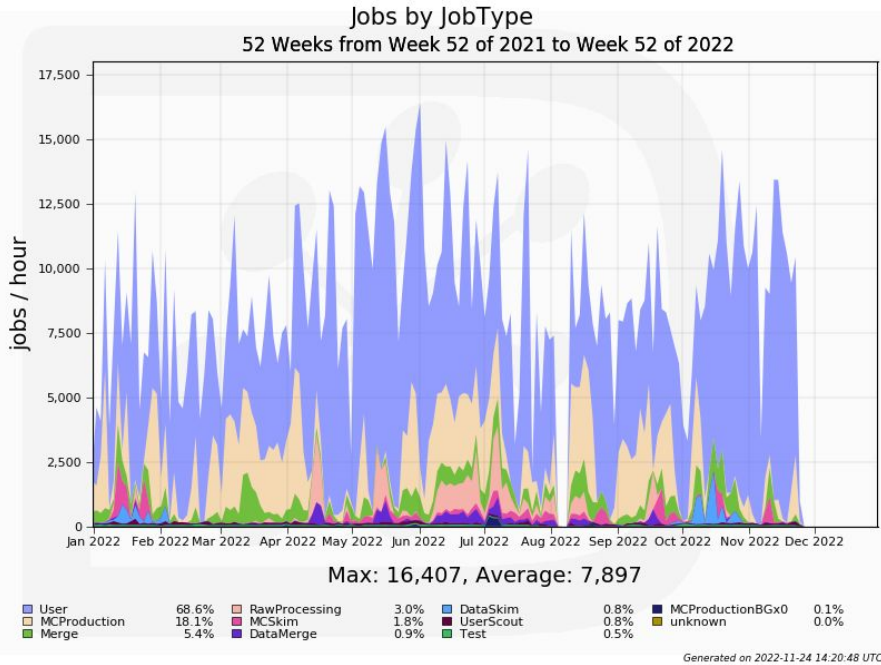
Generated on 2022-11-24 14:00:45 UTC

Cumulative Jobs by JobType
13 Weeks from Week 39 of 2022 to Week 52 of 2022



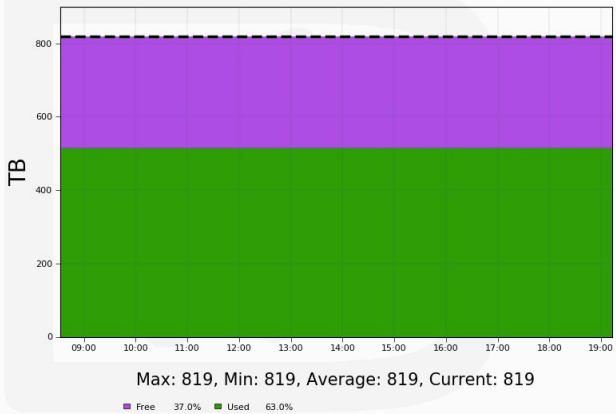
Generated on 2022-11-24 14:01:58 UTC

Job execution rate elevata



CNAF-TMP-SE

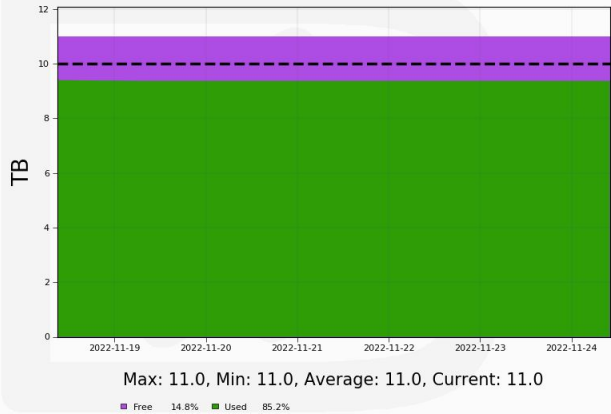
10 Hours from 2022-11-24 08:32 to 2022-11-24 19:12 UTC



Generated on 2022-11-24 10:31:41 UTC

Frascati-TMP-SE

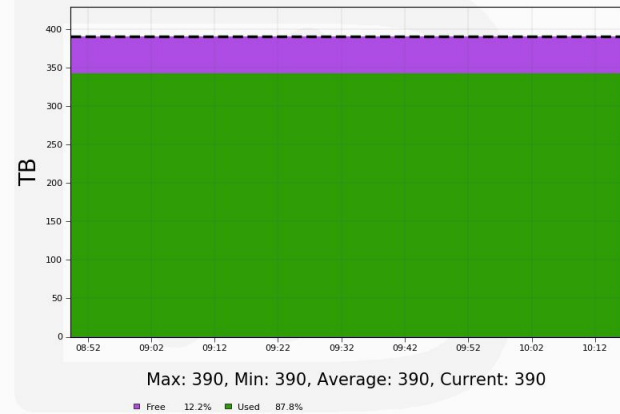
145 Hours from 2022-11-18 08:56 to 2022-11-24 10:01 UTC



Generated on 2022-11-24 10:31:21 UTC

Napoli-TMP-SE

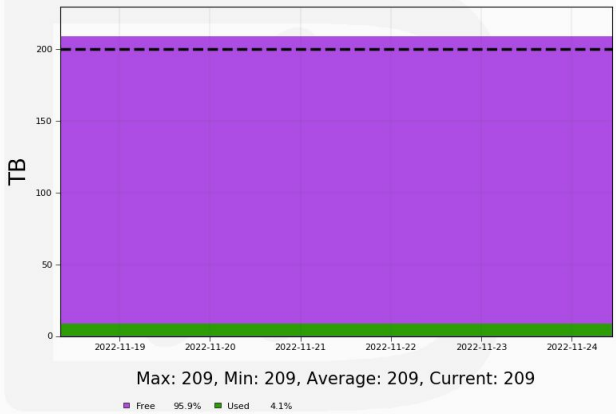
1 Hours from 2022-11-24 08:49 to 2022-11-24 10:16 UTC



Generated on 2022-11-24 10:25:57 UTC

Pisa-TMP-SE

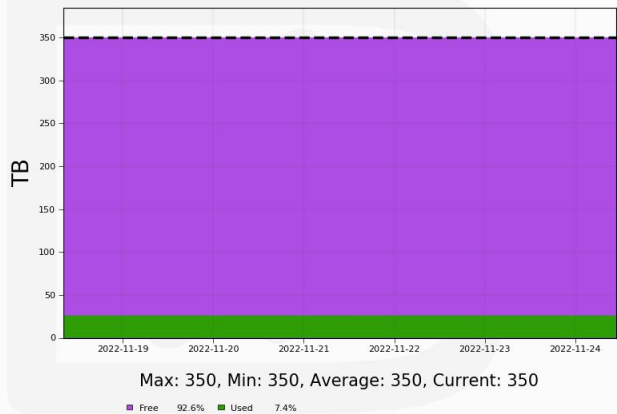
146 Hours from 2022-11-18 08:08 to 2022-11-24 10:42 UTC



Generated on 2022-11-24 10:34:30 UTC

Torino-TMP-SE

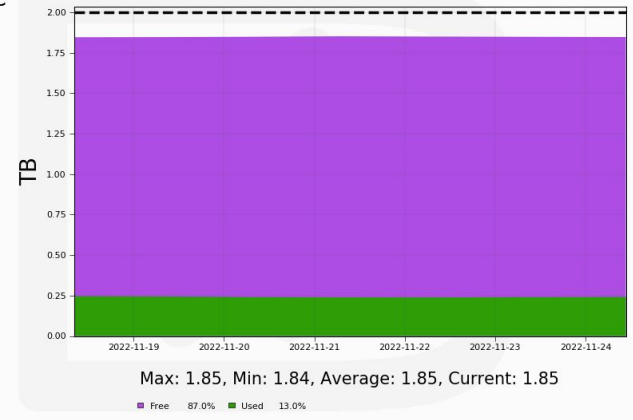
146 Hours from 2022-11-18 08:10 to 2022-11-24 10:43 UTC



Generated on 2022-11-24 10:33:10 UTC

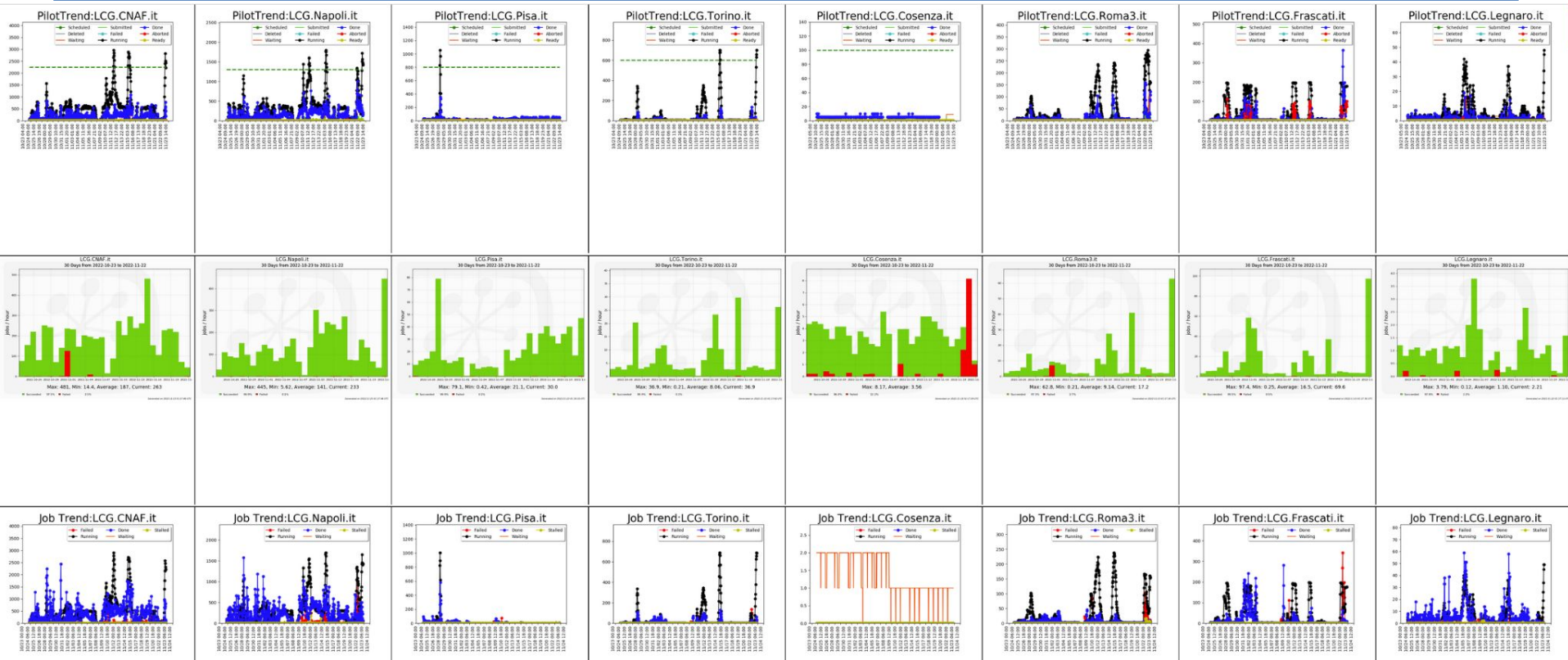
Roma3-TMP-SE

146 Hours from 2022-11-18 08:08 to 2022-11-24 10:42 UTC



Generated on 2022-11-24 10:34:44 UTC

Utilizzo siti italiani nell'ultimo mese



Alcuni limiti e problematiche note

Configurazione DIRAC: Necessaria l'ottimizzazione dei parametri di distribuzione dei dati e delle varie tipologie di job sui siti. Aperta discussione, working session al B2GM di febbraio.

Errori sugli storage: Frequenti sono i casi in cui gli utenti non possono prelevare gli output dei job. Responsabilità dei siti che ospitano gli storage, rispondere con celerità ai GGUS ticket aperti per non bloccare i ricercatori.

SandBox: Un numero troppo elevato di job brevi (tipicamente i job utente) producono un overload della SandBox di DIRAC. Attualmente si riescono a processare circa 10-15kJobs. In studio strategie per incrementare la scalabilità e spostarsi su un nuovo punto di lavoro.

Nuovi certificato server per VOMS DESY

In questo mese:

DESY ha cambiato il DN del certificato del VOMS server.

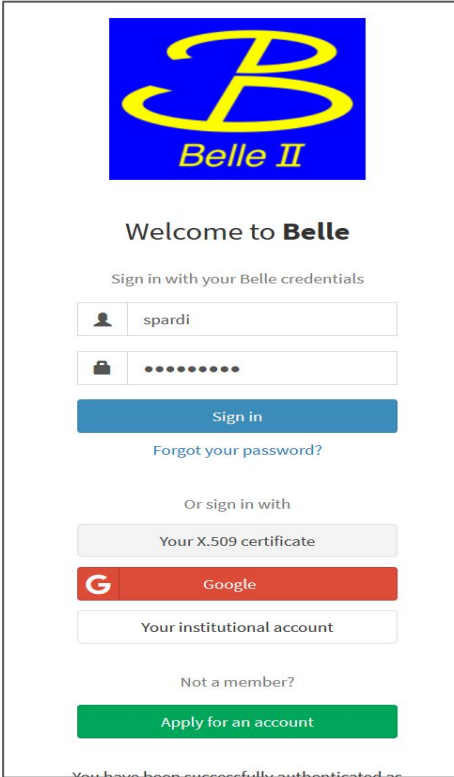
Questo ha richiesto una ri-configurazione di tutti i siti. Una verifica globale è ancora in atto.

I siti italiani sembrano tutti ok.

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 login interface for Belle II. 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". At the very bottom, the text "You have been successfully authenticated as" is partially visible.

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

HTCondor GSI support EOL has been postponed to Feb 2023

- CE token support deployment campaign on EGI launched June 1:
70+%
 - HTCondor v9.0.x with tokens for ATLAS and CMS, others later
 - ARC CE REST interface, in particular to support job submissions via HTCondor-G
- Another campaign on EGI will be needed early 2023 to get all HTCondor
 - CEs on supported versions > v9.0.x
 - Also EGI Check-in tokens should work by that time

Storage

- Workflow details involving Rucio/DIRAC and/or FTS vs. SEs have mostly been identified and implemented to various extents
 - May need to be re-discussed if major implementation or operational hurdles are encountered
- The token testbed covers basic functionality and interoperability
 - Most endpoints pass most tests
- Rucio and DIRAC should drive this → implications for the FTS
 - Will see further progress expected in the next months
- SEs typically need to support concurrent use of X509 and tokens

Job Multicore

In testing la possibilità di sottomettere job multicore nei RAW DC

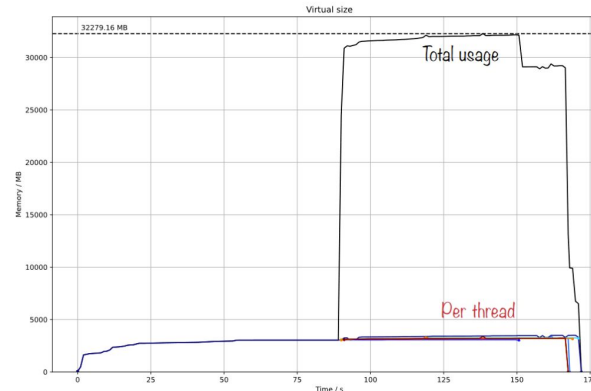
Inviati alcuni job a UVic, BNL, e KIT.

Configurata una coda DIRAC per il CNAF.

Local execution with multicore configuration

Memory consumption of 8-core job

execution time : 172.00 sec
Virtual size : **max=32279.16 MB**, avg=15266.56 +- 14033.21 MB
Resident memory : max=13590.64 MB, avg= 6243.65 +- 5887.30 MB
Proportional memory : max= 3808.78 MB, avg= 2026.52 +- 1230.53 MB



Exceeds memory usage
of single-core x8

Agenda pre-B2GM

Belle 2 General Meeting di febbraio in persona.

Molti temi aperti, ho segnalato l'importanza di verificare le configurazioni dei siti per ottimizzarne l'utilizzo.

WEDNESDAY, 8 FEBRUARY

- 10:00 AM** → 12:00 PM **Site Reports: Items for Site Reports**
What we ask to be included in the next site reports
Convener: Silvio PARDI (INFN - Napoli)
- 2:00 PM** → 3:00 PM **Site Configuration: Site parameters vs Site reports**
Reviewing, updating and/or tuning the parameters
Convener: Silvio PARDI (INFN - Napoli)
- 3:00 PM** → 5:00 PM **Site Configuration: Site parameters vs Performance**
Reviewing, updating and/or tuning the parameters
Convener: I. Ueda (KEK)
- 5:00 PM** → 6:00 PM **Infrastructures: WLCG Data Challenges**
Convener: Hiro Ito

THURSDAY, 9 FEBRUARY

- 10:00 AM** → 12:00 PM **Data Management for Production System: Rucio-DDM interface**
Interface between Rucio and Production System
Convener: Cedric SERFON (Brookhaven National Laboratory)
- 2:00 PM** → 4:00 PM **Data Management for Production System: Staging Subsystem**
Interface between Rucio and Production System
Convener: Ruslan MASHINISTOV (Brookhaven National Laboratory)
- 4:00 PM** → 6:00 PM **Data Management for Production System: Metadata registration by Production System**
Interface between Rucio and Production System
Convener: Hideki Miyake (KEK-IPNS)

FRIDAY, 10 FEBRUARY

- 10:00 AM** → 12:00 PM **Rucio and Jobs: Metadata and Traces**
Metadata registration and Traces sending from Jobs
Convener: Anil PANTA (University of Mississippi)
- 2:00 PM** → 6:00 PM **Tokens: IAM and Tokens in DIRAC**
Transition from X509 proxy to Tokens
Convener: Michel Hernandez Villanueva (DESY)

BACKUP

Site Report 2022

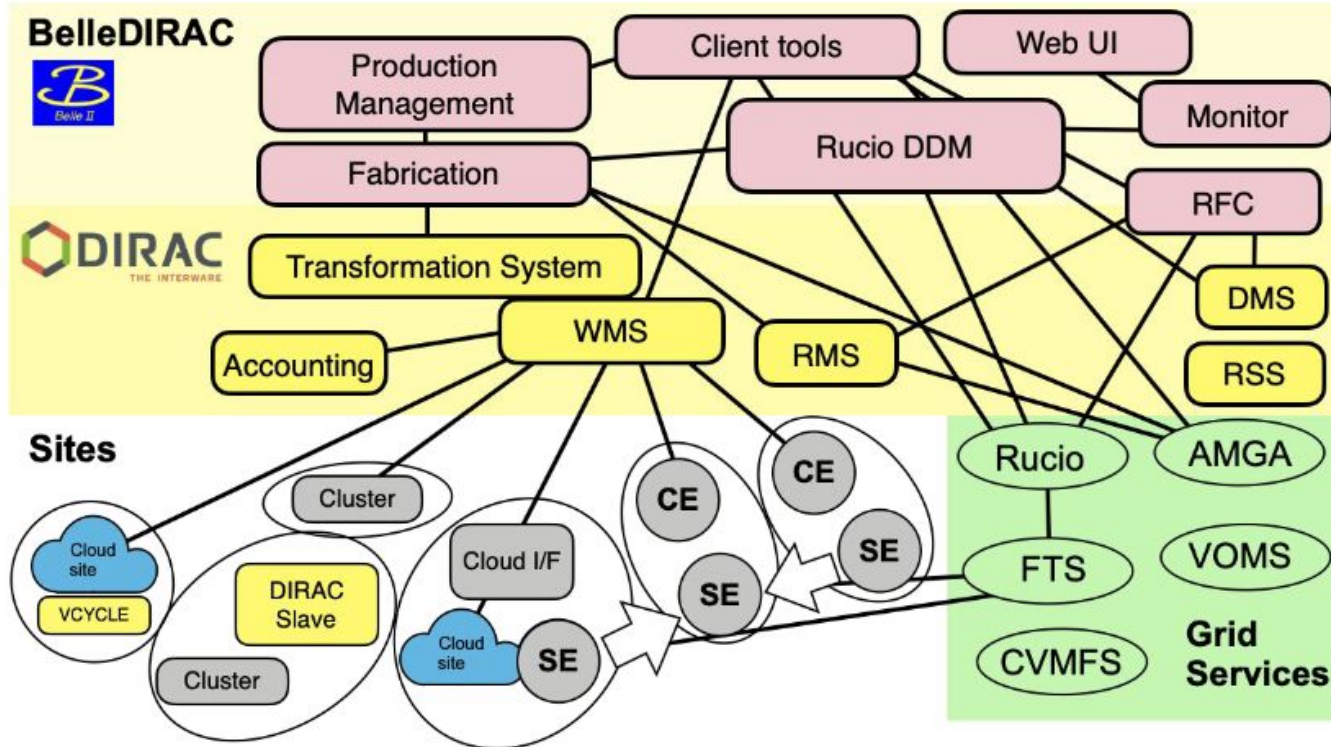
Resources	NOTE	CPU Deployed guaranteed (kHS06)	CPU Deployed guaranteed jobslots	CPU Opportunistic (kHS06)	CPU Opportunistic jobslots	Total CPU (kHS06)	Total Jobslots	Storage DISK	Tape
Production	Total Opportunistic CPU include the BNL core for calibration. CNAF opportunistic are estimated a 10% of declared	466	32k	386	32k	852	64k	15.5PB	12.4PB

Resources	NOTE	CPU Deployed guaranteed (kHS06)	CPU Deployed guaranteed jobslots	CPU Opportunistic (kHS06)	CPU Opportunistic jobslots	Total CPU (kHS06)	Total Jobslots	Storage DISK (TB)	Tape (TB)
Calib/Recalibration	DESY and BNL	36,7	3.1k	0	0	36,7	3.1k	500TB	600TB

Resources per Country 2022

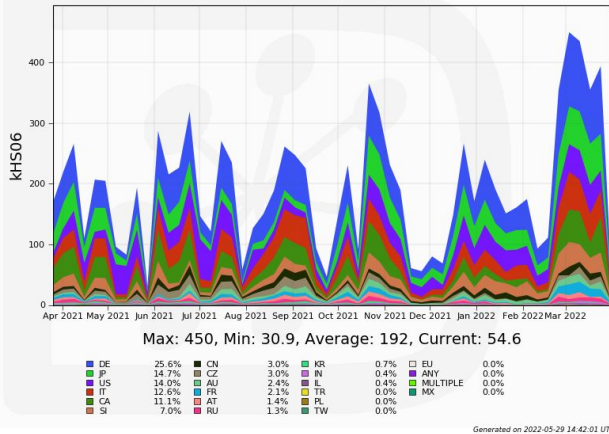
RESOURCES FOR PRODUCTION									
	CPU Deployed guaranteed (kHS06)	CPU Deployed guaranteed jobslots	CPU Deployed Opportunistic (kHS06)	CPU Deployed Opportunistic jobslots	Total Deployed CPU (kHS06)	Total Jobslots	Storage (TB)	TAPE (TB)	Notes
Australia	18	900	10	1000	28	1900	50	0	
Austria	4,8	480	0	0	4,8	480	250	0	
Canada	80	4000	20	1000	100	5000	600	100	N.B. There is not tape but 100TB for RAW Data
China	15	856	0	0	15	856	260	0	
France	11,8	890	2,2	180	14	1070	403	179,22	
Germany	78,02	6424	102,5	8146	180,52	14570	4070	1830	
India	19,58	1100	5,14	161	24,72	1261	0	0	
Israel	2,7	168	0	0	2,7	168	60	0	
Italy	55	5050	95,6	8849	150,6	13899	1772	650	CNAF reported 427 Opportunistic, for this computation considered 10%
Japan	60,3	3256	43,8	2526	104,1	5782	3468	5550	
Korea	0,32	36	1	56	1,32	92	0	0	
Mexico	2,4	144	0	0	2,4	144	0	0	
Poland	2	200	0	0	2	200	10	0	
Russia	13	1156	5	500	18	1656	0	0	
Slovenia	22,5	1800	16	1200	38,5	3000	1210	0	
South Korea	8,576	544	0	0	8,576	544	100	0	
Taiwan	18,33	410	0	0	18,33	410	791,95	0	
The Czech Republic	4,1	400	12,3	1200	16,4	1600	100	0	
Turkey	0,938	128	0	0	0,938	128	130	0	
USA	49,4	4300	73	7000	122,4	11300	2312	4100	Calibration CPU are included also among the opportunistic resources
Production TOT	467	32.242	387	31.818	853	64.060	15.587	12.409	
RESOURCES FOR CALIBRATION									
Germany	3	100			3	100		600	at DESY
USA	33,7	3000			33,7	3000	500		at BNL
Calibration TOT	37	3.100	-	-	37	3.100	500	600	

DIRAC Infrastructure

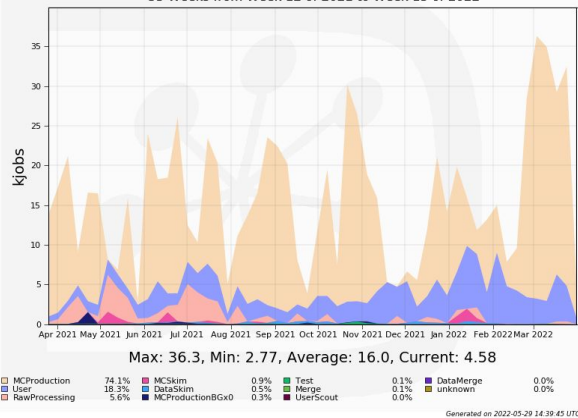


Overall activity in 2021 JFY

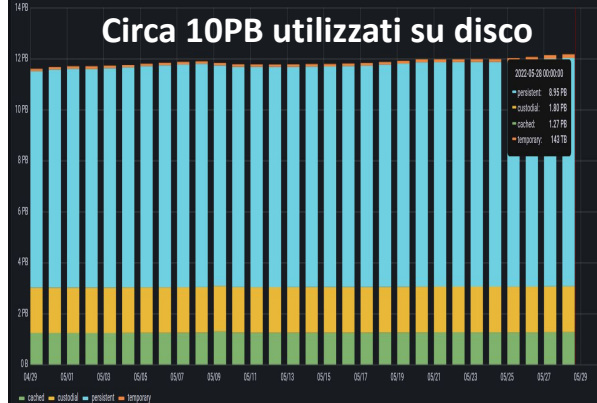
Normalized CPU usage by Country
53 Weeks from Week 12 of 2021 to Week 13 of 2022



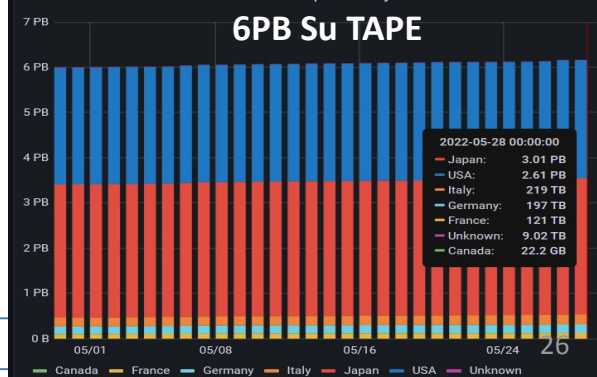
Running jobs by JobType
53 Weeks from Week 12 of 2021 to Week 13 of 2022



Volume per Custodially



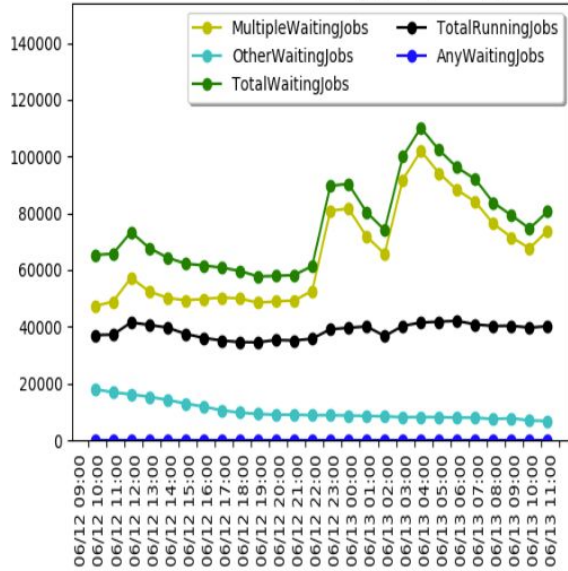
Volume per Country



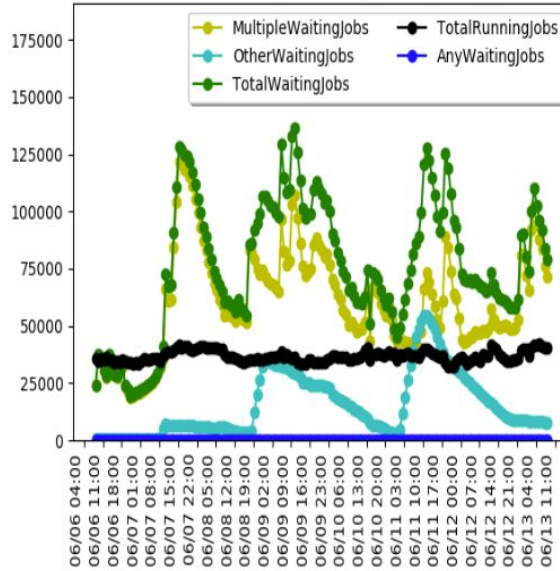
Italian Share 12.6 (Milestone 2021 -11%)

- Attività aumentata rispetto al 2020
- Picchi di ~40 k jobs running
- 31kJobSlots Pledged molte CPU Opportunistiche
- Current User Job 24% (increasing)

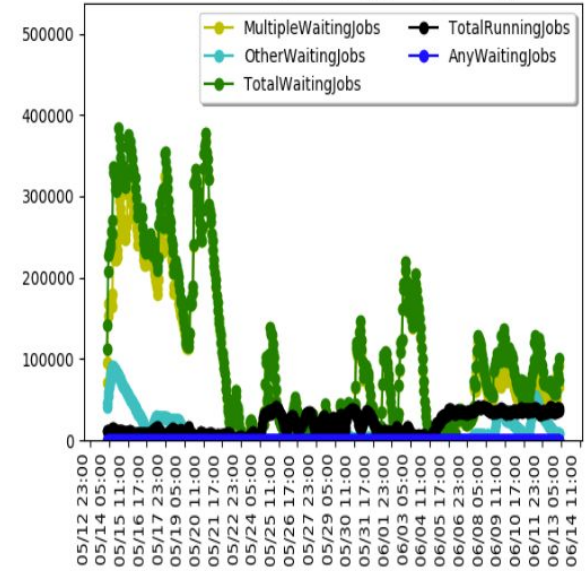
Total Running/Waiting jobs (1 day)



Total Running/Waiting jobs (7 day)



Total Running/Waiting jobs (30 day)



Total Running Job = 40184

Total Waiting Job = 80564