Integration of the Barcelona Supercomputing Center for CMS computing: Towards large scale production

C. Acosta, A. Delgado, J. Flix, J.M. Hernández, A. Pérez-Calero, E. Pineda, I. Villalonga

CHEP 2023, May 11th 2023

GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA E INNOVACIÓN

Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

cfp CIEMAT física de partículas

PIC port d'informació científica

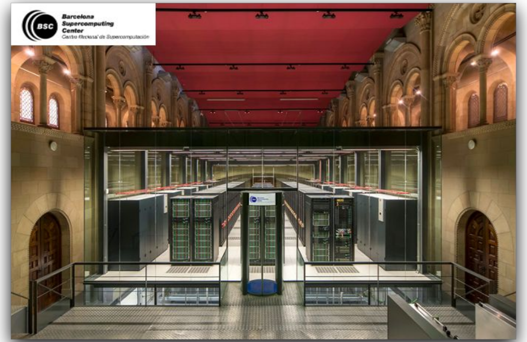
Motivation and context

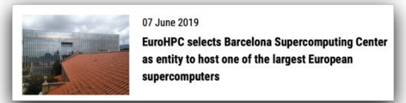
HPC facilities are attractive for **WLCG** computing

- To help cover increasing CPU needs despite flat funding
- Lot of public funding went to HPCs
- Several ongoing HPC integration efforts in CMS

BSC - Barcelona Supercomputing Center

- Largest HPC center in Spain
- Current MareNostrum 4 (*MN4*) general-purpose cluster:
  - 11.5 Petaflops (166k CPU cores), 390 TB RAM, 24 PB disk local SSD disk
  - SLURM as batch system, SUSE Linux Enterprise as OS
  - 15 PB GPFS as storage back-end (mounted on login/compute nodes)
- Next MareNostrum5 (~17xMN4, ~200 petaflops), available after Summer 2023
  - One of Europe’s first pre-exascale supercomputers


Barcelona Supercomputing Center
Centro Nacional de Supercomputación


07 June 2019

EuroHPC selects Barcelona Supercomputing Center as entity to host one of the largest European supercomputers

https://eurohpc-ju.europa.eu/about/our-supercomputers_en#marenostrum-5

PIC
port d'informació
científica

Motivation and context

In 2020 BSC designated LHC computing as a **strategic project**

- Agreement promoted by WLCG-ES community and funding agency

Allocations* of up to a **7% share of MN4 for LHC** [100M coreHours max/year]

- ~70M coreHours/year
- ~20M hours allocation for CMS in 2022

* Submission of proposals for time allocation every 4 months

Potentially, very **significant contribution** for LHC computing in Spain

- Comparable e.g to all ATLAS+CMS+LHCb simulation needs in the country

PIC
port d'informació
científica

The challenges for CMS

BSC imposes very **restrictive network connectivity** conditions

- No incoming *or outgoing* connectivity from compute nodes
- Only incoming SSH/SSHFS communication through login nodes
- A shared disk (GPFS) mounted on compute nodes and login machines - accessible from outside via sshfs
- No services can be deployed on edge/privileged nodes

This is a **major obstacle for CMS** workloads

- Pilot with late binding model execution of payloads
  - Workload management system (glideinWMS - HTCondor services)
- Access to external services
  - Application software (CVMFS) & Conditions data (FrontierDB)
- Consuming and producing experiment data
  - Input/output data files (Storage Elements)

Solutions applied (in detail)

PIC
port d'informació
científica

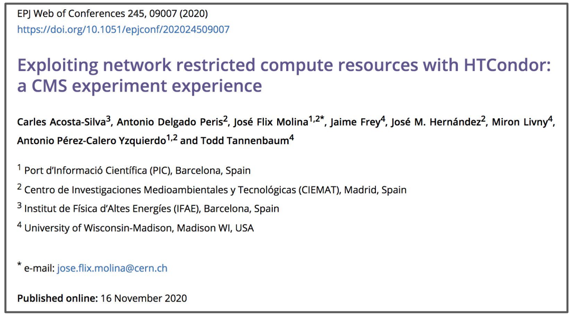
HTCondor split-starter dev.

HTCondor major development: modify CMS resource provisioning, job scheduling and execution framework (HTCondor) to use a **shared file system as communication layer**

- Split-starter model, presented at **CHEP19** and **CHEP21**
- Requires a bridge service at PIC Tier-1 to connect CMS WMS and BSC

A collaboration was formalized during the September'18 RAL HTCondor workshop

[Miron Livny, Todd Tannenbaum, Jaime Frey, Antonio Pérez-Calero, Carles Acosta, José Flix]



EPJ Web of Conferences 245, 09007 (2020)
https://doi.org/10.1051/epjconf/202024509007

Exploiting network restricted compute resources with HTCondor: a CMS experiment experience

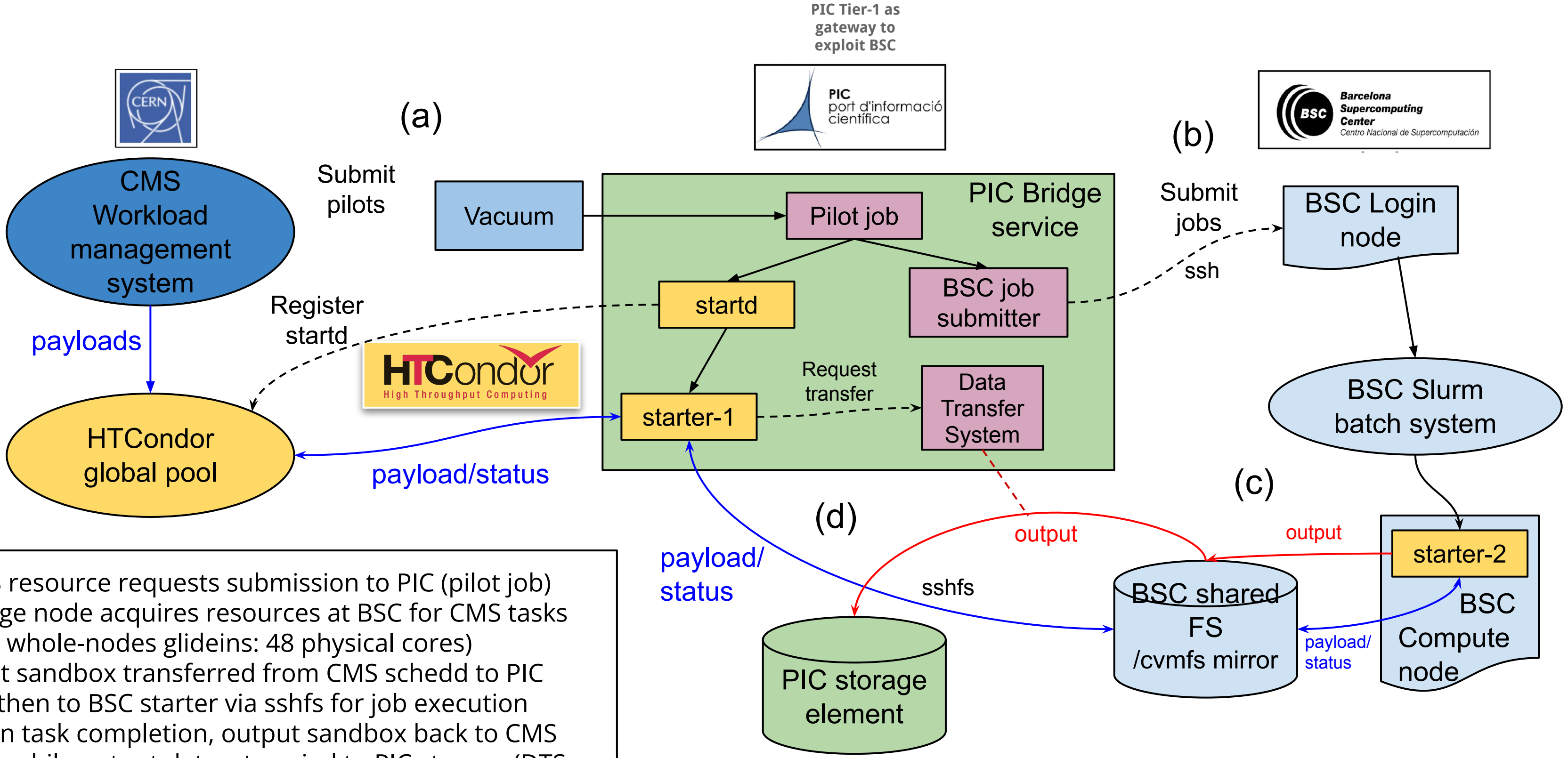
Carles Acosta-Silva[3], Antonio Delgado Peris[2], José Flix Molina[1,2*], Jaime Frey[4], José M. Hernández[2], Miron Livny[4], Antonio Pérez-Calero Yzquierdo[1,2] and Todd Tannenbaum[4]

[1] Port d'Informació Científica (PIC), Barcelona, Spain
[2] Centro de Investigaciones Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[3] Institut de Física d'Altes Energíes (IFAE), Barcelona, Spain
[4] University of Wisconsin-Madison, Madison WI, USA

[*] e-mail: jose.flix.molina@cern.ch

Published online: 16 November 2020

HTCondor split-starter + DTS



(a) CMS resource requests submission to PIC (pilot job)
(b) Bridge node acquires resources at BSC for CMS tasks
(BSC whole-nodes glideins: 48 physical cores)
(c) Input sandbox transferred from CMS schedd to PIC startd, then to BSC starter via sshfs for job execution
(d) Upon task completion, output sandbox back to CMS schedd, while output dataset copied to PIC storage (DTS acting as third party copy manager)

PIC
port d'informació
científica

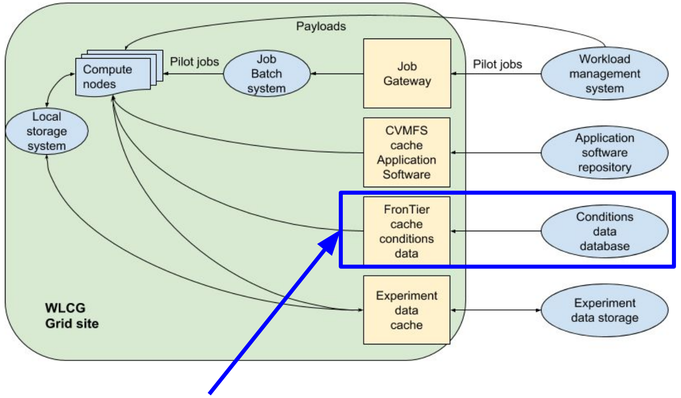
Access to conditions data

The **problem**:

- CMS jobs need to **access conditions data at run time**
- No outgoing or incoming network connectivity on the worker nodes
- No edge nodes with outgoing connectivity (not even login node)
- Thus jobs cannot read conditions data from Frontier service

Attempted solution:

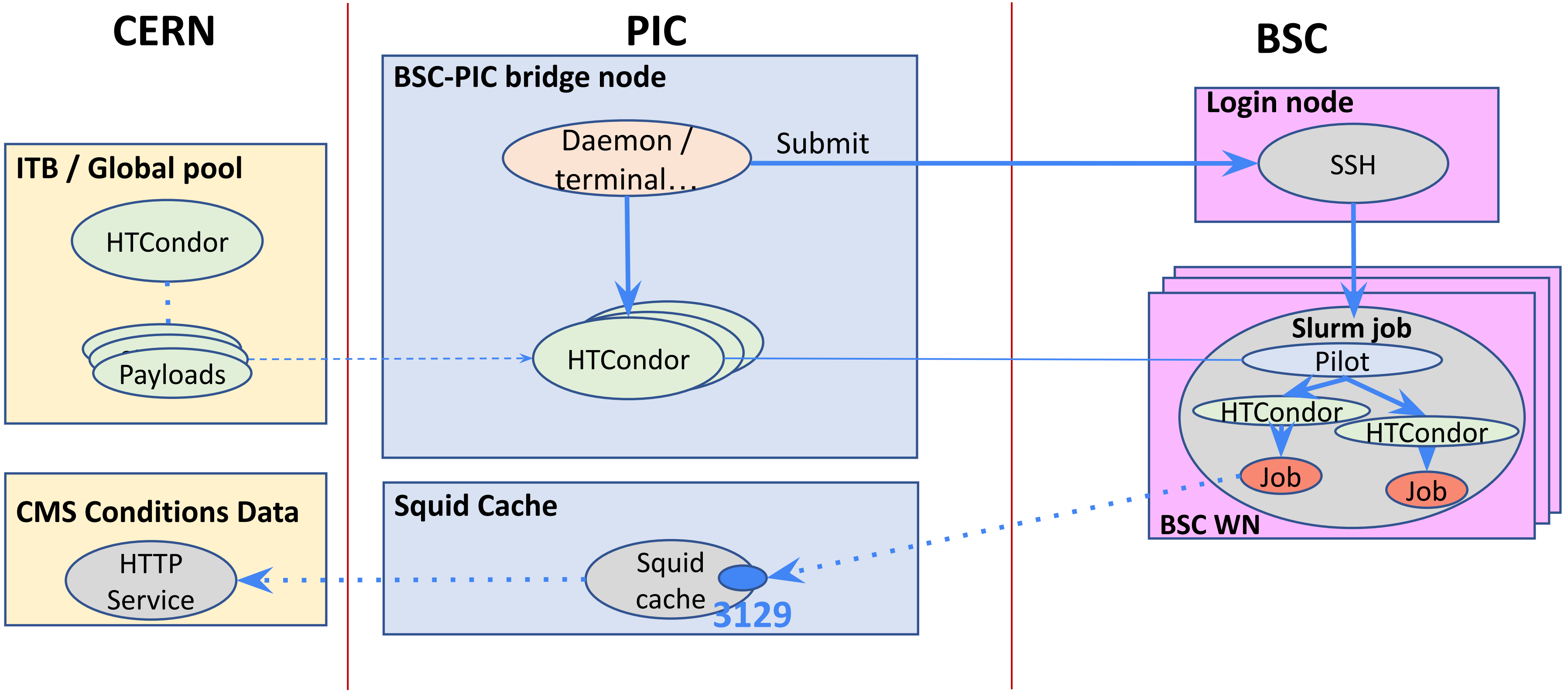
- Read conditions data from pre-placed static **SQLite files**
- Required some development in CMSSW



In practice, **commissioning** of data processing workflows at BSC proved **difficult**:

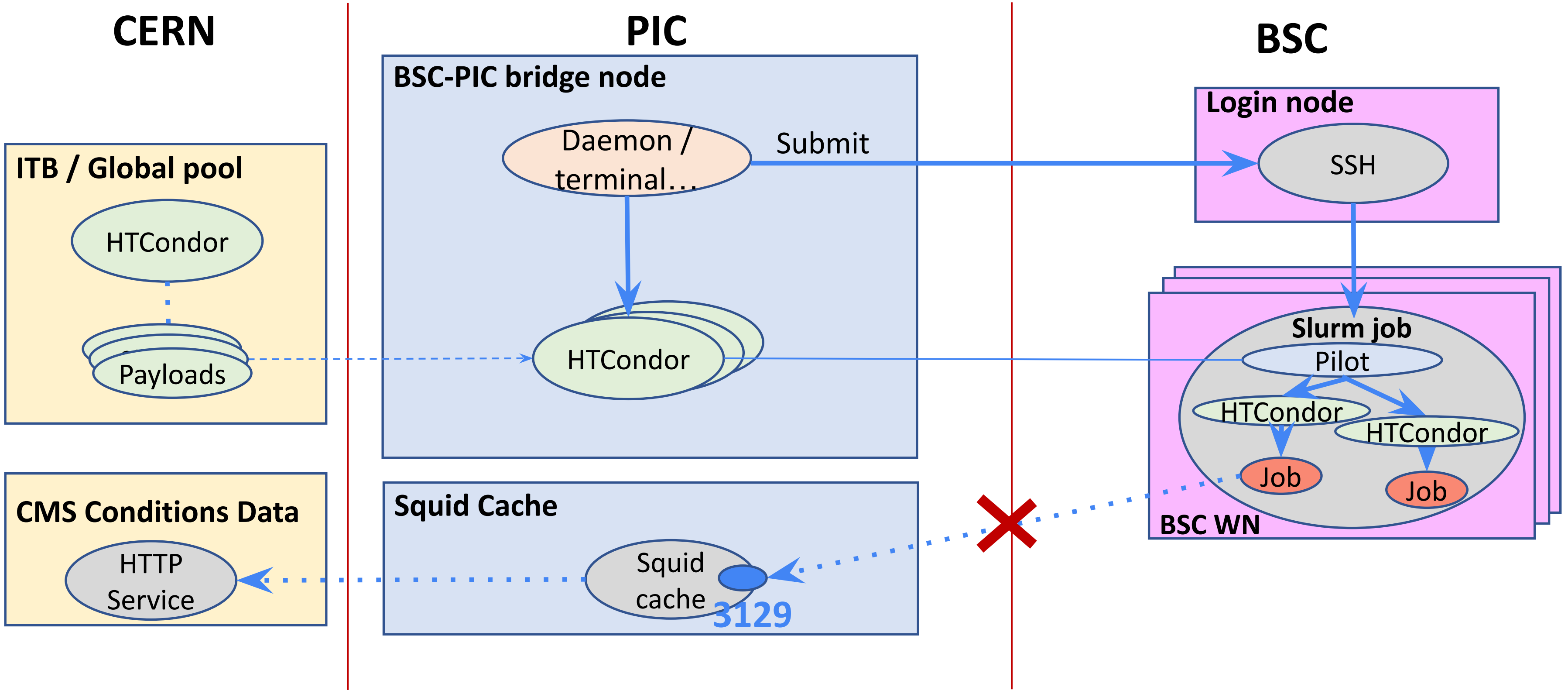
- Required identifying suitable workflows (proper CMSSW version)
- Manual **production of conditions data files was non-trivial**
- **Certain conditions files were huge**, and its production required large amounts of memory

The problem (again)



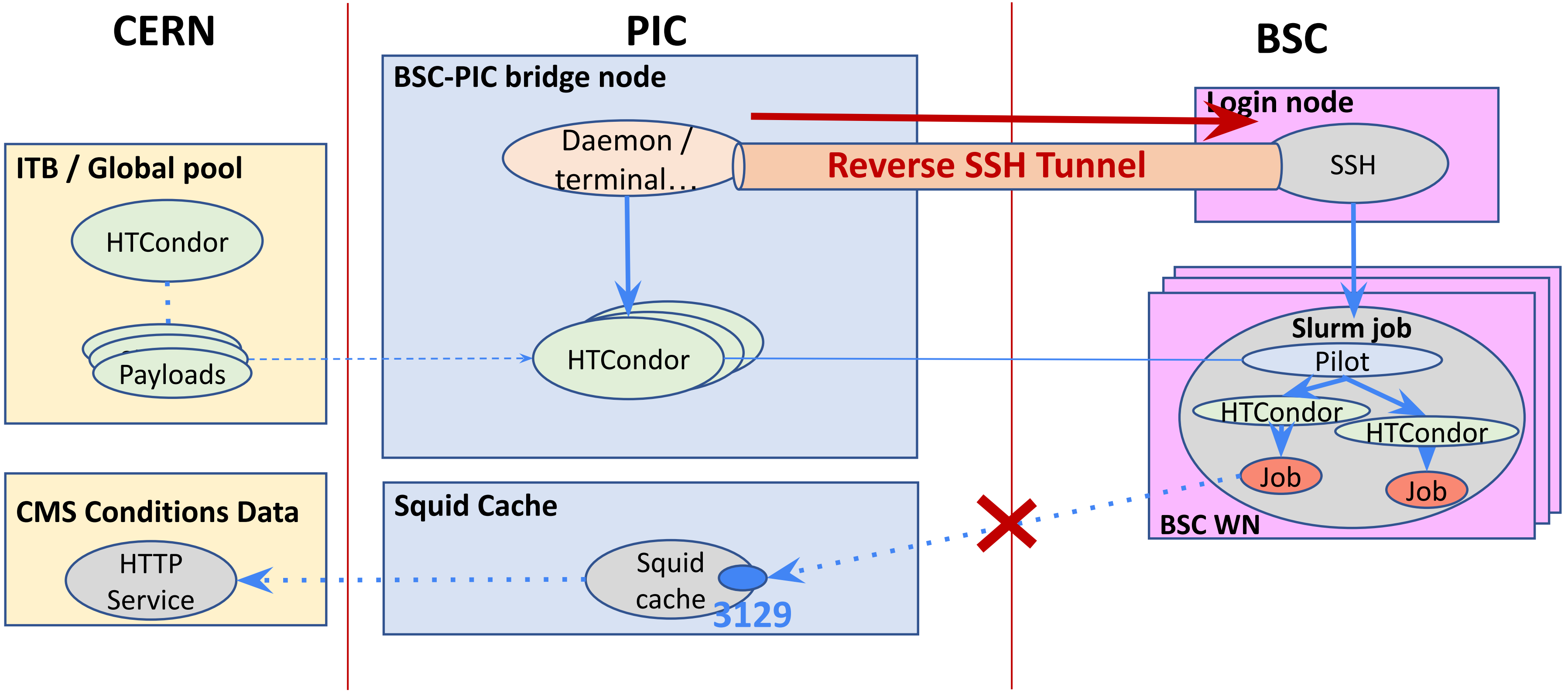
PIC
port d'informació
científica

The problem (again)



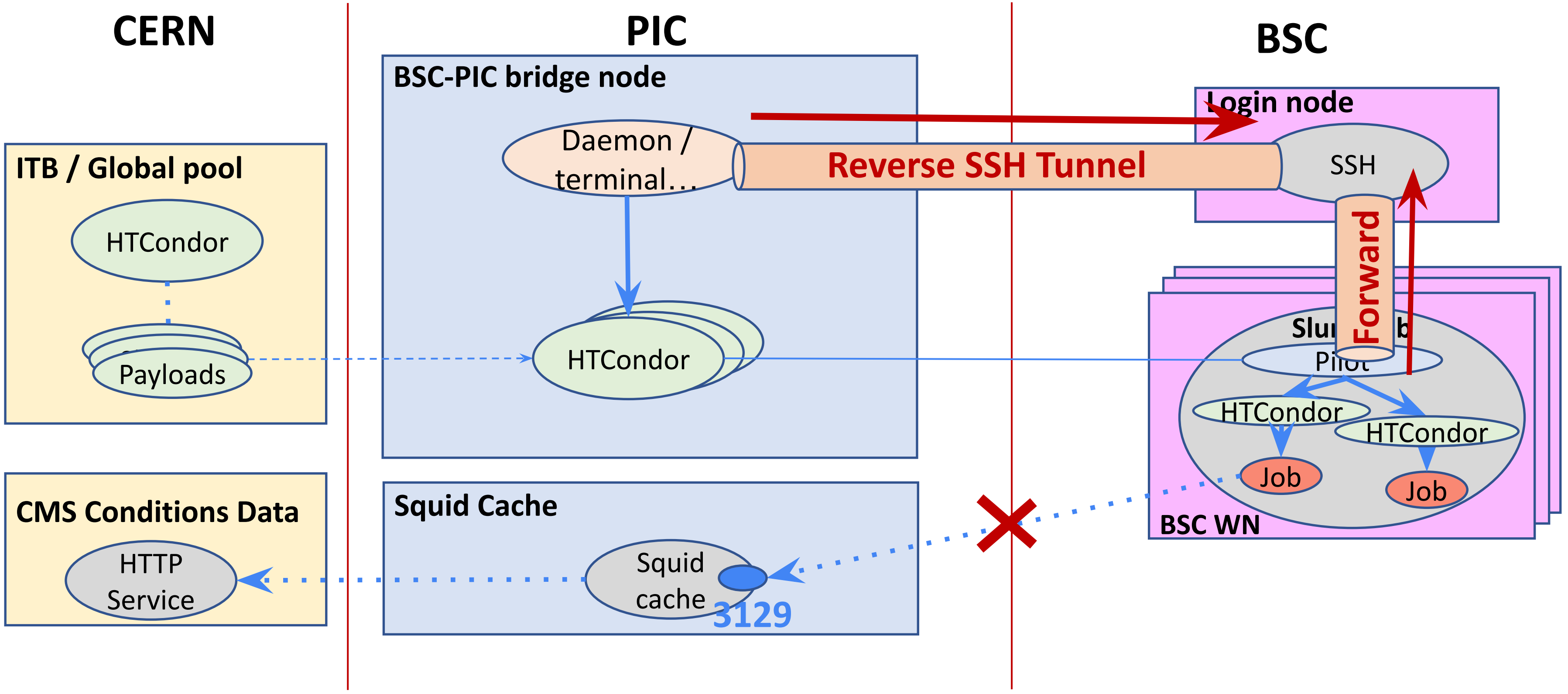
PIC
port d'informació
científica

The solution



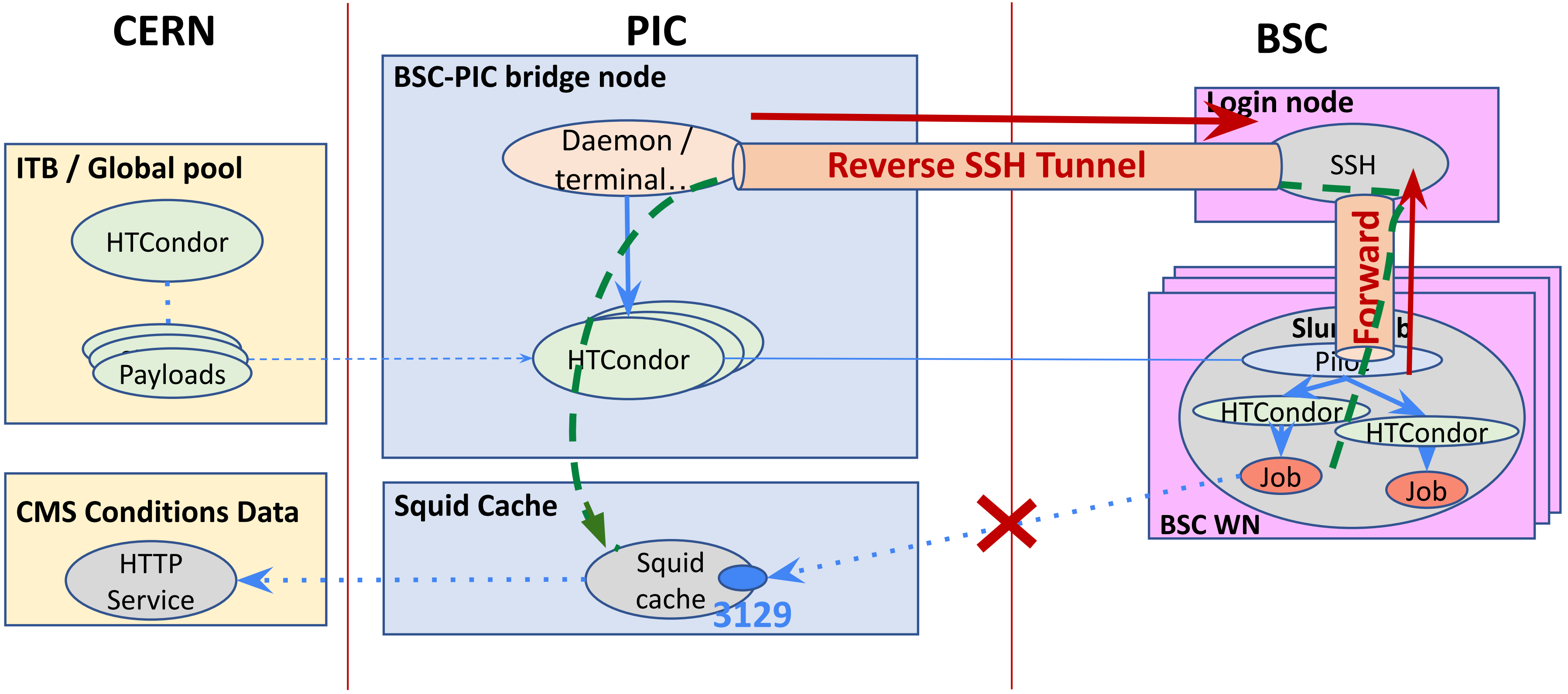
PIC
port d'informació
científica

The solution



PIC
port d'informació
científica

The solution



PIC
port d'informació
científica

Access to conditions using tunnels

The described solution using **SSH tunnels** has been **key to finally move full production at scale**

- Makes it easier to find suitable workflows for BSC
- Relieves CMS from the task of producing SQLite conditions files
- All in all, lighter and more standard operation of workloads

Couldn't the same mechanism be used for output data or CVMFS?

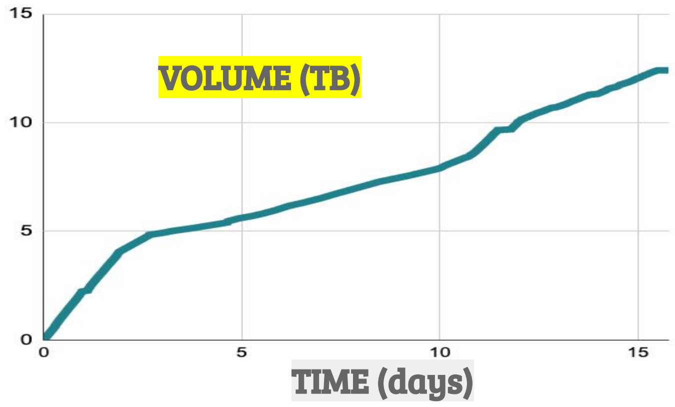
- **No! SSH connections are restricted.** Only a small amount of data can be moved through them, since processes on login nodes are allowed only 5 minutes of CPU time

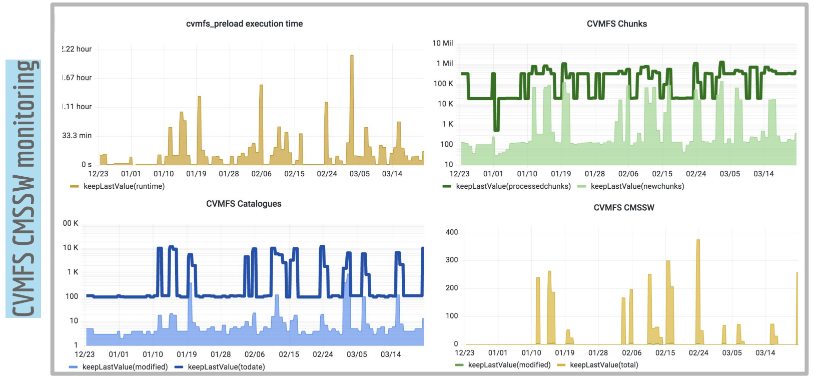
Production-ready **details**:

- Two tunnels are created for each Slurm job, and one is set as back-up proxy for Frontier
- If a tunnel dies for some reason (or killed), it is re-established automatically
- The pilot checks that the tunnels are working properly before starting each new payload

Deploying full CMS CVMFS tree

- Initial copy of the whole *cms.cern.ch* repo:
  - 12.6 TB, **183M files** (37M files de-duplicated)
- Used *cvmfs_preload* tool
  - Avoids duplication, skips already present files
  - Run at PIC, directly into an SSHFS mount of BSC shared filesystem
- Took ~2 weeks to complete
  - After initial phase with frequent transfer errors
    - Intervention at stratum 0 was required
  - Directories with large files showed higher rates
- Following **periodic updates (*deltas*) is much faster**





The Data Transfer Service (I)

A custom Data Transfers Service (DTS) has been designed and implemented in order to manage **output data transfers** from BSC to PIC storage ("mini-FTS" attached to jobs)

- **DTS make asynchronous data transfers but jobs wait until the transfers are completed**
- DTS executes **parallel scp transfers** between BSC and PIC storage. File copy typically adds a few seconds to the total task execution time
- DTS controls **concurrency** and implements **sanity checks** (size, checksum)
- The transfer is **transparent to CMS WM** services, as data is already at PIC when job finishes, then proceeds to register the produced dataset to PIC_Disk RSE
- Fractional output data files (**unmerged files**) generated at BSC are copied to PIC, then standard CMS merge jobs are executed at PIC to produce full size dataset files

Initially, **focus on workflows with no input** (GEN-SIM). The rest of the chain (DIGI-RECO, I/O intensive and needs to read input data files) can then be executed at PIC (or other sites remotely reading those inputs from PIC)

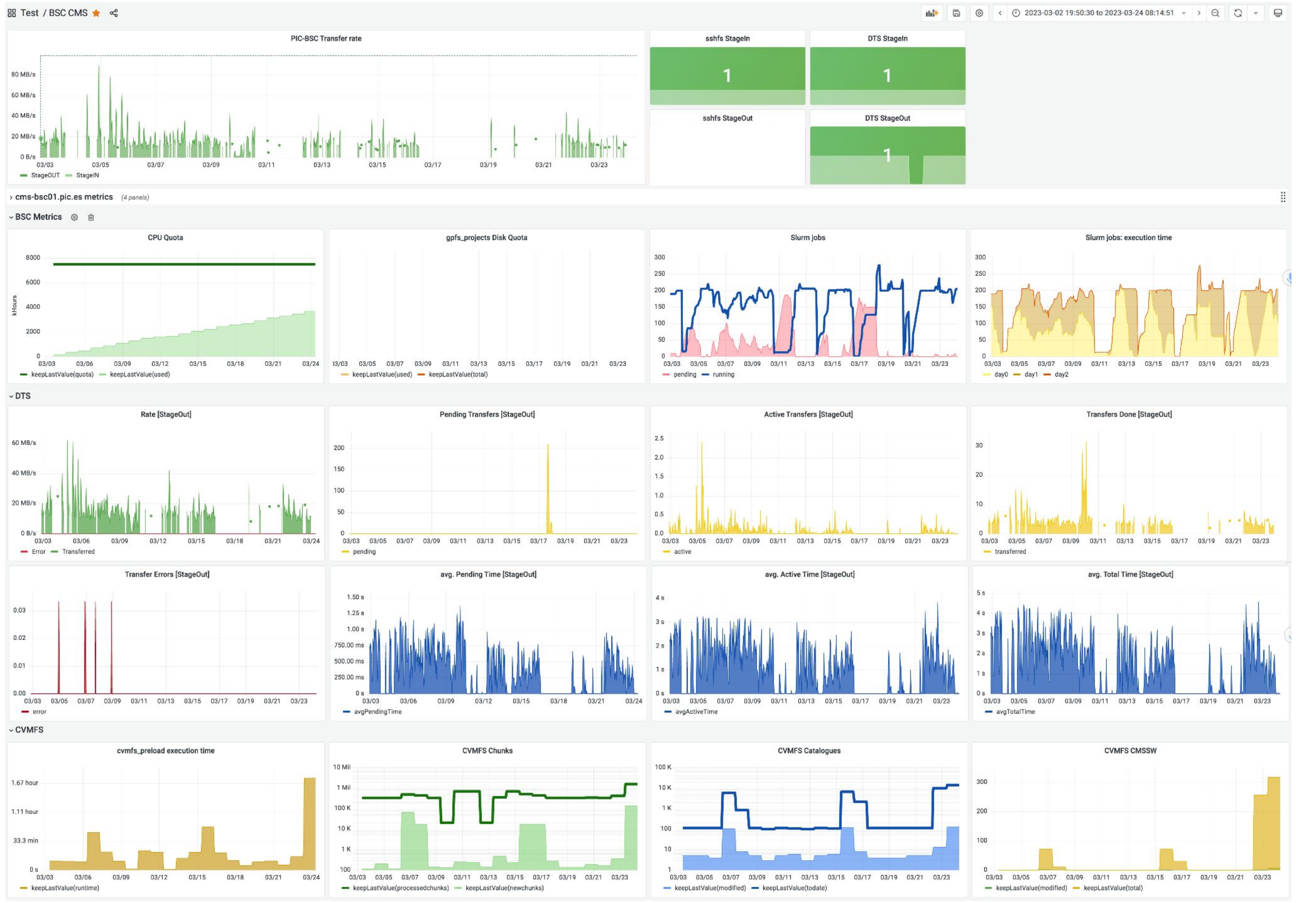
- Singularity container has an **xrootd-server that intercept any xrootd data requests at runtime**
- **Input data** could be served from the DTS, functionality tested, but not yet in production

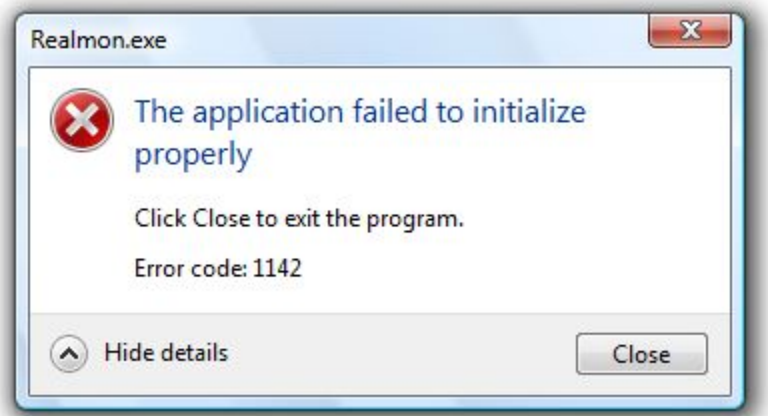
10 Gbps available at BSC. Scale tests showed **we can saturate for hours at 800 MB/s**

PIC
port d'informació
científica

The Data Transfer Service (II)

Monitoring and alarms deployed for the new data service




Realmon.exe

The application failed to initialize properly

Click Close to exit the program.

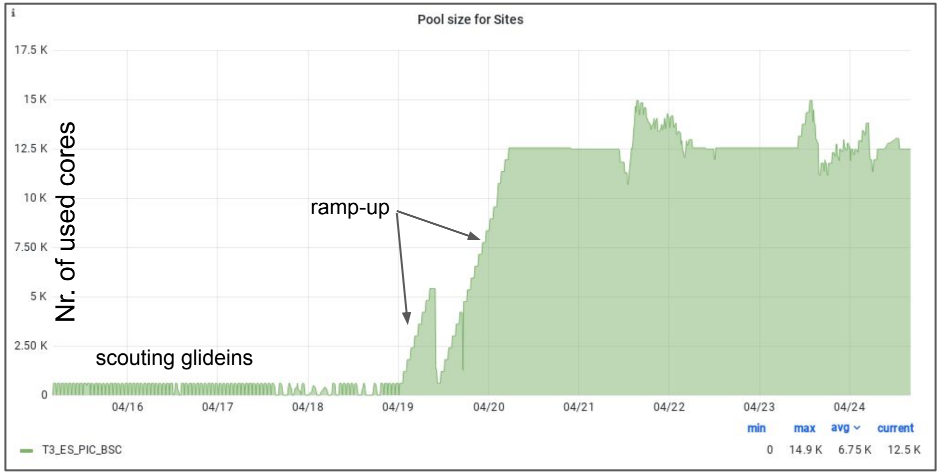
Error code: 1142

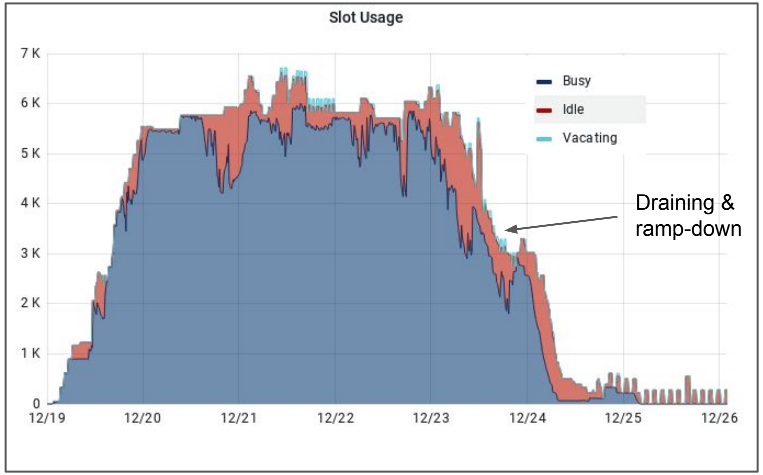
Operations phase

Vacuum model glideins at work

Automated operations relying on the vacuum model pilot jobs

- A batch of scouting glideins is periodically **launched**, acquiring CPU from BSC, launching a HTCondor ***startd process*** that connects to the **CMS HTCondor Global Pool**
- If no **compatible tasks** are present in the CMS workload schedulers (based on HTCondor START expression), glideins **expire** and **CPUs are returned** to BSC scheduler
- When suitable workload is **matched** to the slots, the bridge node sends further resource requests to BSC.
- Ramp-up is conditional to **efficiency of slot utilisation** and a given **maximum total CPU** (270 glideins ~15000 cores)
- Finite **lifetime** pilots at 48h: an internal **draining phase** starts N_hours before pilots max runtime (tuned to 4h)



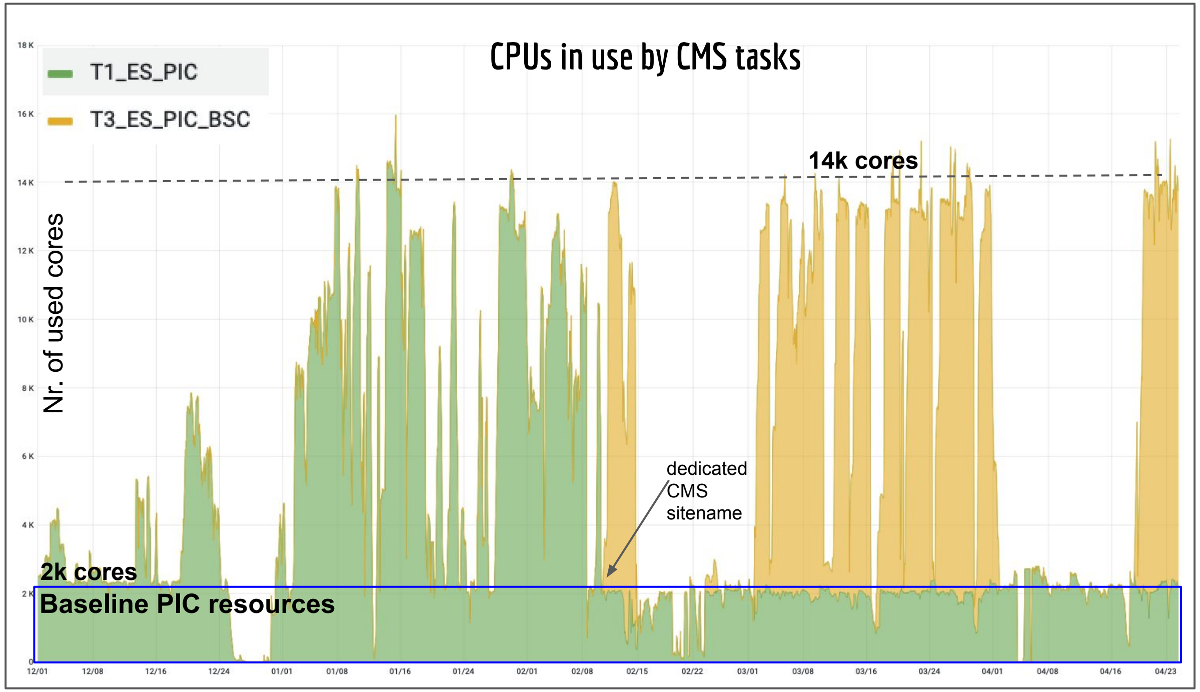


PIC
port d'informació científica

Exploitation phase since Dec. 2022

Use of the BSC resources in CMS under two different models, both employed regularly for HPC integration in CMS:

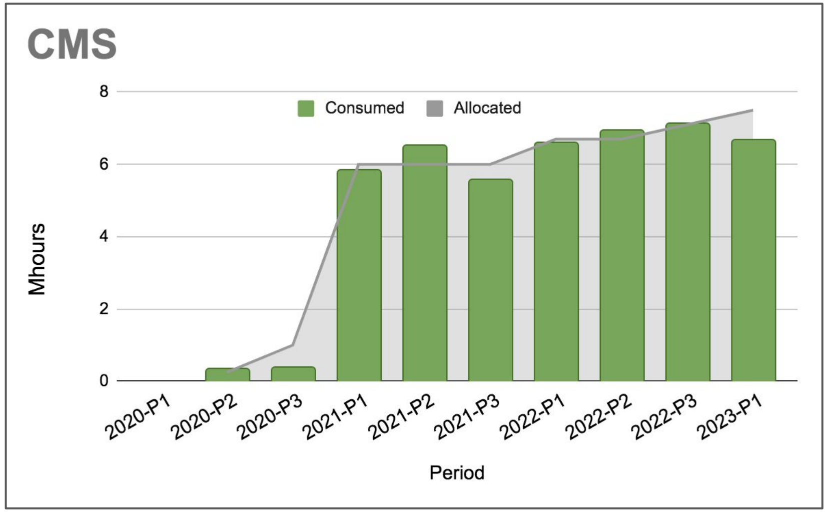
- **Transparent extension of the existing WLCG site** (T1_ES_PIC)
- **Dedicated site name**, used by CMS Computing Operations in order to have the possibility to assign workloads **specifically to the HPC**, i.e. BSC but not PIC (T3_ES_PIC_BSC)
- Can be **reconfigured** with every batch of new glideins, **according to CMS's needs**

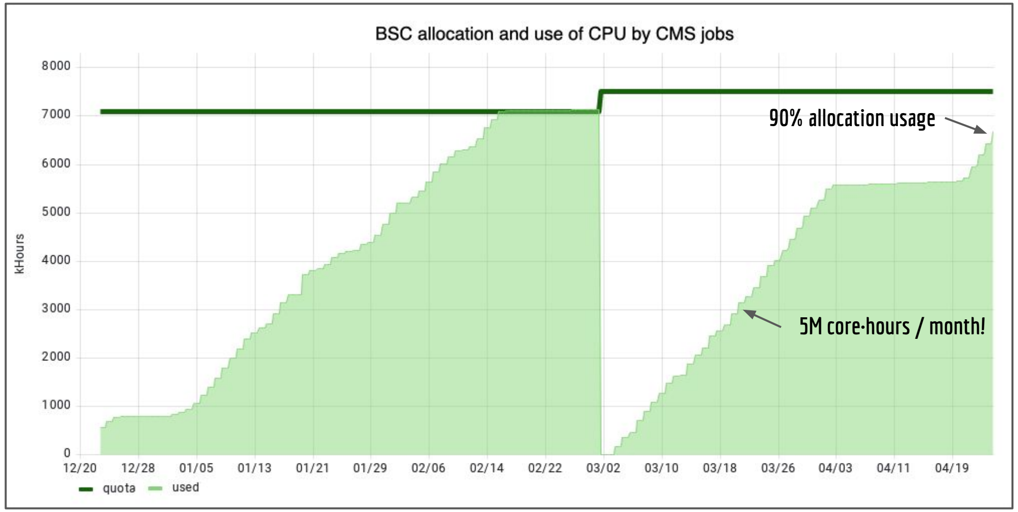


PIC
port d'informació
científica

BSC resource utilization

Operations have progressed successfully, making **fruitful use of the resources** that BSC has been providing to CMS via PIC gateway





PIC
port d'informació
científica

Summary

- Given the network constraints at the HPC, **big efforts have been invested** in the integration of BSC CPU resources for CMS use
    - **HTCondor team**: development of split-starter mode
    - **CIEMAT/PIC team**: interface with CMS and BSC, bridge service deployment, configuration, testing, handling of output datasets, etc
    - **CMS**: Handling of conditions data via files (not anymore a demand). This mode of operation is transparent for workflows that separate the GEN-SIM part (executed in the BSC) and the DIGI-RECO part (executed in the PIC)
- A **fully working system** has been tested **at large scale** and it is now regularly **running CMS simulation production workloads**
    - We can run at saturation, given the BSC maximum number of allowed slurm jobs we can run in parallel as a user - we are not limited by our setup at PIC-CMS!
    - Placed at the top of user's ranking at BSC

Next steps

Still, further refinements can be applied to our setup to improve automated operations and quality of the service

- Enabling and testing all these functionalities in the **upcoming MN5 facility** [link]
- Expected enhanced **network connectivity** between BSC and PIC
  - PIC already at 200 Gbps. BSC expects 100 Gbps along 2023 (current 2x10 Gbps)
- Improving **monitoring**, **alarms** and **automated recovery** for the services
- **Optimize resource usage efficiency** with closer connection to CMS WMS
  - Moving from vacuum-like glideins to GlideinWMS pilot jobs under consideration
- Possible use of a **DTS to manage also input datasets** would increase the usability of the resources by CMS
  - i.e. data reprocessing and analysis tasks
- Automated **accounting** algorithms to properly report BSC usage by CMS

Thanks!

Projects FPA2016-80994-C2-1-R, PID2019-110942RB-C21, and PID2020-113807RA-I00 funded by:

MINISTERIO DE CIENCIA E INNOVACIÓN

UNIÓN EUROPEA

FONDO EUROPEO DE DESARROLLO REGIONAL

"Una manera de hacer Europa"

AGENCIA ESTATAL DE INVESTIGACIÓN

and the Red Española de Supercomputación (RES) through grants FI-2020-2-0025, FI-2020-3-0017, FI-2021-1-0005, FI-2021-2-0010, FI-2021-3-0025, FI-2022-1-0001, FI-2022-2-0026, and FI-2022-3-0036

Backup slides

PIC
port d'informació
científica

Abstract

The CMS experiment is working to integrate an increasing number of High Performance Computing (HPC) resources into its distributed computing infrastructure. The case of the Barcelona Supercomputing Center (BSC) is particularly challenging as severe network restrictions prevent the use of CMS standard computing solutions. The CIEMAT CMS group has performed significant work in order to overcome these constraints and make BSC resources available to CMS. The developments include adapting the workload management tools, replicating the CMS software repository to BSC storage, providing an alternative access to detector conditions data, and setting up a service to transfer produced output data to a nearby storage facility. In this work, we discuss the current status of this integration activity, and present recent developments, such as a front-end service to improve slot usage efficiency, and an enhanced transfer service that supports the staging of input data for workflows at BSC. Moreover, significant efforts have been devoted to improving the scalability of the deployed solution, automating its operation, and simplifying the matchmaking of CMS workflows that are suitable for execution at BSC.

PIC
port d'informació
científica

PIC as gateway to exploit BSC

March 2023

CPU: 114 kHS06
Disk: 18.3 PB
Tape: 69.5 PB

ESCAPE
WLCG
Worldwide LHC Computing Grid
ARCHIVER
CERN
CMS
ATLAS EXPERIMENT
LHCb
MAGIC
cta
cherenkov telescope array
Euclid consortium
T2K
PAU
VIP
VOXEL IMAGING PET
DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science
DARK ENERGY SURVEY
LIGO
VIRGO
DUNE
DEEP UNDERGROUND NEUTRINO EXPERIMENT
MICE

The **Port d’Informació Cientifica** (PIC) is the largest Worldwide LHC Computing Grid centre in Spain
PIC hosts a **Tier-1** site for ATLAS, CMS and LHCb experiments & ATLAS **Tier-2** and **Tier-3**
PIC also supports many **other data-intensive scientific research fields** (Neutrinos, Astroparticle Physics, Cosmology, etc…)

PIC
port d'informació
científica

Solutions applied

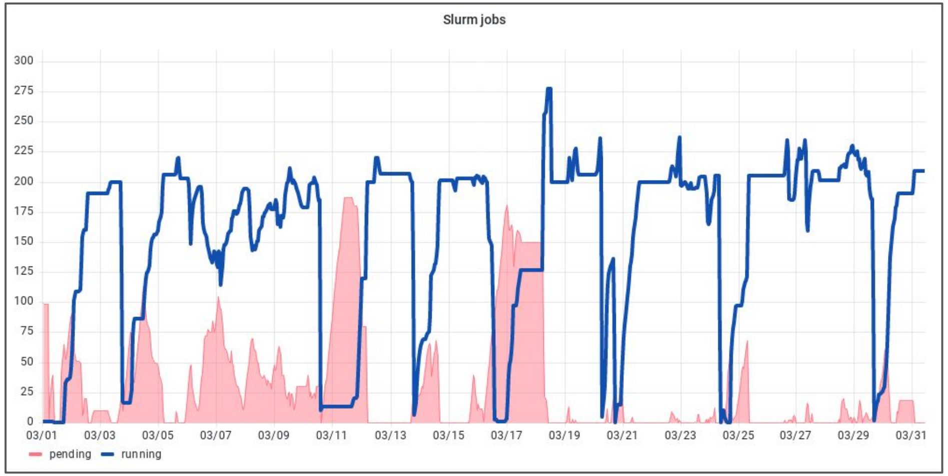
- HTCondor major development: modify CMS resource provisioning, job scheduling and execution framework (HTCondor) to use a **shared file system as communication layer**
  - **Split-starter model,** presented at CHEP19 and CHEP21
  - Requires a **bridge service at PIC** to connect CMS WMS and BSC
- CMS software (**CMSSW**) deployed to BSC environment via CVMFS pre-loaded replica
- **Conditions data** read from sqlite files, pre-placed in BSC GPFS + ssh tunnels
- Local environment configuration for CMS tasks (e.g. where to write output data)
- **Develop** and **operate** a **custom data transfer service** (DTS) for output data migration from BSC to PIC storage

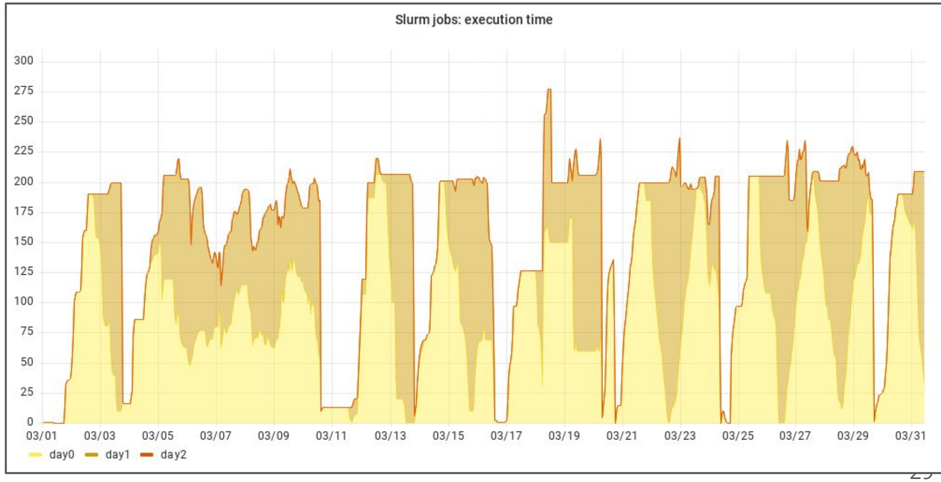
PIC
port d'informació
científica

Operations on BSC slurm

While there is sufficient compatible CMS workload in the queues, a typical operations cycle lasts 48h, as we **request**, **acquire**, **use** and **return** CPU from/to BSC:

- Ramp-up phase is typically fast, acquiring up to the predefined max. CPU allocation target: O(200) glideins
- Our bridge service monitors running glideins age and starts requesting new glideins when previous batch is in draining phase





PIC
port d'informació
científica

Selection of compatible workloads

The START expression of the HTCondor startd process running in our BSC glideins is configured in order to ensure the matchmaking of compatible workloads:

- MC production step not involving pile-up (requires random remote reads at job runtime)
- Considered also data reprocessing, using the DTS as StageIN service (disabled at the moment)

```
START = (ifthenelse(DESIRED_Sites isnt undefined, stringListMember(GLIDEIN_CMSSite,DESIRED_Sites), undefined))
&& ((CMS_RequestType =?= "TaskChain" && CMS_JobType =?= "Production")) && ((GLIDEIN_ToRetire =?= undefined) ||
(CurrentTime < GLIDEIN_ToRetire)) && (true)
```

PIC
port d'informació
científica

Scalability results

- Once the **solutions** described were **deployed**, CMS workflows producing MC simulated datasets were assigned to run at BSC in order to **fully commissioning the new setup**.

- These test workloads were finally executed at BSC to completion successfully, with **comparable error rate and CPU efficiency** to CMS tasks on regular WLCG slots

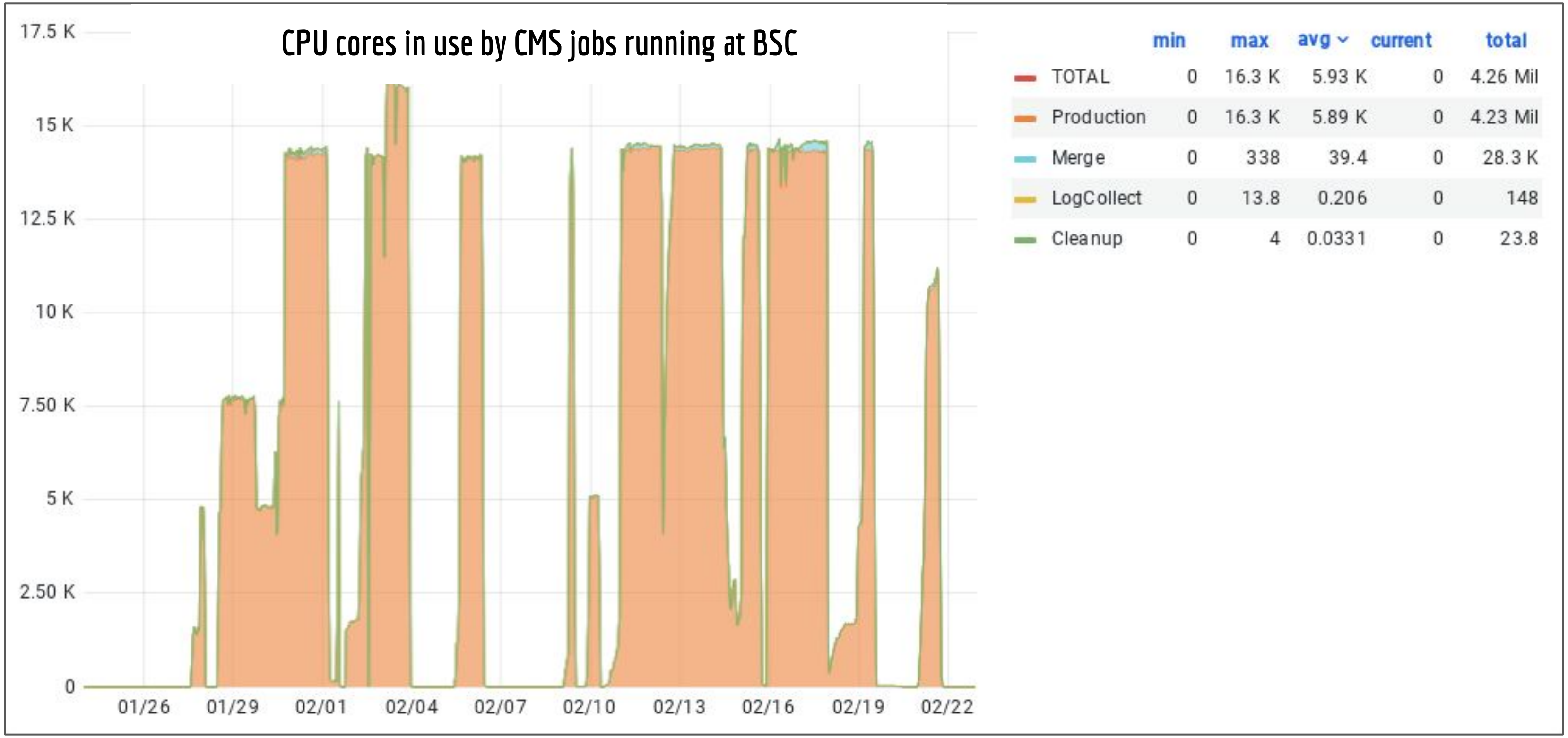
- The new DTS correctly handled file transfers into CMS disk storage at PIC, without introducing additional CPU inefficiency

- The infrastructure and services deployed were proven capable to sustain a scale of ~**15k CPU cores in BSC's MN4** (**500 MB/s aggregate output rate**)

PIC
port d'informació
científica

Execution of CMS workloads at BSC

Integration and tests at scale executed during the Jan-Feb of 2022, with successive corrections and improvements

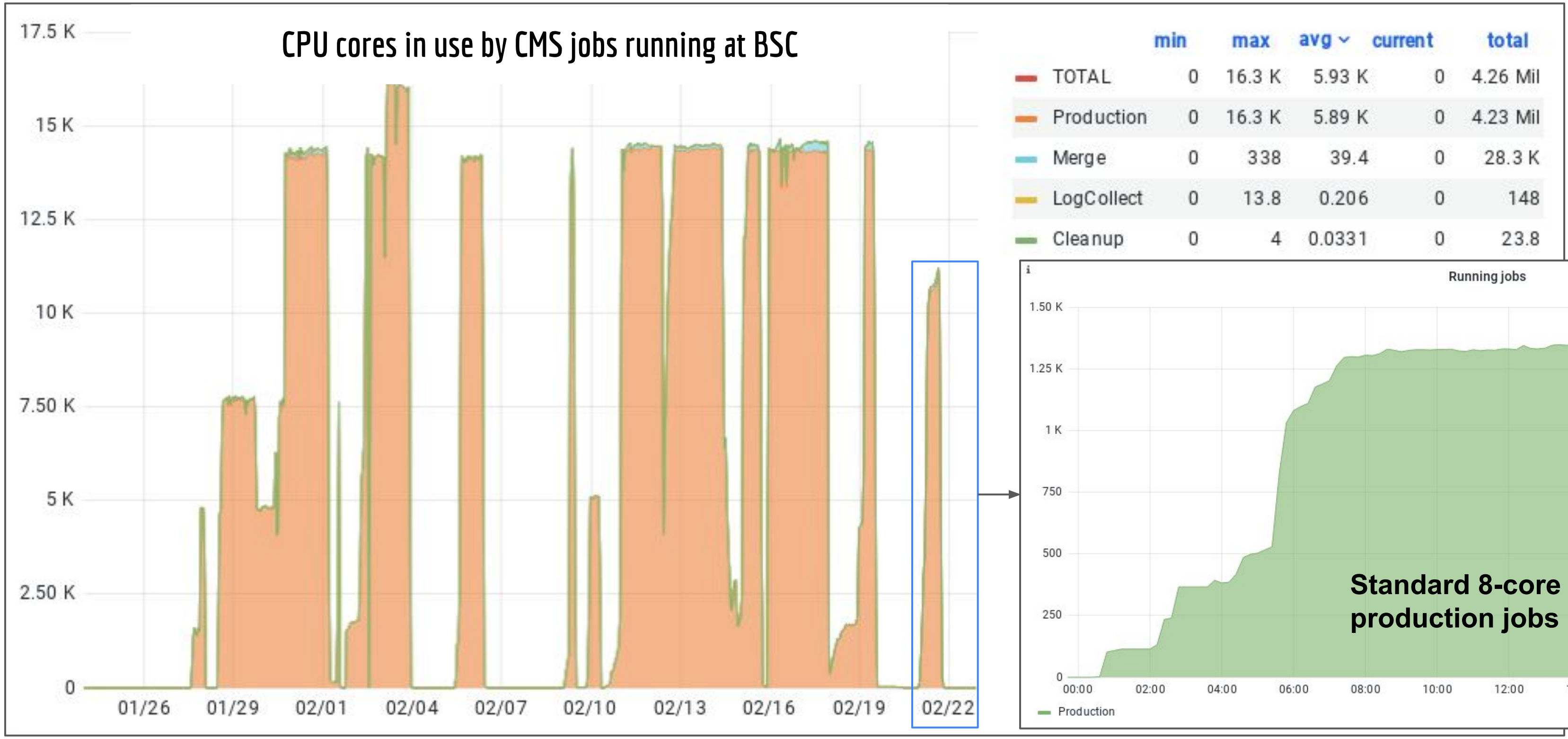


	min	max	avg	current	total
TOTAL	0	16.3 K	5.93 K	0	4.26 Mil
Production	0	16.3 K	5.89 K	0	4.23 Mil
Merge	0	338	39.4	0	28.3 K
LogCollect	0	13.8	0.206	0	148
Cleanup	0	4	0.0331	0	23.8

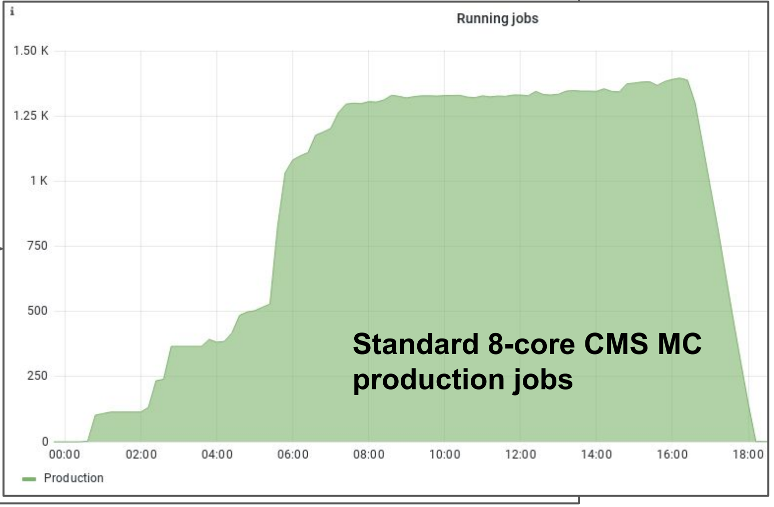
PIC
port d'informació científica

Execution of CMS workloads at BSC

Integration and tests at scale executed during the Jan-Feb of 2022, with successive corrections and improvements



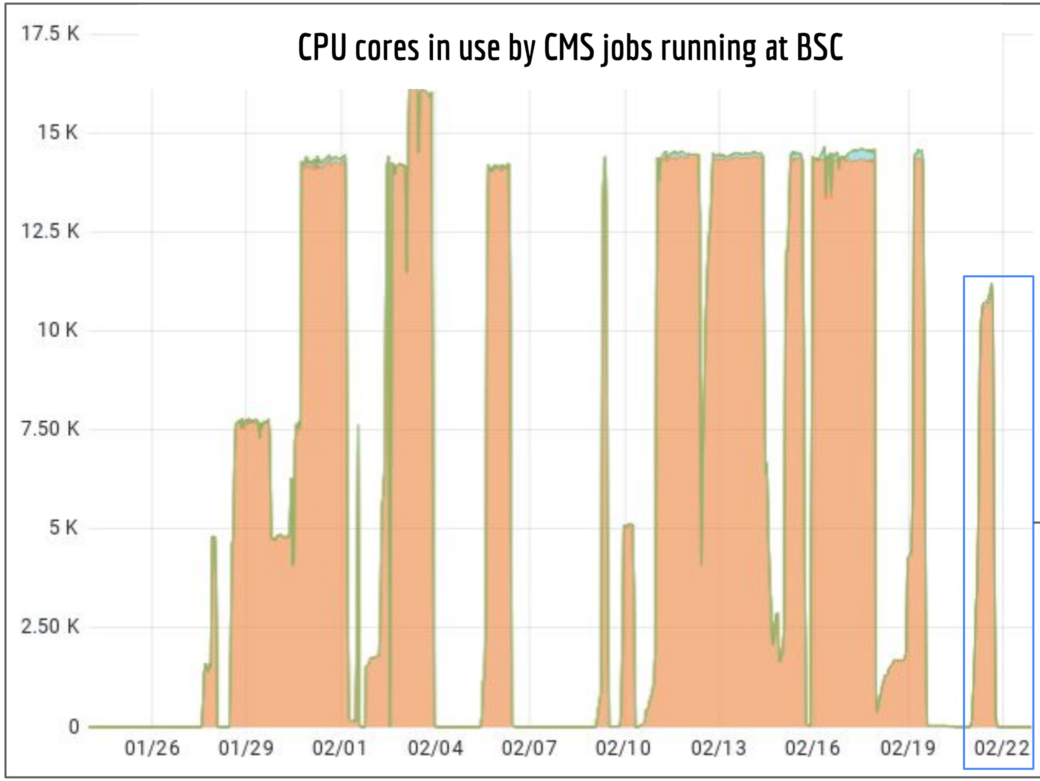
	min	max	avg	current	total
🟥 TOTAL	0	16.3 K	5.93 K	0	4.26 Mil
🟧 Production	0	16.3 K	5.89 K	0	4.23 Mil
🟦 Merge	0	338	39.4	0	28.3 K
🟨 LogCollect	0	13.8	0.206	0	148
🟩 Cleanup	0	4	0.0331	0	23.8



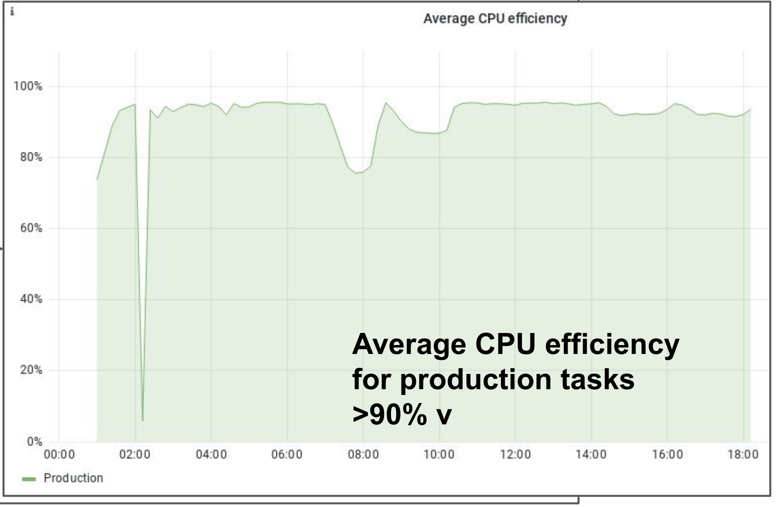
Standard 8-core CMS MC production jobs

Execution of CMS workloads at BSC

Integration and tests at scale executed during the Jan-Feb of 2022, with successive corrections and improvements



	min	max	avg	current	total
🟥 TOTAL	0	16.3 K	5.93 K	0	4.26 Mil
🟧 Production	0	16.3 K	5.89 K	0	4.23 Mil
🟦 Merge	0	338	39.4	0	28.3 K
🟨 LogCollect	0	13.8	0.206	0	148
🟩 Cleanup	0	4	0.0331	0	23.8

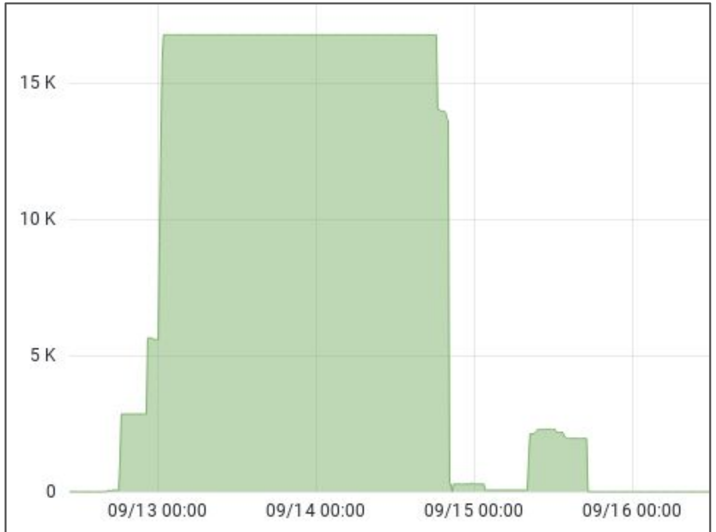


Average CPU efficiency for production tasks >90% v

PIC
port d'informació
científica

Scalability results

A fully working **prototype was tested at scale** with realistic CMS workloads (GEN-SIM, no input data required) and the infrastructure and services deployed were proven capable to sustain a scale of **~15k CPU cores** in BSC's MN4 (**500 MB/s aggregate output rate**)



CPU cores in use during scale tests by CMS @ BSC

A slurm job takes a whole BSC compute node (48-cores)

Run standard 8-core CMS MC production jobs

Average CPU efficiency for production tasks >90%

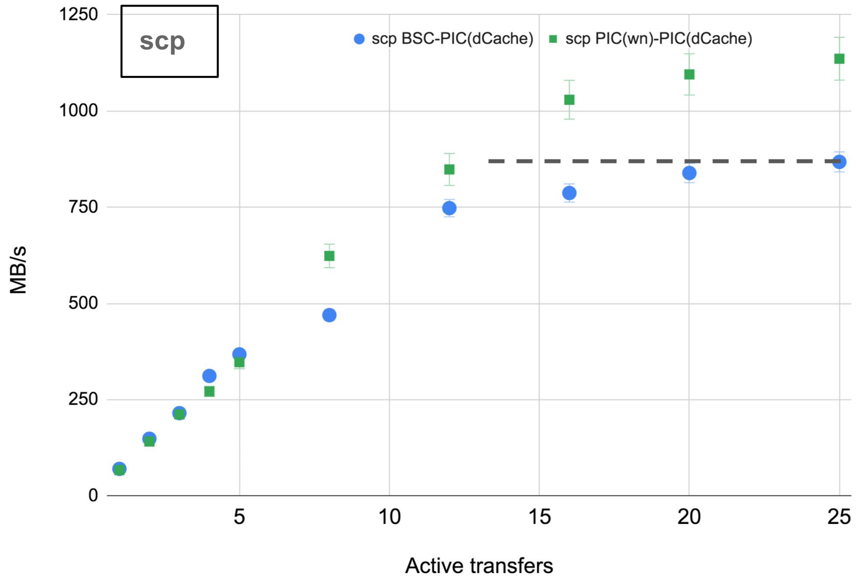
A single bridge node at PIC capable to handle this load and the data transfers from BSC to PIC (AMD EPYC 7452. 64 cores (w/HT) - 128 GB RAM - Network: 25 Gbps)

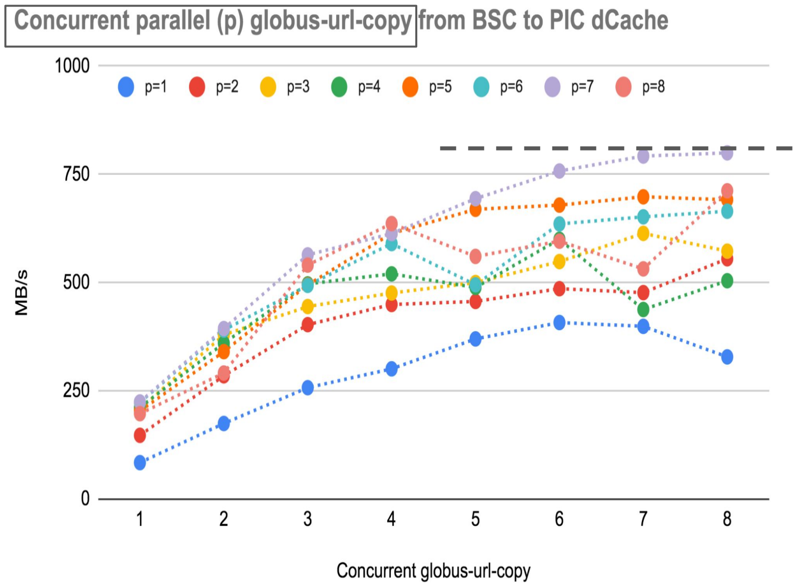
The remaining simulation parts should be executed at PIC (addition of pile-up samples to get the full simulated chain)

PIC
port d'informació
científica

The Data Transfer Service (scp's)

- Tested **single stream multi-file-copies** and **multithreaded** methods (scp's, cp's, globus-url-copy).
- Transfers from BSC **saturating** at ~800 MB/s
  - On a **10 Gbps network link**, so about ⅔ of the max theoretical capacity in use
  - **Similar saturation** regardless of the copy mechanism
- For simplicity (e.g. no extra authentication step), **scp** selected as the DTS method to copy files from BSC GPFS to PIC dCache

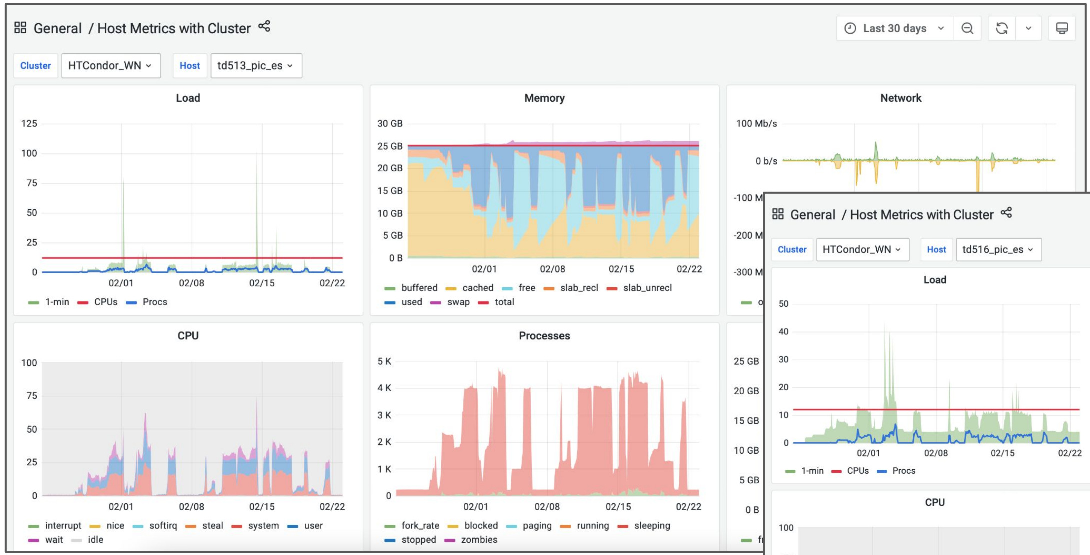




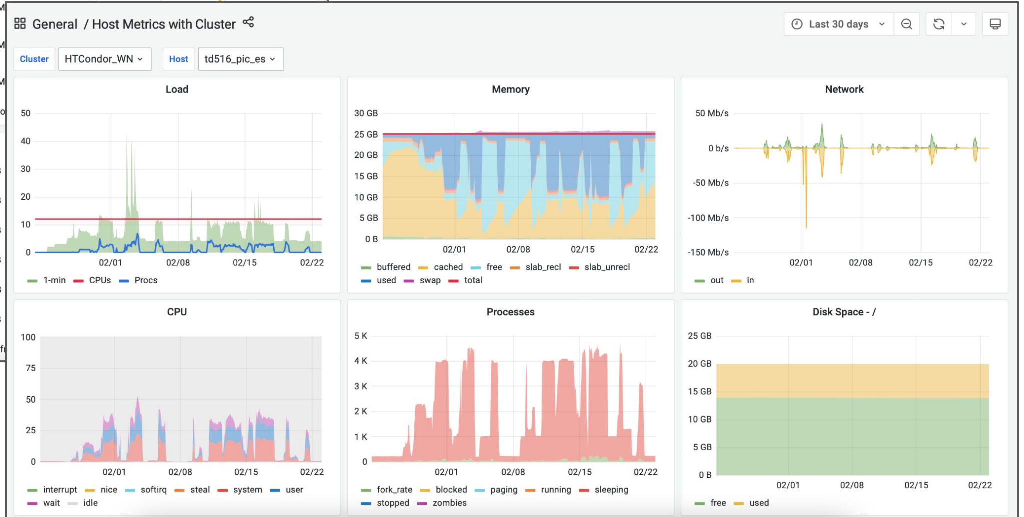
Scalability of the bridge service

The bridge service can be easily **scaled horizontally**, with **multiple bridges acting in parallel to support BSC usage**

- For these tests, bridge deployed on 2 old to-be-decommissioned WNs at PIC (12 CPU cores and 25 GB RAM each)
- Bridge nodes performance carefully watched during the tests
- Results indicate even this modest setup is sufficient to manage **15k BSC CPU core peaks** with **no saturation detected**



General / Host Metrics with Cluster
Last 30 days
Cluster HTCondor_WN
Host td513_pic_es
Load
1-min CPUs Procs
Memory
buffered cached free slab_recl slab_unrecl used swap total
Network
CPU
interrupt nice softirq steal system user wait idle
Processes
fork_rate blocked paging running sleeping stopped zombies
02/01 02/08 02/15 02/22

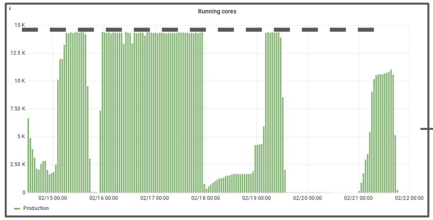


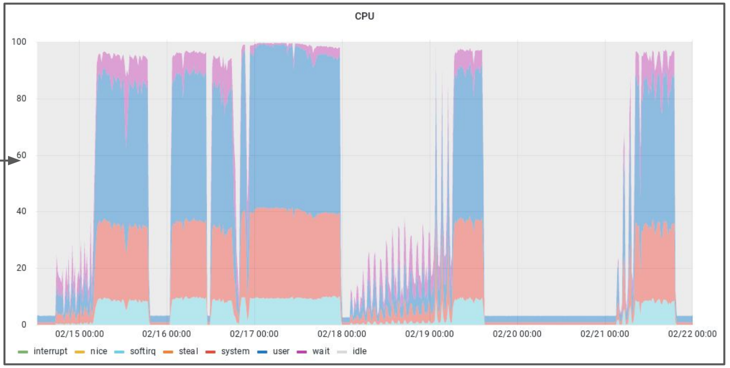
General / Host Metrics with Cluster
Last 30 days
Cluster HTCondor_WN
Host td516_pic_es
Load
1-min CPUs Procs
Memory
buffered cached free slab_recl slab_unrecl used swap total
Network
out in
CPU
interrupt nice softirq steal system user wait idle
Processes
fork_rate blocked paging running sleeping stopped zombies
Disk Space - /
free used
02/01 02/08 02/15 02/22

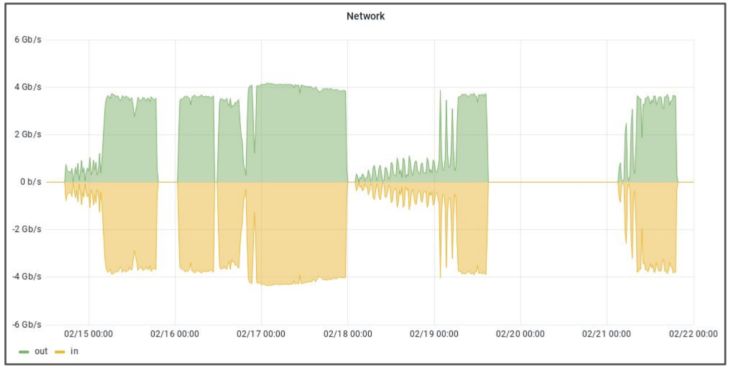
DTS test results (I)

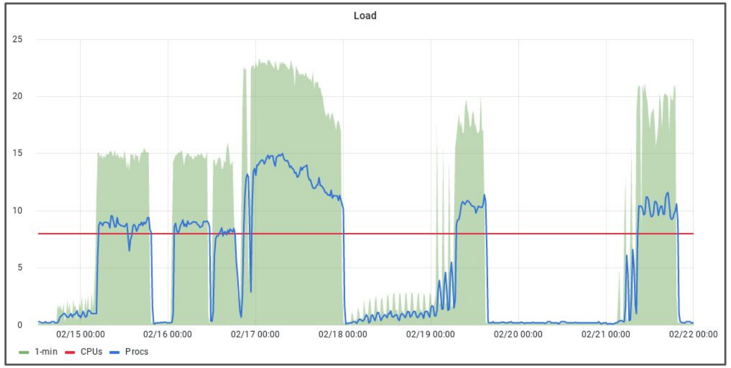
During the tests, the **DTS was deployed on a 8-core 16 GB server with 10 Gbps link**. Running production jobs at maximum scale in BSC (15k CPU cores), the DTS shows some signs of **saturation** (**100% CPU usage**). With **output files at 2 GB**, the effective file transfer to PIC's dCache proceeds at **500 MB/s**

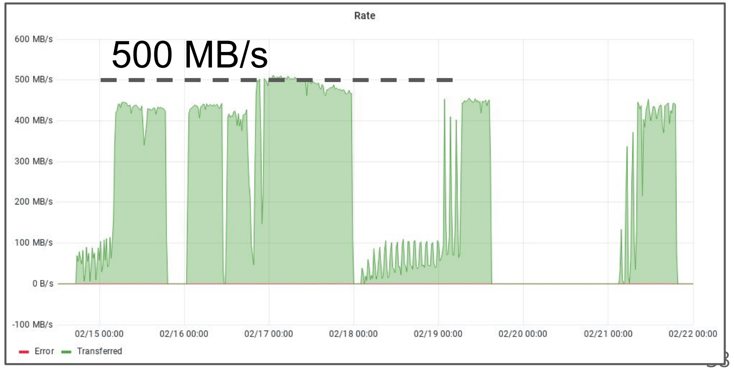
~15k CPU cores in use







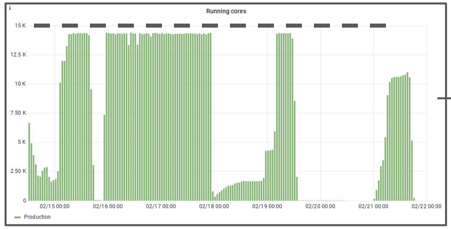


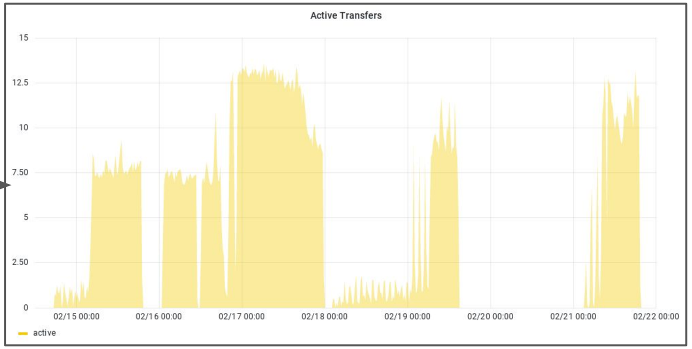


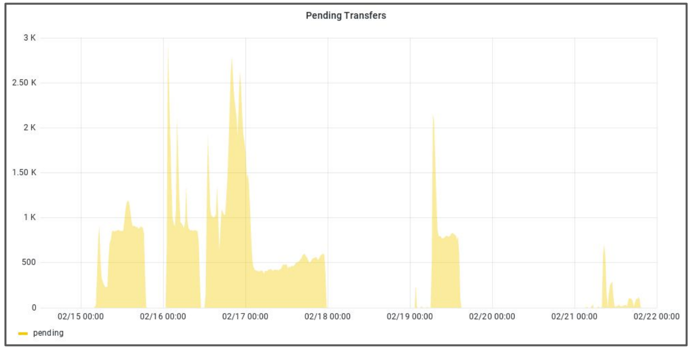
DTS test results (II)

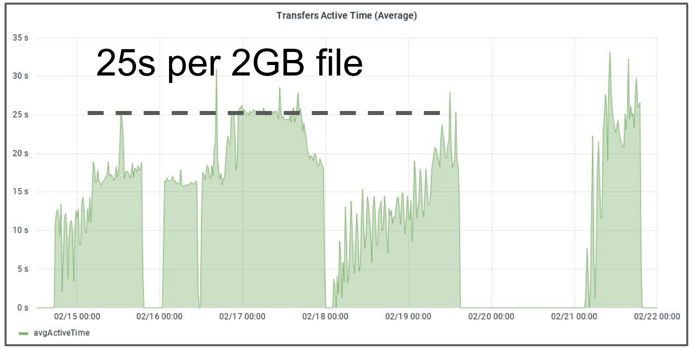
- Producing 2 GB output files, the effective **transfer time** to PIC's dCache is **~25s per job**, while the DTS manages up to **15 concurrent transfers**
- During each of the test rounds, a large number of jobs start and **end their execution close in time.** This causes an **initial saturation** of the DTS that results in **transfer waiting times up to ~2h** while the backlog is processed. Transfers stay **queued until time out, then jobs are marked as error**
- As the test round progresses, **job completion times are randomized** and **transfer waiting time is substantially reduced**

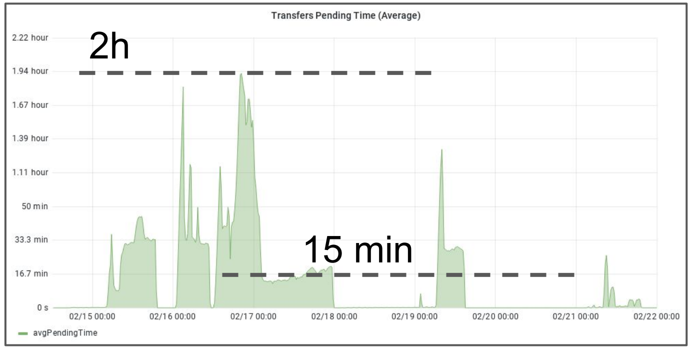
~15k CPU cores in use







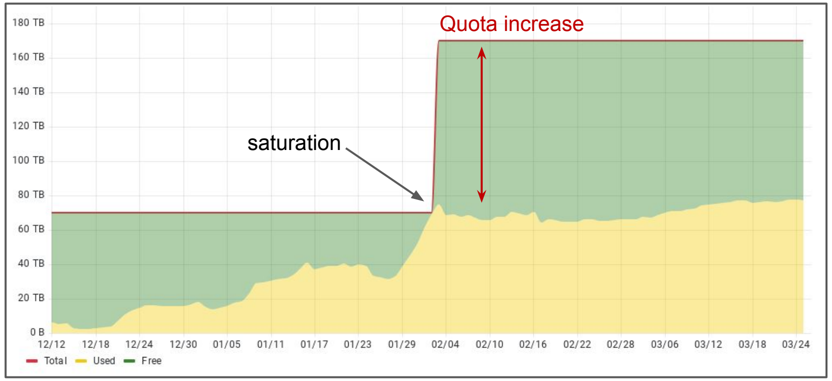




DTS at work

- Our custom-built DTS has operated in general in **StageOut mode** for MC simulation jobs output to PIC storage
- However it can be also configured **StageIn mechanism for data processing tasks**
  - Payload singularity container has an xrootd-server which can be used to intercept any of the xrootd data requests at runtime
  - Problem with saturation of the bridge network capacity, interfering with fast StageOut function
  - Proof of principle achieved, although currently not in use, requires careful tuning of simultaneous data reprocessing jobs
- StageOut to **PIC's CMS temp storage area** risking saturation, due to much larger scale of operations compared to PIC baseline resources
  - Required **quota increase** and active management of temporary files produced by jobs at BSC

Unmerged (temp) volume for CMS tasks at PIC

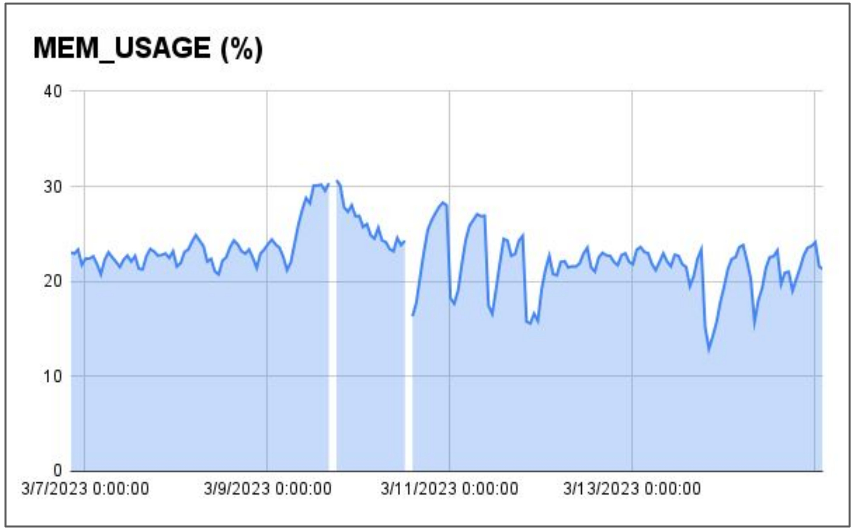


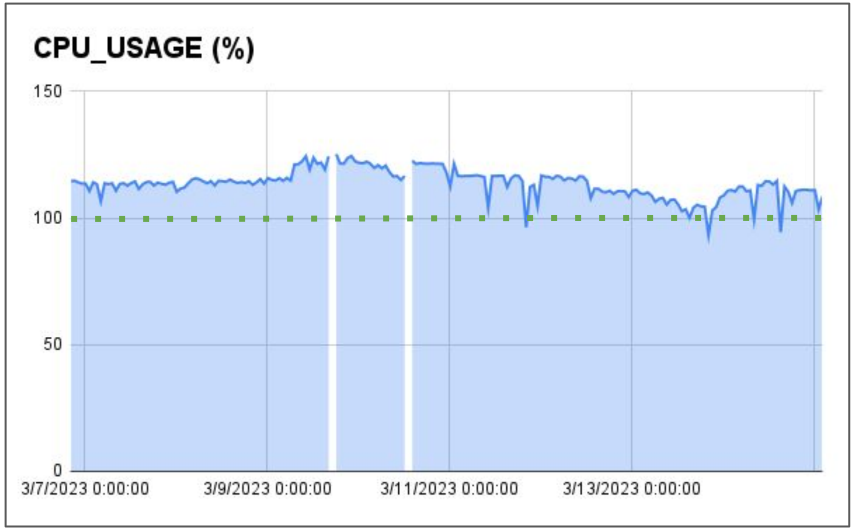
PIC
port d'informació
científica

Performance monitoring

Careful **monitoring and tuning** of CMS whole-node glideins running at BSC **to maximize task execution by CMS**

- Typically, CMS 8-core tasks running on 48 CPU core nodes, but **overcommitted to 56-cores**
- **Maximize CPU usage efficiency while keeping memory utilization under control**





Thanks!

Projects FPA2016-80994-C2-1-R, PID2019-110942RB-C21, and PID2020-113807RA-I00 funded by:

MINISTERIO DE CIENCIA E INNOVACIÓN

UNIÓN EUROPEA
FONDO EUROPEO DE DESARROLLO REGIONAL
"Una manera de hacer Europa"

AGENCIA ESTATAL DE INVESTIGACIÓN

and the Red Española de Supercomputación (RES) through grants FI-2020-2-0025, FI-2020-3-0017, FI-2021-1-0005, FI-2021-2-0010, FI-2021-3-0025, FI-2022-1-0001, FI-2022-2-0026, and FI-2022-3-0036