The Evolution of the CMS Computing Model towards the HL-LHC

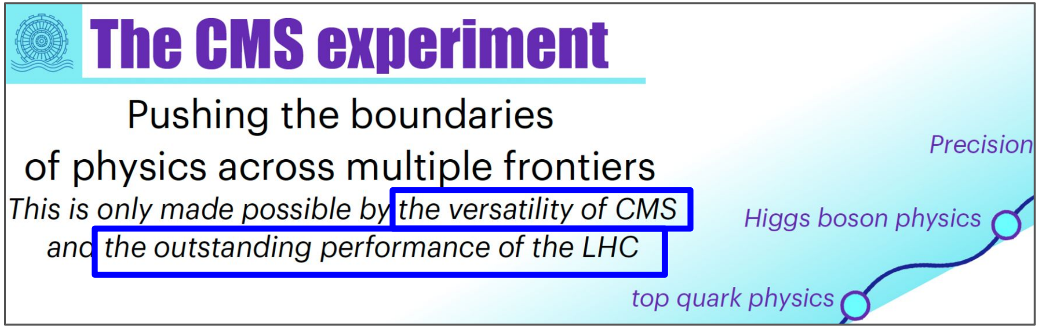
Antonio Pérez-Calero Yzquierdo on behalf of the CMS Collaboration
CIEMAT and PIC

EPS-HEP 2025, Marseille July 10th 2025

Outline of the talk

- Motivation and HL-LHC context
- The evolution of the CMS Computing model for LHC Phase II
- Conclusions and outlook

The role of Computing as enabler of physics!



Monday plenary talk “Highlights from CMS” at EPS-HEP R. Salerno ref.

PIC port d'informació científica

The role of Computing as enabler of physics!

The CMS experiment

Pushing the boundaries
of physics across multiple frontiers

This is only made possible by the versatility of CMS and the outstanding performance of the LHC

Precision

Higgs boson physics

top quar

…and a scalable, flexible and capable Computing infrastructure!

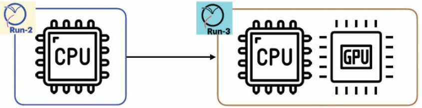
Monday plenary talk "Highlights from CMS" at EPS-HEP R. Salerno ref.

CMS as a technology driver

CMS-DP-2024-082
CMS-DP-2025-034

Heterogeneous computing

Successful R&D effort to run the High-Level Trigger reconstruction on heterogeneous hardware in production since the start of Run-3



Faster and more power-efficient HLT reconstruction

- **+50%** event processing throughput
- **+15-25%** performance per kW

Portable heterogeneous software (Alpaka) deployed

same code base can run seamlessly on CPUs and GPUs

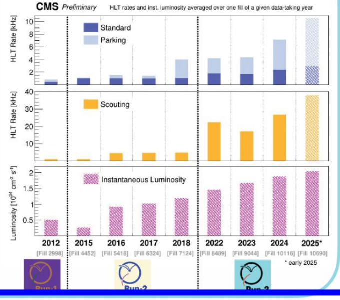
Multiple data-taking strategies

"Standard", "Parking" and "Scouting" pushing CMS trigger coverage far beyond Phase-1 design values

Parking is now integral to the core physics program (Higgs, searches, etc.), and opens large phase space for flavour physics.

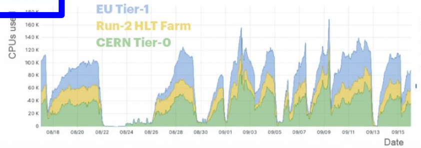
HLT scouting runs at ~30 kHz (1/3 of accepted L1T events)

Using HLT objects for physics analyses in an even wider phase space



Big computing infrastructure

Parking data is processed quickly without delaying, using also the Run-2 HLT Farm and European Tier-1 for prompt reconstruction

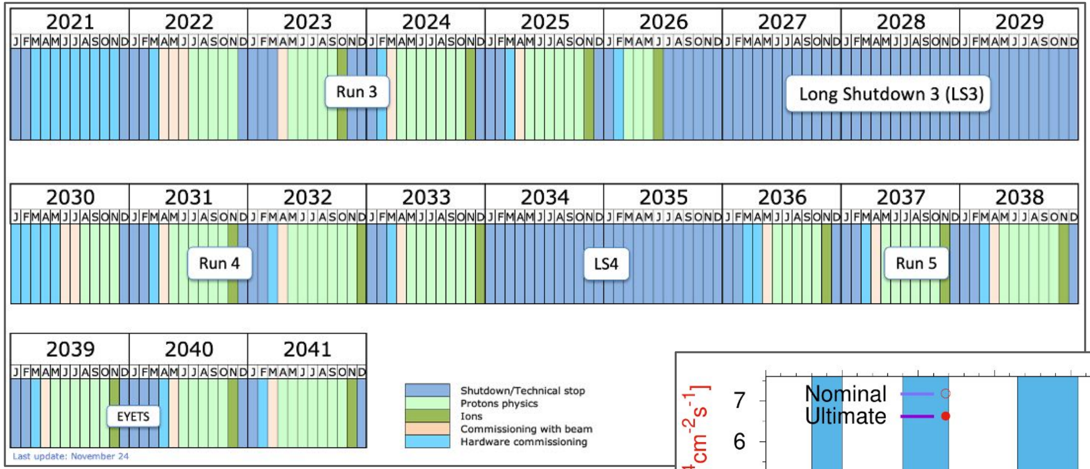


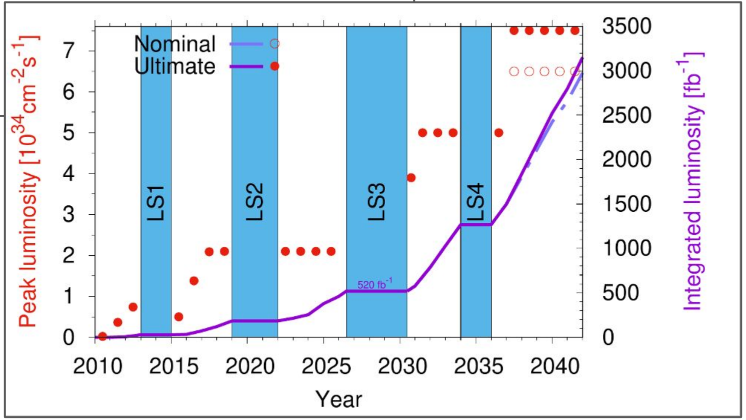
Opportunistic resources, as High-Performance Computing (HPC) facilities, are used for MC production.

In 2024, CMS can generate 1.5B events per week.

The approaching LHC Phase 2: Impact on CMS Computing

HL-LHC: **larger datasets** with **more complicated events** at higher pile-up!

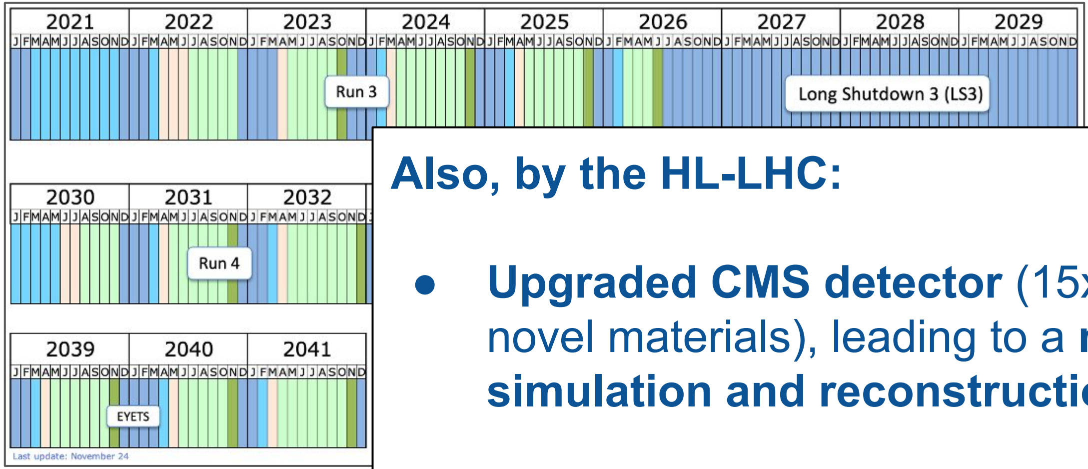




[Ref from Chamonix 25]

The approaching LHC Phase 2: Impact on CMS Computing

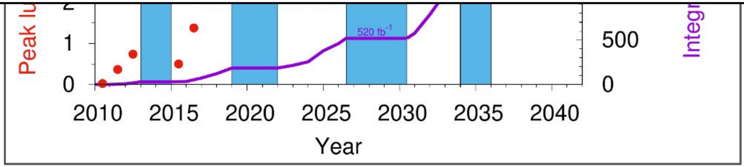
HL-LHC: **larger datasets** with **more complicated events** at higher pile-up!



[Ref from Chamonix 25]()

Also, by the HL-LHC:

- **Upgraded CMS detector** (15x more readout channels, novel materials), leading to a **more expensive simulation and reconstruction**
- **Computing landscape** driven by **external agents** (industry, FAs)



The evolving CMS Computing Model

CMS Computing mission: enable the reconstruction, simulation and analysis of this abundant stream of data, leading to its successful exploitation for the best possible physics results.

Intensive R&D projects required to evolve the CMS **Computing Model (CM)**, making computing **affordable** and **enabling** the collaboration reach its scientific goals.

The evolving CMS Computing Model

CMS Computing mission: enable the reconstruction, simulation and analysis of this abundant stream of data, leading to its successful exploitation for the best possible physics results.

Intensive R&D projects required to evolve the CMS **Computing Model (CM)**, making computing **affordable** and **enabling** the collaboration reach its scientific goals.

CMS is currently finalizing a extensive **Conceptual Design Report** (CDR):

- Demonstrate at the conceptual level that the proposed CM evolution fulfills the requirements from the HL-LHC physics program in the context of the collider and CMS detector scenarios
- Updated strategy from a previous (2022) document (ref)
- Technical Roadmap outside of the scope of this document
    - expect CDR to be followed by a full Technical Design Report (TDR)

Update on ESPP Symposium

"Completing the full HL-LHC programme is essential and must remain a high priority for CERN"

"It is paramount to fully exploit the High-Luminosity LHC (HL-LHC) to maximise scientific returns"

"Key messages from the Symposium", K. Jakobs ref

Computing as a pillar for the success of HL-LHC physics

Future HEP Computing

The goal of the ongoing software and computing development is to facilitate the successful and timely delivery of the best possible physics results

- HL-LHC computing is not an easy journey, but it should be feasible, provided an **intense and uninterrupted R&D** is carried out.
- Likewise, for any future project chosen, the required software and computing should be feasible, **provided adequate person power and funding are available**
- **Successful projects and collaborations** such as WLCG, GEANT, ROOT, KEY4HEP, … and their deliverables and services **need to be sustained**
- Make sure not to miss the opportunities to **adopt new new technologies**, particularly GPUs and AI, as well as Quantum Computing on a longer timefram
- We may not have a clear picture what our computing environment will look like 40 years from now, **but we have an excellent record to adapt and use whatever is available to its full potential**.

K. Jakobs, ESPP Open Symposium, 27th June 2025

Update on ESPP Symposium

"Completing the full HL-LHC programme is essential and must remain a high priority for CERN"

"It is paramount to fully exploit the High-Luminosity LHC (HL-LHC) to maximise scientific returns"

"Key messages from the Symposium", K. Jakobs ref

Computing as a pillar for the success of HL-LHC physics

Future HEP Computing

The goal of the ongoing software and computing development is to facilitate the successful and timely delivery of the best possible physics results

- HL-LHC computing is not an easy journey, but it should be feasible, provided an **intense and uninterrupted R&D** is carried out.
- Likewise, for any future project chosen, the required software and computing should be feasible, **provided adequate person power and funding are available**
- **Successful projects and collaborations** such as WLCG, GEANT, ROOT, KEY4HEP, … and their deliverables and services **need to be sustained**
- Make sure not to miss the opportunities to **adopt new new technologies**, particularly GPUs and AI, as well as Quantum Computing on a longer timefram
- We may not have a clear picture what our computing environment will look like 40 years from now, **but we have an excellent record to adapt and use whatever is available to its full potential**.

K. Jakobs, ESPP Open Symposium, 27th June 2025

Update on ESPP Symposium

"Completing the full HL-LHC programme is essential and must remain a high priority for CERN"

"It is paramount to fully exploit the High-Luminosity LHC (HL-LHC) to maximise scientific returns"

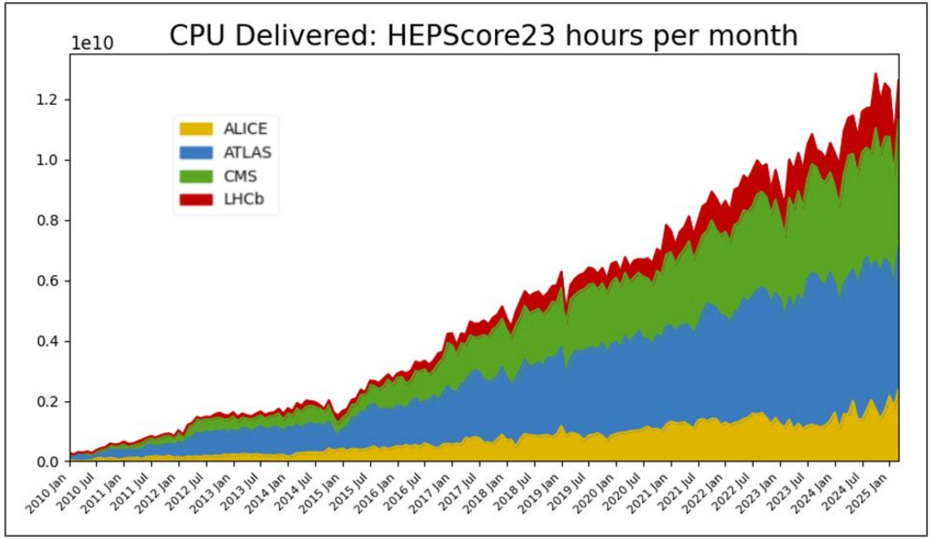
"Key messages from the Symposium", K. Jakobs ref

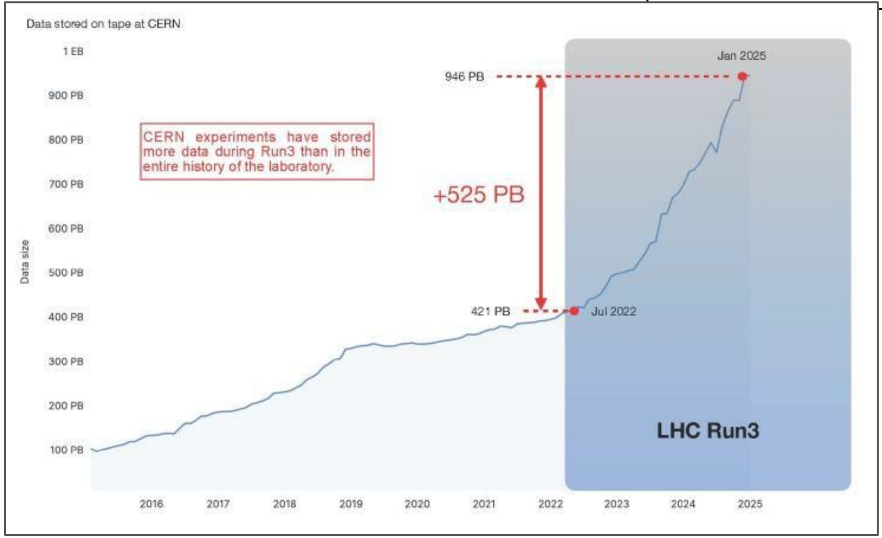
It is crucial for CMS to promote an evolution of our computing model in support of HL-LHC era research

A story of success: WLCG over the LHC lifetime

- Since 2004, the **WLCG has evolved as an infrastructure**, incorporating a diversity of resources (clouds and HPC)
- **Continuous improvements and growth** has been instrumental in the **successful programs of the LHC experiments**
- Achieved while keeping the **system** uninterrupted, **always in production**
- Large efforts from the experiments to **optimize their processing needs without impacting the physics program**
- WLCG **capacity has grown** while **funding has been stable at most**, relying on continuous **model improvements**, **technology evolution** and **integration of accessible opportunistic resources**

S. Campana ref, ref





Evolution of the CMS Computing model

Distributed infrastructures

- Current CMS CM is already quite evolved from the initial Monarc (ref) tiered model: flexibility and technology evolution have created a **convergence of sites "roles"**
  - Data transfers complemented by data streaming from remote storages enabling processing **any data, anytime, *anywhere*** (AAA, ref)
- Potential move towards a reduced number of larger facilities
  - Concentration of compute on large HPCs?
- Data custody problem: compute can be largely externalized, but we want data to remain under direct management of the collaboration
  - But use of caches to reduce latency (***data lakes***)
- Critical dependence on key network infrastructures and their growth (GEANT, ESNet)
  - E.g. **transatlantic traffic**

Event generation and simulation

High-precision measurements require improved precision in physics generator calculations

External R&D efforts in performance improvements for **GEN** and **SIM** code, e.g. vectorization and porting of ME calculations to GPUs,

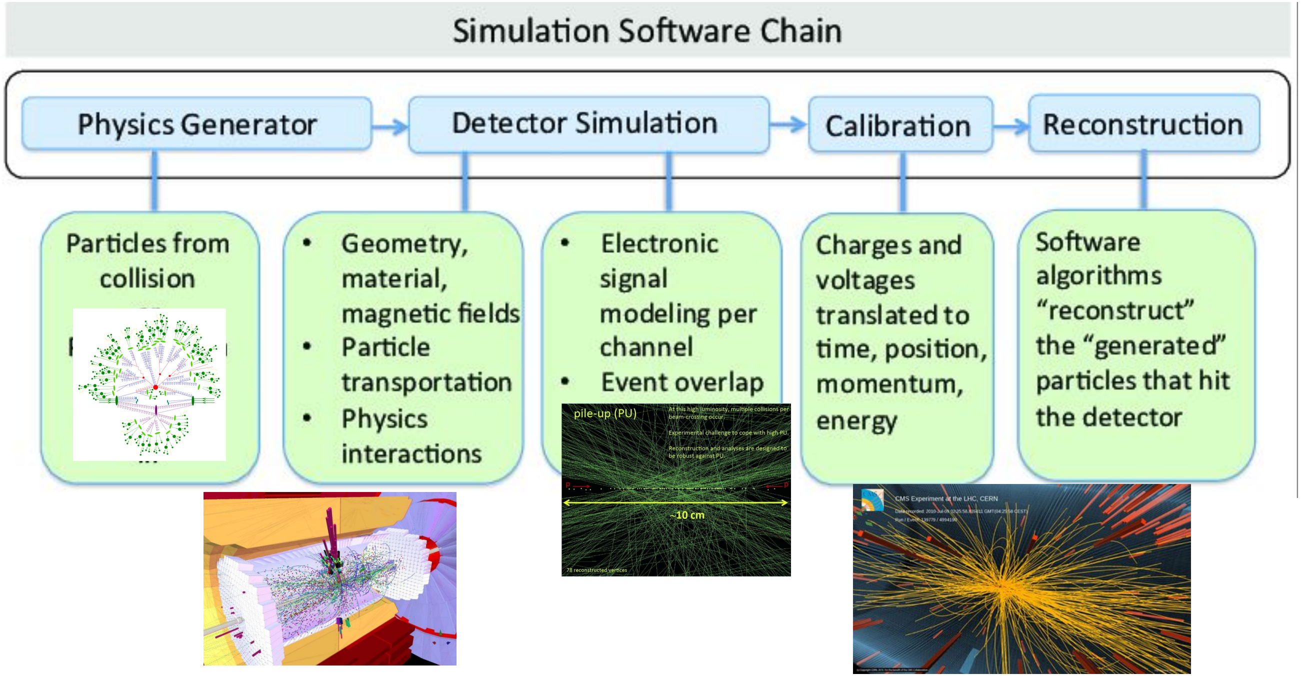
Internally, optimizing CMS-specific code, e.g. move to **FastSim** or **FlashSim**

Streaming premixed **pile-up** events from remote storage

RECO time linear growth with pileup, thus R&D on code improvements and porting to GPUs of vertex, tracking and calorimeter reco.

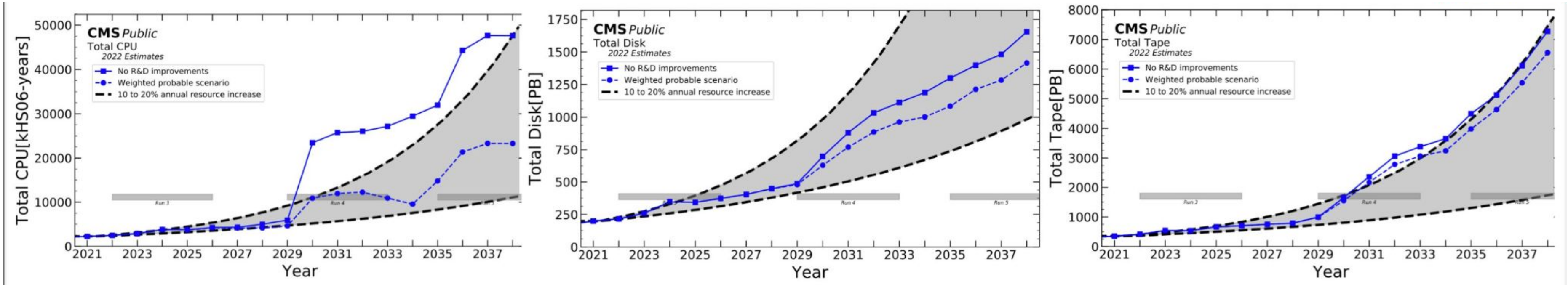
Production of CMS full or reduced analysis formats (**AOD, MiniAOD, NanoAOD**)

StepChain: combined tasks of full simulation chain into **single *job***



Resource needs projections for HL-LHC

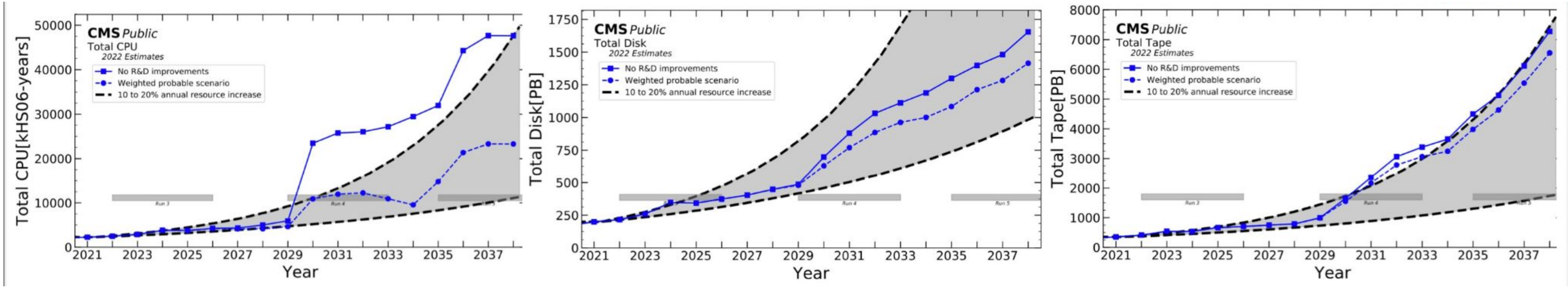
CMS simulates future CM with variations to predict computing resource needs.
Latest projections for the HL-LHC years were made public in 2022 (ref).



- Compare **predicted technology-driven capacity growth** to two basic **CM scenarios**:
    - **Continuation** of the current model
    - Successful **completion of diverse R&D projects** to **minimize computing needs**, use all available **resources effectively** and minimize human **operational costs**

Resource needs projections for HL-LHC

CMS simulates future CM with variations to predict computing resource needs.
Latest projections for the HL-LHC years were made public in 2022 (ref).



- Compare **predicted technology-driven capacity growth** to two basic **CM scenarios**:
  - **Continuation** of the current model
  - Successful **completion of diverse R&D projects** to **minimize computing needs**, use all available **resources effectively** and minimize human **operational costs**
- **Current CDR updates these projections in diverse scenarios of increasingly more aggressive R&D,** to bring the compute needs down to **affordable levels**.
  - **To be released this year**.

Evolution of computing resource costs

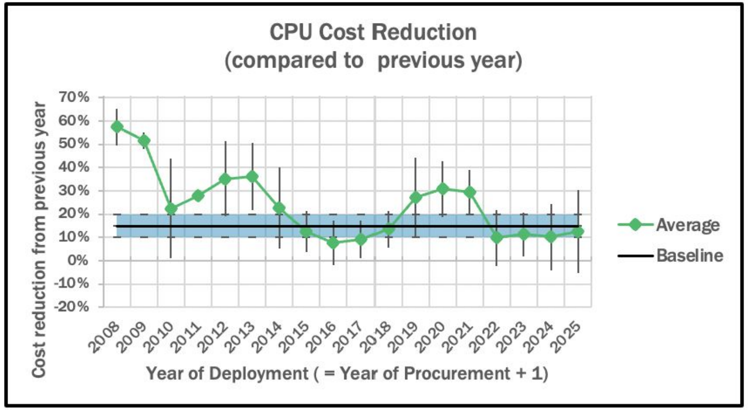
The most recent projected capacity increments based on flat budget might be showing signs of a slow down!

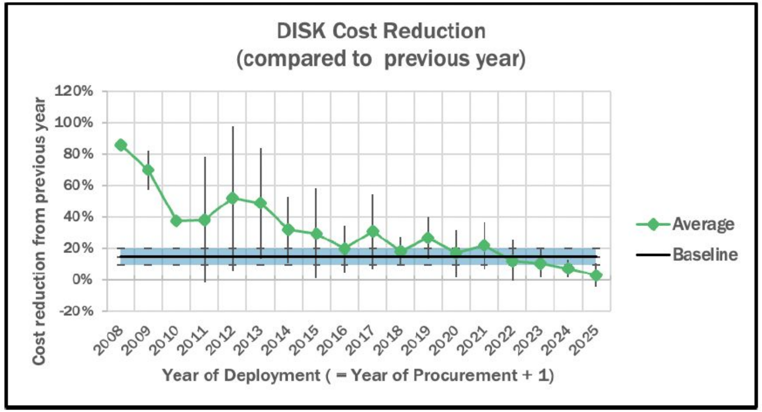
Average cost reduction for CPUs:

last 5 years: 15% - last 3 years: 11% - last year: 13%

Average cost reduction for disk:

last 5 years: 11% - last 3 years: 7% - last year: 3%



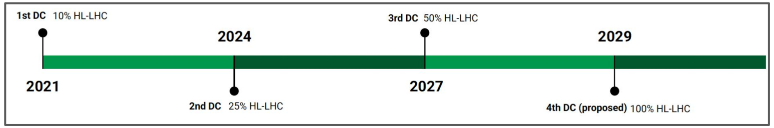


Considerable consequences for resources HL-LHC capacity building!

S. Campana ref
A.. Scieba ref

Data management

Data Management policies evolving and adapting number of **replicas** per data tier, **custodiality**, fraction of prompt reco,



Data transfers::

- **Rucio**, a community-developed product, was adopted by CMS before the Run 3 as key element in the DM model

- CMS DM has been exercised with incrementally harder ***Data Challenges***, planned until 2030
  - **DC24** at 25% of HL-LHC projected needs, went smoothly for CMS (ref)
  - Next stage planned for 2027 at 50% of the required throughput (ref)
  - Mini challenges to exercise concrete components (e.g. data recall from and write to tape endpoints at Tier-1s)

Remote data access in streaming mode is already critical for CMS:

- Very large collections of pile-up events for simulated collisions only available at CERN and FNAL due to constraints in disk capacity availability
- Use of caches for temporary replicas of interesting datasets
- Challenging to monitor and account for!

Technology and market trends being watched in relation to **disks** and **tapes** as basis for our future data storage capacity:

- Limited vendors, market contraction, evolution of costs…

Workload Management

- The CMS **Workload Management** (WM) systems is a set of central services and a pool of distributed agents that execute computing workflows built to satisfy requests from different types of datasets used in data analysis.
- The **Submission Infrastructure** (SI) provides resources for all computing tasks, implementing resource to workload scheduling according to the experiment priorities

Workload Management

- The CMS **Workload Management** (WM) systems is a set of central services and a pool of distributed agents that execute computing workflows built to satisfy requests from different types of datasets used in data analysis.
- The **Submission Infrastructure** (SI) provides resources for all computing tasks, implementing resource to workload scheduling according to the experiment priorities

- The system is currently undergoing an **intense review process** to assess and improve:
  - **Sustainability** and **scalability** of the systems
    - Improved scalability of WM and SI, including bursts
    - Examples in many-core workflows combined with whole-node scheduling, massive use of StepChain type workflows, etc
  - Successful **integration** and **exploitation** of **heterogeneous** computing resources
    - Improved heterogeneous resource and job req. description for improved matchmaking
    - Step-chain workflows with GPUs and/or SONIC
    - Use of non-X86 platforms
  - **Efficiency**
    - Reduced workflow time to completion
    - Automation and improved monitoring for reduced operational cost

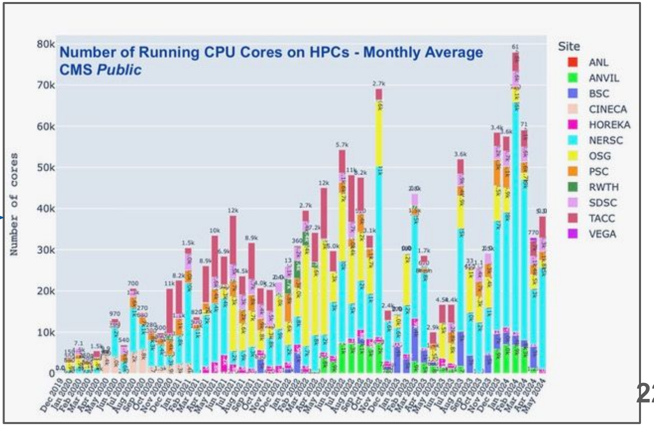
PIC port d'informació científica

The integration of HPCs into CMS model

- **HPCs** identified as **available compute resource** that can **complement** owned centers, presenting a big opportunity for HEP experiments
  - Funding agencies in favor of using these resources, reducing overall computing costs
  - Opportunity to access to new key technologies

- However, **challenging integration** into CMS distributed computing model
  - **Access** and **security policies** differ from site to site
    - Handle technical integration on **case-by-case!**
  - HPC well suited for CPU-intensive workflows (e.g. MC simulation production) but **challenging for data-intensive workflows**
    - Limited/no **network connectivity in compute nodes**
    - Limited storage for caching I/O event data files

- Despite this, use of **HPCs** is **growing** in CMS!



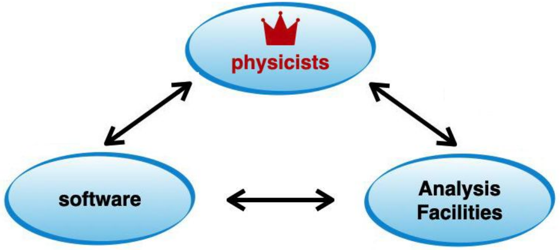
Use of heterogeneous resources

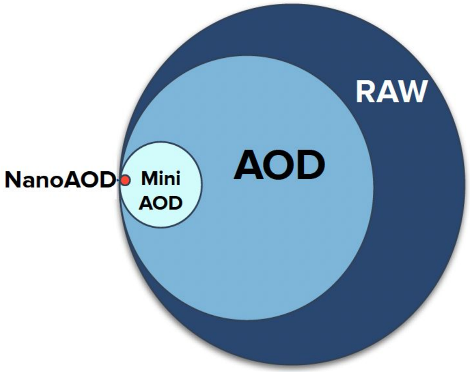
- The experience in use of GPUs in the CMS Run 3 HLT suggests their use can have a **significant impact on offline** reconstruction, tracking and PID algorithms.
  - Also via the use of AI/ML applied in multiple stages of the data simulation, reconstruction and analysis
- Furthermore, the LHC experiments are subject to the **policies in funding agencies** and **market trends**:
  - Push for the use of **accelerators in large HPC facilities**

- CMS SW, written over 20+ years, needs to be **adapted** for diverse heterogeneous HW platforms
  - Use of **performance portability solutions**
  - Need for **experts**, impact on **training** new people
- Modify CMS workload management and resource allocation systems to the new landscape and demands
  - Co-schedule CPU + GPU
  - Heterogeneous computing as-a-Service (SONIC, ref)

Analysis in the HL-LHC era: Analysis Facilities

- **Analysis in the HL-LHC era**: **dimensions** of the problem still being **explored** (ref)
  - Linear with integrated lumi? (ref)
  - Existing "benchmarks": AGC, 200 Gbps Challenge (ref)
- **Analysis Facilities** (AF): concept evolving over recent years to cover expected needs and **reduce *time-to-insight*** (ref, ref) :
  - **Dedicated resources** where users can explore *interactively* large datasets, but also batch
  - **Caches** for local replicas of popular datasets
  - Custom data augmentation
  - Evolution in **analysis software**: python notebooks with dask distributed workload, columnar analyses
- Prioritize centrally-produced **reduced analysis data formats** (estimated sizes for PU 200, approx x5 from today):
  - AOD (~2MB/evt): single replica on tape storage.
  - MiniAOD (~250kB/evt): mainly as input to custom ntuples, few replicas on disk.
  - **NanoAOD** (~4kB/evt): only high level physics objects, N replicas.





relative sizes of CMS data formats

AI for HEP

Boccali, ref

Machine Learning and in general **AI** techniques are **already applied** in CMS in all aspects of the experiment.

Some examples on how AI makes processing tasks

- **Faster/Cheaper**: AI applied to generation and simulation tasks
- **Better**: AI for classifiers, tagging, background rejection
- **New/More**: AI for anomaly detection at trigger level
- **Easier**: LLMs for better coding and lower entry barrier to analysis for students
- **Scalable**: AI applied to computing operations, workload and data management, monitoring and debugging complex systems

Note: big **challenge** in **deploying** and **accessing** performant infrastructures to support the various AI/ML experiment's needs (from WLCG, HPC, AFs, etc)

Environmental Sustainability

Strategic area with multiple initiatives for a **greener**, **more sustainable** computing model to support LHC experiments in general, and CMS in particular.

Summarized in a recent WLCG workshop (ref, ref, ref)

- **More efficient software, computing model, facilities and hardware**
  - **Code** improvements
  - More efficient **hardware** (HS023/watt)
  - Power Usage Effectiveness (**PUE**) improvements for our data centers (e.g. **cooling**)
  - Use of renewable **low-carbon energy sources**, **adjusting capacity** to the local and variable (time-dependent) energy mix
- Be more **efficient** overall!
  - Include power **consumption** of non-CPU components (e.g. **storage**)
  - Overall **accounting** in terms of **consumed power** in e..g failed or incorrectly configured workflows, unnecessary data operations, etc
  - **Train** teams to be **conscious** about sustainability
- Develop metrics that can be widely used: consider **physics output per energy unit**?

Conclusions & Outlook

Conclusions

- **Intense R&D programme** needed in all aspects of the CMS computing model to evolve it and ensure the experiment can reach the scientific goals by the HL-LHC.

- **Flat-budget** model based capacity growth seems to be **slowing down** motivating the collaboration to **double the efforts to bring computing needs to affordability**.

- CMS is **finishing the CDR** (to be released in 2025) with **update resource needs** projections and **R&D actions**. It will likely be followed by a **full TDR**.

- **Continuous evolution** of Computing during the LHC lifetime as a key enabler of physics, plus a highly capable and motivated **team** of experts demonstrate **It can be done!**

Thanks!

Project co-funded by the Recovery and Resilience Facility (Spain), and the European Union – NextGenerationEU

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Financiado por la Unión Europea
NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

Research projects PID2019-110942RB-C21, PID2020-113807RA-I00, and PID2022-142604OB-C21 funded by:

MINISTERIO DE CIENCIA E INNOVACIÓN

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN

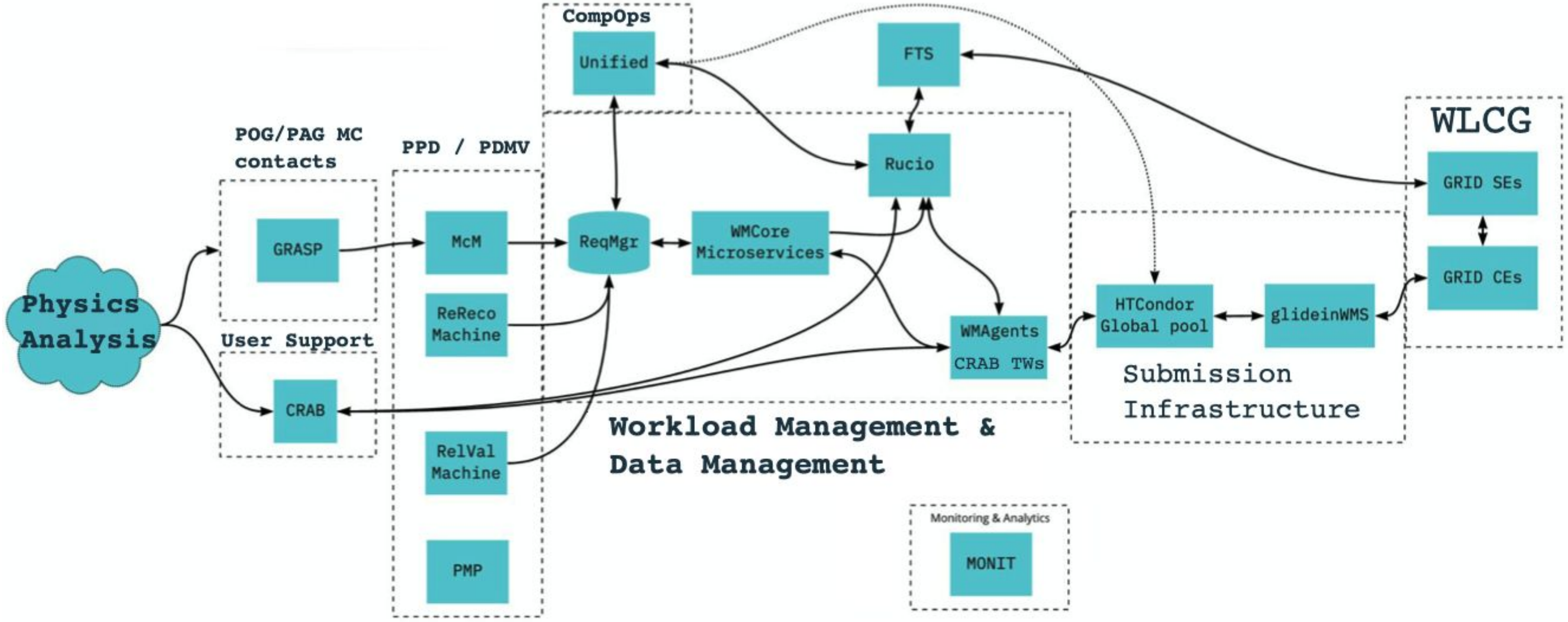
Backup slides

Abstract

With the approaching High Luminosity phase of the LHC programme, scheduled to start in 2030, the Offline Software and Computing group of the CMS collaboration is reviewing the experiment's computing model to ensure its readiness for the computing challenges the HL-LHC poses. An in-depth revision of the current model, tools and practices is being carried out, along with a programme of R&D activities to evolve them, with the goal of effectively managing the available computing resources towards a successful exploitation of the data abundance provided by the HL-LHC.

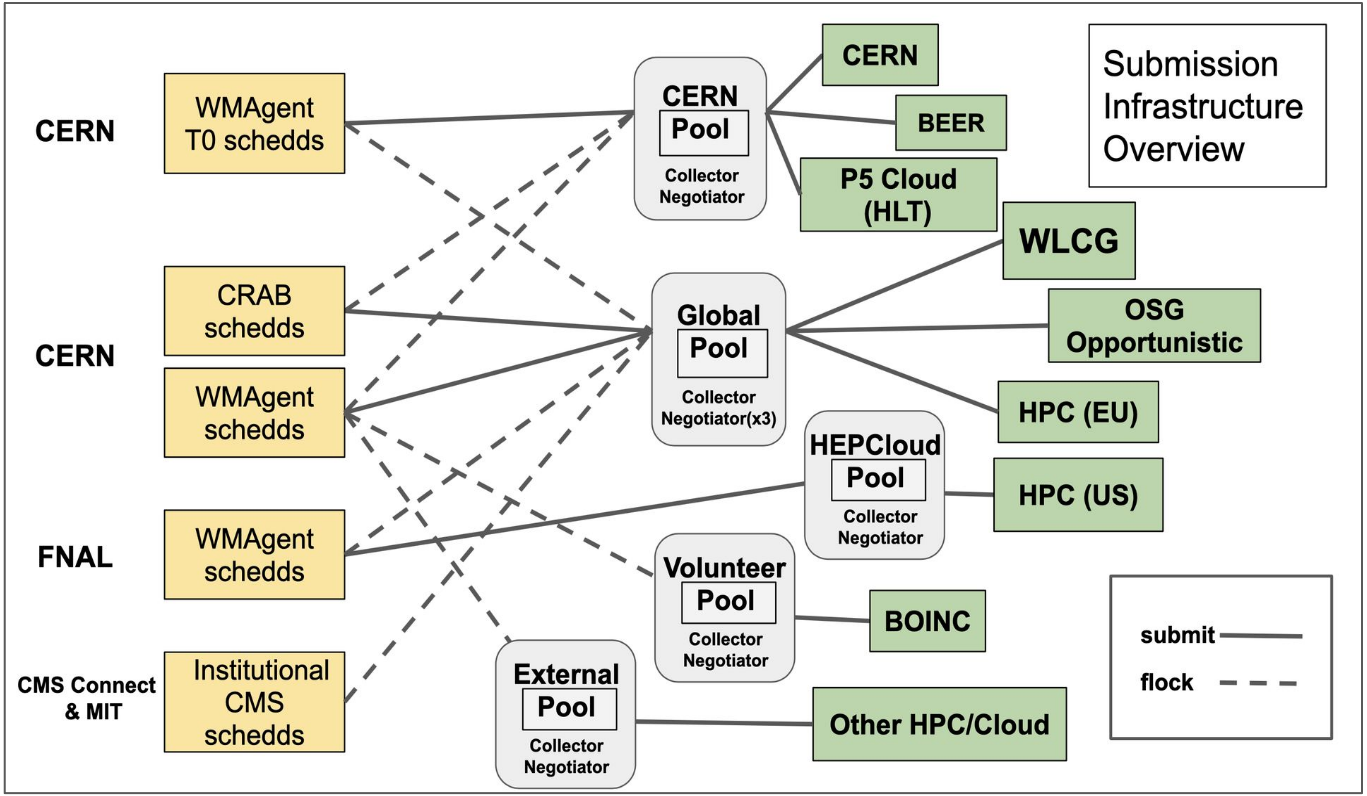
This evolution includes a review of the data management and workflow management systems, to ensure their scalability to the growing data volumes and level of resources required to store, process and analyze them. Larger and more heterogeneous computing resources, both from grid and high-performance computing centers, are being integrated to the global CMS computing pool. Critical progress is being made also in the adoption of new data analysis infrastructures and paradigms. Additionally, further integration of novel technologies in multiple computing areas is being explored. This contribution will present the status of the CMS computing model evolution in the aforementioned key areas, among others.

The current CMS WM system

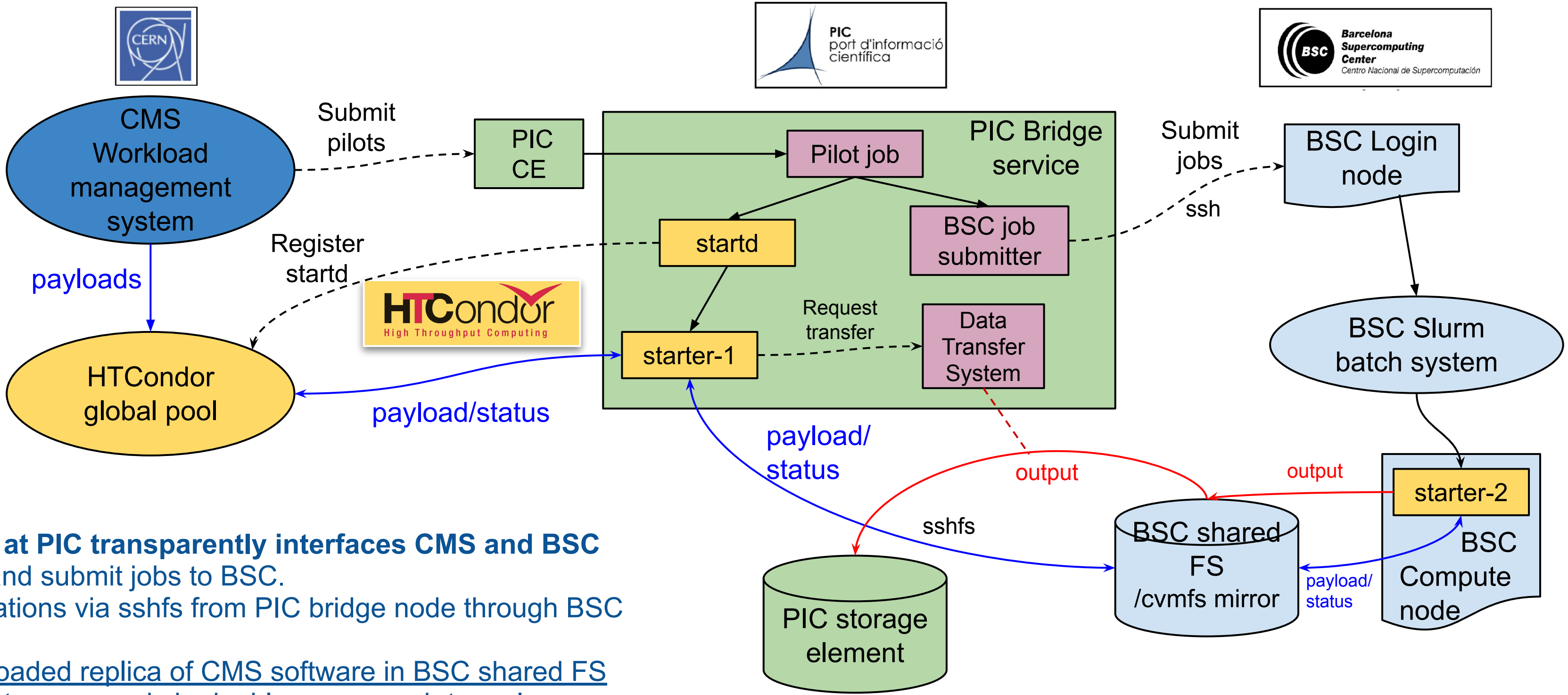


The CMS SI: federated HTCondor pools

- Types of access point
- Types of execution point



Imaginative HPC integration solutions



- **Bridge node at PIC transparently interfaces CMS and BSC**
  It runs pilots and submit jobs to BSC.
  All communications via sshfs from PIC bridge node through BSC login node
- CVMFS pre-loaded replica of CMS software in BSC shared FS
- Conditions data accessed via double reverse ssh tunnels
- Custom data transfer service in bridge node at PIC to move output data files to CMS storage

CMS AF Architecture example: CIEMAT