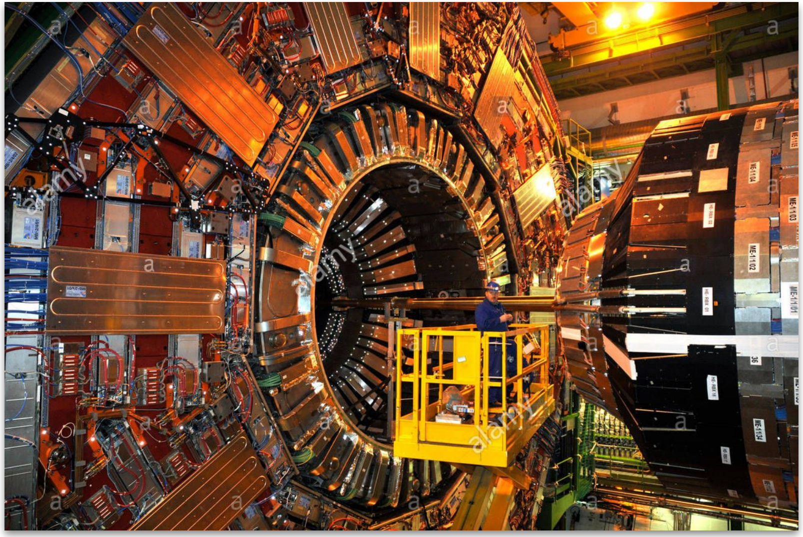
Reaching new scales in the CMS Global Pool

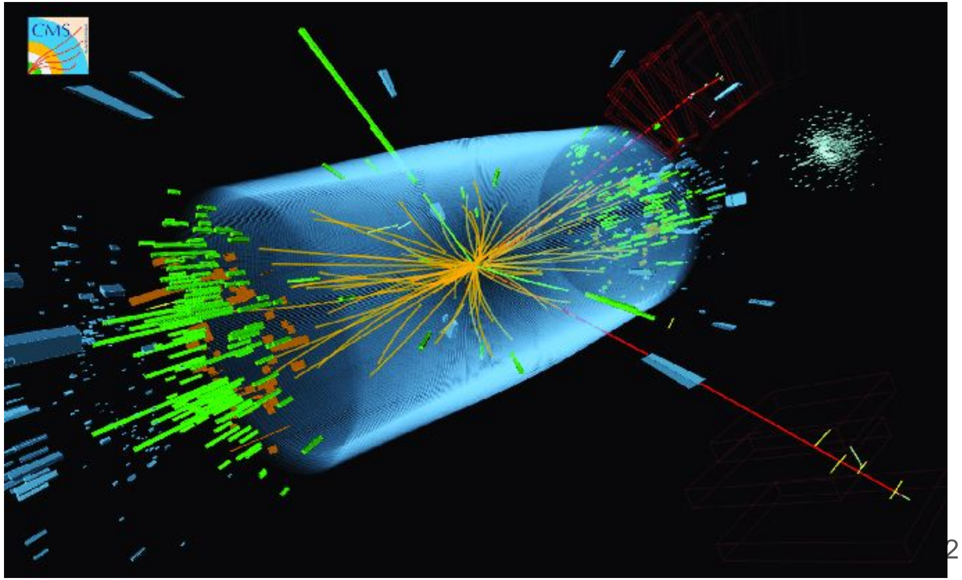
Marco Mascheroni (UCSD), Antonio Pérez-Calero Yzquierdo (CIEMAT-PIC), Edita Kizinevic (CERN), Farrukh Khan (Fermilab), Hyunwoo Kim (Fermilab), Nikos Tsipinakis (CERN)

Throughput Computing 2023, July 13th 2023

The CMS experiment at CERN



- **High Energy Physics general-purpose experiment** recording proton-proton collisions **at the LHC** at CERN



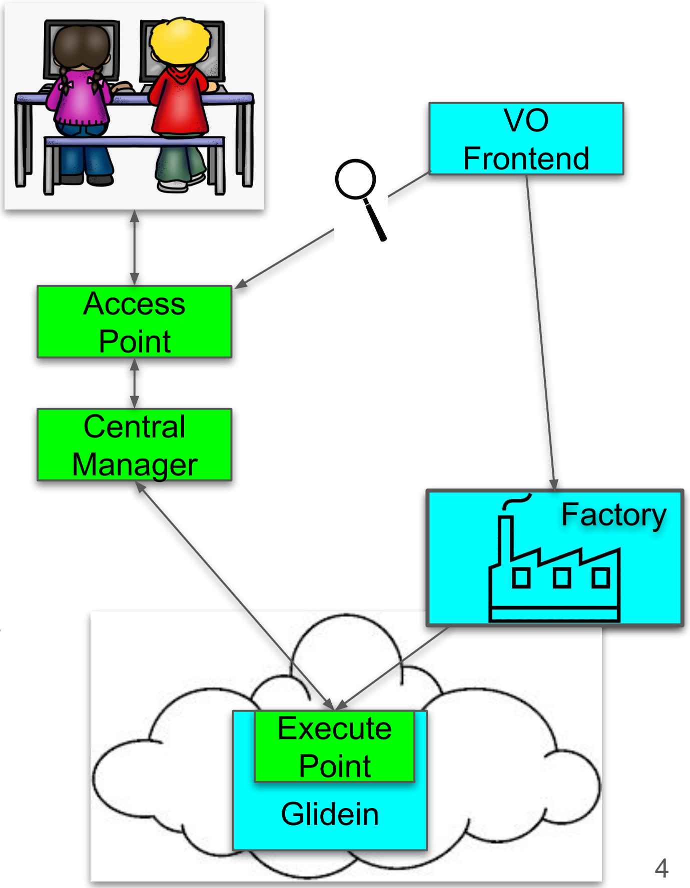
- Experimental data is stored, distributed, reconstructed, and analyzed, comparing to simulated data (Monte-Carlo)
    - **Hundreds of PBs per year**

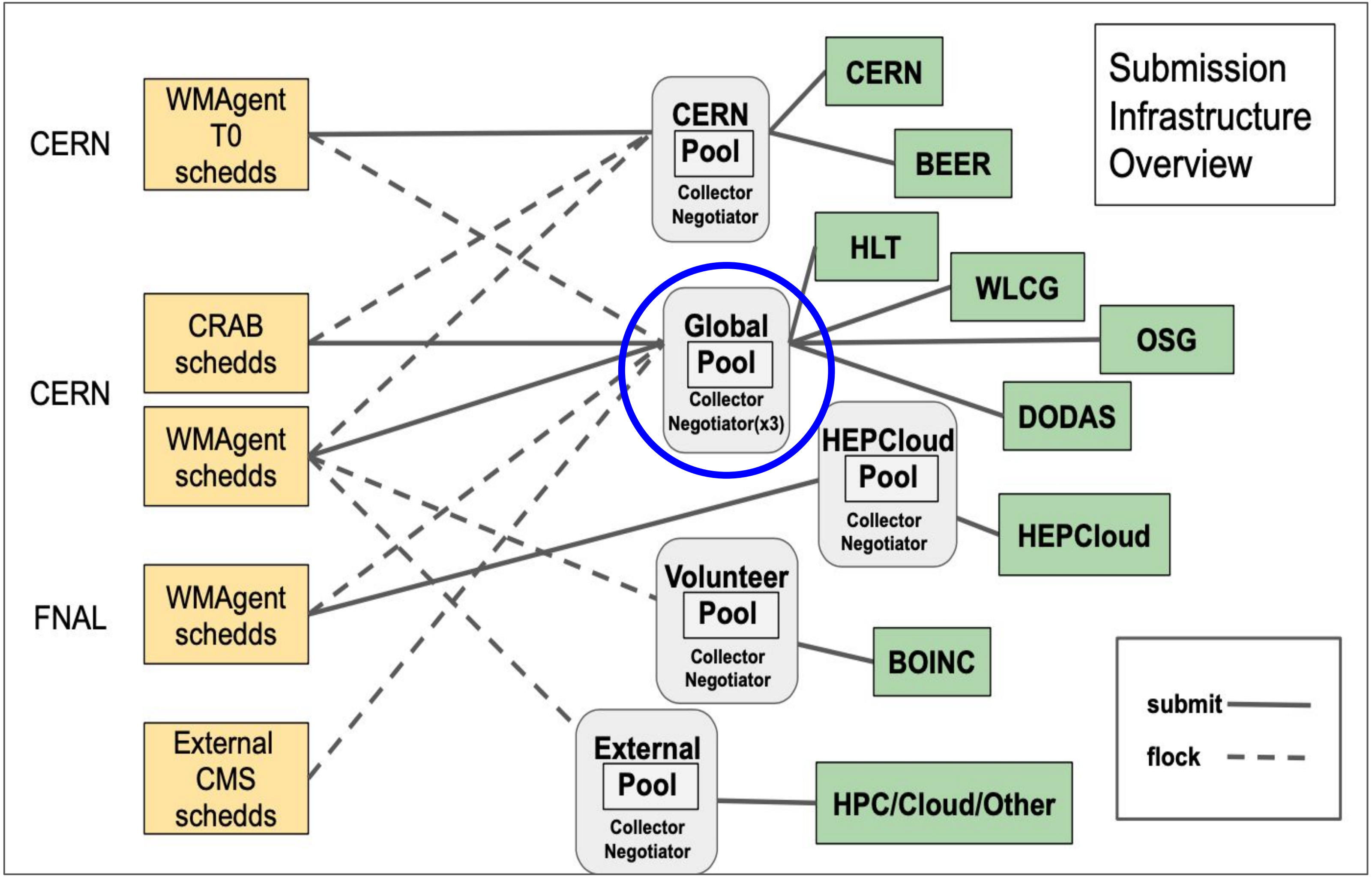
The computing landscape - the WLCG

- Data traditionally analyzed using **Worldwide LHC Computing Grid (WLCG) resources**
  - Global collaboration of around 170 computing centers
  - Access based on dedicated resources (**pledges)**
  - **Nearly 1M CPU cores and 1 EB in data storage**

The CMS Submission Infrastructure Group

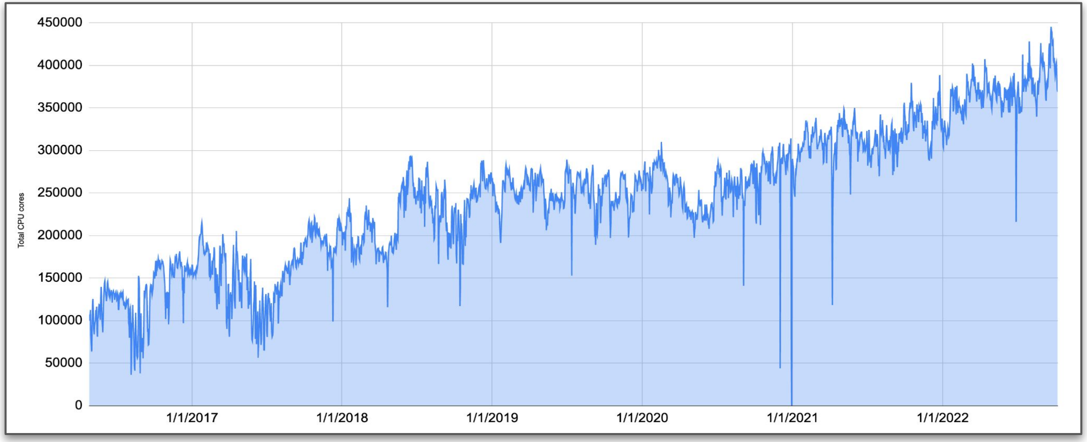
- Part of CMS Offline and Computing in **charge** of:
  - Organizing **HTCSS** and **GlideinWMS** **operations** in CMS, in particular of the **Global Pool**, an infrastructure where reconstruction, simulation, and analysis of physics data takes place
  - Communicate CMS **priorities to the development teams** of **GlideinWMS** and **Condor**
- In practice:
  - We operate a set of federated pool of resources **distributed over 70 Grid sites, plus non-Grid resources**
  - Join them into a Global Pool of resources managed by HTCondor

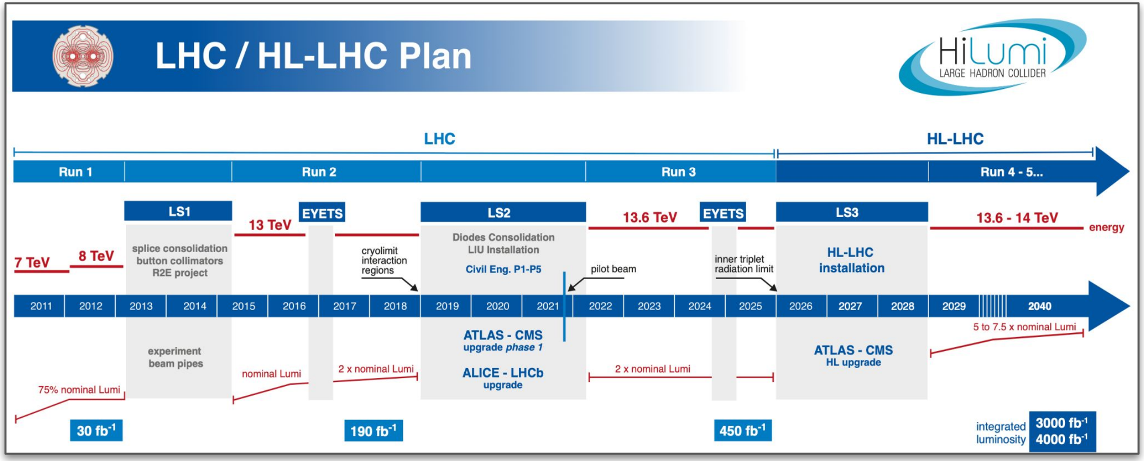



Submission
Infrastructure
Overview
CERN
WMAgent
T0
schedds
CERN
CRAB
schedds
WMAgent
schedds
FNAL
WMAgent
schedds
External
CMS
schedds
CERN
Pool
Collector
Negotiator
Global
Pool
Collector
Negotiator(x3)
HEPCloud
Pool
Collector
Negotiator
Volunteer
Pool
Collector
Negotiator
External
Pool
Collector
Negotiator
CERN
BEER
HLT
WLCG
OSG
DODAS
HEPCloud
BOINC
HPC/Cloud/Other
submit
flock

The scalability challenge (I)

CPU cores allocated to CMS over the past ~6 years (daily averages)



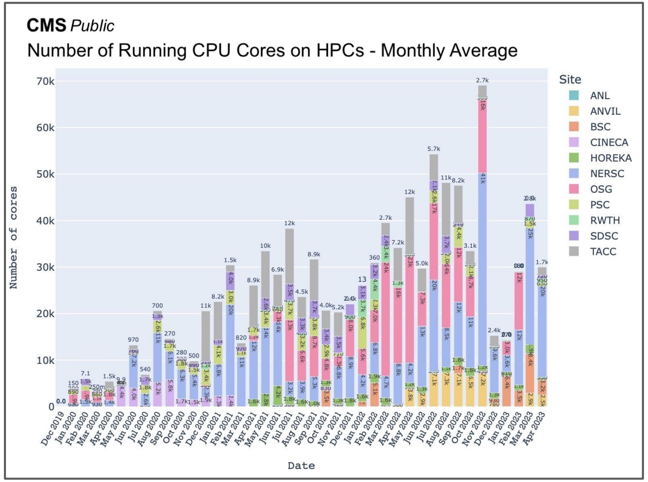


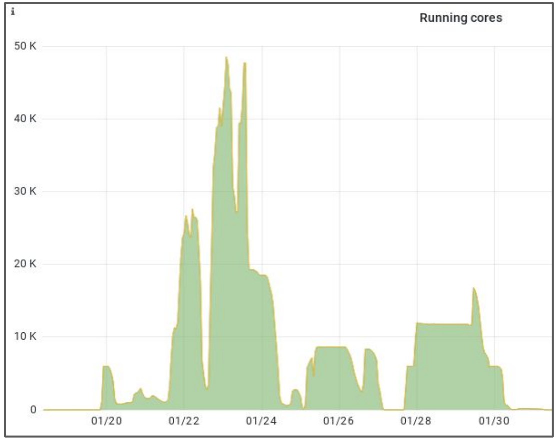
Need of resources expected to grow, especially in the next High Luminosity phase of the LHC program

The scalability challenge (II): HPC component

HPCs can help supporting the computing needs of CMS:

- Substantial investments, so capacity is growing
- HPCs are part of the scientific computing infrastructure, and there to stay
- Several HPC integration efforts in CMS
- Compute capacity used by CMS at HPC facilities could be maintained at the current level, if not higher
- **Scalability:**
  - **Sufficient WM and SI capacity to manage and supply workloads to fill increasingly large pools of resources**
  - **HPC vs WLCG provisioning: WM and SI capacity should be dimensioned to profit from HPC resource peaks, in addition to WLCG resource baseline**
- Stability and flexibility of the WM infrastructure and services
  - To ensure effectively using enlarged compute capacity with HPCs





Global Pool scalability

- Hundreds of thousands of execute points **geographically distributed**
- Tens of CMS **access points** centrally managed (**horizontal scaling**)
- A single central manager for the Global Pool (one collector, 3 negotiators)
  - **Stability of the central manager is key** to CMS success

Central Manager optimizations

- Continuously detecting and solving bottlenecks to our Global Pool, with the support of the HTCondor and glideinWMS developers teams, over the years
  - Multiple **negotiator** daemons, running in **multithreaded** mode
  - Hierarchy of **secondary collectors** connected to the main top collector process
  - **Optimized slot update conditions** (filter on update triggers, use UDP instead of TCP, enlarged UDP buffer)
  - **Classify queries** reaching the collector from the **negotiator as high-prio**, as opposed to those from the GlideinWMS FE and CMSWM and monitoring services
  - **Redirect non-high prio** queries to the slave secondary collector (HA infrastructure at FNAL)

- Most of these actions directed at avoiding the saturation of the **top collector**
  - Essential to build an exhaustive and reliable monitoring infrastructure

Scale tests objectives

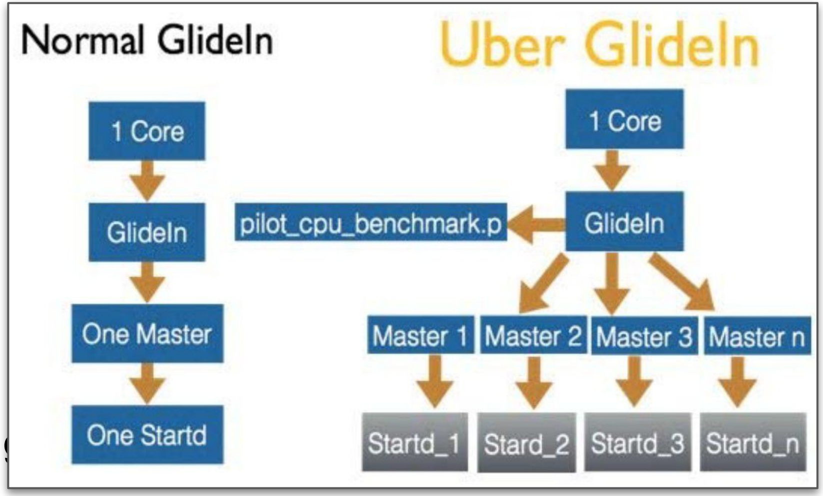
- **Guarantee we operate away from any scalability limiting factor**
  - Critical aspect for a system that is designed to perform in a dynamic environment, adapting itself to growing **resource demands** by CMS, **resource availability** in the WLCG and the **mix of workloads** it has to manage

- **Proactively find those limits**, in every direction, and evolve the infrastructure to push them further away:
  - Total computing power our HTCondor pools can harness and use efficiently
    - Collector capacity to process the stream of **slot updates** and keep resource status fresh
    - Negotiator **matchmaking** cycle time under control
  - Total number of workflows we can manage and jobs we can run simultaneously with our pool of scheduler nodes

The SI 2023 scale tests

Following updates since our latest tests (2021):

- Evolution in HTCondor software
  - Tested version 10.0.1
  - New feature for optimizing slot updates
- A **new physical host** for the central manager
  - AMD EPYC 7302 at 3 GHz
- **Token**-based



As in previous tests

- Über-glideins used to
  of our Global Pool
- Secondary collector service in the pool as a source of monitoring data, to reduce stress from non-essential queries on the main collector

Initial results and mitigations

- Found **first bottleneck** in the total capacity of our **pool of schedds** to supply jobs to the resource pool:
    - At 1 MB of RAM per running job, our 10 schedds with 50 GB of memory saturated at 50k running jobs each
    - Saturation at 0.5M running jobs, comparable to our 2021 result
    - Schedd pool capacity subsequently enlarged to, in principle, be able to support nearly 1M running jobs
- **Second bottleneck** in the HTCondor Connection Brokering (**CCB) service**, hosted in a VM, limiting the maximum number of TCP connections
    - Moved the CCB service to a physical node

Final results

- Pushed the scalability of our Global Pool to about **800k simultaneously running jobs**
  - Factor of 8 away of current pool size considering 4 core jobs

Conclusions and Future work

- The CMS SI team, in close collaboration with the HTCondor and GlideinWMS developers, **periodically evaluates the scalability** of our infrastructure, in order to **anticipate** and **remedy** future scaling and stability problems and stay off the bleeding edge of limitations

- Our 2023 scalability tests showed the **capacity** of our SI to support up to **0.8M simultaneous running jobs**

- The LHC program extends well into the future, so we need to **continue pushing the SI for higher scales**, as required by CMS needs, while maintaining stability and efficiency