Repurposing of the Run 2 CMS High Level Trigger Infrastructure as a Cloud Resource for Offline Computing

Saqib Haleem[1], Marco Mascheroni[2], A. Pérez-Calero Yzquierdo[3], Edita Kizinevic[4], Farrukh Aftab Khan[5], Hyunwoo Kim[5], Nikos Tsipinakis[4], Daniele Spiga[6] and Christoph Wissing[7]
on behalf of the CMS Collaboration

1. National Center for Physics (PK), 2. Univ. of California San Diego (US), 3. CIEMAT and PIC (ES), 4. CERN, 5. Fermi National Accelerator Lab. (US), 6. Universita e INFN, Perugia, 7. DESY

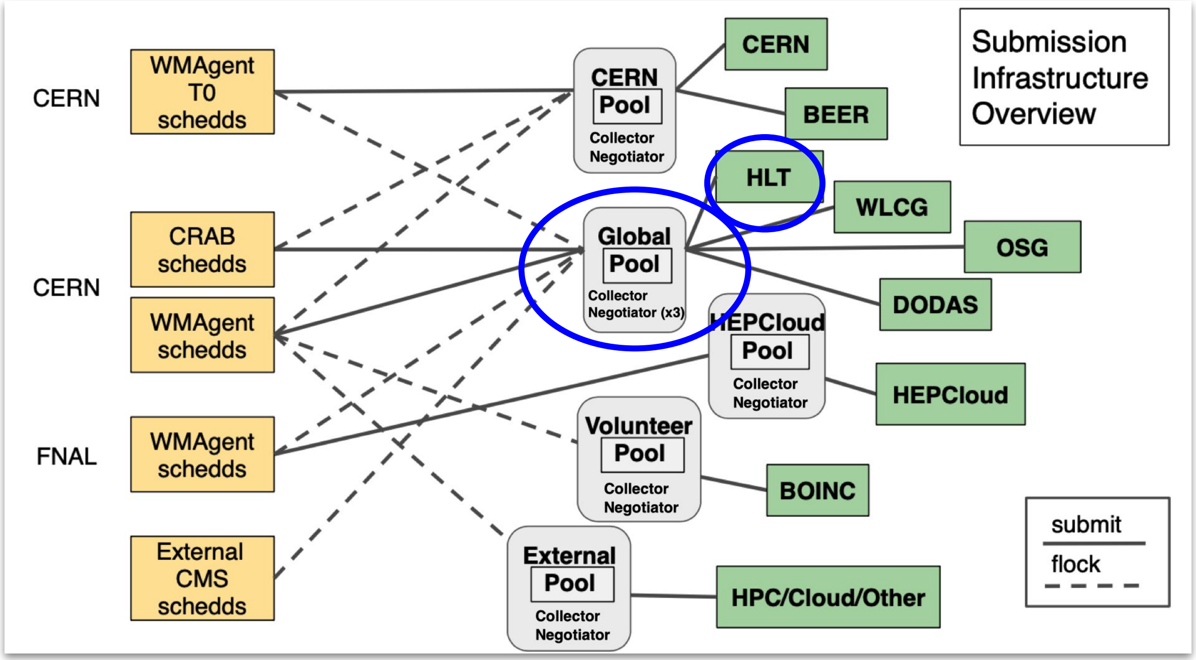
CHEP 2023, May 11th 2023

Outline

- HLT resources for offline computing during Run 2
- Current Hardware Status
  - *Run 2* HLT farm
  - *Run 3* HLT farm
- HLT Integration into CMS Pool (legacy system of static condor service)
- Issues/limitations in VMs with static condor service
- Vacuum model glideins (New approach)
- Implementation and experience
- Prospects of new HLT usage in Interfill mode
- Conclusion

The HLT farm as part of CMS offline computing

- During the Run 2 years, the HLT farm resources at the LHC P5 were configured as execution nodes in the CMS HTCondor Global Pool (details)



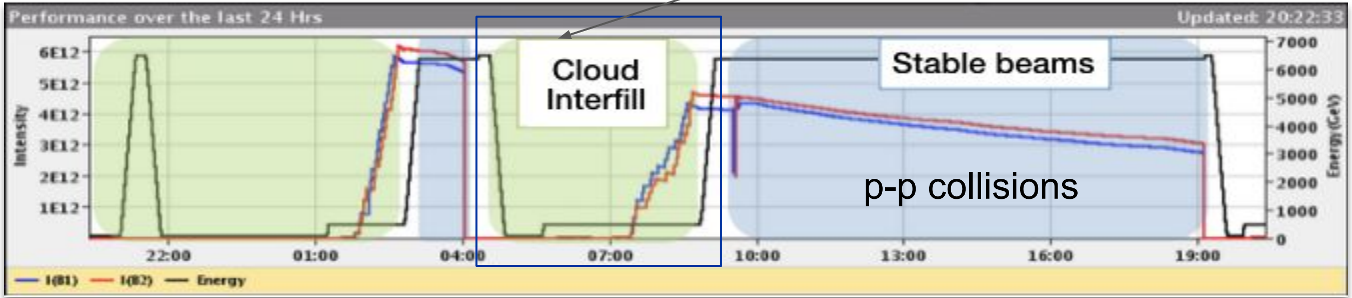
The HLT farm as part of CMS offline computing

- The **CMS Online Cloud** (deployed on the HLT farm resources) was **commissioned during Run 2** to dedicate **HLT resources for offline data processing** when not needed to support data taking (ref)

- Launch **Openstack VMs** running the HTCondor startds when CPU is available (**Interfill + Fill modes**)
  - VMs suspended when HLT needs resources back
  - A **`MaxHibernationTime`** delay (24h) applied at both CCB and WMAgents schedds in order to ensure job execution can be resumed when the VM is back into the Global Pool

Full HLT resources available for Global Pool

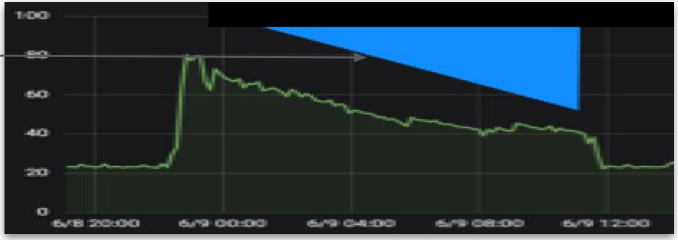
INTERFILL



LHC operation cycles

HLT resources % available for Global Pool

DURING FILLS



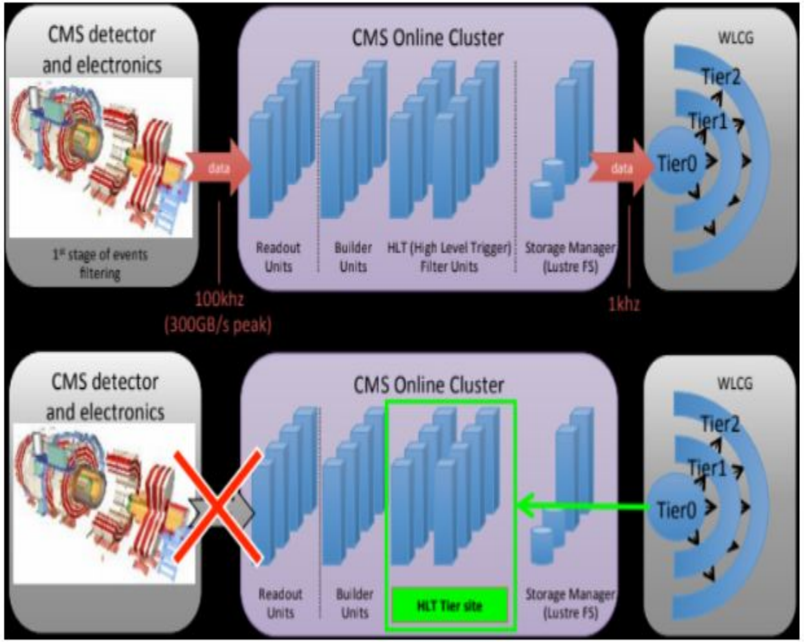
HLT Slots usage in LS2

- During the **LS2**, the Run 2 HLT farm was **intensively used** as opportunistic resource for offline processing.
- See CPU cores in use by CMS MC production and data processing jobs in 2019, 2020 and 2021

HLT Cloud for offline Computing

A **new farm** has been commissioned **for online** usage during the **Run 3**.

Two cloud based **HLT farms** are currently available in CMS:



Run 2 HLT farm (aka Permanent cloud)

- **~25K cores,** fully repurposed for offline computing.
- Renamed to **new** CMS site name **T2_CH_CERN_P5**

Run 3 HLT farm (aka Opportunistic cloud)

- **Opportunistic cloud with ~ 30K cores.**
- Will only be used for offline computing during “no LHC data taking” period
- **T2_CH_CERN_HLT** CMS site name now reserved for new HLT

HLT farm Hardware Status

Permanent Cloud

- Hardware out of warranty.
- Still fits with the current DAQ/HLT power budget for the data centre and space also available. Therefore it can be used until the end of Run 3
- Repair of servers is best effort (resources may shrink)
- Will be decommissioned by the end of Run 3

Opportunistic Cloud

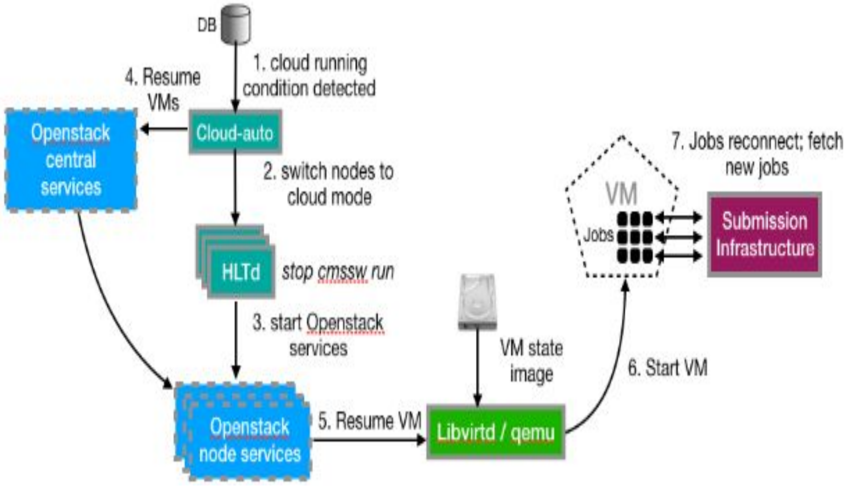
- New HLT nodes from Q1 2022, with default Warranty of 5 year
- Equipped with dual NVidia T4 GPUs
- Will be kept in use for trigger until end of Run-3
- Afterwards, projected to become a permanent cloud for offline

Servers	CPUs	Disk (GB)	RAM (GB)	VMs/ Nodes	Total CPus
Maguway	32	400	64	400	12800
Megware	24	300	56	219	5256
Action	28	400	56	280	7840
Total					~25K
Dell	128	-	256	260	~33K

Ref: https://twiki.cern.ch/twiki/bin/viewauth/CMS/Ocloud

Original configuration of Run 2 HLT farm

- Fast turnaround of resources and jobs are done with “cloud auto” and “HLTd” daemons(shown in diagram)
- The VM images are built with a *static* condor configuration provided by the Submission Infrastructure team
- Those configurations allow long lived (weeks) condor Startd daemon to connect with CMS pool
    a. To be compared with 48h lifetime for regular glideins
- Any change/optimization or tweak needs coordination with SI and requires a new image.



Ref: Andre, Jean-Marc, et al. "Experience with dynamic resource provisioning of the CMS online cluster using a cloud overlay." *EPJ Web of Conferences*. Vol. 214. EDP Sciences, 2019.

Issues/limitation: long lived condor (Startd)

- Configuration synchronization issue with rest of the SI setup.
  - Noticed outdated singularity wrapper script pointing to obsolete and missing missing EL8 images .
- No resource validation prior to connectivity with the pool.
  - Observed high number of job failure due to cvmfs issue.
- Starvation of high priority jobs due to ineffective defragmentation with long lived condor slots.
  - Defragmentation of the partitionable slots relies on pilot job renewal
  - With long-lived slots, no defragmentation from slot rotation, thus high prio 8-core jobs (e.g. T0 PromptReco) remain queued for several days
  - Required the introduction and tuning of induced defragmentation (HTCondor defrag daemon)
    - Not needed in the case of ~48h lifetime pilots
- Finally, SI has no direct control to tune HLT slot parameters
  - Any intervention requires coordination with DAQ team

Solution: Glideins in a Vacuum

- To Overcome the limitations "**Glideins in a Vacuum**" method implemented:
    - Launch glideins directly from the VM, without the need for any compute element or batch system to manage pilot jobs on to WNs

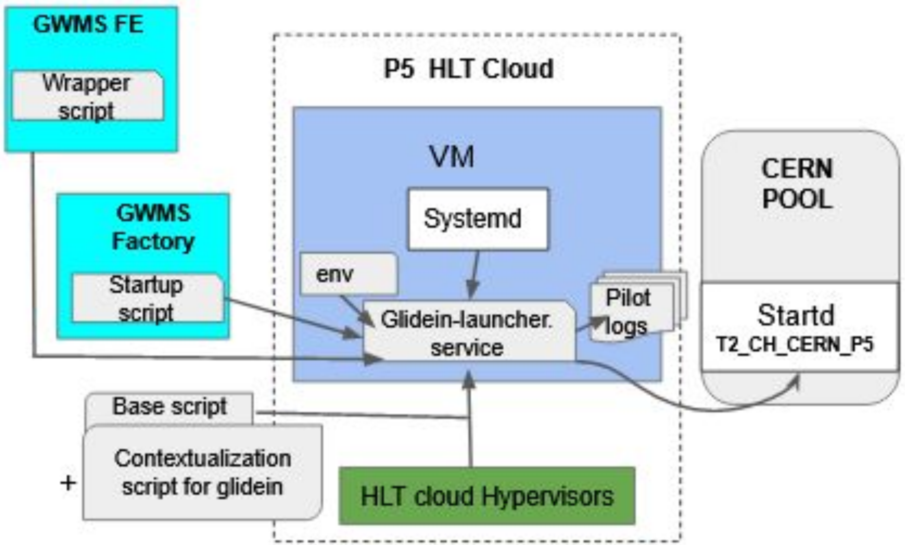
- This idea has already some success history in the WLCG context [*]

- Their configuration is otherwise fully "standard Global Pool" glideins, which results in a number of integrated advantages:
    - Relatively short-lived pilots provide a constant & natural renewal of defragmented slots. Ultimately reduce starvation. (**GLIDEIN_Max_Walltime**)
    - Resource validation before connecting to pool.
    - Reduced operational cost & control:
        - New configurations are automatically applied to pilots upon FE/factory reconfig.

[*] https://www.researchgate.net/publication/263050237_Running_Jobs_in_the_Vacuum

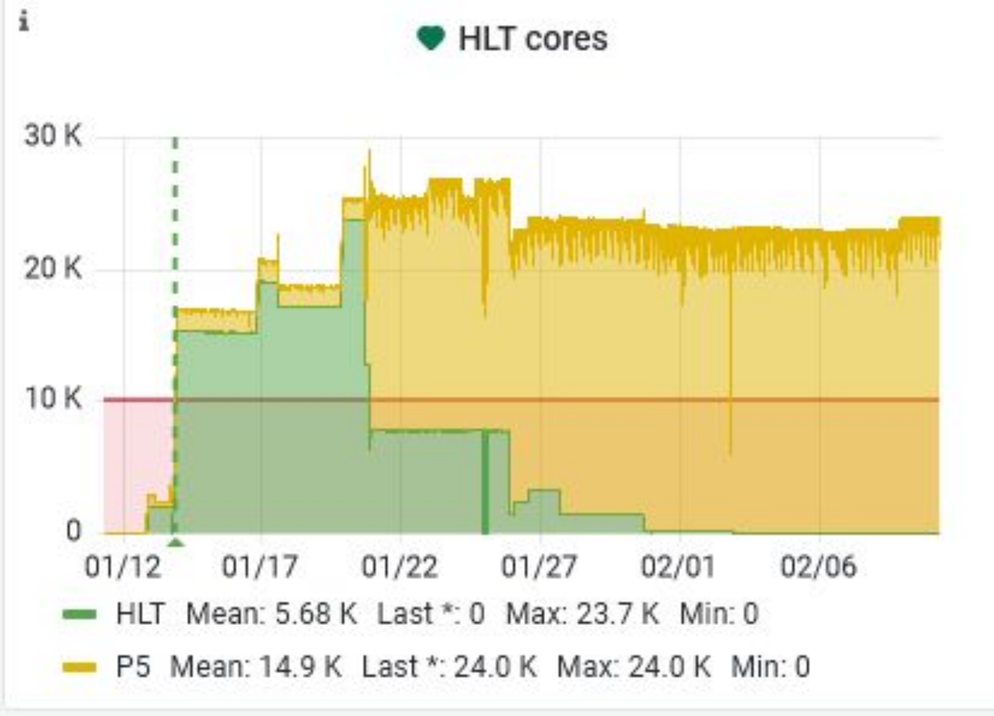
Implementation: Glideins in a Vacuum

- Setting-up new site name/endpoint "**T2_CH_CERN_P5**" for Glideins in vacuum implementation.
- Glidein-launcher.service script is pushed to VM through contextualization script without modifying DAQ provided original VM image.
- Glidein-launcher.service, (systemd based service) pull the wrapper script from GWMS FE whenever it is updated, then launch glideins.
- SI Generates and publish wrapper for (T2_CH_CERN_P5) using Post reconfigure hook of GlideinWMS FE.



Transition of HLT resources to Glideins approach

- Transition of Glideins based site started in January and completely switched to **T2_CH_CERN_P5** in Feb.



What Improved: T2_CH_CERN_P5

- Now configuration/tweaking and optimization of slots can be done directly by the submission infrastructure team & relatively fast (without needing to coordinate with DAQ for draining pushing update/rebuild VM image. Few examples:
  - `GLIDEIN_Max_Idle= 3600  ( 1 Hour)`
  - `GLIDEIN_Max_Walltime=171000 (~48hr)`
  - `GLIDEIN_CPUS= auto ( Whole node pilots)`
  - `MaxHibernateTime = 86400`

  Allow only Production and tier0 jobs (controlled through GWMS FE)
  - ```
    start_expr='
    ifthenelse(DESIRED_Sites isnt undefined,
    stringListMember(GLIDEIN_CMSSite,DESIRED_Sites), undefined)....) &&
    (GLIDEIN_CMSSite =!= "T2_CH_CERN_P5" || WMAgent_AgentName =!= UNDEFINED)'
    ```

- Validation (prerequisite) check, before starting pilot. Reduced chances of job failures.

- Low Idle time observed for high prio jobs.
    ```

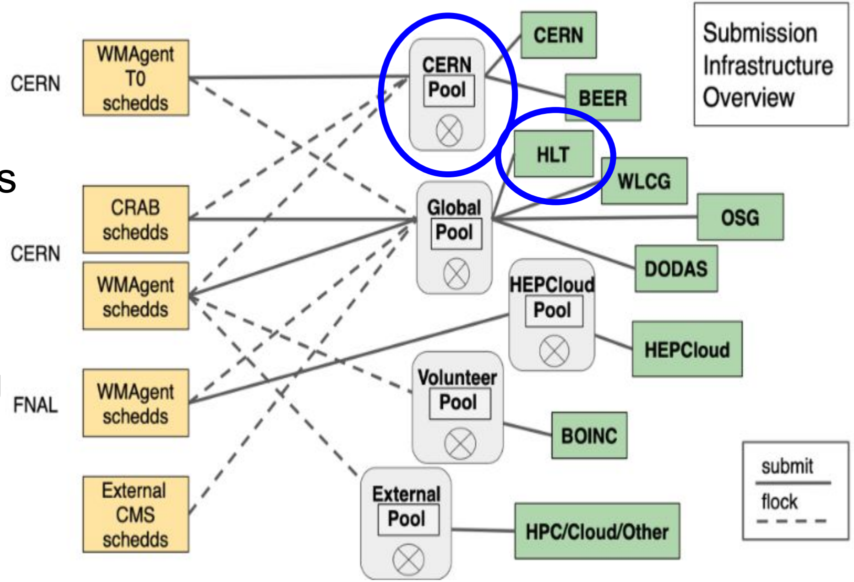
T2_CH_CERN_P5 moved to CERN Pool

_P5 was moved from the Global to the CERN pool, for the following advantages:

- Similar to T2_CH_CERN in shares and priorities between Tier 0, production and analysis jobs, (except we don't allow analysis jobs in P5)

- Now it is isolated from potential issues affecting the global pool (**increased resilience** for the resources mainly responsible for running T0 tasks)

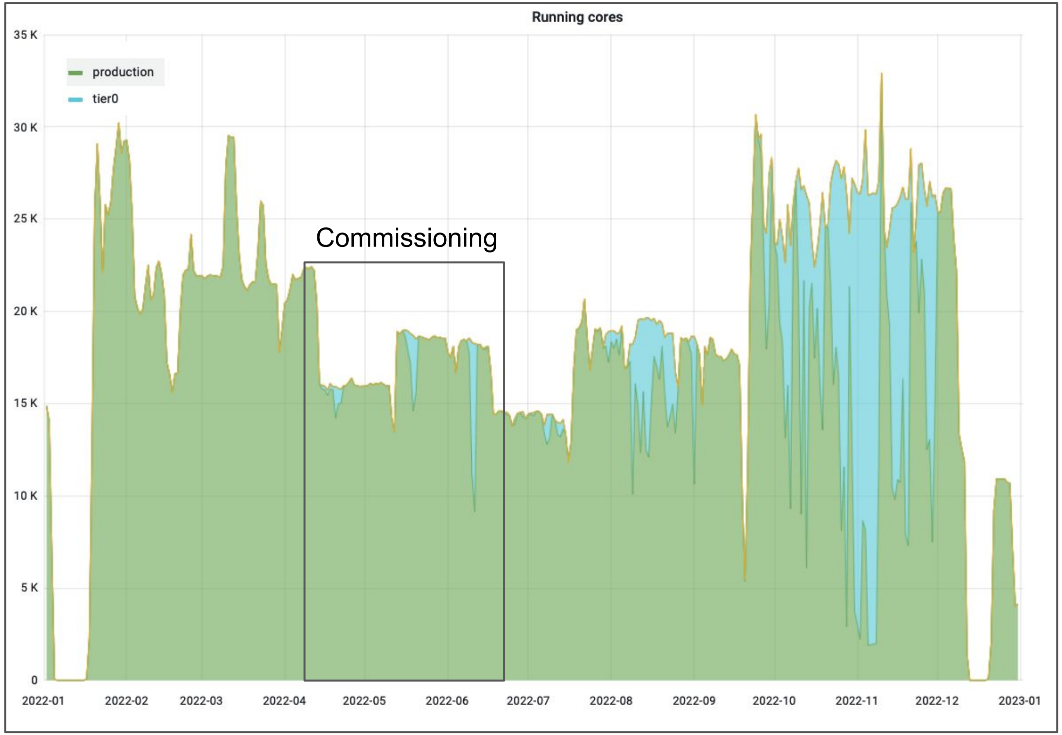
- **Reduce** the increased slot **pressure** on the **global pool collector** and the global pool T2 negotiator



CMS SI federated pools

Tier 0 jobs in the HLT farm in Run 3

The Run 2 HLT farm continued to be used opportunistically in the first year of the LHC Run 3, this time adding support for the Tier 0 in prompt reconstruction tasks



- Commissioning phase in Spring 2022 to ensure T0 tasks can reach old HLT farm and acquire slots at sufficient rate
  - Partitionable slots in long-lived VMs tend to get fragmented by low-core-count production jobs
  - In order to execute PromptReco, 8-core slots are required.
  - Induced defragmentation of the HLT slots (HTCondor defrag daemon) was configured and optimized

- Once stable beams were declared, the resource was regularly used as extension of the Tier 0 capacity, as demanded by the data taking rates

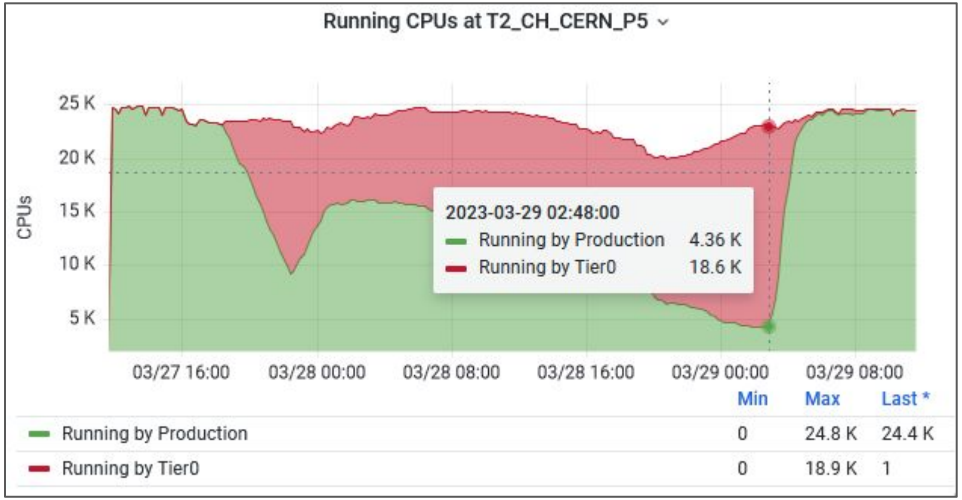
Tier 0 Jobs on T2_CH_CERN_P5

- Run 2 HLT included in the pool of resources for T0 prompt reconstruction job execution in 2022

- Commissioning of the new site name and glidein mechanism with T0 jobs has been successfully tested in preparation to the 2023 LHC data taking season.

- Lately large scale T0 test performed.
  - 1 core jobs: ~20k
  - 8 core jobs: ~57k (456k Cores)

- CPU cores for tier0 (high priority) jobs quickly ramped up. (as shown in plot)



Run 3 HLT farm usage in CMS Offline

- Opportunistic use of Run 3 HLT farm will be during:
  - LHC technical stop,
  - Interfill mode (small intervals (i.e. few hours) when no physics data is generated by LHC)

- All the ingredients for the interfill mode to allow usage for offline computing should already exists, i.e.
  - VM suspension and disk state saving feature (DAQ side)
  - Job suspension and resumption feature for interfill mode use (SI side)

- A few nodes from the Run 3 HLT farm are available for their testing and integration into the CMS offline resource pool

Conclusions

- HLT resources integral part of CMS offline computing
  - As many cores as a **big T1**s
- Identified advantages motivating the transition of the P5 resources for offline computing to a **glideins driven acquisition and management** approach.
- Implemented in full **since February**, working ok as expected
  - Already benefited from the foreseen **advantages** that motivated the transition, e.g. auto reconfiguration of pilots.
- The same pilot-based approach **will be tested in the new (Run 3) HLT** farm, configuring glideins with necessary parameters to account for intermittent nature of the resource
  - Keep using the already reserved sitename “T2_CH_CERN_HLT”

Acknowledgements

Projects FPA2016-80994-C2-1-R, PID2019-110942RB-C21, BES-2017-082665 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

Ciemat Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas

cfp CIEMAT física de partículas

PIC port d'informació científica

US National Science Foundation under Grant No. 2121686

UC San Diego

Backup slides

Abstract

The former CMS Run 2 High Level Trigger (HLT) farm is one of the largest contributors to CMS compute resources, providing about 30k job slots for offline computing. The role of this farm has been evolving, from an opportunistic resource exploited during inter-fill periods in the LHC Run 2, to a nearly transparent extension of the CMS capacity at CERN during LS2 and into the LHC Run 3 started in 2022. This “permanent cloud” is located on-site at the LHC interaction point 5, where the CMS detector is installed. As a critical example, the execution of Tier 0 tasks, such as prompt detector data reconstruction, has been fully commissioned. This resource can therefore be used in combination with the dedicated Tier 0 capacity at CERN, in order to process and absorb peaks in the stream of data coming from the CMS detector, as well as contributing to the prompt reconstruction of a substantial fraction of the “parked data sample”, dedicated primarily to B physics studies. The initial deployment model for this resource, based on long-lived statically configured VMs, including HTCondor execution node services connected to the CMS Submission Infrastructure (SI), provided the required level of functionality to enable its exploitation for offline computing. However, this configuration presented certain limitations in its flexibility of use in comparison to pilot-based resource acquisition at the WLCG sites. For example, slot defragmentation techniques were required to enable matching of Tier 0 multicore jobs. Additionally, the configuration of fair-share quotas and priorities for the diverse CMS tasks could not be directly managed by the CMS SI team, in charge of enforcing the global CMS resource provisioning and exploitation policies. A new configuration of this permanent cloud has been proposed in order to solve these shortcomings. A vacuum-like model, based on GlideinWMS pilot jobs joining the CMS CERN HTCondor Pool has been prototyped and successfully tested and deployed. This contribution will describe this redeployment work on the permanent cloud for an enhanced support to CMS offline computing, comparing the former and new models’ respective functionalities, along with the commissioning effort for the new setup.

T2_CH_CERN_P5 vs T2_CH_CERN: are they equivalent from ComOps perspective? Issues from WM or PnR?

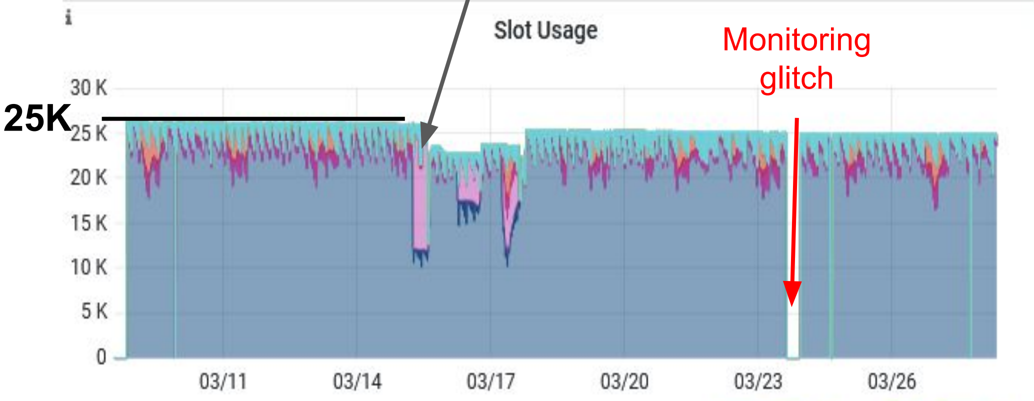
- From WMAgent Point of view both sites are equivalent. If a job is assigned to T2_CH_CERN only , it can go to T2_CH_CERN_P5.

- However, from PnR perspective both sites are not equivalent. T2_CH_CERN_P5 is a special site, and it might not be able to handle every job. e.g jobs with remote access needs.

T2_CH_CERN_P5 Usage (Last 20 days)

- **~25K cores are running with maximum core usage** & without any reported issue.
- Data centre interventions or VM image upgradation requires proper draining of jobs/slots before shutdown
  - previously done with `condor_off -peaceful`
  - Now glidein is drained using standard approach common to all sites i.e. by setting `SiteWMS_WN_Draining as True`.using LCG Machine/Job feature.[1]

Glideins put to drain to perform VM image upgradation



	min	max	avg	current
Busy	0	25.8 K	22.0 K	23.5 K
Glidein Draining	0	13.9 K	470	0
Idle	0	6.24 K	1.49 K	536

[1] https://twiki.cern.ch/twiki/bin/view/LCG/MachineJobFeatures

Job hibernation Implementation for interfill mode

- Custom parameter "`MaxHibernateTime`" added to the VMs. This determines the maximum VM hibernation time a job should wait for it to come back up.

- Submit nodes are configured to publish this in Job's Ad.

```
SYSTEM_JOB_MACHINE_ATTRS =
$(SYSTEM_JOB_MACHINE_ATTRS) MaxHibernateTime
```

- On the WMAgent side, "JobLeaseDuration" has been set to be an expression.

```
 isUndefined(MachineAttrMaxHibernateTime0) ?
1200 : MachineAttrMaxHibernateTime0
```

- Dedicated range of secondary collectors/CCB with longer sweep interval and security session.

```
COLLECTOR973x_ENVIRONMENT =
"_CONDOR_CCB_SWEEP_INTERVAL=86400
_CONDOR_SEC_DEFAULT_SESSION_LEASE=86400"
```