Running GPU-enabled CMSSW workflows through the production system

CHEP2023: 26th International Conference on Computing in High Energy & Nuclear Physics

Charis-Kleio Koraka

on behalf of the CMS collaboration

May 8th - 12th 2023
Overview

- Motivation
- Offline CMS workflows
- CMS production system
- Integrating GPUs
Motivation

- At High Luminosity LHC (HL-LHC):
  - Increase in instantaneous luminosity and pile-up by more than a factor of 2
  - Upgraded detectors with higher granularity
- Will pose significant computing challenge
Motivation

Alternative approaches in computing model needed to mitigate risk:

- GPUs and heterogeneous computing
  - Cope with the higher throughput
  - Keep energy consumption low
  - Allow to utilize High Performance Computing (HPC) to address scientific challenges

Might need a factor of up to 20 increase in computing resources to keep similar physics reach.
Many Tier-1 and Tier-2 sites are already equipped with GPUs.

Most are opportunistic but a few are dedicated to CMS.
GPUs at the High Level Trigger

CMS has leveraged GPUs for the online reconstruction at High Level Trigger (HLT) starting from the beginning of Run-3 (2022-today)

What has been offloaded to GPU:

- ~25% of online reconstruction:
  - Pixel track reconstruction
  - ECAL & HCAL local reconstruction

*Source CERN-CMS-DP-2023-004*
GPUs at the High Level Trigger

Outcome of offloading:

- Higher throughput & equivalent results to CPU
- Equal or better physics performance

*Source CERN-CMS-DP-2023-004

Talk by Felice in WLCG workshop
Talk by Ganesh later today
**Typical CMS offline workflow chains**

- **Generation**
  - GEN-SIM (Simulation)

- **Digitization**
  - DIGI

- **Reconstruction**
  - RECO

- **Reduced data formats**
  - miniAOD/ nanoAOD

**CMS experiment HLT farm**

**Different types of workflow chains:**
- Monte Carlo (McM)
- Re-reconstruction (ReReco)
- Release Validation (RelVal)
Typical CMS offline workflow chains

- Generation
- Simulation
- Digitization
- Reconstruction
- Reduced data formats

GEN-SIM → DIGI → RECO

CMS experiment HLT farm

- Reconstruction uses the CMS software framework (CMSSW)
- Part of the offline reconstruction can be offloaded to GPUs inheriting from what is done online@HLT:
  - ~10% of offline reconstruction (not yet utilized in production)
  - Ongoing R&D to offload more offline code to GPUs in the near future
Workflow management

Submitters
- McM
- ReReco
- RelVal

Global Workqueue

Pool of WMAgents
- WMAgent
  - HTCondor
- WMAgent
  - HTCondor

Worker nodes

GlideinWMS/Global Pool

External flow

WMAgent flow

WMCore flow

Unified
Workflow management

- **Submitters**
  - McM
  - RelVal

Workflow is submitted and directed to the request manager

- **ReqMgr2**

- **Global Workqueue**

- **Pool of WMAgents**
  - WMAgent
    - HTCondor
  - WMAgent
    - HTCondor

- **GlideinWMS/Global Pool**

- **Worker nodes**

- **External flow**
- **WMAgent flow**
- **WMCore flow**
Workflow management

The request manager:
- Creates the request
- Validates request parameters.
- Places the requests into the global WorkQueue.

Submitters:
- McM
- ReReco
- RelVal

Pool of WMAgents

GlideinWMS/Global Pool

Worker nodes

External flow
WMAgent flow
WMCore flow
GPU Workflow specifics

**Specific parameter requests** for workflows that require GPUs
- **GPU site Whitelisting**
  - Provide a list of sites that are equipped with GPUs
- **Introduction of GPU parameters**
  - RequestGPUs, RequiresGPU, GPUMemoryMB, CUDACapability, CUDARuntime
- **TaskChain workflow**
  - Only the task (reconstruction step) that runs on GPUs will request/run on these resources
Example of GPU-enabled RelVal workflow request

Tasks 1 & 2 (GEN-SIM & DIGI) do not require GPUs

Tasks 3 (RECO) requires GPUs. The GPU parameters are also specified.

Example:

```python
'RequestType': 'TaskChain',
'SubRequestType': 'RelVal',
'Task1': {
    ...
    'RequiresGPU': None,
    'TaskName': 'ZMM_14TeV_TuneCP5_2021_GenSim',
},
'Task2': {
    ...
    'GPUParams': None,
    'RequiresGPU': 'required',
    'TaskName': 'Digi_2021',
},
'Task3': {
    ...
    'GPUParams': {'CUDACapabilities': [7.5],
                  'CUDADriverVersion': '',
                  'CUDARuntime': '11.2',
                  'CUDARuntimeVersion': '',
                  'GPUMemory': '8000',
                  'GPUName': ''},
    'RequiresGPU': 'required',
    'TaskName': 'Reco_Patatrack_PixelOnlyGPU_2021',
    'TaskChain': 3,
    'TimePerEvent': 10
}
```
GPU Workflow specifics

- Workqueue elements are split into jobs and sent to the condor pool
- Workflow level GPU configuration gets reflected in the job classad description
- HTCondor matchmaking where resources are attached to jobs

Submitters
- McM
- ReReco
- RelVal

Global Workqueue

ReqMgr2

Unified

External flow
WMAgent flow
WMCore flow

Worker nodes
Usage of GPUs in production

- Example of the usage of the GPU cluster at the Wisconsin Tier-2 site

Large scale validation of CMS reconstruction code before it was deployed and used for data-taking

Release Validation workflows to validate updates in GPU reconstruction code

Test GPU enabled offline workflows are also being submitted and tested at HPC centers
Usage of GPUs in production

Example of output from one of the latest GPU enabled RelVal workflows

- Comparison of different quantities when reconstructed on GPU and CPU
Summary

- High Luminosity LHC will pose a significant computing challenge for the CMS experiment
  - CMS is exploring the usage of heterogeneous hardware, among other things, to address this challenge.
- Part of the online reconstruction that runs at the High Level Trigger has already been ported to GPU
  - Improved physics performance & throughput during data taking
  - Gain experience towards a wider use of GPUs for offline reconstruction
- The infrastructure to run GPU workflows offline is already in place and has been in use since mid-2022 and successfully used to commission the CMS HLT for Run 3
Summary

- High Luminosity LHC will pose a significant computing challenge for the CMS experiment
  - CMS is exploring the usage of heterogeneous hardware, among other things, to address this challenge.
- Part of the online reconstruction that runs at the High Level Trigger has already been ported to GPU
  - Improved physics performance & throughput during data taking
  - Gain experience towards a wider use of GPUs for offline reconstruction
- The infrastructure to run GPU workflows offline is already in place and has been in use since mid-2022 and successfully used to commission the CMS HLT for Run 3

Thank you!

Questions?
BACK-UP