GPU resources at CC-IN2P3

And how to use them

Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

Overview

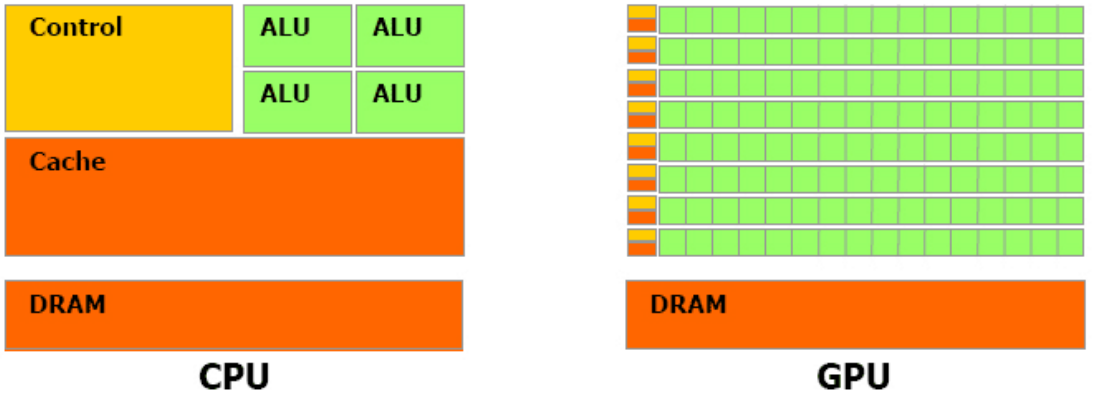
I. Why GPUs?

II. Hardware

III. Software setup

IV. Submitting jobs

Why GPUs?



- Our users have use cases for Single Instruction Multiple Data (signal/image processing, particles propagation, ...)
- GPGPU is one of the competing solutions
- Already adopted by some users
- Documentation and trainings widely available

Why GPUs?

- Something new to most users
- Different requirements for different users
- A few GPUs already installed, thanks to our partnership with Dell
- NVIDIA hardware was necessary to run CUDA

➡ GPU cluster now running since last september

- Servers:
  - 10 Dell C4130

    - 2 Xeon E5-2640v3 @2.6 Ghz (8 cores)
    - 128 GB RAM
    - SSD
    - 1 Gb/s NIC, 10 Gb/s projected
    - 2 Nvidia Tesla K80

      - ➡ 4 GPUs Nvidia GK210 per node, 12 GB DDR5 each
  - Cluster totaling 40 GPUs
- Interconnect:
  - InfiniBand QDR

Software setup

- CentOS 7
  - low latency optimized
- CUDA 7.5
- OpenCL 1.2
- Univa Grid Engine

Submitting jobs

- Batch workers:
  - “qsub” to submit jobs
- Interactive workers:
  - “qlogin” to get a shell on a GPU worker
    - To launch short interactive jobs, tests, compile programs, ...
- Syntax:
  - -l os=cl7,GPU=*x*
  - -q *queue_name*
  - -pe multicores *y*

    https://doc.cc.in2p3.fr/en:ge_submit_a_job_qsub#gpu_jobs
- MPI on multi-nodes soon available

GPU production cluster available

Designed to fit user needs, still evolving to achieve it

Feedbacks welcome!