

Integration of GPUs in WLCG

J. Flix

GDB
13/April/2022



Ciemat Centro de Investigaciones
Energéticas, Medioambientales
y Tecnológicas



Integration of GPUs in WLCG

The WLCG MB asked the GDB to bring back information from the LHC experiments on the integration of GPUs in the WLCG computing infrastructure. Feedback to be used for April 2022 C-RRB.

A twiki was made to collect information from the experiments:

<https://twiki.cern.ch/twiki/bin/view/LCG/GPUsPlans>

QUESTIONS

1. Which is the current status of integration of GPUs for **offline** computing, at the software level?
2. Are you already using GPUs for **offline** computing activities? Please, comment on the experience.
3. If using these resources for offline, do you *account* for the GPU usage in some way?
4. Are there any future plans on GPU resource demands or plans for future utilization at sites?
Please, indicate mid-term and long-term plan (if available).
5. How many GPUs are needed for offline for the next two years?
6. Any other plans, e.g. on FPGAs?
7. Other comments or questions?

(main) conclusions derived from answers

ALICE

Runs offline software on GPUs at O2, a common framework for online and offline computing. The software is compatible with a variety of GPU models. ALICE plans to use the GPUs extensively for the offline processing from Run3 on. After first year of data taking, it's on the plan to use some GPU resources off-O2

ATLAS

R&D projects are ongoing investigating GPUs. GPUs used for some analysis (ML training), resources available at some sites. Large-scale GPU deployment at sites would be needed by the start of HL-LHC (~2028), assuming the offline software has been developed to use them

(main) conclusions derived from answers

CMS

Support from CMSSW to run on GPUs, both at the HLT and Grid sites, in close collaboration with the HTCondor team. GPUs are used opportunistically for some machine learning studies, and user jobs. CMS plans to propagate to Phase-2 and Run 3 offline workflows the algorithms currently used at the HLT on GPUs and to add more during Run 3

LHCb

Plans to use the HLT to run on GPUs to emulate the HLT1 in simulation. Software is ready for this. A few LHCb analyses that use GPUs for e.g. training of Machine-Learning estimators or maximum-likelihood fits, use resources on institute clusters which are outside our distributed computing infrastructure. No plans for GPU resource demands at sites yet.

(main) conclusions derived from answers

- GPU usage marginal (user analysis). No short-term plans to include GPUs from sites (in general)
- FPGAs only for online, no plans for offline yet
- No benchmarks for GPU atm, which affect the accounting
- All of these resources treated as opportunistic, for the moment, waiting for guidance from WLCG, though the usage is not yet at scale
- The survey will be conducted again in 1 year, to know what's changed, prior the Spring's C-RRB