

GPU-enabled LCG Stack

Pere Mato

LIM Meeting, 11 December 2018

Motivation

- ❖ Make available a full LCG software stack ready for GPU (Cuda)
 - ❖ People want to profit from the speedup offered by GPUs for ML applications
- ❖ IT has procured four GPU servers each containing four Nvidia P100 GPUs
 - ❖ They should be operational in Q1 2019
- ❖ For the time being targeting two packages specifically
 - ❖ ROOT (TMVA)
 - ❖ TensorFlow

How is it implemented?

- ❖ Fine tuning versions is critical
 - ❖ E.g. `tensorflow-gpu` version 1.12.0 only works with Cuda 9.0
- ❖ New **full stack label** and not a new **platform**
 - ❖ E.g. `dev3cuda9` (equivalent to `dev3python3`)
 - ❖ Later we will have `LCG_95cuda9` or similar
 - ❖ 99% of the packages are shared between **dev3** and **dev3cuda9**
- ❖ The stack includes the Cuda development toolkit
 - ❖ To run on a target system will only require the NVIDIA driver to be installed (kernel module)
- ❖ Using the binaries for Cuda, tensorflow, tensorboard
 - ❖ This simplifies installation but add some limitations (e.g. SLC6)

How is it used?

```
source /cvmfs/sft.cern.ch/lcg/views/dev3cuda9/latest/x86_64-centos7-gcc62-opt/setup.sh
```

- ❖ Known limitations
 - ❖ One single platform for the time being: x86_64-centos7-gcc62-opt
- ❖ Feedback is most welcome
 - ❖ Try it out in your preferred GPU equipped server
- ❖ Later will be used in SWAN when GPU nodes will be added to the service