



The CNRS logo, consisting of the letters 'cnrs' in a white, lowercase, sans-serif font inside a dark blue circle.

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

# GPU Status at CC-IN2P3

April 2021

- ▶ Hardware
- ▶ Software
- ▶ Farm & Neighbourhood

## K80



- ▶ **10 workers:**
  - 2 Intel(R) Xeon(R) CPU E5-2640 (8 cores)
  - 128GB RAM
  - SSD disk
  - 2 Nvidia Tesla K80 cards  
(4 GPU Nvidia GK210, 12 Go DDR5 each)
- ▶ **40 GPU (total)**
- ▶ **Network**
  - Infiniband interconnection
- ▶ **OUT OF WARRANTY IN JUNE**

## V100



- ▶ **12 workers:**
  - 2 Intel(R) Xeon(R) Silver 4114 (10 cores)
  - 2 Intel(R) Xeon(R) Silver 4214R (12 cores)
  - 192GB RAM
  - SSD M2 disk
  - 4 Nvidia **Tesla V100** 32GB PCIe cards
- ▶ **48 GPU (total)**
- ▶ **Network**
  - NO Infiniband interconnection!

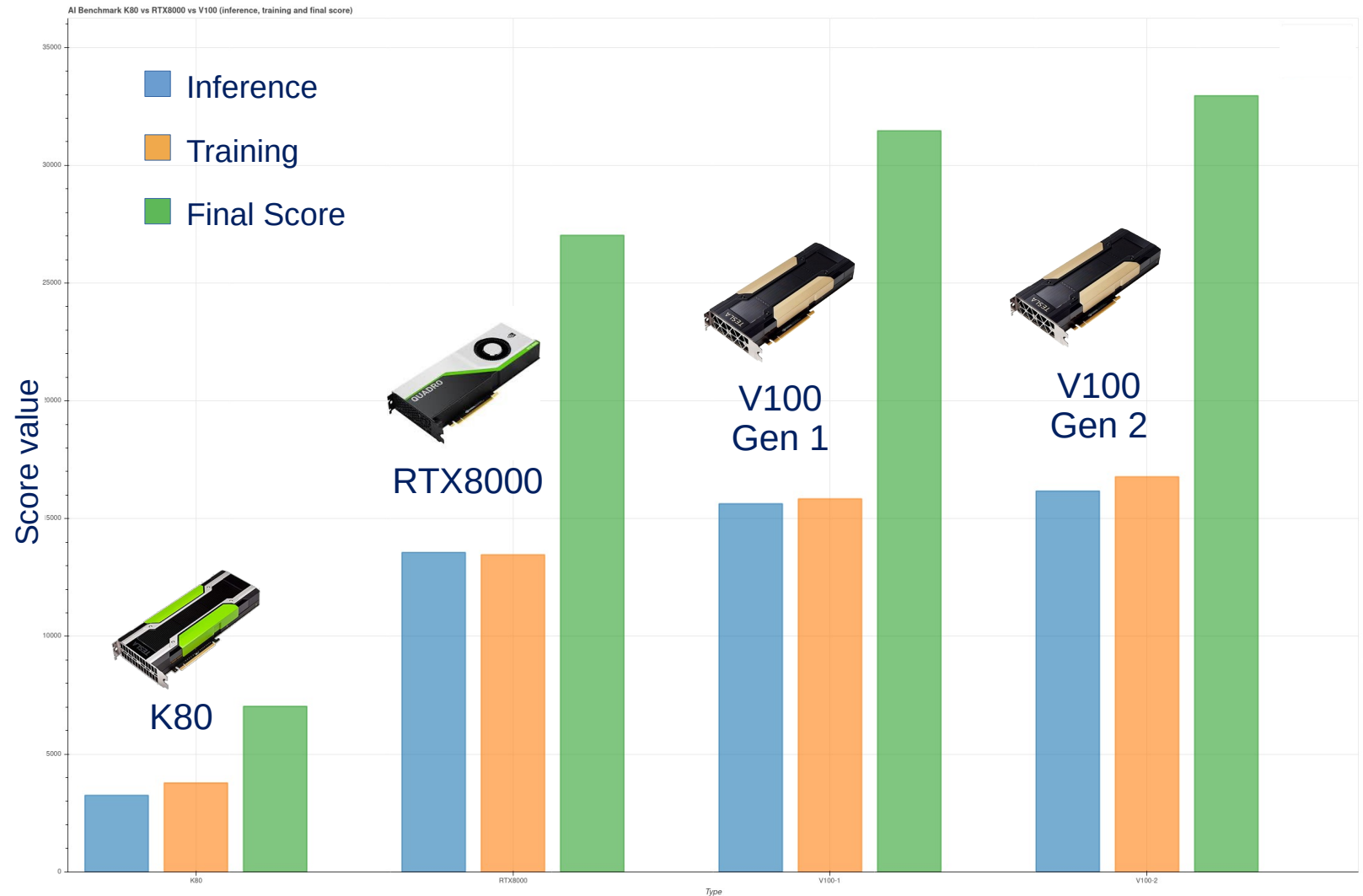
# K80 vs V100

▶ AI Benchmark used at CC (taken from <https://pypi.org/project/ai-benchmark/#modal-close>)

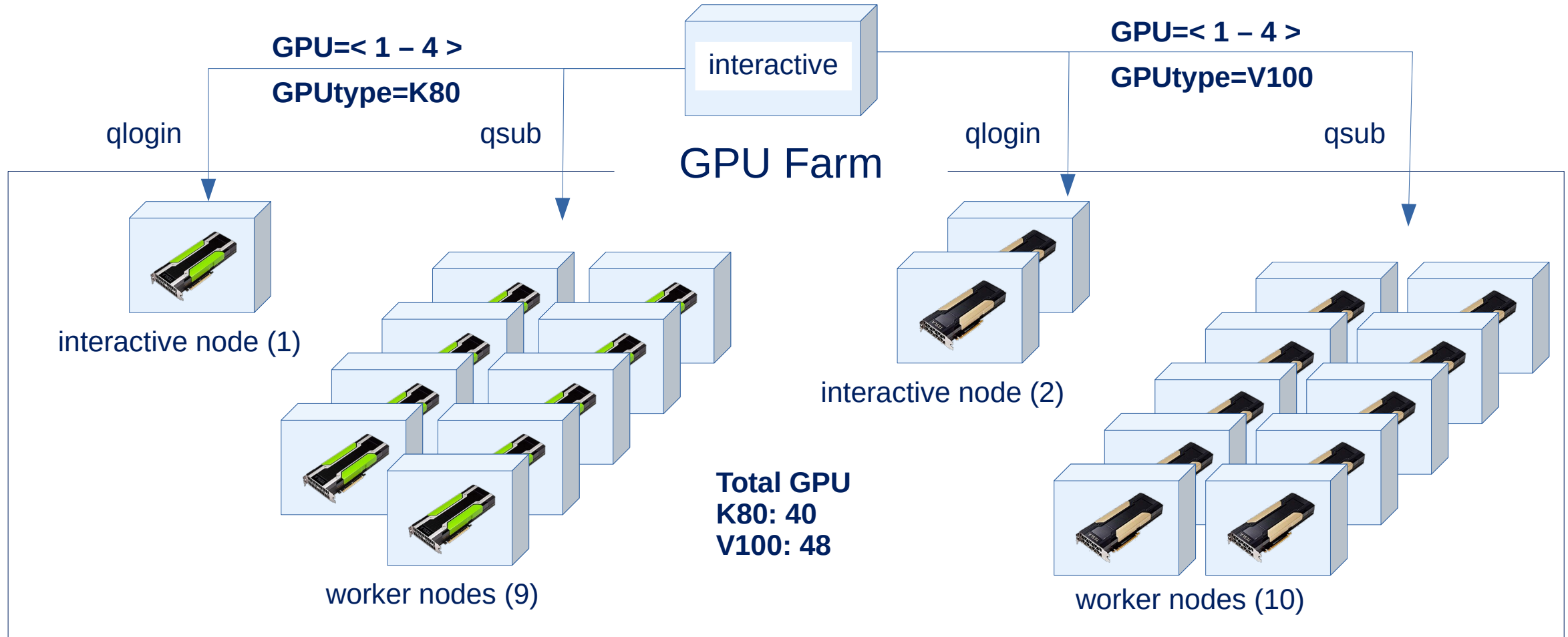
Using **Tensorflow**

**42 tests** on Deep Learning models:

- MobileNet-V2 [classification]
- Inception-V3 [classification]
- Inception-V4 [classification]
- Inception-ResNet-V2 [classification]
- ResNet-V2-50 [classification]
- ResNet-V2-152 [classification]
- VGG-16 [classification]
- SRCNN 9-5-5 [image-to-image mapping]
- VGG-19 [image-to-image mapping]
- ResNet-SRGAN [image-to-image mapping]
- ResNet-DPED [image-to-image mapping]
- U-Net [image-to-image mapping]
- Nvidia-SPADE [image-to-image mapping]
- ICNet [image segmentation]
- PSPNet [image segmentation]
- DeepLab [image segmentation]
- Pixel-RNN [inpainting]
- LSTM [sentence sentiment analysis]
- GNMT [text translation]



- ▶ First, request an account (authorization required)
- ▶ Classical submission on Grid Engine (<https://doc.cc.in2p3.fr/fr/Computing/job-types/job-gpu.html>)



## Available libraries



OpenCL



OPEN MPI

## Customized software



Provided by Singularity containers

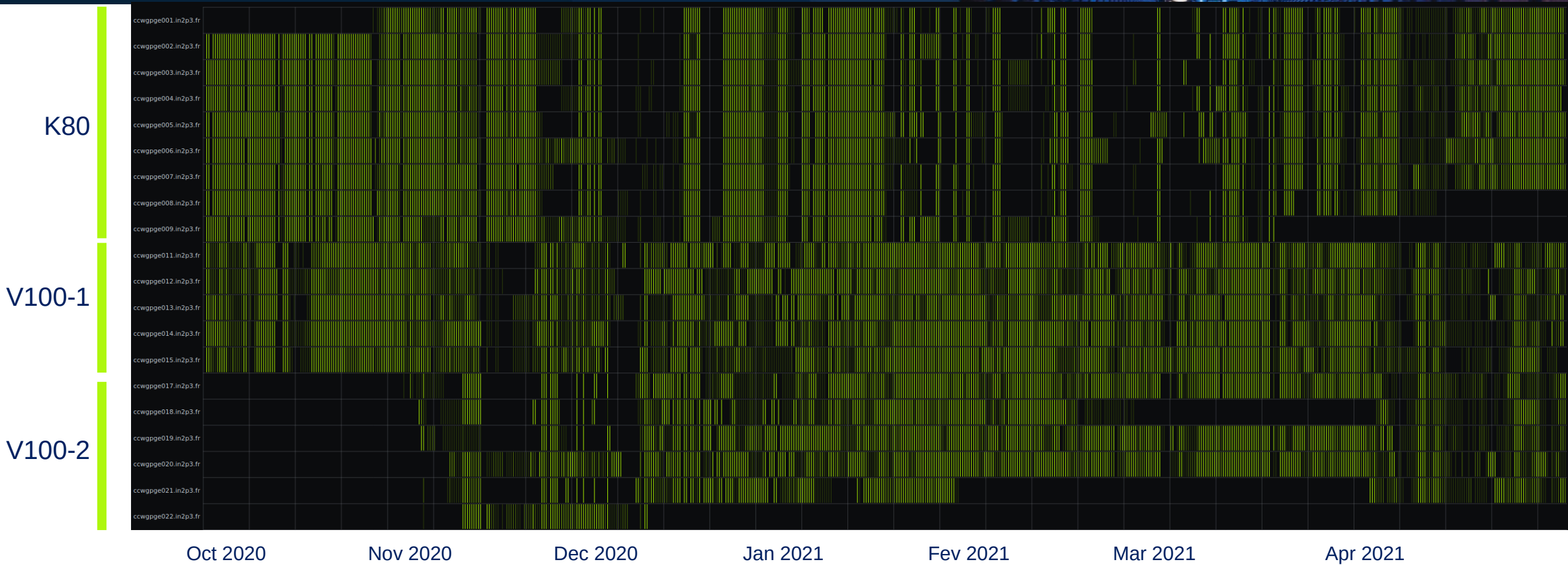


## Current utilization

```
$ gpu_info  
  
K80    worker : used/total/queued = 1/36/0  
K80    interactive : used/total/queued = 1/4/0  
V100   worker : used/total/queued = 23/40/0  
V100   interactive : used/total/queued = 6/8/0
```

Useful CLI command from Interactive machine ([cca.in2p3.fr](http://cca.in2p3.fr))

# GPU Farm Utilization



- ▶ Significant global slowdown in utilization from end of 2020 to March 2021
- ▶ Clear K80 desertion during this period

■ gpu in use  
■ gpu not used



## RTX8000

- ▶ Providing RTX8000 with 48GB memory for answering to a special use case driven by L2IT Atlas Team

CC-IN2P3 is able to study specific needs. Just let us know!

## JNP (Jupyter Notebooks Platform)

- ▶ CC-IN2P3 service
- ▶ Considering the possibility to propose K80 servers to the Jupyter Notebooks Platform
- ▶ Planning tests

Want to try it? Just let us know!

CC-IN2P3 is able to study specific needs.  
Just let us know!



## Questions?

[bertrand.rigaud@cc.in2p3.fr](mailto:bertrand.rigaud@cc.in2p3.fr)

## More questions?

<https://cc-usersupport.in2p3.fr>

Thanks for your attention.