LITO

Start on Jean Zay

Research and application in Artificial Intelligence

Inserm
La science pour la santé
From science to health

Presentation FLASH

Content

- Jean-Zay computational power.
- Main Characteristics.
- Request for access to computing resources.
- Consumption of hours.
- Computing environment: modules and virtual environments.

Computational Power

- Jean Zay is an HPE SGI 8600 calculator composed of two partitions: - a partition scalar or "CPU" containing 61,120 computing cores

  → 1528 scalar nodes (192GB memory, 40 cores @ 2.5GHz, Intel CSL 6248)

- A score converged or "GPU" containing 1292 GPU

  → 261 converged nodes quadri- GPU (192GB scalar node + 4 32GB Nvidia V100 GPUs and 4 OPA links)

  → 31 converged nodes octo- GPU (384GB or 768GB scalar node + 8 Nvidia V100 32GB GPUs and 4 OPA links)

Main Characteristics

- Cumulative peak power: (15.9 Pflops /s)
- Intel Omni-Path Interconnect Network (100Gb / s bandwidth)
- IBM Spectrum Scale Parallel File System (ex-GPFS)
- Two parallel storage devices:

  → 1.3Po on SSD disks (300Gio / s) and 300Po on rotating disks (100Gio / s)
- 5 frontal nodes

Request to connect

- The whole process takes ten days (an additional one to two months if you are subject to a security investigation).
- Dynamic access has been set up to request access to the computing resources of Jean Zay's GPU partition.

  → these accesses are intended for researchers and developers in Artificial intelligence.

  → requests can be made at any time of the year.
- An allowance is valid a year from the opening of the project.

Consumption of Hours

•The calculation of hours of calculation h is done as follows:

→ h = number of GPUs reserved × time elapsed

• Sometimes a knot is reserved exclusively : if more than one node is requested, or if the Slurm option is activated. In this case, the hours are counted as follows:

→ h = number of reserved nodes × 4 GPU × time elapsed

• To track your consumption:

```
idracct
```

Computing Environment

- Software for Artificial Intelligence are installed in virtual environments Anaconda, for Python 2 and Python 3

  → the Python 2.7 version is no longer maintained by the community since 01/01/2020

- Installed software: TensorFlow, PyTorch, Keras and Caffe.

  → the environments are activated when loading the modules

- Parallel management software: Horovod installed for TensorFlow and PyTorch.
- Development software: Jupyter Notebook (with TensorBoard) and JupyterLab.

Our Experiment

- Training the 'VOCALE' classification network with dedicated Nvidia Quadro p5000 and 8gb ram requires 8 to 10 hours to complete 100 epochs.

- The same experiment with same data set on the Jean-Zay server took only 23 minutes to complete 100 epochs.

Commands

cd $WORK /*work directory data input/output and command sources*/

idracct /*to see how many resources have already been used*/

sbatch -A kva@gpu /gpfswork/rech/kva/usf47gi/TrainVocale.sh /*to run batch*/

squeue -u usf47gi /*to know if batch is running*/

JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
541740 gpu_p1 Mnist_ex usf47gi PD 0:00 1 (Resources)
541872 gpu_p1 VOCALE usf47gi R 19:11 1 r6i6n5

Commands

Copy results on local disk

scp usf47gi@jean-zay.idris.fr:/gpfswork/rech/kva/usf47gi/*.* .
usf47gi@jean-zay.idris.fr's password:

MyModel.h5 100% 57MB 97.2MB/s 00:00
MyModel.hdf5 100% 169MB 95.0MB/s 00:01
MyModel.yaml 100% 14KB 1.0MB/s 00:00
TrainVocale.py 100% 5168 263.2KB/s 00:00
TrainVocale.sh 100% 394 431.9KB/s 00:00
VOCALE_example541872.out 100% 354KB 14.0MB/s 00:00
VOCALE_exemple541872.out 100% 5762 289.7KB/s 00:00

The Jean-Zay supercomputer, acquired in January 2019, is located at the CNRS Institute for Development and Resources in Computer Science (Idris).