

Special hardware available (GPUs) & how to use it

October 14, 2025

Rank	Nation	Machine	Performance	Accelerators
1.		El Capitan	1742 PFLOPs/s	AMD MI300A
2.		Frontier	1353 PFLOPs/s	AMD MI250X
3.		Aurora	1012 PFLOPs/s	INTEL DatCntGPUMax1550
4.		Jupiter Booster	793 PFLOPs/s	NVIDIA GH200
5.		Eagle	561 PFLOPs/s	NVIDIA H100
6.		HPC6	478 PFLOPs/s	AMD MI250X
7.		Fugaku	442 PFLOPs/s	
8.		Alps	435 PFLOPs/s	NVIDIA GH200
9.		Lumi	380 PFLOPs/s	AMD MI250X
10.		Leonardo	241 PFLOPs/s	NVIDIA A100

Partition	Model	#Cores	Clock Freq (GHz)	Memory (GB)	Bandwidth (GB/s)	TDP (W)	FP32/FP64 (GFLOPs/s)
zen2_0256_a40x2	45 × 2 × A40	5376/336	1.74	48	696	300	37000/578
zen3_0512_a100x2	60 × 2 × A100	6912/432	1.40	40	1600	400	20000/10000
gpu_rtx2080ti ^[†]	19 × 1 × rtx2080ti	4352/544	1.45	11	616	255	13400/400

[†] private nodes, available only at idle times

1. VSC-5> `salloc -N 1 -p zen3_0512_a100x2 --qos=zen3_0512_a100x2 --gres=gpu:2`

1. VSC-5> `salloc -N 1 -p zen3_0512_a100x2 --qos=zen3_0512_a100x2 --gres=gpu:2`
2. VSC-5> `queue -u $USER` (... to figure out which node has been assigned)

1. VSC-5> `salloc -N 1 -p zen3_0512_a100x2 --qos=zen3_0512_a100x2 --gres=gpu:2`
2. VSC-5> `queue -u $USER` (... to figure out which node has been assigned)
3. VSC-5> `ssh n3071-003` (... or whatever else node had been assigned)

1. VSC-5> `salloc -N 1 -p zen3_0512_a100x2 --qos=zen3_0512_a100x2 --gres=gpu:2`
2. VSC-5> `squeue -u $USER` (... to figure out which node has been assigned)
3. VSC-5> `ssh n3071-003` (... or whatever else node had been assigned)
4. `cuda-zen sh n3071-003:~$ spack unload`
`cuda-zen sh n3071-003:~$ spack load cuda@11.8.0%gcc@9.5.0/ananl33`
`cuda-zen sh n3071-003:~$ cd ~/examples/09_special_hardware/matrixMul`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ nvcc -arch=native`
`./matrixMul.cu`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ ./a.out`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ cd`
`cuda-zen sh n3071-003:~$ cd ~/examples/09_special_hardware/matrixMulCUBLAS`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ which nvcc`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ export`
`LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/gpfs/opt/sw/cuda-zen/spack-0.19.0/opt/spack/linux-`
`almalinux8-zen/gcc-9.5.0/cuda-11.8.0-ananl33lrrpp33xetcoltkbbbfuxoez/lib64`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ nvcc -arch=native`
`./matrixMulCUBLAS.cu -lcublas`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ ./a.out`

1. VSC-5> `salloc -N 1 -p zen3_0512_a100x2 --qos=zen3_0512_a100x2 --gres=gpu:2`
2. VSC-5> `queue -u $USER` (... to figure out which node has been assigned)
3. VSC-5> `ssh n3071-003` (... or whatever else node had been assigned)
4. `cuda-zen sh n3071-003:~$ spack unload`
`cuda-zen sh n3071-003:~$ spack load cuda@11.8.0%gcc@9.5.0/ananl33`
`cuda-zen sh n3071-003:~$ cd ~/examples/09_special_hardware/matrixMul`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ nvcc -arch=native`
`./matrixMul.cu`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ ./a.out`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ cd`
`cuda-zen sh n3071-003:~$ cd ~/examples/09_special_hardware/matrixMulCUBLAS`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ which nvcc`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ export`
`LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/gpfs/opt/sw/cuda-zen/spack-0.19.0/opt/spack/linux-`
`almalinux8-zen/gcc-9.5.0/cuda-11.8.0-ananl33lrrp33xetcoltkbbbfuxoez/lib64`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ nvcc -arch=native`
`./matrixMulCUBLAS.cu -lcublas`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ ./a.out`
5. `cuda-zen sh n3071-003:~$ nvidia-smi`

1. VSC-5> `salloc -N 1 -p zen3_0512_a100x2 --qos=zen3_0512_a100x2 --gres=gpu:2`
2. VSC-5> `squeue -u $USER` (... to figure out which node has been assigned)
3. VSC-5> `ssh n3071-003` (... or whatever else node had been assigned)
4. `cuda-zen sh n3071-003:~$ spack unload`
`cuda-zen sh n3071-003:~$ spack load cuda@11.8.0%gcc@9.5.0/ananl33`
`cuda-zen sh n3071-003:~$ cd ~/examples/09_special_hardware/matrixMul`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ nvcc -arch=native`
`./matrixMul.cu`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ ./a.out`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMul$ cd`
`cuda-zen sh n3071-003:~$ cd ~/examples/09_special_hardware/matrixMulCUBLAS`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ which nvcc`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ export`
`LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/gpfs/opt/sw/cuda-zen/spack-0.19.0/opt/spack/linux-`
`almalinux8-zen/gcc-9.5.0/cuda-11.8.0-ananl33lrrp33xetcoltkbbbfuxoez/lib64`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ nvcc -arch=native`
`./matrixMulCUBLAS.cu -lcublas`
`cuda-zen sh n3071-003:~/examples/09_special_hardware/matrixMulCUBLAS$ ./a.out`
5. `cuda-zen sh n3071-003:~$ nvidia-smi`
6. `cuda-zen sh n3071-003:~$ deviceQuery` (... from the SDK now to be self-compiled)

```
#!/bin/bash
#
# usage: sbatch ./gpu_test.scrpt
#
#SBATCH -J a100
#SBATCH -N 1
#SBATCH -p zen3_0512_a100x2
#SBATCH --qos=zen3_0512_a100x2
#SBATCH --gres=gpu:2

spack unload
spack load cuda@11.8.0%gcc@9.5.0/ananl33

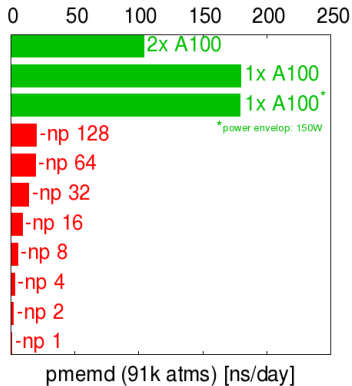
nvidia-smi
```

Using interactive mode or batch submission, figure out whether we have ECC enabled on GPUs of type A100 ?

Hint: `nvidia-smi -i 0 -q` will query all settings on the first device, ie number 0

Example: AMBER-22

Performance:



Power Efficiency:

