1 HPCG

1. What is the theoretical peak bandwidth and DP FP Performance of a single A100 80G card on Alex?
2. What do the three values nx, ny, nz in the HPCG.dat file stand for? How can you calculate the memory usage?
3. Measure and plot the performance of 1, 2, 4 and 8 GPUs.
   - To Download NVIDIAs HPCG and HPL Benchmarks you need to create an account on their NGC website. After Registration you need to go to your Account → Setup → API-Key and generate a new API-Key. Copy this key to a File, you need it later to download the container.
   - You can install the Container by following the singularity instructions in the Container Description at: https://catalog.ngc.nvidia.com/orgs/nvidia/containers/hpc-benchmarks
   - Derive a suitable script from the instructions in the file RUNNING inside the benchmarking container. Set the time for each run to 5 minutes.
4. What happens if you vary nx, ny or nz to be larger or smaller than the other two?
   - Test with a difference of 64, eg. 256 256 320.
5. Use nvidia-smi to measure the power consumption during one benchmark run. As peak power draw will be important in the competition, state the peak power measured.
   - Use sensible time intervals between measurements.
   - By dividing our Performance by the power we derive a key metric for HPC systems, Performance per Watt. State the performance per Watt.

2 MLPerf Inference

1. Make yourself familiar with the MLPerf Inference Benchmark. What is its aim?
   - Hint: The Introduction to MLPerf for SCC at SC22.
2. Build the MLPerf Inference Benchmark for Image Detection and Object Classification Tasks.
   - List what steps were necessary to build it.
   - State changes you made to configuration files, etc...
3. Run the Benchmark as CPU using onnx and one other backend of your choice.
   - Use at least two different models (e. g. Resnet50, Retinanet).
   - Document what steps you took to get these running.
   - Depending on the Dataset, limit the runtime with "–time <s>"
   - State your obtained results from the "–accuracy" pass option.