Containerized GPU

Maruthi S. Inukonda

26th Jun 2018

Agenda

- GPU Dockers & Benchmarking
- Multi-tenancy & CUDA

GPU Dockers

nvidia-docker

- To launch a nvidia GPU docker, use
  `nvidia-docker run -it --name <name> <image>:<tag> <program>`

```
$ docker images
REPOSITORY              TAG                 IMAGE ID            CREATED             SIZE
nvidia/opencl           latest              b546828c2b30        12 days ago         116MB
nvidia/cuda             latest              9337ecb4311e        7 weeks ago         2.24GB
ros                     kinetic-ros-base    2e1693285910        8 weeks ago         1.18GB
autoware/autoware        1.6.0-kinetic      8fc60a26cc84        6 months ago        7.65GB
```

```
$ nvidia-docker run -it --name nvcuda1 -v /mnt:/mnt nvidia/cuda:latest  bash
root@ba47b6c2ad70:/#
```

Performance difference between bare metal and container

A mean pooling program is run on a matrix of 4096x4096 256 times.
An element is updated with mean of its left, right, top and bottom elements.

```
On bare metal:

$ time ./cs18resch01001_Prog
real    0m14.549s
user    0m9.768s
sys    0m4.743s

In container

# time ./cs18resch01001_Prog
real    0m14.656s
user    0m2.057s
sys    0m12.603s
```

Multi-tenancy & CUDA

Launching multiple nvidia/cuda containers

```
$ nvidia-docker run -it --name nvcuda1 -v /mnt:/mnt nvidia/cuda:latest  bash
root@ba47b6c2ad70:/#

root@ba47b6c2ad70:/#  time ./cs18resch01001_Prog
real    0m28.756s
user    0m4.020s
sys    0m24.702s


$ nvidia-docker run -it --name nvcuda2 -v /mnt:/mnt nvidia/cuda:latest  bash
root@22dd291ccb73:/#

root@22dd291ccb73:~#  time ./cs18resch01001_Prog

real    0m28.778s
user    0m4.281s
sys    0m24.507s
```

Q & A