WIR SCHAFFEN WISSEN – HEUTE FÜR MORGEN

Marc Caubet Serrabou :: Systems Engineer :: Paul Scherrer Institute

NVIDIA DGX A100 Status

Merlin – PSI HPC Local Cluster

From Ubuntu to RHEL (I)

Why RHEL?

- **Ubuntu** not officially supported at PSI → no central installation, provisioning, repositories, etc.
- **RHEL** official PSI Linux distro with official RHEL support:
  - **RHEL7** fully in production, <u>widely used at PSI</u>.
    - Central installation (with custom PXE/TFTPBoot).
    - Provisioning with Puppet
  - **RHEL8** early stage of development
    - Central installation (with RHEL Satellite)
    - Provisioning with Ansible
      - Currently providing some basic configurations.
      - <u>Not yet ready for HPC systems.</u>

From Ubuntu to RHEL (II)

"NVDIA DGX A100"@PSI – status:

- **Firmware upgraded** ("nvfw-dgxa100:20.12.3.3") via docker container.
- **Mirrored NVIDIA DGX repositories** (Mellanox repositories already mirrored)
- **RHEL v7.9** (kernel 3.10.0-1160.15.2.el7.x86_64):
  - System **LVM partitioning** (on top of a RAID1) with multiple partitions (NVIDIA default → /):
    - `/, /var, /var/log, /boot/efi, /var/cache/afs, ...`
    - `/scratch` (LVM on top of a RAID5) vs NVIDIA's default `/raid` disk (RAID5 / RAID0)
  - **Shared storage access**:
    - <u>AFS</u> (Pmodules / software, AFS user / project directories, etc.).
    - PSI <u>NFS</u> user home directories.
    - <u>GPFS</u> user, project, shared-scratch directories (Infiniband)
- Integration with PSI <u>central installation</u> system and automatic <u>provisioning</u> via Puppet
  - **easy redeployment from scratch**
- <u>Integration</u> with Central services: alarming system (Icinga), AD (SSSD), etc.
- <u>Integration</u> with HPC monitoring (<u>Grafana</u>+InfluxDB)
- <u>Integration</u> with **Slurm** (see next slide)

Slurm integration

- Allows easy **resource sharing** of the machine as well as **fair** access and use of resources.
- Slurm upgraded yesterday (17.03.2021)
  - Running the latest **Slurm v20.11.5** (from v20.02.6) → fixing NVIDIA AutoDetect issues
- Integrated as a Slurm **GPU computing node** in a new `gmerlin6` cluster (`gwendoline` *partition)*
  - **NVIDIA autodetect**.
  - **MIG not ready** yet → feature has to be implemented in Slurm.
  - **MPS** possible (but not fully configured yet in our cluster).
  - Local `/scratch` 10TB
- **Integrated with central PSI software** (Pmodules) → i.e. OpenMPI with CUDA support
  - Flexibility and **easy code portability** to other PSI systems
- Dedicated <u>full access for specific set of users</u> (owning the machine)
- Open but <u>restricted access to other PSI users</u> (use case example: BIO)

Current issues and questions

- NVIDIA DGX **repositories not in sync** with Mellanox repositories
  - i.e. ucx-cuda requires CUDA 10.2, NVIDIA DGX repos offering 10.0, 10.1, 11.0
- Official **NVIDIA DGX repositories older than PSI software repositories**, but also:
  - Understanding future DGX upgrades and NVIDIA's upgrade policy.
  - **Example**: Running software based on CUDA 11.1.0 caused some problems
    - Solved by switching to CUDA 11.0.3
  - **Understanding CUDA compatibility matrix** (CUDA vs driver version)
  - Two **contradictory sources**:
    - https://docs.nvidia.com/deploy/cuda-compatibility/index.html
    - https://docs.nvidia.com/cuda/cuda-toolkit-release-notes/index.html
- **Container technology**? Singularity vs ENROOT vs ~~Sarus~~ → integrating containers in Slurm
  - Which container types is NVIDIA officially supporting?
- **Understanding NVIDIA DGX A100 architecture** for optimal use.
  - 8 NUMA nodes, 4 NUMA nodes with 2 GPUs + 2 IB ports each

Future work

- Network upgrade:
  - Adding a second **25Gbps** port (2x25Gbps LACP) for network redundancy.
  - Adding (at least) **1 Infiniband cable per NUMA node** with GPUs (currently, 2 cables connected)
- Setting up **Slurm + Containers**:
  - Not integrated yet at PSI and need to decide which container technology will be used.
  - Alternative (at present): running containers in a batch script or by allocating resources.
    - Singularity and ~~docker~~ already available... ENROOT?
  - Exploring **NVIDIA NGC** and integrating with users workflows:
    - Dealing with **user management**.
- **Multi-Instance GPU** in Slurm.
- **Running software optimally.**
- Migrating to **RHEL8**

My thanks go to

- The HCPE Team
  - Achim Gsell
  - Derek Feichtinger
  - Spencer Bliven
- HPC & Science-IT, specially
  - Hans-Christian Stadler
  - Alain Studer
- The Linux Team
- The DataCenter Team