

# Nvidia DGX A100

Im Herbst 2020 wurden die ersten Nvidia DGX A100 kurz nach der Veröffentlichung von Nvidia in Hilbert integriert. Die Systeme eignen sich besonders für extreme KI-Anwendungen durch ihren hohen Speicher pro Karte und die sehr hohe Anzahl an CUDA-Cores pro System. Zudem sind die Systeme mit 4x100Gbit/s Infiniband an den restlichen Cluster und an den Storage angeschlossen.

Die Systeme verfügen über einen lokalen Cache für Daten die über mehrere Jobs hinweg mehrfach gelesen werden müssen um die Latenzen beim Laden von Daten zu minimieren.

## PyTorch

Es wird mindestens Version 1.8.0 benötigt um auf den Karten zu rechnen

```
pip install --user --pre -i http://pypi.repo.test.hhu.de/simple/ --trusted-host pypi.repo.test.hhu.de /software/pytorch/torch-1.8.0.dev20201102+cu110-cp36-cp36m-linux_x86_64.whl /software/pytorch/torchvision-0.9.0.dev20201102+cu110-cp36-cp36m-linux_x86_64.whl
```

## CMake

Please use Cmake 3.19 or newer when compiling software which uses FindCUDA to generate PTX-Code for Compute Capability 8.0 and above.

## Hardware

### Nvidia DGX A100

<b>Jahr</b>	2020
<b>CPU-Architektur</b>	AMD EPYC 7742, 2.25GHz
<b>Cores</b>	2x64
<b>RAM</b>	2 TB DDR4
<b>Netzwerk</b>	1Gbit/s Ethernet 4x100Gbit/s Infiniband EDR
<b>Beschleuniger</b>	8x <a href="#">Nvidia A100 SMX4 40GB</a> 40GB Memory per GPU
<b>Peak Performance (CUDA)</b>	19.5 TFLOPS (FP64, double precision)
<b>PBS</b>	accelerator_model=a100
<b>Nodes</b>	4Nodes, hilbert[400-403]