**The AI@AU GPU Cluster**

Currently, the AI@AU GPU Cluster is composed of a total of 24 GPUs. Sixteen of the GPUs belong to two Lambda Hyperplane 8x NVIDIA A100 machines. The last 8 GPUs belong to a Lambda Hyperplane 8x NVIDIA H200 machine (an NVIDIA H200 is approximately 8x faster than an NVIDIA A100 and has twice the memory). The machine specifications for the three Hyperplanes are shown below. The 24 GPUs are integrated using a Bright Cluster Manager.

**Lambda Hyperplane #1 (8-A100 GPUs)**

|  |  |
| --- | --- |
| GPUs | 8x NVIDIA A100 (80GB) SXM4 |
| Processor | 2x AMD EPYC 7763: 64 cores, 2.45GHz, 256MB cache, PCIe 4.0, Milan |
| System Memory | 1024 GB: DDR4-3200 ECC RDIMM |
| Operating System | Ubuntu 22.04: Includes Lambda Stack for managing TensorFlow, PyTorch, CUDA, cuDNN, etc. |
| OS Drives | 2x 1.92 TB M.2 NVMe: Data center SSD, 1 DWPD, PCIe 4.0 |
| Data Drives | Data drives: 1x 7.68 TB U.2 NVMe: Data center SSD, 1 DWPD, PCIe 4.0 |

**Lambda Hyperplane #2 (8-A100 GPUs)**

|  |  |
| --- | --- |
| GPUs | 8x NVIDIA A100 (80GB) SXM4 |
| Processor | 2x AMD EPYC 7763: 64 cores, 2.45GHz, 256MB cache, PCIe 4.0, Milan |
| System Memory | 1024 GB: DDR4-3200 ECC RDIMM |
| Operating System | Ubuntu 22.04: Includes Lambda Stack for managing TensorFlow, PyTorch, CUDA, cuDNN, etc. |
| OS Drives | 2x 1.92 TB M.2 NVMe: Data center SSD, 1 DWPD, PCIe 4.0 |
| Data Drives | Data drives: 1x 7.68 TB U.2 NVMe: Data center SSD, 1 DWPD, PCIe 4.0 |

**Lambda Hyperplane #3 (8-H200 GPUs)**

|  |  |
| --- | --- |
| GPUs | 8 x NVIDIA H200 SXM5 |
| Processor | 2x Intel Xeon Platinum (8570 56-core) |
| System Memory | 2048 GB |
| Operating System | Ubuntu 22.04 - Includes Lambda Stack for managing TensorFlow, PyTorch, CUDA, cuDNN, etc. |
| OS Drives | 2x 1.92TB M.2 NVMe |
| Data Drives | 2 x Samsung PM9A3 NVMe Gen4 U.2 SSD - 3.84TB |