Wikiprint Book

Title: System overview

Subject: DEEP - Public/User_Guide/System_overview

Version: 47

Date: 22.07.2024 17:36:15

Table of Contents

System overview

This page is supposed to give a short overview on the available systems from a hardware point of view. All hardware can be reached through a login node via SSH: **deep@fz-juelich.de**. The login node is implemented as virtual machine hosted by the master nodes (in a failover mode). Please, see also information about getting an account and using the batch system.

DEEP-EST Modular Supercomputer (prototype system)

The DEEP-EST system is a prototype of Modular Supercomputing Architecture (MSA) consisting of the following modules:

- Cluster Module (CM)
- Extreme Scale Booster (ESB)
- Data Analytics Module (DAM)

In addition to the three compute modules, a Scalable Storage Service Module (SSSM) provides shared storage infrastructure for the DEEP-EST prototype (`/usr/local`) and is accompanied by the All Flash Storage Module (AFSM) leveraging a fast local work filesystem (`/afsm`) on the compute nodes. All modules are connected via a 100 Gbp/s EDR IB network in a non-blocking tree topology accompanied by a Gigabit Ethernet service network. In addition the system is connected to the Jülich storage system (JUST) to share home and project file systems with other HPC systems hosted at Jülich Supercompting Centre (JSC).

Cluster Module

It is composed of 50 nodes with the following hardware specifications:

Cluster [50 nodes]: `dp-cn[01-50]`: • 2 Intel Xeon 'Skylake' Gold 6146 (12 cores (24 threads), 3.2GHz) • 192 GB RAM • 1 x 400GB NVMe SSD • network: InfiniBand EDR (100 Gb/s)	

Extreme Scale Booster

It is composed of 75 nodes with the following hardware specifications:

Extreme Scale Booster [75 nodes]: `dp-esb[01-75]` • 1 x Intel Xeon 'Cascade Lake' Silver 4215 CPU @ 2.50GHz • 1 x Nvidia V100 Tesla GPU (32 GB HBM2) • 48 GB RAM • 1 x 512 GB SSD • network: IB EDR (100 Gb/s)	

Data Analytics Module

It is composed of 16 nodes with the following hardware specifications:

Data Analytics Module [16 nodes]: `dp-dam[01-16]` • 2 x Intel Xeon 'Cascade Lake' Platinum 8260M CPU @ 2.40GHz • dp-dam[01-08]: 1 x Nvidia V100 Tesla GPU (32 GB HBM2) • dp-dam[09-12]: 2 x Nvidia V100 Tesla GPU (32 GB HBM2) • dp-dam[13-16]: 2 x Intel STRATIX10 FPGA (32 GB DDR4) • 384 GB RAM + 3 TB non-volatile memory • 2 x 1.5 TB Intel Optane SSD (1 for local scratch, 1 for BeeOND) • 1 x 240 GB SSD (for boot and OS) • network: IB EDR (100 Gb/s)	

Scalable Storage Service Module

It is based on spinning disks and composed of 4 volume data server systems, 2 metadata servers and 2 RAID enclosures. The RAID enclosures each host 24 spinning disks with a capacity of 8 TB each. Both RAIDs expose two 16 Gb/s fibre channel connections, each connecting to one of the four file servers. There are 2 volumes per RAID setup. The volumes are driven with a RAID-6 configuration. The BeeGFS global parallel file system is used to make 292 TB of data storage capacity available.

Here are the specifications of the main hardware components more in detail:

SSSM [6 servers]: `dp-fs[01-06]`: • 2 Intel Xeon Silver 4114 (20 cores, 2.2 GHz) • 96 GB RAM • 2 x 240 GB SSD • (additional 2 x 480 GB SSD in `dp-fs[01-02]` for metadata) • network: IB EDR (100 Gb/s) SSSM [2 EUROstor ES-6600 RAID enclosures]: `dp-raid[01-02]`: • 24 x 8 TB SAS Nearline • 2 x 16 Gb FC connector	

All Flash Storage Module

It is based on PCIe3 NVMe SSD storage devices. It is composed of 6 volume data server systems and 2 metadata servers interconnected with a 100 Gbps EDR-InfiniBand fabric. The BeeGFS global parallel file system is used to make 1.3 PB of data storage capacity available.

Here are the specifications of the main hardware components more in detail:

AFSM [2 metadata servers]: `dp-afsm-m[01-02]`: • 2 Intel Xeon Scalable Gold 6246 (12 cores (24 threads), 3.30 GHz) • 192 GB RAM • 25.6 TB SSD PCIe3 NVMe (using 8 x 3.2 TB Intel SSD DC P4610) • network: 1 x 100 Gbps EDR-InfiniBand HCA (PCIe3 x16) AFSM [6 volume data servers]: `dp-afsm-o[01-06]`: • 2 Intel Xeon Scalable Gold 6226R (16 cores (32 threads), 2.90 GHz) • 384 GB RAM • 308 TB SSD PCIe3 NVMe (using 24 x 15.36 TB Intel SSD DC P4326) • network: 1x 100 Gbps EDR-InfiniBand HCA (PCIe3 x16)	

Interconnect

As shown in the system overview an EDR IB non-blocking fat tree is used as fast interconnect inside and between all modules along with a Gigabit Ethernet service network (used for administration). The IB fat tree is composed of 6 spine and 10 leaf switches:

Rack plan

This is a sketch of the available hardware reflecting the current rack layout.

SSSM rack

This rack hosts the master nodes (frontends), SSSM file servers and the storage as well as network components for the Gigabit Ethernet administration and service networks. Users can access the login node via **deep@fz-juelich.de** (implemented as virtual machine running on the master nodes). The rack is air-cooled.

CM rack

Contains the hardware of the DEEP-EST Cluster Module including compute nodes, a management node for this module, network components and a liquid cooling unit.

DAM rack

This rack hosts the compute nodes of the Data Analytics Module of the DEEP-EST prototype, a management node for this module, network components and 4x BXI test nodes plus switch. The rack is air-cooled.

SDV rack

Along with the actual prototype system several test nodes and so called software development vehicles (SDVs) have been installed in the scope of the DEEP(-ER,EST) projects. These are located in the SDV rack (07). KNL and ml-GPU nodes can be accessed by the users via SLURM. Access to the remaining SDV nodes can be given on demand:

Prototype DAM [4 nodes]: `protodam[01-04]`

- 2 x Intel Xeon 'Skylake' (26 cores per socket)
- 192 GB RAM
- network: 1 Gigabit Ethernet

Old DEEP-ER Cluster Module SDV [10 nodes]: `deeper-sdv[01-10]`

- 2 Intel Xeon 'Haswell' E5-v2680 v3 (2.5 GHz)
- 128 GB RAM
- 1 NVMe with 400 GB per node(accessible through BeeGFS on demand)
- network: 1 Gigabit Ethernet
- deeper-sdv[09,10]: 1 x Arria 10 FPGA PAC

KNLs [4 nodes]: `knl[01,04-06]`

- 1 Intel Xeon Phi (64-68 cores)
- 1 NVMe with 400 GB per node (accessible through BeeGFS on demand)
- 16 GB MCDRAM plus 96 GB RAM per KNL
- network: 1 Gigabit Ethernet

GPU nodes for Machine Learning [3 nodes]: `ml-gpu[01-03]`

- 2 x Intel Xeon 'Skylake' Silver 4112 (2.6 GHz)
- 192 GB RAM
- 4 x Nvidia Tesla V100 GPU (PCIe Gen3), 16 GB HBM2
- network: 40GbE connection inbetween, 1 GbE external

Further information

- Information about the batchsystem
- Filesystems
- Information on available software and tools
- Use the DEEP-EST Cluster Module
- Use the DEEP-EST Data Analytics Module
- Use the SDV Cluster
- Use the SDV KNLs