HPC Validated Design for AI and Data Analytics

Design and Deployment Guidance

October 2021

H18161

Architecture Guide

Abstract

This architecture guide provides high-level guidance for building a converged architecture that allows customers to run HPC, AI and data analytics workloads on a single infrastructure.

Dell Technologies Solutions

DELL Technologies

Copyright

Contents

Introduction

Solution overview

The current intersection of high-performance computing (HPC), artificial intelligence (AI), and data analytics workloads stems from the convergence of workflows that require different techniques to solve complex problems. AI and data analytics are being used to augment traditional HPC workloads. Combining AI and data analytics methods with traditional HPC workflows can speed scientific discovery and innovation processes. Data scientists and researchers are developing new processes for solving problems at massive scale, requiring compute resources such as those of HPC systems. AI and data analytics workloads benefit from an HPC infrastructure that can scale up to improve performance.

This convergence of workloads is causing customers to look for a unified architecture that supports all the workloads. Traditional HPC workloads such as simulation and modeling are compute-intensive and benefit from fast interconnects and high-performing file systems. Traditional HPC workloads are submitted through a batch scheduler, taking hours or days to run. However, AI and data analytics workloads are data-intensive and require tools from data ingest to data science workbenches. AI and data analytics workloads are interactive; data scientists are iterative when building and training models. Typically, AI and data analytics workloads are initiated and terminated repeatedly, depending on the development stage of the data workflow.

The variance in HPC, AI, and data analytics workloads can lead customers to believe that they need three separate environments; however, this belief is not accurate. Customers can design a unified architecture with multipurpose, balanced nodes to support all workloads. Also, with the correct software and tools, customers can satisfy traditional HPC users as well as AI and data analytics users without forcing them to learn new skills and adjust to a new operational model.

Integrating all three workloads on a single architecture presents challenges, however. Customers must consider that:

- HPC users' skills are different from the skills of AI and data analytics users.
- Resource management systems and schedulers are not interoperable.
- Software and frameworks are not all integrated in a single platform.
- Ecosystems require different tools and features.
- Workload stacks and performance needs are different.

The Dell Technologies HPC Validated Design for AI and Data Analytics takes a building-block design approach, giving customers flexibility in architecture choices that best meet their environmental needs. The architecture design encompasses:

- Server building blocks
- Networking building blocks
- Storage and file system building blocks
- Software environment building blocks, as described in Deployment guidance

The HPC Validated Design for AI and Data Analytics provides an enterprise supported version that uses Bright Cluster Manager for Data Science.

Bright Cluster Manager for Data Science enables you to manage data science clusters as a single entity. You can provision the hardware, operating system, HPC, Big Data software, and deep learning software from a single interface. For customers who want an open-source option, Dell Technologies provides Ansible playbooks that provide a convenient way to install the software for the HPC Validated Design for AI and Data Analytics.

Document purpose

This document is intended to provide high-level guidance on architecture design and deployment for a converged architecture. The document provides the following information:

- Architecture overview—server, network, storage, and software
- Configuration guidance—server, network, storage, and software
- Performance optimization guidance
- Overview of Ansible playbooks for Kubernetes and Docker

Audience

This document is intended for solution architects and IT operations personnel who want to deploy a single environment to run multiple workloads for HPC, AI, and data analytics. The architecture includes Kubernetes and Docker for AI and data analytics workloads.

We value your feedback

Dell Technologies and the authors of this document welcome your feedback on the solution and the solution documentation. Contact the Dell Technologies Solutions team by email or provide your comments by completing our documentation survey.

Author: Armando Acosta

Contributors: Frank Han, Luke Wilson, John Lockman

Note: For links to additional documentation for this solution, see the Dell Technologies Solutions Info Hub for AI & Data Analytics.

Design overview

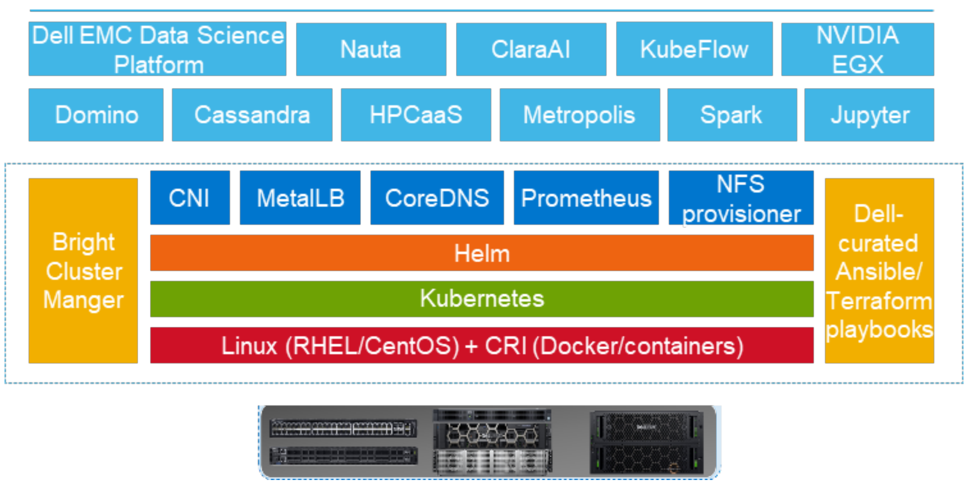
Introduction

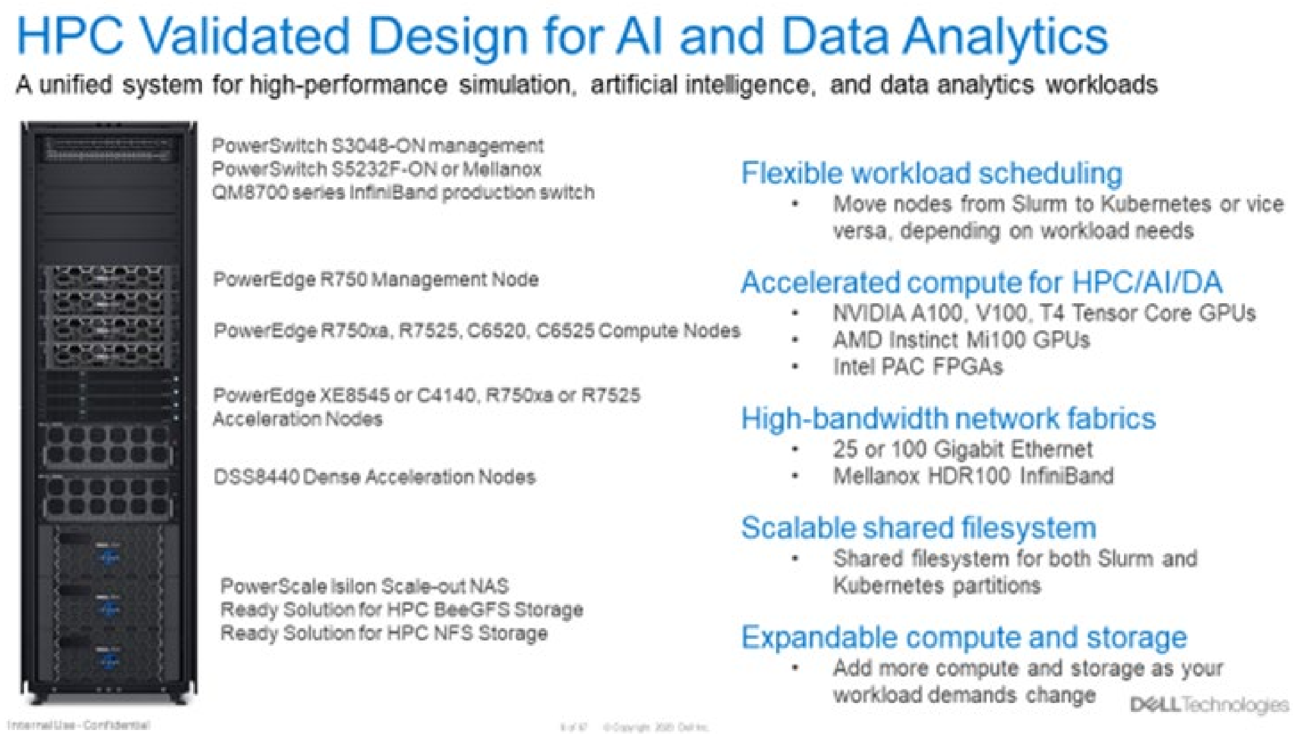
The HPC Validated Design for AI and Data Analytics is a single environment that is tailored for a wide variety of HPC, AI, and high-performance data analytics workloads. The architecture provides a single infrastructure pool. You can assign the pool either to HPC jobs that are managed by an HPC resource manager like Slurm or to containerized AI and data analytics workloads that are orchestrated by the open-source Kubernetes container orchestration system.

The HPC Validated Design for AI and Data Analytics is supported by validation, interoperability testing, and benchmarking. The solution design uses a building-block approach, giving you flexibility in choosing the compute, network, storage, and software features that are best suited for the use cases and workloads that they will run in your environment. Customer conversations indicate that a no-one-size-fits-all approach works.

The following figures show the components and building blocks of the HPC Validated Design Stack:

Figure 1. HPC Validated Design for AI and Data Analytics high-level architecture





Server building blocks

For compute power, the architecture offers multiple servers, including the PowerEdge R750 server, which is an all-in-one building block for Spark, machine learning, deep learning training and inferencing. The architecture also provides higher density options like the PowerEdge XE8545 server, which is well suited for large-scale HPC simulation workloads, machine learning training, deep learning training, and inferencing. Each server offers unique features that make it viable based on the customer use case.

The architecture is defined by node types: management, compute, high-performance acceleration, and dense acceleration, as described in the following table:

Table 1. Node roles

Management	Compute	High-performance acceleration	Dense acceleration
• **Enterprise**—Runs Bright Cluster Manager and Bright Cluster Manager for Data Science • **Open source**—Runs cluster management tools, Dell EMC Data Science Portal, and Ansible playbooks for Kubernetes and Docker	• Provides the balanced building block for AI, machine learning, and deep learning workloads • Supports NVIDIA A100, V100 PCIe or AMD Instinct Mi100 GPUs for object detection and image recognition • Supports A10, and FPGA for machine learning, inferencing • Supports high memory configuration for in-memory processing for Spark	• Provides NVLink enabled, high-performance GPU building block for HPC simulation and deep learning training • Supports NVIDIA A100 or V100 NVLink for simulation and image recognition	• Provides the PCIe-enabled, dense all-in-one building block for large-scale HPC simulation workloads, deep learning training, machine learning training, and inferencing • Supports NVIDIA A10 for machine learning and natural language processing • Supports NVIDIA V100 PCIe for simulation and image recognition
Recommended server: PowerEdge R750	PowerEdge R750xa, R7525, C6520, C6525 servers	PowerEdge XE8545 or with NVLink R750xa or R7525	Recommended server: DSS8440

The following tables provide configuration details about the supported server building blocks:

Table 2. PowerEdge R750 management node

Component	Configuration details
Processor	2x Intel(R) Xeon(R) Gold 6342 CPU 2.8 G, 24C/48T
Memory	16x 64GB RDIMM, 3200 MT/s, dual rank
Hard drive	16x 960 GB SSD SATA read Intensive 6 Gbps
RAID	H755
Network	1x Mellanox ConnectX-6 100G Single port HDR

Table 3. PowerEdge R740 management node

Component	Configuration details
Processor	2x Intel Xeon Gold 6248R 3.0G, 24C/48T, 10.4GT/s, 35.75M cache
Memory	24x 64 GB RDIMM, 2933 MT/s, dual rank
Hard drive	24x 3.84 TB SSD SATA read Intensive 6 Gbps
RAID	PERC H740P RAID controller
Network	Mellanox ConnectX-6 Single port HDR100

Table 4. PowerEdge R740XD compute node

Component	Configuration details
Processor	2x Intel Xeon Gold 6248R 3.0 G, 24C/48T, 10.4 GT/s, 35.75M cache
Memory	24x 16 GB RDIMM, 2933 MT/s, dual rank
Hard drive	2x 960 GB SSD SATA read Intensive 6 Gbps
RAID	PERC H730P RAID controller
Network	Mellanox ConnectX-6 Single port HDR100
Accelerator options	2x NVIDIA A100, 3x NVIDIA A10 PCIe, 2x NVIDIA V100, 2x AMD Instinct Mi100, or 6x NVIDIA T4, or 6 x Intel FPGA

Table 5. PowerEdge R750XA compute node

Component	Configuration details
Processor	2x Intel(R) Xeon(R) Gold 6342 CPU 2.8 G, 24C/48T
Memory	16x 32GB RDIMM, 3200 MT/s, dual rank
Hard drive	1x 960 GB SSD SATA read Intensive 6 Gbps
RAID	No PERC
Network	2x Mellanox ConnectX-6 100G Single port HDR
Accelerator options	4x NVIDIA A100 PCIe, 4x NVIDIA A10 PCIe, 4x NVIDIA A40 PCIe, 8x NVIDIA T4, 4x AMD Instinct Mi100

Table 6. PowerEdge R7525 compute node

Component	Configuration details
Processor	2x AMD 7543 2.8GHz,32C/64T,256M
Memory	24x 16 GB RDIMM, 3200 MT/s, dual rank
Hard drive	PERC H745 Controller
RAID	1x 960 GB SSD SATA SED read Intensive 512e
Network	Mellanox ConnectX-6 Single port HDR
Accelerator options	2x NVIDIA A100 PCIe, 3x NVIDIA A10 PCIe, 2x AMD Instinct Mi100, 6 x NVIDIA T4

Table 7. PowerEdge C6520 node

Component	Configuration details
Processor	2x Intel(R) Xeon(R) Gold 6326 CPU 2.9 G, 16C/32T
Memory	16x 64GB RDIMM, 3200 MT/s, dual rank
Hard drive	1x 960 GB SSD SATA read Intensive 6 Gbps
RAID	No PERC
Network	1x Mellanox ConnectX-6 100G Single port HDR

Table 8. PowerEdge C6420 compute node

Component	Configuration details
Processor	2 x Intel Xeon Gold 6248R 3.0 G, 24C/48T, 10.4 GT/s, 35.75M cache
Memory	24 x 16 GB RDIMM, 2933 MT/s, dual rank
Hard drive	2 x 960 GB SSD SATA read Intensive 6 Gbps
RAID	PERC H730P RAID controller
Network	Mellanox ConnectX-6 Single port HDR100

Table 9. PowerEdge C6525 compute node

Component	Configuration details
Processor	2x AMD 7543 2.8GHz,32C/64T,256M
Memory	24 x 16 GB RDIMM, 3200 MT/s, dual rank
Hard drive	1 x 960 GB SSD SATA SED read Intensive 512e
RAID	PERC H745 Controller
Network	Mellanox ConnectX-6 Single port HDR

Table 10. PowerEdge XE8545 compute node

Component	Configuration details
Processor	2x AMD 7543 2.8GHz,32C/64T,256M
Memory	16x 64GB DIMM, 3200 MT/s, dual rank
Hard drive	No PERC
RAID	2x 960GB Intel SSD SATA SED read Intensive 512e
Network	2x Mellanox ConnectX-6 Single port HDR
Accelerator options	4x NVIDIA A100 SXM4

Table 11. DSS8440 dense GPU node

Component	Configuration details
Processor	2 x Intel Xeon Gold 6248R 3.0 G, 24C/48T, 10.4 GT/s, 35.75M cache
Memory	12 x 32 GB RDIMM, 2933 MT/s, dual rank
Hard drive	PERC H730P RAID Controller, 2 GB NV cache
RAID	1 x 960 GB SSD SATA SED read Intensive 512e
Network	Mellanox ConnectX-6 Single port HDR100
Accelerator options	10 x NVIDIA V100 PCIe, 10 x NVIDIA T4

Networking building blocks

The HPC Validated Design for AI and Data Analytics offers these options:

- Dell EMC PowerSwitch S5232F-ON switch is a high-performance Ethernet building block well suited for HPC simulation workloads, deep learning training, machine learning training, and inferencing. It supports up to 6.4 Tbps of switching I/O bandwidth for high performance and low latency.
- The NVIDIA Mellanox QM8700 switch is a high-performance InfiniBand building block is well suited for large-scale HPC simulation workloads and deep learning training. It supports up to 16 Tbps of nonblocking bandwidth with sub 130 ns port-to-port latency for performance and low latency.

The following figure shows the Ethernet switch architecture:

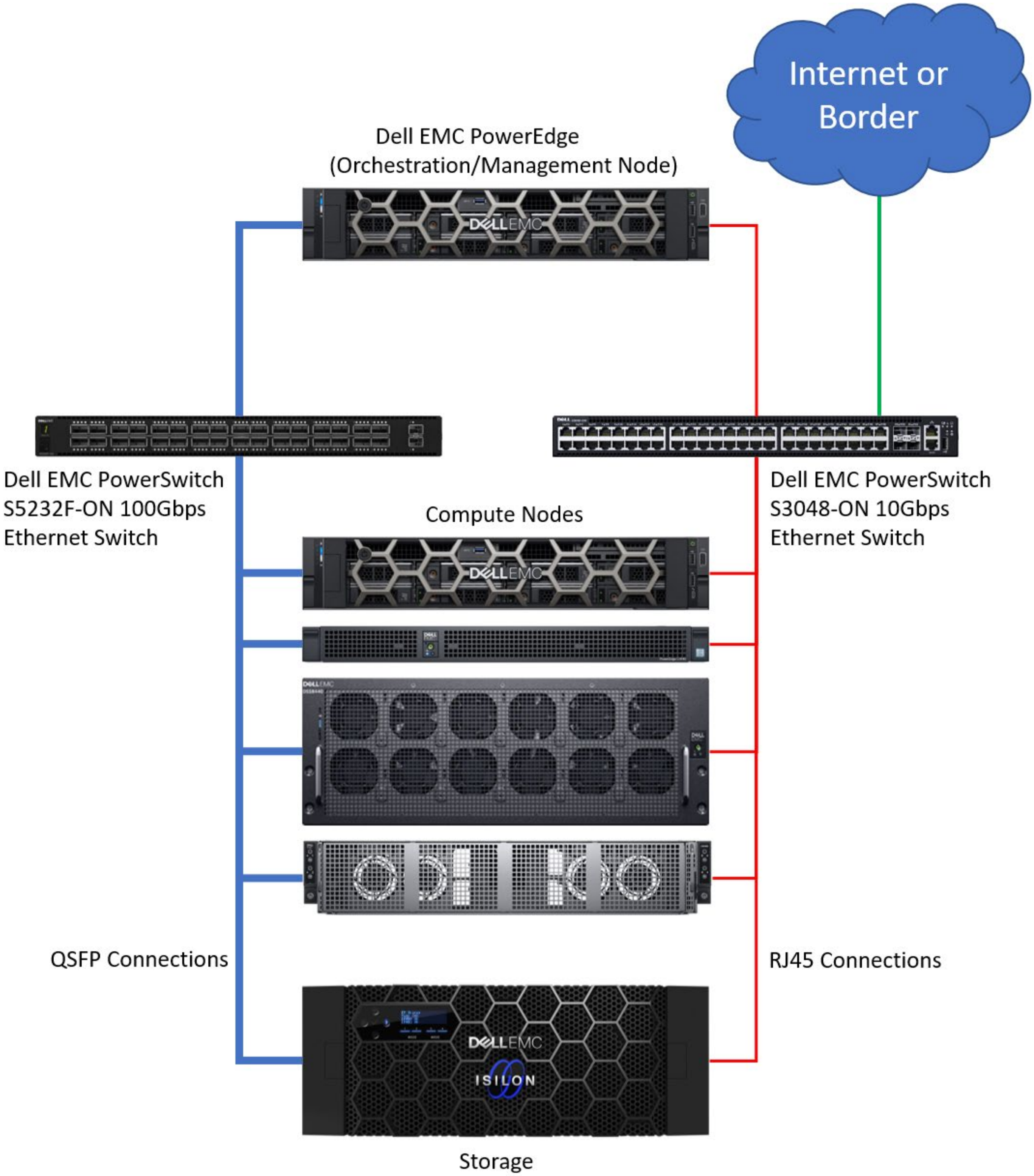


Figure 1. HPC Validated Design for AI and Data Analytics with Ethernet switches

The following figure shows the InfiniBand switch architecture:

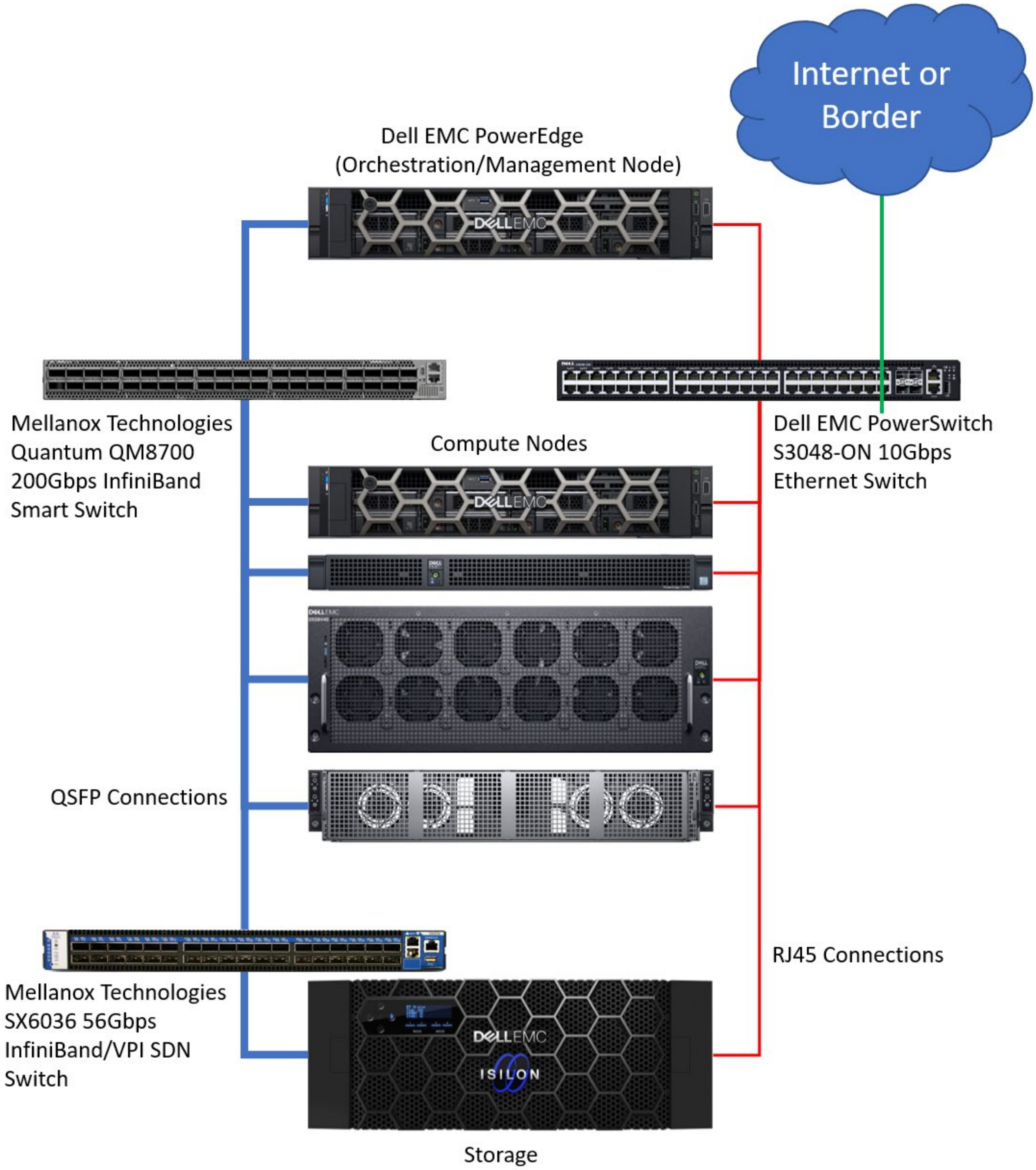


Figure 2. HPC Validated Design for AI and Data Analytics with InfiniBand switches

Choose a networking switch that enables running multiple workloads at high bandwidth and low latency for best performance. The networking switch must have a fast connection to compute and storage resources for best performance.

Storage and file system building blocks

Customers want to have multiple storage and file system choices to best address their workload needs. The HPC Validated Design for AI and Data Analytics offers multiple storage building blocks to meet customer performance needs based on their environment.

The architecture features:

- **Dell EMC PowerScale family with Isilon F800**—An all-flash system that provides a high-performance Network File System (NFS) storage building block that is well suited for large-scale HPC simulation workloads, deep learning training, machine learning training, and inferencing. Isilon F800 all-flash storage is powered by the Dell EMC OneFS operating system. The Isilon F800 provides a powerful yet simple scale-out NFS storage architecture to speed access to massive amounts of unstructured data while dramatically reducing cost and complexity.
- **Dell EMC PowerVault ME4084**—A DAS storage building block that is well suited for HPC simulation workloads, machine learning, and deep learning training. It delivers all-around simplicity, high performance, and high-capacity entry storage. Designed for versatility, the ME4 Series supports various drive types, multiple protocols, and all-inclusive software features.

The architecture also supports multiple parallel cluster file systems, which are designed to manage I/O-intensive workloads in performance-critical environments. One option is BeeGFS, which transparently spreads user data across multiple servers. File system performance and capacity can be scaled out by increasing the number of servers and disks in the system. Because high-performance I/O is a primary goal, the system has been designed as a high-speed scratch solution. At the core of the solution are high-speed NVMe SSDs that offer high bandwidth and low latency by removing the scheduler and queueing bottlenecks from the block layer. The BeeGFS file system also supports high aggregate I/O throughput.

The Validated Design for HPC NFS Storage provides a scalable, high-availability NFS storage service that can be directly integrated into the system fabric. The core of the solution is an HA cluster, which provides a highly reliable and available storage service to HPC compute clusters by using a high-performance network connection such as InfiniBand or 10/25 Gb Ethernet (10/25 GbE).

Deployment guidance

The HPC Validated Design for AI and Data Analytics has two deployment options:

- **Enterprise**—Runs Bright Cluster Manager and Bright Cluster Manager for Data Science, and Data Science Laboratory
- **Open source**—Runs cluster management tools, Data Science Laboratory, and Ansible playbooks for Kubernetes and Docker

The Data Science Laboratory is available with Bright software. In the open-source environment, Ansible playbooks have deployment scripts to help deploy Data Science Laboratory on the management node.

Bright Cluster Manager and Bright Cluster Manager for Data Science

Bright Cluster Manager and Bright Cluster Manager for Data Science contain tools and applications to facilitate the installation, administration, and monitoring of a cluster. Also, Bright Cluster Manager provides users with an optimal environment for developing and running applications that require extensive computational resources.

Bright Cluster Manager for Data Science lets you administer data science clusters as a single entity, provisioning the hardware, operating system, HPC, data analytics, and deep learning software from a single interface. This capability makes it easier to build a reliable, clustered data science infrastructure. When your system is up and running, the Bright cluster management daemon monitors every aspect of every node. The daemon reports any problems that it detects in the software or the hardware so that you can act and keep your infrastructure healthy.

To prepare systems for software deployment using Bright software, ensure that each server is racked, powered, and networked so that the server can download software packages from the Internet or from a full Red Hat Enterprise Linux mirror.

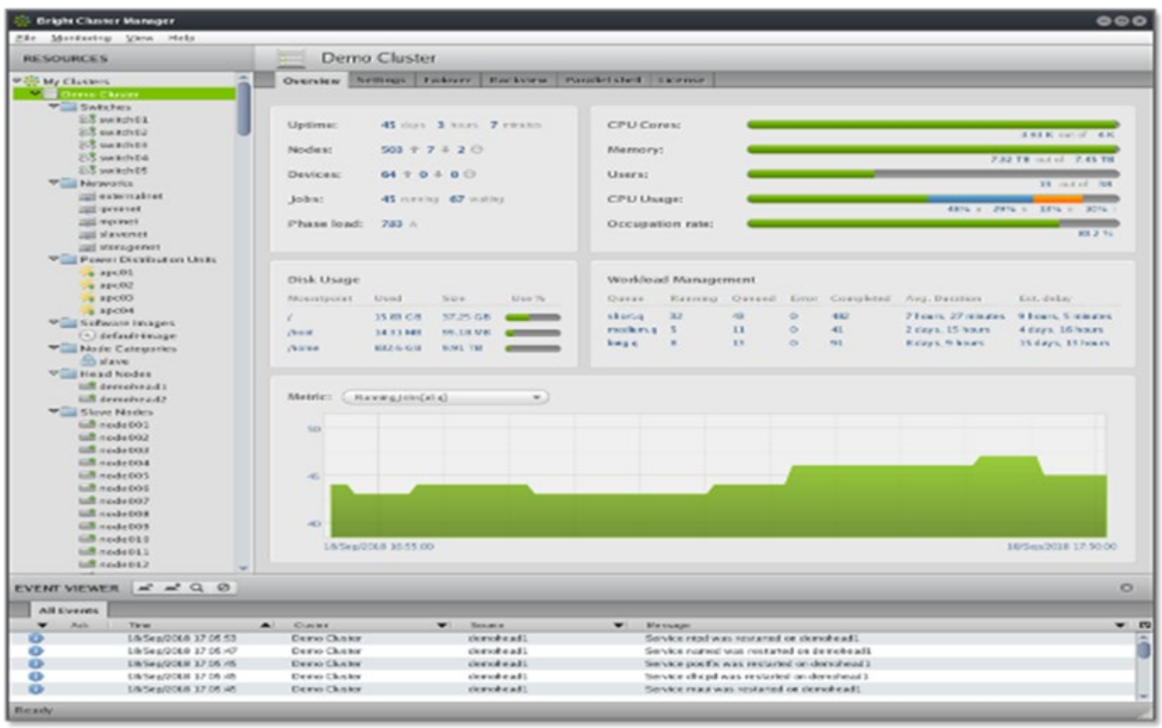


Figure 3. Bright View GUI for Slurm job scheduling and cluster management

Other features of Bright Cluster Manager and Bright Cluster Manager for Data Science include:

- **Modern deep learning environment**—Provides everything that you require to quickly set up a complete cluster-ready and easy-to-manage machine learning environment for data scientists.

- **Choice of machine learning libraries and frameworks**—Includes the most popular machine learning libraries and frameworks such as NVIDIA cuDNN, CUB, CUDA, TensorRT, DyNet, Fastai, JupyterHub, NCCL2, MXNet, PyTorch, Chainer, CNTK, Horovod, Keras, Opencv3, Protobuf3, Torch, TensorFlow, Theano, and XGBoost (for more information, see Documentation on the Bright Computing website).
- **JupyterHub scale-out**—Makes data science easier. Bright provides custom JupyterHub spawners that allow notebooks to be scheduled through the HPC scheduler or Kubernetes.
- **Support for NVIDIA GPU Cloud (NGC)**—Makes NVIDIA NGC deep learning containers easy to use in many ways—run them in containers, through a batch scheduler, on physical nodes, or in a cloud. NGC enhances job metrics, accounting, and reporting functionality, making it possible to determine GPU usage on a per-job basis over time.
- **Simple deployment process**—Eases installation and configuration. After you answer a few questions about your cluster, the Bright software is installed, and the necessary configuration files are created. The software builds your new cluster correctly the first time and every time on bare metal or in the hybrid cloud.
- **Flexible workload scheduling**—Enables administrators to move nodes between Slurm and Kubernetes, based on user demand. You switch nodes from Kubernetes to Slurm and back by moving them between configuration overlays.

  The Cluster Management Shell (CMSH) provides the `movenodes` command to move a node:

```
[demo->configurationoverlay]% movenodes
Name:
          movenodes - Move nodes from one overlay to another
...
Examples:
          movenodes overlayA overlayB node001         Move node001
from overlayA to overlayB
          movenodes overlayA overlayB -n node00[1-3]  Move
node001..node003 from overlayA to overlayB

[demo->configurationoverlay]%
```

  Switching a node from Slurm to Kubernetes and back takes approximately 5 to 10 seconds.

For more information, see Documentation on the Bright Computing website.

Playbooks

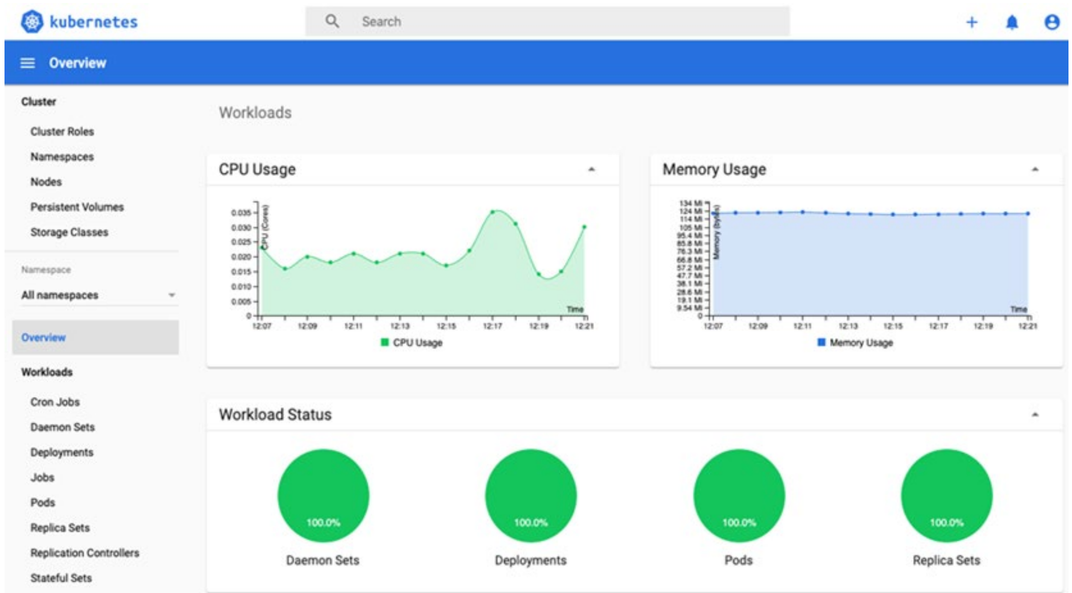


Figure 4. Kubernetes management using Dashboard

Dashboard enables you to manage and monitor nodes, pods, services, persistent volumes, and more from a single interface. You can also run scripts that enable HPC nodes to be converted to Kubernetes nodes and back again for flexible workload scheduling on a single pool of resources. To enable dynamic scheduling, run the partition-switching script from the CLI. Ansible deploys the cluster similarly to the Bright deployment. The cluster will have both Slurm and Kubernetes personalities.

Ansible playbooks provide a convenient way to install the software for the HPC Validated Design for AI and Data Analytics by using PowerEdge servers with factory-installed CentOS images. Playbooks are available on GitHub (https://github.com/dellhpc/omnia).

The following tables provide details about the Ansible playbook environment:

Table 12. Ansible playbook environment capabilities

Capability	Technology
Container runtime with accelerator support	Docker containers
Container orchestration	Kubernetes
System monitoring	Prometheus
CNI-compliant software-defined network (SDN)	Flannel and Calico
Service discovery	CoreDNS
Ingress and proxy	Nginx

Table 13. Ansible playbook environment components – Omnia 1.0

Component	Version
Operating system (management node)	CentOS 8.4
Operating system deployed by Omnia on bare-metal servers	CentOS 7.9 2009 Minimal Edition
Cobbler	3.2.1
Ansible AWX Version	19.0.0
Slurm Workload Manager	20.11.2
Kubernetes Controllers on Management Station	1.21.0
Kubernetes Controllers on Manager and Compute	1.16.7 or 1.19.3
Kubeflow	1.0
Prometheus	2.23.0
Helm	3.0.1

To prepare systems for software deployment by using Ansible, ensure that each server is racked, powered, and networked so that the server can download software packages from the Internet or from a full CentOS mirror.

The HPC Validated Design for AI and Data Analytics assumes that you have two networks:

- A management network that uses the integrated Ethernet for iDRAC-based management
- A high-bandwidth fabric that is built on either 100 GbE or endpoint detection and response (EDR) InfiniBand

Hosting the management network and the high-speed fabric on two separate, private IP spaces is a best practice. As an example, the management network might use 192.168.*x.x*, while the high-speed fabric uses 10.1.*x.x*. Also, assign hostnames to systems. You can assign both names and IP addresses manually or by using Dynamic Host Configuration Protocol (DHCP).

The Ansible playbooks assume that, at a minimum, each node has SSH access on the high-speed fabric. Ansible uses the IP addresses of the high-speed fabric to establish the SDN for the Kubernetes installation.

Ansible uses roles to customize installation on different servers. Each server is given a specific role using the Ansible inventory file. Servers can be master nodes, unaccelerated compute nodes, or accelerated compute nodes.

The master role must list a single node to be used for Slurm and Kubernetes scheduling and orchestration, as well as for managing and monitoring the system. It does not require any accelerators and is not used for compute work.

List unaccelerated (that is, CPU-only) compute nodes in an inventory file. Because GPU enablement is as an add-on to the basic compute node provisioning process, list GPU-accelerated compute nodes in the unaccelerated section of the inventory file.

Here is a sample inventory file:

```
[master]
master

[compute]
compute[000:005]

[gpus]
compute001
compute002

### DO NOT EDIT BELOW THIS LINE ###
[workers:children]
compute
gpus

[cluster:children]
master
workers
```

You can install Ansible on the master node by using the yum package manager (as root):

```
yum install ansible
```

To download Ansible playbooks, go to the following GitHub page:
https://github.com/dellhpc/omnia

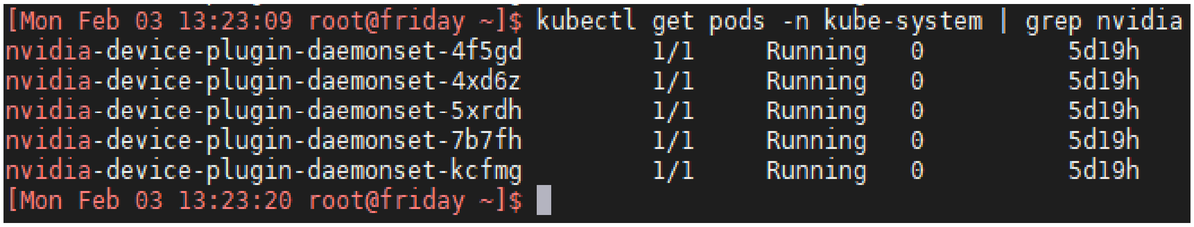
When networking is set up, Ansible is installed on the master node and the inventory file is generated. Run the `build-cluster.yml` playbook to deploy the cluster:

```
ansible-playbook -i host_inventory_file build-cluster.yml
```

The playbook installs all necessary dependencies on the master and compute nodes, and it ensures that nodes are joined to the Kubernetes cluster. The process takes approximately 30 minutes, depending on your Internet connection speed.

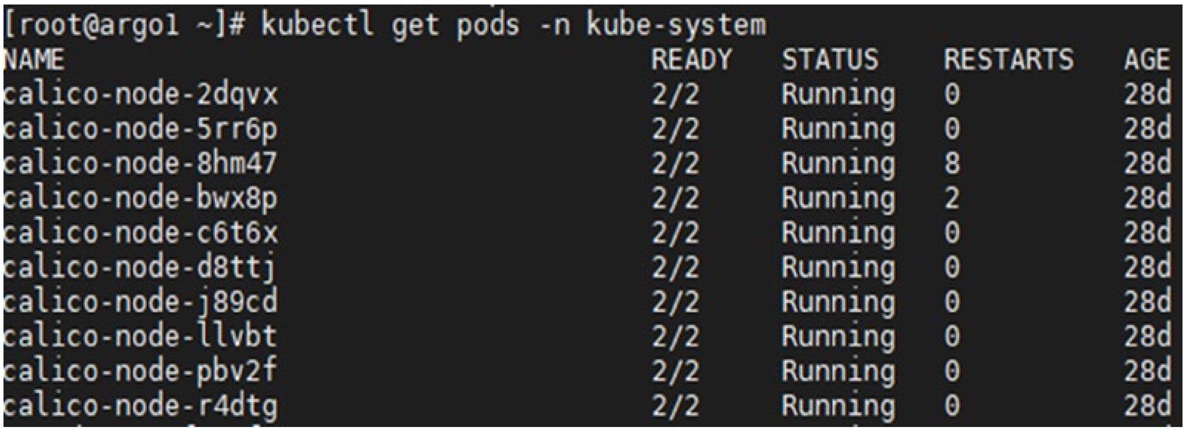
When the cluster is set up, you can install additional applications on the Kubernetes partition, by using Helm, and on the Slurm partition, by using yum.

The following examples show validation that the services are running on the cluster when you use this architecture:



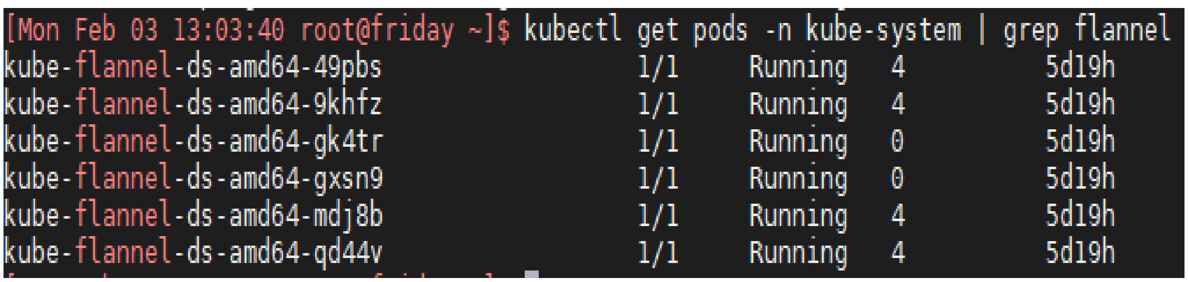
```
[Mon Feb 03 13:23:09 root@friday ~]$ kubectl get pods -n kube-system | grep nvidia
nvidia-device-plugin-daemonset-4f5gd       1/1     Running    0          5d19h
nvidia-device-plugin-daemonset-4xd6z       1/1     Running    0          5d19h
nvidia-device-plugin-daemonset-5xrdh       1/1     Running    0          5d19h
nvidia-device-plugin-daemonset-7b7fh       1/1     Running    0          5d19h
nvidia-device-plugin-daemonset-kcfmg       1/1     Running    0          5d19h
[Mon Feb 03 13:23:20 root@friday ~]$
```

Figure 5. Container runtime with accelerator support



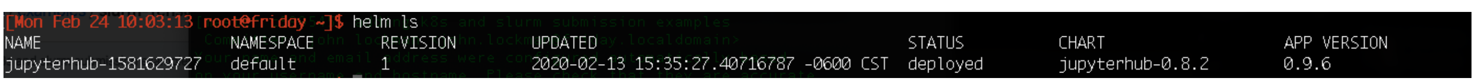
```
[root@argo1 ~]# kubectl get pods -n kube-system
NAME                    READY   STATUS    RESTARTS   AGE
calico-node-2dqvx       2/2     Running   0          28d
calico-node-5rr6p       2/2     Running   0          28d
calico-node-8hm47       2/2     Running   8          28d
calico-node-bwx8p       2/2     Running   2          28d
calico-node-c6t6x       2/2     Running   0          28d
calico-node-d8ttj       2/2     Running   0          28d
calico-node-j89cd       2/2     Running   0          28d
calico-node-llvbt       2/2     Running   0          28d
calico-node-pbv2f       2/2     Running   0          28d
calico-node-r4dtg       2/2     Running   0          28d
```

Figure 6. SDN with Flannel/Calico



```
[Mon Feb 03 13:03:40 root@friday ~]$ kubectl get pods -n kube-system | grep flannel
kube-flannel-ds-amd64-49pbs       1/1     Running   4          5d19h
kube-flannel-ds-amd64-9khfz       1/1     Running   4          5d19h
kube-flannel-ds-amd64-gk4tr       1/1     Running   0          5d19h
kube-flannel-ds-amd64-gxsn9       1/1     Running   0          5d19h
kube-flannel-ds-amd64-mdj8b       1/1     Running   4          5d19h
kube-flannel-ds-amd64-qd44v       1/1     Running   4          5d19h
```

Figure 7. Flannel SDN running on an Ansible-deployed system



```
[Mon Feb 24 10:03:13 root@friday ~]$ helm ls
NAME                     NAMESPACE   REVISION   UPDATED                                   STATUS     CHART               APP VERSION
jupyterhub-1581629727    default     1          2020-02-13 15:35:27.40716787 -0600 CST    deployed   jupyterhub-0.8.2    0.9.6
```

Figure 8. Helm chart configuration

Data Science Laboratory

The Data Science Laboratory, developed by Dell Technologies, simplifies your work by allowing model implementation, training, and inference tests to be run in containerized JupyterLab or directly from the Linux terminal. Containerized JupyterLab enables data scientists to use notebooks for interactive development in both Python and R, as shown in the following figure:

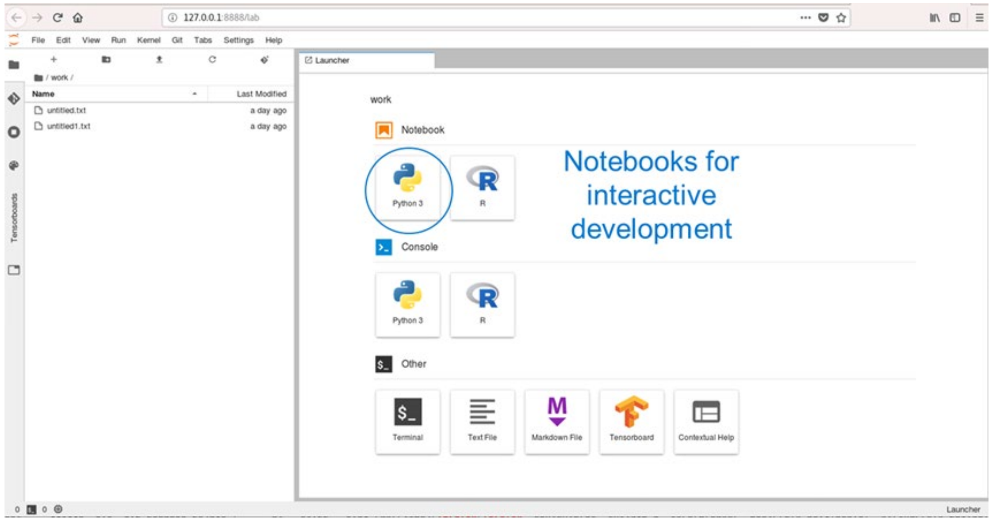


Figure 9. JupyterLab interactive notebooks

Data Science Laboratory integrates TensorBoard for monitoring deep learning training jobs, as shown in the following figure:

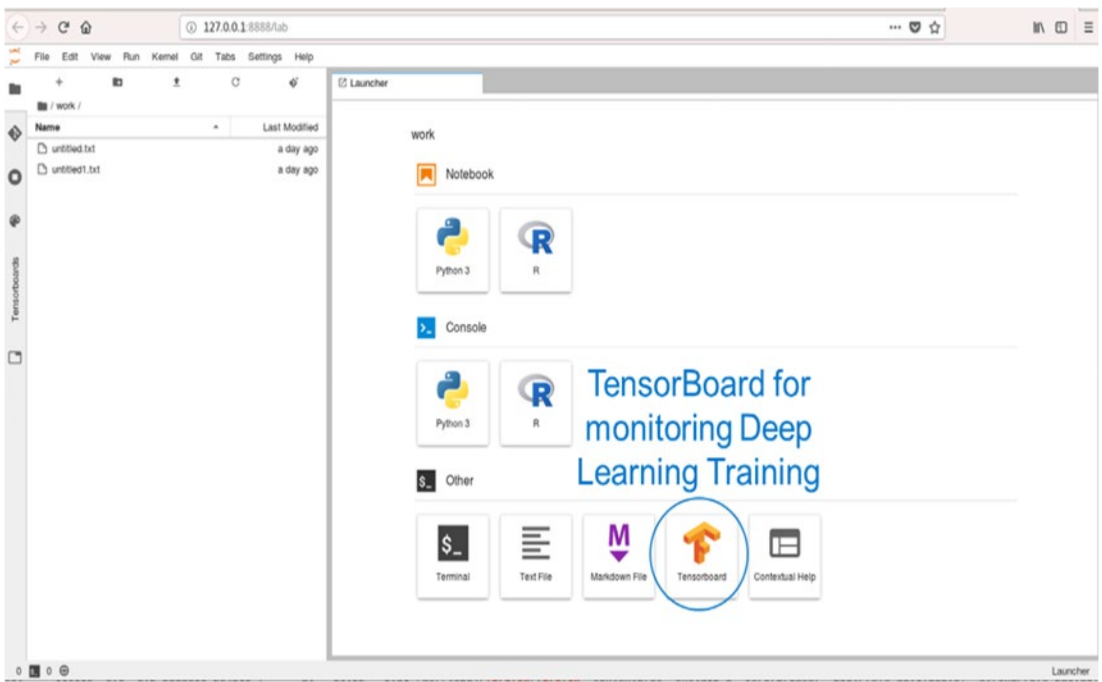


Figure 10. Data Science Laboratory TensorBoard

TensorBoard enables the visualization and tools that are needed for deep learning experimentation. TensorBoard enables features that are used as a visualization toolkit. You can track and visualize metrics for each job, viewing histograms of weights, biases, or other tensors as they change over time.

Data Science Laboratory provides text and markdown editors to help you build documentation on each project. Also, it provides integrated git management for source control, as shown in the following figure:

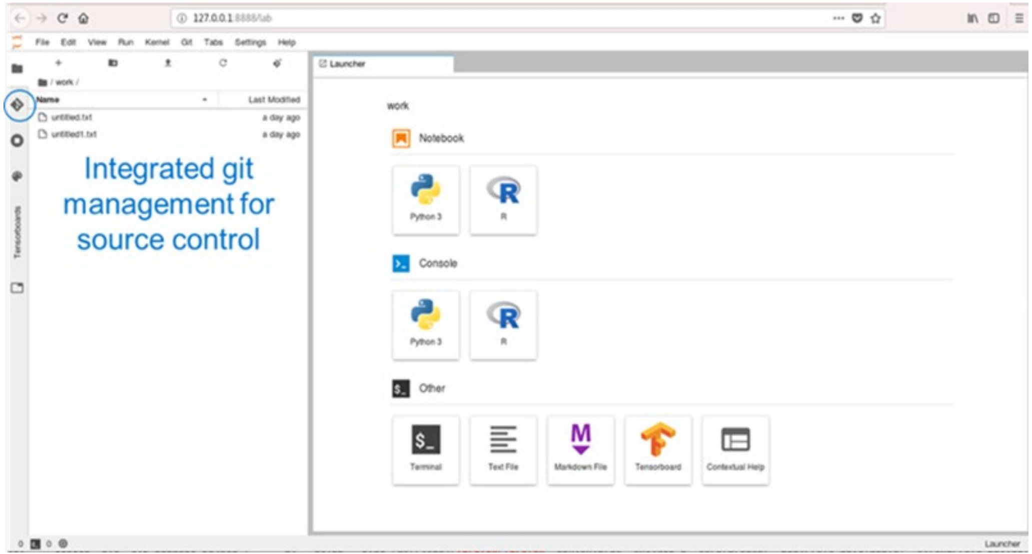


Figure 11. Integrated git management for source control

Source control integration with a git source control repository allows developers to save and manage multiple versions of a model from a nonproduction instance. Data scientists can track their models throughout the model development cycle and maintain version control.

Summary

The HPC Validated Design for AI and Data Analytics is a converged infrastructure solution. With this solution, you can design a unified architecture with multipurpose, balanced nodes to support all your HPC, AI, and data analytics workloads.

The solution employs a building-block approach—supporting multiple servers, networking, storage, and software environment options—so you can design the architecture that best meets your needs. Deployment options include an enterprise option that uses Bright Computing software and an open-source option that runs cluster management tools, Dell Technologies Data Science Laboratory, and Ansible playbooks for Kubernetes and Docker.