SANDIA REPORT

SAND2023-10451
Printed September 2023

Sandia National Laboratories

Glinda: An HPDA Cluster with Ampere A100 GPUs and BlueField-2 VPI SmartNICs

Craig Ulmer, Jerry Friesen, and Joseph Kenny

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

Glinda: An HPDA Cluster with Ampere A100 GPUs and BlueField-2 VPI SmartNICs

Craig Ulmer
Dept. 08753, Scalable Modeling and Analysis
Sandia National Laboratories
cdulmer@sandia.gov

Jerry Friesen
Dept. 08753, Scalable Modeling and Analysis
Sandia National Laboratories
jafries@sandia.gov

Joseph Kenny
Dept. 08753, Scalable Modeling and Analysis
Sandia National Laboratories
jpkenny@sandia.gov

SAND2023-10451

ABSTRACT

Sandia National Laboratories relies on *high-performance data analytics* (HPDA) platforms to solve data-intensive problems in a variety of national security mission spaces. In a 2021 survey of HPDA users at Sandia, data scientists confirmed that their workloads had largely shifted from CPUs to GPUs and indicated that there was a growing need for a broader range of GPU capabilities at Sandia. While the multi-GPU DGX systems that Sandia employs are essential for large-scale training runs, researchers noted that there was also a need for a pool of single-GPU compute nodes where users could iterate on smaller-scale problems and refine their algorithms.

In response to this need, Sandia procured a new 126-node HPDA research cluster named *Glinda* at the end of FY2021. A Glinda compute node features a single-socket, 32-core, AMD Zen3 processor with 512GB of DRAM and an NVIDIA A100 GPU with 40GB of HBM2 memory. Nodes connect to a 100Gb/s InfiniBand fabric through an NVIDIA BlueField-2 VPI SmartNIC. The SmartNIC includes eight Arm A72 processor cores and 16GB of DRAM that network researchers can use to offload HPDA services. The Glinda cluster is adjacent to the existing Kahuna HPDA cluster and shares its storage and administrative resources.

This report summarizes our experiences in procuring, installing, and maintaining the Glinda cluster during the first two years of its service. The intent of this document is twofold. First, we aim to help other system architects make better-informed decisions about deploying HPDA systems with GPUs and SmartNICs. This report lists challenges we had to overcome to bring the system to a working state and includes practical information about incorporating SmartNICs into the computing environment. Second, we provide detailed platform information about Glinda's architecture to help Glinda's users make better use of the hardware.

Acknowledgement

We gratefully acknowledge that a number of people within Sandia helped us to bring Glinda to life. Philip Kegelmeyer, Noël Nachtigal, Todd S. Jones, Rudy Garcia, Ron Oldfield, and Karim Mahrous provided valuable insight into upcoming programmatic needs they envisioned in their respective mission spaces. Tom Klitsner and the Mission Computing Council worked with us to ensure the Glinda platform fit properly into the overall Sandia platform portfolio. Robert Clay, Jeremiah Wilke, and Tricia Gharagozloo championed the work and provided oversight during different periods of the procurement.

HPDA data scientists such as Kasimir Gabert, Jon Bisila, and Stefan Seritan performed early evaluations of the hardware and provided information on how we could help data scientists become more productive on the platform. Sam Knight helped with the initial deployment of the system and did extensive work with system services to allow Kahuna and Glinda to be managed through the same infrastructure. Gavin Baker extended Kahuna's Ceph and NFS capabilities to allow users to transparently access data on both platforms.

We also acknowledge that several people from outside of Sandia helped make Glinda's deployment successful. Once the procurement contract was awarded, Bart Willems from Atipa Technologies helped us resolve multiple technical issues and ensured the system met our operational requirements. Kurt Rago at NVIDIA helped us troubleshoot multiple issues related to the BlueField-2 card and helped us get the InfiniBand environment into a working state.

CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Sandia National Laboratories employs multiple types of computing platforms to serve the different needs of its national security missions. While high-performance computing (HPC) platforms that are optimized for compute-bound problems dominate the landscape, the need to analyze massive datasets has driven researchers to develop high-performance data analytics (HPDA) platforms that are optimized for *data-intensive* problems. In 2015 Sandia established the *Kahuna* HPDA cluster to help analysts from different mission spaces store and process large, unclassified datasets. While this platform has been extremely successful, it lacks GPU resources and is therefore insufficient for many data science workloads. As a means of remedying this problem, we designed and procured a new system named *Glinda*. In contrast to other GPU-enabled systems at Sandia such as the Synapse DGX platform, Glinda follows a "thin-slice" design philosophy, where the system is composed of many nodes, each with a single-socket CPU and a single-card GPU. The intent of this design is to increase data scientist productivity by making it easier to acquire a GPU-enabled node to develop prototypes. We expect that users will leverage Glinda to develop and debug their work at small to medium scale and then move on to Sandia's larger DGX platforms to run at scale.

This report captures information about the creation of Glinda and documents our experiences during the first two years of its use. This section provides background information about Sandia's computing environments and summarizes lessons learned from the Kahuna HPDA platform. We then discuss our motivation for procuring a new system and list the design requirements that were provided to vendors for the Glinda platform.

1.1. Computing Needs at Sandia National Laboratories

Sandia National Laboratories performs work in a wide variety of mission areas that have different computational needs. Sandia's *Institutional Computing* effort currently deploys and supports at least four types of computing platforms:

High-Performance Computing (HPC): The bulk of computing infrastructure at Sandia is optimized to allow large-scale, MPI-based simulation and analysis tools to run as efficiently as possible. These platforms employ thousands of compute-optimized servers that communicate through a low-latency communication fabric. Due to the large number of compute nodes used in individual jobs, HPC platforms generally perform batch processing where jobs are queued up and run as compute nodes become available.

High-Performance Data Analytics (HPDA): Many mission problems at Sandia center around collecting and analyzing large amounts of data. Increased demand for quicker analytics has

resulted in the construction of high-performance data analytics (HPDA) platforms that are optimized for solving data-intensive problems. These systems are typically designed to maximize the capacity and performance of storage, and rely on data-parallel compute resources to accelerate calculations. Compute nodes are often equipped with NVMe storage to enable users to mitigate I/O overheads in their workflows. HPDA platforms are typically much smaller than HPC platforms and focus on user productivity. While some jobs involve batch processing, many are run through interactive sessions where a user interrogates a dataset through high-level languages or processing frameworks.

Cloud Computing: Cloud Computing platforms provide a general solution for users that need an easy way to define, standup, and run custom infrastructure for both long- and short-duration tasks. Clouds are especially useful for groups that need to run reoccurring tasks (e.g., nightly report generation, continuous integration and testing, etc.). Clouds also offer a place for developers to build custom software stacks and are frequently used to prototype information systems locally before they are transitioned to a production environment.

Emulytics: Emulytics platforms provide a place for security researchers to provision bare-metal hardware and conduct different types of network experiments. While similar in nature to Cloud Computing, Emulytics platforms are designed to provide researchers with dedicated access to bare-metal resources in order to guarantee experiments operate at high fidelity and are faithful to real-world constraints.

1.2. Kahuna: An Institutional Computing HPDA Platform

In 2015 Sandia's Mission Computing Council (MCC) conducted a survey of different computing users across the enterprise and found that a number of centers were fielding and maintaining small cluster computers to support specific mission needs. While these systems were critical to the success of different efforts, they burdened centers with operational overheads and created inconsistencies between platforms for users. As a means of remedying these problems, the MCC funded two institutional research clusters, Ray and Kahuna, to help users refocus their efforts on data analytics. While Ray provided an early cloud resource for long-running data engineering projects, Kahuna was constructed to serve as a flexible, data analytics platform that could solve a variety of problems using a traditional SLURM [1] job management system.

The initial Kahuna procurement included 120 compute nodes, 8 Ceph nodes with 1.5PB of networked storage, and 3 login nodes. Each compute node contained 2x14 cores, 256GB of DRAM, 700GB of local NVME, and a 56Gb/s InfiniBand NIC. In later years the Ceph nodes were expanded to house 3.2 PB of raw storage on 13 OSD nodes. Ten nodes with NVIDIA K80 GPUs were also added to the system to allow users to experiment with accelerators in deep learning and machine learning workloads. Based on Kahuna's eight years of operation, we attribute the success of the platform to several characteristics:

Flexibility: Users like that Kahnua nodes can readily be adapted to perform different workloads in a way that is fair to all users. The well-known SLURM reservation system does not require users to embrace a particular programming paradigm the way that some big-data

frameworks do. Users simply request an allocation of nodes for a period of time, launch services and applications on the nodes as needed, and then allow SLURM to cleanup the allocation when the reservation expires. In addition to traditional MPI jobs, Kahuna users routinely launch userspace applications and frameworks in their jobs, including Spark [2], Dask [3], ElasticSearch [4], TensorFlow [5], and Jupyter Notebooks [6].

Accessibility: Given that many of Kahuna's users are *not* HPC developers, we explored alternate programming interfaces into the cluster that could help users become more productive. The cluster's JupyterHub interface has been particularly successful, as it provides a means for users to connect to the cluster through a web browser and interactively run Jupyter notebooks on cluster nodes. The cluster also provides a small number of front-end workstations that allow users to obtain a full GNOME desktop environment in a browser window via FastX [7]. These workstations enable users to launch long-running Jupyter sessions that do not get reset when the user disconnects.

Abundant Storage: The Kahuna platform is connected to a multi-petabyte Ceph storage system that is used for hosting project data as well as datasets that are of general interest to the HPDA community. Abundant and easy-to-use storage allows users to stockpile and share datasets that might otherwise be lost over time.

Research Support: One of the unique aspects of Kahuna is that most of the administrators also use the platform to conduct work in other research projects. This trait provides the administrators with extra motivation to keep the platform stable, performant, and secure. Many of Kahuna's improvements have come about from the administrators working with users to find better ways of doing different types of work on the platform,.

1.3. Next Generation HPDA Platform

In 2021 we recognized that while Kahuna was still an active platform, its lack of GPU accelerators limited how much impact it could have on modern deep learning and machine learning workloads. As a means of evaluating how we could better serve these communities, we conducted interviews with key program leads from different mission spaces and gathered information about what GPU capabilities researchers would like to see in a next-generation HPDA platform.

One observation from these discussions is that data science researchers expressed an interest in having two types of GPU platforms at the labs. First, users require a high-end, multi-GPU plaform where a large number of GPUs can be combined to work on a single, massive problem. The Synapse platform provides this capability at Sandia, and users had overwhelmingly positive responses about how this platform impacts their work. Second, researchers expressed a need for an HPDA platform composed of many single-GPU nodes that could be used for solving smaller problems. The researchers explained that many tasks in their day-to-day work require some GPU acceleration, but do not offer the scale to make the overhead of working on a shared, multi-GPU system worthwhile. Supplementing an existing HPDA platform with GPUs provides a space for data scientists to more easily prototype and debug new algorithms before running at scale on a capability system.

Based on the information gathered from different stakeholders, we proposed a new Institutional Computing procurement to the Mission Computing Council. The key architectural points of this proposal are as follows.

Single GPU per Node: The overall goal of the platform is to provide users with a large pool of single-GPU nodes that they can use in their day-to-day development and small scale work. Based on Kahuna's usage we estimate that approximately 100-150 compute nodes is an appropriate size for Sandia's unclassified HPDA work.

Single-Socket CPU per Node: The intent of this platform is to shift computational workloads from the CPUs to a GPU accelerator. Given that AMD and Intel now offer server processors that have 32 physical cores, we estimate that a single-socket motherboard will be sufficient for most workloads that run on these nodes.

NVIDIA Ampere A100 GPU: A performance evaluation completed in February of 2021 [8] found that the NVIDIA Ampere A100 GPU provided significant performance in a number of HPC and HPDA tasks. While a great deal of GPU attention in the HPC space is currently focused on AMD GPUs, users are comfortable with NVIDIA's CUDA libraries and can leverage the hardware without significant effort.

PCIe Gen4: A server node with a GPU accelerator and an HPC NIC must be equipped with a high-bandwidth peripheral device interconnect in order to rapidly move data between host and card memory. PCIe Gen4 servers are now available from vendors and effectively double the amount of I/O throughput the host can support compared to Gen3 systems.

100Gb/s InfiniBand Network: InfiniBand provides a cost-effective way to move large amounts of data between cluster resources at 100Gb/s speeds. InfiniBand is well understood at Sandia and is the backbone for both HPC and storage in the Kahuna environment.

BlueField-2 SmartNICs: Multiple researchers are investigating how programmable network interface cards (or SmartNICs) can be used to offload different tasks in HPC platforms. The new BlueField-2 VPI card provides an InfiniBand NIC that compute nodes can use for their communication, and presents opportunities for researchers to evaluate how SmartNICs can be integrated into the HPC environment. To our knowledge, this is one of the first SmartNIC deployments that is larger than 100 nodes.

Expandability: While the focus of the current system is on a single-GPU architecture, we require that additional computational accelerator bays be available in the architecture for later expansion. These open bays provide us with the opportunity to consolidate existing cards or purchase alternate accelerators later in the system's lifetime.

1.4. Procurement System Specifications

A Statement of Work was written to document the new system's requirement to potential vendors [9]. The key requirements in this document are as follows:

The cluster will consist of seven scalable compute units (SCUs):

Racks	Each SCU rack will contain 18 compute nodes and three network fabrics
	Racks must be 42U x 24”W x 45”D
	Each rack must contain two Raritan 3PH, 400V AC, 60A PDUs

The cluster will employ three separate networks:

Network	• A 100Gb/s HDR InfiniBand network for IPC and I/O • A 25Gb/s Ethernet boot network for user access • A 1Gb/s IPMI network for management

Each compute node will have:

CPU	A single-socket AMD EPYC Processor • Baseline: EPYC 7502P 32-cores, 2.5GHz • Option 1: EPYC 7543P 32-cores, 2.8GHz
Memory	512GB of DDR4 3200MHz
I/O	PCIe Gen4 with four x16 expansion slots
Disk	1TB NVMe PCIe Gen4 solid state disk
GPU	NVIDIA A100 40GB PCIe Gen4 GPU Accelerator
Network	NVIDIA MBF2H516A-EEEOT BlueField-2 SmartNIC P-Series DPU
	A 25Gb/s Ethernet card
Power	Dual-circuit Power supply, Capable of 1600W 80+ Platinum efficiency

1.5. Contract Award

Following the review of all proposals received from vendors, the contract to build the system was awarded to Atipa Technologies. Atipa’s proposal met all of the requirements. The following details provide additional information about their implementation:

AMD EPYC 7543P Option: Atipa was able to offer both the 7502P (Zen2) and 7543P (Zen3) processors for the procurement deadline. We selected the 7543P because it offers a clock speed that is roughly 12% faster than the 7502P.

Atipa Procyon SE218-8G4 Nodes: Atipa’s compute nodes are 2U servers for a single-socket AMD EPYC processors. As seen in Figure 1-1, a compute node provides three PCIe gen4 x16 bays for double-width cards in the front and two x16 cards in the back. The node uses a 1600W 80+ Platinum efficiency power supply with dual circuits. Atipa’s Procyon is based on the Gigabyte G252-Z11 server.

ConnectX-4 25Gb/s OCP 3.0 Card: The Procyon node does not include built-in networking ports beyond the IPMI management ports. Instead, the node is equipped with a ConnectX-4 Ethernet card (Figure 1-2, right) that connects to the system through the new OCP 3.0 standard. While we had not previously deployed systems with the OCP interface, we have a positive attitude towards this approach as it replaces proprietary “add-on-module” (AoM) interfaces often found in server systems. One of the additional benefits of the OCP interface is that cards can be replaced without having to physically open up the chassis.

Figure 1-1. The front of a Glinda node can house up to three GPUs

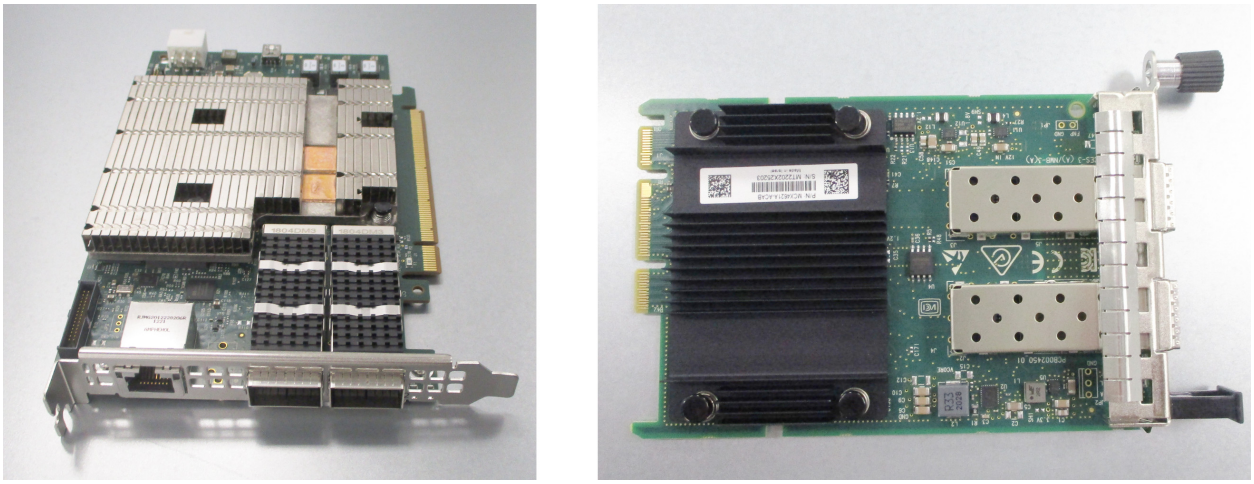


Figure 1-2. Glinda compute nodes include a BlueField-2 SmartNIC (left) for 100Gb/s InfiniBand networking and a 25Gb/s OCP Ethernet card (right) for general traffic

1.6. Glinda and the Great Book of Records

The name *Glinda* was selected for the cluster as a nod to Glinda the Good Witch from *The Wizard of Oz*. Glinda's ever-updating Great Book of Records resonated with the HPDA team. Prior systems in the California data center have loosely followed a mystical theme. The Kahuna name was selected because the system needed to be flexible enough that it could be adapted to solve a variety of problems for many people. In Hawaiian culture, a Kahuna is a shaman that is an expert in many fields. The Carnac cluster needed to be a monumental gathering place for different cyber security researchers to meet. Carnac Stones are an ancient megalithic site in France, where large stones surround a hallowed ground.

2. GLINDA NODE ARCHITECTURE

A Glinda compute node is an Atipa Procyon SE218-8G4 2U server. As illustrated in Figure 2-1, a compute node provides a single-socket EPYC processor, 512GB of memory, an NVMe disk, a BlueField-2 DPU, and an NVIDIA A100 GPU. This section lists the details for each of these components and outlines opportunities for upgrading the hardware in later years.

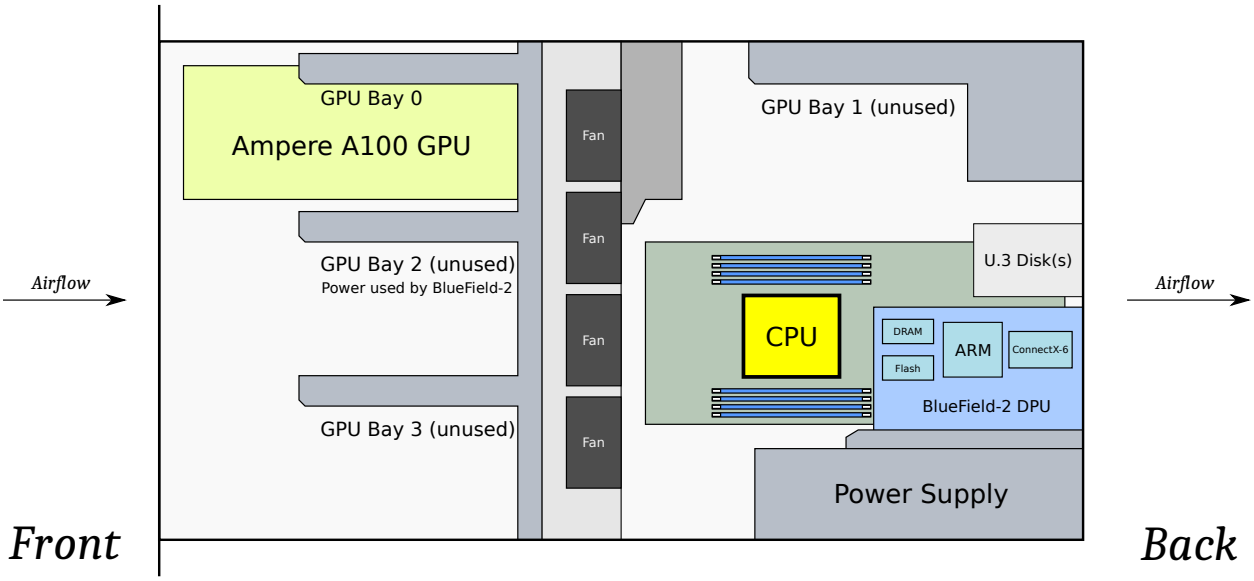


Figure 2-1. Overhead view of a Glinda compute node

2.1. Base System Specifications

The Atipa Procyon SE218-8G4 compute node is a 2U server that is based on a Gigabyte G242-Z11-00 motherboard. Basic system properties are summarized in Table 2-1.

Table 2-1. Glinda node base specifications

Server	Atipa Procyon SE218-8G4
Motherboard	Gigabyte G242-Z11-00
Power Supply	Lite-On Power 1600W (PS-2162-6L2) 80+ Platinum
Base Network	Mellanox ConnectX-4 25Gb/s Ethernet (OCP 3.0)

2.2. AMD EPYC 7543P Specifications

A Glinda compute node features a single AMD EPYC 7543P processor with 32-physical cores. This processor is based on AMD's third generation of the Zen architecture (i.e., EPYC 7003) and has the specifications listed in Table 2-2.

Table 2-2. Glinda processor details

Part Number	AMD EPYC 7543P
Physical Cores	32
Total Threads	64
Base Clock	2.8GHz
Max Boost Clock	3.7GHz
L1 I-Cache	1MB (32 x 32KB 8-way set associative)
L1 D-Cache	1MB (32 x 32KB 8-way set associative)
L2 Cache	16MB (32 x 512KB 8-way set associative)
L3 Cache	256MB (8 x 32MB 16-way set associative)
Memory Channels	8
Memory Type	DDR4-3200 w/ ECC
Memory Bandwidth	204.8GB/s
PCIe	PCIe 4.0 x128
Socket	SP3, LGA-4094
TDP	225W

2.3. Memory Specifications

Each Glinda node is populated with 512GB of DRAM. The Gigabyte G252-Z11 motherboard provides 8 DDR4 DIMM slots that are populated with SK Hynix HMAA8GR7AJR4N-XN parts. An individual DIMM has the datasheet properties listed in Table 2-3.

Table 2-3. Glinda memory details

DIMM Capacity	64GB
Speed	DDR4 3200MT/s
CAS Latency (CL)	22 cycles
RAS to CAS Delay (tRCD)	22 cycles
Row Precharge Time (rTP)	22 cycles
Max Temperature	85°C
Form Factor	RDIMM

2.4. NVIDIA Ampere A100 40GB PCIe GPU Specifications

Each Glinda node contains a single NVIDIA A100 GPU [10] for accelerating data-intensive operations. NVIDIA offers multiple variations of the A100 that differ based on the host interface (PCIe or SXM) and memory capacity (40GB or 80 GB). Glinda employs the PCIe version of the card with 40GB of HBM2 memory. The A100 40GB PCIe card's properties [11] are summarized in Table 2-4.

Table 2-4. Glinda GPU details

Part Number	NVIDIA A100-PCIE-40GB
GPU Chip	Ampere GA100
Streaming Multiprocessors (SMs)	108
CUDA Cores	6,912
Base Clock	765MHz
Boost Clock	1.410GHz
L1 Cache	192KB per SM
L2 Cache	40MB
Memory	40GB HBM2
Memory Bandwidth	1,555GB/s
Peak FP64	9.7 TFLOPS
Peak FP32	19.5 TFLOPS
Peak FP64 Tensor Core	19.5 TFLOPS
Peak FP16 Tensor Core	312 TFLOPS
Peak INT8 Tensor Core	624 TOPS
PCIe	PCIe 4.0 x16 (32GB/s)
TDP	250W
Max Temperature	50°C

The A100 cards consume a significant amount of power and require an additional 8-pin power supply connection to function. Each A100 card includes an NVLink interface to allow two neighboring cards to share data directly using the NVLink protocol. Unfortunately, the GPU bays of the Glinda nodes are oriented in such a way that a bridge connector cannot be installed (i.e., GPUs in Glinda's front bays are parallel to the motherboard. Bridge connectors would require the cards to be perpendicular to the motherboard.) [12].

2.5. NVIDIA BlueField-2 DPU Dual-Port VPI Specifications

Each Glinda node is equipped with an NVIDIA BlueField-2 DPU for high-speed communication and in-network data processing. The DPU occupies a single PCIe x16 slot in the back of the node and requires a supplemental 6-pin auxiliary power connector to function. NVIDIA's DPU specifications [13] state that the host must provide "a minimum of 75W or greater system power supply through the PCIe x16 interface, and an additional 75W through he supplementary 6-pin ATX power supply connector." We were initially concerned that this statement implies the card may consume 150W of power. However, NVIDIA has indicated that this is not the case, and that in additional documentation the maximum power usage of a 16GB P-series card is 63W. NVIDIA's documentation makes it clear that substantial airflow through the chassis is required to properly cool the DPU. Table 2-5 lists key characteristics of the BlueField-2 DPU.

Table 2-5. Glinda SmartNIC details

Part Number	MBF2H516A-EEEOT (P-Series)
Physical Cores	8 Armv8 A72 (64-bit)
Total Threads	8 (1 per core)
Core Clock	2.75GHz
L1 I-Cache	384KB
L1 D-Cache	256KB
L2 Cache	4MB (1MB per 2 cores)
L3 Cache	6MB
DDR Memory	16GB DDR4-3200 w/ ECC
eMMC Storage	64GB
Network Interface	ConnectX-6 Dx
Network Ports	Two 100Gb/s Ports (InfiniBand or Ethernet)
Management	1Gb/s Ethernet
Accelerators	Encryption, Compression, RegEx
PCIe	PCIe 4.0 x16 (32GB/s)
TDP	63W
Max Operating Temperature	105°C

2.6. NVMe Storage Specifications

Each Glinda node provides a single 960GB NVMe device for users to stage intermediate results at the local node. Table 2-6 lists the key statistics for the device, as reported in a KIOXIA product brief [14].

Table 2-6. Glinda storage details

Vendor	KIOXIA
Part Number	KCD6XLUL960G
Capacity	960GB
Flash Memory Type	BiCS 96-layer 3D TLC
Stated Max Data Rate	5.8GB/s (Read), 1.3GB/s (Write)
Stated 4KB Random IOPS	700K (Read), 30K (Write)
Stated Power Consumption	13W (Active), 5W (Ready)
PCIe	PCIe 4.0 x4 (8GB/s)
NVMe	1.4
Max Temperature	70°C

2.7. Processor Comparison

Table 2-7 summarizes the key characteristics of the three types of processors available in a Glinda node. From these numbers we see that while the clock speeds and memory interfaces are similar between the SmartNIC and the host processor, the host features more advanced processor cores, more (4x) cores, more (40x) data cache, and substantially larger (8x) and faster (8x) memory. As such we expect the host to have an order of magnitude more performance than the SmartNIC for compute tasks. Similarly, we see that the GPU has a distinct computational advantage over the host CPU as the GPU has more (3x) "cores", substantially more parallelism per core, and much more (7x) memory bandwidth than the host CPU.

Table 2-7. Glinda processor comparison

Feature	**SmartNIC**	**Host CPU**	**GPU**
Processor	Arm A72	EPYC 7543P	Ampere GA100
Physical Cores	8	32	108 SMs
Base Clock	2.75GHz	2.8GHz	765MHz
L1 I-Cache	384KB	1MB	-
L1 D-Cache	256KB	1MB	192KB/SM
L2 Cache	4MB	16MB	40MB
L3 Cache	6MB	256MB	-
Memory Capacity	16GB	512GB	40GB
Memory Channels	1	8	-
Memory Type	DDR4-3200	DDR4-3200	HBM2
Memory Bandwidth	25GB/s	204GB/s	1,555GB/s
TDP	63W	225W	250W

2.8. Upgrade Opportunities

While the Glinda nodes are expected to be sufficient for our HPDA users' needs for several years, there are multiple ways that the nodes could be upgraded as the system ages. The two open and powered PCIe x16 bays in the node provide an opportunity to supplement a node with additional GPUs. The current power supply and cooling should be sufficient for adding at least one and possibly two A100-generation cards. While these cards cannot take advantage of NVLink bridging due to the physical arrangement of the cards, the system does provide dedicated PCIe 4.0 x16 lanes to each slot. It is unlikely that Hopper-generation cards will be compatible with the Glinda nodes. In NVIDIA's current list of certified data center servers[1], the Atipa Procyon SE218-8G4 does not include the H100 as a supported NVIDIA GPU. Even if an H100 does work in a Glinda node, it will operate at reduced performance because (1) the H100 has a 100W increase in TDP and would throttle performance to meet Glinda's lower power envelope and (2) the H100 would only be able to exchange data with the host at half of its potential bandwidth due to the use of PCIe 4.0 instead of 5.0.

In terms of network improvements, the second 100Gb/s network port of the BlueField-2 is not currently connected due to the expense of cables and network switches. Glinda's InfiniBand network infrastructure could be improved to allow the second port to be connected. This improvement would be valuable for researchers that are investigating how multiple GPU-nodes can be leveraged to scale data-intensive problems. Alternatively, the second network port could be used to boost a Glinda node's Ethernet performance from 25Gb/s to 100Gb/s. We have confirmed that the BlueField-2 can be configured to run one port as InfiniBand and the other as Ethernet. The benefit of this upgrade is that it should allow the node to improve the rate at which users can fetch data from the platform's Ceph storage server.

NVMe storage for the Glinda nodes could be upgraded with minimal cost and effort. The second NVMe drive bay is currently unoccupied and can be accessed without opening the chassis. The primary benefit of placing a second drive in a node is that it would double the bandwidth and capacity of local storage. Users that cache large datasets on their compute node to avoid network overheads would see modest I/O gains without having to change their existing workflows. Another option recently proposed by SmartNIC researchers is to use a second NVMe drive as storage for an ephemeral parallel file system that is maintained by the SmartNICs. Early work has confirmed that the SmartNIC can access the host's NVMe device without involving the host via the NVMe-over-Fabric offload capability provided by the ConnectX hardware. Per-job ephemeral storage may allow users to stage distributed data in compute nodes efficiently.

Other upgrades to the Glinda node architecture are not expected to offer a high return on investment. All DIMM sockets are currently filled with high-capacity DIMMs. The current CPUs are a reasonable speed and are expensive to replace. The BlueField-2 card could be upgraded to the BlueField-3, but there are not enough researchers that take advantage of these cards to justify the disruption. Finally, the Glinda node could house other PCIe accelerator cards. However, doing so complicates the software stack and must be mindful of the power budget of the node.

[1] `https://docs.nvidia.com/certification-programs/nvidia-certified-systems/index.html`

3. PHYSICAL INSTALLATION IN THE 902 DATA CENTER

The Glinda cluster was delivered to Sandia at the end of FY21 and installed into the new Building 902 Data Center (Figure 3-1). This section provides an overview of this building, and describes the physical environment where Glinda operates.

Figure 3-1. Sandia California's new 902 Data Center

3.1. Building 902 Data Center

In calendar year 2020, Sandia National Laboratories, California constructed building 902 to serve as a new data center for the site. Essentially a warehouse with power and cooling, the building is dedicated to equipment, not humans. The largest part of the building is allocated to server deployment, with supporting power distribution and dense fiber connectivity to the rest of the site and elsewhere. A staging area is available for equipment entrance/exit, and a small network

monitoring and console function room is present. The data center itself is a vault-type room (VTR), approved for storage of both classified and unclassified equipment.

The new data center follows traditional computer deployment: servers are mounted in rows of racks on a concrete slab floor with overhead power and above rack cabling. Computer heat dissipation is performed through cold air cooling and hot aisle air containment. Power is distributed via 1200A/480V StarLine Track Busways, with a single dedicated circuit for each row in the data center. Each HPDA system rack contains two Raritan PDUs for "power supply" redundancy. This redundancy provides power for loss of an individual server's internal power supply, but not the loss of site power. HPDA racks that contain compute servers are deployed with 60A/3P PDUs. Other racks that support storage functions in the data center have 30A/3P PDUs.

As California is susceptible to earthquakes, safety measures dictate that we prepare for these infrequent, but potentially crippling events. WorkSafe Technologies' ISO-Base™ Platforms are deployed under each rack to allow racks to move during an earthquake without incurring or creating damage (Figure 3-2).

Figure 3-2. Example WorkSafe Technologies ISO-Base™ Platform

At 5,000 ft^2 of usable space, the data center is configured with 12 rows of 15 racks (Figure 3-3). Each rack has 42 rack units available for computer mounting. The middle rack (8th in each row) is reserved for corporate infrastructure equipment, including networking and dark fiber patching. Infrastructure networking is typically 10 GbE to local equipment, with 100 GbE uplinks. The data center provides control and monitoring infrastructure to make it possible for administrators to interact with their hardware remotely. This infrastructure includes a network of Raritan KVM console switches that allow administrators to obtain console access to specific servers in the data center. While most of these KVM connections are hard-wired to administrative servers, there are additional KVM dongles to allow at least one machine per rack to be connected into the infrastructure.

Additionally, there is an ongoing effort to collect power utilization statistics from all PDUs and StarLine Track Busways. Currently, there are a small number of PDUs that are reporting usage, but the corporate infrastructure to access this data is still under development.

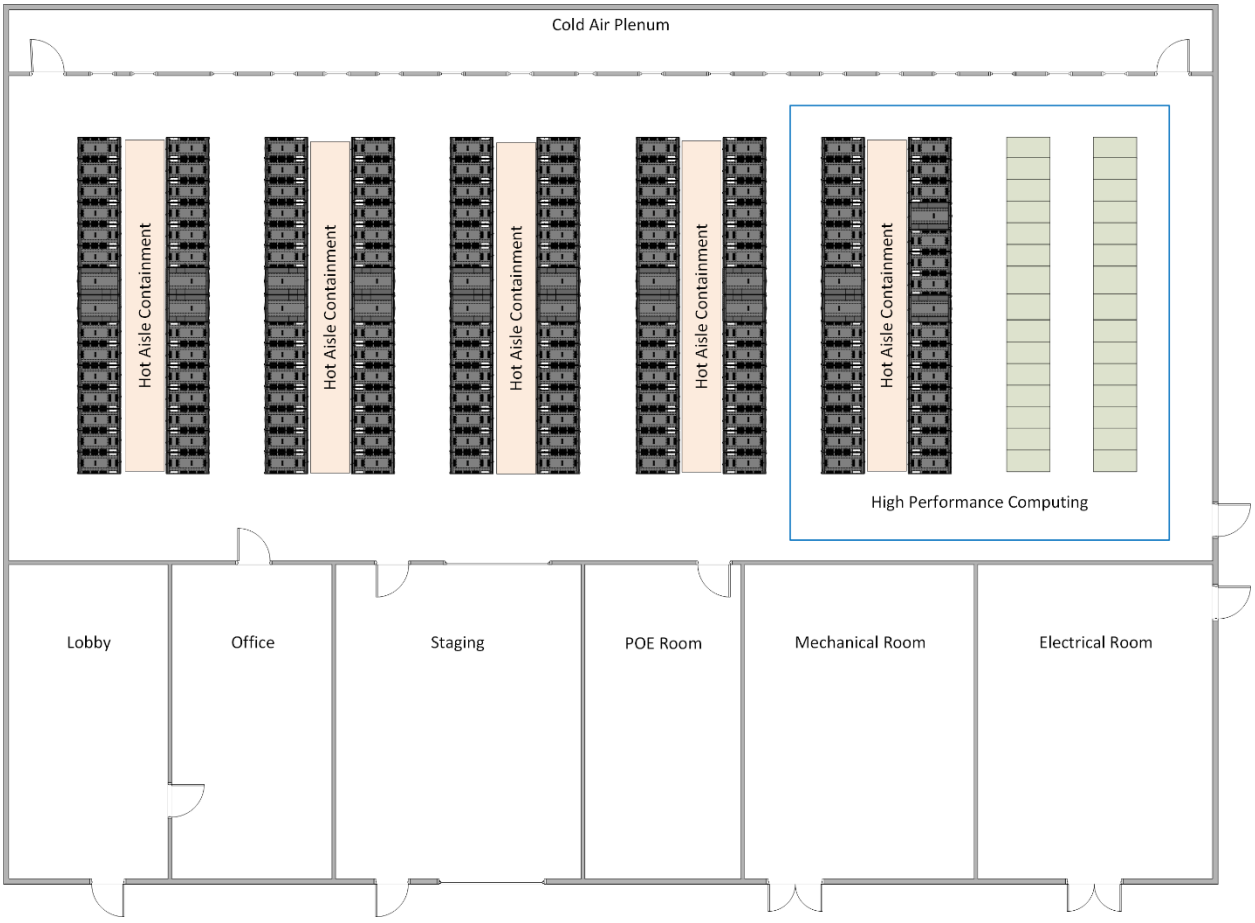


Figure 3-3. The data center is organized into 12 rows of racks

As mentioned earlier in this document, HPDA equipment has been deployed and supported for several years at the California site. The Kahuna and Carnac clusters have been well received by users, and there is a collection of storage and backup servers, GPU servers, admin servers, and testbeds that are used in the operation of the HPDA environment. All this equipment was migrated from the old data center in the basement of building 912 to the new data center during June-December 2021. The HPDA team was allocated four rows of rack real estate in the new data center, of which two rows have power and networking infrastructure available and two are waiting for expansion. After migration of the active HPDA equipment, seven contiguous racks were left vacant in preparation for a new cluster.

3.2. Networking

Networking infrastructure is critical to the functionality of high-speed clusters. The HPC clusters in CA share a common management network for access to service processors, PDUs, and switches. This is a low bandwidth network (1GbE) which is isolated from all other Sandia networks. Cluster nodes boot over a higher bandwidth "boot" network, which is also the primary access point to running nodes and site file storage. The Kahuna and Carnac clusters run at 10GbE; the Glinda cluster runs at 25GbE. Each cluster also has access to a higher-speed private network. As the Carnac cluster is targeted for Emulytics work, its high-speed network is 100Gb Ethernet (using an Arista switch). The Kahuna cluster was deployed as a more traditional HPC cluster, so its high-speed network better supports MPI computing and uses InfiniBand (Voltaire director switch). Glinda is a hybrid, with dual port NICs that support either Ethernet or InfiniBand. The initial deployment for most nodes will be 100Gb InfiniBand using Mellanox switches, but the team plans to test the Ethernet capabilities as well based on our prior success with RoCE [15].

An interesting aspect of the Glinda cluster is the use of "splitter" cables for networking. For the boot network, 100GbE Mellanox switches (SN2100) are deployed, but the cables are 4x splitters, converting a single 100Gb port to 4x25Gb ports. For the InfiniBand network, 200Gb HDR Mellanox switches (QM8700) are used, with 2x100Gb splitters. The use of splitter cables effectively increases the number of ports in a switch while providing the desired speed of the NIC.

Another networking item to note with the Glinda deployment is the use of "top of rack" InfiniBand switches. In this deployment, the initial configuration is heavily tapered. There are 18 nodes connected to a switch, but only a single uplink. The nodes are at 100Gbps and the uplink is at 200Gbps, so the tapering is 18:2. Users that run jobs that fit within a rack will see full network bandwidth as the switch bandwidth is high, but jobs that span multiple racks (and switches) will see degraded bandwidth. This can be addressed using additional uplinks and upgrading the core switch, but until we see this bottleneck affecting the usefulness of the system the configuration will be unchanged.

One final issue of note is that high-speed networking cables are expensive and were in short supply during the COVID-19 pandemic.

3.3. Physical Layout

The Glinda cluster is deployed in 7 racks with 18 nodes/rack. Figure 3-4 depicts the layout, with top-of-rack switches for the boot and management networks, and middle-of-rack switches for InfiniBand. The InfiniBand switches were placed mid-rack to reduce cable lengths. This was a cost and availability decision.

Figure 3-5 is a photograph of the end installation of the Glinda cluster. This picture highlights the overhead cable management and slab flooring used in the data center. The doors on the left side of the cluster help contain Glinda's heat exhaust within the hot aisle.

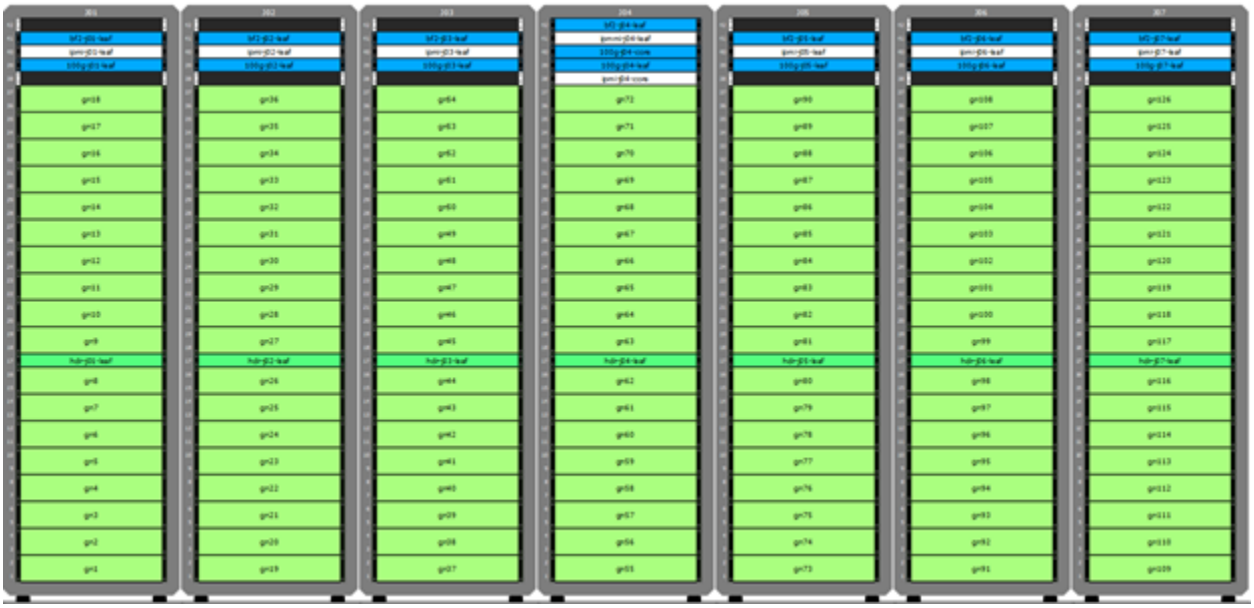


Figure 3-4. Compute nodes are stored in seven racks, with networking at the top

Figure 3-5. The Glinda HPDA Cluster

4. HOST CONFIGURATION

Once the Glinda hardware was physically installed in the data center, the next task in standing up the cluster was to configure our cluster management software to recognize the new hardware and modify our current OS stack to boot on the Glinda nodes. This section provides details about the cluster management environment we use for HPDA systems in the 902 data center and steps through the process by which we adapted our OS stack to boot the nodes into a functional state.

4.1. HPDA Cluster Management Infrastructure

Cluster computers require a small amount of management infrastructure to enable compute nodes to function as a single, parallel-computing platform. As illustrated in Figure 4-1, this infrastructure typically includes (1) one or more administrative nodes for controlling how the compute nodes are configured and hosting services such as DNS, DHCP, and TFTP, (2) storage appliances for hosting home and project directories, (3) an out-of-band IPMI network to monitor resources and power nodes on/off, and (4) login nodes that give users a portal for connecting to the cluster from the enterprise network.

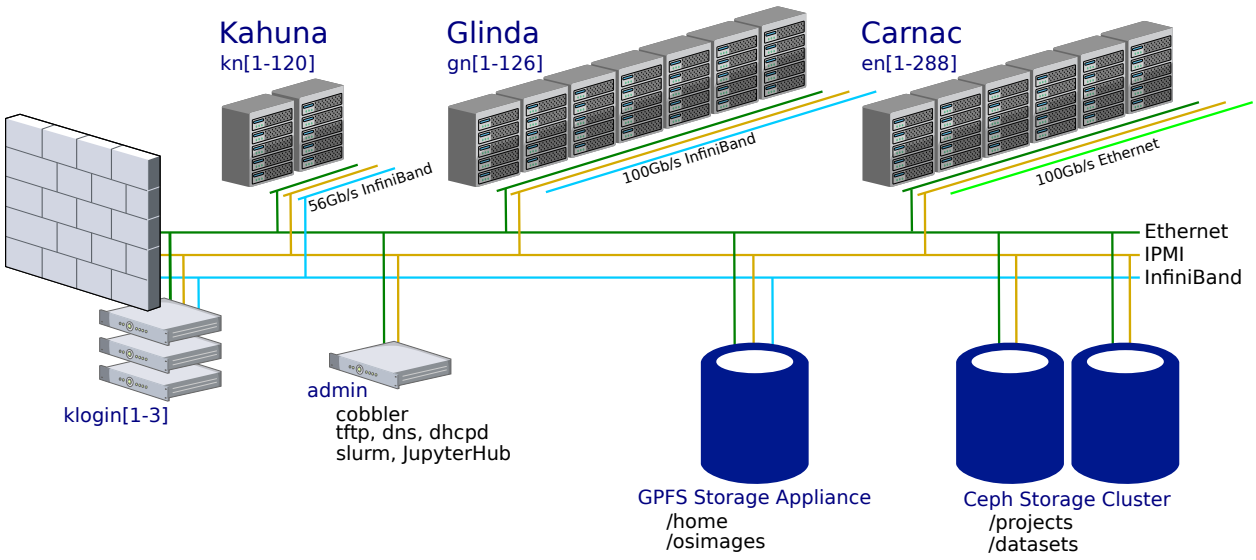


Figure 4-1. Kahuna, Glinda, and Carnac infrastructure

The clusters in the 902 data center use a single instance of Cobbler [16] to manage how all nodes in the cluster boot. While Cobbler is dated, it provides a convenient web interface that allows administrators to provision different OS stacks on a collection of compute nodes as desired.

Cobbler accomplishes its work by writing configuration information to local DHCP, DNS, and TFTP services. Compute nodes are configured to always boot over the network using the PXE protocol. Once powered on a node uses DHCP to find its IP address and determine where the TFTP server is located. The node then uses the TFTP server to retrieve PXELinux, boot information for the node, and the kernel/intramfs data for booting an OS. While most Cobbler deployments use this process to launch a kernel/initramfs pair that installs an OS over the network, Sandia uses this process to boot a specially-designed OS image that resides on an NFS mountpoint.

4.2. Modifications to the oneSIS OS Image

Sandia has constructed a reusable OS image based on CentOS 7.9 that is capable of booting the different compute and login nodes of the Kahuna, Carnac, and Glinda clusters. This image uses oneSIS [17] to customize how each node boots and removes conflicts that arise when multiple compute nodes share the same OS on an NFS share. After a node PXE boots a stock kernel, a custom initramfs is used to mount an NFS share as a read-only root file system. The oneSIS daemon then creates a ramdisk and uses it to maintain a symlink farm for different directories and files in the OS image. While creating a new oneSIS image is challenging, the benefit of this approach is that it is straightforward to boot all the compute nodes once the OS image is configured to work with the new hardware. Additionally, changes made in an image are immediately available on all compute nodes.

A small number of Glinda nodes were used to determine what changes would need to be made in the oneSIS image to support the new hardware. These nodes were manually added to Cobbler and configured to boot into a clone of the current OS image. Through different experiments, we determined the following changes needed to be made to the OS image:

Update OFED Drivers: While Linux distributions have basic support for InfiniBand hardware, NVIDIA maintains an optimized set of drivers and tools in its OFED (OpenFabrics Enterprise Distribution) release. We needed to update the OS image to the latest version of OFED (5.7-1.0.2.1) in order for the BlueField-2 cards to initialize properly.

Update NVIDIA GPU Drivers: The NVIDIA A100 GPU cards require vendor-supplied device drivers in order to access the card for CUDA work. In previous work [8] we updated the oneSIS image's CUDA drivers from 10.2.89 to 11.0.2 to support both the Volta and Ampere cards. We have since updated the core image to CUDA 11.7. Updates involve downloading and installing a "Data Center Driver" RPM from NVIDIA's website. This RPM builds a driver for the installed kernel and needs to be run any time the OS image's kernel is updated.

Update Initram Drivers: During the installation process, the OFED drivers build updated device drivers for the OS image's kernel. When the OpenIB service starts, it verifies the drivers supplied by the initram match the drivers in the OS image and reloads them if necessary. In other systems this reload did not affect operation because the InfiniBand network was not initialized until the OpenIB service was started. Unfortunately, on Glinda the Ethernet boot network is also an NVIDIA NIC and the reload causes the boot network to unmount the

oneSIS OS image and crash the node. The fix in this scenario was simply to rebuild the oneSIS initramfs and ensure it has the latest InfiniBand drivers.

Update oneSIS Configurations: The Glinda nodes required simple modifications to oneSIS configuration files to reflect hardware differences from previous platforms. In addition to defining network parameters for the BlueField card's `tmfifo_net0` port, the oneSIS configuration needed to be updated to reflect the host's GPU and NVMe capabilities.

4.3. Preparing Nodes for Booting

After verifying that our OS image worked properly with the new hardware, we proceeded to prepare all of the new compute nodes so that they could boot the OS image. Preparing compute nodes can be a tedious task, as the work involves configuring each node's BIOS settings to boot over the network and inserting MAC address information for all the nodes into Cobbler. Vendors typically simplify this work by configuring the BIOS for us during burn in and by supplying a list of all MAC addresses in the system. Unfortunately, supply chain issues and the pressure to deliver hardware on schedule forced the vendor to ship some hardware components directly to Sandia without their standard in-house testing procedures. We worked with the vendor to integrate all hardware components on site and enumerate the network devices. This task was tedious, as the manual process for preparing an individual node is:

- Power on the node
- Enter the BIOS setup menu
- Switch the BIOS from EFI to Legacy mode
- Set the network card as the first item in the boot order
- Exit the BIOS setup and reboot the node
- Record the MAC address of the NIC when it performs a DHCP query
- Create a new system in Cobbler and set its MAC and IP addresses
- Resync Cobbler to flush changes to the DHCP service

Once nodes were visible on the IPMI network and manually powered on to boot our image, we discovered that the IPMI login was set with an unknown administrator password. Fortunately, this password can be reset from the OS image by using `ipmitool user` commands. Subsequent changes to the bios settings could then be made to all machines over the IPMI network through carefully constructed curl commands to the Redfish API.

5. SMARTNIC CONFIGURATION

SmartNICs are programmable network interface cards that allow users to place application-specific code at the network's edge. While SmartNICs have previously been deployed in enterprise networks to solve security problems, researchers are still assessing what role, if any, they should have in parallel computing architectures. This section provides background information about the NVIDIA BlueField-2 VPI SmartNIC and summarizes our efforts to integrate these devices into the Glinda system architecture.

5.1. Mellanox/NVIDIA SmartNIC Evolution

Prior to their acquisition by NVIDIA in 2019, Mellanox Technologies, Ltd. was the dominant vendor of high-speed InfiniBand network equipment. Mellanox developed multiple network interface cards (NICs) that employed the ConnectX family of ASICs to rapidly migrate data between host applications and the network fabric. While the ConnectX chips are optimized for data transfer and are not generally programmable, Mellanox has constructed specialty NICs such as the Innova Flex to allow security researchers to monitor and manipulate network traffic via a user-programmable FPGA. In response to requests by commercial cloud vendors to provide a programmable NIC that could help secure cloud infrastructure, Mellanox developed the BlueField SmartNIC product line. The original BlueField SmartNIC supplements a ConnectX-5 network chip with 8-16 Arm processors cores and 16GB of DDR DRAM. While these processors have limited performance due to power and thermal constraints, multiple researchers have demonstrated that useful work can be offloaded into the SmartNICs [18, 19].

The BlueField-2 SmartNIC was announced in 2019 and made available in late 2020. While the BlueField-2 employs only half the processor cores of its predecessor, the cores run at three times the clock rate (2.75GHz vs 800MHz). The BlueField-2 cards also include custom hardware to accelerate cryptography, compression, and regular expression operations. NVIDIA offers multiple variations of the card with different network speeds (25Gb/s-200Gb/s), network fabrics (Ethernet or Ethernet/InfiniBand), and processor speeds (2.0GHz-2.75GHz). NVIDIA also produces the BlueField-2X converged card, which combine a BlueField-2 DPU and an Ampere A100 GPU into a single PCIe card. The converged card may be desirable in situations where several GPUs are attached to a system through InfiniBand and users simply need a minimal hardware path for accessing remote GPU resources.

Sandia initially procured a pair of BlueField SmartNICs in 2019 to assess their capabilities. Additional experiments were conducted by a student in CloudLab's BlueField-2 Ethernet cards in 2021 [20]. These experiments found that while the BlueField-2 processors were significantly

faster than the previous generation of SmartNICs, users should expect that they are an order of magnitude slower than desktop processors.

5.2. BlueField-2 Operating Modes

The Arm processors on the BlueField-2 SmartNIC have dedicated memory and boot an operating system that is independent of the host. As illustrated in Figure 5-1, the BlueField-2 can be configured to operate in one of two modes:

Embedded Function Mode (default): In Embedded Function Mode traffic between the host and the network is routed through software that runs on the Arm processor. Packet processing software and applications such as OpenVSwitch can be used to inspect, manipulate, and route packets on behalf of the host. This mode is commonly used in network security applications where the SmartNIC serves as an embedded hypervisor for protecting the host system.

Separated Host Mode : The BlueField-2 SmartNIC can alternatively be configured to run as an independent host that shares the node's network connection. In this mode, the host exchanges data directly with the ConnectX network interface. The host may communicate with the Arm processors through traditional network operations, such as sockets or RDMAs. Note: the regular expression hardware does not currently work when the card is configured to be in separated host mode.

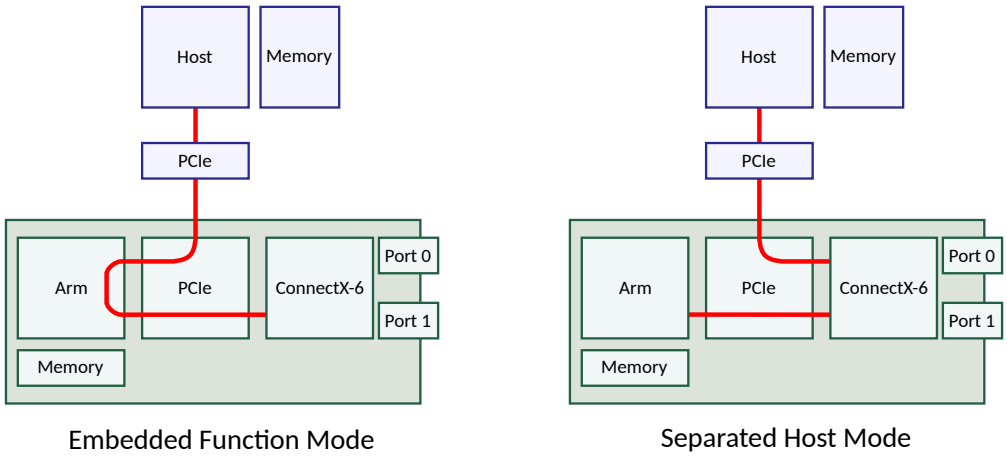


Figure 5-1. BlueField-2 operating modes

An administrator can use the `mlxconfig` tool to change the operating mode for the card. Setting `INTERNAL_CPU_MODEL=0` and power cycling the machine will cause the BlueField-2 to boot into separated host mode. Similarly, setting `LINK_TYPE_P1=1` sets the first port to InfiniBand (or 2 for Ethernet). The Glinda nodes are currently configured to boot in separated host mode and use InfiniBand for the network fabric type.

5.3. General Questions and Answers

One of the challenges of being an early adopter of BlueField-2 equipment was that there was a shortage of general information about the hardware. Table 5-1 provides a short list of questions and answers about the BlueField-2 hardware, based on our experiences.

Table 5-1. General questions about the BlueField-2

Question	Simple Answer
Software	
Will software compiled on hosts with A72 Arms run on the BlueField-2?	yes
Is the BlueField-2's compression accelerator compatible with DEFLATE software?	yes
Can the compression hardware work with streaming data?	yes
Can the host access the compression hardware?	yes
Resilience	
Is the BlueField-2's OS disrupted if the host reboots?	no
Is the host's OS disrupted if the BlueField-2 reboots?	no
Does the BlueField-2 lose network access while the host reboots?	no
Networking	
Does host-to-card InfiniBand work if the port is physically unplugged?	no
Are there (documented) mechanisms for the Arm to sense the host's traffic?	no
Does the host's network traffic get priority over the Arm's?	no
Can the BlueField-2 VPI use InfiniBand on one port and Ethernet on the other?	yes
Will the host and Arm both be visible if the SM does not support 2 LIDs/port?	no
Does InfiniBand work properly when the card is in Embedded Compute Mode?	no
PCIe	
Does the Arm list the host's PCIe cards in `lspci`?	no
Can the Arm access the host's NVMe devices through NVMe-over-Fabric?	yes
Power	
Does the BlueField-2 have a documented, on-card power monitor?	no
Does the BlueField-2 really need the power/cooling listed in the specifications?	yes
What is the stated maximum power of the BlueField-2?	63W
What is the measured idle power use of a BlueField-2 in Glinda?	30W
What is the measured maximum power use of a BlueField-2 in Glinda?	42W

1. We have not tested Embedded Compute Mode extensively, but in our firmware versions the InfiniBand port was missing or nonfunctional.

5.4. Management Interfaces

NVIDIA provides an *rshim* device driver on the host to serve as a means for administrators to interact with the local SmartNIC. The rshim driver provides a number of useful functions:

Console Access: The `/dev/rshim0/console` file provides console access into the Arm processors. The `screen` command can be used to interact with the console and monitor the system's boot process.

Boot Management: The `/dev/rshim0/misc` file displays information about the state of the device and can be used to manage how the system boots. Writing to the `SW_RESET` value triggers a reboot of the card.

Image Update: The `/dev/rshim0/boot` file can be used to write a new OS image to the card's flash. The `bfb-install` tool provides a safe mechanism for writing images.

Local Network: The rshim driver also creates a local, point-to-point network connection between the host and the card. The `tmfifo_net0` network port is much faster than console access, but requires the host and Arm's network settings to be configured properly. The default setting is for the host and Arm to have IP addresses 192.168.100.1/24 and 192.168.100.2/24.

When a new card is being brought online, the easiest way for an administrator to monitor the boot process and log into the card is through the rshim console. The local rshim network port provides a faster way to interact with the card, though it requires the administrator to first activate the port on the host. Finally, the out-of-band (oob) network port provides a more flexible way to log into the card if it is connected to the network and configured appropriately. In the Glinda environment, the oob network port is connected to Glinda's Ethernet network and is used to mount `/home`, `/projects`, and `/datasets`.

The BlueField-2 card provides pins that allow system owners to attach an additional interface for accessing the card's baseband management controller (BMC). While Glinda's BMC connector is not currently populated, owners can purchase adapters that allow users to connect to the BMC and obtain additional information about the status of the card. The documentation indicates that the BMC can be used to control low-level settings of the card, view system error log messages, and access additional IPMI sensors (e.g., temperature and voltage sensor readings).

5.5. Re-imaging the Arm's OS

When powered on, the Arm processors on the BlueField card use EFI to boot an operating system. While a card can be configured to PXE boot an OS image over the network, the default option is to load the Ubuntu 20.04 OS that occupies a portion of the card's local flash storage. This OS installation contains most tools that users will need and is configured to use the rshim interface to obtain DNS, NTP, and `apt` updates through the host.

Given the large number of nodes in the cluster, it is impractical to rely on `apt` to update the Arm's OS when a large number of new packages are available. Similarly, it is essential for an administrator to have a means of resetting the Arm OS to a known-good state. NVIDIA provides OS images for the BlueField-2 Arm processors in the form of BlueField bootstream (BFB) files. A BFB image is a few gigabytes in size and takes approximately 10 minutes to load onto a card from the host. The `bfb-install` tool in the host's installation of the Mellanox OFED tools is

used to perform an installation. An additional configuration file can be supplied to this tool to specify the password to use for the default user account. BFB images include copies of the latest firmware for the BlueField cards. As such, it is useful to log into the card and run `mlxfwmanager` to update the firmware when a new BFB is installed.

Administrators can customize their own BFB installation images through the `bfb-build` project[1]. This project is a set of scripts that use Docker and QEMU to generate a customized OS image for the BlueField. The current repository includes support for multiple OSs, including Ubuntu and CentOS. Platform-specific customizations can be inserted into the Docker file for a particular OS. Building an image may take a half hour or longer on an x86 system due to the need to emulate an Arm processor with QEMU.

5.6. Software Development and the NVIDIA DOCA SDK

Once a BlueField-2 DPU is properly installed in a system, developers can log into it and use it in a manner that is very similar to other hosts in the platform. The Ubuntu OS on the card includes a `gcc` 9.4.0 compiler and build tools, and can be easily updated with other tools from Ubuntu's repositories. We have successfully built a full suite of additional tools and libraries from source using LLNL's Spack[2]. Host applications do not require special libraries to interact with the BlueField-2 other than libraries used for traditional network operations (e.g., TCP/IP sockets or ibverbs). The Data Plane Development Kit (DPDK[3]) does include device-specific support for the BlueField-2 that streamlines IP communication with the card and enables users to take advantage of the compression accelerator. We explored trade-offs with the compression hardware and found that it greatly accelerated the serialization process when working with particle flows [21].

NVIDIA provides the DOCA SDK as a means of simplifying development costs when constructing applications that leverage the BlueField-2's capabilities. DOCA is a collection of host and card libraries that target a mix of different use cases. While a central part of DOCA focuses on creating a trusted environment for offloading security operations, the SDK includes libraries that help optimize host/card data transfers and simplify access to the card's accelerators. However, users should carefully *review the DOCA EULA*[4] before committing to DOCA. Item 4(c) of the current EULA has the following restriction:

> You may not disclose the results of benchmarking, competitive analysis, regression or performance data relating to the SOFTWARE without the prior written permission from NVIDIA Mellanox.

[1] `https://github.com/Mellanox/bfb-build`
[2] `https://spack.io`
[3] `https://www.dpdk.org/`
[4] `https://docs.nvidia.com/doca/sdk/eula/index.html`

6. STRESS TESTING AND POWER MEASUREMENTS

As part of our acceptance process we conducted a series of burn-in tests to ensure that the Glinda hardware functioned properly over long periods of time. During these tests we monitored power use and heat generation to verify that the the cluster did not exceed the capabilities of the data center. Our tests leveraged a variety of tools to generate load on different system components:

- **`stress-ng`**: `stress-ng`[1] is a collection of micro-benchmarks that evaluate system performance in different scenarios. Several of these tests place considerable load on the CPU and are an effective way to force all cores to run at peak levels for a sustained amount of time. Given that `stress-ng` supports different CPU architectures, it can be used to place work on both the host and SmartNIC processors.
- **`dcgmi`**: NVIDIA's Data Center GPU Management[2] software provides the `dcgmi` tool for controlling different aspects of a GPU-based system. Running a "Targeted Power" diagnostic[3] generates work that maximizes the GPU's power use.
- **`ib_send_bw`**: The InfiniBand `ib_send_bw` tool provides a convenient means of generating a large amount of network traffic. Our stress tests set up continuous transfers between different pairs of host and SmartNIC processors.
- **High Performance Linpack (HPL)**: HPL [22] is a well-known HPC application that places a significant amount of strain on a platform's processors, memory, and interconnect. Given that Glinda's InfiniBand interconnect was initially deployed with a highly tapered topology, our HPL stress tests focused on loading all systems with a node-local workload.

After the successful completion of several small-scale tests, we launched a full-scale test that ran continuously for two days. The nodes did not exhibit any problems with power or cooling during these tests.

6.1. Power Monitoring

The baseboard management controller (BMC) in a Glinda node provides access to a variety of sensors that are distributed throughout the chassis. An administrator can easily query the sensors of a node via IPMI using `ipmitool`[4]. The list of sensors in a Glinda node are as follows.

[1] `https://github.com/ColinIanKing/stress-ng`
[2] `https://docs.nvidia.com/datacenter/dcgm/`
[3] `dcgmi diag -r "Targeted Power"`
[4] `ipmitool -H gn78.ipmi -U admin -P redacted sensor`

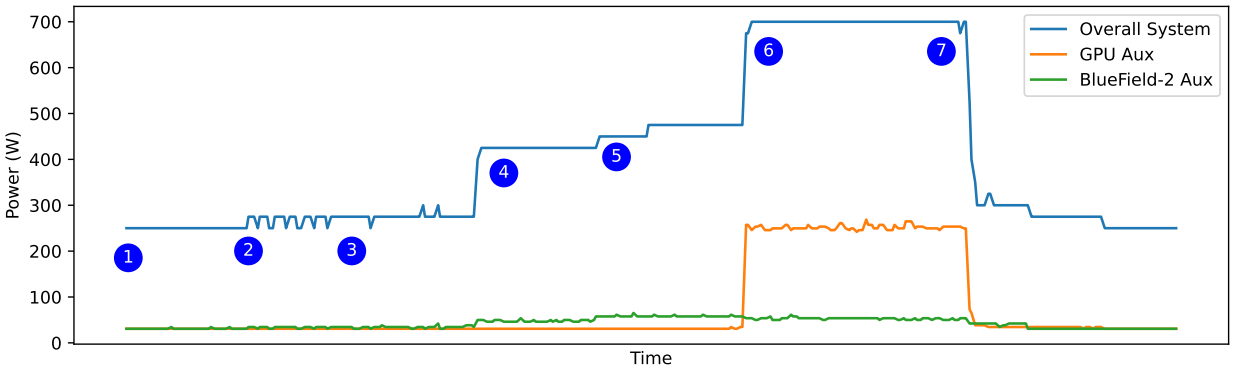
Sensor	Value	Unit	Sensor	Value	Unit
12V_GPU0	2.560	Amps	P0_VDDCR_CPU	0.560	Volts
12V_GPU1	0.000	Amps	P0_VDDCR_SOC	0.721	Volts
12V_GPU2	2.880	Amps	P0_VDD_18	1.833	Volts
12V_GPU3	0.000	Amps	P0_VDD_18_DUAL	1.823	Volts
BPB_FAN1	8400.000	RPM	P0_VPP_ABCD_SUS	2.437	Volts
BPB_FAN2	8400.000	RPM	P0_VPP_EFGH_SUS	2.450	Volts
BPB_FAN3	8400.000	RPM	PS1_Status	0x0	discrete
BPB_FAN4	8400.000	RPM	PS2_Status	0x0	discrete
BPB_FAN5	6900.000	RPM	PSU1_HOTSPOT	31.000	degrees C
CPU0_DTS	68.000	degrees C	PSU2_HOTSPOT	31.000	degrees C
CPU0_Status	0x0	discrete	P_12V	12.025	Volts
CPU0_TEMP	32.000	degrees C	P_1V2	1.232	Volts
DIMMG0_TEMP	30.000	degrees C	P_3V3	3.391	Volts
DIMMG1_TEMP	39.000	degrees C	P_5V	5.037	Volts
GPU0_PROC	26.000	degrees C	P_5V_STBY	5.037	Volts
GPU1_PROC	na	degrees C	P_VBAT	3.159	Volts
GPU2_PROC	na	degrees C	SEL	0x0	discrete
GPU3_PROC	na	degrees C	SLOT1_TEMP	43.000	degrees C
Inlet_Temp	21.000	degrees C	SLOT2_TEMP	59.000	degrees C
MB_TEMP1	34.000	degrees C	SYS_POWER	250.000	Watts
MB_TEMP2	22.000	degrees C	Watchdog	0x0	discrete
NVMe0_TEMP	39.000	degrees C			
NVMe1_TEMP	na	degrees C			

There are several useful sensors in this list:

- **`SYS_POWER`**: The `SYS_POWER` sensor provides the overall amount of power the entire node is currently consuming in watts. This value unfortunately is coarse grained and has a resolution of 25W.
- **`12V_GPUx`**: The Glinda node has four DC power connectors that are intended to be connected to GPUs in the node's four, larger PCIe bays. Each power connector splits into a small connector for a riser and a standard 8-pin auxiliary power connector for a GPU. The vendor has indicated that current sensor measurements cover both connections. Each current sensor has a resolution of 0.32A (i.e., 3.84W at 12V DC). `12V_GPU0` is connected to the A100 card in the front-right bay. `12V_GPU2` from the front-middle bay has been routed to the BlueField-2 DPU in the back-middle bay of the chassis.
- **Temperatures**: There are a variety of temperature sensors throughout the chassis. From the above listing of an idle node, we see an example of how cool air at the inlet (21°C) passes through the GPU (26°C), CPU (32°C), DIMMs (39°C), NVMe (39°C), power supply (31°C), and PCIe slots (59°C).

6.2. Aggregate Power Test

As a means of illuminating how much power is consumed by different components in the system, we collected power measurements from the BMC while different stress tests executed in parallel on the node. For this experiment we launched each test individually and waited approximately 30 seconds before starting the next test. The overall system power usage (i.e., `SYS_POWER`), GPU auxiliary power (i.e., $12V \times$ `12V_GPU0`), and BlueField-2 auxiliary power (i.e., $12V \times$ `12V_GPU2`) were captured during seven stages of activity, listed below.



1. **Idle**: The host initially starts in an idle state with all hardware components booted.
2. **Host-to-Host Network Traffic**: Next, the `ib_send_bw` tool is started on the host to continuously push data to a neighboring node.
3. **Host-to-SmartNIC Network Traffic**: A second instance of `ib_send_bw` is then started to push data to the local SmartNIC CPUs.
4. **Host `stress-ng`**: A stress test is run on the host to maximize host CPU activity on all cores.
5. **SmartNIC `stress-ng`**: An additional stress test is then run on the SmartNIC to place load on its Arm cores.
6. **GPU Load Test**: The NVIDIA `dcgmi` tool is launched to maximize GPU activity.
7. **Shutdown**: Finally, all tests are stopped to allow the system to return to an idle state.

Figure 6-1. Power measurements for a Glinda node during stress tests

Power measurements during the stress test are presented in Figure 6-1. The Glinda node in this test consumed 250W while idle and 700W with all resources active. The largest jumps in power took place when the GPU test 6 (+225W) and host stress test 4 (+150W) activated. The SmartNIC stress tests 5 (+50W) and InfiniBand operations 2 3 (+25W) had less impact on the overall power consumption of the node.

It is important to note that there are multiple sources of uncertainty that cloud an assessment of power use in a Glinda node. First, the 25W resolution of the `SYS_POWER` sensor makes it difficult to get an accurate measurement of power use in the system. As the InfiniBand transfer tests indicate, there are several important operations in the node that fall below a 25W cutoff. Second, it was unclear to us whether a `12V_GPUx` sensor monitors the PCIe card's auxiliary power connector, the riser's power connector, or both. Finally, we noticed that the BlueField-2's power use increased at times when the card was idle (e.g., 4). These measurements imply that the

12V_GPU2 measurements may include more than just the BlueField-2's power usage.

In order to gain more insight into the Glinda node's power characteristics, we examined the A100 and BlueField-2 cards individually.

6.2.1. Ampere A100 Power Use

The Ampere A100 card is listed as having a maximum sustained power consumption of 250W. The IPMI sensor for the GPU's auxiliary power connector reported that an idle card used 30.72W and a fully-loaded card used up to 268W. The A100 card has additional, internal power and temperature sensors that can be queried through the CUDA libraries. The nvidia-smi tool reported the card used 250W while running the stress test. An inspection of the front riser card that holds the A100 reveals that it is a minimal circuit board that merges two PCIe data cables from the motherboard and one power connector into a standard PCIe slot. Given that current measurements for an unloaded front riser board were 0A and the power measurements of an active A100 match the A100's internal estimates, we conclude that the 12V_GPU0 sensor provides an accurate measurement of the A100 card's total power use.

6.2.2. BlueField-2 Power Use

The power specifications for a P-Series BlueField-2 indicate that a 16GB card has a maximum power consumption of 63W. Our initial power measurements with the 12V_GPU2 sensor reported that the BlueField-2 consumed 30.72W when the node was idle, 42.24W when the SmartNIC alone was active, and 65.28W when the CPU, SmartNIC, and GPU were fully loaded. However, the 65.28W was also observed in instances when the host CPU was active and the BlueField-2 was idle. An inspection of the back-middle riser card revealed that it was more complex than the front riser: in addition to being able to host two PCIe cards, the riser connects to the motherboard through a dedicated slot. While the back riser uses the same 12V power connector as the front risers, we suspect that the system can move power from motherboard to riser and riser to motherboard as needed. As such, the 12V_GPU2 sensor is not a good indicator of BlueField-2 power use when the CPU is active.

Table 6-1. Power use for individual slots in different scenarios

Slot	Idle	CPU Active	CPU + SmartNIC Active
Empty Front Slot	0W	0W	0W
SmartNIC in Back Slot	30.72W	65.28W	65.28W
SmartNIC in Front Slot	30.72W	30.72W	42.24W
Empty Back Slot	11.52W	26.88W	26.88W

To test this hypothesis we moved the BlueField-2 to the front-left bay and attached the 12V_GPU3 connector to it while leaving 12V_GPU2 connected to the back riser card. Power measurements for different workloads are presented in Table 6-1. As we suspected, the BlueField-2's idle power consumption remained at 30.72W whether the host CPU was active or not. Placing load on the

Arms increased power use to 42.24W. In contrast the empty, back-middle riser card jumped from 11.52W to 26.88W when the host CPU changed from idle to active. Given that an empty front slot did not consume power, we expect that the 42.24W measurement is a realistic estimate of the Arm's active power use.

Additional attempts to push the BlueField-2 power consumption closer to its 63W limit were not successful. We experimented with oversubscribing the cores, executing continuous RDMA network transfers, and launching additional tasks such as hashing random data, but none of the additions increased the power use beyond the value observed during `stress-ng`'s execution. The hardware accelerators (e.g., compression) were not explored in these tests and may contribute to the device's 63W limit.

7. CHALLENGES AND SOLUTIONS

Every cluster deployment comes with its own set of unique challenges that achitects must overcome. This section provides a description of the main problems we encountered and overcame while standing up Glinda.

7.1. NVIDIA A100 Half-Width Problem

During our initial burn-in of the Ampere A100, we noticed that in the `dcgmi` tests, a small number of cards warned that the PCIe connector was only using half the PCIe lanes that were available to the card. While the BIOS settings for these cards did not reveal anything suspicious, `lspci` confirmed the cards used 8 lanes instead of 16. While the cards passed tests, we could not accept a system issue that causes a 50% reduction in PCIe bandwidth.

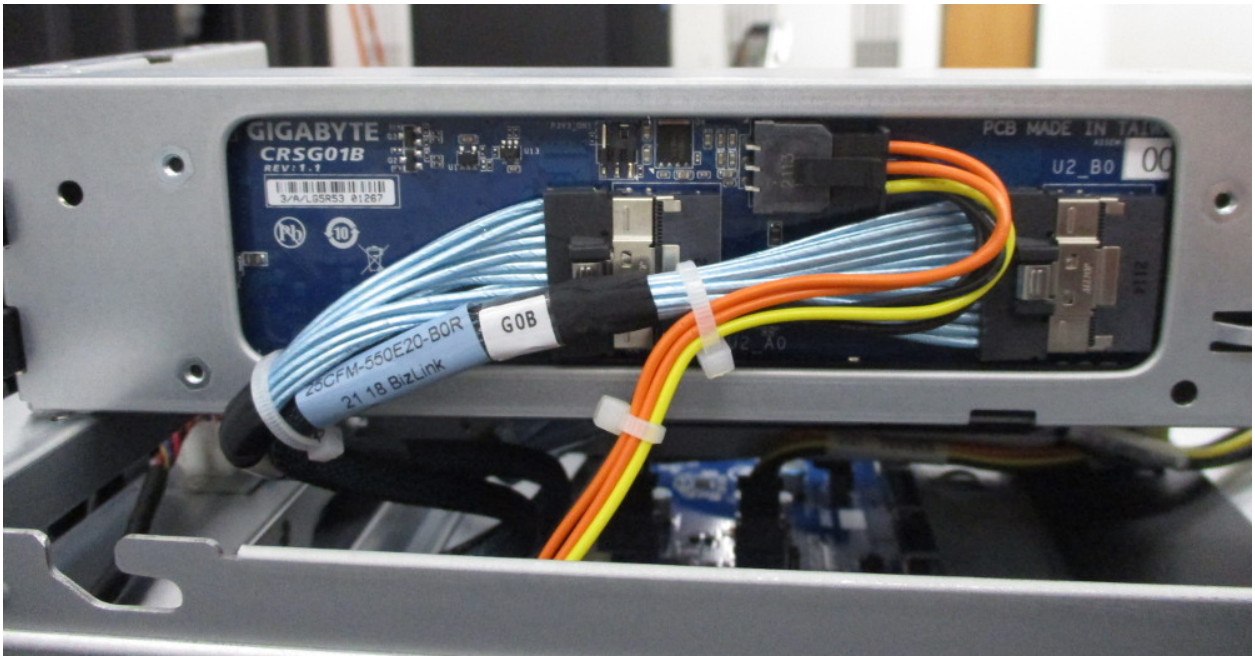


Figure 7-1. Pressure against the right PCIe cable (blue) can cause the left PCIe cable to become unplugged

The vendor visited the site to inspect the bad nodes and determine the cause of the slowdown. After reseating cables and cleaning connections, the vendor discovered that one of the two PCIe cables routed to the A100 riser card had become unplugged. As seen in Figure 7-1, one of the

blue PCIe cables was situated close to the connector of the other cable. When the card was pushed into its bay, the cable pressed against the release post of the other cable and caused it to become unplugged. Simply reconnecting the plug and adjusting the routing of the cable when placing the GPU in its bay was sufficient for fixing this problem.

7.2. NVIDIA A100 ECC Problems

Over the first two years of use, we observed approximately a dozen situations where the A100 cards reported memory corruption problems. In these situations, users typically reported that either a job failed or a card could no longer be detected in the node. For the former, ECC information can be obtained by running the `nvidia-smi` tool and examining the "Volatile Uncorr. ECC" entry in the output table.

```
[root@gn118 ~]# nvidia-smi
Wed Mar 22 08:11:01 2023
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 515.65.01    Driver Version: 515.65.01    CUDA Version: 11.7     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  NVIDIA A100-PCI...  On   | 00000000:01:00.0 Off |                 1120 | <-
| N/A   28C    P0    33W / 250W |      0MiB / 40960MiB |      0%      Default |
|                               |                      |             Disabled |
+-------------------------------+----------------------+----------------------+
```

In situations where volatile errors are reported, NVIDIA recommends issuing a reset to the card:

```
systemctl stop nvidia-dcgm
systemctl stop nvidia-persistence
nvidia-smi -r
GPU 00000000:01:00.0 was successfully reset.
All done.
systemctl start nvidia-persistence
systemctl start nvidia-dcgm
```

Over the last year there have been two instances where A100 cards have completely failed. In these situations the host reports that it cannot access the GPU (e.g., `nvidia-smi -q` returns "No devices were found"). We were not able to revive the cards through reboots, power cycles, or reseating. We RMA'd the cards to Atipa and were supplied with replacements that functioned properly.

7.3. BlueField-2 Driver Replacement Problem

The oneSIS OS image that we use to boot clusters relies on a custom initram image to transition from a PXE Linux boot environment to the full-fledged OS that boots an OS image that resides on an NFS mount point. While the initial OS image that was used to boot Kahuna worked on the Glinda nodes, the system crashed when we upgraded the OFED device drivers in the image. The problem is that at boot, the OFED scripts will inspect the running InfiniBand device drivers and

attempt to replace any old drivers with entries provided in OFED. In systems such as Kahuna, this is not a problem because the replacement is performed at a point in the boot process where the InfiniBand drivers have not mounted any remote file systems. Unfortunately, Glinda's Ethernet boot network that provides the root NFS image is a Mellanox Ethernet card. When the InfiniBand drivers attempt to unload the InfiniBand drivers, they cause the Ethernet card and OS image to go offline, causing the system to hang.

The fix for this problem is to place an updated version of the Mellanox drivers in the oneSIS initram to ensure that the OFED software does not detect an outdated driver. One of the downsides of this approach is that it means the oneSIS initram will need to be regenerated every time there is an update to the kernel or the OFED installation.

7.4. BlueField-2 Corrupted OS Image

While troubleshooting issues with different BlueField cards, we found three systems where the host's networking worked, but the Arms were completely unresponsive. Issuing a software reset of the card using the rshim module did not fix the problem. However, connecting to the Arm through the rshim's console revealed the card was stuck at a grub menu and could not find an OS to boot. Theorizing that there was some form of corruption in the flash memory holding the Arm's OS, we reimaged the SmartNIC with the stock NVIDIA image. This approach is straightforward and fixed the problem. However, we have since noticed that at least one of the nodes has become stuck in a similar way after some hard power shutdowns. We are concerned that the host may power down in a way that corrupts the SmartNIC's flash.

7.5. BlueField-2 Not Detected by Motherboard

When Glinda was first powered on, a significant number of nodes did not detect their BlueField-2 cards. Given that the card was absent from both BIOS and `lspci` listing when Linux was loaded, we suspected the problem was either due to hardware or firmware. Simply rebooting the nodes did not fix the problem, but by power cycling machines multiple times, we could eventually get most of the computers to recognize their cards. Atipa worked with both Gigabyte and NVIDIA to resolve the timing issue. Both vendors provided firmware updates that eventually fixed the problem and stabilized the system. Compute nodes now consistently recognize their BlueField-2 cards on power up.

We have had mixed success with detecting the BlueField-2 cards in other server systems. While some systems never have a problem, others consistently fail to detect a card. NVIDIA is very clear on their website that the BlueField-2 has higher power requirements than other cards and that they may not work in some systems. We recommend consulting NVIDIA's list of systems where cards are known to work before procuring any equipment.

7.6. InfiniBand Routing Issues for the BlueField-2

Our initial plan for integrating Kahuna and Glinda was to connect the InfiniBand switches of the two systems and rely on the subnet manager of Kahuna's core switch to route the entire network. We first observed some routing problems when we connected the BlueField test cards to the older infrastructure: while the subnet manager would detect and route either the host or the SmartNIC Arms, it would not handle both. Sensing that the Kahuna switch's subnet manager may be out of date, we ran a separate subnet manager on one of the hosts. This subnet manager fixed the routing problem. However, over time the two subnet managers would eventually diverge and cause routing errors that crashed other hosts in the system. We promptly detached the SmartNICs from production hardware and created a network island for Glinda experiments.

The subnet managers in Glinda's new switches exhibited the same behavior as Kahuna's switch in regards to detecting both the host and the Arm endpoints. NVIDIA helped us resolve the problem. We had to upgrade the switch's softwate to the latest release, increase the number of LIDs per port, and configure the subnet to enable virtual host support. Our understanding is that by default, the switch only expects to see one local identifier (LID) per port. The router must be informed that there may be virtual hosts on each link at that they each need their own LID routing. As seen in Figure 7-2, the `LIDs for normal ports` is assigned to `2 LIDs per-port` and we show how to enable virtual host support in the lower box.

Figure 7-2. Adjusting the subnet manager to support BlueField-2 use

7.7. Procurement During a Pandemic

Glinda was a unique deployment because the design, procurement, and stand up took place entirely during the COVID-19 pandemic. As a protective measure, Sandia closed its worksites to everyone except essential workers and switched to a remote work environment where employees stayed connected through corporate communication tools. The Glinda team met regularly during the design phase of Glinda and conducted a number of phone interviews with different groups at Sandia to obtain a better understanding of the technologies users would require over the next five years.

The pandemic added a great deal of chaos during the procurement phase of this work due to uncertainties in the market. Sandia queried multiple vendors to get a better idea of what products might be available by the end of the year. We expect that several vendors ultimately did not participate in the contract bidding process because of supply chain uncertainty. As the year moved on we saw shifts in what was plentiful and what was sparse. The A100 GPUs and AMD Zen3 processors proved to be more abundant than the media had indicated. However, some items like NICs and high-speed network cables were extremely hard to obtain due to limited production runs and hoarding.

The pandemic introduced notable delays in standing up and testing the overall system. Delays in network hardware meant that the vendor could not perform a complete integrated test of all the hardware at their site before delivery. Instead, the vendor tested the host and GPU components at their location, and then worked with Sandia on performing full tests on the integrated hardware after it was assembled at Sandia. All in-person meetings at this time followed Sandia's requirements for masking, maintaining distance, and minimizing contact. As seen in Figure 7-3, staff wore protective equipment, even when working in the hot aisles of the data center.

Figure 7-3. Mask and earplug protective equipment

8. REMAINING WORK AND CONCLUSION

The Glinda system was successfully installed in the 902 data center and configured to run the existing Kahuna software stack. The platform is available to users that have access to Sandia's restricted network and has been leveraged by a variety of projects from different mission spaces. This section provides an outline of work the institutional computing team is planning to improve Glinda's computing environment.

8.1. Integrating Glinda's SmartNICs into the Slurm Environment

Access to the Arm processors on Glinda's SmartNICs is currently restricted to a small number of network researchers that have a need to run software on the SmartNICs. Our intent is to make the software environment for these processors robust enough that traditional HPC users can leverage the Arm processors that are available in an allocation of compute nodes. There are multiple policy and technical challenges that will need to be resolved, including:

How should SmartNICs be managed by Slurm? Given that the SmartNIC's Arm processors appear as an independent compute node in the network, some BlueField cluster deployments simply expose SmartNICs as additional hosts that can be scheduled by Slurm. While this approach is straightforward, allocation requests may be too confusing for users to leverage in a practical manner. Additionally, it may undesirable for users to be able to allocate SmartNIC resources that reside in another user's job when both share access to the same network link. Our current philosophy is that users should automatically be granted access to the SmartNICs that reside in the compute nodes they have allocated.

How can access be extended from the host to the SmartNIC? Glinda Slurm installation currently uses Munge to permit users to log into compute nodes they have reserved. Glinda's SmartNICs currently use kerberos and a limited user list to permit access to the Arm processors. The team will need to develop a more robust means of controlling access to the Arms once a broader policy is defined for the system.

8.2. Modernizing the OS Stack

The OS image used with Glinda and Kahuna is currently based on a CentOS 7 OS that uses oneSIS to allow multiple nodes to PXE boot off the same NFS image. While this approach has provided a stable environment for many years, there are multiple operational concerns that motivate us to consider other alternatives for booting the nodes. From an OS perspective, CentOS is no longer a viable operating system for us because Red Hat has switched to a rolling release

process that is incompatible with Sandia's policies. Similarly, oneSIS is no longer being developed and lacks features that other cluster management tools offer. Finally, new Intel chipsets have deprecated support for Legacy mode and can no longer be booted with our current boot system.

As a means of modernizing our boot process, we have successfully constructed a prototype boot environment that will be sufficient for replacing our environment. In this prototype we have replaced CentOS 7, oneSIS, legacy boot mode, and Cobbler with Red Hat Enterprise Linux 8, the read-only root service, UEFI PXE booting, and native TFTP, DHCP, and DNS daemons. We anticpate updating the Kahuna and Glinda clusters to the new OS image at the end of FY23.

8.3. Broader Collection of Software Modules

Glinda currently provides two collections of software modules that users can access to customize their environment: the original software modules developed for Kahuna and the gcc11 set of modules optimized for the Zen3 processor architecture. While the Kahuna modules cover a broad range of libraries and tools, they have not been updated in over a year and use a build process that is no longer supported at Sandia. In contrast the modules customized for Glinda's Zen3 processors only cover a narrow selection of tools and libraries. We are currently developing a new software stack that uses a more modern version of Spack to generate the clusters' tools and libraries.

8.4. Container Integration

While many of our users rely on our software modules to provide standard tools and libraries for their work, we recognize that efforts to make Docker [23] containers more user friendly have made significant strides in recent years. Our experiments with Podman [24] have confirmed that it is possible for users to manage their own Docker images on compute nodes without requiring root access. However, additional experiments with other tools indicate that Apptainer [25] may provide a better mechanism for users to access a container's software stack. We anticipate making Apptainer available to users in the software stack update and will provide stock containers that include common data science tools.

8.5. Conclusion

The Glinda cluster is a new HPDA platform at Sandia that provides users with a heterogeneous architecture for processing large amounts of data. Data scientists from across Sandia are drawn to Glinda because it provides a large pool of individual GPUs that can be used to prototype new data analytics for different mission spaces. The system is also one of the first platforms to employ SmartNICs at a scale that is larger than 100 nodes, and has already been used to demonstrate that SmartNICs can perform useful operations for scientific workflows [26]. Glinda was designed, procured, and brought to life during a stressful time in the COVID pandemic, and is available for general use at Sandia.

Figure 8-1. The SNL/CA Glinda Team (minus Sam Knight) in front of Carnac and Kahuna (pre-COVID)

REFERENCES

[1] Andy B Yoo, Morris A Jette, and Mark Grondona. Slurm: Simple Linux utility for resource management. In *Workshop on job scheduling strategies for parallel processing*, pages 44–60. Springer, 2003.

[2] Matei Zaharia, Mosharaf Chowdhury, Michael J Franklin, Scott Shenker, and Ion Stoica. Spark: Cluster computing with working sets. In *2nd USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 10)*, 2010.

[3] Matthew Rocklin. Dask: Parallel computation with blocked algorithms and task scheduling. In *Proceedings of the 14th python in science conference*, volume 130, page 136. Citeseer, 2015.

[4] Clinton Gormley and Zachary Tong. *Elasticsearch: the definitive guide. A distributed real-time search and analytics engine*. " O'Reilly Media, Inc.", 2015.

[5] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. TensorFlow: a system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, 2016.

[6] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B Hamrick, Jason Grout, Sylvain Corlay, et al. *Jupyter Notebooks-a publishing format for reproducible computational workflows.*, volume 2016. 2016.

[7] Fastx. `https://www.starnet.com/fastx/`. Accessed: 2023-06-27.

[8] Stefan Seritan and Craig Ulmer. Benchmarking the nvidia a100 graphics processing unit for high-performance computing and data analyics workloads. Technical Report SAND2021-1220, February 2021.

[9] Jerry Friesen. Data analytics and emulytics cluster FY21 acquisition statement of work. Technical Report SAND2021-4500, May 2021.

[10] Jack Choquette, Wishwesh Gandhi, Olivier Giroux, Nick Stam, and Ronny Krashinsky. NVIDIA A100 tensor core gpu: Performance and innovation. *IEEE Micro*, 41(2):29–35, 2021.

[11] NVIDIA. NVIDIA A100 tensor core gpu architecture. Technical report. Accessed: 2023-09-01.

[12] NVIDIA A100 40GB PCIe GPU accelerator product brief. https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/A100-PCIE-Prduct-Brief.pdf. Accessed: 2023-09-01.

[13] NVIDIA. Bluefield-2 dpu vpi specifications. Technical report, 2023.

[14] KIOXIA CD6-R series product brief. https://americas.kioxia.com/content/dam/kioxia/shared/business/ssd/data-center-ssd/asset/productbrief/dSSD-CD6-R-product-brief.pdf. Accessed: 2023-09-01.

[15] Joseph P Kenny, Jeremiah J Wilke, Craig D Ulmer, Gavin M Baker, Samuel Knight, and Jerrold A Friesen. An evaluation of ethernet performance for scientific workloads. In *2020 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)*, pages 57–67. IEEE, 2020.

[16] Cobbler. https://www.cobblerd.org/. Accessed: 2023-06-28.

[17] onesis. http://onesis.org/. Accessed: 2023-06-28.

[18] Ming Liu, Tianyi Cui, Henry Schuh, Arvind Krishnamurthy, Simon Peter, and Karan Gupta. Offloading distributed applications onto smartnics using ipipe. pages 318–333, 2019.

[19] Jongyul Kim, Insu Jang, Waleed Reda, Jaeseong Im, Marco Canini, Dejan Kostić, Youngjin Kwon, Simon Peter, and Emmett Witchel. Linefs: Efficient smartnic offload of a distributed file system with pipeline parallelism. In *Proceedings of the ACM SIGOPS 28th Symposium on Operating Systems Principles*, pages 756–771, 2021.

[20] Jianshen Liu, Carlos Maltzahn, Craig Ulmer, and Matthew Leon Curry. Performance characteristics of the bluefield-2 smartnic, 2021.

[21] Jianshen Liu, Carlos Maltzahn, Matthew L Curry, and Craig Ulmer. Processing particle data flows with smartnics. In *2022 IEEE High Performance Extreme Computing Conference (HPEC)*, pages 1–8. IEEE, 2022.

[22] Jack J Dongarra, Piotr Luszczek, and Antoine Petitet. The linpack benchmark: past, present and future. *Concurrency and Computation: practice and experience*, 15(9):803–820, 2003.

[23] Dirk Merkel et al. Docker: lightweight Linux containers for consistent development and deployment. *Linux j*, 239(2):2, 2014.

[24] Podman. https://podman.io/. Accessed: 2023-06-28.

[25] Apptainer. https://apptainer.org/. Accessed: 2023-06-28.

[26] Craig Ulmer, Jianshen Liu, Carlos Maltzahn, and Matthew L Curry. Extending composable data services into SmartNICS. In *2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 953–959. IEEE, 2023.

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.