The Koa Ecosystem: Supporting Data Intensive Computing and Education in Hawaii

DAVID SCHANZENBACH, University of Hawai'i – System, USA
RON MERRILL, University of Hawai'i – System, USA
SEAN B. CLEVELAND, University of Hawai'i – System, USA

The Koa ecosystem is a comprehensive computational ecosystem built on open source software and developed by the University of Hawai'i to enhance data-intensive research across its campuses. The Koa ecosystem addresses critical needs for computational resources, high-performance storage, and sustainable infrastructure to support multiple scientific disciplines. Over two years, the Koa ecosystem has significantly boosted research output, as evidenced by numerous publications acknowledging it and the research community's continued investment in expanding the resource.

CCS Concepts: • **Applied computing → IT architectures**; *IT governance.*

Additional Key Words and Phrases: High Performance Compute, HPC, Storage System, Lustre, System Architecture, Cluster Management, Open Source, Virtualization

ACM Reference Format:
David Schanzenbach, Ron Merrill, and Sean B. Cleveland. 2024. The Koa Ecosystem: Supporting Data Intensive Computing and Education in Hawaii. In *Practice and Experience in Advanced Research Computing (PEARC '24), July 21–25, 2024, Providence, RI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3626203.3670611

1 INTRODUCTION

The University of Hawai'i (UH) has long recognized the importance of advanced computation to research across its ten campuses, which has resulted in a history of strategic investments in High Performance Computing (HPC) resources, advanced networks, software frameworks, visualization environments, and professional staff. In 2013 UH Information Technology Services (ITS) accelerated these efforts by establishing the Cyberinfrastructure (CI) group. With this new CI team, UH received a series of National Science Foundation (NSF) grants to enhance its CI capabilities which included centralized HPC resources, workforce development, software engineering, international networking, and advanced CI support for multidisciplinary research.

Despite investments in CI resources between 2013-2020 the UH system still faces several computational resource impedances that hinder scientific aspirations. One major challenge is the need for readily available computational resources to support research across all campuses to support deep learning and artificial intelligence workloads. Another issue is the lack of sufficient computational scratch storage to aggregate inputs and outputs for large datasets and simulations. Furthermore, poor Input/Output (I/O) scratch storage performance of the existing storage investments impact researcher computational throughput and workflow end-to-end efficiency. Additionally, challenges with research storage with capacity changes for campus Google Drive drive a need for a place to store PBs of existing research data.

Manuscript submitted to ACM

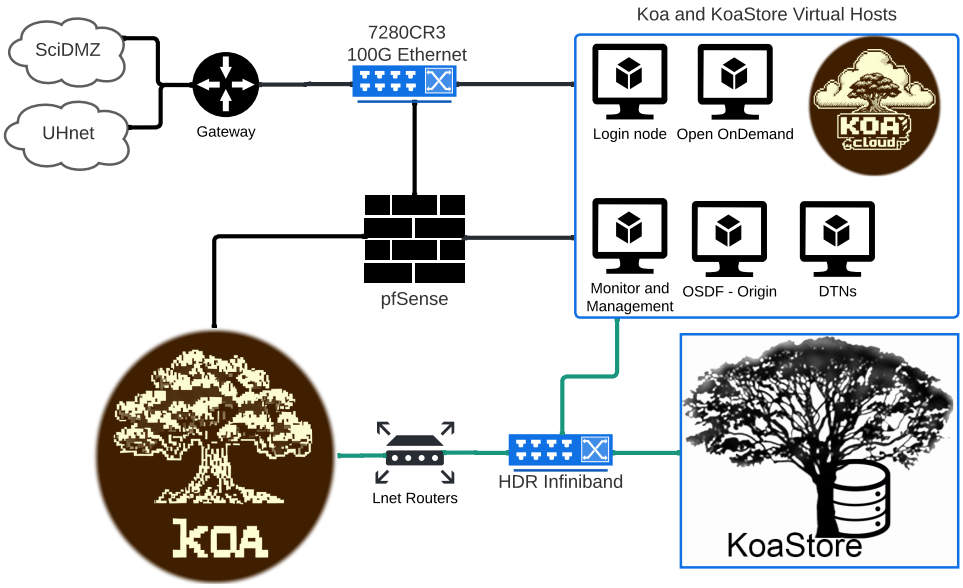

Fig. 1. The Koa ecosystem

To address these issues the Koa (NSF 2201428) and KoaStore (NSF 2232862) awards were used to create a complete computational ecosystem which we henceforth refer to as the Koa ecosystem (Figure 1). The Koa ecosystem addresses the following challenges that researchers at UH face: (1) make additional computational resources available; (2) make additional high performance scratch and permanent storage available; (3) bootstrap a sustainable storage system that UH researchers could buy into and expand; (4) integrate with the large research institutes and labs infrastructure to provide additional storage; (5) bootstrap a Nextcloud deployment for UH researchers to use in place of Google Drive; (6) expand educational opportunities for students through hands-on training workshops.

The Koa ecosystem is managed by the UH Information Technology Services (ITS) Cyberinfrastructure team. Not only does the Koa ecosystem support research but it also provides a platform for advanced technical training and expanded educational opportunities for students and staff in various UH specialties such as astronomy, atmospheric science, ocean science, microbiome studies, and computer and data science. By focusing on these areas, we aim to provide a positive impact on scientific discoveries and advancements at UH. This paper provides details on the Koa ecosystem and its design, implementation and impacts as an advanced CI resource.

2 DESIGN AND DEPLOYMENT OF THE KOA ECOSYSTEM

The Koa ecosystem was designed to serve as a shared inter-campus resource, providing free access to high computational power and storage capabilities for all researchers within the 10 campus UH system. The Koa ecosystem design included incorporating some of the existing Mana [3] cluster's compute, storage and networking infrastructure. The Koa ecosystem replaced Mana's management hardware and open source software systems. These refreshed management systems facilitate continued operation and evolution of the Koa ecosystem.

Qty	Specifications
	Computational Resources
5	Intel Gold 6342, 256 GB RAM 2x A30
2	Intel Gold 6342, 256 GB RAM 10x RTX A4000
8	Intel Gold 6342, 256 GB RAM
	Metadata
4	AMD EPYC 7543P, 512 GB RAM 2x HBA9500-8e 1x ConnectX-6 HDR200
2	Astek A54812-SW-03
2	SMC 216BE2C-R609JBOD w/ 24x 15.36TB SAS SSDs
	Object Storage - SAS3
4	AMD EPYC 7543P, 512 GB RAM 2x HBA9500-8e 1x ConnectX-6 HDR200
4	Astek A54812-SW-03
1	SMC 216BE2C-R609JBOD w/ 12x 15.36TB SAS SSDs
8	WD Ultrastar Data60 w/ 60x 18-24TB SAS HHDs
	Object Storage - NVMe
2	AMD EPYC 7502, 256 GB RAM 1x HBA9500-16e 1x ConnectX-6 HDR200
24	3.84TB SAMSUNG MZWLJ3T8HBLS-00007 NVMe
1	SMC 136R-NEL32JBF w/ 32x 15.36TB NVMe E1.L
	Lustre Router
4	AMD EPYC 7313P, 128 GB RAM 2x ConnectX-6 HDR200
	Virtual Host
8	AMD EPYC 7543P, 256 GB RAM 2x ConnectX-6 100G Eth&HDR200 4x 480GB NVMe
4	AMD EPYC 7543P, 256 GB RAM 2x ConnectX-6 100G Eth&HDR200 4x 960GB NVMe
	Networking
1	Arista 7280CR3-32D4
1	Arista 7050CX3-32S
1	Netgear M4300-48XF
1	Netgear M4300 24x24F
2	Netgear M4300-52G
4	Nvidia QM8700 HDR IB Unmanaged

Table 1. Components purchased by the CC* Koa and KoaStore grants enabling Mana to Koa ecosystem transformation.

2.1 Hardware

As seen in Table 1, the Koa ecosystem includes 8 CPU-only nodes with 48 Intel CPU cores each, providing ample processing capacity for scientific simulations and data analysis tasks. Additionally, it features two types of GPU nodes: five balanced GPU nodes with two A30s and two GPU nodes each with ten RTX A4000 GPUs. Existing Mana cluster components were combined with the new Koa components to provide a total of 275 nodes, with 53TB of aggregate RAM, 7,683 total CPU cores spanning Intel Ivy-bridge up to sapphire rapids and AMD EPYC Milan, and a total of 148

Support and Service Software	
Function	**Software**
Virtualization	ProxmoxVE 8.2
VM image storage	Ceph on Vhosts
KoaStore OS	Rocky 9.3
KoaStore file system	Lustre 2.15.3
Cloud storage	Nextcloud
Management	xCAT 2.17.0
Koa HPC OS	Rocky 9.3
Scheduler	Slurm 23.11.6
Monitoring	Prometheus, Loki, Thanos, Grafana
Exporters	lustre, slurm, dcgm, promtail, node, smart, ZFS

Table 2. The open source software components that make up the management, storage and computational systems that power the Koa ecosystem.

GPUs ranging from Nvidia GeForce RTX 2070 up to Nvidia Tesla L40 GPUs. Koa, like its predecessor Mana, is designed to support future growth via researcher investment by condo nodes and to date 5 additional nodes including 2 H100 GPUs have been purchased for inclusion in Koa this summer.

On the storage side, the Koa ecosystem utilizes a virtual machine (VM) to provide user home storage over NFS, both over TCP and RDMA. The homes are stored as a file based ZFS file system that is backed onto a larger Lustre [1] file system, named KoaStore, which has taken its design cues from Stanford Oak [4] and is used directly for scratch storage for our Koa HPC cluster, permanent storage for various research groups and the backend storage for our Nextcloud instance named KoaCloud. KoaStore consists of four metadata servers connected to two 24 disk JBOFs utilizing two 12 port SAS3 switches. In total, we have 16 15.36TB SSDs that are split into four different metadata targets that utilize ZFS with zstd-3 compression and mirrored stripes. Each of KoaStore's metadata targets provides 27.6 TB of space. KoaStore's scale out design enables researchers to contribute additional storage via a condo model by purchasing hardware that contributes to the overall storage system.

The object storage of KoaStore is more heterogeneous in configuration. Currently, it consists of two sets of two object storage servers (OSS) connected to two SAS3 switches. These servers are connected to Western Digital Data 60 enclosures consisting of 18 to 24 TB drives, each divided into 6 object storage targets (OSTs) utilizing ZFS with zstd-3 compression and raidz-2 (8+2). In addition to high density hard drives, we also have a half-populated JBOF that consists of 12x 15.36TB SSDs configured as two OSTs using mirrored stripes. KoaStore also integrates a pair of OSSs that were part of a Cray/HPE E1000, in order to both reduce the file system choices for our users and bring NVMe into KoaStore. At this time, we have two NVMe OSTs that are configured with ZFS and zstd-3, as raidz-2 (10+2). In total, KoaStore currently provides UH users 6.6PB of object storage and 110TB of metadata storage.

2.2 Software

As listed in Table 2, at the heart of the Koa ecosystem lies a suite of functional open source software, encompassing a wide array of tools tailored to meet the diverse needs of researchers. From virtualization to parallel file systems, cloud storage, and management utilities, each component plays a pivotal role in enhancing productivity and performance.

2.2.1 Virtualization. A flexible, yet cost efficient virtualization platform is paramount for effective management of the Koa ecosystem. To provide, this we utilize ProxmoxVE 8.2 for our virtualization infrastructure to ensure optimal resource utilization with high levels of reliability and security. In concert with ProxmoxVE, we utilize the integrated

version of Ceph running on each of the vHosts to provide a resilient and scalable storage solution for VM images. Ceph's distributed architecture ensures data integrity, redundancy, and high availability, making it an ideal choice for storing critical VM images in our environment. Wanting to eliminate VMWare licensing costs, we chose the free version of ProxmoxVE as a number of our collaborators at UH use it.

2.2.2 Storage system. KoaStore is the primary file system in the Koa ecosystem. It is a Lustre parallel file system running Lustre 2.15.3 on OS Rocky 9.3 servers accessing ZFS storage targets. Lustre was chosen based on our prior experience with two different commercial Lustre systems our appreciation of the attributes scalable and low cost attributes of Stanford Oak storage system.

Whether handling large-scale simulations, data analytics, scientific computing workloads, or cloud storage, KoaStore ensures efficient data access and management across our array of services to meet the performance and capacity needs of our researchers. KoaStore is accessible both on the Koa HPC and as our cloud storage service.

The cloud storage that we provide is facilitated with Nextcloud. This addition of cloud storage extends our storage system beyond traditional HPC paradigms, offering users seamless access to cloud-based storage and collaboration tools. Nextcloud empowers researchers to securely store, sync, and share data, facilitating collaboration and enhancing productivity.

2.2.3 Koa HPC. At the core of our Koa HPC cluster is the need to provide a stable environment while minimizing the time required by system administrators. To streamline system administration tasks Koa uses xCAT [5] 2.17.0, a comprehensive management tool that simplifies cluster provisioning, configuration, and maintenance. xCAT enables centralized management of HPC resources, ensuring optimal performance and resource utilization across the system. For our users, an environment that is stable and will not undergo a complete refresh is paramount in reducing the disruption to continued research efforts. to provide this environment, the Koa HPC utilizes the Rocky 9.x operating system. Rocky 9 provides a robust and secure computing environment tailored to the unique requirements of high-performance computing workloads. Rocky offers stability, performance, and a rich ecosystem of software packages, empowering users to tackle complex computational challenges with confidence. The final component of the Koa HPC is the job scheduler. In order to effectively schedule jobs on a highly heterogeneous set of resources that make up the Koa HPC, we utilize the Slurm [2] job scheduler, a highly scalable and efficient workload manager. Slurm optimizes resource utilization, prioritizes job execution, and ensures fair access to computing resources, enabling our researchers to maximize productivity and throughput.

2.2.4 Monitoring. While monitoring and software constantly evolve, the Koa ecosystem has a comprehensive monitoring stack comprising Prometheus, Loki, Thanos, and Grafana. Together, these tools offer real-time insights into system performance, resource utilization, and workload metrics. With support for various exporters, including lustre, slurm, dcgm, promtail, node, smart, and ZFS. This monitoring stack enables the UH ITS CI team to provide proactive system management and troubleshooting, ensuring optimal system health and performance for researchers.

3 BENCHMARKS

The Koa ecosystem realized many of the performance metrics goals we were targeting to improve. Our previous infrastructure, Mana, primarily utilized an NFS based file system, which sometimes exhibited degraded performance under heavy loads. Table 3 provides a comparison of several different metrics that we were interested in improving. For the IO500 tests, we utilize 7 nodes each with 8 tasks with 5 CPU cores per task. In both tests, we utilize the

Benchmark	Mana Results	Koa ecosystem Results
IO500: ior-easy-write	6.23 GB/s	23.24 GB/s
IO500: ior-easy-read	6.99 GB/s	38.08 GB/s
IO500: mdtest-easy-write	18.48 kIOPs	44.24 kIOPs
IO500: mdtest-easy-stat	72.52 kIOPs	26.84 kIOPs
MLPerf HPC: Cosmoflow	2.8726 Hrs/epoch	1.43 Hrs/epoch
Globus: ESNet Houston to UH	13.76 Gbit/s	33.2 Gbit/s
Globus: ESNet Starlight to UH	7.3 Gbit/s	31.76 Gbit/s

Table 3. Benchmarks performance goals set by Mana's NFS file system compared to performance results obtained using KoaStore.

same hardware, that is HDR100 connected, just with a variation in the file system being used. While we see large improvements in the read and write tests, we do note that the metadata stat test did perform worse. We believe this difference could be related to the difference in how NFS and Lustre are designed and operated. To see if our machine learning users would benefit from the additions of the Koa ecosystem we utilized the cosmoflow benchmark found in MLPerf. These tests were performed utilizing 8 SMX2 connected v100 GPUs on a single node. As we can see, utilizing the file system added by the Koa ecosystem, the benchmark sees a two times speed up when compared to when it ran on Mana. Our final set of benchmarks utilizes the four virtual DTNs we have in the Koa ecosystem. These DTNs are 100Gbit connected to the world and 200Gbit connected to KoaStore. Both the Ethernet and Infiniband network cards are presented to the VMs using SR-IOV. Previous read tests from the Houston and Starlight ESNet test points show that we could not exceed around 13Gbit/s. This result we believe was due to performance limitations on the NFS file system on Mana. On the other hand, the new file system in the Koa ecosystem is capable of almost three times that speed.

All of these increases in performance will translate to researchers at UH being able to perform more experiments at scale and acquire and disseminate their data and results much faster than they could prior to the Koa ecosystem coming online.

4 USAGE

The Koa ecosystem came into full production at the start of 2024 with the completion of assimilating the Mana HPC. This process took place from May to December 2023. If we compare the usage stats from Quarter 1 2023 of Mana to Quarter 1 of 2024 for the Koa ecosystem, we see that we had 1.31 times more active users, 1.05 times more CPU hours and 1.58 times more GPU hours utilized by researchers at UH. We also saw an increase in the amount of storage utilized by researchers at UH where previously, only 1 TB of scratch was used by users on Mana compared to 270 TB used today. Permanent storage has also grown from Mana's maximum of 1 PB in 2023 to today's total which is over 2 PBs.

5 EXTERNAL PARTNERSHIPS

The Koa ecosystem aims to seamlessly integrate and provide resources to the Open Science Grid (OSG) and the Open Science Data Federation (OSDF), allowing for untapped resources in the Koa ecosystem to be utilized by other researchers outside of UH. From February 12, 2024 to April 25, 2024, the Koa ecosystem has provided 457,419 CPU hours OSG. The OSG glide in is limited to 300 concurrent jobs, each providing 1 CPU core and 6 GB of RAM. For OSDF, since November 2023, the Koa ecosystem has provided 800 TB of storage from our KoaStore storage system as a data origin. Current use cases for this storage center on distributing multi-terabyte data products produced by the UH Institute for Astronomy.

6 CONCLUSION

Over the past 2 years, the Koa ecosystem has proved to be a catalyst for research and innovation at UH. In this time, more than 20 publications have been made by researchers at UH that utilize the Koa ecosystem. The demand for compute in the research community at UH, illustrated by researchers' continued investment in the Koa ecosystem computational and storage infrastructure, highlights the importance of advanced computing resources in the eyes of researchers.

ACKNOWLEDGMENTS

This work was made possible by grant funding from the following: NSF OAC-2201428 CC* Compute: Koa - A High Performance and Flexible Research Computing Resource and NSF OAC-2232862 CC* Data Storage: KoaStore: A High Performance and Flexible Research Storage Resource.

REFERENCES

[1] Peter J Braam and Rumi Zahir. 2002. Lustre: A scalable, high performance file system. *Cluster File Systems, Inc* (2002).

[2] Morris A. Jette, Andy B. Yoo, and Mark Grondona. 2002. SLURM: Simple Linux Utility for Resource Management. In *In Lecture Notes in Computer Science: Proceedings of Job Scheduling Strategies for Parallel Processing (JSSPP) 2003.* Springer-Verlag, 44–60.

[3] Ron Merrill, David Schanzenbach, Sean B. Cleveland, and Gwen A. Jacobs. 2023. Mana - Bringing Accessible HPC to Hawai'i. In *Practice and Experience in Advanced Research Computing* (Portland, OR, USA) *(PEARC '23)*. Association for Computing Machinery, New York, NY, USA, 86–93. https://doi.org/10.1145/3569951.3593611

[4] S. Thiell and K. Cavalotti. 2016. Cheap 'n' Deep - The Holy-Grail of Storage. https://www.eofs.eu/wp-content/uploads/2024/02/07_thiell_cheap_n_deep.pdf. (Accessed on 04/25/2024).

[5] xCAT. 2020. xCAT | Extreme Cloud Administration Toolkit. https://www.xcat.org/. (Accessed on 02/07/2020).