LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-871269

HPC Center of the Future: R&D Acquisition Intent

T. Gamblin, B. de Supinski, B. Van Essen, A. Bertsch, T. D'Hooge, M. Leininger, J. Hill, B. Behlendorf, T. Quinn

November 14, 2024

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

HPC Center of the Future
R&D Acquisition Intent
DRAFT TECHNICAL REQUIREMENTS

November 19, 2024

Lawrence Livermore National Laboratory
Department of Energy / National Nuclear Security Administration

Table of Contents

This work is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

1.0 Introduction

This document contains intended technical requirements for an anticipated future procurement for Lawrence Livermore National Laboratory (LLNL), hereafter referred to as “LLNL” or “the Laboratory”, which seeks to fund a few new and innovative proposals for developing concepts and performing studies that would lead to products available in the 2029-2030 timeframe. Our interest is to incentivize explorations and evaluations of technical capabilities in support of our vision for a High Performance Computing (HPC) center of the future at Livermore Computing (LC), the HPC center at LLNL. To realize this vision a new approach to our approach to acquisition is being considered by LC. The approach is designed to decrease risks for the vendors, while encouraging a broader set of vendors able and we hope willing to partner with LLNL and LC. The future HPC Center vision and acquisition approach was conceived to meet the future mission needs of the Advanced Simulation and Computing (ASC) Program within the National Nuclear Security Administration (NNSA).

Additional information on proposal preparation and related matters will be addressed in **an anticipated future formal Request for Proposal (RFP)**, **which will supersede this document**..

The following sections outline the research areas of interest for a new kind of HPC center. LLNL envisions a center where vendors are able to compete for multiple capabilities or a single capability. The capabilities will be integrated to create a scalable and flexible, and yet tightly coupled computing center capable of integrated HPC, AI, and cloud workloads.

2.0 Program Overview and Mission Needs

2.1 National Nuclear Security Administration (NNSA)

The NNSA, an agency within the DOE, is responsible for the management, security, and modernization of the nation's nuclear weapons, nuclear nonproliferation, and naval reactor programs. The NNSA Strategic Plan supports the Presidential Declaration that the United States will maintain a "safe, secure, and effective nuclear stockpile." The Plan includes ongoing commitments to:

- Understand the condition of the nuclear stockpile;
- Sustain the current stockpile of U.S. nuclear warheads; and
- Undertake comprehensive weapons modernization.

The NNSA Stockpile Stewardship Program, which underpins confidence in the U.S. nuclear deterrent, has been successful since its inception in 1995, largely as a result of HPC-based modeling and simulation (ModSim) tools used in the NNSA's Annual Assessment, as well as solving issues arising from Significant Finding Investigations. HPC tools have increasing roles in understanding evolving nuclear threats posed by adversaries, both state and non-state, and in developing national policies to mitigate these threats.

The NNSA's Advanced Simulation and Computing (ASC) Program provides the computational resources that are essential to enable nuclear weapon scientists to fulfill stockpile stewardship and modernization requirements through simulation without underground testing. Modern simulations on powerful computing systems are key to ensuring that we do not return to testing, that we develop and deploy cost-effective and high quality solutions, and that the stockpile can address an evolving threat landscape.

2.2 Evolving Mission Needs

The stockpile continues to move further from the nuclear test base, through aging of stockpile components and modifications involving system refurbishment, reuse, or replacement. The realism and accuracy of ASC simulations must continue to increase over time to track the aging stockpile through development and use of improved physics models and solution methods, which require orders of magnitude greater computational resources than are currently available. In the coming decade, weapon modernization efforts are expected to become a much larger fraction of the NNSA workload, and simulation teams at design agency (DA) sites not only must achieve higher fidelity but also must closely collaborate with production agency (PA) sites to understand their processes, capabilities, and manufacturing constraints.

We expect simulations increasingly to use data and models from across the NNSA complex to ensure that designs are optimized for real-world production facilities. HPC use cases will no longer be confined to studies or workflows conducted at a single site at a time; they will span a web of connected sites across the complex, and sites will iterate on the design process. As the need for HPC expands across the complex, we will need to integrate an increasing number of capabilities into our multi-physics codes, to make these capabilities available to more users across the complex, and to accelerate turnaround times for simulation and analysis runs while ensuring that they reflect data from PA sites so that not only are they consistent with manufacturing processes but can guide them.

High performance accelerators, particularly GPUs, have become an essential tool for NNSA codes, enabling simulations to complete in hours when previously they required weeks. We expect to continue to exploit GPUs for physical simulations into the foreseeable future. However, we must also pursue other ways to accelerate mission work, including artificial intelligence (AI) and increased automation.

AI is emerging as a tool to guide and to accelerate large simulation workflows, to understand and to model complex scientific phenomena better, and to automate critical non-simulation tasks (e.g., image or document analysis) in mission workflows. Fundamentally, AI will increase our ability to synthesize knowledge from our data. Increasingly, NNSA workloads will be coupled with AI, leading to much more complex data-centric workloads in our HPC centers.

Next-generation NNSA workloads will integrate large-scale simulations and simulation ensembles alongside AI training, AI inference, complex data lakes, live data streams, services, storage, and extensive automation. NNSA HPC workloads will be central to a connected NNSA complex where users at remote sites leverage simulations, models, and databases hosted at LLNL. Conversely, we also expect LLNL-hosted simulations to leverage models and data from production facilities, enabling more accurate simulation scenarios through rapid iteration between the DA and PA sites. For the vast majority of our AI use cases, we anticipate the need to generate training data with simulations. No other means exists to get enough data to use modern AI techniques. As such, we will need to couple simulation ensembles tightly with AI training—data generated by simulations must be available to train models rapidly and iteratively.

These scenarios depend increasingly on the ability of users and facility staff to leverage automation. These scenarios depend increasingly on the ability of users and facility staff to leverage automation. Users must automate large scientific workflows. Developers must automate building, testing, and deployment of their code. The facility and users must automate infrastructure and service deployment. Code teams must automate their development and testing workflows to implement new versions of codes, which are regularly deployed to the center. These workflows will need to be augmented with ML-ops, so that developers can as easily depend on versions of ML models as they currently depend on software packages. The scientific and AI library landscape is constantly changing, and users must also be able to test rapidly with the latest versions of internal and external scientific and AI packages. Launching of complex workflows with multiple services and applications must be automatic and secure. Further, to serve remote customers, users and staff must be able to deploy web services and other front-ends to our simulation infrastructure rapidly. Regular updates and rapid development will be critical to exploit the full power of next-generation NNSA systems without sacrificing correctness or end user productivity.

3.0 Future Infrastructure Vision

With over 3.25 double precision exaflops in peak compute capability, around 100 MW of electrical power and over 30 kilotons of cooling being delivered to our computer room floors, LC is a world-class data center. While LC currently meets its mission need, we see ways to improve LC to meet those mission needs better. The current state of the HPC center is one in which we procure entire integrated systems and the many systems that we site largely operate independently. Incremental upgrades of center resources are difficult and often entail changes to other resources just to accommodate the desired improved capability. Ensuring the security of the resources and the jobs that run on them often requires choices that limit center-wide flexibility and impede meeting the center's overall purpose. Users must be fully aware of the resources on which their jobs run. They must submit jobs through batch schedulers and decomposing their workflows into multiple interacting components is, at best, poorly supported. Execution of overall workflows that require disparate resource types must be manually coordinated and optimized. Ultimately, overall center efficiency suffers from the state of HPC center system management.

We envision an evolved HPC data center that facilitates greater efficiency in using center resources and, most importantly, greater productivity of its users. To support NNSA's anticipated mission needs, LLNL

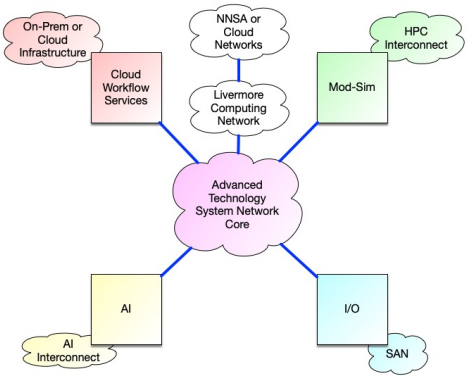

Figure 1. Conceptual Center Decomposition

aims to transform LC into a converged data center for cloud and HPC-style workloads. HPC, AI training, AI inference, web services, analytics, and continuous integration (CI) jobs should all be well supported, and users should be able to employ heterogeneous compute and storage resources seamlessly to support complex workloads and orchestrated workflows. Future systems should enable composable, and yet tightly integrated technologies to serve multiple purposes. Effectively, the center should become the system and the system should not be optimized only for any one workload (e.g., ModSim or AI) although large fractions of it may be. For example, the system may have a GPU partition that works well for both AI and HPC, and another CPU-only partition that is optimized for service workloads. Center operations should support incremental updates that target specific improvements in capabilities and procurement of homogeneous, tightly integrated resources should be strictly in order to exploit opportunities for greater efficiency in the use of the procured resources. Diverse systems resources should be able to be co-scheduled and orchestrated using common interfaces. Ultimately, users should not care what resources are used to execute portions of their workflows, only that their workflows are executed quickly and efficiently.

To support efficient execution of the wide range of workloads that meet LLNL's mission need, future computing capabilities must support automated infrastructure provisioning so diverse compute, storage, and networking resources can be assembled and tailored to particular jobs, applications, and workflows. Some workload scenarios may require the computing power of the entire platform, taking days or weeks to complete, but also must simultaneously store, consume, analyze, transform, and train on data from such jobs. Other use cases may require millions of small jobs that use the same computational resources. Users will need to provision and to manipulate the system programmatically through well-defined APIs. The system must support rapid development of new codes, services, and workflow tools, and web interfaces must support secure availability of these persistent services to users across the NNSA complex.

We see the required transformation of HPC centers as their convergence with cloud technologies. We anticipate funding developing concepts and studies that will contribute not only to extending LC but also enhancing it with next-generation high-performance cloud capabilities. These R&D technologies technologies are intended to contribute to the transformation of LC into a hybrid high-performance cloud provider. In this model, LLNL will integrate the components and operate the data center, using a common, open-source software stack to provision hardware resources flexibly and securely. LLNL cannot do this alone; it must procure *components* of an HPC/cloud data center from hardware vendors. However, these components must work together as an integrated whole, *even if they come from separate providers*. LLNL anticipates issuing a future formal RFP seeking R&D proposals for technologies that enable this model.

While LC's transformation will eventually enable more focused procurements, LLNL anticipates anticipates the initial step being a future large, unified procurement, with the bulk of resources being sited in the FY29-31 time frame. This is in direct contrast to past large LC procurements that were awarded one company to deliver a single computer with associated infrstructure. This new approach to procurements will support multiple awards that target specific resource types in order to establish a clear baseline from which the LC hybrid HPC cloud center can evolve to meet NNSA ongoing needs. The procurement that is

the subject of this document will inform LC's initial vision laid out here and influence the approach to our future procurements for hardware. Figure 1 depicts a conceptual decomposition of the resources to be deployed under that procurement, with major components including: a next-generation, integrated ModSim capability; a component to train AI models efficiently and to use those models for inference in conjunction with ModSim activities, enabling a Cognitive Simulation (CogSim) workflow; compute resources highly optimized to support persistent data-intensive services; center-wide storage resources; and a high-speed system-wide network to support the composition of these resources to execute complex workflows, and jobs within those workflows, efficiently.

LLNL welcomes responses that detail support for the desired transformation of HPC centers. The remainder of this section details some specific topics to be addressed to support it.

3.1 Security as a First-Class Citizen

Security is paramount in this transformation. HPC centers traditionally implement security using OS-level controls and disparate network zones, but this solution is too coarse-grained for our future needs. Future systems must incorporate robust security measures at every level, from hardware to software, ensuring data integrity, confidentiality, and availability. This incorporation will include (but is not limited to):

- **Multi-Tenancy**: Efficient use of the system requires that resources can be used flexibly and securely by multiple users at different security levels. Users should be able to compose arbitrary subsets of system resources to run jobs, workflows, and services as needed for different workloads, and no portion of the system should be dedicated to any one security level. While we still expect to use an air gap to separate classified from unclassified workloads, on either side of the air gap we should be able to use strong *logical* rather than physical separation to ensure secure separation between jobs across the center and within a node.
- **Strong Isolation**: Serving a diverse set of users and missions requires that users are isolated from each other, that is, that one user cannot observe not only the data of another user but also whatever that user is doing without permission. This model must support fine-grained partitioning of system components (e.g., nodes, GPUs, CPUs, networks, and storage) while maintaining this heightened level of security for each partition.
- **Flexible access control**: Modern workflows across diverse teams require that users have greater flexibility and control of access to their data while still ensuring that the data is secure. Advanced authorization and authentication mechanisms must protect data, must allow the owner of the data to determine who has access and must ensure that only those authorized users can access it. In addition to traditional user-based authentication, role-based authentication is also a critical capability to enable data-centric workflows. Filesystems and other system-level data services should support user-level authorization mechanisms so that a compromise of any node does not result in a breach of an entire filesystem.

In short, *resources in the computing center should be easily usable for any workload*. The type of job should not matter—a given piece of hardware should be allocatable to a batch job, on-demand and orchestrated jobs, a persistent service or an untrusted CI job without changing the architecture or security model of the data center. The size of a job should not matter—users should be able to allocate small jobs (down to fractions of a CPU or GPU) and large jobs (all nodes in the system). The *user* running a job should not matter—we should have confidence that users on the same node cannot access each other's data without authorization. Storage resources should be usable for large-scale parallel filesystems (e.g., Lustre), for small block volume claims, or for object storage. Multiple tenants, potentially from different

programs or laboratories, should be able to run separate jobs on the same network or within a single node. Users should be able to build secure, persistent services without assistance from facility administrators.

NNSA developers work across a wide range of environments, from completely open GitHub repositories to internal unclassified machines to air-gapped systems. NNSA software necessarily leverages many open-source libraries in addition to internally developed libraries and tools, and often developers need to run tests with on-prem hardware in response to new releases or changes in external software. Such integration testing ensures that on-prem applications continue to function reliably. Security restrictions currently prevent us from running untrusted tests on-prem, but we expect that next-generation systems will need sufficient isolation (e.g., network and compute virtualization or containerization) to allow it.

Cloud systems provide this degree of isolation for their users, and overheads of isolation technologies like virtualization, software defined networking, trusted execution environments, and encryption have become low enough to be considered in HPC environments. That traditional HPC environments lack the guarantees that these technologies provide is a major barrier to unifying on-prem infrastructure with cloud technologies, and we welcome research directions that lower the performance overheads and costs of using these technologies, as we expect them to be critical for future mission workloads.

3.2 Integration and Flexibility

LLNL will integrate cloud and data services into a leadership-class HPC center using a common open-source software stack, common system administration practices, and industry-standard APIs where available. We seek R&D into open standards that support seamless integration of heterogeneous hardware and software components from multiple vendors. While we will build on existing standard APIs when possible, new open, standard APIs may be necessary. All APIs should support integration with existing open-source projects (from others as well as LLNL). Most importantly, they must enable selection of all components independently while resulting in a fully integrated data center, enabling separate incremental upgrades of each component of the data center.

3.3 Base Software Environment

LLNL uses TOSS, the Trilab Operating System Stack, which augments Red Hat Enterprise Linux (RHEL) with large-scale system management and configuration support. We expect the guest OS on any new system to continue to be based on RHEL version N or N-1, where N is the latest RHEL release. Any software required to enable hardware should be released in a form that allows LLNL to compile it from source against the TOSS kernel and to include libraries and tools as needed. This software will be distributed with TOSS outside of LLNL, as TOSS is used by other NNSA laboratories and external sites. We strongly prefer that software is provided through a third-party hosting site like GitHub, where LLNL can work directly with the vendor on issues. Software should be clearly version-tagged and should have a well-structured release process. Early and frequent engagement with kernel.org to upstream any needed hardware support patches is required.

LLNL will use Flux (flux-framework.org) to allocate resources and to schedule HPC jobs on all future systems. LLNL also uses OpenShift and Kubernetes extensively. Any APIs or administration interfaces that reserve or manage compute or storage hardware must be open and well specified, and must be manageable by Flux, Kubernetes, and other standard infrastructure management tools (e.g., TerraForm or OpenTOFU).

3.4 Infrastructure-as-a-Service (IaaS)

LLNL aims to make on-prem hardware manageable through low-level IaaS calls similar to those in the cloud. We require provisioning of resources at the center to be manageable through open, industry standard APIs, e.g., Sunfish, Redfish, or de-facto standard APIs like S3 and TerraForm providers. If an industry standard does not exist for a given API, we expect the API to be open and well documented. APIs will support management of nodes, VMs, network, storage, AI processors, volumes, and other system hardware. Both in-band and out-of-band management support are required for finer grain access control with strong isolation.

We seek to build the baseline services required to run an HPC center like an on-prem cloud. The on-prem cloud does not need to include *all* services that major industry cloud providers provide, as the support burden and required investment would be too high. However, we will provide *at least* baseline storage, compute, and network allocation services from which higher-level services and/or Platform-as-a-Service (PaaS) systems can be constructed, either by users or facility staff. For example, ideally facility staff will maintain resources for large, center-wide filesystem services using the same underlying interfaces that a user would leverage to construct their own smaller, job-specific filesystem. These capabilities will enable facility users to construct their own persistent services, frameworks, ensemble runs, and workflows rapidly. Center users should be able to find TerraForm scripts, Helm charts, and other Infrastructure-as-Code (IaC) solutions online and to adapt them quickly to provision their own services and HPC workflows securely at our center.

3.5 Flexible Networking and Storage

We anticipate that future LC workloads will be increasingly data-centric, and users will need to manage large, centralized data sets. Often these data sets will be generated by simulation codes as part of a workflow, and jobs that use the data (e.g., AI training or analysis) may differ significantly from jobs that produce the data (e.g., simulation codes). Over time, we expect to add new types of hardware into the center and we must be able to share existing data sets with those nodes.

To support these use cases, we anticipate that we will need both flexible networking and flexible storage hardware. Flexible partitioning of bandwidth to different parts of the system and smooth upgrades of the fabric over time will be essential as requirements for that fabric change for the evolving hardware that comprises the overall system. The network should be virtualizable, so that traffic from multiple tenants can flow securely within and between partitions without risk of compromise. QoS features at all levels will be necessary, e.g., to allow for simultaneous interactive use and bulk data transfer. Our current center networking model deploys high-speed interconnects (e,g,, Infiniband, OmniPath, Slingshot) within each cluster, and from each cluster to parallel filesystems. For future workloads, we will need to integrate networks across system components over time, integrating high-speed interconnects across systems.

In addition to balancing network bandwidth throughout the data center, we require flexible and varied use of storage resources. Currently, on-node storage in our systems is mainly used for filesystems, and users cannot easily assemble or allocate their own storage services. To satisfy data-centric computing needs, users should be able to allocate nodes with local storage to compose their own storage services, for use by applications or by several applications in a workflow. Data storage should be able to be allocated independently of compute resources to support data ingestion pipelines from edge experimental facilities such as NIF, the National Ignition Facility. If a user needs a filesystem, a database, object store, or simply near-node storage, flexible storage should be available with sufficient bandwidth to compute resources.

3.6 Revolutionizing Productivity

A converged infrastructure center will enable a revolution in developer productivity and time-to-solution. Armed with the automation, collaboration, and security capabilities of the cloud, we will have the tools to integrate and to accelerate AI and traditional simulation, leveraging the latest data and the fastest hardware for each application. While exascale brought accelerated computing into the mainstream of HPC, the post-exascale era will bring the software tools we need to use diverse hardware productively in workflows.

4.0 Future HPC Center Use Cases

This section describes several emerging workflows that our future HPC Center will support. This scenario is not intended to be all encompassing. It is an example intended to give readers an idea of how the envisioned technologies will be used in practice.

4.1 Modeling and Simulation coupled with AI

4.1.1 Cognitive simulation

We expect that traditional numeric ModSim codes will be enhanced, both with embedded AI surrogate models tightly incorporated into multi-physics, multi-scale simulations, and with AI orchestration models driving simulation campaigns. We call this “cognitive simulation”. Emerging workflows will require a blend of traditional double precision floating point operations for ModSim codes, tightly interwoven with AI-centric, low-precision floating point for “small” model inference, likely on the same node. The orchestrating AI model that launches ModSim jobs will also need to run on AI-capable hardware, at modest to large scale.

4.1.2 Digital Twins

In addition to AI integrated directly into the simulation workload, we anticipate increased use of digital twins to model components produced at PA sites (e.g., Y-12, Kansas City National Security Complex (KCNSC), Pantex). These sites manufacture components according to designs produced by DA sites. Differences often arise between designed components and manufactured components. A digital twin of a 3-D printed part can be used as an input to a traditional simulation, enabling parts to be “born certified” via increased simulation accuracy. We expect that AI models will serve two roles for digital twins. First, AI models will be used to monitor physical systems and to generate streams of measurements based on the monitoring data. Second, AI models will be used to monitor and to integrate multiple streams of measurements to annotate existing digital models. In both of these use cases, the computational needs of the AI models are more focused on real-time execution and on-demand scheduling to be consistent with live experiments at the production facility.

4.1.3 Inverse Design

Generative AI models can be used for inverse design capacity, specifically to explore a design space rapidly and to identify key regions that should be examined with further ModSim runs. Typically in this type of workflow, thousands or more traditional ModSim runs are orchestrated with scientifically varied inputs. The outputs of the simulations are collected and the full input and output data are used to inform a human designer or to train a surrogate model of the simulation. These types of generative models can be substantially more expensive to execute than a typical surrogate model, but lighter weight than the orchestration models discussed in Section 4.1.1. These generative models could benefit from AI accelerators near the compute processors or at the rack-level.

4.2 AI Workloads and their Computational Motifs

We expect that our users will begin to exploit AI in even more scenarios than the three mentioned in 4.1. A more complete list with computational characteristics is:

1. Cognitive simulation (4.1.1): HPC ModSim integrated with AI such that many small inference requests short circuit calculations, tightly coupling AI and 64-bit interaction;
2. AI-augmented simulation campaigns (4.1.1): Using an AI model (possibly a large reasoning model) to orchestrate modeling and simulation, which will include LLM inference, traditional ModSim with coupled surrogate models, surrogate model re-training, and orchestration and fine-tuning of LLMs;
3. Inverse-design (4.1.3): LLM or other model to orchestrate batched jobs, modeling and simulation jobs along with AI surrogates;
4. Specialized foundation model development: Large-scale, compute-intensive training of transformer models for a specialized scientific domain, possibly coupled with modeling and simulation to generate data;
5. Creation of data-surrogate models: Train a DSM (domain specific model) to represent and to compress a multi-modal data set (including rare events); and
6. HPC Code assistant: LLM assists with porting HPC code, likely run as a persistent service on AI-capable nodes.

For each of the models above, the training requirements vary substantially, from hundreds of compute hours to potentially exaflop days for the largest models. The data I/O requirements for hybrid AI workflows differs substantially from a traditional ModSim checkpoint paradigm:

1. Data sets will range from Terabytes (TB) to Petabytes (PB) and contain up to billions of trillions of samples / tokens;
2. Storage systems will be required to serve complex sets of this data in a read-mostly, near random-access pattern, or allow for in-situ ingestion of data streams for online training;
3. Provenance of the data will become a first class property that is critical for understanding the fidelity and veracity of trained models;
4. Data may be sourced from real-time edge experimental facilities, such as NIF, the Advance Manufacturing Lab (AML), the Vera Rubin telescope array, and the Scorpius laser facility; This live data is crucial for digital twins, but also the calibration of ModSim codes; and
5. Finally, as models become part of CogSim workflows, reproducibility and auditability require the ability to recreate exact workflows, including specific variants of the models.

Models that are used for inference and training will range from millions of parameters that run efficiently on portions of a modern GPU or AI accelerators, up to trillions of parameters that require hundreds to thousands of accelerators just to load, let alone train efficiently. Training costs of the most sophisticated models will stretch into many exaflop days, with a exaflop month(s) of ModSim runs and large surrogate model inferences required to generate the supporting training data set.

4.3 Services and Cloud Capabilities

A majority of these workloads require significantly more services to orchestrate large sets of runs, training, inference, and model updates than HPC centers have previously encountered. Unlike prior workflows, these scenarios require the HPC user to manage the dynamic scheduling of their own ensemble runs. Moreover, the workflows require an intelligent system that can place different types of jobs on the most appropriate resources, at different scales, in conjunction with needed data. These types of workflows require thinking about more than MPI batch jobs—services communicate through many more network layers and libraries. In many cases, the granularity of computational work is much finer than what we have seen in the past. In large ensembles, the training jobs may be small and we may need to run several tens, hundreds, or thousands of small simulations to amass sufficient training data.

Our anticipated future workflows clearly need co-scheduled services, deep resource awareness, and strong user isolation. The center must ensure that services running do not unintentionally expose or leak data to other users, and orchestrators on the new system will need to ensure that analysis, training, and inference jobs from the same workflow are efficiently co-scheduled.

4.4 Data-centric Computing in a Connected Complex

For the digital twin use case, we must deploy persistent services, not only to orchestrate jobs *within* our center but to connect us with production agencies *outside* it. We will need frequent updates of data from the PA sites sent to persistent services to update and to redeploy our models.

Nearly all of the new ML components require careful placement near training or analysis data. Currently, our large data sets reside in filesystems and tape archives across sites, but we anticipate that new workloads will need to manage data sets differently. We will keep persistent data warehouses, data lakes, and large data sets that will need to be used across multiple sites. Regular cross-site replication, versioning, and updates of these data sets will be critical, as will staging of compute jobs near data and data near appropriate compute resources.

Many NNSA data sets are subject to need-to-know restrictions, so flexible access controls as well as flexible data movement will be essential. We cannot allow unauthorized access to data sets, but we do not want users who should have access to struggle to gain it. Sharing should be simple, fine-grained, and should be possible without excessive copying or wait times. The connected complex requires that we enable users to work together efficiently within and across sites.

4.5 Developer and Operational Workflows

ML models and simulations must be versioned and managed like code to support these workflows, and the deployment tools for developer and operational workflows are typically distributed services. Developers need Continuous Integration (CI) to trigger easily both for code and model changes that occur inside the center, at trusted sites outside the center, and for support libraries on external sites like GitHub. ML workloads require frequent model updates and redeployments, and often require a human in the loop. Engineers use tools like Jupyter, Colab, and SageMaker to create, to tune, and to deploy AI models, and ensuring that this type of productive iteration is possible at our HPC center is a critical motivator.

5.0 Topic Areas of Interest

To achieve the vision and notional use case outlined above, LLNL anticipates issuing a future RFP for R&D into technology that could enable LLNL to construct a large-scale, converged hybrid HPC/AI cloud data center. For that RFP, an offeror will be able to respond to *one or more* of the below capability

requirements. The number of capability requirements that an offeror may address will not be limited, and the Laboratory anticipates considering proposals that satisfy some or all capabilities.

We specifically *do not* seek hardware R&D that would entail detailed designs and extensive evaluations through simulation or construction of prototypes. Rather, we are looking for identification of gaps of existing roadmap items towards meeting our goals, along with potential high-level hardware directions that could enable cloud/HPC center capabilities and high-level evaluations of their costs and benefits. Similarly, we *do not* seek software implementations that would entail extensive developer effort. Rather, we are looking for identification of gaps in existing APIs and software with high-level of designs of APIs and software solutions that fill those ga.ps

5.1.1 Modeling and Simulation Software and Hardware

Traditional ModSim will remain the central and likely largest (in terms of cycles) portion of our workload even in our converged data center. We still expect to run full-scale (or at least full scale for the compute-capable portion of the machine), and we expect to be able to run ensembles of millions or even billions of smaller simulation jobs. To support these ModSim workloads, the anticipated future formal RFP will seek proposals for concept development or studies related to, but not limited to the following topics:

1. Descriptions of planned next-generation processors, GPUs, and accelerators optimized for high-precision simulation workloads in a coupled HPC/AI environment;
2. High-level designs of next-generation software technologies that would enable us to partition large GPU nodes to serve many small, isolated jobs better (see also Section 5.1.3); and
3. High-level designs of next-generation software libraries, programming models, compilers, and other technologies that enable the hardware technologies above.

We expect the compute hardware in next-generation systems to be able to interface, via the common network layer described in 5.1.5, with other components and capabilities listed here. The compute portion of the system must ultimately meet traditional procurement benchmark criteria (e.g., the CORAL2 benchmarks or a derivative) for ModSim workloads.

5.1.2 AI Hardware and Software

In addition to an HPC capability, we expect our future system to have an AI capability, which may or may not be distinct from the ModSim compute capability described in Section 5.1.1. The AI capability must support large-scale AI training, suitable to build and to tune in-house, proprietary foundation models as well as domain adaptation of premier models from vendor partners. The AI capability of our next-generation systems must also be capable of running rapid inference with a pre-trained model, potentially as a service for other jobs and other users in the system. The anticipated future formal RFP will seek proposals for R&D into technologies including but not limited to the following areas:

1. Descriptions of planned next-generation AI processors, spatial-dataflow accelerators, potentially for lower-precision or AI-specific acceleration that advance our future HPC center goals;
2. Discussion of modular, chiplet-based accelerators that would enable the composition of domain-specific or bespoke accelerators to be integrated with traditional compute hardware; and
3. High-level designs of novel software techniques or libraries that enable AI hardware to function in a converged data center, or to integrate with ModSim workloads better.

The AI capabilities to be investigated may leverage the same hardware as the compute capability described in Section 5.1.1, or it may be distinct. If it is the same, offeror should describe why an integrated solution would provide greater overall value to NNSA. If it is distinct, why a distinct solution can provide greater overall value should be discussed, where overall value is for the *complete* converged HPC/cloud data center, not value specific to AI or ModSim workloads.

5.1.3 Cloud Capabilities and Services

Throughout our next-generation converged system, we expect strong security and strong isolation at all levels. This requirement is likely the biggest gap between current HPC centers and modern cloud offerings. The anticipated future formal RFP will seek R&D proposals including but not limited to the following areas to enable converged HPC/cloud centers:

1. Roadmap descriptions for next-generation, low-overhead isolation and virtualization hardware or software technologies that could enable users to partition nodes, CPUs, GPUs, AI accelerators, memory and any other relevant resources in a multi-tenant system;
2. Roadmap descriptions for encryption or confidential computing technologies with low enough overhead for use in an HPC or AI-oriented system;
3. Roadmap descriptions for network virtualization technologies that enable isolated (cryptographically or otherwise), user-and job-specific silos within a larger multi-tenant system, such as SmartNICs, DPUs, or encryption techniques; Low or zero-overhead solutions are of particular interest;
4. APIs and design of technologies that enable composable storage in an HPC/AI compute environment; Of particular interest are solutions that enable users to allocate raw storage or compose storage services for particular jobs or workloads, while also enabling the facility to host traditional large-scale storage services; and

- Open APIs, control planes, and their implementations to facilitate automation throughout a large-scale cloud/AI data center, or within a node; We require interoperability and extensibility with existing industry standards and, where standards do not exist, APIs that could establish industry standards are of particular interest.

We are interested in how the *overhead* of virtualization technologies can be lowered to reduce the performance penalty for using these technologies for HPC and AI jobs to as low as possible, while still enabling infrastructure flexibility. Non-performant solutions are unlikely to be usable in LLNL systems.

5.1.4 I/O, Storage, and Composable Storage Services

Just as ModSim will remain the bulk of the workload of the converged HPC/AI data center, we expect a large fraction of I/O to continue to be satisfied by traditional large-scale parallel filesystems, and we aim to continue to maintain and to develop software our existing tape archive system. However, we see an increasing need for a continuum of storage technologies within the data center. A filesystem is, at its core, a service that runs on nodes with local storage, and this storage can be used for much more than one large filesystem. Without losing the full performance of our large-scale parallel filesystems, we aim to allow storage nodes to be allocated by users or facility staff to compose storage services, either for the whole center (or a large fraction of it) or for user-specific jobs and services. Further, storage does not have to be dedicated to one part of the center. It could be concentrated into specific racks, as it is currently, or spread across different components of the center to allow faster access, e.g., from compute nodes.

The anticipated future formal RFP will seek proposals for R&D into technologies that could enable a continuum of composable storage services, including, but not limited to:

1. Work on the continuum of storage services for an HPC/cloud data center:
    - Next-generation large-scale parallel filesystems handle the demanding I/O requirements of HPC and AI workloads (e.g., innovations in Lustre);
    - Innovative near-node storage solutions (e.g., node-local filesystems over PCIe or CXL);
    - Next-generation object storage solutions (e.g., MinIO, Ceph, DAOS);
    - Next-generation filesystem capabilities for large-scale systems; and/or
    - Databases, key-value-stores, and other services;
2. Simple storage services/APIs that can be used as building blocks for larger services, e.g.:
    - Network filesystem/object services (e.g. NFS, SMB, S3);
    - Block mount services (e.g., EBS, Google Block, Azure Block, Oxide Crucible);
    - User-level persistent volume claim services;
3. Improvements to provisioning speed of existing solutions to enable them to work better in IaaS scenarios; For example, mounting Lustre filesystems can be slow, but mounting and unmounting Lustre filesystems for specific jobs instead of mounting Lustre across LC is desirable so improving Lustre mount time (or mount/initialization time for other services) is of interest.
4. Services that can intelligently allocate storage mounts to VMs;
5. Infrastructure for managing heterogeneous storage resources across a data center; and
6. Software innovations to support new use cases for composing storage.

5.1.5 Data Center Interconnection Network

As mentioned above, we aim to build our converged HPC/AI data center on a common fabric that supports the gradual upgrade of the center over time and allows flexible partitioning of bandwidth to match compute and data needs throughout the center. Historically, the design of LC has focused on high speed interconnects for intra-cluster and cluster-filesystem traffic. We anticipate that we will need higher bandwidth between different portions of the next-generation data center, but that these connections will need to change over time. Moreover, we anticipate that traffic will need to be virtualized and isolated both within and between homogeneous portions of the larger heterogeneous data center. To facilitate this, the anticipated future formal RFP will seek proposals for R&D enabling technologies including, but not limited to the following areas:

1. High-level designs of novel networking fabrics, switches and architectures that support flexibly partitioning bandwidth throughout an entire data center;
2. High-level designs of novel network security solutions or implementations that can enable software-defined networking and isolated network silos within a larger data center;
3. Discussions of RDMA (e.g., RoCE, Ultra Ethernet, next-generation Infiniband, etc.) support between endpoints that may be located in different clusters or partitions in the next gen center;
4. High-level designs of technologies that enable fabric upgrades over time as the center changes;
5. Exploration of techniques to support very low-overhead network isolation, including but not limited to cryptographic isolation, SmartNIC or DPU-enabled isolation;
6. High-level designs of new technologies to facilitate provisioning of storage (persistent volumes, remote filesystems or services, etc.) across the fabric and within partitions;
7. High-level designs of novel QoS features to enable interactive and latency-sensitive workloads alongside bulk data transfers;

8. High-level designs of novel technologies that could bridge the next-generation network with other fabrics (e.g., Slingshot, Infiniband, OmniPath, Ultra Etherne);

- High-level designs of novel technologies to understand, quantify, verify, and ensure the reliability of large-scale data center networks; and

9. Software designs to enable these capabilities or the management of such a network.

5.2 Provided Open Source System Software Stack

LLNL leverages an open source software stack to administer the current HPC data center. The stack includes:

- TOSS (a derivative of RHEL) as the base operating system and the expected guest operating system for HPC and AI workloads;
- Spack to manage user-level software above TOSS;
- Flux as the system-wide resource manager, integrating with Kubernetes, OpenShift, or other mechanisms for service job and storage deployments;
- RAJA, CHAI, Umpire, CAMP, and OpenMP for performance portability;
- Other open source tools such as Ansible for configuration management and GitLab for source control and DevOps workflows and
- In-house development of Lustre and ZFS on Linux for parallel filesystems.

LLNL aims to expand the open source management framework to include tools and APIs needed to administer the converged data center. As mentioned above, any libraries needed for hardware support on HPC resources should be able to compile with TOSS (N or N – 1 RHEL kernel release). We strongly preferr that any additional tools, hypervisors, low-level libraries, IaaS implementations, system services, etc. be implemented as open source and also available from well-known open repositories (e.g., GitHub, GitLab) for open collaboration between LLNL staff and respondents. Responses must discuss how proposed infrastructure can integrate into LLNL's open source facility management framework and should leverage and contribute to LLNL open source if it addresses part of the solution.

6.0 Response Structure and Requirements

The anticipated future formal RFP will include technical requirements with value / priority designations, defined as follows:

(a) Mandatory Requirements designated as (MR)

Mandatory Requirements (designated MR) are performance features that are essential to the Laboratory's requirements, and an Offeror must satisfactorily propose all Mandatory Requirements to have its proposal considered responsive.

(b) Target Requirements designated as (TR).

Target Requirements are features, components, performance characteristics, or other properties that are important to the Laboratory, but that will not result in a nonresponsive determination if omitted from a proposal.

6.1 Requirements for R&D Investment Areas

Detailed requirements for each of the targeted R&D areas of investment are discussed in this document. A single proposal may address multiple areas of investment, that is, an Offeror need not submit a unique proposal for each area of investment on which it chooses to propose. Each proposal shall address all of the common MRs listed below. All of the MRs in each area of investment shall be included in the proposal.

6.2 Mandatory Requirements

The following items are mandatory for all proposals. That is, they must be present in any proposal for that proposal to be considered responsive and eligible for further evaluation.

6.2.1 Solution Description (MR)

Offeror shall describe the proposed R&D, with emphasis on how it will provide improvement in the targeted R&D area. Offerors shall discuss the innovative nature of the proposed R&D. Work that funds engineering for a company's current roadmap is not acceptable. Technology acceleration is acceptable if there is a clear benefit, and it is part of a broader strategy. The primary intent is to fund long-lead-time R&D objectives where significant advances can be made during the term of this program.

6.2.2 Research and Development Plan (MR)

Respondents to the anticipated future formal RFP shall provide a plan for conducting the proposed R&D, including timelines, milestones, and proposed deliverables. Deliverables shall be meaningful and measurable. Pricing shall be assigned to each milestone and deliverable. A schedule for periodic technical review by the DOE laboratories shall also be provided.

Projects may propose a high-level hardware design that explores the proposed concept, but detailed hardware evaluations and prototypes are beyond the scope and intent of the anticipated future formal RFP. We seek to leverage existing planned hardware capabilities and leverage them for this vision. High-level analyses that assess the impact (or feasibility) of a proposed development are within the scope.

We recognize that innovation involves risk. Proposals shall discuss technical and programmatic risk factors and the strategy to manage and to mitigate risk. If the planned R&D is not achieving the expected results, what alternatives will be considered? The amount of risk must be commensurate with the potential impact. Higher risk projects may be acceptable if the impact of the project is also high.

6.2.3 Staffing and Partnering Plan (MR)

Offerors shall describe staffing categories and levels for the proposed R&D activities. Any collaboration with other industry partners and/or universities shall be identified.

6.2.4 Project Management Methodology (MR)

Offerors shall provide a project management plan that includes quarterly milestone status updates.

6.3 Target Requirements

The following items are not mandatory but will contribute to a strong proposal.

6.3.1 Productization Strategy (TR)

Offeror shall describe how the proposed technology will be commercialized, productized, or otherwise made available to customers. Offerors should identify target customer base/market(s) for the technology.

Offerors should describe impact specifically on the HPC market as well as the potential for broad adoption. Solutions that have the potential for broader adoption beyond HPC are highly desired. Offerors should indicate a projected timeline for productization.