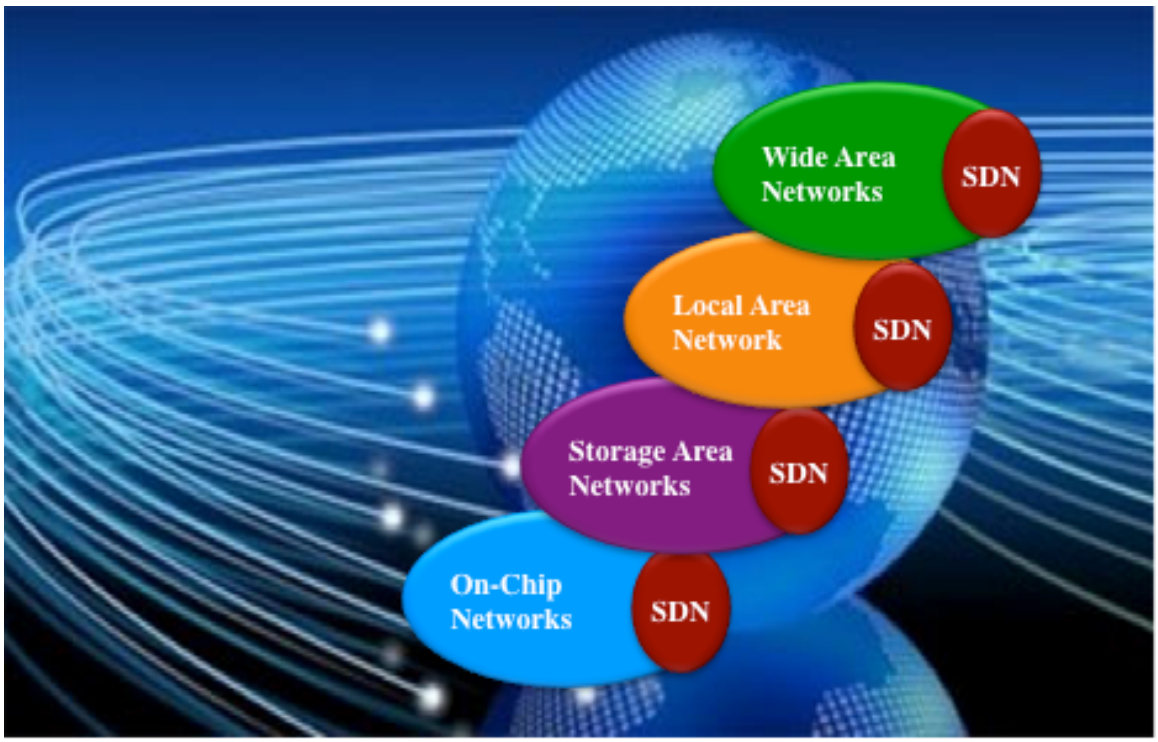
Software Defined Networking for Extreme-Scale Science: Data, Compute, and Instrument Facilities



Report of
DOE ASCR Intelligent Network Infrastructure Workshop
August 5-6, 2014

U.S. DEPARTMENT OF ENERGY | Office of Science

ASCR

Editors:

Tom Lehman, University of Maryland at College Park
Inder Monga, ESnet
Nagi Rao, Oak Ridge National Laboratory
John Wu, Lawrence Berkeley National Laboratory

Table of Contents

Executive Summary

The performance of several DOE Extreme-scale science workflows critically depends on robust, adaptive, high-performance network connections over open data infrastructures, all optimized for scientific productivity. These workflows span a variety of tasks, ranging from providing access to large simulation datasets to science users across the globe, to utilizing remote supercomputers to drive experiments at large science facilities. The extreme-scale of these tasks presents unprecedented challenges to networking technologies in achieving (i) high data rates for massive data (Exabyte) transfers, and (ii) performance predictability and stability for complex computing and instrument workflows. In particular, they require flexible, high-performance end-to-end solutions that integrate diverse components such as large science instruments, supercomputers, multi-domain local/wide/storage area network connections, and distributed storage and file systems. Empowered by *virtualization* technologies, recent developments in large-scale data centers have led to unprecedented levels of flexibility, scale and automation in their deployment and operation. Indeed, they made it practical to install and operate large complexes of servers and storage systems in flexible and agile configurations, using powerful automated software. In networking infrastructures, analogous developments in the form of *Software-Defined Networks* (SDN) represent an enormous potential for substantial productivity improvements in DOE Extreme-scale science workflows.

A broad spectrum of DOE distributed science workflow capabilities can be potentially enabled by SDN technologies: (A) *cyber-powered* experiments can utilize networks to transfer large datasets to remote analysis sites and return complex computational results in time for next steps, (B) *distributed, replicated repositories* can exploit networks and distributed storage to house structured datasets for global access by user hierarchies, and (C) *coordinated, computing and analyses* schemes can combine high bandwidth and stable data flows to remotely on-the-fly steer computations and store datasets. Each such capability can benefit multiple DOE science areas, but requires a specific combination of networking, data and facilities solutions, which must be optimized to DOE environments. The underlying networking solutions, however, require complex SDN technologies; in particular, some of them are beyond the projected trajectories of commercial and general SDN technologies, and are unlikely to be solved as their by-products. Indeed, they require *Software-Defined Networks for Extreme-scale Science* (SDN-ES) technologies that seamlessly integrate networks, experimental facilities, and computing and data systems at the extreme-scale in terms of data volumes and workflow complexity.

There is a need to thoroughly evaluate the emerging SDN technologies to leverage and enhance them for optimized science network solutions that are easily deployed and maintained across the DOE complex. Research and development efforts must encompass multi-site, multi-domain connections for distributed storage and file systems, computing systems, and experimental facilities, which must be co-scheduled and complemented by high performance data transport tools and systems. In particular, these SDN-ES technologies must be developed and optimized specifically for DOE science applications by establishing a focused task force and a dedicated experimental testbed, and also by engaging sites and science users to establish adoption paths for production environments.

1 Introduction

Extreme-scale DOE science applications of DOE include the fundamental study of the complex natural, scientific and technological phenomena [8]. Such scientific activities are frequently carried out by worldwide collaborations of researchers utilizing geographically dispersed and diverse resources such as large science instruments, data storage systems, leading-edge supercomputers, and visualization and other facilities [1]. This type of complex workflows are critical to many DOE science areas including biology, climate science, chemistry, materials science, nuclear science, computing, physics, and others [3,5,6,10,11]. Many of these distributed scientific workflows rely on the DOE network resources such as the ESnet wide area network and the individual laboratory networks [2,9]. The overall ecosystem for these collaborative science efforts also involves the higher education network infrastructure consisting of Internet2, the regional networks, and the site infrastructures. In particular for DOE extreme-scale science, these high-performance networks are critical for carrying out complex workflows by providing access to science instruments, supercomputers, and complex datasets. Indeed, robust and well-engineered high-performance networks enable scientists to collaborate effectively, and move large data sets over long distances.

The DOE domain science disciplines are continuously evolving and innovating, and are reaching unprecedented complexity and scale as progress is made towards Extreme-scale data and computing [1]. The result is increasingly sophisticated and complex science workflows that subsequently require unprecedented capabilities from the network infrastructures. We are in a particularly intense phase of this innovation cycle now with the combined emergence of Big Data, extreme-scale computing, and sophisticated large science instruments. However, recent innovations in high-performance networks are lacking when compared to those in high-performance computing and related computational science enabling technologies. Despite the extraordinary advances in computing technologies in the past two decades, basic communication network capabilities have not changed significantly; networks operations and multi-domain connections involve significant effort. Scientists continue to spend considerable effort and time moving a few terabytes of data or integrating data movement in network-intensive workflows. Furthermore, today's networks are rigid, over-provisioned, and unable to effectively meet the requirements of high-end, network-intensive science applications, especially at extreme-scale. The potential of networks is locked into switching devices that are inaccessible to third party technologies. In this approach, network innovations are limited to vendors' proprietary solutions that often do not take into account science users' evolving requirements, especially those of extreme-scale science.

Unlocking the immense potential of communications networks to support network-intensive workflows is critical to enabling continued scientific innovation. There are many challenges that have to be overcome in order to achieve this objective. Among the many challenges that have to be addressed are: a) refactoring existing statically configured networks to provide agile, intelligent, and guaranteed services across autonomous network domains; b) making the abundant bandwidth available in backbone networks accessible to network-intensive, extreme-scale applications; c) reducing network complexity and make them easy to use and available to scientific applications and workflows.

The DOE community is not alone in its observation that networks now represent a major bottleneck with respect to provisioning agility and resource management flexibility. Empowered by *virtualization* technologies, large-scale data centers have reached unprecedented levels of flexibility, scale and automation in their deployment and operation [16]. Indeed, it is now practical to install and operate large complexes of servers and storage systems in flexible and agile configurations, using powerful automated software. Having realized great benefits from the virtualization innovation in the end-system, compute, and storage spaces, the commercial industry is now turning its attention to the network infrastructure. The emerging *Software-Defined Networking* (SDN) technologies are part of a network infrastructure innovation cycle that holds an enormous potential to close this gap [15]. SDN is expected to greatly change the way networks are constructed and operated in the future. The high level objective is to apply virtualization concepts to networks with hopes to realize innovations similar to what has been seen in the host and storage space where these technologies resulted in new paradigms and use models.

The result is that network architectures, designs, and feature sets are currently on the edge of a paradigm shift which is more significant than anything that has happened in networking since the wide spread deployment of Dense Wavelength Division Multiplexing (DWDM) in the late 1990's. Commercial Data Centers have been the main driving use case for adoption of this SDN innovation and evolution. These technologies are now making their way to other network infrastructures such as large operators who can afford to build their own Greenfield networks. It is unclear how these technologies will be deployed in the more traditional network infrastructures such as enterprise, metropolitan, and wide area networks. However, it seems clear that a shift is on the way for these network infrastructures as well, and indeed *they represent an enormous potential for substantial productivity improvements in DOE Extreme-scale distributed science workflows*.

As the momentum for SDN based network technologies is building in the commercial space, there is a need for the DOE network community to thoroughly evaluate how these next generation network technologies can be leveraged to support DOE extreme-scale science through next generation network infrastructures. The needed networking solutions, however, require complex SDN technologies beyond the projected trajectories of commercial and general SDN technologies; indeed, the full promise of these networks is very unlikely to be reached as their by-products. Indeed, they require *Software-Defined Networks for Extreme-scale Science* (SDN-ES) technologies that *seamlessly integrate networks, experimental facilities, and computing and data systems at the extreme-scale*. As a result, the DOE needs to develop focused expertise in SDN-ES areas within their network community, and also to conduct research and trials to determine best strategies for Extreme-scale science.

To address these issues, the DOE Advanced Scientific Computing Research (ASCR) Intelligent Network Infrastructure Workshop brought together experts from academia, industry, and national laboratories to: (a) assess current network technologies in light of extraordinary advances in computing and networking technologies and the emergence of Big Data driven workflows, and (b) identify the challenges and opportunities in transforming current network and data technologies to provide intelligent and automated services to accelerate scientific discoveries in the era of distributed extreme-scale computing and pervasive data-centric science.

2 Extreme-Scale Science Drivers

DOE distributed science workflows arise from a broad spectrum of application areas including climate science, high energy physics, material science, chemistry and others [1]. And, the most challenging ones are due to extreme-scale in data volumes and sheer complexity of workflows [7,8]. We first briefly describe two use cases that capture the essence of these challenges and then describe the generic network capabilities that can be empowered by SDN-ES technologies. Additional use cases are in Appendix A.

2.1 Science Use Cases: Networking Challenges

Earth System Grid Federation: The global climate research community has a shared goal of producing a comprehensive assessment of anticipated changes in climate. To aid the access and dissemination of large-scale climate data, the Earth System Grid Federation (ESGF) was formed as a coordinated multi-agency, international collaboration of institutions that continually develop, deploy, and maintain software needed to facilitate and empower the study of climate change. Through ESGF, thousands of users from around the world access, analyze, and visualize data using a globally federated collection of networks, computers, and software. The key functionality of ESGF is to act as the distributed data repository for climate research. One of the routine tasks that require heavy network use is replicating data between sites, for example, the data node at Lawrence Livermore National Lab (LLNL) frequently receives over 5 TB a day from three-dozen worldwide sites. These subsets of data could be defined by a variety of conditions such as climate models used, geographical locations, or simulated time covered, and therefore present a range of different demands on the network infrastructure that supports the distributed data repository. A key challenge is to provide robust, on-demand network connections between the user and repository sites. The current strategy of overprovisioning rigid network infrastructures that span multiple domains is not scalable as the climate datasets reach extreme-scale. The SDN-ES technologies promise flexible network infrastructures, wherein optimal connections can be automatically provisioned at scale for site-to-site and user-to-site connections.

Spallation Neutron Source: DOE operates a number of large science instrument facilities such as the Spallation Neutron Source (SNS) located at Oak Ridge National Laboratory (ORNL). A variety of sophisticated science experiments are conducted at SNS and the neutron events are acquired by the data acquisition system. This stream of events is sent to compute resources co-located with the Oak Ridge Leadership Computing Facility (OLCF). At OLCF, it is translated and stored on a parallel file server which supports live viewing and monitoring. The data rate per instrument is approaching 1 Gbit/s for each of the 19 currently operating instruments. A completed data set after an experiment can be on the order of 100 GBytes. Once on the PFS, data is accessed from multiple systems, from a cluster of high performance computing machines. A number of users are planning for experiments that require sub-minute turn-around time for the data analysis operations. Users are also downloading their data to their own computers outside of OLCF. In the near future, SNS plan to support higher performance data transfer options such as using ESnet data transfer nodes. Currently, various parts of the workflow are configured rigidly over site and long haul networks, which are under different operational domains. And this process is repeated for all workflows which have highly varied data and computing requirements. Using SDN-ES technologies, the required complex connections can be automatically and quickly provisioned to optimize the performance across the experimental facility, supercomputers and data systems.

2.2 Network Capabilities for Distributed Extreme-Scale Science

Use cases described above and also in Appendix A allows us to identify several capabilities and feature sets that will significantly enhance the performance of next generation DOE science workflows, particularly at extreme-scale. We focus here on the network infrastructure as an integral part of a holistic view that considers network, computing systems, storage, and instrument resources together to support science workflows.

Interactive and Adaptive Network Enabled Scientific Workflows: As science workflows become more sophisticated, an ability to interact and adapt in near realtime with the key resources is becoming increasingly important. In this context these resources are typically a workflow specific combination of compute, storage, and instrument resources. The timescales for this interaction and adaptation are typically minutes to hours and may involve some preliminary analysis of data, in order to adjust a compute process, instrument setting, or data access/storage action. This type of run-analyze-adjust-run method has always been part of the science process. However, currently this series of steps often includes long delays associated with offline data movement (i.e., FedEx) methods or over the network data transfer rates that effectively reduce the workflow to a non-realtime process. The next generations of science workflows need to transition to a true near realtime interactive environment. This will require network infrastructures to be more flexible, adaptive, and intelligent as part of its role in providing these capabilities.

Intelligent Data Movements: Each of the capability sets described require fast data movement in support of their specific focus areas. Whether the need is based on near realtime interaction and adaptation, support of a distributed data infrastructure, or an attempt to get compute and data resources together in a timely fashion, there is a need to maximize the throughput for data movement. These types of operations are a limiting factor and key bottleneck for today's workflow. This will be an increasingly limiting factor as the volume of data, and degree of resource and scientist distribution are expected to greatly increase. Much progress has been made in this area with well-engineered edge resources such as ScienceDMZ and the Data Transfer Node (DTN) as two example technologies. However, these capabilities must continue to improve to enable the next generation of science workflows. In particular a true end-to-end data movement paradigm needs to be developed which includes not only the wide-area network and site DMZ resources, but also extends to the local area network, storage area networks, compute, storage, instrument and data storage systems. A tighter integration and coordination between networking, storage area networks, storage systems, and project unique resources will likely be needed.

Smart Services for Distributed Science Big Data: Many of the domain science applications and workflows depend on project specific data distribution, replication, archiving, and access. This typically requires a data storage and distribution infrastructure that allows for discovery and access of project specific data. This data is typically replicated and stored in multiple data depots or repositories that are geographically distributed. Advanced network infrastructures and services that allow for increased flexibility and end-to-end capability awareness are needed to optimize these infrastructures. Intelligent cyber-infrastructure services which can facilitate decisions regarding where to store and how to access data in a workflow context are needed to enable enhancements in these areas.

High Performance Computing and Big Data Integration: The timely combination of compute and data resources is a persistent problem within the domain science application community. In this context, computation may be performed on leadership class supercomputers at DOE Laboratories, distributed compute environments such as Open Science Grid (OSG), or local compute resources. Future science applications and workflows need greatly improved capabilities for flexible integration of data and compute resources. Solutions will likely involve multiple approaches including improved mechanisms to: moving data to the compute; moving the compute to the data; improving remote access of data; feature extraction and data reduction at the sources to reduce the volume of data to be transferred. These new capabilities will leverage the other capabilities described previously in this section. These services will have to be developed in the context of the compute job execution environment and access mechanisms.

Real-time Interaction and Adaptation: Real-time interaction across distance for computer-to-computer or human driven remote control applications are expected to be of increasing interest. These interaction timescales could be an order of magnitude smaller than the near realtime Interactive and Adaptive Workflow Support scenario discussed earlier. The types of applications and workflows for which this will be important will be based on computer-to-computer interactions or ones where there is realtime human interaction such as remote steering operations. Even within these two categories, there are orders of magnitude differences in latency and responsiveness requirements. Both of these types of realtime interactions scenarios are considered longer-term requirements and goals.

3 SDN Enabled Extreme-Scale Distributed Science

A review of the current state of the art of network services highlights the fact that network architectures and services have been relatively stagnant as compared to the innovation that has occurred in the compute and storage system space. Furthermore, a large number of science application domains, in particular DOE science, have not seen big benefits from the last several rounds of networking technology improvements. This has occurred during a time of great compute and storage system innovation, realized primarily in the supercomputing's Exascale efforts and virtualization and cloud spaces for Big Data. As a result, network infrastructures are far behind host and storage system technology with respect to dynamic resources instantiation and provisioning agility. At the same time, DOE extreme-scale science workflows are becoming increasingly distributed and complex, thereby requiring flexible, adaptive and optimizing high-performance networks. The emerging SDN technologies are part of a network infrastructure innovation cycle to help close this gap [15,16]. SDN is expected to greatly change the way networks are constructed and operated in the future. The high level objective is to apply virtualization concepts to networks with hopes to realize innovations similar to what has been seen in the host and storage space where these technologies resulted in new paradigms and use models.

The future network capabilities of next-generation science domain applications (section2) require flexible and seamless integration across multiple resources, namely compute, storage, instruments, and networks. This need is motivated by several paradigm shifts in the science domain application spaces. The first is big data driven, wherein the number of domain science unique considerations is increasing with respect to the location, volume, mobility, and persistence requirements. Another important characteristic is the increasingly distributed nature of the resources (storage, compute, and instrument) needed

by science workflows. While these resources have always been physically distributed, the science workflow use cases are rapidly evolving to require real-time adaptations to adjust the specific resource set they are using. This will require specific capabilities from the network with regard to bandwidth, latency, and rapid re-provisioning.

Increasing scale and complexity of science workflows is driving a need to revaluate the concept of end-to-end, which traditionally focused on network resources. For science workflows, the end-to-end includes all the systems between the data source and sink: SAN, LAN, ScienceDMZ, Regional Network, Wide Area Network, and end-systems. This end-to-end view should be a focus for the network community as part of the development of new architectures to support big data driven science. One goal should be to provide applications and workflows with a "deterministic performance" environment. That is, while applications will not always be able to have all the resources or end-to-end performance they would like, it should be possible for critical applications to determine what level of performance they can expect on an end-to-end basis. This will allow applications to optimize their workflows for the operational environment.

While this need for flexible resource integration is not new, the technology and capability advances in each of these resource realms represents a paradigm shift where this lack of integration is now a limiting factor While this broad consensus is indeed becoming more in focus, there are still many unknowns about what it really means to seamlessly integrate data, compute, and networking in a manner which provides the flexibility and simplicity that domain science applications require. Advanced networking infrastructures and capabilities are the cornerstone technology to enable this integration. Network attachment is a common and unifying feature around which subsequent resource integration and coordination activities can be organized. For future network infrastructures for DOE science all point to the need for networks to evolve into a flexible, agile, and programmable infrastructure, scalable to extreme-scale. Networks to be able to participate in science application workflows operations as a first class resource on the same level as compute, storage, and instrument resources.

3.1 Software Defined Networking Approach

The emerging SDN paradigm is a promising mechanism to help meet the requirements for the next generation DOE distributed science workflows, if the extreme-scale requirements are adequately met. The key to transforming networking lies in the programmability introduced via the SDN framework. A key focus of SDN is the decoupling of the network control plane from the underlying hardware or data plane layer. The control plane can then be presented as a programmable system. Some are viewing this as network operating system, providing a computing environment that allows traffic control and management functions to be programmed and reprogrammed as needed, to continuously add new software-based innovations. With programmability, advanced concepts such as machine learning, parallel algorithms, and resource virtualization can be leveraged through intelligent automation to improve network performance and ease of use for scientist. Figure 3a depicts this transition from closed proprietary systems to open API based programmable networks.

Fortunately, the SDN is currently a key focus for the commercial network industry. These technologies and feature sets are in the early stages of a new innovation cycle. However, SDN technologies are being developed in the commercial space with a different set of application drivers. While it appears that these technologies can be utilized in science workflows, the DOE network community needs to evaluate, innovate, and test them to determine how to apply these to the distributed science use cases. The opportunities for leveraging the commercial sector expertise and influencing the designs and standards are available with the proper engagement by the DOE network and science community. Furthermore, certain extreme-scale requirements of science workflows may be beyond the expected trajectories of commercial technologies, for example, integration of science instruments and supercomputers within a single workflow. These requirements must be identified so that they can be met by focused R&D strategies.

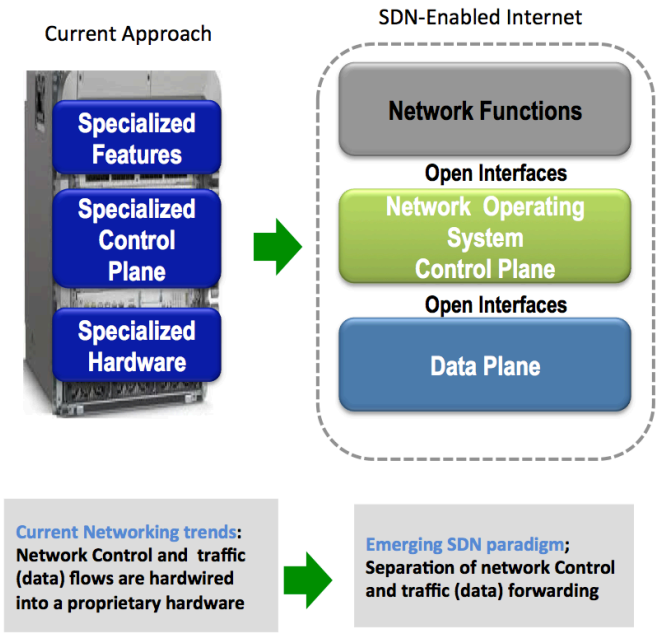


Figure 3a Transition to a SDN Based Network Infrastructure

SDN is a broad term from which many individual technologies are emerging. There are three main concepts that have been defined which capture the SDN core features. A brief overview of these is provided below.

- *Software-Defined Networking*: The focus for SDN is the decoupling of the network control plane from the data plane. This enables network programmability thru an SDN Controller that interacts with the data plane element forwarding engines. OpenFlow is one example of a technology based on the SDN paradigm.
- *Network Virtualization (NV)*: These technologies are focused on two modes. The first is an environment where Virtual Machines (VMs) need to be interconnected in

an agile and dynamic fashion. NV provides mechanisms to build tunnels, or overlays, across an existing network infrastructure to create unique VM interconnection topologies. This mode of NV typically relies on technologies such as Virtual Extensible LAN (VXLAN) or NVGRE (Network Virtualization using Generic Routing Encapsulation) to construct the overlay networks. The other NV mode is one where the network is truly sliced at a network dataplane technology level such as VLANs, DWDM, or some other flowspace. This allows multiple "virtual" networks to exist on one physical infrastructure.

- *Network Functions Virtualization (NFV)*: These technologies build on the NV paradigm to add network functions such as firewalls, intrusion detection, or load balancing features within the virtual or sliced environment.

SDN also provides an opportunity to better manage the heterogeneous nature of the underlying network data plane. The current data plane consists a variety of technologies that includes satellite, wireless, fiber optic wavelength based transport, best effort packet based switching/routing, dynamic circuits, and others. These data plane elements are deployed in many complex environments that often include multi-layer, multi-technology, and multi-vendor configurations. Many of the core feature sets are often locked within their layer/technology/vendor regions. SDN's programmability can be leveraged to manage this complexity and facilitate the design and operation of agile networks suitable for distributed science. There are many areas that require further research and development as it relates to SDN control for multi-layer, multi-technology, and multi-vendor environments. This includes options for a single SDN controller versus a hierarchical or collaborative SDN controller system. Figure 3b shows how an SDN enabled networks may be applied to multi-layer networks.

DOE science applications workflows are generally distributed and multi-domain. A typical workflow includes resources across DOE Laboratories, wide area networks, regional networks, and university campuses. As a result, Federated and multi-domain SDN technologies will be needed. In this environment, autonomous SDN domains will need mechanisms to interact with each other, or with higher-level workflow agents in order to coordinate operations that cross multiple domains. Past experience indicates that commercial development efforts may not focus on these issues due to the business considerations associated with the multi-provider and multi-vendor Internet topology. The R&E community is well positioned to address these issues associated and it is necessary that solutions be developed in these areas. SDN Exchange Point (SDX) is a concept that has been discussed as a mechanism to facilitate multi-domain services. SDXs are well defined points of peering which may offer opportunities to realize a rich policy based automation of network peering and services exchange. SDXs are also envisioned as a mechanism to facilitate the transition to multi-domain SDN infrastructures where non-enabled SDN networks may need to interconnect with SDN enabled networks. Figure 3c depicts a possible federated multi-domain SDN environment.

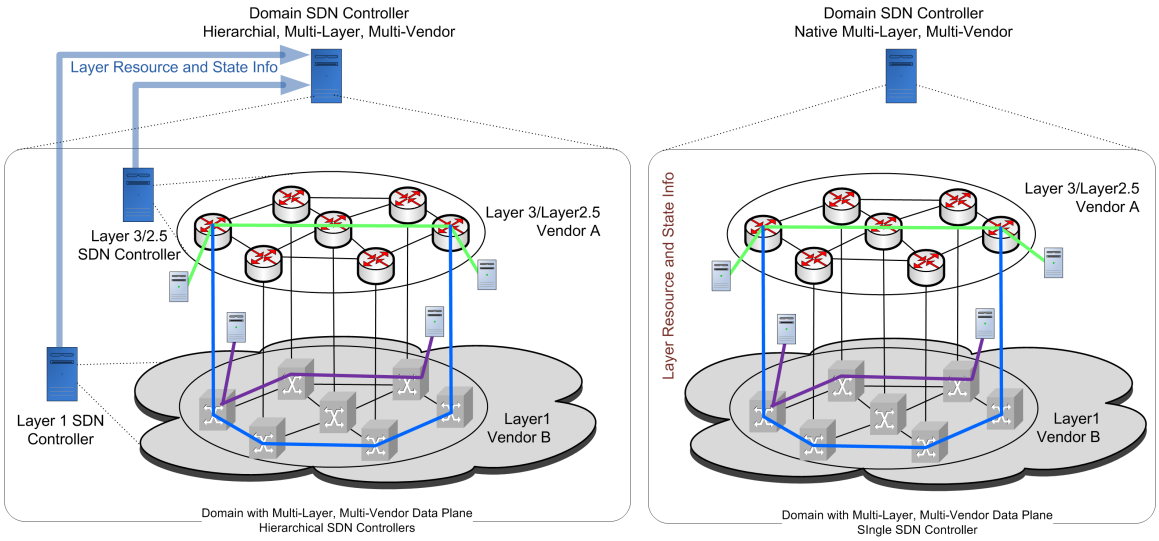


Figure 3b SDN in Multi-Layer, Multi-Vendor Environments

3.2 Intelligent Service Plane

While SDN may be a necessary technology to realize this larger vision of end-to-end flow management across domains and resources, it will likely need some additional capabilities. In order to provide services from the perspective of the user, i.e., the science applications, an "intelligence layer" may also be required. The vision for this capability is a degree of end-to-end resource and performance awareness that is not currently available in networks today. The SDN revolution is largely about separation of the control and data planes. A key benefit of this architectural shift is that it provides the opportunity to apply orders of magnitude more computational power to the control plane process. No longer limited by the computation resources in expensive and specialized network elements, there are many options for ingesting control plane data, computing solutions, and interacting with applications in the context of real time network operations.

This provides an opportunity to develop "Intelligent Network Services" which can allow applications to ask abstract questions about network and other resource availability and status. These may be key features needed in order to realize the amount of real time adaptability and interactive operations that the next generation of science workflows will require. Combining these per domain Intelligent Services, with proper multi-domain protocols and AAA (Authentication, Authorization, Accounting) mechanisms can allow intelligent services which will offer a new paradigm and capabilities for end-to-end multi-resource flow control and management. In addition, network operations agents may be developed improve user experience via background actions without direct user involvement utilizing these features sets. Figure 3d shows how this Intelligent Service Plane would relate to the other SDN architectural constructs.

Figure 3e shows how an SDN enabled intelligent service plane may interact with a complex set of data plane technologies and coordination agents to provide multi-domain SDN-enabled Intelligent Services.

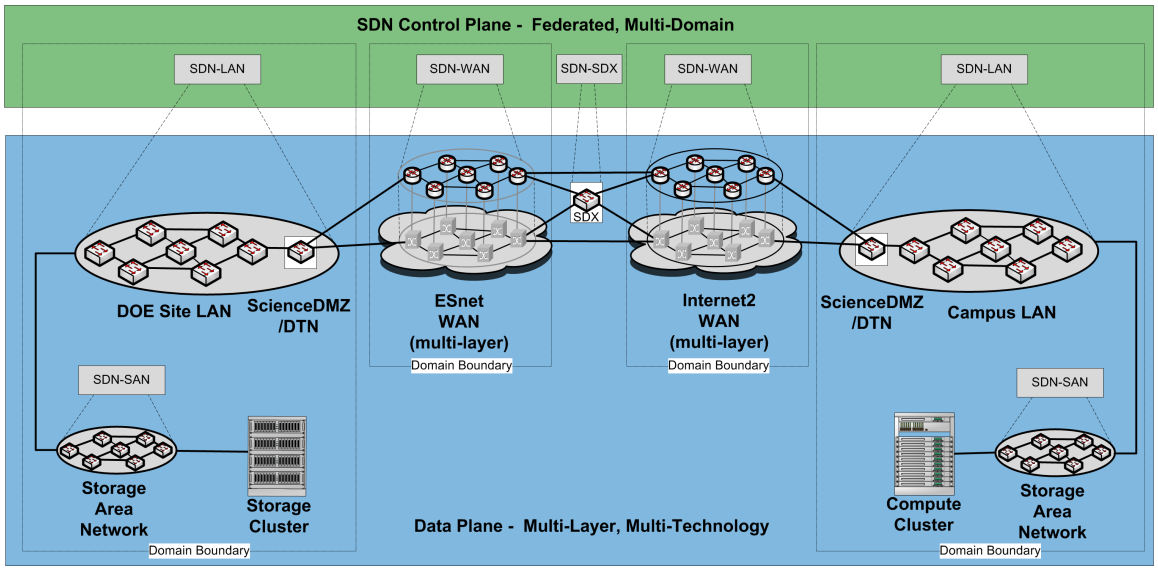


Figure 3c Federated Multi-Domain SDN Environment

3.3 Key SDN-ES Research Areas

SDN is a nascent and promising communication network paradigm. Its foundational underpinnings are not fully understood, validated, secured, and tested. The industry focus on SDN is on exploring its full potential for advanced Internet services. These SDN-enabled Internet services will be inadequate for distributed extreme-scale science, which leads us to SDN-ES. However, it appears that these technologies can be adapted and applied to the domain science requirements and use cases. During this time of active development by the commercial sector is an ideal time for DOE to evaluate how these technologies can be utilized or adapted to the DOE uses. The opportunity to leverage the commercial sector expertise and influence the designs and standards are both available. In addition, DOE extreme-science tasks require complex SDN technologies beyond the projected trajectories of commercial and general SDN technologies due the sheer data volumes and complex workflows that span experimental facilities and supercomputers. Indeed, the full promise of SDNs for science workflows is very unlikely to be reached as their by-products commercial efforts alone. Instead, we require SDN-ES solutions that are designed and optimized for Extreme-scale Science that seamlessly integrate networks, experimental facilities, and computing and data systems at the extreme-scale.

In many ways, this network architecture paradigm shift is unique, because it is happening in parallel with a similarly momentous change in application and workflow designs being driven by big data. There is a synergistic and iterative relationship between the emerging SDN network infrastructures and the big data driven applications. The requirements of these next generation big science applications will drive the next generation network infrastructures and services. The SDN based next generation networks and services will drive what new and innovative workflows operations domain science applications can develop.

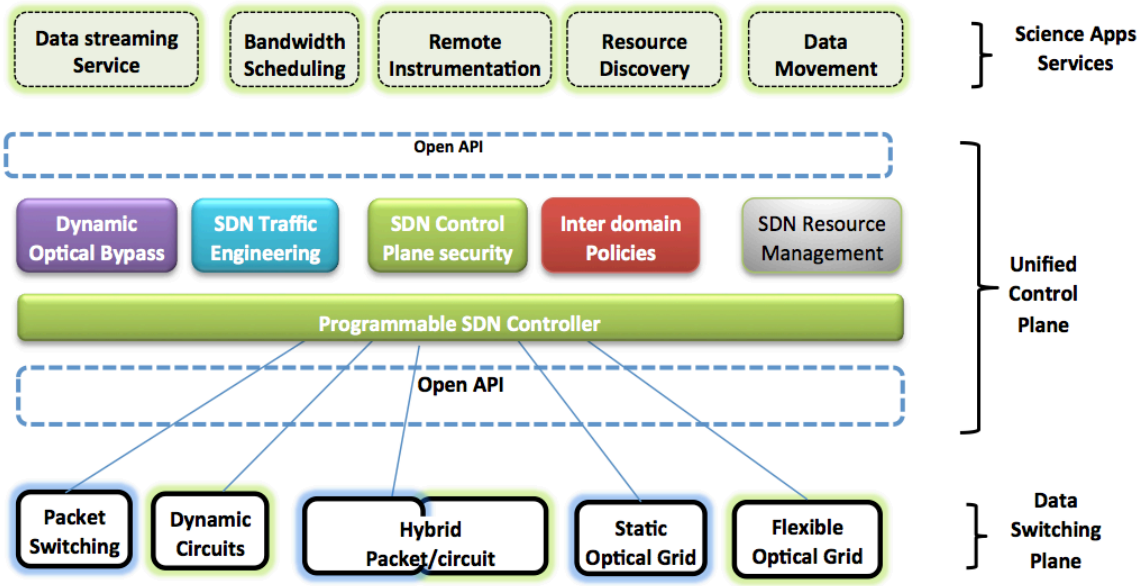


Figure 3d SDN Ecosystem of Services and Heterogeneous Data Plane

The result is that a new and important service boundary layer can be identified which sits in between the next generation network infrastructure and the next generation big data driven domain science applications. The requirements and designs for this service boundary and the associated features sets required by both the networks and the applications/workflows are currently undefined. As a result, a group of researchers will be needed who can work collaboratively across this boundary to maximize the benefit for the network operators and the domain science application and workflow developers.

The following topics are identified as key SDN-ES research areas as it relates to application of these technologies to the DOE use cases.

SDN-ES Based Network Architecture Definition: Define the key architectural principals and design features required to refactor the existing statically configured networks to provide agile, intelligent, and guaranteed services across autonomous network domains. A key focus should be on reducing network complexity and greatly increasing the ease of use. Another important consideration is develop interfaces and extensions to seamlessly integrate experimental facilities, supercomputers and data systems into workflows along with network connections.

SDN-ES Data Plane Control: Leverage SDN's open APIs and programmability to more intelligently control and monitor heterogeneous data plane environments. This should include enhanced intelligence and integrated provisioning across multi-layer networks that typically operate via independent/autonomous technology and vendor regions, and special provisions are provided to support workflows that involve instruments and supercomputers at the edges of networks. Application of SDN to emerging technologies such as flexible optical grids and terabit networks should also be evaluated.

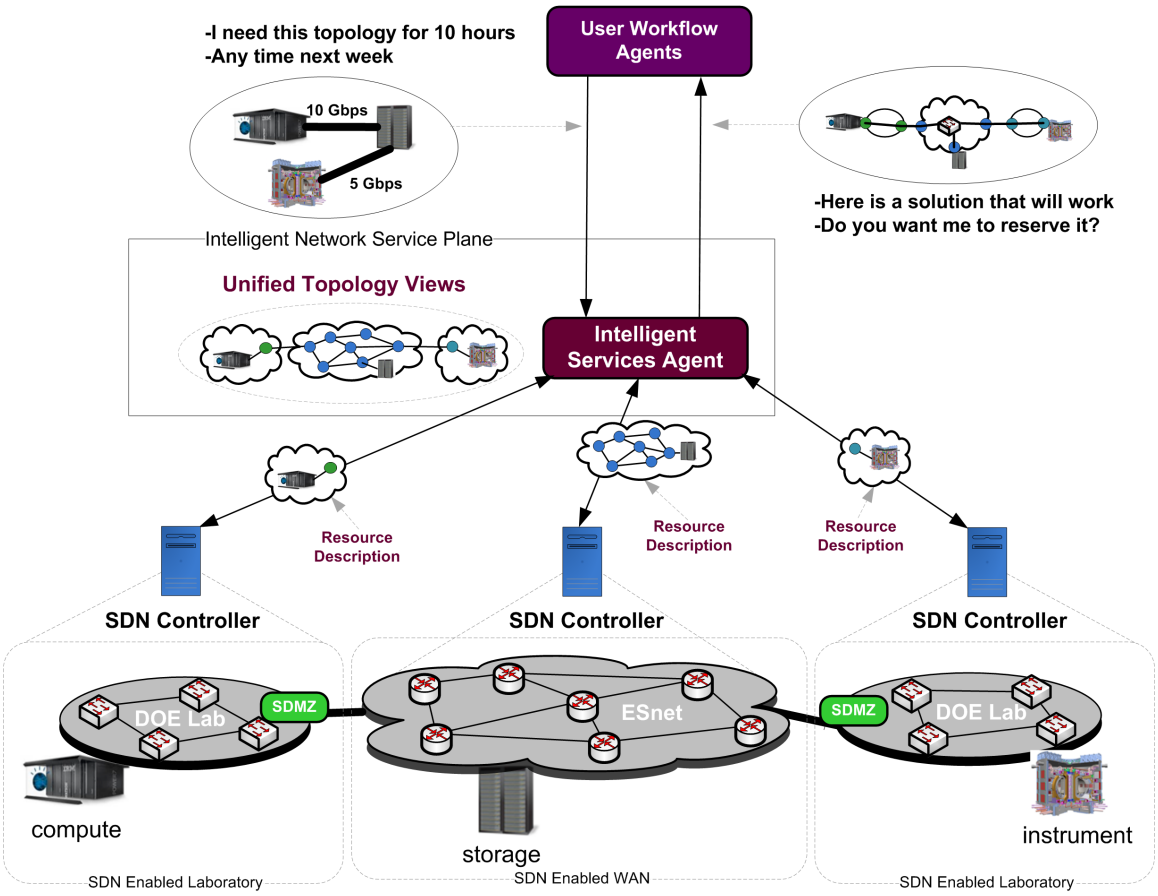


Figure 3e Intelligent Service Plane based on SDN

Network Virtualization (NV), Network Function Virtualization (NFV): Evaluate NV and NFV technologies in the context of R&E networks and distributed science support. Define and develop associated technologies tailored to the DOE use cases. Incorporate these technologies into the overall SDN enabled network architectures.

Federated Multi-Domain SDN Networks: Basic architecture research is needed to define the options and required features sets for science application driven multi-domain and federated SDN system designs. This should include options for multi-domain protocols and support for higher-level orchestration agents as mechanism to realize multi-domain services. Methods for network topology description and sharing which also allows operators to control and configure abstraction levels should be part of this work.

SDN Exchange (SDNx): This is a relatively new concept around on which there is currently much discussion, concept development, and debate regarding the purpose and architecture for this new paradigm. One view of a SDX is as an enhanced Internet or network eXchange Point (XP), where the XP switch is replaced with an SDN-capable device. This enhancement would allow for an enhanced set of policies and peering technologies. An SDN enabled XP could allow for richer, complex, and custom peering configurations which could span Layer 3, Layer 2, Layer 1, and flow level parameters. Extending this model beyond just network

elements to include host, computation, and storage resources, leads to an expanded view of this paradigm referred to as Software Defined Infrastructure eXchange (SDIX). The SDIX vision starts with the richer network connectivity options enabled by SDX and augments this further with host, computation, and storage resources available for use by the exchange point peers. Research is needed into what services SDNx should provide, and how they may be a part of the larger distributed SDN enabled infrastructure.

SDN based Traffic Engineering and End-to-End Flow Management: Research is needed in to methods and technologies for SDN based traffic engineering. The primary objective of this work is to determine how SDN programmability can enhance the ability for more fine-grained management of single and aggregated network flows possibly from instruments and supercomputers. This work will need to consider both intra and inter-domain services as well as sites that house science instruments and supercomputers.

Performance, reliability, security, and survivability: Research into models to predict, measure, and diagnose the performance, reliability, security, and survivability of federated SDN enabled networks.

Intelligent Network Services: Leverage the SDN enabled functionality to add an intelligence layer to facilitate science workflows real-time interactions with the network and other resources, particularly, high-performance computing systems, science instruments, and distributed data and file systems.

SDN-ES based Services Orchestration: The distributed nature of SDN administrative domain combined with the desire to manage a heterogeneous set of resources will likely require higher-level orchestration technologies. These may be in the form of the science application workflows, or more likely intelligent middleware that will provides services to the user applications. This is an area that will require extensive research and development to realize the needed functionality to support distributed science.

4 Recommendations

The workshop recommendations include technology and programmatic suggestions, all focused on development of next generation networks to support distributed domain science. The first category addresses key SDN related research and development topics for future R&E network architectures. The second category addresses the need to develop an Intelligent Service Plane, built on top of SDN enabled infrastructures, that can provide sophisticated services to science workflows. The remaining categories addresses the need to form teams of researchers that can work across the discipline boundaries of network, storage, compute, instrument, and domain science applications.

<u>Recommendation I: SDN-SE Research and Development</u>

Conduct research and development in the area of SDN Enabled Networks for Distributed Science. This work should strive to leverage, extend, and influence the building SDN development emerging in the commercial space. Specific research and development topics are discussed in Section 3.3 and noted below:

- SDN Based Network Architecture Definition
- SDN Data Plane Control

- Network Virtualization (NV), Network Function Virtualization (NFV)
- Federated Multi-Domain SDN Networks
- SDN Exchange (SDNx)
- SDN based Traffic Engineering and End-to-End Flow Management
- Performance, reliability, security, and survivability
- SDN based Services Orchestration

The overall objectives of this work should include the following:

- Exploit the network programmability framework introduced in the emerging SDN to develop a new generation of intelligent terabit networks to support distributed extreme-scale science.
- Leverage SDN programmability to transform current networking infrastructures (protocols, architectures, traffic engineering, federated network control) into intelligent and agile infrastructures that are responsive to new innovations and rapidly changing science application requirements;
- Develop performance models, metrics, and dynamic optimization frameworks to measure and predict the performance and behavior of agile terabits optical networks under various network conditions, including normal operating conditions, catastrophic failures, and cyber attacks.
- Reducing network complexity and make them easy to use and available to scientific applications and workflows.

<u>Recommendation II: Intelligent Network Service Plane Research and Development</u>
Conduct research and development in the area of Intelligent Network Service Plane and associated services. This work should build on the overall SDN work to build a set of intelligence functions that will allow application and workflow agents to interact with the network using sophisticated and purpose driven requests. It is anticipated that this research area will be beyond what is being considered in the commercial space currently.

The overall objectives of this work should include the following:

- Develop smart APIs and automated network services to embedded intelligence in scientific workflows to simplify network service discovery, invocation, and orchestration for scientists.
- Develop orchestration protocols and mechanisms to enable multi-domain services
- Develop the architecture and mechanisms so that these services can be included as part of a set of multi-resource intelligent services via extension to storage, compute, and instrument resources. Collaboration with domain experts from these resource domains will be needed in order to realize this part of the end-to-end vision. This is addressed in the next recommendation.

<u>Recommendation III: Cross-Disciplinary Cyberinfrastructure Research and Development</u>
The combination of SDN enabled network infrastructures and intelligent services will provide a basis to support new innovations in the science application domains. As discussed in Section 2, the emerging application workflows need an expanded end-to-end paradigm that includes the compute, storage, and instrument resources along with the network elements. There are many reasons for this requirement, with the desire for realtime resources need identification and adaptation being the driving factor.

These emerging workflow requirements, combined with the expectation that SDN will transform the network into a programmable resources on the same level as host and storage resources, points to a need for cross-disciplinary research teams. This may require a new class of expertise, perhaps called a "cyberinfrastructure engineer" who can work across all aspects of networks, data, compute, instrument, and application spaces. It is expected that individuals with this skill set and desire will often emerge from an advanced network technology expert base, due to the distributed and heterogeneous system nature of modern network architectures. The following recommendations are noted in this area:

- Form research and development teams that include experts from across the networks, data, compute, instrument, and application domains.
- Develop integrated "Software Defined Resource" functionalities which combine the SDN programmability, Intelligent Network Service Plane, and other resource programmability functions into a set of multi-resource intelligent services
- Formalize a process for the formulation of cross-disciplinary teams to work on purpose-driven science application problem solving and new functionality development.

<u>Recommendation IV: Experimental Testbed and Focused Areas Development</u>
The list of research areas broad and it will likely be beneficial to identify a subset for initial focus. The following recommendations are noted:.

- Task Force: A Small Task force may be established to interface with R&D and Science communities to focus on a subset of SDN-ES areas that are specific to DOE use cases and facilities.
- Experimental Testbed: A testbed spanning multiple DOE sites connected over ESnet may be established to support the development, maturation, and testing steps for DOE-specific SDN-ES areas.

<u>Recommendation V: Programmatic and Strategic Areas</u>

- DOE Site Collaborations: Site and multi-domain technology adaptions may be developed to ensure the operational stability and performance together with security of SDN deployments.
- Multi-Agency Investments Leverage: Strategies may be developed for DOE to leverage other agency (NSF, DOD, etc.) investments in SDN areas, and vice versa.
- Operations and deployments may be enhanced by establishing collaborations between vendors, facility providers, users and R&D communities.

References

1. DOE ASCR 2011 Scientific Collaborations for Extreme-Scale Science (SCESS) Workshop, December 6-7, 2011, Gaithersburg Marriott Washington Center, Gaithersburg, MD - SCESS Workshop (Report in PDF)
2. Terabits Networks for Extreme-Scale Science, February 16-17, 20011, Rockville Hotel & Executive Meeting Center, MD – (Report in PDF)
3. Data and Communications in Basic Energy Sciences: Creating a Pathway for Scientific Discovery Workshop, October 24-25, 2011, Bethesda Marriott Hotel and Conference Center, Bethesda, MD – PDF Report
4. DOE Exascale Workshop on Data Analysis, Management, and Visualization Workshop, February 22-23, 2011, Hilton Hotel, Houston, TX – PDF Report
5. Fusion Energy Network Requirements Workshop, December 2011 - Final Report, ESnet Network Requirements Workshop, December 8, 2011, (Report.pdf)
6. Nuclear Physics Network Requirements Workshop, August 2011 - Report.pdf
7. *Scientific Grand Challenges: Architectures and Technology for Extreme Scale Computing*, December 2009,http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Arch_tech_grand_challenges_report.pdf
8. *Scientific Grand Challenges: Crosscutting Technologies for Computing at the Exascale*, February 2010, http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Crosscutting_grand_challenges.pdf
9. Science Driven R&D Requirements for ESnet Workshop, April 23-24, 2007 - Report (pdf)
10. Networking Requirements Workshop- Office of Biological and Environmental Research, April 29-30, 2010- Report (pdf)
11. Networking Requirements Workshop- Office of Basic Energy Sciences - Report (pdf)
12. ESnet On-Demand Secure Circuits and Advanced Reservation Systems Federation Networking, Report (ppt)
13. *Scientific Discovery at the Exascale: Report from the DOE ASCR 2011 Workshop on Exascale Data Management, Analysis and Visualization*, Houston, TX, http://science.energy.gov/~/media/ascr/pdf/program-documents/docs/Exascale-ASCR-Analysis.pdf
14. *Synergistic Challenges in Data-Intensive Science and Exascale Computing, DOE ASCAC Data Subcommittee Report*, March 2013, http://science.energy.gov/~/media/ascr/ascac/pdf/reports/2013/ASCAC_Data_Intensive_Computing_report_final.pdf
15. T. D. Nadeau and K. Gray, Software Defined Networks, O'Reilly publishers, 2013
16. Software Defined Networks Special Issue, IEEE Computer, November 2014.

Appendix A: Extreme-Scale Science Application Drivers

An important part of this workshop activity was to discuss the next generation advanced network infrastructure in the context of the big data driven science that will rely on new and advanced network services. Multiple specific examples were discussed during the workshop and overviews of these were presented in Section 2 of this report. These include the following:

- Earth System Grid (ESG)
- Spallation Neutron Source (SNS)
- SLAC National Accelerator Laboratory Linac Coherent Light Source
- Astrophysics Surveys: Palomar Transient Factory (PTF) and Large Synoptic Survey Telescope (LSST)

This Appendix contains more detailed summary of how these domain science application workflows operate across today's Research and Education (R&E) high performance network infrastructures. In addition to a description of the current workflow, a forward looking vision is also presented for how advanced network services could enable greater innovation and scientific accomplishment.

A.1. The Earth System Grid (ESG)

The study of climate change, including an evaluation of its impact on Earth's ecosystem and human society, is one of the most important scientific challenges of our time. Because the physical processes governing Earth's climate are extremely diverse and complex, this research involves sophisticated model simulations, which generate an unprecedented amount of output data, as well as the collection of observational data from multiple sources (such as remote sensors, in situ probes, and vertical profiles) on a global scale. These data sets are managed and stored at multiple geographic locations across the globe; yet, they need to be discovered, accessed, and analyzed as if they were stored in a single centralized archive.

To aid the climate community in the access and dissemination of large-scale climate data, the Earth System Grid Federation (ESGF) was formed as a coordinated multi-agency, international collaboration of institutions that continually develop, deploy, and maintain software needed to facilitate and empower the study of climate change. Through ESGF, users access, analyze, and visualize data using a globally federated collection of networks, computers, and software.

ESGF allows users to interactively explore and analyze large data sets over U.S. DOE and university networks for intercomparing highly valued distributed archives, such as the Coupled Model Intercomparison Project (CMIP) used for the Intergovernmental Panel on Climate Change (IPCC) assessment reports. Currently, ESGF is managing and storing petabytes of observational and climate model output for scientists contributing to international assessment reports, such as the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5).
ESGF is a leader in climate data discovery and knowledge integration. Its distributed and federated architecture is geographically distributed around the world.

One of the routine tasks that require heavy network use is replicating data between sites, for example, the data node at Lawrence Livermore National Lab (LLNL) frequently receives over 5 TB a day from three-dozen worldwide sites. Figure A-1 shows the file-based workflow for moving data between modeling groups and data centers. Users access the portals to transfer data to their site. Now that much of the data has been replicated at LLNL from easy to access sites, we are looking at replicating data to LLNL from hard to reach places with unreliable networks. On average, we are downloading data from Korea, Japan, and China at rates averaging 250 GB/day. Another way of looking at this is that most individual server sites provide 1-10 MB/s by HTTP. In some cases, we are only receiving 0.1 MB/s. The fastest networks provide 100 MB/s, using GridFTP.

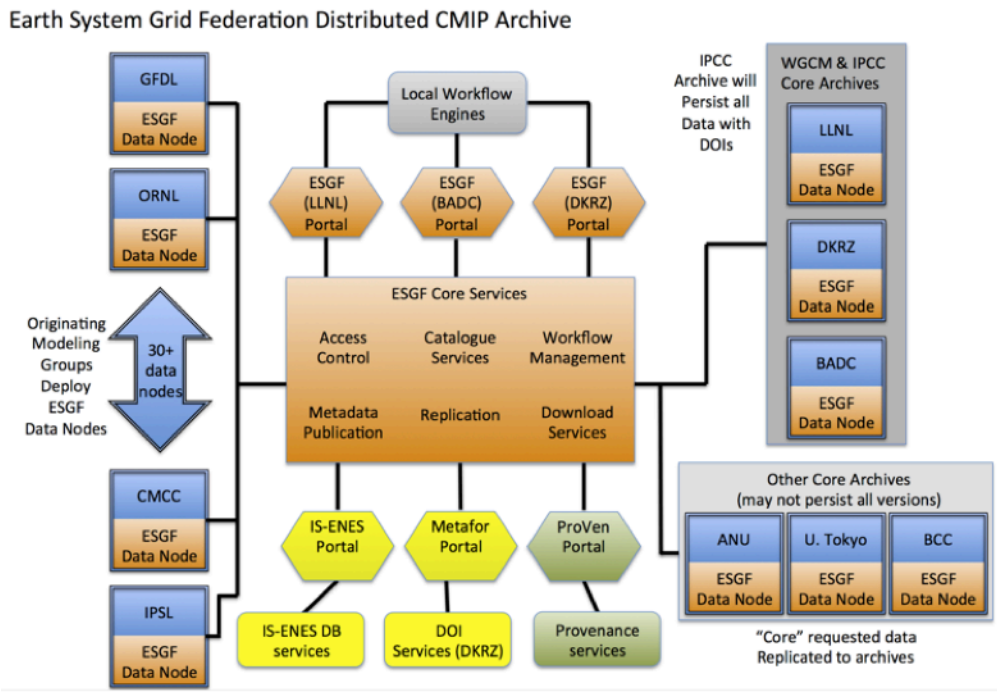


Figure A.1. A network of geographically distributed Nodes are built into a globally Federated "built-to-share" scientific discovery infrastructure. By federating these Nodes using network, independent data warehouses deliver seamless access to vast data archives to scientists and their specialized client applications. Experts (e.g., model developers, climate researchers) and non-experts alike need fault-tolerant end-to-end system integration and large data movement, and benefit from rich data exploration and manipulation in the process moving vast amounts of data to and from data sites around the world.

Unfortunately, access to the ESGF system by potentially 25,000 global users is constrained by limited local network bandwidth. For example, the ESGF data center at LLNL is connected to ESnet's Bay Area Metropolitan Area Network (BAMAN), a shared 20 Gb/s protected ring. In a highly collaborative, decentralized problem-solving environment like

ESGF, a faster network—utilizing individual connections 10 to 100 times faster than what is available today—is needed to make efficient use of the data and tools available to scientists. Climate researchers want the ability to quickly and seamlessly combine multiple data sets, up to 300 TB each, for analysis – one which is not feasible using today's networks. The only way to ensure the ESGF architecture scale to meet its users' requirements is to utilize next generation networks of the highest speeds—somewhere around 10 to 100 Gb/s. As a result, the ESGF project is proposing in this call to leverage a new coordinated effort of networks to make high-speed data federation a reality.

A.2. Spallation Neutron Sciences (SNS)

Neutron scattering is a critically importance in nuclear engineering. It is used in physics, chemistry, biophysics, and materials research. Because of its usefulness, DOE operates a number of national user facilities based on neutron scattering. Next, we will use the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory as an example of such a user facility. A typical SNS data flow is shown in Figure A-2.

In this data flow, first neutrons events are acquired by the data acquisition system, which gives each event a time stamp and an id corresponding to the hardware pixel where it was captured. Time based metadata is also combined into a stream of events at this point. This stream of events is sent to compute resources co-located with the Oak Ridge Leadership Computing Facility (OLCF) on the main Oak Ridge National Laboratory (ORNL) campus. There it is translated to a NeXus file and stored on a parallel file server (PFS). The process of transporting the events to, and storing them on, the PFS is handled with the Adara Package, developed at ORNL. This package also provides access to the stream for live viewing and monitoring of the data. The data rate per instrument is less than, but approaching 1 Gbit/s for each of the fastest of the 19 currently operating instruments. When the SNS is filled out, its second target station completed, and the instruments from HFIR included, the number of instruments will nearly triple. The variability of data rates between instruments means an optimized network that requires less than a 1 Gbit line per instrument could be developed.

Once on the PFS, data is accessed from multiple sources. Primarily from a cluster of high performance computing machines at the SNS, but also external users download there data using scp related utilities and, in the near future, will access the data through an Esnet data transfer node. A completed data set after an experiment can be on the order of 100 GBytes.

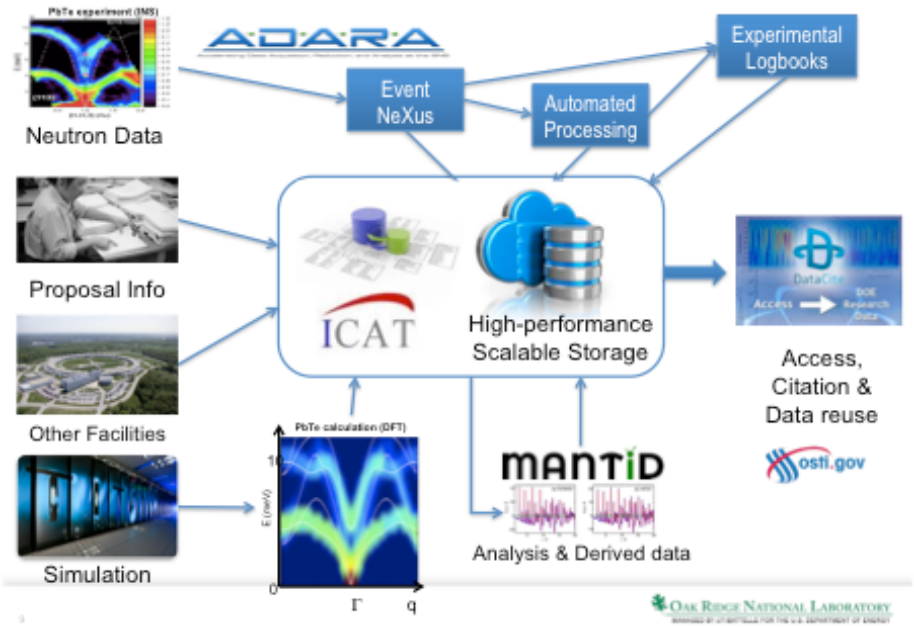


Figure A.2: Data Flow at the SNS

Advance data analysis work flows start starting to incorporate first principle simulations to verify the experimental measurements or to guide the understanding of the experiments. Such simulations could produce petabytes of data and would also need to access the terabytes of experimental observations. Currently, these simulation and analysis tasks have been using the OLCF resources and the SNS. However we envision users that want to use a code finely tuned to a HPC resource at their own institution or one of the other DOE supercomputing centers. Furthermore, as is being demonstrated through the Oak Ridge BES data pilot project for xray and neutron pdf measurements, users are moving towards pulling results from multiple, facility scale techniques. A network with a single sign on, that could configure itself to access the facility data, simulation data, and analysis packages that are needed for a given experiment in a manner that is invisible to the end user, would be the biggest benefit of an SDN to sciences at the DOE BES user facilities.

A.3. SLAC National Accelerator Laboratory Linac Coherent Light Source (LCLS)

Another powerful tool for studying materials is to use X-ray. The strong X-ray sources at various DOE light sources allow scientists to study advanced materials such as soft matters (proteins), nanocrystals, and nanoelectronic memories (memreister). These new materials could revolutionize biofuel production, and provide a new ways to build computers. Similar to the neutron sources, these light sources are also national users facilities with hundreds of users from around the country. The data produced from these facilities are often shared with collaborators around the word. As these light sources upgrade their instruments, more data is produced and shared among their community of users. To illustrate the data distribution, analysis and sharing requirements, we next examine the data flow of LCLS as a concrete example if Figures A.3 and A.4.

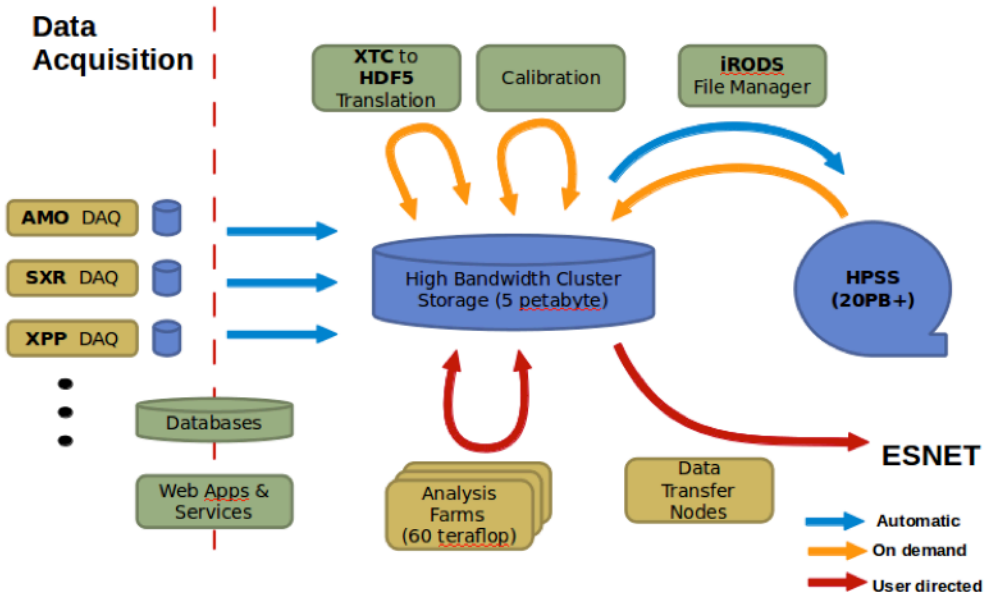


Figure A.3. Schematic of the LCLS data management system. The automatic dataflow is operated by the core data management systems, without user intervention. The on-demand dataflow is triggered by the users, but handled by the core system. The user directed dataflows are entirely managed by the users.

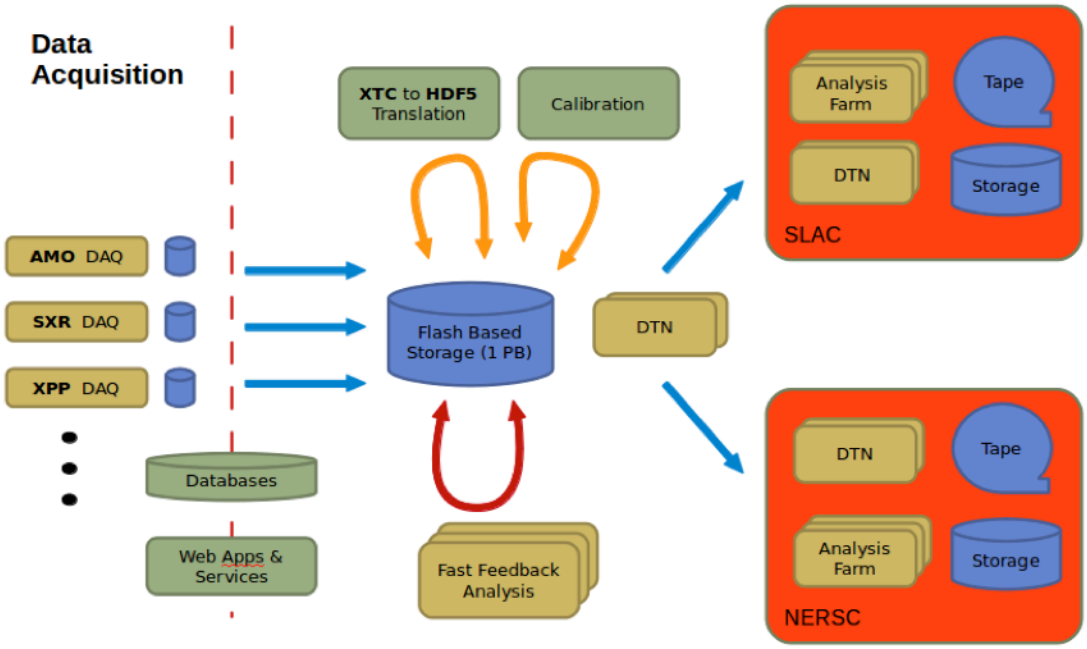


Figure A.4. Schematic of the evolution of the LCLS data management system. The new system will be able to offload some of the data processing and data storage capabilities to external data centers, eg NERSC.

Each LCLS instrument has a dedicated data acquisition (DAQ) network. Each network is built around one 10Gbps high performance L2 edge switch in the hutch and one dedicated 10Gbps switch in the server room. Directly connected to each DAQ network are two consoles for the control room DAQ operators, a variable number of readout nodes (typically 10 to 20 per hutch), anywhere from two to six monitoring nodes, two to twelve data cache nodes, and between two and twelve fast feedback nodes.

The DAQ system can currently acquire up to 5GB/s (10GB/s in CXI, 1GB/s in MEC) without introducing dead-time in the system. A minimal amount of online analysis is performed in the DAQ system to monitor the safety related issues and gather metadata for the recorded data. The bulk of the analysis operations are performed offline.

The offline analysis system, which is shared among the different instruments, is made up of medium-term storage, long-term storage, and a processing farm. The current size of the medium-term storage, which is disk-based, is 5PB. Each PB has a maximum throughput of 12GB/sec. Long-term storage uses the tape staging system in the SLAC central computing facilities, and it can scale up to several petabytes. Science data files are kept in medium-term storage for two years, with a per experiment quota enforced after 6 months, and are kept on tape in long-term storage for 10 years. Access to the data for each experiment is granted only to the members of that experiment. Experimenters are able to transfer their data files to their home institution if they decide to do so. Multiple specialized data mover nodes are allocated for that purpose. Medium-term storage is co-located in the experimental areas and communicates with the tape staging system in the SLAC central computing facilities through one dual 10Gbps link. An additional dual 10Gbps link between the NEH and the SLAC central computing is used to transfer the data off-site.

The processing farm is based on a batch pool and an interactive pool. They consist of 3000 and 400 cores, respectively, with fast access to the science data files in medium-term storage. An Infiniband QDR network connects the processing nodes and the processing nodes to the storage.

A.4. Astronomical Surveys: Palomar Transient Factory (PTF) and Large Synoptic Survey Telescope (LSST)

The PTF science relevant to the DOE-HEP program is split into two areas. The primary, and most active, is in the study of Type Ia supernovae (SNe Ia) for cosmology. The second is in the creation of target selection lists of quasars and galaxies for the BOSS and proposed DESI experiments.

PTF is a comprehensive transient detection system including a wide-field survey camera, an automated realtime data reduction pipeline, a dedicated photometric follow up telescope, and a full archive of all detected sources. The survey camera achieved first light on 13 Dec 2008; it completed commissioning in 1 Mar 2009; and will finish its original survey 31 Dec 2012 with continued operations until the 2016 timeframe.

The Large Synoptic Survey Telescope, which will see first light in 2021, is similar to PTF in that it will have a dedicated real-time transient program with an emphasis on SNe Ia. In addition, LSST will use three other techniques to constrain the properties of the expansion and the large scale structure of our universe: Baryon Acoustic Oscillations, Counts of Galaxy

Cluster and Weak Lensing. These three programs introduce an entirely different set of workflow pipelines, with a focus on understanding the systematics in data reduction in concert with simulation data, over multiple PB sized data sets.

As PTF is an ongoing experiment, we are able to use it as a perfect model to verify and validate our workflow models in preparation for the ~$1B LSST project. Thus we focus on it here in the following section but note that LSST will generate more than 2 orders of magnitude more data and, even more importantly, will have 3 orders of magnitude more people interacting with the data and workflow.

During the normal operation of PTF, the active area for the search (including all overheads) is 2700 square degrees and the observations generate up to 150GB of raw data per night. Data taken with the camera is transferred to two automated reduction pipelines as illustrated in Figure A-5. A near-realtime image subtraction pipeline is run at NERSC/LBNL for identifying optical transients within minutes of images being taken. The output of this pipeline is sent to UC Berkeley where a source classifier determines a set of probabilistic statements about the scientific classification of the transients based on all available time-series and context data.
On few-day timescales the images are also ingested into a database at the Infrared Processing and Analysis Center (IPAC). Each incoming frame is calibrated and searched for objects, before the detections are merged into a comprehensive database.

At NERSC the Real-time Transient Detection Pipeline makes use of the IBM iDataPlex supercomputer Carver, a high-speed parallel filesystem and sophisticated machine learning algorithms to sift the data and identify events for scientists to follow up on. On the software front, there is a tight coupling of a PostgreSQL database involved in tracking every facet of the image processing, reference building, image subtraction and candidate detection. This database now contains over 1 billion candidate sources. This database is queried by both humans and machines nearly continuously 24/7 and is one of the bottlenecks in the PTF pipeline.
Each clear night PTF pushes 100-150GB of data to NERSC. This data needs to be processed, distributed, and then archived.

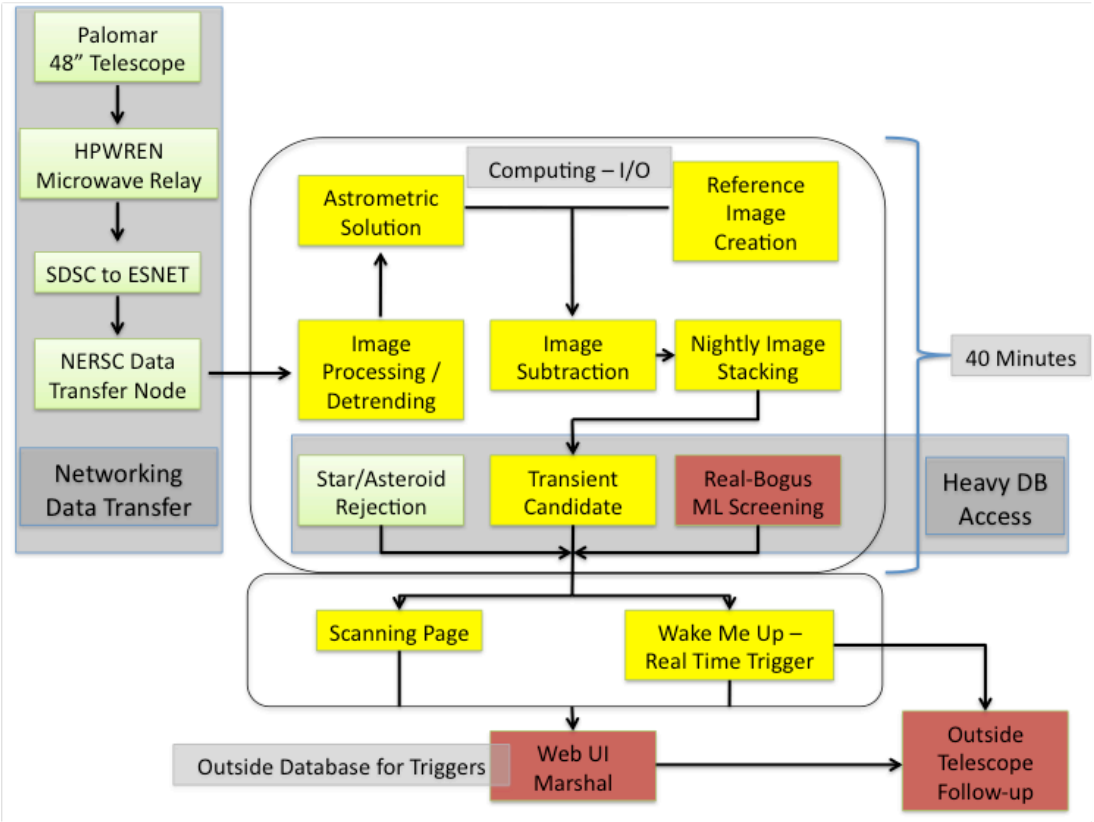


Figure A.5 An Overview of the PTF project data flow

Appendix B: Workshop Discussion Summaries

The workshop activities included a variety of discussions that ranged from high-level architectural models to lower level specific research and development tasks. The higher-level discussions and sense from the group were summarized in the previous sections of this document. In this section, additional details regarding the workshop discussions, identified issues, and recommendations are provided. These discussions were focused on identifying the key issues and tasks that needed research and development attention in order to realize the next generation network services in support of the DOE science use cases and missions.

These discussions were organized into three main discussion areas, as listed below:

- Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking
- Federated Science Big Data infrastructures with SDN-enabled Services and Workflows
- Cross-Cutting Areas which span and rely on the integration of the above two topics.

The key discussions points are summarized below.

B.1 Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking Workshop Breakout Session Summary Points

This group's discussion focused on end-to-end networking in the context scientific applications and workflow requirements. Key discussion topics addressed how the emerging paradigm of SDN could enhance user experience and network operations. Most of the key discussion points are reflected in sections 3 and 4 of this report. This section summarizes some of the specific research and technical task related to those broader discussions.

(A) End-to-End in the Science Application Context: The definition of end-to-end has typically focused on the network resources. As discussed earlier, this end-to-end model needs to now be expanded to include all of the systems and infrastructure in the end-to-end path. This may include host, SANs, LANs, ScienceDMZs, regional networks, wide area networks, and instruments. Figures B-1 and B-2 illustrate the many of these components that may be in an application end-to-end path. It should be noted that in this context, application is a broad term, which may refer to a scientific workflow, a data mover middleware, an orchestrator on behalf of an application, and anything thing else that is utilizing the network.

The challenge for researchers working on next generation network architectures is to adopt a more application centered end-to-end view. This will require an approach that focuses on how new technologies, such as SDN, can improve performance from a user experience perspective.

(B) Software Defined Networking (SDN): As discussed in section 4, SDN is an important emerging technology. Network Virtualization (NV) and Network Functions Virtualization (NFV) are other concepts that have closely related to SDN. For the R&E community, the focus relating to these technologies will need to go beyond the network, and be multi-resource based to reflect the integration with compute, storage, and instrument resources. As a result, the term Software Defined Ecosystem (SDE), in the context of domain science application support, may be a better way to refer to this technology development activity. A brief overview of these concepts is provided below.

- Software Defined Networking– The focus for SDN is the decoupling of the network control plane from the data plane. This enables network programmability thru an SDN Controller that interacts with the data plane element forwarding engines. OpenFlow is one example of a technology based on the SDN paradigm.
- Network Virtualization - These technologies are focused on two modes. The first is an environment where Virtual Machines (VM) need to be interconnected in an agile and dynamic fashion. NV provides mechanisms to build tunnels, or overlays, across an existing network infrastructure to create unique VM interconnection topologies. This mode of NV typically relies on technologies such as Virtual Extensible LAN (VXLAN) or NVGRE (Network Virtualization using Generic Routing Encapsulation) to construct the overlay networks. The other NV mode is one where the network is truly sliced at a network dataplane technology level such as VLANs, DWDM, or some other flowspace. This allows multiple "virtual" networks to exist on one physical infrastructure.
- Network Functions Virtualization– These technologies build on the NV paradigm to add network functions such as firewalls, intrusion detection, or load balancing features within the virtual or sliced environment.
- Software Defined Ecosystem (SDE)– This is the combination of SDN for network programmability and virtualization and the corresponding technologies for the host and storage infrastructure provisioning.

In this paper, the term SDN is used generally to capture the group of related technologies that are under development including Network Virtualization (NV), Network Functions Virtualization (NFV), and Software Defined Ecosystem (SDE). A key challenge for R&E network community is to determine what network services will the next generation of big data driven science applications need. In addition, it will be important to consider how emerging paradigms like SDN, SDE, NV, and NFV may be able to help satisfy some of these requirements.

In pursuit of the more application focused end-to-end model, "SDN like" capabilities may also needed for the end systems and infrastructures such as SANs, Storage Systems, and Hosts. This may represent the transition from SDN to Software Defined Ecosytem (SDE). The combination of SDN for network infrastructures (LAN/ScienceDMZ/MAN/WAN), and for storage system resources (SAN, Storage Systems, Hosts) should allow for end-to-end services not available today such as:

- end-to-end capability discovery, negotiation, and instantiation
- impedance matching for improved performance
- automated monitoring and troubleshooting

An important part of realizing this SDE vision may be SDN Enabled Storage Area Network (SAN) technologies. To further evaluate this, investigations into extending SDN capabilities into the SAN should be pursued as a mechanism for complete end-to-end flow management. Current mechanisms to extend flow management across the SANs and into the storage file systems are limited to vendor proprietary functions applicable to the local SAN environment only. There are two key drivers of the need for SDN enable SANs. The first is to maximize and tailor the performance across SAN infrastructure among storage targets and initiators (including per flow and per virtual interconnect fabrics). The second is to work toward a converged infrastructure where compute, storage, and enterprise/datacenter networks can all utilized the same underlying SDN enabled infrastructure. Currently these use case are divided amongst multiple technologies including Fibre Channel, InfiniBand, and Ethernet. In addition, these end-to-end technology development activities should includes the various types of High Performance File Systems (HPFS) such as Lustre and Ceph which are often layered on top of these systems.

The integration of SDN with memory system evolutions is another area where further research and development may be of value. Current system memory hierarchy consists of volatile Dynamic random-access memory (DRAM) for high speed access on the system bus, and Non-volatile random-access memory (NVRAM) and Disk Drives for slower speed access via the system I/O bus. There is some expectation that as NVRAM capacities increase and access times decrease, that the traditional separation of the system bus from the I/O bus may be able to converge or become more tightly coupled. This presents opportunities for the data to move between the system memory bus and I/O bus connected memory in a faster and more seamless manner. Technologies, such as this, or others, may provide new options for moving data between systems in a more direct system memory to system memory fashion. Integrating these types of technologies with SDN techniques for local or wide area transfers should be investigated as a possible new mechanism for distributed data systems.

(C) Intelligent Service Plane: The emerging control plane paradigms discussed above are expected to be an important part of developing end-to-end flow management across domains and resources. However, in order to fully realize this vision, additional capabilities will likely be needed. Providing these types of services from the user perspective, i.e., the science applications, an intelligence layer will also likely be required. The vision for this capability is to provide a level of end-to-end resource and performance awareness that is not currently under development or consideration in the commercial space in the context of SDN or similar emerging network technologies. The vision is for a set of services and functions based on an "Intelligent Service Layer" which will enable applications and workflows to better manage the user experience. This may include an ability for science applications and workflows to ask abstract or science specific questions in pursuit of specific workflow objectives. In addition, network operation agents may improve user experience via background actions without direct user involvement utilizing these features sets.

Multiple possible features set of this Intelligent Service Plane were identified:

- An ability to discover the available end-to-end paths, options, associated resources, and capabilities

- An ability to select specific end-to-end paths, or specific performance profiles for an end-to-end path, for a specific data flow
- An ability to interact with the network and storage system infrastructure an in "intelligent" and "automated" manner. This may include the ability to ask "what is possible" questions via an API to determine best workflow options.
- An ability to obtain "deterministic" service. That is, the exact resources desired may not be available, but it should be possible for the application to know what it has, so workflow operations can be planned.
- Automated monitoring and debugging feature sets so that per flow problems can be isolated without requiring human involvement
- Intelligent services may include applications interacting with the network in service oriented or abstract manners. Examples may be an application asking for the lowest latency path between two end sites. Or asking for options for moving a large data set to any of three possible computation sites, during a specific range of dates.
- Network services may need to go beyond just data movement, and may need to tightly embedded capabilities in the network. Embedded network storage or compute systems maybe used as an infrastructure for applications construct pipelined realtime data processing. Intelligent services will be needed to help workflows discover and utilize these resources.

Several specific functions of this Intelligent Service Plane were identified as follows:

- Capability Discovery
- Topology Description (with feature sets for policy based abstraction and export)
- Current Utilization
- Resource Provisioning mechanism
- Multi-phase negotiation and commit features including scheduling
- Monitoring
- Performance Measurement and SLA conformance
- An AAA architecture in the context of all of these functions
- Some entity may be needed to manage/broker the end-to-end service, across a federated multi-domain environment. This may be a broker agent that utilized the features of the intelligent service to provide turnkey multi-domain end-to-end services to applications.

There are other standards emerging which focus on the existing IP infrastructures in addition to emerging network technologies. This includes the Next Generation Service Overlay Network (NGSON) standard. NGSON is addressing IP-based service overlay networks. This includes a set of context-aware, dynamically adaptive, and self-organizing networking capabilities, including advanced routing and forwarding schemes to support dynamic construction of application services. These may also be evaluated for their applicability to an intelligent service layer.

(D) Multi-Domain and Federation: As advanced services are defined, it is very important that these be architected with multi-domain and federation in mind. The distributed nature of DOE science requires that advanced cyber-infrastructure services be able to operate on a multi-domain basis. Toward this goal, some model of Federation amongst resource owners seems to be necessary. A fine grained per flow, per user authentication

and authorization capability should be included. A SDN Exchange Point (SDX) is a paradigm that has been discussed as a mechanism to facilitate multi-domain services. SDXs may offer opportunities to realize a rich policy based automation of network peering and services exchange. This concept should be evaluated further in the context of intelligent services. Basic architecture work is needed to define the options and required features sets for science application driven multi-domain and federated system designs. Evaluation of components that are emerging in the commercial cloud space should be part of this analysis. This includes capabilities such as hybrid cloud, intelligent edge boxes, and an API focused model for coordination across administrative and technology demarcation points.

These demarcation points will be based on technology transitions and administrative regions. Examples of technology demarcation points across and end-to-end path is shown in Figure B-3. Examples of administrative demarcation points are shown in Figure B-4. For each of these demarcation points, next generation architectures will need to include mechanism for end-to-end flow management which can progress thru these boundaries. This will likely require well defined APIs and security mechanisms which specifically deal with these boundary locations. These demarcation points imply that protocols, APIs, AAA mechanisms are needed to deal with this multi-dimensional Federated infrastructure.

(E) Security: The envisioned set of network and cyber-infrastructure next generation services is quite advanced as compared to current basic Layer 3 and Layer 2 services. The security architecture and technologies will need to also advance to a similar degree to secure these infrastructures. This need for new security features as part of an evolution to SDN and Network Services provides challenges and opportunities. The challenges revolve around the fact that a key goal is to make the network more agile and flexible. These features will allow a more responsive network from an application perspective. However, these same features will open up new attack vectors for the malicious actor. Software bugs and/or human error will also be of more concern with a more intelligent and dynamic network infrastructure. The opportunities revolve around the fact that the SDN based architecture is in an early stage of development, and security components and technologies can be included as a native part of the transition.
There are a few general architectural constructs about SDN that will require a new approach to the associated security mechanisms. The first is the separation of the control plane from the data plane. This is one of the main benefits of SDN, in that a separated control plane greatly increases the flexibility and computation abilities. However, this creates a new security concern regarding authentication and trust between the control plane and data plane.

The second SDN characteristic is that of a more centralized architecture. The current expectation is that at some level, perhaps a domain or a region within a domain, there will be an area of "centralized" SDN control over some set of data plane elements. This centralized controller will have a unique perspective on the capabilities and real time status of the resources for which it has responsibility. This information will allow the controller to provide intelligent services not available today, but will also expose some more powerful attack vectors. The SDN architecture needs to include strong security mechanisms to monitor and prevent these types of attacks.

The other aspect of SDN that needs special security research and development revolves around multi-domain services. While SDN is expected to have some level of centralized control on a per domain or sub-domain basis, the larger global system will still be a distributed SDN architecture. As a result, security mechanisms need to be created to address multi-domain SDN service provisioning. In addition, the expectation is that SDN will enable a new set of services that are fundamentally different from the multi-domain services we see today. Multi-domain services which allow for a fine-grained policy application on both a user and flow basis is expected to be required.

The conclusion is that security architectures and technologies should be included a core component of the overall SDN architecture. This should included single domain/region and multi-domain architectures. The security features sets should address all of the following items: access control, redundancy, monitoring, forensics, troubleshooting, policy frameworks, user authentication, and user authorization.

(F) Current State-of-the-Art: The current state of the art for advanced network infrastructure and services was also discussed. The focus here was on the science application flows and how they use the networks today. Best effort routed services are still the most common method for science applications to move data. Research networks like ESnet and Internet2 include advanced dynamic network services that allow for on demand provisioning in support of science application flows. These dynamic network capabilities are primarily available in the wide area networks. The provisioned services typically do not extend across the regional networks and local area resources located at laboratories, campuses, and enterprises. As an example, file transfer applications, such as GridFTP, do not have plug-in modules to use the currently available dynamic network service offered by systems such as OSCARS. The extension of dynamically provisioned network resources to local resources is typically a manual process for which only the largest science application communities can take advantage. End systems, storage nodes, LANS, and SANS, are generally not provisioned and managed on a per flow basis. As a result science applications do not have mechanisms to obtain repeatable or deterministic services across the end-to-end path. This is a problem for both single domain and inter-domain flows.

The current state of the art for high performance R&E networking revolves around the following deployed capabilities and features sets:

- 100 Gigabit/second (Gbps) core links: Most of the wide area links and some of the regional network links are now operating at 100 Gbps. The transition from 10 Gbps links to 100 Gbps is expected to continue as traffic profiles require. In addition, the R&E wide-area network providers like ESnet and Internet2 have acquired access to nationwide dark-fiber, enabling them to provide multiple 100Gbps wavelengths across the nation to aggregate traffic from multiple national labs and universities
- 100 Gigabit/second (Gbps) access links: Driven by the raw data throughput needs of science workflows, a lot of the large national labs and universities are deploying 100Gbps access links to their campus. For example, the three supercomputing centers at LBL, Argonne, and Oakridge have 100Gbps to their campus. Similarly, the CC-NIE/CC-IIE NSF program has funded more than 80 campuses to upgrade their access links to 100Gbps in support of domain science. This forces the WAN providers to build support for large end-to-end data flows.

- Layer 3 IP Routing: Best effort routed services are the most common method for science applications to move data. Large science collaborations like LHCONE have built their own virtual routed network specifically to exchange traffic related to High-energy physics research on data from the Large Hadron Collider (LHC).
- Layer 2 path provisioning across core networks: DOE has been a pioneer in the use of Layer 2 provisioned paths for data movement across the ESnet infrastructure using OSCARS (On-Demand Secure Circuits and Advance Reservation System). This provides mechanisms for a science application to obtain an isolated Layer 2 path across the wide area network in an automated fashion. Extending this Layer 2 path across the regional, laboratory, or campus network typically requires manual configurations, and usually the end-network does not support mechanisms to provide the right QoS/guarantees .
- Science DMZ: At the edge of the laboratories and campuses networks, a Science DMZ is often deployed to manage the large bandwidth flows from/to the wide area networks. The Science DMZ architecture provides the capability to bypass performance-limiting firewalls and enterprise network gear while locating high-performance storage and data transfer nodes (DTNs) on the perimeter of the laboratory/campus network. The Science DMZ architecture also advocates deploying perfSONAR nodes so that end-to-end performance can be monitored actively.
- 10 and 40 Gbps end system interfaces: The current standard for network interface speed for end-systems is 10 Gbps. End systems, especially purpose-designed data transfer nodes (DTNs), with 40 Gbps interfaces are becoming more common, but use of parallel data movement to multiple 10Gbps connected end systems is still the most common method for moving large amounts of data.
- End System Software and Protocols: To this network infrastructure, the domain science communities connect end systems configured with various middleware, data movement protocols, storage and compute systems, and domain science specific applications and workflows. The data movement protocols are typically based on TCP and UDP, although increased experimentation is ongoing using protocols such as iSCSI and RDMA over Ethernet.

Several observations can be made regarding this networking current state of the art as it relates to the support for domain science applications and workflows:

- Because of the reliance on best effort IP routed services and/or partial path Layer 2 circuit provisioning, there is very little "end-to-end" path or network resource awareness from an application (or network operators) perspective.
- Obtaining consistent and deterministic end-to-end performance is a complex endeavor which requires close collaboration between the network operators and the domain scientist on a per application and flow basis. Performance troubleshooting requires a high level of network and end-system expertise and system access.
- Future science application will need the definition of End-to-End to include the storage, compute, and instrument systems attached to the network, as opposed to just the network segment as has typically been the case for testing and evaluation in the past.

- This method of engineering high performing end-to-end performance is human and equipment resource intensive, and typically only the larger domain science communities have the required resources.
- There are few to no network features or services that allow the domain science workflows to adapt their resource topology on an end-to-end basis with quality of service guarantees.
- The large domain science communities approach to overcoming these weaknesses has been to build special compute and storage facilities that are well connected to the regional and wide area infrastructures. These communities are also able to build well-engineered end systems that can maximize performance on an end-to-end basis.
- The future domain science research requirement are expected to be dominated by increasingly distributed resources, order of magnitude increase in data volumes and data mobility, and increased flexibility requirements with regard to real-time adaptation for where and how data and compute resources are accessed.
- The current method of integrating network with the storage, compute, and application resources will not scale in this emerging environment, even for the large, well resourced, domain science communities.
- The smaller community and independent domains science researchers are typically not able to get their resources well connected and struggle to obtain good end-to-end performance in today's environment. They will have even more difficulty as the degrees of distribution and data volume greatly increases.
- There are projects that are working on building infrastructures for data movement as a service. The emergence of these types of infrastructures is important and exactly what is needed as part of expanding the services model for compute, storage, and data movement. However, these types of application services rely on the current suite of network services and are subject to the set of issues discussed above.

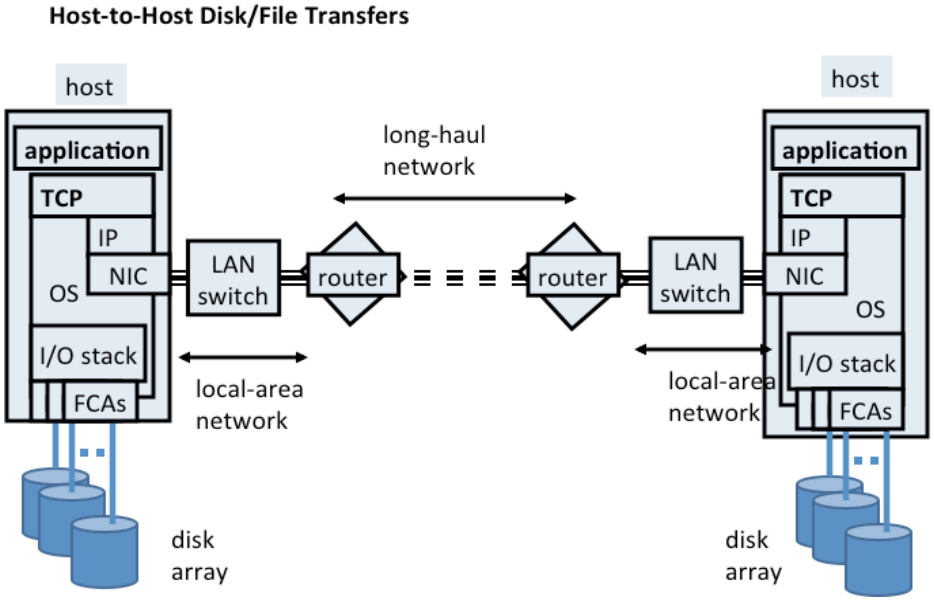


Figure B-1 End-to-End Resources – Host to Host

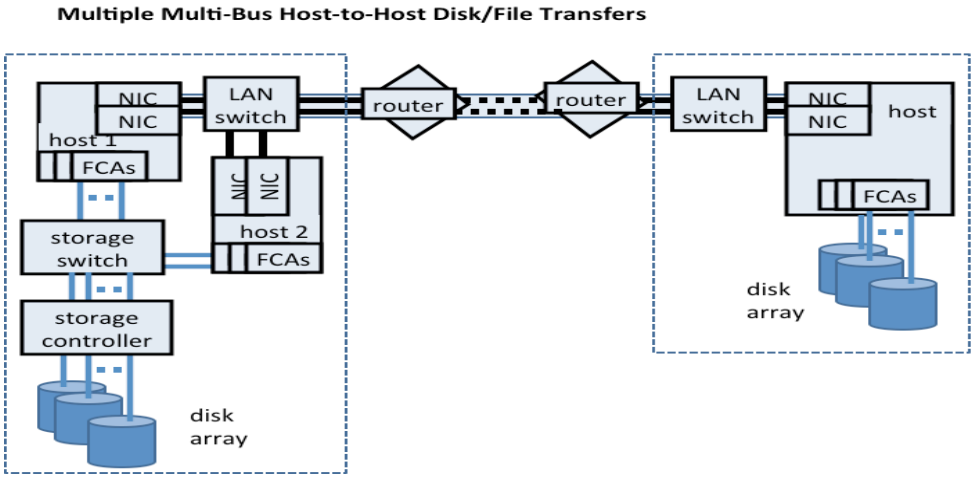
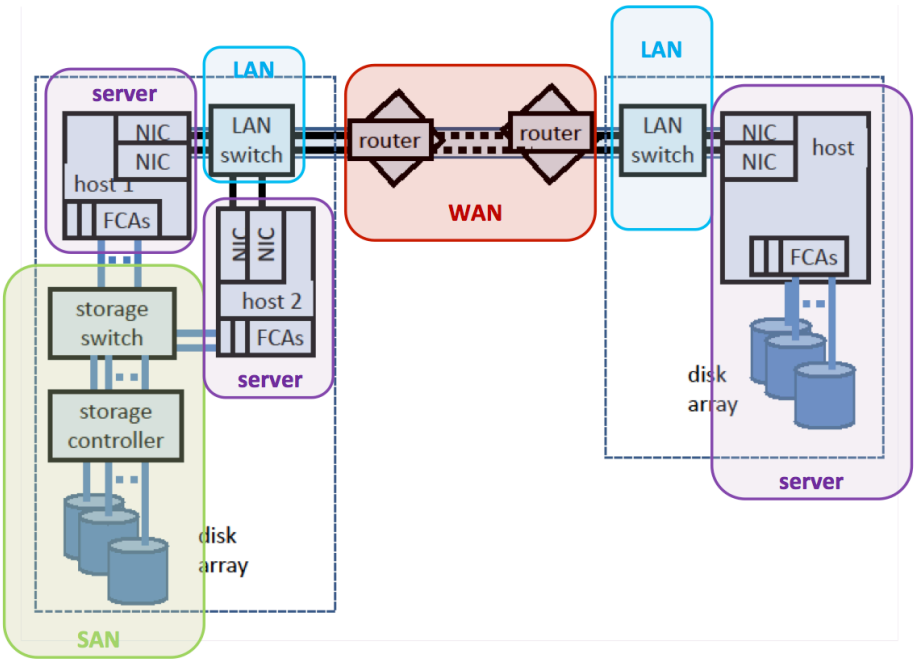
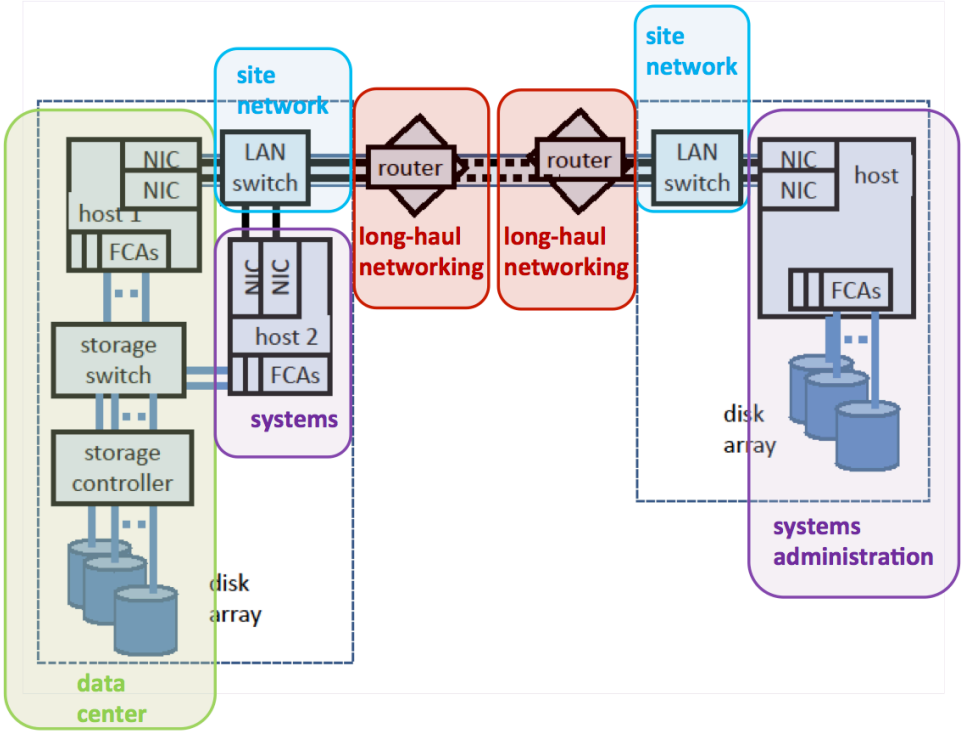


Figure B-2 End-to-End Resources – Storage to Host



B-3 Technology Demarcations for Multi-Domain Transfers



B-4 Administrative Demarcations for Multi-Domain Transfers

B-2 Federated Science Data infrastructures

The distributed data infrastructures are composed of a variety of complex, component systems: storage arrays and file systems that provide the repository capabilities; data transfer nodes that typically consists of multi-core, multi-bus, multi-NIC servers; supercomputers and science instruments that generate datasets; and SANs and LANS that connect the complexes of these systems to WANs. Advances in these components systems would be needed to achieve the data volumes that match WAN throughputs of multiple 100Gbps and beyond; in addition, they must also provide stable and dynamic performances needed for interactive and control operations. Furthermore, these systems provide appropriate provisioning, monitoring and performance optimization APIs and tools to effectively integrate them into SDEs.

(A) Disk arrays and hierarchical storage systems: Methods are needed to utilize arrays of disks and other storage system to store and retrieve large files, and controller that provide multiple IO streams to sustain high throughputs and stable, predictable performances.

(B) Distributed files systems and metadata servers: Placement of file servers, in particular, metadata services, is extremely important in providing the site-level access to computing systems, instruments and data transfer nodes. In particular, methods are needed to provide uniform high-performance file systems both at the site and across the networks as a part of SDE. Specifically in DOE environments, solutions are needed to

provide transparent file access across multiple files systems such as Lustre, GPFS and others, that may be accessed by a single science workflow.

(C) Uniform middleware for user access to different data services and systems: Due to the large variety and highly specialized nature of components of data systems, it is essential that virtualizations and middleware (in terms of APIs and tools) are developed to make them available under SDE framework. Using these technologies, the data systems can be provisioned and operated in concert with network services.

(D) SAN and LAN connections and protocols for high-performance dataflows: The local connections and protocols to storage and data systems could be based on Infiniband, Ethernet and FibreChannel technologies, and may supported by TCP/IP, RDMA transport methods.

(E) Protocols and APIs for supercomputer data transfers: Due to the inherent complexity of datapaths to/from supercomputers, methods are needed to establish, monitor and optimize these dataflows under SDE frameworks.

(F) Data Transfer Nodes: Multi-core, multi-NIC, multi-bus data transfers server systems are used to implement data transfers between data systems and networks. Methods are needed to implement efficient data transitions between HCA/HBA and NICs, to aggregate IO data flows in to high-performance network flows. Methods are needed locate NIC and HCA/HBA on appropriate buses to eliminate data flow bottlenecks, and IRQ mapping from them to appropriate processor cores. Also, high-performance protocols are needed to sustain performance over various connections.

B-3 Cross-Cutting Areas and Operational Aspects

The solutions based on the networking and data technologies must be integrated in under a unified SDE framework utilizes diverse technology components which are under different operational domains. As illustrated in the federated SDN environment presented in Figure B-5, a composition of solutions from different technology areas is essential. In this case, optimally matching disk and file systems and serves to SAN, LAN and WAN connections is essential to achieving 100Gbps transfers rates. Furthermore, these component systems are typically different operational domains at DOE sites. Typically, servers, storage systems and networks are maintained by different operational groups, and the experimental facilities and supercomputers may be maintained by yet different groups -. Such solutions require a number of cross-cutting technologies.

(A) Flow Compositions and Realizations: Methods are needed to design complex-end-to-end flows that optimally align data systems and networks to minimize impedance mismatched and achieve sustained performance. Also, methods and tools are needed to implement these designs using the SDE tools.

(B) Co-Scheduling and Provisioning Methods: On-demand and advanced reservation and resource management systems are essential so that data systems and networks can be provisioned.

The distributed science applications typified by the use cases in Appendix A require a wide variety of solutions that integrate networking, data and computing components. This broad spectrum of solutions can be distilled into three main capabilities, each of which encompasses multiple use cases and yet requires a specific combination of networking and data solutions.

A. Cyber-Enabled Big Data Intensive Sciences: Instruments of DOE experimental facilities such as observational telescopes of LSST and SNS beam lines are expected to generate large volumes of data, which is processed by complex computational codes at remote sites. The outputs from these computations must sent back promptly to the facility sites to setup next set of observations or experiments. The feedback response time can vary from minutes to hours, for example, in setting up microscopes to capture a transient phenomenon. The observation datasets could be massive and must be transported to remote sites at high data rates, and the computational results must be sent back to meet strict time deadlines. Experiments empowered by cyber, networks and computations, include SNS, LSST, LTTC, APS and others. The computational facilities can be reached over LAN in some cases (e.g. SNS) or may be several hundred miles away over WAN connections. Also, the data may be filtered on-site and sent for further processing at remote sites. In terms of networking, these work flows require high bandwidth connection in the forward direction and low latency in the feedback direction. Furthermore, these capabilities are needed at certain times and for only short durations in some cases. It would be inefficient to build dedicated infrastructures for these capabilities. Rather, these network connections may be provisioned either on-demand or in-advance, and may be returned to the resource pool to support other traffic. Technologies that can provide such capabilities are currently very limited or non-existent.

B. Distributed-Replicated Data Repositories: Data generated by codes on supercomputers and observations from instruments may be provided for access by science users distributed across the globe. These datasets are suitably replicated and provided at sites across the globe to provide easy and efficient access to the large user community with varying network connectivity. The climate datasets of ESG represent such a use case of distributed data across the globe accessed by different PIs. In DOE applications such as LHC and ITER, the datasets are organized into a hierarchy so that there transferred and stored as per a multi-tier architecture. In these applications, larger datasets are transferred and stored at sites at the higher levels, thereby providing effective access to users at these sites. The ability to setup on-demand or in-advance network and storage connections from users to the data sites can enable optimal access times. Such workflows require network connections that provide sufficient capacity and latency guarantees needed for such tight control of computations and challenging network connections – data from observations, weeks to months; schedule detection and feedback within few hours, new supernova; transients in the sky to support quick follow-up.

C. Coordinated Computing, Analyses and Archival: Complex computational workflows involve analyzing the data generated by supercomputer codes, either on-line or during post-processing at remote analysis or visualization facilities. The results of these analyses may be utilized for setting-up parameters for next set of computations, or for closed-loop steering of on-going computations. Furthermore, the analyses may incorporate data from auxiliary sources such as instruments. The examples of such workflows include supernovae computations, and climate computations. Furthermore, the outputs of computations and analyses may be archived to be made available for subsequent access by the wider science community.

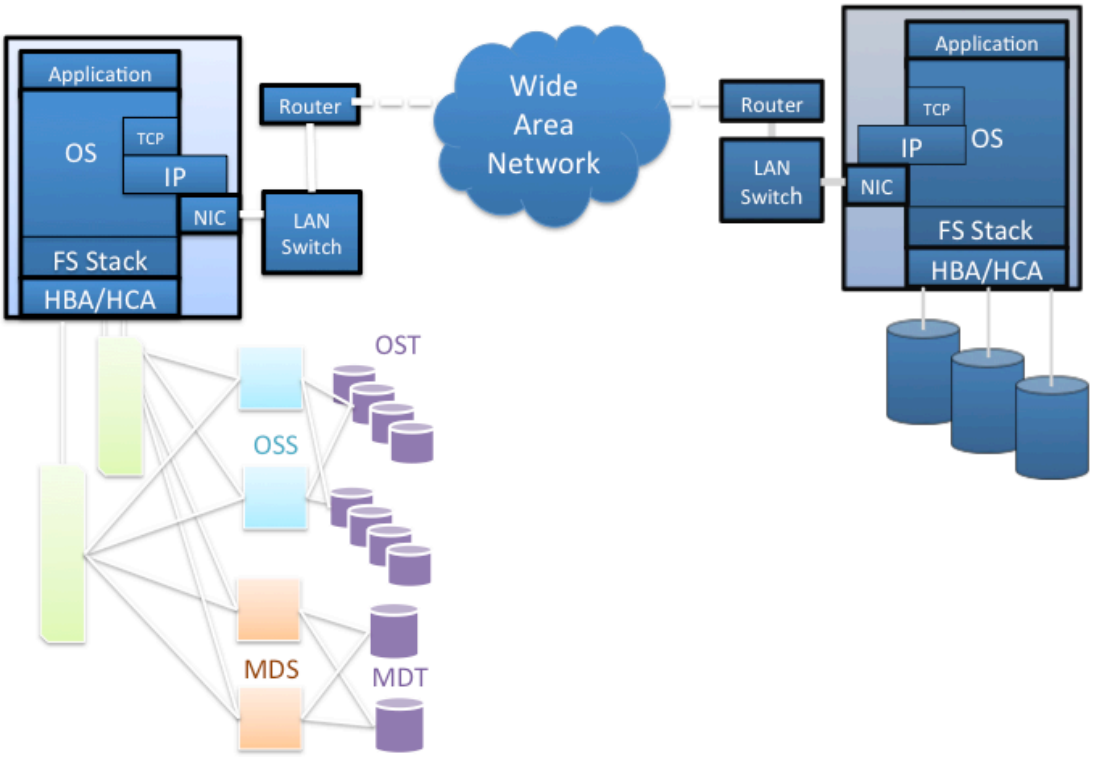


Figure B-5 Components of end-to-end file transfer

Across the three capabilities described above, there is need for advanced SDE solutions for transporting large datasets from storage systems or computer memory to remote sites. One of the simple scenarios required in all three capabilities is illustrated in Figure B-5, which involves transferring a large file from a source site to a destination site. On the source side, the file is stored on an array of disks supported by Lustre file system, which itself consists of data and metadata nodes. The file transfer is carried out by a multi-core, multi-NIC server. At the destination, the file is received by a multi-core, nulti-NIC server which is connected to an array of disks via multiple HCAs. Thus, effective file transfer requires appropriate provisioning and configuration of storage and file systems and LAN and WAN connection, by optimally placing the IO and network processes on the source and destination servers. However, if the transfers are to be carried to and from a supercomputer, the task now becomes increasingly more complex as illustrated in Figure B-6. Here the data flows originating from computing nodes must be appropriately realized through the system interconnect to data transfer nodes. In addition, if the datasets stored on the local storage system may be retrieved and utilized by the computation, in which case both local data flows from local file systems need to be supported simultaneously with remote memory flows through complex combination of components systems.

Apart from these "bandwidth" needs that have been the focus of most recent works in network and storage flows, the categories A-C are also characterized by low latency and fast dynamic responses. Also, current infrastructures are mainly static and often require provisioning by human operators; consequently, they are often over-provisioned to accommodate high performance that may be needed for very short durations. The promise of SDEs is the ability to dynamically configure the resources for the dedicated purposes and return them to the default general pool. However, such the dynamically provisioned SDEs may have to support radically different "response" time-scales, and at each such scale may involve:

- networking at all levels, namely, SAN, LAN and Wan using SDN technologies
- effective composition of disparate, distributed data, storage, file and computing systems.

In the next section, we describe various technical components of the solutions needed to realize the capabilities A-C.

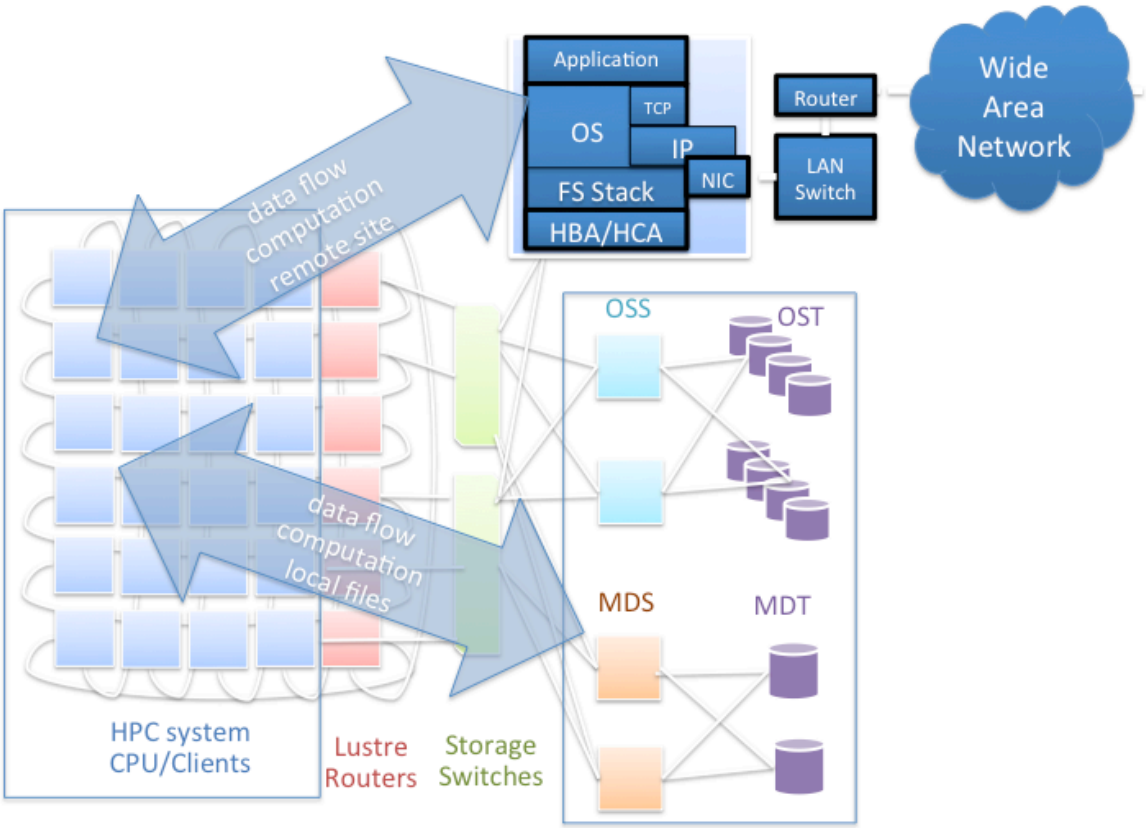


Figure B-6 Data transfer from a supercomputing system

Appendix C: Workshop Agenda

ASCR Intelligent Optical Network Infrastructure Workshop Agenda

(Smart Networks & Middleware Services for Open Distributed Science BigData Infrastructures)

August 5-6, 2014

Hilton Hotel, Gaithersburg, MD

August 5, 2014: Day 1: SDN, Data Services for Science Applications

08:00 - 08:15 AM	Workshop Logistics
08:15 - 08:30 AM	Welcome – T. Ndousse-Fetter (ASCR)
08:30 - 09:00 AM	Plenary Talk 1: David Cohen (Intel)
09:00 - 09:45 AM	Plenary Talk 2: Peter Nugent (LBNL) Title: Real-Time Data Analysis in Astrophysical Surveys: 2014 and Beyond
09:45 - 10:30 AM	Coffee Break
10:45 - 12:00 AM	Parallel Breakout Sessions

- **Breakout Session #1: Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking**
- **Breakout Session #2: Federated Science BigData infrastructures with SDN-enabled Services and Workflows**

12:00 - 01:00 PM	Lunch
01:00 - 02:00 PM	Panel #1:

- **Panel 1 Title: Intelligent Networking with SDN: Implications for end-to-end Internetworking - Current State of the Art, Emerging Technologies, and Future Visions**

2:30 - 3:00 PM	Coffee Break
3:00 - 5:00 PM	Parallel Breakout Sessions

- **Breakout Session #1: Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking**
- **Breakout Session #2: Federated Science BigData infrastructures with SDN-enabled Services and Workflows**

5:00 - 6:00 PM	Joint session
6:00 PM	Adjourn for the day

August 6, 2014: Day 2: DOE-Specific Data and Areas

08:00 - 08:15 AM	Workshop Logistics
08:15 - 08:30 AM	Remarks/Workshop progress – T. Ndousse-Fetter (ASCR)
08:30 - 10:00 AM	Panel #2:

- **Panel 2 Title: SDN-Oriented Networking and Distributed Data management: Implication for Science networking Infrastructures and Distributed Data-intensive Science**

10:00 - 10:15 AM	Coffee
10:15 - 12:00 PM	Parallel Breakout Sessions

- **Breakout Session #1: Intelligent Network and Middleware Services: SDN-Oriented End-to-End Networking**
- **Breakout Session #2: Federated Science BigData infrastructures with SDN-enabled Services and Workflows**

12:00 - 01:00 PM	Lunch
01:00 - 02:00 PM	Summary of Breakout Sessions: Session Leads
02:00 - 03:00 PM	Workshop report & Wrap-up Session
3:00 PM	Adjourn of main workshop
3:00 – 5:00 PM	Report planning session – Organizers/Breakout leads

Appendix D: Workshop Participants

Jeff Boote, Sandia National Laboratory
Todd Bowman, Los Alamos National Laboratory
Donagh Buckley, EMC
Dave Cohen, Intel
Paul Curtis, Ciena
Phil DeMar, Fermilab
Parks Fields, Los Alamos National Laboratory
Andrea Fumagalli, University of Texas Dallas
Garrett Granroth, Oak Ridge National Laboratory
Chin Guok, ESnet
Deniz Gurkan, University of Houston
Shu Huang, RENCI
Robert Jacob, Argonne National Laboratory
Dimitrios Katramatos, Brookhaven National Laboratory
Greg Lauer, BBN
Tom Lehman, University of Maryland
Tony Lentine, Sandia National Laboratory
Inder Monga, ESnet
Thomas Ndousse-Fetter, DOE
Peter Nugent, Lawrence Berkeley National Laboratory
Dhabaleswar K. Panda, Ohio State University
Amedeo Perazzo, SLAC
Donald Petravick, Oak Ridge National Laboratory
Nagi Rao, Oak Ridge National Laboratory
Steve Schwab, University of Southern California/Information Sciences Institute
Brian Tierney, ESnet
Malathi Veeraraghavan, University of Virginia
Venkat Vishwanath, Argonne National Laboratory
Vinod Vokkarane, UMass Lowell
Bing Wang, University of Connecticut
Dean Williams, Lawrence Livermore National Laboratory
John Wu, Lawrence Berkeley National Laboratory
Wenji Wu, Fermilab
Xi Yang, University of Maryland
S. J. Ben Yoo, University of California Davis
Dantong Yu, Brookhaven National Laboratory