Synergistic Challenges in Data-Intensive Science and Exascale Computing

Summary Report of the Advanced Scientific Computing Advisory Committee (ASCAC) Subcommittee

Draft version, February 28, 2013

Contents

List of Figures

List of Tables

Chapter 1

Executive Summary

The ASCAC Subcommittee on Synergistic Challenges in Data-Intensive Science and Exascale Computing has reviewed current practice and future plans in multiple science domains in the context of the Big Data and the Exascale Computing challenges that they will face in the future. The review drew from public presentations, workshop reports and expert testimony. Data-intensive research activities are increasing in all domains of science, and exascale computing is a key enabler of these activities. We briefly summarize below the key findings and recommendations from this report from the perspective of identifying investments that are most likely to positively impact both data-intensive science goals and exascale computing goals.

Finding 1: There are opportunities for investments that can benefit both data-intensive science and exascale computing. There are natural synergies among the challenges facing data-intensive science and exascale computing, and advances in both are necessary for next-generation scientific breakthroughs. Data-intensive science relies on the collection, analysis and management of massive volumes of data, whether they are obtained from scientific simulations or experimental facilities or both. In both cases (simulation or experimental), investments in exascale systems or, more generally, in "extreme-scale" systems[1] will be necessary to analyze the massive data involved in DOE's science missions.

The Exascale Computing Initiative [8] envisions exascale computing to be a sustainable technology that exploits economies of scale. An industry ecosystem for building exascale computers will necessarily include the creation of higher-volume extreme-scale system components which will be beneficial for data analysis solutions at all scales. These components will include innovative memory hierarchies and data movement optimizations that will be essential for all analysis components in a data-intensive science workflow in the 2020+ timeframe.

For example, high-throughput reduction and analysis capabilities are essential when processing large volumes of data generated by science instruments. While the computational capability needed within a single data analysis tier of an experimental facility may not be at the exascale, extreme scale processors built for exascale systems will be well matched for use in different tiers of data analysis, since these processors will be focused (for example) on optimizing the energy impact of data movement. Further, there is a clear synergy between algorithms for near-sensor computing in experimental facilities and algorithms for in-situ analysis in simulations.

The Exascale Computing Initiative has also identified the need for innovations in applications

---

[1]As in past reports, we use "exascale systems" to refer to systems with an exascale capability and "extreme-scale systems" to refer to all classes of systems built using exascale technologies which include chips with hundreds of cores and different scales of interconnects and memory systems.

and algorithms to address fundamental challenges in extreme-scale systems related to concurrency, data movement, energy efficiency and resilience. Innovative solutions to these challenges will jointly benefit analysis and computational algorithms for both data-intensive science and exascale computing. Finally, advances in networking facilities (as projected for future generations of ESNet [7]) will also benefit both data-intensive science and exascale computing.

Finding 2: Integration of data analytics with exascale simulations represents a new kind of workflow that will impact both data-intensive science and exascale computing.
In the past, the computational science workflow was represented by large-scale simulations followed by off-line data analyses and visualizations. Today's ability to understand and explore gigabyte and some petabyte spatial-temporal high-dimensional data in this workflow is the result of decades of research investment in data analysis and visualization. However, exascale data being produced by experiments and simulations are rapidly outstripping our current ability to explore and understand them. Exascale simulations require that some analyses and visualizations be performed while data is still resident in memory, so-called *in-situ* analysis and visualization, thus necessitating a new kind of workflow for scientists. In addition, we need new algorithms for scientific data analysis and visualization along with new data archiving techniques that allow for both in-situ and post processing of petabytes and exabytes of simulation and experimental data. This new kind of workflow will impact data-intensive science due to its tighter coupling of data and simulation, while also offering new opportunities for data analysis to steer computation.

In addition, in-situ analysis will impact the workloads that high-end computers have traditionally been designed for. Even for traditional floating-point-intensive applications, the addition of analytics will change the workload to include (for example) larger numbers of integer operations and branch operations than before. Design and development of scalable algorithms and software for mining big data sets, as well as an ability to perform approximate analysis within certain time constraints will be necessary for effective in-situ analysis. In the past, different assumptions were made for designing high-end computing systems vs. analysis and visualization systems. Tighter integration of simulation and analytics in the science workflow will impact co-design of these systems for future workloads, and will require development of new classes of proxy applications to capture the combined characteristics of simulations and analytics.

Finding 3: There is an urgent need to simplify the workflow for data-intensive science.
Analysis and visualization of increasingly larger-scale data sets will require integration of the best computational algorithms with the best interactive techniques and interfaces. The workflow for data-intensive science is complicated by the need to simultaneously manage large volumes of data as well as large amounts of computation to analyze the data, and this complexity is increasing at an inexorable rate. These complications can greatly reduce the productivity of the domain scientist, if the workflow is not simplified and made more flexible. For example, the workflow should be able to transparently support decisions such as when to move data to computation or computation to data. The recent proposal for a Virtual Data Facility (VDF) will go a long way in simplifying the workflow for data-intensive science because of its integrated focus on data-intensive science across the DOE ASCR facilities.

Finding 4: There is a need to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing. Earlier workflow models allowed for a separation of concerns between computation and analytics that is no longer possible as computation and data analysis become more tightly intertwined. Further, the separation of

concerns allowed for science to progress with personnel that may be experts in computation or in analysis, but not both. This approach is not sustainable in data-intensive science where the workflow for computation and analysis will have to be co-designed. There is a need for investments to increase the number of computer and computational scientists trained in both exascale and data-intensive computing to advance the goals of data-intensive science.

Recommendation 1: The DOE Office of Science should give higher priority to investments that can benefit both data-intensive science and exascale computing so as to leverage their synergies. The findings in this study have identified multiple technologies and capabilities that can benefit both data-intensive science and exascale computing. Investments in such dual-purpose technologies will provide the necessary leverage to advance science on both data and computational fronts. For science domains that need exascale simulations, commensurate investments in exascale computing capabilities and data infrastructure are necessary for advancement. In other domains, extreme-scale components of exascale systems will be well matched for use in different tiers of data analysis, since these processors will be focused on optimizing the energy impact of data movement. Further, innovations in applications and algorithms to address fundamental challenges in concurrency, data movement, and resilience will jointly benefit data analysis and computational techniques for both data-intensive science and exascale computing. Finally, advances in networking (as projected for future generations of ESNet technology) will also benefit both data-intensive science and exascale computing.

Recommendation 2: DOE ASCR should give higher priority to investments that simplify the science workflow and improve the productivity of scientists involved in exascale and data-intensive computing. We must pay greater attention to simplifying human-compute-interface design and human-in-the-loop workflows for data-intensive science. To that end, we encourage the recent proposal for a Virtual Data Facility (VDF) because it will provide a simpler and more usable portal for data services than current systems. A significant emphasis must be placed on developing a collection of scalable data analytics and data mining algorithms and software components that can be used as building blocks for sophisticated analytics pipelines and flows. We also recommend the creation of new classes of proxy applications to capture the combined characteristics of simulation and analytics, so as to help ensure that computational science and computer science research in ASCR are better targeted to the needs of data-intensive science.

Recommendation 3: DOE ASCR should adjust investments in programs such as fellowships, career awards, and funding grants, to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing. There is a significant gap between the number of current computational and computer scientists trained in both exascale and data-intensive computing and the future needs for this combined expertise in support of DOE's science missions. Investments in ASCR such as fellowships, career awards, and funding grants should look to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing.

Chapter 2

Introduction

2.1 Big Data and the Fourth Paradigm

Historically, the two dominant paradigms for scientific discovery have been theory and experiments, with large-scale simulations emerging as the third paradigm in the 20th century. In many cases, large scale simulations are accompanied by the challenges of *data-intensive computing* . Overcoming the challenges of data-intensive computing has required optimization of data movement across multiple levels of memory hierarchies, and these considerations have become even more important as we prepare for exascale computing. The approaches taken to address these challenges include (a) fast data output from a large simulation for future processing/archiving; (b) minimization of data movement between cache and main memory; (c) optimization of communication across nodes using fast and low-latency networks, and communication optimization; and (d) effective co-design, usage and optimization of system components from architectures to software.

Over the past decade, a new paradigm for scientific discovery is emerging due to the availability of exponentially increasing volumes of data from large instruments such as telescopes, colliders, and light sources, as well as the proliferation of sensors and high-throughput analysis devices. Further, data source, analysis devices, and simulations, connected with current-generation networks that are faster and capable of moving significantly larger volumes of data than in previous generations. These trends are popularly referred to as *big data*. However, generation of data by itself is of not much value unless the data can also lead to knowledge and actionable insights. Thus, the *fourth paradigm*, which seeks to exploit information buried in massive datasets to drive scientific discovery, and has emerged an essential complement to the three existing paradigms. The complexity and challenge of the fourth paradigm arises from the increasing velocity, heterogeneity, and volume of data generation. For example, experiments using the Large Hadron Collider (LHC) currently generate tens of petabytes of reduced data per year, observational and simulation data in the climate domain is expected to reach exabytes by 2021, and light source experiments are expected to generate hundreds of terabytes per day. The heterogeneity of the data adds further complexity; for example, data may be spatio-temporal, unstructured, or may be derived data with complex formats.

Analysis of this large volume of complex data to derive knowledge, therefore, requires *data-driven* computing, where the data drives the computation and control including complex queries, analysis, statistics, hypothesis formulation and validation, and data mining. Figure 2.1 illustrates a typical knowledge discovery life-cycle for big data [3], which consists of the following steps:

1. Data Generation: Data may be generated by instruments, experiments, sensors, or supercomputer simulations. In fact, a supercomputer can be thought of as another type of instrument,

which generates data by simulating mathematical models at large-scale. However, to derive knowledge from the simulations, the data (which is typically spatio-temporal) must be processed, mined and visualized, like data from any other instrument.

2. Data Processing and Organization: This phase entails (re)organizing, processing, deriving subsets, reduction, visualization, query analytics, distributing, and many other aspects. This may also include combining data with external data or historical data, in essence creating a virtual data warehouse. For example, in LHC, this phase may include common operations on and derivations from raw data that is generated with appropriate creation of metadata. This in turn, becomes a virtual warehouse of data used by thousands of scientists for their knowledge discovery process. Similarly, a collection of simulation and observational data in the climate application case may be used by thousands of scientists for their discovery process.

3. Data Analytics, Mining and Knowledge Discovery: Given the size and complexity of data and the need for both top-down and bottom up discovery, scalable algorithms and software need to be deployed in this phase. The discovery process typically becomes very specific based on the scientific problem under consideration. For example, the data set may be used to predict extreme events, or for understanding long term macro climate patterns. Repeated evaluations, what-if scenarios, predictive modeling, correlations, causality and other mining operations at scale are part of this phase.

4. Actions, Feedback and Refinement: Insights and discoveries from previous phases help close the loop to determine new simulations, models, parameters, settings, observations, thereby, making the closed loop a virtuous cycle for big data.

While Figure 2.1 represents a common high-level approach to data-driven knowledge discovery, there can be important differences among different science domains as to how data is produced, consumed, stored, processed, analyzed and mined. The options to perform these activities can also depend on where the data originates. For example, on an exascale system, there may be significant opportunity to perform in-situ analytics using the same (or part of) system as that used for generating the data. Other scenarios that requires combining the analysis of historical data and simulation data may require a different approach.

2.2 Intertwined Requirements for Big Data and Big Compute

Though the third and fourth paradigms mentioned earlier depend on "Big Compute" and "Big Data" respectively, it would be a mistake to think of them as independent activities. Instead, their requirements are tightly intertwined since they both contribute to a shared goal of scientific discovery. In the 2022 timeframe that this report is focused on, Big Compute will be exemplified by exascale systems or, more generally, "extreme-scale" systems. (As in past reports, we use "exascale systems" to refer to systems with an exascale capability and "extreme-scale systems" to refer to all classes of systems built using exascale technologies which include chips with hundreds of cores and different scales of interconnects and memory systems.) Many science applications increasingly compute ensembles to study "what-if" scenarios, as well as to understand potential behaviors over a range of different conditions. Whereas traditional approaches for analysis typically facilitate examining one dataset at a time, the growth in ensemble collections of simulations makes it increasingly important to quantify error and uncertainty within and across ensembles.

For example, data-intensive simulations on Big Compute exascale systems will be used to generate volumes of Big Data that are comparable to the data volumes generated by many scientific
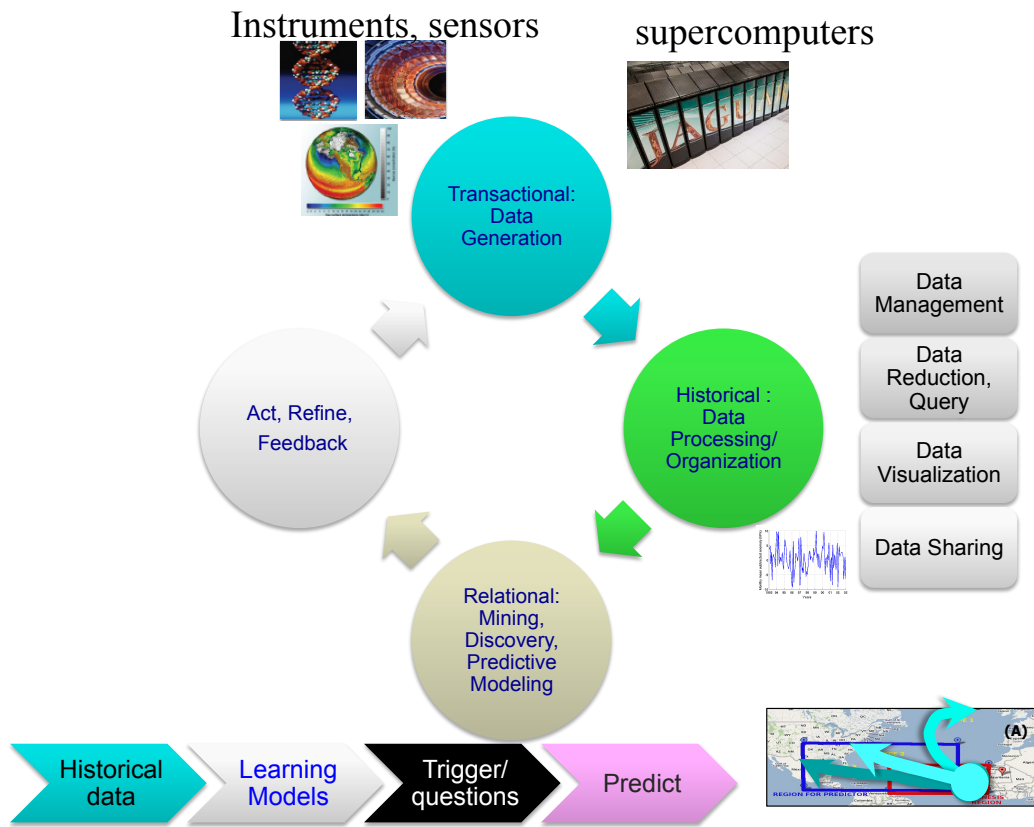
Figure 2.1: A knowledge-discovery life-cycle for Big Data

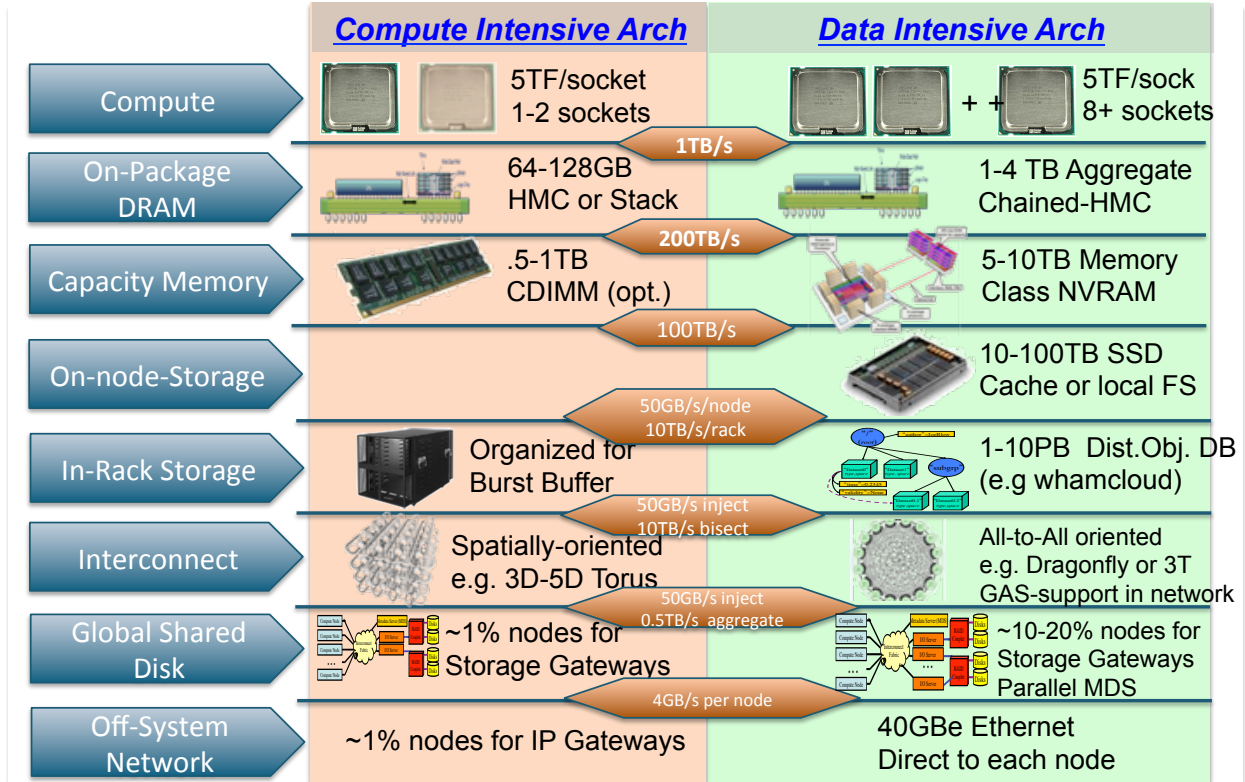|	Compute Intensive Arch	Data Intensive Arch
Compute	5TF/socket 1-2 sockets	+ + 5TF/sock 8+ sockets
	1TB/s	
On-Package DRAM	64-128GB HMC or Stack	1-4 TB Aggregate Chained-HMC
	200TB/s	
Capacity Memory	.5-1TB CDIMM (opt.)	5-10TB Memory Class NVRAM
	100TB/s	
On-node-Storage		10-100TB SSD Cache or local FS
	50GB/s/node 10TB/s/rack	
In-Rack Storage	Organized for Burst Buffer	1-10PB Dist.Obj. DB (e.g whamcloud)
	50GB/s inject 10TB/s bisect	
Interconnect	Spatially-oriented e.g. 3D-5D Torus	All-to-All oriented e.g. Dragonfly or 3T GAS-support in network
	50GB/s inject 0.5TB/s aggregate	
Global Shared Disk	~1% nodes for Storage Gateways	~10-20% nodes for Storage Gateways Parallel MDS
	4GB/s per node	
Off-System Network	~1% nodes for IP Gateways	40GBe Ethernet Direct to each node

Figure 2.2: Strawman compute-intensive vs. data-intensive computer architectures in the 2017 timeframe. (Figure courtesy of NERSC.)

instruments. Likewise, the volumes of Big Data generated by the data-driven paradigm will need to be analyzed by Big Compute exascale or extreme-scale systems. Historically, the data-intensive and data-driven approaches have evolved somewhat independently of each other, and each has not leveraged many optimizations developed in the context of other, even though they are faced with similar challenges related to optimizing data movement, data management, and data reduction. As we head to the exascale timeframe, it will be critical to exploit synergies between the two approaches, even though some fundamental differences may remain between the two approaches.

Further, since the data-driven paradigm is relatively recent, current systems have been designed using workloads that are focused more on computation requirements than on data requirements. Figure 2.2[1] compares strawman computer architectures in the 2017 timeframe built for compute-intensive vs. data-intensive requirements. The compute-intensive architecture follows current supercomputing trends and aims to maximize the computational density and memory bandwidth, assuming that the working set data will mostly fit in main memory. The data-intensive architecture will focus on tighter integration of storage and computation, to accommodate workloads with data sets that exceed main memory capacities. Exascale systems will need to span the range of compute-intensive and data-intensive architectures, though early exascale systems will likely be aligned with compute-intensive designs.

The need for data-intensive computer architectures is motivated by the fact that past architecture designs have been based on established workloads that pre-date the fourth paradigm. For

---

[1]This figure and analysis was obtained by courtesy of NERSC.

example, compute-intensive architectures have evolved over decades to provide increased optimization of floating-point computations in HPC workloads. Likewise, commercial servers have been optimized for transactional and query-based workloads, and mobile processors have been optimized for media processing. The characteristics and requirements of data analytics and mining applications that are central to the fourth paradigm have not as yet had a major impact on the design of computer systems, but that is likely to change in the near future. For example, as early as 2015-2016 with the DOE Office of Science Trinity/NERSC 8 procurements visualization and data analysis capabilities will be built into the main compute resources. For fast I/O near compute nodes, burst buffers will be included. They serve as both the primary resiliency mechanism and as persistent, 'slow memory' near data for in-transit analytics/visualization.

In the commercial world, the idea of building data-intensive computer architecures at the datacenter scale has well established for a few years now [1]. Even there, it was recognized that the data and computing requirements of a data center cannot be addressed effectively with off-the-shelf software and systems. For example, Google has its own software stack that extends from the kernel up through multiple levels of storage and networking into the specialized infrastructure layers provided by families of applications such as Gmail, Google+, Youtube, and Docs. Many of these applications rely on the BigTable structure. The largest BigTable cluster in 2009 at Google held 70+ petabytes of data, with a sustained I/O rate of 30+ gigabytes/second. To handle larger scales of data, the Spanner [4] system was recently introduced with the goal of scaling "up to millions of machines across hundreds of datacenters".

Some preliminary work has been performed to understand the differences between the requirements of data mining and analytics applications compared with other kinds of workloads [13]. This analysis used a variety of application suites including integer application benchmarks (SPEC INT from SPEC CPU2000), floating point application benchmarks (SPEC FP), multimedia application benchmarks (MediaBench) and decision support application benchmarks (TPC-H from Transaction Processing Council. The MineBench benchmark[2] was used as a representative of data mining and analytics applications. The results of the comparison are shown in Table 2.1. For example, one attribute that signifies the uniqueness of data mining applications is the number of data references per instruction retired. This rate is 1.10, whereas for other applications, it is significantly less, clearly demonstrating the data intensive nature of these applications. The L2 miss rates are considerably high for data mining application as well. The reason for this is the inherent streaming nature of data retrieval and computation dependent access patterns, which do not provide enough opportunities for data reuse. These characteristics suggest that memory hierarchies in exascale and extreme-scale systems should be made more flexible so as to support both Big Compute and Big Data applications. Given the preliminary nature of the study in [13], more work is clearly needed in developing proxy applications to ensure that co-design methodologies include requirements from compute-intensive, data-intensive and data-driven applications in guiding the design of future computing facilities for scientific discovery.

2.3 Interpretation of Charge

The subcommittee appreciated the timeliness of the charge, a copy of which is included in Appendix A. Data-intensive research activities are increasing in all domains of science, and exascale computing is a key enabler of these activities. To focus our efforts in this study, we made the following assumptions when interpreting the charge:

---

[2]See `http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html` for details.

Parameter	Benchmarks				
	SPECINT	SPECFP	MediaBench	TPC-H	MineBench
Data References	0.81	0.55	0.56	0.48	1.10
Bus Accesses	0.030	0.034	0.002	0.010	0.037
Instruction Decodes	1.17	1.02	1.28	1.08	0.78
Resource Related Stalls	0.66	1.04	0.14	0.69	0.43
ALU Instructions	0.25	0.29	0.27	0.30	0.31
L1 Misses	0.023	0.008	0.010	0.029	0.016
L2 Misses	0.0030	0.0030	0.0004	0.0020	0.0060
Branches	0.13	0.03	0.16	0.11	0.14
Branch Mispredictions	0.0090	0.0008	0.0160	0.0006	0.0060

Table 2.1: A comparison of selected performance parameters for different benchmarks with data analytics and mining workloads [13]. The numbers shown here for the parameters are values per instruction.

- There are several Federal government programs under way to address the challenges of the "big data" revolution [9] to further scientific discovery and innovation. Likewise, there are multiple efforts in industry to address big data challenges in the context of commercial data, including the use of cloud computing to enable analysis of big data. We explicitly restricted the scope of our study to the intersection of big data challenges and exascale computing in the context of data-intensive science applications that are part of the Office of Science's mission.

- This scope includes experimental facilities and scientific simulations that exemplify the Office of Science's unique role in data-intensive science. While some of our conclusions may be more broadly applicable e.g., to Department of Defense exascale applications, we focused on scenarios that centered on the Office of Science's mission needs. Other areas (e.g., cyber defense) were ruled to be out of scope for this study.

- The charge did not specify a timeframe to be assumed for our recommendations. However, the focus on synergies with exascale computing suggests that a 10-year timeframe should be our primary focus, since the current roadmap for exascale computing estimates the availability of exascale capability approximately in the 2022 timeframe. Nearer term (e.g., 5-year) considerations also deserve attention since they will influence the migration path to exascale computing. Likewise, we do not wish to ignore the potential impact of longer-term technology options (e.g., in the 20-year timeframe), but their study represents a secondary goal for this study.

- The charge highlighted the importance of identifying investments that are most likely to positively impact both data-intensive science research goals and exascale computing goals. Since other studies are investigating the prioritization of investments across the Office of Science (e.g., the recent Facilities study), this study will focus on investments that can leverage synergies between data-intensive science and exascale computing.

Chapter 3

Data Challenges in Science

In this chapter, we briefly summarize data challenges arising in multiple science domains, with the goal of identifying cross-cutting data requirements from different domains to guide our study. Though the set of science domains studied may not be complete, our aim was to cover enough of a variety so as to understand the range of requirements underlying data-driven science.

Figure 3.1 summarizes projected growth rates for data acquisition from sequencers and detectors relative to those for processor for memory capacities. It conveys the general trend facing multiple science domains. Even if the absolute data requirements do not appear to be a significant obstacles for some science domains today, future trends indicate that most science domains will become data-intensive in the exascale timeframe (and many well before then, including in the current timeframe).

3.1 High Energy Physics

High Energy Physics (HEP) is inherently a data-intensive domain because of the fundamental nature of quantum physics. Advances in HEP require the measurement of probabilities of "interesting" events in large numbers of observations *e.g.,* in 10^{16} or more particle collisions observed in a year. Different experiments can be set up to generate collisions with different kinds of particles, and at different energy levels.

The ATLAS and CMS experiments at the Large Hadron Collider are at the energy frontier and present the greatest data challenges seen by HEP. The detectors in these experiments generate massive amounts of analog data, at rates equivalent to petabytes per second running round the clock for a large fraction of each year.

These data must of necessity be reduced by orders of magnitude in real time to a volume that can be stored at acceptable cost. Today, about 1 gigabyte per second or around 10 petabytes per year is the limit for a major experiment. Further, this data is also required to be distributable worldwide at an acceptable cost, so as to be shared among scientists in different laboratories and countries.

After the initial reduction, data volumes are inflated by storing derived data products, replication for safety and efficient access, and by the need for storing even more simulated data than the experimental data. ATLAS currently stores over 100 petabytes and this volume is rising rapidly.

3.1.1 Real-time data reduction

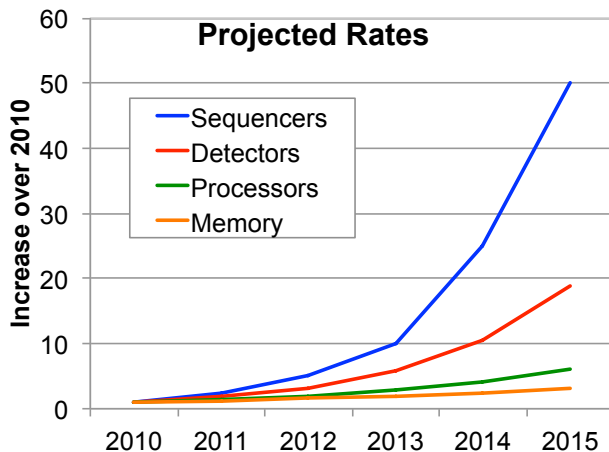Data reduction relies on two strategies:

Figure 3.1: Projected growth rates for data acquisition and processor/memory capacities. (Figure courtesy of Kathy Yelick.)

- Zero suppression, or more correctly "below threshold suppression". Data from each collision are produced by millions of sensitive devices. The majority of the devices record an energy deposit consistent with noise and inconsistent with the passage of a charged particle. These readings are cut out at a very early stage. This is not a trivial choice. For example a free quark might deposit 11% of the energy expected for a normal charged particle, and would very likely be rendered invisible by the zero suppression. However, discovery of a free quark would be a ground-breaking science contribution that could well be worthy of a Nobel prize.

- Selection of the "interesting" collisions. This proceeds in a series of steps. The first step is to decide which collisions merit the digitization (and zero suppression) of the analog data in the detection devices. This produces hundreds of gigabytes per second of digital data flow on which further increasingly compute-intensive pattern recognition and selection is performed by trivially parallel "farms" of thousands of computers.

Typically, the group responsible for each major physics analysis topic is given a budget of how many collisions per second it may collect and provides the corresponding selection algorithms. Physicists are well aware of the dangers of optimizing for the "expected" new physics and being blind to the unexpected. For real time data reduction, all the computing is necessarily performed very close to the source of data.

3.1.2   Distributed Data Processing and Analysis

Construction of each LHC detector necessarily required that equipment costing hundreds of millions of dollars be funded and built by nations around the world and assembled in an underground pit at CERN. The motivation for this consolidation of experimental equipment is clear. However, there is no corresponding motivation to consolidate computing and storage infrastructure in the same manner. Instead, for both technical and political reasons, a distributed computing and storage infrastructure has been the only feasible path to pursue despite some of the complexities involved.

The workflow for HEP scientists involved in an experiment like ATLAS or CMS typically consist of the following tasks:

1. Reconstruction of likely particle trajectories and energy-shower deposits for each collision from the individual measurements made by the millions of sensitive devices.

2. Simulation of the detector response, followed by reconstruction, for all the expected collisions plus a range of new physics hypotheses. Given the cost of acquiring the real collisions, it is cost effective (but still expensive) to generate several times as much simulated data as real data. The detector response simulation is extremely compute-intensive.

3. Selection of a subset of the collisions, and often a subset of the reconstructed objects to create relatively compact datasets for use by particular physics analysis groups.

4. Detailed analysis of the subset datasets in comparison with the relevant simulated datasets to derive publishable science contributions.

Several hundred national and university-based computing centers form the Worldwide LHC Computing Grid (WLCG). They bring a total of almost 300,000 compute cores and hundreds of petabytes of disk and tape storage capacity. Much of the communications flows over the LHC Optical Private Network (LHCOPN), composed typically of dedicated wavelengths on national research networks plus some explicitly purchased international links (e.g. $6x10$ Gbps between CERN and the US). The rising capabilities of general-purpose research and education networks worldwide has begun to make "anywhere to anywhere" high-bandwidth data transfer feasible. LHC distributed computing is evolving rapidly to exploit this capability.

WLCG resources are made available via Grid middleware developed in the US and Europe. The European and US middleware have some differences, but interoperate successfully, providing the basic utilities for authentication, data movement, job submission, monitoring, accounting and trouble ticketing. The US Open Science Grid and European Grids are also open to other sciences either as opportunistic users or as stakeholders sharing their own resources. The Grid approach pre-dates virtualization and Cloud technology. As virtualization is increasing deployed there is an increasing convergence with the Cloud world.

3.1.3   Challenges and Opportunities

Data Storage and Access Technologies: The million-fold reduction of data performed in real time is driven by the limitations of data storage and access technology, not by science. It amounts to a blinkered view of nature — no problem if you look in the right direction, but with a danger of being unable to see the unexpected. If technology allowed storing and analyzing one thousand times more data at tolerable cost, this would immediately become part of the process of this science.

Efficient Execution on Future Computer Architectures: A majority of all available computing power and storage is currently devoted to detailed simulation of how the detector and the reconstruction software respond to collision products. The ratio of simulated to real data is barely sufficient now and is at risk of further deterioration as the LHC collision rate increases as planned. Unlike many simulations, HEP event simulation and reconstruction is not naturally vector-processor or GPU friendly. The main hope of achieving acceptable simulation statistics is, nevertheless, to understand how to exploit upcoming extreme-scale architectures while still retaining the millions of lines of collaboratively written code in a practical manner.

Towards Widely Applicable Distributed Data Intensive Computing: The distributed computing environment for the LHC has proved to be a formidable resource, giving scientists access to huge resources that are pooled worldwide and largely automatically managed. However, the scale of operational effort required is burdensome for the HEP community, and will be hard to replicate in other science communities. Could the current HEP distributed environments be used

as a distributed systems laboratory to understand how more robust, self-healing, self-diagnosing systems could be created?

3.2 Climate Science

Climate science is a prominent example of a discipline in which scientific progress is critically dependent on the availability of a reliable infrastructure for managing and accessing of large and heterogeneous quantities of data on a global scale. It is inherently a collaborative and multi-disciplinary effort that requires sophisticated modeling of the physical processes and exchange mechanisms between multiple Earth realms (atmosphere, land, ocean and sea ice) and comparison and validation of these simulations with observational data from various sources, possibly collected over long periods of time. The climate community has worked for the past decade on concerted, worldwide modeling activities led by the Working Group on Coupled Modeling (WGCM), sponsored by the World Climate Research Program (WCRP) and leading to successive reports by the International Panel on Climate Change (IPCC). The fifth assessment (IPCC-AR5) is currently under way and due out by the end of 2013. These activities involve tens of modeling groups in as many countries, running the same prescribed set of climate change scenarios on the most advanced super-computers and producing several petabytes of output containing hundreds of physical variables spanning tens and hundreds of years. These data sets are held at distributed locations around the globe but must be discovered, downloaded, and analyzed as if they were stored in a single archive, with efficient and reliable access mechanisms that can span political and institutional boundaries (see Figure 3.2).

The same infrastructure must also allow scientists to access and compare observational data sets from multiple sources, including, for example, Earth Observing System (EOS) satellites and Advanced Radiation Measurements (ARM) sites. These observations, often collected and made available in real-time or near real-time, are typically stored in different formats and must be post-processed to be converted to a format that allows easy comparison with models. The need for providing data products on demand, as well as value-added products, adds another dimension to the needed capabilities. Finally, science results must be applied at multiple scales (global, regional, and local) and made available to different communities (scientists, policy makers, instructors, farmers, and industry). Because of its high visibility and direct impact on political decisions that govern human activities, the end-to-end scientific investigation must be completely transparent, collaborative, and reproducible. Scientists must be given the environment and tools for exchanging ideas and verifying results with colleagues in opposite time zones, investigating metadata, tracking provenance, annotating results, and collaborating in developing analysis applications and algorithms.

3.3 Combustion

3.3.1 Motivation

Over 70% of the 86 million barrels of crude oil that are consumed in this nation each day are used in internal combustion engines. The nation spends about 1 billion dollars a day on imported oil. Accompanying the tremendous oil consumption is the undesirable emissions nitric oxides, particulates, and CO2 production. To mitigate the negative environmental and health implications, there is legislation that mandates reductions in fuel usage per kilometer by 50% in new vehicles by 2030 and greenhouse gases by 80% by 2050. Although this may appear to be far off in time, the time required to bring new vehicle technologies to market and to permeate the overall fleet is also lengthy.
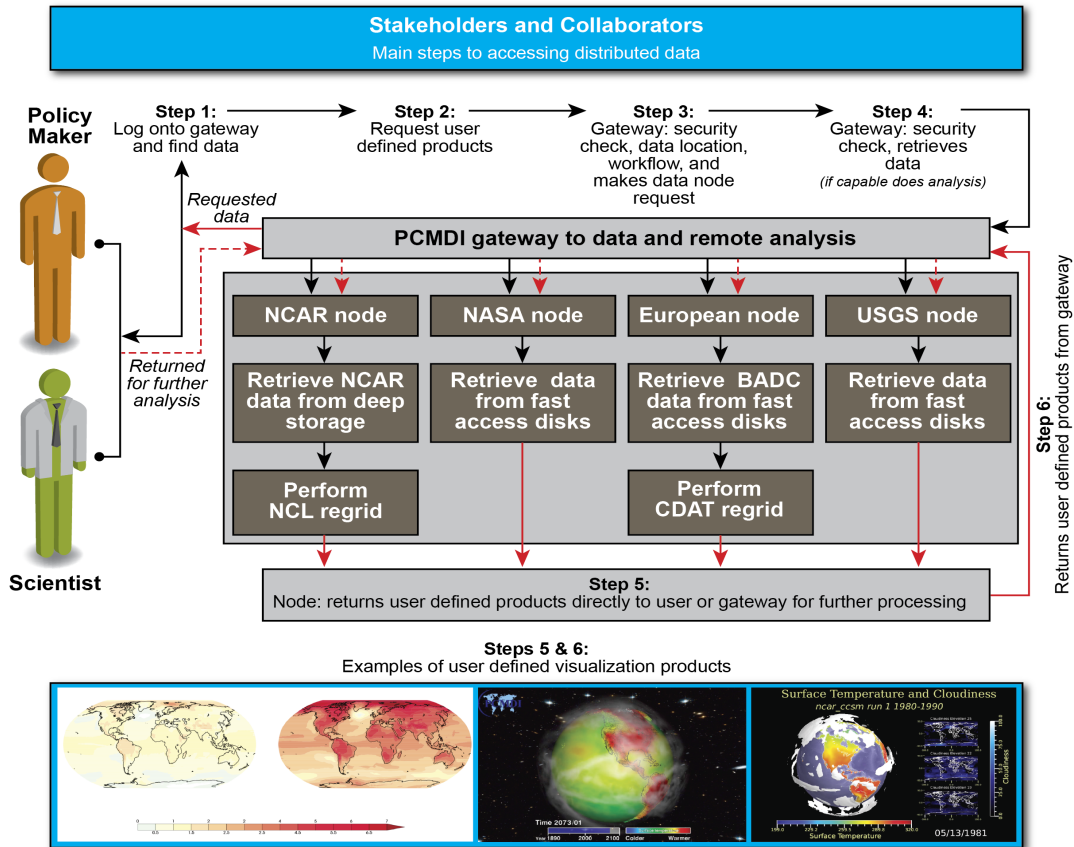
Figure 3.2: Scientists and policy makers remotely access simulation and observation data and seamlessly preform analysis to uncover climate change results.

Hence, the urgent need for a concerted effort to develop non-petroleum-based fuels and their efficient, clean utilization in transportation is warranted by concerns over energy sustainability, energy security, and global warming. Drastic changes in the fuel constituents and operational characteristics of automobiles and trucks are needed over the next few decades as the world transitions away from petroleum-derived transportation fuels. Conventional empirical approaches to developing new engines and certifying new fuels have only led to incremental improvements, and as such they cannot meet these enormous challenges in a timely, cost-effective manner. Achieving the required high rate of innovation will require computer-aided design, as is currently used to design the aerodynamically efficient wings of airplanes and the molecules in ozone-friendly refrigerants. The diversity of alternative fuels and the corresponding variation in their physical and chemical properties, coupled with simultaneous changes in automotive design/control strategies needed to improve efficiency and reduce emissions, pose immense technical challenges.

A central challenge is predicting combustion rates and emissions in novel low temperature compression ignition engines. Compression ignition engines have much higher efficiencies than spark ignited gasoline engines (only 20% efficiency) with the potential to increase by as much as 50% if key technological challenges can be overcome. Current diesel engines suffer from high nitric oxide and particulate emissions requiring expensive after-treatments. To reduce emissions while capitalizing on the high fuel efficiency of compression ignition engines constrains the thermo-chemical space that advanced engines can operate in, i.e. they much burn overall fuel-lean, dilute and at lower temperatures than conventional diesel engines. Combustion in this new environment is governed by previously unexplored regimes of mixed-mode combustion involving strong coupling between turbulent mixing and chemistry characterized by intermittent phenomena such as auto-ignition and extinction in stratified mixtures burning near flammability limits. The new combustion regimes are poorly understood, and there is a dearth of predictive models for engine design operating in these regimes. Basic research in this area is underscored in the Department of Energy Basic Energy Sciences workshop report on "Basic Energy Needs for Clean and Efficient Combustion of 21st Century Transportation Fuels" which identified a single overarching grand challenge: to develop a "validated, predictive, multi-scale, combustion modeling capability to optimize the design and operation of evolving fuels in advanced engines for transportation applications."

3.3.2 Exascale use case

Exascale computing offers the promise of enabling combustion simulations in parameter regimes relevant to next-generation combustors burning alternative fuels that are needed to provide the underlying science required to design fuel efficient, clean burning vehicles, planes, and power plants for electricity generation. Use cases based on this target have been designed to provide a focal point for exascale co-design efforts. One representative exascale combustion use case corresponds to an advanced engine design, homogeneous charge compression ignition, which relies on autoignition chemistry at high pressures to determine the combustion phasing and burn rate. We estimate that to describe this system would require a mesh with 10^{12} points with more than 100 variables per point to describe the fuel chemistry. Simulation of this case would require memory usage of around 3 PB integrated for 1.2 million time steps representing 6 ms of physical time, requiring 5×10^{11} cpu-hrs and generating 400 PB of raw data (1 PB per 0.5 hour).

3.3.3 Comprehensive combustion workflow design space exploration

In addition to understanding the specific performance characteristics of individual algorithms, we need to understand the overall workload and impact of various designs in the context of the end-

to-end combustion workflow. The design space is large, including such options as placement of analysis tasks, if/when/how/where to move or make data available from the solvers to the analysis tasks, scheduling analysis tasks, and availability of hardware resources.

Consider, for example, the selection of cores for analysis, and how they interact with the main simulation. In one case, the analysis operations are delegated to discrete cores on the simulation node (staging cores), and compare this to time sharing with the simulation. In the first case, the analysis operators will access the data asynchronously and will not impact the data locality optimizations for the simulation.

However, this specialization will reduce the peak compute capability available to the simulation. In the second case, the simulation can access all the available compute resources, but will have to wait for the analysis tasks to complete before making progress. Moreover, the sharing of resources could have a larger impact on the simulation performance due to cache pollution. These design choices must be evaluated in the context of specific use cases and for a variety of hardware platforms, both current and future using hardware simulators.

Figure 3.3 illustrates the rich co-design space of end-to-end science workflows on exascale machines where tradeoffs have to be made between selection of optimal analysis compute resources, synchronization and scheduling analysis and solve, data access, placement and persistence.
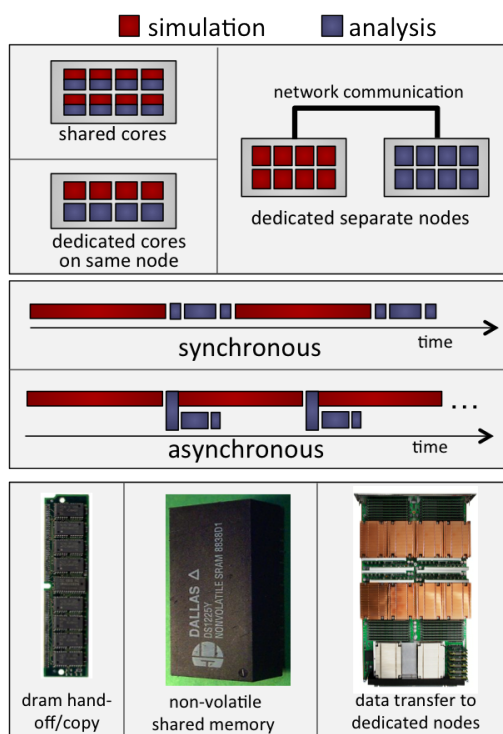

Figure 3.3: Rich co-design space of end-to-end workflows on exascale machines.

3.4 Biology and Genomics

The vision of the grand challenges confronting the biology community over the next few years showcases several different types of data management challenges arising from the extreme volumes

of data being produced by simulation, analysis, and experiment. Ranging from the understanding of the fundamental connections be- tween the chemistry of biomolecules and the function of whole organisms, to the processing and interpretation of complex reaction pathways inside organisms used in bioremediation, to the simulation and understanding of the human brain itself, these complex challenge problems use data in ways that extend the requirements for extreme-scale data management and visualization. In particular, the Opportunities in biology at the Extreme Scale of Computing report [5] highlights 'Grand Challenge Issues' in biology for exascale computing:

- Biophysical simulations of cellular environments, either in terms of long time dynamics, crowded environments and the challenges therein, or coarse graining dynamics for entire cellular systems.

- Cracking the 'signaling code' of the genome across the tree of life, implying a reconstruction of large-scale cellular networks across species, and across time points and environmental conditions.

- Re-engineering the units of biological function at a whole organism scale.

- Correlating observational ecology and models of population dynamics in microbial community systems, and from this basis, to apply these models to microevolutionary dynamics in microbial communities.

- Reverse engineering the brain to understand complex neural systems, as well as to provide new methods to integrate large-scale data, information and simulation dynamics, and ultimately to provide societal benefits for health. And perhaps most importantly, to learn how nature's design of a biologically-based knowledge machine beats all estimates for low power systems to compute or store information when compared to even the most power efficient hardware systems now being conceived for construction in the future.

3.5   Light Sources

Unlike the previous sections which focused on scientists in specific domains, users of light sources can be very diverse. They come from many different science domains (*e.g.,* ranging from paleontologists to material scientists), and in many cases, have little or no familiarity with high performance computing.

The Advanced Photon Source (APS) at ANL provides this nation's brightest x-ray beams for research in multiple scientific disciplines. It includes about 65 beam lines organized in 34 sectors, each with different data rates. For example, a tomography beam line can generate 1 terabyte of data per day. While many users still bring in their own external storage devices to capture experimental data, this will become increasingly untenable with increasing data rates when some future light sources are expected to generate 1 terabyte of data per second! Likewise, the historical trend of individual users performing manual analysis of their data will not be feasible with increasing data sizes. While some communities (*e.g.,* protein crystallographers) have data analysis user groups that contribute to an ecosystem for data analysis tools, there is no standard science workflow for users of light sources. The availability of GridFTP and GlobusOnline services has been helpful in supporting some of the APS users. However, in general, the primary focus on maximizing utilization of APS as a scientific facility has made it challenging to justify resources to build software services and capabilities that are necessary to increase the scientist's productivity when using a light source.

In contrast to APS, the Linac Coherent Light Source (LCLS) is a relatively new facility that was able to take a different approach to data acquisition. It provides users access to an offline storage

facility with a capacity of approximately 2.5 petabytes, where data is stored for 2 years, and an on-line cache of approximately 50 terabytes, where data is stored for 5 days [14]. While these data volumes do not sound prohibitive in size, they are expected to increase dramatically in the future with the introduction of larger multi mega-pixels cameras and an increased number of stations. All data management can be controlled by the user via the LCLS web portal, including archiving to tape (HPSS) which is implemented as an iRODS service. The success of centralized data services at LCLS bodes well for the recent proposal for a Virtual Data Facility (VDF) to simplify the workflow for scientists. However, the data analysis ecosystem still lags the data management ecosystem; LCLS users need a wide range of data analysis software functionalities, including advanced algorithms and interactive frameworks. In general, there is a strong need for a collaborative, high performance, open source, software ecosystem for data analysis across all the light source facilities.

3.6 Identification of cross-cutting requirements from different domains

Data may be generated by instruments, experiments, sensors, or supercomputer simulations. In fact, a supercomputer can be thought of as another type of instrument, which generates data by simulating mathematical models at large-scale. Data can be processed, organized, mined concurrently with the data generation phase (in-situ or on-line) to some extent depending on the scenario and application. Subsequently, raw data and/or derived data must be stored, organized, managed, distributed and shared for future use (post-processing). The extent and applicability of these options depends on the applications as well as technology and infrastructures (including cost) capabilities to handle volume, velocity, variety and other parameters.

Table 3.1 shows four scenarios as viewed from the data generation phase — two each for simulations and instruments. The table identifies similarities and differences among different scenarios and domains. Scalable data analysis, mining, storage and visualization prove to be critically important in all scenarios. The requirements for storage (particularly for distribution and future usage), sharing, management, analytics etc. varies. As an illustration, consider the two scenarios of exascale simulations generating data. In the point design oriented scenarios (e.g., combustion), a lot of analytics and processing can be done within the framework of data generation phase using in-situ processing. On the other hand, for applications such as climate, in addition to point analysis, which can leverage in-situ analytics, accessing historical data, observational data or other ensembles may be important to derive insights. Furthermore, a lot more data may need to be stored, derived and organized in order for a much larger community to use the data and its derivation for many applications within each domain. For example, spatio-temporal slices to execute regional assessments, understand extreme events or evaluate broader climate trends. The communities, applications, processing, mining and discovery applications may vastly vary, thereby, requiring a much more extensive and sophisticated management, storage, distribution and provenance capabilities.

Data Generation Phase (scenarios)	Comments (Overview)	Transactional / in situ processing requirements	Storage for Post processing	Sharing and distribution	Visualization
Exascale Simulations (1) Design Oriented (e.g., combustion, CFD)	Generation of data from simulations.	(1) Data reduction for post processing (2) feature detection & tracking (3) advanced analytics	Reduced data	Low (Only the producer or a few scientists may analyze data in the future)	In-situ, interaction, feature display, uncertainty, visual debugging
Exascale Simulations (2) (Science Discovery Oriented (e.g., Climate, Cosmology)	Data Generation from simulations. Data Generation from instruments Integration	(1) Data reduction for post processing (2) time series (3) statistics (4) advanced analytics	(1) Raw data (2) Well organized (DB) (3) Enabled for queries	High (A large number of scientists, geographically distributed)	InfoVis and SciVis, pattern detection, correlation, clustering, ensemble vis, uncertainty
Large instruments (1) (e.g., LHC)	Data generation from large devices Extremely high rates Centralized, coordinated/-controlled access	(1) HW/SW for high-rate data processing (2) derived data (3) metadata (4) Extensive queries	(1) Raw data (2) Different forms of derived data (3) Lots of distributed copies	High (A large number of scientists, geographically distributed), different sets defined by queries and other parameters	Custom user interfaces enabling query visual analysis, trajectory vis/analysis, user driven data triage/-summarization
Instruments (2) (sensors, devices) Examples: field work, biology, observation sensors, internet	Data generation from massive number of distributed devices, sensors, crowd	(1) Local processing and derivations (2) Local analytics (3) Integration of massive data (possibly at an exascale level system, data centers)	(1) Raw data (2) Derived data and subsets (3) Distributed copies	High (A large number of scientists, geographically distributed)	InfoVis, high dimensional vis, large-scale graphs, patterns, clustering, scalability

Table 3.1: Data-generation requirements for different domains

Chapter 4

Data Analysis and Visualization

4.1 Introduction: From Big Data to Big Information

Since the advent of computing, the world has experienced an explosion of data that can be viewed as an "information big bang". Information is being created at an exponential rate such that since 2003, new information is growing at an annual rate that exceeds the information contained in all previously created documents; digital information now makes up more than 90% of all information production, vastly exceeding paper and film. The coming of exascale computing and data acquisition from high-bandwidth experiments across the sciences is creating a phase change. Our ability to produce data is rapidly outstripping our ability to use it. As Herbert Simon, Nobel Laureate in economics, noted: "A wealth of information creates a poverty of attention and a need to allocate it efficiently." With exascale datasets, we will be creating far more data than we can explore in a lifetime with current tools. Yet, exploring these dataset is the essence of the fourth paradigm of scientific discovery.

As such, one of our greatest scientific and engineering challenges will be to effectively understand and make use of this rapidly growing data; to create new theories, techniques, and software that can be used to make new discoveries and advances in science and engineering [10,11]. Data Analysis and Visualization are key technologies for enabling future advances in simulation and data-intensive based science, as well as in several domains beyond the sciences.

4.2 Challenges of in-situ analysis

Science applications are impacted by the widening gap between I/O and computational capacity. I/O costs are rising, so, as simulations increase in spatiotemporal resolution and higher-fidelity physics, the need for analysis and visualization to understand results will become more and more acute. As I/O becomes prohibitively costly, the need to perform analysis and visualization while data is still resident in memory during a simulation, so-called *in-situ* analysis, becomes increasingly necessary. In-situ processing has been successfully deployed over the past two decades in specific instances *e.g.,* see Figure 4.1. However, in-situ analysis is still not a mainstream technique in scientific computing, for multiple reasons:

1. There are software development costs for running in situ. They include costs for instrumenting the simulation and for developing in-situappropriate analysis routines.

2. There can be significant runtime costs associated with running in-situ. Running analysis in-situ will consume memory, FLOPs, and/or network bandwidth, all of which are precious

to the simulation. These costs must be sufficiently low that the majority of supercomputing time is devoted to simulation.

3. At the exascale, resiliency will be a key issue; in-situ analysis software should not create additional failures, and it should be able to perform gracefully when failures occur.

Further, to make in-situ processing a reality, we must fundamentally rethink the overall scientific discovery process when using simulation and determine how best to couple simulation with data analysis. Specifically, we need to answer several key questions and address the corresponding challenges:

- To date, in-situ processing has been used primarily for operations that we know to perform a priori. Will this continue to be the case? Will we be able to engage in exploration-oriented activities that have a user in the loop? If so, will these exploration-oriented activities occur concurrently with the simulation? Or will we do in situ data reduction that will enable subsequent offline exploration? What types of reductions are appropriate (e.g., compression, feature tracking)?

- How do simulation and visualization calculations best share the same processor, memory space, and domain decomposition to exploit data locality? If sharing is not feasible, how do we reduce the data and ship it to processors dedicated to the visualization calculations?

- What fraction of the supercomputer time should be devoted to in situ data processing/visualization? As in-situ visualization becomes a necessity rather than an option, scientists must accept embedded analysis as an integral part of the simulation.

- Which data processing tasks and visualization operations are best performed in-situ? To what extent does the monitoring scenario stay relevant, and how is monitoring effectively coupled with domain knowledge-driven data reduction? If we have little a priori knowledge about what is interesting or important, how should data reduction be done?

- As we store less raw data to disk, what supplemental information (e.g., uncertainty) should be generated in-situ?

- What are the unique requirements of in-situ analysis and visualization algorithms? Some visualization and analysis routines are fundamentally memory-heavy, and some are intrinsically compute-heavy. Thus some are not usable for in-situ processing. We will need to reformulate these calculations. Furthermore, some analysis requires looking at large windows of time. We may need to develop incremental analysis methods to meet this requirement.

- What similarities can be exploited over multiple simulation projects? Can the DMAV community develop a code base that can be re-used across simulations? Can existing commercial and open-source visualization software tools be directly extended to support in-situ visualization at extreme scale?

4.3   Climate exemplar

Climate science is a prominent example of a discipline in which scientific progress is critically dependent on the availability of a reliable infrastructure for managing and accessing of large and heterogeneous quantities of data on a global scale. It is inherently a collaborative and multi-disciplinary
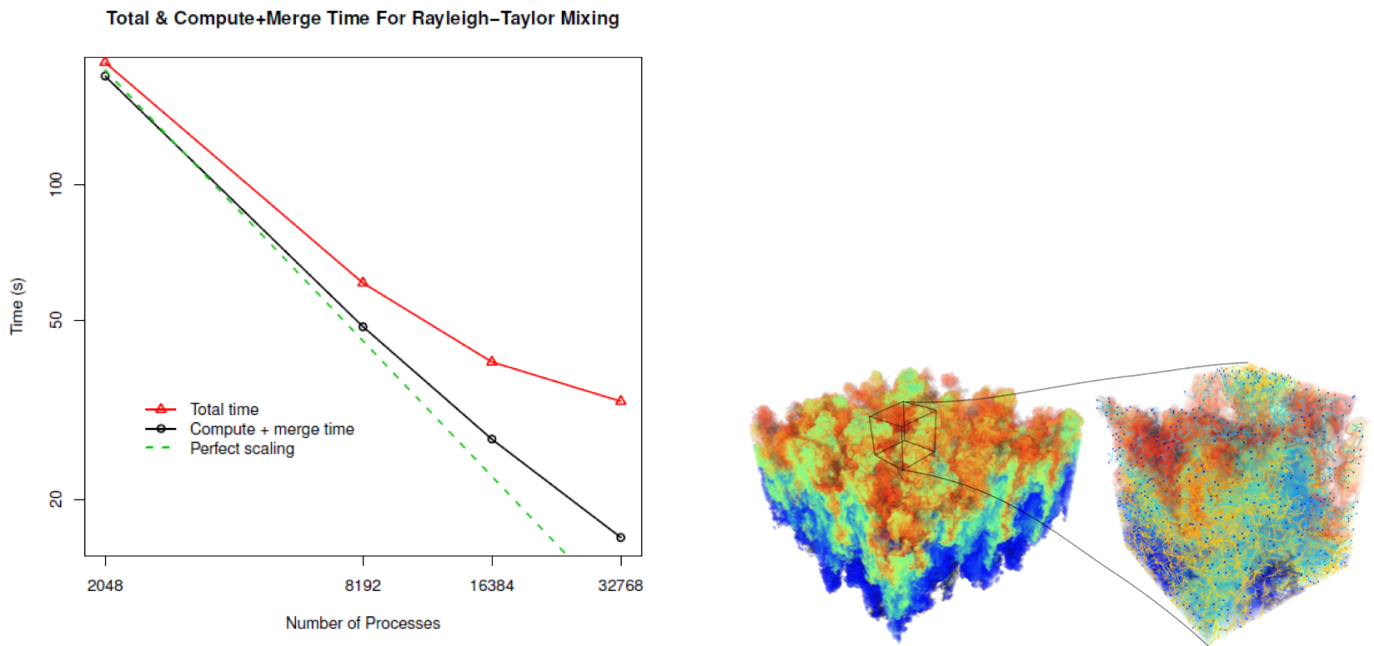
Figure 4.1: In-situ Visualization Scaling Graph (Left) and In-situ Visualization Example (Right). Topological representations can provide insight into complex phenomenon, such as identifying bubbles in a Rayleigh Taylor instability. This figure shows an example of computing the Morse Smale complex, traditionally restricted to small data because the serial nature of existing algorithms, using a newly introduced parallel technique, achieving 40% end to end strong scaling efficiency on 32k Nodes of BG/P.

Figure 4.2: Federated data access in climate workflow

effort that requires sophisticated modeling of the physical processes and exchange mechanisms between multiple Earth realms (atmosphere, land, ocean and sea ice) and comparison and validation of these simulations with observational data from various sources, possibly collected over long periods of time. The climate community has worked for the past decade on concerted, world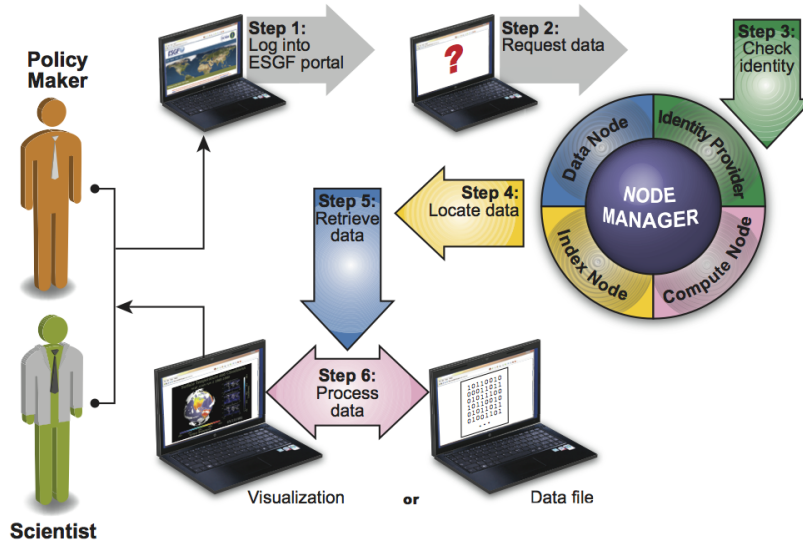wide modeling activities led by the Working Group on Coupled Modeling (WGCM), sponsored by the World Climate Research Program (WCRP) and leading to successive reports by the International Panel on Climate Change (IPCC). The fifth assessment (IPCC-AR5) is currently under way and due out by the end of 2013. These activities involve tens of modeling groups in as many countries, running the same prescribed set of climate change scenarios on the most advanced super-computers and producing several petabytes of output containing hundreds of physical variables spanning tens and hundreds of years. These data sets are held at distributed locations around the globe but must be discovered, downloaded, and analyzed as if they were stored in a single archive, with efficient and reliable access mechanisms that can span political and institutional boundaries (see Figure 4.2).

The same infrastructure must also allow scientists to access and compare observational data sets from multiple sources, including, for example, Earth Observing System (EOS) satellites and Advanced Radiation Measurements (ARM) sites. These observations, often collected and made available in real-time or near real-time, are typically stored in different formats and must be post-processed to be converted to a format that allows easy comparison with models. The need for providing data products on demand, as well as value-added products, adds another dimension to the needed capabilities. Finally, science results must be applied at multiple scales (global, regional, and local) and made available to different communities (scientists, policy makers, instructors, farmers, and industry). Because of its high visibility and direct impact on political decisions that govern human activities, the end-to-end scientific investigation must be completely transparent, collaborative, and reproducible. Scientists must be given the environment and tools for exchanging ideas and verifying results with colleagues in opposite time zones, investigating metadata, tracking provenance, annotating results, and collaborating in developing analysis applications and algorithms.

ANALYSIS	UNDERLYING ALGORITHM	COMMUNICATION PATTERN	ALTERNATIVES QUALITY VS COST TRADEOFF	DATA REDUCTION, IN TRANSIT, MEMORY USAGE, OFFLINE	DATA SIZE INPUT	FREQUENCY, SYNCHRONICITY
Spectra (Offline)	Spatial convolution	Global all-to-all + reduction across integral scale O(1/10) domain	Temporal convolution –no global comm, must be done every dt	Global reduction can be done in transit Can be reduced to small number of sample points, global spectra	Input: Grid size; Output: Grid size, dimension increased by O(nx) points	diagnostics; sample at outer timescale
timescaleScalar field comparison	Pointwise difference	None	n/a	Output can be reduced to statistical description.	Input: 2 fields Output: 1 field	diagnostics; accumulate over outer timescale, sample at inner timescale
Chemical analysis (CEMA)	Pointwise.	Access entire state (pointwise)	n/a	Possibly.	Input: Multiple fields Output: Vector fields.	diagnostics and analysis; sample at outer timescale, async
Conditional moment – single pass	Weighted (conditional)summation.	Global allreduce.	n/a	In situ accumulation, aggregation could be done in transit.	Input: Multiple fields; Output: N~O(1) dimensional field with O(50-100) points in each dimension.	diagnostics &analysis; accumulate over outer timescale, sample at inner timescale, async
Statistical dimensionality reduction (joint pdfs)	Conditional summation; pointwise + aggregation.	Global allreduce.	n/a	In situ accumulation, aggregation could be done in transit.	Input: Multiple fields;Output: N~O(1) dimensional field with O(50-100) points in each dimension.	diagnostics and analysis; accumulate over outer timescale. Sample at inner timescale. Asynchronous.
Shape analysis	Eigenvalue decomposition	requires results of feature extraction. when done intransit each feature can be analyzed independently in parallel	n/a	Can be done intransit after completion of feature extraction algorithm	Input and output sizes dependent on number of features O(f)	analysis, asynchronous, (inner timescale)*(feature size)/(grid size)
Feature tracking	Pointwise comparisons	requires results of feature extraction. when done in transit each pair of timesteps can be computed independently in parallel	n/a	Can be done intransit after completion of feature extraction algorithm	Input and output sizes dependent on number of features O(f)	For analysis, asynchronous, (inner timescale)*(feature size)/(grid size)
Level set features, ie. Merge trees /contour trees	(Multiple) Global Union-find with history	Global gather followed by global scatter with most nodes potentially idle at some point	n/a	Depending of feature of interest significant data reduction after data parallel computation. Potential for in-situ – in-transit split	Input and output sizes dependent on number of features O(f) and their spatial extent	Timescale of phenomena of interest, meaning dependent of the ratio of feature size vs. expected speed
Multivariate volume and particle rendering	Ray casting, point sprites and image compositing	Global reduce	n/a	Rendering is mainly done in-situ, can be done in-transit after data reduction. In-situ or in-transit image compositing	Input: Multiple fields Output: 2D images	For diagnostics and analysis. Asynchronous.
Lagrangian particle querying and analysis	Range query	Global gather and/or global reduce	n/a	Querying is mainly done in-situ depending on particle number. In-situ or in-transit analysis	Input: entire particle data Output: query specific	For diagnostics and analysis, accumulate over outer timescale of simulation. Sample at inner timescale. Asynchronous
Distance field	Level set; pointwise	Global all-gather followed by global all-to-all	n/a	Can be done in-transit after feature extraction	Input and output sizes depend on specific features defining level set	For analysis. Timescale of the phenomena of interest. Asynchronous
Spectra (In situ)	Spatial convolution	Global all-to-all + reduction across integral scale O(1/10) domain	Temporal convolution –no global comm, but must be every dt	In transit global reduction; Can be reduced to small number of sample points, global spectra	Input: Grid size, Output: Grid size, dim increased by O(nx) points.	For diagnostics; sample at outer timescale of simulation.
Filtering (in place)	Spatial convolution	Global all to all	Necessary for some simulation algorithms. Truncated filter.	No	Input: 1 fields Output: 1 field	diagnostics (sample at outer timescale, async), analysis (resolve many timescales, async), and test subgrid models in situ every substep, sync
Filtering (with decimation)	Spatial convolution	Global all to all.	Truncated filter; all to few	Aggregation in transit	Input: 2 fields; Output: Decimated field	diagnostics (sample at outer timescale),analysis (resolve many timescales), async
Temporal filtering	Temporal convolution	Need to buffer moving window of filter size.	Truncated filter; limit window size, do for subset of domain.	In situ accumulation, aggregatioin in transit. Could be decimated in time.	Input: Time series; Output: Time series	diagnostics (accum over outer timescale), analysis (resolve inner timescales), at subset of spatial locations (outer spatial scale) or along feature trajectory. Asynch.
Conditional moments - multipass	Weighted (conditional)summation.	Buffering first pass; aggregate in place; global allreduce.	Form single pass alternative.	In situ accumulation, aggregation could be done in transit.	Input: Multiple fields across multiple timesm Output: N~O(1) dimensional field with O(50-100) points/dim	diagnostics and analysis; accumulate over outer timescale. Sample at inner timescale. Synchronous +async 2nd pass
Gradient features, ie Morse / Morse-Smale complex	Global breadth-first traversal	Global gather	On-the-fly simplification &/or pre-filtering to reduce data. Limit feature size to reduce comm	Data reduction through pre-filtering. Potential to store partial or sub-complex	Unless input is pre-filtered need all data. Output dependent on feature density and simplification level	Timescale of the phenomena of interest, meaning dependent of the ration of feature size vs. expected speed

Table 4.1: Categorization of representative combustion analytics where green, yellow and red denote increasing degree of difficulty in implementation at the exascale, respectively.

4.4 Combustion exemplar

4.4.1 In-situ analytics for exascale combustion use case

The exascale simulation summarized in Section 3.3 itself is only one component of the overall combustion workflow; the resulting simulation data must also be analyzed. The types of analysis are dictated by the characteristics of the combustion and turbulence and ranges from topological feature segmentation and tracking, level set flame distance function and normal analysis, statistical analysis including moments and spectra, and visualization of scalar, vector and particle fields. The range of analysis are presented in Table 4.1. These representative analysis are being characterized (loads, stores, memory movement) and studied in light of opportunities for new programming languages that would expose the semantics to enable automatically exploring scheduling placement, i.e. that would optimize among storage, communication, and computation on different exascale resources. Current combustion practice at the petascale is to write data to persistent disk and then read data later for analysis.

This approach won't scale to exascale due to the widening gap between computational power and machine memory compared to the available I/O bandwidth. If current projections hold true, the classical paradigm of simulation followed by post-processing will become infeasible. At exascale, simulation codes will not be able to write a sufficient number of time steps to permanent storage to ensure a reliable analysis. Therefore, we anticipate that much of the current analysis and visualization will need to be integrated with the simulation codes into a comprehensive combustion workflow. However, data analysis algorithms typically have characteristics that are very different

from typical simulation codes. For example, they may be branch heavy with comparatively few floating point operations; many algorithms are I/O bound and/or heavily dependent on the data layout of their respective input data; and the performance of most algorithms is strongly data dependent (e.g. depends on the distribution of features in the data). By contrast, combustion solvers are typically memory bandwidth bound. As a result, many analytics algorithms are difficult to scale even to current architectures let alone projected exascale systems. Fortunately, analysis tasks in general tend to be computationally less expensive than the simulation itself, making integrated analysis still a viable option.

To understand the data analysis needs and potential trade-offs at exascale there are three areas that need attention: first, a wide range of potential execution models and the necessary APIs to explore them need to be defined; second, proxy or mini-applications for a variety of representative data analysis algorithms are needed to characterize fundamental, machine-independent behaviors such as memory accesses and/or data movements; and third, these algorithms need to be integrated into a comprehensive combustion workflow to study the interactions between the different components and evaluate design-space trade-offs inherent in the system.

4.4.2 Execution models

As mentioned above the two primary characteristics of many data analysis algorithms are that they are computationally inexpensive (compared to the simulation) but data intensive. The former provides significant freedom in how algorithms can be implemented while the latter puts strong restrictions on where they can be executed given the expense of moving data.

Traditionally, a simulation will use all available cores on its given partition of a machine and the most straight forward integration of an analysis algorithm is to do the same: as the simulation advances it directly calls (an) analysis routine(s), waits for it to finish, and proceeds. However, assuming that any given analytics algorithm may not need or cannot reasonably utilize all cores on an exascale machine, this introduces a number of viable alternative execution models (as indicated earlier in Figure 3.3). For example, the analysis could use only some of the cores on a given node, or only some nodes of the partition, or even nodes on a different machine. Each approach has advantages and disadvantages trading off between, for example, code scalability, wall-clock time, impact on the simulation, and required data movement. A closely connected choice is how the analysis accesses the simulation data. For example, one may directly access the simulation data structures or agree on an in-memory hand-off requiring a copy.

Alternatively, there may exist sufficient amounts of non-volatile memory on-node to store medium term copies of the simulation state or send parts of the state to adjacent nodes. Finally, one can process the data either synchronously or asynchronously. Combining these choices results in large design space that can be adapted to different hardware configurations but can also be used to inform hardware. Using representative data analysis algorithms described in Table 4.1, as well as the exascale use cases, it is possible to explore this space of potential execution models.

For example, a large amount of non-volatile memory coupled with an asynchronous computation on a single core per-node is an attractive model, especially for slowly changing quantities of interest. However, a background analysis process may interfere with the carefully tuned caching behavior of the simulation. Similarly, a data-parallel filtering or compression may allow sending all data required for analysis to dedicated analysis nodes or even an external analysis cluster, yet may cause unexpected congestion on the network.

Some of these design trade-offs are being explored with analytics proxy applications: RTC, SSA, and PVR operating in-situ with the petascale combustion solver, S3D. For example, it was shown that by dividing each task into in-situ and in transit components, the frequency of analysis
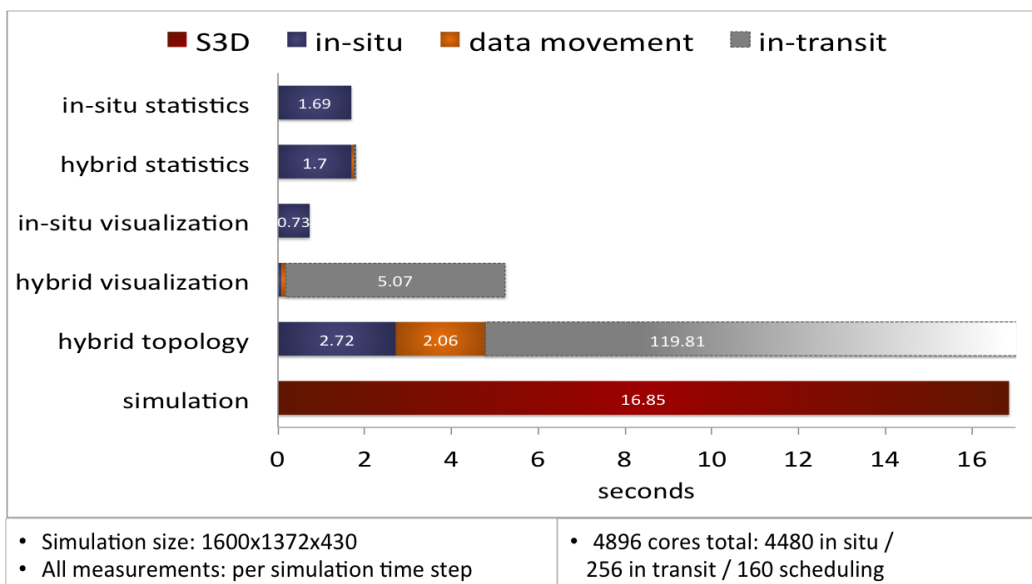
Figure 4.3: The timing breakdown for in situ, in transit, and data movement for the simulation and the various analytics algorithms using 4896 cores on Jaguar, the Cray XK6 at Oak Ridge National Laboratory's National Center for Computational Sciences. 4480 cores were used for the simulation and in-situ processing, 256 cores were used for in transit processing and 160 cores were used for task scheduling and data movement. The simulation grid si ze was 1600x1372x430 and all measurements are per simulation time step.

was increased without negatively impacting the performance of the solver on current machines [2]. In this approach, secondary compute resources were used to hide the less scalable global reduction portions of analysis tasks. Figure 4.3 shows the experimental setup as well as timings for the various stages for each analysis algorithm.

4.4.3 Proxy Applications for Analytics/Visualization

To fully quantify expected workloads and enable exploration of design space options for a combustion workflow, the behavior of representative analytics tasks needs to be characterized. There are many analytics tasks commonly performed by combustion scientists including the following, all of which exhibit a range of characteristic algorithmic behaviors:

1. volume rendering of scalar, vector, and particle data (see Figure 4.4);

2. Lagrangian particle query and analysis;

3. level set distance function and normal analysis;

4. topology driven feature segmentation and characterization based on merge trees, Morse Smale complex (see Figure 4.5);

5. feature tracking over time (see Figure 4.6);

6. shape analysis;
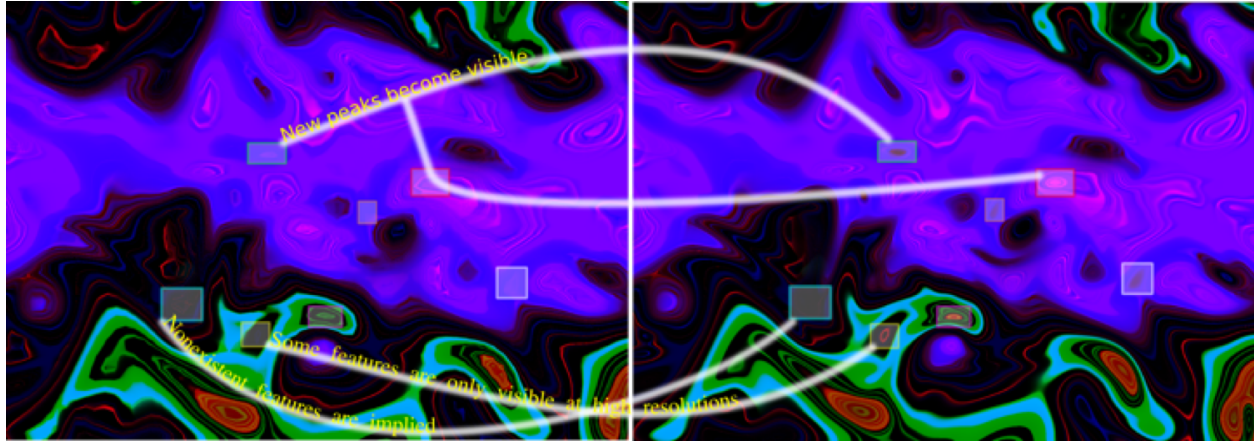
7. dimensionality reduction;

Figure 4.4: The need for high resolution when exploring small-scale features. The figure on the right shows a volume rendering of one time step of a large-scale combustion simulation while the figure on the left has been down-sampled. New peaks become visible only in the full, high resolution visualization and some features that are implied in the down-sampled visualization are shown to be artifacts of the down-sampling and do not appear in the high resolution visualization.

8. conditional moments of various orders;

9. spectra and filtering algorithms with spatial convolutions.

Some of the spectra and filtering operations involve spatial convolutions that imply global communications patterns that are expected to be very expensive and will be highly influenced by the network infrastructure. Feature extraction based on the full Morse-Smale complex computation, involves unevenly distributed, data dependent computations that are infeasible in postprocessing and highly unbalanced when executed in-situ. As discussed below, even reduced topology computations will involve hierarchical communication patterns that can take advantage of a network layer with low latency for small messages (e.g. achievable with on chip network interface). Feature tracking that resolves properly the fine time scales of intermittent phenomena seems feasible only in situ, but requires storing additional information that may go against the tight limits imposed by the exascale architectures and would advocate for extra memory per node (even if slower). Shape analysis involves eigenvalue decomposition for regions of interests whose distribution is data dependent and can only be determined after completing a full topological segmentation, which indicates that it may scale to exascale if properly coupled with the topology code. Chemical explosive mode analysis (CEMA), based on an eigenvalue analysis of reaction rate Jacobians, on the other hand, can be performed as a point-wise computation and is expected to scale easily, although it may still involve challenges.

4.5   Biology exemplar

Data visualization, analysis, and management play a key role in addressing the grand challenges in the Biology domain summarized in Section 3.4. In all of the key biological research areas addressed, there is an important intersection between simulation data and experimental data. Thus analysis and visualization tasks oriented around large data integration and verification are prominent in the requirements. In fact, one of the key concerns highlighted in the report has to do with innovations
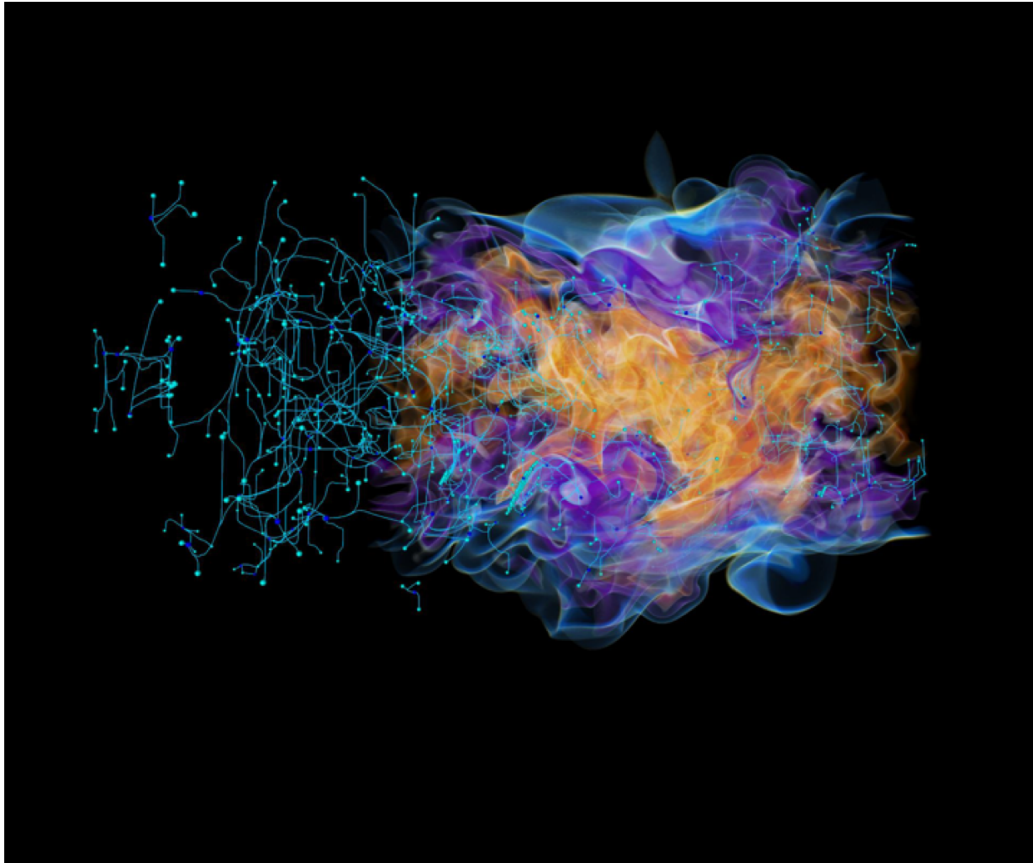
Figure 4.5: A visualization of a topological analysis and volume visualization of one time step in a large-scale combustion simulation. The topological analysis identifies important physical features (ignition and extinction events) within the simulation while the volume rendering allows viewing the features within the spatial context of the combustion simulation.
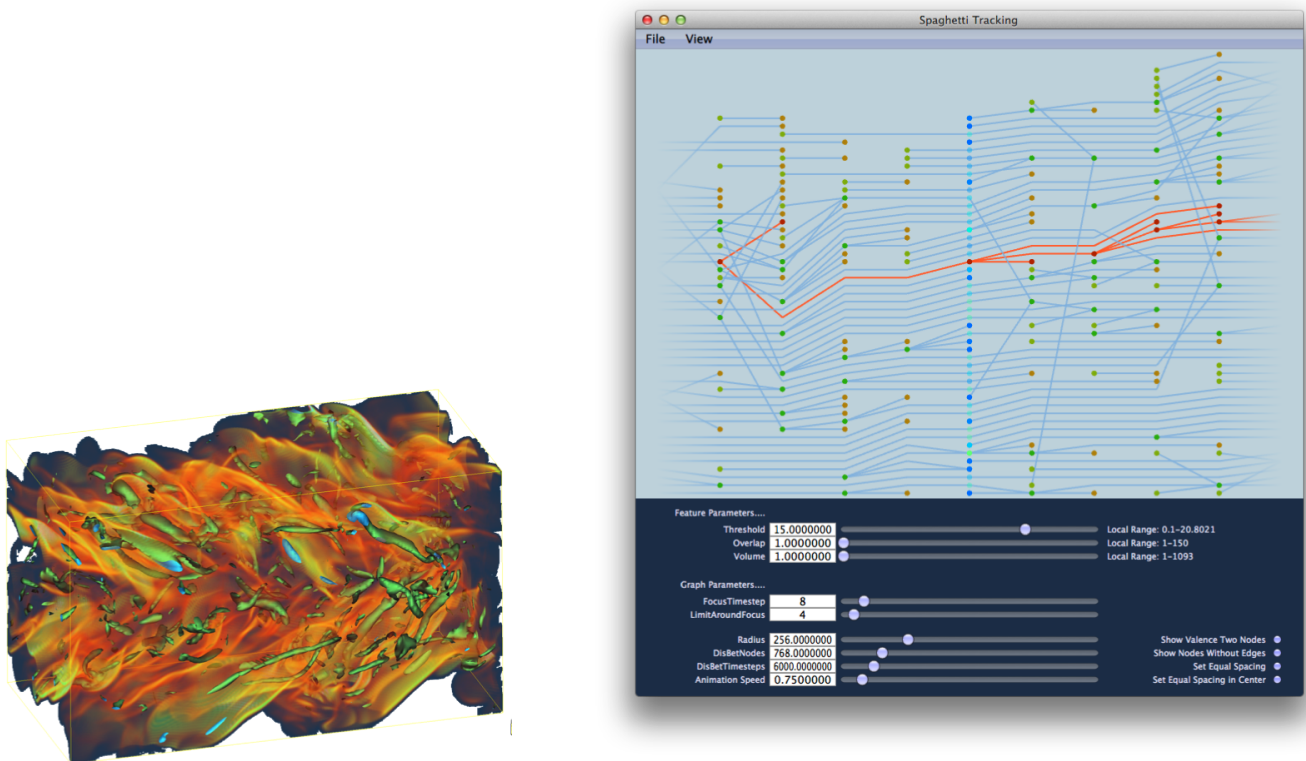
Figure 4.6: The visualization on the left shows features for a single time step within a large-scale combustion simulation while the figure on the right shows a graph that tracks features over time. Combustion scientists can follow specific events over time.
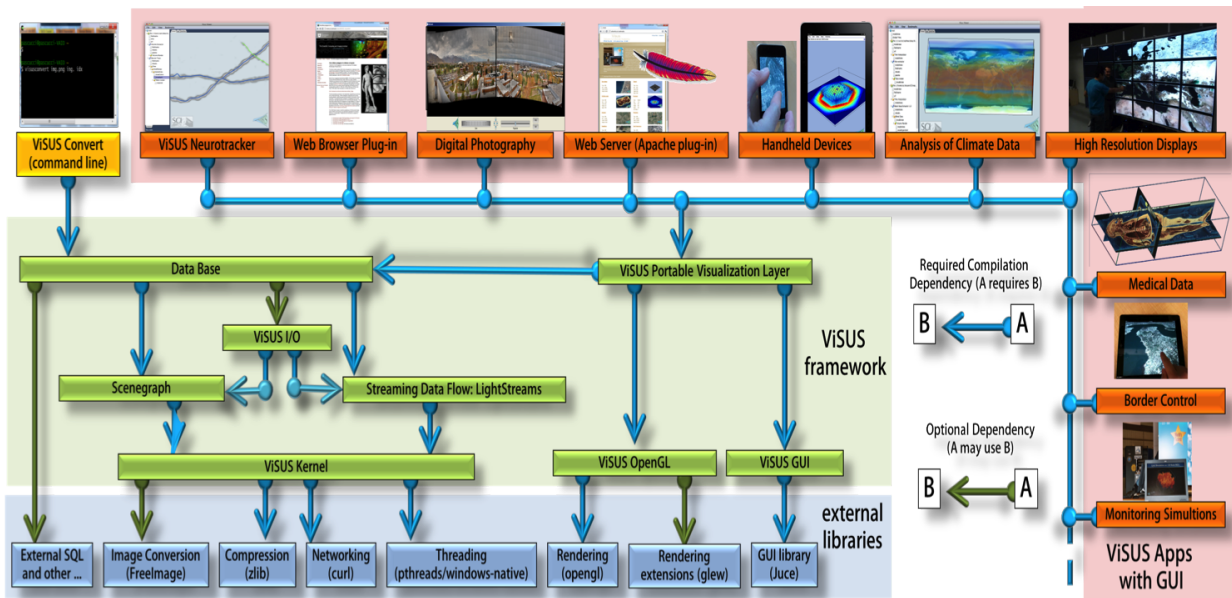
Figure 4.7: An example visual data analysis application framework called ViSUS. ViSUS is a scalable data analysis and visualization framework for processing large scale scientific data with high performance selective queries. It combines an efficient data model with progressive streaming techniques to allow interactive processing rates on a variety of computing devices ranging from handheld devices like an iPhone, to simple workstations, to the I/O of parallel supercomputers. Arrows denote external and internal dependencies of the main software components. Additionally we show the relationship with several applications that have been success-fully developed using this framework.
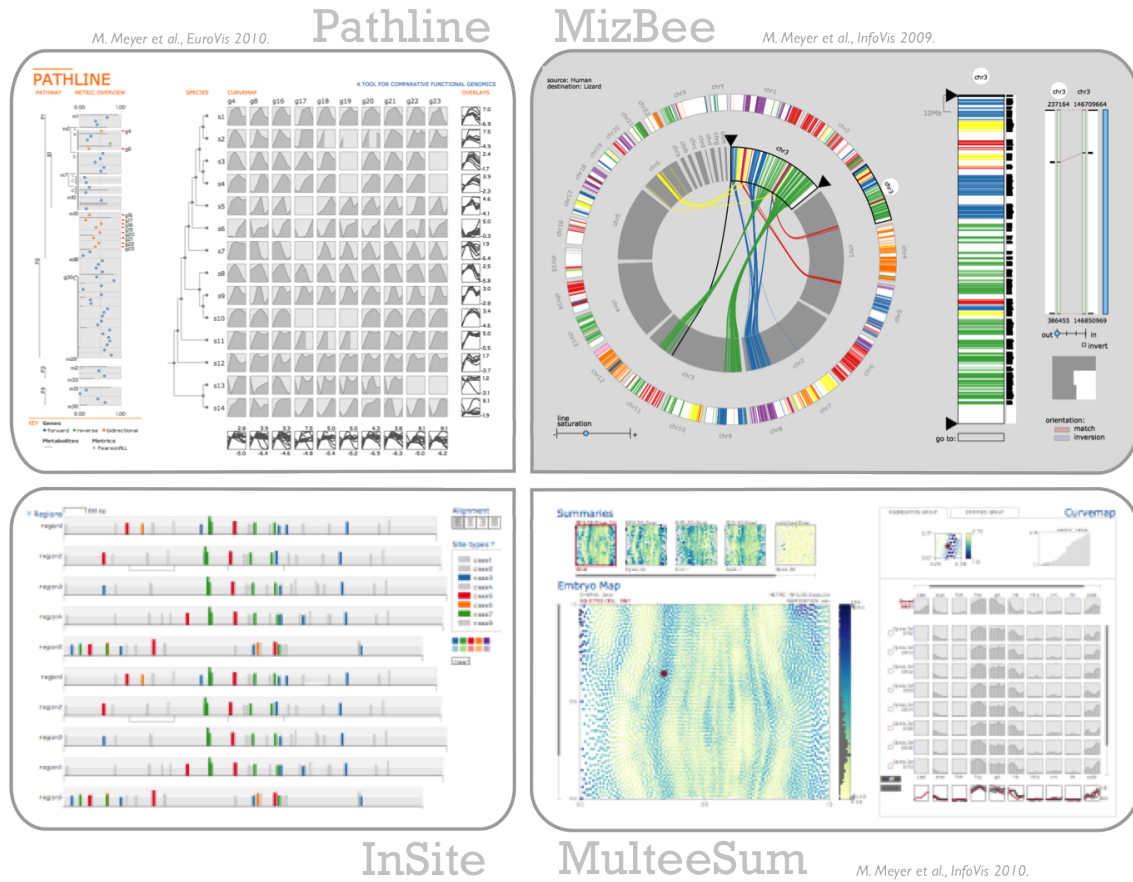
Figure 4.8: Biological visualization examples

in high-throughput sequencing equipment; it is not concerned solely with large simulation data volumes. From the perspective of the report, the extreme-scale computing challenges in biology will depend heavily on the ability to adequately support data management, analysis, and visualization.

Figure 4.8 contains some examples of biological visualization. The rise of the fields of computational biology and bioinformatics has brought significant advances in algorithms for processing biological datasets. However, deciphering raw data and computational results through visual representations is too often done as the final step in a complex research process, with tools that are rarely specific to the task. Hence, there is a significant opportunity to enhance biological data analysis through a thoughtful and principled investigation of visualization.These four tools are examples of custom, interactive visualization tools designed in collaborations with biologists – they have all been deployed in biological research labs and led to a variety of scientific insights.

Figure 4.9: Accelerating Data Acquisition Reduction and Analysis (ADARA) system architecture

4.6   Data streaming and near-sensor computing

4.6.1   ADARADA data streaming project
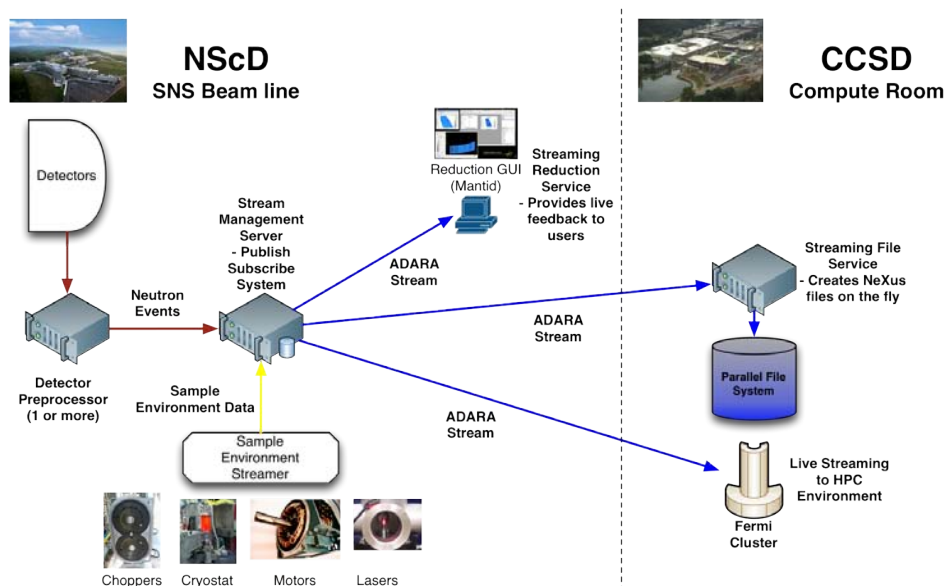
Oak Ridge National Laboratorys Spallation Neutron Source (SNS) is the worlds most intense pulsed, accelerator-based neutron source for scientific research and development[1]. From a high level perspective, the process of a neutron scattering experiment involves 3 stages, (i) data acquisition, (ii) data reduction and (iii) data analysis. The raw experimental data can be very large (10s of Gigabytes to Terabytes) is in instrument units (detector pixels and time of flight etc.) and must be converted to physical units, e.g. energy and wavevector (momentum) etc., and also corrected for instrumental/measurement artifacts. This data reduction process usually results in significantly smaller data sets that are feasible for data analysis. As an example the NOMAD beam line at SNS can produce raw data files that are Terabytes in size, but in many cases these reduce to an analysis data file that is only a 1-D spectrum of a few 1000 values, *i.e.,* order of kilobytes. The final stage, data analysis, then involves fitting, or modeling, the reduced files that are in physical units against atomistic models of the materials.

Traditionally, reduction and visualization of some of the large SNS data sets take hours after the data has been collected. This impedes the scientist's workflow, where the analysis of the data often strongly influences the next step in the experimental investigation. The Accelerating Data Acquisition Reduction and Analysis (ADARA) project (Figure 4.9) has developed a streaming data system for real-time feedback from experiments and a software-as-a-service HPC infrastructure to enable efficient reduction and analysis of the data generated from these experiments. The streaming data infrastructure provides in-situ reduction of the data as it is generated from the instrument, allowing users to visualize their data in energy/momentum space as the experiment is underway.

Many of the technological advancements required in exascale computing are needed for the productive use of the data generated by the SNS. These advances span system architecture to advances

---

[1]The material in this section was obtained courtesy of Galen Shipman at ORNL.

in simulation and data analysis/visualization software. Explanatory and validated materials science simulation software optimized for time-to-solution is required in order to provide timely feedback during experiment. These improvements are non-trivial, requiring strong-scaling codes and a corresponding scalable system architecture capable of providing time-to-solution improvements of up to $1000\times$. Advances in in-situ data processing, particularly in streaming data processing will require lightweight, composable data analysis software optimized for use on next-generation systems.

4.6.2 Near-sensor computing and in-situ analysis

The previous sections have established the requirement for performing data reduction and analysis close to the data source in many science domains. This form of near-sensor computing has been under way in High Energy Physics for many years and will become increasingly important for DOE light sources such as ALS and LCLS. The key features of near-sensor computing are as follows[2]:

1. Triggers to instantly detect when something interesting is occurring, so that information recording can begin. This requires very fast sensing and detection mechanisms – often utilizing FPGAs or custom logic embedded into the sensor due to the extremely low latencies required for the trigger.

2. Filters are applied upon the data to reduce its size using analysis, selection, compression, or other forms of data summarization. This is performed as close to the instrument as possible so as to minimize the cost of downstream data transport and data storage. These mechanisms are preferably lossless in the sense that they are targeted at reducing or removing information that would not need to be revisited later. This often involves streaming processing using FPGAs, GPUs, and/or clusters, all of which will take the form of extreme scale systems in the future. Examples can be found in radio telescopes, X-ray sources, adaptive optics for telescopes and other high-throughput sensor systems.

3. After the data has been reduced by the triggers and filters, it is curated and archived for re-processing and re-analysis by end users. This minimum subset of data must be preserved because the cost of repeating the experiment far exceeds the cost of maintaining the data in archive. Curation and data provenance controls are essential for this archived data.

There is an interesting correspondence between near-sensor computing and what is needed for simulation data. The data production rates of exascale simulations will be severely limited by I/O rates. Our current paradigm for data storage is to store all simulation data, and decided later what is needed. However, the increased cost of purchasing sufficient back-end disk resources will make this approach untenable in the future. As discussed earlier in this chapter, the solution to this challenge lies with in-situ analysis, an approach in which data summarization and analysis is placed as close as possible to the data production source. Thus, similar approaches and patterns can be used for both near-sensor computing and in-situ analysis. For example, certain events of interest in the simulation can serve as triggers, and filters can be applied to simulation data to summarize regions of interest. Finally, in an increasing number of domains (*e.g.,* climate simulation), it is impractical to re-execute millions of CPU-hours of calculations invested in a specific simulation. So, the data produced by the simulation must be curated and stored (as is done for experimental data) so that it can be re-analyzed multiple times by the scientific community that generated it.

This intersection in the needs of reducing and analyzing experimental vs. simulation data is another example of the synergies between data-driven science and exascale computing.

---

[2]The material in this section was obtained courtesy of John Shalf at LBNL.

Chapter 5

Synergies with Exascale Computing

5.1 Challenges and opportunities of exascale computing

For the past several years, the high performance computing research community has been focused on the challenges associated with creating exascale computers – those with 10^18 operations per second, about 100 times faster than today's fastest machines. Machines with this much computational power would be of enormous value to a wide variety of scientific, commercial and national security applications. But the path to exascale will be difficult and will require significant advances in a variety of technologies. It is uncertain whether exascale is achievable without disruptive changes in the way we build and use computers. Many of these challenges are detailed in the DARPA Exascale report [6].

Progress towards exascale will play out in the context of the broader computing ecosystem. The high performance computing business is too small to drive the agenda of large processor, memory, or computer systems vendors. Instead, as has been true for two decades, the HPC community will need to leverage technologies developed for larger market segments. A key challenge for the HPC community will be to optimize the use of comparatively small investments to tweak these technologies to meet HPC needs.

Currently a major driver in the larger computing world is the explosive growth of data-centric applications. These include a vast array of business analytics and web services, as well as the kinds of scientific applications detailed elsewhere in this report. Popular computational approaches include large shared-memory databases, computing clouds, and map-reduce approaches to analytics. The giants of the internet economy have set up enormous server farms with aggregate computational capability that used to be found only at leading HPC sites. IBM and Cray, traditional leaders in high end computing for scientific applications, have both asserted that the data-computing market is their principal focus.

One implication of the conclusions of [6] is that path to exascale will likely require technology disruptions. With most of the computing community focusing on data-centric opportunities, the technologies advances available to HPC will probably emerge from data-computing efforts. To the extent possible, it will be advantageous for the exascale community to avail itself of the momentum behind data-computing. When appropriate, exascale architects should use technologies supported by the data-computing community as these will likely be advancing rapidly. In addition, where the exascale community needs new technologies, these will be easiest to realize if they also benefit data applications and so have a broader commercial viability.

In this section, we review some of the exascale challenges that are aligned with the needs of data-centric computing. The examples below are meant to be illustrative and by-no-means

comprehensive.

5.2   The power challenge

Probably the greatest impediment to building and operating an exascale computer is the power it will consume. The Exascale report [6] concludes that an exascale machine built out of extrapolations of current technologies would consume several hundred megawatts. The report offers no path towards resolution of this challenge and acknowledges that " The energy and power challenge is the most pervasive of the four [exascale challenges], and has its roots in the inability of the group to project any combination of currently mature technologies that will deliver sufficiently powerful systems in any class at the desired power levels. "

Intriguingly, the report also concludes that the power consumed by moving data will be significantly larger than the power for performing computations. " A key observation of the study is that it may be easier to solve the power problem associated with base computation than it will be to reduce the problem of transporting data from one site to another ?- on the same chip, between closely coupled chips in a common package, or between different racks on opposite sides of a large machine room

Power is also an issue of growing concern to the data-computing community. Large data centers are generally built in locations where inexpensive electricity and cooling are available. Furthermore, data-centric computing (at any scale) tends to be less computationally intensive than scientific computing, and so the power required for data movement will dominate overall power consumption.

So both exascale computing and data-centric computing have strong interest in technologies that can reduce power consumption – particularly for data movement. Several emerging technologies have the potential to help with this problem and so provide opportunities for cross-community leverage. One illustrative example is stacked memory with through-silicon vias which allow more data to be kept closer to processors thereby reducing aggregate data movement. But the potential power advantages of stacked memory go well beyond this. The Hybrid Memory Cube (HMC) Consortium /citeHMMC is developing technology in which the stacked memory is combined with processing capability. By doing some computation very close to the memory itself, data transfers to the CPU can be avoided altogether. This can significantly reduce power consumption while simultaneously boosting performance. The HMC Consortium reports that, relative to DDR3 memory technology, the HMC can improve performance by a factor of 15 while reducing power by 70/

5.3   Interconnect and bandwidth

Data movement between chips is essential to both exascale computing and large-scale data-centric computing. Both communities have an interest in technologies that improve performance and the power efficiency of data transfers. Off-chip communication is a primary scalability limiter for scientific computing. Similarly, given the large amount of communication involved in typical map-reduce computations, it is likely that off-chip communication is the rate limiter for many large-scale data analysis jobs as well. In addition to limiting performance, off-chip communication also consumes considerable power. With current technologies, the power required to move a bit off-chip is nearly two orders of magnitude larger than that required to move a bit on-chip.

Silicon photonics is a maturing technology that has the potential to improve performance and reduce power consumption for both the scientific and data computing communities. Photons are currently much more power efficient than electrons for moderate and long distance communication

and are widely used when communicating over distances larger than several meters, e.g. rack-to-rack communications. The advantages of optics include higher wiring density and a power cost that does not grow with distance. But the power required to convert between electrons and photons has limited the advantages of optical communication for shorter links (e.g. within a rack, board or chip). Approaches to address these limitations are active areas of research in both components and systems, and promising prototypes have been demonstrated in laboratory settings. Progress in accelerating the maturation of optical networking would be beneficial to multiple computing communities.

5.4   Storage and data management hierarchies

Typical scientific computing application involve a small amount of data input, and a potentially large amount of data output. Data-centric computing usually displays the opposite behavior. Large amounts of data (e.g. from an experimental facility) are read into a computer and then reduced or analyzed to generate a small set of output. Both the scientific and data-centric computing communities will benefit from advances in storage and data management hierarchies, although they are likely to exploit them in different ways. The data computing community needs to provide access to data coming into a processing unit, while the scientific computing community needs to manage data exiting (e.g. via burst buffers).

An important emerging technology in storage is solid state memory. NAND FLASH is a non-volatile, solid-state memory that is being used to provide a higher performance alternative to disk in some settings. It supports random access and low latency memory operations. However, it remains more expensive than disk on a per-bit basis, and so seems most useful as an intermediate level in the storage hierarchy.

An important shortcoming of current NAND FLASH technology is its limited endurance. The technology degrades with use and can only sustain 10K to 100K read-write cycles. A promising alternative is phase-change memory which is expected to provide much greater endurance. The emergence of solid-state memory provides new opportunities for constructing storage hierarchies, to the benefit of both the scientific and the data-centric computing communities.

Chapter 6

Cross-Cutting Issues

Rapid advances in experimental, sensor, computational technologies, and techniques are driving exponential growth in the volume, acquisition rate, variety, and complexity of scientific data. This new wealth of scientifically meaningful data has tremendous potential for scientific discovery. However, to achieve scientific breakthroughs, these data must be exploitablethey must be analyzed effectively and efficiently, and the results shared and communicated easily with the wider community. The explosion in data complexity and scale makes these tasks exceedingly difficult to achieveparticularly given that an increasing number of disciplines are working across techniques, integrating simulation and experimental or observational results. Consequently we need new approaches to software data management, analysis, and visualization that provide research teams with easy-to-use, end-to-end solutions. These solutions must facilitate (and where feasible, automate) every stage in the data lifecycle, from collection to management, annotation, sharing, discovery, analysis, and visualization. Hereby, core functionalities required are the same between different science communities but require customization to adapt to gaps in their specific needs and fit into their research and analysis workflows.

6.1 Data sharing and lifecycle management

6.1.1 Data Retention: what to store and what to discard

Already, in data-intensive science, the financial and technical challenges of data storage and access can match, or even dwarf the challenges of computation. The LHC with its petabyte per second data rate is currently the outstanding example. All but 0.001% of the data must be examined and discarded in real time, and the discarding of data, especially derived and simulated data, continues throughout the years of analysis. The data-retention challenge divides into two closely related challenges, deciding what to keep (data retention) and keeping it in a viable form (data preservation). We first discuss data retention considerations for different classes of fata.

Unique data: Examples include observations of cosmic, or earth-systems happenings, such as supernovae or climate sensor data. The universe, or the earths climate, are not readily reproducible systems and these data, once obtained are normally retained indefinitely.

Hard to reproduce data: Data from frontier particle physics experiments are (statistically) reproducible in principle, but, somewhat like the moon explorations of the 1970s, most such experiments are not expected to be reproduced for the foreseeable future. Occasionally, old particle physics data are reduced to negligible value by repeating the experiment in factory mode with many orders of magnitude higher statistics and better detectors. The hard-to-reproduce data must be retained for the foreseeable future.

Reproducible data: Reproducible experimental data includes much of the data captured at light and neutron sources. While reproducing these data may not be trivial, at any instant in time it is possible to assess the cost of reproducing the data and compare it with the cost of retention for future use. For many decades, and probably for several more, the decreasing unit cost of storage has meant that any experimental and observational data retained for a few years might just as well be retained indefinitely, provided the other challenges of data preservation can be addressed.

Simulated and derived data: Data that are a result of simulation, or are derived from retained experimental or observational data, can all be recreated in principle. The retain/discard decision is (or should be) an economic one, sometimes distorted by knowing that although the probability of needing to recreate discarded data would be small, obtaining the funds to do so would be a major distraction. In practice these decisions are hard to address algorithmically and it is often convenient to make a large, but easily swamped, pool of storage available to a group of scientists and let them manage the retain/delete decisions as best they can.

The imperative to retain "just in case" gets escalated to panic level if inadequate recording of the provenance of derived and simulated data makes their reproduction close to impossible. The ability to make good human or automated decisions on retention requires the rigorous recording of provenance, leading to the technical ability to reproduce anything if resources are made available. This is a mature concept that has, nevertheless, been only patchily implemented by (or for) the sciences that need it now, or will need it soon.

6.1.2 Data Preservation

Experimental or observational data that are impossible or impossibly costly to reproduce should be retained indefinitely. However, keeping the bits intact loses all value if nobody knows what they mean. Gallileos notebooks are readily intelligible 400 years after he made his experimental measurements. The bits emerging from modern high-rate data acquisition systems are unintelligible in themselves and may require the combined knowledge of tens of scientist for correct interpretation. This is the challenge of data preservation, it requires the retention of the data, the metadata/provenance, and the recipes, often encoded into hundreds of thousands of lines of partially commented software, used to extract information and knowledge from the bits. Often non-algorithmic human processes, such as internal peer review by colleagues, are a vital part of the production of knowledge.

Repeated experience has shown that, where this preservation can be achieved, the old bits can be mined for new, even unexpected information and knowledge, but the experience has also highlighted the many daunting difficulties. Those with experience conclude that the preservation of original experimental and observational data requires planning from the outset and will consume resources that are significant on the scale of the gathering and analysis of the original data.

6.2 Software Challenges

Both exascale computing and data-driven science will require new approaches to software. ASCR's recent X-stack and OS/R research programs are focused on addressing the software challenges of exascale computing, including concurrency, energy efficiency, and resilience. Many of these challenges were described in [15], and we include a brief summary here.

There are several reasons for paying attention to software in the development of extreme scale systems. First, the exascale systems of 2022 will be dramatically different from todays systems and will require correspondingly fundamental changes in the execution model and structure of system software. Second, while there has been significant innovation at the hardware and system level for todays systems, previous approaches (prior to the X-stack and OS/R programs) have not paid

much attention to the co-design of multiple levels in the system software stack (operating system, runtime, compiler, libraries, application frameworks) that is needed for exascale systems. Third, while certain execution models such as Map- Reduce in cloud computing and CUDA in GPGPU data parallelism have demonstrated large degrees of concurrency, they haven not demonstrated the ability to deliver a thousand-fold increase in parallelism to a single job with the energy efficiency and strong scaling fraction necessary for extreme scale systems. Finally, applications can only be enabled for exploiting extreme scale hardware by exploring a range of strong scaling and innovative weak scaling techniques, but only with attention to efficient parallelism and data movement. Operating system-related challenges include scalability, spatial partitioning, direct hardware access for inter-processor communication, and asynchronous rather than interrupt-driven events. There are additional challenges in runtime systems for scheduling, memory management, communication, performance monitoring, power management, and resiliency, all of which will be built atop future extreme scale operating systems.

Earlier in Section 4.2, we referred to software challenges for in-situ analysis. These challenges arise from two kinds of costs: (1) the costs to couple simulation and analysis routines and (2) the costs to make analysis routines work at extremely high levels of concurrency. These areas are discussed in more depth in the following paragraphs. It takes considerable effort to couple the parallel simulation code with the analysis code. There are two primary approaches for obtaining the analysis code: writing custom code or using a general purpose package. Writing custom code, of course, entails software development, often complex code that must work at high levels of concurrency. Often, using a general purpose package is also difficult. Staging techniques, in which analysis resources are placed on a separate part of the supercomputer, requires routines for communicating data from the simulation to the analysis software. Co-processing techniques, which place analysis routines directly into the memory space of the simulation code, requires data adapters to convert between the simulation and analysis codes data models (hopefully in a zero-copy manner) as well as a flexible model for coupling the two programs at runtime. Further, making analysis routines work at very high levels of concurrency is an extremely difficult task. In-situ processing algorithms thus must be at least as scalable as the simulation code on petascale and exascale machines. Although some initial work has been done that studies concurrency levels in the tens of thousands of MPI tasks, much work remains. This work is especially critical, because slow performance will directly impact simulation runtime. As an example of the mismatched nature of analysis tasks and simulation tasks, consider the following: the domain decomposition optimized for the simulation is sometimes unsuitable for parallel data analysis and visualization, resulting in the need to replicate data to speed up the visualization calculations. Can this practice continue in the memory-constrained world of in-situ processing

6.3   Technology disruptions

Recently, significant attention is being paid to the challenges and opportunities related to exascale computing the necessary hardware advances and potential impacts to science. However, comparatively little consideration has been given to the resulting disruptive changes in areas such as data analysis and visualization which have become increasingly vital for scientific progress. In particular, it is becoming increasingly clear that data analysis and visualization will have to adopt fundamentally new strategies to remain viable at exascale. On a high level the challenges can be divided into two areas: Data availability; and Data size and complexity. The former is directly linked to increasing spread between the amount of available memory and practical file I/O rates. Effectively, as simulations are becoming more detailed in both space and time less and less data (relative to the

size of the problem) can be permanently stored. Already, current simulations are reaching the limit at which too little data is stored to allow accurate temporal analysis. Furthermore, as simulations explore ever-finer time scales it become increasingly likely that important events in a simulation are not at all represented in the data stored for analysis. To address this challenge new techniques are being developed to enable in-situ data analysis. The goal is to perform all necessary analysis or visualization concurrently with the simulation saving only the results, which is expected, is orders of magnitude smaller than the raw data. However, this approach faces some difficult challenges. For example, typical post-processing analysis tools often do not scale well or depend on a large number of unknown parameters that must be manually tuned to achieve the proper results. Neither of this is feasible for a concurrent analysis. Instead, DOE together with collaborators at other national laboratories and academia is developing topology-based tools. These have been demonstrated to achieve the necessary performance to avoid any significant impact to the primary simulation and enable parameter independent analysis allowing a one-pass processing. The second major change in the traditional analysis and visualization pipeline is the size and complexity of the data. The largest current simulations already produce data that pushes the boundary on what a human observer can comprehend in a single image let alone what common hardware can process. Furthermore, these data sets often require understanding complex inter-relation between different species or across time not well represented in existing techniques. To address this challenge development in fast multi-resolution visualization techniques provide a quick way for users to get an overview of their data while allowing interactive zooming into details. Furthermore, developed new graph drawing and high dimensional analysis techniques to address the rise in complexity through interactive layouts and new visual metaphors. Going forward it is generally accepted that advancing data analysis and visualization to exascale will require significant shifts away from the traditional algorithms and techniques towards novel ways of analyzing and interacting with massive data sets.

6.4 Provenance, metadata, security, privacy

Open-source scientific provenance and workflow management systems that are currently being deployed by science communities support data exploration and visualization. Workflows are traditionally used to automate repetitive tasks. As researchers generate and evaluate hypotheses about data under study, they adjust a workflow in an interactive process, in effect creating a series of different workflows. For data description and discovery, metadata is important. Metadata are based on two complementary themes: descriptive and structural. Metadata catalogs allow users to identify and locate data sets of interest from multiple data centers and archives. Today, most scientific applications maintain a centralized metadata catalog. For security reasons, descriptive metadata may be restricted to certain groups of users. The thought here is to prevent users from discovering data that is not accessible for their use. However, discovery metadata across centers during the search process is based on structural metadata in the form of metadata schemas. For general use, metadata schemas must be extensible to accommodate new types of data and resources at appropriate levels of center operations. In addition to being flexible across multiple science domains, metadata and catalog services are also used to keep track of data replications and the physical locations of actual data files and their accessibility. Secure data access to resources requires the ability to specify and enforce group policy controls. This involves the implementation of Authorization services to grant permission to would be users and Authentication Services to verify that a user is who they claim to be. This process enables users access to remote data and resources under a multitude of scenarios and conditions. For example, data transfer protocols and servers such as GridFTP, HTTP, WGET, etc.), as well as client analysis and machine resources. Underlying authentication services could rely

on the popular single sign-on access to system resources. Security machinery obtaining public key certificate credentials require overhead normally associated with certificate generation. In light of more stringent security systems, security and private policy issues have become more daunting and data and resource providers are requiring security measure to protect and limit data and resource access. In light of the above and as science demands intensify, innovative approaches are needed to improve efficiency and transparency in research. Four very important requirements have been identified:

- Integrated remote data access and analysis methods to enable researchers (and non-researchers) to better collaborate and learn from each other.

- A provenance and workflow environment to easily reproduce analyses and products for anyone requesting results from articles, books, workshops, or reports.

- Descriptive and structural metadata services to capture the data content and structural specification of the data.

- Security and privacy to foster open, secure communication channels and collaborative work environments.

As a result of these requirements, support for community research is expanding and, in particular, will include the delivery of ultra-large data and diagnostic products to a broader research and non-research community. In addition to making document workloads and provenance more transparent, common workflows are being used to optimize access to the data and the analyses involved, reducing the data transmission load on the data infrastructure. The broad community of researchers and non-researchers can thus efficiently access the most popular data products and trace how they were produced. Future requirements call for greatly enhanced remote data access and analysis capabilities, which can be provided via browser-based or client-based access. As we have seen from a range science domains, the real power comes from access to catalogs and higher-level services that are harnessed remotely by a variety of popular client-side applicationssignificantly streamlining the user experience and reducing the volume of data transmitted over the network. Keeping track of the large volumes of data and analysis methods for process is non-trivial because of: Existing disparate metadata, file formats, and conventions. Numerous existing client analysis tools and their underpinning languages. Aggregated data sets located on different archival storage systems. Managing hundreds (if not thousands) of standard or custom analysis workflows with limited federated resources. Data and resource authentication and authorization. System discovery and management of available (or updated) data and resources.

In this new environment, running analysis at remote data centers will be the norm, and provenance recordings will be of utmost importance. That is, users will be interested in maintaining a detailed record of where their data was generated and how it has been processed throughout the workflow process. Metadata catalogs and services will accommodate heterogeneous centralized and decentralized databases. As the number of database rise, scalability and overhead for managing communication will slow. To combat this situation, the peer-to-peer technique has been proven to optimize distributed services with no single point of failure.

Provenance metadata will allow scientists to understand the origin of their results, repeat their experiments, and validate the processes that were used to create or derive data products. The process of conducting data-intensive science research will be primarily (or be more commonly) performed at remote data centers. The goal is to have large-scale analysis processes co-located where ultra-scale data reside. At these data centers, provenance will be recorded at every step in the process and archived as a workflow configuration co-located with the data product. Later, other

scientists can run the same analysis from the workflow descriptor to confirm the results, and they can expand on early findings by running different variations of a processing algorithm or using different input data sources. Keeping track of high-performance computing and clusters, data movement, and data ontology represents knowledge as a set of concepts within the provenance domain and the data relationships among those concepts. For successful comprehensive provenance metadata capture, it will be crucial to capture all aspects of the end-to-end data intensive enterprisefrom hardware to softwarefor an overall increase in scientific productivity and new knowledge discovery.

6.5 Expertise and skills gap

A strong research program cannot be established without a complementary education component, which is as important as adequate infrastructure support. A continuing supply of high quality computational scientists available for work at DOE laboratories is critical. For example, the DOE Computational Science Graduate Fellowship (CSGF) program has successfully provided support and guidance to some of the nations best scientific graduate students, and many of these students are now employed in DOE laboratories, private industry, and educational institutions. However, in order to meet the increasing need for computer and computational scientists trained to tackle exascale and data intensive computing challenges, there is a significant need for a similar program supporting training in exascale and data intensive computing and related areas as outlined previously in this report. The DOE High-Performance Computer Science Fellowship formed by Los Alamos National Laboratory, Lawrence Livermore National Laboratory, and Sandia National Laboratories to foster long-range computer science research efforts in support of the challenges of high-performance computing was a step in right direction. Unfortunately, these fellowships have been discontinued. A DOE Graduate Fellowship in High-Performance Computer Science is needed that trains people in exascale and data intensive computing as well as large-scale data analysis, visualization, scientific data management and high- performance software and hardware [11]. It should be emphasized that the demand for the DOE CSGF is extremely high. In recent years, each year there have been upwards of 600 applicants for 20 or fewer slots. By conservative estimate the program could be significantly increased in size with no diminution of the extraordinary quality of successful applicants.

In addition to the DOE CSGF Program, another excellent program that could help meet the increasing demand for computer and computational scientists trained in exascale and data intensive computing is the DOE Early Career Research Program. In the past three years, the program received more than 3700 proposals from laboratory and university scientists. Of the 200 awards made, only 18 were awarded to ASCR investigators [12]. Again, this program could be significantly increased in size with no diminution of the extraordinary quality of successful applicants.

Another nascent example that could be replicated to help train both new scientists, as well as existing scientists about exascale computing is the Argonne School on Extreme-Scale Computing. The first offering of the School on Extreme-Scale Computing will take place this summer in July 2013. The purpose of the Argonne School on Extreme-Scale Computing is to train computational scientists on the key skills, approaches, and tools that are necessary for implementing and executing computational science and engineering (CS&E) projects on the current and next generations of leadership computing facility (LCF) systems. The two-week program will involve daily lectures and hands-on laboratory sessions, and will culminate in a final exam. Target participants are doctoral students, postdocs, and computational scientists with substantial experience in MPI and/or OpenMP programming, and who have used at least one HPC system for a reasonably complex application and are preparing to conduct CS&E research on large-scale computers. Principal inves-

tigators of Leadership-Class facility-scale computational science projects and researchers/developers of programming models, algorithms, and tools for leading- edge systems may also apply.

Expanding or replicating such summer schools on exascale computing and introducing similar courses on data intensive computing could help train both a new generation of scientists capable of tackling the challenges of exascale and data intensive computing, but also update and upgrade the skills of existing scientists and computer and computational scientists in these important areas as well.

Chapter 7

Findings

7.1 Opportunities for investments that can benefit both Data-Intensive Science and Exascale Computing

There are natural synergies among the challenges facing data-intensive science and exascale computing, and advances in both are necessary for next-generation scientific breakthroughs. Data-intensive science relies on the collection, analysis and management of massive volumes of data, whether they are obtained from scientific simulations or experimental facilities or both. In both cases (simulation or experimental), investments in exascale systems or, more generally, in "extreme-scale" systems[1] will be necessary to analyze the massive data involved in DOE's science missions.

The Exascale Computing Initiative [8] envisions exascale computing to be a sustainable technology that exploits economies of scale. An industry ecosystem for building exascale computers will necessarily include the creation of higher-volume extreme-scale system components which will be beneficial for data analysis solutions at all scales. These components will include innovative memory hierarchies and data movement optimizations that will be essential for all analysis components in a data-intensive science workflow in the 2020+ timeframe.

For example, high-throughput reduction and analysis capabilities are essential when processing large volumes of data generated by science instruments. While the computational capability needed within a single data analysis tier of an experimental facility may not be at the exascale, extreme scale processors built for exascale systems will be well matched for use in different tiers of data analysis, since these processors will be focused (for example) on optimizing the energy impact of data movement. Further, there is a clear synergy between algorithms for near-sensor computing in experimental facilities and algorithms for in-situ analysis in simulations.

The Exascale Computing Initiative has also identified the need for innovations in applications and algorithms to address fundamental challenges in extreme-scale systems related to concurrency, data movement, energy efficiency and resilience. Innovative solutions to these challenges will jointly benefit analysis and computational algorithms for both data-intensive science and exascale computing. Finally, advances in networking facilities (as projected for future generations of ESNet [7]) will also benefit both data-intensive science and exascale computing.

---

[1]As in past reports, we use "exascale systems" to refer to systems with an exascale capability and "extreme-scale systems" to refer to all classes of systems built using exascale technologies which include chips with hundreds of cores and different scales of interconnects and memory systems.

7.2 Integration of Data Analytics with Exascale Simulations represents a new class of workflow

In the past, the computational science workflow was represented by large-scale simulations followed by off-line data analyses and visualizations. Today's ability to understand and explore gigabyte and some petabyte spatial-temporal high-dimensional data in this workflow is the result of decades of research investment in data analysis and visualization. However, exascale data being produced by experiments and simulations are rapidly outstripping our current ability to explore and understand them. Exascale simulations require that some analyses and visualizations be performed while data is still resident in memory, so-called *in-situ* analysis and visualization, thus necessitating a new kind of workflow for scientists. In addition, we need new algorithms for scientific data analysis and visualization along with new data archiving techniques that allow for both in-situ and post processing of petabytes and exabytes of simulation and experimental data. This new kind of workflow will impact data-intensive science due to its tighter coupling of data and simulation, while also offering new opportunities for data analysis to steer computation.

In addition, in-situ analysis will impact the workloads that high-end computers have traditionally been designed for. Even for traditional floating-point-intensive applications, the addition of analytics will change the workload to include (for example) larger numbers of integer operations and branch operations than before. Design and development of scalable algorithms and software for mining big data sets, as well as an ability to perform approximate analysis within certain time constraints will be necessary for effective in-situ analysis. In the past, different assumptions were made for designing high-end computing systems vs. analysis and visualization systems. Tighter integration of simulation and analytics in the science workflow will impact co-design of these systems for future workloads, and will require development of new classes of proxy applications to capture the combined characteristics of simulations and analytics.

7.3 Urgent need to simplify the workflow for Data-Intensive Science

Analysis and visualization of increasingly larger-scale data sets will require integration of the best computational algorithms with the best interactive techniques and interfaces. The workflow for data-intensive science is complicated by the need to simultaneously manage large volumes of data as well as large amounts of computation to analyze the data, and this complexity is increasing at an inexorable rate. These complications can greatly reduce the productivity of the domain scientist, if the workflow is not simplified and made more flexible. For example, the workflow should be able to transparently support decisions such as when to move data to computation or computation to data. The recent proposal for a Virtual Data Facility (VDF) will go a long way in simplifying the workflow for data-intensive science because of its integrated focus on data-intensive science across the DOE ASCR facilities.

7.4 Need for Computer and Computational Scientists trained in both Exascale and Data-Intensive Computing

Earlier workflow models allowed for a separation of concerns between computation and analytics that is no longer possible as computation and data analysis become more tightly intertwined. Further, the separation of concerns allowed for science to progress with personnel that may be

experts in computation or in analysis, but not both. This approach is not sustainable in data-intensive science where the workflow for computation and analysis will have to be co-designed. There is a need for investments to increase the number of computer and computational scientists trained in both exascale and data-intensive computing to advance the goals of data-intensive science.

Chapter 8

Recommendations

8.1 Investments that can benefit both Data-Intensive Science and Exascale Computing

The DOE Office of Science should give higher priority to investments that can benefit both data-intensive science and exascale computing so as to leverage their synergies.

The findings in this study have identified multiple technologies and capabilities that can benefit both data-intensive science and exascale computing. Investments in such dual-purpose technologies will provide the necessary leverage to advance science on both data and computational fronts. For science domains that need exascale simulations, commensurate investments in exascale computing capabilities and data infrastructure are necessary for advancement. In other domains, extreme-scale components of exascale systems will be well matched for use in different tiers of data analysis, since these processors will be focused on optimizing the energy impact of data movement. Further, innovations in applications and algorithms to address fundamental challenges in concurrency, data movement, and resilience will jointly benefit data analysis and computational techniques for both data-intensive science and exascale computing. Finally, advances in networking (as projected for future generations of ESNet technology) will also benefit both data-intensive science and exascale computing.

8.2 Simplifying Science Workflow and improving Productivity of Scientists involved in Exascale and Data-Intensive Computing

DOE ASCR should give higher priority to investments that simplify the science workflow and improve the productivity of scientists involved in exascale and data-intensive computing.

The findings in this study have identified multiple such opportunities including a) leveraging commonalities in visualization and analytics requirements across multiple science domains, b) integration of simulation and experimental studies, and c) expanding the scope of co-design to include scenarios from data-intensive science.

Todays ability to understand and explore gigabyte and some petabyte spatial-temporal three-dimensional data and higher dimensional data is the result of decades of research investment by DOE, NSF, DARPA and other agencies in data analysis and visualization . However, exascale data being produced by experiments and simulations are rapidly outstripping our ability to explore and understand them. As such, we need new algorithms for scientific data analysis and visualization along with new data archiving techniques that allow for both in-situ and post processing

of petabytes and exabytes of simulation and experimental data. Analysis and visualization of increasingly larger-scale data sets will require integration of the best computational algorithms with the best interactive techniques and interfaces. We must pay greater attention to human computer interface design and human in the loop workflows.

We must pay greater attention to simplifying human-compute-interface design and human-in-the-loop workflows for data-intensive science. To that end, we encourage the recent proposal for a Virtual Data Facility (VDF) because it will provide a simpler and more usable portal for data services than current systems. A significant emphasis must be placed on developing a collection of scalable data analytics and data mining algorithms and software components that can be used as building blocks for sophisticated analytics pipelines and flows. We also recommend the creation of new classes of proxy applications to capture the combined characteristics of simulation and analytics, so as to help ensure that computational science and computer science research in ASCR are better targeted to the needs of data-intensive science.

8.3 Recommendation for Building Expertise in Exascale and Data-Intensive Computing

DOE ASCR should adjust investments in programs such as fellowships, career awards, and funding grants, to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing.

There is a significant gap between the number of current computational and computer scientists trained in both exascale and data-intensive computing and the future needs for this combined expertise in support of DOE's science missions. Investments in ASCR such as fellowships, career awards, and funding grants should look to increase the pool of computer and computational scientists trained in both exascale and data-intensive computing.

Chapter 9

Conclusions

9.1 Summary of report

This report reviewed current practice and future plans in multiple science domains in the context of the Big Data and the Exascale Computing challenges that they will face in the future. The review drew from public presentations, workshop reports and expert testimony. Data-intensive research activities are increasing in all domains of science, and exascale computing is a key enabler of these activities. The report includes key findings and recommendations from the perspective of identifying investments that are most likely to positively impact both data-intensive science goals and exascale computing goals.

TO BE COMPLETED

9.2 Synergies between Data-Driven Science and Commercial Big Data Systems

While the scope of this study was focused on synergies between data-driven science and exascale computing, the subcommittee encountered a few cases where there were also synergies with commercial big data systems.

TO BE COMPLETED

9.3 Broader impact

While the scope of this study was focused on DOE Office of Science's unique role in data-intensive science vis-a-vis other agencies, some of the findings in this report may be relevant to other federal agencies and also to commercial applications.

TO BE COMPLETED

Appendix A

Charge to Subcommittee

July 25, 2012

Professor Roscoe Giles, ASCAC Chair
Department of Electrical & Computer Engineering
Boston University
8 St. Mary's Street
Boston, MA 02215

Dear Professor Giles:

Thank you for the recent Advanced Scientific Computing Advisory Committee (ASCAC) report on the Computational Sciences Graduate Fellowship. The report was thorough, informative and very timely.

Overcoming the challenges of managing data rates and movement of data in an exascale computing environment will likely require significant research investments. In addition to the challenges and opportunities of exascale computing, the Office of Science is facing related challenges from data-intensive research activities, such as the growing volumes of data generated at our next generation scientific user facilities and by the new genomics-based technologies that are enabling a revolution in systems biology research. The Linac Coherent Light Source, for example, currently generates several petabytes of data each year and the National Synchrotron Light Source II, currently under construction and scheduled to begin operations later this decade, is expected to generate hundreds of petabytes of data each year. In order to maximize the return on our limited federal resources, we need to understand the similarities among and differences between these data challenges and the potential to leverage research investments to address issues spanning both exascale and data-intensive science.

By this letter, I am charging the ASCAC to assemble a subcommittee to examine the potential synergies between the challenges of data-intensive science and exascale. The subcommittee should take into account the Department's mission needs, which define the Office of Science's unique role in data-intensive science vis-a-vis other agencies. The subcommittee should specifically address what investments are most likely to positively impact both our exascale goals and our data-intensive science research programs, including data management at our next generation facilities.

I would appreciate the committee's preliminary comments by November 2012 and a final report by March 30, 2013. I appreciate ASCAC's willingness to undertake this important activity.

2

If you have any questions regarding this matter, please contact either Daniel Hitchcock, the Associate Director of the Office of Science for ASCR or Christine Chalk, the Designated Federal Official for the ASCAC.

Sincerely,

W. F. Brinkman
Director, Office of Science

Appendix B

Acknowledgments

The subcommittee would like to thank the following experts for their input provided in teleconference calls attended by subcommittee members:

- Jacek Becla, SLAC.

- Amber Boehnlein, SLAC.

- Roscoe Giles, ASCAC Chair.

- Barbara Helland, ASCR.

- Chris Jacobsen, APS.

- Lucy Nowell, ASCR.

- Sonia Sachs, ASCR.

- Nicholas Schwarz, APS.

- Rick Stevens, ANL.

We are also grateful to Vincent Cave and Shams Iman from Rice University for their help with the document infrastructure used to produce this report, and to Nathan Galli from University of Utah for his design of the report cover.

Appendix C

Subcommittee Members

The ASCAC Subcommittee on Synergistic Challenges in Data-Intensive Science and Exascale Computing consisted of the following members:

- Jacqueline Chen, Sandia National Laboratory, ASCAC member.

- Alok Choudhary, Northwestern University.

- Stuart Feldman, Google.

- Bruce Hendrickson, Sandia National Laboratory.

- Chris Johnson, University of Utah.

- Richard Mount, SLAC.

- Vivek Sarkar, Rice University, ASCAC member (subcommittee chair).

- Victoria White, FermiLab, ASCAC member.

- Dean Williams, LLNL, ASCAC member.

Bibliography

[1] Luiz André Barroso and Urs Hölzle. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Synthesis Lectures on Computer Architecture. Morgan & Claypool Publishers, 2009.

[2] J. C. Bennett, H. Abbasi, P. Bremer, R. W. Grout, A. Gyulassy, T. Jin, S. Klasky, H. Kolla, M. Parashar, V. Pascucci, P. Pbay, D. Thompson, H. Yu, F. Zhang, and J. Chen. Combining in-situ and in-transit processing to enable extreme-scale scientific analysis. In *ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis (SC)*, November 2012.

[3] Alok Choudhary. Discovering Knowledge from Massive Networks and Science Data Next Frontier for HPC. Invited talk at the 2012 DOE CSGF Annual Conference, July 2012.

[4] James C. Corbett, Jeffrey Dean, Michael Epstein, Andrew Fikes, Christopher Frost, J. J. Furman, Sanjay Ghemawat, Andrey Gubarev, Christopher Heiser, Peter Hochschild, Wilson Hsieh, Sebastian Kanthak, Eugene Kogan, Hongyi Li, Alexander Lloyd, Sergey Melnik, David Mwaura, David Nagle, Sean Quinlan, Rajesh Rao, Lindsay Rolig, Yasushi Saito, Michal Szymaniak, Christopher Taylor, Ruth Wang, and Dale Woodford. Spanner: Google's globally-distributed database. In *Proceedings of the 10th USENIX conference on Operating Systems Design and Implementation*, OSDI'12, pages 251–264. USENIX Association, 2012.

[5] M. Ellisman, R. Stevens, M. Colvin, T. Schlick, E. Delong, G. Olsen, J. George, G. Karniakadis, C.R. Johnson, and N. Sematova. Scientific grand challenges: Opportunities in biology at the extreme scale of computing. Technical report, August 2009.

[6] Peter Kogge et Al. Exascale computing study: Technology challenges in achieving exascale systems. *Technical Report, AFRL contract Number FA8650-07-C-7724*, 2008.

[7] Roscoe Giles et al. Draft ASCAC Facilities Letter, February 2013.

[8] William Harrod. A Journey to Exascale Computing. Invited plenary talk at SC12: The International Conference for High Performance Computing, Networking, Storage and Analysis, July 2012.

[9] The White House. Fact Sheet: Big Data Across the Federal Government, March 2012.

[10] C.R. Johnson, R. Moorhead, T. Munzner, H. Pfister, P. Rheingans, and T.S. Yoo. NIH/NSF visualization research challenges report. Technical report, 2006.

[11] C.R. Johnson, R. Ross, S. Ahern, J. Ahrens, W. Bethel, K.L. Ma, M. Papka, J. van Rosendale, H.W. Shen, and J. Thomas. Visualization and Knowledge Discovery: Report from the

DOE/ASCR Workshop on Visual Analysis and Data Exploration at Extreme Scale. Technical report, October 2007.

[12] DOE Office of Science. DOE Office of Science Early Career Research Program. http://science.energy.gov/˜/media/early-career/pdf/FAQ_FY13.pdf, 2013.

[13] Berkin Özisikyilmaz, Ramanathan Narayanan, Joseph Zambreno, Gokhan Memik, and Alok N. Choudhary. An architectural characterization study of data mining and bioinformatics workloads. In *IISWC*, pages 61–70, 2006.

[14] Amedeo Perazzo. ASCR/BES Data Pilot: FEL Facilities Perspective. Presentation given at Facility Directors Meeting, December 2012.

[15] Vivek Sarkar, William Harrod, and Allan E. Snavely. Software Challenges in Extreme Scale Systems. January 2010. Special Issue on Advanced Computing: The Roadmap to Exascale.