DDN STORAGE

SUCCESS STORY

ACCELERATE: LIFE SCIENCES

DDN's Powerful, Scalable Storage Helps Virginia Bioinformatics Rapidly and Accurately Predict the Spread of the World's Largest and Deadliest Outbreak of Ebola

CHALLENGES

- Rapid response to emerging Ebola crisis required expedited, data-intensive outbreak modeling
- Highly scalable storage needed to support the creation of highly detailed synthetic global populations
- Dozens of computational models required high-performance, flexible data-heavy and I/O parallel processing with flexible, high-performance storage to support dozens of models with hundreds of simulations

SOLUTION

DDN's Storage Fusion Architecture® GRIDScaler Appliances with 1PB of storage with industry-leading density, performance and reliability

RESULTS

- High-performance storage fueled rapid-response Ebola outbreak modeling with recommendations to DOD within 48 hours
- Flexible storage platform supported both data- and compute-intensive computational epidemiology

VIRGINIA BIOINFORMATICS INSTITUTE (VBI) IS A WORLD-CLASS RESEARCH INSTITUTE DEDICATED to the study of information biology. Founded in 2000, more than 250 scientists and staff at state-of-the-art facilities on Virginia Tech's Blacksburg campus and the National Capital Region Research Center in Arlington, Va., strive to uncover the complex relationships between human health, well-being and sustainability.

VBI develops genomic, proteomic and bioinformatics tools, databases and computational models to study genomes and diseases for the discovery of new vaccine, drug and diagnostic targets as well as better understand how to respond to disease outbreaks. Infectious diseases account for more than 13 million deaths each year, so the ability to quickly and accurately understand and predict their spread is crucial, especially during emerging crises.

Since July 2014, computational epidemiologists at VBI's Network Dynamics and Simulation Science Laboratory (NDSSL) have been on the frontlines in combatting the world's largest and deadliest outbreak of Ebola. A team of 30 researchers and scientists on VBI's Ebola rapid response team worked with the U.S. Department of Defense's Defense Threat Reduction Agency (DTRA), National Institutes of Health (NIH), World Health Organization (WHO) and West Africa's Ministries of Health (MOH) to forecast the disease spread and mitigate risks of the outbreak moving outside of the infected regions into other parts of the world.

According to statistics released by the U.S. Centers for Disease Control and Prevention, there have been 22,000 reported cases and Ebola has already claimed more than 9,000 lives as of February 2015 from the 2014 Ebola Outbreak in West Africa. Continued study into how Ebola passes among individuals, as well as monitoring outbreaks and infection clusters, is critical in assisting with rapid response efforts mitigating the spread outside of infected regions and into other parts of the world, including the United States.

Using a combination of team science and sophisticated modeling techniques to predict the spread of the epidemic, VBI scientists have created an adaptable set of global synthetic populations with detailed demographics, family structures, travel patterns and activities to help model what could happen as the outbreak spreads. The synthetic data is created in such a way that it omits personally identifiable information but mirrors actual census, social, transit and telecommunications data patterns from the targeted population. Millions of simulations are applied to dozens of computational models to fuel rapid response efforts.

"Through computational modeling, we can better understand how Ebola is transmitted, predict infection 'hot spots' and provide vital data to assist with policy recommendations and interventions," says Bryan Lewis, computational epidemiologist and leader of VBI's response team at NDSSL. "As with all epidemics, time is of the essence. The ability to have all the necessary data at our fingertips is essential to delivering rapid answers to a series of tough questions."

> *With DDN, we've attained a fast, reliable parallel file system to handle all our different workloads. DDN was best suited for our mix of computational models while delivering both the capacity and performance demanded by multiple users who need to access different data at the same time.*

Kevin Shinpaugh
PhD, director of IT and High Performance Computing at Virginia Bioinformatics Institute

> *It was critical to provide answers within the DoD's decision cycle as a late answer was the same as no answer at all. Our robust storage and compute infrastructure enabled us to run additional models with more variables for consideration, giving us confidence in the quality of our response.*

Keith Bisset
Senior Research Scientist, Virginia Bioinformatics Institute

THE CHALLENGE

Producing in-depth models and analyses of disease and crisis processes are familiar to NDSSL's computational epidemiologists. For the past decade, VBI has collaborated with scientists around the world as part of the institute's work with the NIH and its Models of Infectious Disease Agent Study (MIDAS) network along with DTRA and WHO. Over the years, VBI has supplied significant research and informatics to support pandemic influenza planning for H1N1 and H5N1, shed light on optimizing household responses to epidemics and informed decision makers about likely challenges of resource scarcity during pandemics.

According to Kevin Shinpaugh, PhD, director of IT and High Performance Computing at VBI, the ability to produce intricate and accurate computational models of global populations is part of what differentiates VBI in the research world. "We are unique in our ability to create models with incredible breadth and depth of detail," he explains. "What this means on the computational side is that we must keep up with a constant appetite for compute and storage capacity to produce unmatched levels of information."

In fielding DTRA's request for modeling the Ebola outbreak, VBI needed a robust cyber-infrastructure capable of handling both data and compute intensive demands. "We built virtual cities on local, regional and global levels, infused with a tremendous amount of demographics collected from census, power, cell phone, transportation, social networks and other data," says Keith Bisset, senior research scientist for VBI. "Our ability to expand our population framework enabled us to move to Ebola modeling very quickly."

After receiving the call from DTRA, teams from VBI's NDSSL and Advanced Computing and Informatics Laboratories (ACIL) ramped up the institute's Comprehensive National Incident Management Systems (CNIMS). The highly scalable and agent-oriented system supports very large, complex event characterization, situational assessments and decision support for adaptation to a broad range of possible national-scale crisis events.

To support Ebola outbreak modeling, VBI required a mix of computations, which meant its compute and storage would have to be both powerful and flexible enough to handle various workloads. Some models require a lot of data while others have a constant stream of information requiring parallel processing. "We have three or four main simulation engines," says Lewis. "One is fast but assumes the social network is just a graph, whereas the slower one checks the social network every second and is more versatile in its representation. Understanding all the nuances is the art behind which simulation engine we choose to perform different computational models."

VBI needed its storage infrastructure to keep pace with escalating growth. "Just one instance of our global population requires 10TB," notes Bisset. "And, we might run the simulation 15 different times with variant conditions, which means we need ample storage capacity to accommodate the multiple runs." Storage scalability is a top requirement as some parts of the Institute's storage infrastructure grow by 100 percent each year. "As soon as I release more storage, it's quickly used," adds Shinpaugh. "So, I need to ensure that we can address unfolding requirements with a robust storage platform that can grow in place with efficiently and cost-effectively."

THE SOLUTION

In addressing the critical challenges related to Ebola outbreak modeling, VBI took advantage of the robust performance of its Shadowfax compute cluster with 2,500 cores and nearly 1PB of high-performance storage from DataDirect Networks (DDN). The institute deployed a powerful storage cluster, built on two DDN Storage Fusion Architecture (SFA) GRIDScaler™ Appliances which have DDN's edition of IBM® GPFS™ parallel file systems embedded within the storage controllers for ease of management and lower data access latency. These are used for home and scratch file storage.

"With DDN, we've attained a fast, reliable parallel file system to handle all our different workloads," says Shinpaugh. "DDN was best suited for our mix of computational models while delivering both the capacity and performance demanded by multiple users who need to access different data at the same time."

A team of 30 researchers and scientists on VBI's Ebola rapid response team initially provided DTRA and West Africa's Ministries of Health (MOH) with short-term forecasts of disease spread and answers to questions about vaccine production. As the number of Ebola cases began to climb, VBI moved to agent-based computational modeling to provide more in-depth analysis, including regional travel and social interactions, to deliver more insight into the movement of the outbreak.

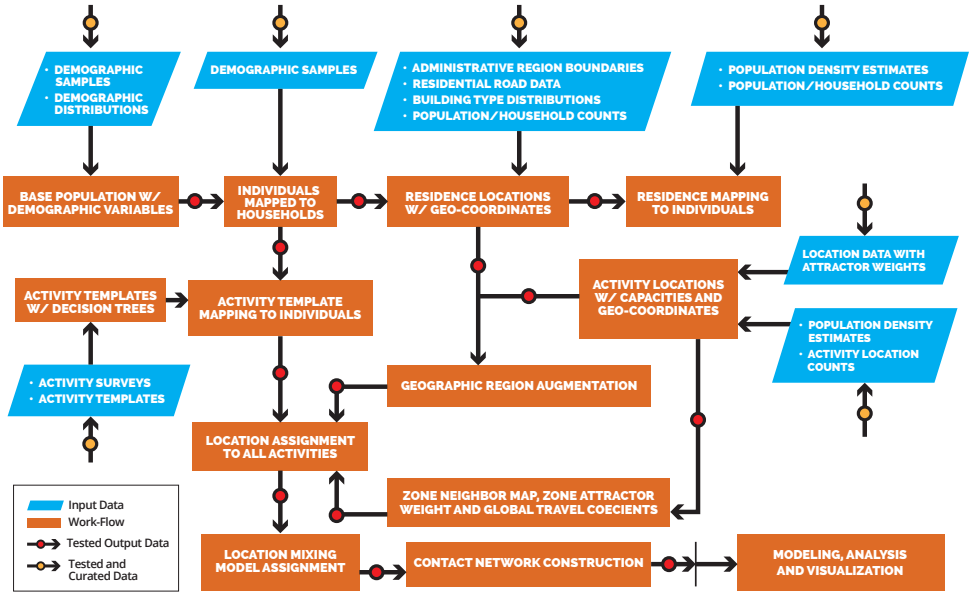
BUSINESS BENEFITS

- VBI created an ever-expanding synthetic global population, which now contains three-quarters of a billion people for computational models to support outbreak spread predictions and interventions
- Unique breadth and depth of modeling detail enabled VBI to provide DOD with recommendations on where to place emergency treatment units in West Africa within 48 hours
- The ability to continually add to VBI's synthetic population framework ensures the institute can accelerate the production of computational models with better input and more answers

> *With robust storage from DDN, VBI was able to continually add to our synthetic population framework and produce models that made it easier to offer rapid recommendations on next steps. We now have the ability to offer better input and more answers much faster.*

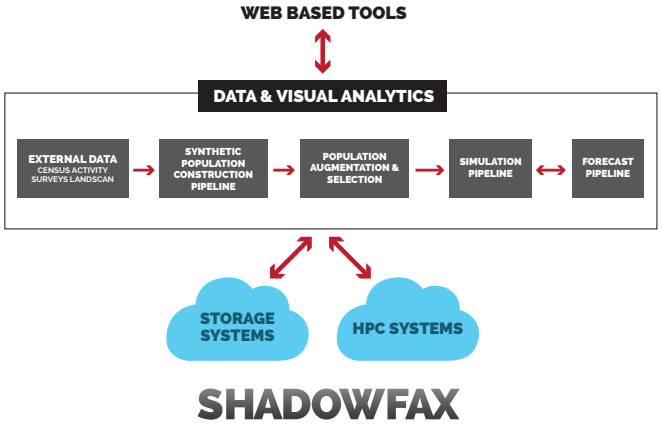
Bryan Lewis
Computational Epidemiologist, Virginia Bioinformatics Institute

BUILDING AND MODELING WITH A SYNTHETIC POPULATION



The institute relied on internally developed HPC modeling tools, including EpiFast, a very fast simulation for producing a broad class of interventions; EpiSimdemics, for yielding highly-detailed and flexible, complex dynamic interventions with rich disease specifications; and Indemics, interactive, SQL-based queries of simulations and very fast specifications of novel interventions. Additionally, VBI leveraged Panda and Python open-source data analysis tools to help DTRA and the MOH determine the resources needed to combat the outbreak.

All of VBI's computational modeling tools have been designed to scale rapidly. For example, EpiSimdemics was used initially to create a population model of the U.S., which researchers used to simulate 120 days of activity. Creating this synthetic population took 12 seconds; expanding the model to accommodate a global population was projected to take six minutes.



The team also relies on a variety of computational epidemiology applications, including Sibel, a public health response tool; Epinome, a public health training tool; SIV population visualization; Flu Caster, a crowd-sourced surveillance tool; and Virus Tracker, a live action simulation application. "The variety of data we gather as part of our modeling process drives the incredible amount of detail within our models as well as the output of each model," notes Shinpaugh. "With DDN storage, we're confident we can scale data storage to address both current and future modeling demands while expediting accurate responses during an emerging crisis."

When the number of cases in Sierra Leone jumped exponentially in September, scientists added more details to their population model to assist with increased contact tracing and spread predictions. "Since census data in parts of West Africa was lacking, we structured models with social triggers and hooks based on behavior," says Lewis. "All this modeling and constant feedback generated a massive amount of data and DDN's high-performance storage met these unique demands."

TECHNICAL BENEFITS

- Centralized storage with an embedded parallel file system makes it easy to add storage where needed—in support of both data and compute intensive requirements—and then manage it all through a single pane of glass
- DDN storage cluster offers scalable capacity to accommodate both home and scratch file storage
- Seamless support of diverse computational epidemiology applications and data analysis tools accommodates increasing modeling demands
- DDN's WOS and iRods support is ideally suited to address increasing collaboration and data protection needs

THE BENEFITS

Armed with detailed computational models, VBI met DTRA's needs for expedited answers. "With robust storage from DDN, VBI was able to continually add to our synthetic population framework and produce models that made it easier to offer rapid recommendations on next steps," Lewis adds. "We now have the ability to offer better input and more answers much faster."

For example, VBI received a call from the DOD on a Friday requesting input on where to place new emergency treatment units (ETUs) by Monday morning, when the military transport planes were taking off. The scientists quickly looked at many variables, including the road infrastructure in West Africa. They also identified "hot spots" where additional outbreaks were considered imminent. "It was critical to provide answers within the DOD's decision cycle as a late answer was the same as no answer at all," recalls Bisset. "Our robust compute and DDN storage infrastructure enabled us to run additional models with more variables for consideration, giving us confidence in the quality of our response."

VBI has continued to build out its synthetic population, which now numbers more than three-quarters of a billion people. In its quest to build highly accurate profiles, genomics data and other details generated by the institute's Social Analytics and Decision Laboratory are added continually. "Our synthetic global population forms the basis for a lot of our research both now and in the future," adds Bisset. "We have the ability to quickly react when a situation arises to offer insight based on hundreds of unique combinations of parameters."

This underlying framework is bolstered by VBI's converged storage infrastructure from DDN, which continues to play a pivotal role in accelerating and simplifying computational epidemiology research and collaboration. "Our ever-expanding synthetic population and computational modeling will drive constant storage consumption," says Shinpaugh. "With DDN's scalable storage, we can keep pace with aggressive, rapid data growth."

VBI strives to make its synthetic global population and other informatics research available to other researchers. Currently, information is shared with approximately 75 scientists and computational epidemiologists as part of VBI's work with the MIDAS network. The institute also is exploring the use of DDN's WOS® Object Storage to more efficiently share, protect and distribute huge volumes of modeling data, including unstructured social analytics like Twitter.

In the coming year, VBI will further investigate DDN's ability to embed applications like iRODS—an open source data grid solution for managing large sets of data files—within its storage system. "The flexibility of our DDN infrastructure gives us options for sharing and safeguarding more data as it needs to move around as well as pull in information from other databases," concludes Shinpaugh. "The solid, performance of our DDN system has exceeded our expectations while providing great value in terms of price and performance. With DDN, we can stay ahead of whatever comes next."

ABOUT DDN®

DataDirect™ Networks (DDN) is the world leader in massively scalable storage. Our high performance, high efficiency data storage solutions and professional services enable content-rich and high growth IT environments to achieve the highest levels of systems scalability, performance, efficiency and simplicity. DDN enables enterprises to extract value and deliver business results from their information.

Our customers include the world's leading online content and social networking providers, high performance cloud and grid computing, life sciences, media production, and security and intelligence organizations.

Deployed in thousands of mission critical environments worldwide, DDN's solutions have been designed, engineered and proven in the world's most scalable data centers to ensure competitive business advantage for today's information powered enterprise.