

# Large Data Federated Distributed Infrastructure

www.datadirectnet.com

## LARGE DATA

Joint Capabilities Technology Demonstration  
USSTRATCOM - NGA - NRL - DISA - INSCOM - DIA and Others



DataDirect Networks and SGI deliver advanced image acquisition, storage, retrieval, and processing enabling real-time intelligence on the battlefield in the Naval Research Laboratory's Large Data Joint Capabilities Technology Demonstration



# Large Data Federated Distributed Infrastructure

[www.datadirectnet.com](http://www.datadirectnet.com)

## Challenge

Battlefield imagery from satellites, UAVs, and first responders must be gathered, shared, and analyzed in real-time in order to safeguard U.S. forces and achieve mission objectives.

## Application

The Large Data JCTD is a pioneering storage, computing, and communications network designed to handle massive inflows of image and video data and allow analysts to operate on it in real-time from remote locations as if the information were local to their desktops.

## Solution

NRL and the Large Data JCTD turned to DataDirect Networks and SGI to deliver the critical storage and computing infrastructure. DataDirect's Infiniband-based storage systems provide low-latency, high speed, shared access to a Petabyte-scale storage pool, while SGI's Altix servers enable real-time visualization and manipulation of the incoming images

NRL uses an innovative unified Infiniband architecture to provide high-speed, high-capacity, low-latency storage and a single-wire computing and networking infrastructure enabling the warfighter in the Large Data Joint Capabilities Technology Demonstration.

For several years now the military has used satellite, manned airborne, and Unmanned Aerial Vehicle (UAV) photography to gain insight to the battlefield. However, the continuing challenge is obtaining rapid access to the information being collected, sharing it among analysts, planners, and decision makers, and using it to provide a decisive advantage. As sensors gain in numbers and analysis is performed in multiple spectrums, the amount of data being generated has grown tremendously, requiring new technologies to retrieve, store, move, and make sense of it.

At the Naval Research Laboratory (NRL), this challenge is being met through the Large Data Joint Capability Technology Demonstration (Large Data JCTD) project. NRL's team of scientists and engineers have architected a revolutionary high-speed data acquisition, storage, retrieval, and analysis system based on proven, low-latency Infiniband technology that enables military planners to more efficiently utilize the mountain of data from UAVs, first responder video, and satellite imagery, making it rapidly accessible from their desktops regardless of where the information is physically stored.

The Large Data JCTD is a robust, reliable, and scalable storage, computing, and distribution network built across geographically dispersed sites interconnected by multiple, high-speed, low-latency Infiniband data links. Images are streamed in real-time to a data center at rates that can exceed one Terabyte per hour per sensor. Individual file sizes grow to hundreds of gigabytes, and multiple images are often digitally "stitched" together to provide a larger, coherent view of an area, and to layer multi-

spectral data such as visible light, infrared, and radar. With potentially dozens or hundreds of sensors streaming data in an active theatre, the amount of raw data to ingest, store, process, and share can easily grow into Petabytes (1,000 Terabytes) or even Exabytes (1,000,000 Terabytes).

Large Data JCTD must not only ingest and store massive amounts of data, but must also deliver real-time processing and distribution of the information wherever it is needed. Raw data is collected and mirrored in real time between multiple special-purpose, high-speed Infiniband-connected storage arrays specifically designed by DataDirect Networks to handle large files and high bit-rate streaming. A range of SGI high-performance computing systems and clusters, process the information, running sophisticated search engines to automatically index the data as it is acquired. The storage systems and compute nodes in and between the Large Data JCTD sites are interconnected using ultra low-latency Infiniband technology with caching replication, enabling any user of the system to view data as if it were locally stored and processed, regardless of the actual distance to the computing and storage resources. All data in transit across the network is encrypted using NSA approved encryption technology.

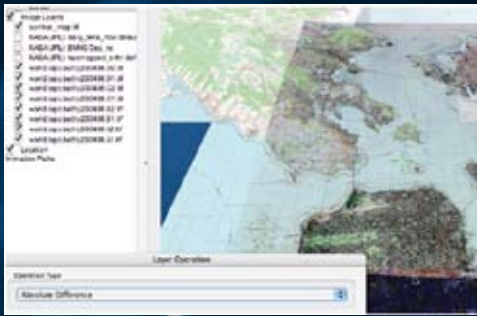
## Three Challenges: Storage, Transport, and Visualization

Large Data JCTD's project requirements set forth several new challenges. First was the need to handle massively large data files and total data sets. Even in trials the data would reach nearly a Petabyte per site and require ingest and output rates exceeding 3GB/s.

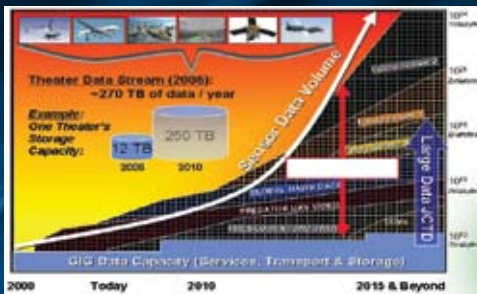
# Large Data Federated Distributed Infrastructure

www.datadirectnet.com

## Large Data JCTD Real-time Image Visualization



## A Mountain of Data Requires New Analysis Capabilities



As data is acquired, it may need to be automatically replicated between each Large Data JCTD site. At the data rates required by the project, this presented challenges in both WAN transport and encryption technologies.

Image indexing, processing, and manipulation represented the last major challenge. Even with the images available locally, combining multi-spectral data, automatically searching the images for points of relevance, and rapidly visualizing the images for human analysts was key to success.

### Solution: DataDirect Networks & SGI

The storage challenge of Large Data JCTD was met by DataDirect Networks. The company's S2A storage systems were purpose built for speed and scalability and have been deployed in the world's leading supercomputing sites, where bandwidth and massive data storage are the norm. DataDirect Networks supplied S2A storage appliances to each Large Data JCTD site, managing disk arrays that are scalable to nearly a Petabyte each in a dense footprint using only two data center racks.

The S2A storage solution is unique because it can handle the rapid data ingest rates of today's sensors, while scaling to allow for future growth in bandwidth and capacity as sensors improve.

SGI provided Large Data JCTD with an ar-

ray of Infiniband-connected Altix servers as well as its CXFS high performance clustered file system. The SGI compute nodes provide all image processing, data management, and automated indexing functions. Infiniband provides a single wire solution for the computing, storage, and communications networks.

DataDirect Networks and SGI worked together to integrate their respective systems with WAN encryption and replication technology supplied by additional vendors.

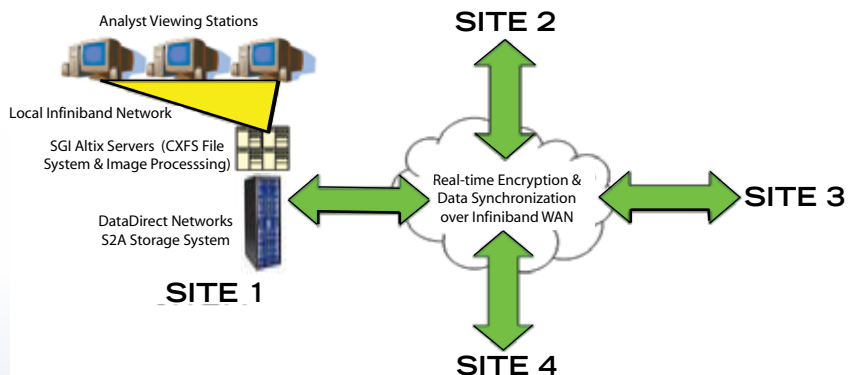
### Large Data JCTD Provides Critical Battlefield Support

The Large Data project is fundamentally about saving lives by exploiting the time value of information. When soldiers are supported by real-time situational analysis, an advantage is gained.

“Safeguarding the warfighter requires relevant information to be delivered in real-time. The historical process of offline acquisition and analysis is not enough in today's theatres. The Large Data system proves that there is a better way.”

— Hank Dardy, Chief Scientist for the Center for Computational Science at the Naval Research Center

### The NRL Large Data JCTD Network Links all Sites for Real-time Analysis of Large Image Sets



DataDirect Networks is the leading provider of open, scalable storage systems for performance and capacity driven applications. DataDirect Networks' S2A (Silicon Storage Architecture) appliance enables modern applications such as video streaming, content delivery, modeling and simulation, backup and archiving, cluster and supercomputing, and real-time collaborative workflows, that are driving the explosive demand for storage performance and capacity. DataDirect Networks' S2A technology and solutions solve today's most challenging storage requirements, including providing shared, high-speed access to a common pool of data, minimizing data center footprints and storage costs for massive archives, reducing simulation computational times, and capturing and serving massive amounts of digital content.

Major corporations, supercomputing centers and rich media organizations, including AOL, Ascent Media, Boeing, CINECA, CGGVeritas, CNN, Disney, Federal Reserve Board, Ford, Hess, Kodak Gallery, Lawrence Livermore National Laboratories, Mediaset, Microsoft, MTV, NASA Ames, NBC Universal, RIOT, Sandia National Laboratories, Shutterfly, Slide.com, Sony, Technical University Dresden, Technicolor, Time Warner, Thomson, Total, and Trinity College Dublin utilize DataDirect Networks high performance, high capacity solutions.

SGI is a leader in high-performance computing. SGI delivers a complete range of high-performance server, storage and visualization solutions along with industry-leading professional services and support that enable its customers to overcome the challenges of complex data-intensive workflows and accelerate breakthrough discoveries, innovation and information transformation.

SGI helps customers solve some of their toughest computing challenges whether it's enhancing the quality of life through drug research, designing and manufacturing safer and more efficient cars and airplanes, studying the global climate, providing technologies for homeland security and defense, or helping enterprise manage large sets of data. With offices worldwide, the company is headquartered in Sunnyvale, California, and can be found on the Web at [www.sgi.com](http://www.sgi.com).

**DataDirect**  
NETWORKS  
Performance. Capacity. Innovation.

9351 Deering Avenue . Chatsworth . California 91311  
phone +1.800.TERABYTE (837.2298) . fax +1.818.700.7601  
[sales@datadirectnet.com](mailto:sales@datadirectnet.com)  
[www.datadirectnet.com](http://www.datadirectnet.com)