

Scale Bigger

Store Smarter

Accelerate Everything

A High-Performance Storage and Ultra-High-Speed File Transfer Solution for Collaborative Life Sciences Research

BlueArc Storage Platforms with Aspera

Overview

A growing number of organizations with data-intensive application environments in research, life sciences, high-performance computing, financial services, media and entertainment, have adopted file-based workflows for the delivery and creation of data and media. As a result, location independence is a driving force in the way business and research is conducted. Files of virtually every imaginable size and format need to be stored and transported between facilities, delivered to distribution hubs, backed up and archived, over national and global distances. But moving many gigabytes or terabytes of data over public or private IP networks can introduce great inefficiencies in a digital workflow. Data throughput and delivery times are susceptible to TCP-based bottlenecks that get worse as distances increase, can seriously impact productivity and hamper the need to meet aggressive schedules, time to market and revenue goals.

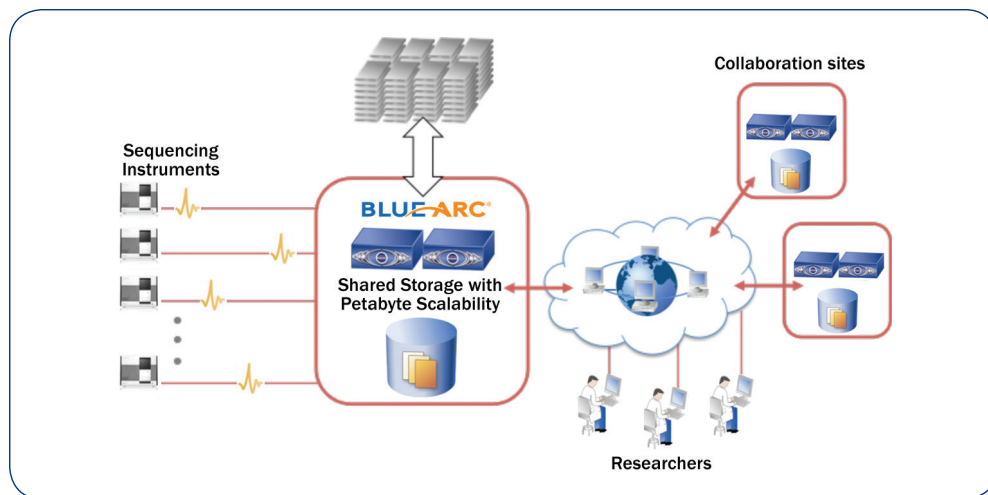


Figure 1. Collaboration and sharing in Genome Sequencing environments

Collaborative research is common in Life Sciences, where geographically dispersed sites within a single organization or a group of organizations share data, results and resources. The data often originates from genome sequencing instruments and go through multiple heavy computational analysis stages to extract the desired information, resulting in a wide range of file sizes and types. Data and results can be shared at any step in the process depending on how the workflow is set up and how the research facilities are collaborating and sharing resources. Science and methodologies change rapidly and require the infrastructure to be flexible and support changing workload and communication needs.

Aspera's *fasp* transport technology and BlueArc high-performance storage platforms create a formidable combination: lightning-fast data access matches storage performance.

A Fully Optimized 6 Gbps Distance-independent WAN Transfer Solution

Aspera's patented *fasp* protocol eliminates the devastating effects of packet loss and latency that plague standard TCP transport solutions over long distances. *fasp* is capable of exceeding the WAN transfer performance and scalability limits of traditional TCP-based protocols; specifically NFS/CIFS in network-attached storage environments, and FTP, HTTP, SCP and RSYNC in application environments.

Aspera software runs natively on all major operating systems (Windows, Linux, Mac) and leverages existing client/server hardware architectures and networking technologies. Aspera solutions integrate fully with BlueArc's Mercury and Titan storage platforms and suite of optimization tools for data management. With BlueArc's patented Silicon Accelerated File System architecture with its efficient metadata handling, exceptional performance can be achieved from even the slowest disk type or tier of storage, while Aspera *fasp* transport enables distance-neutral data access that scales linearly with WAN bandwidth.

Testing High-Performance Data Storage and Transport

To evaluate BlueArc storage platforms and Aspera *fast* transport technology, a test environment was used to simulated real-world conditions. The primary goal in this testing was to measure the performance of the Aspera and BlueArc solution, and how Gbps transfer rates could be achieved under varying WAN conditions (packet loss and latency combinations). To stress both the storage and transfer protocols, 200+ GB files and a great number of small files averaging 5MB in size were transferred between BlueArc storage nodes. This reflects the typical mix of files often found in Life Sciences related research environments – with data originating from Next Generation Sequencing instruments are managed and analyzed, generating derivative data sets and results of different types.

Joint lab testing demonstrated WAN transfer speeds up to a blistering 6.1Gbps with the wide range of file sizes, structure and types. This throughput was observed over a variety of adverse network conditions, including worst-case global Internet scenarios with very high latency and packet loss (5% packet loss and 300 ms latency). The combined BlueArc/Aspera storage and transfer system is capable of transferring 1TB of data in about 20 minutes, *regardless* of the WAN distance.

The 10Gbps test topology is shown below: real-world WAN conditions (latency and packet loss) were achieved using the Apposite Linktropy 10Gb Ethernet WAN emulator. Commodity network and server systems were used to route data and run Aspera server software (responsible for initiating and managing the file transmission from source to destination). BlueArc’s high-performance network storage platforms served as the send/receive storage nodes to accommodate the high-speed data transfer applications.

The Test Environment

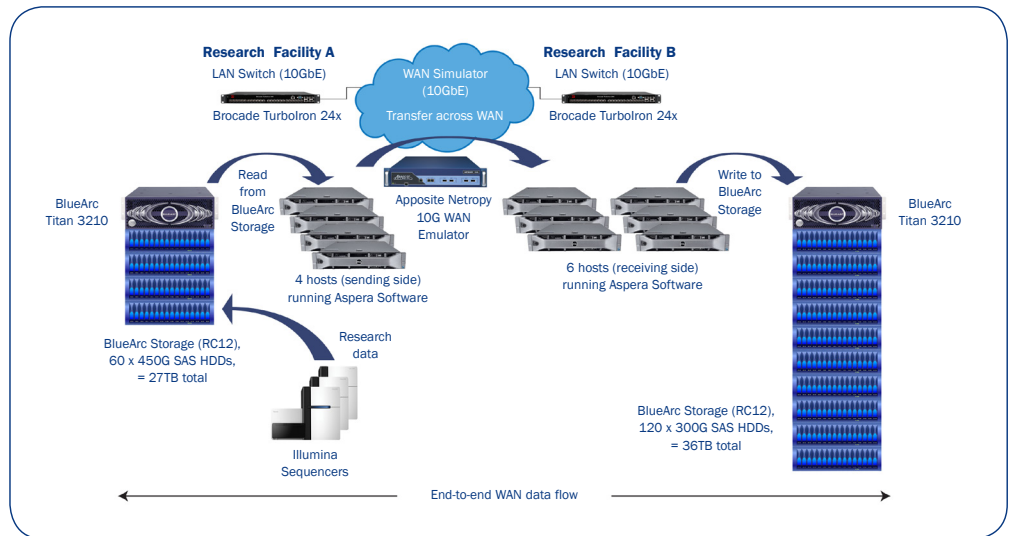


Figure 2: 10Gbps network topology, server, instrument and storage configuration

Testing Detail

All tests were performed over a 10Gbps Ethernet WAN environment. The Apposite Linktropy WAN emulator introduced 0.1%, 1.0% and 5% packet loss, with 10ms, 100ms and 300ms round-trip time (RTT). All packet loss and latency combinations were tested, simulating metro, national and global WAN distances with best-case private WAN performance and worst-case Internet loss. Output performance metrics were used to calculate transfer times, receive/send efficiency and data throughput speeds.

Transfer times and send/receive efficiency. Because standard TCP-based protocols are greatly affected by network latency and packet loss in long distance data transfers, WAN emulation hardware is essential in objectively measuring Aspera's *fasp* performance versus that of TCP-based protocols such as CIFS, NFS, FTP and HTTP.

Throughput. BlueArc storage platform throughput was measured by transferring a data set designed to be representative for multiple next generation sequencing data sets, including the changing characteristics due to changes in instruments and applications:

- multiple data files averaging 5MB in size and totaling 203GB
- multiple data files averaging 5GB in size and totaling 209GB
- multiple data files averaging 250MB in size and totaling 1TB
- multiple data files averaging 5MB in size and totaling 1TB
- individual 201GB data files

File types and sizes varied from large image files, over intermediate analysis results to smaller text files. In each case, the files were segmented and transferred using 96 concurrent sessions, demonstrating the storage platform's ability to handle IOPS-intensive data sourcing.

The Results: An average of 5Gbps or more for all WAN combinations

BlueArc storage transfer throughput over 10GbE WAN under moderate packet loss and latency conditions reached a sustained 5.5Gbps. Under worst-case Internet packet loss and latency conditions (5% loss / 300 ms latency), throughput was sustained at 4.3Gbps.

At these throughput rates, transferring 1TB over any metro, national or global WAN would take a little more than 20 minutes. Moving the same amount of data using NFS, CIFS, FTP, HTTP, RSYNC, and SCP—would take 24 hours, under the best-case WAN condition. Under worst-case network conditions, it is likely that the same transfers would never successfully complete.

The diagram below shows the Aspera/BlueArc performance compared to TCP-based protocols, over the different WAN latency and packet loss combinations. The performance of Aspera's *fasp* is nearly linear over all WAN conditions, yielding a near 3000x improvement over TCP.

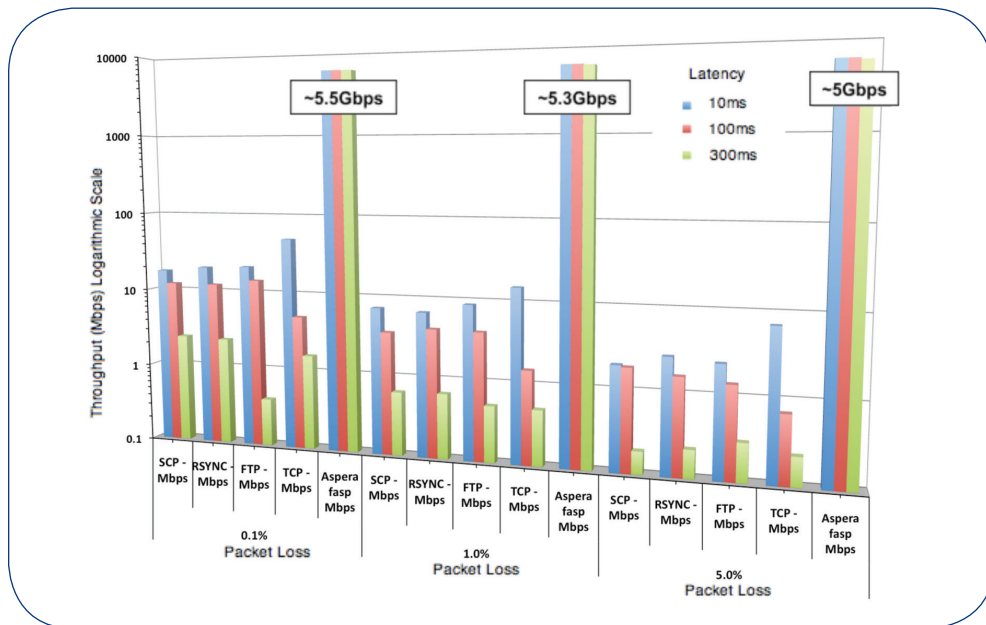


Figure 3: 10Gbps BlueArc storage WAN throughput: Aspera *fasp* compared to TCP-based transport solutions.

Summary

In a variety of enterprises and research organizations, ranging from life sciences, financial services, and high-performance computing to media and entertainment, incredible amounts of data need to be stored, exchanged and distributed quickly, predictably and reliably. Legacy file transport and storage technologies cannot keep up with growing file sizes, high-speed WAN deployments and the need for global, dynamic data access.

BlueArc storage platforms allow organizations to incorporate their existing storage infrastructure in a dynamic, tiered high-performance storage environment that can be scaled to accommodate future growth. Aspera's *fast* ultra-high-speed transfer technology scales to take full advantage of any WAN bandwidth, independent of distance or packet loss. Using existing infrastructures, this breakthrough technology combination enables massive volumes of information to be accessed, managed and moved on a global basis, at speeds that meet the mission-critical nature of today's data access.

Deployed in conjunction with Aspera software, BlueArc storage delivers real solutions that allow organizations to meet the demand for data access and ensure business, time to market and revenue goals are met.

TEST SYSTEM CONFIGURATION

- BlueArc storage node
- File system consisted of 120 LSI 300GB, 15,000-rpm SAS write drives, RAID 5+1
- One file system and one storage pool consisting of 120 LSI RC12 300GB, 15,000-rpm SAS drives using RAID5 5+1@64K volume groups with 20 total
- 16 mount points per client (16 independent TCP connections between each storage client and the storage server)
- Linux servers running CentOS release 5.5 (Final) and Red Hat Enterprise Linux Server release 5.5 (Tikanga):
 - One Dell R710 with x2 Quad-Core Xeon 2.5 GHz processors, 24 GB RAM;
 - 2 SuperMicro Intel x5680 @ 3.33 GHz 2 x6 core 48 GB RAM
 - 1 SuperMicro Intel x5420 @ 2.5 GHz 2x4 core 16 GB RAM
- Dual-port 10GbE NIC per host server
- Apposite Linktropy 10G WAN Emulator
- Brocade TurboIron 10GbE switch with VLAN configuration
- FC connectivity to the LSI storage array
- x1 file system and one storage pool consisting of 60 LSI RC12 450GB, 15,000-rpm SAS drives using RAID5 5+1@64K volume groups with 10 total
- x1 10TB file system, 24 mount points per client (24 independent TCP connections between each storage client and the storage server)

CLIENT SYSTEM CONFIGURATION

- BlueArc storage node
- x6 Linux servers running CentOS release 5.5 (Final) and Red Hat Enterprise Linux Server release 5.5 (Tikanga)
- x2 Dell R710 with 2 Quad-Core Xeon 2.5 GHz processors, 24 GB RAM
- x2 SuperMicro E5640 @ 2.67 GHz
- x2 SuperMicro E5540 @ 2.53 GHz
- Dual port 10 GbE NIC per host server



WAN EMULATION CONFIGURATION

- Brocade TurboIron 10GbE switch with VLAN configuration
- Apposite Linktropy 10G WAN emulator

ASPERA SERVER SOFTWARE

- Aspera Enterprise Server 2.6

About Aspera

Aspera is the creator of next-generation transport technologies that move the world's data at maximum speed regardless of file size, transfer distance and network conditions. Based on its patented *fast* protocol, Aspera software fully utilizes existing infrastructures to deliver the fastest, most predictable file transfer experience. Aspera's core technology delivers unprecedented control over bandwidth, complete security and uncompromising reliability. Over one thousand organizations across a variety of industries on six continents rely on Aspera software for the business-critical transport of their digital assets. Please visit <http://www.asperasoft.com> for more information.

About BlueArc

At BlueArc, we know that lightning fast storage and large capacities in a standards based file storage environment enables new research and collaboration. In Life Sciences and Research we deliver the performance and scalability needed to produce results quickly and handle an ever increasing flow of data and analysis results. With methodologies and applications changing much quicker than IT infrastructures, we deliver storage solutions that are architected from the start to thrive in those situations. BlueArc provides industry-leading file storage at scale with the simplicity for the end user of having a single, global namespace and performance to handle the demands from researchers. While also providing the control, flexible architecture and scalability needed to optimize for cost and efficiency.

About BlueArc

BlueArc is a leading provider of high performance unified network storage systems to enterprise markets, as well as data intensive markets, such as electronic discovery, entertainment, federal government, higher education, Internet services, oil and gas and life sciences. Our products support both network attached storage, or NAS, and storage area network, or SAN, services on a converged network storage platform.

We enable companies to expand the ways they explore, discover, research, create, process and innovate in data-intensive environments. Our products replace complex and performance-limited products with high performance, scalable and easy to use systems capable of handling the most data intensive applications and environments. Further, we believe that our energy efficient design and our products' ability to consolidate legacy storage infrastructures, dramatically increases storage utilization rates and reduces our customers' total cost of ownership.



BlueArc Corporation
Corporate Headquarters
50 Rio Robles Drive
San Jose, CA 95134
t 408 576 6600
f 408 576 6601
www.bluearc.com

BlueArc UK Ltd.
European Headquarters
Queensgate House
Cookham Road
Bracknell RG12 1RB, United Kingdom
t +44 (0) 1344 408 200
f +44 (0) 1344 408 202