Bioscience

Introduction

Over the past several years there has been in a shift within the biosciences research community, regarding the types of computer applications and infrastructures that are deployed to sequence, analyze, compare, and model gene information. Today there exists large petabytes of already sequenced data that are contained in publically and privately available databases. Many researchers are leveraging these databases for further analysis, whether it is to explore specific mutations, model biological systems, explore the side effect of new drugs, or make protein structure predictions.

This large growth in already sequenced data, and the mining of this data for further research is driving the need for state-of-the-art computing clusters. And while state-of-art computing clusters have been leveraged for decades within life science, the growth of highly distributed databases, increased time to market pressures, and tighter R and D budgets, are driving fundamental changes.

These changes include an increased dependency on distributed storage, fast localized search technologies, outsourcing to external hosting providers, and operational excellence specific to 7x24 uptime and green energy efficiencies. Moreover, these clusters are purpose built for many of the open source and commercially available applications within this market including BLAST, PBLAST, KEGG, GSEA, Brita etc. These applications benefit from several network optimization and data transfer technologies currently available.

High speed network switching in which traffic is switched within nanoseconds per switch node (port-to-port) and in microseconds end-to-end (server to server, server to storage) is the most cost effective and proven technology for bioscience research clusters.

The Importance of the Network

Today's state-of-the-art clusters need to move large volumes of data at high speeds, with ultra low transfer times, between server nodes, storage nodes, and often across the Internet. Searches are requiring increased topology and data intelligence for minimizing traffic and reducing search times. Data transfers between server and storage nodes, and any combination there-of, require ultra low latency and high network bandwidth for completing a variety of research jobs. Stability, reliability and upgrading are also important to insure 7x24 availability for the next job in queue. And the ability to scale network bandwidth bisectionally as more server and storage devices are added (new capacity) is also paramount to insure investment and growth protection. In a word the network is one of the more critical asset for delivering on these requirements.

The following summarizes many of the networking requirements

Network Switching Requirements

High speed network switching in which traffic is switched within nanoseconds per switch node (port-to-port) and in microseconds end-to-end (server to server, server to storage) is the most cost effective and proven technology for bioscience research clusters. The following outlines six on the most important functions of networking switching technologies.

1. *Server-to-Server Communications*: Network switching interconnects servers together when there are high computation requirement and the computation can be completed much faster when processed across many hundreds of server nodes. This is the most common bioscience server architecture, and the most well known role of switching today within this market. What is new over the last three years is the growing adoption of multi-core, multi-socket servers, and the amount of data processing these servers can handle. These multi-core servers can easily over subscribe 1 Gbps network interfaces.

2. *High Throughput, Low Latency*: Life science applications are latency-sensitive (tolerate little delay). Additionally, life science applications require a lot of movement of data between compute and storage nodes. Network switching must be optimized for ultra fast delivery with leading edge transport speeds (10 Gbps and above). This requires specialized cut through switching architectures and fine-tuning with many of biosciences application specific driver technologies.

3. *Server-to-Storage Communications*: Network switching interconnects servers to storage, especially as storage becomes distributed and parallel, with a growing need for very fast retrieval, search, and read/writes in out of the databases storing petabytes of data. The majority of network traffic is no longer isolated between servers; increasingly traffic is between servers and storage thus driving the need for greater bandwidth aggregation (bi-sectional) and seamless movement of packets between these two resource classes.

4. Scalable Bandwidth: Network switching must offer a very scalable architecture. As large amounts of data are moved across the network, there are aggregation points in which the communication crisscrosses. These aggregation points must offer the ability to add more link, failover and forwarding capacity as more end node capacity is added. This is often referred to as the ability to scale bandwidth bisectionally.

5. 7x24 Uptime: Researchers within the biosciences market, whether academic, government or private community-related are under increased pressure to produce results. Wait times for results are now measured in hours, weeks and months and not in years. And there is always the next job in queue with little to no expected downtime. Network switching must offer 7x24 uptime reliability, not only from a failover and recovery high availability requirements, but operationally, when there is maintenance required, whether it is adding a new line card, a new server/storage rack, or there is a need to perform software upgrades. The cluster must remain fully operational.

6. Open Programmatic Interfaces: A majority of the biosciences applications, GUI's and tools are based on open source software technologies and programs. Many academic and private research communities are adding their own algorithms, interfaces and/or specialized utilities unique to their projects. Unfortunately, network switching traditionally has had closed proprietary interfaces with no ability to modify the switching behavior unique to these highly purposed applications, especially those related to very targeted search and database retrieval requirements, such as the case with big data Hadoop clusters. Network switching platforms must offer open interfaces for programming specific traffic flows, providing management data, or assisting in search functions.

Network Cost Considerations:

As with most industries requiring IT technologies, bioscience researchers have tight budgets and need to operate within the constraints of these budgets. Network switching needs to affordable, manageable, and operationally efficient. Network switching costs need to be in line with the server and storage acquisition costs. Equally important network switching should be easy to manage, and very efficient in terms of the monthly power, cooling and general maintenance costs (operation costs).

This drives the need for a common network architecture that can equally move large data packets between servers, between servers and storage, and between storage for back-up and data recovery operations. Any speed, protocol, packet, or network management conversions resulting from different networking technologies between these resources will significantly impact performance while adding unwanted complexities. As a result the most optimized network is one that is common across all resources within and across clusters.

High speed Ethernet switching, starting at 10 Gbps today, with 40 Gbps aggregation capability, and 100 Gbps in the next 2-3 years offers the best choice in this market. High speed Ethernet switching has been optimized with bioscience specific server-to-server driver enhancements near equal to the ultra-low latency delivery capabilities of InfiniBand.

All server and storage suppliers offer 10 Gbps Ethernet interfaces on their chassis with many of these vendors including these interfaces as part of their base system. This insures an end-to-end high speed Ethernet switching architecture without the need for specialized gateways. And there have been significant price drops with 10 Gbps switching, routing, and server/storage node interface costs. An average 10Gbps Ethernet switching is 2X the price of 1 Gbps switching, with 10 times the bandwidth, and exponential improvements in latency. This offers a 5X benefit minimally. Latency delays with 10 Gbs switching are now being measured in nano seconds, in comparison to milliseconds 5 years ago.

For those concerned operationally, from a power, cooling, rack density, and management end, there have been substantial improvements in 10Gbps transceiver, cabling and switching chip technologies. Researchers can achieve rack densities of up to 64 ports per 1 RU rack slot, can budget 2 watts per port, and can write management utilities (either in our out of band) with many of the latest open source tools (Python, Bash, XML, NetConf). Moreover, server and storage admins can leverage the centralized switching intelligence for re-booting failed servers via PXE boot, and for performing intelligent database search functions as integrated with Hadoop clustering technologies.

Biosciences Server, Network, and Storage Design Considerations

The following should be considered when choosing high speed Ethernet switching for bioscience research clusters.

Server Interface Capacity Considerations

As stated in previous sections a majority of research clusters are based on X86 distributed server architectures, with ultra low latency networks that interconnect the servers together. This approach offers the most cost effective approach for computational intensive applications.

High volume X86 servers, on average, have 8 CPU's per node (2 CPU's on average, with 4 CPU cores) and are growing to between 16 to 32 CPU's per node over the next 2-3 years. This increased CPU density drives the need for greater networking capacity (I/O) where a 1 Gbps Ethernet interface is no longer sufficient for moving

Top Networking Challenges and Solutions

Challenges	Solutions
Current server I/O interfaces insufficient for increased CPU density and processing power.	10 Gbps Ethernet server interfaces offer best price/performance option for high performance multi-core X86 server applications.
Highly distributed bioscience applications consume precious server CPU and memory resources when transferring datasets inefficiently between compute nodes.	Optimized 10 Gbps Ethernet server driver technologies with specialized network interface cards (NICs) offload CPU and memory processing from the server. These NIC offload cards also perform memory-tomemory transfers within the application user space. Significant server performance enhancements equal to InfiniBand are achieved via these interface technologies.
Growing dependencies on the re-use of sequenced data, the need for fast read/write storage access, and fast search technologies that can quickly scan petabytes of stored information.	High speed 10/40 Gbps data center class switching significantly improves search, write and retrieval times for Hadoop clusters and centralized network attached storage systems.
General-purpose data center network switching chassis not designed for ultra low latency applications.	Use of 10 Gbps Ethernet cut through switching technologies, especially between server peers addresses the ultra low latency switching needs (600 nano seconds port to port) required for optimizing the completion times of running large computation centric clusters.
Bisectional network bandwidth in sufficient to meet the bioscience scalability requirements of large 1000 node plus compute and storage clusters.	Leaf/spine Ethernet network switching designs, with link aggregation, deep traffic buffering, load balancing, and the option of layer-2 or layer-3 forwarding address the traffic scalability requirements and 1000 node plus clusters.
Around the clock cluster availability based upon increased time to completion research and commercial completion goals.	Cost efficient high availability cluster and network design with no single points of failure. These designs are well supported with Ethernet switch redundancy, hitless software upgrading, proactive monitoring, and the ability to boot servers from a centralized network control point.
Closed proprietary interfaces restrict any application customization/optimization.	Modular Ethernet switch operating system, based on a non-modified Linux kernel, with open API's allows network and application engineer's to modify switch behavior specific to their application needs.

bioinformatics data in and out of the server equal to the rate in which these multi-core servers can process the data. A 16 core server (2 processors with 8 cores each) minimally requires 2.5 Gbps per second I/O data rates, and can easily burst to 10 Gbps depending on how distributed the applications and search requirements are.

Researchers should be specifying 10 Gbps Ethernet interfaces with any server that is multi-core, especially if they are using clustered applications that scale horizontally across many compute and storage nodes. There will be billions of data transfers in/out of each server. In 2012, many multi-core servers will include as part of the standard offering, 10G baseT interfaces. In other cases, the researcher may need a 10 Gbps Ethernet plug in option, as the server may not have this as part of the base system, and/or they made need several I/O offloading capabilities, in which the plug in Network Interface Controllers (NICs) help process the transfer of data in and out of the server at the application layers.

Low Latency Performance Optimization

Many of the biosciences applications move large amounts of data between servers and storage. This data movement is very latency sensitive. It has been proven in several tests (see Intel white paper) that when the transfer and latency rates are reduced by milliseconds per transaction, given the billions of transactions that take place between these resources when sequencing data, that the overall job completion times drop exponentially.

Unfortunately, these latency sensitive data transfer requirements place a lot of overhead on the server CPU and the server memory to process. As a result the server can be bogged down in processing the data transfers to and from the network interface, up through the operating system, and to up the application layers. This can have a negative impact on server and network performance.

To overcome this overhead, several network drivers have been developed to optimize the data transfers and reduce latency. These drivers are tuned for applications within the bioscience market. These drivers include Remote Direct Memory Access (RDMA), Message Passing Interface (MPI), and iWarp (RDMA over Ethernet). Additionally, these drivers have been designed to run on 10 Gbps Ethernet plug-in network interface card within the server, thus offloading these data transfer and data communication tasks from the server CPU and memory. These drivers are standards based, have been tested by several well-known server interface providers, are compatible with many open source commercial available Linux operating systems, and have been proven to significantly reduce the latency and processing within the server.

NetEffect Server Adapters from Intel: Low Latency for High Performance Clusters

High performance cluster computing has made it possible to achieve breakthrough advances in research areas such as bioinformatics, computation chemistry, and gene sequencing. These large-scale clustering workloads demand an efficient fabric, such as iWARP, that is capable of providing low-latency and network scalability.

iWARP delivers converged, low-latency fabric services through Remote Direct Memory Access (RDMA) over Ethernet. The iWARP specification, maintained by the Internet Engineering Task Force (IETF), supports transmissions over TCP and is implemented on top of IP networks using an existing Ethernet infrastructure.

These key iWARP components deliver low-latency: *Kernel Bypass*. Placing data directly in user space avoids kernel-to-user context switches, which add additional latency and consume CPU cycles that could otherwise be used for application processing.
Direct Data Placement. Data placed directly in application buffers rather than being copied multiple times to driver and network stack buffers frees up memory bandwidth and CPU compute cycles for the application.
Transport Acceleration. Transport processing performed on the network controller instead of the host processor frees up valuable CPU cycles for application compute processing.

NetEffect Server Cluster Adapters from Intel use iWARP to virtually eliminate processor overhead associated with Ethernet networking and deliver a fabric suitable for high performance clustering networks. By addressing the key sources of Ethernet overhead, NetEffect™ Ethernet Server Cluster Adapters from Intel deliver these benefits:

Fabric consolidation. With iWARP technology running on Ethernet infrastructures, cluster-specific fabrics and related cables, switches, and adapters can be eliminated.
IP-based management. Network administrators can use standard IP tools to manage traffic in an iWARP network, taking advantage of existing skill sets and processes to reduce the overall cost and complexity of operations.
Native routing capabilities. Because iWARP uses Ethernet

Distributed File System for Improved Performance

Biosciences applications are growing in complexity. Researchers are now beginning to leverage datasets that have already been sequenced, globally. This requires fast access into petabytes of already processed data, with localized searches, localized analysis, and fast writes back into the database. Moreover, rather than moving the data where the processors reside, many new designs are deploying Hadoop clusters; these cluster designs move the database searches and computation closest to where the data resides within storage, and where a tree search can be run locally. This requires a parallelization of the data, storage, and analysis algorithms. Hadoop technologies, either open source or through several commercial providers, is becoming the preferred approach for complex data searches.

From a networking perspective, 10 Gbps Ethernet switching offers the best performance, price, and broadest technology adoption benefits for connecting storage arrays, and cluster file systems together. All of the major storage vendors now offer their storage devices with standard 10 Gbps Ethernet interfaces. And Hadoop with MapReduce, as open source technology can be tightly integrated with Ethernet switching, for discovering database search tree structures and more intelligently placing the searches closest to where the data resides. This requires open interfaces that can be programmatically access specific to topology data, and best path forwarding tables.

Since much of the research going forward within the biosciences market will be based on database searches, researchers should be specifying Hadoop clustering technologies as part of their application environment, and should ensure that their 10 Gbps Ethernet switches offer open interfaces for providing topology and best search placement information.

Scalable, Redundant Leaf/Spine Networking Architecture

Scaling to thousands of server and storage devices within a bioscience compute cluster requires a highly efficient, low latency, fully redundant switch networking design. In this design the number of switch layers between one node and another should be kept to a minimum as this reduces latency. Further, there should be redundant links within the backbone layer to insure no single point of failure, and to also load balance traffic between these redundant links as way to increase bandwidth and throughput (bisectional bandwidth scaling). And for applications that require latency measured within nanoseconds, the switches should be able to operate in cut through mode when locally switching within the same server rack.

EMC² ISILON®

Isilon Scale-Out NAS Storage Platform

Enabling Greater Research Productivity Dramatic advances in next gen sequencing technologies; ever-greater microscopy resolutions and a many-fold increase in the number of medical imaging devices are driving unprecedented growth of critical filebased research data. Isilon scale-out NAS solutions dramatically increase the productivity of mission-critical life science and bioinformatics research applications and workflows while driving down complexity and cost.

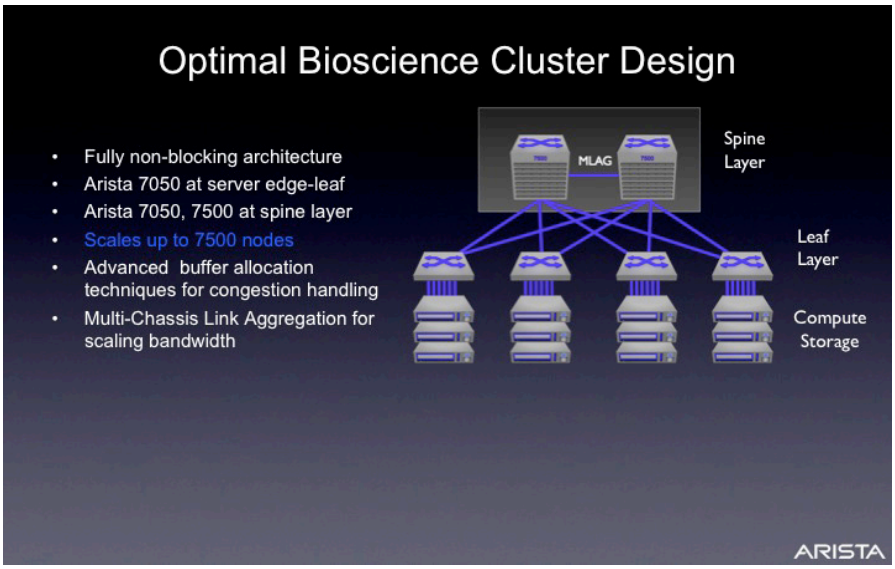
Making a High-Growth Environment Simple & Powerful Life science and bioinformatics organizations pursue their research objectives in myriad ways, but their storage challenges have much in common. They require massive scaling headroom, significant storage performance, accommodation of various file types and access patterns, and multi-protocol accessibility. And most of all, as organizations often perceived as "data rich" but "IT-light", they need their large-scale storage to be simple. Simple to buy, build, maintain, and grow.

Isilon's life sciences and bioinformatics customers benefit from Isilon Scale-out NAS ability to:

- Easily and non-disruptively scale a single file system/single volume on a "pay-as-you-grow" basis from 10 terabytes to over 15 petabytes.
- Leverage high speed, low latency networking technologies between compute and storage
- Provide simplicity and ease-of-use that extend the productivity of one FTE to multi-petabyte scale.
- Accommodate a wide range of file types and access patterns with multi-protocol network transports
- Enable raw storage utilization rates to over 80 percent.
- Reduce technology "lock-in" risk with multi-protocol access options.
- Facilitate massive concurrent read/write access.
- Drive productivity in a wide range of workflows-from the highest performance HPC environments to massive data archives.

Complete and Cost-Effective Scale-out NAS for Maximum Workflow Productivity

Isilon scale-out NAS solutions, designed specifically to meet the challenges typical of life science and bioinformatics organizations, offer the only "pay-as-you-grow" model for adding storage capacity quickly and easily. Arista's high-speed Ethernet switching platforms offer a perfect compliment for this pay-as-you-grow storage solution.

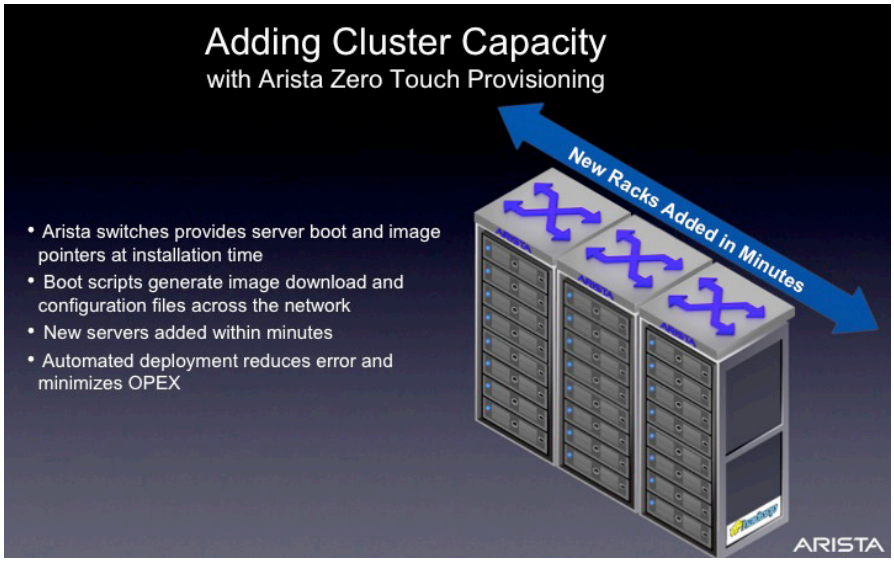


A leaf/spine topology design for optimizing many of the above requirements (see diagram) is recommended. This design places at most three switch hops between any two end points. Additionally, there should be redundant load balanced links between the leaf and the spine. This includes multi-link aggregation (MLAG) for eliminating single points of failure, and for utilizing all of the bandwidth in a non-blocking layer-2 or layer-3 topology design. When specifying the topology design, researchers should consult with their IT organization regarding the number of layer-3 addresses available to consume. If addresses are limited and restricted they are better off scaling with layer 2 topologies. If IP address pools are widely available then a layer-3 design is recommended. To insure maximum flexibility Ethernet switches should provide both layer 2 and layer 3 forwarding options.

At the aggregation/spine layer, modular based switching is recommended as this addresses the need for high port density (bisectional bandwidth) with port to port switching across a non blocking switching backplane. This is better than adding many smaller chassis as each additional smaller chassis adds another latency hop, and increases the number of devices needing to be managed.

Network Operation Efficiencies

For many IT organizations the cost of operating a data center is growing at a faster rate, than the upfront costs of purchasing and installing the server, network, and storage equipment. Over 80% of the total cost of ownership is based on monthly re-occurring costs including data center rack space, power, cooling, on going maintenance, redundant systems, and system and support contracts. Many researchers over look these costs as they have never operated, managed or been charged operationally for their compute clusters. As the operation costs continue to rise, more and more the consumers of these clusters will get charged monthly for managing these clusters, whether via their own staff, or cross charged by their IT organization. As such it is imperative to consider these costs during the design and equipment specification/acquisition stages.



Researchers should specify networking switches that offer high port densities, both within fixed top of sever rack 1 RU form factors, or within modular plug in chassis. Further researchers should consider factors such as power consumption, compatibility with their server, storage and data center cooling plans, and they should consider the cost of cabling and what offers the best option based upon distance, performance and cost. And researchers should consider the amount of 7x24 uptime required.

Conclusion

While there is a significant shift from gene sequencing to many different types of analysis, modeling, and comparing within the several large petabytes of stored sequence data, the computing, network, and storage needs to process and produce results continues to grow exponentially. Moreover, there is increased pressure especially within the private sector to produce results faster, file a patent and get to market with new treatments whether therapeutic or drug related. It is well known that first to market yields high margins and profits for a number of years thereafter.

- The network, and the network I/O is taking on a increased role by performing the following
- Server off load functions
- Switching in ultra low latency cut through mode
- Providing topology intelligence on where to perform the best database search operations
- Offering open interfaces for application forwarding customization
- Offering bandwidth scalability via link aggregation
- Insuring 7x24 uptime with high availability features
- Minimizing the cost of operations via low power consumption, high chassis and line card densities
- Offering easily patched, non-operationally disruptive, upgradeable software. paragraph spacing.

Santa Clara—Corporate Headquarters
5453 Great America Parkway,
Santa Clara, CA 95054

Phone: +1-408-547-5500
Fax: +1-408-538-8920
Email: info@arista.com

Ireland—International Headquarters
3130 Atlantic Avenue
Westpark Business Campus
Shannon, Co. Clare
Ireland

Vancouver—R&D Office
9200 Glenlyon Pkwy, Unit 300
Burnaby, British Columbia
Canada V5J 5J8

San Francisco—R&D and Sales Office 1390
Market Street, Suite 800
San Francisco, CA 94102

India—R&D Office
Global Tech Park, Tower A & B, 11th Floor
Marathahalli Outer Ring Road
Devarabeesanahalli Village, Varthur Hobli
Bangalore, India 560103

Singapore—APAC Administrative Office
9 Temasek Boulevard
#29-01, Suntec Tower Two
Singapore 038989

Nashua—R&D Office
10 Tara Boulevard
Nashua, NH 03062