

## **The Power of InfiniBand and Clustered Storage**

December, 2005



It is official: InfiniBand® (IB) is on the rise in the corporate data center. Once restricted to exotic supercomputing deployments, we now see IB making a solid push into enterprise computing environments across a broad range of industry verticals. While still in its early stages of adoption, enterprise IB is already delivering powerful results in the realm of enterprise clustered storage. Because clustered storage solutions rely on communications across large numbers of atomic nodes federated together, the appeal of a high-speed interconnect like IB has long been obvious. Until recently, though, no storage player had found a way to cost-effectively architect an IB-based storage cluster that could scale into the 100s of terabytes and beyond. All that changed earlier this year when Isilon® Systems emerged as a clear leader in the IB-based clustered storage market with a screamingly fast and scalable IB-based version of its Isilon IQ clustered storage systems. The kicker: Isilon priced their IB offering the same as their GigE version! Based on our analysis and discussions with Isilon customers, the benefits of an IB-based storage cluster are too compelling to pass up: massive end-to-end performance gains, superior scalability, and extremely competitive ROI. Isilon's forward-looking move proves that while InfiniBand has long been the darling of the research and HPC communities, it has now become a viable mainstream enterprise technology. For a deeper look at what InfiniBand will be doing in your data center in the coming years, we encourage potential customers to understand what Isilon is doing today with their IB-based Isilon IQ platform.

### **InfiniBand and the Enterprise**

If you are running an enterprise data center in 2006, why should you even care about InfiniBand (IB)? For the past five years, the answer *has* been that, in most cases, you did not need to concern yourself with this high-speed interconnect. InfiniBand was originally designed to provide servers with a new high-speed serial I/O interconnect that would replace the aging shared bus architectures of the ISA and PCI paradigm. Alas, after a late 1990s cycle of innovation, hype, and politicking, IB had been effectively relegated to the niche world of high performance computing environments, where its impressive transport capabilities were both

needed and appreciated. Meanwhile, all of us in the enterprise world essentially forgot about the promise of the InfiniBand interconnect. In fact, we can honestly say that until early 2005, nobody had identified an enterprise usage case for InfiniBand that uniquely solved an immediate burning IT issue, deployed easily, and met enterprise standards for cost-effectiveness.

This has all changed and changed quickly. Indeed, we believe that you *should* now care about InfiniBand because it will be coming on strong in your enterprise data center for the foreseeable future. And not surprisingly, InfiniBand is making its strongest beachhead at the storage layer. Specifically, we believe

## TECHNOLOGY BRIEF

that InfiniBand is on its way to becoming a powerful new standard for interconnecting the dozens, hundreds, or thousands of atomic storage nodes that comprise the most performant clustered storage infrastructures.

Because the storage nodes within a clustered environment are essentially thin servers with internal storage federated together into a unified cluster, it should come as no surprise that InfiniBand is being applied to the task. To answer our own question posed above: *We believe that you should care about InfiniBand because it will accelerate and improve your storage infrastructure, increasing enterprise collaboration and productivity.*

To understand why InfiniBand is gaining traction in the world of enterprise clustered storage, it makes sense to first take a look at what makes InfiniBand so different as an interconnect technology.

### **InfiniBand Unique Differences**

InfiniBand is typically defined as “a high-speed, low-latency, serial I/O interconnect for computing resources.” However, because of the very robust nature of IB and its unique design approach, its implications for the data center are far greater than those of any typical ‘interconnect’ and therefore merit deeper exploration.

#### **InfiniBand: Serial and Switched**

Most critically, the serial nature of IB enables bi-directional communication that avoids signal skew challenges common in parallel interconnects. Also, this serial bus approach lends itself readily to “siliconization” of the

entire IB bus architecture, a critical advance for enabling a high-speed Infiniband *switched* fabric where multiple Infiniband devices are sharing the same network I/O.

Infiniband utilizes channel-based connectivity, enabling very flexible point-to-point connectivity between any resources residing on the same IB fabric. Serialization, channel-based connectivity and a switched topology all distinguish IB from the much slower and more rigid shared bus PCI architecture that has become a standard over the past 15 years.

#### **InfiniBand: Unleashing Server Power**

The implications of InfiniBand are significant for server deployments: by leveraging dedicated silicon Host Channel Adaptors (HCAs) deployed on servers, InfiniBand-enabled resources effectively offload server I/O capacities into a shared I/O fabric. This creates several advantages: it frees the servers’ CPU cycles to do processing and, equally important, enables I/O to be freed from dedicated server relationships. Thus, Infiniband decouples the historically inefficient marriage of server processing power and I/O capacity. With IB fabrics, servers can be added to scale up based on pure processing power while shared I/O capacity is held constant in the switched fabric (or vice versa, if so desired.) This kind of interconnect flexibility is totally unprecedented in open systems high performance computing.

#### **InfiniBand: High Performance**

Of course, the real power of InfiniBand is the performance gains that it enables as a result of this architectural approach. Current IB

## TECHNOLOGY BRIEF

raw signaling rates for a standard architecture are 30 Gbps, tracking soon to 120 Gbps over copper. This kind of speed represents a 10 to 100x gain over standard 1 Gigabit Ethernet or 10 Gigabit Ethernet deployments.

Another major difference and an enabler of IB performance is the interconnect's support for RDMA (Remote Direct Memory Access). With RDMA, the computing resources in an InfiniBand fabric can communicate directly between respective server memories without impacting CPUs or caches. This kind of parallelization is very useful in applications that require high throughput, such as those commonly found in large node-count clusters.

While already fast, the roadmap for IB speed increases over the coming years is quite awe inspiring. Some researchers foresee a point within a few years where IB interconnects will be operating at close to internal server speeds, effectively making the interconnect a transparent "ultra-speed" bridge between discrete processors.

Because of all of these factors, we are confident that InfiniBand will continue to gain traction in the enterprise market.

### **The Path to IB Storage Clusters**

Because of its flexibility and raw power, InfiniBand spent a fruitful early adoption stage (1999-2004) driving High Performance Computing (HPC) deployments, enabling large collections of smaller computers to be grouped into a common resource for solving large computational problems. IB has

received high technical kudos in this role for several reasons: it eliminated the need for less efficient proprietary server interconnects, it demonstrated great scaling capabilities, and it clocked I/O speeds that were otherwise impossible to achieve.

### **Rise of Enterprise Clustering**

While IB continues to expand its presence in HPC, a parallel trend has taken hold over the past 18 months; a trend that ultimately brought IB into a new market opportunity altogether. Specifically, the enterprise community has become seriously interested in clustering technologies. Driven by thin server scale-out architectures and advanced file system technologies, enterprises are now thinking about how larger numbers of smaller components could both drive performance and improve economics. And within that general trend, enterprises have begun to pay particular attention to *clustered storage* offerings.

### **The Rise of Clustered Storage**

A form of clustered computing, enterprise clustered storage exists *solely* to provide high-speed, high scale, protected storage capacities to its client machines. Similar in principle to an HPC environment, a clustered storage solution harnesses the power of dozens or hundreds of individual servers to create one large, flexible storage pool, typically exposing NAS protocols (NFS, CIFS) to the clients it supports. Some of the defining characteristics of enterprise clustered storage are:

- Symmetric Clustered Architecture
- Scalable Distributed File System
- Inherent High Availability
- Unified Cluster Management

**T E C H N O L O G Y B R I E F**

- **Linear Scaling Performance**

As a clustered storage system adds new nodes, CPU and storage capabilities then scale with it. Clustered storage solutions have received interest from enterprise environments requiring maximal end-to-end throughput for rapidly scaling capacities. Clustered storage is being deployed across a wide range of industries, including life sciences, oil and gas, media and entertainment, manufacturing, government, universities, and Internet services. It is not uncommon to see customers in this market category desiring *file system sizes* ranging from the dozens of terabytes to over 100 terabytes in size, all without sacrificing performance or reliability.

At Taneja Group, we have seen clustered storage solve data challenges that are prohibitively expensive or simply impossible to manage for traditional storage architectures. There is no question that clustered storage addresses a burgeoning market with needs that can not be met by traditional NAS and SAN storage architectures.

### **InfiniBand, Meet Clustered Storage**

At this point, it should be obvious where InfiniBand enters the clustered storage equation: *By replacing the GigE node-to-node cluster interconnect with InfiniBand, a clustered storage solution can create an even faster, more scalable storage platform.* In fact, if architected correctly, such a clustered storage solution would likely eclipse the best possible storage performances and scalability ever achieved.

While this potential vision of an IB-based storage cluster has been conceptually understood for several years, design issues and cost limitations have prevented any vendors from making the leap. That is, until now. Isilon Systems, based in Seattle Washington, has delivered a true InfiniBand-based clustered storage solution. We feel it merits a deeper inspection, given the impact this technology is likely to have on both the storage industry and enterprise data center operations in coming years.

### **Isilon IQ Goes InfiniBand**

The Isilon IQ clustered storage solution has been getting high marks in the storage industry for several years now. The purpose of this brief is not to provide details on that well regarded offering, but to talk instead about a unique move that Isilon recently made relative to its Isilon IQ product line.

In April 2005, Isilon Systems announced that their entire Isilon IQ storage product line would be available in an InfiniBand-based version, replacing the traditional Gigabit Ethernet node-to-node connectivity that traditionally drove their offerings. All Isilon customers now have a choice as to whether they want Gigabit Ethernet or InfiniBand. And perhaps most importantly, Isilon has made their InfiniBand version of IsilonIQ available at the *same price as GigE*. The result: Isilon reports an InfiniBand adoption rate of over 90% in its customer base. The reasons are not hard to comprehend, as we will now explore.

## **Isilon's InfiniBand Architecture**

From an architectural standpoint, it is not surprising that Isilon would be the first high-scale/high-performance storage provider to move to an IB-based offering. Recall that the Isilon IQ platform is entirely based on the assembly of atomic membership nodes that each adds incremental CPU, memory, bandwidth and storage capacity to one single expandable storage pool. This type of modular clustering approach is as close to the ideal "InfiniBand-friendly" server/storage fabric as any that could exist in the storage industry today.

The only thing that changes in the IB version of Isilon IQ is that the customer also deploys small 1U *Intracuster InfiniBand Cluster Switch* (see diagram, page 8) that drive the IB-switched fabric communication between Isilon IQ nodes. The only other noticeable difference is that Isilon ships their Isilon IQ nodes equipped with Host Channel Adaptors (HCAs) which are required to enable the IB switching fabric. Based on customer feedback, Isilon IQ with IB installs every bit as easily as Isilon IQ with GigE..

## **IB Impact for Isilon IQ**

The impact of having InfiniBand in the Isilon IQ backplane creates differentiation for the company and its customers on several fronts:

### **IB Isilon IQ Impact: Scalability**

Prior to deploying InfiniBand, Isilon had already demonstrated impressive linear scalability and high aggregate throughput in their Isilon IQ offerings. However, with the addition of the IB version of Isilon IQ, their

scalability metrics have improved even more. Based on our analysis of internal Isilon IQ metrics, combined with field reports from Isilon IQ customers, IB-based Isilon IQ clusters are providing higher aggregate throughput with fewer nodes than their Ethernet-based counterparts. The end result is an extremely high level of scalability, with the Isilon IQ 6000i ("i" designating "InfiniBand") supporting a potential single cluster size of 250 TB. This manner of "performance to scale" metric is a standout in the storage industry today. Frankly, we believe it would not be possible for the vast majority of enterprise customers to achieve these scalability levels without InfiniBand.

### **IB Isilon IQ Impact: High Performance**

Related to the scalability increases, the performance gains that Isilon reaps from the Infiniband version of its offerings are quite impressive. We see that a 10 node IsilonIQ 1920i ("i" designating "Infiniband) cluster is capable of supporting aggregate throughput of approximately 7200 Mbps *whereas the same 1920 with GigE can support approximately 4800 Mbps*. Both are impressive aggregate throughput numbers for a clustered storage solution, but its obvious to see that the approximately 33% gain from the Infiniband interconnect makes it a superior choice in terms of performance.

We also spoke with an Isilon IQ 1920i customer in the media production space who is now achieving an aggregate throughput of 5000 Mbps out of their 8 node Isilon IQ IB cluster. We collectively estimated that had they chosen to utilize GigE instead for the same workload, they would likely have managed an impressive 3500 Mbps

**T E C H N O L O G Y B R I E F**

aggregate throughput. Obviously, getting an extra 1500 Mbps for the same price, from the same footprint, with virtually zero impact on their management processes seemed to make good sense.

**IB Isilon IQ Impact: High Efficiency**

One of the major concerns that we know most IT shops have is the risk of exposing themselves to a yet another new technology. In the case of InfiniBand in particular, we know that many enterprises are still concerned about the potential impact this new interconnect might have on the efficiency of their existing management operations. Fortunately, in the case of Isilon IQ, the answer is that it will have virtually no managerial impact whatsoever while still providing all of the other gains we have discussed above.

Because Isilon leverages the IB interconnect as the backplane for intracluster communication between nodes, it is completely transparent to clients and administrators alike. In all other respects, the Isilon IQ offering will provide the IT team with exactly the same kind of efficiencies they expect from their GigE-based clustered storage environments.

**IB Isilon IQ Impact: Strong Economics**

The Isilon IQ offering was already considered an extremely cost-effective storage solution, with market-leading price/performance in its category. Now we feel that the ability to get anywhere from an additional 20-40% or more aggregate throughput out of that same investment places Isilon in a very unique position in the clustered storage market. For customers exploring the ROI of a clustered

storage solution, taking a hard look at how IB will stretch their return makes good sense.

**Taneja Group Opinion**

The management of Isilon Systems made a significant bet when they went down the path of deploying InfiniBand ahead of the pack. We believe they absolutely have made the right bet, and the proof is in the scalability, performance, and rapid customer adoption we are now seeing in the IB-based storage offerings the company delivers.

As we indicated above, Isilon was perhaps the most optimally positioned storage provider to make the early move to an IB-based backplane. Given the strong analog that their clustered storage architecture shares with other high performance computing clusters, it only made sense that they would want to find ways to leverage their atomic, modular architecture to provide even more throughput and higher scalability.

Another impressive point for us is that Isilon Systems is delivering their IB-based offerings at the same price as their GigE-based offerings. This was a very smart move, and their 90% customer uptake for IB over GigE proves the interest customers have in exploiting IB-based technology. Of course, the unspoken reality behind this economic move from Isilon is that InfiniBand has actually matured to the point where it is financially feasible for a vendor to offer price-equivalency between GigE and InfiniBand. Just two short years ago, such an option would not have been financially feasible.

**T E C H N O L O G Y B R I E F**

This leads us to the very interesting topic of where InfiniBand will continue to go within the enterprise market. Additionally, we ask what it will mean for the storage industry and Isilon in particular. We believe that InfiniBand will become an increasingly common switched fabric for clustered deployments in the enterprise market. Certainly it has gotten a start by providing a backplane to clustered solutions like Isilon IQ, but we expect to see IB-based technology reach out from this beachhead and go both wider and higher in the infrastructure.

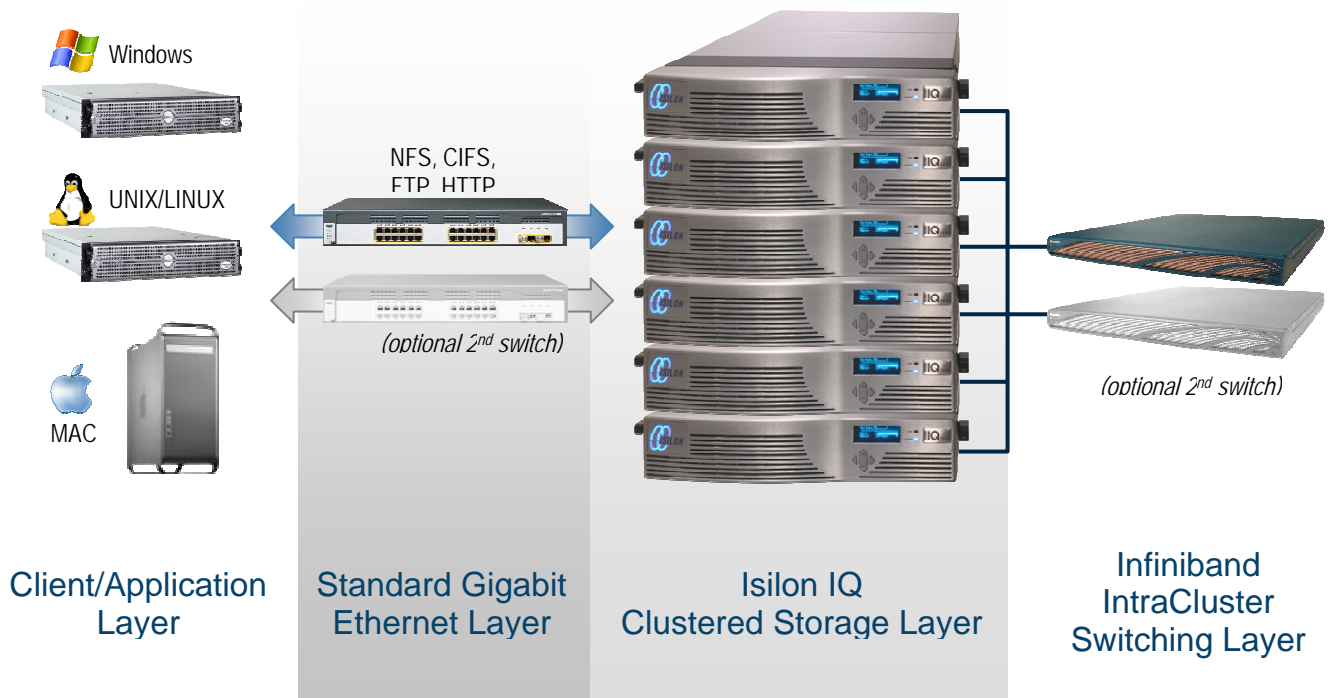
As InfiniBand speeds increase even more and IB-enabled technologies proliferate throughout the IT stack (servers, network switches, etc.) we believe that the storage applications for InfiniBand will truly blossom. For example, it is not hard for us to imagine a time 24 months from now when an Isilon IQ IB cluster is capable of supporting high performance database operations surfaced through a file system interface. Assuming that Isilon and other clustered storage vendors continue to develop their file system technologies to support the power of the IB interconnect, the addressable applications for their solutions could eventually become all encompassing.

Clustered storage has proven to be a good “first home” for InfiniBand within the greater enterprise community. We are very excited to see this technology finding new applications outside of its early adoption HPC usage models. The fact that clustering technologies are also taking the enterprise market by storm indicates to us that InfiniBand will continue to expand its presence in the enterprise market for the foreseeable future.

For companies like Isilon Systems that are already leveraging InfiniBand to their own competitive advantage, we expect to see them continue to extend their investment in this interconnect, ultimately using it to open up entirely new applications for clustered storage outside of its traditional usage cases. Indeed, InfiniBand could be their key to the larger enterprise storage kingdom if they continue to aggressively invest in the technology as they have to date.

So far, no one can argue with the fact that Isilon’s InfiniBand bet is paying off handsomely. And judging by the enthusiastic response we’re seeing from their customers, we’re willing to bet that this is just the beginning of Isilon’s cultivation of this critical new enterprise technology. Stay tuned: we expect that 2006 will be a very strong year for both Isilon and InfiniBand for clustered storage.

## Isilon Systems Isilon IQ and IB Architecture



***NOTICE:** The information and product recommendations made by the TANEJA GROUP are based upon public information and sources and may also include personal opinions both of the TANEJA GROUP and others, all of which we believe to be accurate and reliable. However, as market conditions change and not within our control, the information and recommendations are made without warranty of any kind. All product names used and mentioned herein are the trademarks of their respective owners. The TANEJA GROUP, Inc. assumes no responsibility or liability for any damages whatsoever (including incidental, consequential or otherwise), caused by your use of, or reliance upon, the information and recommendations presented herein, nor for any inadvertent errors which may appear in this document.*