

iSCSI Technology Brief

Storage Area Network using Gbit Ethernet

The iSCSI Standard

On February 11th 2003, the Internet Engineering Task Force (IETF) ratified the iSCSI standard. The IETF was made up of the industries leading IT technology companies including IBM and Cisco.

The iSCSI specification defined a protocol that allows SCSI commands to be encapsulated in TCP/IP to give computers access to storage devices over common IP networks. Low cost IP Networks are then used as the backbone to connect centralised or distributed storage resources to servers anywhere that an IP connection is available.

Having established the standard, operating system vendors were encouraged to develop iSCSI initiator software (for the computer server) and storage vendors produced iSCSI target storage systems. Any computer with an iSCSI initiator could gain access to the resources of any iSCSI target device through a low cost Gbit Ethernet connection with all the storage provisioning and security features supported by the server operating system.

Free Software

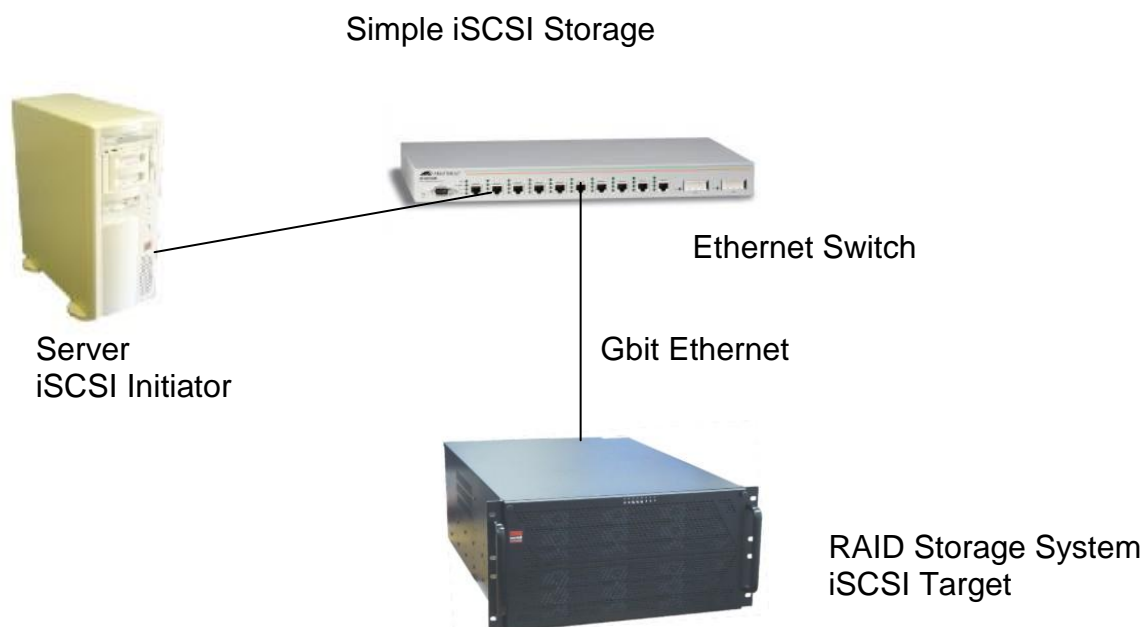
Microsoft had always been an active supporter of iSCSI and it demonstrated this on June 30th 2003 when it announced the free availability of iSCSI initiator code for Windows 2000, Windows 2003 and XP Pro. By downloading and installing this software from the Microsoft web site all applicable Windows servers, desktops and laptops became capable of receiving storage network services through their existing NIC interfaces. Without any hardware changes or additional cost millions of computers running Windows have become SAN ready clients.

Free iSCSI initiator software for servers using other operating systems (including many Linux and UNIX flavours) are widely available and the principles involved are the same. Mixed host operating systems are also entirely compatible with iSCSI storage systems.



Architecture

The beauty of an iSCSI based storage system is its simplicity. Any target storage capacity provisioned to a server (initiator) simply appears to that server as if it had a new disk directly inside that server, even though it may be hundreds of metres away or even over a wide area network at another location if bandwidth permits.



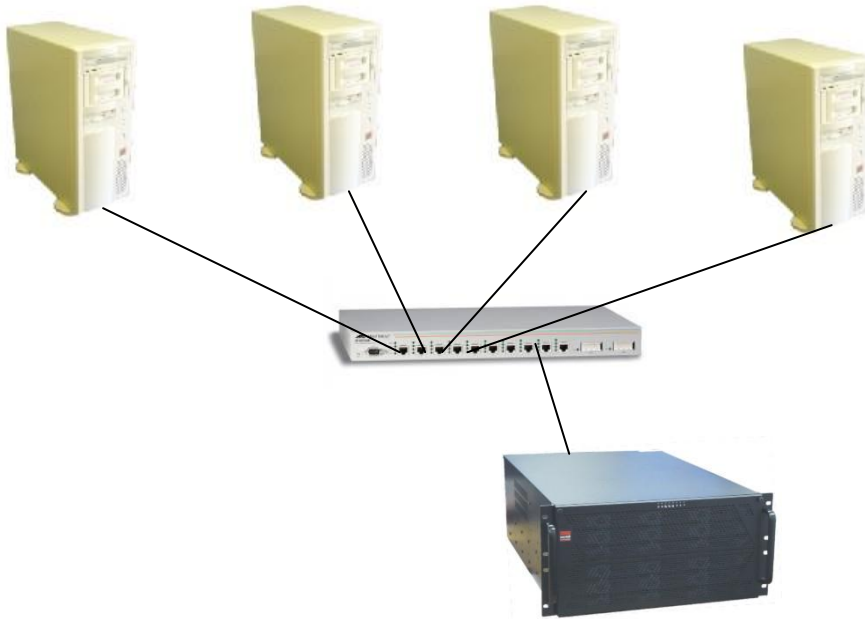
Anything that can be done with an internal disk can also be done with a remote storage pool provisioned using iSCSI over an IP connection.

So the "disk" can be backed up, mirrored, a point in time copy of the changes (snapshot) taken or the disk replicated without using expensive proprietary solutions.

A single iSCSI target storage system may provide storage capacity to one server, to a few servers, or to many servers. Each server would be allocated a part of the iSCSI's target storage capacity, and that part would be secure and owned only by the initiating server.

Westek offers iSCSI target storage systems to allow this scaleable, flexible and cost effective storage architecture to be deployed on a small, medium or large scale.

Multiple Servers using centralised storage



Connection and setup is very simple, and the provisioned storage capacity may be expanded as required. Multiple Gbit connections may also be made to the servers and to the iSCSI target storage device if required. Performance is greatly enhanced by connecting servers through a non-blocking Gbit switch rather than connecting directly to the iSCSI unit.

Multiple Servers using multiple centralised or distributed storage



Performance

In terms of performance, a single Gbit Ethernet connection is capable of providing a data rate of a little over 120MBytes per second which is excellent performance compared with many direct attached storage devices.

SANs may be built using Gbit Ethernet switches to allow connection to multiple servers and multiple iSCSI storage systems. The switch needs to be non-blocking and support “Jumbo Frames” to provide the best level of performance and ideally support channel aggregation so that multiple Gbit links can be teamed to form a higher bandwidth pipe for servers that require higher read/write performance than can be provided by a single Gbit link..

The Gbit Ethernet port connections on the host computers ideally also need to be high performance. If you are adding a Gbit Ethernet card to your host computer system, then it should be either a 64bit PCI-X card or PCI-Express and use a fast slot (133MHz for PCI-X and x4 or higher for PCI Express) if possible for best performance.

iSCSI Target Architecture In A Windows Environment

There is a key difference between an iSCSI Target and any other SAN software, in that implementing an iSCSI Target extends the existing Windows storage features instead of offering alternatives to them. By adopting this unique approach it effectively extends Windows to be a SAN aware operating system.

Each server that requires storage to be provisioned to it requires simple and free iSCSI initiator software to be added which may be installed without even rebooting the computer. Each iSCSI based storage unit (The “iSCSI target”) is then used to provide storage capacity to that server via simple low cost Gbit Ethernet. Each iSCSI target storage system may be provisioned to provide storage capacity to a large number of server initiators. Similarly many iSCSI target storage systems may be used to provide scalable storage capacity. The architecture is flexible and may be one to many, many to one, or many to many with no additional licensing costs

The iSCSI target storage system connects seamlessly with the freely available iSCSI Microsoft initiator code to form the foundation building block of the Windows SAN and extending its capabilities further.

Examples of the solutions enabled are:

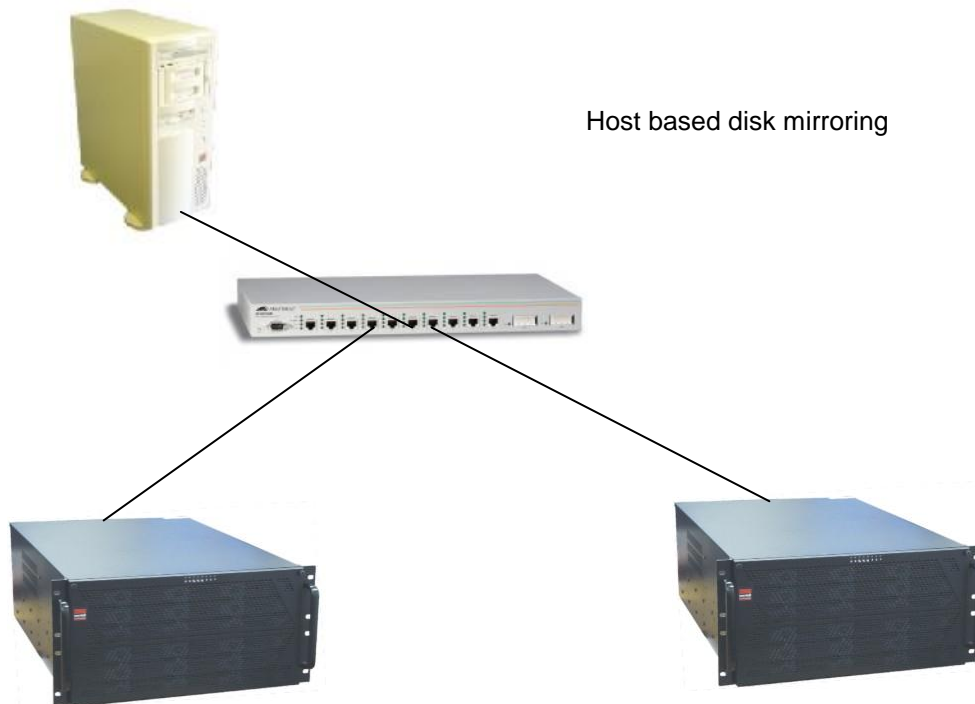
- Extending disk management to provision disks network wide
- Extend VSS (Volume Shadow copy Service - Microsoft’s snapshot feature in W2003) to take point in time copies anywhere and at any time
- Extend disk mounting capability to allow remote file systems to be viewed locally
- Leverage of Windows security (Microsoft standards CHAP & IPsec) to protect the SAN.

Data Security

There are a number of ways to provide added data security and data recovery using iSCSI based storage.

1. Disk Mirroring

One simple way is to allocate storage capacity on two separate iSCSI target storage systems, and simply mirror the data using the features supported under your client systems operating system. In order to setup a disk mirror under Windows the disks need to be configured as dynamic. This feature needs to be supported by the iSCSI initiator that you are using, or you may need to use an iSCSI HBA card in your server that does support dynamic disks. Once the RAID 1 mirror is setup, data is duplicated in real time onto both storage devices, and is still available to the server should the data on either one of the storage systems become unavailable. Unlike the replication suite in section 3, mirroring does place a burden on the host since data is being written twice as a RAID 1 mirror.



Host based disk mirroring

2. Snapshot and backup/restore

Snapshot is a feature to enable a point in time copy of iSCSI target data to be created. The snapshot copy will contain only changes made to the data since the previous snapshot was created. Using this snapshot, data can be reviewed and restored back to an earlier version to overcome data deletion, data corruption and virus infected files. The snapshot can be rolled back to the original state entirely or individual files restored to a previous version. The snapshot can also be used to backup the iSCSI target even if the data is live. As the snapshot is a copy of the

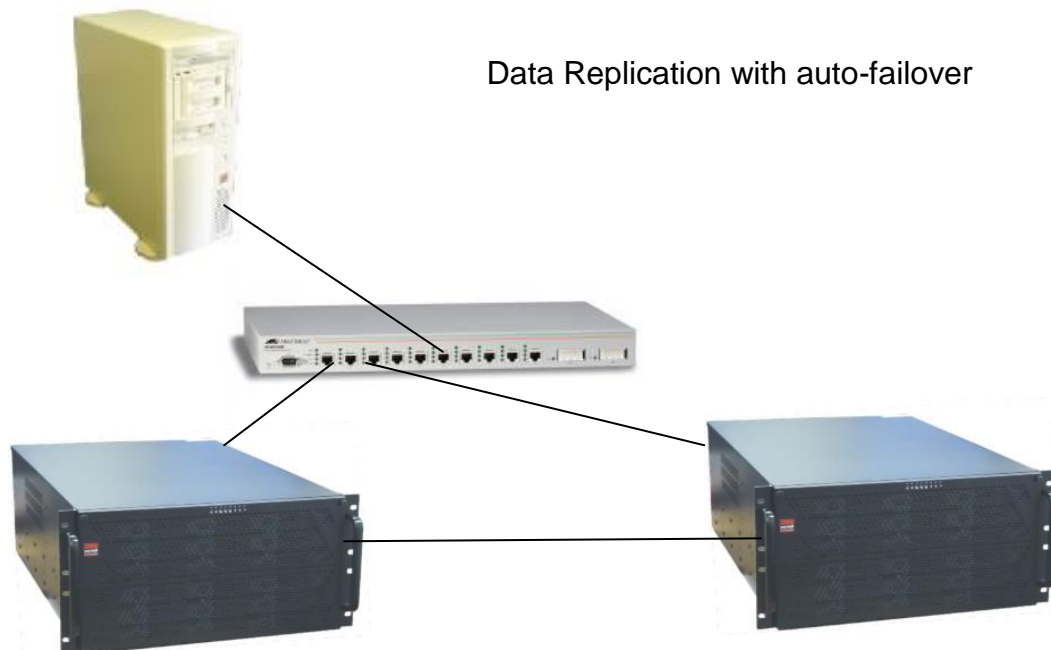
changes made to the live data it can be presented to any iSCSI initiator or backup server elsewhere on the network and used to backup the volume in full.

Automatic time stamped data volume snapshots at the block level is available as an optional feature on the iSCSI target storage system. A snapshot typically takes just a few seconds to execute and takes up virtually no disk space. Snapshots may be scheduled to happen automatically at preset intervals or manually executed at any time under the system administrators control.

3. Block level data replication with seamless failover

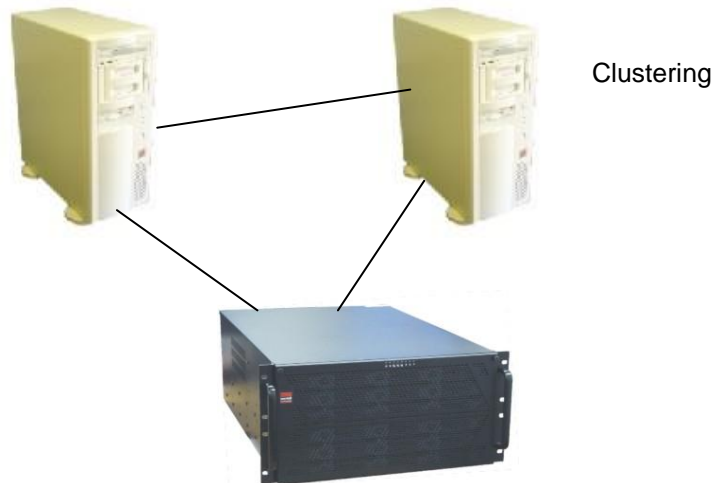
Data replication is another option that allocates one iSCSI target as the primary data storage device, and asynchronously writes data to a second iSCSI target device using a dedicated and separate Gbit connection. Should the primary unit fail for any reason, the secondary unit takes over in a seamless fashion using its inbuilt failover capabilities. This does not involve the host computer at all, unlike disk mirroring, and may be configured to be entirely automatic. Furthermore the host is only writing one set of data to the primary storage device (unlike host based mirroring) and the replication is taken care of by the iSCSI targets themselves. IP connections are seamlessly transferred to the secondary preserving data availability.

Only new data block level changes are written to the secondary unit, which may be located somewhere else on the same site or even remotely over a WAN link. Since the replication is pure block level, the data can be anything including database SQL / Exchange / Oracle etc.



4. Clustering

The iSCSI Target is cluster aware such that a single iSCSI Target storage system may be connected separately to two servers setup in a cluster. Access to data is available should either server fail. Clustering is provided by the host operating system (such as Microsoft's Advanced Server edition of Windows 2003)



Is iSCSI the right solution for my data storage needs?

iSCSI is a block based storage technology, just like direct attached storage (DAS) or a fibre channel SAN system. It is therefore entirely compatible with database applications including SQL and Microsoft Exchange, unlike Network Attached Storage (NAS) file based storage systems. iSCSI is not data or operating system dependant as long as the initiator is available for the host system, and most are now supported.

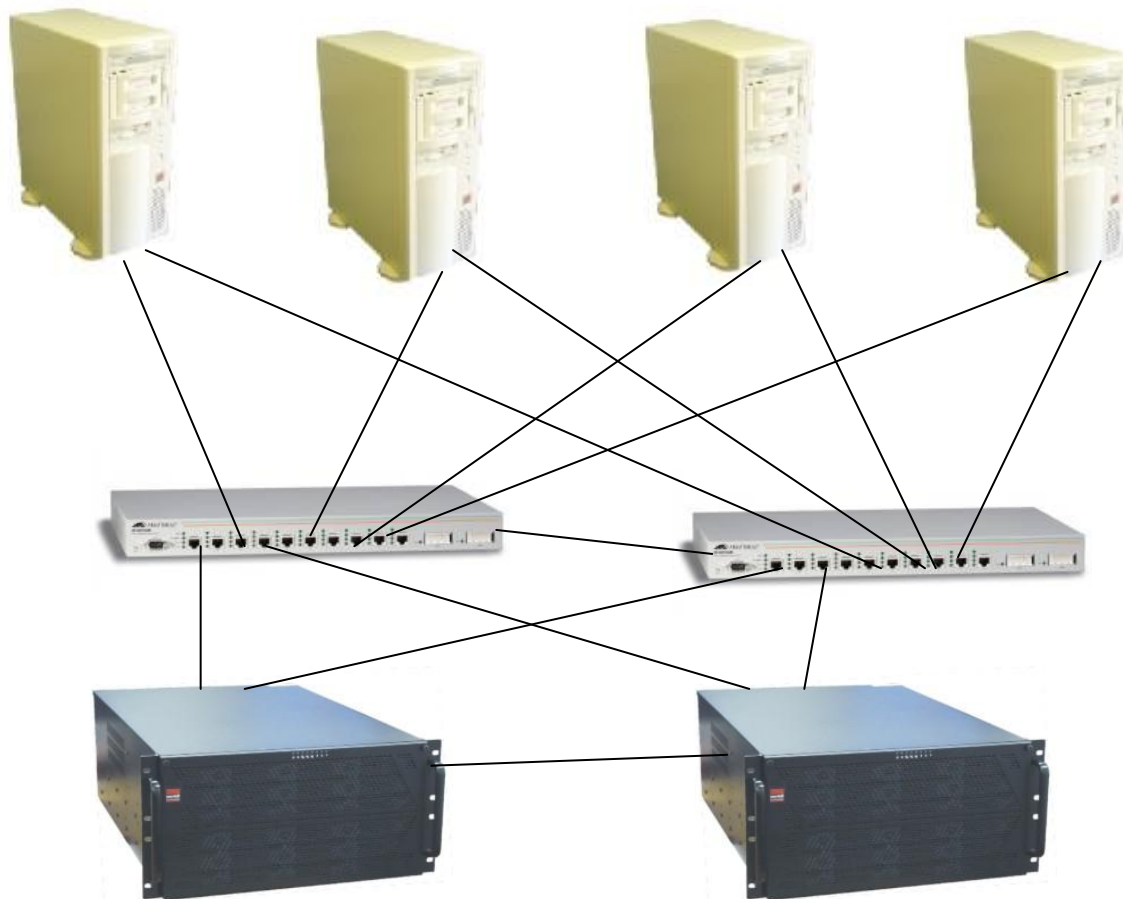
NAS systems offer IP connections to a central file based storage system. NAS systems are operating system and file system dependant. They are great for offering a file based shared storage system, but performance is usually lower than a DAS system and they are useful really only for file based applications and not for use with applications such as an SQL database or Microsoft Exchange data.

Today DAS systems are available with either SCSI or Fibre channel host connections. The latest U320 SCSI provides data rates in theory up to 320MBytes per second. The disadvantages are that a relatively expensive U320 SCSI interface is required in each host computer and the storage system must be within a few meters of the server – clearly not good for building a SAN or where the server systems are in an office environment.

Fibre channel provides up to 400MBytes per second on the latest 4Gb/s systems and allows the storage system to be remotely located and a SAN to be built. However fibre channel infrastructure interfaces and switches are expensive and require specialist management suites further increasing the cost of ownership.

The iSCSI initiator driver does add a small load to the host cpu, since if using a standard TCI/IP Ethernet card, then the host CPU performs the iSCSI encapsulation part of the protocol process. In most systems this is not an issue, and, if it is, then "TOE" cards which off load the IP encapsulation are available to replace the standard Gbit Ethernet interface. The client machine should use a separate dedicated Gbit Ethernet connection to the iSCSI storage network device and not share this connection with the normal LAN.

Fully redundant data replication



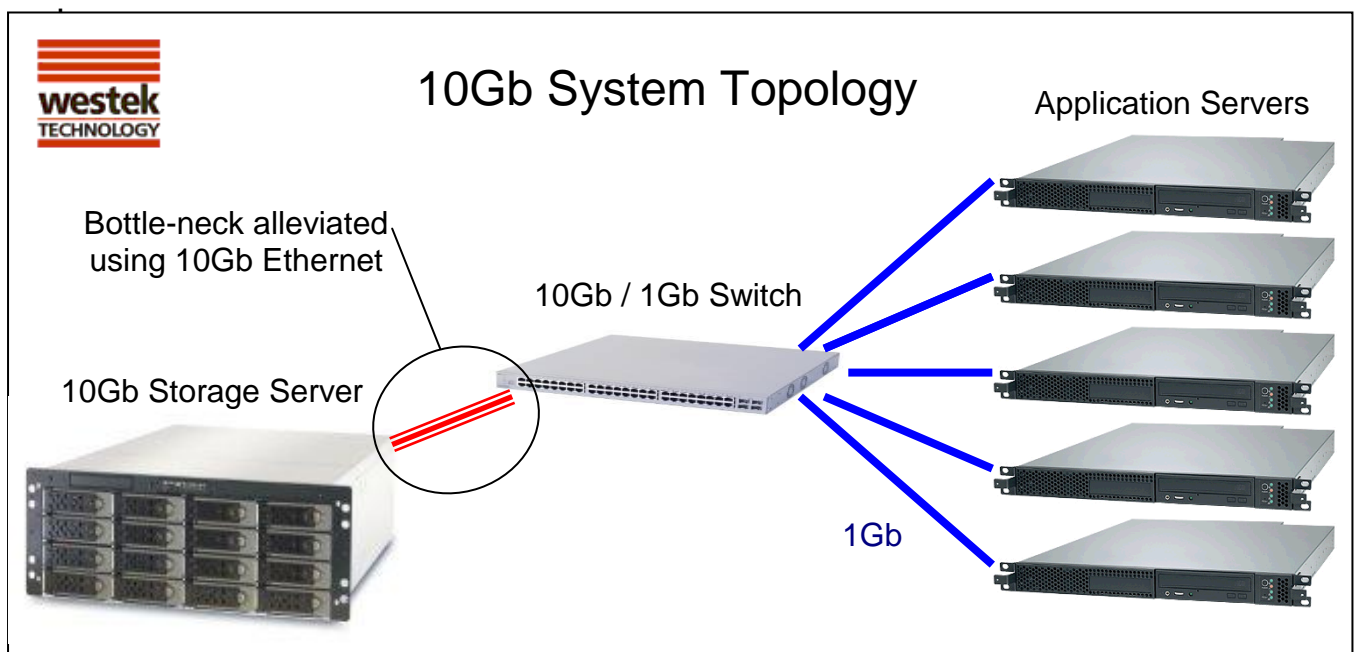
Higher Performance : 10Gbit

Current Ethernet interfaces are typically 1Gbit per second, and around 120MBytes per second is achievable over a single Gbit link. 10Gbit Ethernet adapters and switches are now available from a number of vendors and the cost of 10Gbit Ethernet infrastructure has decreased significantly since its introduction in 2005. 10Gbit interfaces provide iSCSI storage solutions with a strong future by providing very high performance data transfer rates exceeding even the fastest fibre channel products available today

Unless you have exceptional data throughput needs it is unlikely that 10Gbit connections are needed to the actual servers on your iSCSI storage area network. However using one or more 10Gbit links between the Ethernet switch and the iSCSI storage system will alleviate the potential bandwidth bottleneck at this point for systems with a large number of initiator servers.

A single 10Gbit Ethernet connection will provide up to 1GByte per second performance. If this is shared between ten iSCSI initiators all reading data at once the architecture below would still provide over 100MBytes per second to each server, assuming the iSCSI storage device itself is capable of supporting such a high performance.

10Gbit Interfaces are available in relative low cost CX4 copper based connectivity for short distance connections or optical fibre for longer distances.



Summary

iSCSI provides a cost effective way to provide scaleable storage for many applications. Even over standard 1Gbit Ethernet links its performance is high enough to support intensive real time data, even for high definition TV or multiple real time video streams. It is simple to setup and manage, is highly scaleable and with 10Gbit interfaces it offers serious performance at low cost compared with other SAN technologies.

An iSCSI based storage SAN is an excellent choice to build a long lasting reliable and scaleable storage solution to meet your business needs.

For more information on iSCSI storage products please visit our web site at www.westekuk.com or call +44 (0)1225 790600