

SANRAD White Paper
IP SAN Performance – Best Practices
WP-013-02

SANRAD
US Tel: +1-800-471-2616
International Tel: +972-3-767-4800

Copyright SANRAD 2006

All rights reserved. The copyright and all intellectual property rights in this article belong to SANRAD. It is strictly forbidden to copy, duplicate or otherwise use this article or any part thereof in any way shape or form without the prior written consent of SANRAD.

IP SAN Performance – Overview

SAN benefits of improved storage utilization, high availability and data protection are well understood. Today there are two protocols available for building block-based SANs, FC and iSCSI. Both protocols use SCSI commands generated by the file systems of the servers. These SCSI commands are converted by the iSCSI or FC protocol so they can move thru a network to and from centralized disk storage systems where the commands are executed. In the case of FC the network equipment is specific to the protocol. In the case of iSCSI the network equipment type is anything that will handle IP packets, 1GB Ethernet is the most popular.

There are many IT professionals considering iSCSI based IP-SAN's as a means for centralizing storage for their application servers. As opposed to FC-SANs, IP-SANs have the benefit of being based on TCP/IP allowing businesses to use standard Ethernet equipment, NICs, tools and the knowledge base within their IT staff. But when data is being read and written across an IP-SAN and not to internal disk drives, users are concerned that network latencies will degrade server performance. Similar to an FC-SAN, when using an IP-SAN the server, the network and the storage system all play a part in application performance and client satisfaction. It's important to understand how to identify and eliminate latency bottlenecks to ensure superior application performance. In many cases, a properly designed IP-SAN can deliver better performance than internal disk drives.

Attaching a server to the IP-SAN – Server CPU induced Latency

Today most popular operating systems support an iSCSI software initiator. The iSCSI initiator is software responsible for encapsulating the SCSI command into TCP/IP and placing it onto the network. iSCSI in itself is fairly low CPU intensive software and even during heavy loads uses very little of the CPU power. But TCP/IP can consume noticeable CPU resources. If you want to eliminate latency at the server layer and not dedicate much of the CPU/s for driving IP-SAN traffic, it is recommended to use an iSCSI TCP/IP TOE NIC. A TOE (TCP/IP Offload Engine) NIC is a special interface card specifically designed for interfacing a server to the IP-SAN and will offload iSCSI as well and TCP/IP encapsulation from the server CPU/s. As a general rule an iSCSI TOE NIC (also called an iSCSI HBA) is recommended if your average CPU utilization before the use of iSCSI is higher than 50% during usual business hours, above 65% during peak use periods, or nearing 75% during backup or mirroring operations. Most iSCSI TOE NICs come with their own initiators so you'll want to check compatibility with your target operating systems before selecting a TOE. Network boot is also a feature supplied by some iSCSI TOE NICs.

Attaching a server to the IP-SAN – Saturating the Ethernet link to the LAN

“Server Fan-in” describes how many servers you can run thru a single GB Ethernet port within an IP-SAN before you start experiencing latency caused by excessive storage traffic. Average business servers do not generate from 50 to 200 IOP's (input/output operations) toward the storage drives. And in most cases do not generate more than 5 megabytes of storage traffic. As a general rule, assume a GB connection can support 80 megabytes of storage traffic and 10,000 IOPs. Using these figures you can attach up to 16 servers on a single GB connection assuming each server is not generating more than 5 megabytes per second of storage traffic. It's important to note that iSCSI is encapsulated into TCP/IP and thus any network that supports TCP/IP can be used as part of an IP-SAN to move storage traffic including 10/100 connections, wireless, infrared LAN/MAN/WAN and even the internet. Naturally performance across these types of networks will vary greatly depending on connection speed but all have been tested and work. For new IP-SAN deployments, 1Gb is recommended.

As you can see, understanding your server statistics relative to storage IO, MBs and CPU usage is very important when determining what equipment to purchase and what size to make your IP-SAN solution. Most operating systems have performance monitoring tools and logs you can use to collect statistics on server CPU and storage usage. For example Windows performance monitoring can be done using “perfmon” and in Linux you can use “sysstat”. This information in addition to your expansion plans will help you to determine the proper configuration and equipment needed to build an IP-SAN solution that will exceed the performance requirements of your business application servers and be able to scale into the future.

IP-SAN Network Speed – Selecting an Ethernet Switch

The main criteria for an Ethernet switch is that the switch is non-blocking. It can be either a layer 2 or layer 3 switch. Having a layer 3 switch is generally preferred for easier SAN management and monitoring. Today’s IP networks are extremely fast and scalable. For example, it takes only a quarter second to send and receive a “ping command” from half way around the world. This is minuscule when compared to the time it takes to seek and read a 10K file from any disk drive. In all shared iSCSI/ IP storage models, the performance of the network is usually insignificant when compared to the performance of the disk storage system. (Usually 5 to 10 times faster than the disk storage system)

IP-SAN Network Speed – Selecting an Intelligent IP-SAN Switch

Intelligent IP-SAN switches are designed to sustain very high levels of random read and write IOPs operations. Intelligent storage switches manage the IP-SAN and eliminate the need for 3rd party software and agents. Intelligent switches have high-speed internal architectures utilizing network processors, real-time operating systems and 25Gb backplanes. They provide necessary storage services like security, virtual disk creation, multi-pathing, failover and mirroring for high availability, protocol conversion to use basic SCSI or FC disk arrays, virtual disk resizing, backup and data replication. A single IP-SANs switch generally has many Ethernet ports and can sustain 300 megabytes per second or 600 megabytes per seconds when clustered of random read and write requests and over 60,000 IOPs. This delivers raw random read or write performance and complete storage services for well over 100 standard business application servers assuming all the servers do not simultaneously generate more than 5 megabytes per second and 600 IOPs. If needed, more servers can be attached assuming that all the servers not peek at exactly the same time. The transverse would also be true if you had high-end servers generating over 25 megabytes each and 3000 IOPs, the IP-SAN could reasonably only support 20 such servers. Like arrays, intelligent IP-SAN switches are available in different sizes.

Selecting the Appropriate Centralized Disk Storage System

IP-SANs can utilize any type of storage system. This allows the IT professional to select the storage system/s that best fit the performance and reliability needs of the organization. Moreover you can select different classes of storage. For example you can use an attached FC array rated for 200 MBs and 20,000 IOPs for servers requiring high performance (10 to 20 megabytes per second / 2000 IOPs) and use a lower cost array with slower drives and interface rated for 40MBs and 3000 IOPs for applications requiring less performance (under 5 megabytes per second / 500 IOPs). Because the IP-SAN can use different grades of storage, it’s easy to construct a SAN with primary, secondary and even tertiary storage.

In addition to selecting different classes of storage (different cache and drive types), IP-SANs can simultaneously read and write data to multiple independent storage systems. Unlike individual storage arrays that cannot address more than their own internal storage capacity without causing significant performance

degradation, IP-SANs can simultaneously read and write to multiple independent storage arrays. By spreading volumes across independent storage systems and being able to directly access those storage systems without having to pass thru another control layer, IP-SAN's can maintain line speed performance to the storage systems (up to 2GBs, 200MBs, 20,000 IOPs per storage array) regardless of the location of the data. Moreover, since the storage systems are independent of the intelligent storage switches, capacity can be increased with additional arrays without degrading performance or having to suspend application servers.

In selecting a storage system for any SAN it's important to understand that the array/s will be shared among all the servers. By the nature of a shared storage system, random read and write performance is significantly more important than sequential performance specifications. This is obvious since the storage solution is shared among many application servers, each using unique volumes and reading and writing data whenever and wherever required. In general, published MBs (megabyte per second) specifications for disk drives and arrays are for sequential large block writes (1028k). This provides the optimum performance and is usually what is published. But within a SAN random IO and use of smaller block sizes (1k to 16k) is common. It's very important to work with your storage and IP-SAN supplier to understand the performance of the array within a random access environment and how to configure the array or optimum performance.

Two common applications are Microsoft SQL and Exchange. The following test results demonstrate the performance of an IP-SAN relative to these popular applications.

Microsoft SQL Test

The Microsoft SQL test demonstrates the performance differences between an FC-SAN and an IP-SAN. The test environment included a Windows 2003 server running SQL 2000. The server was attached to the IP-SAN using a regular 1 GB NIC and the standard MS iSCSI initiator software. The same server was attached to the FC-SAN using a FC HBA.

In the case of the IP-SAN the server had a connection to an intelligent IP-SAN switch, which was using storage from an attached mid-range FC disk storage system. The disk storage system was configured with two LUNs, each configured with RAID 0+1. On the intelligent IP-SAN switch two volumes were created, one for the database and one for the transaction log.

In the case of the FC-SAN test the server was connected to same the disk storage system, through a fibre-channel switch.

Performance was tested using the Microsoft Database Hammer Utility from the SQL 2000 resource kit. This utility created a database with 10 million rows and simulated SQL 2000 activities with different numbers of connections. The test was first conducted on the FC-SAN and then the IP-SAN. As you can see from the results the FC-SAN did perform better than the IP-SAN but the difference in latency and latch time was negligible and became relatively less as the number of connections increased.

Number of Connections	IOPs		Average Disk Latency (Database volume) (ms)		SQL Server Avg Latch (ms)	
	FC-SAN	IP-SAN Intelligent Switch	FC-SAN	IP-SAN Intelligent Switch	FC-SAN	IP-SAN Intelligent Switch
100	934	879	21	27	47	51
200	1009	976	29	35	51	55
300	1134	1118	39	43	62	69
400	1401	1355	48	55	78	81
500	1603	1546	59	66	96	103

Microsoft Exchange Test

The Microsoft Exchange test demonstrates the ability of an IP-SAN to perform within the recommended performance boundaries recommended by Microsoft. The test environment included a Windows 2003 server running Exchange. The server was attached to the IP-SAN using a regular 1 Gb NIC and the standard MS iSCSI initiator software. On the intelligent IP-SAN switch two volumes were created from an attached mid-range FC disk array, one for the database and one for the transaction log. The following table describes the “JetStress” stress mode results for Exchange supporting different numbers of users. The important parameter to look for is average disk latency. The table presents latency in milliseconds for read IO and write IO for the database and transaction log volumes. The test results show IP-SAN Exchange performance well within the limits recommended performance by Microsoft.

<u>Volume</u>	<u>500 users</u>	<u>1000 users</u>	<u>1500 users</u>	<u>2000 users</u>	<u>Maximum recommended by Microsoft</u>
Database Volume Read	15 ms	17 ms	18 ms	20 ms	40 ms
Database Volume Write	7 ms	8 ms	9 ms	11 ms	40 ms
Log Volume Read	0 ms	0 ms	0 ms	1 ms	20 ms
Log Volume Write	2 ms	2 ms	2 ms	4 ms	20 ms

Conclusion

Building an IP-SAN on TCP/IP allows businesses to use standard Ethernet equipment, NICs, tools and the knowledge base within their IT staff. IP-SANs are easily deployed so they exceed the performance requirements for the majority of small to medium business application servers. By following basic guidelines for connecting the servers and selecting the correct network equipment and storage arrays, any business can benefit from the cost savings, ease of management and reliability of an IP-SAN.

By Zophar Sante'
VP of Market Development
SANRAD Inc.