StarWind iSCSI SAN for Microsoft Windows: Best Practices Guide

INTRODUCTION

StarWind Windows-based iSCSI SAN software is designed specifically for the Windows 2003/Vista/2008/2008 R2 platforms and enables IT departments to extend Windows storage capabilities to provide cost effective network storage solutions. It supports all Windows ertified storage devices and hardware such as Fibre Channel, SCSI, IDE and SATA. StarWind's ability to work well with any hardware components makes the software extremely flexible. End users can configure StarWind using hardware that meets the minimum system requirements, or it can be installed on even the fastest hardware available on the market. Given the wide range of possible hardware configurations, this document provides guidelines on how to setup a StarWind server for optimal performance and availability.

CONFIGURING FOR PERFOMANCE

Operating System

Taking full advantage of StarWind's capabilities starts with the operating system. StarWind Software recommends a server class network operating system such as Windows 2003 Server, Windows Server 2008 or Windows Server 2008 R2. A server operating system is preferred for two reasons: first, a server operating system is optimized for a multi-user environment. One aspect of this optimization is the smaller CPU time slice or quantum granted to each process. Since StarWind typically handles multiple initiators concurrently, the smaller quantum means that requests are handled quicker resulting in lower latency. Secondly, the operating system gives priority to a service over a foreground application. StarWind directly benefits from this because its core runs as a Windows service. Having higher priority over foreground applications ensures that StarWind gets as much processing time as it needs.

Processor And Physical Memory

The system processor plays an important role in StarWind's performance. Unlike many hardware implementations of iSCSI targets, StarWind relies on the host CPU to process all iSCSI requests. A slow processor delays StarWind from servicing incoming requests in a timely manner. If the processor is not fast enough to handle incoming packets, as the traffic increases the number of packets waiting to be processed continue to grow. This bottleneck not only drastically reduces performance but also causes network congestion.

Insufficient memory can also severely hinder performance. When the operating system cannot allocate memory from the physical memory, it swaps data between memory and disk via the pagefile. If StarWind has to wait for its data to be read from disk, this will result in increased latency which will result in lower I/O rates.

StarWind leverages advanced I/O features and multi-thread capabilities available on the Windows platform which allows it to use the processor and memory sparingly. As a rule of thumb, one iSCSI session typically consumes about 10 MB of memory. For optimal performance, StarWind Software recommends configuring the server to use a Xeon 2.8 GHz CPU with 512 MB of RAM.

Network Speed

iSCSI allows storage commands to be transmitted over a TCP/IP network. TCP/IP is the virtual "wire" between a hard disk and a hard disk controller. Performance of TCP/IP can be measured in two ways: throughput and latency. Throughput is the amount of data transferred in a time period, usually defined in bytes per second. Latency is defined as the amount of time between the request of the data, and when the data becomes available, usually defined in milliseconds. In this case iSCSI is the communication medium between the initiators and StarWind Disks. In order to gain the maximum performance from the disk, the overhead of the network "wire" (latency) needs to be negligible while the transfer rate (throughput) is high.

As such, optimal iSCSI performance is obtained by using a Gigabit or 10 GB network. A Gigabit or 10 GB network has low latency while providing maximum throughput. Gigabit Ethernet latency is typically less than the seek time of today's hard drives, while Gigabit Ethernet throughput (1,000Mbps or 130MBps) is more than adequate for most existing drive technology.

Network Adapter

In a traditional network adapter, the TCP stack uses the system processor to break a TCP segment down into Ethernet frames before passing the data to the adapter. This requires a large amount of processing time, especially in a Gigabit network. Furthermore, each frame that is passed to the adapter requires an interrupt to be raised, utilizing even more processor cycles. Large send/TCP Segmentation offload moves this computation from the TCP stack onto the network adapter. This means the entire buffer is sent to the adapter using just one interrupt; and since the work of segmenting the buffer into Ethernet frames is done by the adapter, the processor is available to perform other tasks. In the case of incoming traffic, Ethernet data is consolidated into a large buffer and passed to the operating system using a single interrupt.

Every TCP packet contains a checksum. This binary value is used to ensure that data is not corrupted along the path from the client to the server. The data integrity check requires processing time which can also be done on the adapter.

StarWind Software recommends the use of network adapters with hardware offload functionalities such as large send/TCP segmentation offload and checksum offload. These features are typical in most server adapters available today.

Network Switch

Network components can be connected using a point-to-point configuration or a broadcast configuration. In a point-to-point configuration, each node in the network has its own path to the server. A direct benefit of this model is that each client has the entire network bandwidth available. Furthermore, it reduces the amount of spurious traffic resulting in a lower number of collisions. This is critical to the efficiency of a high traffic network.

In a broadcast configuration, all nodes share one path to the server, which means the bandwidth is shared by all nodes. Collisions are more likely to occur because of the shared traffic. When network collisions are high, the overhead incurred in retransmitting packets drastically reduces the performance of StarWind.

Point-to-point configuration is accomplished by using network switches and broadcast configuration is available via network hubs. StarWind Software recommends the use of a switch when connecting network components involving StarWind.

Disks

Reading and writing to the hard disk accounts for half of StarWind's processing activities. Three types of disk drives are widely in use today that are relatively inexpensive: ATA, SATA, and SCSI. All three standards feature what is known as tagged command queuing (TCQ) which enables a disk to efficiently service multiple reads and writes simultaneously. A TCQ enabled disk and/or controller can re-order requests to minimize head seek time resulting in lower latency. This is greatly beneficial to StarWind's ability to service many clients at the same time while minimizing the number of iSCSI requests waiting on the network.

To further improve disk performance, the disks can be set up in a RAID configuration to provide striping. (Fault-tolerance with disk is discussed in the RAID Volumes section below.) In this configuration, data are spanned across multiple disks. Seek time during reads and writes is minimized because the operations are performed at the same time across the disks. The net results are lower latency and faster disk operations. StarWind Software recommends using disks with tagged command queuing and RAID capabilities since they are better suited for high performance applications such as StarWind.

CONFIGURING FOR HIGH AVAILABILITY

Operating System

In addition to being built for high performance, Windows Server is also designed for high availability environments. Without any additional configurations, the server has a downtime of less than an hour per year. Having a server that requires reboots infrequently means that StarWind is available to service storage requests nearly 24x7x365.

Raid volumes

The most powerful functionality that StarWind provides is the ability to consolidate storage. All the storage resides on the StarWind server. There are many advantages to this, some of which are ease of management and dynamic provisioning. StarWind can provide disk space on an "as you need it" basis. There is however a downside to centralize storage; it is the risk of data loss and server downtime. If the disk on the StarWind server fails, servers that utilize StarWind Disks such as Exchange, IIS, file and print server cannot access their data. And worse, data could be lost. Utilizing RAID eliminates this problem.

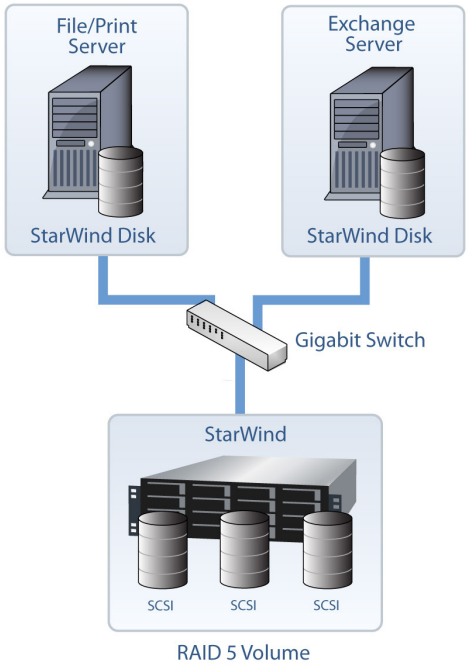

Figure 1 - StarWind Disk backed by RAID 5 volume.

The easiest way to provide fault-tolerance to important application data is by using RAID-1 (mirroring) functionalities. StarWind is fully capable of handling any RAID configurations supported by Windows. When StarWind Disks are created using mirrored volumes, the Disks are available for use by clients such as Exchange Server with virtually zero downtime. The Windows Logical Disk Manager ensures that StarWind continues to read and write to the disk even when one of the mirrors fails. The mirror is re-established when the failed disk is replaced making the volume fault-tolerant once again. In this scenario, the system incurred no downtime and no data loss. To gain both the performance benefits of RAID-0 and the fault-tolerance of RAID-1, configure StarWind with RAID-5 volumes, see Figure 1. See Disks section above for RAID-0 discussion.

Hot Swappable Hardware

To further minimize downtime in the event of a hardware problem, deploy StarWind on a server in combination with hot-swappable backplane hardware. In the unlikely event of a hard disk failure for instance, the disk can be replaced without having to shut the system down. It is also advantageous when hardware needs to be upgraded. Expanding the storage capacity of a server is as simple as replacing the existing disk with a higher capacity hard drive. Such procedures do not have to involve scheduling a downtime for the system.

Redundant Hardware

Redundancy is vital to any high availability systems. This is achieved with the hard disk by using multiple disk drives in a RAID configuration. Replacements for other internal components that cannot be made redundant such as CPU and RAM should be available and ready in the event that these parts need replacing. Even though it does not provide continuous uptime for the system in the event of major failure, this practice significantly reduces the downtime of the StarWind server. Another factor that needs to be considered is the power supply; perhaps the weakest link in the high availability chain. To that end, the power supply feeding the system should also have a failover solution. Having the system on a UPS or a secondary power supply solves this problem. This helps ensure that power fluctuations do not affect the availability of the server.

NETWORK SEGMENTATION

Network Topology

When deploying StarWind in a SAN environment, separate the network segment from other traffic in the enterprise. This helps to minimize network congestion. Since the segments are separated, network packets destined for the StarWind nodes do not have to compete with other corporate traffic for bandwidth. Another advantage of this type of network segmentation is the ease in security and network management. Troubleshooting errors, detecting network failures, and auditing security is simpler since problems can be easily isolated to the SAN segment. See Figure 2 below.

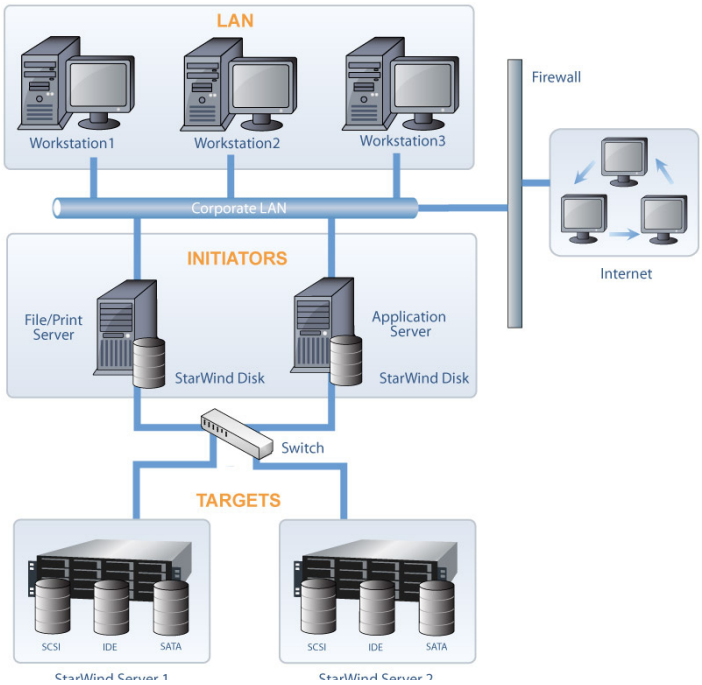

Figure 2 - Network Segmentation.

ISNS

Secondly, the SAN should be set up to use iSNS. iSNS is analogous to DNS. Just as DNS provides name service for servers and workstations in a LAN, an iSNS server provides name service for initiators and targets in a SAN. This makes the task of managing the storage network easier because the data are centralized on a server.

Furthermore, StarWind fully supports integration with iSNS server. StarWind also registers itself with the iSNS server allowing initiators to easily locate the storage server. Using iSNS greatly simplifies the management of initiators and targets as the SAN grows.

CONCLUSION

This document is intended as a guide for deploying StarWind. Many variables are involved in finding the best way to deploy StarWind in a particular environment. Some of which are network environment, storage capacity, and budget constraints. Use this reference as a starting point for implementing and deploying network storage solutions with StarWind.