

SQL Server 2000 I/O Configuration in a SAN/NAS Environment

Edward Whalen, Performance Tuning Corporation

March 2002

INTRODUCTION

Some of the most common SQL Server performance problems involve the I/O subsystem. Since SQL Server's main function is to manipulate data, and that data resides either in memory or on the I/O subsystem, any I/O performance problems will result in SQL Server performance problems. Much of the design of the SQL Server RDBMS is intended to make accessing the I/O subsystem as efficient as possible.

In this article the fundamental concepts and tuning of an I/O subsystems will be explored. By understanding the limitations of the I/O subsystem, you will be able to design and properly size it so that performance can be optimized. This article will start by describing the basics of how a disk drive works and the limitations of a disk drive. Next, RAID subsystems will be described and how to properly configure and optimize them. Finally advanced I/O subsystems such as SAN and NAS storage will be covered.

If you understand how SQL Server and the I/O subsystem interact you can better configure your I/O subsystem for optimal performance. A properly configured I/O subsystem will allow SQL Server to perform optimally. A poorly configured I/O subsystem can easily become a bottleneck and can severely affect performance.

Why is I/O Performance Important?

The performance of the I/O subsystem is key to SQL Server performance. Further, read performance is critical to SQL Server performance, write performance is secondary. When a query is executed the user waits on reads to complete before the system responds with the data that is requested. When modifications are done to the database, the lazy writer will write that data out at a later time, so, although write performance is important, no users ever wait on writes to occur (except to the transaction log). Let's look at a couple of examples:

Indexed reads

When a query is able to read from an index, it must make many trips to the I/O subsystem. The root page of the index is read and SQL Server makes a decision what next page in the index is read. That page then incurs an I/O and SQL Server again must make a decision and read again. With an efficient index that is not too deep, this might take 50-100 I/O's in order to find the data. Let's say for example that each I/O takes 10 ms (milliseconds). This index access will take 500ms – 1sec. That's pretty quick.

Let's now assume that the I/O subsystem is performing poorly, say each I/O takes 40ms. (I've seen as high as 450ms in production environments). In this case, the same index lookup would now take 2sec – 4sec. This is bordering on unacceptable, but this is often the case.

Table Scans

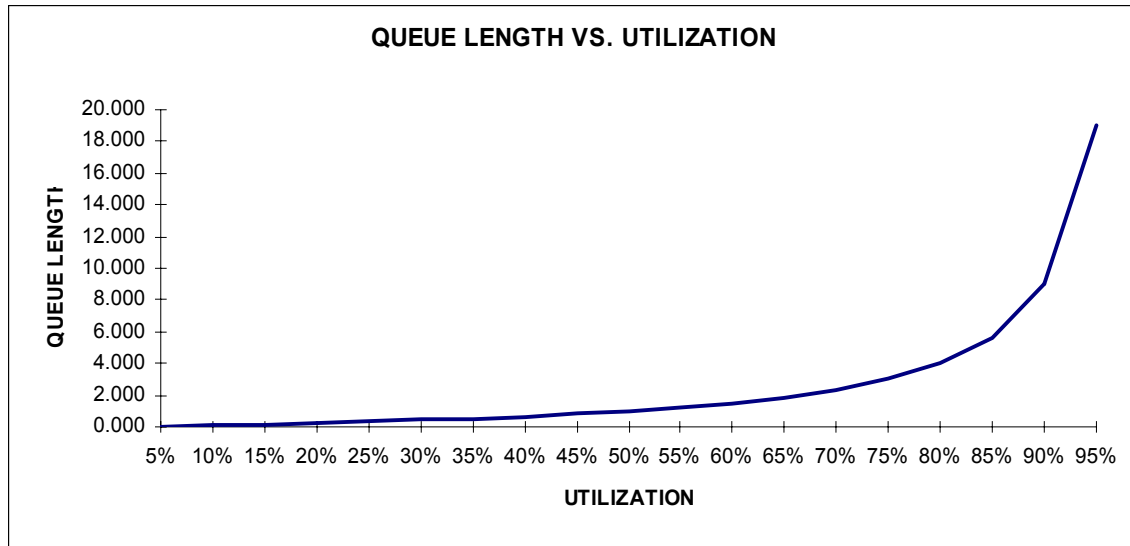
Now let's look at the case of the table scan. Here you will not be generating tens or hundreds of reads, but 10,000 or 100,000 or 1,000,000's of reads. In this case SQL Server is more efficient and will read multiple pages at a time, but queries that would take 10s of seconds on a well-tuned I/O subsystem could take 100s of seconds. A difference of 4x in read performance is quite noticeable.

This is why we are very concerned about I/O performance when we are tuning a SQL Server system. When a query takes excessive time to read due to a poorly tuned I/O subsystem, it may not only be affecting itself, but may be blocking other users as well. This leads to blocking and even to deadlocks. So, a well tuned I/O subsystem is critical.

In order to properly tune the I/O subsystem you must first learn about what affects the performance of the I/O subsystem. The smallest distinct piece of the I/O subsystem is the disk drive. The performance is also affected by your choice of RAID level. In this article you will see why this is so, and how to tune it.

DISK DRIVE PERFORMANCE

A standard 15,000 RPM disk drive is a finite component, only able to perform a finite number of IOPS (I/O's Per Second) without experiencing performance problems. This performance limitation is basically caused by the number of seeks per second that the disk drive can do. A top of the line disk drive takes approximately 6 ms (milliseconds) on average to move from where the disk heads currently are to where the desired data is. This information is based on disk drive specifications and experimental data. The 6 ms seek latency corresponds to 166 IOPS. However, when you get near to this limit, disk queuing occurs and latencies can increase exponentially. This is shown in the graph below.



In order to keep I/O performance reasonable, you should not exceed 75% of the maximum capacity of the disk drive (166 IOPS), which is approximately 125 IOPS. So, I/O tuning can be related to sizing. If your disk drives are configured to run within the specified limits, performance problems will be reduced.

If the number of IOPS issued to each disk drive exceeds the capacity, the latencies (response time) will increase. In fact, in practice it is not uncommon to see the normal 10ms-20ms latency increase to 40ms, 100ms or even worse in overloaded disk subsystems. These extreme latencies can significantly affect performance.

RAID PERFORMANCE

In order to size your disk drives properly you must also keep in mind the overhead incurred by RAID striping. Most systems use some sort of RAID striping in order to avoid the loss of data in the event of the loss of a disk drive. RAID overhead usually does not come into play during read operations, but the write overhead for RAID 1, and RAID 0+1 is 2x and for RAID 5 is 4x. Thus if you are calculating the required number of IOPS for RAID subsystems you must take into account the RAID overhead.

RAID Level		Notes
RAID 0		No fault tolerance, one fault and you're out.
RAID 1 or RAID 0+1	2 x (1 logical I/O = 2 physical I/Os)	Good fault tolerance.
RAID 5	4 x (1 logical write = 2 physical reads and 2 physical writes)	OK fault tolerance. Poor write performance. Most economical.

Note: No RAID levels incur any read performance. The overhead that you incur by using RAID is only reflected in writes.

Tuning the I/O Subsystem

To tune the I/O subsystem you must make good choices. These choices are; how many disk drives do you need and how they are to be configured. Choosing the correct RAID level and the number of disk drives is the most important thing that you can do when designing your system. In addition, I/Os must be monitored in order to determine if there are problems, and if there are problems, the I/O subsystem should be modified.

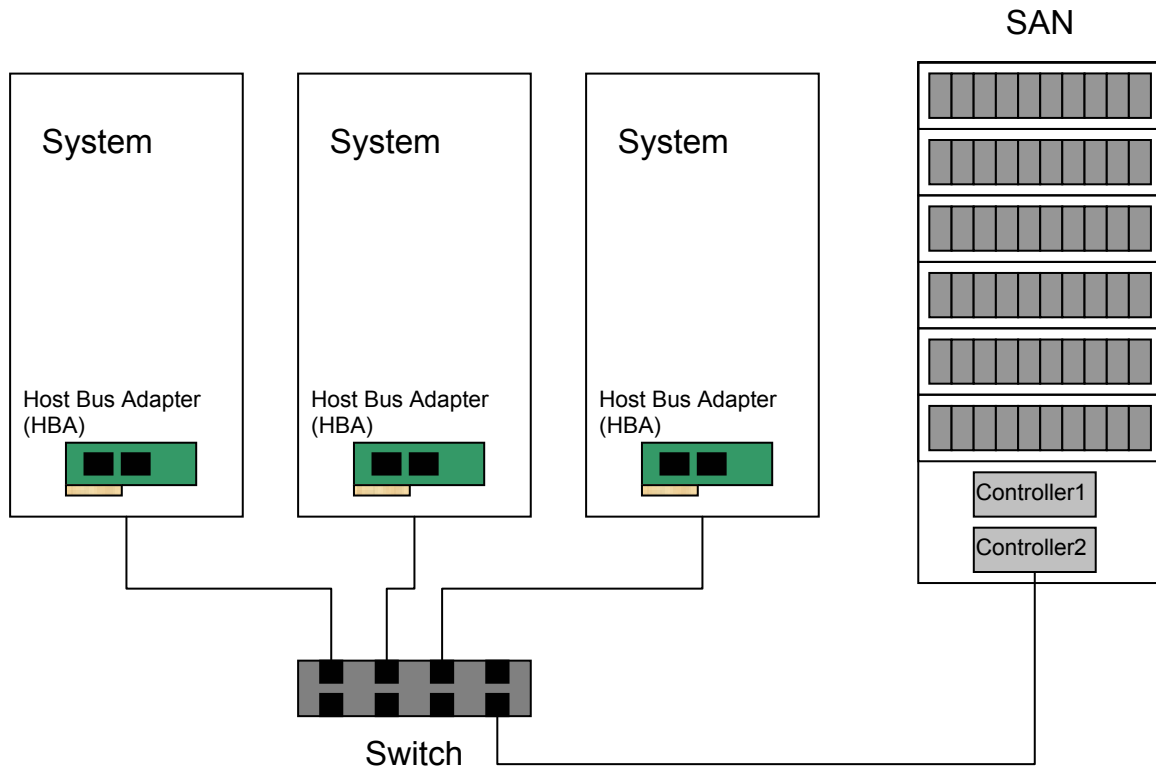
Of course there are other factors, such as RAID stripe size, RAID controller CPU utilization, controller caches, data file placement, etc., but they are beyond the scope of this article.

SAN AND NAS SYSTEMS

The SAN and NAS systems provide different types of storage from traditional internal or direct attached storage. A SAN (Storage Area Network) is designed to provide access to storage over a private fibre channel network. A NAS (Network Attached Storage) is used to provide storage over a standard network.

SAN SYSTEMS

A SAN system is an external storage system that allows multiple computer systems to access the same storage. The RAID controller inside the external storage system is able to take requests for different logical volumes within the storage system from different HBAs (Host Bus Adapters). This allows for several different features. One of the most common uses of a SAN is for storage consolidation. This is where multiple systems share the disks in the external storage subsystem. This allows for consolidation of storage resources and management. An example of this is shown in the figure below.



A SAN System

With storage consolidation, even though the storage in the external disk subsystem is shared among the different systems, it is not entirely accessible to all systems. Logical disks are carved out of the physical disk drives and allocated to each of the computer systems. Only one system can access a particular disk volume.

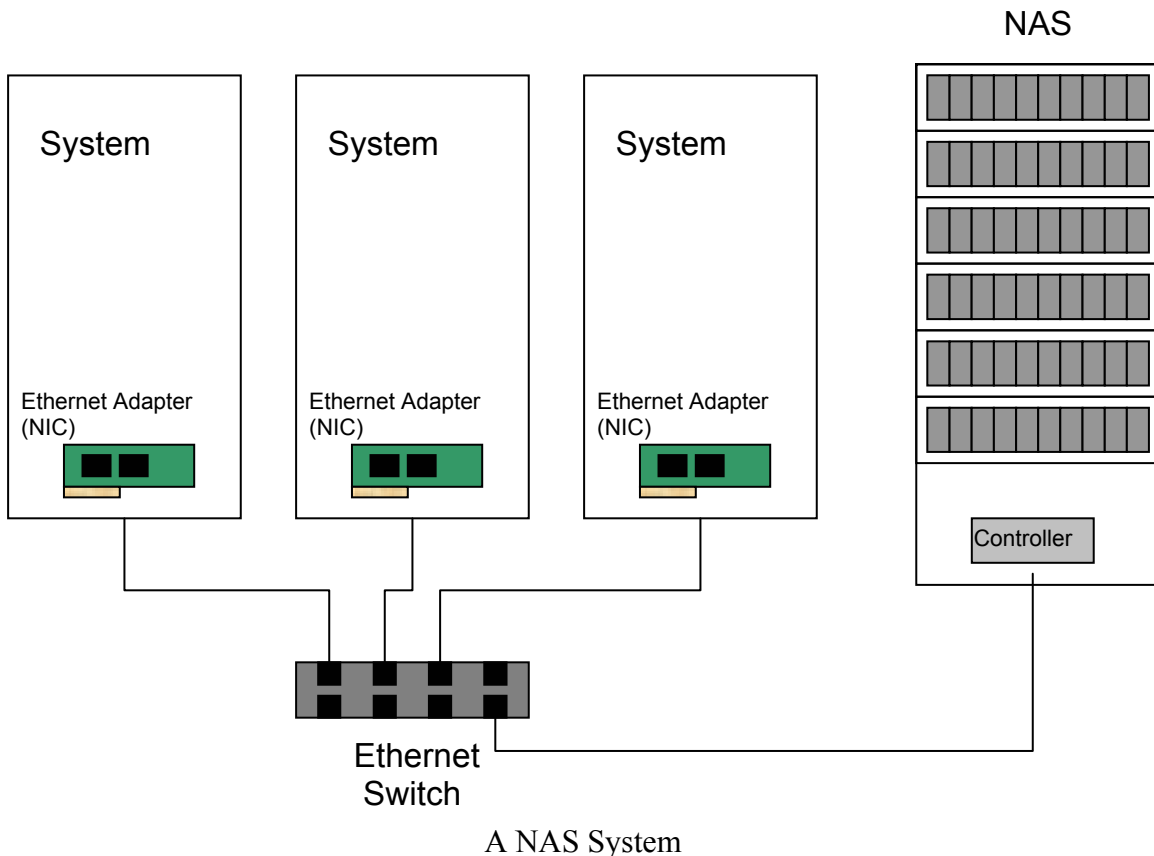
Another use for a SAN system is for clustering. Failover clusters use a shared disk subsystem that allows one of two systems to access the same storage. Even though the storage is accessible by both systems to access the storage, only one will be used at the same time. In a failover cluster, if one server were to fail, the second server could resume operation by taking control of the storage and the SQL Server database that resides on that shared storage.

SAN Performance Considerations

When putting together a SAN system you must not only look at the I/O traffic that is being generated by one system, but the I/O traffic being generated by all systems in the SAN. So in addition to sizing the I/O subsystem disk drives and RAID levels, you must look at the traffic on the SAN itself, as well as keeping in mind that other systems might be accessing the same RAID controllers. It is also possible to run into bandwidth limitations on the SAN itself, since fibre channel has a limited bandwidth.

NETWORK ATTACHED STORAGE (NAS) SYSTEMS

Network attached storage is similar to the SAN system in that the brains of the storage is external to the computer system. However, unlike the SAN system where the storage is connected via a fibre channel connection, a NAS system is accessed via the network. This is illustrated here.



Although the NAS system is supported under SQL Server, the NAS system usually cannot support the performance required by SQL Server unless you use a sufficiently fast network interface. The speed of the NAS is usually limited by the speed of the network interface.

NAS Performance Considerations

The NAS is different from the SAN in that multiple systems can access the same NAS storage simultaneously, whereas in a SAN, typically only one system accesses a particular filesystem at a time. So, the NAS may have performance problems related to too many people accessing the storage at the same time.

However, the main issue with NAS is that the path that the I/O must take typically is much longer than that of a SAN. This is due to the fact that the NAS has network protocols involved that must be processed as well as filesystem overhead. So, in a NAS environment it may not be possible to achieve 10ms I/O accesses, depending on the NAS. This might adversely affect your performance.

SUMMARY

The basic performance problems and solutions of running SQL Server in a SAN/NAS environment are related to the I/O capacity of the disk drives, the fibre channel network, the LAN and software overhead. Rigorous tuning should be done on a regular basis in order to make sure that the capacity of the I/O subsystem is not overloaded. In the case where the I/O subsystem is overloaded you may need to add disk drives, change RAID levels, add RAID controllers or a combination of all of these. SQL Server performance is all about I/O performance. I/O performance is mainly affected by RAID level and by the number of disk drives and the rate the I/O subsystem is driven at. An overloaded disk subsystem can cause severe performance problems.

About the Author

Edward Whalen is vice president and principal consultant at Performance Tuning Corporation (www.perftuning.com). Performance Tuning Corporation provides database performance tuning, load testing and troubleshooting services on MS SQL Server. Edward Whalen was a co-author on for SQL Server books from Microsoft Press; SQL Server 7 Administrator's Companion, SQL Server 7 Performance Tuning Technical Reference, SQL Server 2000 Administrator's Companion and SQL Server 2000 Performance Tuning Technical Reference. Edward Whalen has also authored four Oracle books. Edward Whalen is considered a leader in database performance tuning.