ORACLE9I I/O CONFIGURATION IN A SAN/NAS ENVIRONMENT

Edward Whalen, Performance Tuning Corporation
August 2001

INTRODUCTION

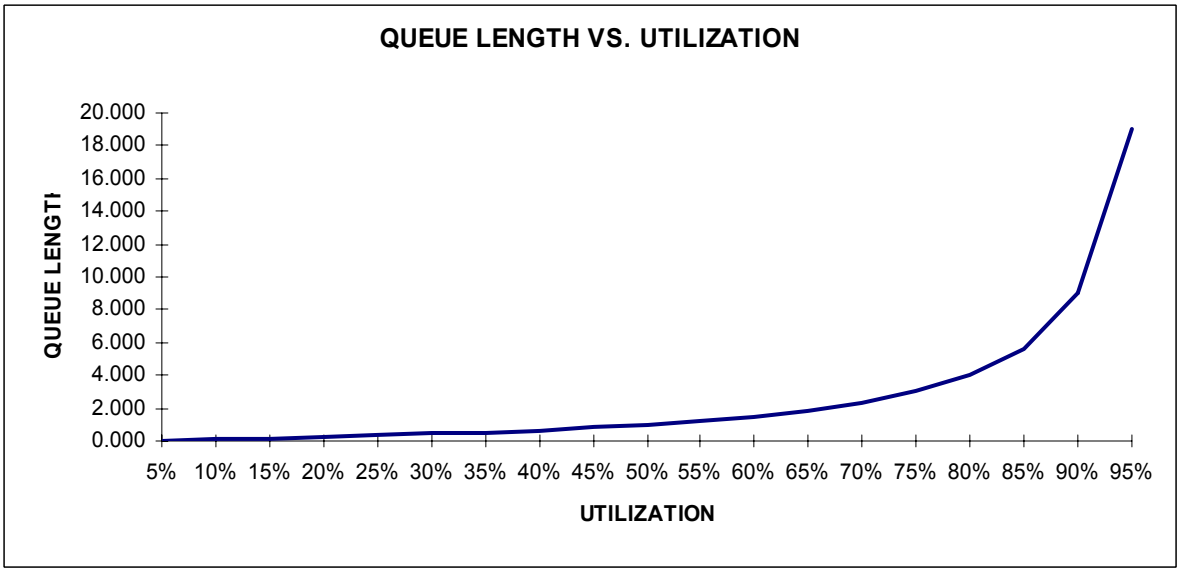
Some of the most common Oracle performance problems involve the I/O subsystem. Since Oracle's main function is to manipulate data, and that data resides either in memory or on the I/O subsystem, any I/O performance problems will result in Oracle performance problems. Much of the design of the Oracle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

In this paper the fundamental concepts and tuning of 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 paper 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.

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

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 apx. 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

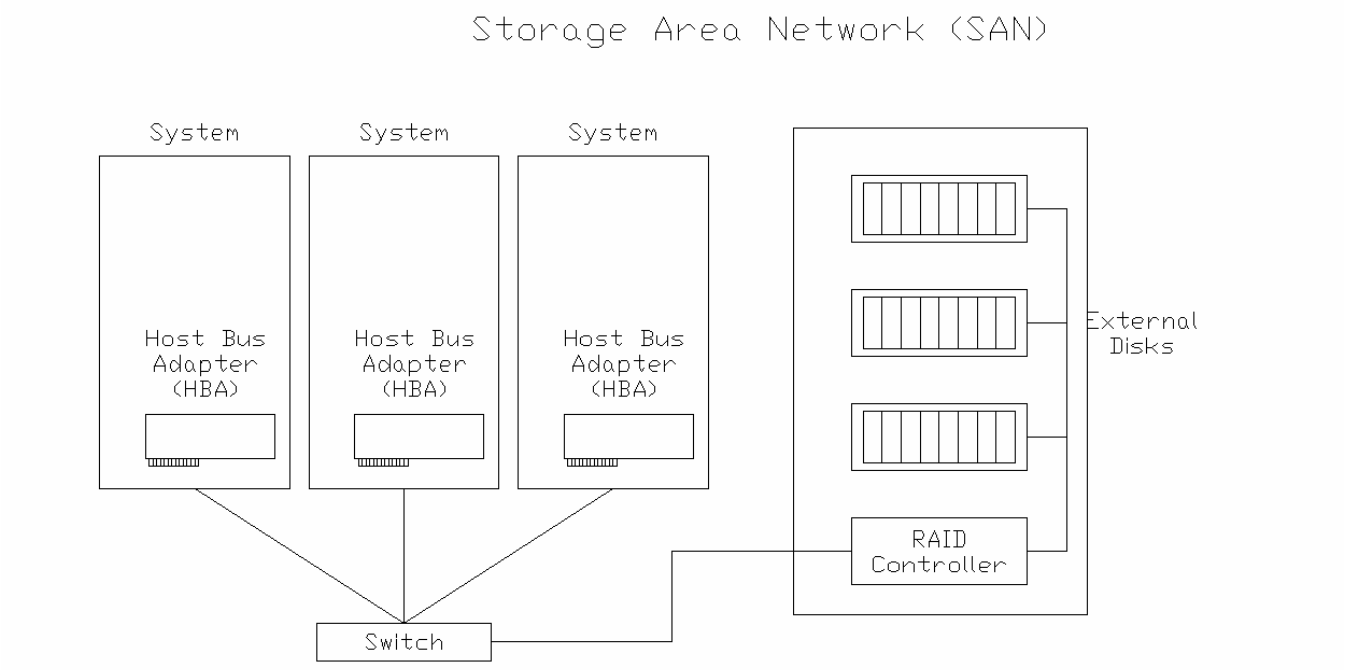
In order to size your disk drives properly you must also keep in mind the overhead incurred by RAID striping. Most systems use some sort of RAID striping in order to avoid the loss of data in the event of the loss of a disk drive. RAID overhead usually does not come into play during read operations, 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.

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. There are several different types of clusters: failover clusters and shared disk clusters. These different clusters use the SAN in slightly different ways and have different requirements.

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 Oracle database that resides on that shared storage. Oracle Fail Safe for Windows NT is a failover cluster that Oracle uses.

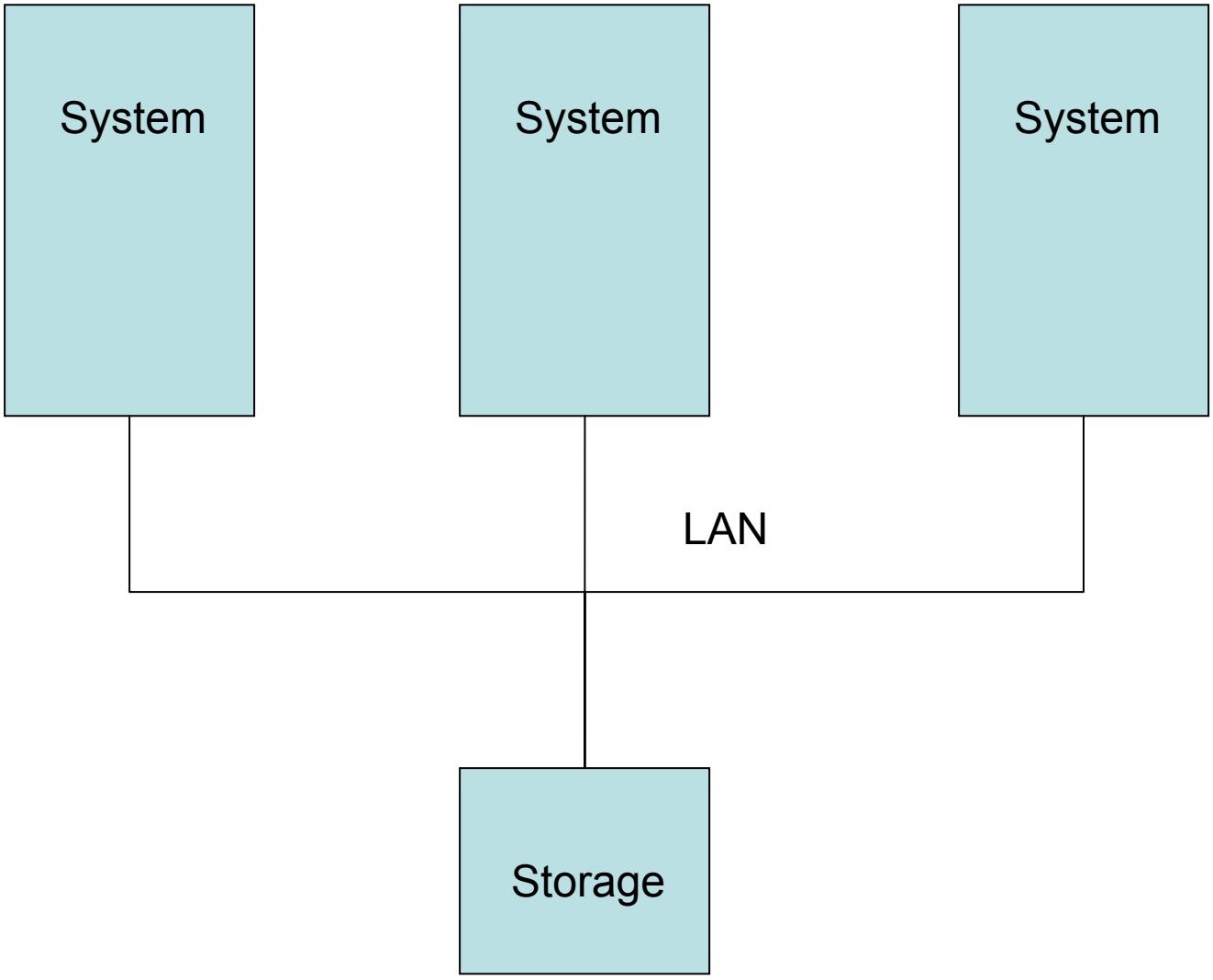
Oracle9i RAC (Real Application Clusters) is the latest version of Oracle Parallel Server (OPS). With an OPS system multiple Oracle instances access and use the same database (and datafiles). Oracle Parallel Server (OPS) and Oracle9i Real Application Clusters (RAC) both require a shared disk subsystem. Typically a SAN is used for this shared disk subsystem. With OPS and RAC the same disk subsystem must be accessible simultaneously from multiple servers as shown here.

OPS and RAC systems are used for both failover and for performance. Since multiple systems can access the same database, you can achieve more performance than a single system. RAC is not covered here.

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.

Network Attached Storage (NAS)



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

SUMMARY

The basic performance problems and solutions of running Oracle 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.

About the Author

Edward Whalen is vice-president and founder of Performance Tuning Corporation (www.perftuning.com), a consulting company specializing in database performance, administration, and backup/recovery solutions. He has extensive experience in database system design and tuning for optimal performance. Mr. Whalen has written two books on the Oracle RDBMS and four books on MS SQL Server and is currently completing an Oracle9i book. He is recognized as a leader in Oracle performance tuning and optimization.