Storage Update and Storage Best Practices for Microsoft Server Applications

Dennis Martin

President, Demartek

January 2009

Agenda

- Introduction
- Storage Technologies
  - Storage Devices
  - Storage Interfaces
  - RAID
  - Storage Architectures
- Microsoft Server Application Best Practices and Recommendations

Introduction

- Storage is where the data resides
- Access to storage in Exchange Server, SQL Server and SharePoint Server is of critical importance
- Understanding storage technologies is important

Storage Devices

- Solid State Disk (SSD)
- Flash memory (removable, non-volatile)
- Magnetic disk
- Removable disk
- Others not discussed today
  - Optical disk
  - Magnetic tape

Solid State Disk (SSD)

- Uses memory technology designed to appear as an online storage (disk) device
  - DRAM, NAND flash or combination
- Extremely fast
- Capacities vary from 8 GB to 1 TB+
- Expensive (although prices dropping)
- DRAM-based storage almost always includes battery-backup and disk-drive for safety

DRAM SSD Details

- IOPS (I/O per second) range from 70K to 3M+
- Latencies measured in microseconds
- Ideal for:
  - Database indexes
  - TempDB
- Can be used as a cache in front of other storage

NAND Flash Memory

- Non-volatile (mostly)
- Removable
- Often used for desktop computer data transport (USB flash drives, etc.)
- Also used in consumer devices (cameras, MP3 players, etc.)
- IOPS of 10K – 35K reads, writes are slower
- Capacities available up to 256 GB
- Variety of interfaces

NAND Flash Memory Types

- Single-Level Cell (SLC)
  - One bit per cell, faster, lower capacity
  - Lower error probability and longer life (100,000 cycles)
  - More suited to enterprise-class applications
- Multi-Level Cell (MLC)
  - Multiple bits per cell, slower, higher capacity
  - Higher error probability and shorter life (10,000 cycles)
  - More suited to consumer applications

SSD Comments

- Various form-factors:
  - 2.5 inch disk drive
  - Directly in PCI-Express bus slot
  - Other internal and external connection types
  - Consumer devices: USB, camera cards, etc.
- Expect some overlap between enterprise-class and consumer-grade technologies
- NAND flash is quiet, low-power, low-weight, low-heat
- Current issues: usable life of NAND-flash
- New memory technologies on the horizon
- Is SSD Technology Ready for the Enterprise?
  www.demartek.com/Demartek_SSD_Article_2008-12-19.html

Magnetic Disk

- Disk drive technology is well-known
- Market requirements dictate differences in drive types, from enterprise to consumer devices
- Fast
- Good pricing with steady price declines
- Price decline curve steeper than tape

Disk Drives: Enterprise

- Rotation speeds of 10K and 15K RPM
- Dual processors
- Command Queuing (TCQ & NCQ)
- Can tolerate higher vibration in racks
- Designed for 7x24x365 operation
- Moving to 2.5 inch form factor
- MTBF: 1M+ hours
- Warranty: 5 – 7 years

Disk Drives: Desktop

- Rotation speeds of 5400 and 7200 RPM
- Single processor
- Native Command Queuing (SATA only)
- 3.5 inch form factor
- Large capacities (currently up to 1.5 TB)
- Some are only designed for 8x5 operation
- Less expensive than enterprise drives
- Warranty: 3 – 5 years

Disk Drives: Notebook

- Rotation speeds of 4200, 5400 & 7200 RPM
- Single processor
- 2.5 inch and 1.8 inch form factor
- Lower power consumption than enterprise or desktop drives
- Designed to handle higher physical shock
- Warranty: 1 – 5 years

Disk Drives: Consumer

- Rotation speeds of 3600 & 4200 RPM
- Single processor
- 1.8 inch and 1.0 inch form factor
- Very low power consumption
- Very light weight
- Smaller capacities
- Warranty: 1 – 3 years

Disk Drives: Comparison Chart

Device	Enterprise	Desktop	Notebook	Consumer
Avg. seek time	3 – 5 ms	8 – 11 ms	10 – 15 ms	12 – 15 ms
Xfer rate (MB/s)*	70 – 170	60 – 120	30 – 80	6 – 40
RPM (K)	10, 15	5.4, 7.2, 10	4.2, 5.4, 7.2	3.6, 4.2
Capacities	Large	Very large	Medium	Small
Processors	2	1	1	1
Cmd. Queuing	TCQ or NCQ	NCQ	NCQ	-
Power need	High	Medium	Low	Very low
Warranty	5 – 7 years	3 – 5 years	1 – 5 years	1 – 3 years

* Maximum device transfer rate in megabytes per second from drive surface to buffer. Sustained rates are lower. This is not the same as the interface transfer rate.

Disk Drives: Recommendations

- Microsoft Server Applications, only consider enterprise and desktop drives
- Enterprise drives = Performance
  They will run out of capacity before they run out of performance
- Desktop drives = Capacity
  They will run out of performance before they run out of capacity

Removable Disk

- Removable data cartridges
- Used primarily for backup and archiving
- Have hard disk drives inside
- Performance characteristics of hard disks
- Available in a variety of capacities
- Docking system is forward-compatible with future disk drive capacities
- Different than external SATA disk drives

Storage Interfaces

- SATA
- SAS
- Fibre Channel
- USB

Storage Interfaces: SATA

- Used in desktops, laptops and low-end servers
- Low-cost
- High-volume
- Used for SSD, hard drives, CD-ROM, DVD
- SATA (serial ATA) replaces the older parallel IDE and ATA interfaces
- Speeds available: 1.5 and 3 Gbits/sec

Storage Interfaces: SAS

- Used in higher-end workstations and servers
- Higher cost than SATA
- Lower volume than SATA
- Used for primarily for hard drives today
- Replaces older parallel SCSI interface
- SAS and SATA interfaces use similar connectors for interoperability
- Speeds available: 3 and 6 Gbits/sec

Storage Interfaces: Fibre Channel

- Designed for high-performance, high reliability and extended distance
- Can be used in a SAN
- Three configurations
  - Point to point
  - Arbitrated Loop
  - Fabric
- Speeds available: 1, 2, 4, and 8 Gbits/sec

Storage Interfaces: USB

- Available for a variety of devices
- Storage applications include portable data, backup data and consumer data
- Speeds available:
  - Low speed: 1.5 Mbits/sec
  - Full speed: 12 Mbits/sec
  - High speed: 480 Mbits/sec
  - SuperSpeed: ~5 Gbits/sec

SuperSpeed USB (USB 3.0)

- USB 3.0 – SuperSpeed USB
- Specification completed November 2008
- Approximately 10x faster than USB 2.0
- First SuperSpeed USB controllers to become available 2H2009
- SuperSpeed consumer products expected in 2010
- First devices to include storage devices
- More information: www.usb.org

Storage Interface Comparison

	SATA	SAS	Fibre Channel	USB
Number of devices	1	16K	16M	127
Maximum distance	1 meter	10 meters	100+ KM	5 meters
Cable type	Copper	Copper	Fiber Optic	Copper, wireless
Interface type	Serial	Serial	Serial	Serial
Transfer speeds (MB/sec)	150, 300	300, 600	100, 200, 400, 800	0.15, 1.5, 50, ~500*

MB/sec = Megabytes per second, which is generally calculated as megabits/second (Mbps) divided by 10 for planning purposes

* *SuperSpeed USB devices expected in 2010*

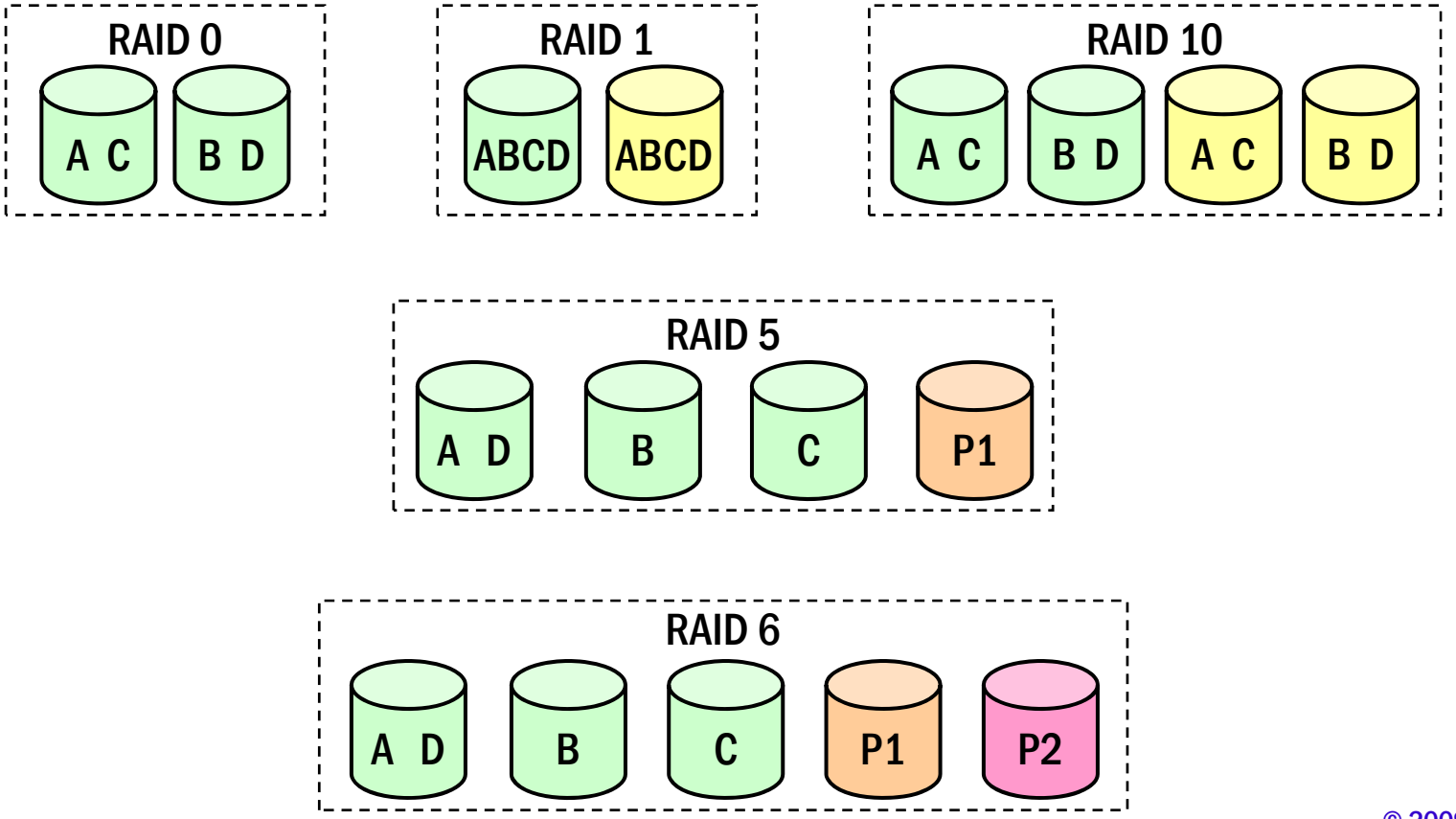
RAID

(**R**edundant **A**rray of **I**ndependent **D**isks)

- RAID 0: Interleaving or “striping” data across two or more disks
- RAID 1: Disk mirroring – same data written on two different disks (data can be rebuilt if drive fails)
- RAID 5: Data striping with parity across multiple disks (data can be rebuilt if drive fails)
- RAID 6: Data striping with double parity across multiple disks (data can be rebuilt if two drives fail)
- RAID 1+0 or RAID 10: combine RAID 1 and RAID 0

RAID Examples

Data: “ABCD”



RAID Comparison Chart

	Transactional I/O Performance	Capacity Utilization	Disk Failure and Rebuild Performance
RAID 0	Good	Best	Poor
RAID 1	Best	Poor	Best
RAID 5	Good	Good	Moderate
RAID 6	Good	Moderate	Good
RAID 10	Best	Poor	Best

RAID Recommendations

- Choose RAID to spread data over multiple disks (“spindles”) to get better performance and reliability than using individual disks
- Best overall performance: generally RAID10
- Best capacity (with recoverability): generally RAID5

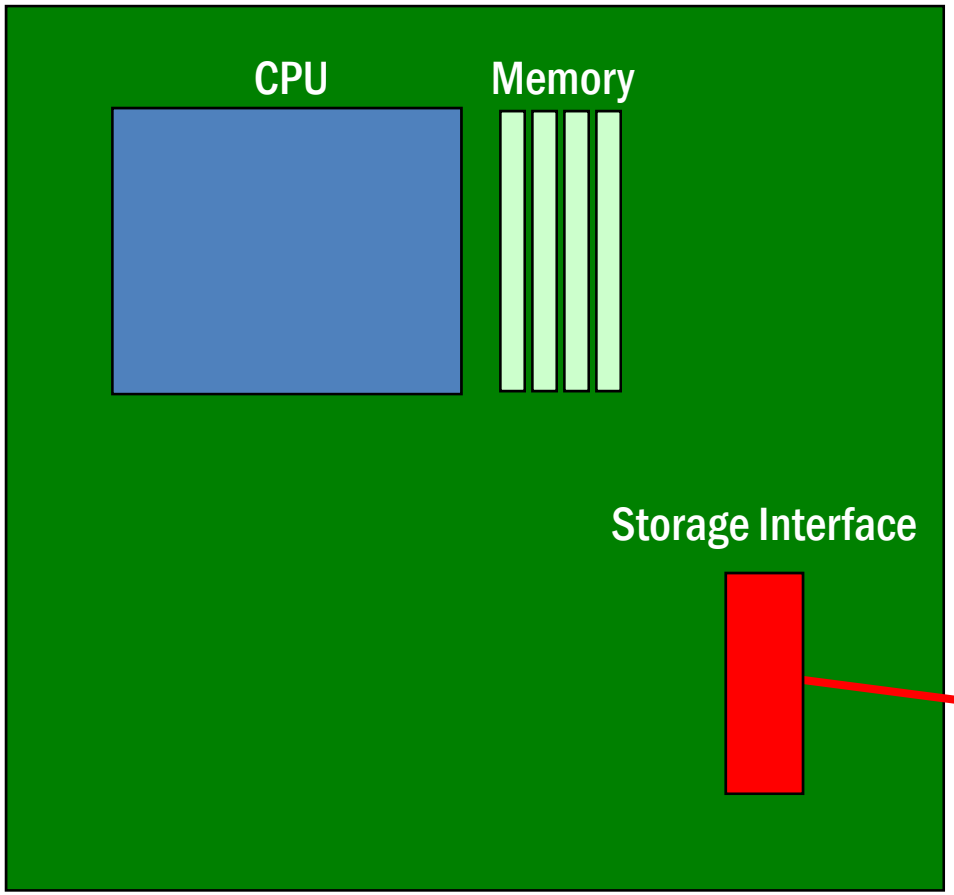
Storage Architectures

- Direct Attached Storage (DAS)
- Network Attached Storage (NAS)
- Storage Area Network (SAN)

Storage Architectures: DAS

- Storage is directly connected to CPU
- Short distance between CPU and storage (inches or small number of meters)
- Storage is “owned” by one host computer only
- Limited number of storage devices can be connected
- Common interfaces are SATA and SAS

DAS Diagram



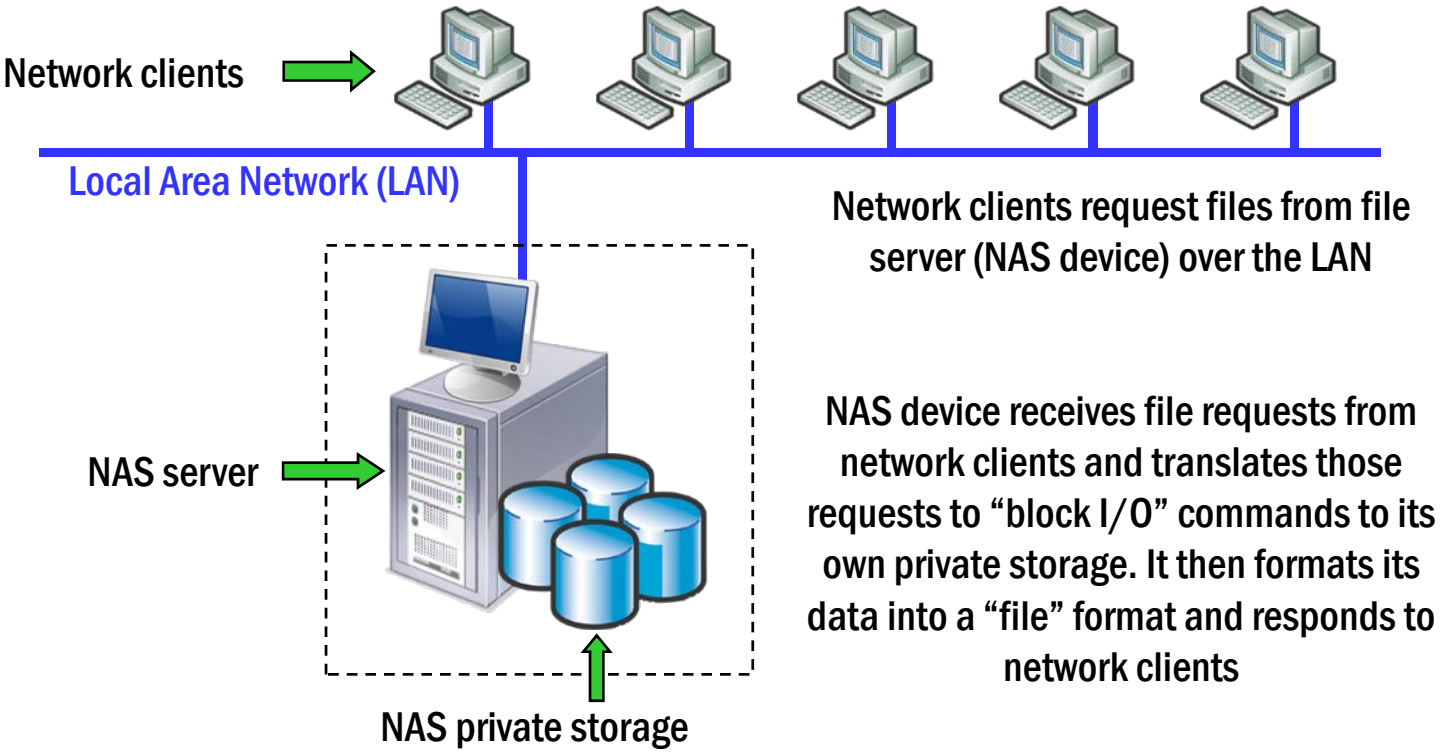
Due to cable length limitations, the storage devices often reside in the same cabinet as the CPU, or in a separate enclosure physically near the CPU cabinet

Host performs "block" I/O directly to device

Storage Architectures: NAS

- One host server shares its resources with other clients on the network
- Clients make requests by filename
- NAS server has its own storage, clients only see “share” names
- Large number of “shares” can be created
- NAS server can be a long distance from clients
- Common protocols are NFS and CIFS

NAS Diagram



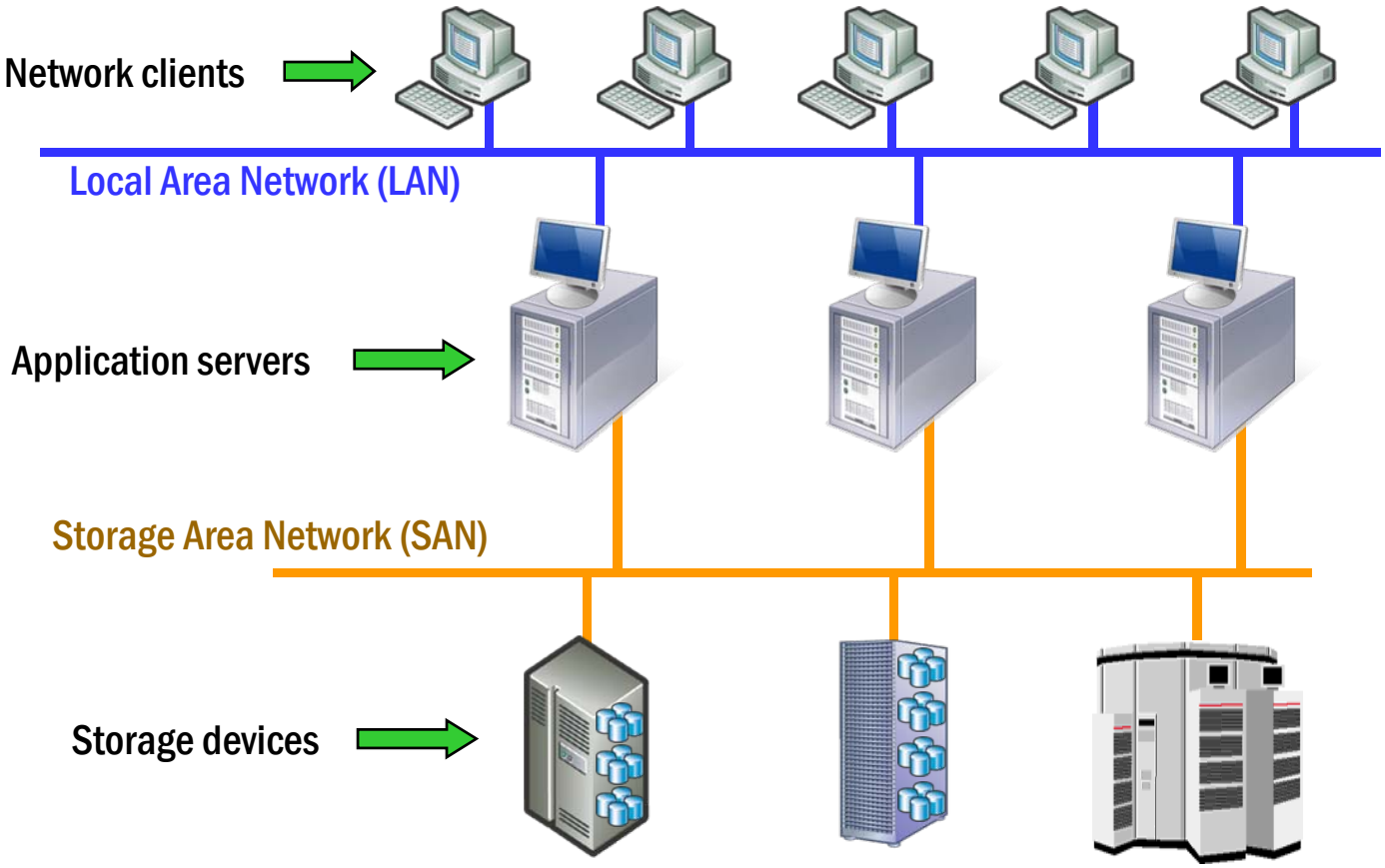
Network clients request files from file server (NAS device) over the LAN

NAS device receives file requests from network clients and translates those requests to “block I/O” commands to its own private storage. It then formats its data into a “file” format and responds to network clients

Storage Architectures: SAN

- Infrastructure allows computers to appear to have direct connection to storage
- Storage can be great distance from computers requiring access
- Clients make “block” I/O requests
- Can be millions of storage devices in a SAN
- Two technologies today: Fibre Channel and Ethernet (iSCSI)

SAN Diagram



Fibre Channel (FC) vs. iSCSI

- Each addresses a different market that has different needs with respect to performance, reliability, scalability and manageability
- Although there are different "plumbing" characteristics between FC and iSCSI, the applications storing data on them can't tell the difference

Fibre Channel (FC) Technology

- FC has been deployed in the vast majority of large IT datacenters
- FC is typically deployed where there is full-time storage staff within IT
- FC is designed for large enterprises
- FC uses HBAs, FC switches and other dedicated technology
- FC is a hardware-intensive technology
  - Host CPU utilization is low because of off-load

iSCSI Technology

- iSCSI is more appealing to small and mid-size IT shops
- iSCSI is typically deployed where there are not full-time storage people
- iSCSI uses existing Gigabit LAN infrastructure, with some caveats
- iSCSI: is it networking or storage? Who manages it?

Storage Architecture Recommendations

- Use DAS for simple applications requiring only one server
- Use NAS for file serving and sharing over a TCP/IP network
- Use SAN to share storage between servers and easy re-allocation of storage as needs change and grow

Futures

- As interface speeds increase, expect increased usage of fiber-optic cables and connectors for most interfaces
  - At higher Gigabit speeds, copper cables and interconnects become too “noisy” except for very short distances
- Expect more SAN-like types of architectures for interfaces such as SAS, USB and perhaps others

Microsoft Server Application Best Practices

- What does all this mean for Microsoft Server Applications?
  - Exchange Server
  - SQL Server
  - SharePoint Server (runs on SQL Server)

Recommendations

- Configure database servers with performance and availability as design criteria
- Use more disks and faster disks for best performance
  - If you choose SATA disk drives, you're emphasizing capacity above performance
- In a heavily read-oriented SharePoint portal site, prioritize data over logs

Exchange Server

- Consider performance before capacity
- Exchange Server is sensitive to disk read and write latencies
  - Exchange Server wants average read and write latencies < 20 msecs.
- Place Exchange logs on lowest latency disks
- Place databases and logs on separate RAID sets

SQL Server & SharePoint

- Ideally, use separate RAID sets for:
  - TempDB: RAID10 (write-heavy)
  - Transaction logs: RAID10 (write-heavy)
  - Search database: RAID10 (read-write mix)
  - Content databases: RAID10 (read-heavy)
- For best performance, the number of TempDB data files should equal the number of CPU cores in the server

Disk Alignment

- Windows Server 2003 or older:
  Align the file system to the disk offset recommended by the storage hardware vendor. If unknown use an offset of 64K.
  - Diskpart command:
    ```
    create partition primary align=64
    ```
- Windows Server 2008 uses default alignment of 1MB

Format Allocation Size

- Exchange Server
  - Databases: 64K
  - Logs: can use default size (typically 4K)
- SQL Server: use 64K for volumes dedicated to SQL Server
  - The SQL Server page size is 8K
  - SQL Server allocates disk from the operating system in units known as “extents” of 8 pages

Stripe Size

- Since SQL Server accesses disk storage in 64K blocks, the optimum disk array stripe size Microsoft Server Applications volumes is 64K
- If there are other applications using the same storage as SQL Server, you may decide to choose a different stripe size
- Similar recommendations for Exchange Server

References – 1

- SQL Server 2005 pre-deployment I/O Best Practices
  http://www.microsoft.com/technet/prodtechnol/sql/bestpractice/pdpliobp.mspx
- SQL Server 2005 I/O Basics
  http://www.microsoft.com/technet/prodtechnol/sql/2005/iobasics.mspx

References – 2

- **Exchange Server 2007 Storage Guidelines**
  http://technet.microsoft.com/en-us/library/bb124518.aspx
- **SharePoint Server 2007 Best Practices**
  http://technet.microsoft.com/en-us/office/sharepointserver/bb736746.aspx

Performance Comparisons

- SSDs did not arrive in time for this presentation
- Look for performance comparisons in February 2009

Contact me

Dennis Martin, President
Demartek
(303) 940-7575
dennis@demartek.com

www.linkedin.com/in/dennismartin