Sizing Exchange 2003 Servers

Steve Tramack
Sr. Solutions Engineering Manager
TSG Solution Alliances – Microsoft Solutions
Hewlett-Packard

Topics

- **Sizing / capacity planning process**
- Exchange server sizing and design
- Basic monitoring
- Summary

What's Out There?

- Benchmarks
  - …Are not the best source of sizing data
- Why?
  - Single server
  - Unrealistic configuration
  - One workload
  - No additional software or redundancies
- Intended to compare server models, not provide sizing guidelines

What Else Is Available?

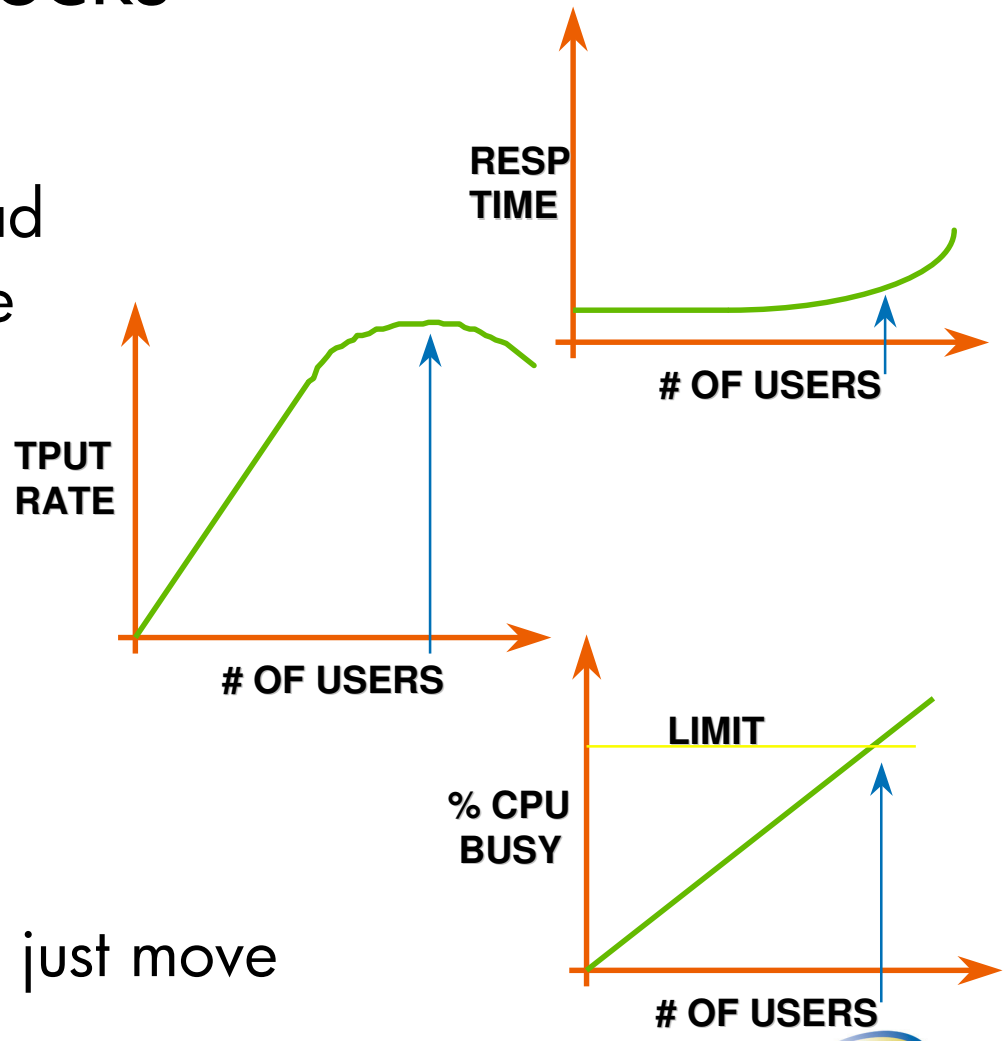
- Sizing tools
  - Varying levels of complexity and comprehensiveness
  - Risk of GIGO as solution grows
- White papers
  - Source reference materials
  - Based on lab testing or "notes from the field"
- Consultants / Solution Architects

Where Do I Start?

- Know thyself
  - Identify your definition of a "user"
  - Understand server and user workload characteristics
  - Identify best practices, business considerations, SLAs, planned architecture / topology
- Establish baselines
- Gather data
  - Personalized benchmarking
  - Understand peaks, averages, percentiles, patterns
  - Pilot, if possible, and monitor

Identifying bottlenecks

- Response time
  - Server's reaction to load
  - Only part of the picture
- Throughput
  - Messages/sec
  - Transactions/sec
- What is a bottleneck
  - High demand resource
  - Workload dependent
  - Never really eliminate; just move



Dealing with Constraints

- Size to peak or average?
- Key System Resources
  - CPU
  - Memory
  - I/O
    - Disks
    - Network
- Scale up versus scale out
- SLAs

Topics

Sizing / capacity planning process

Exchange server sizing and design
- Server role overview
- Mailbox server sizing guidelines
- I/O planning
- Tools

Basic monitoring

Summary

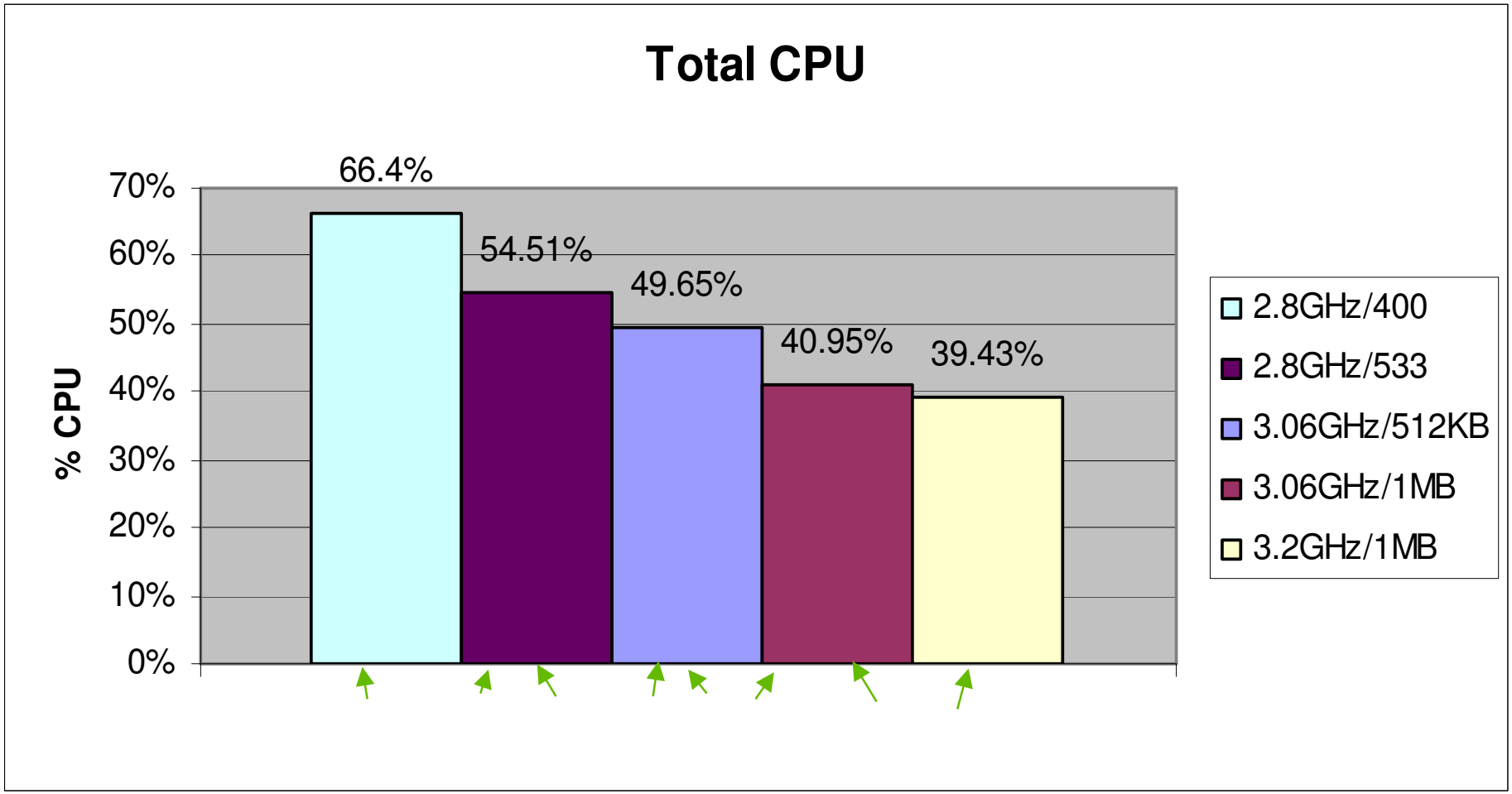
Exchange 2003 Server Roles

- Different requirements based on role
  - Mailbox servers
  - SMTP servers
  - X.400 / legacy connector servers
  - Front end servers (OWA / OMA)
  - Active Directory/Global Catalog
  - Expansion servers
  - Free/busy
  - Public folders

Sizing Exchange: CPU

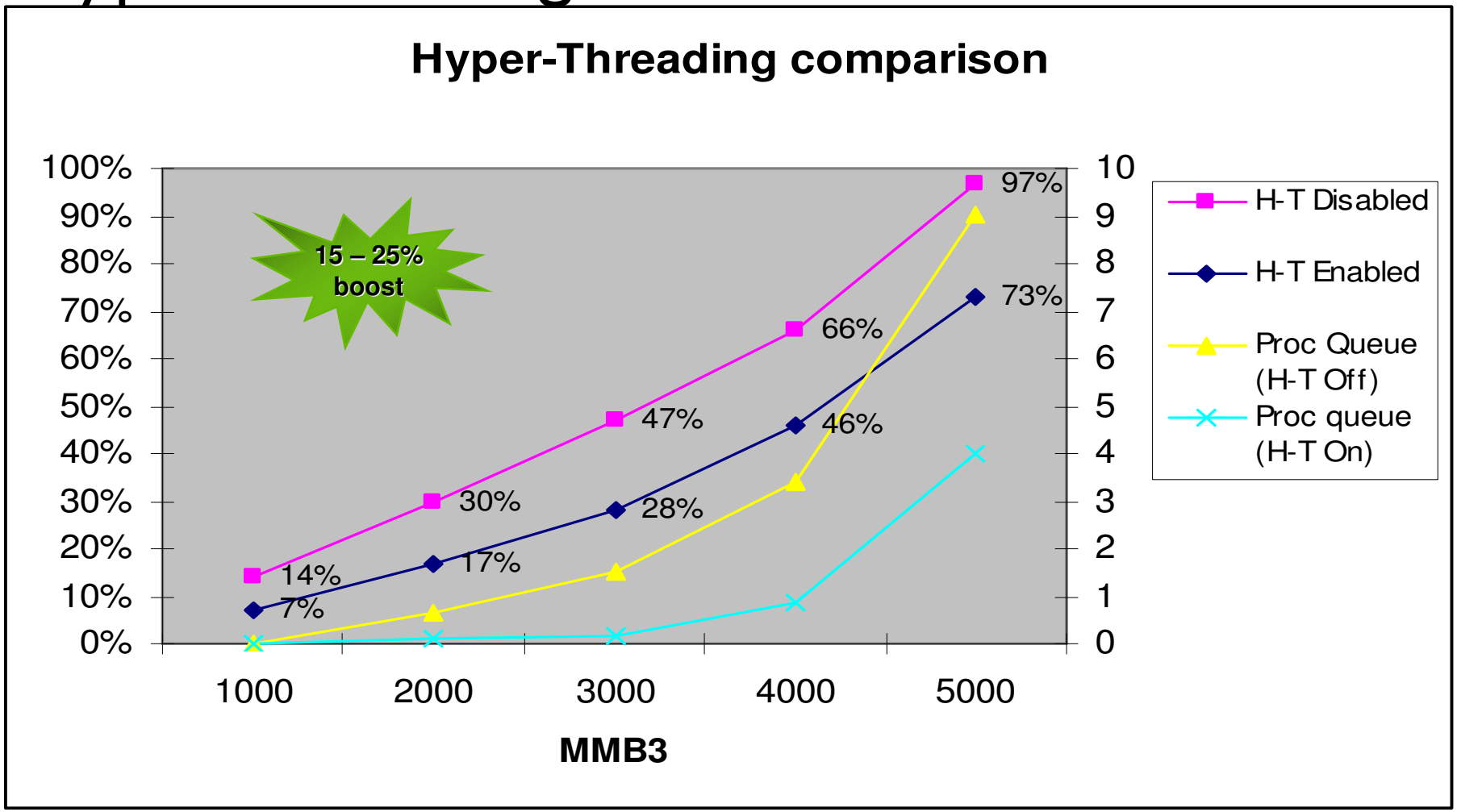
- Scalability
  - Pre-2002 processors: scale through 8P
  - Xeon MP / Opteron: 4P (8P w/SP1? Affinity?)
  - Sweet spot moving down the chain?
  - Goal: burn 80%...a challenge?
  - Check KB 827281
- Scalability factors
  - Hyperthreading impact
  - Clock frequency
  - L2, L3 cache
  - Front-side bus impact
  - Architecture (Xeon / Xeon MP / Opteron)

Frequency, L3 Cache And FSB



Platform: ProLiant DL360G3; 2P, Exchange 2003 RTM; identical configuration.
Workload: 3,000 MMB3. Response Time relatively consistent

Hyper-Threading Results



Platform: ProLiant DL580G2; 4P 2.0GHz/2M cache, Exchange 2003 RTM
Workload: MMB3

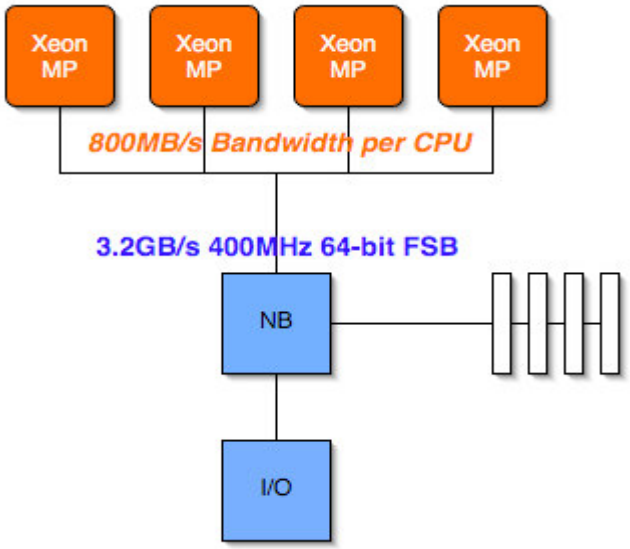
Opteron And Xeon MP Architecture

Xeon MP FSB architecture

All 4 procs share 64-bit connection to external North Bridge

Each proc has access to memory at 400MHz, with a max of 800 MB/s

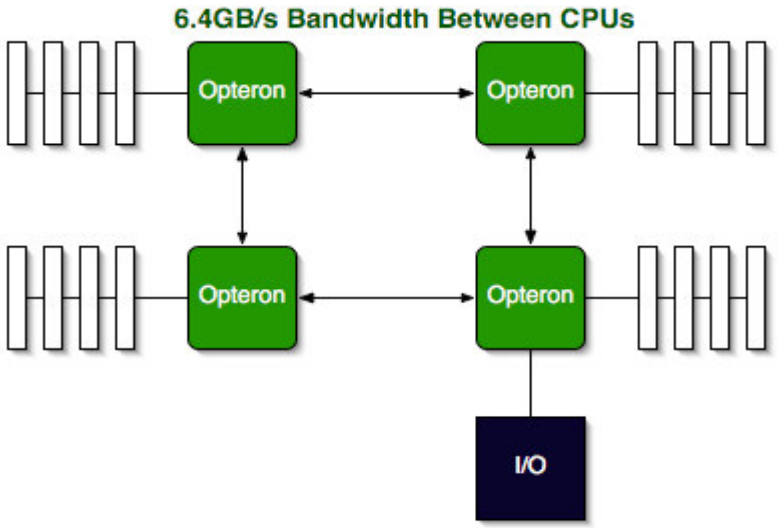
Compensate with larger L3 cache on die



Opteron architecture

North Bridge and memory controller integrated with each processor

Each proc includes 3 point-to-point hyperlinks; 6.4GB full duplex bandwidth



Sizing Exchange: Memory

- Physical versus Virtual
  - 4GB on high-end back-end servers (2,000 mailboxes and above)
  - No PAE / AWE
  - 64-bit extension support (Windows 2003 SP1)?
- /3GB (mbox and PF servers)
  - Boot switch when >1GB of RAM
  - Must be used with /USERVA (Windows 2003) 2970⇔3030 (SystemPages Regkey with W2K)
  - All editions of Windows 2003, Advanced Server and Datacenter of Windows 2000
    - REMOVE on Windows 2000 Standard Edition!
    - REMOVE if DC present on same system
  - Check KBs 815372, 810371

Sizing Exchange: Network

- Back-end
  - Typically, 100Mbit full duplex sufficient
  - Consider Gbit if
    - Network backup/restore
    - iSCSI or NAS storage – TOE support
    - High concentration of OWA / POP / IMAP users
  - Network Teaming
  - Consider NICs with IPSec offload if applicable
  - MAPI compression / buffer packing – 70% improvement in cached mode synchronization
- Front-end
  - Dual switched 100Mbit or Gbit default

Sizing Exchange: I/O The #1 Consideration

- I/O Profiles
  - EDB
  - STM
- The basics
  - Split logs from databases
    - Store: RAID 0+1 (recommended) or RAID 5
    - Log files: Recommend RAID 1 with write-back caching
    - Virtualization impact?
  - Size for capacity *and* I/O (subscribed vs. concurrent)
  - Design for Monday morning peak load

Storage Design: FE Servers

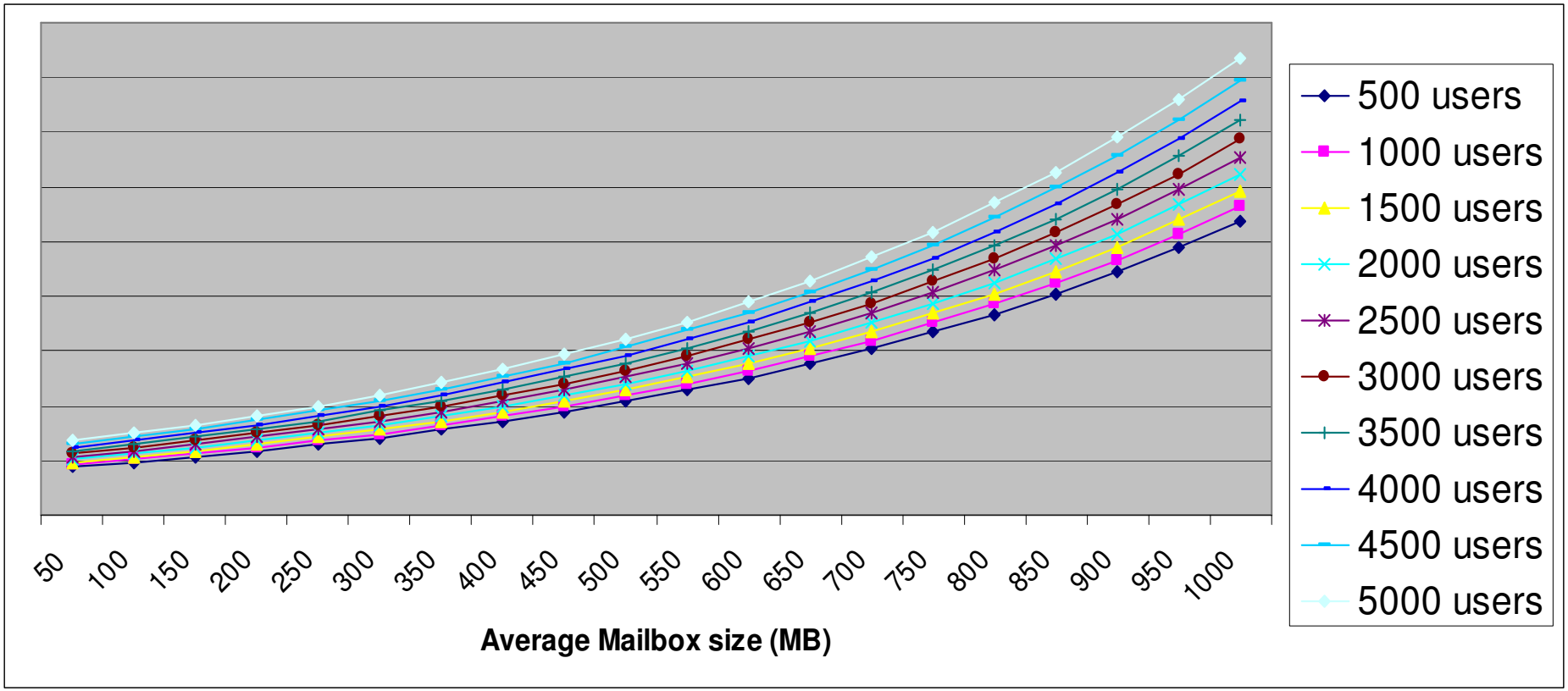
- SMTP
  - Create a single partition; ~30 small msgs / sec / spindle
  - RAID 0+1 recommended; 100% write-back cache
- X.400 and legacy connectors
  - Separate MTA (RAID 5), logs (RAID 1) and page file, if possible
- Other FE servers
  - Not I/O-intensive

Storage Design: BE Servers

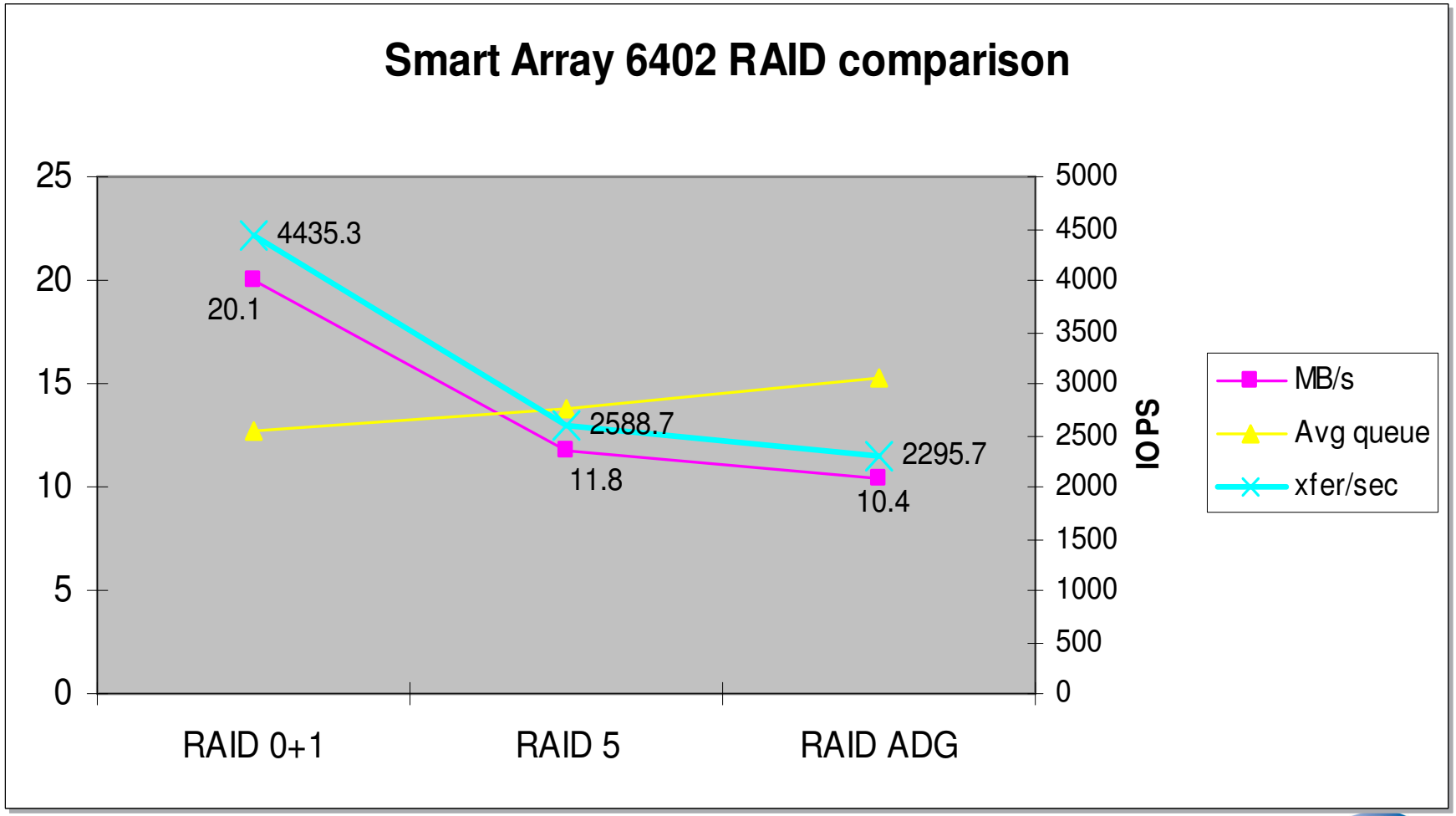
- Many considerations
  - IOPS/u rates
  - Array / LUN design
  - RAID level comparisons and read/write mix
  - Controller cache settings
  - Architecture (direct attached vs. Network)
  - …To mention a few
- TEMP/TMP to "fast" drive
  - Deferred content conversion (MAPI⇔MIME)
  - SMTP, move maibox, ISPs
  - Check KB 317722, 329067
  - In clusters, configure the value to the Cluster Service Account profile

Database IOPS/u Rates

- Exchange 2003 IOPS/u rates (MAPI user, 200MB mailbox)
  - ISP / "light" user: <.3 IOPS/u
  - Medium corporate: .5 IOPS/u
  - Heavy corporate: >.75 IOPS/u
  - Microsoft: 1.2 – 3 IOPS/u
- Rates do not stay constant as load, mailbox sizes increase

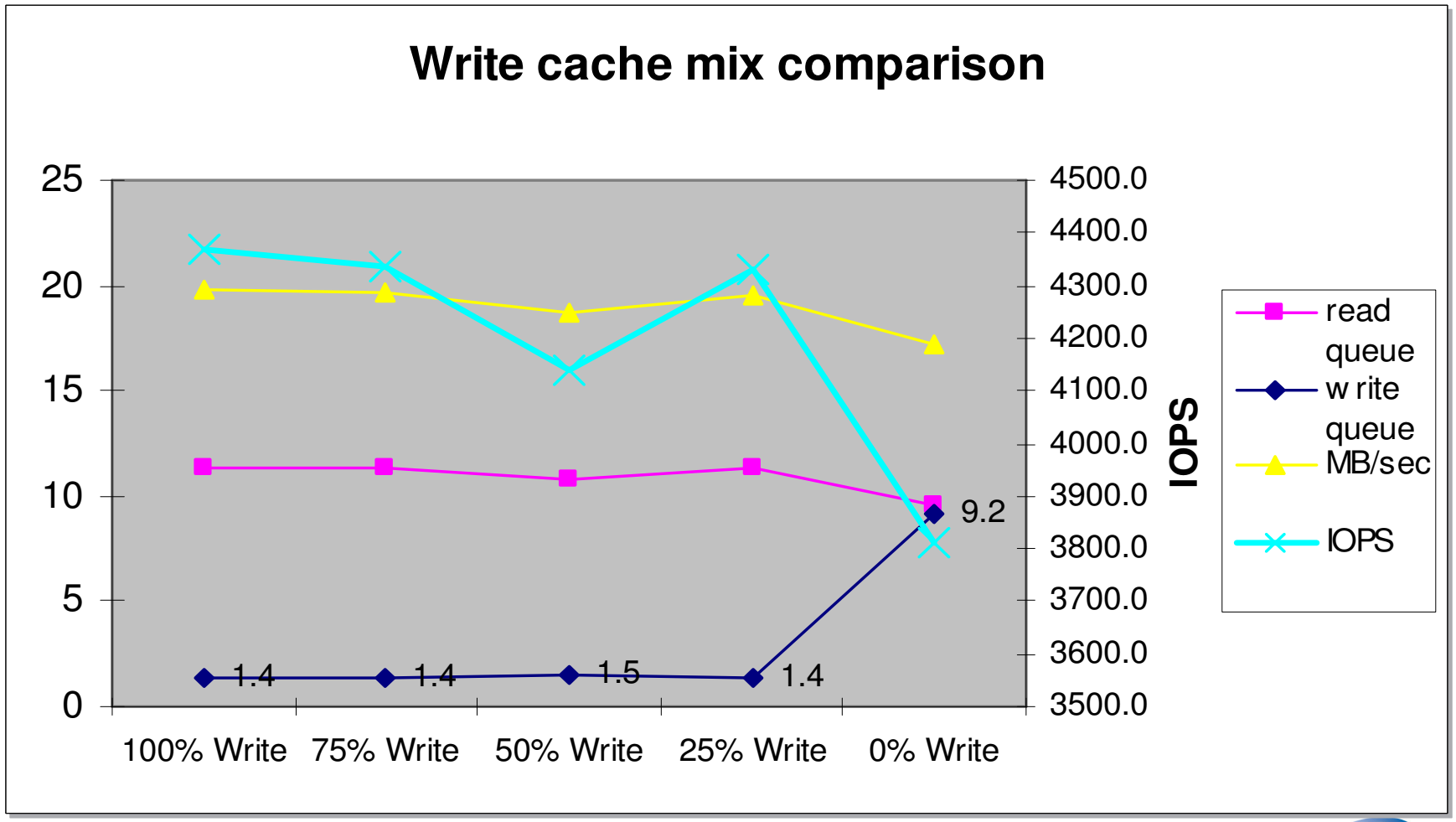


RAID Level Comparison



Smart Array 6402 controller; 12 disks/database; 1 database/LUN/array; consistent load generated by JetStress

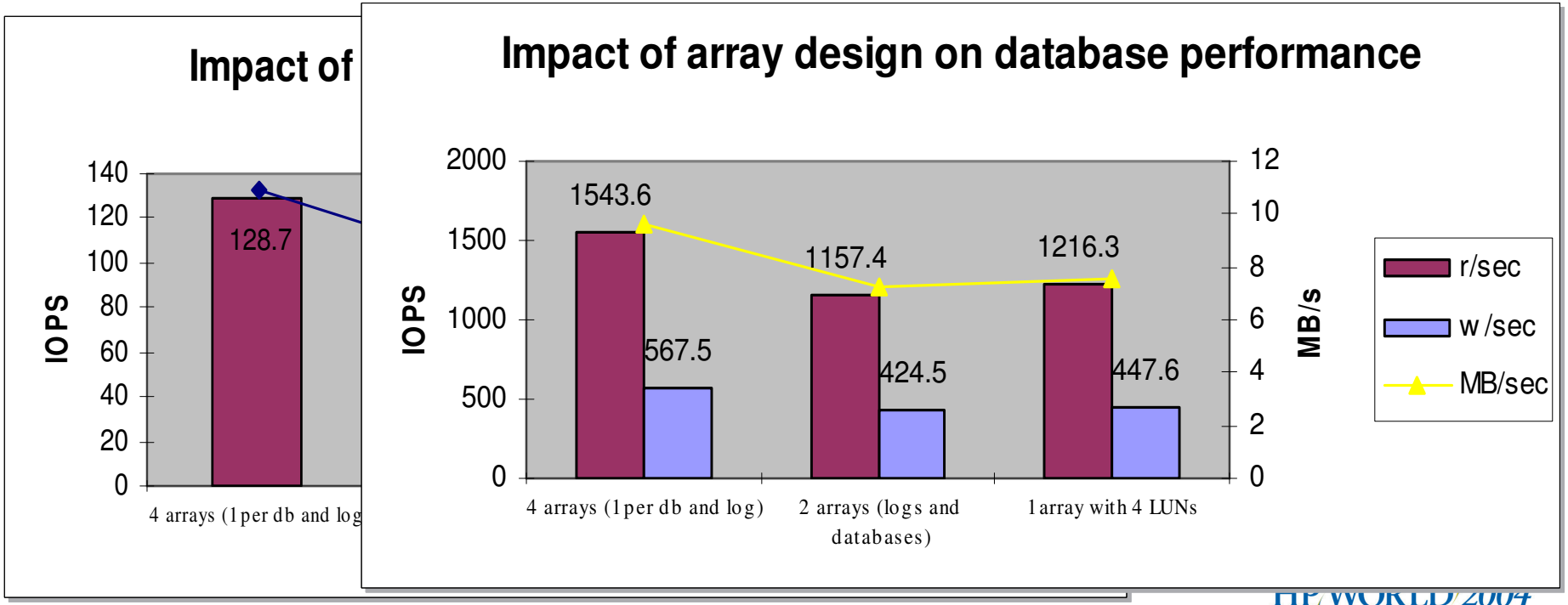
Controller Cache Comparison



Smart Array 6402 controller; 12 disks/database; 1 database/LUN/array; 75/25 R/W workload mix; consistent load generated by JetStress

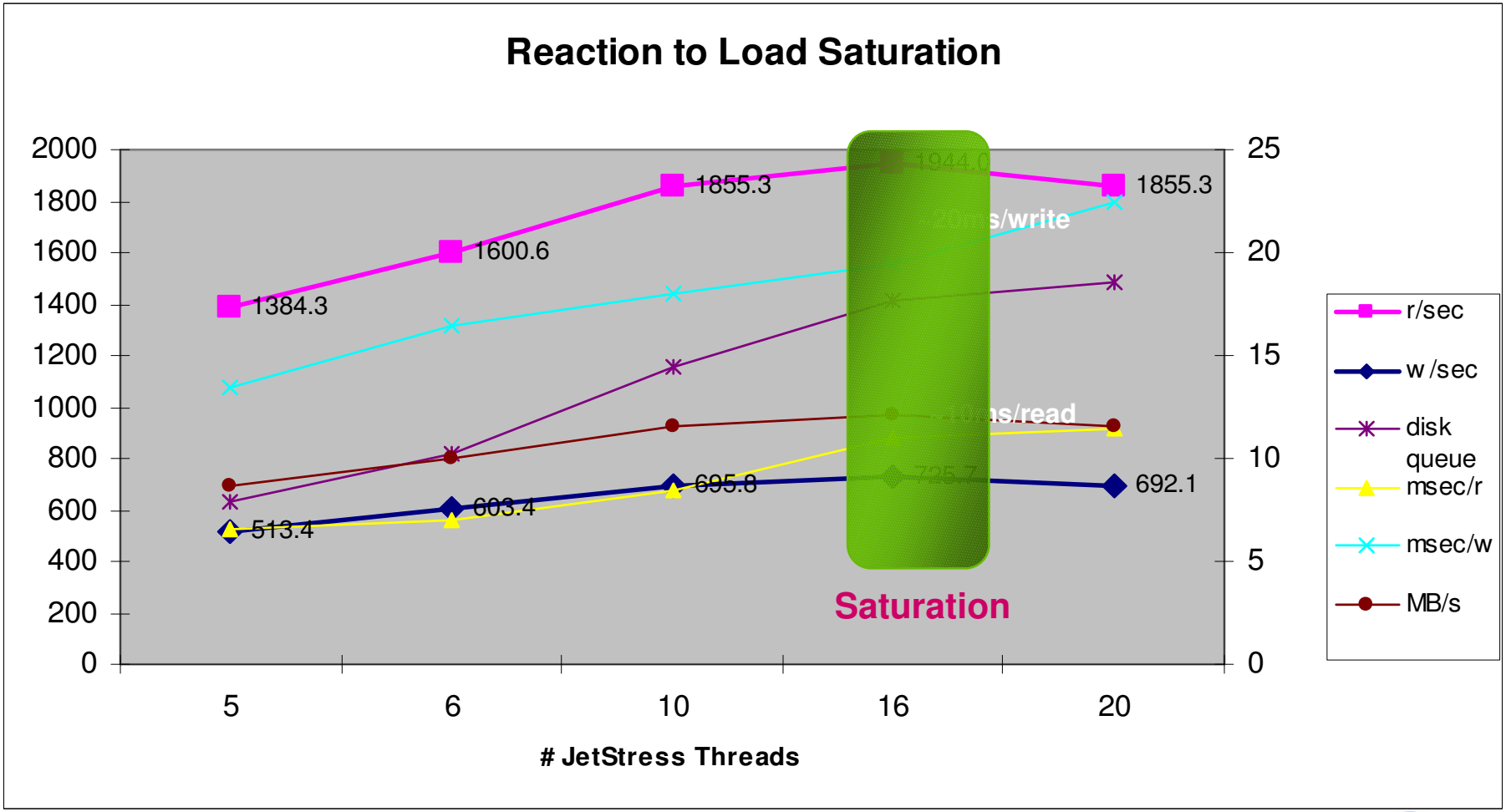
Array Design

- Log placement based on resiliency to failure
- Virtualization benefits (cross-referencing databases and logs in same disk groups) may outweigh penalty of mixing IO types
- Impact of Storage Groups and MDBs
  - Read/write mix, Loss of SIS ratio



Smart Array 6402; 12 disks/database; 1 DB/LUN/RAID 0+1 Array; consistent load generated by JetStress

Anatomy of a disk bottleneck



MSA1000; 16 disks/database; 1 DB/LUN/RAID 0+1 Array; increasing load to saturation generated by JetStress

Sample sizing examples

- IOPS vs Capacity
- Storage Planning Calculator
  - (free) Active Answers download
  - Exchange 2K vs 2003
- "Back of the Napkin"
  - 4,000 "medium" users
  - 8,000 "light" users
  - 4,000 "heavier" users

RAID Level and Capacity Planning BotN

4,000 "medium" users

IOPS

- 0.8 IOPS average; 1.0 IOPS *peak*
  - 4,000 users * 1.0 IOPS = 4,000 IOPS
  - 3:1 ratio: 3,000 read and 1,000 write IOPS
- Backend disk IOPS: RAID 10 calculation
  - Total IOPS = 3,000 + 2 * 1,000 = 5,000
  - 42 drives @ 10K RPM or 34 drives @ 15K
- Backend disk IOPS: RAID 5 calculation
  - Total IOPS = 3,000 + 4 * 1,000 = 7,000
  - 59 drives @ 10K or 47 drives @ 15K

Capacity

- 200 MB mailboxes
  - 4,000 users * 200 MB * 1.5 (rough calc for deleted items, maintenance, etc.) = 1,172 GB
  - 1,172 GB ÷ 66 GB/drive = 18 drives
- Backend disks: RAID 10 calculation
  - 18 drives * 2 (RAID 10) = 36 drives
  - Best match to 42 drives 73GB @ 10K RPM
- Backend disks: RAID 5 calculation
  - 18 drives * 1.2 (RAID 5) = 22 drives
  - Poor match to 59 (10K) or 47 (15K)

RAID Level and Capacity Planning BotN 8,000 "light" users

IOPS

- 0.3 IOPS average; 0.5 IOPS peak
  - 8,000 users * 0.5 IOPS = 4,000 IOPS
  - 3:1 ratio: 3,000 read and 1,000 write IOPS
- Backend disk IOPS: RAID 10 calculation
  - Total IOPS = 3,000 + 2 * 1,000 = 5,000
  - 42 drives @ 10K RPM or 34 drives @ 15K
- Backend disk IOPS: RAID 5 calculation
  - Total IOPS = 3,000 + 4 * 1,000 = 7,000
  - 59 drives @ 10K or 47 drives @ 15K

Capacity

- 200 MB mailboxes
  - 8,000 users * 200 MB * 1.5 = 2,344 GB
  - 2,344 GB ÷ 66 GB/drive = 36 drives
- Backend disks: RAID 10 calculation
  - 36 drives * 2 (RAID 10) = 72 drives @ 73GB
  - Would match 42 drives 146GB @ 10K RPM
- Backend disks: RAID 5 calculation
  - 36 drives * 1.2 (RAID 5) = 44 drives @ 73GB
  - Would match 47 drives 73GB @ 15K

RAID Level and Capacity Planning BotN 4,000 "heavier" users

IOPS

- 1.2 IOPS average; 1.5 IOPS peak
  - 4,000 users * 1.5 IOPS = 6,000 IOPS
  - 3:1 ratio: 4,500 read and 1,500 write IOPS
- Backend disk IOPS: RAID 10 calculation
  - Total IOPS = 4,500 + 2 * 1,500 = 7,500
  - 63 drives @ 10K RPM or 50 drives @ 15K
- Backend disk IOPS: RAID 5 calculation
  - Total IOPS = 4,500 + 4 * 1,500 = 10,500
  - 88 drives @ 10K or 70 drives @ 15K

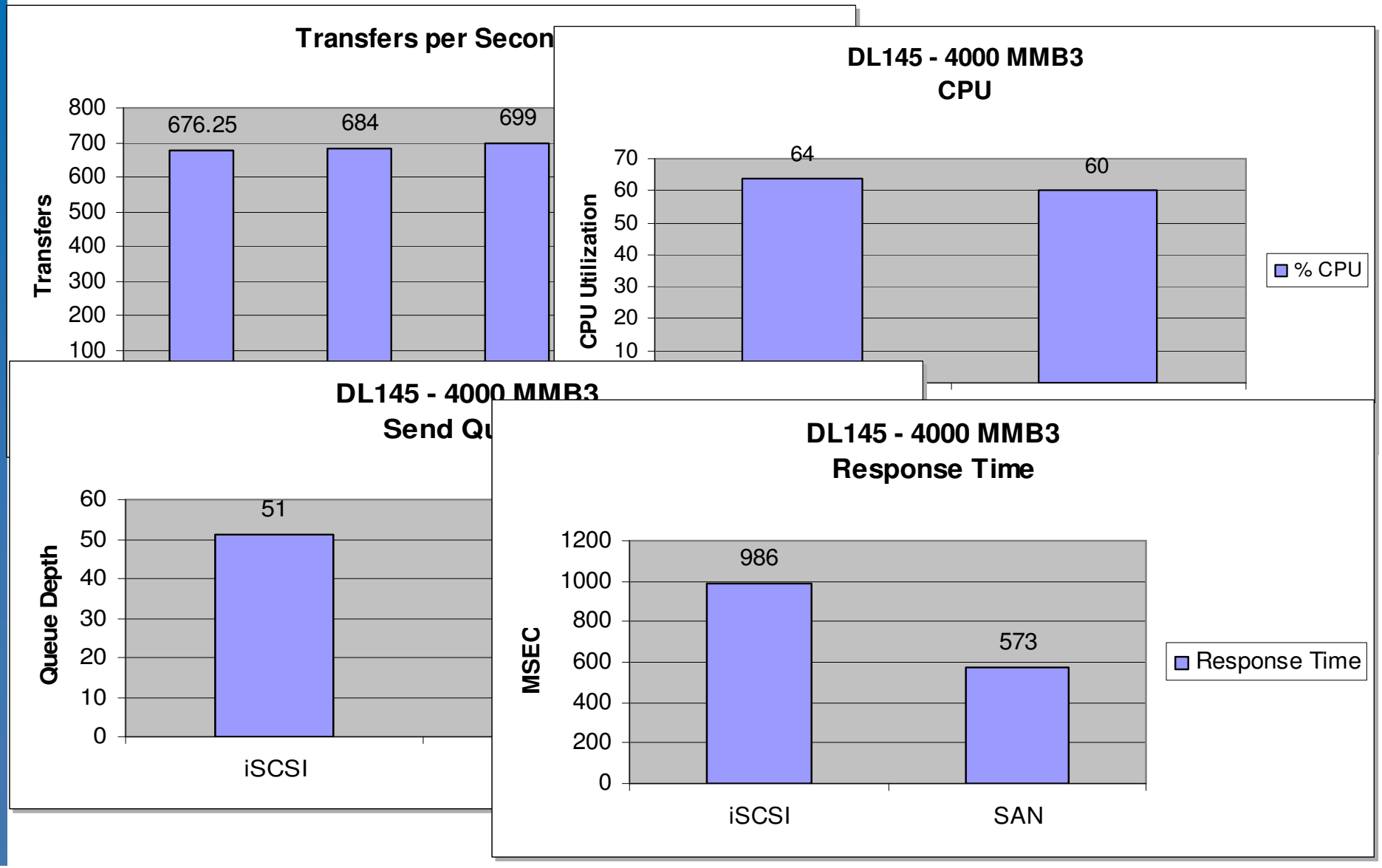
Capacity

- 400 MB mailboxes = "heavier" IO profile
  - 4,000 users * 400 MB * 1.5 = 2,344 GB
  - 2,344 GB ÷ 66 GB/drive = 36 drives
- Backend disks: RAID 10 calculation
  - 36 drives * 2 (RAID 10) = 72 drives @ 73GB
  - Best match 72 drives 73GB @ 10K RPM
- Backend disks: RAID 5 calculation
  - 36 drives * 1.2 (RAID 5) = 44 drives @ 73GB
  - Poor match to 88 (10K) or 70 (15K) drives

Network Storage Alternatives

- Storage Area Network (SAN)
  - Consolidation and availability
    - Multi-node, application fabric
    - VSS support – KB 822896
    - RAIS (SAN booting, recovery)
  - Virtualization benefits (I/O performance)
- iSCSI and NAS
  - KB articles 839686 and 839687
  - Block mode access (no UNC) and WHCL required; Gbit recommended
  - Latency (Sec/read, Sec/write), additional CPU hit are key factors
  - Recommend persistent targets (iSCSI initiator)

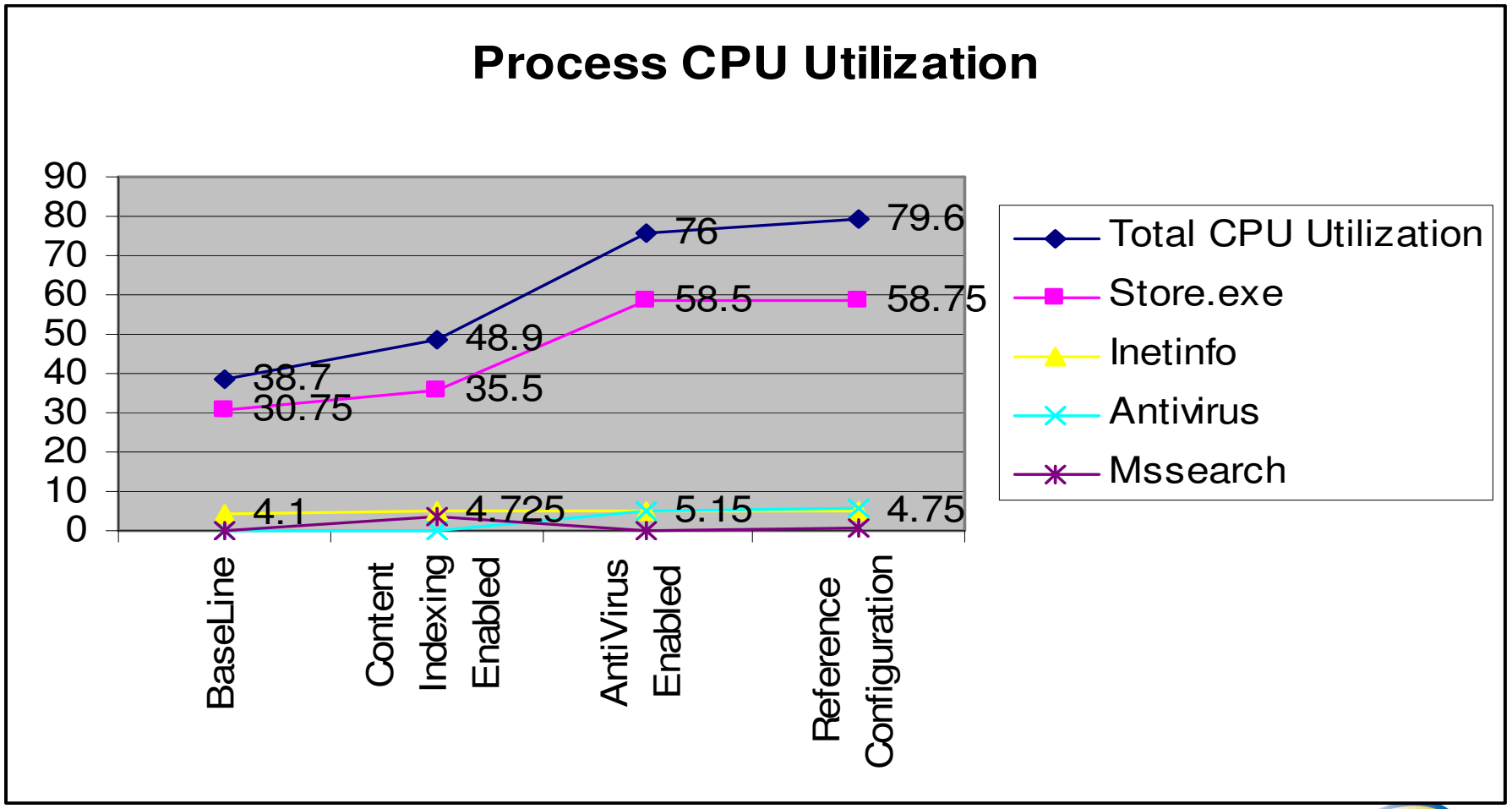
iSCSI, SAN and DAS



Sizing Exchange: Software

- Software component impacting CPU
  - Anti-Virus
  - Exchange components
    - Content indexing
    - DDLs
    - Cached mode
- Mobile device support
  - E.G. 1 Blackberry user = 2.21 MAPI users for CPU and network, but not storage

Example Of Software Impact



Sizing Active Directory

- How many clients per Global Catalog Server?
  - 1:4 GC/Exchange processor ratio
- Outlook's use of NSPI
  - Rule of thumb is 4,000 clients per GC
  - RPC over HTTP: Must used fixed port (registry key through W2K3 reg utility)
- Use /3GB on Global Catalog Server!
  - > 1GB of RAM; AS/DC of W2000
  - Decrease of 20%-40% of disk I/O
  - 512MB → 1GB of ESE Cache
- Consider upgrading to Windows Server 2003 (including 64-bit for large enterprises)

Other Important Considerations

- IS Online Maintenance
  - Checking Active Directory for deleted mailboxes
    - Minimal BE server impact
    - Scheduling important to minimize AD impact
  - Permanently remove mailboxes / messages older than retention policy – disk intensive
  - Online defrag of the data within the database – disk intensive
- Backup window – halts maintenance activities
- Performance impact related to business-driven factors
  - E.g. complete maintenance per SLA

Topics

Sizing / capacity planning process

Exchange server sizing and design

Basic monitoring

- Main Objects and Counters

Summary

Objects And Counters

- Main Exchange Objects
  - Database(s)
  - MSExchangeIS series
  - MSExchangeMTA
  - SMTP Server
  - Exchange Web Mail
  - Process
    - STORE
    - MTA
    - Internet Information Server

If You Were To Pick One Counter?

- MSExchangeIS Mailbox(_Total) | Send Queue Size
- Others
  - Process | STORE | CPU
  - Epoxy | <xxx> Que Len
  - SMTP Server | Categorizer Queue Length
    - Indicates if Active Directory is not handling the demanded workload
  - Process | Store | VM Largest Block Size (>200MB)

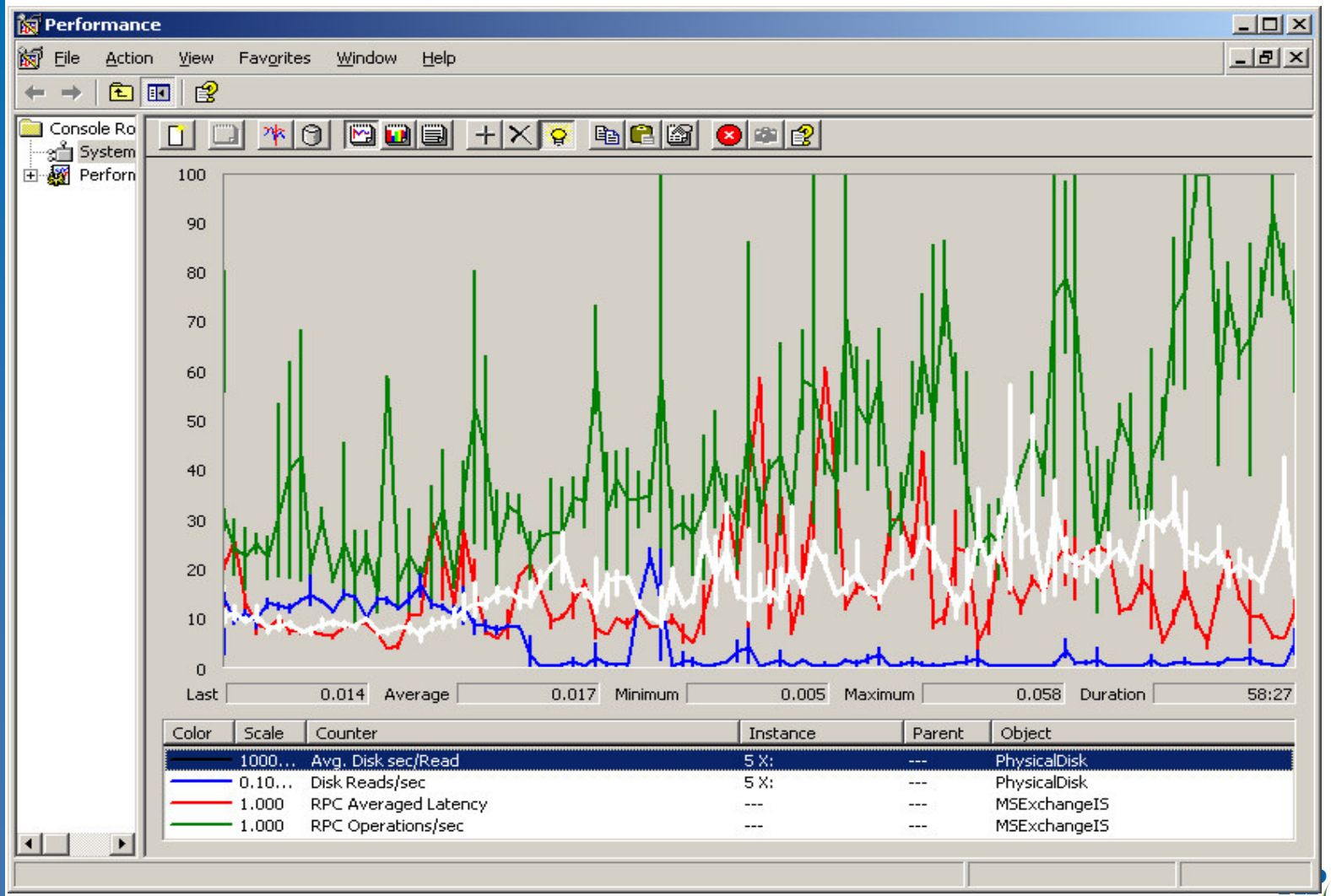
Disk Counters

- Monitor Log drives
  - Database Instances\Log Record Stalls/sec (per Storage Group)
- Monitor latency
  - PhysicalDisk(drive:)\Avg. Disk sec/Read
    - Low latency: <20ms avg
  - PhysicalDisk(drive:)\Avg. Disk sec/Write
    - Low latency: <5ms avg (caching controller)
    - <20ms avg is the goal with a few spikes that don't exceed 50-60ms.
  - Common problems – misconfigured SANS

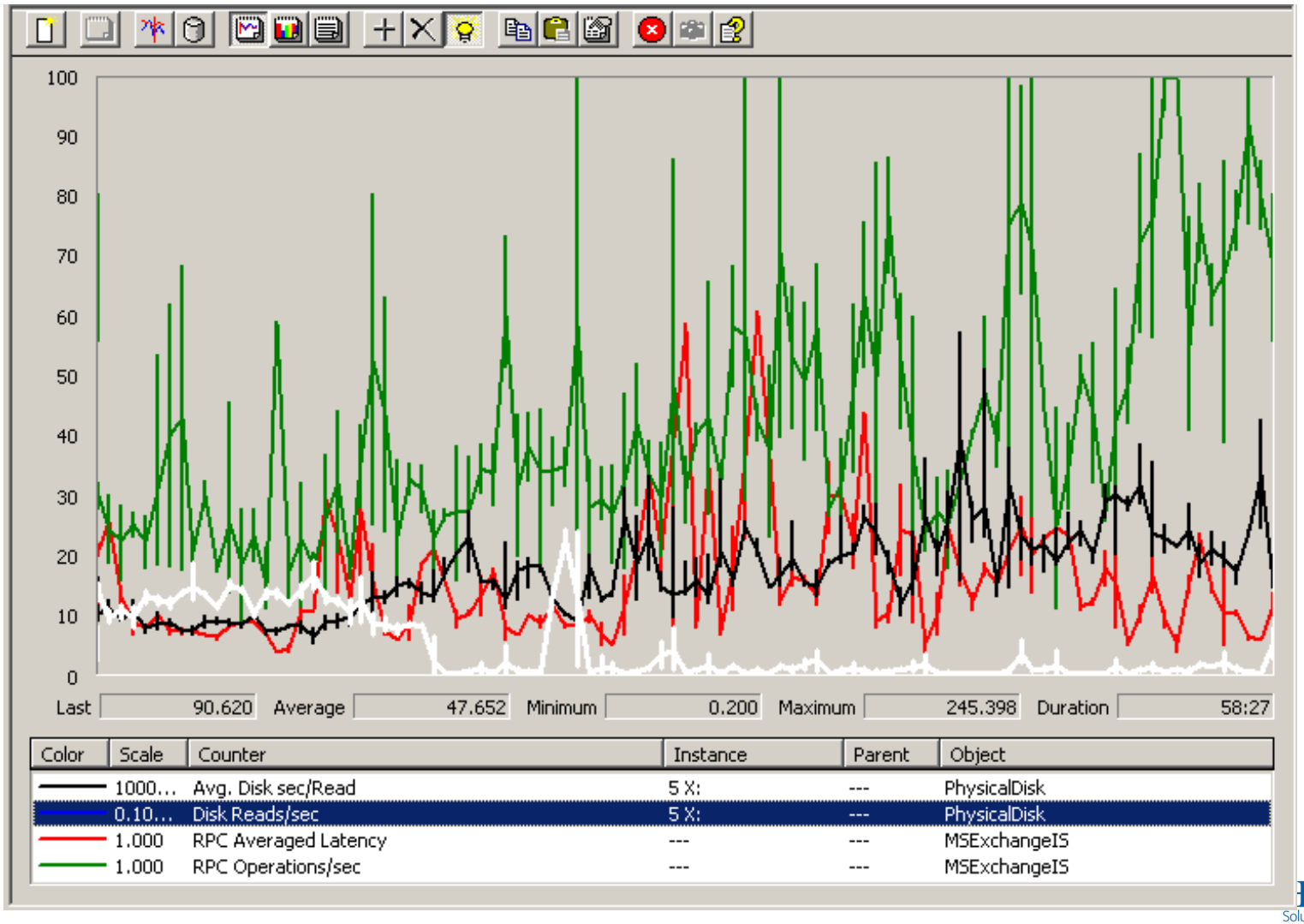
Disk Counters

- Monitor disk queues:
  - PhysicalDisk(drive:)\Current Disk Queue
    - never = 0?
  - PhysicalDisk(drive:)\Avg. Disk Queue
    - > # spindles?
- Monitor throughput
  - LogicalDisk(drive:)\Read/sec (Write/sec)
- Disk space capacity versus I/O Capacity
  - Want to be < 80% max I/O Capacity
  - Use cache controller for low latency
  - Use RAID for high transaction rates and fault tolerance

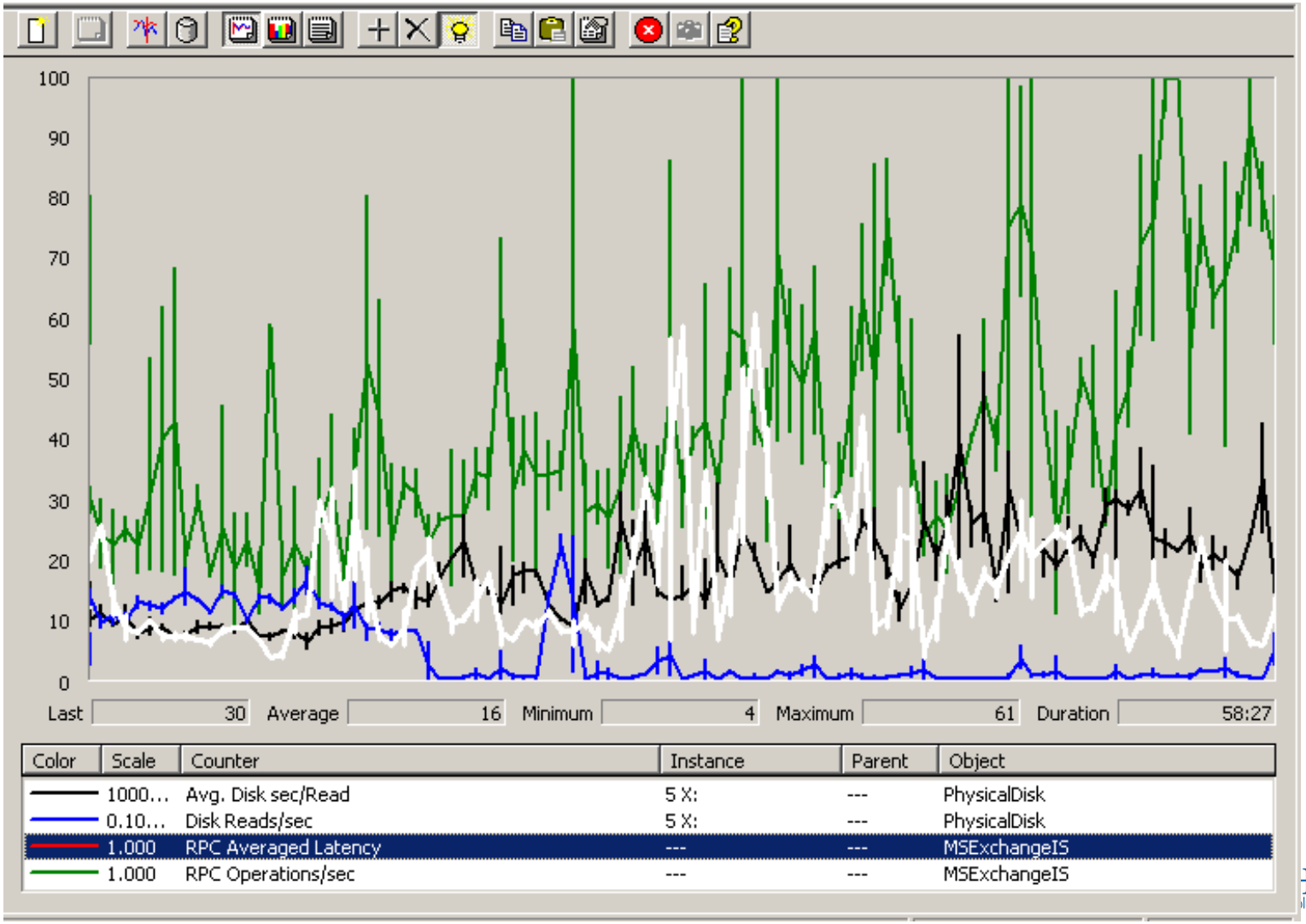
Disk At Capacity: RPC Ops/Sec Increase (Green Line)



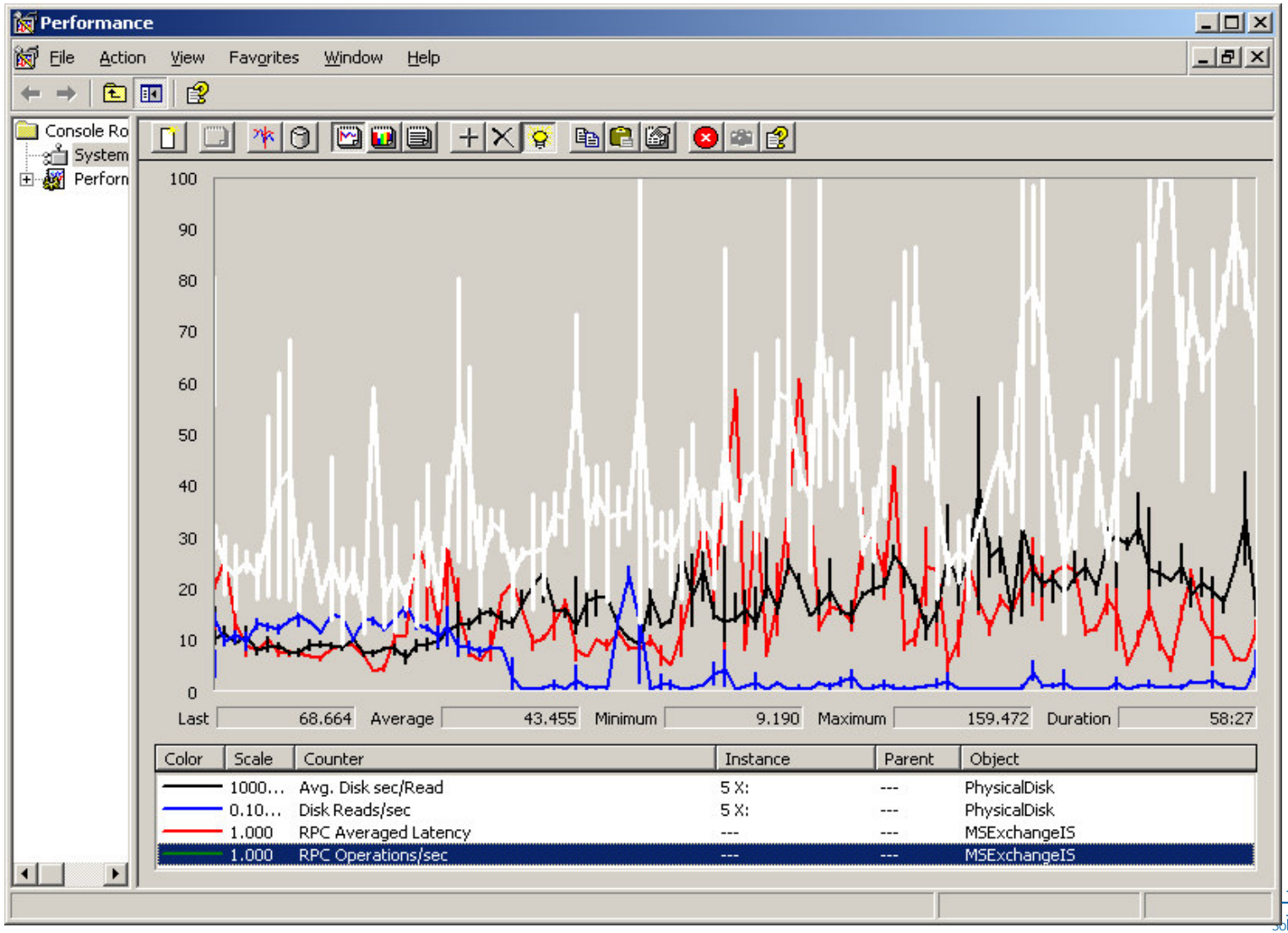
Disk At Capacity – Disk Latency Increases (Red Line)



Disk At Capacity – Disk Throughput Decreases (Disk\reads/Sec) (Blue Line)



Disk At Capacity – RPC Averaged Latency Increases (Red Line)



Topics

Sizing / capacity planning process

Exchange server sizing and design

Basic monitoring

Summary

Summary

- Benchmarking: Caveat emptor: Develop own tests / baselines / acceptance criteria
- Rules of Thumb
  - Separate roles to best consolidate users
  - Mailbox server
    - CPU: 2 – 4. Consider fastest FSB, larger L3 cache over CPU speed
    - Memory: 4GB
    - Network: 100Mbit; dual 100 Mbit for FE
    - I/O: Business factors drive architecture. Consider controller write cache, separate arrays when not virtualizing, split I/O types. Size for I/O *and* capacity
  - Active Directory
    - 1 GC near each Exchange server
    - 1 GC per 4,000 users

HP WORLD 2004

Solutions and Technology Conference & Expo

Co-produced by:

interex

shared knowledge • shared power

encompass

AN HP USER GROUP

RECOMMENDED TRAINING VENUE FOR THE

HP Certified Professional