Understanding Storage to Provide a Solution

Randy Kreiser

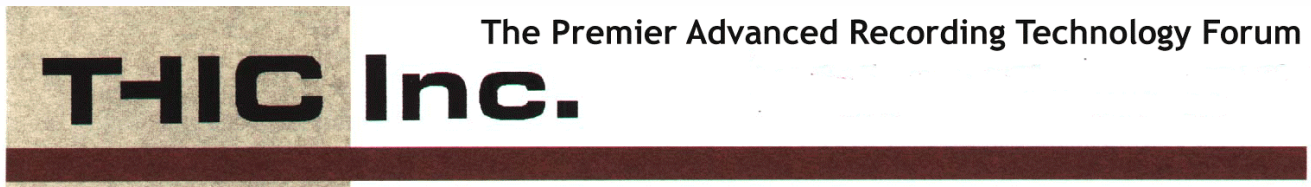
Engenio Information Technologies

12007 Sun Valley Dr, Suite 325, Reston VA 20191-3487

Phone: +1-703-262-5431, e-mail: randy.kreiser@engenio.com

Presented at the THIC Meeting at the Raytheon ITS Auditorium
1616 McCormick Drive, Upper Marlboro MD 20774-5301

October 26-27, 2004



Understanding Storage to provide a Solution

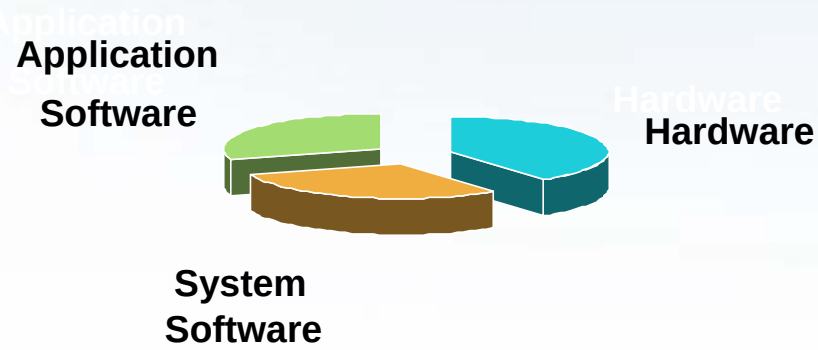
Randy Kreiser

Engenio Information Technologies

Formerly LSI Logic Storage Systems

Understanding Storage to provide a Solution

- 40/30/30 Customer Value Rule
  - 40% Hardware Setup
  - 30% System Software Setup
  - 30% Application Software



Understanding Storage to provide a Solution

- Understand Customer storage requirements
  - Capacity / Lun Size
  - Clustered shared file system
  - NAS or SAN
  - Performance
  - Striping with volume manager
  - Cost
  - Ease of setup / GUI or CLI
  - Path failover
  - Interoperability / heterogeneous environment

Understanding Storage to provide a Solution

- Capacity / Lun Size
  - Customer needs 1TB Lun Sizes
    - Requests 7+1 luns with 146GB drives
    - Performance is necessary for a successful solution (request size = 1MB)
    - Cost is paramount to success
    - Ease of setup preferred
    - Path Failover required
    - Interoperability a plus (opportunity to introduce CXFS/XVM)
- Possible path to solution
  - Understand enough of the application to see how to keep reasonable performance levels
    - Understand Server I/O architecture
    - I/O threads (reads vs writes and percent of mix)
    - Even strip width I/O's preferred
    - Coalesce smaller than even stripe width I/O's in write cache
  - Consult with customer on changing lun size (4+1/8+1 lun) to achieve even stripe width I/O
  - Script lun setup and use cli to implement

Understanding Storage to provide a Solution

- Server architecture is extremely important as the RAID becomes more robust.

  Example:
  - 5884 (TP9500) has 4 - 2Gb WWN connections to a O300
  - Each 2Gb WWN connection on the 5884 is capable of 212.5MB/s raw bandwidth performance
  - 5884 real bandwidth is 630MB/s write & 800MB/s read
  - O300 has two PCI 64bit 66Mhz buses each capable of 512MB/s and a total of 4 slots (IO9 uses one slot)
  - O300 has maximum I/O throughput capability of 1024MB/s
    2 PCI 64/66 buses * 512MB/s = 1024MB/s
  - O300 shares base I/O with one of the PCI buses thus further reducing the maximum I/O throughput capability.
  - O300 may not have enough I/O bandwidth to drive maximum 5884 performance.

Understanding Storage to provide a Solution

- Server architecture is extremely important as the RAID becomes more robust.

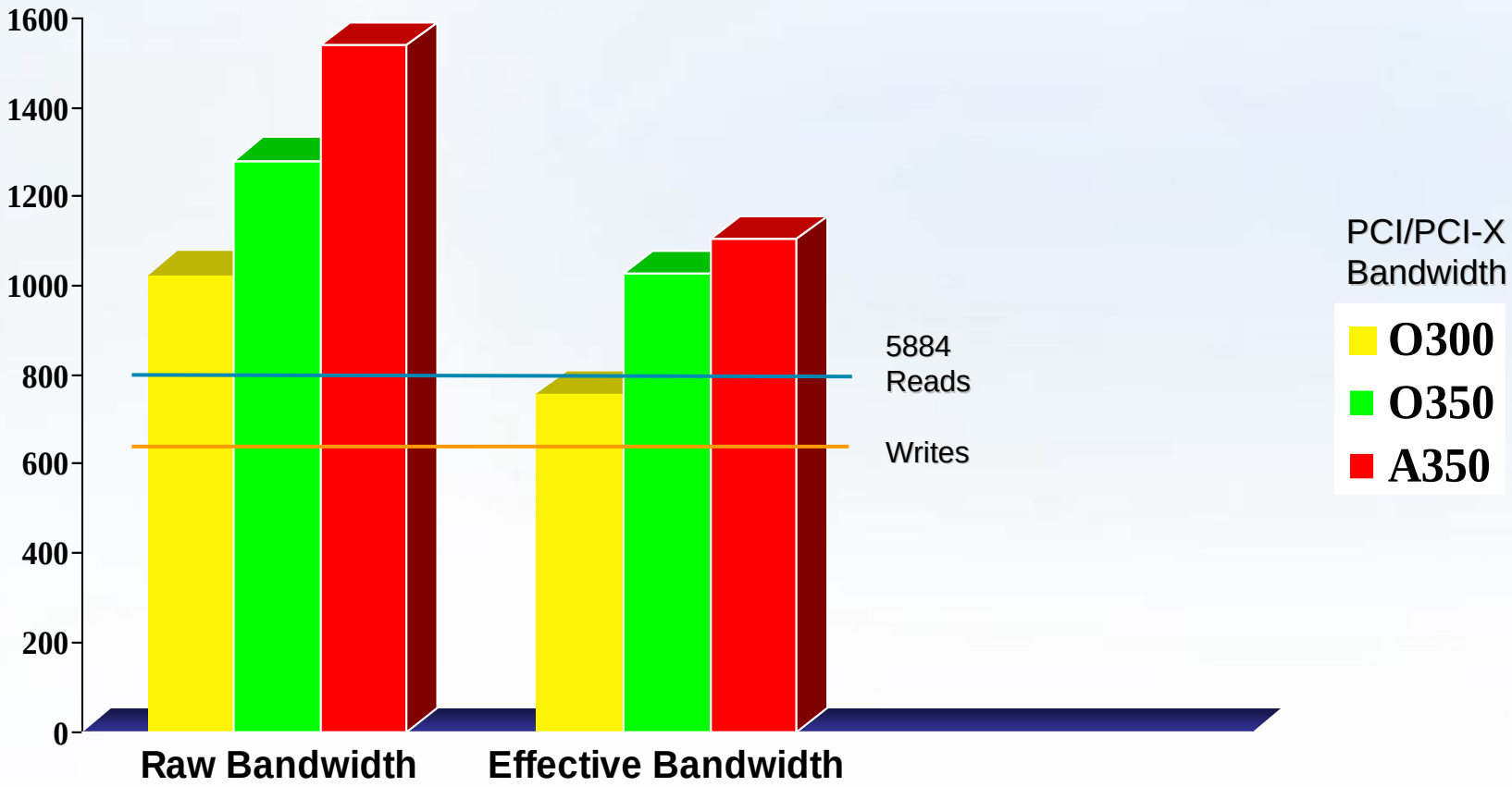
  Example:
  - 5884 (TP9500) has 4 - 2Gb WWN connections to a O350
  - Each 2Gb WWN connection on the 5884 is capable of 212.5MB/s raw bandwidth performance
  - 5884 real bandwidth is 630MB/s write & 800MB/s read
  - O350 has one PCI 64bit 100Mhz bus capable of 770MB/s and one PCI 64bit 66Mhz bus capable of 512Mb/s and a total of 4 slots (IO9 uses one 64/66 slot)
  - O350 has maximum I/O throughput capability of 1282MB/s
    1 PCI 64/100 bus @ 770MB/s + 1 PCI 64/66 bus @ 512MB/s
  - O350 shares base I/O with one of the PCI buses thus further reducing the maximum I/O throughput capability
  - O350 may have enough I/O bandwidth to drive maximum 5884 performance

Understanding Storage to provide a Solution

- Server architecture is extremely important as the RAID becomes more robust.

  Example:
  - 5884 (TP9500) has 4 - 2Gb WWN connections to a A350
  - Each 2Gb WWN connection on the 5884 is capable of 212.5MB/s raw bandwidth performance
  - 5884 real bandwidth is 630MB/s write & 800MB/s read
  - A350 has two PCI 64bit 100Mhz buses each capable of 770MB/s and a total of 4 slots (IO9 uses one slot)
  - A350 has maximum I/O throughput capability of 1540MB/s
    2 PCI 64/100 buses * 770MB/s = 1540MB/s
  - A350 shares base I/O with one of the PCI buses thus further reducing the maximum I/O throughput capability
  - A350 should have enough I/O bandwidth to drive maximum 5884 performance

Understanding Storage to provide a Solution



Understanding Storage to provide a Solution

- Application I/O threads (posix) are very important especially when scaling application I/O. Pay attention to mix of reads versus writes. Use the XVM volume manager.

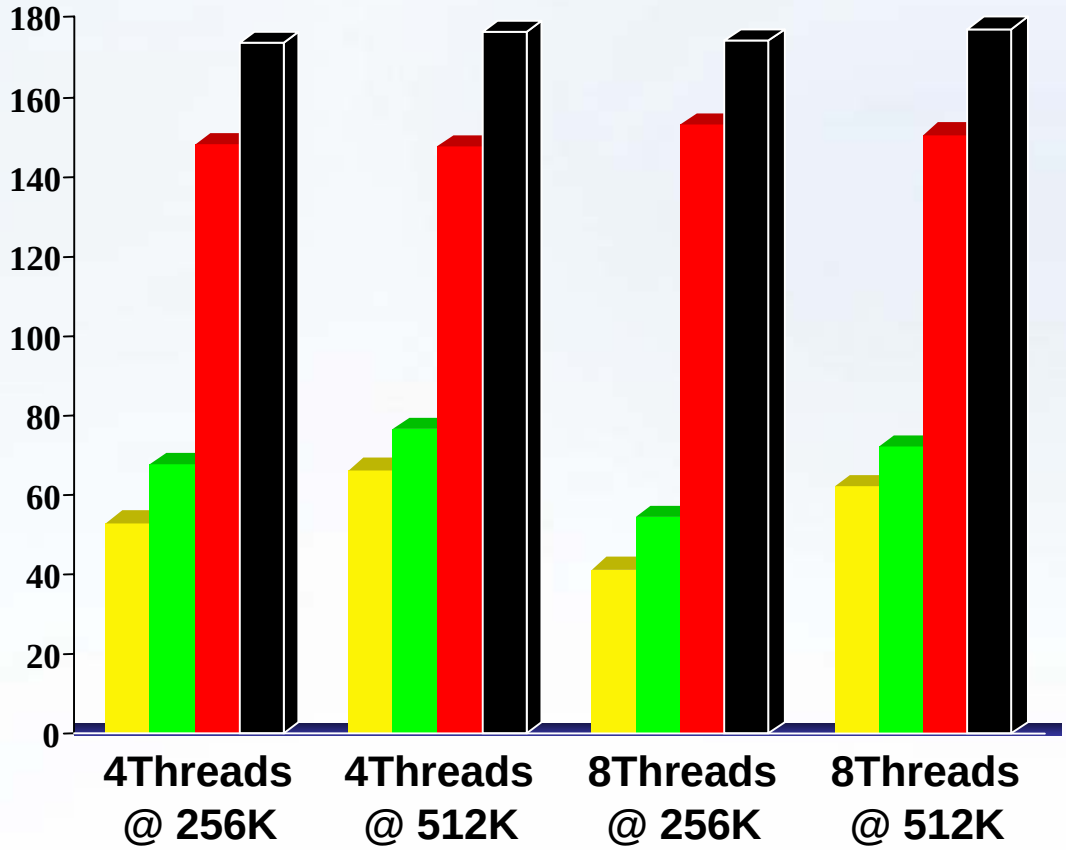
Altix 350
2882
1 RAID 5(4+1) LUN

XFS FS with & w/o XVM volume with 64K segment size; 16K cache block size; Drive read/write buffer=enabled

Read cache enabled; Read ahead multiplier=8; Write cache enabled w/o mirroring; direct sequential I/O

Test	256K	512K	Test	256K	512K
4 Threads			8 Threads		
Writes w/o XVM	52.75	66.12	Writes w/o XVM	41.30	62.11
Writes w/XVM	67.49	76.55	Writes w/XVM	54.38	71.97
Reads w/o XVM	148.04	147.54	Reads w/o XVM	152.85	150.53
Reads w/XVM	173.45	176.30	Reads w/XVM	174.02	177.09

Understanding Storage to provide a Solution



Understanding Storage to provide a Solution

- Application I/O threads (posix) are very important especially when scaling application I/O. Pay attention to mix of reads versus writes. Use the xscsiqueue command to enable CTQ.

Altix 350

2882

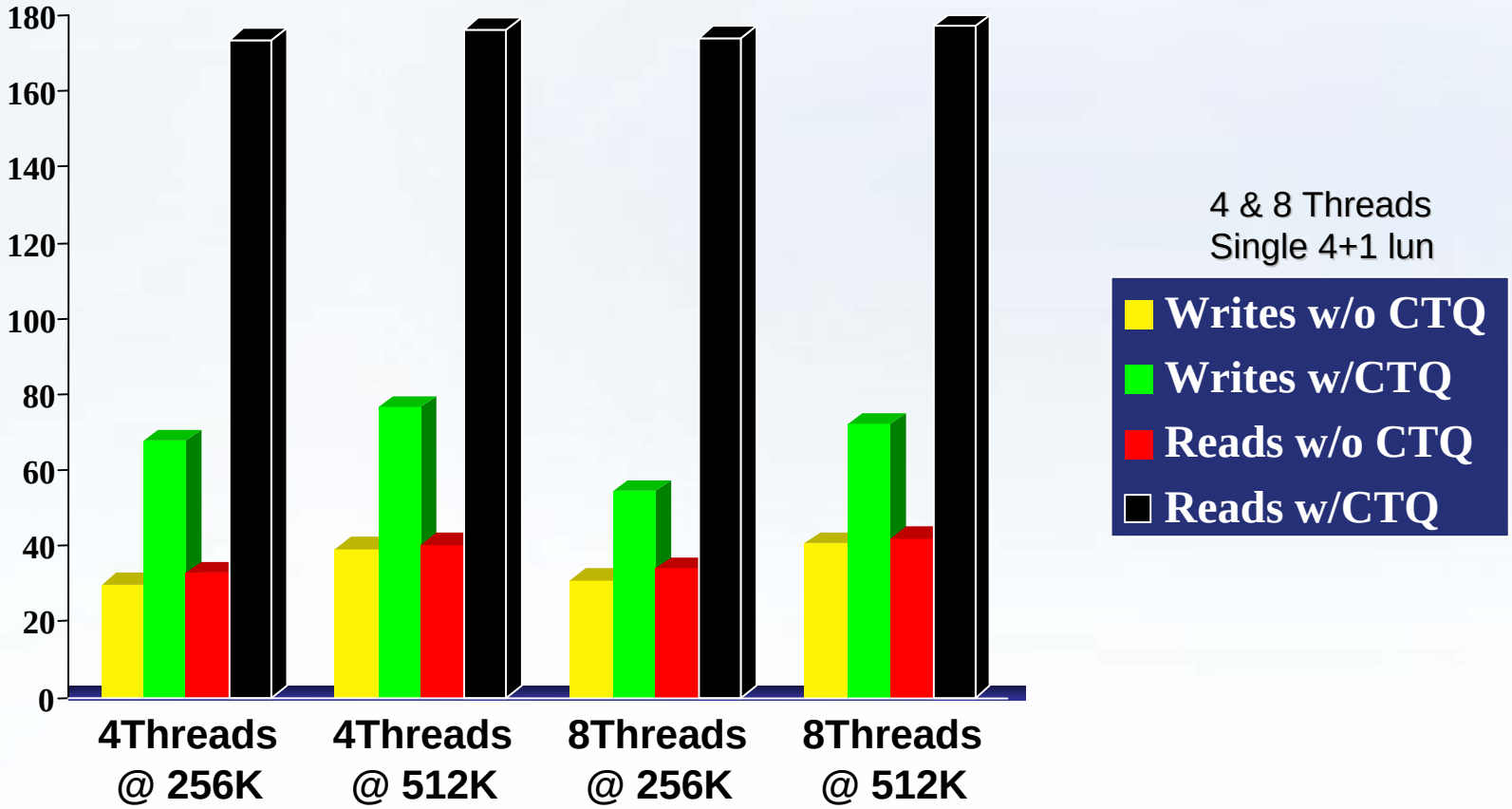
1 RAID 5(4+1) LUN

XFS FS with XVM volume with 64K segment size; 16K cache block size; Drive read/write buffer=enabled

Read cache enabled; Read ahead multiplier=8; Write cache enabled w/o mirroring; direct sequential I/O

Test	256K	512K	Test	256K	512K
4 Threads			8 Threads		
Writes w/o CTQ	29.84	39.01	Writes w/o CTQ	30.74	40.49
Writes w/CTQ	67.49	76.55	Writes w/CTQ	54.38	71.97
Reads w/o CTQ	32.81	40.19	Reads w/o CTQ	33.93	41.92
Reads w/CTQ	173.45	176.30	Reads w/CTQ	174.02	177.09

Understanding Storage to provide a Solution



Understanding Storage to provide a Solution

- Even Stripe Width I/O for an individual RAID 3/5 lun
  - ((# drives – 1) * Segment size) = stripe width of a lun
    - Example:
      4+1 lun with 64K segment size
      ((5 - 1) * 64K) = 256K stripe width
      8+1 lun with 256K segment size
      ((9 - 1) * 256K) = 2048K stripe width

- Even Stripe Width I/O for an application on striped RAID 3/5 luns
  - (((# drives – 1) * Segment Size) * # luns) = stripe width of an application
    - Example:
      2 x 4+1 luns with 64K segment size
      (((5 - 1) * 64K) * 2) = 512K application stripe width
      2 x 8+1 luns with 256K segment size
      (((9 - 1) * 256K) * 2) = 4096K application stripe width

Understanding Storage to provide a Solution

- Even Stripe Width I/O for an individual RAID 1 & 1/0 lun
  - ((# drives / 2) * Segment size) = stripe width of a lun
    - Example:
      RAID 1 1+1 lun with 64K segment size
      ((2 / 2) * 64K) = 64K stripe width
      RAID 1/0 2+2 lun with 256K segment size
      ((4 / 2) * 256K) = 512K stripe width

- Even Stripe Width I/O for an application on striped RAID 1 & 1/0 luns
  - (((# drives / 2) * Segment Size) * # luns) = stripe width of an application
    - Example:
      2 x 1+1 luns with 64K segment size
      (((2 / 2) * 64K) * 2) = 128K application stripe width
      2 x 2+2 luns with 256K segment size
      (((4 / 2) * 256K) * 2) = 1024K application stripe width

Understanding Storage to provide a Solution

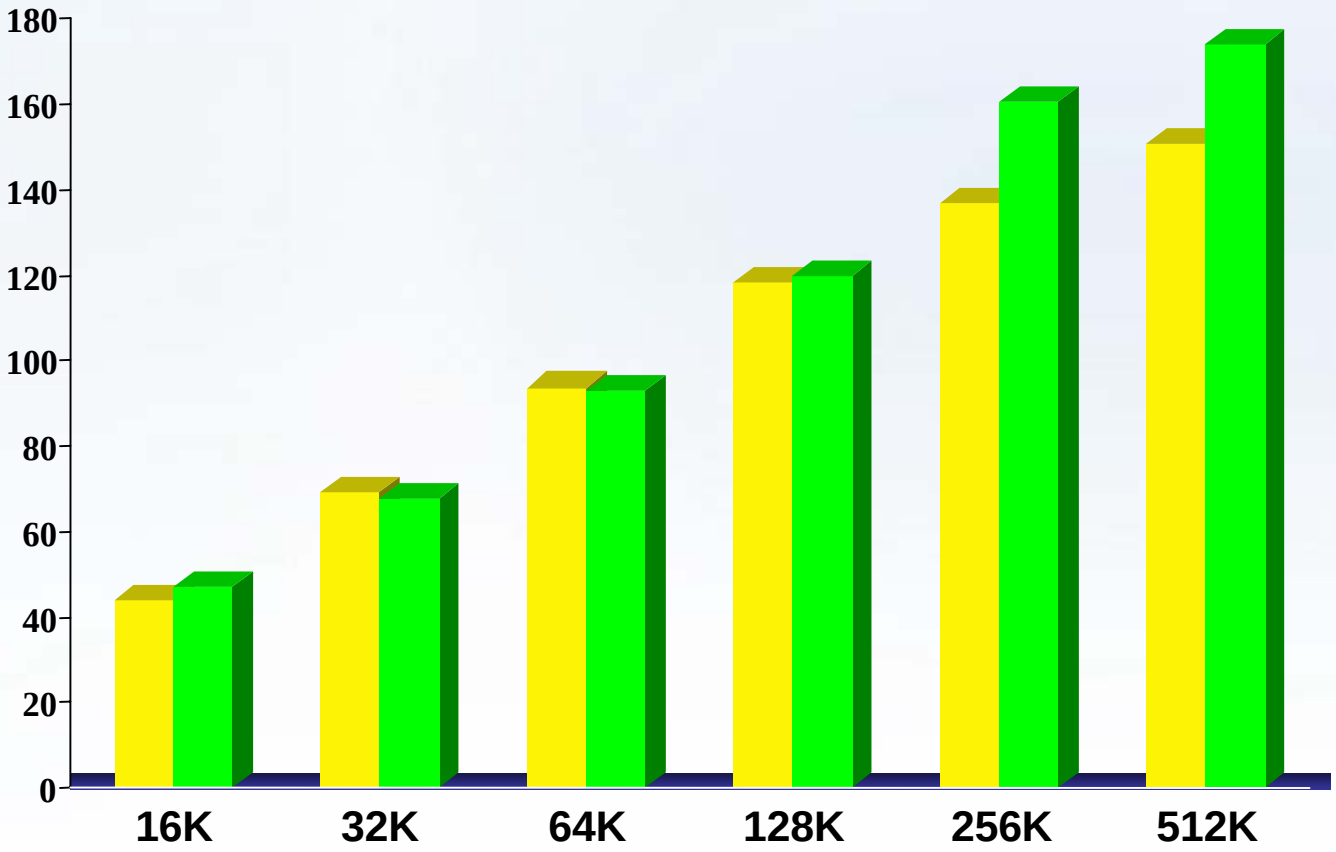
Less than even stripe width 16K to 256K I/O and even stripe width 512K I/O

Altix 350
5884
1 RAID 5(8+1) LUN

XFS File system & XVM volume with 64K segment size; 16K cache block size; Drive read/write buffer=enabled
Read cache enabled; Read ahead multiplier=8; Write cache enabled w/o mirroring; direct sequential I/O

Test	16K	32K	64K	128K	256K	512K
1 Thread						
Writes	43.91	69.19	93.68	118.08	136.73	150.52
Reads	46.97	67.41	92.83	119.60	160.50	174.01

Understanding Storage to provide a Solution



Understanding Storage to provide a Solution

Origin versus Altix Comparison with Even Stripe Width and Multiple Stripe Widths

1 RAID 5(4+1) LUN

XFS File system & XVM volume with 64K segment size; 16K cache block size; Drive read/write buffer=enabled
Read cache enabled; Read ahead multiplier=8; Write cache enabled w/o mirroring; direct sequential I/O

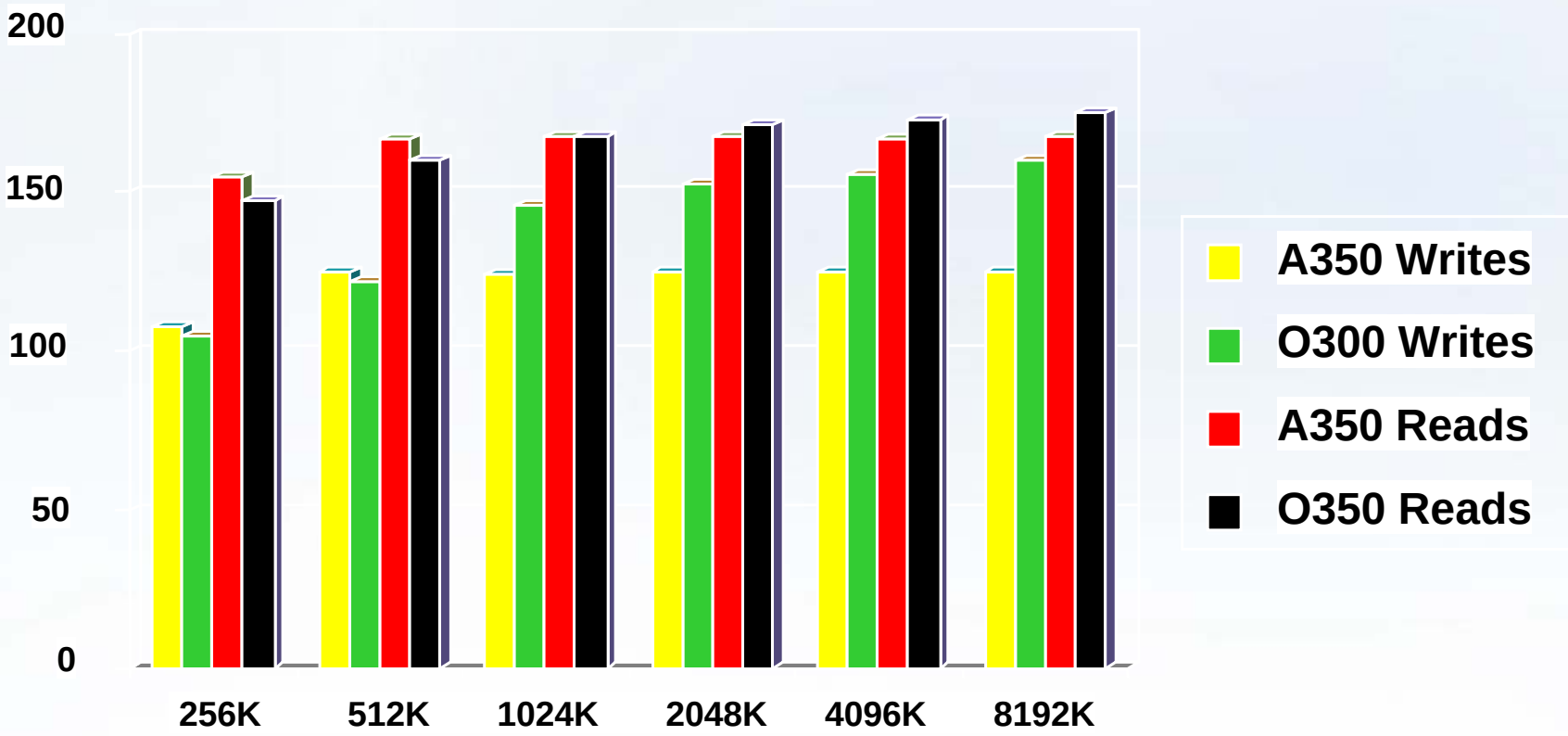
A350

Test	256K	512K	1024K	2048K	4096K	8192K
1 Thread						
Writes	107.28	124.43	124.06	124.65	124.64	124.63
Reads	154.85	167.01	167.50	167.53	166.57	167.17

O300

1 Thread						
Writes	104.20	121.62	145.79	152.21	155.65	160.04
Reads	147.37	159.64	167.26	171.04	172.43	175.02

Understanding Storage to provide a Solution



Understanding Storage to provide a Solution

CTQ Enable

xscsiqueue -e 128 /dev/xscsi/pci02.02.1/target0/lun0/rdisc

xscsiqueue -e 128 /dev/xscsi/pci02.01.1/target2/lun1/rdisc

Write Buffering Enable

xscsimode -P 0x8 -F 8 -V 0x1 -d /dev/xscsi/pci02.02.1/target0/lun0/ds

xscsimode -P 0x8 -F 8 -V 0x1 -d /dev/xscsi/pci02.01.1/target2/lun1/ds

ProPack is a must!