Tape Technology Update: State of the LTO, SDLT and DAT

Thomas Rush, Nearline Storage Technical Architect Glenn Wuenstel, Senior Solutions Systems Engineer

Hewlett-Packard

Today's Agenda

- Current top tape technologies
  - LTO
  - SDLT
  - DAT
- The performance puzzle – will it get any easier?
- Q & A

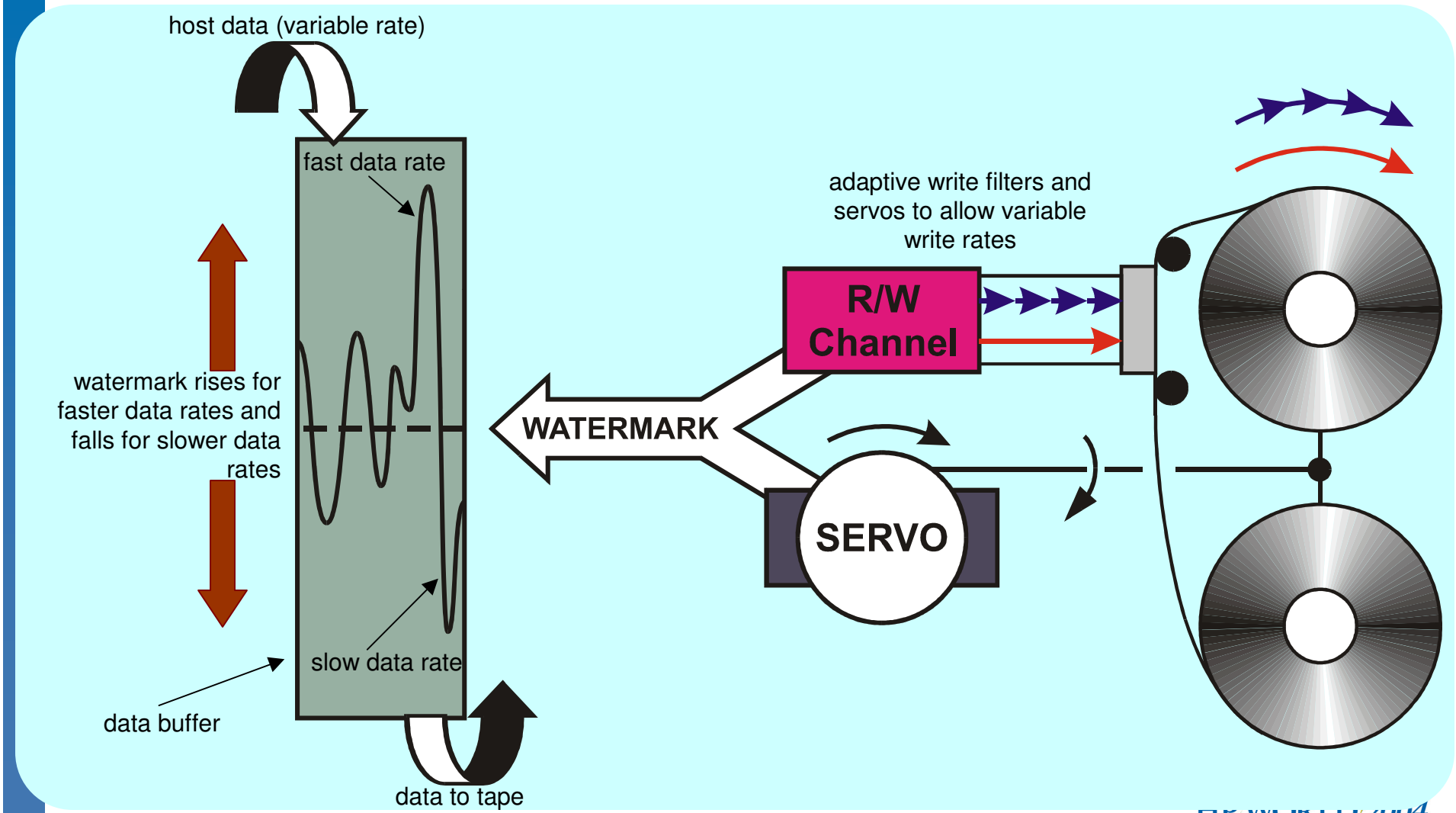
LTO Ultrium Roadmap

		Gen 1	Gen 2	Gen 3	Gen 4
what	availability	2001	2002	2004	2006
	capacity (native)	100GB	200GB	400GB	800GB
	transfer rate (native)	15MB/s	30MB/s	60MB/s	120MB/s
how	media type	MP	MP	MP	MP
	encoding scheme	RLL 1,7	PRML	PRML	PRML
	tape speed	2.7 – 5.4 m/s 106-212 ips	3.8-7.5m/s 150-295 ips	3.8-7.5m/s 150-295 ips	3.8-7.5m/s 150-295 ips
	tape length	580m	580m	800m	800m
	data tracks	384	512	768	1024
	data channels	8	8	16	16

State of LTO

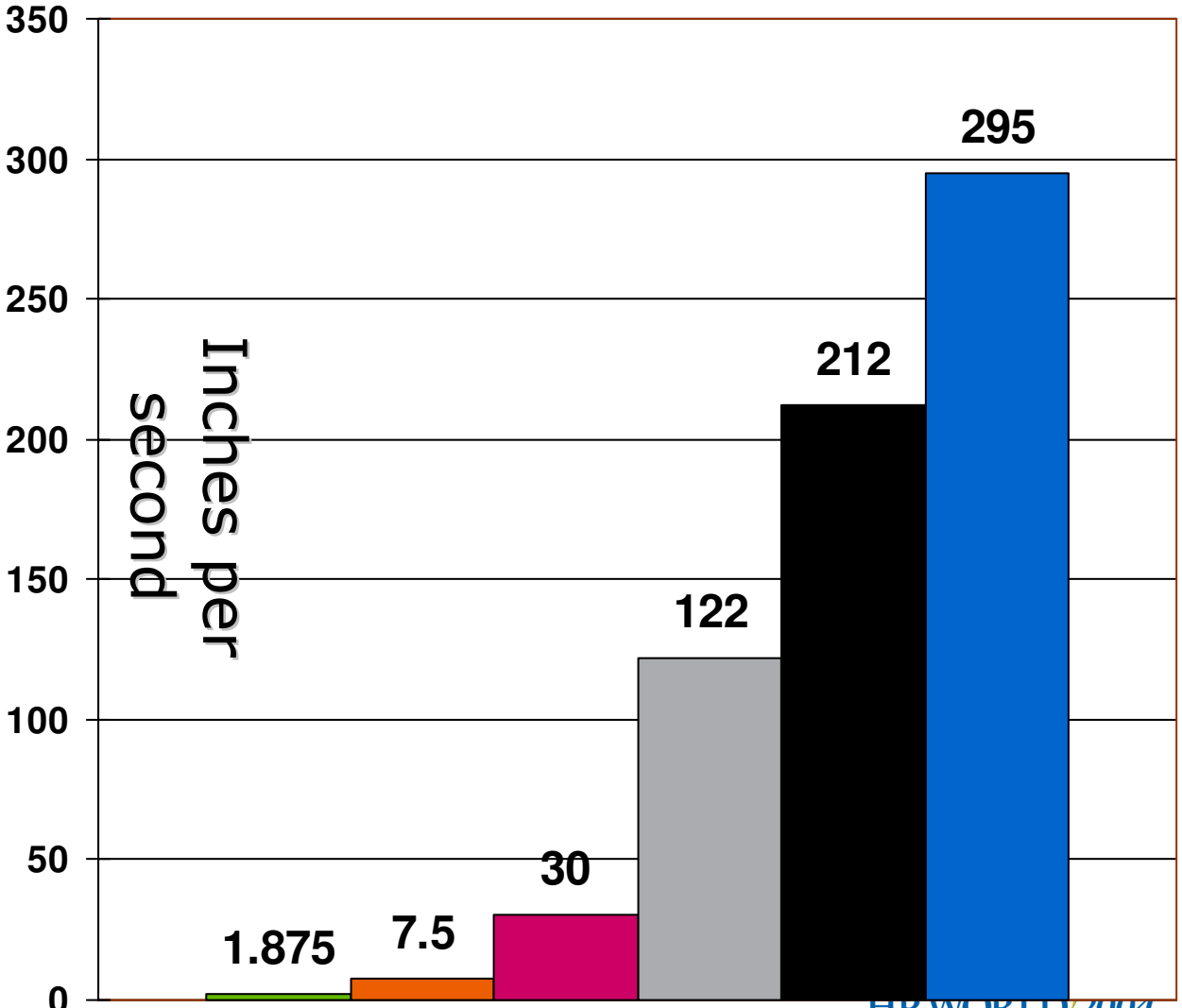
- Manufacturers poised to release LTO Gen 3 drives late this year or early next year
- Plans call for SCSI and FC versions of drives
- LTO standard backward compatibility – read back two generations, write previous generation's media
- Tape transfer rates again outperforming capabilities of hosts to feed data for streaming

Only HP Ultrium technology has Adaptive Tape Speed

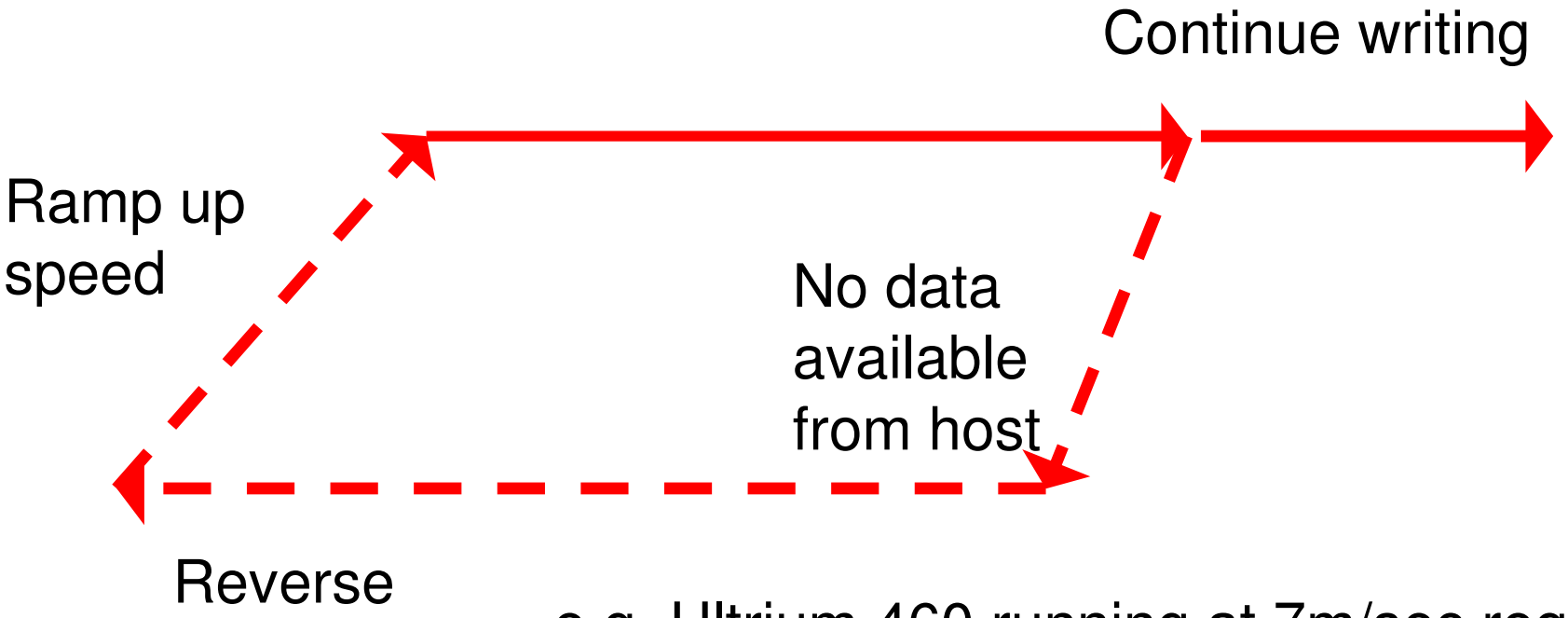


How FAST is the TAPE MOVING –

- Cassette
- Commercial Audio Tape
- High Speed Tape Duplicator
- SDLT
- Ultrium 1
- Ultrium 2

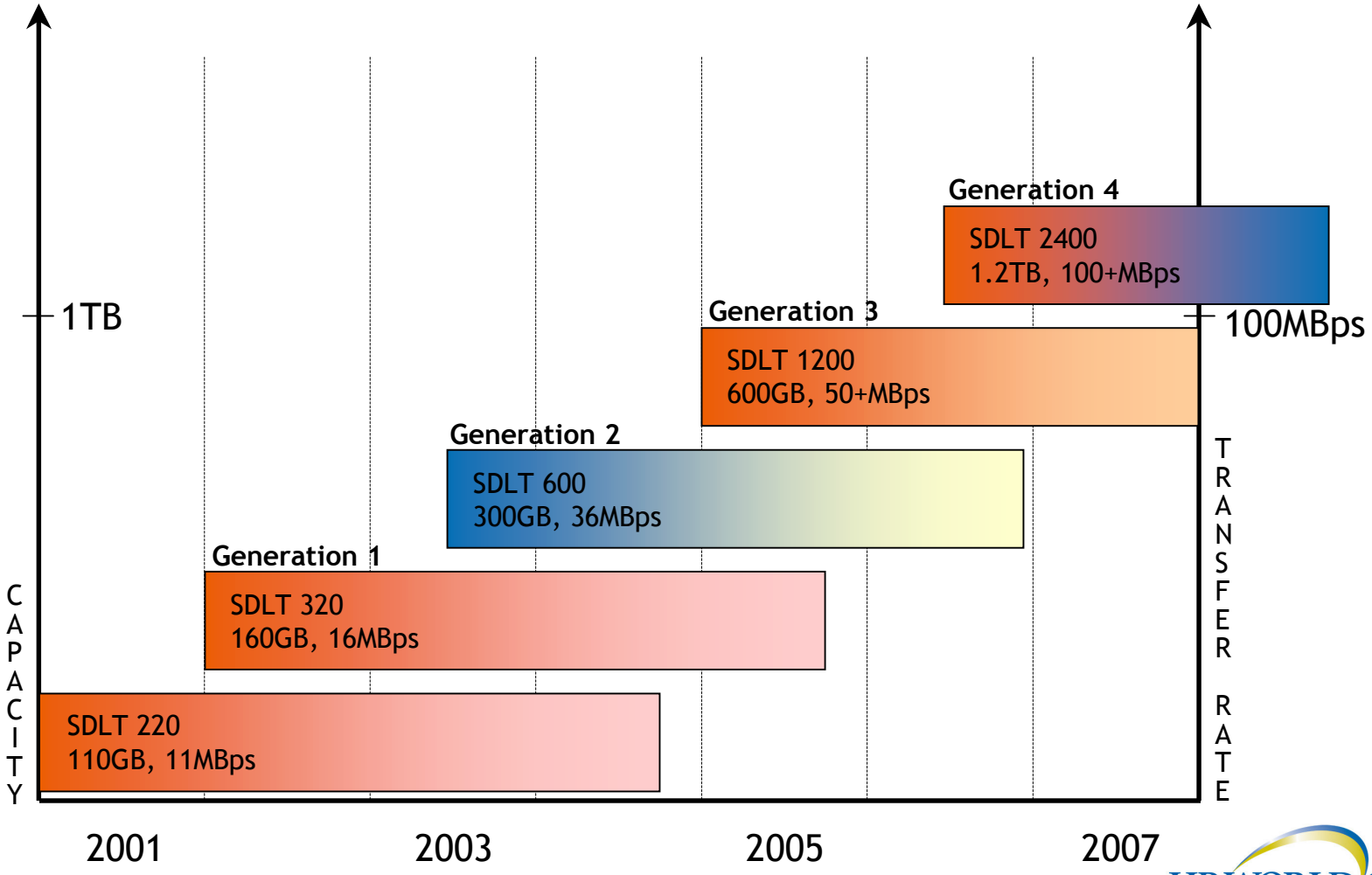


Streaming vs Repositioning



e.g. Ultrium 460 running at 7m/sec requires a full 3 seconds to reposition

Super DLT Road Map

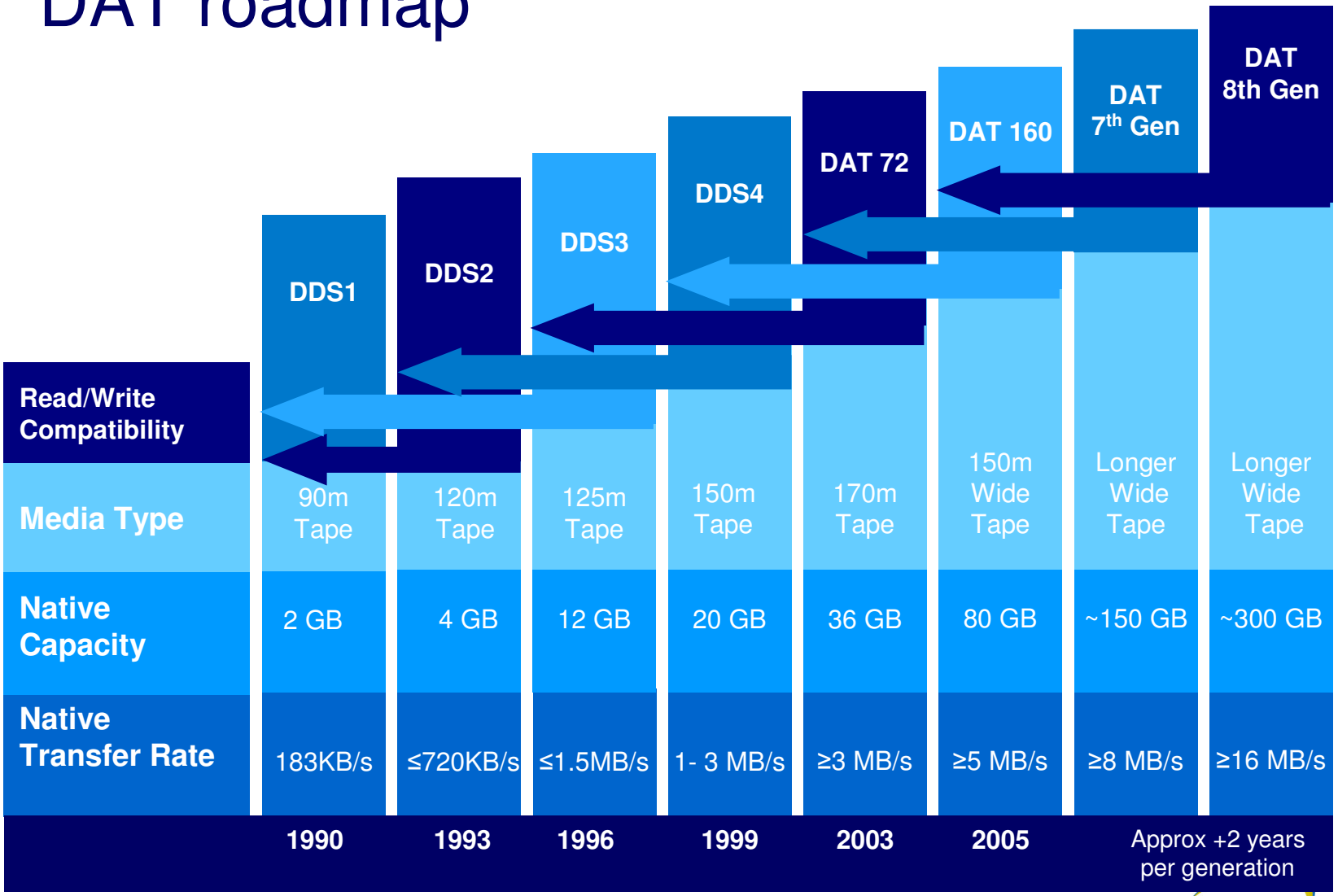


Source : Quantum

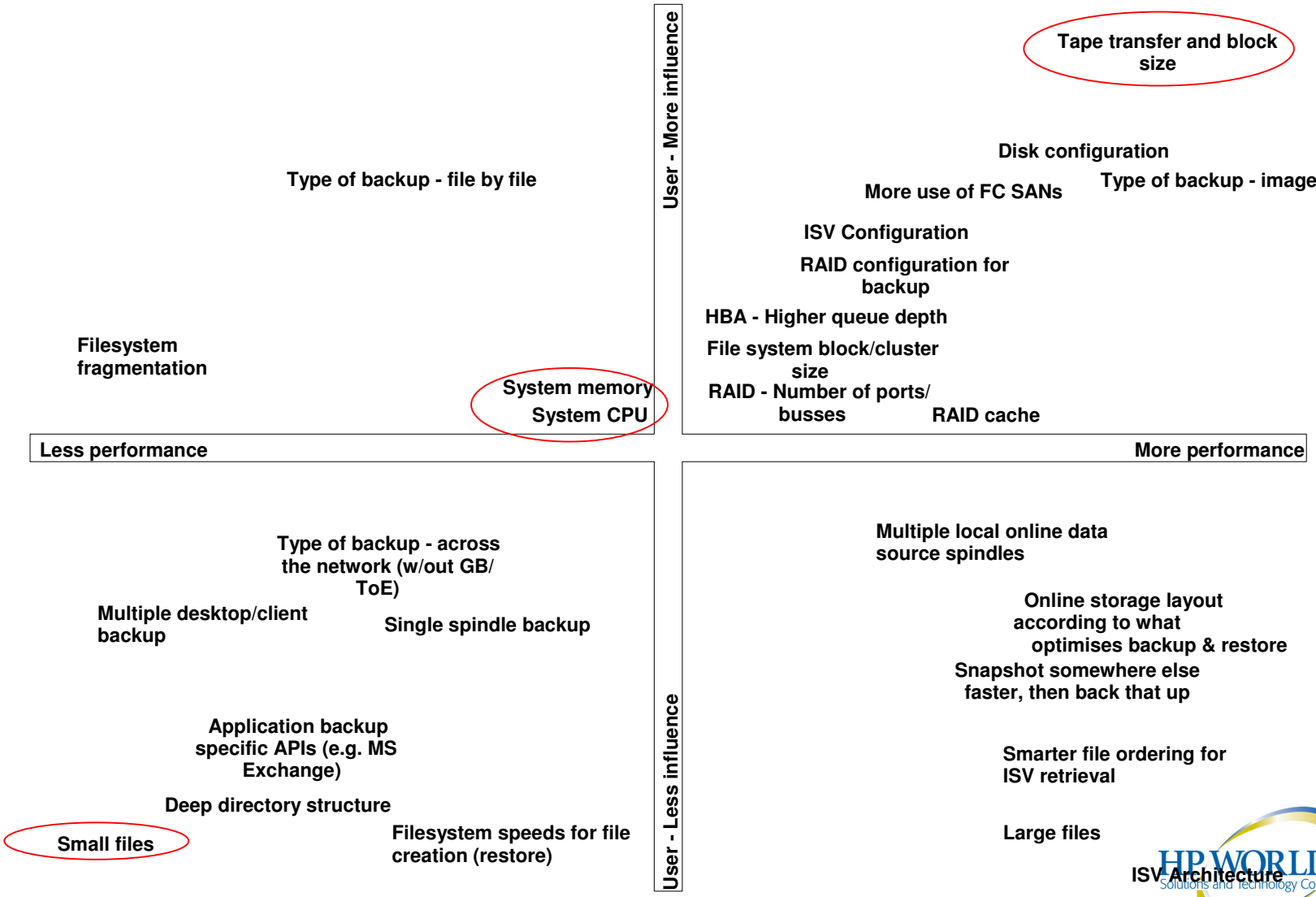
State of SDLT

- Quantum poised to release SDLT 600 drive for volume shipments this year
- Drive available with SCSI interface
- Drive can read SDLT Gen 1 media (written in SDLT 220 and SDLT 320 drives) and DLTtape VS1 media (written in DLT VS160 drives)
- Tape transfer rates of 36 MB/sec

DAT roadmap


DDS1
DDS2
DDS3
DDS4
DAT 72
DAT 160
DAT 7th Gen
DAT 8th Gen
Read/Write Compatibility
Media Type
90m Tape
120m Tape
125m Tape
150m Tape
170m Tape
150m Wide Tape
Longer Wide Tape
Longer Wide Tape
Native Capacity
2 GB
4 GB
12 GB
20 GB
36 GB
80 GB
~150 GB
~300 GB
Native Transfer Rate
183KB/s
≤720KB/s
≤1.5MB/s
1- 3 MB/s
≥3 MB/s
≥5 MB/s
≥8 MB/s
≥16 MB/s
1990
1993
1996
1999
2003
2005
Approx +2 years per generation

Tape backup performance influence map - Summary

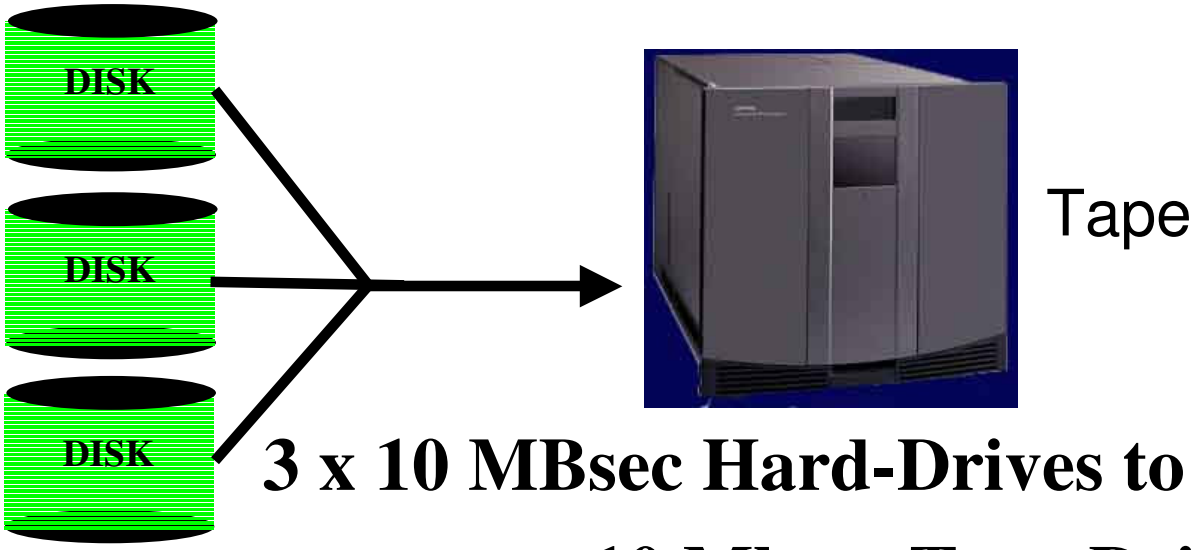


Data Source considerations

- As a rule of thumb you need 2x – 3x tape speed to maintain streaming.
- Conversely your backup speed is around 1/3 of your data source speed.
- Restore speed is generally around 40-60% of Backup speed.

Primary storage performance – the 3:1 rule

- SOURCE TO TARGET SPEED - You usually need 2 to 3 times the source speed as it compares to your desired backup speed.
  - Tape is a typically a streaming device.
    - Buffer Under-runs , Shoe Shining, Back Hitching and other signs of source performance issues.



3 x 10 MBsec Hard-Drives to feed a single 10 Mbsec Tape Drive

Data Source considerations – RAID structure

Structure Considerations

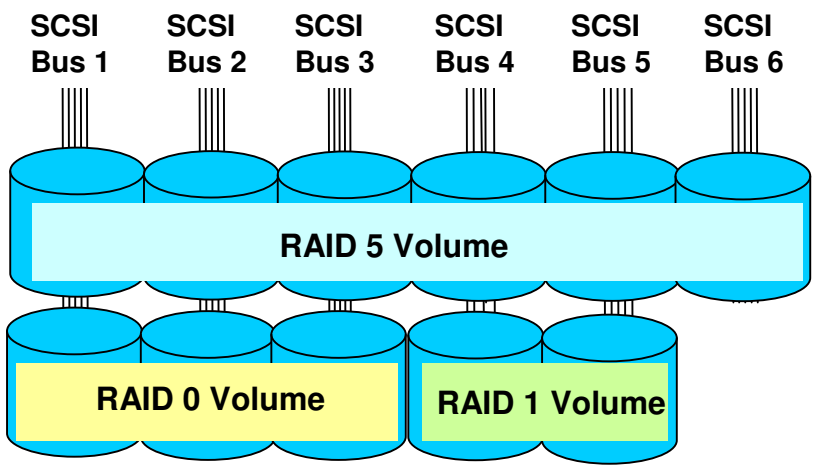
- Striping across multiple arrays controllers improves performance

- Multiple Volumes permit multiple jobs (multiple entry points for backup applications) - Parallelism

Data Source Considerations - RAID

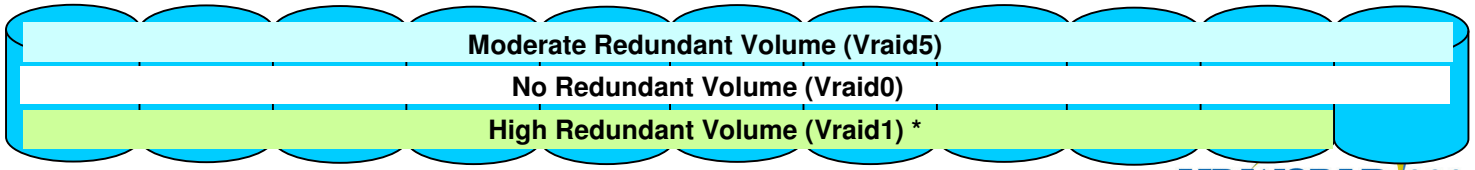
- eliminate throughput bottlenecks
- eliminate load balancing procedures for applications and databases

Yesterdays RAID Sets



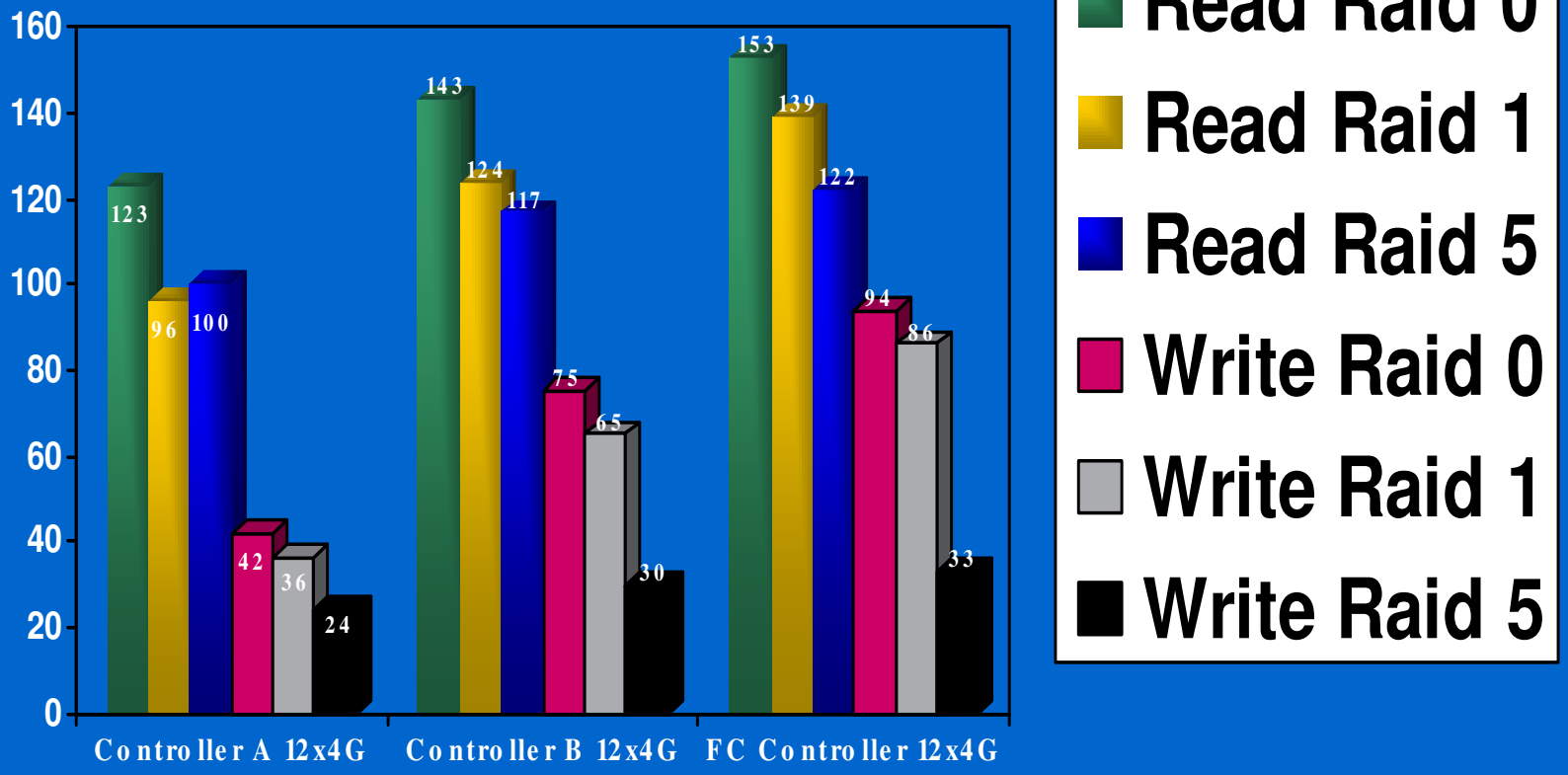
HP's Virtual Disk Technology

Workload is evenly distributed across the spindles in group



*RAID V1 uses even numbers of disks

Data Source Considerations- RAID



Disk Array Performance




	XP1024	EVA5000	EVA3000	MSA1000
# Disks in base unit	1024	240	56	14
Max Throughput MB/sec	2000	628	335	200
# FC ports on base unit	64	16	4	2

Source Data : Best Case vs. Worse Case Data

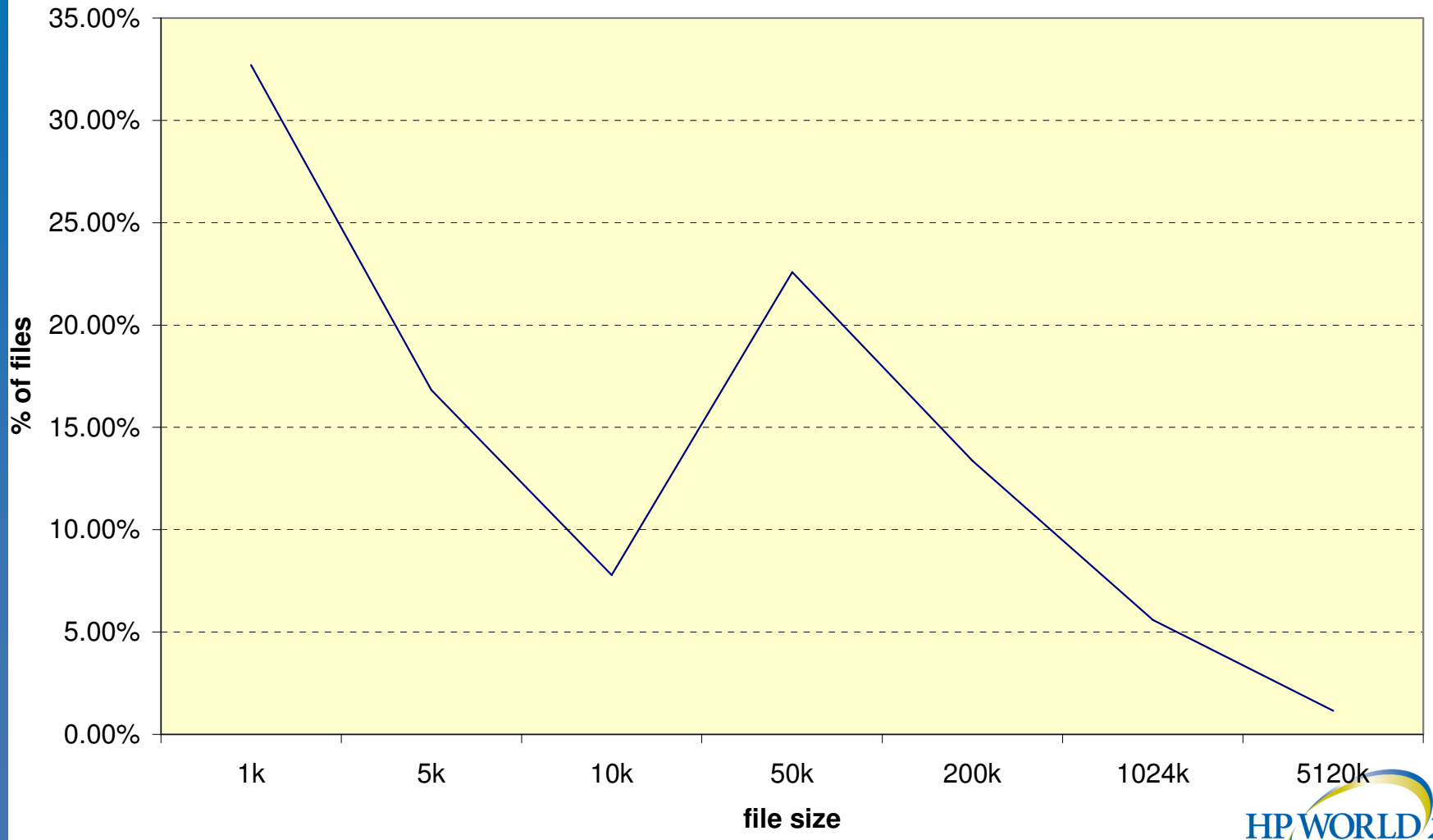
Best Case –

- Low File Count
- Large Files
- Simple File Tree Structure
- Short file names
- Compressible Data (2:1)

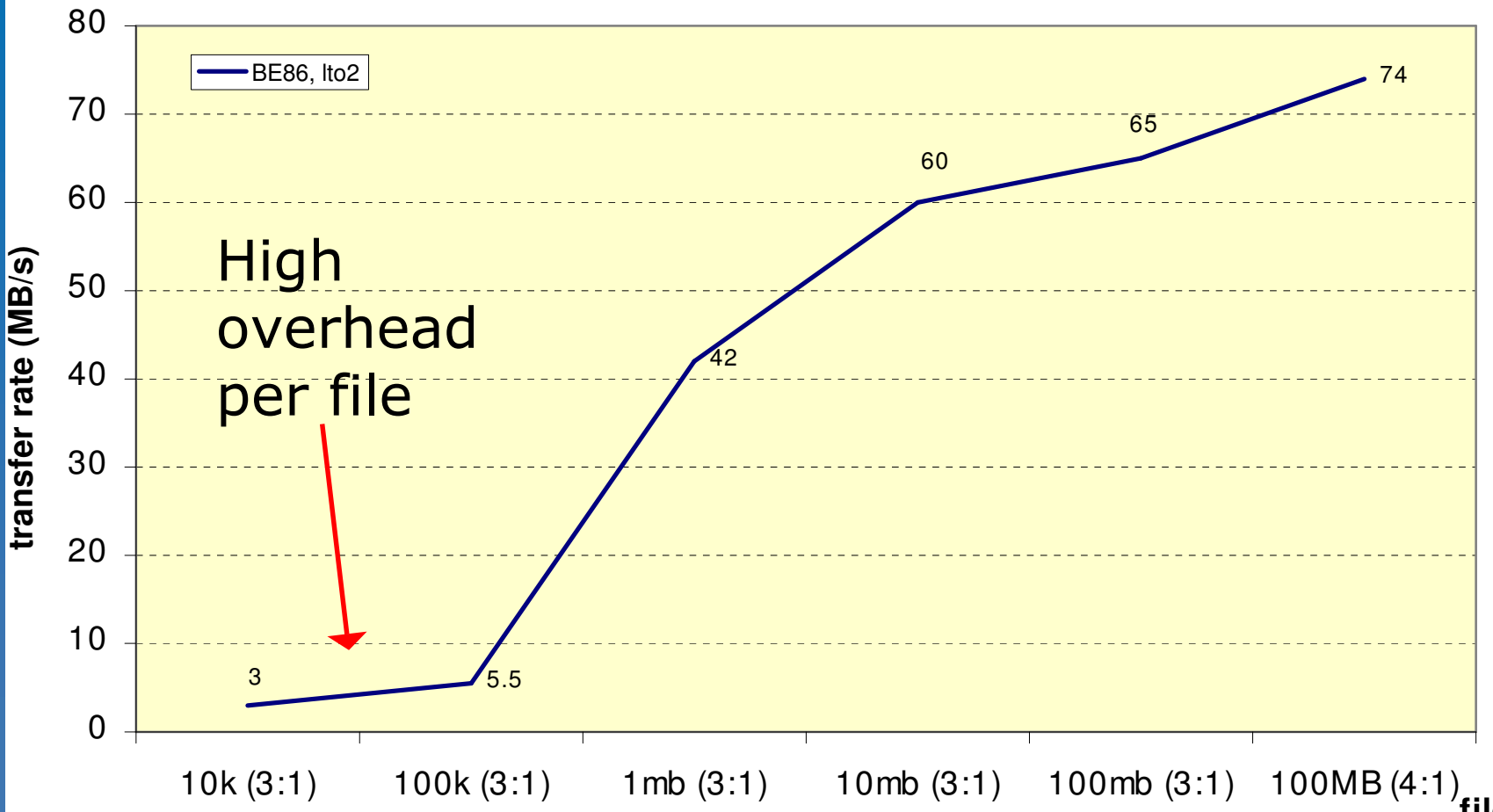
Worse Case –

- High File Count
- Tiny Files (1k byte)
- Complex directory structure
- Long File Names
- Non-Compressible (1:1)

Example file size distribution



File size performance example (NT)



Sizing the Backup Server

- Hardware Considerations (Microsoft W2K or NT)
  - Number of Processors
  - Memory - 256 Megs minimum (app drivers?)
  - Boot Drives (Mirrored RAID-1)
  - Faster C: Drive can help OS and TAPE Software
  - PCI Bus(s) – 32bit, 64bit, 33-66-133Mhz, PCI-X, PCI-X 2.0
  - LAN / SAN connections

- Software Considerations (Microsoft NT/W2K)
  - Don't put the Backup app. database on a slow drive
  - Use the largest block sizes for tape drivers (64K or greater)

PCI shouldn't be the bottleneck

Clock Frequency	Bus Width	Burst Perf	Sustainable Perf
33MHz	32 bit	133 MB/s	115MB/s
33MHz	64 bit	266 MB/s	230MB/s
66MHz	32 bit	266MB/s	230MB/s
66MHz	64 bit	533MB/s	490Mb/s
133MHz (PCI-X)	64 bit	1066MB/s	980MB/s
266 MHzPCI-X 2.0	64 bit	2132MB/s	1960MB/s
533 MHzPCI-X 2.0	64 bit	4264 MB/s	3940MB/s

Sizing a backup server - Summary

Parameter	Rule of Thumb	Comment
Processors	Today a typical 2 proc (> 1GHz) machine can handle up to 5 single data streams @ 20Mb/sec at 75% processor load. From a single datapoint of a single stream = X% processor load, then a concurrent stream = 2X% (irrespective of concurrency value)	Critical at the planning stage to understand likely growth and select a "scaleable" server. Highly scaleable servers also have better memory access speeds and more PCI bus "peer" capabilities.
Memory	1GB should be adequate unless large Nos of parallel streams.	Some Backup apps scale better in performance than others if more memory available.
PCI Architecture	Use servers with "peer" PCI buses not bridged PCI buses. Keep cards in the appropriate slots!	Unlikely that PCI will be the bottleneck in most cases.

The SAN Switch

- **Functions of the SAN Switch**
  - **Interconnection of all SAN components.**
  - **Performance Monitoring device used to analyze backup performance problems.**
  - **Permits “speed matching”, the interconnection of newer 2 Gig devices to older 1 Gig devices without forcing the 2 Gig device to run at a 1 Gig rate all the time.**
  - **Zoning controller/manager – zoning is much like LUN masking, in that it limits visibility of SAN devices to each other.**

The SAN Switch (continued)

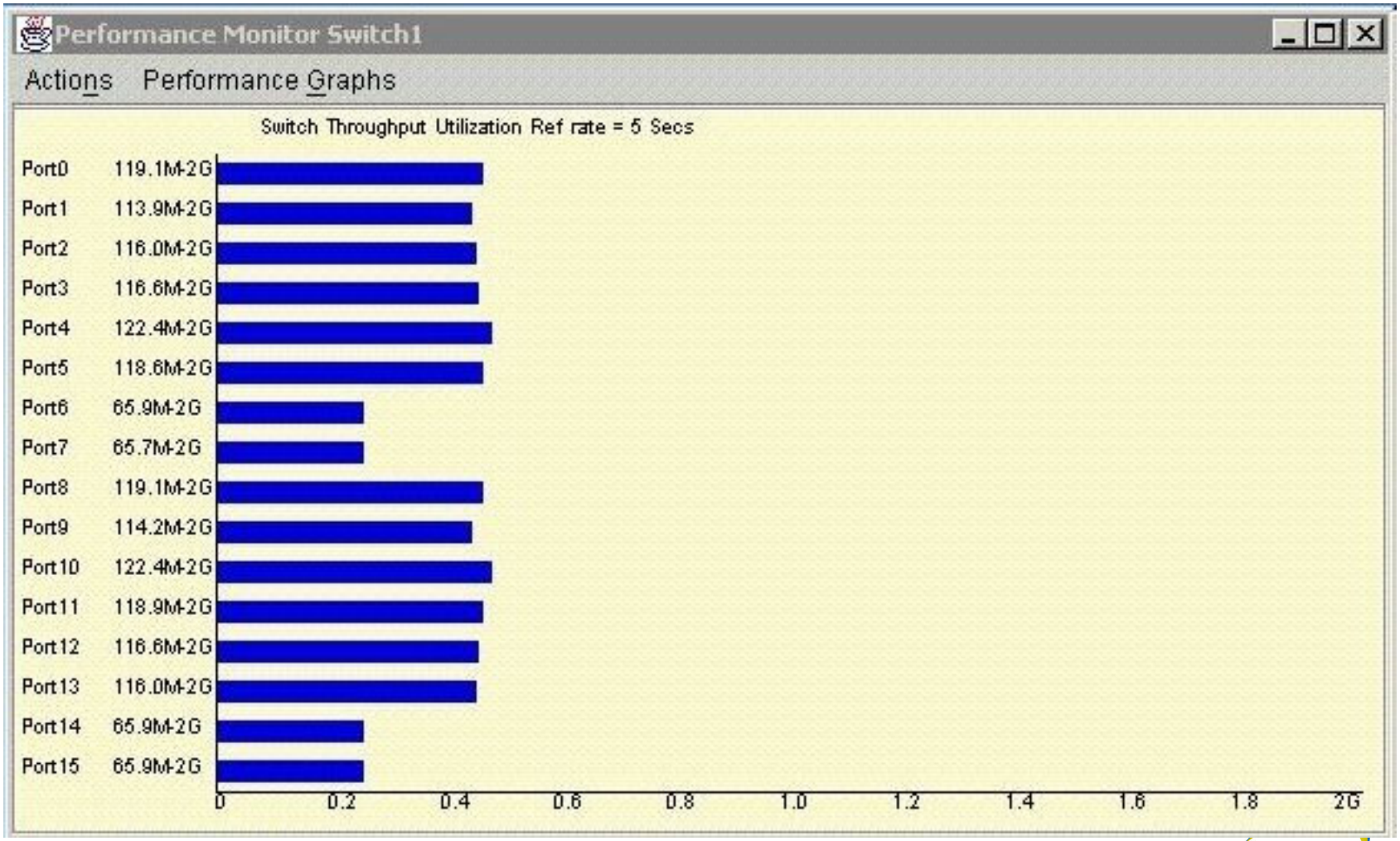
- **SPEED of a Fibre Channel port:**
  - **1.063 Gbit/Sec full duplex (100+ MBytes/Sec)**
  - **2.125 Gbit/Sec full duplex (200+ Mbytes/Sec)**

  Full duplex – an FC adapter can do 2X the above speeds if the adapter card is both transmitting AND receiving at the same time.

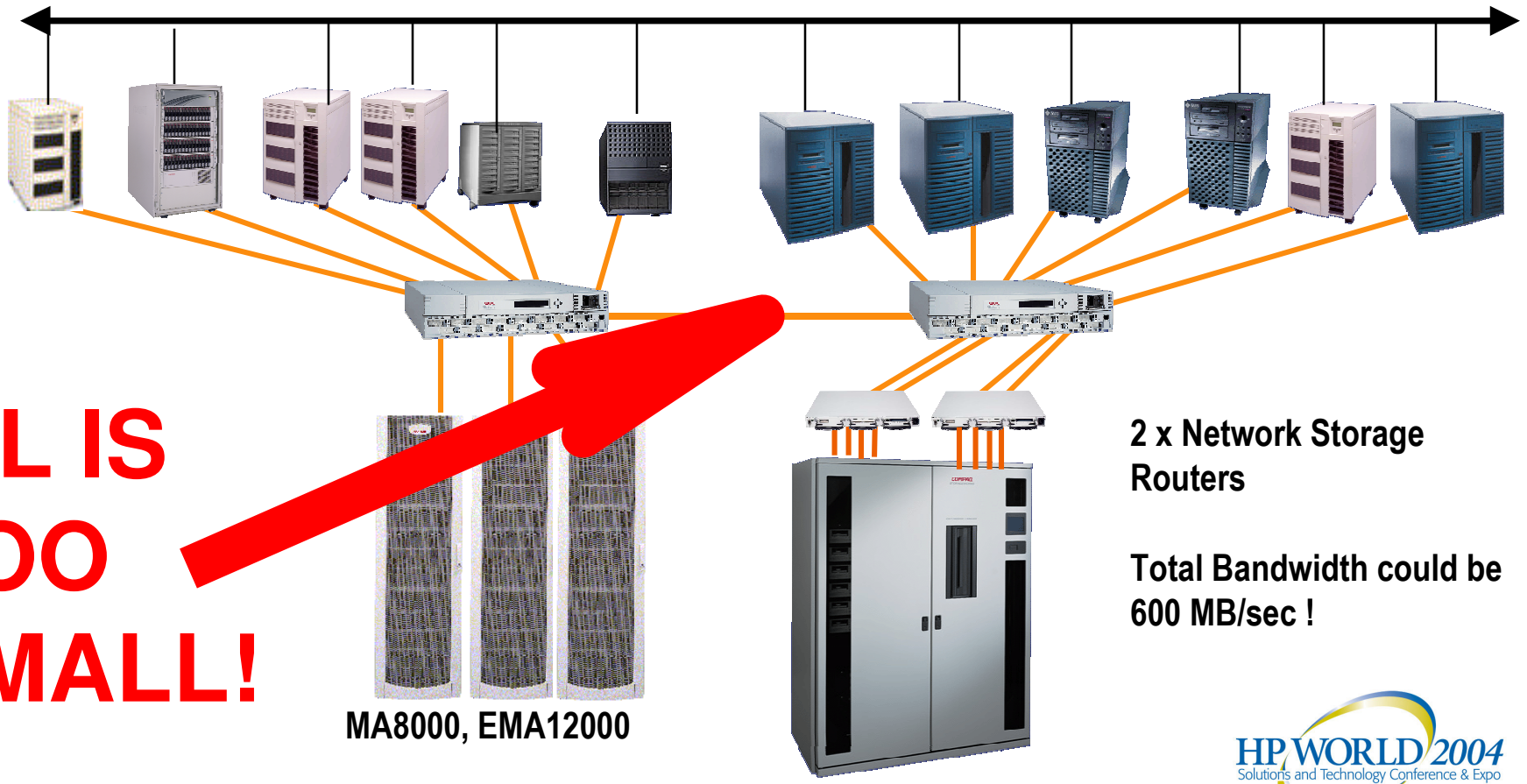
The SAN Switch – TIPS

- **Tuning Considerations: None for individual connections. They usually either WORK WELL or NOT AT ALL!**
- **FABRIC TUNING – The SAN Fabric can consist of many switches that are interconnected via ISL's (Inter Switch Links). Backup tuning should always analyse the pathways from the source to the target (disk to tape) to determine if the paths through the switches are bottlenecked by insufficient ISL's.**

Switch Link Monitoring



What Is WRONG with this Picture?



ISL IS TOO SMALL!

MA8000, EMA12000

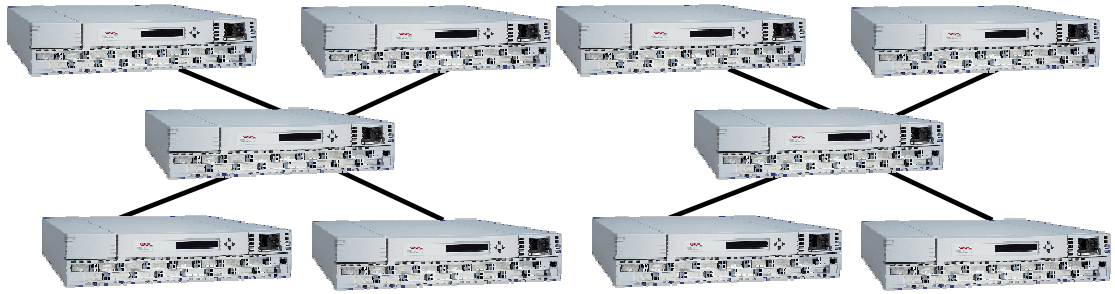
2 x Network Storage Routers

Total Bandwidth could be 600 MB/sec !

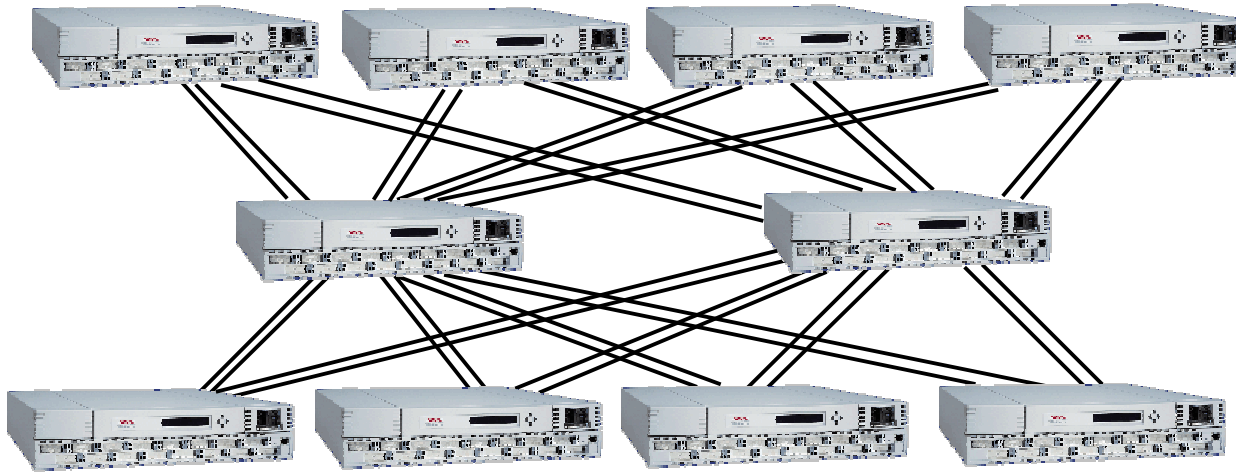
Backbone SAN - Skinny Tree Fabric ISL's

Some methods of Switch Interconnection Techniques

5 Switch Skinny Tree
(56 - 60 F-Ports)



5x2 Switch Skinny Tree (HA - 2 Fabrics)
(112 - 120 F-Ports)



10 Switch Skinny Tree
(96 - 112 F-Ports)

Next generation FC.

Speed	Throughput Mbps (duplex)	Line Rate Gbaud	Release date
1 GFC	200	1.0625	1998
2 GFC	400	2.125	2001
4 GFC	800	4.25	Late 2004
10 GFC	2400	10.5	2004

Recent changes in 4Gb FC

- 4Gb/sec is an extension of 2Gb
- 10Gb is a new development with no backwards compatibility
- Originally 4Gb destined for “intrabox” use only. E.g. inside disk arrays
- Recent FCIA review has decided to move forward with a 4Gb switching network – expect products late 2004
- 10Gb likely to be used for ISL’s only, no “devices” will connect to 10Gb
- Impact for Backup: to avoid wasted bandwidth a controller based approach to libraries is best placed to utilize these enhancements. (see later)

What is an Interface Controller?

- An Interface Controller is
  - NOT just a bridge
  - NOT just a router

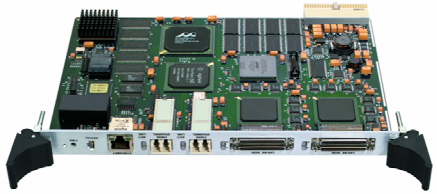
- An Interface Controller
  - HAS the intelligence to actually look at the data, and makes some judgements as to where it will go.
  - HAS the mission to communicate with a management system to provide performance tuning, diagnostic monitoring and reporting abilities.
  - HAS the ability to support advanced functions such as….
    - Selective Storage Presentation
    - Direct (Serverless) Backup
    - Secure Path
    - Partitioning
    - Virtualization

Interface Controller Performance

- Ultra 2 Routers have a total bandwidth of 140MB/sec, 2 SCSI ports

E1200/E2400 – up to 2 x SDLT220/320 or 2 x Ultrium 230 per SCSI port

- s have a total bandwidth of 280MB/sec, 4 SCSI

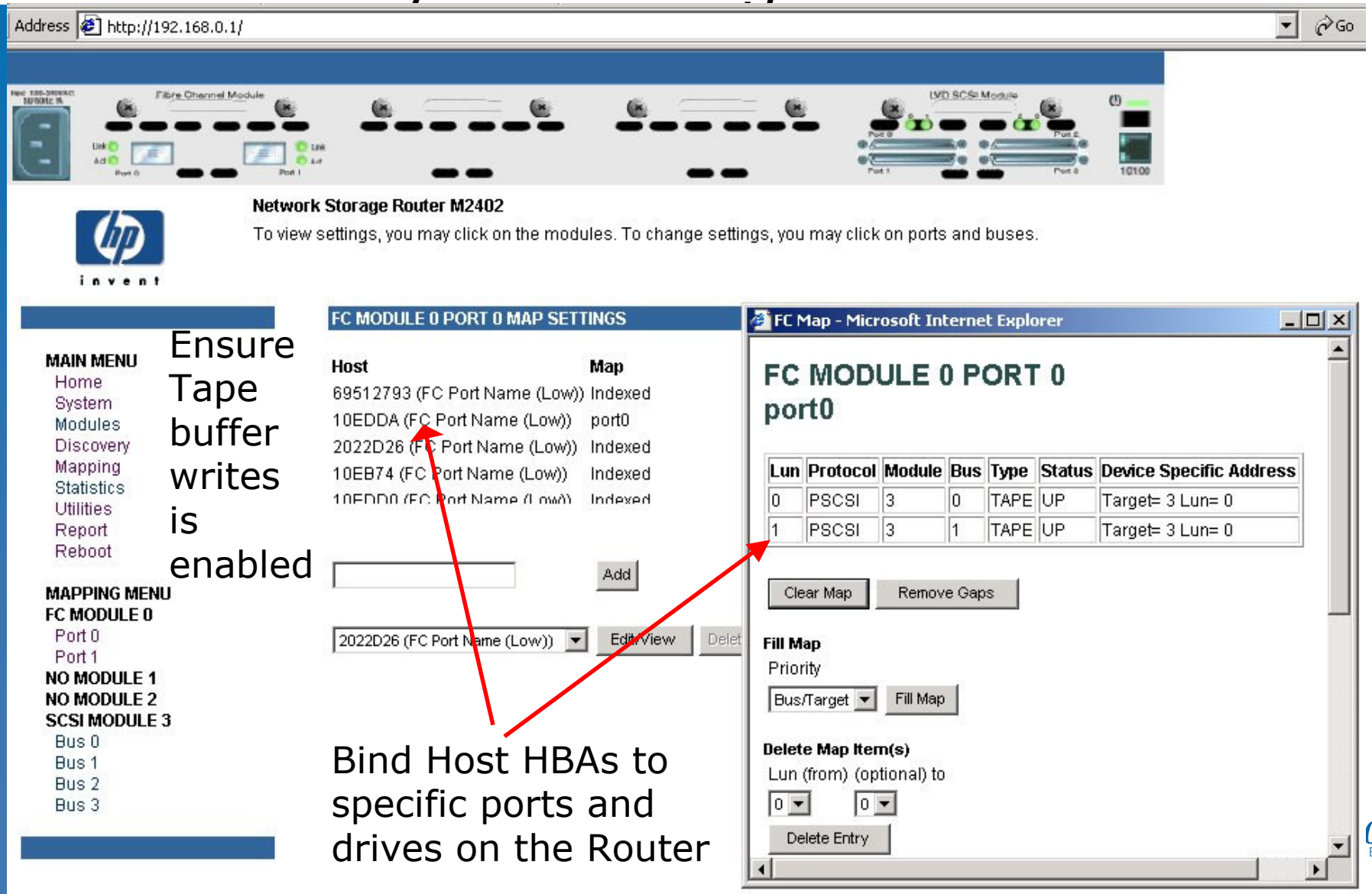


E2400-160 & M2402 – 1 Ultrium 460 per SCSI port

- up to 2 x SDLT220/320 or Ultrium 230 per SCSI port

- **DON'T EXCEED THESE VALUES OR THE ROUTER WILL BECOME THE BOTTLENECK TO BACKUP PERFORMANCE.**

Selective Storage Presentation – optimizes performance by maximizing FC bandwidth



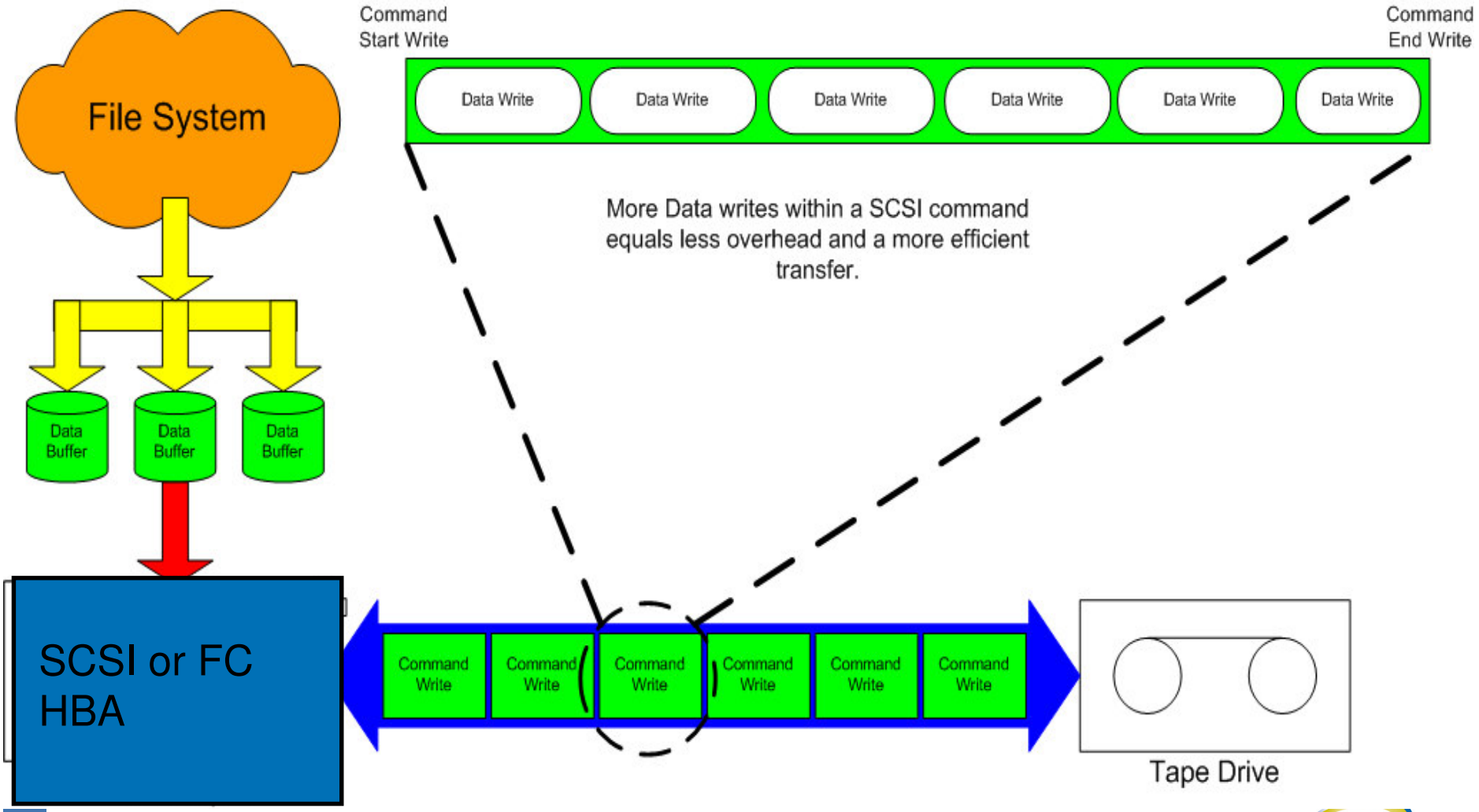
Interface Controller Bottleneck Considerations

- Don't forget to calculate your data throughput rate based on what level of compressibility your data has on your specific tape drive technology. If you don't know, assume at least 2:1, so that you DOUBLE your expected data rates compared to using 1:1 data.
- When you are considering how much data can go through the router at one time, consider analyzing your data sets compressibility and how you might stagger your backup jobs so that data flow based on tested transfer rates (see tools coming up!) are run in a way that doesn't exceed the router's performance limits.

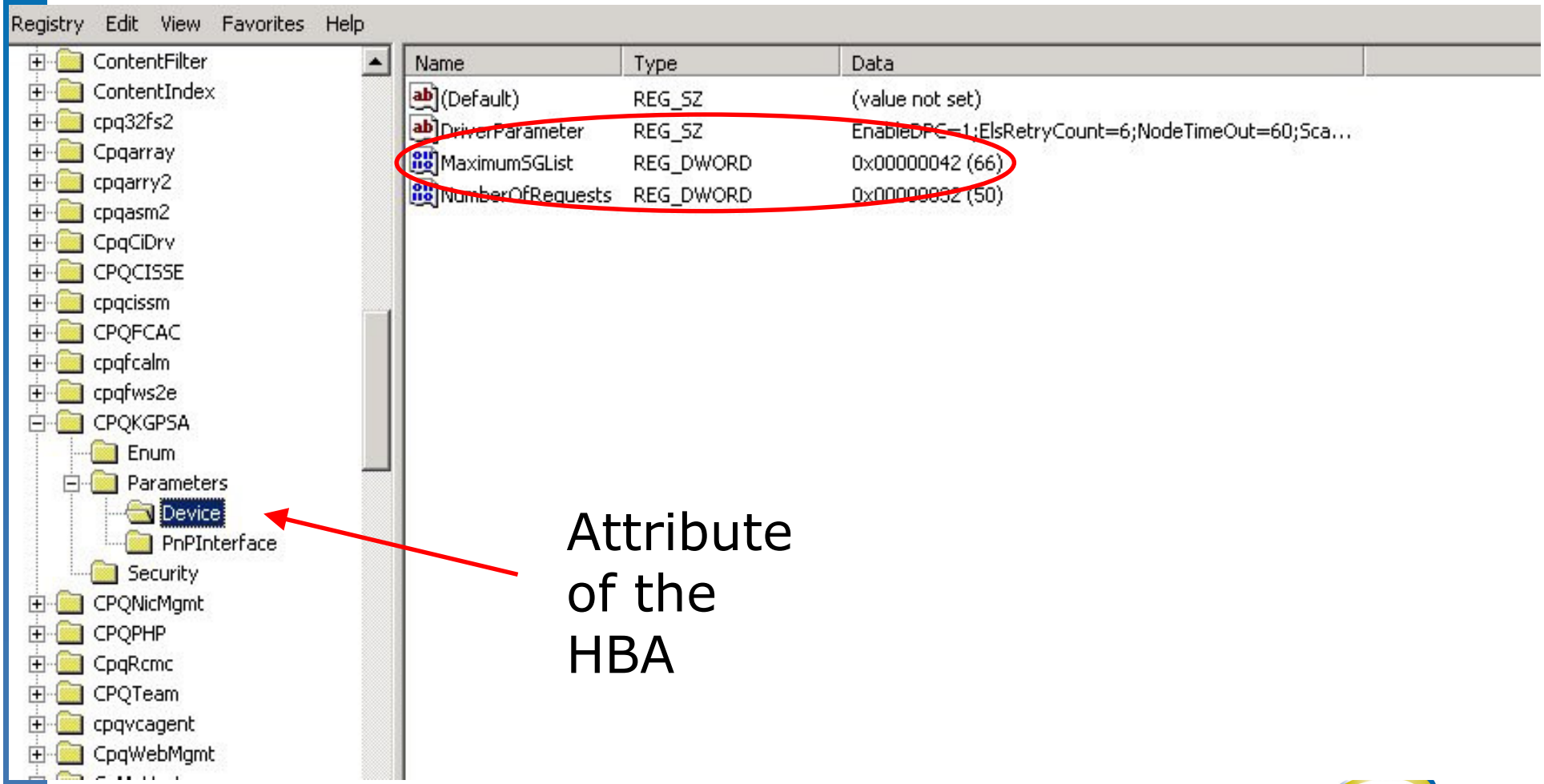
OS Transfer Size

- OS Transfer size is the maximum amount of data the operating system will transfer in a single “operation”
  - Windows is preset to 64K unless MaximumSGList is changed in the registry
  - hp-ux has a 1MB limititation (although 256K at atomic level)
    - kmtune –s scsi_maxphys = <value>
  - More recent Solaris versions have 64K default extendable to 16M via st.conf

Transfer Size – e.g. MaximumSGList



MaximumSGList



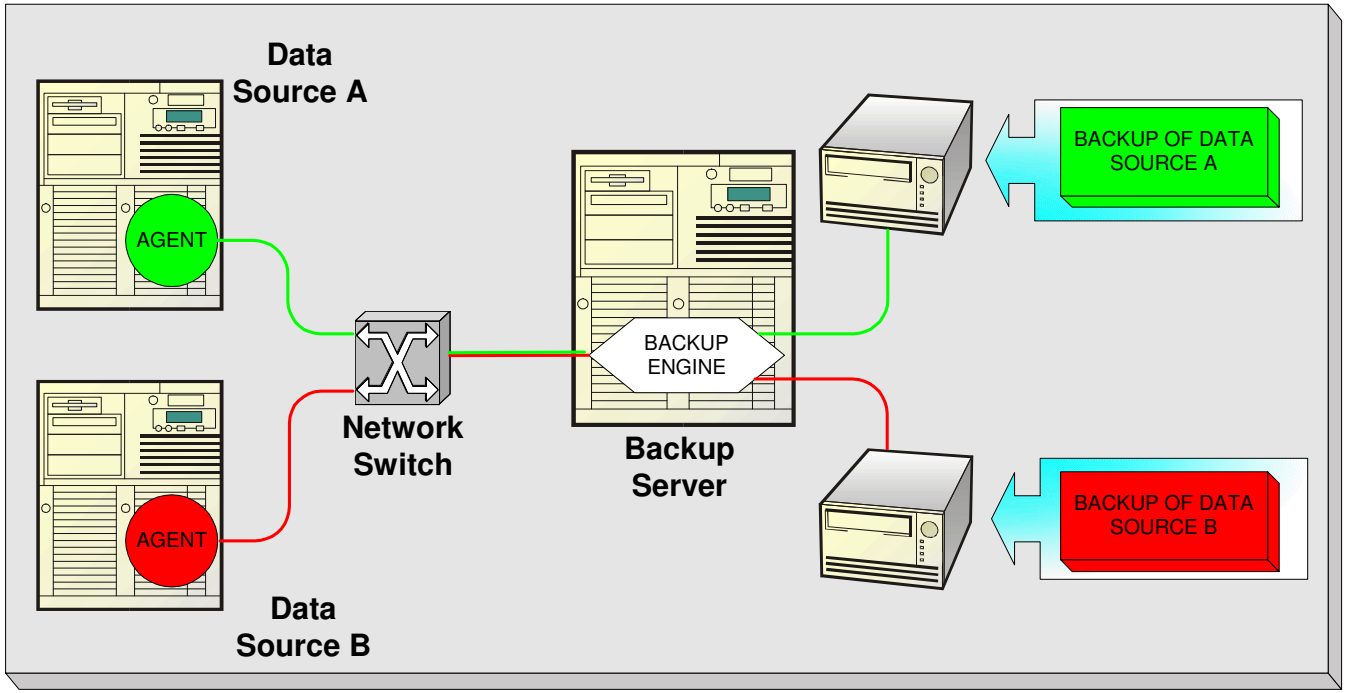
Attribute of the HBA

ISV performance tuning

- Concepts
  - Multiple Streams
  - Multiplexing (optimising tape performance)
  - Image Backup
- Transfer Size
- Database API's (bypass filesystem)
  - RMAN (recovery manager) - Oracle
  - VDI (Virtual device interface) – SQL
  - Limitations of Exchange API
- Samples featuring HP Data Protector.
  - Tape Blocksize/Disk Buffers
  - Concurrency

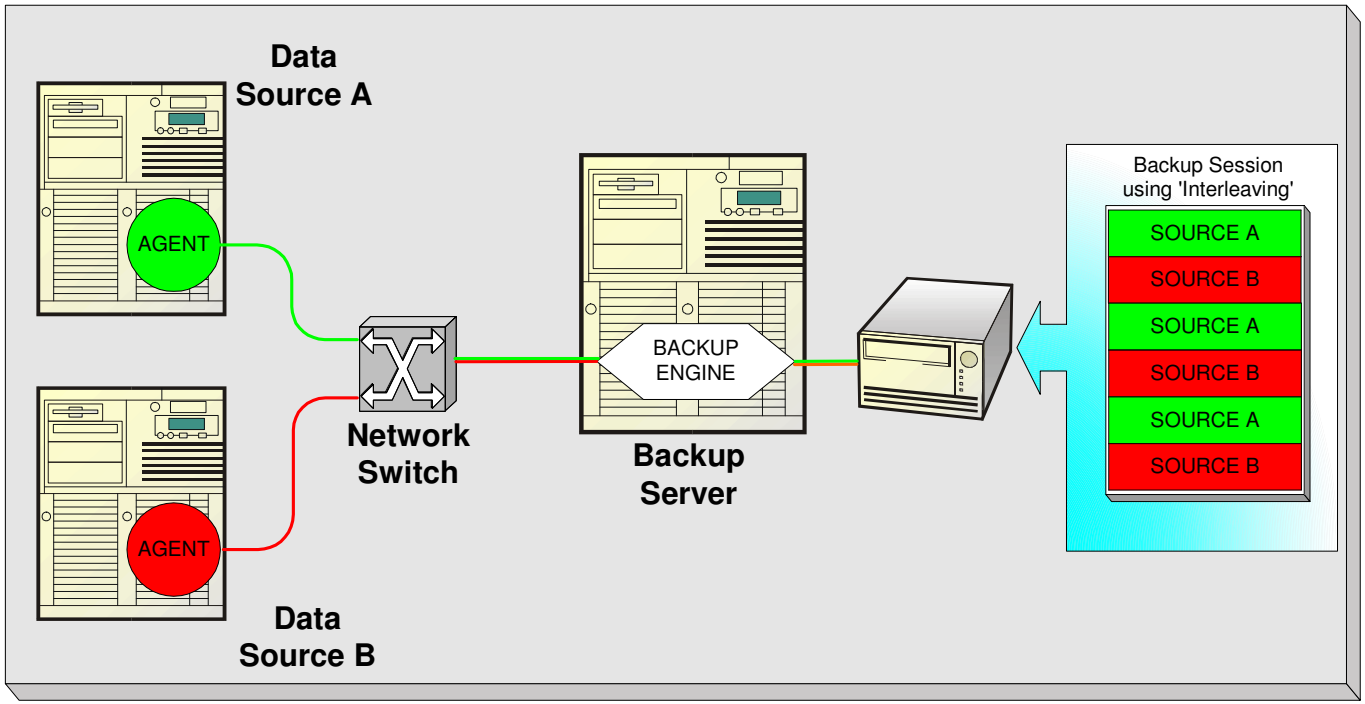
All major Enterprise ISV's support most of these features.

Network Backup with 'Parallel' Data Transfers



- Device Parallelism
  - 'Device Parallelism' or 'Multiple device streams' uses the Principle that data from a specific source system can be routed to a dedicated device in a one-to-one relationship

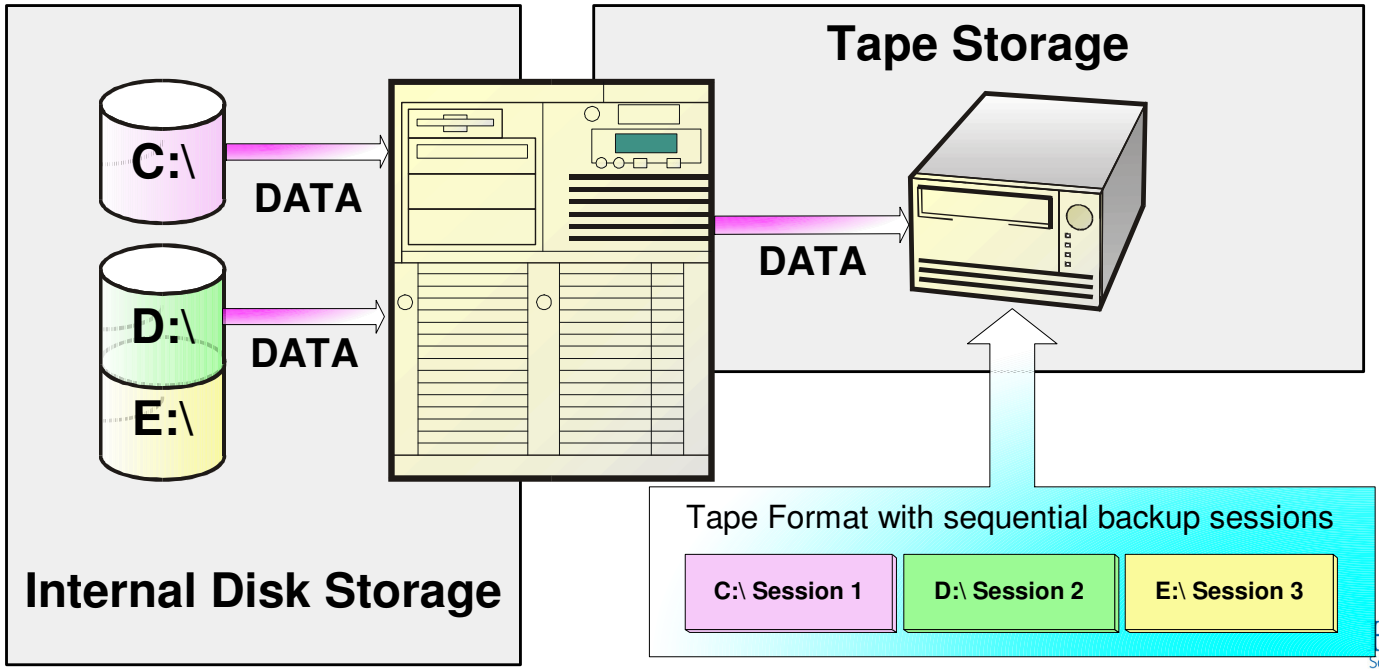
Backup with Concurrency (Interleaving/multiplexing)



- Combines data from multiple Data streams from multiple sources to a ‘specific backup session’ onto a single tape or backup tape set, so we can maintain tape streaming.
- Increases backup performance but can reduce restore performance.

ISV Software 'Features' – Image Option

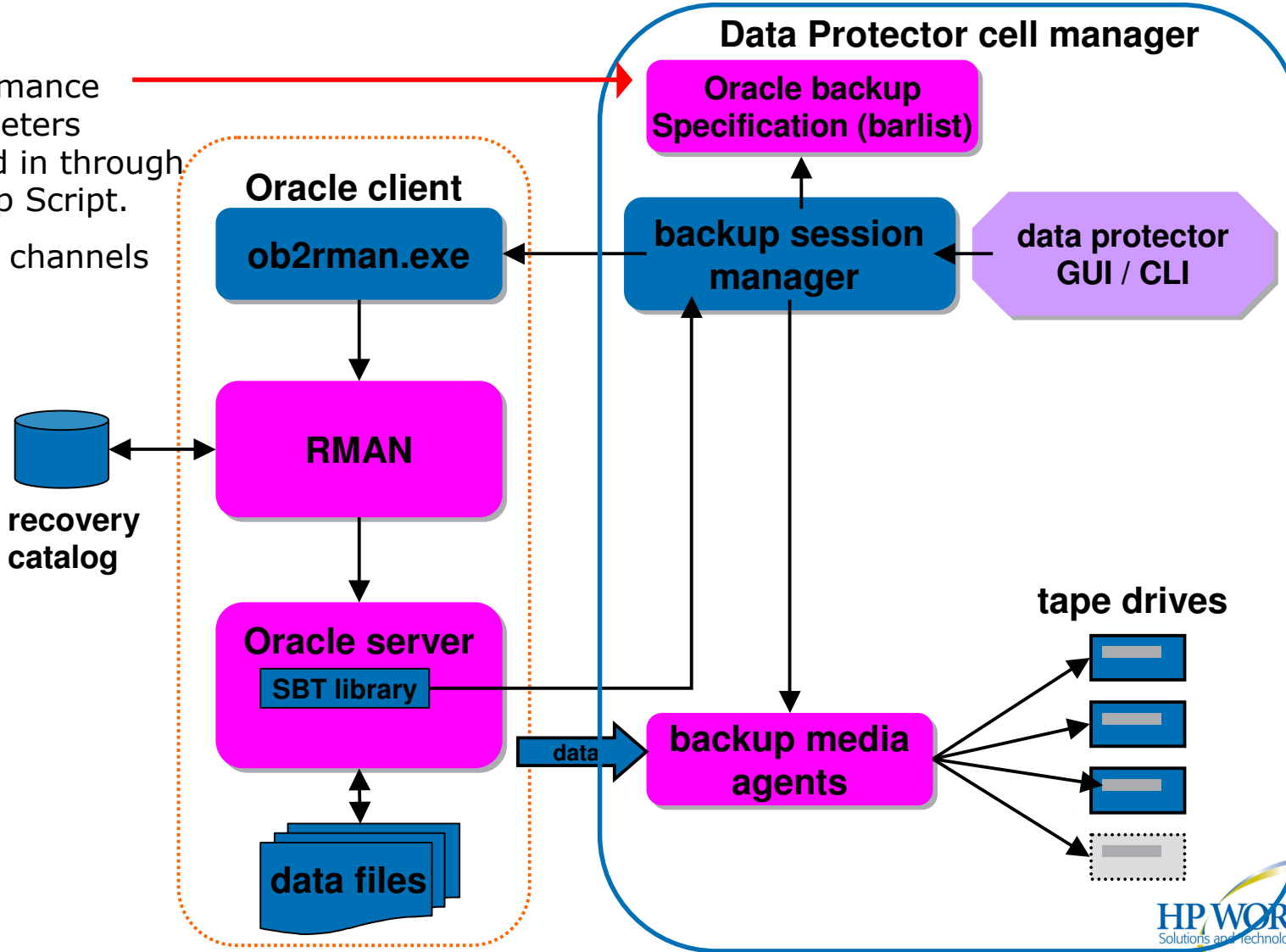
- Image backup operates at the device level rather than the file system level
- The source Drive Must be quiesced
- Lower File-System overhead resulting in higher performance
- Best used when lots of small files would give slow file access.
- Single file restore is possible but slow.



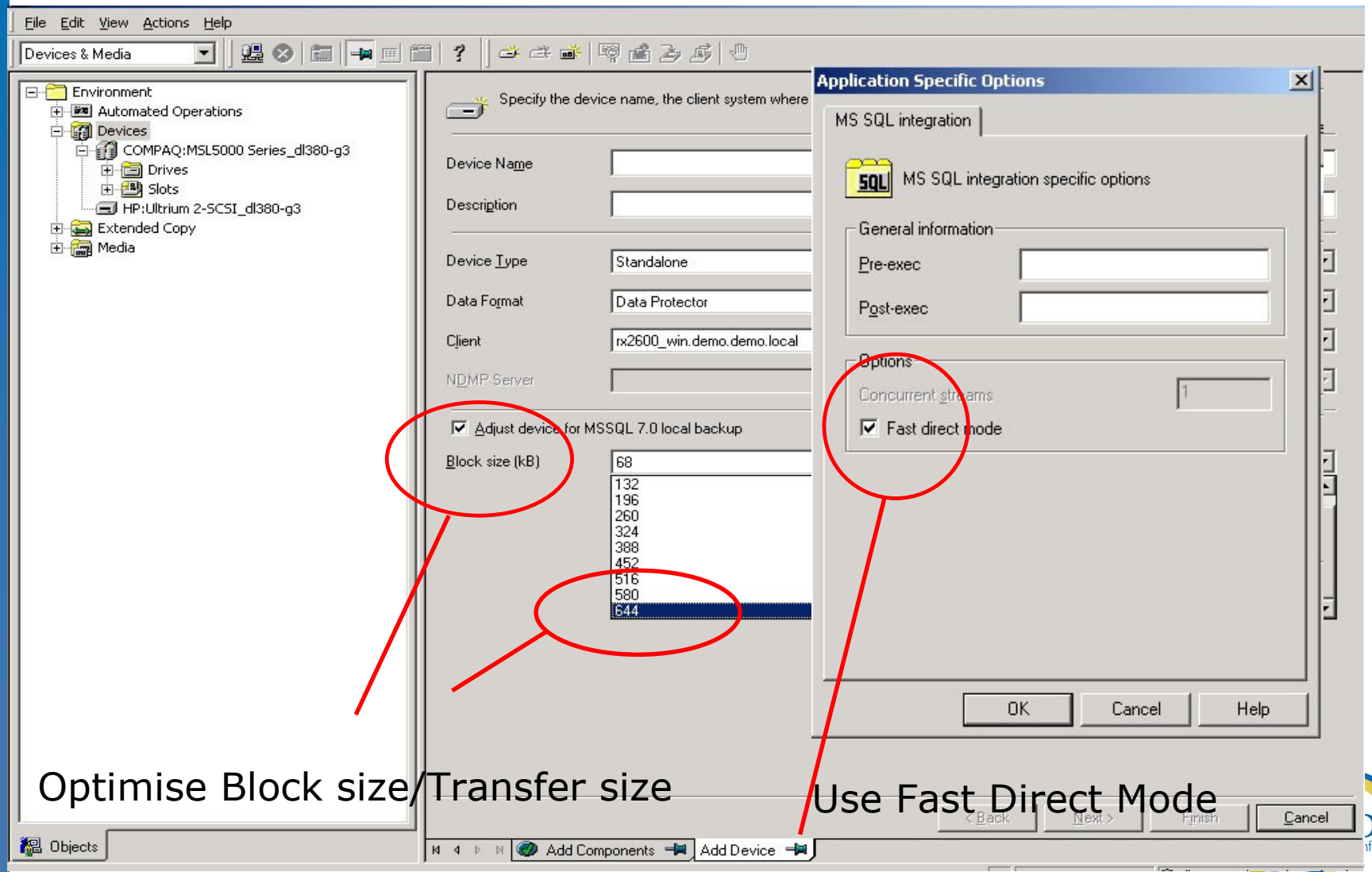
ISV performance tuning – database API's - Oracle

RMAN performance parameters passed in through Backup Script.

E.g. # channels



ISV performance tuning – database API's - SQL



Optimise Block size/Transfer size

Use Fast Direct Mode

ISV performance tuning database API's - Exchange

- No inherent tuning parameters available for Exchange, other tha basic buffer size and # of disk buffers.
- Exchange API is renowned for being a tape performance bottleneck

Example: Veritas NetBackup

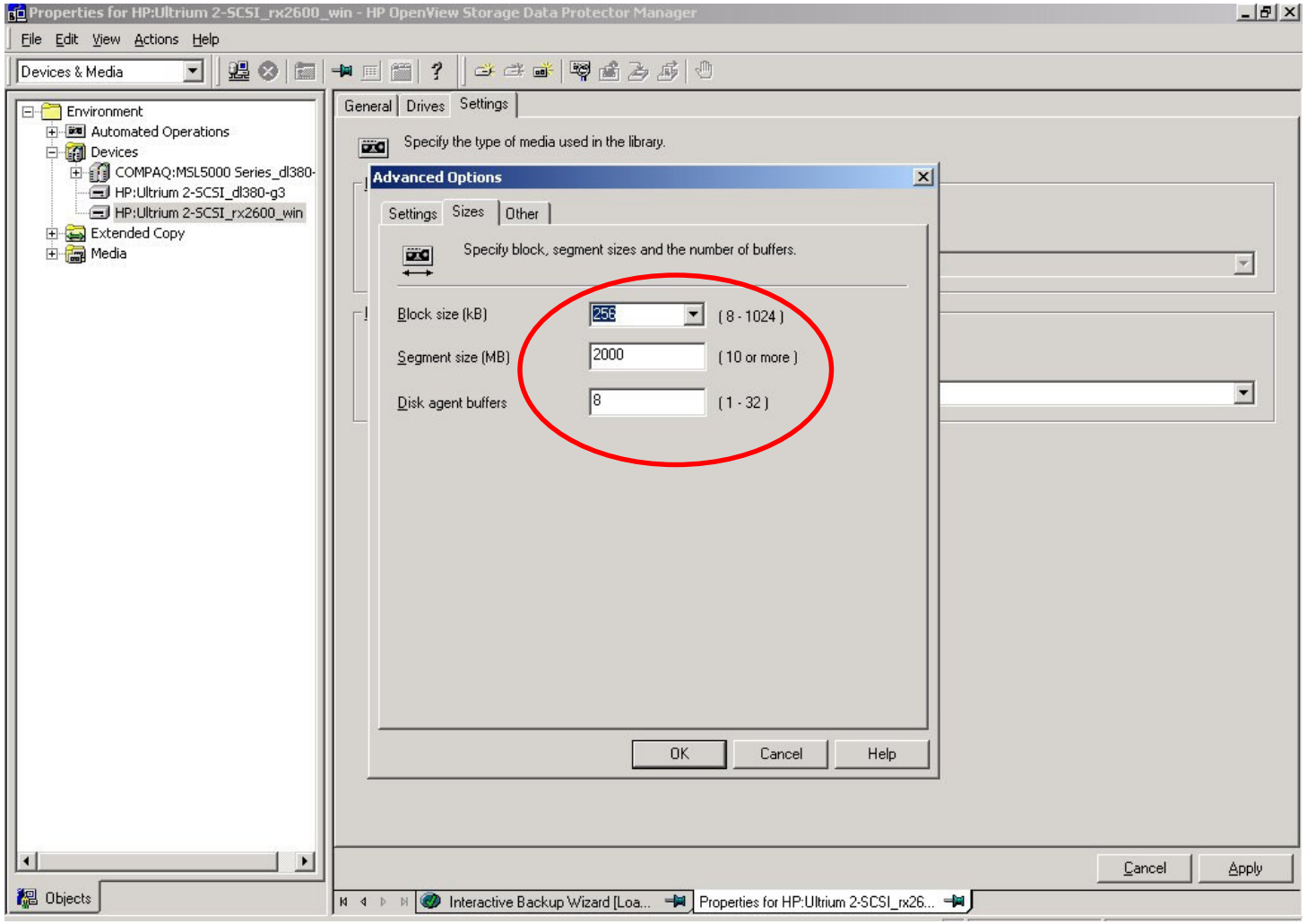
C:\Veritas\Netbackup\db\config\

Two files must be created to override the default NetBackup settings:

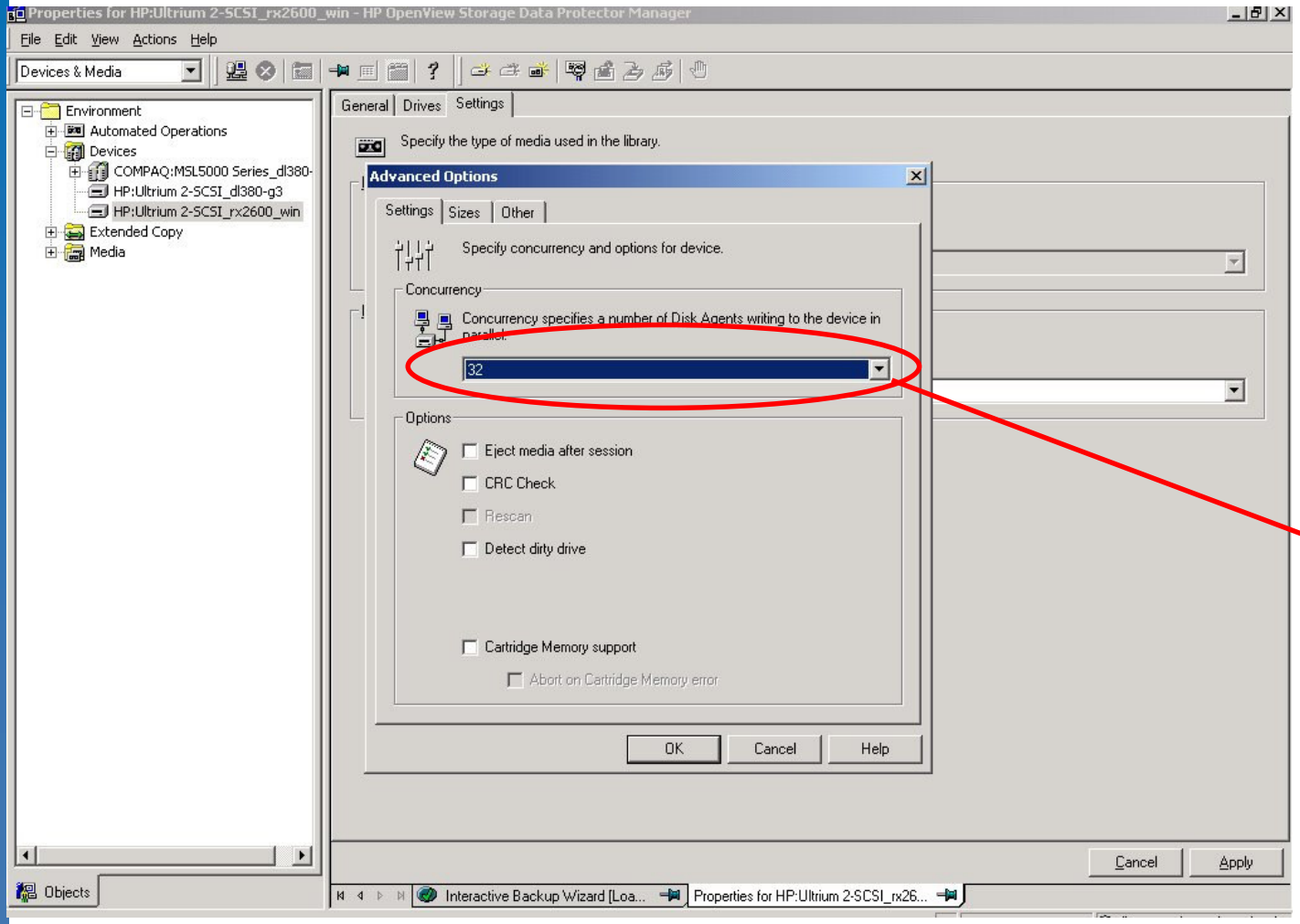
NUMBER_OF_BUFFERS (64)

BUFFER_SIZE (64)

ISV performance tuning
Tape Block Size/Buffers and Segments



ISV performance tuning – Concurrency/Multiplexing



Set at a device level.

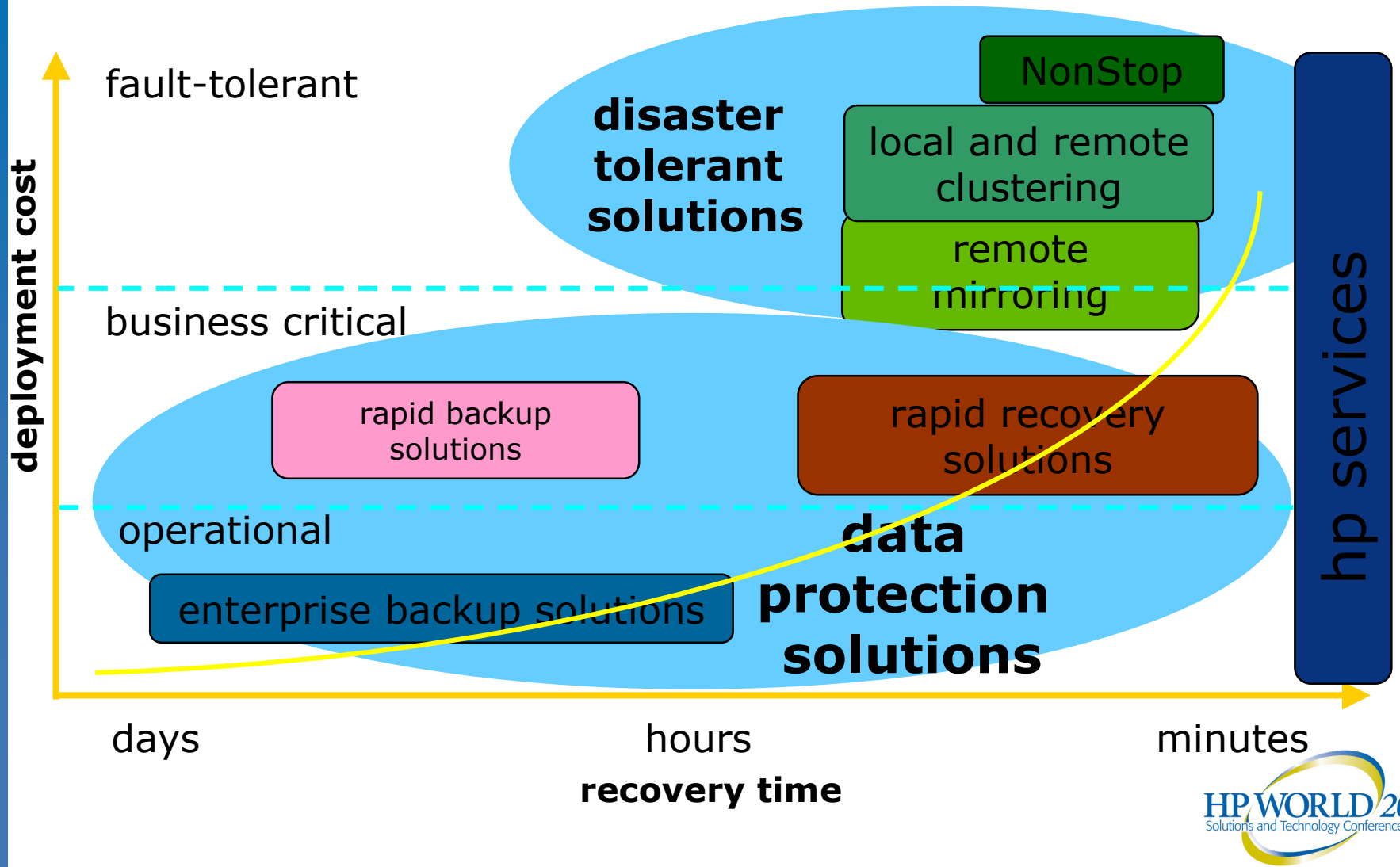
Calculate these concurrency levels as part of an overall backup strategy to meet specific window requirements

ISV features relating to Backup performance

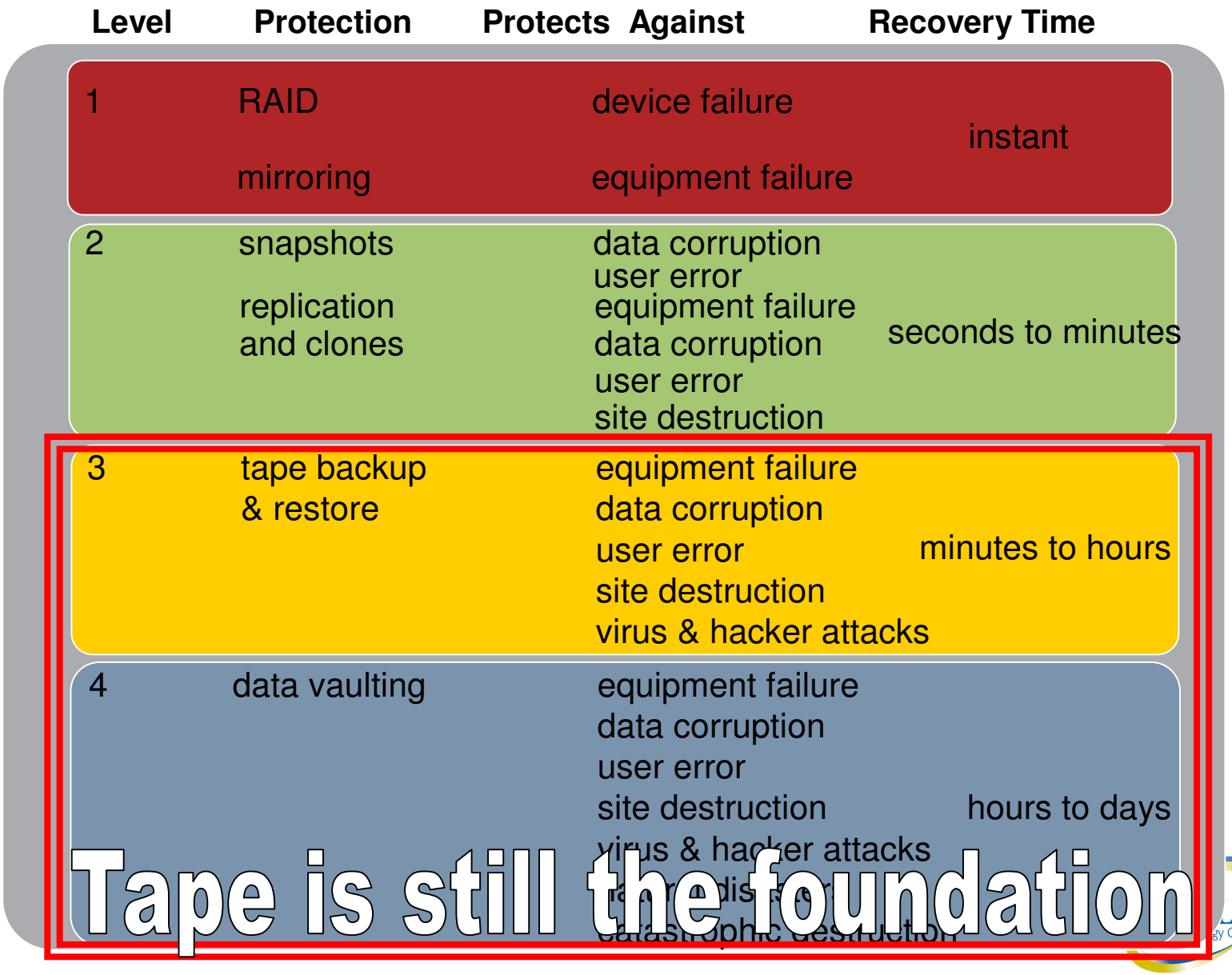
Feature	HP Data Protector	CA BrightStore Enterprise	Veritas NetBackup	Legato Networker
Backup to Disk (fast restore)	✔	✔	✔	✔
3PC/Serverless XCopy/Rapid Backup	✔ (oracle 9i with hp-ux)	✔ (windows & solaris)	✔ (oracle 9i with hp-ux)	✔ (oracle 9i with Network appliance & EMC hw)
Image backup	**X (raw disk via Oracle RMAN)**	✔ **(Windows only)**	✔ (Solaris & hp-ux **)**	✔ **(Windows & unix)**
Parallel Streaming	✔	✔	✔	✔
Concurrency (multiplexing)	✔ (32)	✔ (32)	✔ (32)	✔ (32)

What About Restore Performance ?

HP: A Range Of Business Continuity Solutions



multi-level data protection



Recovery considerations

- With higher performance environments recovery from catastrophic failure can take place at speeds of around 2 Terabytes/Hr Maximum. (16 drives in Library)

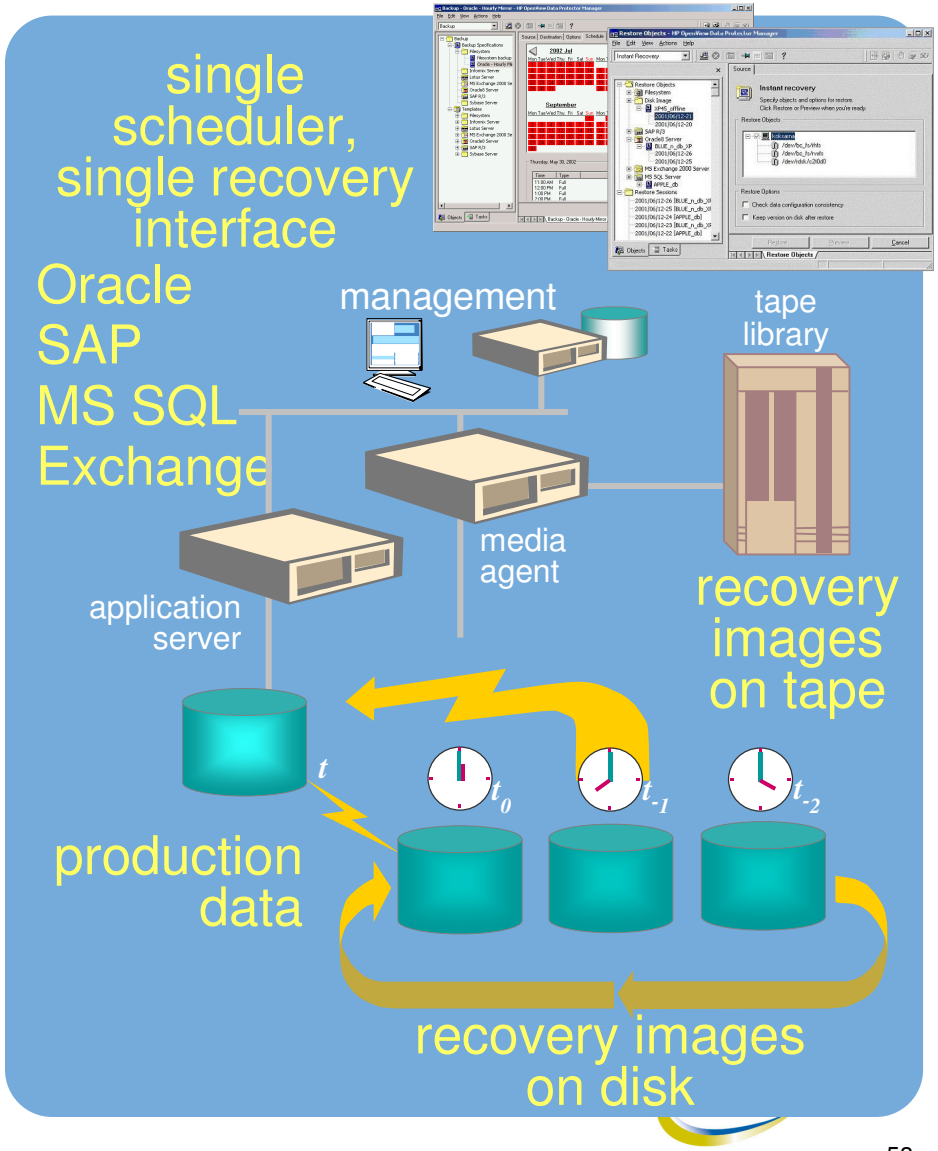
- Main restore bottleneck is in re-creating RAID consistency and creating files through the filesystem.

- Consider Parallelism to improve recovery times.

- Single file recovery performance can also be dependent on Drive search performance and ISV search algorithms.

Integrated fast recovery

- builds on zero-downtime backup techniques to retain multiple images on disk for selective recovery to any point-in-time image
- A fully automated protection process, including creation and rotation of mirrors or snapshots and regular backup to tape.
- for recovery, administrator selects a specific recovery image from the graphical user interface



Pros and cons of fast recovery technologies

- Zero-Downtime Backup
  + no impact on application performance
  - requires specific arrays and software

- Instant Recovery
  + recovery of TBs in minutes
  - requires Zero-Downtime Backup as a basis

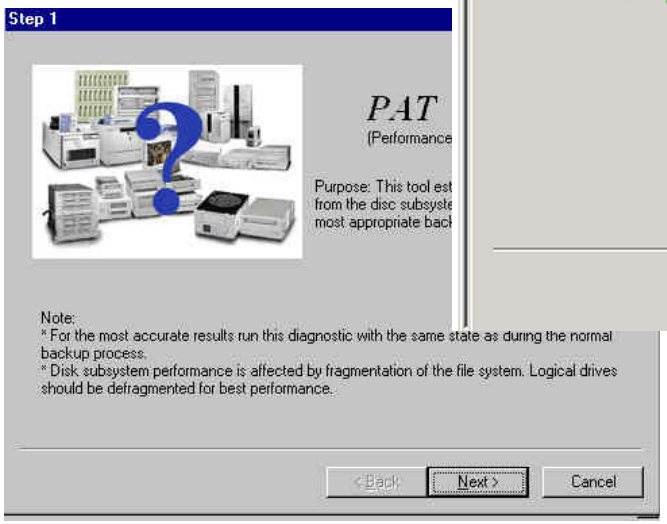
- Volume ShadowCopy Service (VSS)
  + simple mirroring on any disc
  - supported within Windows 2003 only

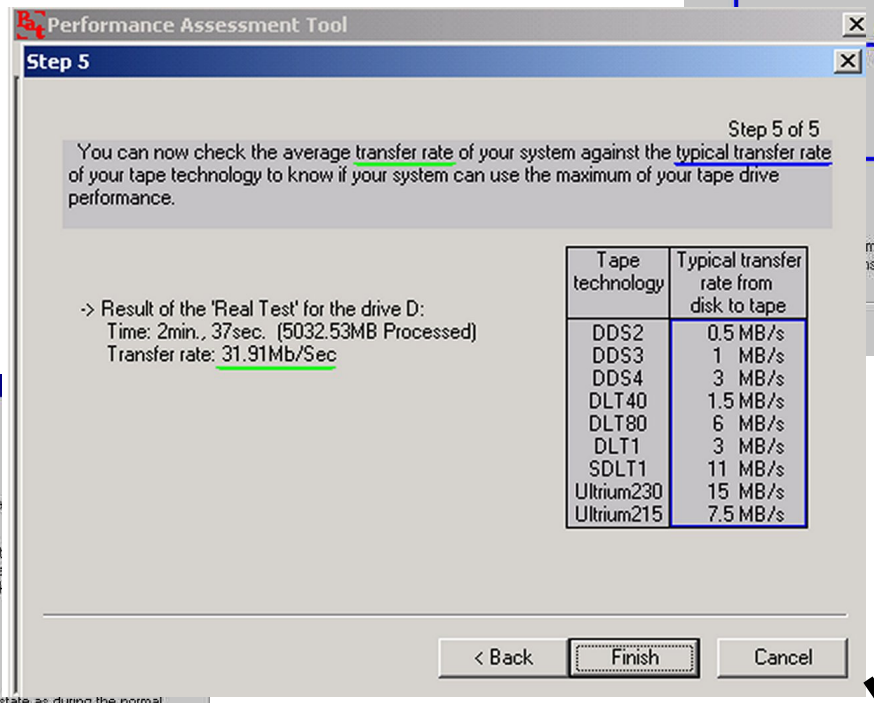
Tools to help you do the job

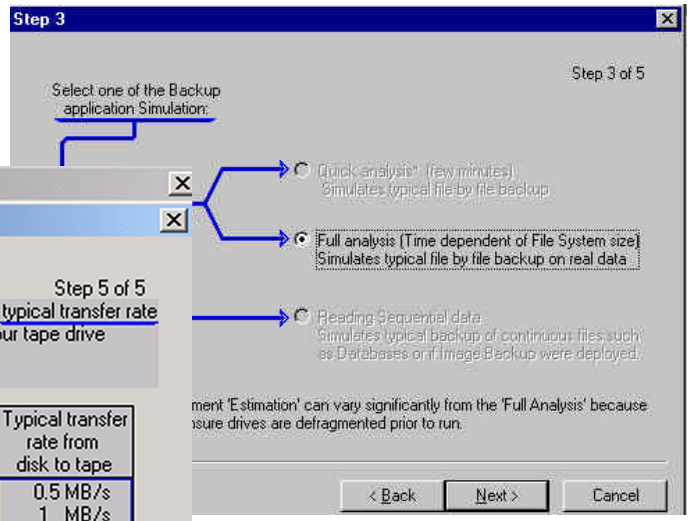
hp tape performance assessment tools for Windows, Linux, HP-UX, Solaris, AIX , Windows(64)

HPTapeperf

PAT







HPCreatedata

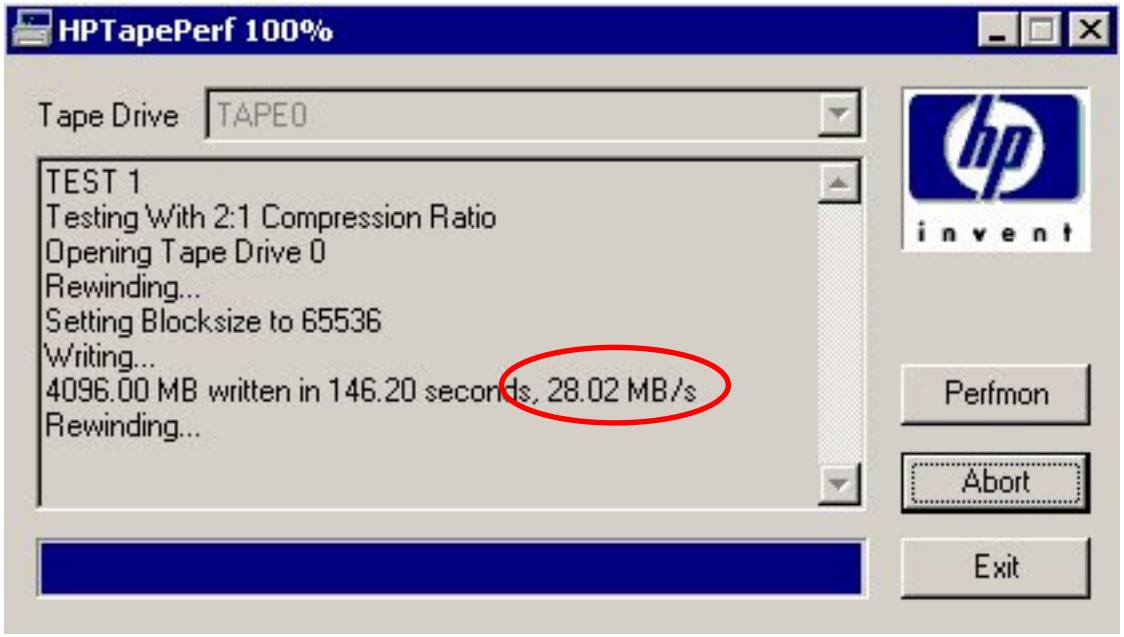
HPScandisk

90% of backup performance issues are NOT the tape drive itself

Free download from …http://www.hp.com/go/tape

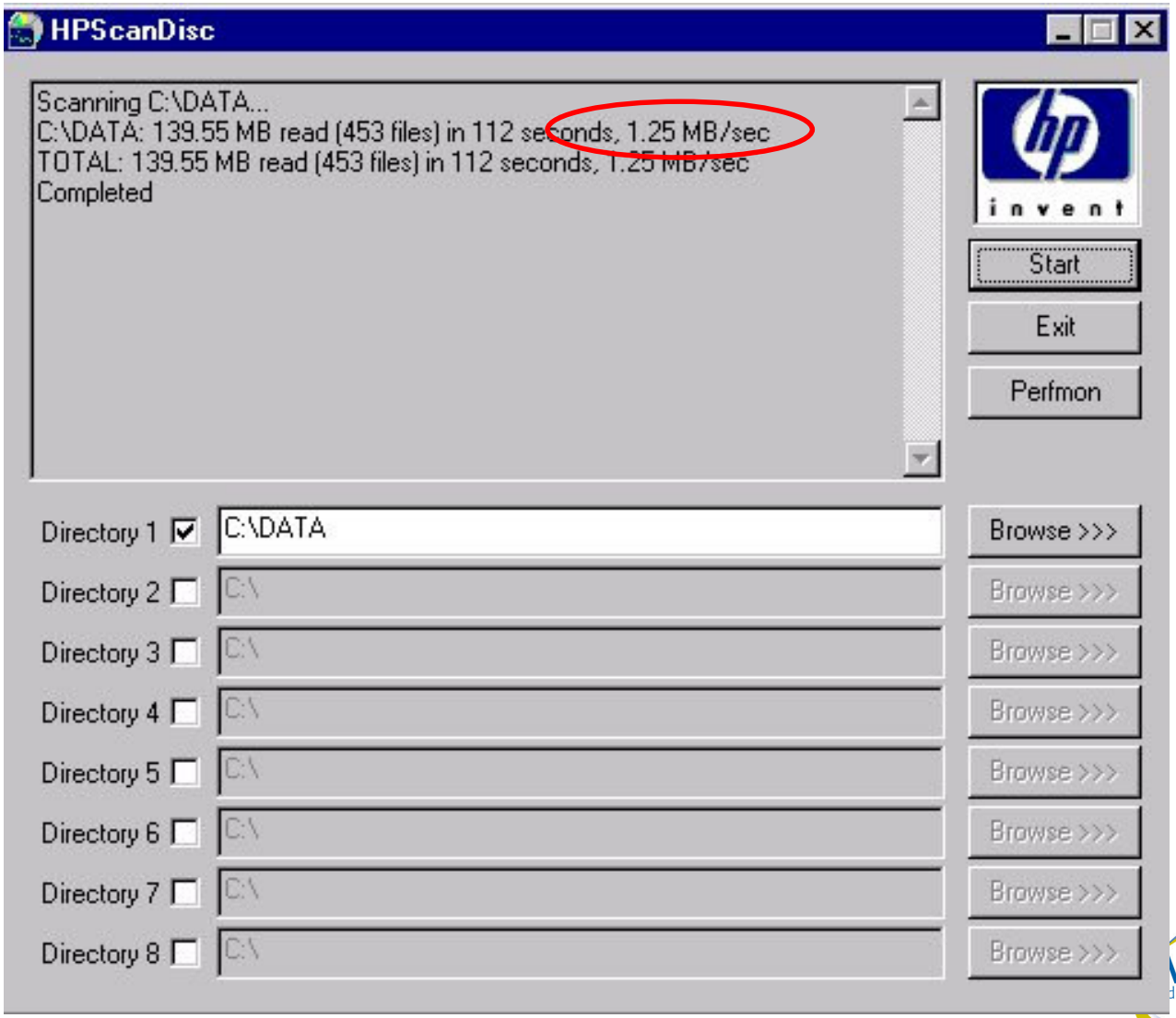
HP WORLD 2004
Solutions and Technology Conference & Expo

HPTapePerf

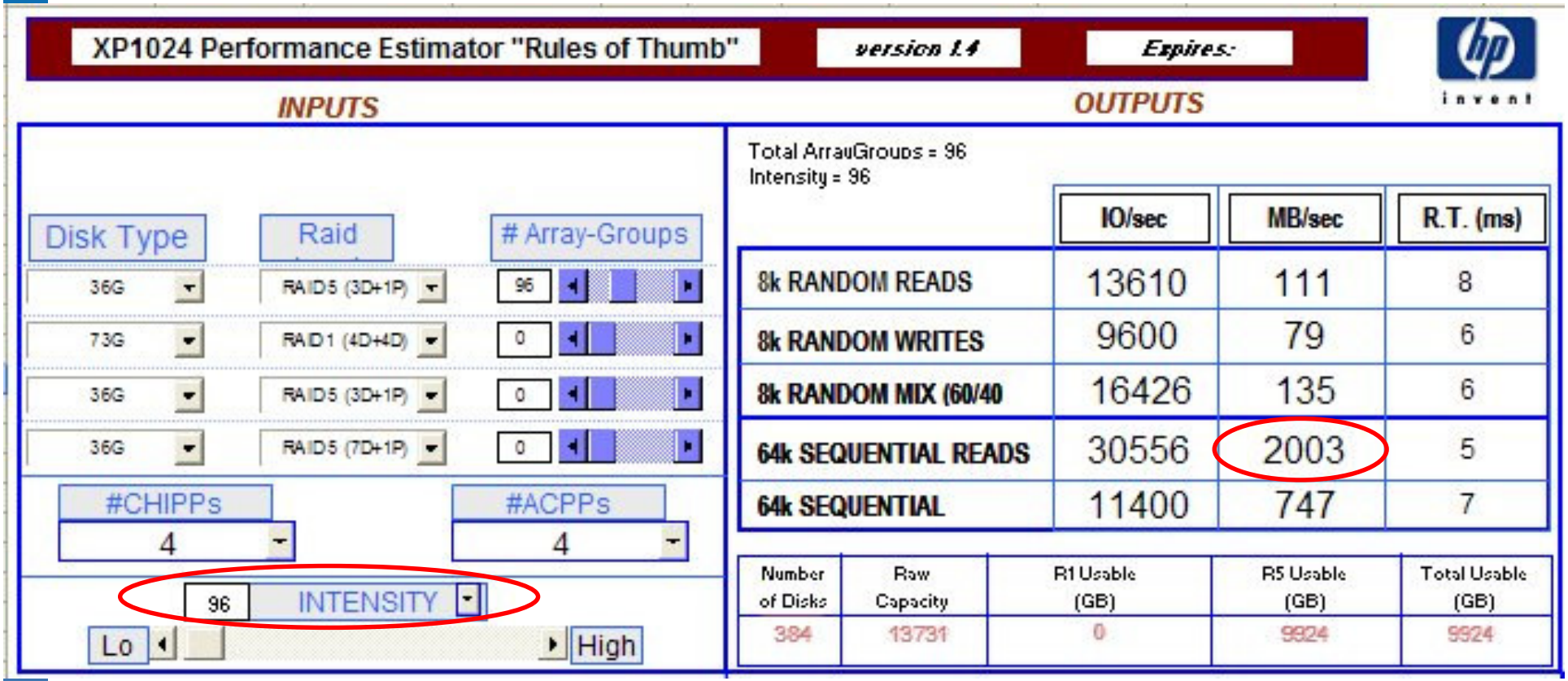


Proves the tape drive is not the bottleneck by writing data from memory direct to tape.

HPScandisc


HPScanDisc
Scanning C:\DATA...
C:\DATA: 139.55 MB read (453 files) in 112 seconds, 1.25 MB/sec
TOTAL: 139.55 MB read (453 files) in 112 seconds, 1.25 MB/sec
Completed
invent
Start
Exit
Perfmon
Directory 1
C:\DATA
Browse >>>
Directory 2
C:\
Browse >>>
Directory 3
C:\
Browse >>>
Directory 4
C:\
Browse >>>
Directory 5
C:\
Browse >>>
Directory 6
C:\
Browse >>>
Directory 7
C:\
Browse >>>
Directory 8
C:\
Browse >>>

XP Disk Performance Estimator.



XP1024 Performance Estimator "Rules of Thumb"

version 1.4

Expires:

INPUTS

Disk Type	Raid	# Array-Groups
36G	RAID5 (3D+1P)	96
73G	RAID1 (4D+4D)	0
36G	RAID5 (3D+1P)	0
36G	RAID5 (7D+1P)	0

#CHIPPs: 4

#ACPPs: 4

96 INTENSITY

Lo — High

OUTPUTS

Total ArrayGroups = 96
Intensity = 96

	IO/sec	MB/sec	R.T. (ms)
8k RANDOM READS	13610	111	8
8k RANDOM WRITES	9600	79	6
8k RANDOM MIX (60/40	16426	135	6
64k SEQUENTIAL READS	30556	2003	5
64k SEQUENTIAL	11400	747	7

Number of Disks	Raw Capacity	R1 Usable (GB)	R5 Usable (GB)	Total Usable (GB)
384	13731	0	9924	9924

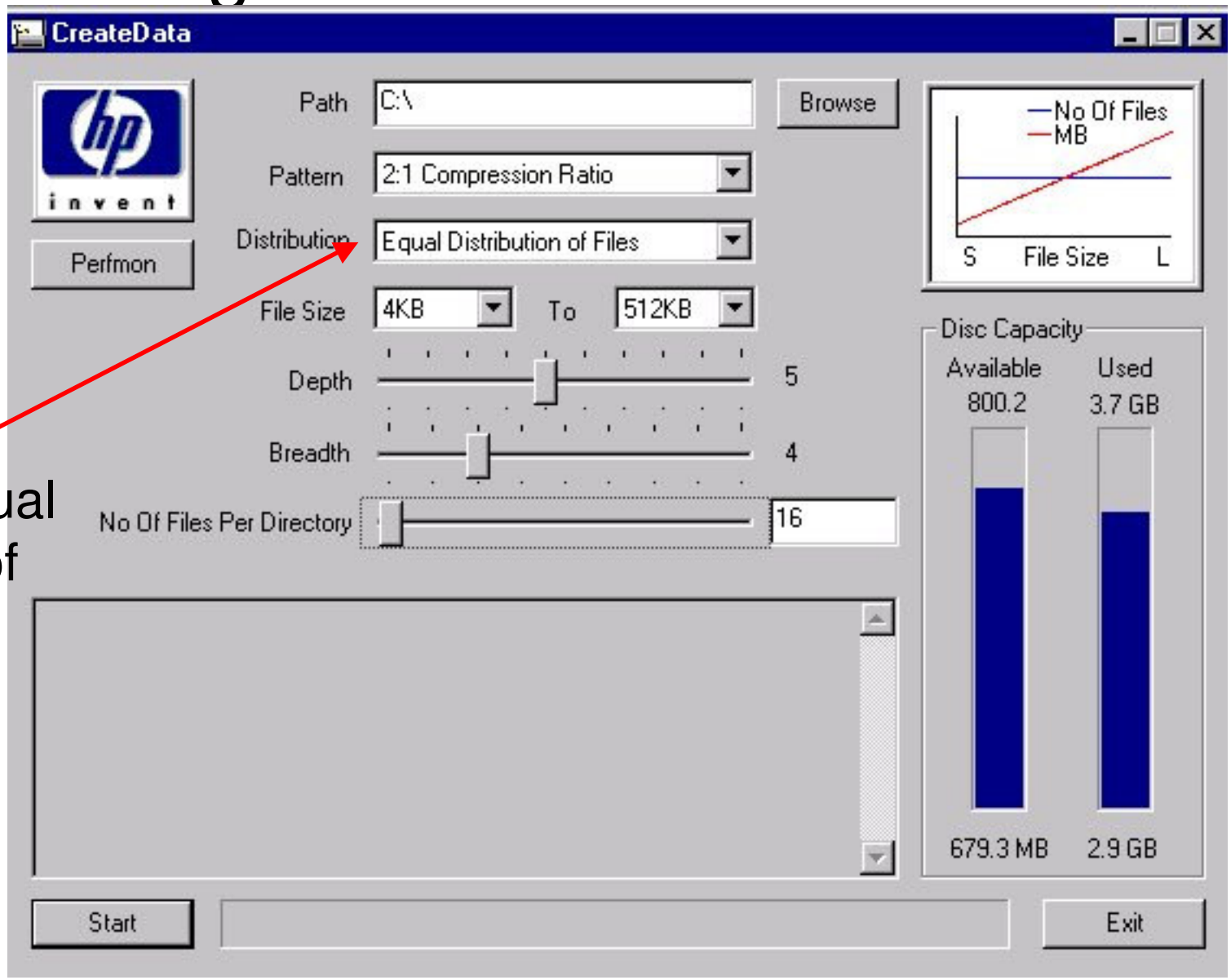
For EVA & MSA – try IOmeter

MSA1000
Access pattern: 64 Kilobyte block size, 100% read, 0% Write,
100% sequential, 0% random

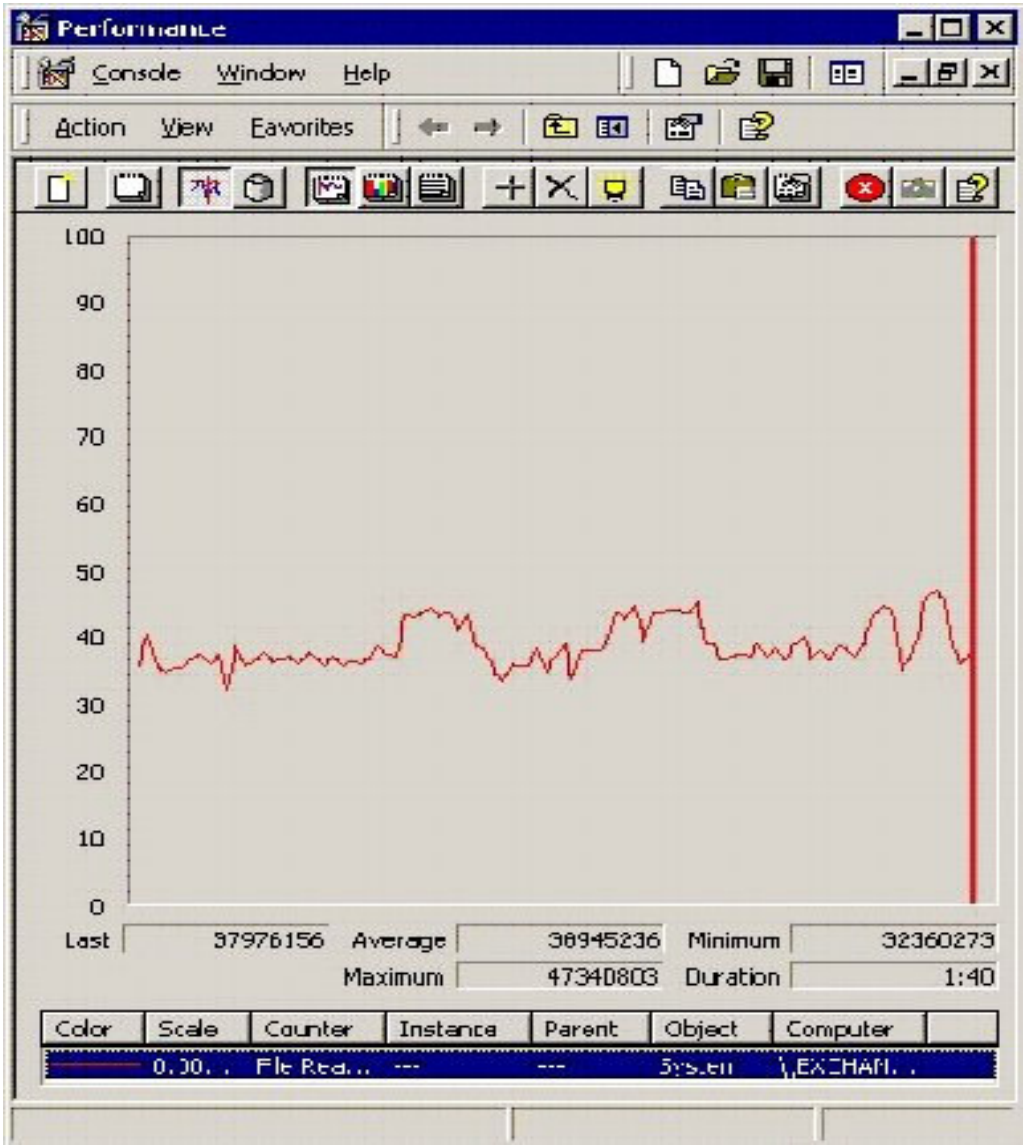
RAID Level	I/O per SEC	MB/sec
0	2445	152.8
1+0	2403	150.2
5	2403	150.2
ADG	2407	150.4

CreateData – for use in Benchmarking

Can also generate equal distribution of MB



There's always performance monitor

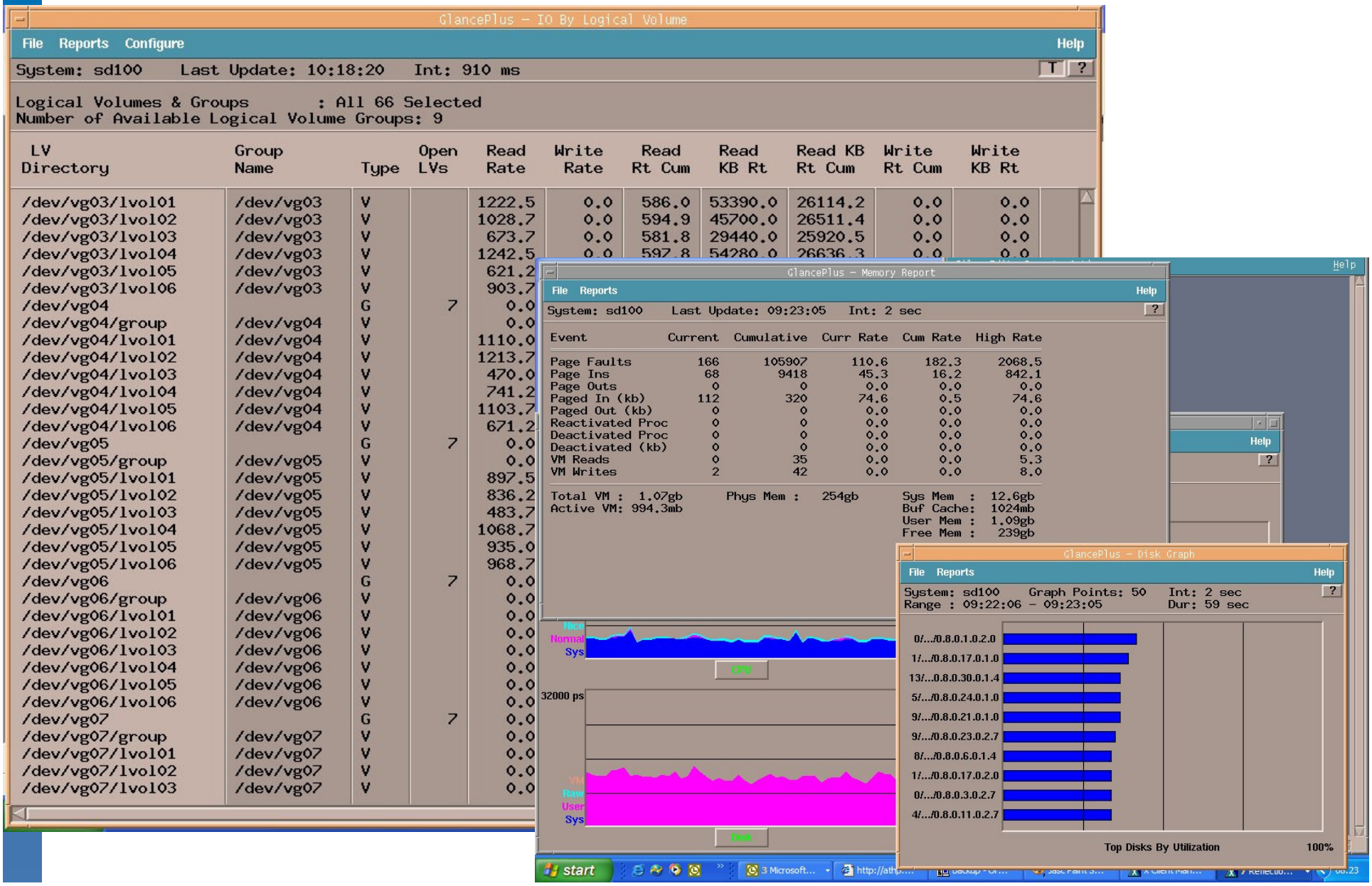


Sometime Backup apps have a lot of pre & post processing that disguise the actual tape backup rate.

On windows platforms

Performance Monitor set to Monitor "system" "File reads" gives a good idea of instantaneous backup rate.

Glance on hp-ux



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