Steven Johnson
June 24, 2013 Colorado Springs

Agenda

- Introduction to SNIA Emerald Hot Banding test
- Concatenated Work Space
- Discussion of the components of the workload
  - Random
  - Five Sequential read streams
  - Three Sequential write streams
  - Four hot bands
  - Complex mix of transfer sizes.

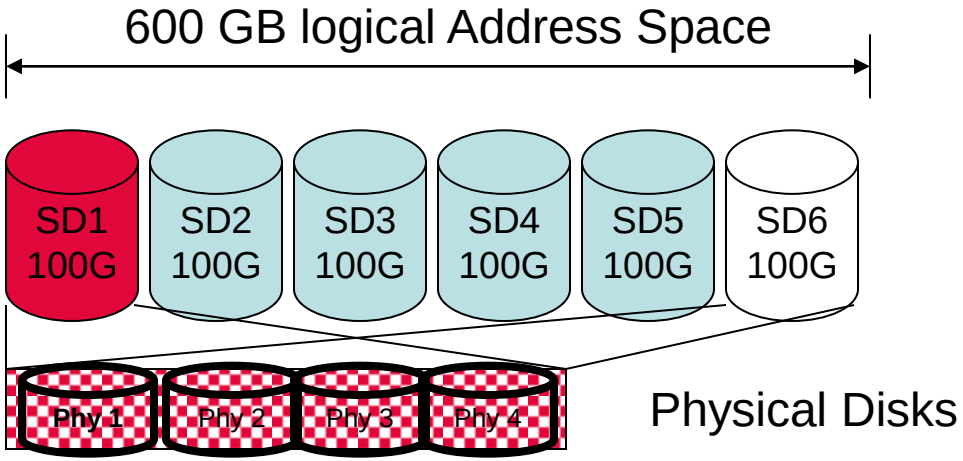
- Scale-ability – able to increase in throughput or performance with increasing application demands
- Utilization – How busy a resource is during a period of time. Generally expressed as a percent from 0 – 100
- Service time – Generally the actual time something take for a specific task
- Response time – Usually considered Service time plus queueing time for resource
- Latency – the period of time one component in a system is waiting for another component
- Data transfer time – The latency required to transfer the requested data from a resource
- Queueing – The natural process of things lining up to be services
- Queueing Theory – The Mathematical study of Queueing systems
- Queue depth – Frequently associated with number of outstanding IOs to a Storage System
- Cache – Placing frequently used things in an easily accessible place. For computers, placing data in a place that has much faster access time.

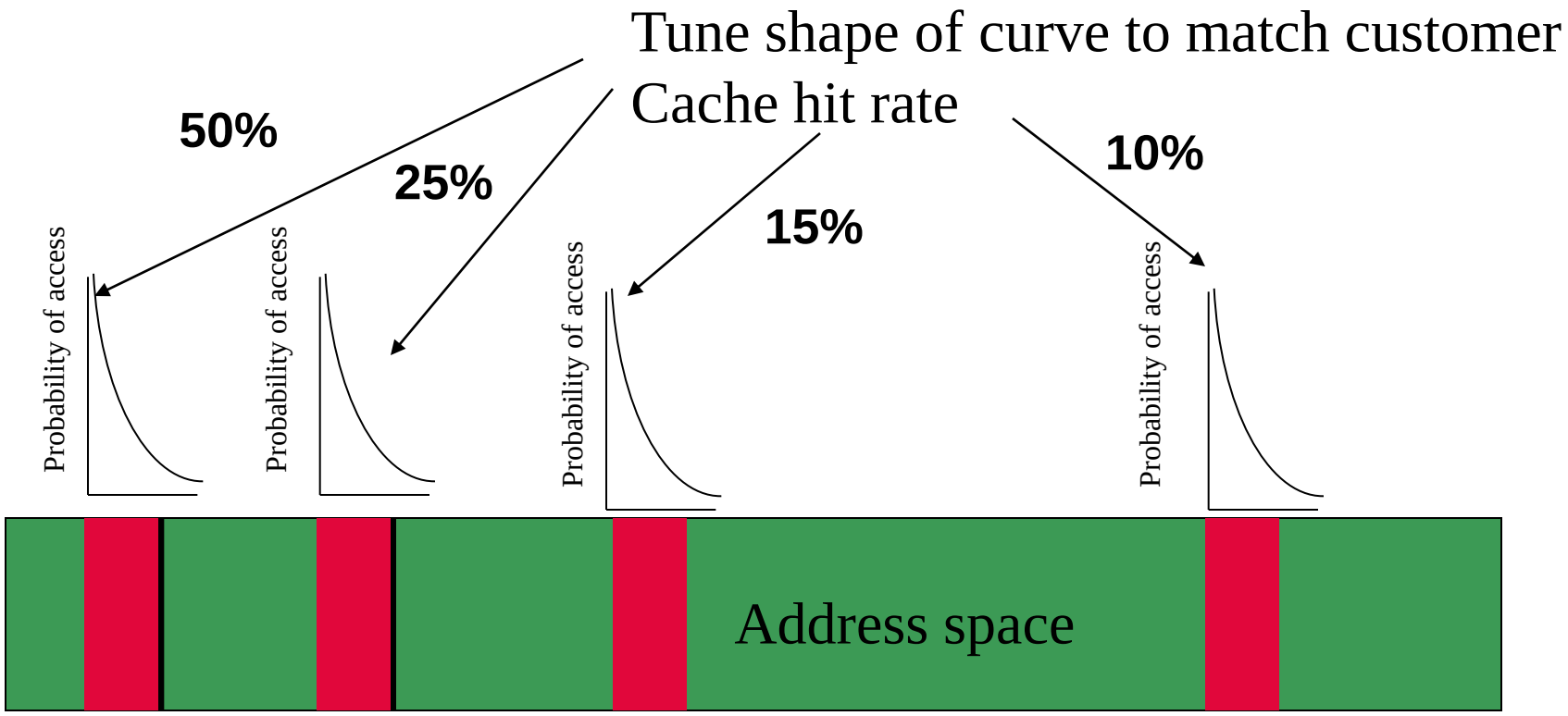
- Cache hit – Information the system is looking for is located in high-speed memory
- Cache miss – Information was not in high-speed memory and had to be found on a slower device
- Sequential – Type of workload that can read or write something one block after another.
- Logically sequential – An application may read or write a file from beginning to end
- Physically sequential – While while an application may think it is reading physically sequential, generally this is not the case. Dd at the raw level can create physically. Seq workloads
- Random – Access pattern moves around a file or physical device
- Locality of Reference – Accesses are concentrated in a particular area (i.e. head of indexes of a data base)
- Solid State Disk (SSD) – A disk drive whose storage capability is provided by solid state storage

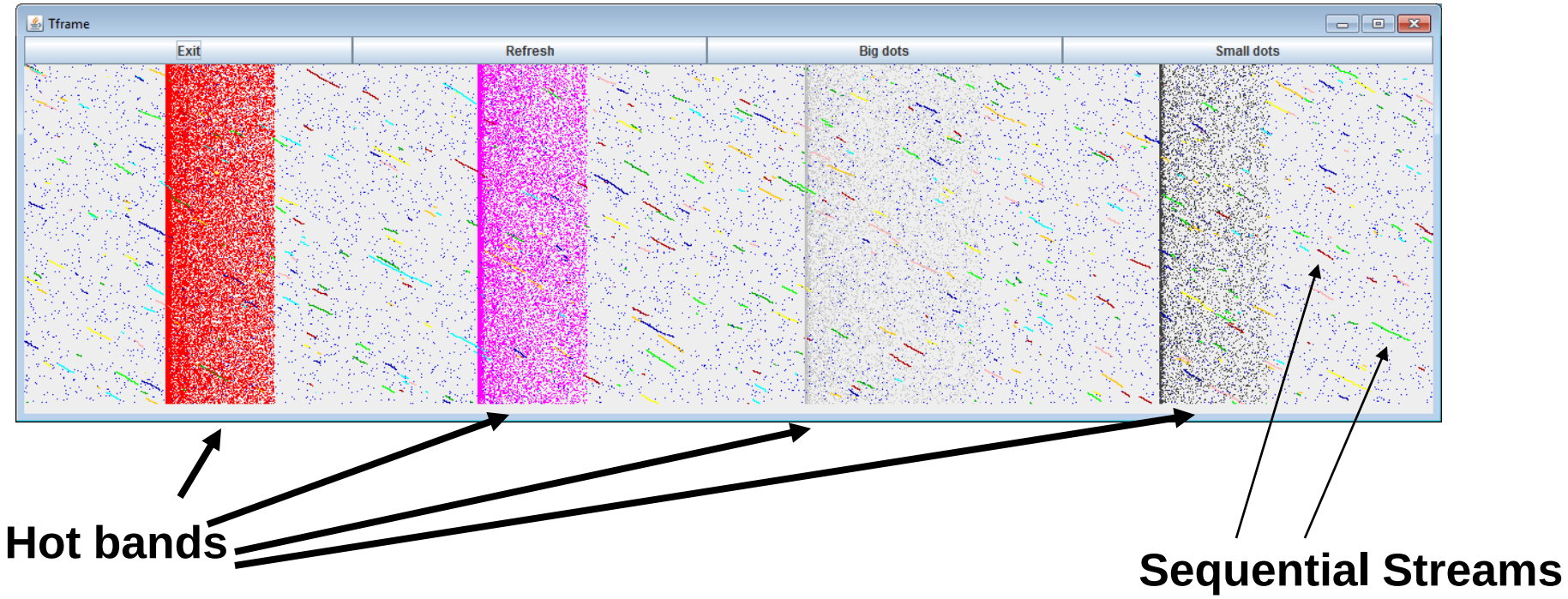
- RAID – Redundant Array Independent (Inexpensive) Disks
- RAID 0 – No Redundancy – maybe striped across many drives (rarely used)
- RAID 1 – Also know as mirroring. Data is mirrored to two drives
- RAID 10 – A variation of RAID 1. Will stripe across more than two drives.
- RAID 5 – A complex scheme of storing Parity blocks to recreate data if one device fails
- RAID 6 – Similar to RAID 5 except there are two parity blocks and can survive a double drive failure. Important to new SATA drive technologies where during the drive rebuild process a second failure is likely.
- Bottleneck – a term used to discuss what is holding the system back from performing better. Bottlenecks can be in Processors, HBAs, Controllers or Disk drives.

- Problem
  - Uniform access across storage does not address performance aspect of cache in customer-like environments
  - Customer environments do not have uniform access across storage unless they are doing a full back up
- End of the day, goal is to create a fairly complex workload to measure the power consumption – not a pure performance benchmark

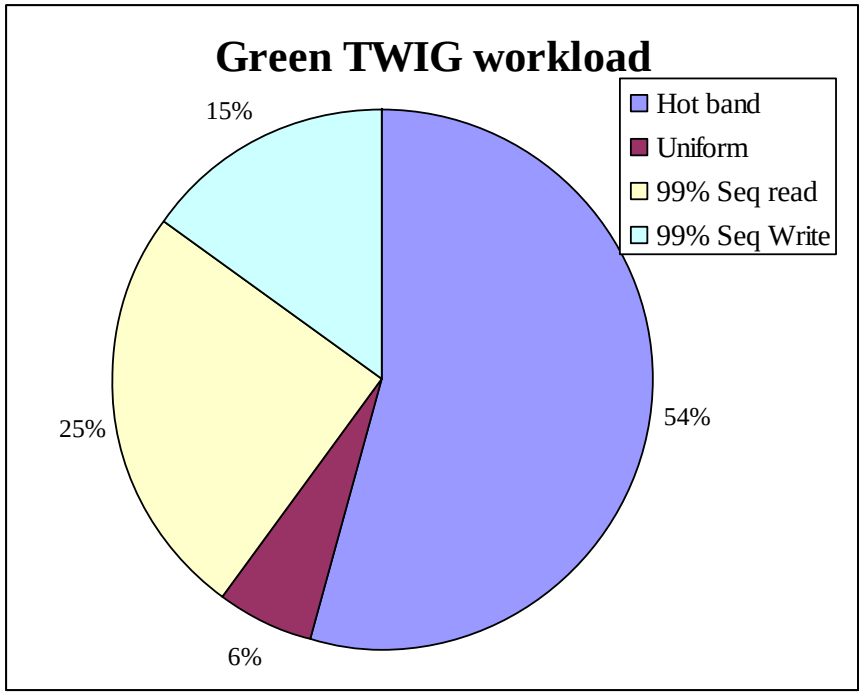
- First part of the process is the configure your storage.
- Determine your optimal configuration (Raid 1, Raid 5, ??, Stripe size, Volume Manager settings, etc)




Tune shape of curve to match customer
Cache hit rate
50%
25%
15%
10%
Probability of access
Probability of access
Probability of access
Probability of access
Address space

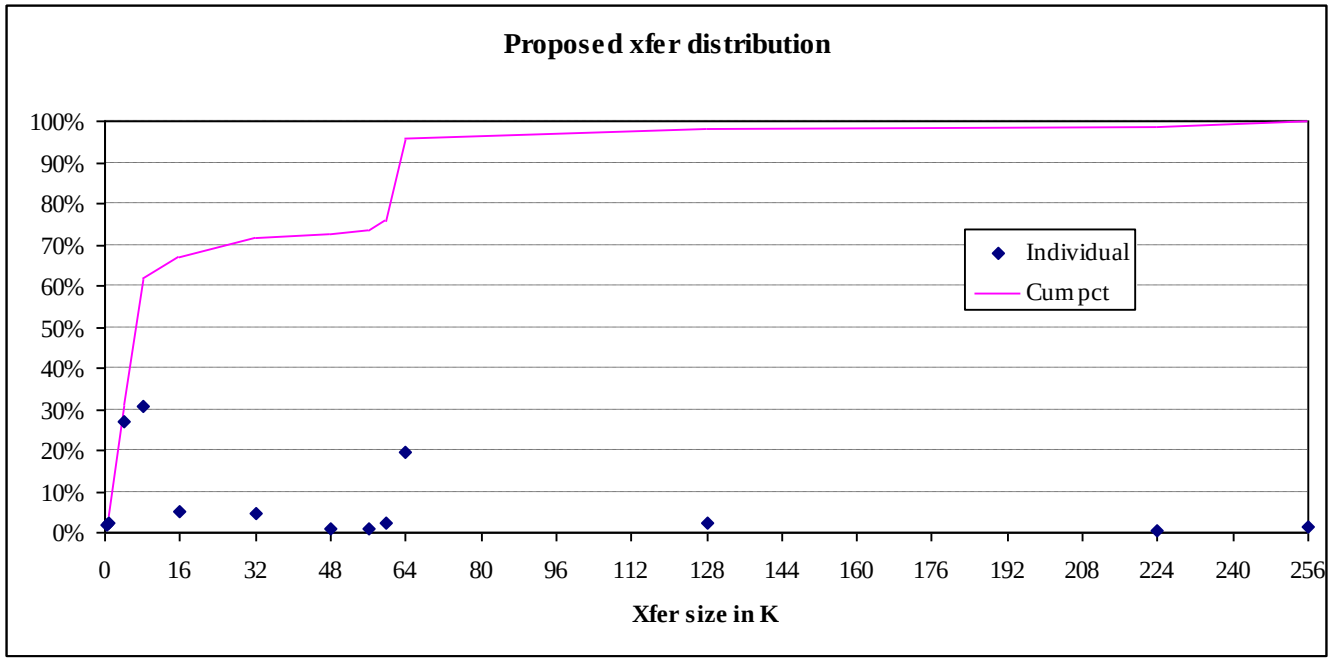

Tframe
Exit
Refresh
Big dots
Small dots
Hot bands
Sequential Streams

- Broad mix of IOS:
  - Workload type % workload
  - Hot band 54%
  - Random 6%
  - 99% Seq read 25%
  - 99% Seq Write 15%
- Over all workload 66% read



K	Individual	Cum pct
0.5	2%	2%
1	2%	4%
4	27%	31%
8	31%	62%
16	5%	67%
32	5%	72%
48	1%	73%
56	1%	74%
60	2%	76%
64	20%	96%
128	2%	98%
224	0%	98%
256	2%	100%

Average = ~27K



compratio=2 # Compression Ratio 2:1

sd=sd1,lun=e:\junk\hotband1,size=500m # Define Storage definition (file for testing purposes)

```
###################
wd=default,xfersize=(8k,31,4K,27,64K,20,16K,5,32K,5,128K,2,1K,2,60K,2,512,2,256K,2,48K,1,56K,1),rdpct=70,th=1
wd=HOTwd_uniform,skew=6,sd=sd*,seekpct=rand,rdpct=50
wd=HOTwd_hot1,sd=sd*,skew=28,seekpct=rand,hotband=(10,18)
wd=HOTwd_99rseq1,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=100
wd=HOTwd_99rseq2,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=100
wd=HOTwd_99rseq3,sd=sd*,skew=5,xfersize=(8k),seekpct=1,rdpct=100
wd=HOTwd_99rseq4,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=100
wd=HOTwd_99rseq5,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=100
wd=HOTwd_hot2,sd=sd*,skew=14,seekpct=rand,hotband=(32,40)
wd=HOTwd_hot3,sd=sd*,skew=7,seekpct=rand,hotband=(55,68)
wd=HOTwd_hot4,sd=sd*,skew=5,seekpct=rand,hotband=(80,88)
wd=HOTwd_99wseq1,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=0
wd=HOTwd_99wseq2,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=0
wd=HOTwd_99wseq3,sd=sd*,skew=5,xfersize=(8k,33,4K,29,64K,22,16K,6,32K,5,128K,3,256K,2),seekpct=1,rdpct=0

## Fill storage workload
wd=wd_fill,sd=sd*,seekpct=eof

## 4 Corners workload
wd=wd_mixed,sd=sd*,seekpct=rand
## zz defines the number of streams across the concatenated space
##  The number of threads for the sequential workload must be a multiple of aa
wd=wd_seq,sd=sd*,seekpct=0,streams=aa
rd=rd1_hband,wd=HOTwd*,iorate=MAX,warmup=30,elapsed=6H,interval=10,pause=30,th=XX
```

Hot Band Workloads and High End HP Storage Products

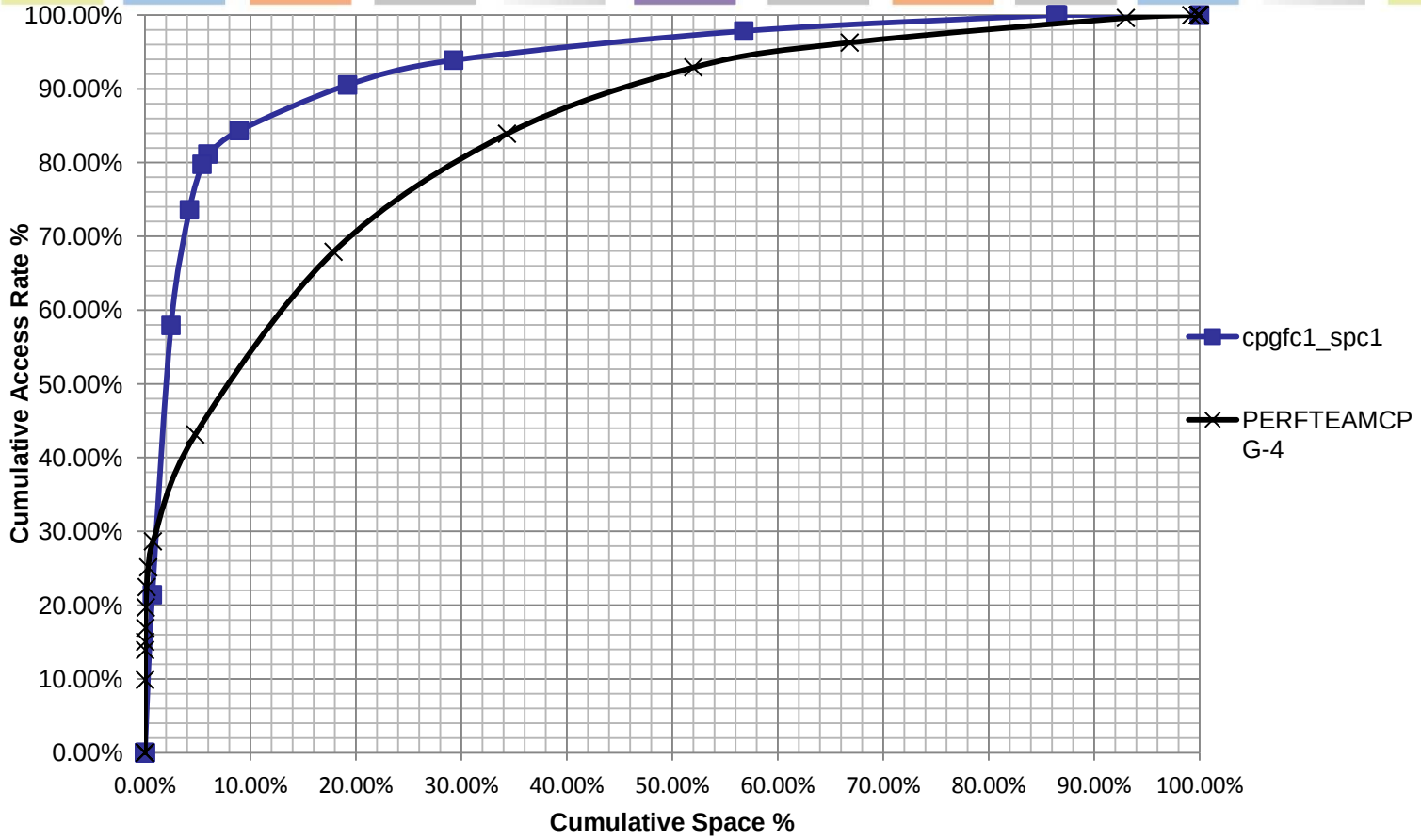
Chuck Paridon, HPES Master Storage Performance Architect

SNIA Emerald™ Training

SNIA Emerald Power Efficiency Measurement Specification, for use in EPA ENERGY STAR®

June 24-27, 2013

SNIA Green TWG Cache Friendly Performance Comparison Chart
Tiered Storage Speedup on 3PAR Storage Products



Source: HP Blue line is the SPC-1 workload IO density data. Black is Hot Band. The steeper the curve and the sharper the knee, the more Tiered Storage can enhance the IO rate. SPC speedup ~50% Hot Band ~30% So total speedup due to both cache and tiering ~ 100 + (40 *0.3) = 112%

Hot Band IO Patterns

- The Hot Band workload is comprised of several different IO streams, some of which contain hot spots, or regions of more intense IO demand.
- This results in varying degrees of logical block address re-referencing in certain regions of the overall IO space that can be either contained with the cache of an array, or can be placed on storage devices that deliver a higher IO rate.

Hot Band IO Patterns

Hot Band IO Profile

IO Profile	% of workload	Read/Write Percentage	IO Size (KiB)	Access Pattern	Usable Address Range
Write Stream 1	5	0/100	See Table 12	Sequential	0-100%
Write Stream 2	5	0/100	See Table 12	Sequential	0-100%
Write Stream 3	5	0/100	See Table 12	Sequential	0-100%
Read Stream 1	5	100/0	See Table 12	Sequential	0-100%
Read Stream 2	5	100/0	See Table 12	Sequential	0-100%
Read Stream 3	5	100/0	See Table 12	Sequential	0-100%
Read Stream 4	5	100/0	See Table 12	Sequential	0-100%
Read Stream 5	5	100/0	See Table 12	Sequential	0-100%
Uniform Random	6	50/50	See Table 12	Random	0-100%
Hot Band 1	28	70/30	See Table 12	Random	10 -18%
Hot Band 2	14	70/30	See Table 12	Random	32-40 %
Hot Band 3	7	70/30	See Table 12	Random	55-63 %
Hot Band 4	5	70/30	See Table 12	Random	80-88 %

Hot Bands concentrate 54% of the IO in 32% of the space

Hot Band IO Patterns

- The Hot Band workload when run on High End Storage demonstrates the power/performance advantage of two product features
  - Array Based Cache
  - Storage Tiering
- Although the initial goal was solely cache focused, there is also a benefit of implementing faster tiers in the product such as HP P9500 Smart Tiering

SNIA Green TWG Cache Friendly Performance Comparison Chart Cache and Tiering Speedup on HP P9500 Storage Product

	Hot IOPS	Hot RT	Rnd IOPS	Rnd RT	C/WS ratio	Cache Hit Ratio Hot Band vs Rnd
Small Array	4,330	32.8 ms	4,130	33.4 ms	<<	N/A
Large Array Cache Assist only	39,900	8.97 m	18,410	22.59 ms	~3%	60%/24%
Large Array (Tiered)	42,870	5.77 ms	N/A	N/A	~3%	60%/24%

- The Workload Analysis Process Consists of Two Steps

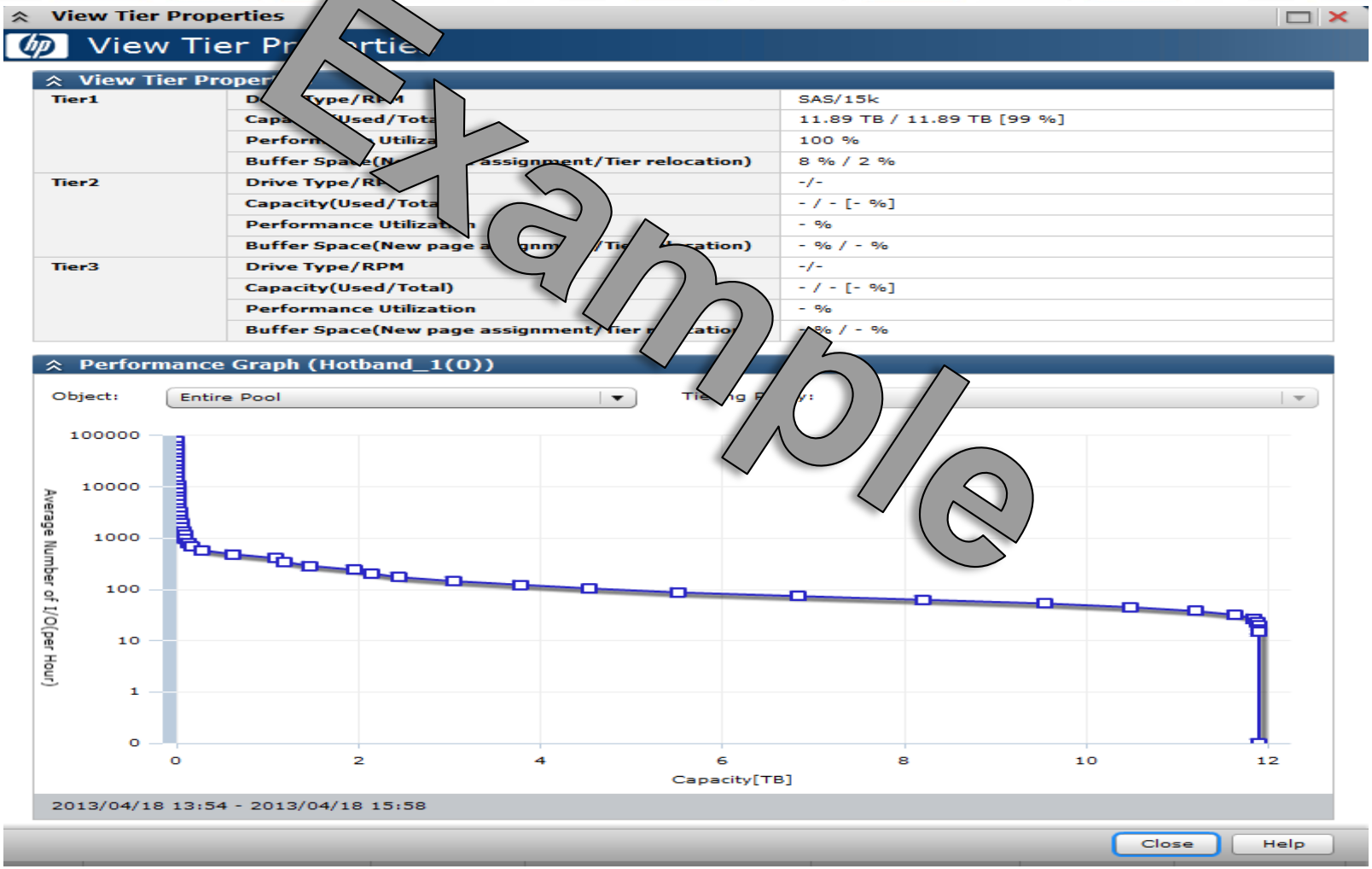
1. Create a single pool large enough to hold the desired working set
2. Run the Hot Band workload on that pool and use the analysis tools to produce a report guiding the composition of the tier(s)

- The tier construction process involves 2 decision points.

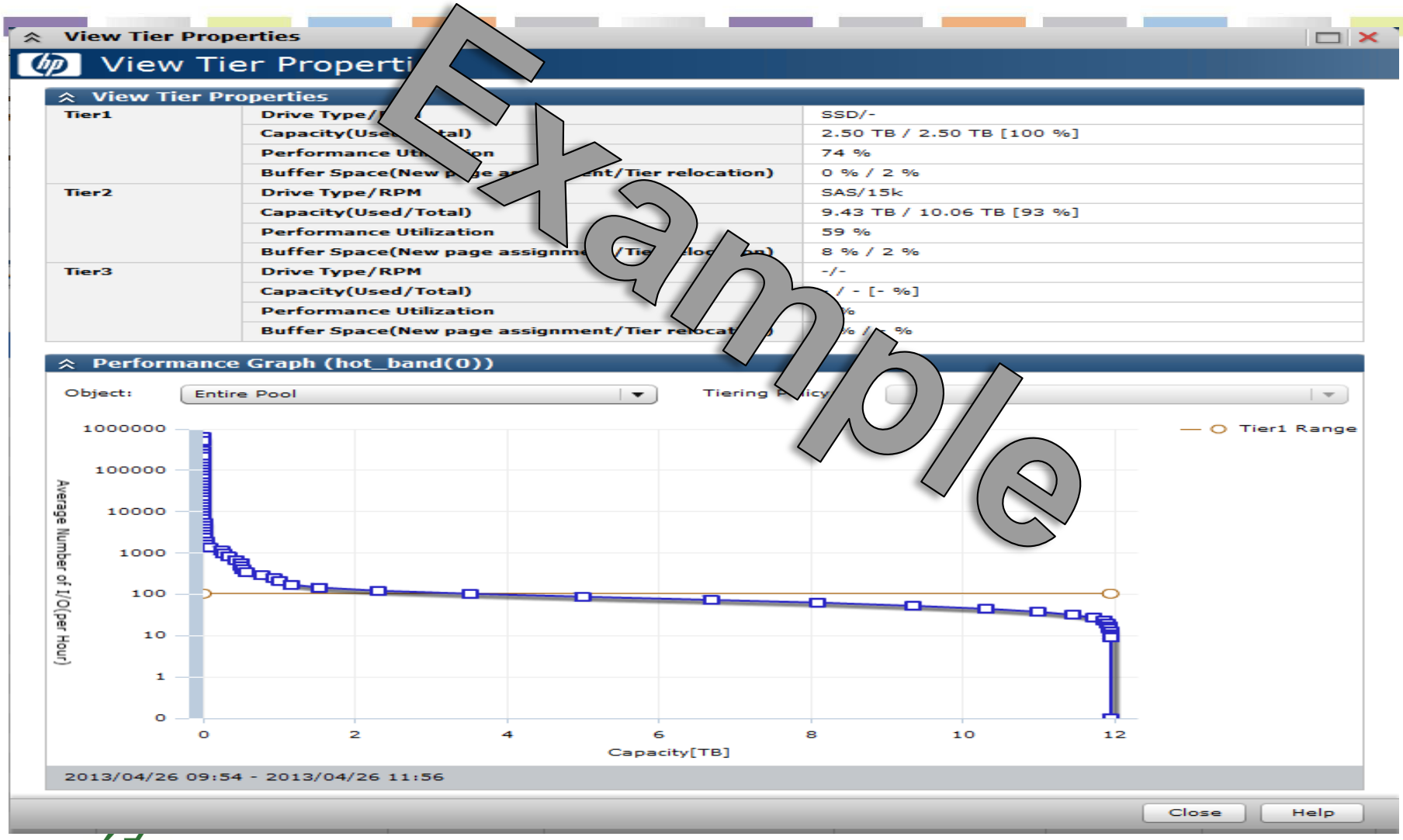
1. Which technology to deploy
2. The capacities of the tier(s)

- The following slides illustrate an example of these activities

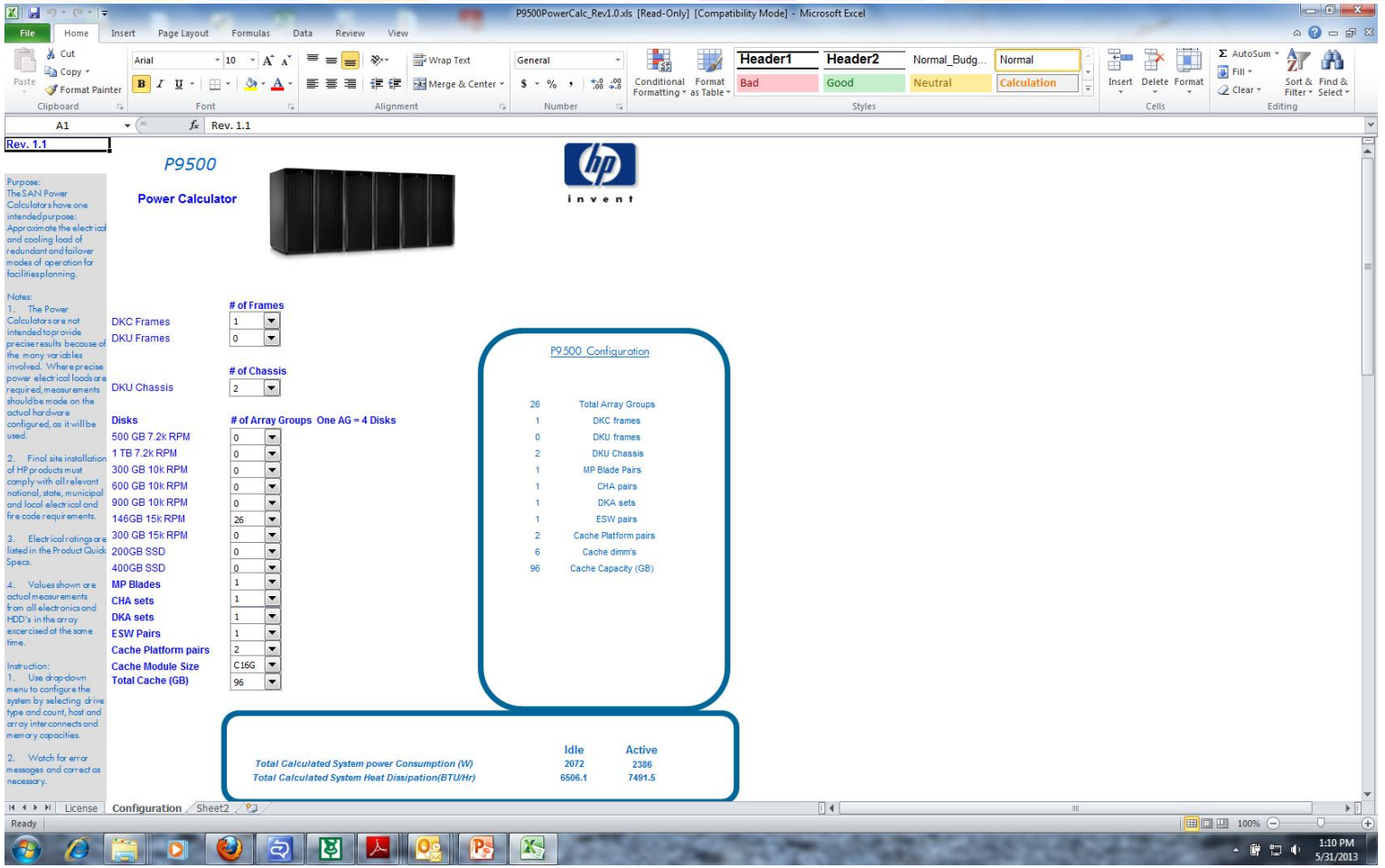
SNIA Green TWG Tiered Storage w/Hot Bands
Initial Tier Property Analysis



SNIA Green TWG Tiered Storage w/Hot Bands Final Tier Property Analysis



The HP P9500 Power Calculator



SNIA Green TWG Tiered Storage w/Hot Bands Primary Metric Comparison

Configuration	Tier Type	Power Consumption (Modeled)	IOPS	IOPS/Watt
Large Array (Initial)	15k RPM	7,491 Watts	18,410	2.457
Large Array (Cache Assist)	15k RPM	7,491 Watts	39,900	5.326
Large Array (Tiered)	[illegible]	7,283 Watts	42,870	5.886

SNIA Green TWG Hot Band Workload Conclusions and Observations

- The Hot Band workload is amenable to performance optimization by both storage subsystem cache and the proper deployment of tiered storage.
- As a result of high cache hit rates (~60%) the overall performance contribution of tiered storage is limited.
- In addition to the increase in IO rate (132%), there is also a corresponding decrease in power consumption from the substitution of SSDs in the configuration
- The net effect of these two parameter changes is a 140% improvement in the SNIA primary active metric (IOPS/Watt) of