SNIA VDBENCH Emerald Overview

Steven Johnson

SNIA Emerald™ Training

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

June 24-27, 2013

SNIA™ Green Storage Initiative

SNIA Emerald™

Agenda

- Introduction to SNIA Emerald
- Phases test
  - Pre-fill
  - Warm up
  - Optionally time for tiering to move things around
  - Hot band test
  - 4 corners test
    - Small block random read and write
    - Large block sequential read and write

Quick overview of performance terms

- 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.

Performance terms (cont)

- 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) – Storage device with no moving parts. A disk drive whose storage capability is provided by solid state storage

Performance terms (cont)

- 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.

Determining target storage

- First part of the process is the configure your storage.
- Determine your optimal configuration (Raid 1, Raid 5, ??, Stripe size, Volume Manager settings, etc)
- Determine the amount of total capacity
- Export 56+% of the storage or all of it and let vdbench execute against 56+% of capacity (sd parameter “range=(0,57)”
- List all the luns for the test and put it in a file. Use format or Disk Manager (OS Dependent) to identify the LUNs

Filling storage

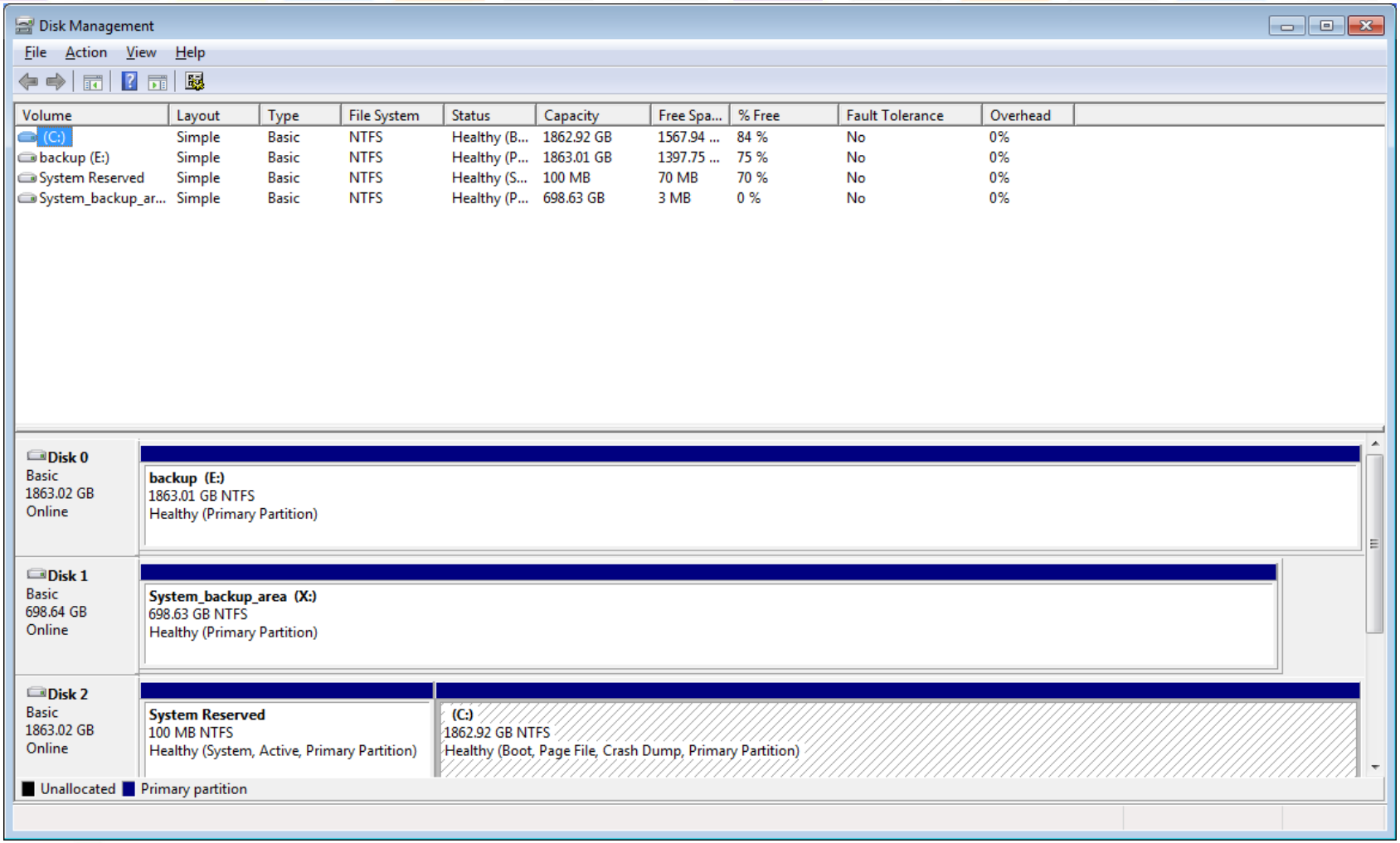
- For the 56+% capacity, all of it needs to be initialized by vdbench
- Vdbench will create a near random pattern that has a characteristic of 2:1 compression
- Have vdbench write from beginning to end to have a known pattern across all active blocks of storage (part of the example script provided)
- No power measurements or performance will be reported on this phase

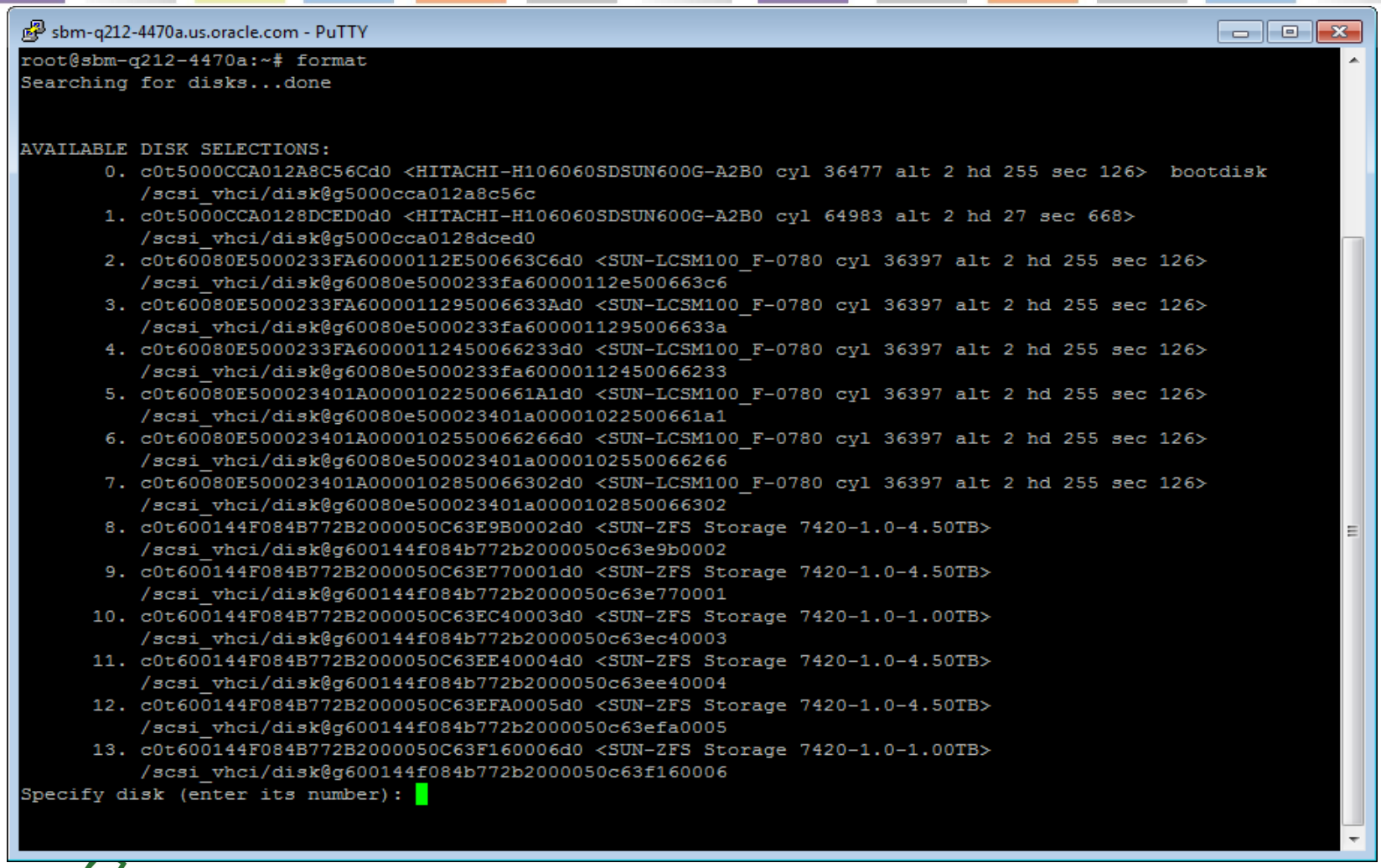
Warm-up storage 6+ hours

- The warm-up phase designed to have the following effect on storage:
  - Get the system to a relatively steady state. No out of the box performance for any component.
  - Cause some amount of fragmentation in systems that do “copy on write”
  - Let intelligent tiering systems study access pattern for moving blocks around
  - Optionally slow period that allows tiering systems have cycles to move data around
- Hopefully system will approximate a customer system after weeks of activity

Hot Band run

- Hot Band workload
  - Hot band is a cache friendly workload
  - Has very hot spots that tend to stay in cache
  - More cache the more cache hit the storage should produce
  - 8 sequential streams, 5 read, 3 write make up 40% of workload
- Intended to simulate a transactional workload


Disk Management
File Action View Help
Volume Layout Type File System Status Capacity Free Spa... % Free Fault Tolerance Overhead
(C:) Simple Basic NTFS Healthy (B... 1862.92 GB 1567.94 ... 84 % No 0%
backup (E:) Simple Basic NTFS Healthy (P... 1863.01 GB 1397.75 ... 75 % No 0%
System Reserved Simple Basic NTFS Healthy (S... 100 MB 70 MB 70 % No 0%
System_backup_ar... Simple Basic NTFS Healthy (P... 698.63 GB 3 MB 0 % No 0%
Disk 0
Basic
1863.02 GB
Online
backup (E:)
1863.01 GB NTFS
Healthy (Primary Partition)
Disk 1
Basic
698.64 GB
Online
System_backup_area (X:)
698.63 GB NTFS
Healthy (Primary Partition)
Disk 2
Basic
1863.02 GB
Online
System Reserved
100 MB NTFS
Healthy (System, Active, Primary Partition)
(C:)
1862.92 GB NTFS
Healthy (Boot, Page File, Crash Dump, Primary Partition)
Unallocated
Primary partition


sbm-q212-4470a.us.oracle.com - PuTTY
root@sbm-q212-4470a:~# format
Searching for disks...done
AVAILABLE DISK SELECTIONS:
0. c0t5000CCA012A8C56Cd0 <HITACHI-H106060SDSUN600G-A2B0 cyl 36477 alt 2 hd 255 sec 126> bootdisk
/scsi_vhci/disk@g5000cca012a8c56c
1. c0t5000CCA0128DCED0d0 <HITACHI-H106060SDSUN600G-A2B0 cyl 64983 alt 2 hd 27 sec 668>
/scsi_vhci/disk@g5000cca0128dced0
2. c0t60080E5000233FA60000112E500663C6d0 <SUN-LCSM100_F-0780 cyl 36397 alt 2 hd 255 sec 126>
/scsi_vhci/disk@g60080e5000233fa60000112e500663c6
3. c0t60080E5000233FA6000011295006633Ad0 <SUN-LCSM100_F-0780 cyl 36397 alt 2 hd 255 sec 126>
/scsi_vhci/disk@g60080e5000233fa6000011295006633a
4. c0t60080E5000233FA60000112450066233d0 <SUN-LCSM100_F-0780 cyl 36397 alt 2 hd 255 sec 126>
/scsi_vhci/disk@g60080e5000233fa60000112450066233
5. c0t60080E500023401A00001022500661A1d0 <SUN-LCSM100_F-0780 cyl 36397 alt 2 hd 255 sec 126>
/scsi_vhci/disk@g60080e500023401a00001022500661a1
6. c0t60080E500023401A0000102550066266d0 <SUN-LCSM100_F-0780 cyl 36397 alt 2 hd 255 sec 126>
/scsi_vhci/disk@g60080e500023401a0000102550066266
7. c0t60080E500023401A0000102850066302d0 <SUN-LCSM100_F-0780 cyl 36397 alt 2 hd 255 sec 126>
/scsi_vhci/disk@g60080e500023401a0000102850066302
8. c0t600144F084B772B2000050C63E9B0002d0 <SUN-ZFS Storage 7420-1.0-4.50TB>
/scsi_vhci/disk@g600144f084b772b2000050c63e9b0002
9. c0t600144F084B772B2000050C63E770001d0 <SUN-ZFS Storage 7420-1.0-4.50TB>
/scsi_vhci/disk@g600144f084b772b2000050c63e770001
10. c0t600144F084B772B2000050C63EC40003d0 <SUN-ZFS Storage 7420-1.0-1.00TB>
/scsi_vhci/disk@g600144f084b772b2000050c63ec40003
11. c0t600144F084B772B2000050C63EE40004d0 <SUN-ZFS Storage 7420-1.0-4.50TB>
/scsi_vhci/disk@g600144f084b772b2000050c63ee40004
12. c0t600144F084B772B2000050C63EFA0005d0 <SUN-ZFS Storage 7420-1.0-4.50TB>
/scsi_vhci/disk@g600144f084b772b2000050c63efa0005
13. c0t600144F084B772B2000050C63F160006d0 <SUN-ZFS Storage 7420-1.0-1.00TB>
/scsi_vhci/disk@g600144f084b772b2000050c63f160006
Specify disk (enter its number):

Red Hat parted

- Linux will have various ways to find the partition table
- Need to figure out your version
- Red Hat has the “parted” command and print will display the drives and partitions on each drive
- See details
  https://access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/s1-disk-storage-parted.html#s2-disk-storage-parted-view-part-table

4 Corners 80 minutes

- Tests the extremes of workloads
  - Random Read and Write 8K transfer
  - Sequential Read and Write 256K transfer

Typical

Phase	Fill	Hot Band Warm-up	4 Corners
Duration	MB capacity/MB rate/3600	~6Hr	80 Minutes

Tiering

Phase	Fill	Hot Band Warmup/learning	Hot Band Semi-idle Move blocks	Hot Band	4 Corners
Duration		~6Hr	TBD	TBD	80 Minutes

Steven.A.Johnson@oracle.com