SDC
STORAGE DEVELOPER CONFERENCE
SNIA SANTA CLARA, 2016

Analysis of SSD Health & Prediction of SSD Life

Dr. M. K. Jibbe
Technical Director
NetApp, Inc

Bernard Chan
Senior Engineer
NetApp, Inc

Agenda

- Problem Statement
- SMART
- Endurance Reporting
- Wear Life Prediction
- Recovery from Common SSD Errors
- Drive Data Migration
- Summary

Problem Statement

- As a storage admin, how can I minimize the impact of SSDs failing?
  - pro-active in handling data in the SSD before it fails?
    - warning when SSD about to wear out
    - copy data off to another SSD
    - fail non-performing SSD
  - maintain data access efficiency & availability

What is S.M.A.R.T.?

Self-Monitoring Analysis and Reporting Technology

- Disk Drive's feature to provide various monitoring indicators of disk reliability
- Intent is to enable the anticipation of hardware failures so data can be copied off to another device prior to drive failure

What is S.M.A.R.T.?

- SSD Program Fail Count (171)
- SSD Erase Fail Count (172)
- SSD Wear Leveling Count (173)
- Erase Fail Count (176)
- Wear Range Delta (177)
- SSD Life Left (231)
- Available Reserved Space (232)
- Media Wearout Indicator (233)

Solid State Media log page (11h)

Table 281 Solid State Media log page

Bit / Byte	7	6	5	4	3	2	1	0
0	DS	SPF (0b)	PAGE CODE (11h)					
1	SUBPAGE CODE(00h)							
2	(MSB)	PAGE LENGTH (n - 3)						
3								(LSB)
SOLID STATE MEDIA LOG PARAMETERS (see table 280)								
4	SOLID STATE MEDIA PARAMETER (first							
. . .	. . .							
n	SOLID STATE MEDIA PARAMETER (last)							

Solid State Media log page (11h)

Table 282 Percentage Used Endurance Indicator log parameter format

Bit / Byte	7	6	5	4	3	2	1	0
0	(MSB)							
1				PARAMETER CODE (0001h)				(LSB)
2	**PARAMETER CONTROL BYTE – binary format list log parameter (see 4.2.2.2.2.5)**							
	DU	Obsolete	TSD	ETC	TMC		FORMAT AND LINKING	
3	PARAMETER LENGTH (04h)							
4								
. . .	RESERVED							
6								
7	PERCENTAGE USED ENDURANCE INDICATOR							

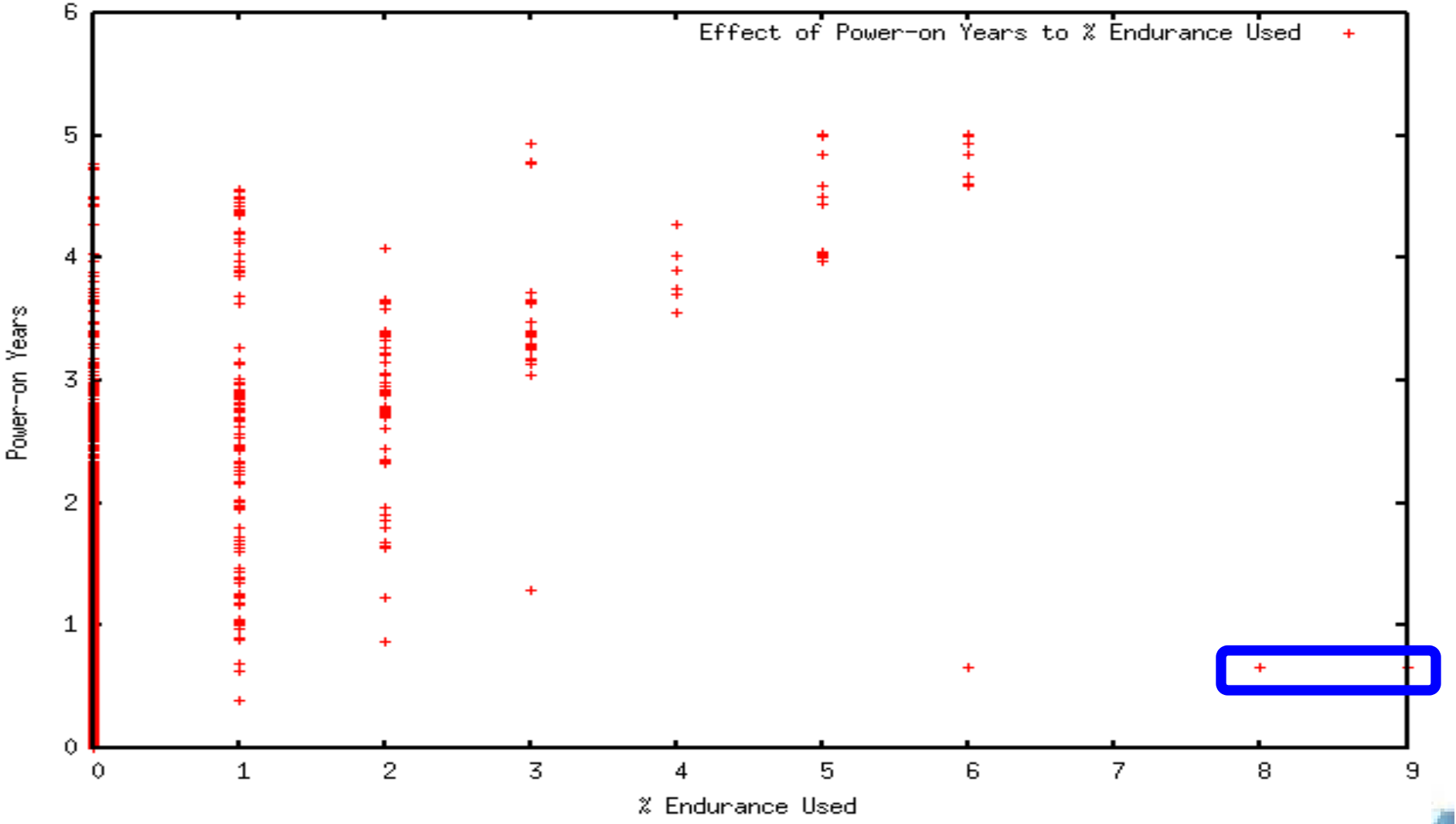
PERCENTAGE USED ENDURANCE INDICATOR field

The PERCENTAGE USED ENDURANCE INDICATOR field indicates an estimate of the percentage of device life that has been used. The value in the field shall be set to zero at the time of manufacture. A value of 100 indicates that the estimated endurance of the device has been consumed, but may not indicate a device failure (e.g., minimum power-off data retention capability reached for devices using flash technology). The value is allowed to exceed 100. Values greater than 254 shall be reported as 255. The device server shall update the value at least once per power-on hour.

Endurance Reporting – NetApp Unique

- The AVERAGE BLOCKS ERASED field indicates the number of blocks erased per die averaged over all of the dies of the drive.
- The SPARE BLOCKS REMAINING PERCENTAGE field indicates the percentage of total spare blocks that remain for the device and is tracked over the life of the device.
- The ENDURANCE REMAINING PERCENTAGE field indicates an estimate of the percentage of device life that has been used.

Effect of Power-on Years to % Endurance Used



High Worn SSD – Commodity Trader

- Powered on for ~240 days
- E2760 with 37TB storage
- Vendor A 400G for SSD Cache
- 10 DWPD eMLC NAND
  - 1 – 8% (1.7PB written / 114TB read)
  - 1 – 9% (1.7PB written / 114TB read)
- Prediction: 5 years = 69% (9% in 240 days)

High Worn SSD – Commodity Trader

- Data ingest per day (Vendor A 400G 10DWPD)
  - 100% = (400G x 10 DWPD x 365 days x 5 years)
  - 1% = ~70TB
  - 9% = 630TB in 240 days = 2.6TB / day
- Vendor B 3.8TB (1 DWPD) vs 3.2TB (3 DWPD)
  - 2.6 TB/day / 3.8TB: 68% worn at end of 5 years
  - 2.6 TB/day / 9.6TB: 27% worn at end of 5 years
- Vendor 15.3TB (1 DWPD)
  - 2.6 TB/day / 15.3TB: 17% worn at end of 5 years

How Does NetApp Predict SSD Health?

- Recovery from Common SSD Errors
  - Power cycle the SSD
  - Check the thresholds;
    - Migrate data if necessary and there is a spare drive
    - Mark Impending drive failure and continue to use the drive if there is no spare

Drive Data Migration Feature

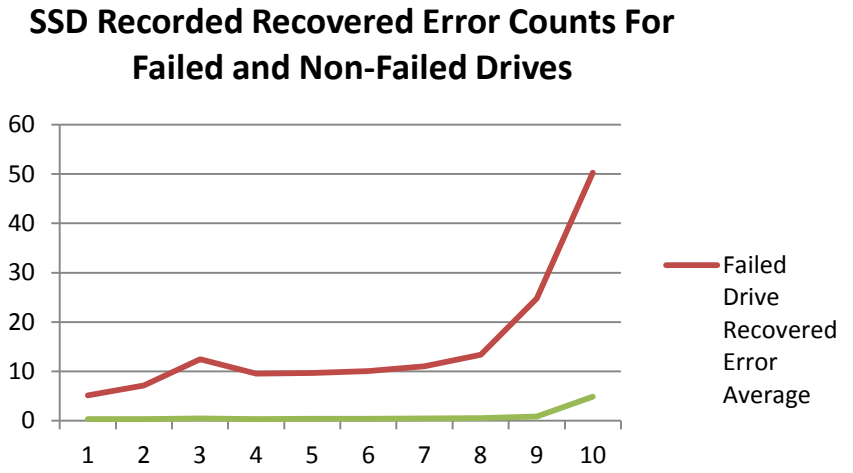
- Purpose: have a mechanism to copy data from an impending failure drive in lieu of fail/reconstruct
  - Faster to copy than reconstruct
  - Less impact on foreground I/Os

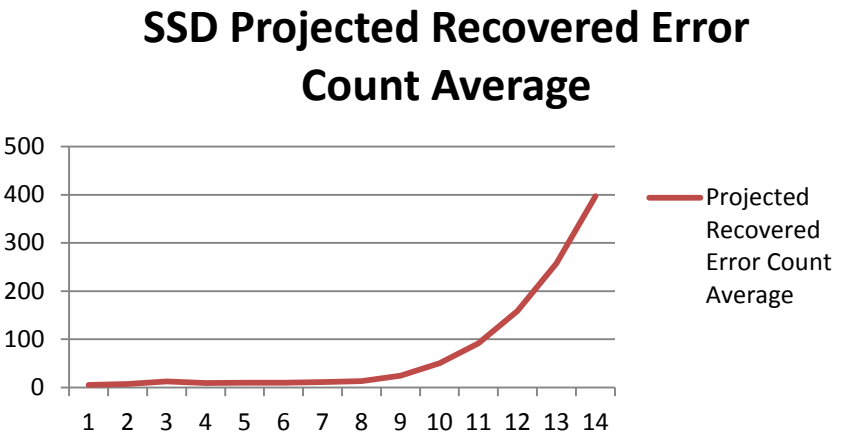
Drive Data Migration Feature

- Triggers on SSD's
  - Predicted Failure Analysis (PFA)
  - Synthesized PFA (SPFA) – error exceeding threshold in a time window
    - Media Error (5 per day)
    - Recovered Error(150 per day)
    - Hardware Error(2 per day)
    - Fast I/O Timeout (8 per day)
    - Stagnant I/O (30 count)

Drive Data Migration “cont’d”

- Drive Data Migration thresholds was based upon analysis of failures in populations of 250k Drives from three different vendors
  - HDD Nearline drives,
  - HDD Enterprise drive,
  - SSD drive from two different vendors with capacity ranging from 200GB – 8TB





The Normal distribution probabilities of the failed drive data shows that the Recovered Error SPFA threshold is increased from 60 to 150.

Summary and Conclusions

- Two methods presented to protect array system from SSD failures
  - SSD parameters are used to predict drive life time
  - SSD error types were used to predict the usability of an SSD drive in a SAN
- Possible SSD PFA and SPFA triggers to migrate the data off a drive which is expected to fail
- Data backup can be on site or to a cloud