SDC
STORAGE DEVELOPER CONFERENCE
SNIA ■ SANTA CLARA, 2015

Right SSD for the Right Array

M. K. Jibbe, PhD
HSG NetApp
Technical Director

Bernard Chan
HSG NetApp
Sr. QA Arch

Brian Mckean
HSG NetApp
Sr. SW Arch

Date: 09/01/2015

Abstract

- Given the explosive growth in SSD adoption it is no surprise that there has been huge advancement in SSD technology such as the invention of vertical stacking of cells which created 3D SSDs and multi-level cells chips as well as a multitude of software techniques to address the inherent challenges of NAND (cell degradation, wear leveling, etc.). SSD per drive capacities are increasing dramatically. It is replacing traditional HDD even at the capacity layer of the storage stack to store warm data. Taking this all into account, it is important for us to choose the right SSD for the right storage array product to guarantee that we are delivering the most compelling Flash storage at the most attractive price and performance without sacrificing overall product quality.

SDC 15

Outline

- SSD cell structure (3D, eMLC, tLC, etc.)
- Storage array product RAID controller performance
  - I/O profile (R/W ratios, block size, etc.)
- RAID level operation amplification
- Deployed storage array historical field data
- SSD MTBF and DWPD
  - Endurance Metrics for the last 2 ½ years(Erase Count and Percent Of Blocks Remaining)
  - Expected Life cycle of different SSD drive types

SDC 15

Approach – SSD Cell Structure

- SLC – highest reliability but too $$$
  - Single cell to store one bit of data
  - Faster and much more reliable
- MLC – reliable & $$
  - Single cell to store two bits of data
  - Lower in price
  - Higher wear rates
  - Lower write performance
- cMLC – reliable enough & cheaper than MLC
- TLC – less reliable but $ – consumer grade

Approach – Array & SSD Performance

- Run industry benchmarking test on new SSDs in a storage array
- Compare performance results with previous drive models, and with other drive vendors
  - Latency spikes
  - Read vs. Write
  - IO sizes
  - IO Randomness
- Expose any Array limitation with newer & faster SSDs

SSD Model Y – FW Z3 relative to Z2

- Rates shown are the peak IO rates measured for all queue depths tested
- FW Z3 peak IO rates are within 1% of FW Z2
- Trending to <1% below MSB2

Vendor X SSD Model Y FW Z3			Max IOP's	
	Seq Wr	Ran Wr	Seq Rd	Ran Rd
512	120923	67853	99115	217632
1K	134158	67795	162114	217371
2K	131428	68207	206399	218491
4K	127644	71108	204100	204226
8K	94807	95209	115672	116300
16K	47609	47877	58344	58325
32K	23981	23992	30533	30523
64K	12014	12015	15472	15403
128K	6015	6014	7804	7799
256K	3009	3011	3930	3927
512K	1508	1506	1966	1965

Vendor X SSD Model Y FW Z2			Max IOP's	
	Seq Wr	Ran Wr	Seq Rd	Ran Rd
512	121486	67847	98866	218150
1K	134030	67840	163128	218068
2K	133741	68216	207568	219243
4K	129666	71063	203449	203852
8K	92966	94743	119133	119671
16K	47629	47955	59948	59887
32K	23996	23994	31072	31021
64K	12040	12023	15515	15512
128K	6026	6033	7774	7770
256K	3022	3014	3915	3913
512K	1509	1507	1955	1954

Max IOP's - Vendor X SSD Model Y FW Z3 relative to Vendor X SSD Model Y FW Z2				
	Seq Wr	Ran Wr	Seq Rd	Ran Rd
512	1.00	1.00	1.00	1.00
1K	1.00	1.00	0.99	1.00
2K	0.98	1.00	0.99	1.00
4K	0.98	1.00	1.00	1.00
8K	1.02	1.00	0.97	0.97
16K	1.00	1.00	0.97	0.97
32K	1.00	1.00	0.98	0.98
64K	1.00	1.00	1.00	0.99
128K	1.00	1.00	1.00	1.00
256K	1.00	1.00	1.00	1.00
512K	1.00	1.00	1.01	1.01

E-Series SSD Technology

- Must meet endurance and performance expectations
  - Characterized by DWPD (drive writes per day)
  - Write endurance is more critical than NAND type
- NetApp® EF-Series SSD is eMLC
  - Rated at 10 DWPD (medium endurance)
  - Warrantied up to 5 years
  - Life expectance well over 5 years
- SSD operation
  - Handles all maintenance operations in the SSD
    - Garbage collection
    - Wear leveling
  - Keeps SSD ready for consistent operations
  - Multiple processing chips eliminates contention
- SSD Technology
  - Dual ported: more throughput, more resiliency
  - SSDs also have DRAM to assist in multiple concurrent operations

Approach – RAID level operation amplification

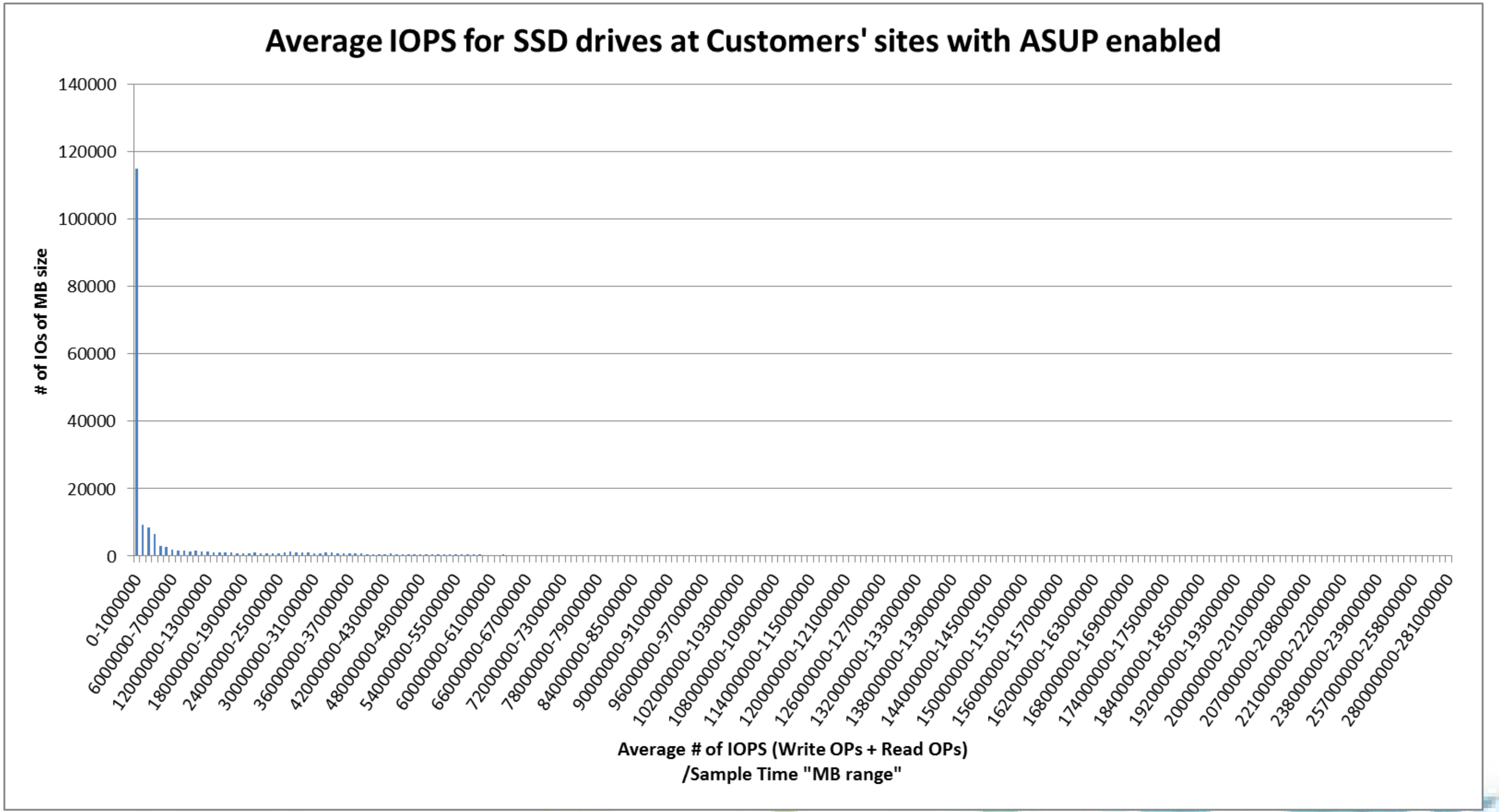
- RAID 0 – no amplification
- RAID 1 or 10 – every Write requires additional Write to mirrored SSD
- RAID 5 or 6
  - Full Stripe Write is good
  - Partial Strip Write triggers additional Read(s) for Parity calculation
  - Parity Write
  - Random small Writes needs to be

Write amplification = committed Write / host writes

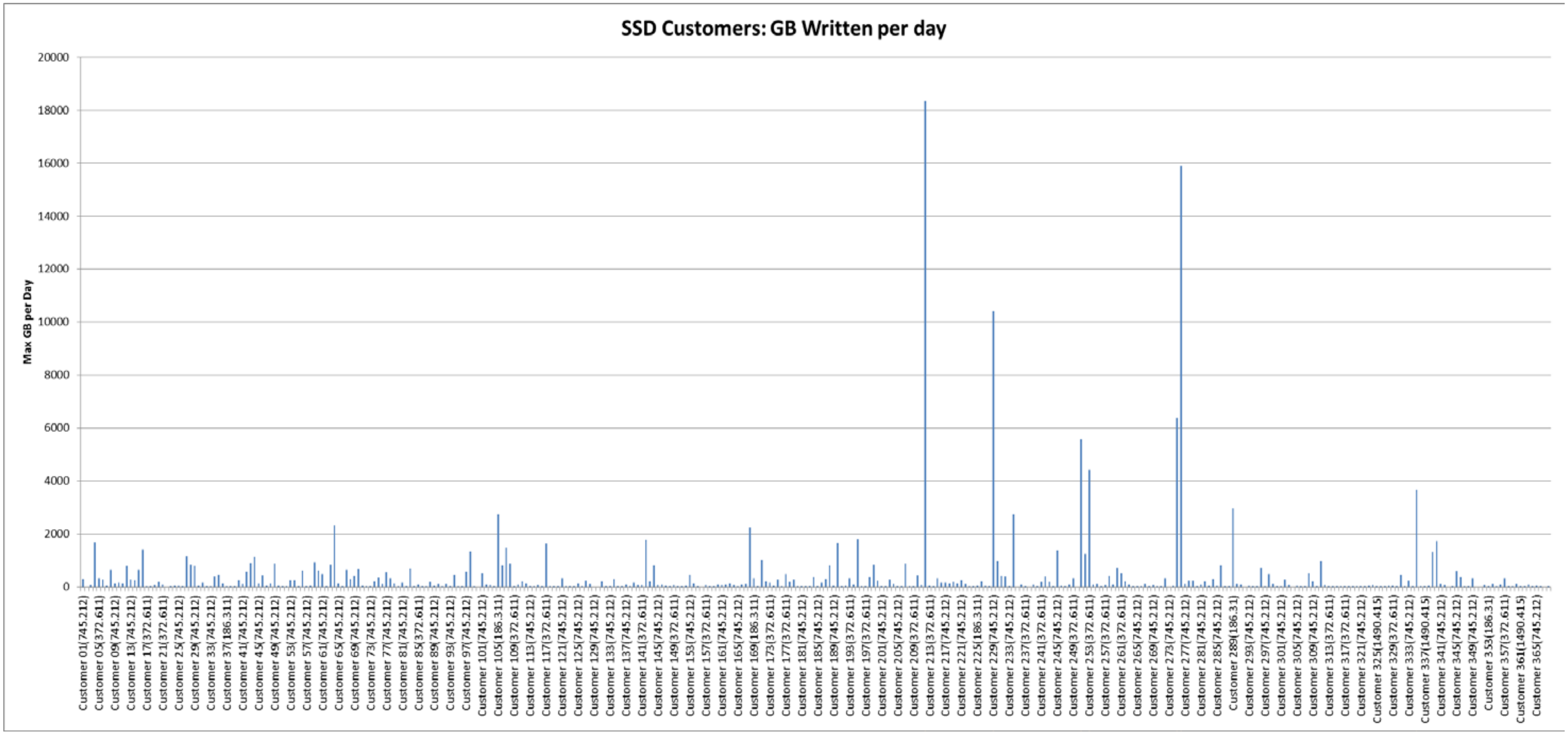
Approach – Deployed storage array historical field data

- How are customers using our Flash Array?
  - Average throughput
  - Volume Written per day

Average number of IOs per second, (WR + RD) / Sample Time (Period 8 weeks)



Statistics of GB written per day for SSD Population (Period 8 weeks)

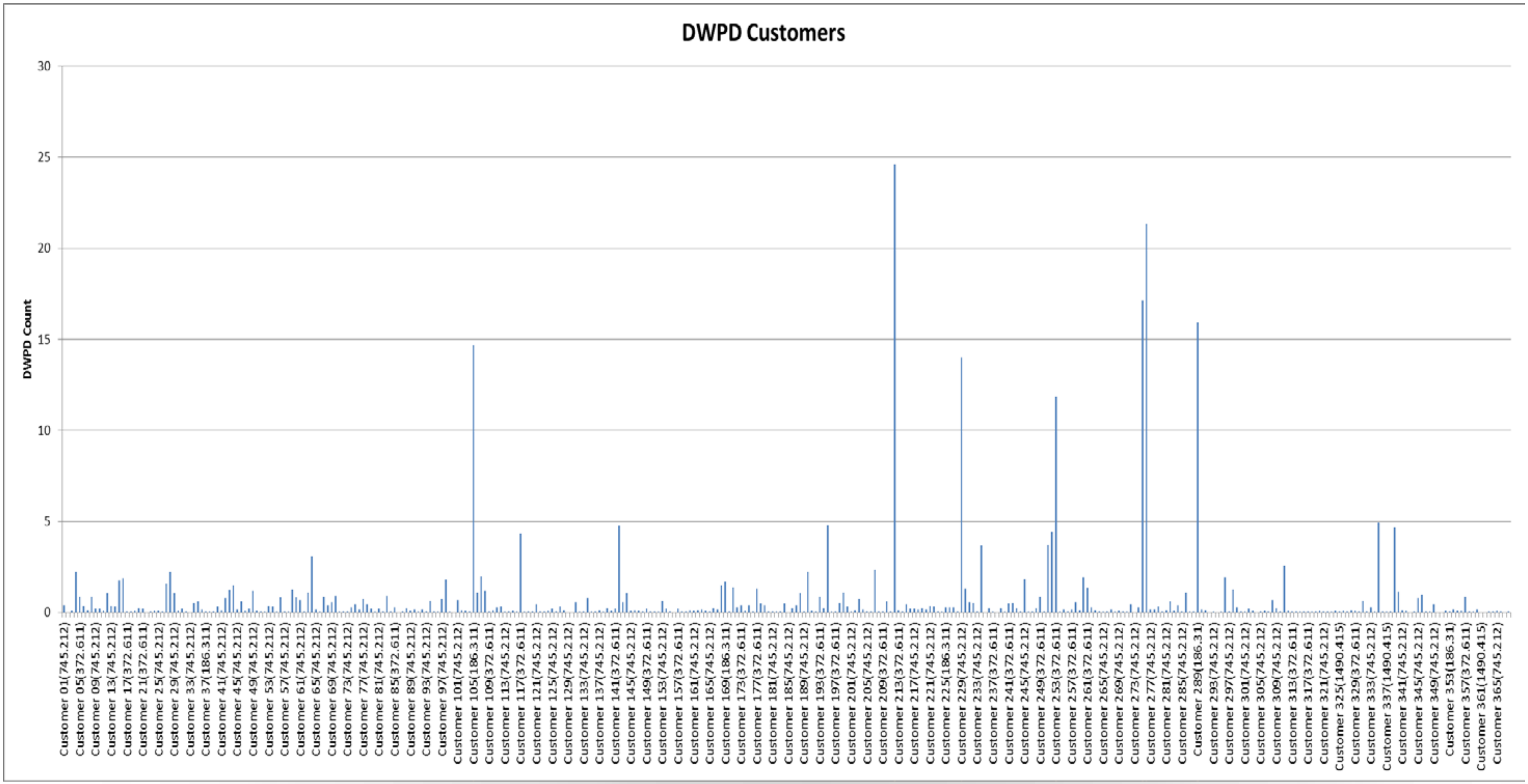


Approach – MTBF & DWPD

- MTBF = Mean Time Between Failure
  - Expected time between two failures for a repairable system

- DWPD = Drive Writes Per Day
  - Daily usage figure in terms of Writes to last the SSD a pre-determined number of years of good operation

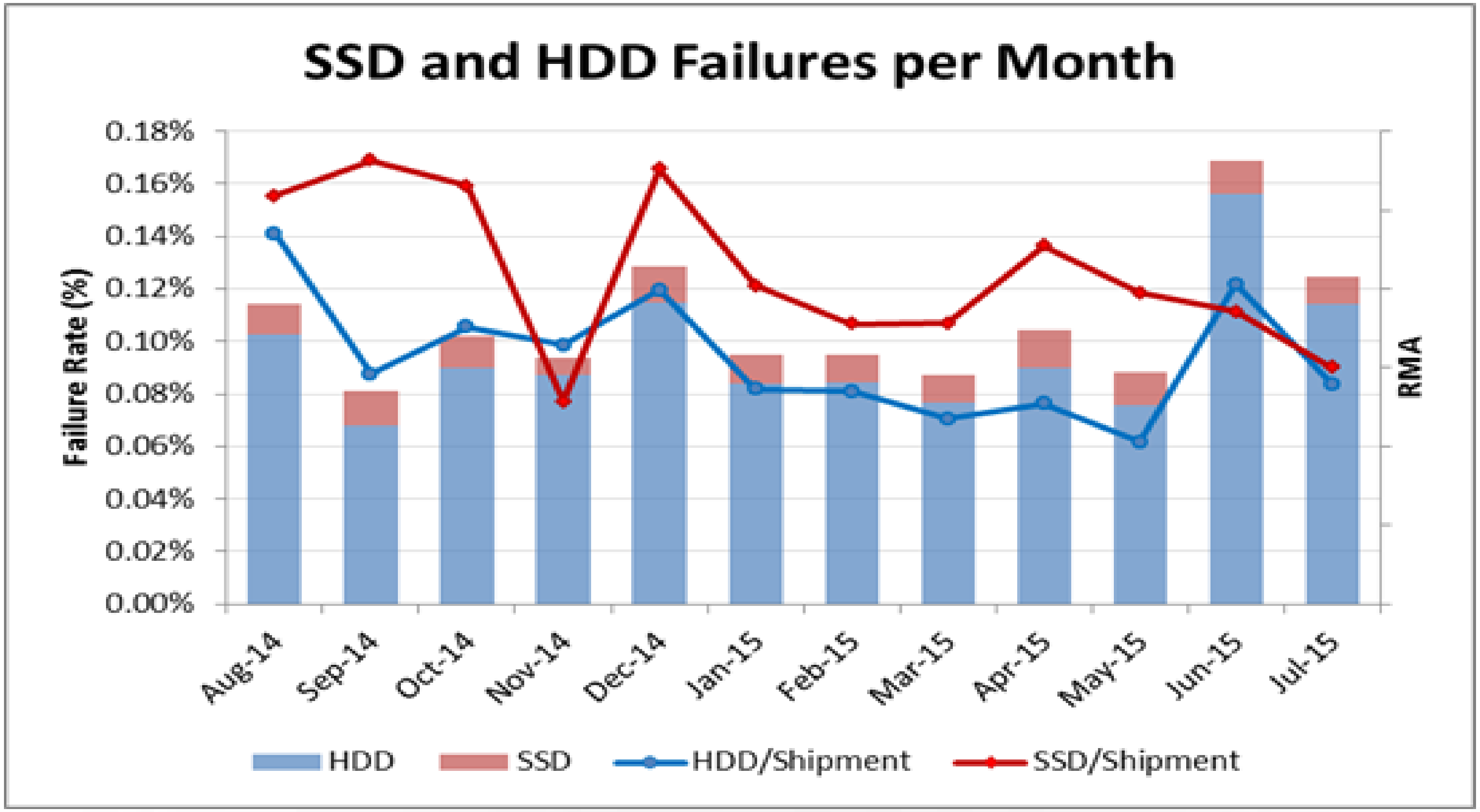
DWPD statistics for our SSD population (SSD are eMLC with 10DWPD → No wearout)



SSD Failure Statistics

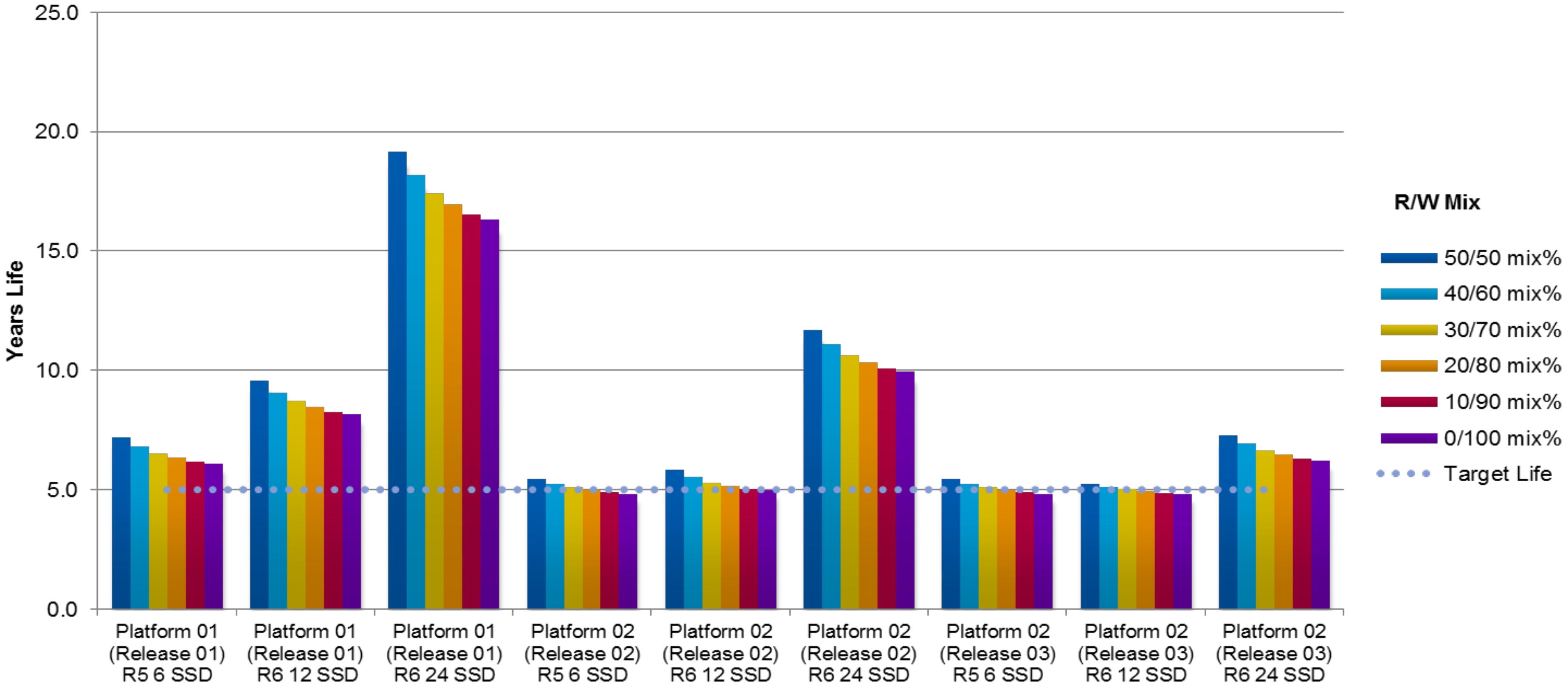
- Customer Internal Testing
  - AFR for Flash is 0.2% vs 1.2% for HDD

- Monitored AFR for SSD vs HDD:
  - HSG AFR for SSD is 0.135% vs 1.75% for HDD



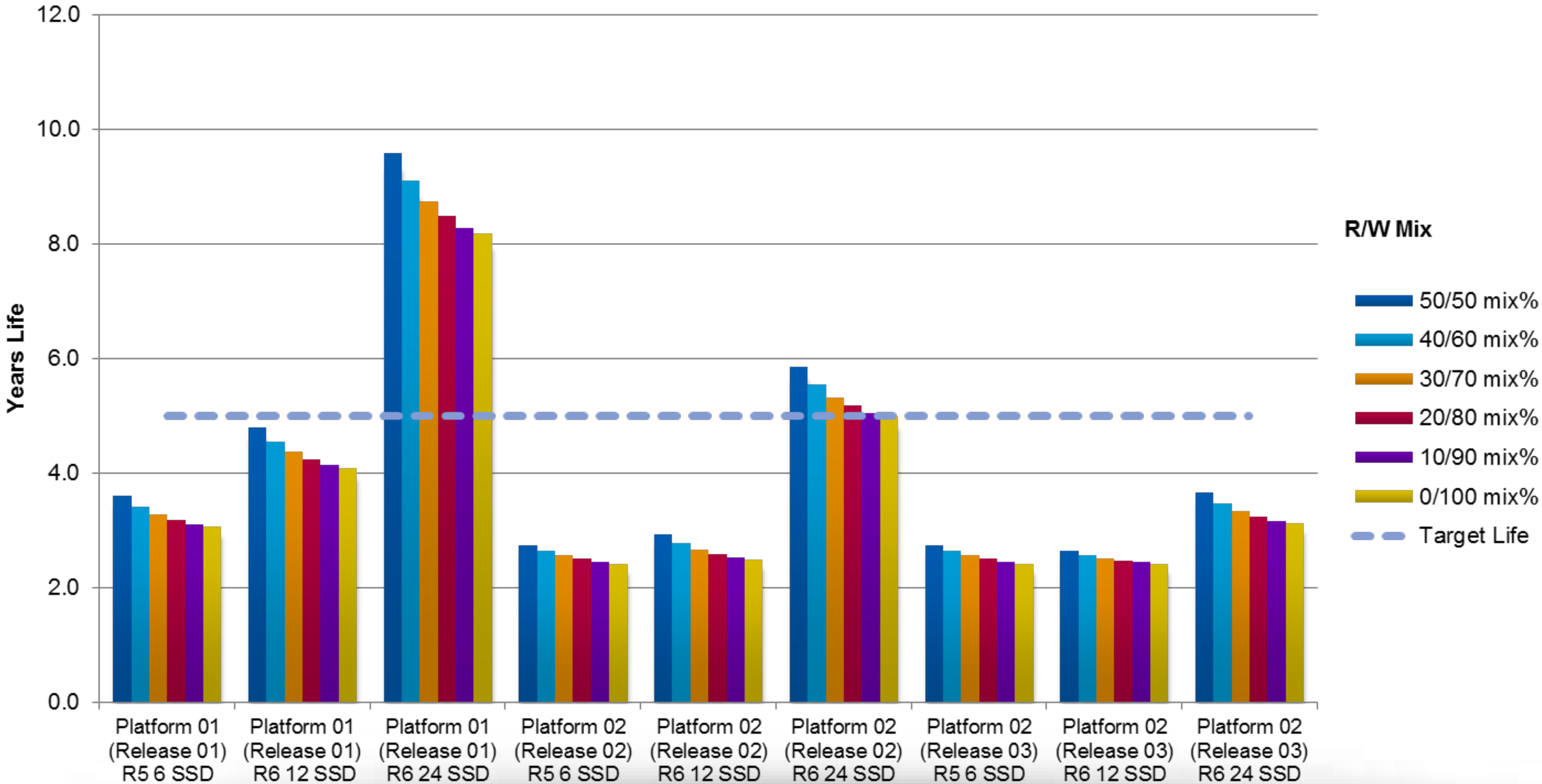
NetApp

1.6 TB 3 DWPD SSD Endurance per Platform versus Host R/W Mix, 60% max IOPs, 8K IO, 70% Duty Cycle (sustained activity 70% of time, 30% no activity), worse case write amplification

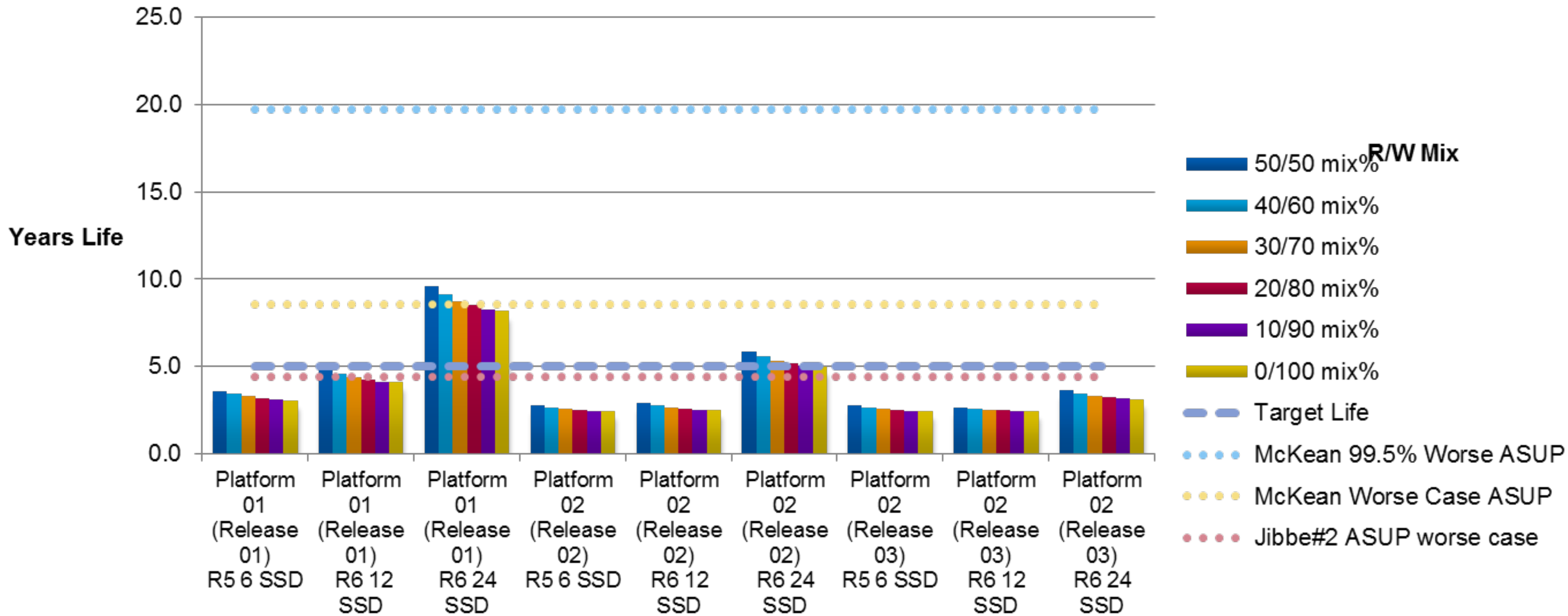


*Platform02 Release03 is drive limited with 6 SSD R5 for all mixes and for 0/100, 10/90 for 12 SSD R6

800 GB 3 DWPD SSD Endurance per Platform versus Host R/W Mix, 60% max IOPs, 8K IO, 70% Duty Cycle (sustained activity 70% of time, 30% no activity), worse case write amplification



800 GB 3 DWPD SSD Endurance per Platform versus Host R/W Mix, <u>60% max IOPs</u>, 8K IO, 70% Duty Cycle (sustained activity 70% of time, 30% no activity), worse case write amplification



Endurance Metrics for the last 2 ½ years
(Erase Count and % of Blocks Remaining)

- Analysis of SSD drives at different customers' sites for a 2.5 year period shows the following:
  - SSD at customer sites are not overloaded with large I/Os
    - 56% of the I/O sizes are <= 1M and 18% where 1M < IO<= 10M, 26.4% where 10M < IO<= 100M, and 0.021% where IO > 200M
  - SSDs at customer sites are not wearing out because
    - a) Erase count is very small
      - 92% of SSDs have not been erase yet because its written data 0 times to the entire SSD.
      - 8% of the SSD have been erase 1 – 11 times
    - b) Majority Drive Write Per Day (DWPD) is below the DWPD of eMLC
      - a) 95% of DWPD is < 3 , 5.3% where 3 < DWPD <= 10, and 0.17% where DWPD > 10.

Conclusion: Moving to cMLC (where DWPD is 10) with our new E-series product releases (HW, CFW, and MSW) is a viable and feasible solution.

Summary – choosing the right SSD

- What matters?
  - Customer usage
  - Storage RAID amplification
  - SSD performance
  - SSD endurance (DWPD) and reliability (MTBF)

SDC 15
NetApp