Leveraging Flash in Scalable Environments:

A Systems Perspective on How FLASH Storage is Displacing Disk Storage

Roark Hilomen, Engineering Fellow

Systems & Software Solutions

May 3, 2016

Forward-Looking Statements

During our meeting today, we may make forward-looking statements.

Any statement that refers to expectations, projections, or other characterizations of future events or circumstances is a forward-looking statement, including those related to product performance, cost, capacity, various use cases, expectations that flash will displace traditional media and HDD, and deployment of flash with Hadoop. Risks that may cause these forward-looking statements to be inaccurate include among others: products may not perform as expected, at the cost expected, in the capacity expected, use cases may not apply as expected, flash may not displace traditional media and HDD as expected, and deployment of flash with Hadoop may not continue as expected or at all; or the other risks detailed from time-to-time in our Securities and Exchange Commission filings and reports, including, but not limited to, our annual report on Form 10-K for the year ended January 3, 2016. This release contains statements from third parties. We undertake no obligation to update these forward-looking statements, which speak only as of the date hereof or the date of issuance by a third party, as applicable.

Speeds and Feeds

Technology	Throughput	Latency (micro)	AFR	
1Gbe Ethernet	80MB/s	400+ (40+ RDMA)		IP based solutions
10Gbe Ethernet	800MB/s	400+ (40+ RDMA)		IP based solutions
25Gbe Ethernet	~2GB/s	400+ (40+ RDMA)		IP based solutions
40Gbe	~3+GB/s	400+ (40+ RDMA)		IP based solutions
HDD Enterprise	~200MB/s	4-6ms / 200-300 IOPS	0.73	Published AFR 0.73% (various by vendor)
HDD Cloud	~200MB/s	4-6ms / 100 IOPS	0.73	
6G SAS SSD	550+MB/s	80-800	0.15	Dual Ported Capable * Latency is controller bound and GC impacted
12G SAS SSD	700MB-1GB/s Per Port	80-600	0.15	Dual Ported Capable * Latency is controller bound and GC impacted
PCIE/NVME SSD	800-2+GB/s	50-200	0.15	
6G SATA SSD	500+MB/s	300-800	0.15	Consumer Grade

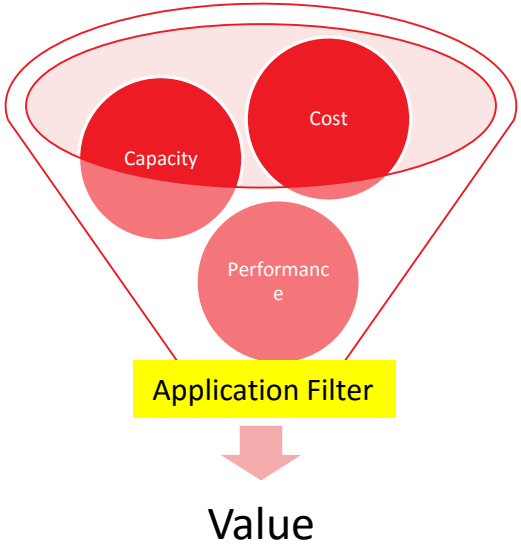
Speeds and Feeds Comparison

Media 1	Media 2	Ratio	Fill Time @ 4TB
6G SAS/SATA	HDD	~2-3x	2 hrs vs 5 ½ hrs
12G SAS SSD	HDD	~5x	1 hr vs 5 ½ hrs)
12G SAS SSD	6G SAS/SATA SSD	~2x	1 hr vs 2 hrs
PCIE/NVME	HDD	10x	½ hr vs 5 ½ hrs
PCIE/NVME	6G SAS/SATA SSD	4-6ms / 200-300 IOPS	½ hr vs 2 hrs
PCIE/NVME	12G SAS SSD	4-6ms / 100 IOPS	½ hr vs 1 hr

Transport	Fill time @ 2TB	HDD	6SAS/SATA	12G SAS	PCIE/NVME
1Gbe IP	~7 hrs	.4	.15	.08	0.04
10Gbe IP	~40 min	4	1 ½	~1	0.4
25Gbe IP	~15 min	10	~3 ½	~2	~1
40Gbe IP	~10 min	15	~5 ½	~3	~1 ½
100Gbe IP	~4min	40	~14 ½	8	4

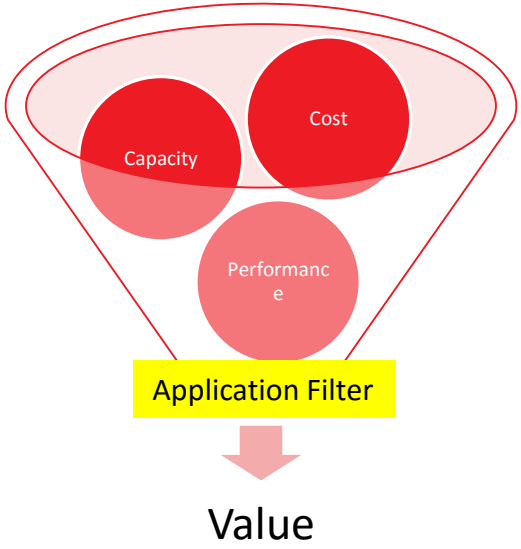
Is it the Economics or is it Really about Value

- Flash Fragmenting into Extremes
  - Lowest cost consumer drives
    - ~42c /GB Street. Prosumer Devices
  - Highest capacity, balanced cost/performance
    - 4TB / 8TB Flash Devices
  - Highest performance, high cost, limited capacity
- Application Filter is a Modifier to Value
  - Real-time application ignores cost for performance
  - Big Data requires capacity and cost at performance penalty
  - DB requires middle ground of capacity, cost, and performance



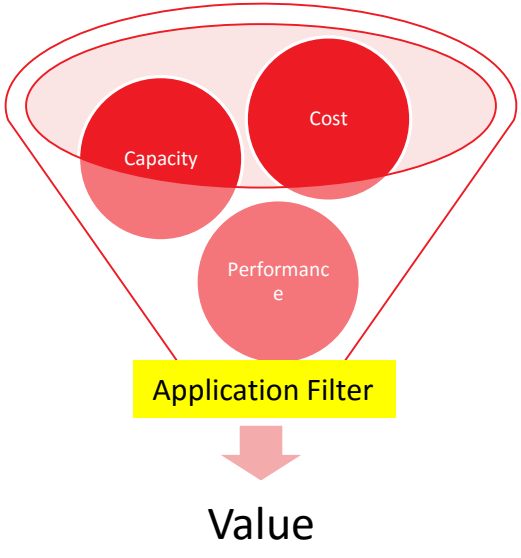
Value – Real Time/Near Real Time

- Characteristics
  - Latency is king (lower the better)
  - High value results (Time to results utmost importance)
  - Completion measured in ms
  - DRAM is expensive and data sets are in 10s or 100s TB
  - PCIe/NVME Cache/Intermediary Store to offset Cost/Performance
- Application
  - In Memory Hyper Cubes
  - In Memory DB (NoSQL)
  - Edge Sensor Analytics
- Values
  - High Rating: Performance (low latency, high IO)
  - High Rating: Capacity in Single TB xx
  - Moderate Rating: Cost



Value – Batch Analytics

- Characteristics
  - Large data sets (100s TB-> 100s PB)
  - Completion measured in seconds/minutes/hours
  - Bandwidth is king, **was** network bottlenecked
- Application
  - Search
  - Commerce/eCommerce
  - Close to core sensor analytics
  - Fraud detection during transaction
- Value
  - High Rating: Bandwidth/Aggregate IOPS
  - High Rating: Capacity
  - High Rating: Cost
  - Moderate Rating: Performance (not sub ms latency bound)



Data Lakes?
Or is it really Ponds, Lakes, Rivers, and Oceans

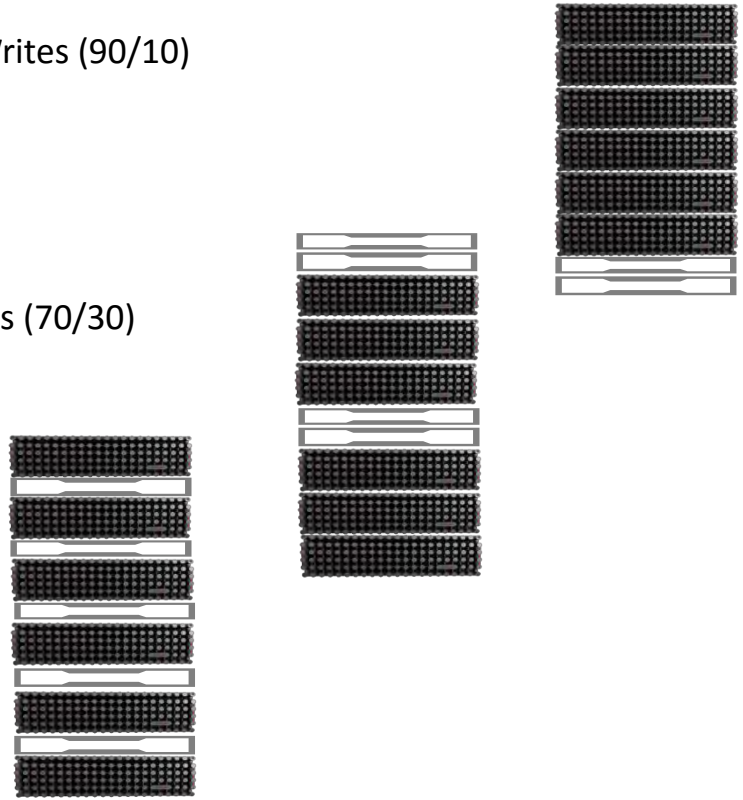
- WAN limitations pushing need for Edge to Core Big Data/Data Reduction
- Common Factor - lack of real estate and power constraints
- As ponds feed lakes, lakes feed ocean
  - Commerce : Retail -> Regional DC -> Source of Truth
  - Sensor Analytics: Critical Infrastructure -> Geolocal DC -> Source of Truth
- LAN getting faster 1->10->25->100->400Gbe striking distance
  - Hyperscale 10Gbe->40Gbe, transceiver change 25Gbe->100Gbe, not a heavy lift
  - WAN still a bottleneck

Big Data Use Cases Where Flash is Displacing Traditional Media

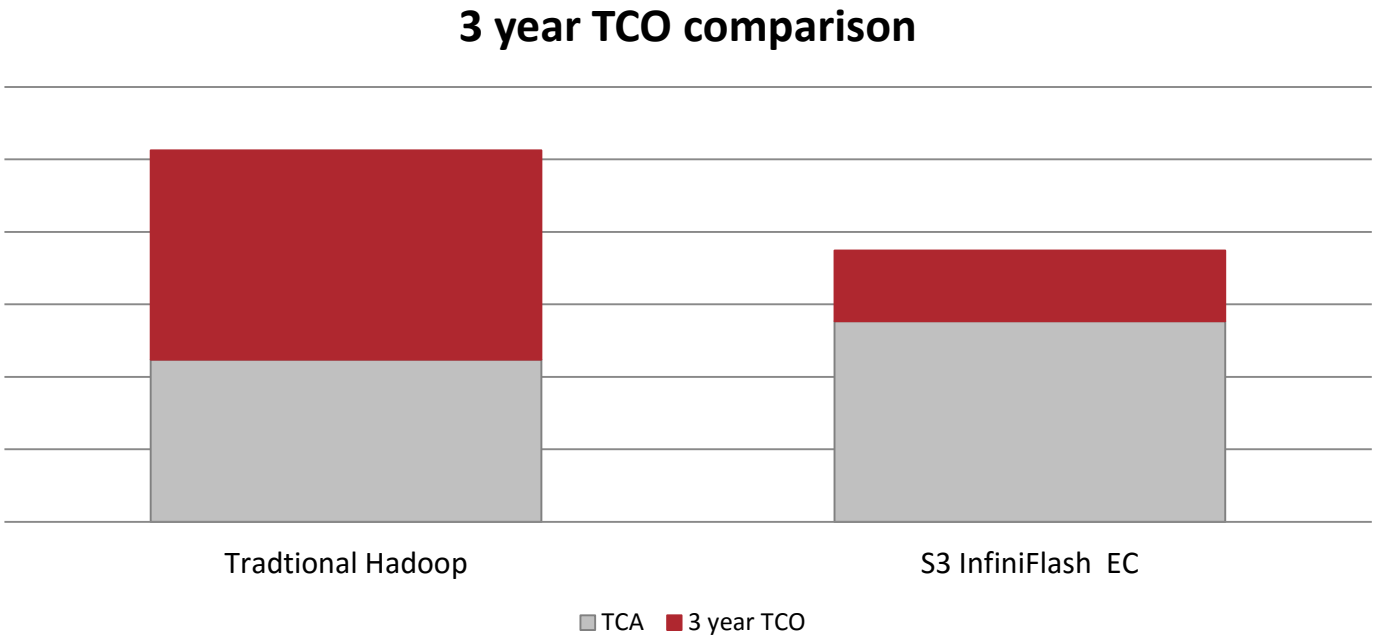
- InfiniFlash™ System - Purpose-built for Big Data
- 64 x 8TB flash cards (500TB)
- 2+ MIOPS 4k Read (aggregate)
- 500+ KIOPS 4k Write (aggregate)
- 15GB/s Chassis level bandwidth (full duplex)
- 1/10 AFR of HDD
  - (20HDD failure for every InfiniFlash card)
- Chassis level idle @ less then 200W, active @ 400-500W
  - (Comparative HDD @700W avg)
- Endurance (multi-exabyte at chassis level)
- 3U

Big Data Use Cases Where Flash is Displacing Traditional Media… Perspective

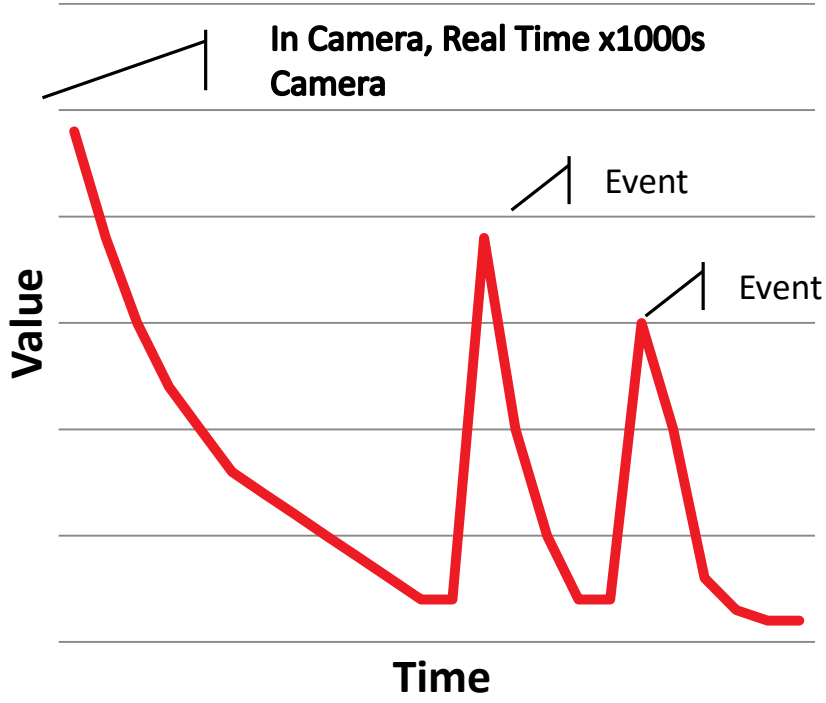
- Active Archive (Rack) - Heavy read, Batch/Periodic Writes (90/10)
  - 15 IF100, 2 Servers
  - Capacity 7.5PB (6.8PB Usable) vs. 8TB HDD usable @4+PB
  - Bandwidth 255 GB/s vs. HDD @ 30GB/s
  - Power 3KW Idle, 7KW all active vs. HDD @ 9+KW
- Data Lake Model (Rack) Heavy Read, Moderate Writes (70/30)
  - 8-12 IF100, 8-12 Servers
  - Capacity 4-6PB
  - Bandwidth up to 180GB/s
  - Power 6KW Active
- Active Analytics Model (Rack) (50/50)
  - 4-8 IF100, 8-24 Servers
  - Capacity 2-4 PB
  - Bandwidth up to 120GB/s
  - Power 10KW Very Active



InfiniFlash™ Acquisition vs. Operations Ratio



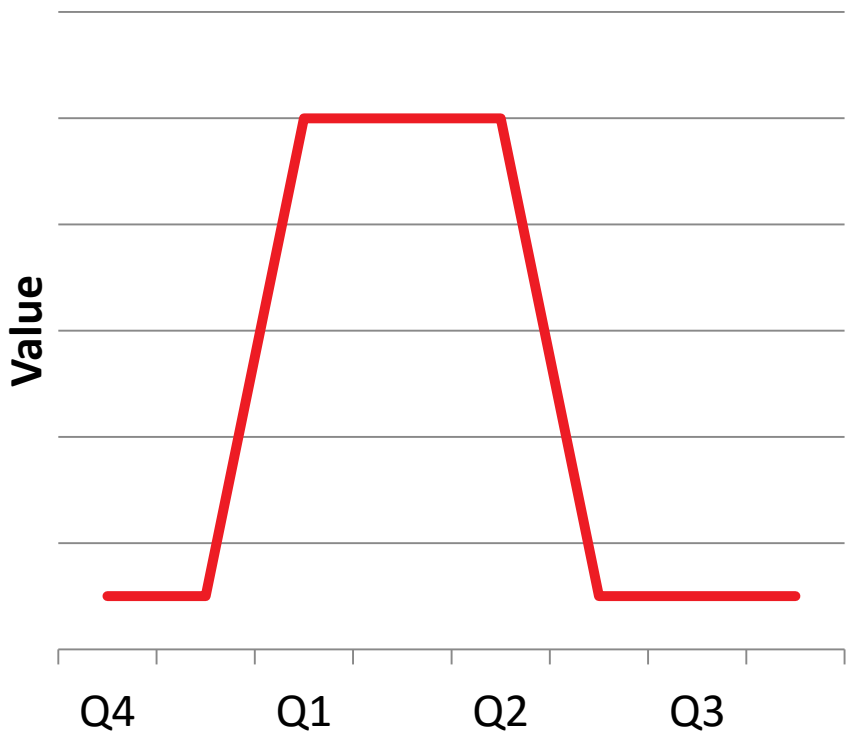
Sensor Analytics / Video Surveillance



Characteristics

- Data set size
  - High def – 100KB/s each
  - 4K – 10s MB/s each
- Source 10,000s or more data generators /sec
- Capacity and ingest bandwidth key
- Constrained by rack space/power

Federal/State, Commerce, eCommerce, Banking

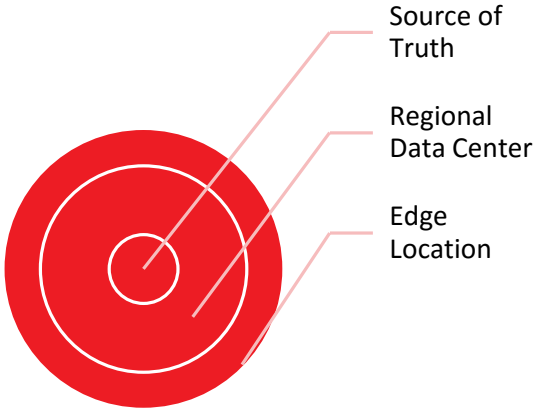


Large Data Set x Millions Individual

- Seasonality
  - Data latent until needed
  - Constant data adjustment
    - Fraud analytics
    - Buying patterns
    - Taxes
  - Resources Highly Inefficient
    - Procured for high watermark
    - 10% utilized on non-peak
    - All data important, but very long tail
    - Access pattern varies

Edge To Core

- Sample use cases
  - Retail/eCommerce
  - Critical Infrastructure
  - Banking/Fraud Detection
  - Oil & Gas
- Commonality
  - Unwanted Data Fidelity Reduction from Edge to Core
  - High Value to Analyze at Edge, Regional, and Core
  - High Value to get to Core, WAN Limitations



Key Technologies for Flash Displacement of HDD

- Thicker network pipes (data pump)
- Erasure coding (thesis: Flash will never get below the price of HDD)
  - Deep archive for HDD (1.5x cost multiplier, fail in place enablement)
  - Equivalent to hybrid-based storage arrays on flash (1.15x cost multiplier)
  - With heavy compute requirements for HDD and less so for flash, ‘price’ within all-flash systems are within 20% or less with better ‘TCO’ advantages and still better performance
- Advent of Big Data flash (devices of 4-16TB changes endurance and IO dynamics
- Throughput optimized flash vs general-purpose flash. Cost savings passed to customer.
- Advent of in-memory applications and fast network

Thank You!

@BigDataFlash
#bigdataflash
http://bigdataflash.sandisk.com

©2016 SanDisk Corporation. All rights reserved. SanDisk is a trademark of SanDisk Corporation, registered in the United States and other countries. InfiniFlash is a trademark of SanDisk Corporation. All other product and company names are used for identification purposes and may be trademarks of their respective holder(s).

Big Data on Flash: a Performance Perspective

- BSC paper: https://www.bscmsrc.eu/sites/default/files/bsc-msr_aloja.pdf
- SNDK: http://www.sandisk.com/assets/docs/increasing-hadoop-performance-with-sandisk-ssds-whitepaper.pdf

Hadoop Execution phases for different disk configs

- For Terasort

URL http://hadoop.bsc.es/perfcharts?benchmarks_length=-1&execs%5B%5D=84766&execs%5B%5D=85322&execs%5B%5D=84824

Hadoop Execution phases for different disk configs (detail)

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

URL http://hadoop.bsc.es/perfcharts?benchmarks_length=-1&execs%5B%5D=84766&execs%5B%5D=84772&execs%5B%5D=84808&execs%5B%5D=84824&execs%5B%5D=84852&metric=Disk