Wikibon

Wikibon.org

Next Generation Flash Architecture & Management

David Floyer

CTO & Co-founder, Wikibon

David.Floyer@Wikibon.org, @dfloyer

March, 2015

All Flash Case Studies

- UK Financial House:
  - Will be 100% Flash in 2015
  - Flash moved bottleneck to Processors – Installed New Faster Servers
  - Every developer has own full copy databases
  - Doubled number of production databases from 25 to 50
  - Expected doubling of development productivity
- US ISV
  - Combined all Production & Development Workloads to Flash
  - Implemented 100% Flash & Continuous Development
  - Increased # Builds/day by 3x, from 600 to 1,800
  - Build failures decreased from 17% to 2%
- US Electronic Distributer
  - Combined all workloads onto Flash
  - 30% increase in Revenue with no additional headcount in 18 months

…They All Removed the Disk Boat Anchor

At the End of this Presentation..

- Plan Implementation of an ***Electronic Data Center*** as a Strategic Imperative
- Measure & Minimize # Physical Copies of Data
- Plan to Combine Transactional, Data Warehouse & Development Data
- Plan to Completely Revamp Application Development Infrastructure & Practice
- Completely Revamp Application Architecture

…by Removing the Disk Boat Anchor □

Agenda: Second Generation Flash Architectures

- Flash vs. HDD Comparison
- Impact of Response Time on People Efficiency
- Impact of Response Time on System Efficiency
- Impact of Data Reduction & Data Sharing on Cost
- Flash Enabled Application Design
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management Requirements for New Generation AFAs
- Conclusions & Recommendations

Agenda: Second Generation Flash Architectures

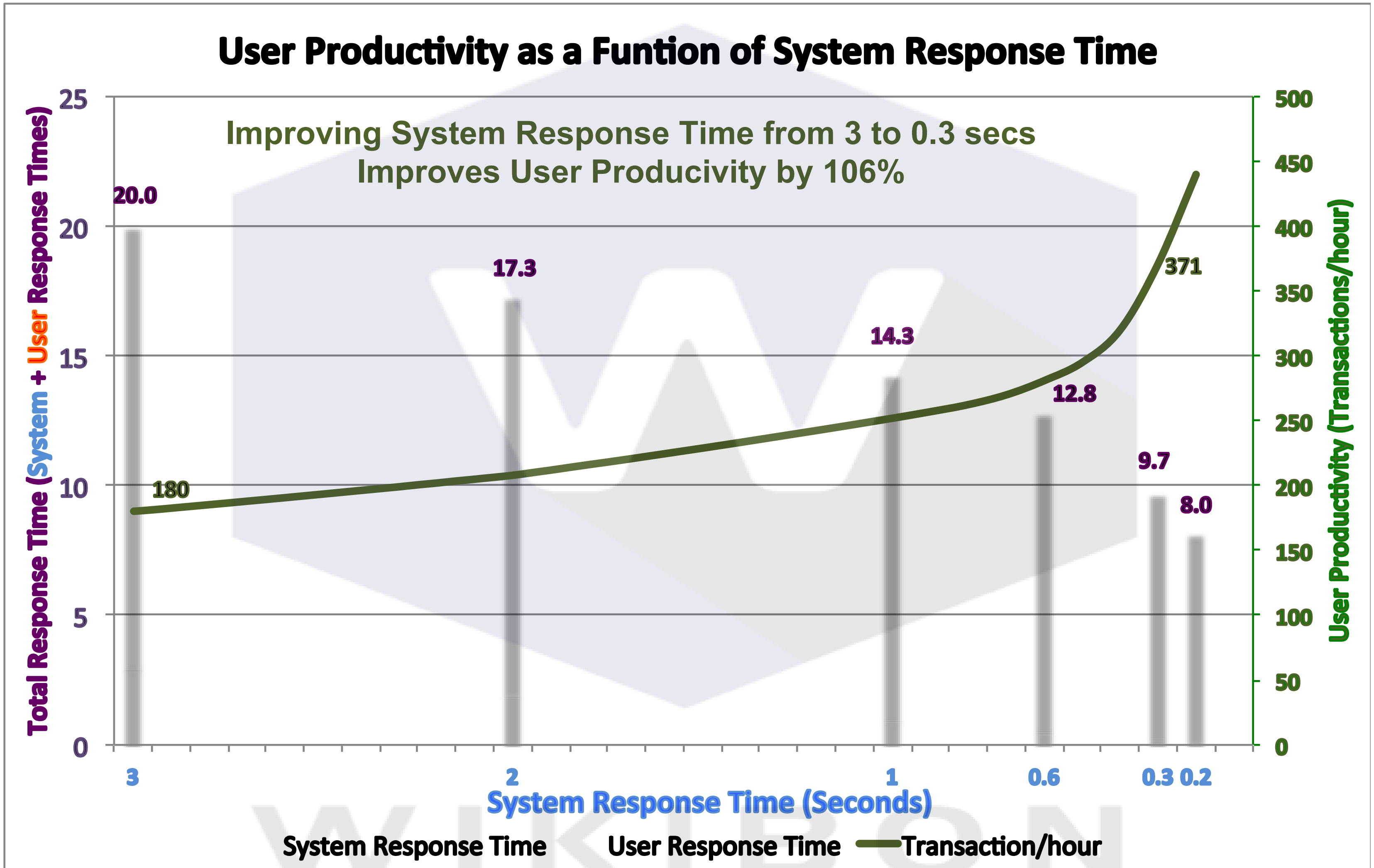
- **Flash vs. HDD Comparison**
- **Impact of Response Time on People Efficiency**
- **Impact of Response Time on System Efficiency**
- Impact of Data Reduction & Data Sharing on Cost
- Flash Enabled Application Design
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management Requirements for New Generation AFAs
- Conclusions & Recommendations

Flash Characteristics compared with HDD

- Flash more expensive per Byte raw
- Flash prices driven by consumer demand (mobile)
- HDD for mobile & desktop rapidly declining market
  - Desktop/Laptop SSD 25% in 2014, 50% in 2018
  - Mobile market 100% Flash
- Flash faster improvement compared with HDD
  - Capacity: Flash ~30% CAGR, HDD ~15% CAGR
  - Bandwidth: Flash ~30% CAGR, HDD <8% CAGR
  - IOPS: Flash ~30% CAGR, HDD <0% CAGR
- HDD characteristics allow very little sharing of data
  - Space-efficient snapshots limited to fast recovery
  - Full copies must be made if data is accessed by multiple applications (e.g., production & development)
- Flash allows true virtualization of data
  - Data can be aggressively reused
  - Fewer full copies need to be made
- HDD is best with sequential workloads, Flash is best with random
  - HDD need large caches & small working sets for random workloads
  - Flash can work with all workloads, including truly random workloads

Flash & Disk Need Completely Different Architecture & Management

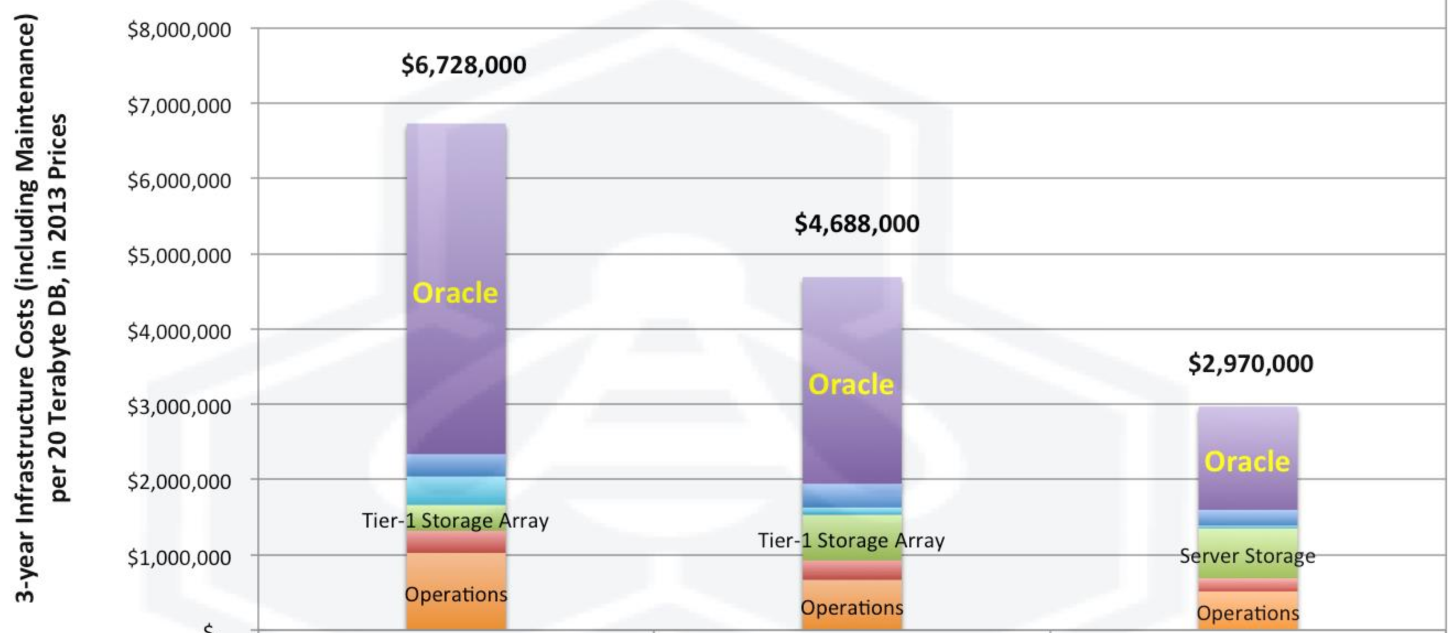
Productivity as a Function of Response Time



Source: Data interpreted by Wikibon from a seminal paper by Walter J. Doherty & Ahrvind J. Thadani, 1982, IBM Thomas J. Watson Research Center, updated to include other studies 1997, downloaded http://www.vm.ibm.com/devpages/jelliott/evrrt.html April 2015

Cost of Database Licenses as a function of IO RT

Impact of Flash on $3-year Cost of 20TB Database Infrastructure



	Traditional (DISK, SCSI)	All or High % Flash (SCSI)	All-Flash (Atomic Writes)
Oracle Database Enterprise Edition	$4,390,000	$2,744,000	$1,372,000
Servers	$296,000	$314,000	$210,000
Environmentals (Power & Space)	$378,000	$98,000	$36,000
Tier-1 Storage or Server Storage	$342,000	$604,000	$664,000
Infrastructure Software	$296,000	$260,000	$174,000
Operations or Dev/Ops	$1,026,000	$668,000	$514,000
Total Cost	$6,728,000	$4,688,000	$2,970,000

Source: © Wikibon April 2013

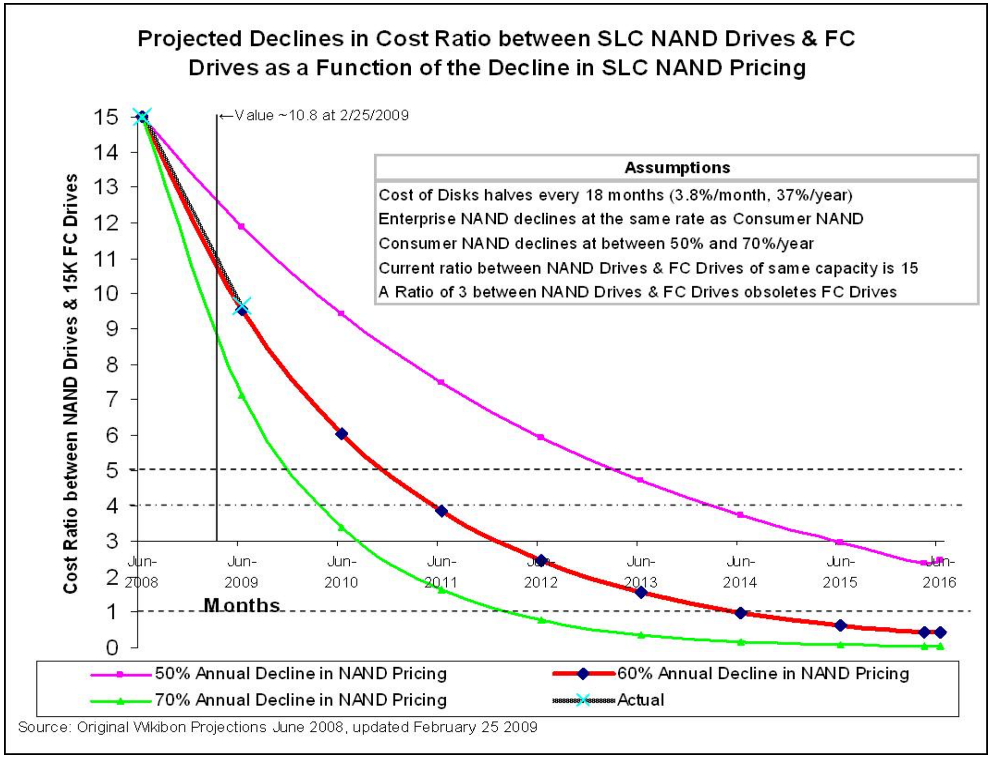
In reality - more Applications for same Oracle Budget

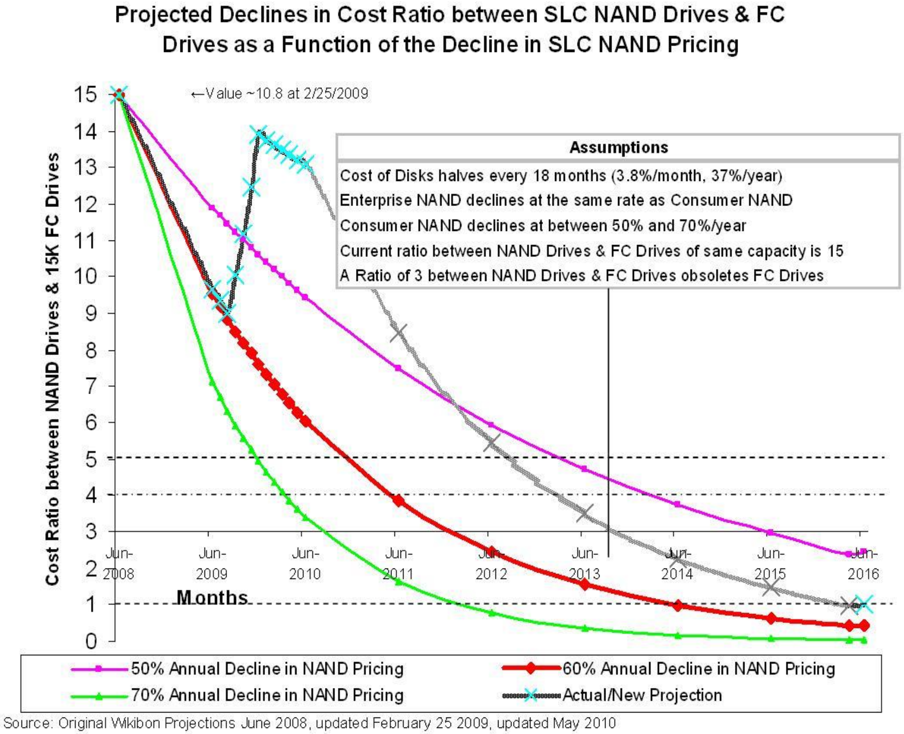
http://wikibon.org/wiki/v/Flash_and_Hyperscale_Changing_Database_and_System_Design_Forever

Agenda: Second Generation Flash Architectures

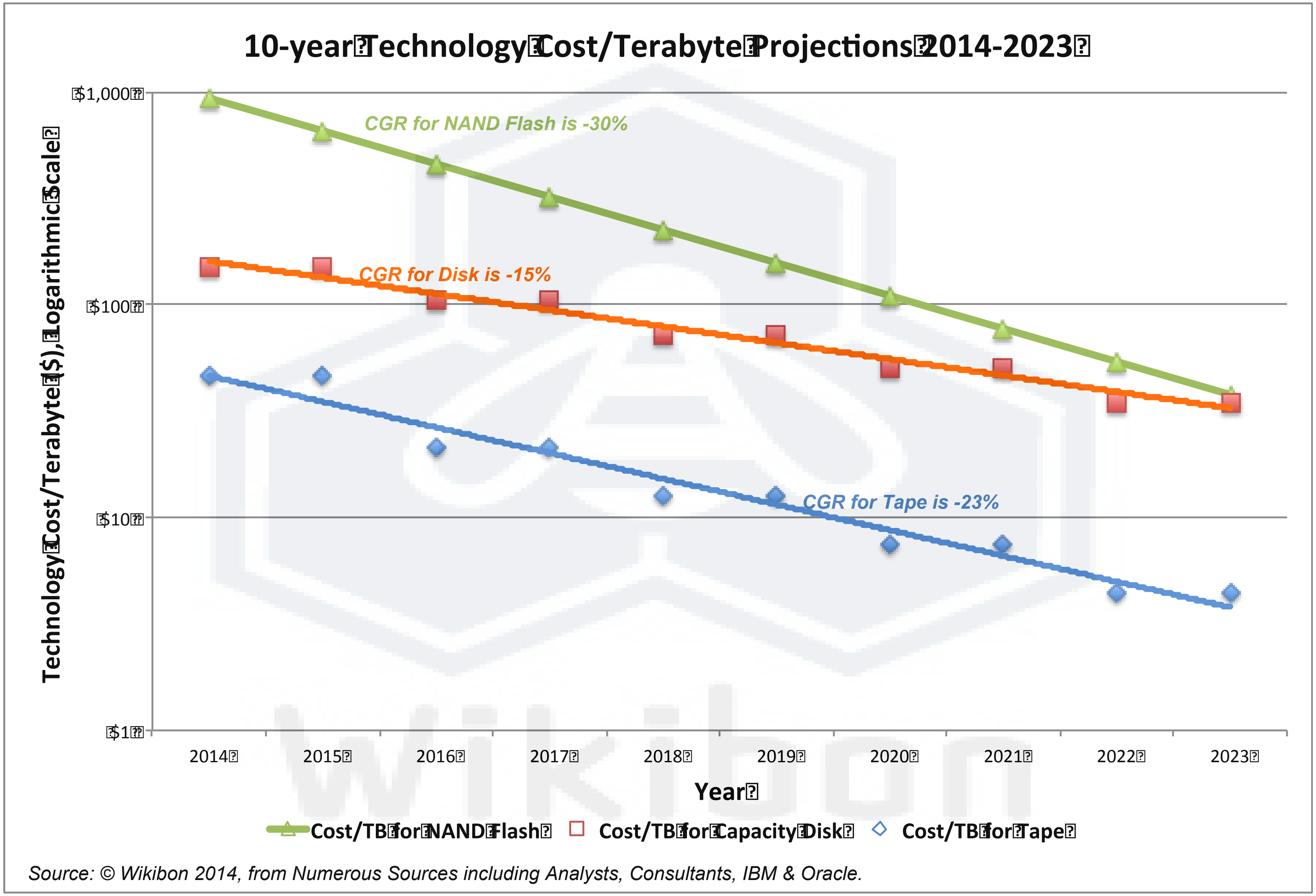
- Flash vs. HDD Comparison
- Impact of Response Time on People Efficiency
- Impact of Response Time on System Efficiency
- **Impact of Data Reduction & Data Sharing on Cost**
- Flash Enabled Application Design
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management Requirements for New Generation AFAs
- Conclusions & Recommendations

Wikibon 2009/2010 Flash Forecasts



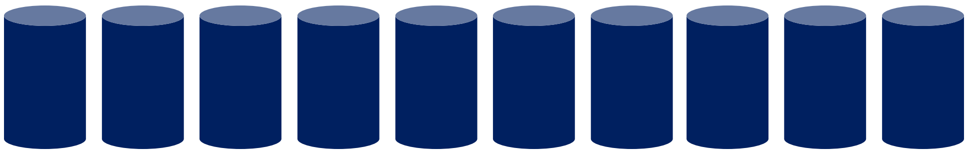


10-year Technology Cost/TB Projections

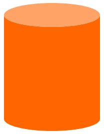


Copy Management

Large Independent Caching



Small Shared Cache



Traditional Disk Array

All Flash Array

90% of Data is a Copy of Original data

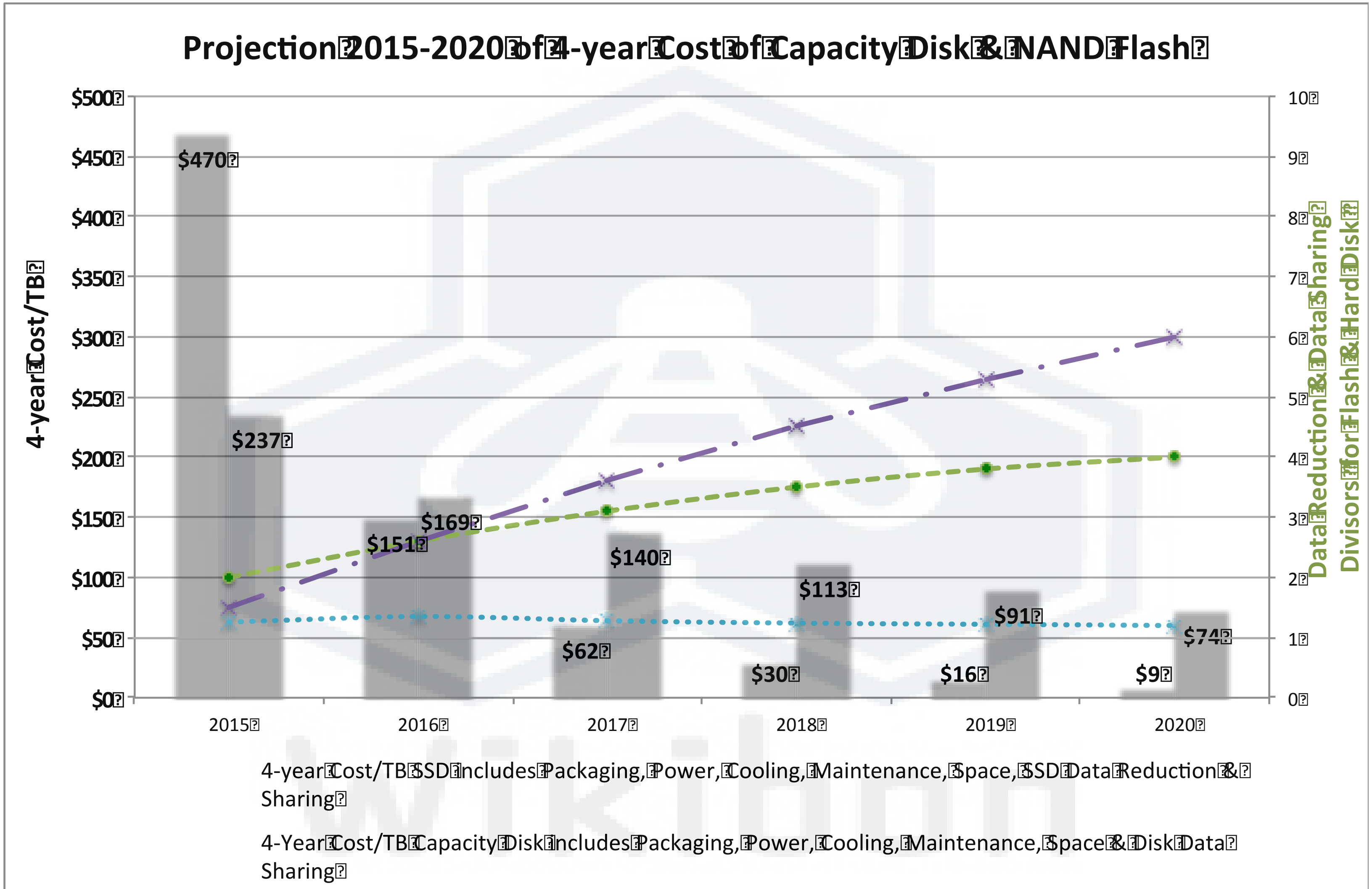
Flash allows Data Reduction & Space-efficient Snapshots allow Data Sharing

Action: Measure & Minimize # Physical Copies of Data

Cost case of AFA

- 6 x reduction in cost from data sharing and copy elimination
- 4 x reduction from compression and de-duplication
- Much faster response time for all applications (end-user productivity)
- Ability to deploy new applications with OLTP mixed with ***Inline Analytics***
- ***Potential 24 x Reduction in Raw Storage Required***

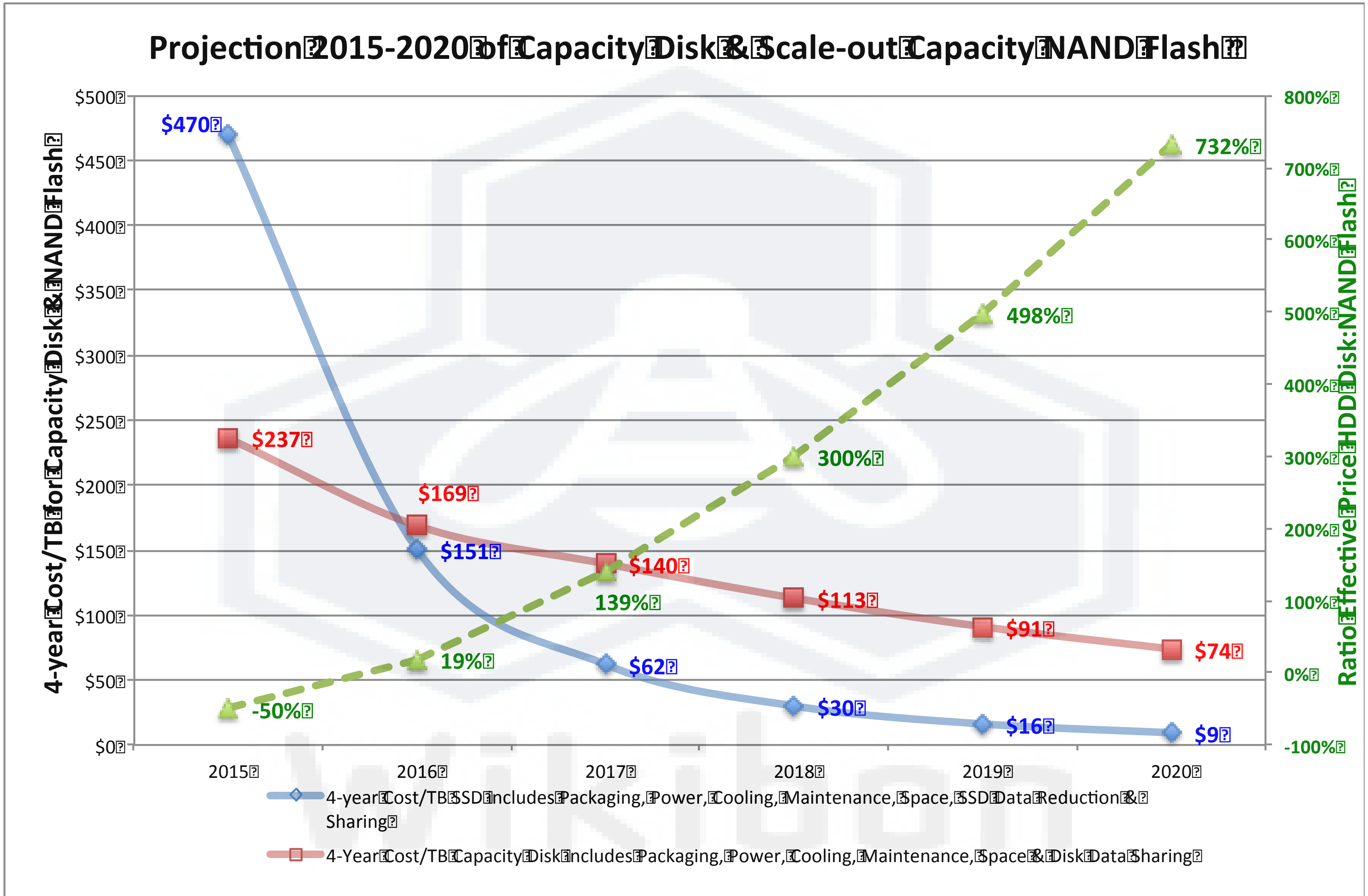
Infrastructure Costs by Technology



Source: © Wikibon 2015. 4-Year Cost/TB Magnetic Disk & SSD, including Packaging, Power, Maintenance, Space, Data Reduction & Data Sharing

http://wikibon.org/wiki/v/Evolution_of_All-Flash_Array_Architectures

Infrastructure Costs by Technology



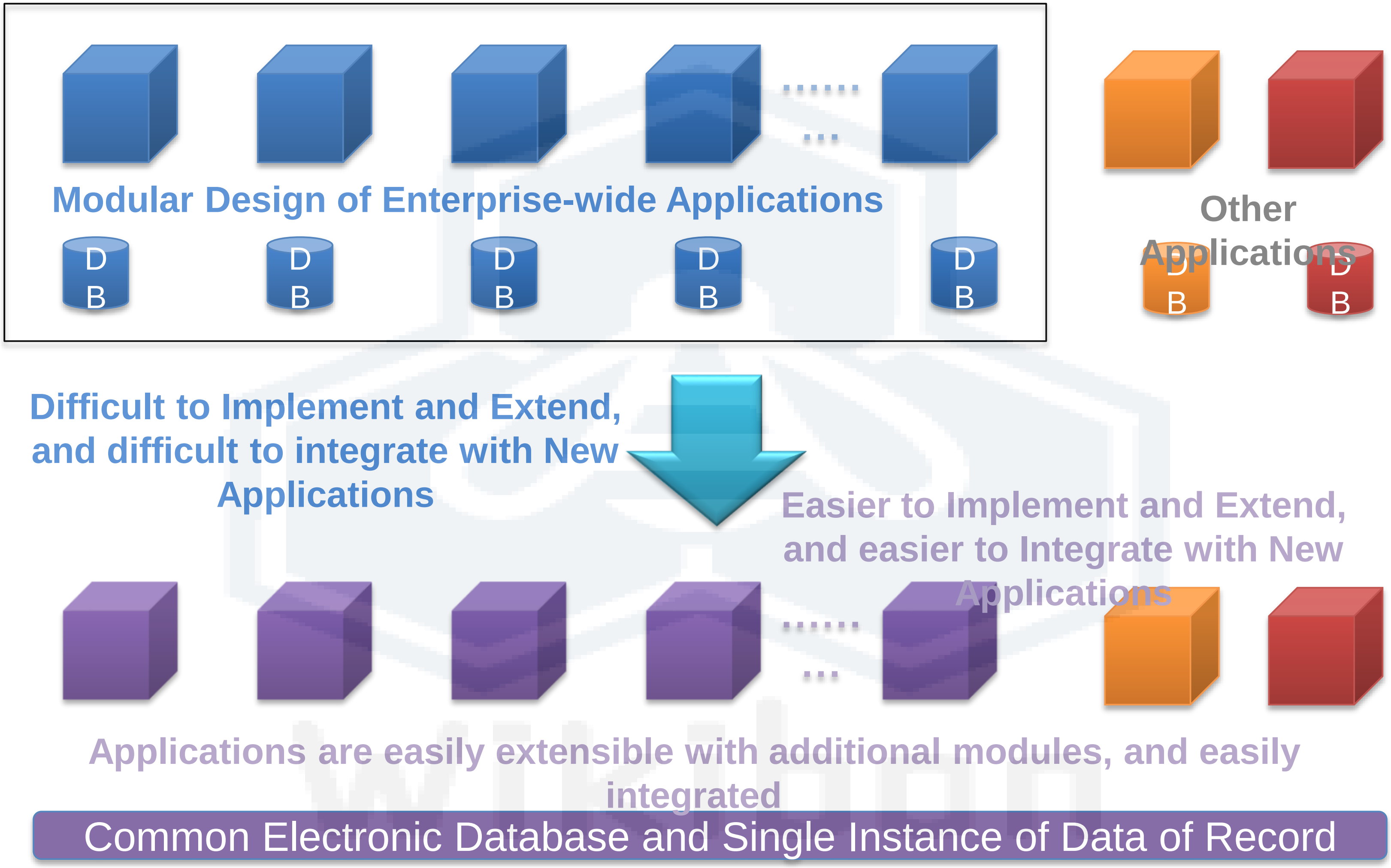
Source: © Wikibon 2015. 4-Year Cost/TB Magnetic Disk & SSD, including Packaging, Power, Maintenance, Space, Data Reduction & Data Sharing

http://wikibon.org/wiki/v/Evolution_of_All-Flash_Array_Architectures

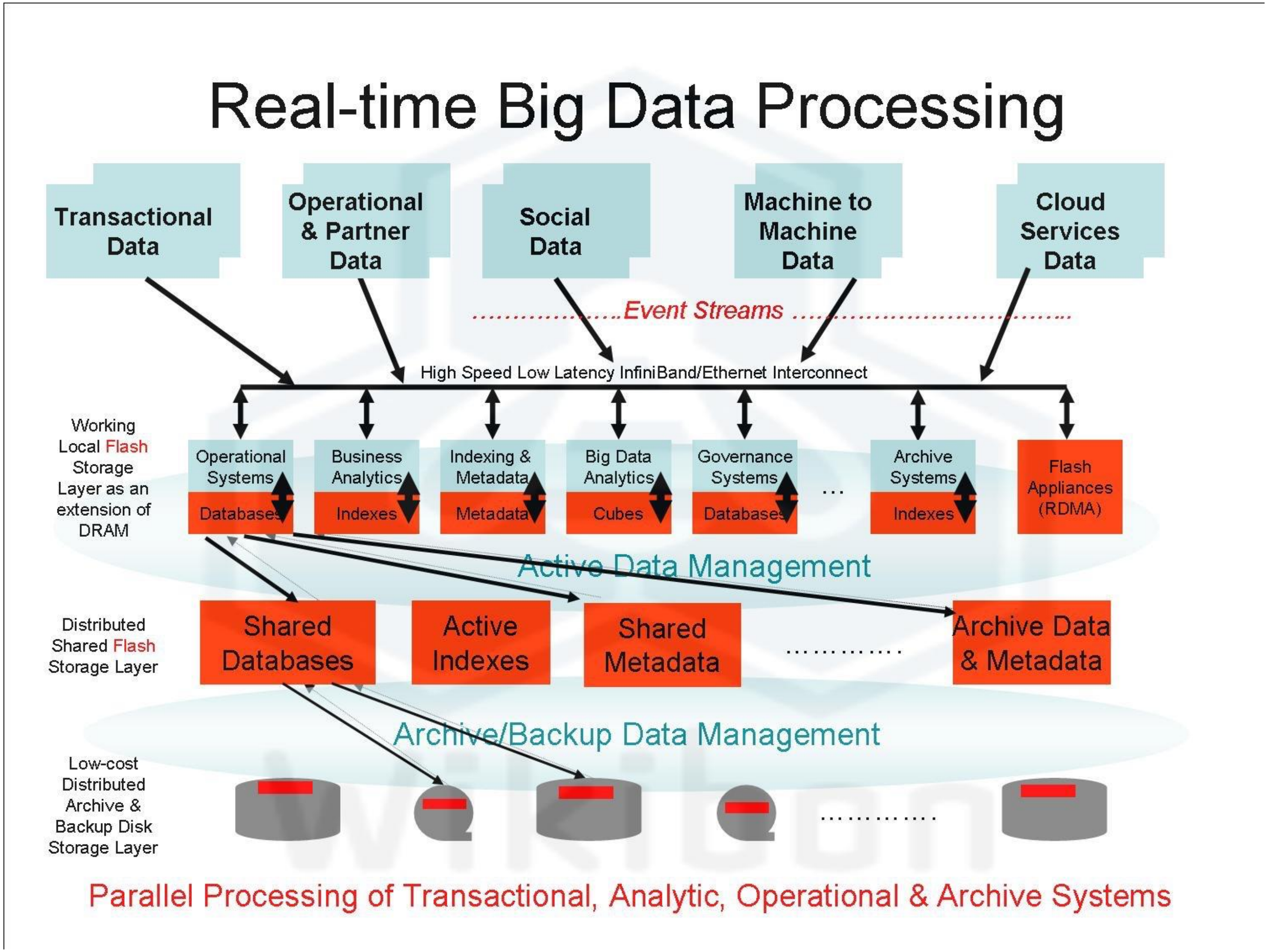
Agenda: Second Generation Flash Architectures

- Flash vs. HDD Comparison
- Impact of Response Time on People Efficiency
- Impact of Response Time on System Efficiency
- Impact of Data Reduction & Data Sharing on Cost
- **Flash Enabled Application Design**
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management Requirements for New Generation AFAs
- Conclusions & Recommendations

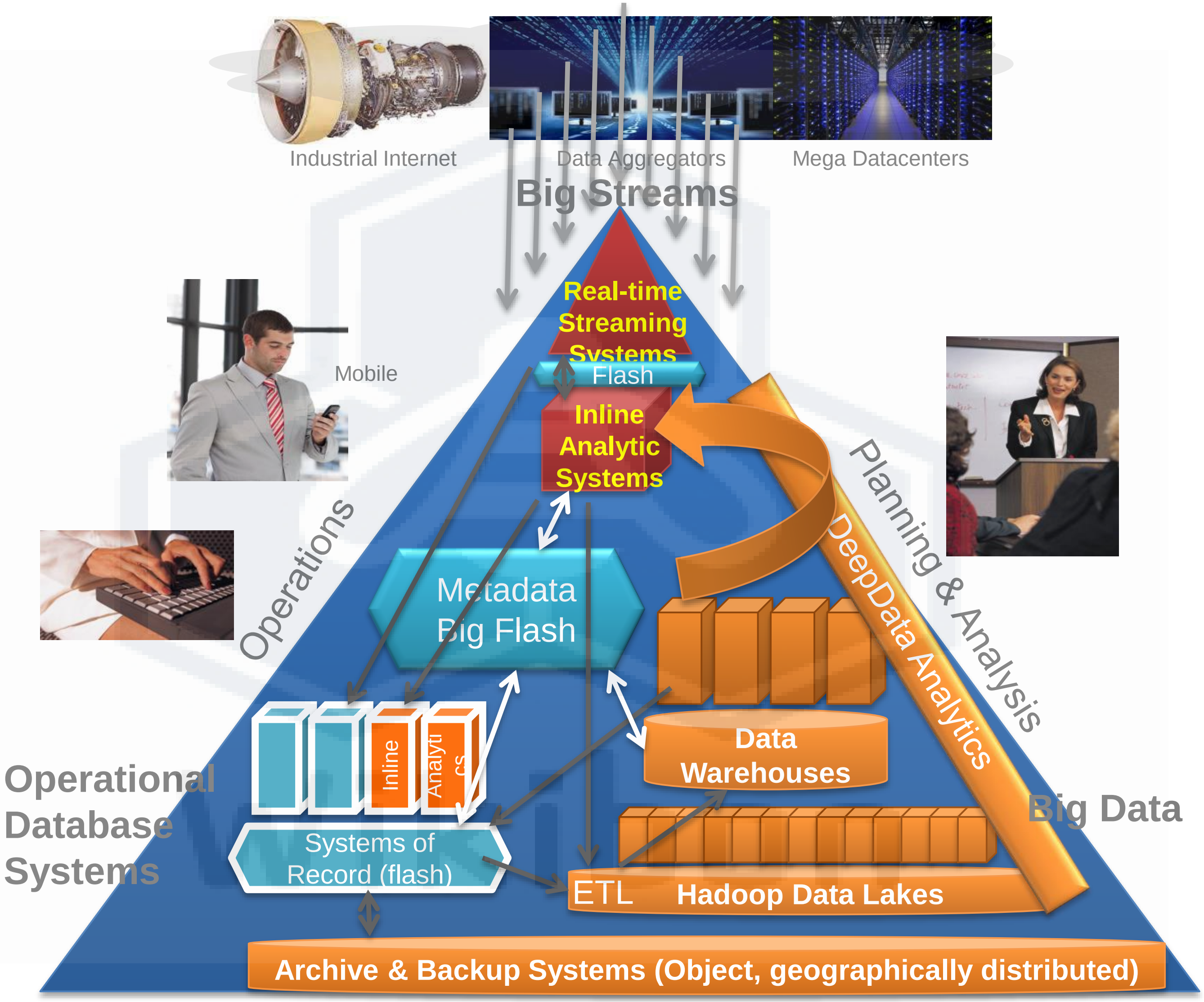
Flash-enabled Application Design



Real-time Big Data Processing



Integrated Transactional, Analytic & Development Data Management



Agenda: Second Generation Flash Architectures

- Flash vs. HDD Comparison
- Impact of Response Time on People Efficiency
- Impact of Response Time on System Efficiency
- Impact of Data Reduction & Data Sharing on Cost
- Flash Enabled Application Design
- **First Generation AFA**
- **Architectural Requirements for New Generation AFAs**
- **Management Requirements for New Generation AFAs**
- Conclusions & Recommendations

1st Generation AFA

- Copy of Traditional HDD Array architecture
- Traditional 2-controller Design
- Traditional Cache management
- Controller speed Constraint for Functionality & Amount of storage
- “Storage Silo” view of world
- Examples:
    - Cisco Whiptail
    - IBM TMS
    - NetApp e-Series
    - Nimbus
    - Pure
    - Skyera
    - Violin

Architecture Requirements for New Generation AFAs

- More data held in Array, greater savings in reducing copies
  - Scale out architecture, Dynamic addition of capacity
- No tiering required for 95%+ of data
- Simple tiering only required for <5% of data with:
  - Very low change rate
  - Low historical data access
  - No dynamic requirement for transfer
- Full storage reduction techniques multiply benefits by amount of reuse
- AFA must use snapshot change management (vs. traditional replication by application and copy of data)
- Virtualization & Sharing of Data requires extremely high levels of metadata protection
  - Accidental loss
  - Microcode failure
  - Technology failure
  - Malicious long-term/short-term hacking

Management Requirements for New Generation AFAs

- Catalog of Data Copies, Snapshots, etc.
  - Catalog shared with Linked & Remote AFA arrays
  - Automated Backup & Recovery system
- Full access to data via Restful APIs for platform integration
- Extensive Quality of service management
  - Minimums & Maximum IOPS, Bandwidth & RT
  - Different QoS for snaps
- Full Application IO view
- Full IO monitoring
  - By application
  - By copy
  - % shared data
  - Etc.
- Automated migration of unsuitable data to HDD
  - Option to retain Metadata at AFA
- Full Orchestration & Workflow Automation support for Platforms

Management Requirements for New Generation AFAs

- Catalog of Data Copies, Snapshots, etc.
  - Catalog shared with Linked & Remote AFA arrays
  - Automated Backup & Recovery system
- Full access to data via Restful APIs for platform integration
- Extensive Quality of service management
  - Minimums & Maximum IOPS, Bandwidth & RT
  - Different QoS for snaps
- Full Application IO view
- Full IO monitoring
  - By application
  - By copy
  - % shared data
  - Etc.
- Automated migration of unsuitable data to HDD
  - Option to retain Metadata at AFA
- Full Orchestration & Workflow Automation support for Platforms

Reasons for Scale-out

- Greater Sharing of Data
- Greater De-duplication
- Fewer Copies
- Simpler Data & Metadata Management
- Allows Migration to Continuous Development
- Allows Migration to Real-time ETL
- Allows Migration to In-line Analytics
- Allows Next-generation Applications with 1,000x Database Calls

Conclusions & Recommendation's

- Plan Implementation of an ***Electronic Data Center*** as a Strategic Imperative
- Measure & Minimize # Physical Copies of Data
- Plan to Combine Transactional, Data Warehouse & Development Data
- Plan to Completely Revamp Application Development Infrastructure & Practice
- Completely Revamp Application Architecture

Business & IT Plan to Double IT Productivity & Double Productivity of Application Users

Appendix I: Cost Assumptions for Flash on Storage Arrays

Wikibon 2015 Q2 Storage Cost Assumptions						
	Storage Type	$/Usable TB without DRe	Data Reduction Ratio (DRe)	Number of Shared Copies	Overall Data Reduction Divisor	Net $/Uable TB
Performance	Very Low Latency Flash without DRe	$16,500	1	1	**1**	**$16,500**
	Low Latency Flash with Compression	$16,500	2	1	**2**	**$8,250**
	Tier 1 Array with Flash Tiering	$8,000	1	1	**1**	**$8,000**
	All-flash Array without DRe Function	$9,000	1	2	**2**	**$4,500**
	All-flash Array with full DRe Function	$15,000	4	2	**8**	**$1,875**
	Tier 1 Array with Magnetic Disk	$1,700	1	1	**1**	**$1,700**
	Tier 2 Array with Magnetic Disk	$1,000	1	1	**1**	**$1,000**
Capacity	Capacity All-flash Array	$900	1	1	**1**	**$900**
	Capacity Magnetic Disk Array	$550	1	1	**1**	**$550**
	Capacity All-flash Array with Sharing	$900	1	2	**2**	**$450**
	Capacity All-flash Array with Compression & Sharing	$900	2	2	**4**	**$225**
	Capacity DAS Magnetic Disk Storage	$150	1	1	**1**	**$150**

Source: © Wikibon 2015

Appendix II: Storage Cost Assumptions

Assumptions & Calculations for 4-year Effective Cost/TB for Magnetic Disk & NAND Flash, 2015 - 2020						
Year	2015	2016	2017	2018	2019	2020
Disk TB/Drive	4	5	6	7	8	10
Raw Cost for Disk/TB	$150	$127	$104	$88	$72	$61
Disk Maintenance % of Disk Cost/year	18%	18%	18%	18%	18%	18%
4-year Disk Maintenance/TB	$127	$85	$61	$41	$30	$20
4-year Disk Power, Cooling & Space/TB (PUE=3)	$19	$16	$13	$11	$9	$7
4-year Cost Disk (including Power, Cooling, Maintenance, Space)/TB	**$296**	**$228**	**$179**	**$140**	**$111**	**$89**
Raw Cost for Flash SSD	$656	$459	$322	$225	$158	$110
Flash Maintenance % of Flash Cost	10%	9%	8%	7%	6%	5%
4-year Flash Maintenance/TB	$223	$138	$84	$52	$33	$22
4-year Flash Power, Cooling, & Space/TB (PUE=3)	$2	$2	$1	$1	$1	$1
4-year Cost SSD (including Power, Cooling, Maintenance, Space)/TB	**$881**	**$599**	**$406**	**$278**	**$192**	**$133**
Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Cost/TB for Flash (including Packaging, Power, Space, Cooling & Maintenance)	$1,410	$958	$650	$445	$307	$213
Cost/TB for Disk (including Packaging, Power, Space, Cooling & Maintenance)	$474	$365	$286	$224	$178	$142
Data Sharing Divisor for Flash	1.5	2.6	3.6	4.5	5.3	6
Data Reduction Divisor for Flash	2	2.6	3.1	3.5	3.8	4
Data Reduction Divisor for Disk	1.25	1.35	1.28	1.24	1.22	1.2
4-Year Cost/TB Capacity Disk includes Packaging, Power, Cooling, Maintenance, Space & Disk Data Sharing	**$237**	**$169**	**$140**	**$113**	**$91**	**$74**
4-year Cost/TB SSD includes Packaging, Power, Cooling, Maintenance, Space, SSD Data Reduction & Sharing	**$470**	**$142**	**$58**	**$28**	**$15**	**$9**
Source: © Wikibon 2015						