# 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
  - Doubled productivity of development
- US ISV
  - Combined all Production & Development Workloads to Flash
  - Implemented 100% Flash & Continuous Development
  - Increased # Updates/Release by 3x, from 600 to 1,800
- 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

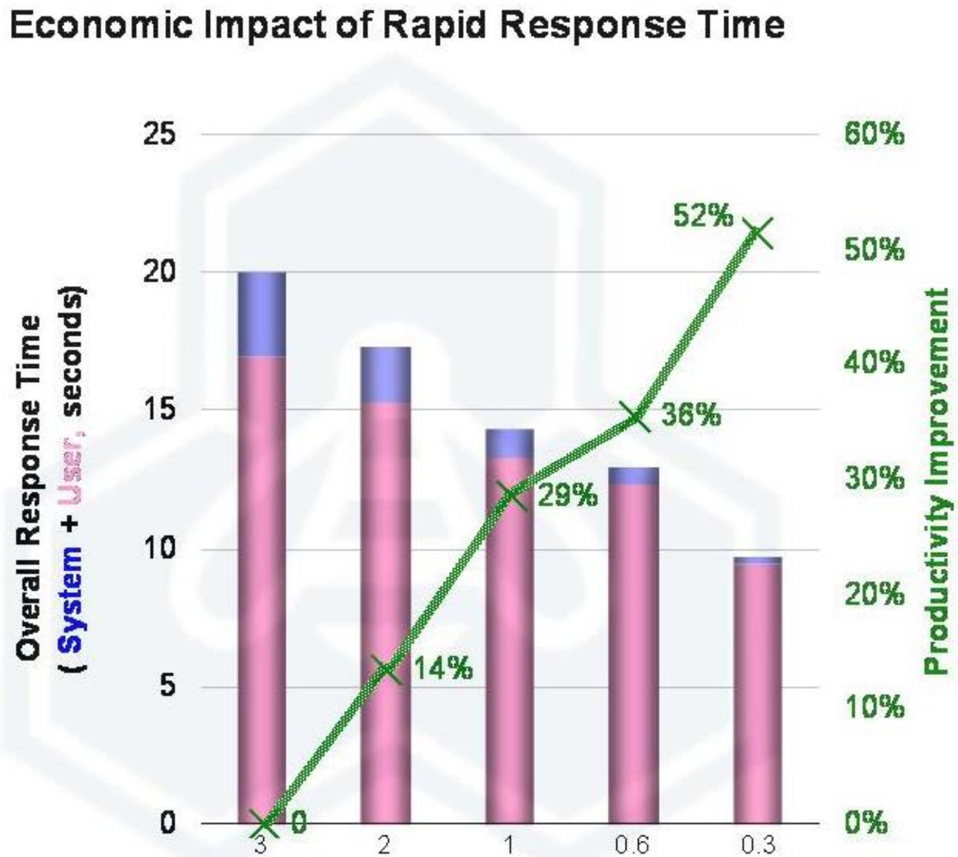
- **Flash vs. HDD Comparison**
- **Impact of Response Time on People Efficiency**
- **Impact of Response Time on System Efficiency**
- Impact of Data Reduction & Data Sharing on Cost
- Flash Enabled Application Design
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management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

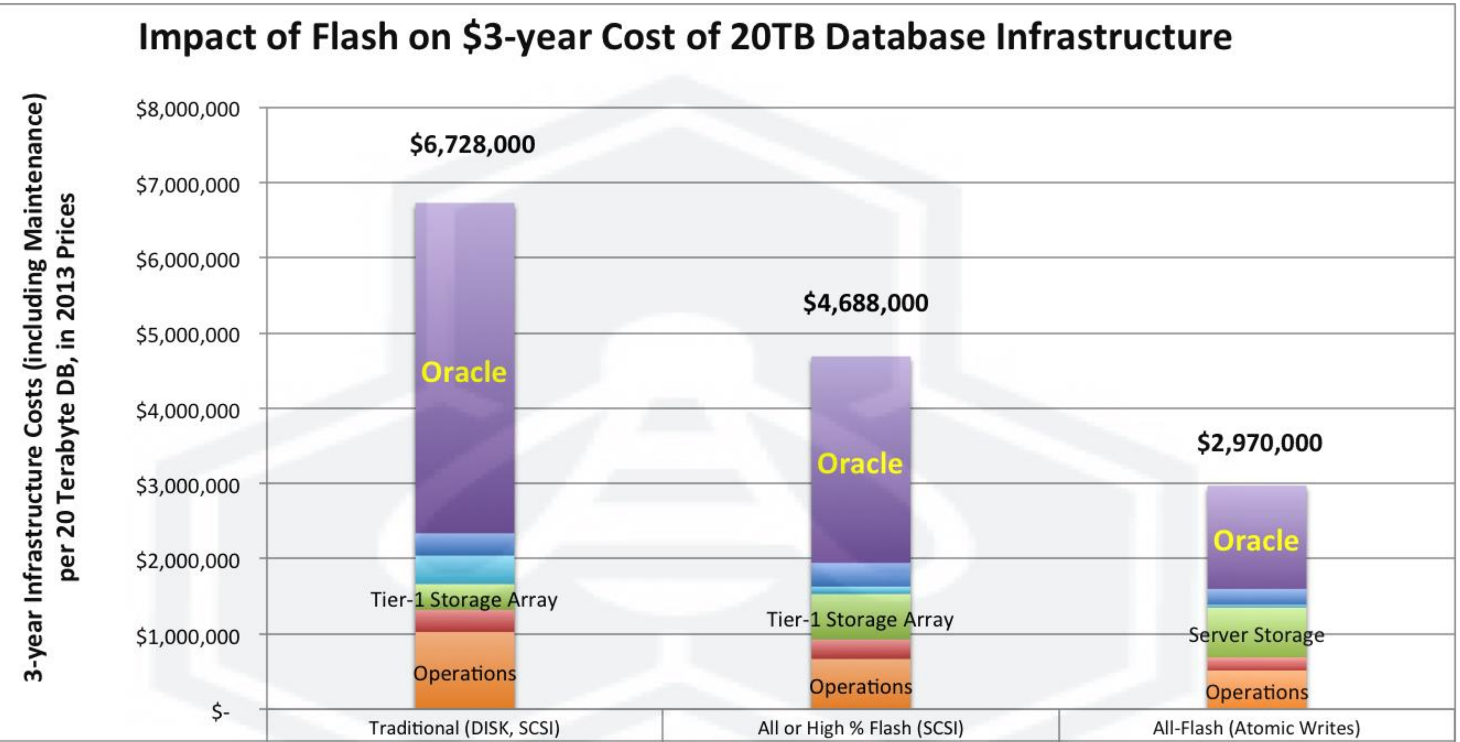



| | 3 | 2 | 1 | 0.6 | 0.3 |
|---|---|---|---|---|---|
| System Response Time | 3 | 2 | 1 | 0.6 | 0.3 |
| User Response Time | 17 | 15.3 | 13.3 | 12.3 | 9.4 |
| % Productivity Improvement | 0 | 14% | 29% | 36% | 52% |

**System Response Time (seconds)**

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

# Cost of Database Licenses as a function of IO RT




| | 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 |

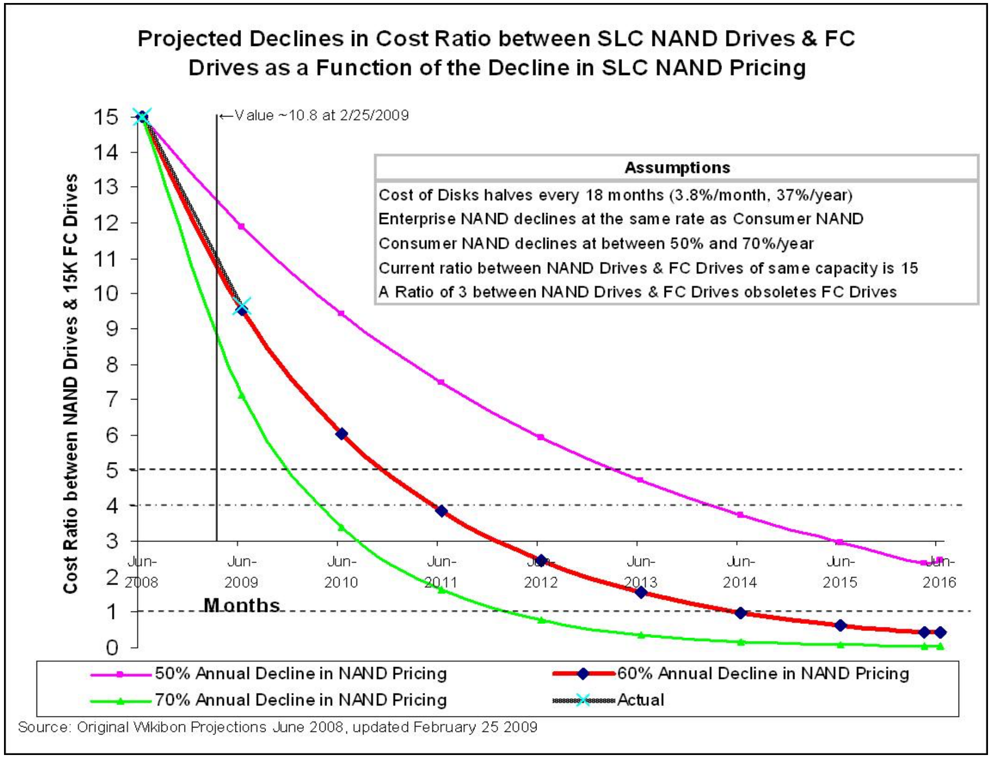
*Source: © Wikibon April 2013*

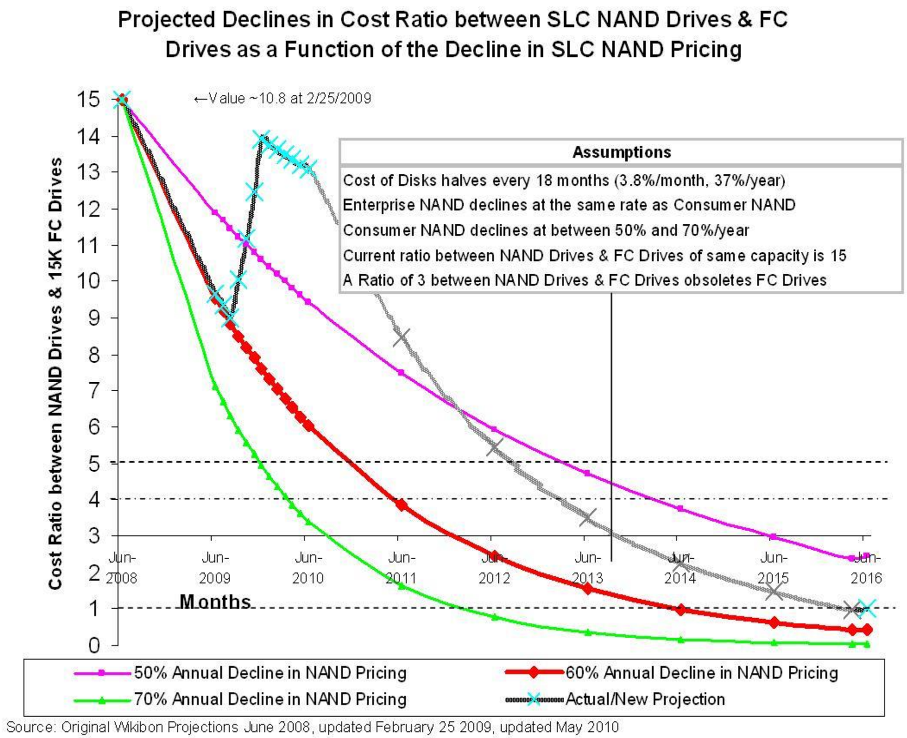
*http://wikibon.org/wiki/v/Flash_and_Hyperscale_Changing_Database_and_System_Design_Forever*

# Agenda: Second Generation Flash Architectures

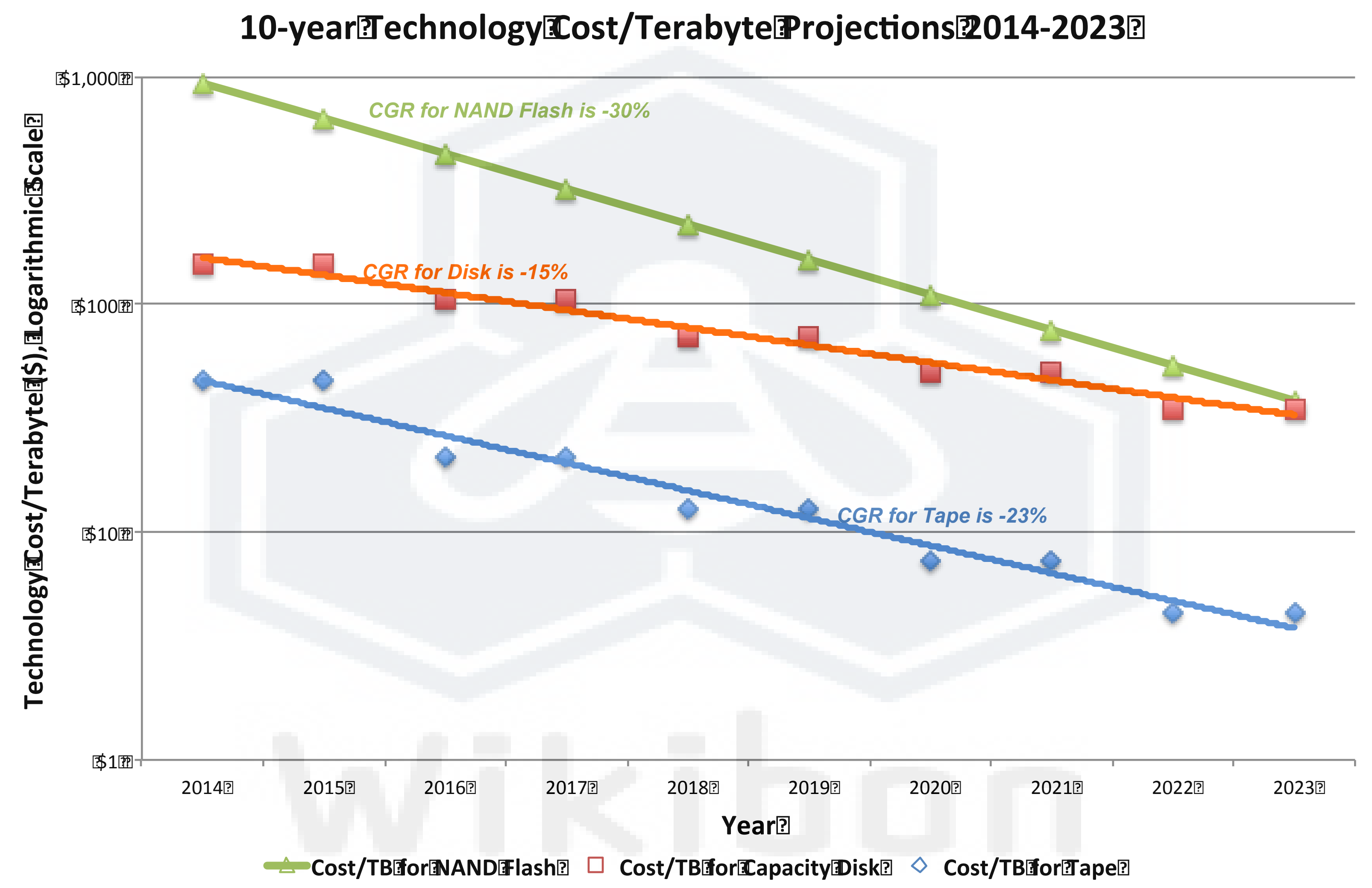
- Flash vs. HDD Comparison
- Impact of Response Time on People Efficiency
- Impact of Response Time on System Efficiency
- **Impact of Data Reduction & Data Sharing on Cost**
- Flash Enabled Application Design
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management Requirements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

# Wikibon 2009/2010 Flash Forecasts



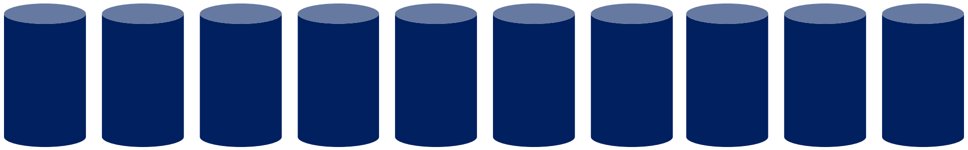
# 10-year Technology Cost/TB Projections




Source: © Wikibon 2014, from Numerous Sources including Analysts, Consultants, IBM & Oracle.

# Copy Management

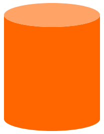
Large Independent Caching



**Traditional Disk Array**

***90% of Data is a Copy of Original data***

Small Shared Cache



**All Flash Array**

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

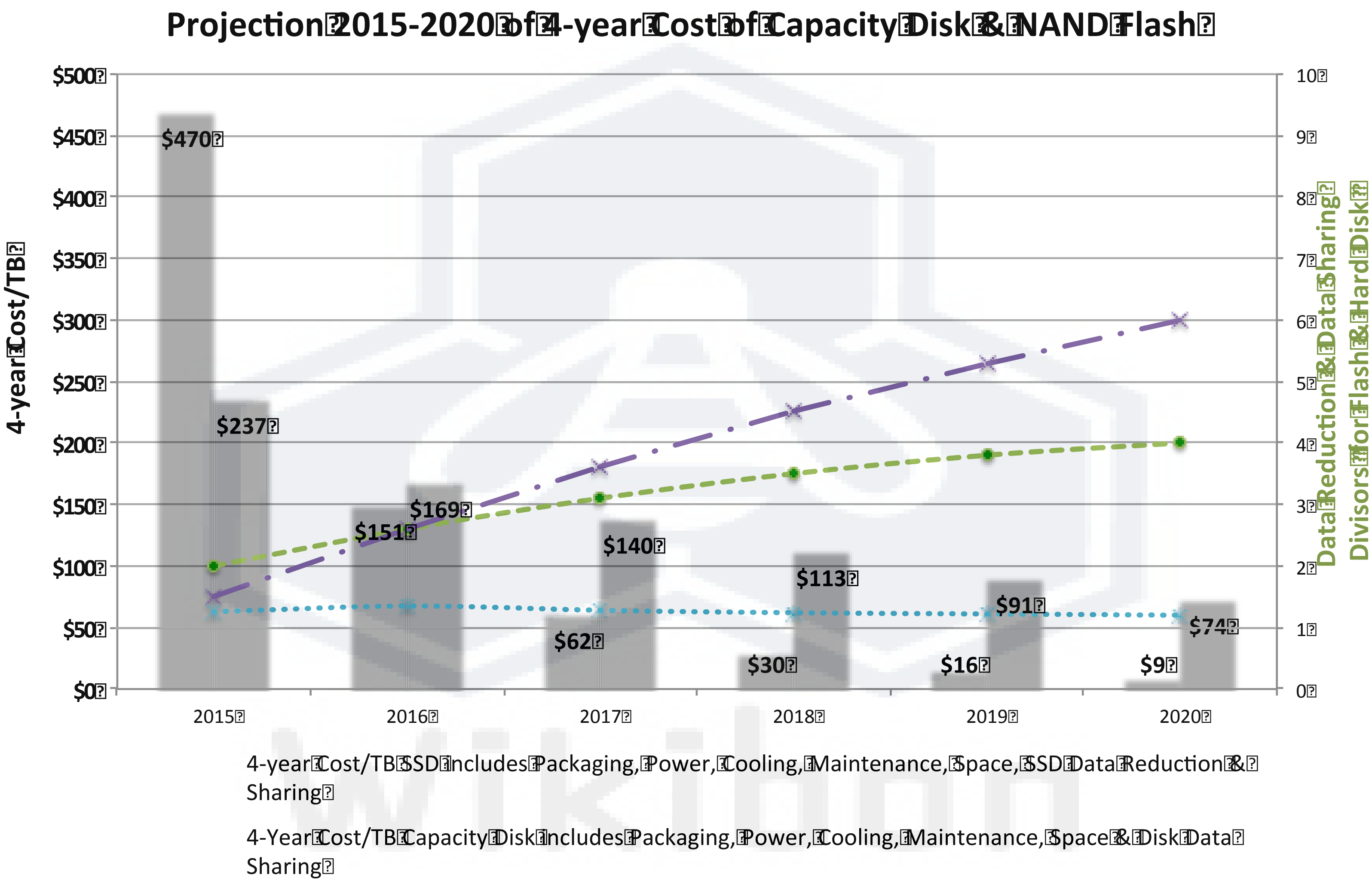
***Action: Measure & Minimize # Physical Copies of Data***

# 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***

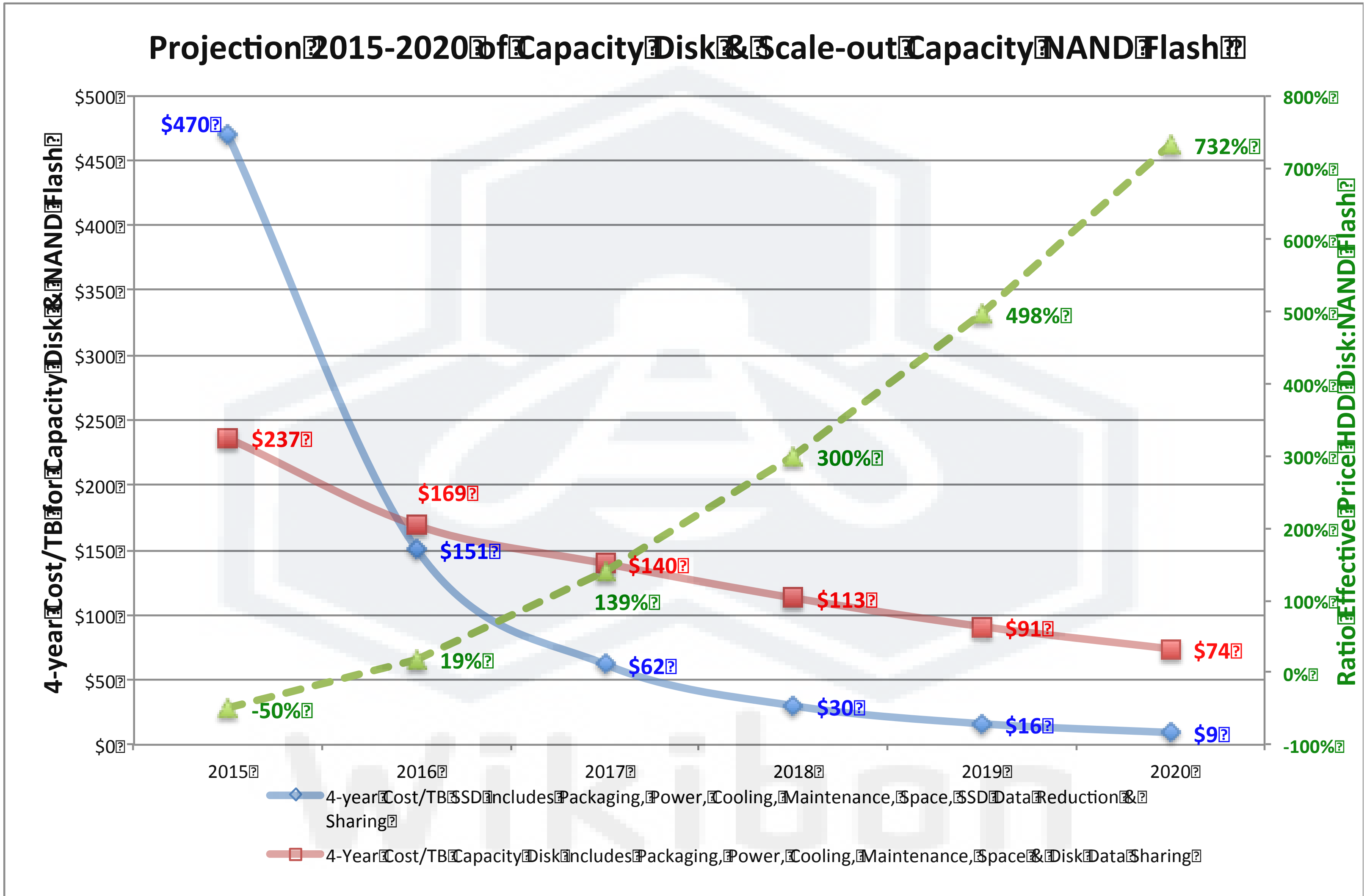
# Infrastructure Costs by Technology




*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



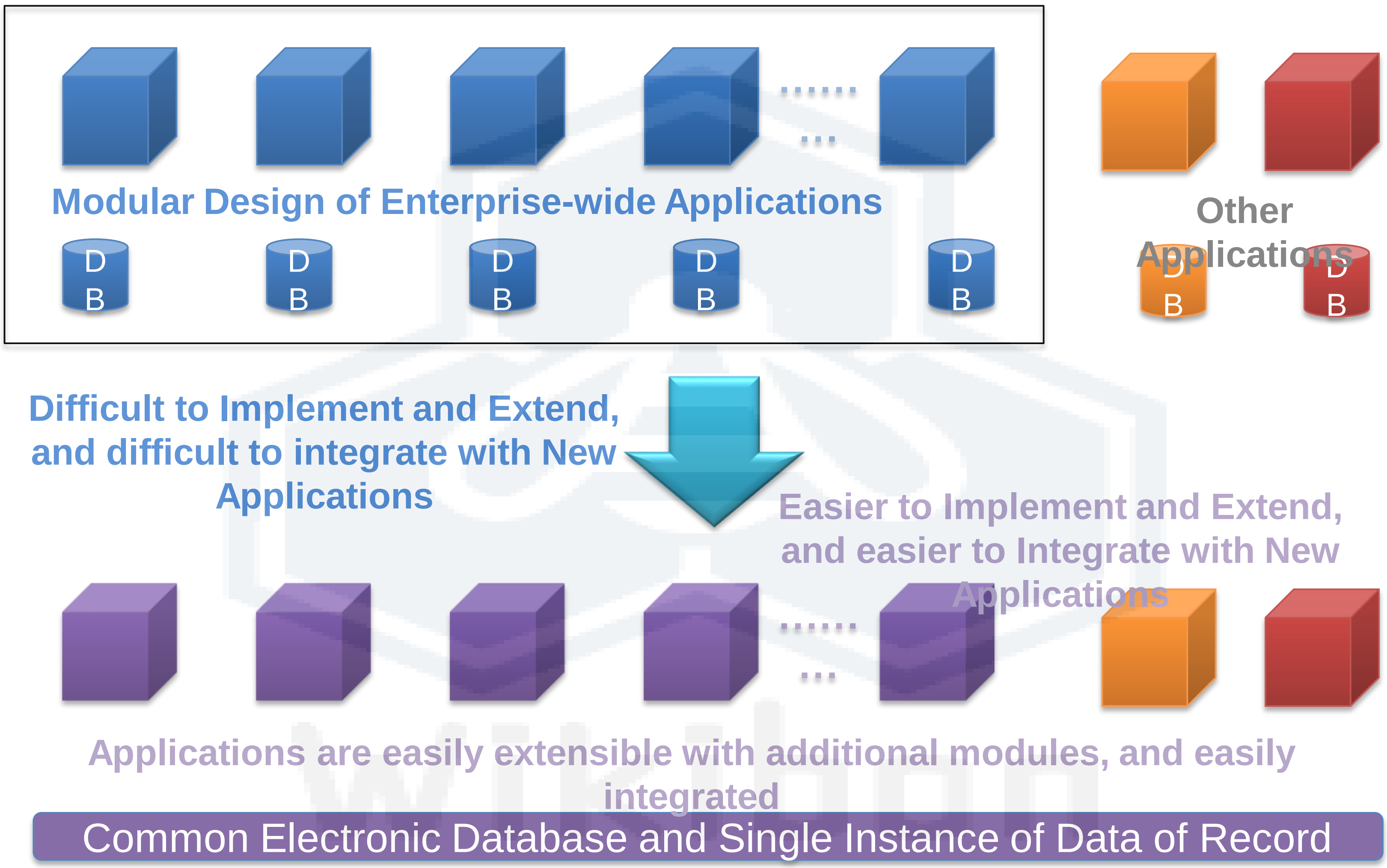

*Source: © Wikibon 2015. 4-Year Cost/TB Magnetic Disk & SSD, including Packaging, Power, Maintenance, Space, Data Reduction & Data Sharing*

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

# Agenda: Second Generation Flash Architectures

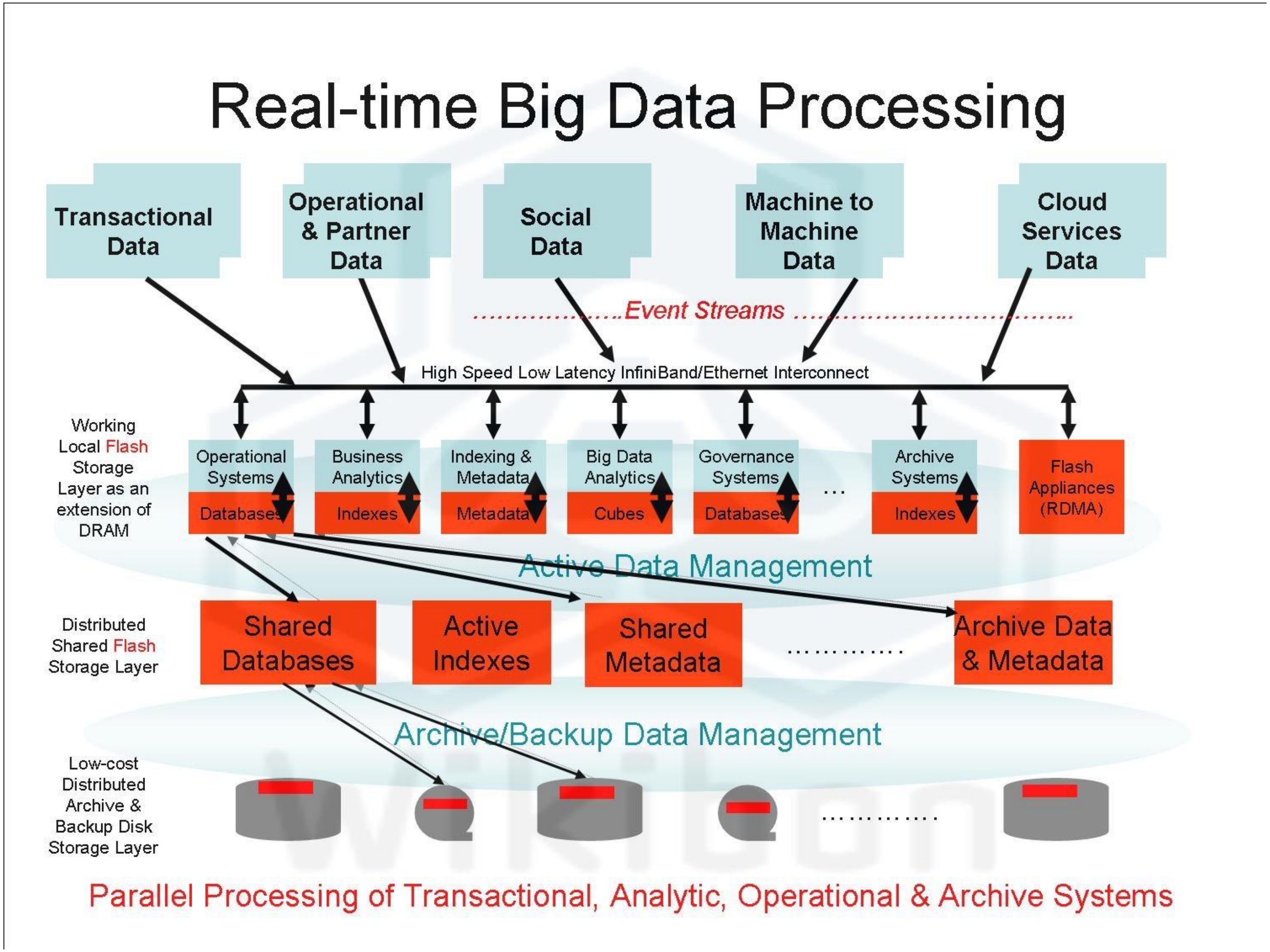
- Flash vs. HDD Comparison
- Impact of Response Time on People Efficiency
- Impact of Response Time on System Efficiency
- Impact of Data Reduction & Data Sharing on Cost
- **Flash Enabled Application Design**
- First Generation AFA
- Architectural Requirements for New Generation AFAs
- Management Requirements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

# Flash-enabled Application Design

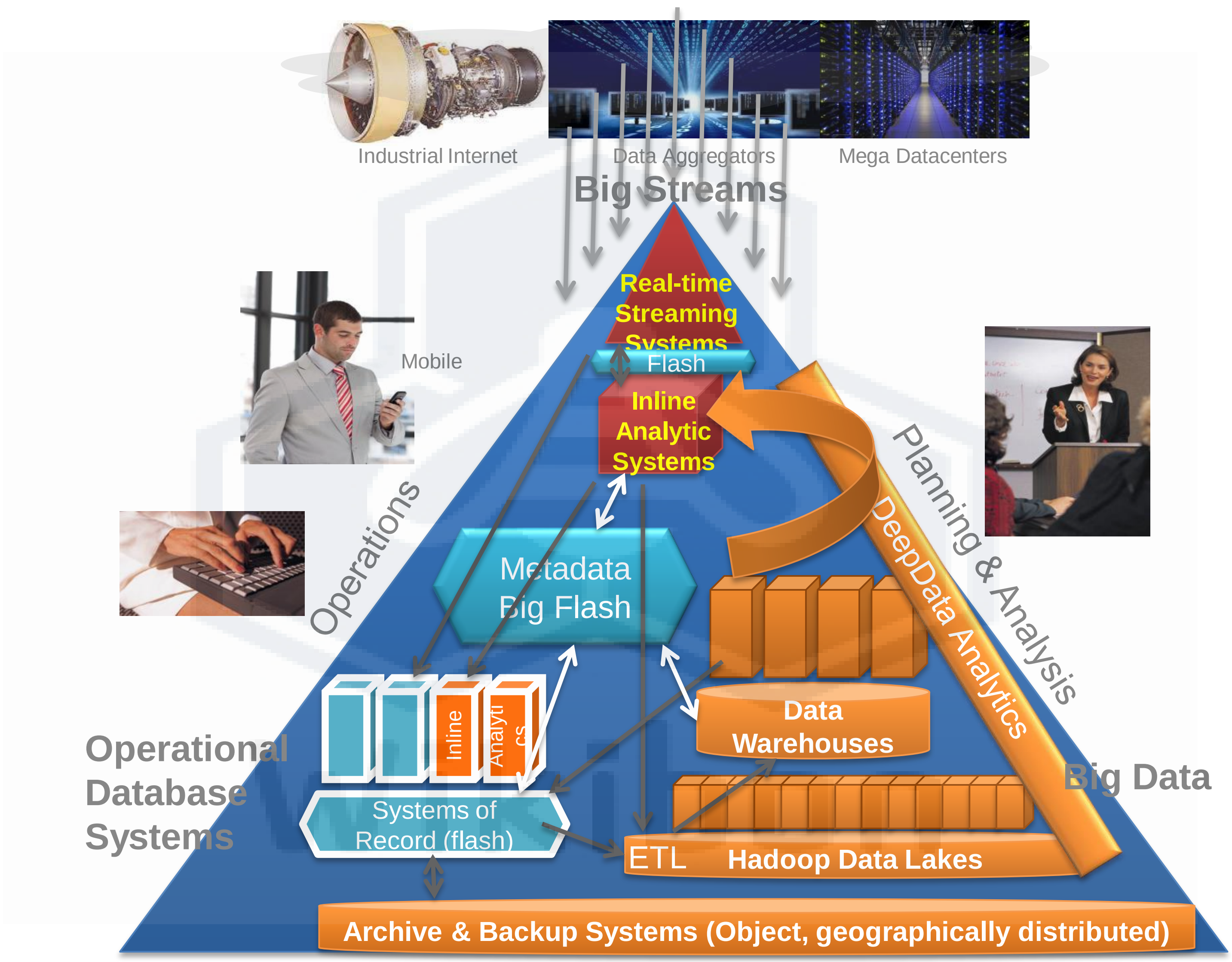



*©Wikibon, 2015*

# 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

# Infrastructure Costs by Technology (No Copy)

**Worldwide All-Flash Array Revenue by Vendor, 1H 2014**

| Vendor | Revenue Jan-June 2014 ($M) | Capacity Jan-June 2014 (TB) | Revenue Share (%) | $/GB | Scale-out | De-Duplication | Compression |
|---|---|---|---|---|---|---|---|
| EMC XtremIO | $112 | 13,405 | 23% | $8.4 | Y | Y | Y |
| Pure Storage | $91 | 7,558 | 18% | $12.0 | N | Y | Y |
| IBM FlashSystems | $83 | 22,773 | 17% | $3.6 | N | N | Y |
| NetApp EF550 | $45 | 5,500 | 9% | $8.2 | N | N | N |
| SolidFire | $36 | 7,526 | 7% | $4.7 | Y | Y | Y |
| Nimbus Data | $34 | 7,501 | 7% | $4.6 | N | Y | N |
| Other | $95 | 19,214 | 19% | $5.0 | N* | | |
| Total | $496 | 83,476 | 100% | $5.9 | | | |

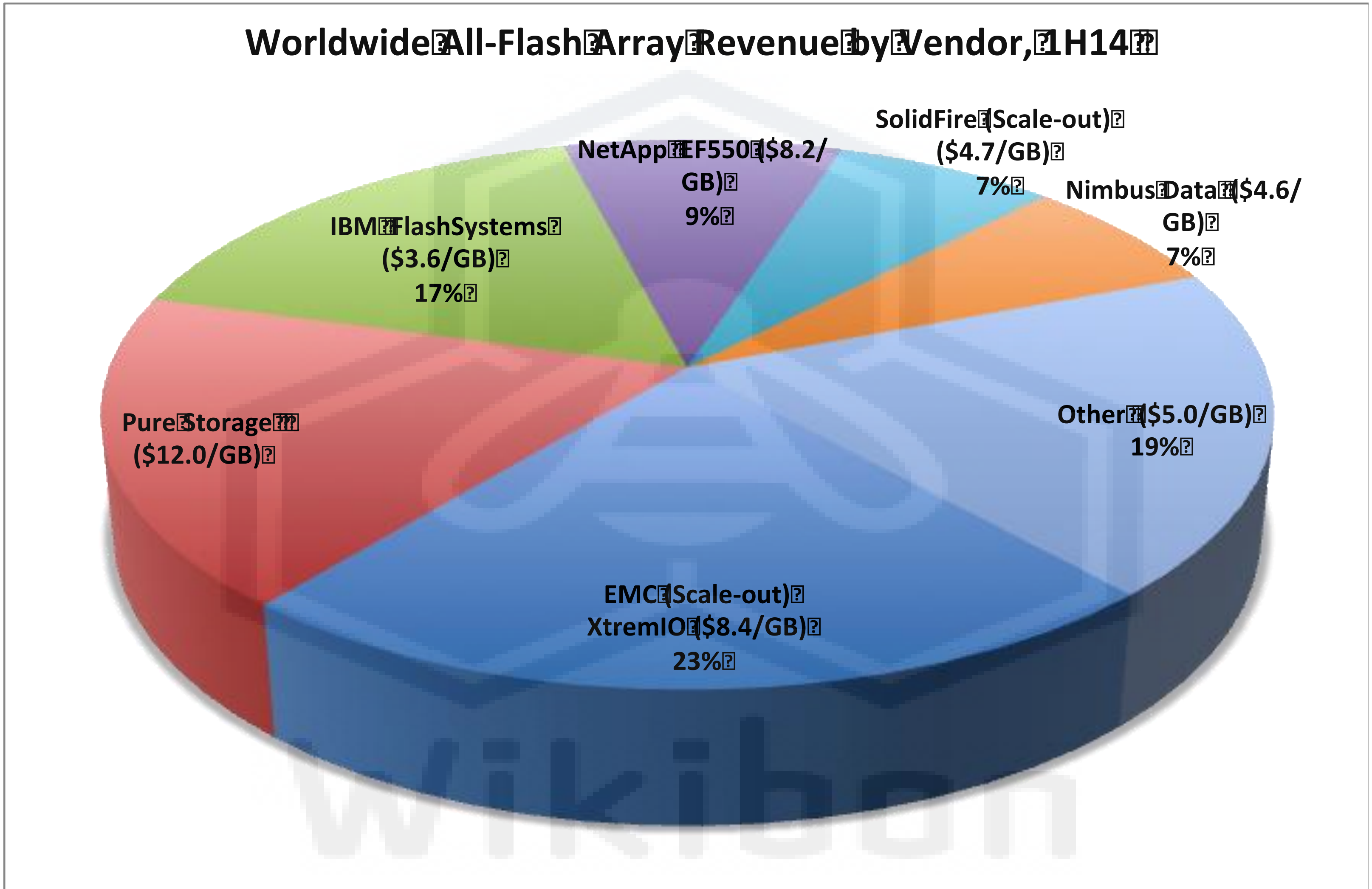
*Source: IDC, 2014 (Report # 252304e, Wikibon Analysis on Tables 1 & 2)*

** All other all-flash arrays are dual controler with the exception of Kaminario, which is scale-out.*

*Notes: Data includes the value of the entire system but excludes channel markup. Texas Memory Systems moved from the "other" category to IBM during CY13.*

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

# Infrastructure Costs by Technology (No Copy)




*Source: IDC, 2014 (Report # 252304e, Wikibon Analysis on Tables 1 & 2). See Table Footnotes-2 in Footnotes below.*

# 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

| | $/Usable TB without DRe | Data Reduction Ratio (DRe) | Number of Copies | $/Usable DRe |
|---|---|---|---|---|
| Cost of Capacity Flash AFA without DRe | $900 | 1 | 1 | $900 |
| Cost of Tier 1 Disk | $1,700 | 1 | 1 | $1,700 |
| Cost of Tier 1 Flash Tiering | $8,000 | 1 | 1 | $8,000 |
| Cost of AFA without DRe Function | $10,000 | 1 | 2 | $5,000 |
| Cost of AFA with DRe Function | $15,000 | 4 | 4 | $938 |
| Cost Very Low Latency Flash without DRe | $16,500 | 1 | 1 | $16,500 |

# Appendix II: Storage Cost Assumptions

| Assumptions for Maintenance, Power, Cooling & Space |
|---|
| Cost of Power is $0.12/kWhour |
| Cooling & power distribution cost is equal to twice equipment power draw |
| Cost of power, cooling & space for disk is 12% of acquisition cost of disk for 4 years |
| Cost of power, cooling & space for flash is 10% of disk power, cooling & space costs |
| Maintenance for disk is 18% of acquisition cost of flash for four years |
| Maintenance of flash is 10% of acquisition cost of flash for four years, reducing by 1%/year and stabilizing at 5% |
| Data reduction divisor & data sharing divisor for scale-out flash are averages for all data |
| Data reduction divisor for disk is average for all data. |
| *Source: Wikibon 2014* |