IBM DS8870 and High-End Storage Overview

Copenhagen
4th November, 2014
Stockholm
6th November, 2014

Josh Krischer & Associates
Growth in capacity is, on average, is ca. 35-40 percent per year, but the disk-technology development pace is fast enough to answer the requirements for capacity and performance.

Growth in storage capacity for System z is much smaller than for other platforms.

The biggest growth in storage is in so called CAMS – Cloud, Analytics, Mobile, and Social workload requirements.

The Storage product revenue is slightly dropping, mainly due to deployment of Storage Economic features such as de-duplication, compression, thin-provisioning and storage tiering.

Storage services and Storage cloud revenue is raising.

The enterprise SSD market is expected to double from $4 billion in 2013 to $8 billion in 2017. Storage Tiering (and Flash storage) with Automated Data Placement is gaining momentum.

Encryption of data at rest becomes necessity.

Demands for disaster recovery tools and services are increasing.

Technology has become complex and specialized.

IBM is leading in delivering functionality for all storage portfolio.

The previous EMC VMAX generations (10K/20K/40K) are retained because VMAX³ has no mainframe support or protocol support for FCoE, iSCSI (and obviously FICON) in the first iteration.
The external disk storage systems market in Europe, the Middle East, and Africa (EMEA) increased 2.6% year on year in terms of user value in 2Q14, according to the *EMEA Quarterly Disk Storage Systems Tracker* from IDC.

<table>
<thead>
<tr>
<th>Vendor</th>
<th>2Q13</th>
<th>Market Share</th>
<th>2Q14</th>
<th>Market Share</th>
<th>2Q14 YoY Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMC</td>
<td>$497.33</td>
<td>27.62%</td>
<td>$493.52</td>
<td>26.72%</td>
<td>-0.77%</td>
</tr>
<tr>
<td>HP</td>
<td>$270.76</td>
<td>15.04%</td>
<td>$298.07</td>
<td>16.14%</td>
<td>10.08%</td>
</tr>
<tr>
<td>IBM</td>
<td>$251.61</td>
<td>13.98%</td>
<td>$272.87</td>
<td>14.77%</td>
<td>8.45%</td>
</tr>
<tr>
<td>NetApp</td>
<td>$259.99</td>
<td>14.44%</td>
<td>$264.30</td>
<td>14.31%</td>
<td>1.66%</td>
</tr>
<tr>
<td>Dell</td>
<td>$120.92</td>
<td>6.72%</td>
<td>$122.32</td>
<td>6.62%</td>
<td>1.16%</td>
</tr>
<tr>
<td>Other suppliers</td>
<td>$399.80</td>
<td>22.21%</td>
<td>$395.87</td>
<td>21.43%</td>
<td>-0.98%</td>
</tr>
<tr>
<td><strong>Grand total</strong></td>
<td>$1,800.41</td>
<td>100.00%</td>
<td>$1,846.94</td>
<td>100.00%</td>
<td>2.58%</td>
</tr>
</tbody>
</table>
“Western Europe's return to growth over the quarter (+2.3%)”

“However, growth remains mixed across vendors and countries, with IBM's strong performance (mainly due to a revival in its DS8000 series) accounting for most of the growth in the sub-region.”
What are users‘ Requirements from Vendors?

- Technology has become complex and specialized which request specialized skills for storage, server, networks, hypervisor, ERP and data base administration
- Customers require from vendors to sell solution and not pieces of equipment or software
- A solution means integration, installation of the applications and management tools such as infrastructure management, monitoring, and reporting
- Vendors are requested to offer innovative financial proposal for the equipment, software, middleware applications and maintenance
- A vendor which can provide the above has significant advantage in comparison to vendors which can deliver only parts of it
What are users’ storage purchase criteria?

- Features & functions
- Reliability/availability
- Performance
- Price
- Service & support
- Vendor viability
- Past experience with the vendor & offering
- Vendor trust/style, culture
- Vendor IT offerings portfolio
- Press, analyst recommendations
- References
- Fashion (everybody buys xyz...)
What are the Differences between Different Products?

- Functionality
- Remote Copy techniques
- RAS features
- Performance
- Number and type of host ports
- Service

Price (but not always the most expensive is the best and the cheap bad!)

All points above depend on CU units structure but structure itself as a selection criteria is irrelevant
<table>
<thead>
<tr>
<th></th>
<th>VSP</th>
<th>VMAX</th>
<th>DS8870</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Architecture</strong></td>
<td>massively parallel crossbar switch</td>
<td>loosely-coupled cluster</td>
<td>Symmetrical Multiprocessor</td>
</tr>
<tr>
<td><strong>Processor design levels</strong></td>
<td>3, Front-end, Back-end control</td>
<td>Each module with own FE &amp; BE</td>
<td>3; Front-end, SMP, Back-end</td>
</tr>
<tr>
<td><strong>Control processors</strong></td>
<td>2-4 quad Core Xeon</td>
<td>Xeon E5-2687-v2 2.7GHz 12-core per module</td>
<td>POWER7+ (4.228GHz) 2, 4, 8 and 16 cores</td>
</tr>
<tr>
<td><strong>Front, and Back-end</strong></td>
<td>quad Core Xeon ASICs</td>
<td></td>
<td>PowerPC and ASICs</td>
</tr>
</tbody>
</table>
Three Layer' Architecture

- **Layer 1: Up to 16 distributed PowerPC / ASIC Host Adapters**
  - Manage the 8Gbps Fibre Channel host I/O protocol to servers and perform data replication to remote DS8000s

- **Layer 2: Centralized Power 7+ Servers**
  - Two symmetric multiprocessing processor (SMP) complexes performing up to 256 concurrent operations, manage two monolithic data caches, and advanced functions

- **Layer 3: Up to 16 distributed PowerPC / ASIC Device Adapters**
  - Manage the 8Gbps Fibre Channel interfaces to internal storage devices, SAS paths, RAID protection, and drive sparing
DS8870 Block Diagram

Intelligent SAN/Host Adapters

Enterprise Controller Hardware

RAID adapters

High Bandwidth, Fault Tolerant Interconnect

N-way SMP

Cache memory

Persistent memory

RAID Adapters

High Speed PCIe IO Fabric

Host Adapters

Host Adapters

Host Adapters

Host Adapters

Host Adapters

RAID Adapters

RAID Adapters

© Josh Krischer & Associates
EMC VMAX Hardware

**VMAX 20K**
- 2.3 GHz Xeon (Harpertown)
- 16 CPU cores
- 128GB Cache Memory (Max.)
- Dual Virtual Matrix
- PCIe Gen1

**VMAX 40K**
- 2.8 GHz Xeon w/turbo (Westmere)
- 24 CPU cores
- 256GB Cache Memory (Max.)
- Quad Virtual Matrix
- PCIe Gen2

© Josh Krischer & Associates
HDS VSP structure (hp XP 9500)

Any-to-Any

• 8 FEDs
• 4 BEDs

HiStar-E Network
4 PCIe Grid Switches (96 ports)

• 8 Cache
• 16 CPU Cores

VSP Single Chassis - Boards

Source: HDS
## Cache

<table>
<thead>
<tr>
<th></th>
<th>VSP</th>
<th>V-MAX</th>
<th>DS8870</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Architecture</strong></td>
<td>Global + Control cache</td>
<td>Distributed cache</td>
<td>allocated as part of the p570 server memory</td>
</tr>
<tr>
<td><strong>Cache levels</strong></td>
<td>1</td>
<td>1</td>
<td>3; L1, L2, Server memory</td>
</tr>
<tr>
<td><strong>Write protection</strong></td>
<td>Allocated from global</td>
<td>Fully mirrored</td>
<td>NVS in other side of the cluster</td>
</tr>
<tr>
<td><strong>Control data</strong></td>
<td>Control cache</td>
<td>Part of the cache</td>
<td>outside</td>
</tr>
<tr>
<td><strong>Effective cache size</strong></td>
<td>Installed</td>
<td>Less than 50%</td>
<td>Equals installed - NVS</td>
</tr>
<tr>
<td><strong>Cache design</strong></td>
<td>Dynamic</td>
<td>Static</td>
<td>Dynamic</td>
</tr>
<tr>
<td><strong>Cache segment</strong></td>
<td>32KB</td>
<td>64K</td>
<td>4KB</td>
</tr>
</tbody>
</table>

© Josh Krischer & Associates
Tightly Coupled, Cache Centric CU Structure

Cache-Centric any-to-any connection

Hosts FC, iSCSI, FICON

Cache

BE

HDDs, SSDs

Cache

Cache

Hosts FC, iSCSI, FICON

Cache

Cache

BE

HDDs, SSDs
Clustered Modules based CU (Bus)

- RapidIO or Ethernet Switch

Diagram shows clusters of modules with interfaces and caches.
DS8870 Functionality Advantages

- Functionality
- Data Mirroring techniques
- GDPS full support
- HyperSwap
- GDPS vs. GDDR
- Open HyperSwap for System p
- Easy configuration in comparison to “BIN” files
- Performance and throughput
- Encryption
The DS8870 supports Advanced Functionality

- Thin Provisioning
- Easy Tiering
- Storage Pool Striping
- Full Drive Encryption (FDE)
- I/O Priority Manager
- Space Saving Mirroring - Snapshot
- Synchronous Remote Copy
- Asynchronous Remote Copy

© Josh Krischer & Associates
Synergy Between DS8870 and the System z

- **High Performance FICON (zHPF)** – basic and extended

- **I/O Priority Manager** supports "importance" and "achievement" information provided by z/OS Workload Manager to manage execution priorities. zWLM can manage host and storage resources end-to-end to optimize the workload execution based on the specified Service Class.

- **Performance – DB2** Specialized cache algorithm can optimize DB2 list prefetch operations by multiple, parallel data fetches. DB2 List Prefetch Optimizer with zHPF, enables the storage system to read all 32 DB2 pages in parallel, transferring all 32 pages back to the host in a single exchange. At any point in time, DB2 has two List Prefetch I/Os outstanding; therefore the DS8870 is always reading 64 DB2 pages in parallel.

- **Performance – IMS** provides enhanced performance for IMS write-ahead data set (WADS).

- **Performance – zDAC** supports optimization to improve performance of z/OS Discovery and AutoConfiguration (zDAC).

- **Volume Management** supports dynamic volume expansion for standard (thick) 3390 volumes, Extended Address Volumes (EAV) – supports 3390 volumes up to 1 TB capacity
Disorganized Index Scans with hot cache, 4K pages

- DB2 9 throughput increases list prefetch by 43%
- DB2 10 throughput increases list prefetch by 111%
- Together DB2 10 zHPF is 11 times faster than DB2 9 FICON
# EMC VMAX 40K and Hitachi VSP Support Matrix (October 14)

<table>
<thead>
<tr>
<th>Supported feature</th>
<th>EMC*</th>
<th>HDS/HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>HyperPAV</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>High Performance FICON (zHPF basic)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>zHPF – multitrack</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>zHPF – QSAM, BSAM, BPAM</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>zHPF – format writes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>zHPF – DB2 list prefetch cache optimization</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Performance – DB2 (cache algorithm)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Performance – IMS</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Volume Management</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Multiple Readers for z/GM</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Large 3390 Volumes /EAV 1 TB</td>
<td>No</td>
<td>NO, SOD</td>
</tr>
<tr>
<td>zWLM integration for I/O Priority Manager</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

*To the best of our knowledge
DS8000 series Local and Remote Copy Techniques

- **Local replications:**
  - FlashCopy and FlashCopy SE
  - Remote Pair FlashCopy (Preserve Mirror)

- **Remote Mirror and Copy:**
  - Metro Mirror (a.k.a. PPRC)
  - Global Copy
  - Global Mirror
  - Global Copy Metro/Global Mirror
  - z/OS Global Mirror (a.k.a. eXtended Remote Copy (XRC))
  - z/OS Metro/Global Mirror

Sources: The economics of IT risk and reputation—What business continuity and IT security really mean to your organization, IBM Global Study, 2013
The most commonly used GDPS implementation is GDPS/PPRC, which includes Parallel Sysplex, PPRC-compatible remote copy and automation technology.

Key functions of GDPS are to maintain data consistency and data integrity. If data is not consistent at the recovery site, a time-consuming recovery is usually required, which could take days.
Example of z/OS Metro/Global Mirror – a three-site configuration
Workload Characteristics

Transaction processing

- Small-block I/O requests, e.g. Oracle DB – 8KB
- OLTP, databases, mail servers majority of ERPs
- I/O response time (< 5ms e.g.) is critical
- Achieved by CU bandwidths, number and speed of disk drives
- SSD or FC/SAS disk drives preferred, small capacity, many actuators
- SATA disk drives not generally recommended
- SPC-1 standard industry benchmark

Batch, sequential

- Typical sequential, large-block I/O requests,
- HPC, seismic processing, data mining, streaming video applications, large file access, e.g. TSM 256 KB
- Backup/restore, batch jobs
- High data transfer rate required, I/O response time less crucial
- Achieved by sequential prefetching, data striping, front-end and back-end Bandwidth and
- SATA disk drives may be a suitable choice
- SPC-2 standard benchmark
Workload skew drives Easy Tier benefits

- 58% of the random IOPS and 33% of the MB from about 5% of the extents!
- 50% of the extents do 10% of the MB and virtually no random IOPS!
DS8870 Easy Tier, VMAX FAST VT

- CU ability to identify hot data
- Automatically migrate that data to and from solid-state and traditional spinning drives
- This automated data relocation can help optimize data placement across tiers of drives with different price and performance attributes
- Provide significant performance benefit to applications with relatively small investment in Flash storage
- Helping clients more effectively balance system price and performance.
- DS8870 – 1 GB blocks
- VMAX – FAST VP-360 MB blocks
Automated Data Placement example (DS8700)
Easy Tier 5 - Heat Map Transfer

Continuous Workload Monitoring

Workload Hotspot Analysis

Smart Data Placement

MM/GC/GM Replication

System Storage DS8900

© Josh Krischer & Associates
IBM DS8000 Family – Generations of Evolution

2004
POWER5
DS8300

2006
POWER5+
DS8300 Turbo

2009
POWER6
DS8700

2010
POWER6+
DS8800

2012
POWER7
DS8870

2013
POWER7+
DS8870

Replication and Microcode Compatibility
DS8000 Enterprise Storage Evolution

Disk: DS8300 (FC), DS8700 (FC), DS8800 (SAS), DS8870 (SAS)

Power: DS8300 (Bulk), DS8700 (Bulk), DS8800 (Bulk), DS8870 (DC-UPS)

CEC: DS8300 (p5/p5), DS8700 (p6), DS8800 (p6+), DS8870 (p7+)

IO Bay: DS8300 (RIOG), DS8700 (PCIE), DS8800 (PCIE), DS8870 (PCIE)

Adapters: 4Gb/2Gb, 4Gb/2Gb, 8Gb/8Gb, 8Gb/8Gb
Multiple Target Peer to Peer Remote Copy: for zero disruption

Multiple Incremental FlashCopy: additional data protection

High Performance Flash Enclosure: 2X flash for hybrid systems

Multi-thread Performance Accelerator: 35% IOPS improvement

Higher capacity, faster drives: 1.6TB flash & 600GB 15K rpm drives

Improved GUI: easier management

Synergy with System z advanced functions:
- Easy Tier 7th Gen for optimal placement of data
- zHyperWrite 43% reduction in DB2 log write time

Release 7.4 Content – GA Dec. 5th, 2014
Multi-target Peer to Peer Remote Copy and Multiple Incremental FlashCopy

- With DS8870 v7.4
- A volume is a source for (up to 12) multiple remote copy relations
- Two configuration options: one synchronous and one asynchronous relation
- Global Mirror or XRC (Global Mirror for z/OS on System z).
- >3 site support is expected to be available in the 1H15.
- Incremental re-synchronization between two targets
- Storage migration without disruption
- HyperSwap allowed to either secondary - 99.9999% availability with the HyperSwap
DS8870 Latest Enhancements – “the BlueHawk“

- New! DS8870 High Performance Flash Enclosure with Micro Tiering
- 1U Drawer with RAID 5 up to 30 x 1.8” SSD (9.2 TB netto)
- 4x faster flash performance in 50% less space than existing flash options
- Accelerate database performance by up to ×3.2
Hybrid DS8870 integrated with High Performance Flash Enclosure (HPFE)

- Performance optimized High Performance Flash enclosure
  - Enclosure is 1U, independently and concurrently installable
  - Higher performance than SSD as the High Performance Flash enclosure extends the PCIe fabric from I/O drawer to the enclosure
    - Directly connects to DS8000 internal PCIe Gen2 fabric via currently open slots in the I/O Bays
    - Each enclosure is it’s own DA pair
  - Enclosure contains up to 30 encryption capable eMLC Flash cards
    - Provides up to 12 TB raw / 9.2TB usable RAID5 capacity
  - Up to 4 enclosures in base (A) frame and first expansion (B) frame
    - Up to 96TB Raw (73TB Usable) Flash capacity in base (A) and expansion (B) frame

- Enterprise class reliability, availability and serviceability
  - RAID5 protected with two spares per enclosure
  - Redundancy provided with dual power supplies, adapters and fans
  - Components are concurrently replaceable

- Easy Tier enabled to exploit High Performance Flash

- Hybrid DS8870 allows up to 240 Flash cards + 1536 2.5” HDD/SSDs
DS8870 Hybrid Scalability with HPFE

High-Performance Flash Enclosure will be supported in base frame and first expansion frame for 4 additional HPFEs

- 240 2.5” drives
- 120 Flash drives

- 336 2.5” drives
- 120 Flash drives

- 480 2.5” drives

- 480 2.5” drives

8* 4/8 port 8Gb HA
16* 4 port 8Gb DA

12 3.5” Nearline drives replace 24 2.5” drives

© Josh Krischer & Associates
The DS8870 code and POWER7 and POWER7+ processors were optimized for greater IOPS performance.

Designed to deliver performance improvements for those analytic clients who desire greater IOPS performance.

**Benefits**

- Up to 35% improvement in maximum IOPS in random I/O workload environments*
- Analytic data processing clients can achieve up to 940,000 IOPS in Database Base Open (DBO) environments (70% read, 30% write, 50% hit ratio)

*All-flash HPFE configuration with Multi-thread Performance Accelerator
Improved DB2 log write performance with IBM zHyperWrite

- Improved DB2 Log Write Performance with DS8870 Metro Mirror - reduce latency by about 300μs
- Primary / Secondary HyperSwap enabled- Media Manager validates Primary / Secondary relationship exists

**Benefits**

- Reduce up to 43% of the DB2 log write time
- Avoids latency overhead of storage based synchronous mirroring
- Reduced DB2 write latency,
- Improved Log throughput
Easy Tier application integration with System z

- New API between System z and DS8870 Easy Tier
- Applications running on the System z send the heat map to the DS8870 system
- Removes requirement from application/administrator to manage hardware resources directly
- Common System z Use Cases
  - DB2 Reorganization and Defragmentation
  - Bulk dataset move and copy
- Benefits
  - Deploy storage more efficiently by enabling applications running on System z to direct more optimal placement of data
## DS8000 All Flash Comparison – OLTP

<table>
<thead>
<tr>
<th></th>
<th>1,056 15k drives (RAID 10)</th>
<th>240 flash drives</th>
<th>All-Flash Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Raw Capacity</strong></td>
<td>152TB</td>
<td>96TB</td>
<td><strong>36.8% Less</strong></td>
</tr>
<tr>
<td><strong>Usable Capacity</strong></td>
<td>72TB</td>
<td>73.6TB</td>
<td>Same</td>
</tr>
<tr>
<td><strong>Response Time</strong></td>
<td>~ 1 ms</td>
<td>~ 0.3 ms</td>
<td><strong>70% Less</strong></td>
</tr>
<tr>
<td><strong>Drive Count</strong></td>
<td>1,056</td>
<td>240</td>
<td><strong>77% Less</strong></td>
</tr>
<tr>
<td><strong>Frames</strong></td>
<td>3 frames</td>
<td>1 frame</td>
<td><strong>67% Less</strong></td>
</tr>
<tr>
<td><strong>Energy Usage</strong></td>
<td>13.9kw</td>
<td>5.3kw</td>
<td><strong>62% Less</strong></td>
</tr>
</tbody>
</table>

© Josh Krischer & Associates
DS8870 – All Flash System

DS8870 All Flash- Enterprise Class Functions & Performance!

- Initial offering: up to 8 Ultra-drawers (73TB usable)
- PCIe direct attached
- 8-Way 256GB Cache or 16 Way 512GB/1TB Cache
- Throughput IOPs > 640,000 70/30, up 1.4M IOPs Read
- 2H14- up to 266 TB usable
- Up to 50% reduction in floor space and 12% in Energy
DS8870 Advantages - summary

- More effective cache utilization
- Advanced Remote Mirroring techniques
- **Flexibility** - each port can also be set to either FCP or FICON
- **I/O Priority Manager** supports "importance" and "achievement" information provided by z/OS Workload Manager to manage execution priorities
- **Performance** – DB2 Specialized cache algorithm can optimize DB2 list prefetch operations by multiple, parallel data fetches (the DS8870 is always reading 64 DB2 pages in parallel)
- Advanced **zHPF** support
- **HyperSwap** for Systems z and p
- System z architecture ownership
- IBM is delivering advanced functionality faster than the competition
- The operation GUI, ported from XIV, delivers the most user-friendly functionality in the industry and is also shared by the complete portfolio of IBM storage solutions.
IBM innovation

- IBM leading recipient of US patents in 2012
  Record 6,478 patents
- IBM inventors have received more than 67,000 patents since 1993
- It is more than HP, Apple, Intel, Amazon, Oracle/SUN, Symantec, Accenture, and EMC together.
- The number 50 received 626 patents, EMC, VMware and other subsidiaries not in the top 50!

It's not just the $6B per year which IBM invest in the future that matters. It's their innovation and the access customers get to IBM's entire ecosystem that matters.
The DS8000 Series, 5 points to remember

The best business continuity and Disaster Recovery capabilities

Advanced technology and functionality

Specialized skills for storage, server, networks, hypervisor, ERP and data base administration

Total IT solution provider

The synergy with System z, p & i

IBM Global Financing
IBM Global Technology Services
Tivoli

© Josh Krischer & Associates
A selection of our coverage areas:

- Procurement & price evaluations
- Enterprise storage
- Mid-range storage
- Disaster recovery techniques
- Data center consolidation
- Data center design
- Mainframes

© Josh Krischer & Associates GmbH

josh@joshkrischer.com
Tel +49 6251 580626
Mobile + 49 172 6203272
Fax +49 6251 67952

http://www.joshkrischer.com