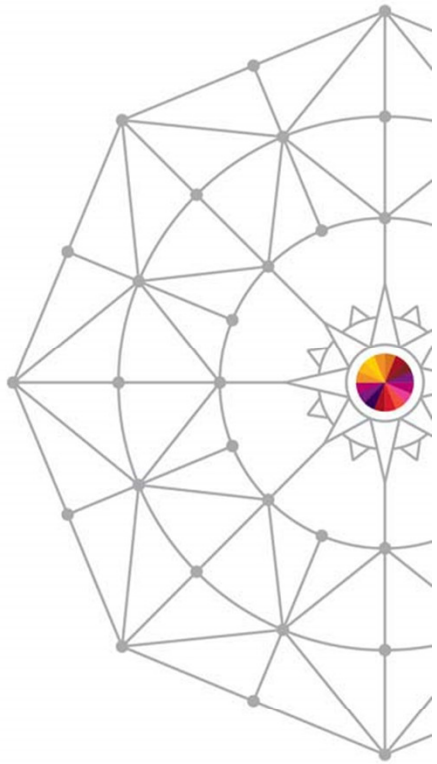
DS8000 and System z
Integrated by Design

Nick Clayton
Solution Architect for DS8000
IBM Systems and Technology Group
claytonn@uk.ibm.com



#SHAREorg

SHARE is an independent volunteer-run information technology association that provides education, professional networking and industry influence.

Agenda

- DS8000 and System z Synergy
- DS8870 review and update
  - High Performance Flash Enclosure
  - Easy Tier and z/OS
- Data replication and resiliency
  - Multi-target Metro Mirror
  - TPC-R and GDPS related enhancements
- Optimising z/OS performance
  - High Performance FICON
  - WLM and IO priority management
  - Managed File Transfer Enhancements
- Summary

System z and DS8000 Synergy

- zHPF Enhancements (now includes all z/OS DB2 I/O, BxAM/QSAM)
- Extended Distance FICON
- Caching Algorithms – AMP, ARC, WOW, 4K Cache Blocking
- DFSMS Recognition of SSDs and Storage Tiering
- Easy Tier – Easy Tier Application and Heat Map Transfer
- z/OS GM Multiple Reader Support
- SSDs + DFSMS + zHPF + HyperPAV + DB2
- zWLM + DS8000 I/O Priority Manager

Performance

- HyperPAV
- GDPS and GDOC Automation
- GDPS Active/Standby
- HyperSwap Technology Improvements
- Remote Pair FlashCopy and Enhancements
- zCDP for DB2, zCDP for IMS – Eliminating Backup windows

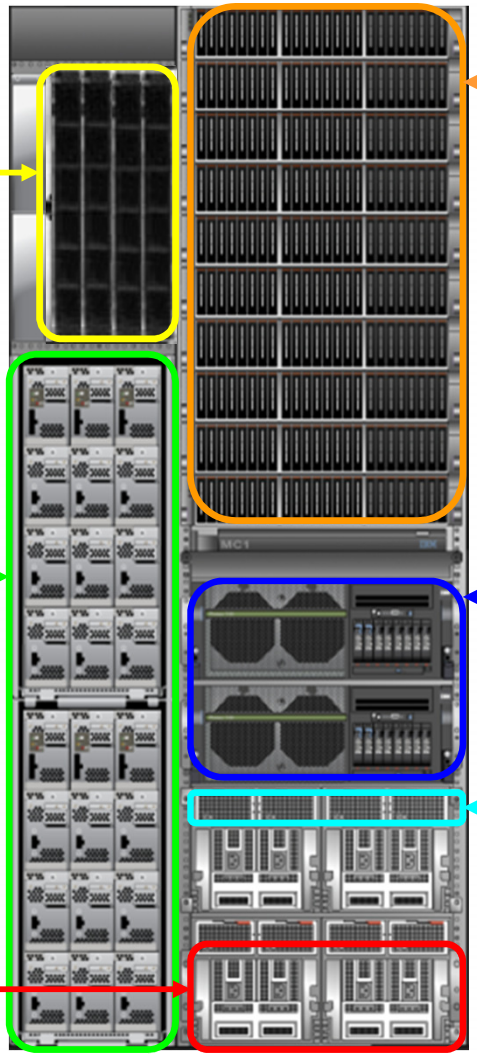
Availability

- 1 TB EAVs
- Quick Init for CKD Volumes
- Dynamic Volume Expansion
- Space Efficient FlashCopy
- z/OS Distributed Data Backup
- System z Discovery and Automatic Configuration (zDAC)
- Alt Subchannel Exploitation
- Disk Encryption
- TPC for Replication, GDPS

Management/Growth

DS8870 Hardware Review

High Performance Flash Enclosures provide Flash optimized RAID and significant performance improvements (new in DS8870 R7.4 - June 2014)



2.5” drives provide improved density, power efficiency and drive reliability (introduced on DS8800)

3.5” Nearline drives also supported.

DC-UPS power supply for significantly improved efficiency and serviceability – up to 20% reduction in power usage (introduced with DS8870)

POWER7+ processors with up to 32 cores and 1TB memory in total

PCI-e fabric provides improved performance, reliability and serviceability (introduced on DS8700)

Dual core 8Gb Host and Device Adapters for significantly improved performance (introduced with DS8800)

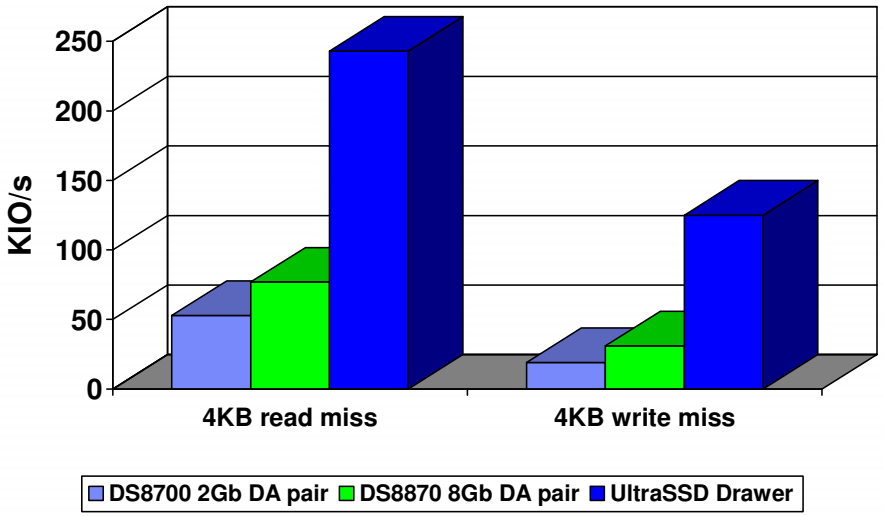
High Performance Flash Enclosure (HPFE)

New RAID Storage Enclosure

- 1U Storage enclosure with integrated RAID
- Developed in collaboration with POWER and Flash Systems
- Up to 30 400GB Flash cards (1.8” form factor) per enclosure
  - 12 TB raw capacity, up to 9.1 TiB usable
- RAID function based on IBM chip, optimized for Flash
- Integrated power and cooling
- Supports encryption

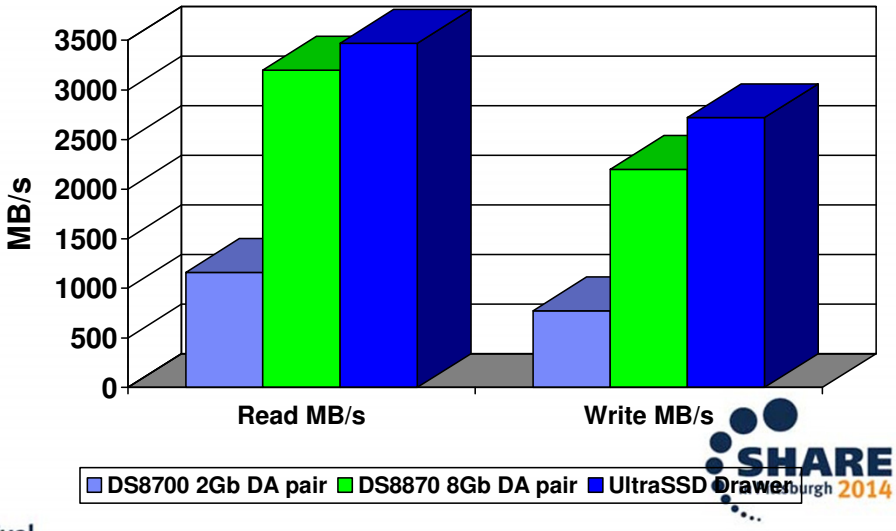
High Performance Flash Enclosure performance

Random IO



- Massive improvements in throughput for random workloads for High Performance Flash Enclosure compared to DS8870 device adapter
  - Up to 5.5x increase for random write
  - Up to 4x increase for random reads

Sequential IO



- Single High Performance Flash Enclosure has modest increase in sequential throughput capability compared to DS8870 device adapter
  - 16% for sequential read
  - 23% for sequential write
- DS8870 device adapter already provided significant sequential throughput improvements

DS8870 scalability with HPFE

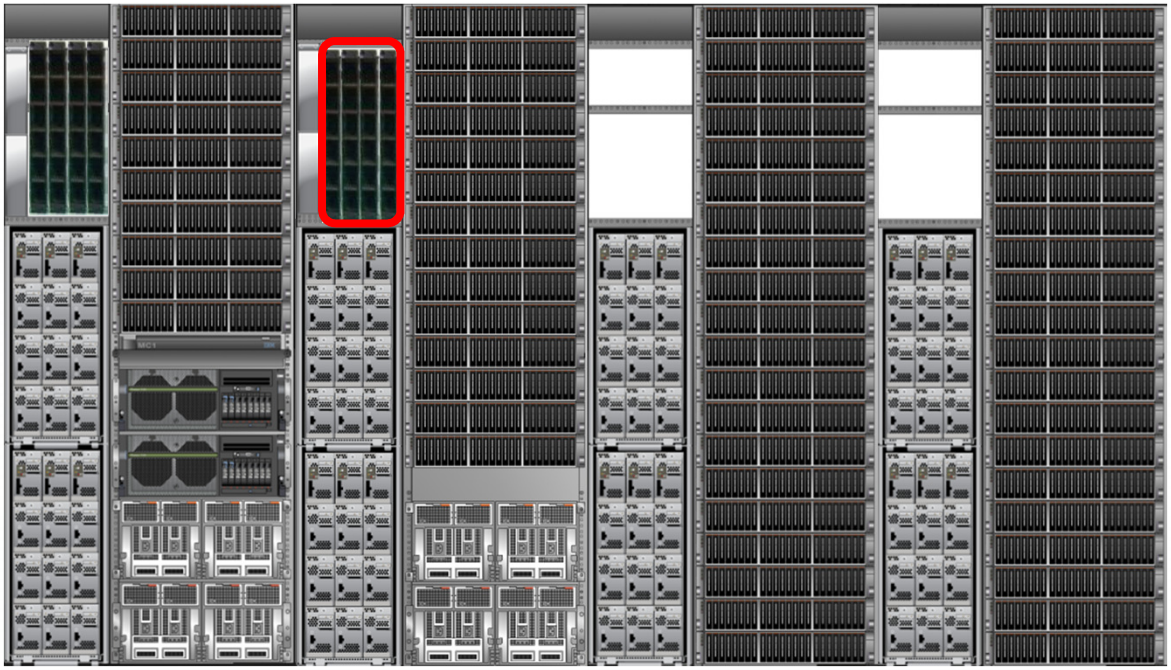
240 2.5” drives +
120 Flash drives

336 2.5” drives +
120 Flash drives

480 2.5” drives

480 2.5” drives

12 3.5” Nearline drives replace 24 2.5” drives



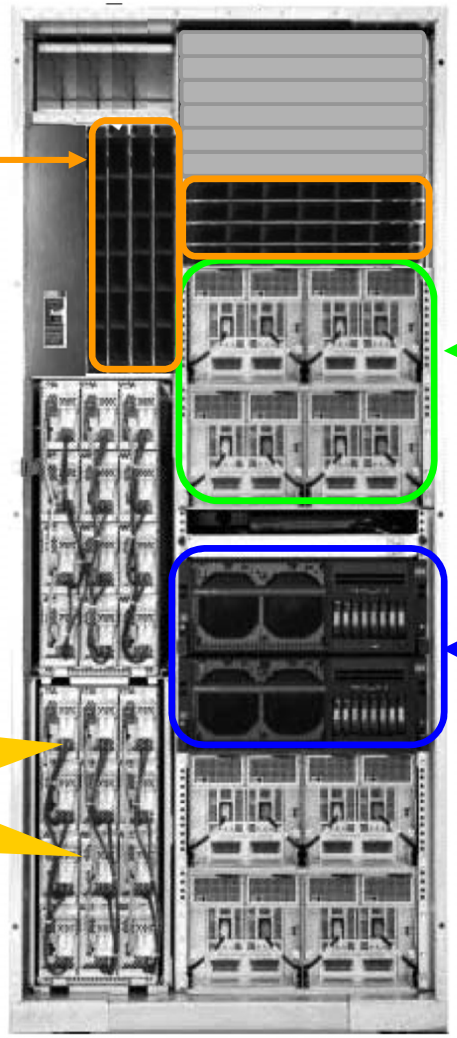
8* 4/8 port 8Gb HA
16* 4 port 8Gb DA

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

Statement of direction for support of 120 Flash drives in B frame later in 2014

All-flash DS8870 configuration

Up to 8 PCIe attached High Performance Flash Enclosures provide up to 96TB raw capacity with 400GB drives



All 8 IO bays installed in base frame for maximum throughput and host connectivity in single frame

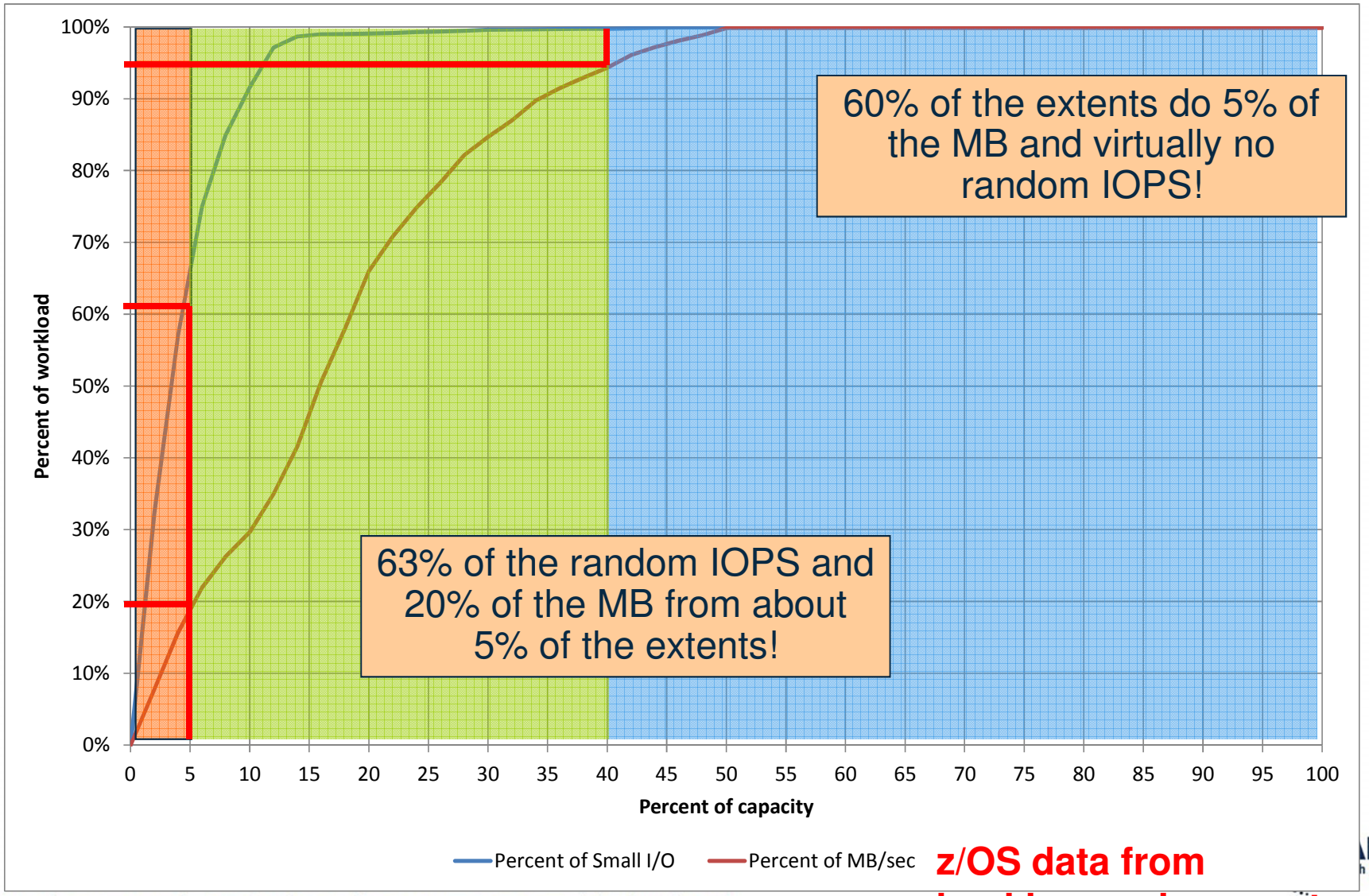
8-core 256GB cache or
16-core 512GB cache or
16-core 1024GB cache

67% reduction in footprint and 69% reduction in power compared to comparable throughput system with Enterprise drives

DS8870 Drive Technology

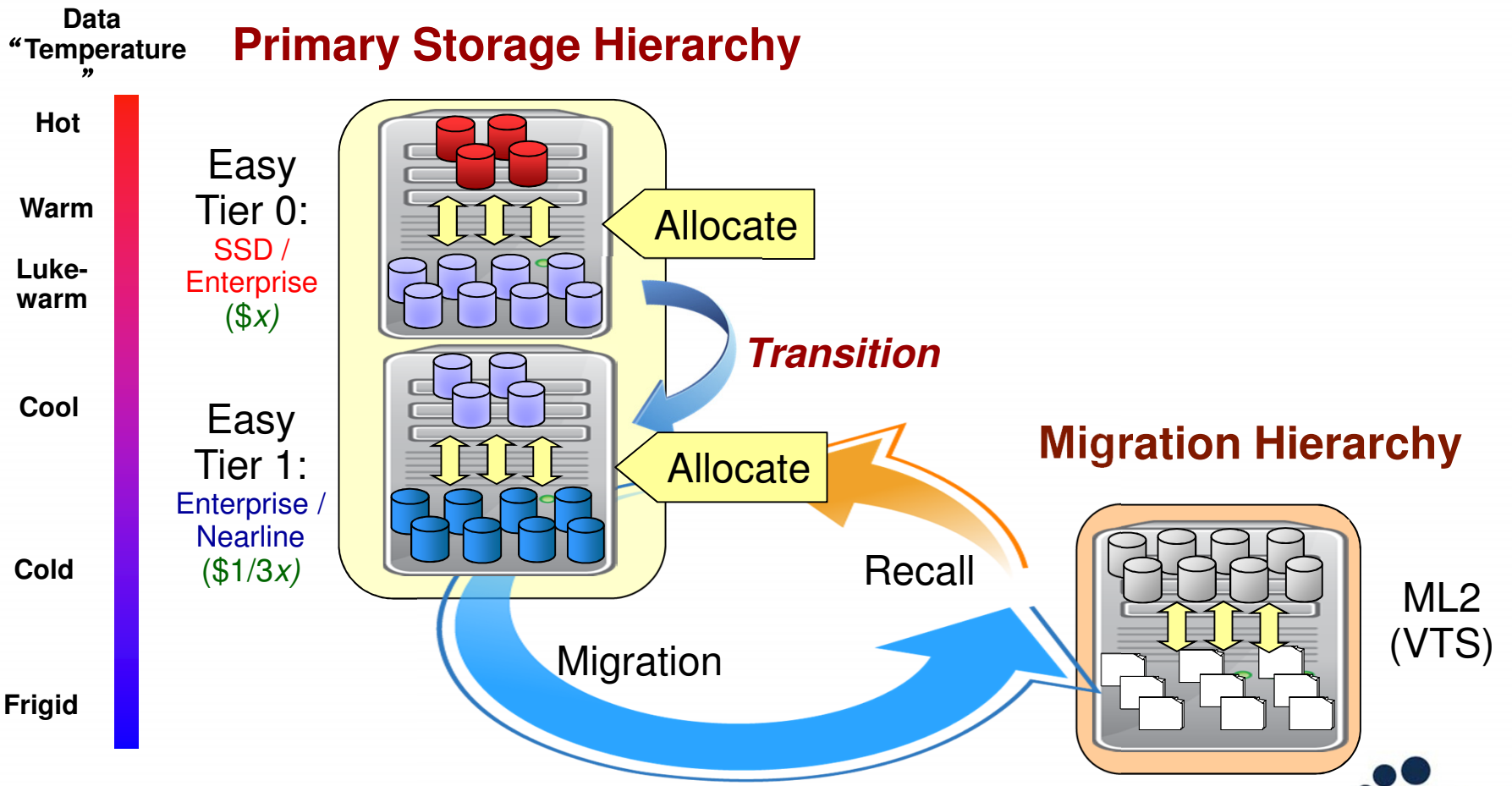
- Flash – 1.8” in High Performance Flash Enclosure
  - 400 GB drive
- SSD – 2.5” Small Form Factor
  - Latest generation with higher sequential bandwidth
  - 200/400/800GB SSD
- 2.5” Enterprise Class 15K RPM
  - Drive selection traditionally used for OLTP
  - 146GB and 300GB drives
- 2.5” Enterprise Class 10K RPM
  - Large capacity, much faster than Nearline
  - 600GB and 1.2TB drives
- 3.5” Nearline – 7200RPM Native SAS
  - Extremely high density, direct SAS interface
  - 4TB drives

z/OS workload skew and Easy Tier



z/OS data from banking environment

Storage Tiering for z/OS – Easy Tier, DFSMS and VTS



Easy Tier Application Technology Demonstration

zEC12 + DB2 for z/OS + DS8000 Directive Data Placement

Use Case

- DB2 for z/OS integration of Directive Data Placement capability in REORG
- Demonstrate DB2 is able to control the data placement of a Shadow data set based on the storage tier of a REORG Target data set

Value: Performance Assurance On Demand

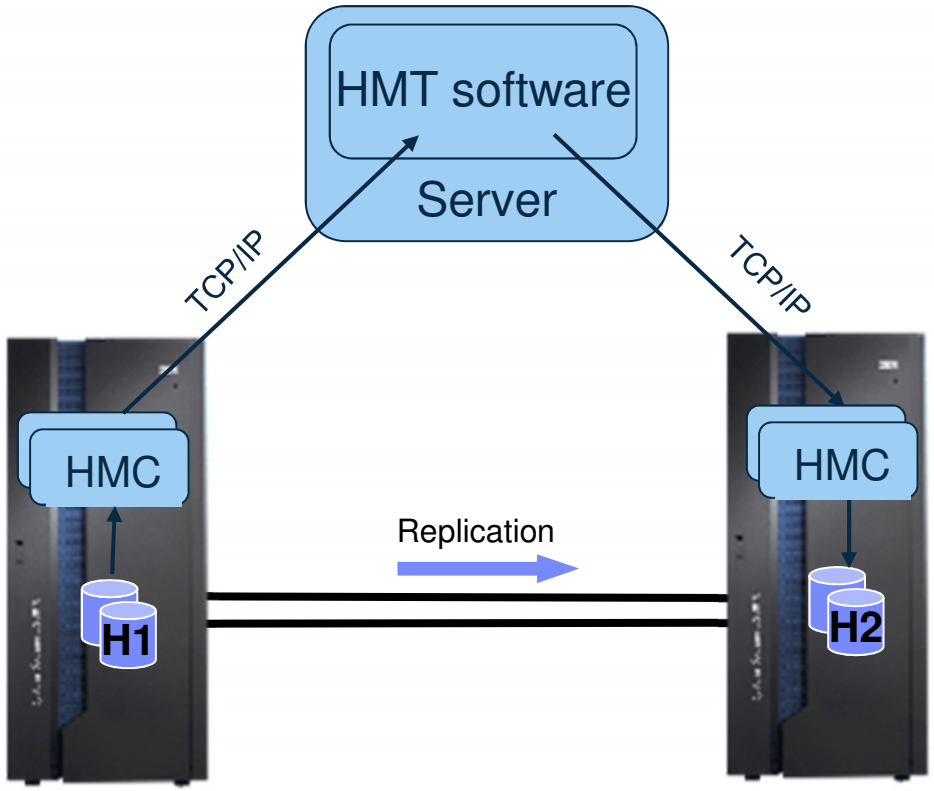
- Just-In-Time Data Placement
  - Based on unique knowledge about data processing task and data classification, application can timely choose data placement decision
  - Easy Tier function continues to manage the rest of data placement decisions in a coordinated fashion
- Storage Performance Assurance
  - Ensure a newly created Shadow dataset is placed at the same tier as the existing target dataset during a typical database operation such as REORG
  - Assure new dataset is performance optimized immediately after the reorganization

Demonstration Execution

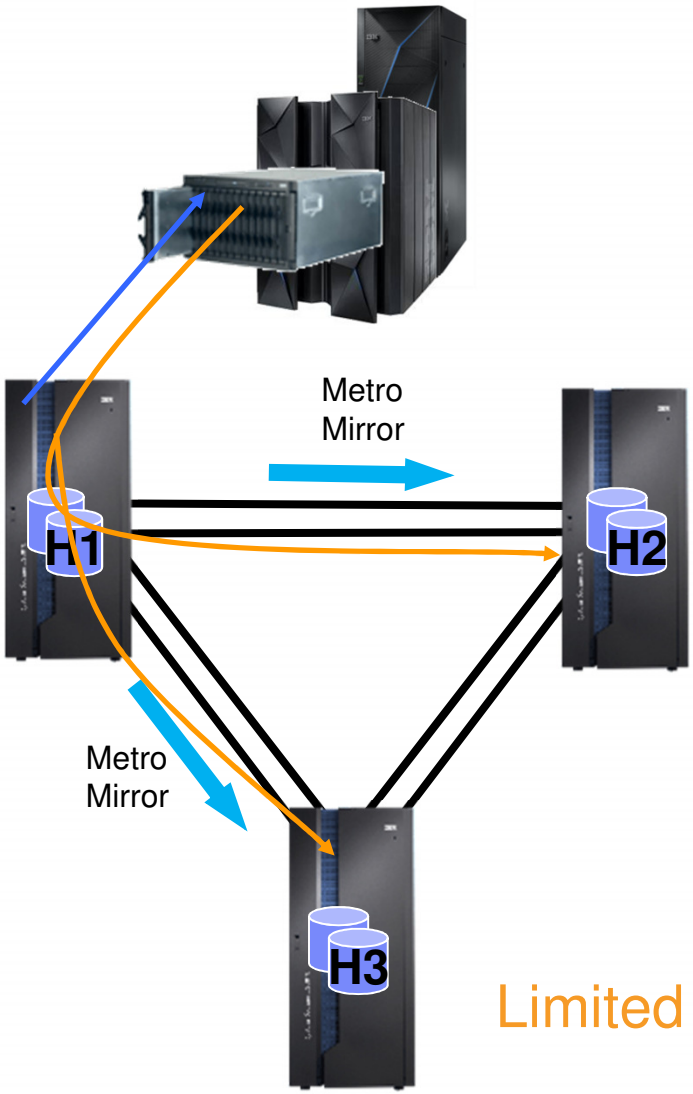
- Apply special DB2 utility, DFSMS, and storage prototype support in demonstration environment
- Perform DB2 for z/OS REORG
- Confirm the storage tier assigned for a newly reorganized data matches original data storage tier

Easy Tier HeatMap Transfer

- Provide capability to transfer Easy Tier HeatMap information for a volume from one disk subsystem to another
- Ensures that replication target is being optimised based on source workload to avoid period of learning when workload is switched
- Primary and secondary configurations are not required to be identical as each will optimised based on configuration and overall workload (local and transferred)
- Out of band software implementation provides flexibility and common approach across different replication types
- TPC-R and GDPS support as well as standalone utility



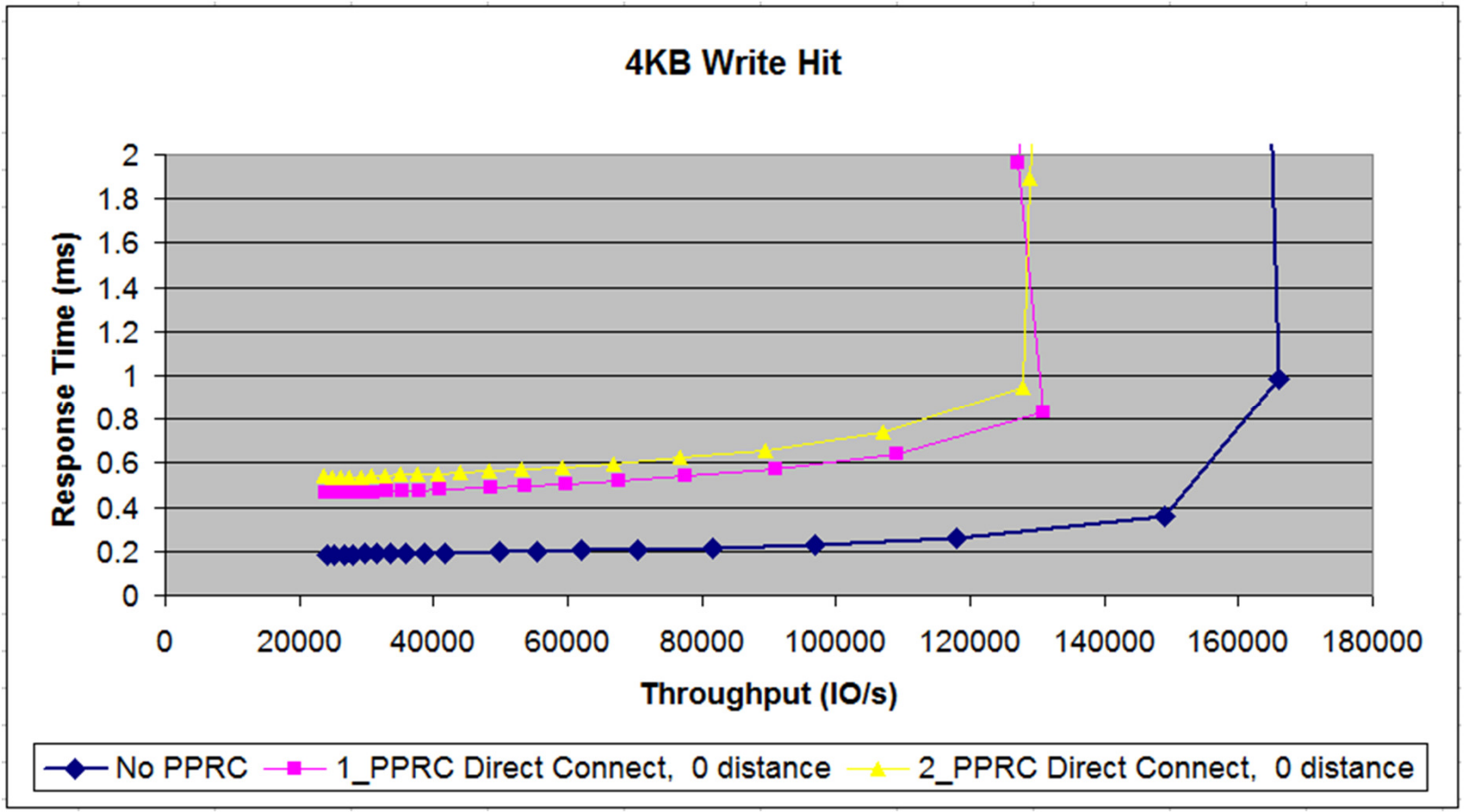
Multi-target Metro Mirror



- Allow a single volume to be the source for more than one PPRC relationship
- Initial support for two Metro Mirror and/or Global Copy relationships
- Statement of direction for Metro Global Mirror support in 2014
- Provides incremental resynchronisation functionality between target devices

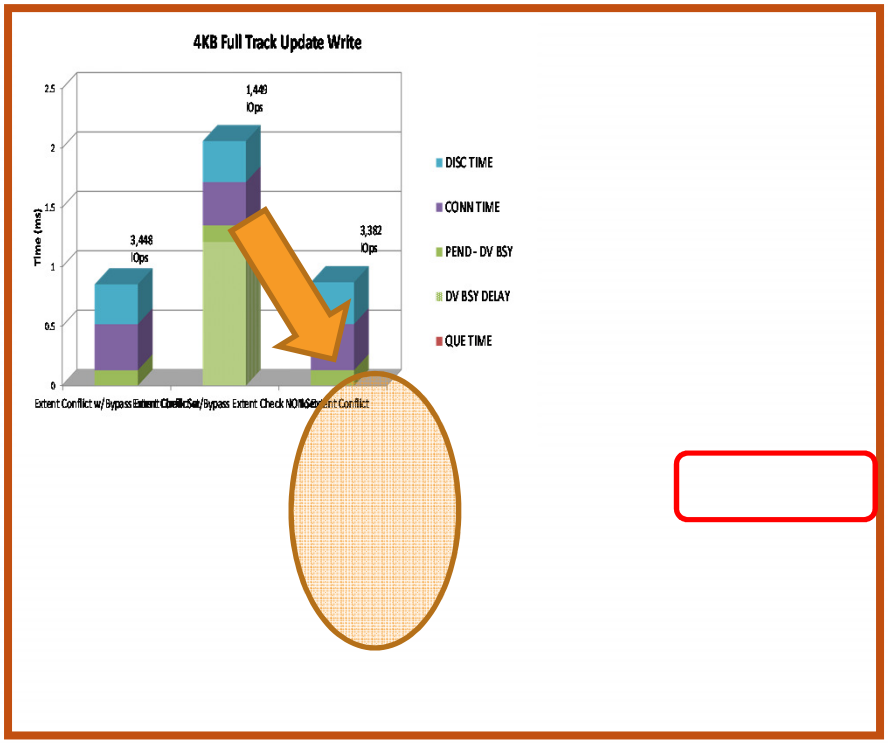
Limited Availability with Release 7.2.7

Multi-target Metro Mirror small write performance



Metro Mirror performance improvements

- z/OS uses the define extent command to control serialization of access to a dataset.
- Certain applications such as JES and DB2 use the bypass extent blocking feature as they have their own serialization
- With R7.2 DS8870 will honor the bypass extent blocking option in Metro Mirror environments
- Applies to specific z/OS workloads



Accelerate throughput for some Metro Mirror environments by up to **100%**

Performance based on measurements and projections using IBM benchmarks in a controlled environment.

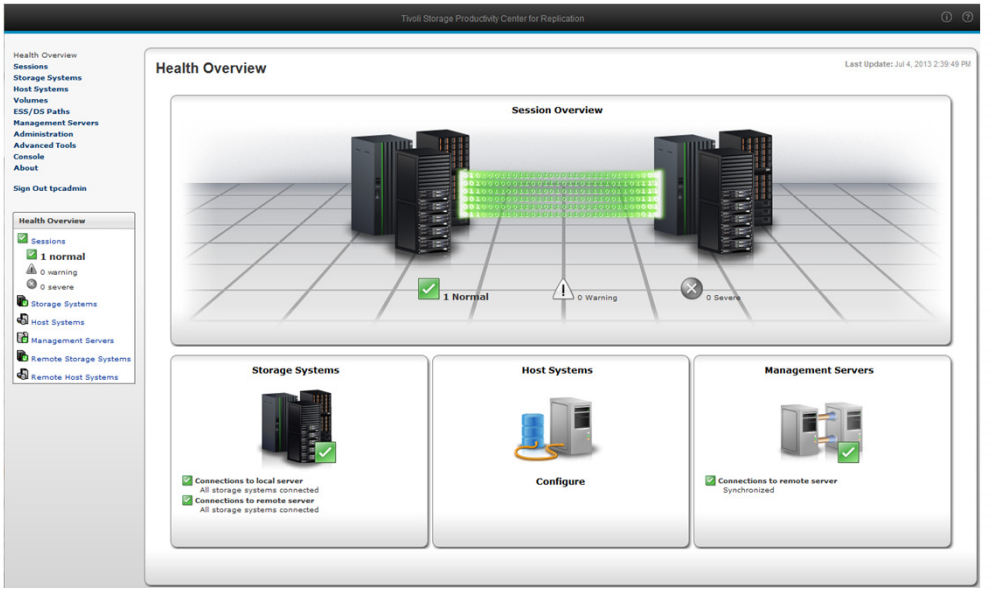
HyperWrite for z/OS and DB2*

- New function for z/OS and DS8870 with GDPS or TPC-R HyperSwap. Initial exploitation designed to help accelerate DB2 Log Writes
  - Expected benefits:
    - Improved DB2 transactional latency
    - Log throughput improvement
    - Additional headroom for growth
    - Improved resilience for workload spikes
    - Potential cost savings from workload consolidation
- Response time reduced up to 61% in prototype testing (final numbers TBD)
  - Longer distance response benefit percentage varies with distance
  - Planned to require:
    - HyperWrite function in z/OS 2.1, with the PTF for APAR OA45662
    - DB2 10 and DB2 11
    - IBM DS8870 Storage Subsystem MCL
- Planned for year end 2014

*** Statements regarding IBM future direction and intent are subject to change or withdrawal, and represent goals and objectives only.**

TPC for Replication

- Volume level Copy Service Management
  - Manages Data Consistency across a set of volumes with logical dependencies
  - Supports multiple devices (ESS, DS6000, DS8000, XIV, SVC, Storwize Family)
- Coordinates Copy Service Functionalities
  - Flash Copy
  - Metro Mirror
  - Global Mirror
  - Metro Global Mirror
- Ease of Use
  - Single common point of control
  - Web browser based GUI and CLI
  - Persistent Store Data Base
  - Source / Target volume matching
  - SNMP Alerts
  - Wizard based configuration
- Business Continuity
  - Site Awareness
  - High Availability Configuration – active and standby management server
  - No Single point of Failure
  - Disaster Recovery Testing
  - Disaster Recovery Management

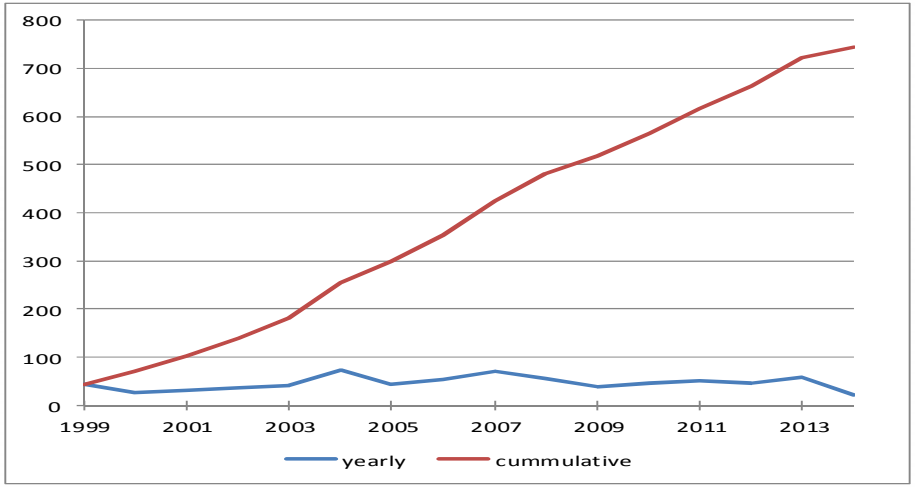


TPC-R z/OS TCP/IP Communication Feature

- Prior to TPC-R 5.2 z/OS HyperSwap management or Hardened Freeze could only be managed by TPC-R running on z/OS
  - Have to purchase the z/OS license and run on a z/OS server.
  - A TPC-R z/OS server can only manage volumes attached to the z/OS system where TPC-R is installed on.
- Enhancement with R5.2
  - Provide an IP based interface into a z/OS system in order to enable HyperSwap or Hardened Freeze configuration from a remote, non z/OS TPC-R server
- Benefits
  - Manage z/OS HyperSwap or Hardened Freeze from any TPC-R installation.
  - Enable the ability for a single TPC-R z/OS server to manage multiple z/OS systems from the same installation.
  - Enable the ability for a TPC-R standby server running on Windows to continue to manage HyperSwap or Hardened Freeze after a takeover on the Standby server

High Availability and Disaster Recovery with GDPS

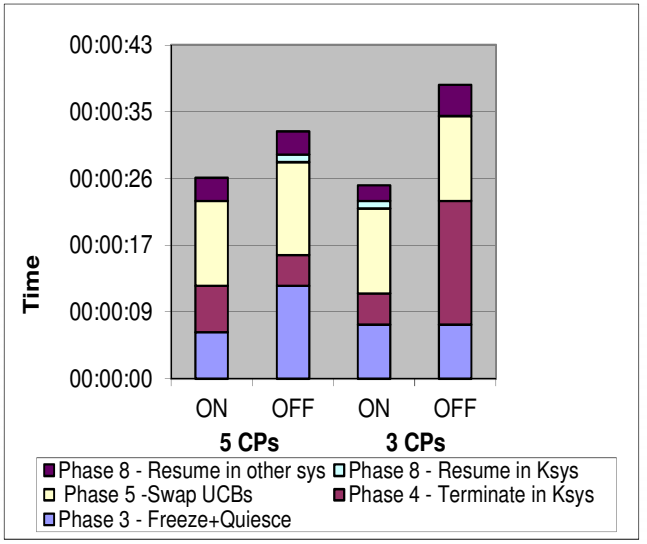
- GDPS provides a complete solution for high availability and disaster recovery in mainframe environments
- Replication management, system management, automated workflows and deep integration with z/OS and parallel sysplex
- DS8000 provides significant benefits for GDPS users with close cooperation between development teams



- Over 740 GDPS installations worldwide with high penetration in financial services and some of the largest mainframe environments
- 102 3-site GDPS installations and 4 4-site GDPS installations
- Over 90% of GDPS installations are currently using IBM disk subsystems

PPRC summary events and HyperSwap

- With larger and larger environments PPRC state change events can generate large amounts of messages consume significant resources in event processing
- This can elongate HyperSwap processing as the state changes are generated at the same time as the HyperSwap is performed
- PPRC summary messages provide one state change and message per LSS rather than one per device
- Testing showed up to 35% reduction in HyperSwap times in large environments



HS Suspend - 40 K volumes / 17 Systems

IEA075I PPRC SUMMARY,SSID=0A03,DEVICE NED=2107.951.IBM.75.0000000AFD51.0300, SUSPENDED=080,PPRC=080,TOTAL=0E0,REASON=SUSPEND(0A),FREEZE

z/OS Soft Fence

- After a HyperSwap or an unplanned site switch, potential exposures exist to systems using the old Metro Mirror (MM) primaries
  - Reading down-level data from the old MM primaries
  - Updates the old MM primaries would be lost when the MM pairs are re-established in the reverse direction

- Example scenarios
  - GDPS/PPRC provides IPL protection, however, if GDPS isn't used to IPL a system image, the IPL could be done against the former MM primaries
  - Systems outside of the Sysplex will continue to access (and perhaps update) the old MM primary volumes following an unplanned HyperSwap

- Soft Fence (SF) will allow a host system to put a volume into a "soft fenced" state and also take it out of the "soft fenced" state. When a volume is in the soft fence state the disk subsystem will prevent all reads and writes to the volume from any host system

Query Host Access and ICKDSF VERIFYOFFLINE

- DS8870 provides information to ICKDSF about every system (including different Sysplexes, Vm, Linux,...) that has a path to this volume
- New ICKDSF parm – VerifyOffline will fail an INIT or REFORMAT if the volume is online to any other system
- If the INIT/REFORMAT is rejected because it is online to some remote system, client can issue Query Host Access from ICKDSF or DEVSERV
- Output listing indicates which host(s) have the volume online

```
HOST ACCESS INFORMATION LSS=01 CCA=07
+--------------------------+----+--------+------+--------+---------+
|       PATH GROUP ID      |    |        |      |        | MAXIMUM |
|------+------+----+--------+    |        |      | DEVICE |NUMBER OF|
|      |      |CPU |CPU TIME|PATH|SYSPLEX |      |RESERVED|CYLINDERS|
|  ID  |SERIAL|TYPE|  STAMP |MODE|  NAME  |ONLINE|  TIME  |SUPPORTED|
+------+------+----+--------+----+--------+------+--------+---------+
|800002|B947  |2827|CA78BC17|S   |N/A     |NO    | ------ |120936   |
+------+------+----+--------+----+--------+------+--------+---------+
|880005|B947  |2827|CAAD6FBA|M   |PROD    |YES   | ------ |FFF0     |
+------+------+----+--------+----+--------+------+--------+---------+
```

Storage Subsystem health determination

- DS8000 will alert z/OS whenever hardware events occur generating a message and ENF signal
  - Provides immediate information to z/OS for failure and recovery events

```
IEA074I STORAGE CONTROLLER HEALTH,MC=20,TOKEN=1004,SSID=AB01, DEVICE
NED=2107.961.IBM.75.0000000ABCD1.0100,PPRC SECONDARY CONTROLLER RECOVERY ACTION
```

- DS8000 also provides an API to query the health status of a disk subsystem

- GDPS exploits storage subsystem health determination
  - Specific DS8000 failure events will immediately trigger a HyperSwap
  - GDPS Freeze and Conditional Stop will query secondary disk subsystem to determine if systems can be allowed to continue in a Freeze event

z/OS Support for Non-disruptive Statesave

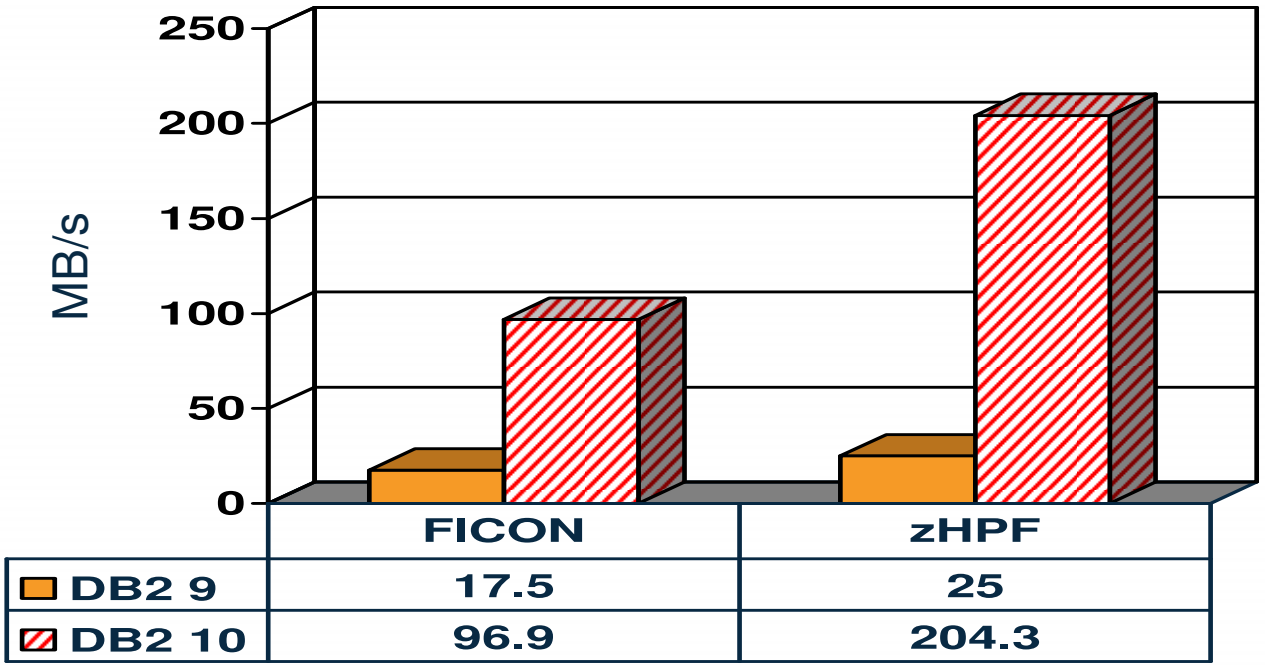
- Adding capability for z/OS Software to issue a command to collect the Non-Disruptive StateSave (NDSS)
- Users can explicitly initiate a state save via the F ANTAS00x,STATESAVE command
  - New “N” option will perform a non-disruptive statesave
- DFSMS System Data Mover (XRC) and GDPS will perform non-disruptive statesaves at appropriate times
  - Certain XRC errors or GDPS freeze events

High Performance FICON architecture levels

- zHPF for Media Manager (phase 0) - DS8000 R4.1
- zHPF Multi-Track support for Media Manager + z196 Support - DS8000 R4.3
- zHPF for QSAM/BSAM (phase 1) - DS8000 R6.2
  - this includes format writes that also accelerate DB2 table space provisioning
- zHPF List Prefetch Optimizer - DS8000 R6.2
- z196 + DS8000 BiDi support for the List Prefetch Optimizer - DS8000 R6.2

zHPF List Prefetch

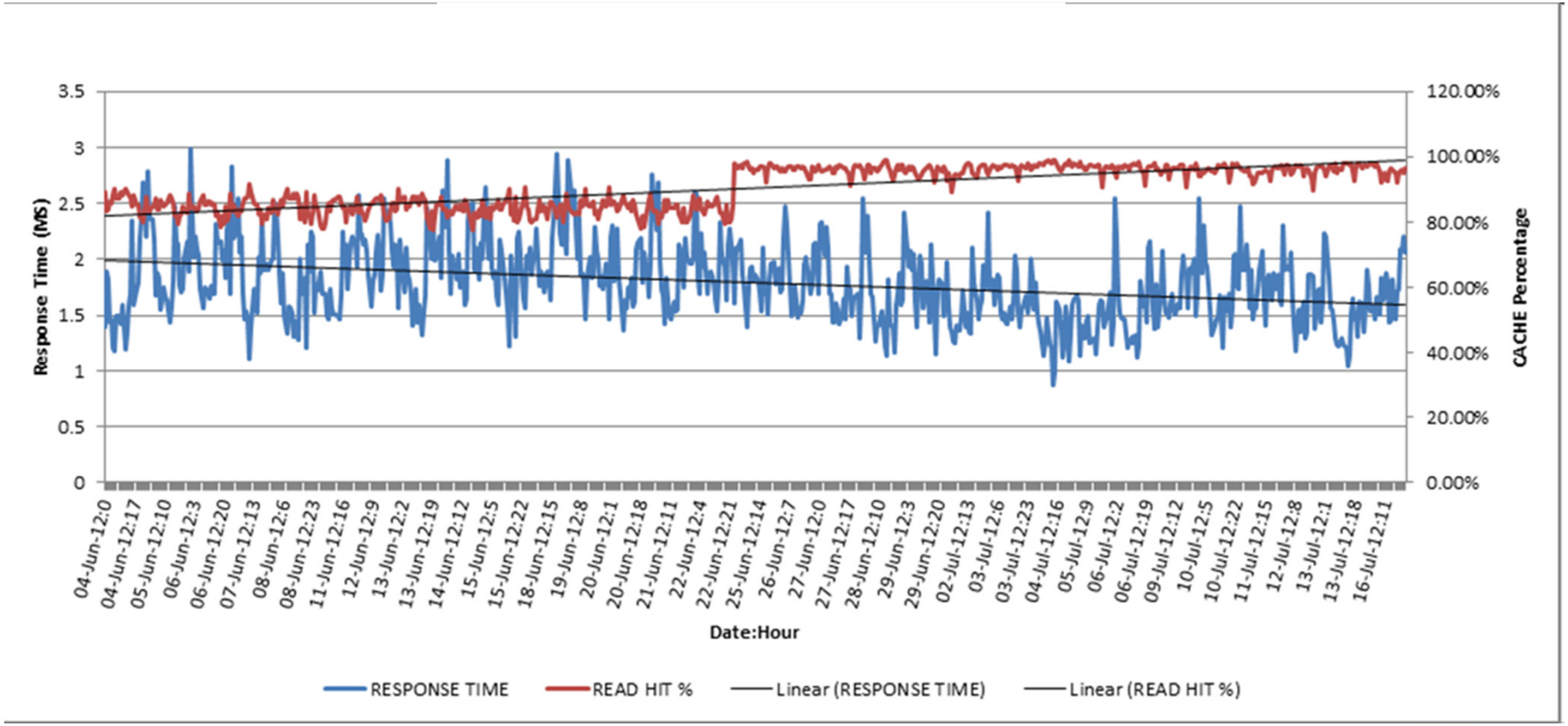
(Disorganized Index Scans with hot cache, 4K pages)



	FICON	zHPF
DB2 9	17.5	25
DB2 10	96.9	204.3

- 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

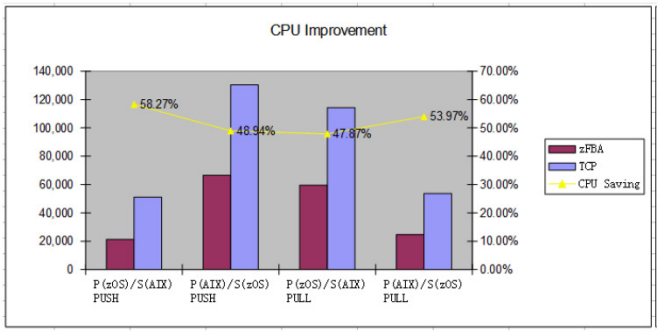
zHPF List Prefetch Optimizer Results



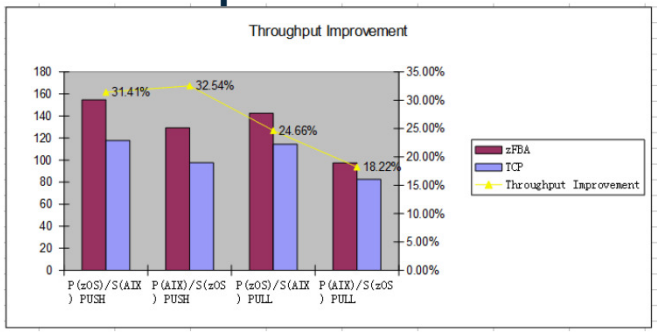
Enhancements for Managed File Transfer

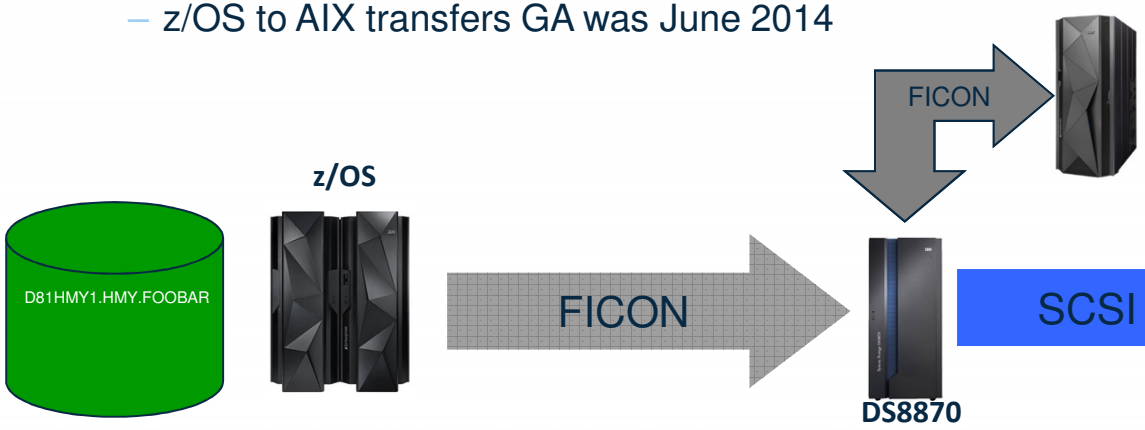
- Lowering the cost of moving data
  - Send the data over a DASD bridge vs. Ethernet network
  - Connect Direct with DS8000 zDDB and new z/OS services
- IBM goal is to improve data transfer between z/OS images and distributed platforms
- IBM expects a 2x improvement in CPU cost and significant reduction in elapsed time for Connect Direct vs. TCP/IP
  - Based on z/OS and Connect Direct performance testing
- Improves data transfer between z/OS images and distributed platforms
  - z/OS to z/OS transfers GA was March 2014
  - z/OS to AIX transfers GA was June 2014

2x CPU Improvement



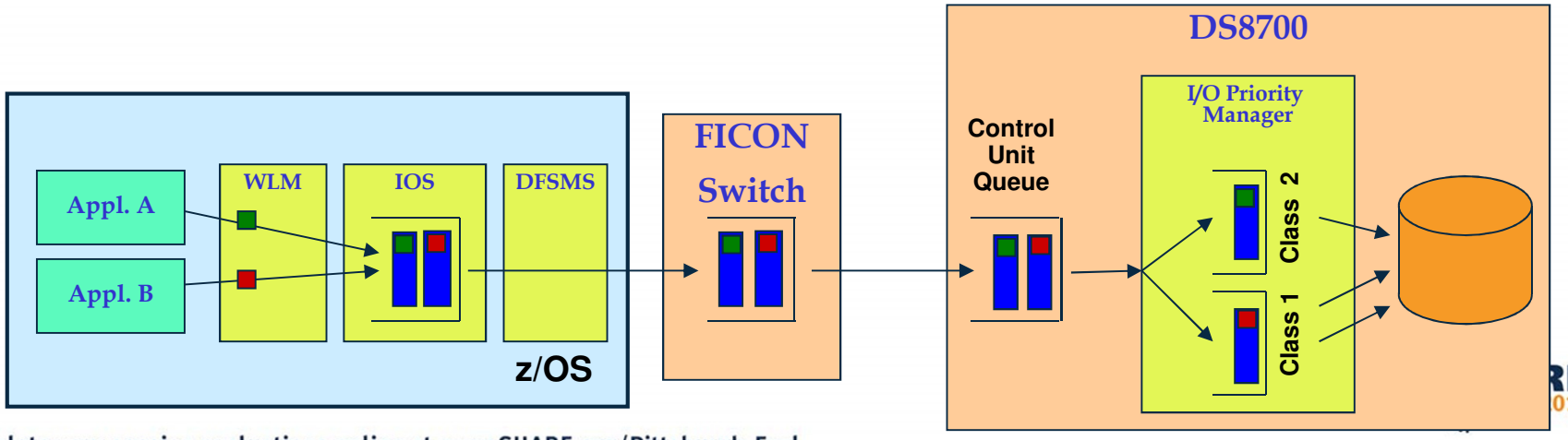
30+% Elapsed Time Reduction





WLM Support for DS8000 IO Priority Manager

- WLM sends I/O Priority Manager information about the goal fulfillment and importance of z/OS workloads (service classes).
- Passing these performance parameters to the storage server enables the I/O Priority Manager to determine which I/O requests are more important than others and which I/O requests need to be processed faster to fulfill the performance goals defined for the corresponding workload in z/OS.
- Using the passed information from WLM, the I/O Priority Manager throttles I/O requests of workloads which exceed their goals to help I/O requests of workloads which do not fulfill their goals.
- New IEAOPT parameter STORAGESERVERMGT={YES|NO}



DS8000 and System z Summary

- Significant investment in synergy between DS8000 and System z
- High Performance Flash Enclosure provides Flash storage improvements for DS8870 and enhances tiered storage with Easy Tier
- Many enhancements related to replication with both GDPS and TPC-R management
- High Performance FICON provides significant enhancements to DB2 workloads with a range of capabilities all supported by DS8000
- Further future enhancements in the pipeline such as DB2 HyperWrite and Easy Tier Application integration with DFSMS