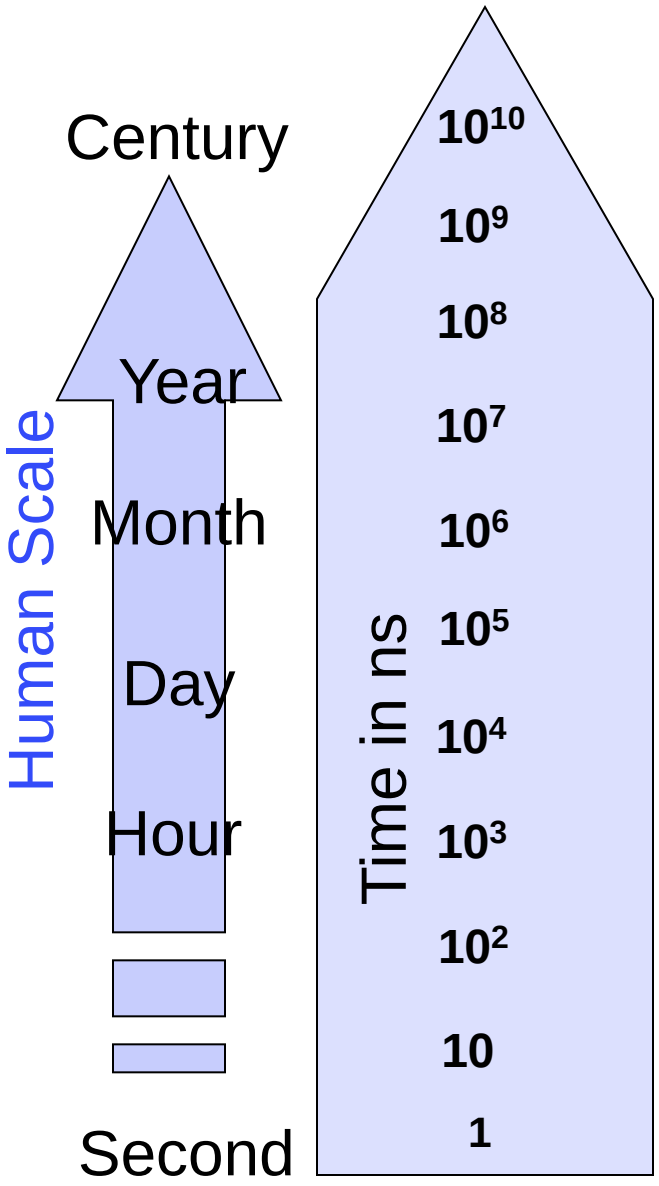
IBM Storage Tiering Technology

Clodoaldo Barrera
IBM Systems Storage

Agenda

- Storage Class Memory Overview and Future
- IBM Storage Tier Technology Overview and Performance Results
  - IBM DS8K Product Overview
  - Observation of Workload
  - IBM SSD Option
  - Performance Management
- IBM Storage Tier Future
  - Cooperative Caching
  - Tape Tier

Performance defines the battle field, …



Get data from TAPE (40s)

Access DISK (5ms)

Storage

Access FLASH (20 us)

SCM

Access PCM (100 – 1000 ns)

Get data from DRAM *or* PCM (60ns)

Get data from L2 cache (10ns)

Memory

CPU operations (1ns)

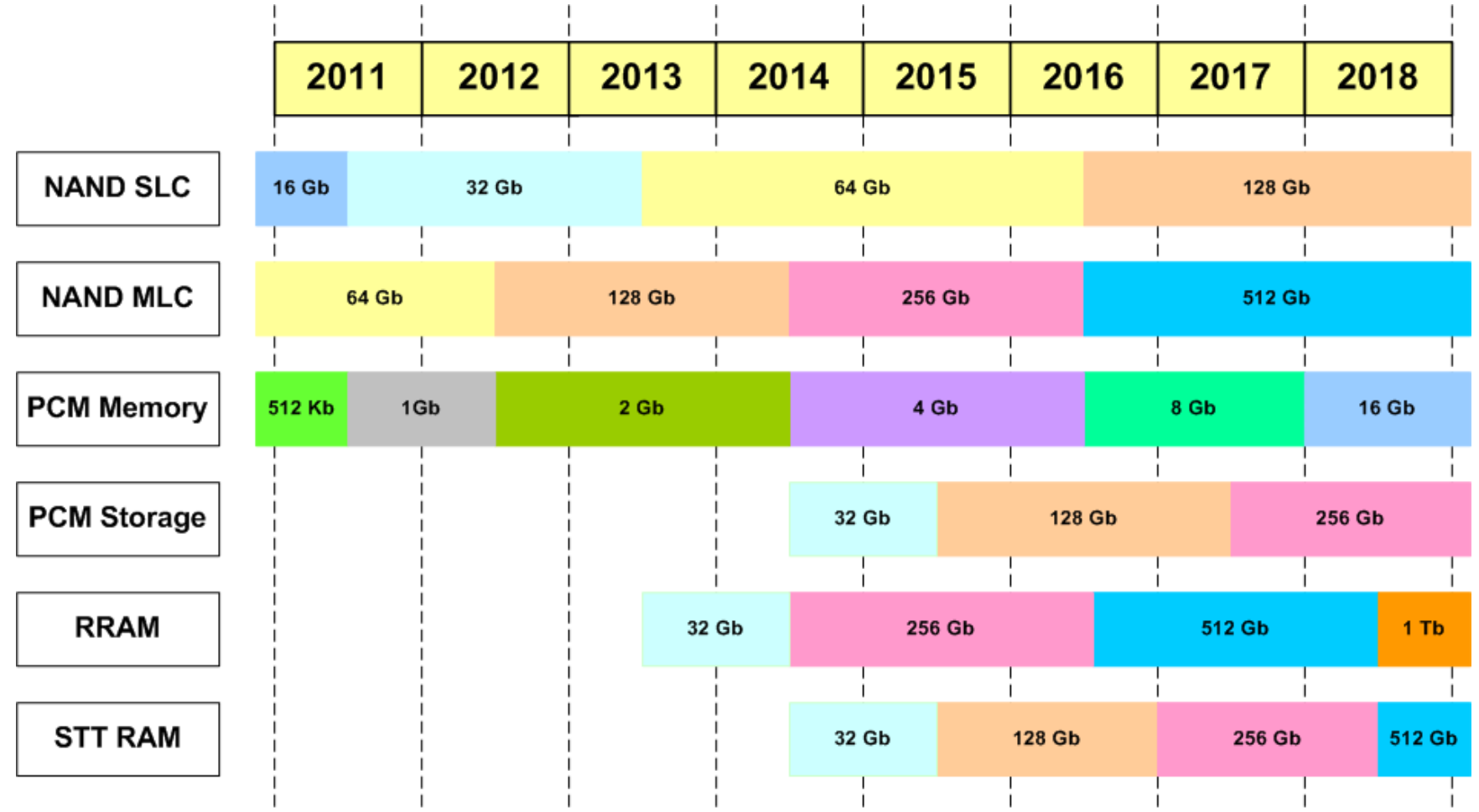
But, density is key to winning

- Cost competition between IC, magnetic and optical devices comes down to **effective areal density**.

Device	Critical feature-size F	Area (F^2)	Density (Gbit /sq. in)
Hard Disk	50 nm **(MR width)**	1.0	250
DRAM	45 nm **(half pitch)**	6.0	50
NAND (2 bit)	43 nm **(half pitch)**	2.0	175
NAND (1 bit)	43 nm **(half pitch)**	4.0	87
Blue Ray	210 nm **(λ/2)**	1.5	10

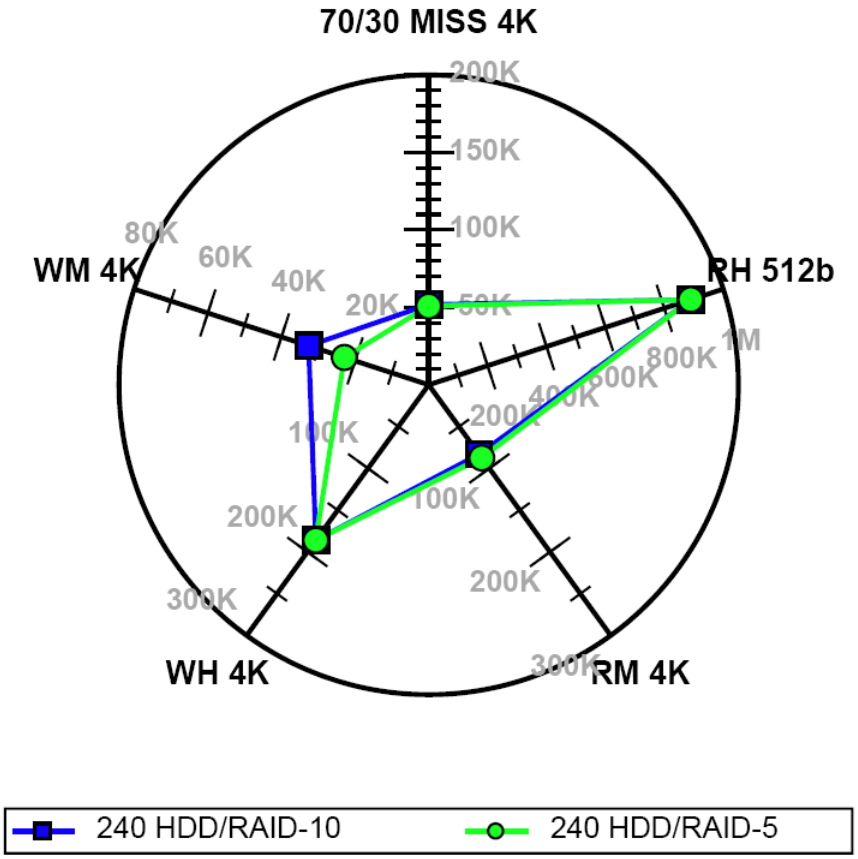
[Fontana:2004, web searches]

Readiness and Availability Road Map

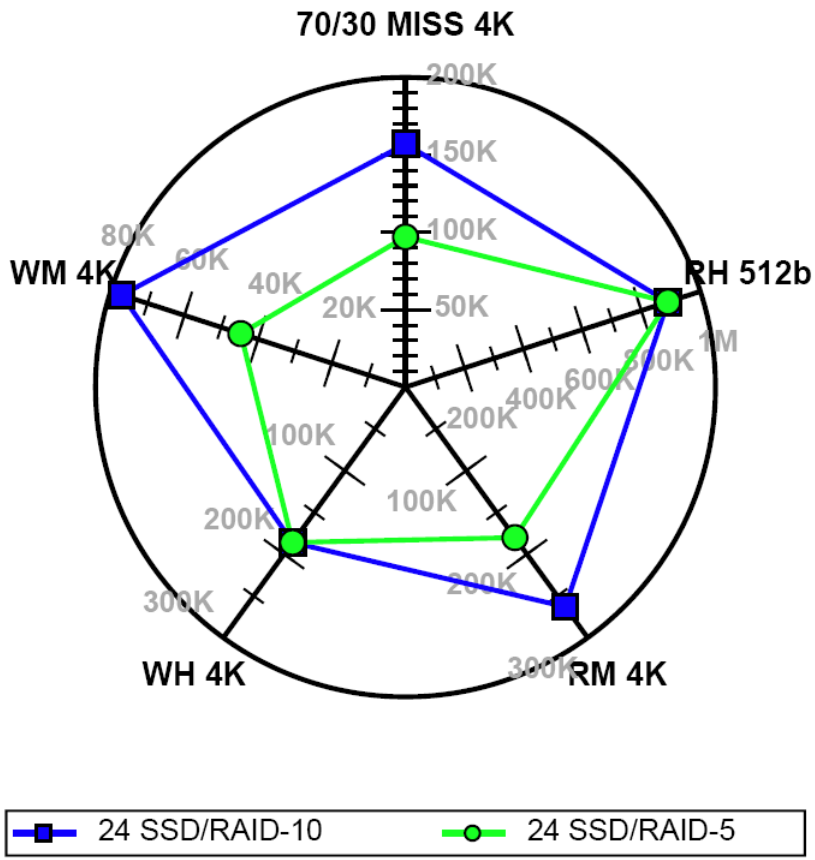


Source: 2010 Non Volatile Memory Conference

Performance – HDD vs SSD



HDD



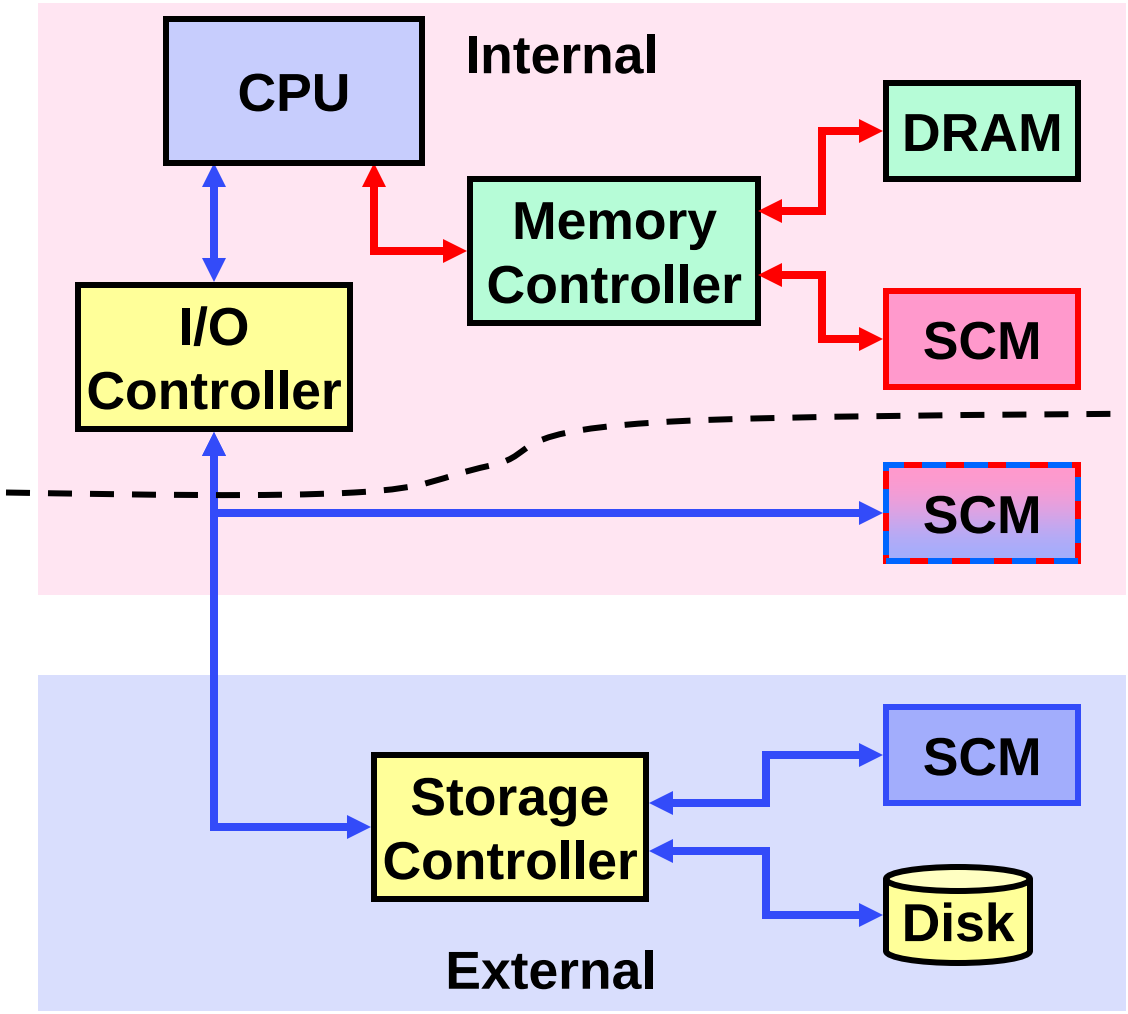
SSD

Attribute claims for Storage Technologies: ~2015

	MLC Flash	PCM	RRAM	STT-RAM	Disk
Capacity M/S	256 Gb	4 Gb / 32 Gb	128 – 256 Gb storage?	? / 32 Gb	3-5 TB
Multi generation Scalability	Yes, but Limit??	Yes	Yes	Yes	Yes 25-40%/yr
MLC	Yes	Yes	Maybe	Maybe	NA
3D	Probably	Probably	Maybe	Maybe	NA
Performance R/W in us	200/500	< .1/ 1-10	< .1/ < 5 Storage?	< .035/ < .035	4000/4000
Write Approach	P/E	WIP	WIP	WIP	WIP
Endurance	10^3	10^8	10^8	10^{15}	10^{11}
Data retention	Good-Fair	Fair (temp)	?	?	Excellent
Power	good	high	good	good	Very High
Maturity	mature	maturing	new	new	mature

Source: drawn from 2010 Nonvolatile Memory Conference 'Developing Successful Strategies for the NVM Revolution: 2010 – 2020' + various other sources + ITRS +interpolation, etc.

Architecture



Synchronous

- Hardware managed
- Low overhead
- Processor waits
- Fast SCM, Not Flash
- Cached or pooled memory

Asynchronous

- Software managed
- High overhead
- Processor doesn't wait
- Switch processes
- Flash and slow SCM
- Paging or storage

Agenda

- Storage Class Memory Overview and Future
- IBM Storage Tier Technology Overview and Performance Results
  - IBM DS8K Product Overview
  - Observation of Workload
  - IBM SSD Option
  - Performance Management
- IBM Storage Tier Future
  - Cooperative Caching
  - Tape Tier

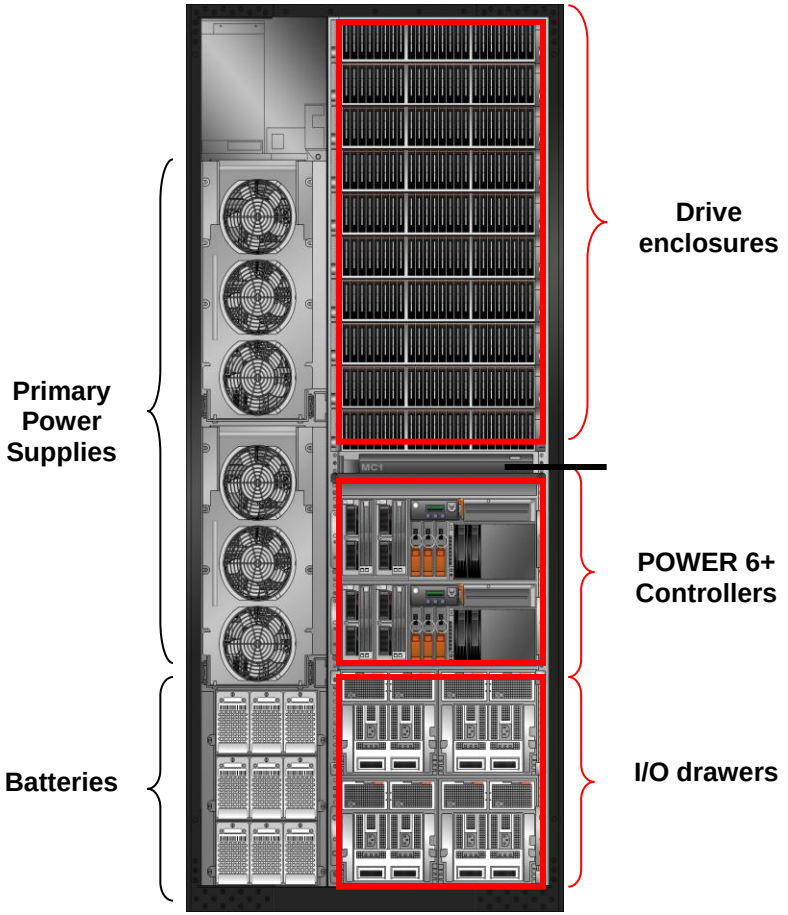
DS8000 sustained growth and leadership

- **State of the business**
  - –Over 30,000 frames and over 17,000 systems sold
  - –Shipped more disks in Q411 than any previous quarter
  - –Shipped more petabytes in 2011 than any other year
  - –Average system in 2011 shipped with 129TB
    - Increase of 70% since 2009

- **Features clients care about**
  - –Highest Gartner rating in MarketScope for Monolithic Enterprise Disk Arrays – **"Strong Positive",** March 2011
  - –Easy Tier / SSD penetration at 60% in Q2 2012
  - –Optimized design synergy with IBM server environments

DS8800 Hardware

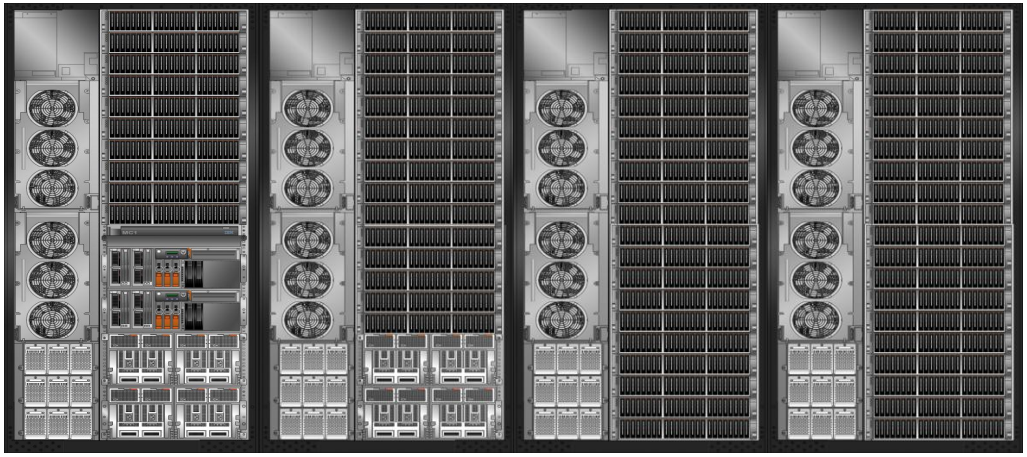
- Compact and highly efficiency drive enclosures
  - New 2.5”, small-form-factor drives
  - 6 Gb/s SAS (SAS-2)
  - New enclosures support 50% more drives
- Upgraded processor complexes
  - IBM POWER6+ for faster performance
- Upgraded I/O adapters
  - 8 Gb/s host adapters
  - 8 Gb/s device adapters
- More efficient airflow
  - Front-to-back cooling
  - Aligns with data center best practices



Front view
(cover removed)

DS8800 configuration limits

- From 32GB to 384GB Processor Memory (128GB on 2-way)
  - 1GB to 12GB NVS based on memory size
- Up to 128 FC or FICON ports (32 on 2-way)
  - 8Gb only (no 2Gb or 4Gb)
- Up to 8 device adapter pairs (2 on 2-way)
  - Device adapter pair for each 48 DDMs for first 384 DDMs
- From 16 to 1536 DDMs (128 on 2-way and 240 on 2-way business class)



IBM System Storage DS8000 R6.2-3 Announcement Highlights

- **Storage Scalability and Resiliency**
  - Third expansion unit (base + three expansion frames)
  - Additional drive options
  - DDM Smart Rebuild
- **System z Synergy**
  - DS8000 I/O Priority Manager
  - Larger Extended Address Volumes
  - Dramatic improvements to High Performance FICON for System z
  - DB2 list prefetch enhancements
  - Quick Initialization for CKD volumes
  - Enhancements in support of HyperSwap

IBM System Storage DS8000 R6.2-3 Announcement Highlights - continued

- **Enhanced Tiering Capabilities**
  - Three tier support
  - Full support for Thin Provisioned volumes (FB only)
  - Auto-rebalance of homogenous pools
- **Thin Provisioning (ESE) Enhancements**
  - Thin Provisioning support for FlashCopy (Open systems only)
  - Thin Provisioning support for Easy Tier pools
- **User Interface Enhancements**
  - DS Storage Manager support for Resource Groups
  - Parallel volume create and delete reduces configuration/reconfiguration times
  - DS Storage Manager GUI can be launched directly from any supported remote browser without having to launch through TPC
- **Encrypted Drives Supported**
  - SSD and HDD Full Disk Encryption capable drives

Easy Tier generational enhancements

- **1st Generation (DS8000, SVC and V7000)**
  - Objective: Optimizing use of expensive SSDs
  - Benefit: Relocating just 5% of the data from HDDs to SSDs, reduced average I/O response time from 9ms to 2ms

- **2nd Generation (DS8000 only)**
  - Objective: Intra-tier rebalancing and support for any 2 tiers (no SSDs)
  - Benefit: Automates performance optimization ***within*** a tier; Relocating less active data to slower nearline drives maintains performance with lower $/GB as data grows

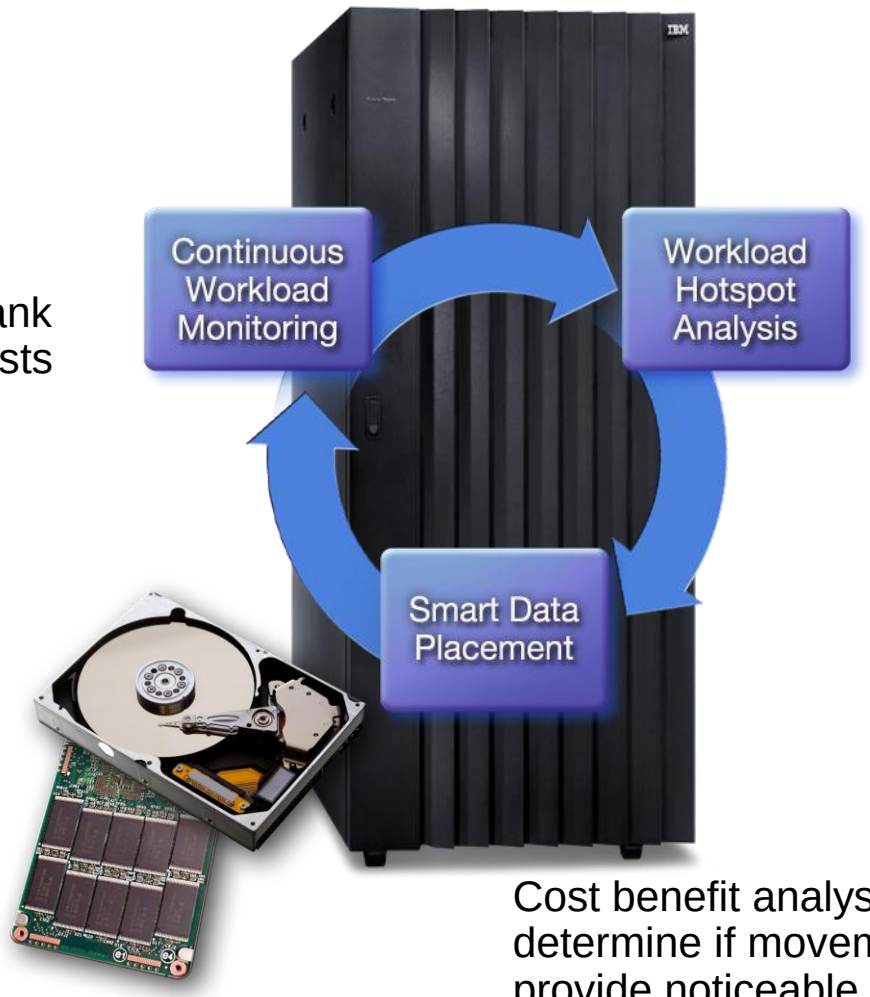
- **3rd Generation (DS8000 only – R6.2)**
  - Objective: Full support across all 3 tiers and support for Thin Provisioned volumes
  - Benefit: Faster performance when and where it's needed with SSDs; Cost savings (reduced footprint and $/GB) for cold data; flexibility to support standard and Thin Provisioned volumes

- **4th Generation (DS8000 only – R6.3)**
  - Objective: Support for Full Encryption
  - Benefit: Combines advanced tiering with superior security for the ultimate in efficiency and data protection

Easy Tier Processing Cycle

Performance data collected every 15 minutes

Data collected is for rank activity not IO from hosts



Workload analysis performed at least ever 24 hours

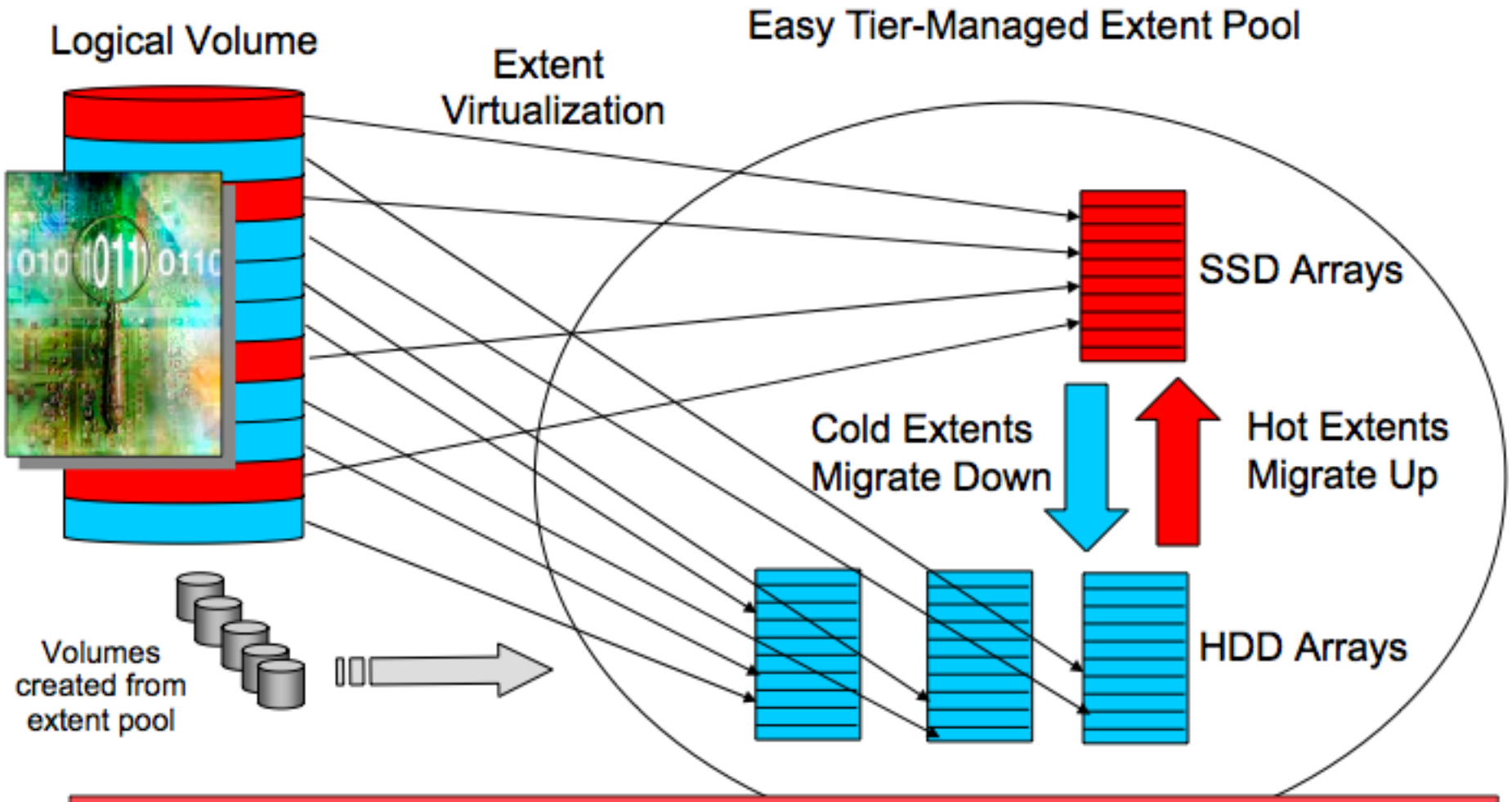
Takes into account both short term and long term data

Extents categorised based on small and large IO activity

Cost benefit analysis performed to determine if movement of data will provide noticeable improvement

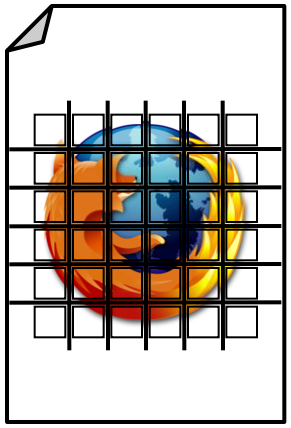
Movement of extents scheduled

How Easy Tier Works

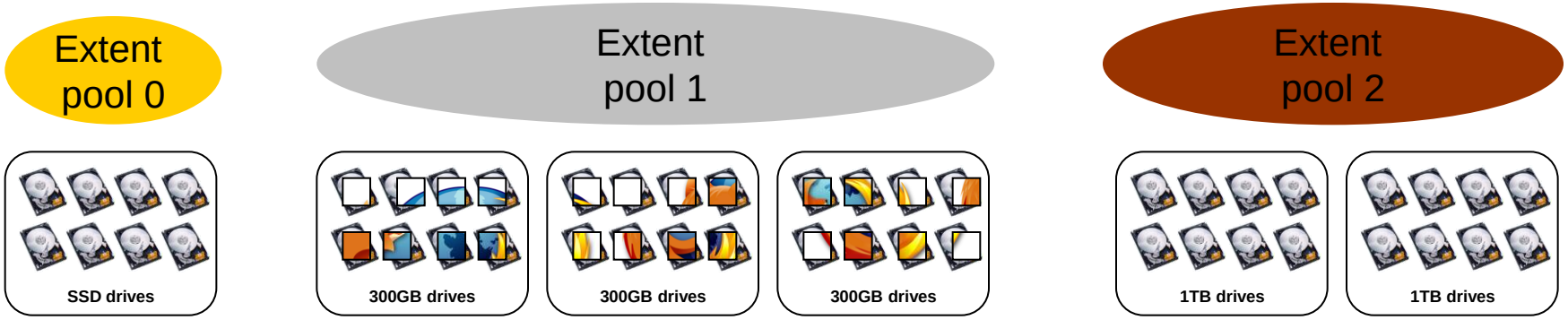


1. Easy Tier monitors performance of each extent to determine the data 'temperature'
2. Easy Tier creates extent migration plan for optimal data placement every 24 hrs based on performance statistics
3. Easy Tier migrates extents within an extent pool according to plan over 24 hour period (limited number of extents are chosen for migration every 5 minutes)

DS8000 virtualisation architecture



- DS8000 virtualisation architecture simplifies management and enables advanced function
  - Arrays are cut into 1GB extents and assigned to extent pools
- Storage Pool Striping rotates the extents related to a device over the arrays in an extent pool
  - Provides even utilisation of drives in the extent pool
- Thin provisioning allocates extents only when the extent is written to
  - Function is optimised for performance giving equal performance to fully provisioned devices



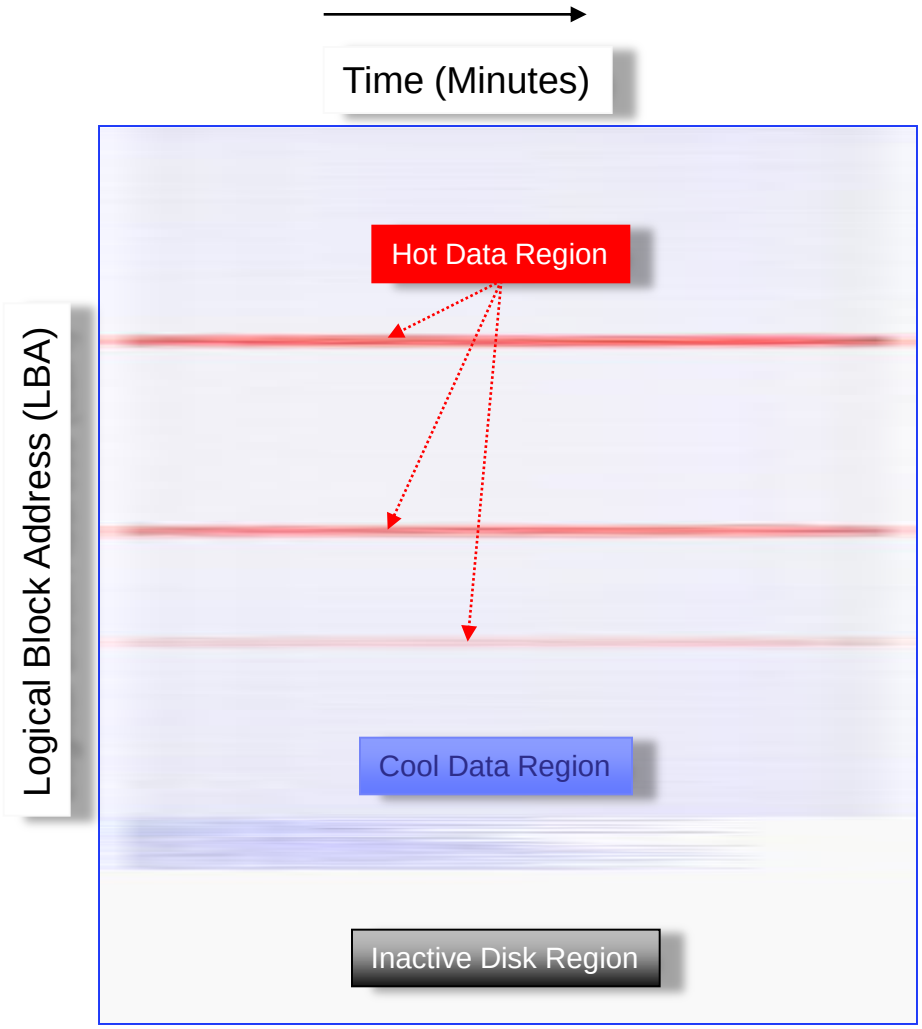
Sample Easy Tier configurations

- 10-20% SSD, 80-90% Enterprise
  - Provides SSD like performance with reduced costs
- 3-5% SSD, 95-97% Enterprise
  - Provides improved performance compared to single tier solution
  - Removes requirement for over provisioning for high access density environments
  - All data guaranteed to have at least enterprise performance
- 3-5% SSD, 25-53% Enterprise, 40-70% Nearline
  - Provides improved performance and density to a single tier solution
  - Significant reduction in environmental costs
- 20-50% Enterprise, 50-80% Nearline
  - Provides reduced costs and comparable performance to a single tier Enterprise solution

Agenda

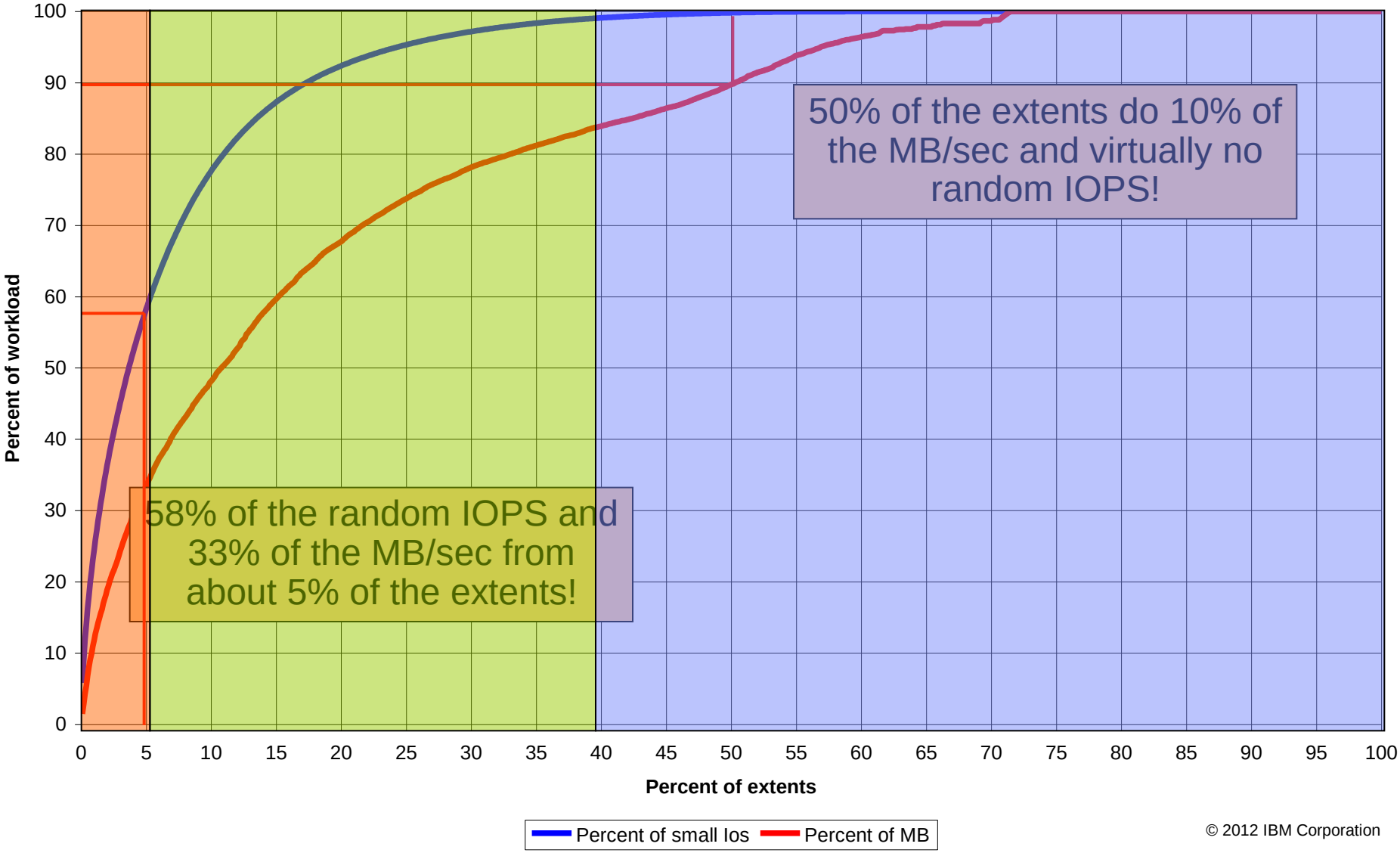
- Storage Class Memory Overview and Future
- IBM Storage Tier Technology Overview and Performance Results
  - IBM DS8K Product Overview
  - Observation of Workload
  - IBM SSD Option
  - Performance Management
- IBM Storage Tier Future
  - Cooperative Caching
  - Tape Tier

Workload Learning



- **Each workload has its unique IO access patterns and characteristics over time.**
- **Heatmap will develop new insight to application optimization on storage infrastructure.**
- **Left diagram shows historical performance data for a LUN over 12 hours.**
  - Y-axis (Top to bottom) LBA ranges
  - X-axis (Left to right) time in minutes.
- **This workload shows performance skews in three LBA ranges.**

Skew in a typical client environment

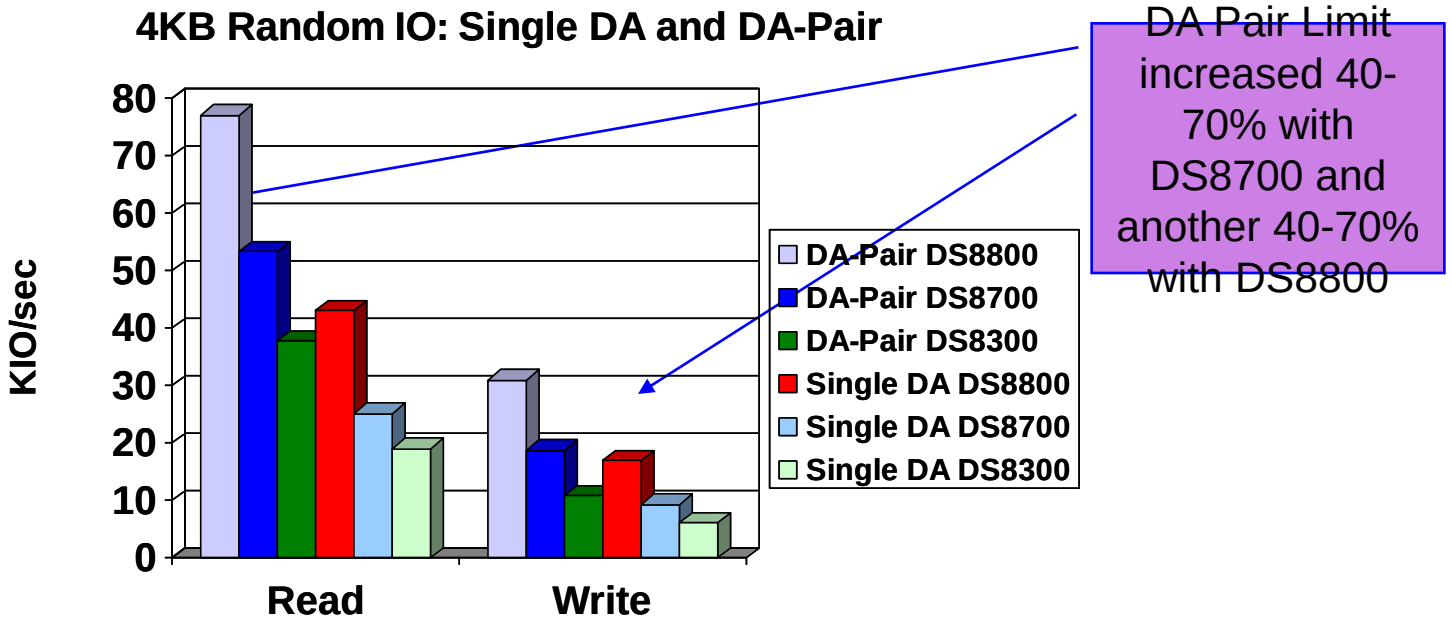


Agenda

- Storage Class Memory Overview and Future
- IBM Storage Tier Technology Overview and Performance Results
  - IBM DS8K Product Overview
  - Observation of Workload
  - IBM SSD Option
  - Performance Management
- IBM Storage Tier Future
  - Cooperative Caching
  - Tape Tier

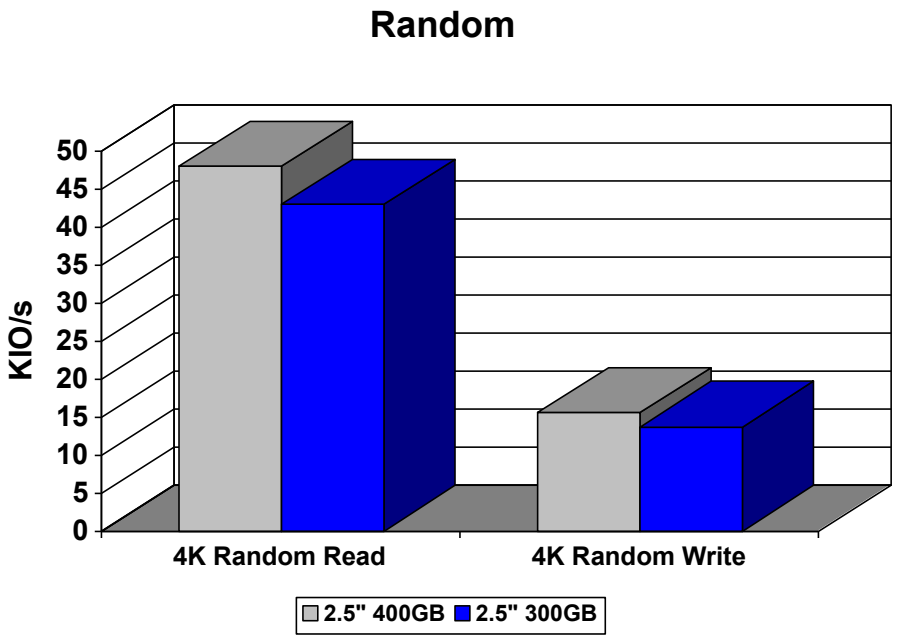
SSDs in DS8700/DS8800

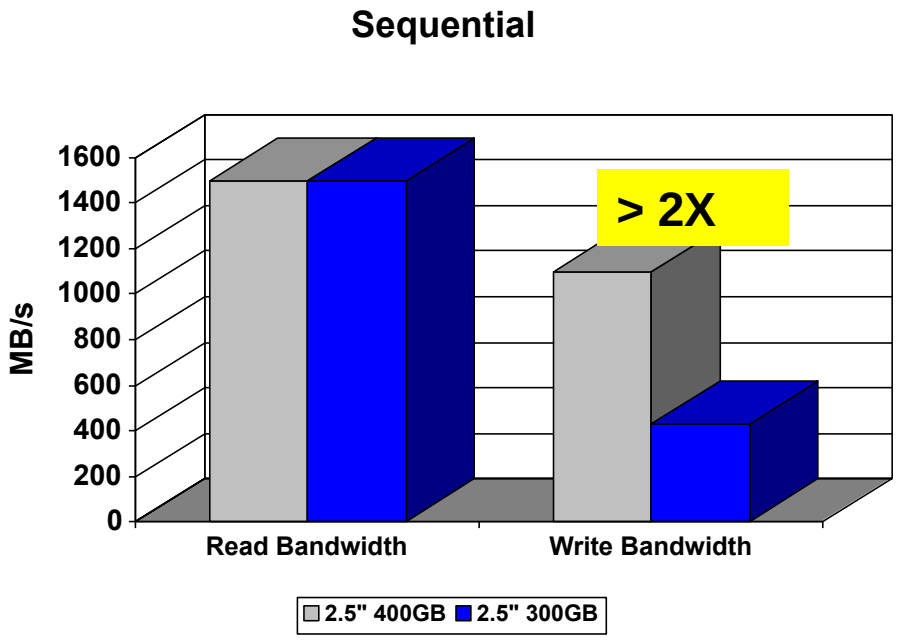
- Up to 256 SSDs available on DS8700 or 384 SSDs on DS8800
  - Minimum of 8 SSDs with recommended minimum of 16
- More powerful Device Adapters on DS8700 and again on DS8800 allowing for higher backend throughputs possible with solid state storage



Another Upgrade: 2.5” 400GB SSD Drive Performance

The new 2.5” 400 GB SSDs on the DS8800 has equal or slightly better performance for 4k random read/write and sequential read comparing with 2.5” 300GB SSDs, however, for sequential write 400GB SSDs perform more than 2X better.
This technology is also available on SVC and V7000.



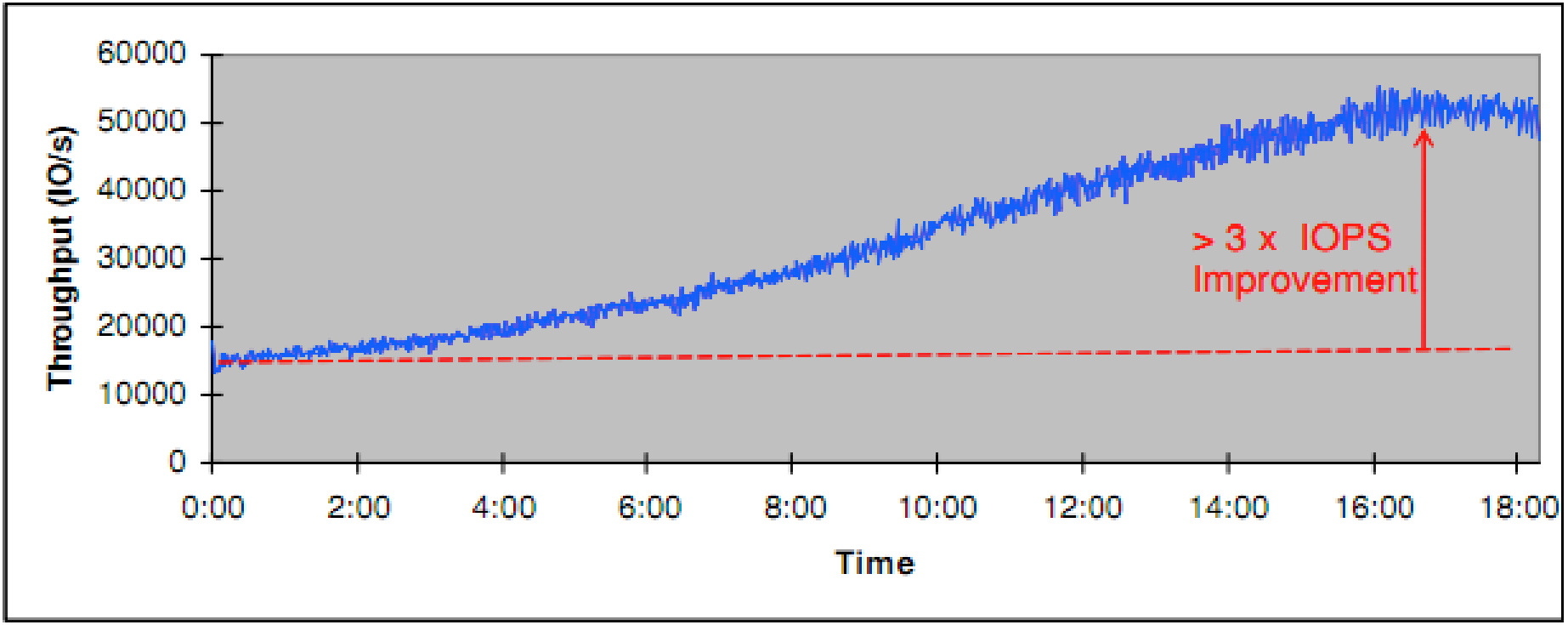


Agenda

- Storage Class Memory Overview and Future
- IBM Storage Tier Technology Overview and Performance Results
  - IBM DS8K Product Overview
  - Observation of Workload
  - IBM SSD Option
  - Performance Management
- IBM Storage Tier Future
  - Cooperative Caching
  - Tape Tier

Demonstration of Data Placement Technology on IBM Enterprise Storage System

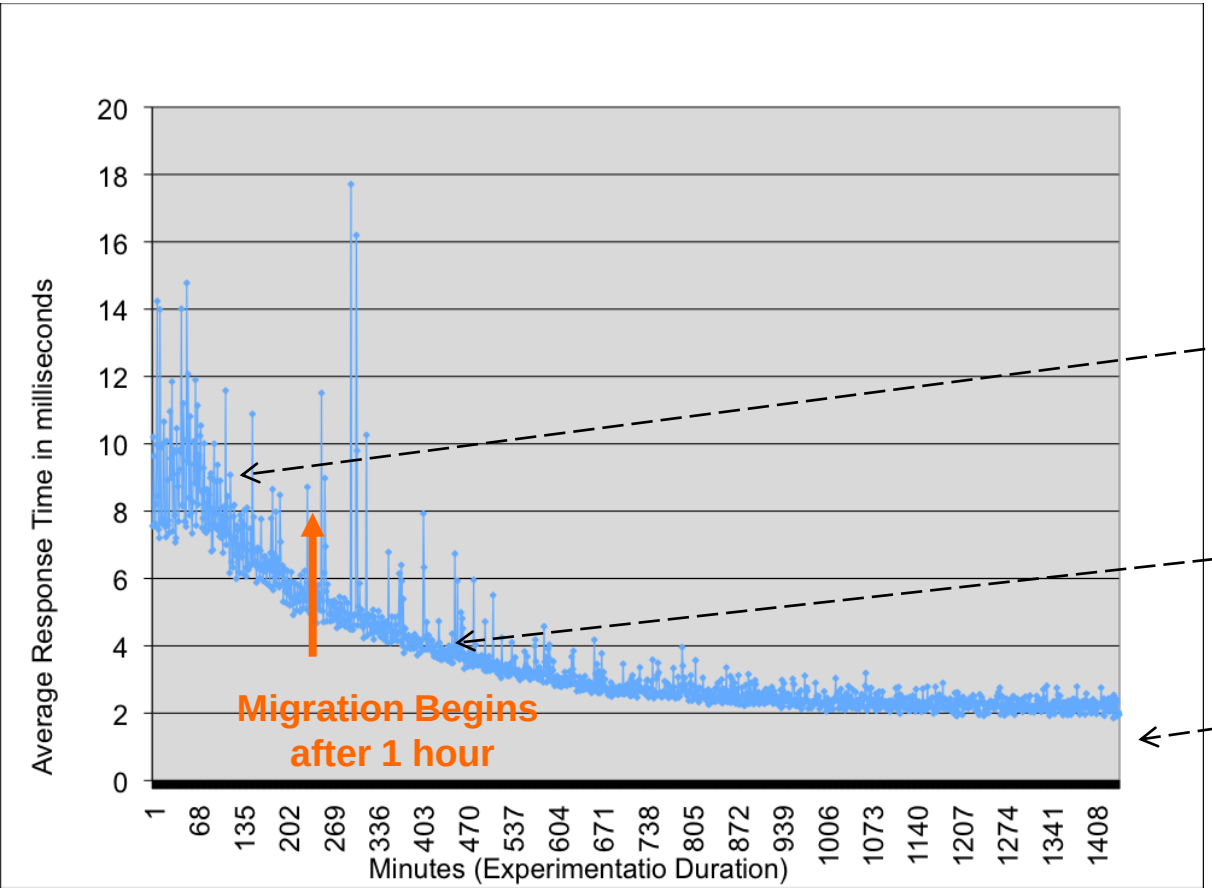
SPC-1 Improvement through Automatic Tiering



http://www.storageperformance.org/results/benchmark_results_spc1#a00092

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/5cb5ed706d254a8186256c71006d2e0a/f1bf1c2fa3250abf86257745004c80a4/$FILE/DS8700%20Performance%20with%20Easy%20Tier.pdf

Average Response Time Shows Significant Improvement with Data Placement and Migration Technology

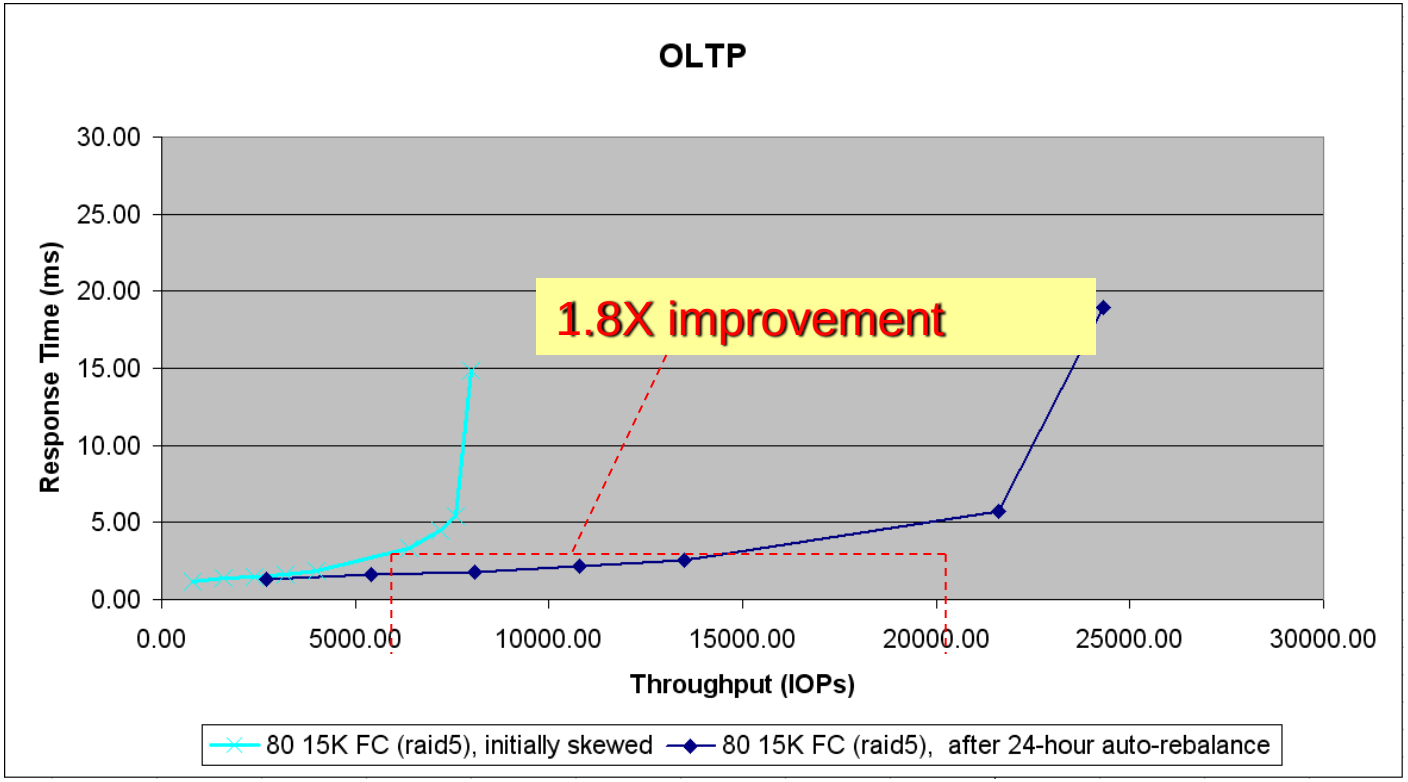


Before Migration:
Avg RT 9.13msec

After 5 hours Migration:
Avg RT 4 msec

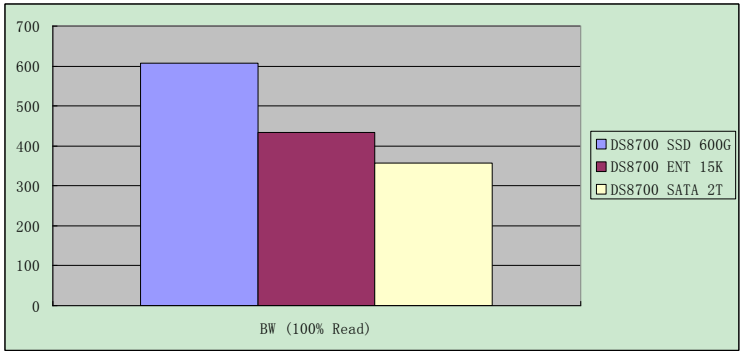
Maximum Improvement
Of Average RT to 2 msec

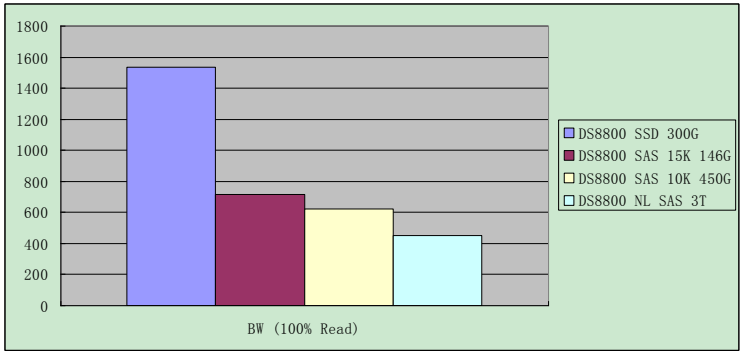
Eliminate performance skew with Auto Performance Rebalance



Leverage SATA tier bandwidth – Expanded Cold Demote

Rank Bandwidth on DS8700/DS8800 (100% Sequential Read)



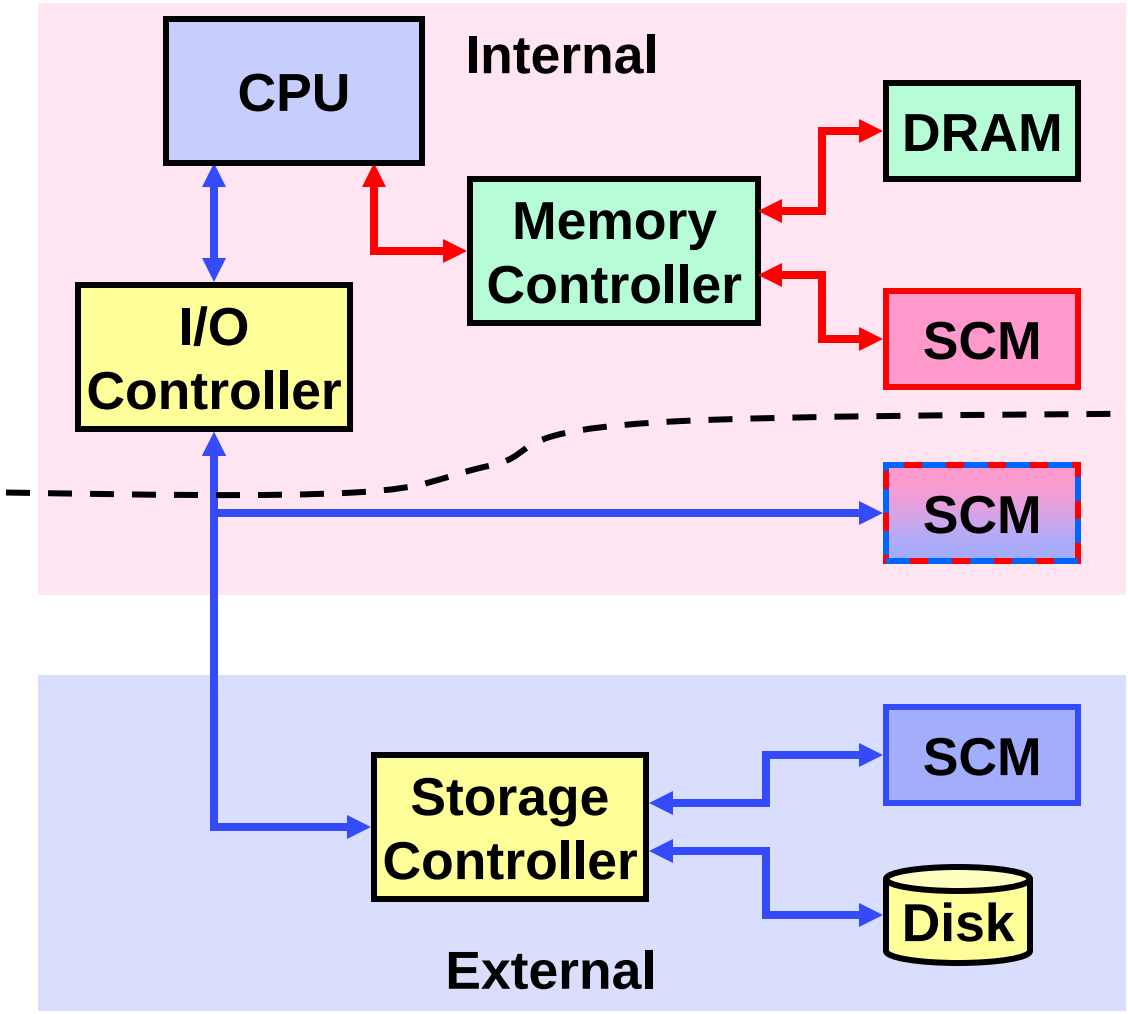


- Bandwidth difference between the enterprise tier and the SATA/Nearline tier is not significant.
- Bandwidth on SATA/Nearline tier can be utilized to balance the sequential workloads.
- Expanded Cold demote will demote:
  - Pure sequential accessed extents (random bucket = 1)
  - Mostly sequential but with few random accessed extents (random bucket =2, sequential bucket >1)

Agenda

- Storage Class Memory Overview and Future
- IBM Storage Tier Technology Overview and Performance Results
  - IBM DS8K Product Overview
  - Observation of Workload
  - IBM SSD Option
  - Performance Management
- IBM Storage Tier Future
  - Cooperative Caching
  - Tape Tier

Architecture



Synchronous

- Hardware managed
- Low overhead
- Processor waits
- Fast SCM, Not Flash
- Cached or pooled memory

Asynchronous

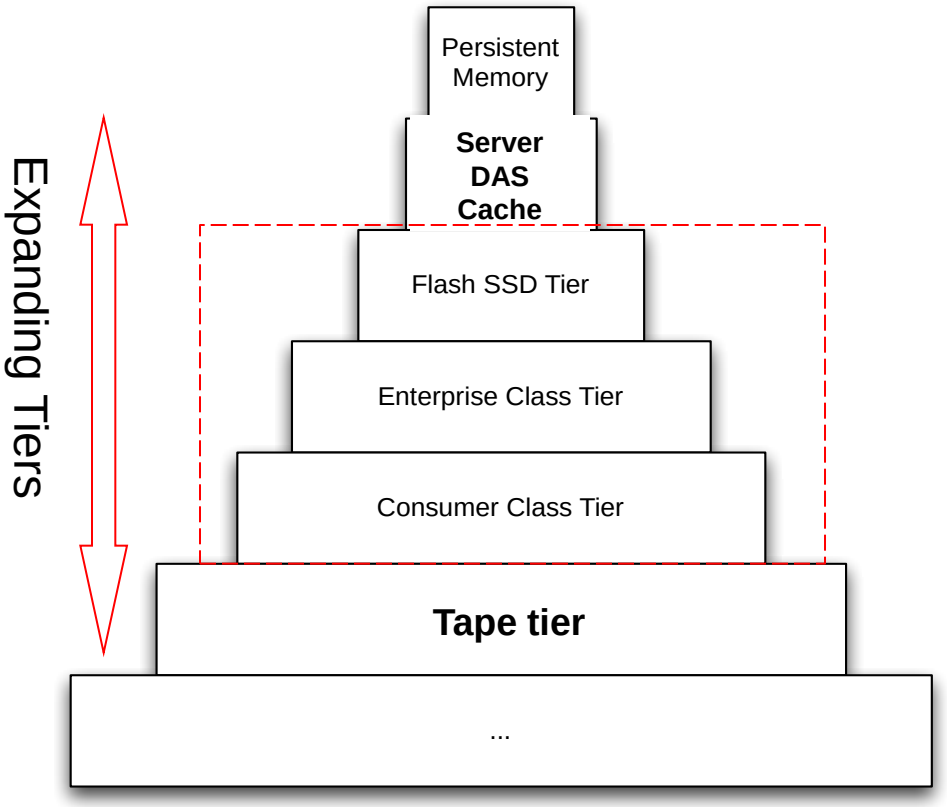
- Software managed
- High overhead
- Processor doesn't wait
- Switch processes
- Flash and slow SCM
- Paging or storage

Expanding tier Up

1. Closer to the computing power**.**
2. Higher performance.
3. Fewer transport layer.

Expanding tier Down

1. Massive capacity**.**
2. Cheaper in $/G.
3. Saving Power (offline storage).



Memory Storage Hierarchy

June 2012 Statements of Direction – Easy Tier

- Advanced Easy Tier™ capabilities on selected IBM storage systems, including the IBM System Storage® DS8000® , designed to **leverage direct-attached solid-state storage on selected AIX® and Linux™ servers**. Easy Tier will manage the solid-state storage as a large and low latency cache for the "hottest" data, while preserving advanced disk system functions, such as RAID protection and remote mirroring.

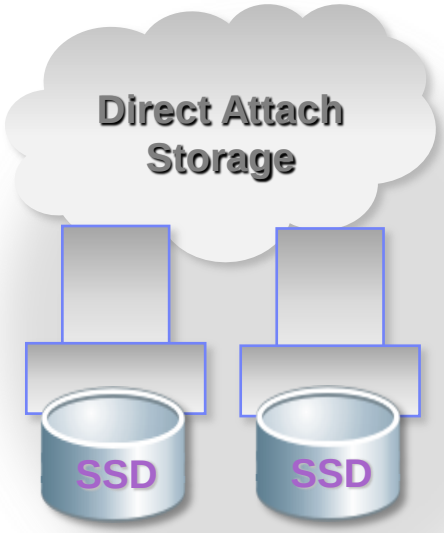
- An **application-aware storage application programming interface (API)** to help deploy storage more efficiently by enabling applications and middleware to direct more optimal placement of data by communicating important information about current workload activity and application performance requirements.

- A new **high-density flash storage module** for selected IBM disk systems, including the IBM System Storage DS8000 . The new module will accelerate performance to another level with cost-effective, high-density solid-state drives (SSDs).

EasyTier Cooperative Caching Technology Demonstrated At IBM Edge Conference

Announcement : http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=ca&infotype=an&appname=iSource&supplier=877&letternum=ENUSZG12-0163

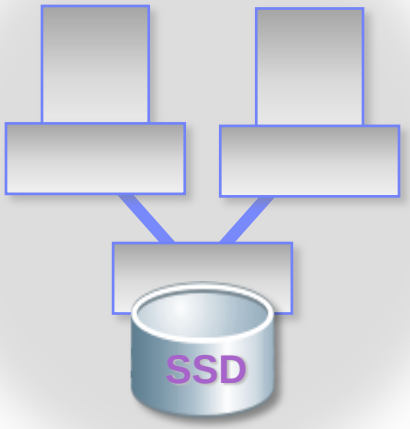
1+1 = 3: Extending Today's Storage Models



Strong Performance

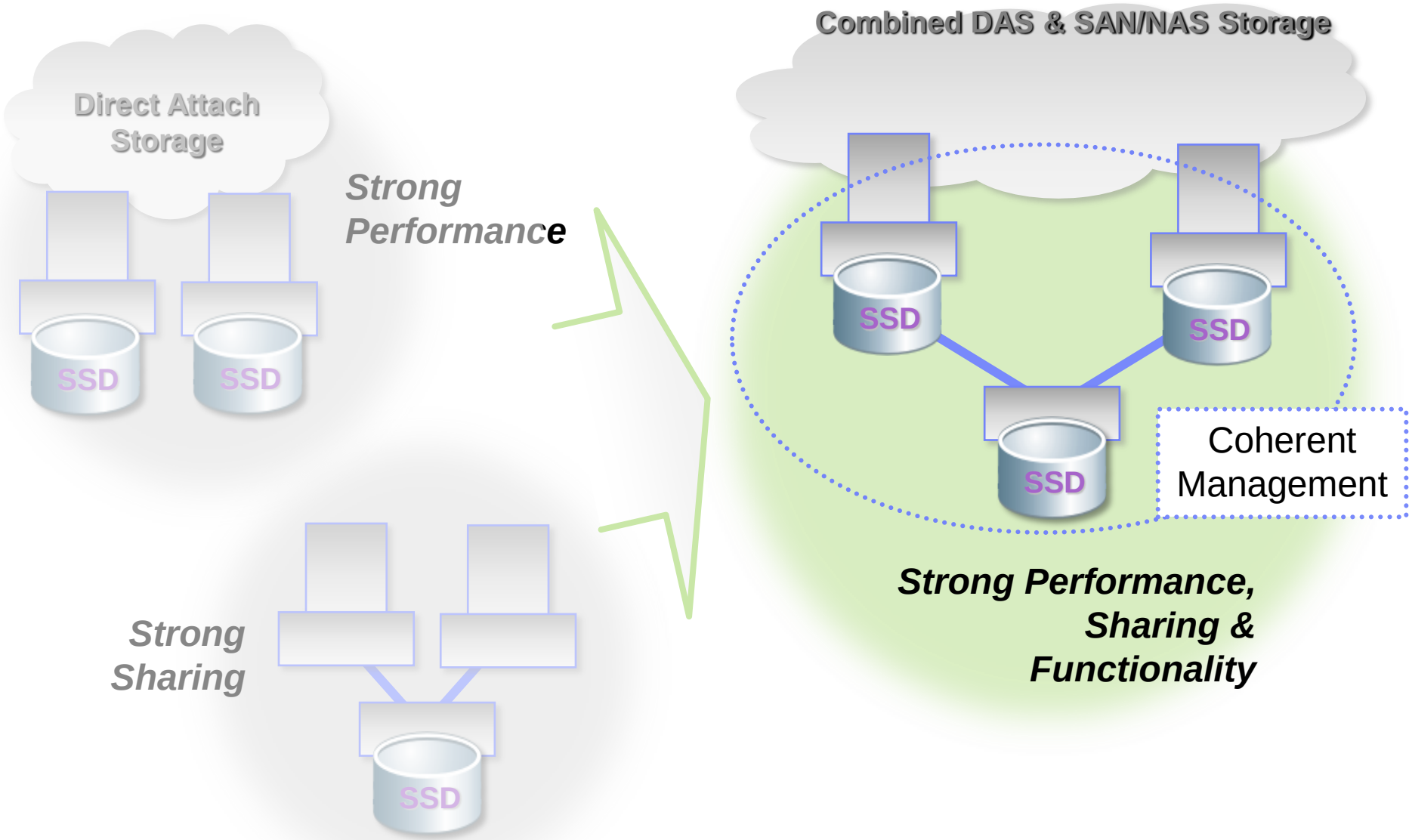
Messaging, social networking, content delivery applications

Strong Sharing & Functionality



Many Enterprise SAN Storage Customers (20+% of DS8K)

1+1 = 3: Extending Today's Storage Models



THANK YOU

Understanding Flash based SSD performance

- Flash media can only do one the following three things: Read, Erase, Program
- IO Read -> Flash Read, IO Write -> Flash Erase and Flash Program
- Erase cycle is very time consuming (in msec)
- Major latency difference for IO Read operation (50-100 usec) versus IO Write (15-200+usec) operation
- Flash based SSD device requires storage virtualization to deal with undesirable flash properties, erase latency and wear-leveling.
- Storage virtualization techniques typically used are : Relocate on write, batch write operation and , over provisioning.

How to manage unpredictable SSD performance?

- Characterize SSD micro-performance.
- Manage different “peaks” and “valleys” of IO performance landscape.
- Develop traffic shaping algorithm to fit different SSD performance profile
- Develop APIs to improve visibility of SSD runtime characteristics with applications and system.