

Scalable Intelligent Video Server System

<i>Title</i>	Integration-ready high density SATA storage system
<i>Revision</i>	B
<i>Deliverable #</i>	D 6.1
<i>Author</i>	Ian Cook, Dave Davis, Will Hughes, Jerry Hoetger
<i>Company</i>	Xyratex
<i>Date</i>	27 th February 2006
<i>Filename</i>	D6.1-v1-0.doc
<i>Dissemination[†]</i>	PU

REVISION	DATE	DESCRIPTION
A	27 Feb 2006	Initial release
B	30-mar-2006	Publishable update

[†] **CO** = Confidential (only for members of the consortium + EC); **RE** = Restricted to a stated circulation list (+ EC)
[replace this footnote with the list]; **PP** = Restricted to other FP6 participants (+ EC); **PU** = Public

“Scaleable Intelligent Video Server System” – FP6 Contract # 002075

TABLE OF CONTENTS

1 INTRODUCTION	3
2 PRODUCT OVERVIEW	3
3 OVERALL PACKAGING	4
4 DISK DRIVES IN CARRIER PACKAGING	5
5 SHOCK AND VIBRATION	6
5.1 CONCERNS IN A SATA ENCLOSURE	6
5.2 SHOCK AND VIBRATION PERFORMANCE MEASUREMENTS	7
6 THERMAL PERFORMANCE	9
6.1 REQUIREMENTS.....	9
6.2 MEASUREMENTS	10
7 PERFORMANCE SUMMARY	14

List of Figures

Figure 2-1 – RA-4835 enclosure – out of rack	4
Figure 2-2 – RA-4835 enclosure – rear view	4
Figure 3-1 – RA-4835 EMC Gaskets	5
Figure 4-1 – RA-4835 Drive Carrier	5
Figure 6.2-1 – Disk Drive location in the RA-4835 enclosure	11
Figure 6.2-2 – Computer simulation of the RA-4835 enclosure air flow analysis.....	11

List of Tables

Table 5.2-1 – Demonstration of typical operations per second (IOPS) in an enclosure with 24 concurrently operational drives	7
-----------------------------------------------------------------------------------------------------------------------------------	---

1 INTRODUCTION

The purpose of this delivery is to describe the development of the highest capacity Serial ATA, (SATA), drive enclosure available in the market. This is the first enclosure to package 48 SATA drives in a 4U enclosure. The enclosure uses N+1, (three), power supplies; each PSU has two high capacity fans at the back of the unit to provide power and cooling to the enclosure. The enclosure is mounted on rails and slides in and out of a standard 483mm, (19”), wide rack for maintenance and can be maintained while operational with Serial Attached SCSI, (SAS), Fibre Channel IO modules or Fibre Channel RAID controllers. Power supplies and IO modules are accessible from the back of the enclosure while locked in to the rack. The 48 drives are vertically mounted and accessible when the enclosure is pulled out of the rack. The resulting package, while operating all 48 SATA drives in a normal air conditioned space, will keep all 48 drives, power supplies, (PSU’s), and IO modules functioning within stipulated thermal and vibration levels, and can withstand documented shock impacts. Non-operational vibration and shock impact environments are also met using Xyratex standards and shipping procedures. This report explains these requirements and how the Xyratex RA-4835 enclosure measured against these requirements.

2 PRODUCT OVERVIEW

The Xyratex RA-4835 enclosure sets a new standard in the design of mass storage enclosures and provides unsurpassed disk storage capacity in a single enclosure. Housing 48 SATA drives in 4U of rack space, the RA-4835 enclosure can store 24TB of data in a single unit utilising SATA drive technology. With a combination of high performance controllers and high density packaging the High Availability, (HA), design of the RA-4835 enclosure increases the flexibility of storage deployment to the end user.

As the importance of Information Lifecycle Management (ILM) grows, the RA-4835 enclosure provides an ideal solution to the secondary disk storage needs of compliance, disk to disk backup, Massive Array of Idle Disks, (MAID), and Content Addressable Storage, (CAS), environments. The innovative enclosure design of the RA-4835 ensures that for the first time customers will have a mass storage enclosure that allows drives to remain on-line and active during servicing or maintenance. In addition, an LCD display located at the front of the enclosure provides real time status and reporting on all aspects of the 48 drives and overall enclosure health.

Each of three power supplies generates 450W of power. Two independent 200 to 240V (50/60Hz) power inputs each feed one of the three power supplies and a redundant AC switch using bifurcated power cords. The redundant AC switch then supplies power to the third power supply. Loss of either power source results in at least two power supplies receiving mains power, without the interruption to the operation of the enclosure.

The feature set of the RA-4835 enclosure includes:

- 48 SATA drives in a 4U enclosure
- FC-SATA JBOD/SBOD configuration
- OEM customizable front panel and LCD display
- Front mounted LCD display of drive status and error conditions
- N+1 (3) power supply units – no impact from a single failure
- HA design – enclosure remains on-line during maintenance
- Fans, power supplies and I/O modules serviceable from rear of unit
- Sliding mechanism provides easy access to drives for maintenance
- Vertically mounted drives with individual drive carriers



Figure 2-1 RA-4835 enclosure – Out of rack

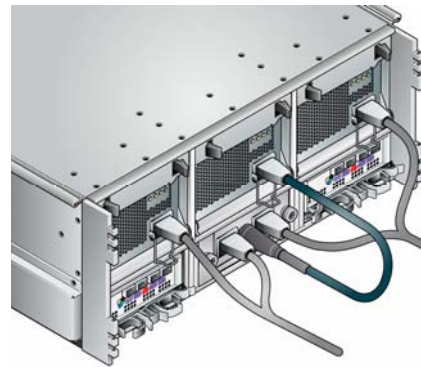


Figure 2-2 RA-4835 enclosure – Rear view

3 OVERALL PACKAGING

The RA-4835 enclosure was designed to meet the most demanding operational and non-operational environments used by customers today.

The enclosure pulls cool air through the front, over and around the drives and down through the centre over the channel card positions. When the enclosure is in the rack a metal lid seals the top via a special gasket on the rim of the side panels, from the bezel to the rear top of the enclosure. This seal ensures a maximum flow of cool air through the unit. The enclosure can be pulled out for maintenance for a maximum of 5 minutes in every 15 minutes, while remaining within required thermal tolerances. A space between each vertically mounted SATA drives provides for air flow between the disk drives. Air is also pulled through the IO modules located below the power supplies and out to the rear of the enclosure through the PSU's.

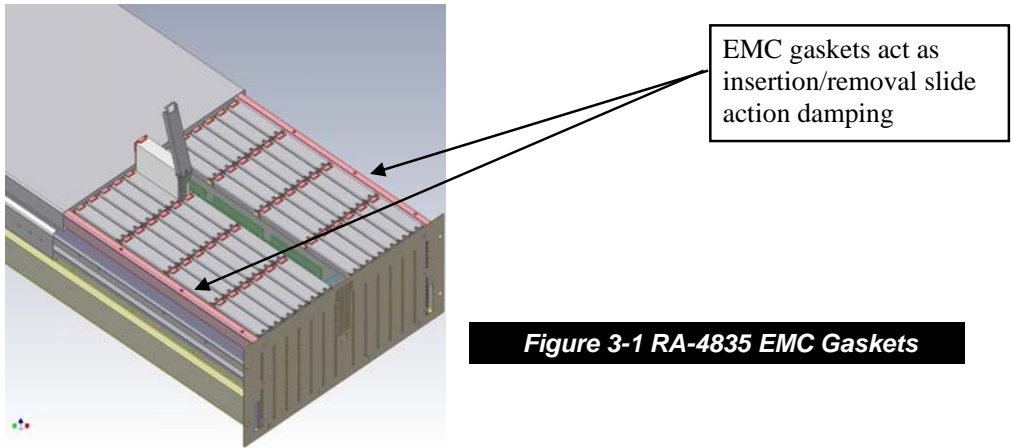


Figure 3-1 RA-4835 EMC Gaskets

4 DISK DRIVES IN CARRIER PACKAGING

Xyratex has patented the design of the low profile SATA drive carrier used in the RA-4835 enclosures. This design was crucial in fitting 48 drives in a 4U package, while meeting cooling, shock, and vibration tolerances. Each drive used in the RA-4835 must be encased in the special Xyratex drive carrier. This low profile drive carrier ensures the secure fitting of the drive in to the enclosure and ensures proper electrical connections to the base plane. A light pipe transmits a light signal from the SCSI Enclosure Services, (SES), manager in the enclosure to the top of the drive carrier when inserted and operational. This light pipe is used for positive identification of drives for maintenance or replacement. A locking mechanism ensures drives in carriers stay properly connected to the base plane. Special constrained layer springs on the side of the carriers minimize the vibration transfer and ensure the enclosure can continue to operate as disk drives spin and actuators move to read and written to the platters in the SATA disk drives.

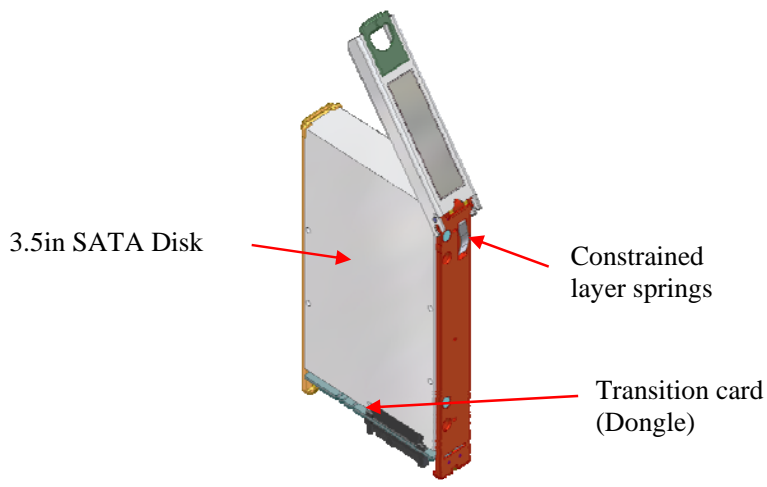


Figure 4-1 RA-4835 Drive Carrier

5 SHOCK AND VIBRATION

5.1 Concerns in a SATA enclosure[‡]

“As SATA based storage networks arrive on the scene, engineers must address system level reliability. Storage enclosures must be carefully managed to guard against subtle design problems caused by vibration performance degradation, signal integrity induced data failure and heat induced catastrophic failure.”

One critical imperative when building a storage chassis that supports a large number of disk drives like the RA-4835 enclosure is properly managing vibration effects generated by the drive's actuator system. When placed in an enterprise environment where millions of I/Os are the daily norm, the continuous disturbance caused by an actuator moving across the disk surface can cripple the drive's ability to relay data quickly and correctly to and from its host. These dynamics not only affect the performance of an individual drive but can also adversely affect the neighbouring disk drives within a storage enclosure.

Understanding the magnitude of this performance impact requires a series of tests that compare the maximum I/O rate of a single disk drive securely constrained to that of the same drive mounted into a populated storage enclosure. The difference of the two measured I/O rates constitutes the percentage of performance degradation contributed by the storage enclosure. In storage industry language, this measurement is called the rotational vibration interference (RVI).

For enterprise based SCSI and Fibre Channel, drive makers have incorporated mechanical and firmware solutions to counter a lot of the influences of drive vibration; these enhancements add cost and contribute to the high price normally accompanying such drives. To maintain a low price point for their serial ATA products some drive makers have opted to leave out pieces of the technology used to counter the effects of drive vibration.

The capabilities of the SATA drives posed a challenge for Xyratex when designing the RA-4835 enclosure. Xyratex needed to design a SATA enclosure as a price-sensitive alternative for near-line applications. No longer is it a case of wrapping some sheet metal around a bunch of disk drives, attention had to be paid to the structural stability of the chassis and to the design of the disk drive carrier and its interface within the assembly.

As serial signalling rates grow, routing and managing the signals as they pass through cables and traces on boards become critical. In the parallel world, these interfaces were quite tolerant and many approaches were taken that diverged from accepted specifications. The RA-4835 enclosure design needed to ensure proper fitting of SATA drives without the luxury of tolerant interfaces and meet or exceed shock and vibration tolerances for qualified disk drives.

[‡] Electronic Engineering Times, “Serial ATA nets need solid chassis design”, April 16 2004. Peter Herz, Mike Bell, 3ware Inc

5.2 Shock and Vibration Performance Measurements

Xyratex engineering evaluated the performance of the Hitachi Ultrastar 7200 RPM 500GB Serial ATA drives in the RA-4835 FC-SATA enclosure.

The first test performed was a mix of read and write IO operations to half of the drives in a fully populated 48 drive RA-4835 FC-SATA enclosure installed in a standard rack. Few applications would be reading and writing to more than 24 SATA drives in one RA-4835 enclosure at the same time. The Hitachi 7200 RPM 500G Ultrastar when ideally mounted performs 69 random IOPS, given a 512 byte transfer size, queue depth of 4, 66% reads, 33% writes workload.

All 24 Hitachi Ultrastar 7200 RPM 500G drives in Xyratex RA-4835 carriers without dual path (active/passive) dongles performed at the expected 69 IOPS. The same test run with Hitachi Ultrastar 7200 RPM 500G drives in Xyratex RA-4835 carriers with dual path (active/passive) dongles resulted in each drive averaging 67.7 IOPS, approximately 98.8% of the ideal performance of that drive. The results are summarised in Table 5.2-1

Type	Connection	Power in Watts	Typical Spec RV Tolerance rad/s/s rms	Typical Operational Shock Tolerance g ms	Typical Operations per second (IOPS)
SATA	Direct Dock or Dual Path Dongle	13	12	10g 11ms 55g 2ms	69

Table 5.2-1 – Demonstration of typical operations per second (IOPS) in an enclosure with 24 concurrently operational drives

The Xyratex guidelines for vibration performance in enclosures are a minimum of 85% of the ideal disk drive performance figures and the rejection of any drive in carrier design that does not perform at this level or greater.

Good performance is specified as any drive in carrier design that performs at 95% of ideal performance. The Hitachi Ultrastar 7200 RPM 500G drive in a RA-4835 carrier and enclosure significantly exceeds that guideline.

A second test in this section was performed to evaluate operational vibration levels when applied to the base of the RA-4835 enclosure in the rack. The same workload was used as in the previously described test, IO operations to all 48 drives in the RA-4835 enclosure, 512 byte data transfer size, 66% reads, 33% writes and a queue depth of 16. The results of this testing, shown in Table 5.2-2, showed no degradation of the drive performance.

Axis	Vibration Level G rms Random	System Operations per second
Benchmark	0	2124
X	0.1	2157
Y	0.1	2150
Z	0.1	2155

Table 5.2-2 – Demonstration of operational vibration levels on system operations per second (IOPS) when applied to three axis

The third test performed measured the change in data transfer response time when applying specific mechanical impact to the RA-4835 enclosure installed in a rack. The workloads applied during tests were 100%, sequential read and write, 256 kbyte transfers. The results of this test, summarised in Table 5.2-3, show that there were no increase in response times due to the applied shock level.

	Max response time ms	
	RA-4835 in Rack Normal	RA-4835 in Rack with 10x impacts
Sequential Write	2078	2021
Sequential Read	1848	1762

Table 5.2-3 – Demonstrating of mechanical impact levels on data transfer response time

The last test in this section measured the shock to the individual disk drives when sliding out and re-inserting the RA-4835 enclosure in a standard rack. The RA-4835 enclosure can be pulled out for maintaining individual disk drives while in operation and transferring data to and from connected host environments. This test is important for users demanding availability to drives in the RA-4835 enclosure while maintaining the vertical mounted SATA drives. Table 5.2-4 provides a summary of the results for this test.

Drawer Shock Levels	Front to Back X axis	Side to Side Y axis	Up and Down Z axis
Shock during drawer removal	5g <1ms	18g <1ms	13g <1ms
Shock during drawer insertion	3g <1ms	6g <1ms	5g <1ms
Drive Operational Shock Limits	55g 2ms	55g 2ms	55g 2ms

Table 5.2-4 – Demonstration of mechanical force applied to disk drives during maintenance

This test proves that normal maintenance operations (sliding out and re-inserting the RA-4835 enclosure in a standard rack) will not cause shocks to any of the individual 48 drives in the enclosure beyond the manufacturer defined threshold of the Hitachi Ultrastar 7200 RPM 500G SATA drive.

6 THERMAL PERFORMANCE

6.1 Requirements

As storage systems migrate out of the controlled environment of a data centre and into densely packed data closets, the efficiency of the enclosure's cooling system also becomes crucial. Historically, designers have tried many approaches to manage airflow and heat dissipation within a storage chassis. Most of these systems use axial or radial fans to pull air across the hot components within the box in an effort to keep the components below their maximum operating temperature. Ideally, a well designed storage enclosure will have a cooling system that produces an unimpeded uniform flow of air through the chassis. It will provide just enough cooling to achieve maximum mean time between failure, (MTBF), of the heat sensitive components housed within the box.

Cooling requirements for SATA drives are no greater than for other drive architectures, so current design practices can be used for SATA enclosures. As long as those established design practices are used, the likelihood of premature drive and system failure can be greatly reduced and the system's online availability increased. The challenge faced by Xyratex engineers when designing the RA-4835 enclosure to work in a +5 C to +35 C environment was to pack 48 drives in a 4U chassis and still meet the cooling requirements of all expected SATA drives; this was achieved by optimizing the drive air gaps to provide the minimum drive temperature rise.

The designers started off by specifying the use of 500GB, 7200 RPM, SATA drives in a RA-4835 enclosure, fully populated with 48 individual drives in carriers with the requirement to:

- 1) meet MTBF requirements at 25 C ambient, and
- 2) not exceed the maximum disk drive case temperature allowable with a maximum of +35 C enclosure ambient temperature, with and without a fan failure.

This was to ensure that the system was maintained within the operating guidelines from the various disk drive manufacturers.

6.2 Measurements

Xyratex measured air flow through enclosures at 2 different ambient temperatures, the consistent air temperature in the chamber where the enclosure resides during the test completely enveloping the enclosure, +25 C and +40 C. The upper test limit of +40 C was chosen to ensure sufficient margin existed above the recommended maximum operational limit of +35 C.

Three test conditions with random workloads were performed; the first with 6 fans running at 50% fan speed, +25 C ambient temperature (normal conditions), the second with 6 fans running at 100% fan speed, +40 C ambient temperature (maximum high temperature conditions without fan failure), and finally with 5 fans running at 100% fan speed and +40 C ambient temperature (maximum high temperature condition with fan failure). Table 6.2-1 shows the actual result that we obtained from these tests.

Workload	Random	Random	Random	Recommended Maximum Operating Temperature (C)
Fans	Normal	Normal	Fan Fail	
Speed	50% X 6	100% x 6	100% x 5	
IOPS	3088	3090	3081	
Ambient (C)	+25	+40	+40	
Heatsink on IO Module	49.7	60.8	68.4	75
Altera Chip	42.5	54.5	61.6	75
Drive location 3,4	47.1	56.1	59.4	65
Drive location 4,1	31.5	44.8	45.8	65
Drive location 6,4	39.6	50.0	53.6	65
Drive location 7,4	50.1	58.5	63.5	65
Drive location 9,3	45.0	54.8	58.5	65
Drive location 9,4	51.3	59.4	64.6	65
Drive location 11,4	49.0	57.5	62.5	65
Exit Air	42.8	55.5	60.6	70
IO Module local	41	53	60	

Table 6.2-1 – Temperature measurements taken at various locations throughout the RA-4835 enclosure

Figure 6.2-1, below, shows the location of the disk drives in the RA-4835 enclosure.

Sumo drive position layout												
RA-4835 rear of enclosure												
1,4	2,4	3,4	4,4	5,4	6,4		7,4	8,4	9,4	10,4	11,4	12,4
1,3	2,3	3,3	4,3	5,3	6,6		7,3	8,3	9,3	10,3	11,3	12,3
1,2	2,2	3,2	4,2	5,2	6,2		7,2	8,2	9,2	10,2	11,2	12,2
1,1	2,1	3,1	4,1	5,1	6,1		7,1	8,1	9,1	10,1	11,1	12,1
Left Hand Side						Right Hand Side						
RA-4835 front												

Figure 6.2-1 – Disk Drive location in the RA-4835 enclosure

All measured values of components in the “normal condition” thermal test were well within limits, as were the test results obtained at the maximum “high” temperature with and without fan failures. These tests indicated that the RA-4835 enclosure has a slight temperature gradient towards the rear of the enclosure; this is as originally designed and aligns with our computer simulation models, diagrams of which are shown in Figure 6.2-2 below.

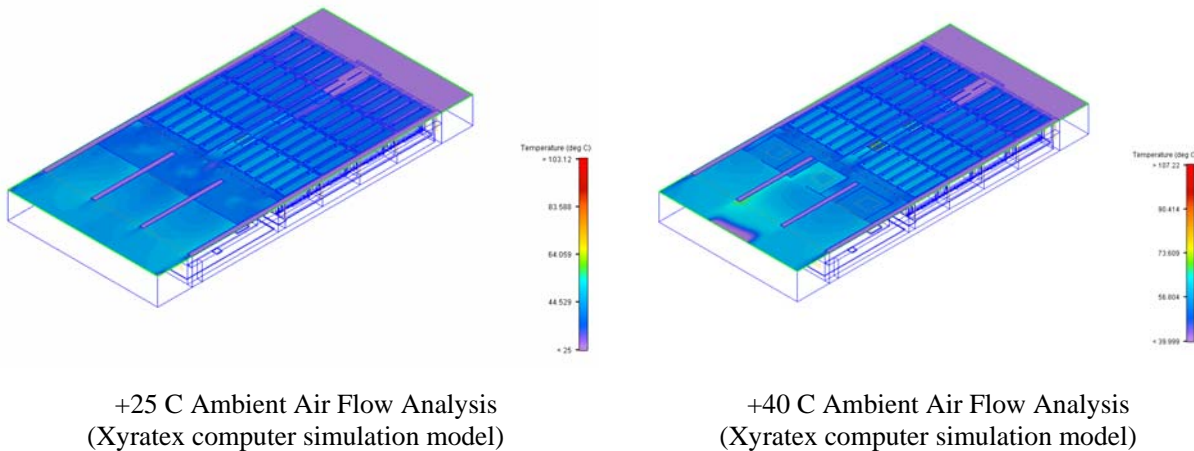


Figure 6.2-2 – Computer simulation of the RA-4835 enclosure air flow analysis

Analysis was also undertaken to determine the thermal characteristics for the RA-4835 enclosures when used in a rack environment. Six enclosures were stacked in an enclosed rack with blanking plates fitted above them to reduce air re-circulation. The temperatures were monitored at the same location in each enclosure during idle operation with the results detailed in Table 6.2-2, showing that there were minor differences.

RA-4835 No.	Controller chip temperature	Internal ambient temperature	Temperature delta
1 (top shelf)	43	29	14
2	43	30	13
3	43	29	14
4	41	29	12
5	42	28	14
6 (bottom shelf)	41	28	13

Table 6.2-2 – Temperature measurements taken at various locations throughout the RA-4835 enclosure

Based on these results enclosures 3 and 5 were selected for further investigation, as they represented the highest temperature deltas in the rack in the idle state, and thermocouples were fitted to disk drive carriers throughout the enclosure.

The enclosures were then tested with four different configurations running IOmeter with the following test conditions:

- Queue Depth – 4
- 67/33 read/write
- 512 byte transfers
- 100% random

The testing was conducted at ambient temperatures of both +25 C and +40 C under normal operating conditions as well as with one forced fan failure. The results obtained from these tests are shown in Table 6.2-3.

Test No.	Fan Speed	Ambient external	RA-4835 Enclosure 3 temperatures (C)					RA-4835 Enclosure 5 temperatures (C)						
			Ambient internal reported	Row				Controller Internal temp	Ambient internal reported	Row				Controller internal temp
				Front		Rear				Front		Rear		
1	2	3	4	1	2	3	4	1	2	3	4			
1	60%	25	26	30	35	41	44	42	25	29	34	39	43	42
2	Fan Fail	25	26	31	36	41	45	40	25	30	35	40	44	42
3	100%	40	41	44	49	53	56	54	40	44	48	51	55	54
4	Fan Fail	40	41	46	51	56	60	55	40	45	50	55	59	57

Maximum permissible controller temperature = +75 C

Maximum permissible drive case temperature = +65 C

Table 6.2-3 – Temperature measurements taken at various locations throughout the RA-4835 enclosure

The conclusion from these tests was that there are no issues seen when running the RA-4835 under any of the normal or forced failure test conditions that were undertaken.

7 PERFORMANCE SUMMARY

The Xyratex RA-4835 enclosure meets or exceeds all Xyratex and drive manufacturer standards using Hitachi Ultrastar 7200 RPM 400G and 500G SATA drives. Several key points are worth special consideration from the Xyratex testing process:

- Normal operation (+25 C ambient), and a random access HDD workload resulted in HDD temperatures between 32 - 50 C with 60% fan speed
- The maximum high heat conditions of +40 C ambient, (+5 C above the recommended operational limit for the enclosure), with a fan failure and a random access HDD workload resulted in HDD temperatures between +46 to +64 C.
- Random HDD workloads resulted in HDD temperatures of +2 C hotter than sequential HDD workloads.
- Sequential HDD workloads resulted in controller temperatures of +2 C hotter than random HDD workloads
- Each row of drives starting at the front of the enclosure progressing backwards towards the rear power supplies showed a gradient casting temperature of +5 C/row or ~15 C front to back
- The FC SBOD module used in the RA-4835 JBOD/SBOD in 40 C fan fail operates within satisfactory limits.
- Multiple enclosures can be used within specification in a six rack configuration.