Project Olympus
DX-88 Disk Expansion Chassis

Author:
Larry Cannon, Senior Mechanical Engineer, Microsoft

© 2017 Microsoft Corporation.

As of November 1, 2017, the following persons or entities have made this Specification available under the Open Web Foundation Final Specification Agreement (OWFa 1.0), which is available at http://www.openwebfoundation.org/legal/the-owf-1-0-agreements/owfa-1-0

Microsoft Corporation.

You can review the signed copies of the Open Web Foundation Agreement Version 1.0 for this Specification at Project Olympus License Agreements, which may also include additional parties to those listed above.

Your use of this Specification may be subject to other third party rights. THIS SPECIFICATION IS PROVIDED "AS IS." The contributors expressly disclaim any warranties (express, implied, or otherwise), including implied warranties of merchantability, non-infringement, fitness for a particular purpose, or title, related to the Specification. The entire risk as to implementing or otherwise using the Specification is assumed by the Specification implementer and user. IN NO EVENT WILL ANY PARTY BE LIABLE TO ANY OTHER PARTY FOR LOST PROFITS OR ANY FORM OF INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES OF ANY CHARACTER FROM ANY CAUSES OF ACTION OF ANY KIND WITH RESPECT TO THIS SPECIFICATION OR ITS GOVERNING AGREEMENT, WHETHER BASED ON BREACH OF CONTRACT, TORT (INCLUDING NEGLIGENCE), OR OTHERWISE, AND WHETHER OR NOT THE OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

CONTRIBUTORS AND LICENSORS OF THIS SPECIFICATION MAY HAVE MENTIONED CERTAIN TECHNOLOGIES THAT ARE MERELY REFERENCED WITHIN THIS SPECIFICATION AND NOT LICENSED UNDER THE OWF CLA OR OWFa. THE FOLLOWING IS A LIST OF MERELY REFERENCED TECHNOLOGY: INTELLIGENT PLATFORM MANAGEMENT INTERFACE (IPMI); I^2C IS A TRADEMARK AND TECHNOLOGY OF NXP SEMICONDUCTORS ; EPYC IS A TRADEMARK AND TECHNOLOGY OF ADVANCED MICRO DEVICES INC.; ASPEED AST 2400/2500 FAMILY PROCESSORS IS A TECHNOLOGY OF ASPEED TECHNOLOGY INC.; MOLEX NANOPITCH, NANO PICOBLADE, AND MINI-FIT JR AND ASSOCIATED CONNECTORS ARE TRADEMARKS AND TECHNOLOGIES OF MOLEX LLC; WINBOND IS A TRADEMARK OF WINBOND ELECTRONICS CORPORATION; NVLINK IS A TECHNOLOGY OF NVIDIA; INTEL XEON SCALABLE PROCESSORS, INTEL QUICKASSIST TECHNOLOGY, INTEL HYPER-THREADING TECHNOLOGY, ENHANCED INTEL SPEEDSTEP TECHNOLOGY, INTEL VIRTUALIZATION TECHNOLOGY, INTEL SERVER PLATFORM SERVICES, INTEL MANAGABILITY ENGINE, AND INTEL TRUSTED EXECUTION TECHNOLOGY ARE TRADEMARKS AND TECHNOLOGIES OF INTEL CORPORATION; SITARA ARM CORTEX-A9 PROCESSOR IS A TRADEMARK AND TECHNOLOGY OF TEXAS INSTRUMENTS; GUIDE PINS FROM PENCOM; BATTERIES FROM PANASONIC. IMPLEMENTATION OF THESE TECHNOLOGIES MAY BE SUBJECT TO THEIR OWN LEGAL TERMS.

Revision History

Date	Description
11/01/2017	Version 1.0 OCP Release

Contents

Figures

Tables

Summary

1.1 Project Olympus DX-88 Overview

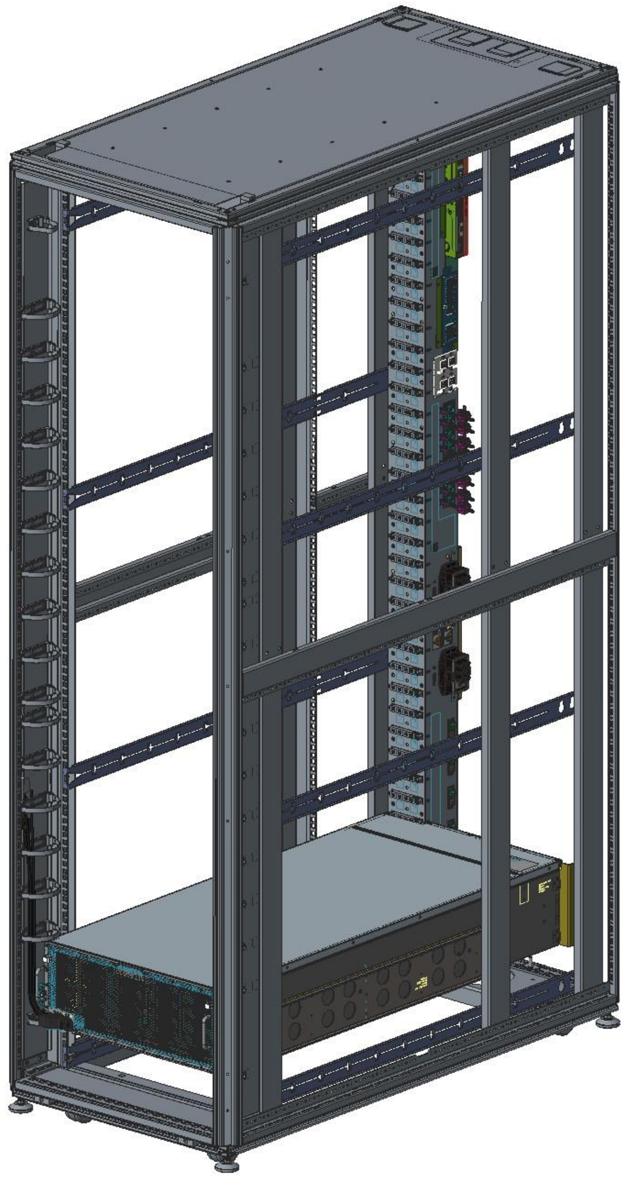

Figure 1: Project Olympus DX-88 Rack View.

The Project Olympus DX-88, short for Disk Expansion 88, is shown in Figure 1, installed in the Project Olympus Chassis with the Power and Management Distribution Unit (PMDU). Details of the Project Olympus Chassis and the PMDU are provided in separate Microsoft OCP documents.

The Project Olympus compute node is a 1U (or in some variants 2U) Sky Lake-based server that can function as a storage head node. The Project Olympus DX-88 is a 4U, high capacity, hot plug device that supports 88 near-line 3.5” HDDs, split logically into four 22 HDD segments. The compute node and the storage drawer blind mate to the PMDU.

The DX-88 contains two hot-plug power supplies that are remotely connected to the PMDU and consume three single-phases of AC power, and provide 12VDC output.

1.2 High Level DX-88 Architecture

The DX-88 is designed to support three high level configurations.

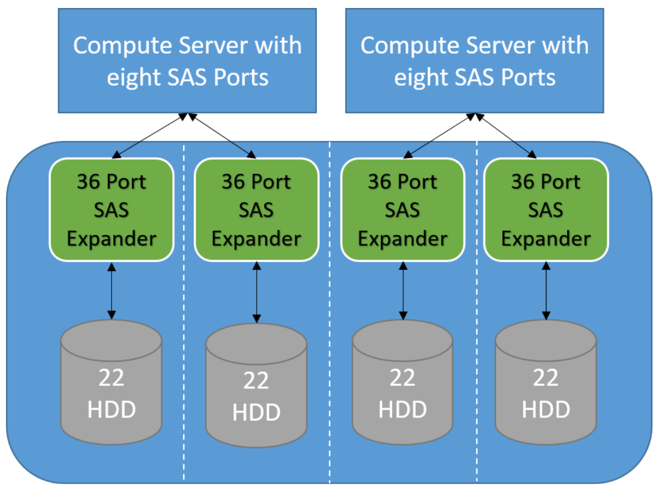

Figure 2: DX-88 Standard Storage

The first configuration is standard storage. In this configuration, the DX-88 is logically split into four twenty-two HDD segments, and each server connects to two of the segments via an eight SAS port host bus adapter.

The second configuration is hot storage. In this configuration, the DX-88 is logically split into four twenty-two HDD segments, and each server connects to one of the segments via an eight SAS port host bus adapter.

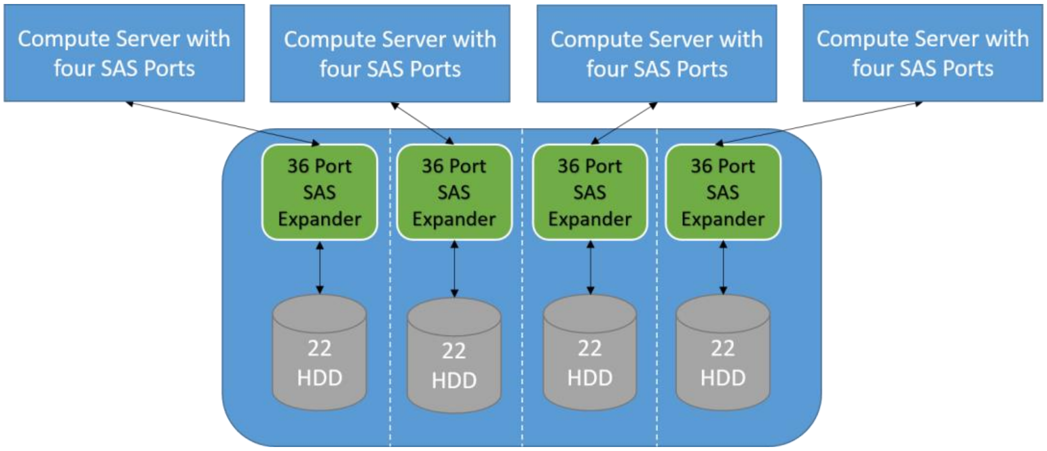

Figure 3: Hot Storage

The third configuration is cool storage, and is architecturally enabled in the design. In this configuration, the DX-88 is logically split into four twenty-two HDD segments, and one server connects to all four segments via a sixteen-SAS port host bus adapter.

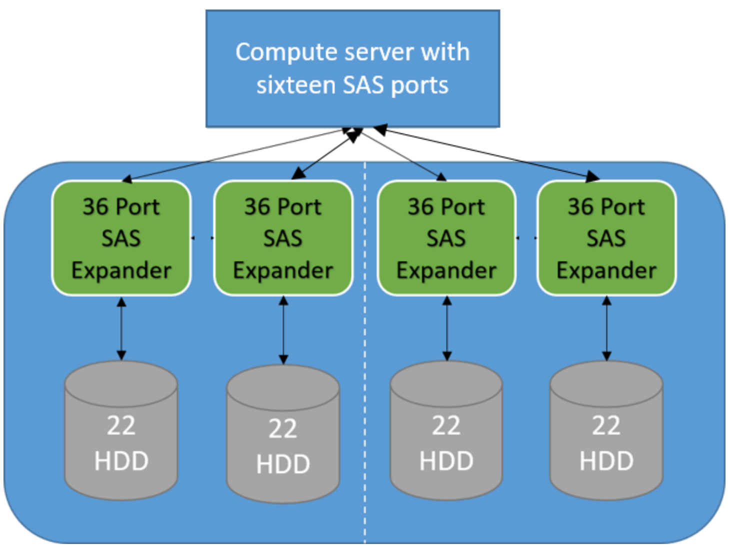

Figure 4: Cool Storage

1.3 DX-88 Features

Table 1: DX-88 Features

Function	Details
HDD count	88 3.5" near-line SATA, split in four 22 HDD segments
Expander	4x 36 port 12Gb expander.
Power	N+N 1650W redundant hot plug power supplies; phase balanced.
Cooling	N+1 hot swap fans
LEDs	LEDs for status of power supplies and individual HDDs.

Size	4U height
Weight	124lbs excluding HDDs and PSUs.

The following is a list of the primary features supported by the DX-88.

- Supports full rack width, 4U height (7”), DX-88 assembly
- Supports blind-mate power with Project Olympus Rack with PMDU
- Supports cold aisle cabling for I/O and Ethernet management
- Supports cold-aisle, tool-less, hot-swap servicing of critical components (HDDs, Fans, PSUs)
- Supports cold-aisle, non-hot-swap servicing of all active PCBAs
- Supports redundant N+1 integrated Project Olympus PSUs
- Supports up to four server nodes using four individual expander segment back planes
- Supports four, dual rotor (N+1) 92mm hot-swappable fans

1.4 Physical Description

The DX-88 is 4U tall. Dimensions are shown in Figure 5. The entire DX-88 drawer slides out to provide access the hot plug HDDs. The mass of a fully populated DX-88 is approximately 285lbs.

Shown in Figure 6 are the locations of 88 HDD slots, 2 PSUs and 4 Fans.

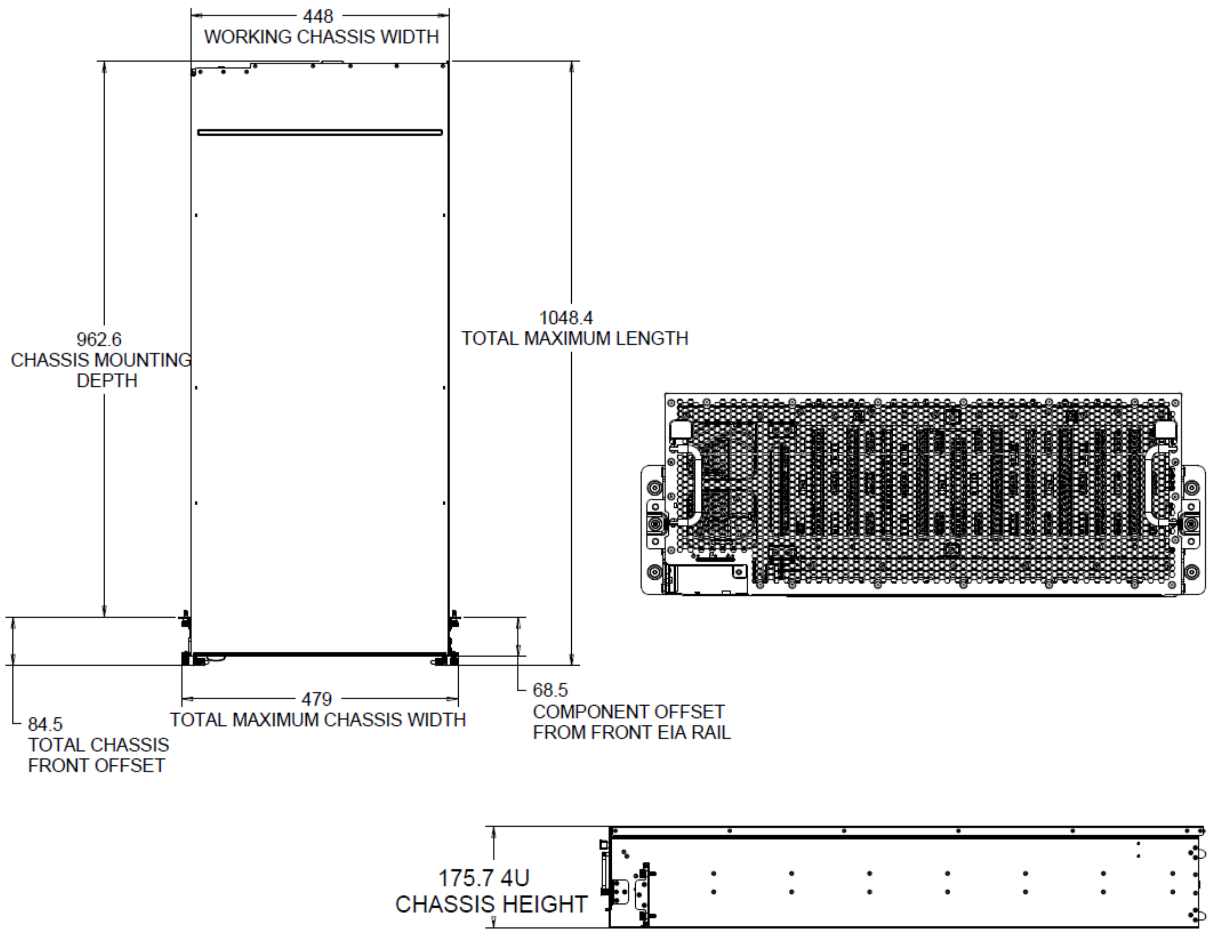

Figure 5: DX-88 Dimensions

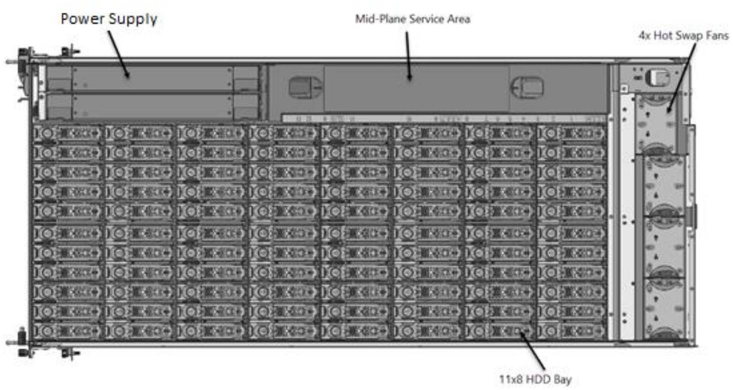

Figure 6: DX-88 Top View (cover omitted for clarity)

1.5 Cable Management

The DX-88 connects to a PMDU at the back left of the rack. When the DX-88 is moved forward to hot service HDDs, PSUs, or fans, a cable management arm is deployed to ensure the DX-88 remains connected to AC power and low-speed management signals.

Each DX-88 connects to a rack management switch (RJ45) and one, two, or four servers. SAS and GbE cables are routed on the left side of the rack (as seen from the front aisle).

Figure 7: DX-88 Top View; Components Omitted for CMA Visibility

1.6 Rear Panel

The Rear Panel is shown in Figure 8. The Rear Panel provides the following features.

- PSU connection to PMDU
- Four 92mm fans
- PSU exhaust

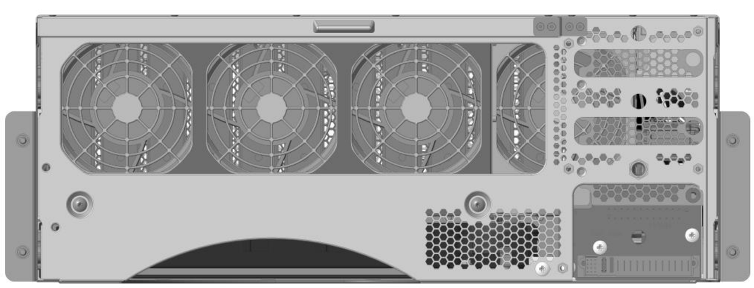

Figure 8: DX-88 Rear Panel

1.7 Blind Rack Installation

During integration the DX-88 chassis slides on top of preinstalled static rails. As the DX-88 slides back into the rack, the rear AC connection blind-mates to a connector on the PMDU.

After connecting the DX-88 to the PMDU, the storage chassis is locked into the rack with a set of toolless fasteners.

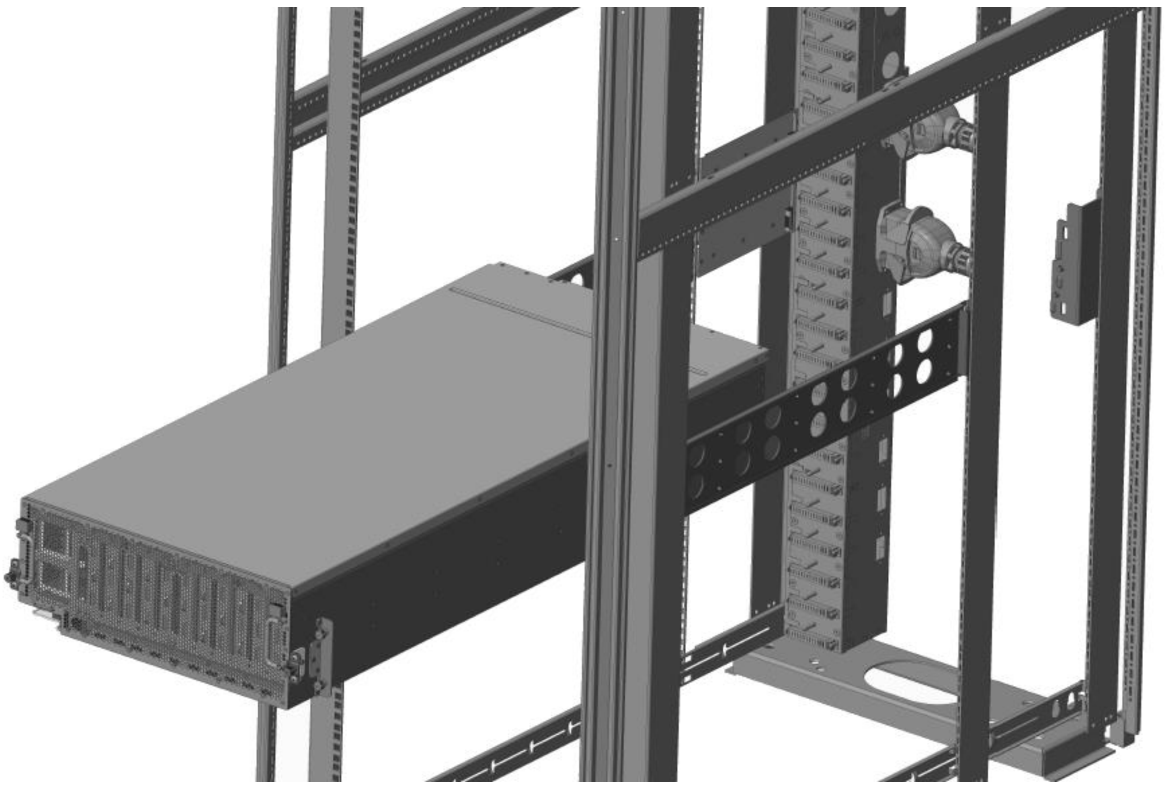

Figure 9: DX-88 Partially Installed in Rack

1.8 Full Cold Aisle Service

Once installed into the rack, the chassis' shell is static and the drawer can be fully extended from the rack on ball bearing slides. When extended, all hot-swap components and active PCBAs can be serviced.

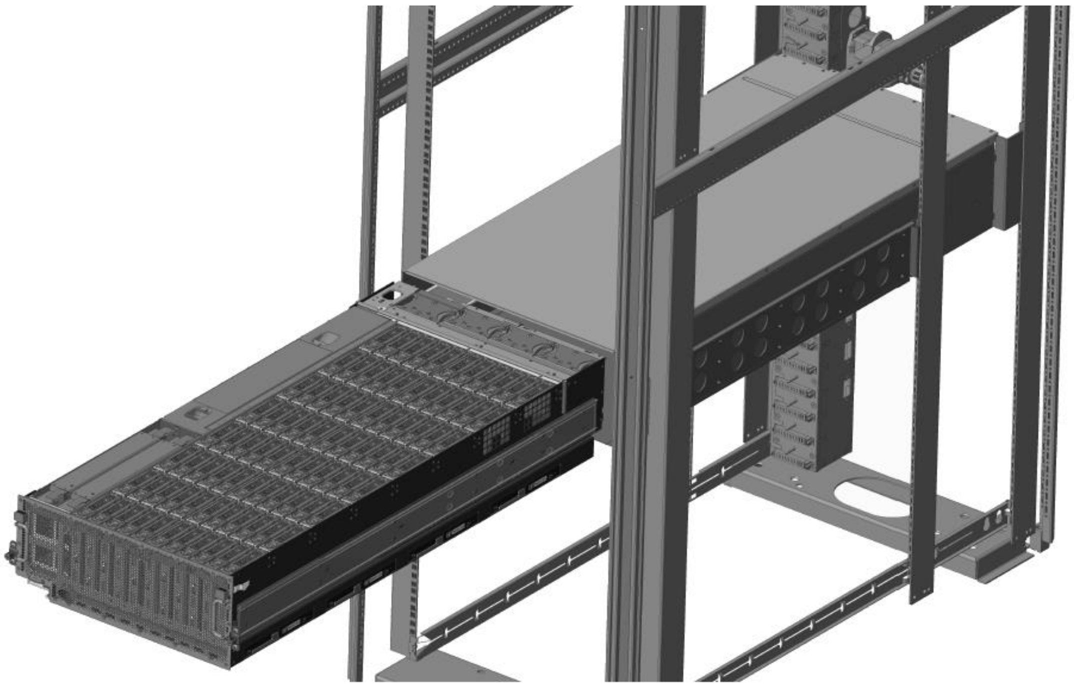

Figure 10: DX-88 Service Position

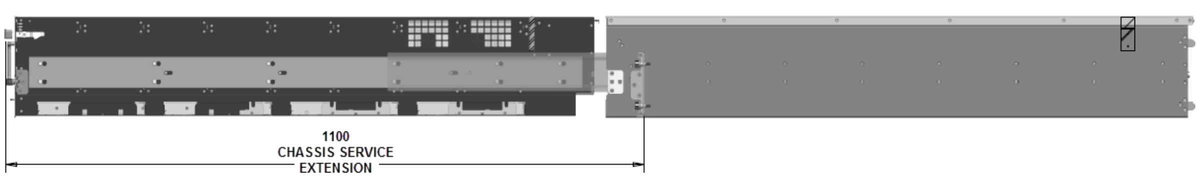

Figure 11: DX-88 Service Extension

1.9 Serviceability

The DX-88 provides toolless access to hot swap components, which include HDDs, PSUs, and fans.

All hot swap components are top-loaded and serviceable from the cold aisle.

SAS expanders can be serviced while the DX-88 drawer is extende. HDDs must be removed from the chassis prior to replacing a SAS expander PCBA. HDDs are removed, rotated 180 degrees in a vertical axis, and repositioned in their slots during SAS expander PCBA replacement. The rotation provides a lock-out that prevents them from being fully inserted into their slots, allowing the SAS expander to be remvoed.

All PCBAs are serviceable while the chassis is installed in the rack.

The DX-88 has a front-accessed pull-out tag that contains label information.

The DX-88 and all of it's supporting hardware can serviced with a T15 Torx driver.

2 DX-88 Chassis Requirements

2.1 HDD Support

The DX-88 supports up to 88 near-line 3.5" HDDs (12W max each). These HDDs are logically split into four equal 22 HDD segments, with each segment connected to a 36 port 12Gb SAS Expander, and each expander provides a 4-lane Mini-SAS cable on the DX-88's front (cold aisle) side.

The DX-88 provides individual power control to each of the 88 HDD slots. This function is controlled by the SAS expander (per DX-88 segment) and the BMC (by sending messages to the expander). During HDD start up, HDDs spin up in batches to avoid system overload.

2.2 Segment Support

The DX-88 segment of 22 HDDs and one SAS Expander are named Segments 1 through 4. Each segment is independently controlled. The design is enabled for SAS-4 expander support, solely by modifying SAS cables and PCBAs. Each HDD segment is mounted onto a tray for service functionality.

Table 2: Segment Features

Function	Details
HDDs	22 HDDs in each logical expander
SAS Expander	Each segment has one 36 port Expander to attach to 22 HDDs
Power Control	Each segment has its own in-rush controller and ability to report power. Segment power control and power metering is implemented via DX-88 BMC.
Fan Control	DX-88 BMC gathers data from each segment (in addition to thermal sensors) and control fan speed

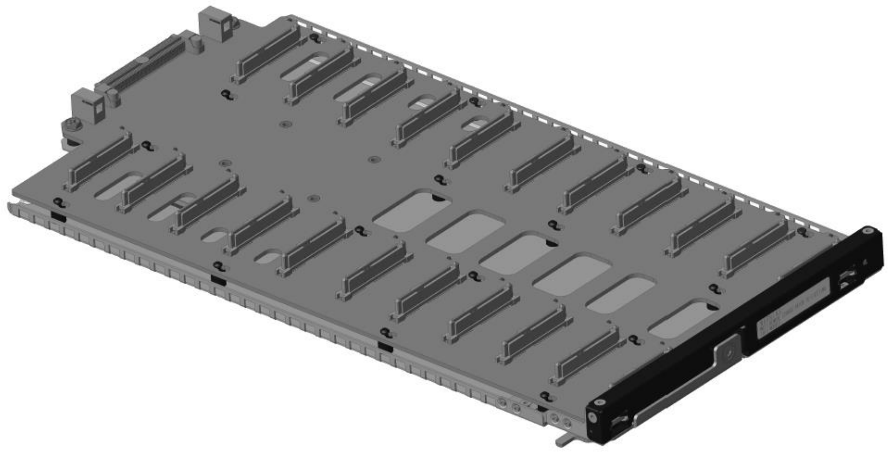

Figure 12: HDD Board Tray

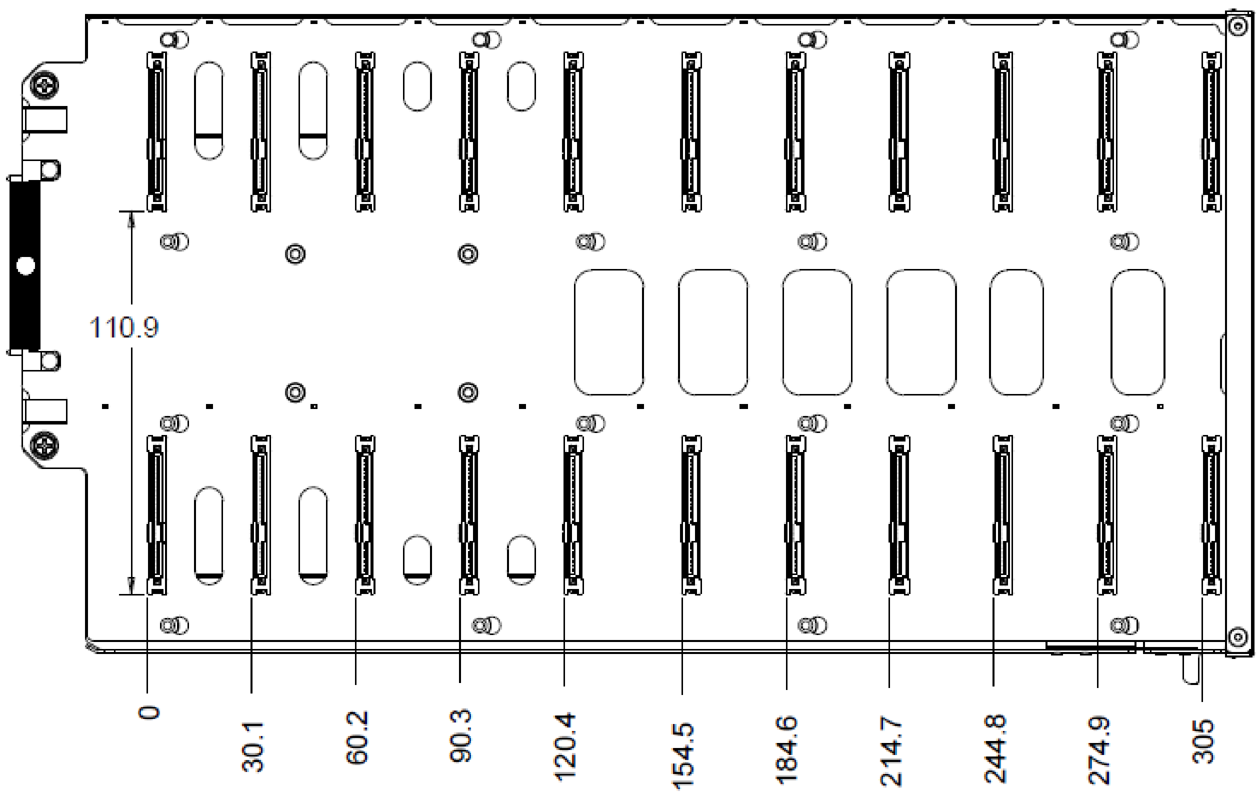

Figure 13: DX-88 HDD Board Dimensions

2.3 CFM and Thermal

The DX-88 is capable of a maximum inlet temperature of 35 degrees Celsius, assuming a maximum power configuration of 88 12W HDDs, and a worst case, single fan rotor failure.

The DX-88 maintains operation for a 5-minute minimum service time (with all drives pulled forward), and maintains operation for at least 2 minutes while a hot serviceable component is removed. The fan algorithm is configured to bring all component temperatures back to their original non-failed stable state within three hours of service.

The storage enclosure uses a bilateral cooling method (Patent Pending) to maintain cooler downstream HDD temperatures. This approach is depicted in Figure 14.

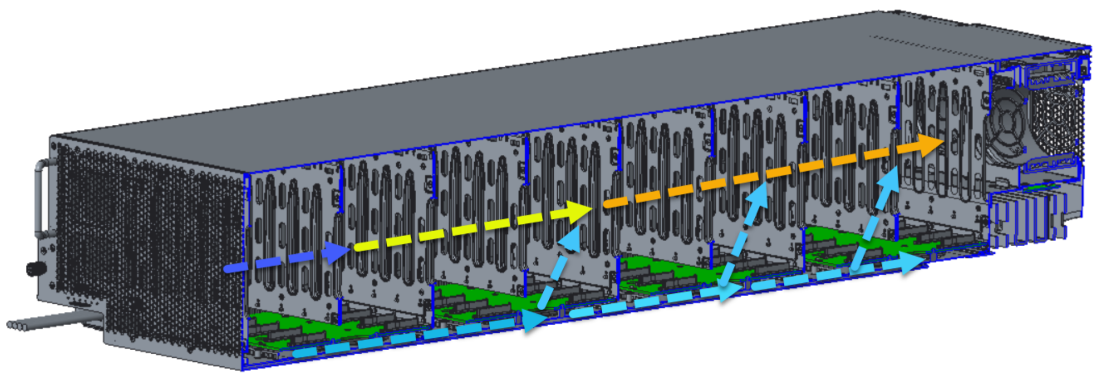

Figure 14: DX-88 Bilateral Cooling

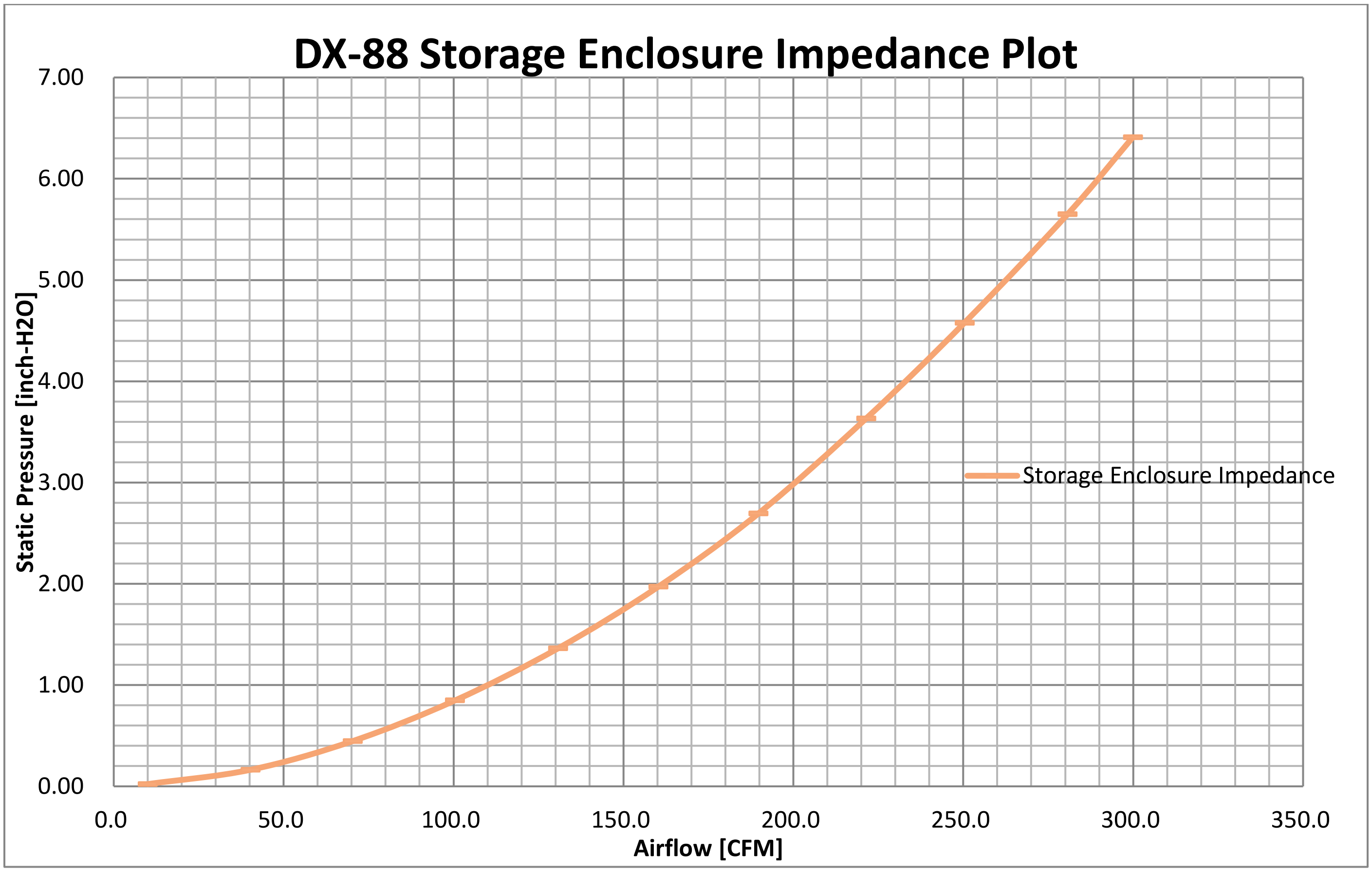

Figure 15: DX-88 Storage Enclosure Impedance Plot

2.4 Acoustic and Rotational Vibration

DX-88 fans and HDD operations do not impact performance of adjacent HDDs by no more than 5% across all operational fan speeds, including fan failure scenarios.

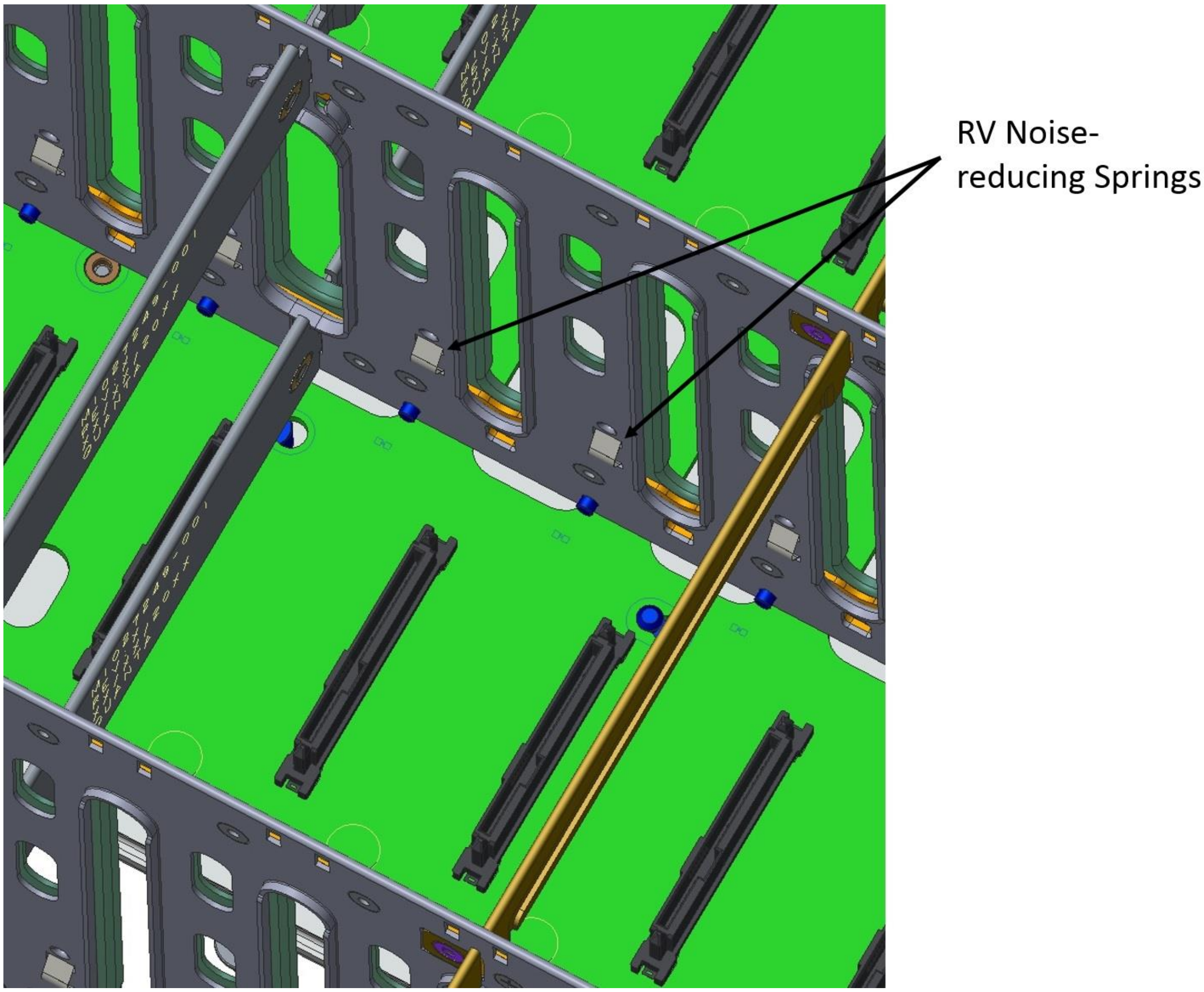

Figure 16: RV Noise-reducing Springs

Individual HDD positions are created in the DX-88 drawer by using sheet metal partitions and flanges around the cooling slots. Transmission of rotational vibration from the drives is reduced by the use of integral leaf springs (Patent Pending), two for each HDD.

To improve isolation between the HDDs and the noise from the fans, there is a plenum at the rear of the drawer separating the drives from the cooling fans. The plenum employs configurable acoustic attentuators (Patent Pending) as shown in Figure 17.

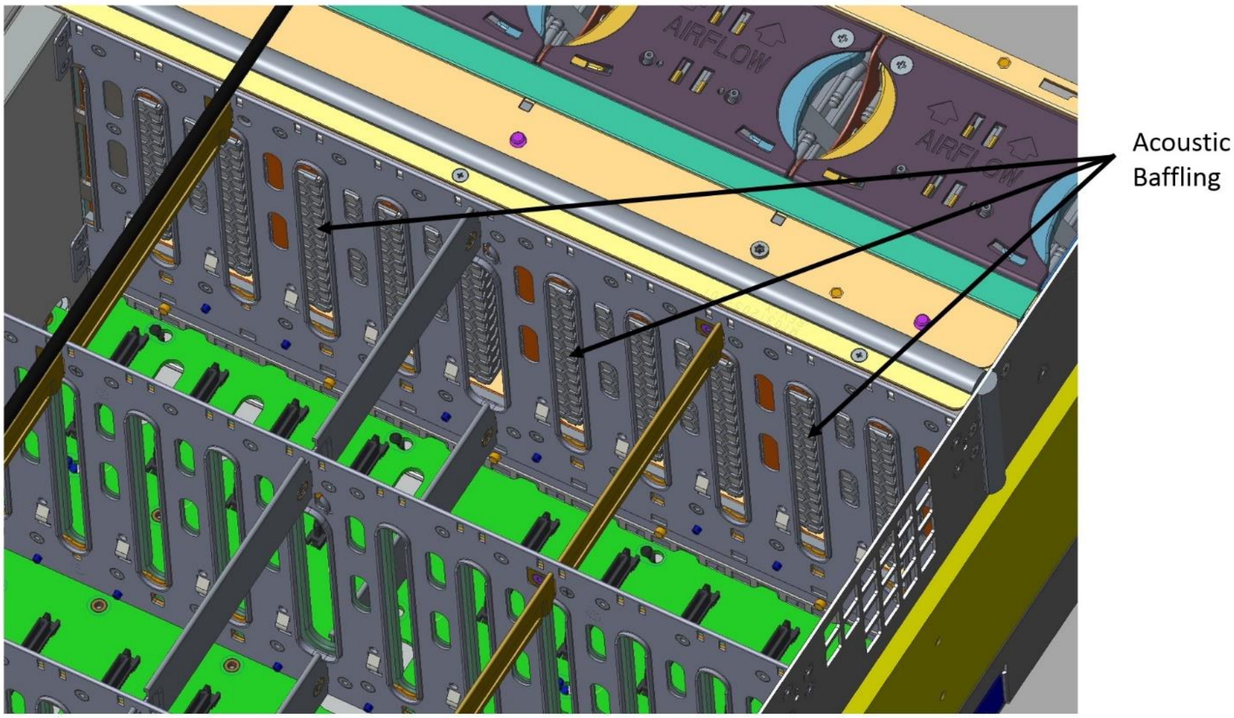

Figure 17: Acoustic Attenuation Baffles

2.5 Environmental

The storage enclosure will be deployed in environmentally controlled locations. The inlet to the unit is typically exposed to the environment described in Table 3. The server provides full functional operation under the these conditions.

Table 3. Environmental Specifications

Specification	Requirement	
Inlet temperature	Operating	• 50°F to 95°F (10°C to 35°C) • Maximum rate of change: 18°F (10°C)/hour • Allowable derating guideline of 1.6°F/1000ft (0.9°C/304m) above 3000 ft.
	Non-operating	• -40°F to 140°F (-40°C to 60°C) • Rate of change less than 36°F (20°C)/hour
Acoustic	Less than 9.2 bells at maximum fan speed operating condition	
Non-Operational Shock and Vibration	The server is capable of rack level transportation via common carrier. Rack level testing complies with ASTM 4169, ISTA 2A, ISTA 2B. Recommended levels for a single DX-88 directly mounted to table: • Shock – Square Wave, 15G, 4.19m/s • Vibration – 1.54 Grms, 10-500Hz, 15 mins per axis	

Variable fan speed capability is implemented. This enables the rack to minimize energy consumption of the air movers and facilities in conditions that permit it. The speed of airflow is based on component temperature requirements within the enclosure.

2.6 Power Supply

The DX-88 power supply features are:

- N + N 1650W DC hot swappable power
- Two three-phase power supplies with the following normal (non -fault) power supply conditions.
  - Phase to phase balance +/- 3%, feed to feed balance +/- 10%
- Top Hot plug, backwards insertion prevention
- The stand alone DX-88 does not consume more than 1500W in any operational or start-up/shut down state.

2.6.1 Power Supply Form Factor

A drawing is shown in the figure below.

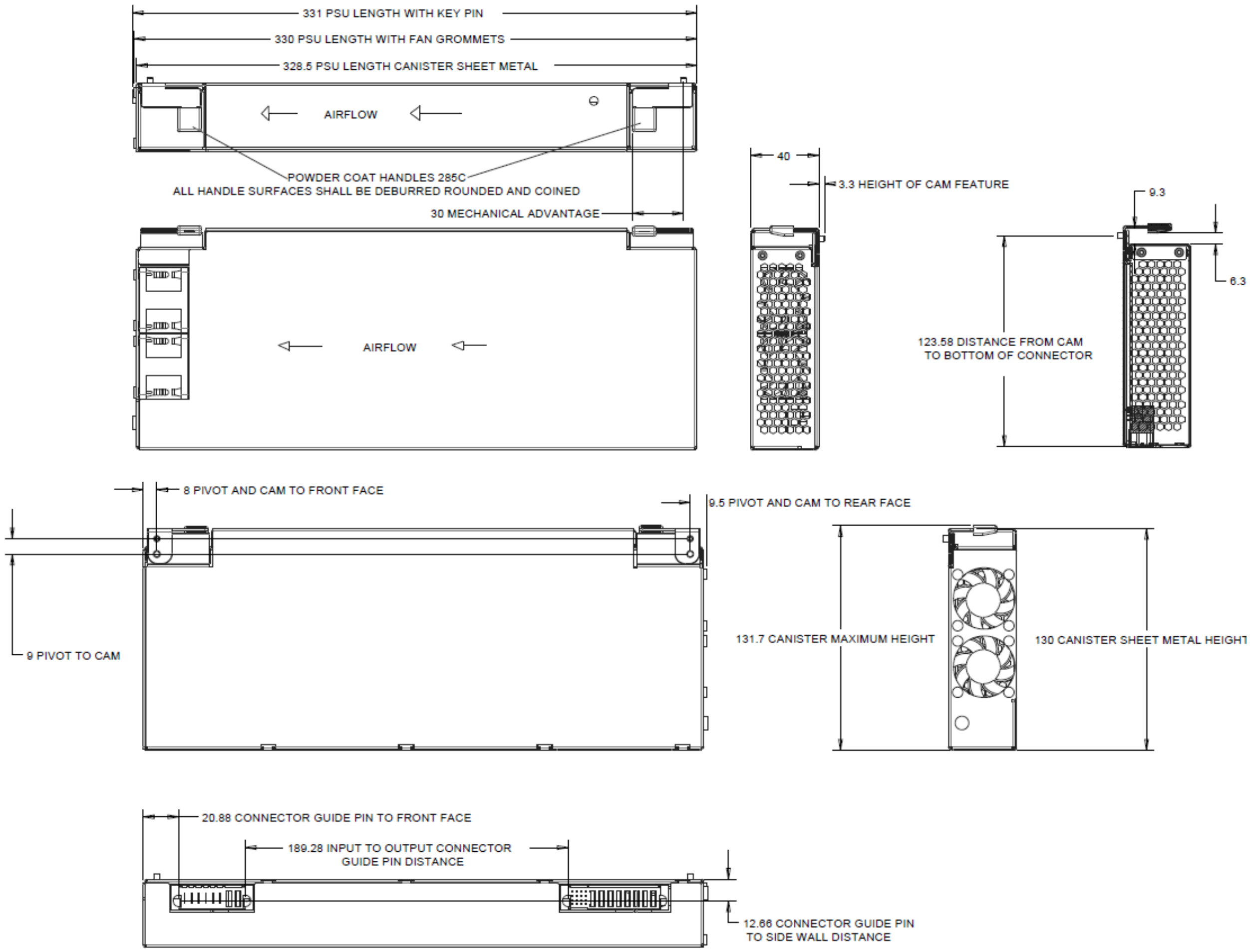

Figure 18: P2020 Form Factor Dimensions

2.7 Fans

The DX-88 supports N+1 hot swappable, 92mm fans. The fans can be serviced from the cold aisle. All fans operate at the same speed.

Figure 19: DX-88 Fan Module

2.7.1 Fan Operation

Lower HDD temperatures are required for reliability, but driving the fans too fast results in too much airflow to the rack and excessive acoustic transmitted vibration inside the chassis. The airflow allowance is 158 CFM per KW at the rack level. The DX-88 BMC (Aspeed 2520) is responsible for setting fan speeds. The fans operate to keep all DX-88 components (PSU, Expanders, HDDs) within thermal limits. The fan control algorithm uses PID Control methods. Below are the fundamentals that drive the PID control.

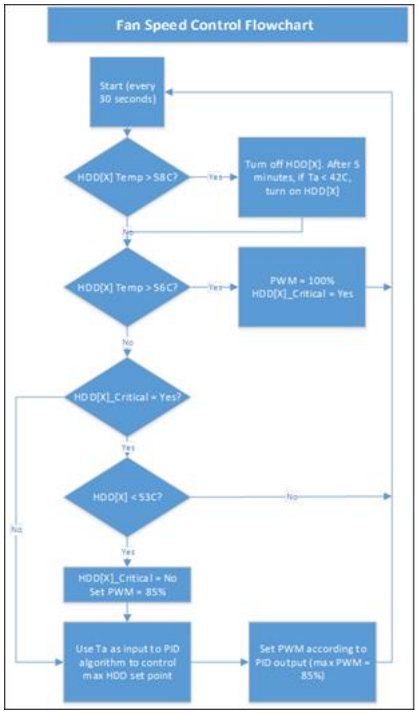

Figure 20: DX-88 Fan Speed Control Flowchart

Set Point Table for PID Algorithm	
Ta (Celcius)	**Max HDD Set Point (Celcius)**
< 31	46
31	47
32	48
33	49
34	50
35	51
36	52
37	53
> 37	54

Table 4: Set Point Table for PID Algorithm

- For Ta, use PDB_Temp sensor
- Fan speeds adjusted every 30s since HDD temperatures are read every 30s
- Max PWM=85%
- IF Max Drive Temp >56°C, Set Fans to 100% PWM
- IF Critical HDD Temp Triggered, wait until Drive Temp <53°C, Reset Max PWM to 85% and resume algorithm

- IF HDD “X” Temp >58°C, turn off slot “X”
- To Turn HDD “X” back on: Set 5min timer, IF Ta<42°C, Power on HDD “X” OK

2.8 HDDs

The DX-88 houses 88 3.5” HDDs in eight rows with 11 drives per row. The HDDs are top-loaded and hot-swappable, and the HDD carrier allows for tool-less operation and cam-action insertion and removal.

The carrier is installed onto the HDD with a set of plastic and metal pins that interface with the 6-32 HDD mounting holes.

The carrier supports one fault indication LED. The LED is amber in color.

Figure 21: HDD Carrier

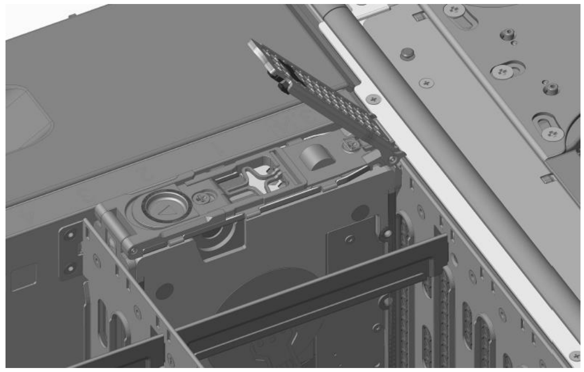

Figure 22: Servicing HDD

2.9 LEDs

The following sections describe the light-emitting diodes (LEDs) used as indicators in the DX-88. Table 5 lists the LEDs and provides a brief description. Greater detail for some LEDs is included in subsequent sections below. The indicator LED is visible at the front of the DX-88 (cold aisle). HDD, PSU, and Fan LEDs are visible when the corresponding devices are in a position to be serviced.

Table 5. LEDs

LED Name	Color	Description	Location
UID LED	Blue	Unit Identification LED	Front Panel
DX-88 Health LED	Red	Indicates something in the DX-88 needs repair	Front Panel
Power Status LED	Green/Amber	Indicates Power State of Enclosure	Front Panel
HDD Repair LED	Amber	Indicates HDD flagged for repair (one per HDD)	HDD Board; Visible on HDD carrier
PSU Health LED	Green/Amber	Indicates PSU flagged for repair (one per PSU)	PSU
Fan Repair LED	Amber	Indicates Fan flagged for repair (one per FAN)	Fan Board
Expander Health LED	Green	Indicates Expander is alive (one per HDD Board)	HDD Board, Visible on side of chassis
BMC Health LED	Green	Indicates BMC is alive	BMC Board
GbE Activity LED	Green	Indicates Management Port activity	Front Panel
GbE Speed LED	Green/Amber	Indicates GbE speed	Front Panel

2.9.1 UID LED

Blue= ID ; OFF= Default

2.9.2 DX-88 Health LED

This is a front visible amber LED that is indicates something inside the DX-88 needs service.

Table 6. DX-88 Health LED Description

LED status	Condition
Off	No attention indicated
Solid Red	Indicates some service required in the DX-88 (HDD, PSU, Fan, or HDD Back Plane)

2.9.3 Power Status LED

Table 7. Power Status LED Description

LED status	Condition
Solid Green	System Power On
Solid Amber	BMC Down

2.9.4 HDD Repair LED

Each HDD has a status LED to support quick and accurate repair. Viewing the LED from the top of an inserted drive is made possible via a light pipe in the drive carrier. Software or firmware will light the LED when service is necessary.

Table 8 describes the operation of the HDD repair LED.

Table 8. HDD Repair LED Description

LED status	Condition
Off	No attention indicated
Solid Amber	Indicates HDD failure

2.9.5 PSU Health LED

The DX-88 supports one status LED per PSU. This LED is viewed on the PSU.

2.9.6 Fan Repair LED

Indicates a fan failure

Table 9. Fan Repair LED Description

LED status	Condition
Off	No attention indicated
Solid Amber	Fan Fault

2.9.7 Expander Health LED

Table 10. Expander Health LED Description

LED status	Condition
Off	Expander is down

Blinking Green	Expander is alive

2.9.8 BMC Health LED

Table 11. BMC Health LED Description

LED status	Condition
Off	BMC is down
Blinking Green	BMC is alive

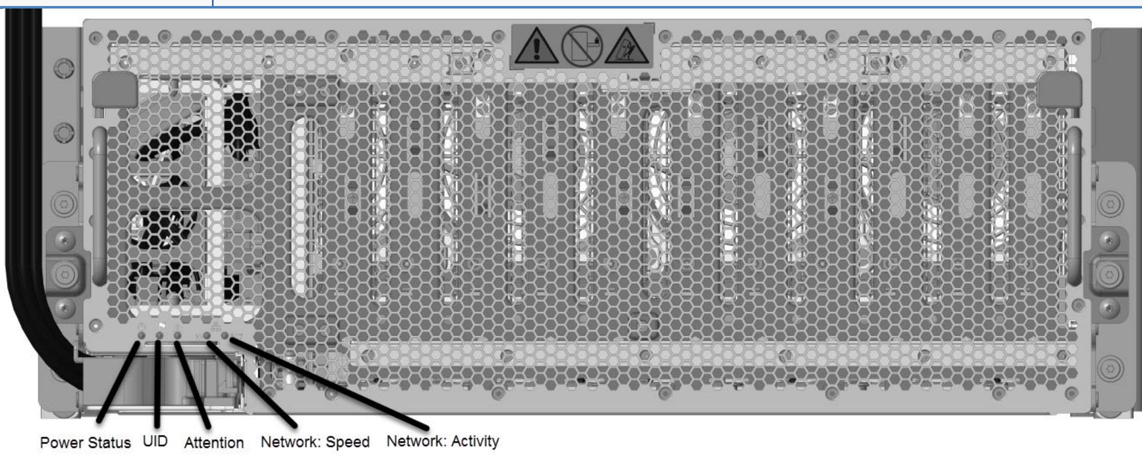

Figure 23: DX-88 Front View LED Location

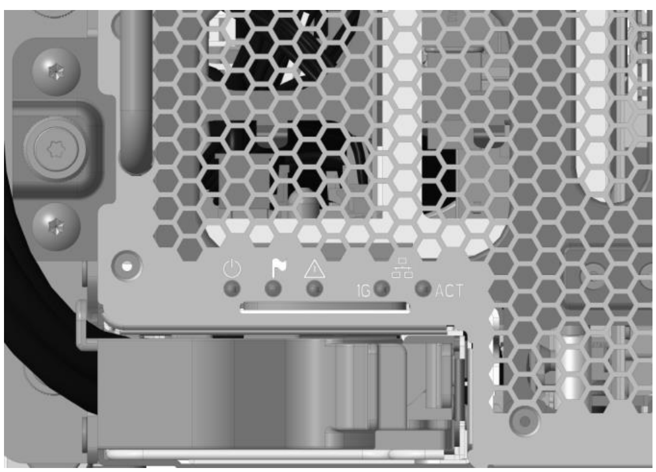

Figure 24: DX-88 Front Panel Zoomed View

Figure 25: Power Status LED Symbol

Figure 26: Attention/DX-88 Health LED Symbol

Figure 27: UID LED Symbol

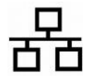

Figure 28: Network Symbol

2.10 Labels

Space to place instructional/warning labels in key service areas is provided.

2.10.1 System Component Map Label

This label includes:

- Component map
- Allowable Hot-Swap Service Time Warning
- LED Color and Operation Definition
- Other Servicability Information

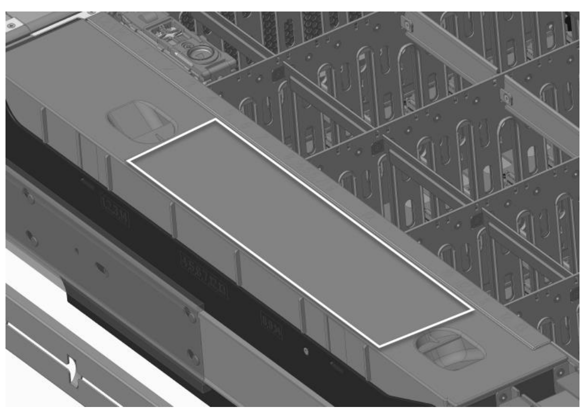

Figure 29: PSU Exhaust Cover

2.10.2 HDD Board Tray Service Warning Label

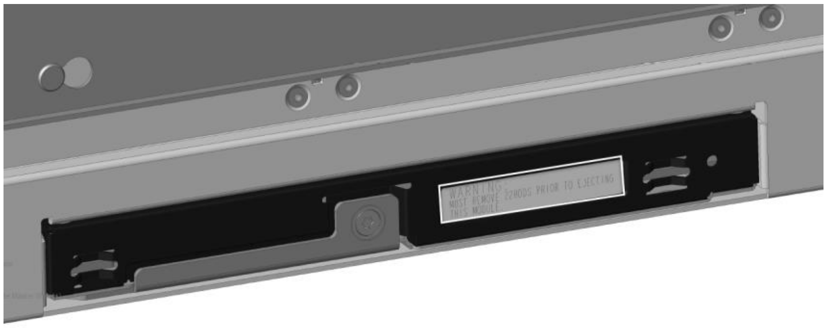

Figure 30: HDD Board Tray

This label includes a warning to remove the segment's 22HDDs prior to service

2.10.3 Tipping Hazard Label

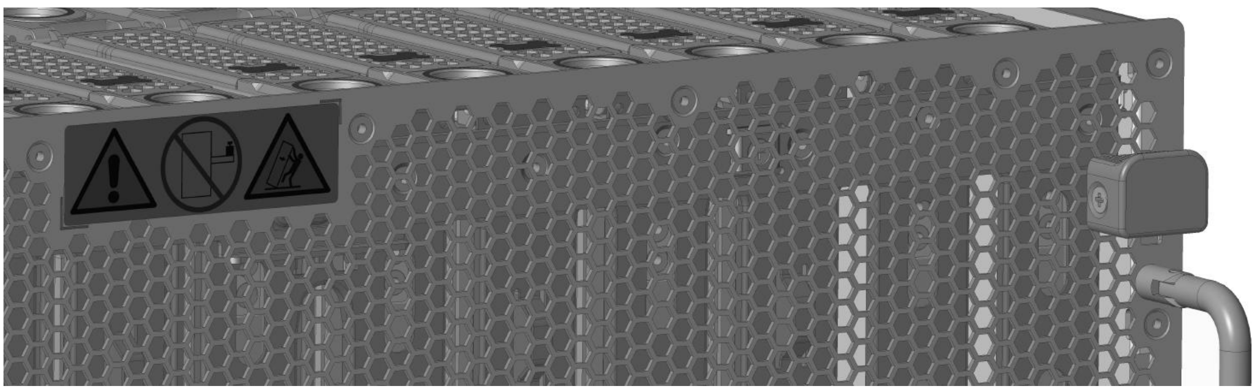

Figure 31: DX-88 Front Panel

2.10.4 Fan Blade Hazard Label

Figure 32: DX-88 Rear Panel

3 DX-88 System Management

3.1 High level DX-88 Management Architecture

There is one Aspeed 2520 BMC in the DX-88 that is connected to all thermal sensors, both power supplies, all hot swap controllers, and all SAS expanders. The logical view can be seen in the following figure:

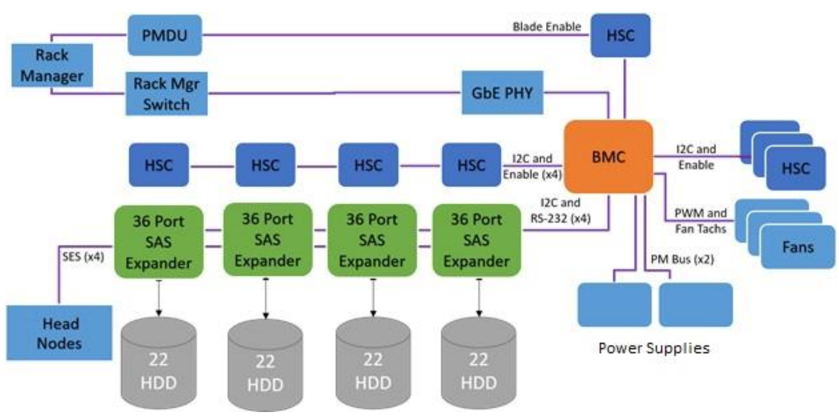

Figure 33: High level Management Architecture

3.2 Rack Manager Data Signals via PMDU

The DX-88 shall connect to the Project Olympus Rack Manager in the PMDU via contacts in the PMDU/device interface connector. The data signals from this connection are routed to the BMC.