

WHITE PAPER

Understanding UCS X-Series & M6 servers

Contents

Introduction 3

Background on previous UCS chassis & server models..... 3

Hardware Refresh of the UCS Platform 4

 Capabilities of X-Series - Compute Nodes4

 Capabilities of X-Series - Connectivity & Bandwidth4

"Architect Once" with FlashStack..... 5

Conclusion 6

About the Author..... 7

Introduction

This document is meant to provide background to help understand the Cisco UCS compute platform, beginning with the history of the original 5108 chassis and B-series blades, then following the product timeline to bring us to the new X9508 modular chassis and X-Series compute nodes.

There are many benefits to businesses with choosing a datacenter platform with such a demonstrated hardware lifespan that can allow for multi-year lifecycles without redesign and new deployments.

Puritans should understand the features available in the new M6 models and the differences in capabilities between the X & B series models.

Also, much greater connectivity capabilities can be implemented in newer fabric interconnects concerning both speeds and protocols.

This guide targets internal personnel without direct hardware experience or extensive knowledge of the Cisco UCS platform.

Background on previous UCS chassis & server models

In March 2009, the Unified Computing System (UCS) platform was first announced with the 5108 chassis focused on addressing the needs for providing a data center platform built upon a unified fabric meant to converge the network, storage, and management into a single fabric.

The core of the platform and management logic was based on the concept of stateless hardware – this means that the hardware is abstracted from the identity of an individual compute node, which is managed as a server profile.

The compute node "personality" identifiers are applied to the hardware when associated with a blade or rack server. It is then that policies of the server and the hardware IDs such as mac addresses and WWPN/WWNN get set, but they are not burned into the hardware as with typical server components.

This means these profiles can be moved to different hardware or disassociated with a server. Any other server profile, or "personality," can be associated with the same hardware. There was now a logical separation of the identity of a server from its hardware.



Cisco changed the way we approached computing and changed the industry with the release of UCS. The platform solved some challenging problems where new approaches brought new capabilities to managing hardware with abstractions, which still simplified the server and datacenter management paradigm.

The 5108 chassis carried the UCS architecture for 12 years, covering 6 generations of server hardware. The hardware that controls the platform and handles the management (the fabric interconnects) has gone through 4 generations. Still, each allowed for upgrades of those dedicated switches without downtime to the platform. Even throughout the lifecycle of each iteration of the management hardware, and the server blades themselves, customers could get use the same chassis throughout these past 12 years. Even the most cost-conscious customers must admit that the ROI for both the hardware and platform installation time has been attained, and the longevity has been proven.

Hardware Refresh of the UCS Platform

More recently, in June 2021, the first update to the chassis for the UCS platform was released. Again, many years of improvements and engineering were built into the newest hardware design. To understand the changes required, only look at some specifications with computing have grown during this time. For example, when the 5108 chassis and the first server generation of M1 servers were released, CPUs had a max 50W power requirement; with the sixth generation of M6 servers being released, CPUs now have a 270W power requirement.

It becomes easy to underestimate the industry changes over 12 years when we consider how fast many of those components have changed over time. Within a platform that addresses compute, network, and storage, it is a very different landscape than in 2009. The changes in the new X-Series chassis are not engineering for engineering's sake, nor a solution just looking for a problem. There were significant changes needed to address a different class of computing/network/storage needs for customers, primarily if this new platform is meant to last for the next decade.

Each generation of servers has come with larger and larger CPU capacity for CPU cores and memory that could fit within each blade or rackmount server. The need for additional resources such as GPU has become more prominent for many customers, and these types of needs led to redesigning the chassis. The new hardware is now even more modular for advanced capabilities, removes the backplane, provides more power, and improves the connectivity options to provide greater bandwidth to the environment.

Capabilities of X-Series - Compute Nodes

By changing to a vertical orientation for compute nodes in the X-Series chassis, rather than having half-width horizontally oriented slots of B-Series, the chassis achieved 33% more room for the nodes, with only 1 additional RU of space. This new X-Series platform can now support 600 cores, and ~100TB of memory, with significant advancements in both power distribution and cooling. There is now 54V power distribution, up from 12V in 5108 chassis, with much greater delivery efficiency. In addition, improved airflow provides higher cooling at lower RPM while having greater reliability of fan modules, along with discrete thermal zones. The cooling changes now give lots of headroom for higher power CPU/GPUs down the road, and the X-Series chassis will also allow for future liquid cooling options.

For a quick comparison, the B-Series chassis with M6 blades maxes out at 480 cores and 64TB of memory. Again, the X-Series platform can now support 600 cores, and ~100TB of memory.

Capabilities of X-Series - Connectivity & Bandwidth

There were drastic changes and improvements to the fabric, which provides connectivity inside and outside the chassis. To provide outbound connectivity, the new Intelligent Fabric Module (IFM) acts as the network fabric, and it delivers 8x 100GbE of upstream connectivity (to uplinks) with 32x 25GbE downstream connectivity (to compute nodes). It delivers 200GbE per compute node with 25GbE links at release, with roadmap plans to support 2x100GbE links in the future. Finally, the X-fabric modules are the chassis' internal fabric, and these modules will deliver future capabilities, such as enabling shared/pooled PCIe Gen4 devices across the chassis.



For a quick comparison, the B-Series chassis combined with 2408 fabric extenders and 6400 fabric interconnects can provide 32x 10G interfaces to blades and 400 Gbps of I/O uplink from the chassis. Again, the X-Series platform can now 200 Gbps per compute node (blade) and 800Gbps of I/O uplink from the chassis.

"Architect Once" with FlashStack

Now that we have explained the history & lifecycle of the Cisco UCS platform, it's time to explain the integration with FlashStack and the future of our joint solution. For the final point of significant changes to the management of the platform, we only need to look at Cisco's quote on this:

"The Cisco UCS X-Series is a new modular compute system configured and managed from the cloud. It is designed to meet the needs of modern applications and improve operational efficiency, agility, and scale through an adaptable, future-ready, modular design."

Cisco UCS has proven its longevity and has demonstrated the capability to perform rolling upgrades for hardware, connectivity, and management components, each independently for a decade of innovation. As a result, Cisco UCS is the logical compute platform to be the core of FlashStack. It truly helps form the basis for our statement that FlashStack allows customers to "Architect Once, Innovate Continuously."

Cisco UCS X-Series is now managed by the Cisco Intersight SaaS product, which delivers cloud-connected management and provides a single interface to manage the global fleet of hardware a customer owns. It is a platform that leverages a modular or "plug-in" style of adding functionality. This model for the Intersight platform and engine means additional products and features can be added to expand management capabilities.

With the reality of it being a SaaS application, these changes can be visible to customers by only logging out and back into the web portal. This quick access to new features is a dramatic shift from the customer-driver, time-intensive updates to the old management platform. The introduction of Intersight is also the first time that Cisco has made the UCS ecosystem truly extensible for partners to be capable of integrating with the management plane for Cisco UCS.

Pure Storage became the first partner to develop integration with Cisco Intersight, which allows for the ability to view the inventory of a FlashArray, or the ability to create and run workflows for managing a FlashArray from within the Intersight portal.



Conclusion

With the top performance of Pure Storage and the proven record of Cisco UCS, FlashStack has become a fantastic platform for customers looking to deliver simplified performance at scale. The work to deliver Validated Designs for many years built upon this platform and the initiative to become the first partner to integrate with the next iteration of management are a great lead-in to help showcase where we will go for the next decade of the FlashStack story.



About the Author

Joe Houghes is a Senior Solutions Architect in the Portfolio Solutions team within Pure Storage, focused on solutions on the FlashStack platform and automation and integration. He has experience from over 20 years in Information Technology across various customer/vendor organizations with architecture and operations. His expertise covers computing, networking, storage, virtualization, business continuity and disaster recovery, cloud computing technologies, and automation and integration across many applications and vendor platforms.

The Pure Storage products and programs described in this documentation are distributed under a license agreement restricting the use, copying, distribution, and decompilation/reverse engineering of the products. No part of this documentation may be reproduced in any form by any means without prior written authorization from Pure Storage, Inc. and its licensors, if any. Pure Storage may make improvements and/or changes in the Pure Storage products and/or the programs described in this documentation at any time without notice.

THIS DOCUMENTATION IS PROVIDED "AS IS" AND ALL EXPRESS OR IMPLIED CONDITIONS, REPRESENTATIONS AND WARRANTIES, INCLUDING ANY IMPLIED WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT, ARE DISCLAIMED, EXCEPT TO THE EXTENT THAT SUCH DISCLAIMERS ARE HELD TO BE LEGALLY INVALID. PURE STORAGE SHALL NOT BE LIABLE FOR INCIDENTAL OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH THE FURNISHING, PERFORMANCE, OR USE OF THIS DOCUMENTATION. THE INFORMATION CONTAINED IN THIS DOCUMENTATION IS SUBJECT TO CHANGE WITHOUT NOTICE.

Pure Storage, Inc.
650 Castro Street, #400
Mountain View, CA 94041

purestorage.com

800.379.PURE

