ıı|ıı|ıı
CISCO
The bridge to possible

Cisco UCS X210c M7 Processor Selection Guide for VDI Using Citrix Virtual Apps and Desktops

Contents

What you will learn

Choosing the CPU for enterprise applications is not an easy task, whether you use a simple back-of-the-envelope calculation to determine your needs or perform exhaustive capacity planning with analytical model simulations for workloads and their future growth. Often, lack of details (or inaccurate information) about the input data and the workload type leads to uncertain estimations about the rightsizing of the hardware (CPU type, core count and clock speed, memory capacity, etc.), leading to suboptimal resource usage.

4th-generation Intel Xeon Scalable Processors

4th Gen Intel® Xeon® Scalable processors are designed to accelerate performance across the fastest-growing workloads in artificial intelligence (AI), data analytics, networking, storage, and high-performance computing (HPC). These processors have the most built-in accelerators of any CPU on the market. They help bring a zero-trust security strategy to life while unlocking new opportunities for business collaboration and insights–even with sensitive or regulated data–with advanced security technologies. You can scale across multiple clouds and edges to meet your deployment needs. And Intel Xeon Scalable processors offer the most choice and flexibility in cloud selection, with smooth application portability.

4th Gen Intel Xeon Scalable processors feature a new architecture with higher per-core performance than the previous generation. They also feature up to 60 cores per socket and one, two, four, or eight sockets per system. To balance those core-count increases, the platform provides accompanying advances in the memory and input/output (I/O) subsystems. DDR5 memory provides up to 1.5x the bandwidth and speed of DDR4, for 4,800 megatransfers per second (MT/s).8 The platform also features 80 lanes of PCIe Gen5 per socket, for dramatically improved I/O compared to earlier platforms.9 It provides CXL 1.1 to support high fabric bandwidth and attached accelerator efficiency. 4th Gen Intel Xeon Scalable processors support technologies that let you scale and adapt as workload requirements change.

To choose the right CPU for virtual desktop infrastructure (VDI), you must consider several factors, including feature sets and hardware requirements. Proper processor selection is crucial. For VDI solutions, given adequate memory, storage performance, and network bandwidth, the CPU is the element that determines user density and performance. Different user types benefit from different processor and memory configurations.

This document provides guidance for starting a configuration using 4th-generation Intel Xeon Scalable Processors that are in the processors found in the Cisco Unified Computing System™ (Cisco UCS®) portfolio for the most common VDI user persona type, known as Knowledge Worker.

Cisco Unified Computing System X-Series M7 Compute Nodes

The Cisco UCS X210c M7 Compute Node is the second generation of compute nodes to integrate into the Cisco UCS X-Series Modular System. It delivers performance, flexibility, and optimization for deployments in data centers, in the cloud, and at remote sites. This enterprise-class server delivers market-leading performance, versatility, and density without compromise for workloads. Up to eight compute nodes can reside in the 7-rack-unit (7RU) Cisco UCS X9508 Server Chassis, offering one of the highest densities of compute, I/O, and storage per rack unit in the industry.

The Cisco UCS X210c M7 provides these main features:

- CPU: Up to 2x 4th-generation Intel Xeon Scalable Processors have up to 60 cores per processor, up to 2.625-MB Level 3 cache per core, and up to 112.5 MB per CPU.

- Memory: The solution offers up to 8TB of main memory with 32x 256-GB Double-Data-Rate 5 (5DDR5) and 4800 dual in-line memory modules (DIMMs).

- Storage: The solution offers up to six hot-pluggable solid-state drives (SSDs) or Nonvolatile Memory Express (NVMe) 2.5-inch drives with a choice of enterprise-class Redundant Array of Independent Disks (RAIDs) or pass-through controllers, and up to two M.2 Serial Advanced Technology Attachment (SATA) drives with optional hardware RAID.

- Optional front mezzanine GPU module: The Cisco UCS front mezzanine GPU module is a passive PCIe 4.0th-generation front mezzanine option with support for up to two U.2 NVMe drives and two HHHL GPUs.

- Modular LAN on Motherboard (mLOM) virtual interface cards:

  - Cisco UCS Virtual Interface Card (VIC) 15420 occupies the mLOM slot of the server, enabling up to 50 Gbps of unified fabric connectivity to each of the chassis intelligent fabric modules (IFMs) for 100-Gbps connectivity per server. Cisco UCS Virtual Interface Card (VIC) 15231 occupies the mLOM slot of the server, enabling up to 100 Gbps of unified fabric connectivity to each of the chassis IFMs for 100-Gbps connectivity per server.

  - Optional mezzanine card: Cisco UCS 5th-generation Virtual Interface Card (VIC) 15422 can occupy the mezzanine slot of the server at the bottom rear of the chassis. The I/O connectors of this card link to Cisco UCS X-Fabric technology. An included bridge card extends 2x 50 Gbps of network connections of this card through IFM connectors, bringing the total bandwidth to 100 Gbps per fabric (for a total of 200 Gbps per server). A Cisco UCS PCI Card for X-Fabric can occupy the mezzanine slot of the server at the bottom rear of the chassis. The I/O connectors of this card link to Cisco UCS X-Fabric modules and enable connectivity to the Cisco UCS X440p PCIe Node. All VIC mezzanine cards also provide I/O connections from the X210c M7 compute node to the X440p PCIe node.

- Security: The server supports an optional trusted platform module (TPM). Additional features include a secure boot field-programmable gate array (FPGA) and ACT2 anti-counterfeit provisions.

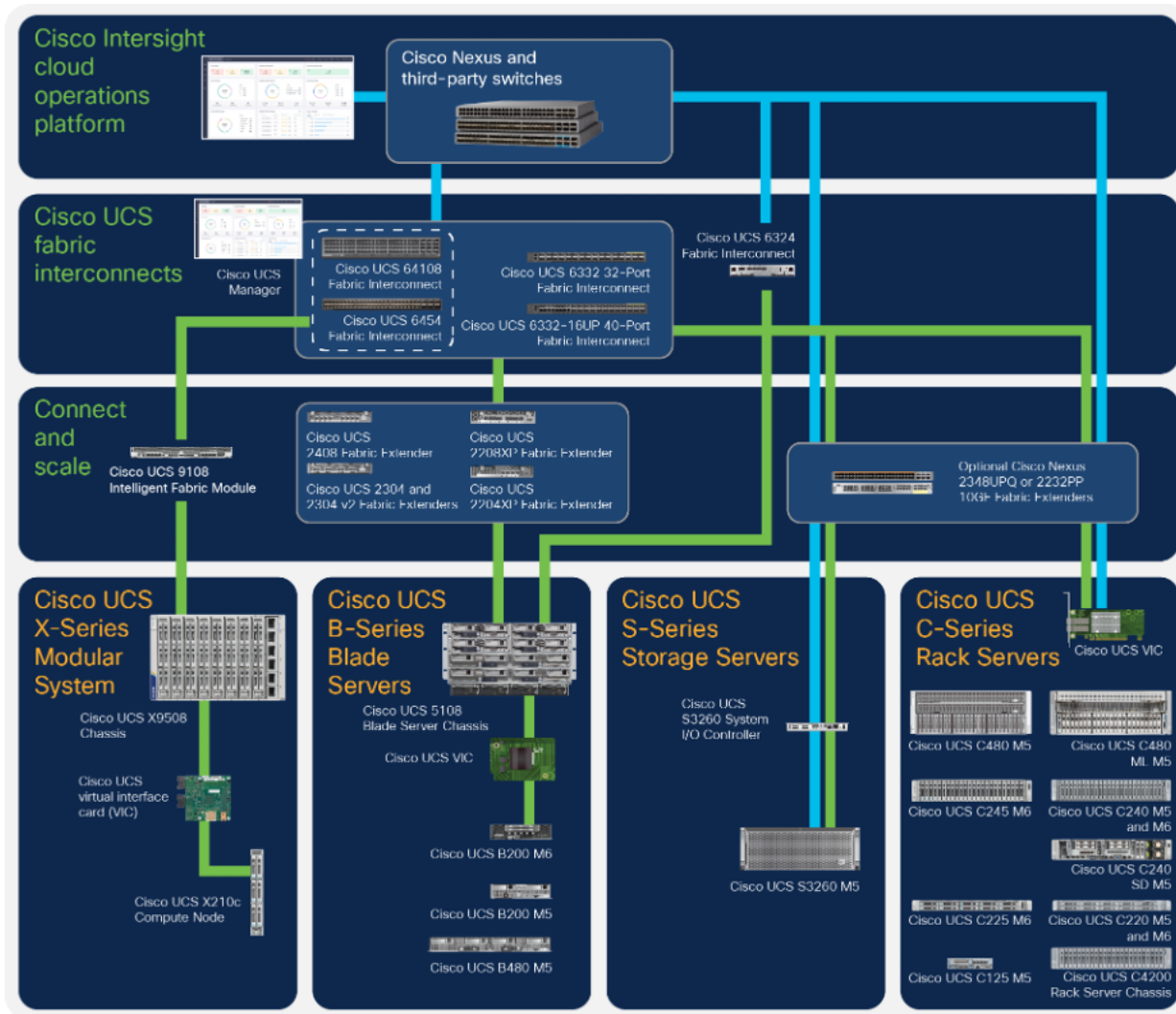Figure 1 shows the components of Cisco UCS.

Figure 1.
Cisco UCS components

The main components of Cisco UCS include:

- Computing: The system is based on an entirely new class of computing system that incorporates rack-mount and blade servers based on the Intel Xeon Scalable Processors product family.

- Network: The system is integrated onto a low-latency, lossless, 11/25/40/100-Gbps unified network fabric. This network foundation consolidates LANs, SANs, and high-performance computing (HPC) networks, which are separate networks today. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables and by decreasing power and cooling requirements.

- Virtualization: The system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.

- Storage access: The system provides consolidated access to both SAN storage and network-attached storage (NAS) over the unified fabric. It is also an excellent system for software-defined storage (SDS). Combining the benefits of a single framework for managing both the computing and storage

servers in a single pane, Quality of Service (QoS) can be implemented if needed to inject I/O throttling into the system as well as workload isolation. In addition, server administrators can pre-assign storage-access policies to storage resources for simplified storage connectivity and management, leading to increased productivity. In addition to external storage, both rack and blade servers have internal storage, which you can access through built-in hardware RAID controllers. With storage profile and disk configuration policy configured in Cisco UCS Manager, storage needs for the host OS and application data are fulfilled by user-defined RAID groups for high availability and better performance.

- Management: The system innovatively integrates all system components to enable management of the entire solution as a single entity by Cisco UCS Manager. Cisco UCS Manager has an intuitive graphical user interface (GUI), a command-line interface (CLI), and a powerful scripting library module for Microsoft PowerShell built on a robust application programming interface (API) to manage all system configuration and operations.

Sustainability in VDI solutions

Some organizations may be focusing efforts on sustainability by looking at modern solutions that are more power efficient and less wasteful during hardware refresh. Cisco is leading the way in this endeavor by engineering our products to be power efficient. We are also working with partners to ensure that our solutions incorporate steps that will guide our customers into maximizing their efficiency while still providing the performance customers expect from Cisco UCS.

With X-Series and VDI solutions, we can run more users on fewer nodes compared to past server models. For example, if we could fit 1,000 virtual desktops on eight Cisco UCS M4/M5 nodes, we can now get 1,000 virtual desktops on four Cisco X-Series nodes.

Cisco UCS sustainability steps follow:

- Cisco Intersight monitors and displays power consumption of X-Series nodes.

- Cisco basic input/output (BIOS) policies now allow balanced or energy-efficient policies to run with VDI workloads, and we have validated that the end-user experience is not negatively affected.

VMware sustainability steps follow:

- Using the host power policy as Balanced, VMware VSphere 8 allows efficient use of host P-states and C-states to provide great performance along with power management.

- Using VMware vRealize Operations Manager, we can monitor and show our sustainability efforts in tangible charts and graphs monitoring metrics such as C02 emissions saved, greener planet contributions, and more. (Refer to Figure 2.)

Citrix sustainability steps follow:

- Using Citrix power management, VDI administrators can create polices that shut down unused machines when not needed, possibly saving quite a bit of power consumption if used properly.
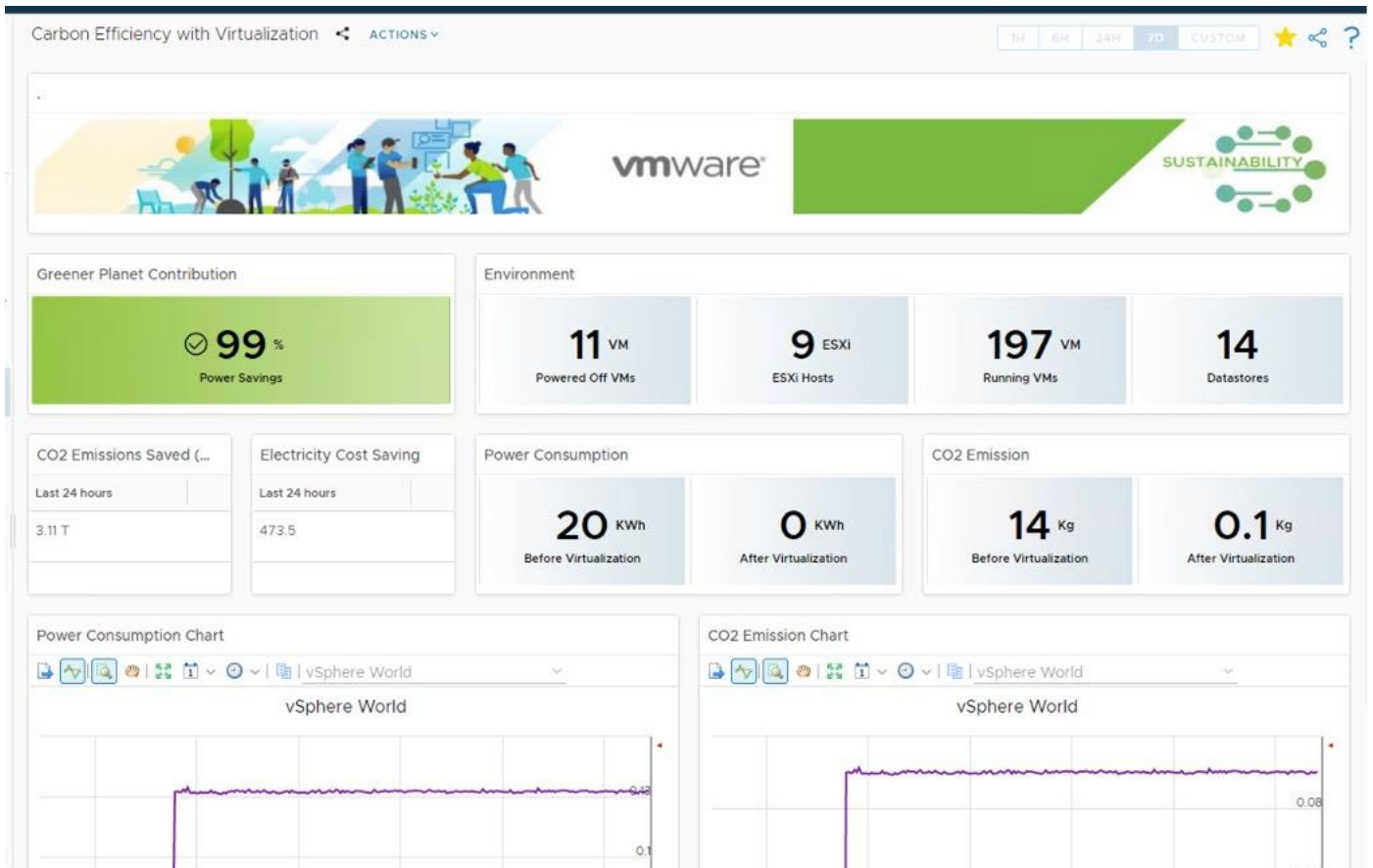
Figure 2.
VMware vRealize Operations Manager

VMware vSphere 8.0

VMware provides virtualization software. The company's enterprise software hypervisors for servers—VMware vSphere ESX, vSphere ESXi, and vSphere—are bare-metal hypervisors that run directly on server hardware without requiring an additional underlying operating system. VMware vCenter Server for vSphere provides central management and complete control and visibility into clusters, hosts, virtual machines, storage, networking, and other critical elements of your virtual infrastructure.

vSphere 8 is the latest major vSphere release from VMware. vSphere 8 has been rearchitected with native Kubernetes to enable IT administrators to use vCenter Server to operate Kubernetes clusters through namespaces. VMware vSphere with Tanzu allows IT administrators to operate with their existing skillset and deliver a self-service access to infrastructure for the Dev Ops teams while providing observability and troubleshooting for Kubernetes workloads. vSphere 8 provides an enterprise platform for both traditional and modern applications, so customers and partners can deliver a developer-ready infrastructure, scale without compromise, and simplify operations.

Additional information about VMware vSphere 8 is available here.

Citrix Virtual Apps and Desktops 2203 LTSR

Citrix Virtual Apps and Desktops are virtualization solutions that give IT control of virtual machines, applications, licensing, and security while providing anywhere access for any device.

Citrix Virtual Apps and Desktops allow:

- End users to run applications and desktops independently of the operating system device and interface

- Administrators to manage the network and control access from selected devices or from all devices

- Administrators to manage an entire network from a single data center

Citrix Virtual Apps and Desktops share a unified architecture called FlexCast Management Architecture (FMA). The important features of FMA features are the ability to run multiple versions of Citrix Virtual Apps or Citrix Virtual Desktops from a single site and integrated provisioning.

For additional details about Citrix Virtual Apps and Desktops and Citrix Provisioning Server LTSR 2203, refer to the Citrix documentation here.

Citrix Virtual Apps and Desktops test platform for 4th-generation Intel Xeon Scalable Processors

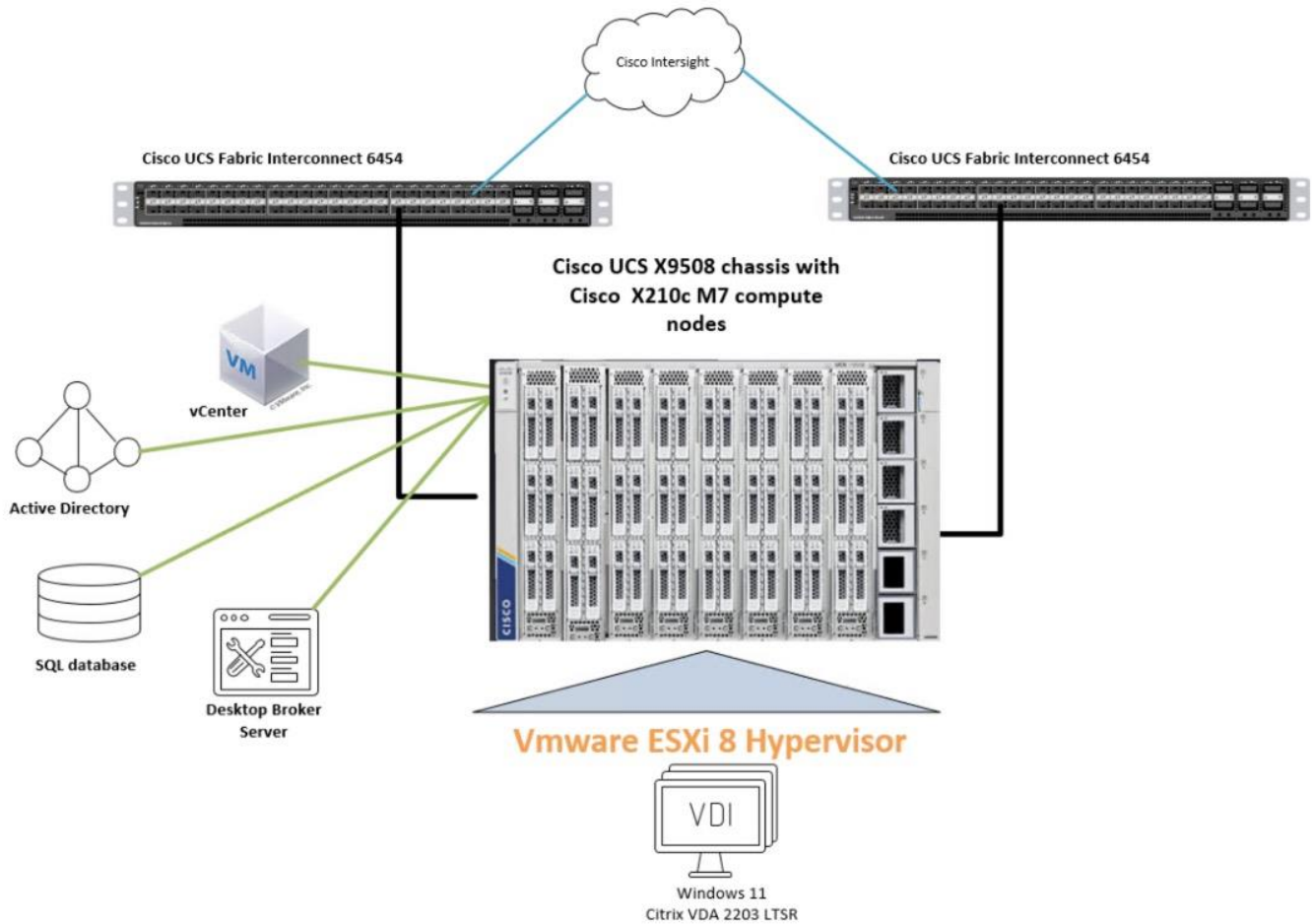Figure 3 provides an overview of the test platform used to evaluate and select processors.

Figure 3.
Citrix VDI test platform for 4th-generation Intel Xeon Scalable processors evaluation

The solution includes the following hardware components:

- Cisco UCSX-210c M7 Compute Node for VDI workload (two 4th-generation Intel Xeon Scalable Refresh Gold Processors with 2 TB of memory (32 GB x 32 DIMMs at 4400 to 4800 MHz) with:

  - Intel Xeon Gold 6442Y Processor

  - Intel Xeon Gold 6448H Processor

  - Intel Xeon Gold 6454S Processor

- Cisco UCS VIC 15231 mLOM (Cisco UCSX-210c M7 Compute Node)

- Two Cisco UCS 6454 Fabric Interconnects (4th-generation fabric interconnects)

- Two Cisco Nexus® 93180YC-FX Switches (optional access switches)

The software components of the solution follow:

- Cisco UCS Firmware Release 5.1

- VMware ESXi 8.0 for VDI host

- Citrix Virtual Apps and Desktops 2203 LTSR

- Microsoft Windows 11 64-bit Operating System

- Microsoft Office 2021

Test strategy

We planned to evaluate the performance of the 4th-generation Intel Xeon Scalable Processors (Table 1) that are available to order for the Cisco UCS M7 servers for use in the VDI environments. Table 2 lists the CPU suffixes.

Features of the standard 4th-generation Intel Xeon Scalable Processors (Sapphire Rapids) include:

- Intel series chipset

- Cache size of up to 112.5 MB

- Up to 60 cores

Table 1. Processors used in the evaluation

Product ID (PID)	Clock Frequency (GHz)	Power (W)	Cache size (MB)	Cores	UPI1 links (GT/s)	Highest DDR4 DIMM clock support (MHz)2
6000 Series processors						
UCS-CPU-6442Y	2.6	225	60	24	3	4800
UCS-CPU-I6448H	2.2	225	60	32	3	4800
UCS-CPU-I6454S	2.2	270	60	32	4	4800

Table 2. CPU suffixes

CPU suffix	Description	Features
S	Max SGX enclave size	This suffix supports the Max SGX stackable managed switches enclave size (512GB) to enhance and protect the most sensitive portions of a workload.
H		Workloads such as database, analytics, and virtualization are all optimized
Y	Speed select – performance profile	Intel Speed Select Technology allows you to set a guaranteed base frequency for a specific number of cores, and assign this performance profile to a specific application or workload to guarantee performance requirements. In addition, you can configure settings during runtime and provide additional frequency profile configuration opportunities.

We evaluated the performance of each processor against the Login VSI knowledge worker test workloads in benchmark mode.

Knowledge workers

Knowledge workers are individuals in an organization who use a large number of applications to perform their duties. Examples of knowledge workers include sales and marketing professionals, business development managers, healthcare clinicians, and project managers.

In some cases, Remote Desktop Hosted Sessions (RDSH) server sessions or published applications can support these workers, but in most cases, organizations provide a medium-capability Windows 11 virtual desktop to these users.

You can find additional information about Login VSI and all the workloads we tested for this document here.

Test methodology

We installed the chosen processor candidates in a Cisco UCSX 210c M7 server and ran Login VSI benchmark mode tests at calculated maximum user densities to determine the actual maximum recommended user density per server. The maximum recommended user density is some number of users who complete the Login VSI workload with all attempted users active and logged off without triggering Login VSImax. In addition, average CPU usage on the host should not exceed 90 percent during the test.

We used the maximum recommended user density achieved to determine server loading in a server maintenance or failure scenario.

We used non-persistent Windows 11 virtual machines in Desktop Pools with Floating Assignment across all processor testing (Table 3).

Table 3. Virtual machine configuration

OS	Virtual CPU	Memory	Virtual NIC
Windows 11	2 vCPUs	4 GB of memory	1 x 25-GB vNIC

Test data

This section presents the data from the test runs for the selected processors.

In addition to the Login VSI test suite, we measured host usage by gathering data from ESXTOP.

Microsoft Windows 11 and Citrix Virtual Apps and Desktops LTSR 2203 single-server synopsis: Intel Xeon— Gold 6442Y Processor

The test results are summarized here and in Figures 4–6.

- Operating system: Windows 11 64-bit with Microsoft Office 2021 and VMware optimizations

- 2 vCPUs; 4 GB of RAM

- Maximum number of users: 250 users running Login VSI Knowledge Worker workload with Windows 11

- Recommended number of users: 225 users running Login VSI Knowledge Worker workload with Windows 11
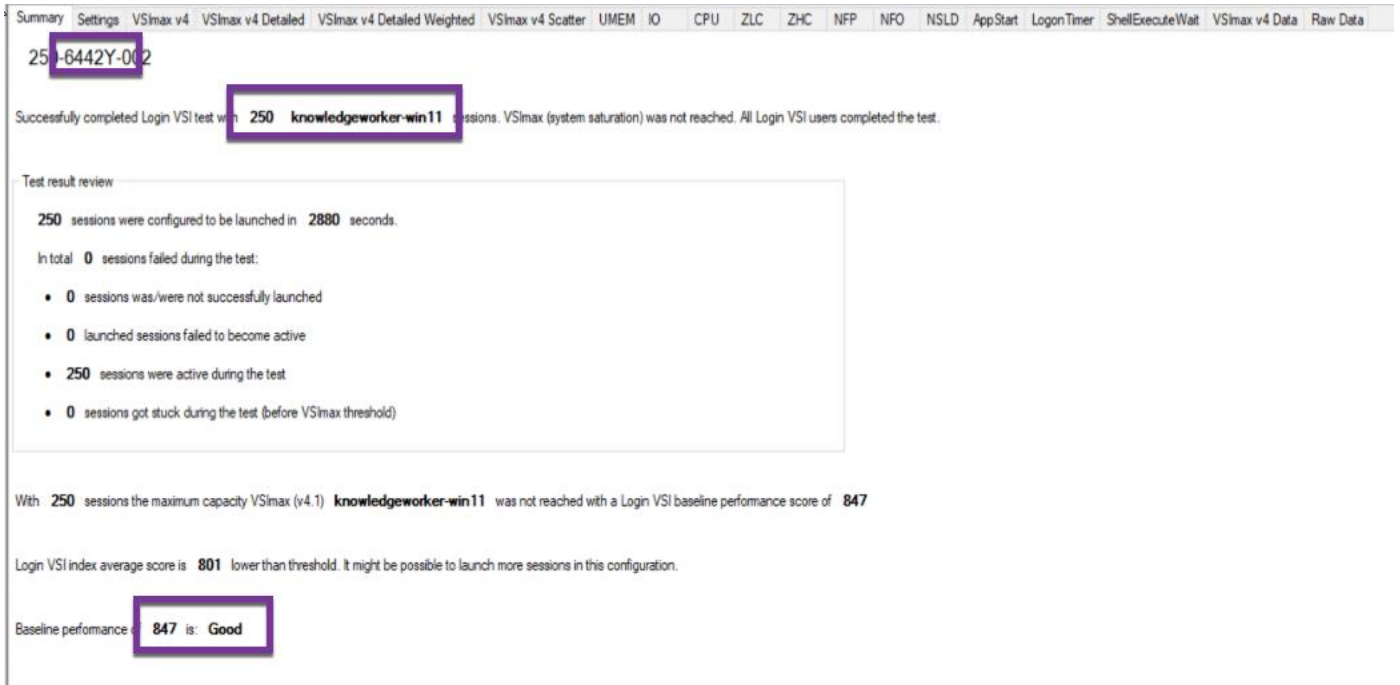
Figure 4.
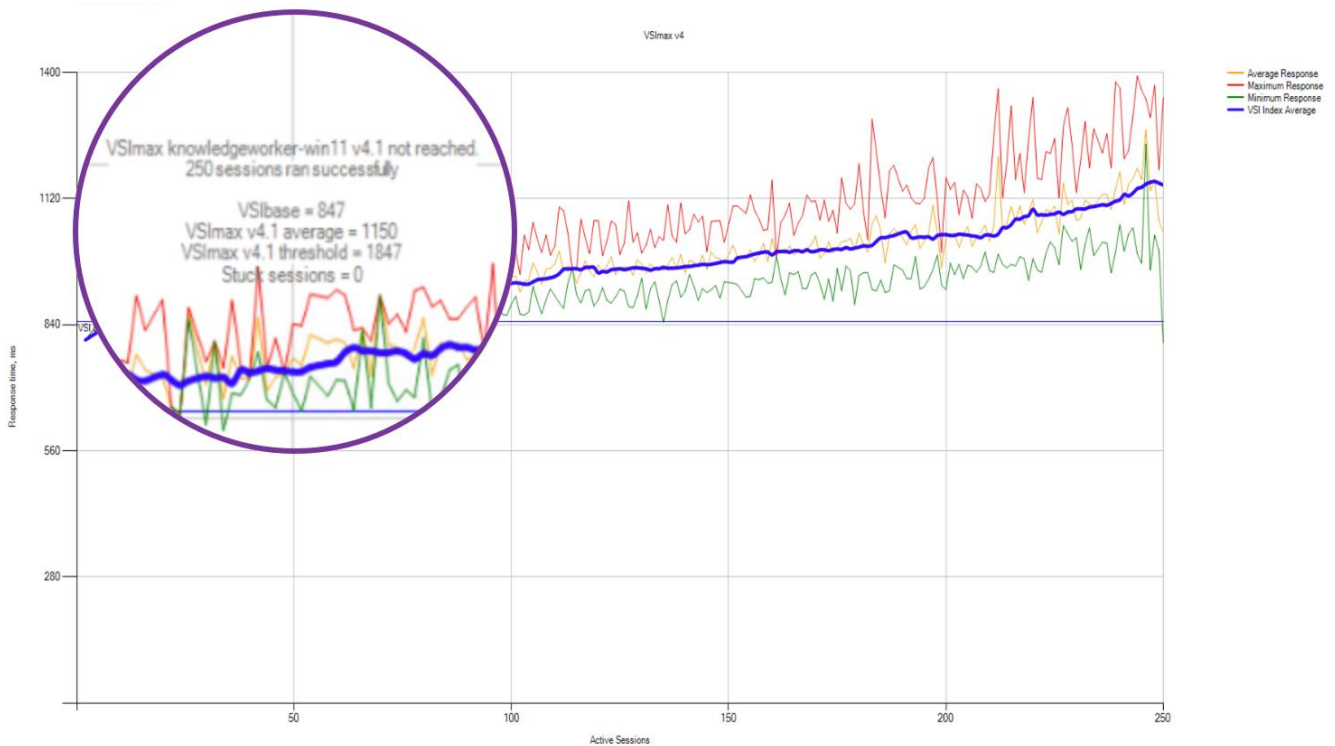Login VSI maximum end-user experience summary for maximum number of users test

Figure 5.
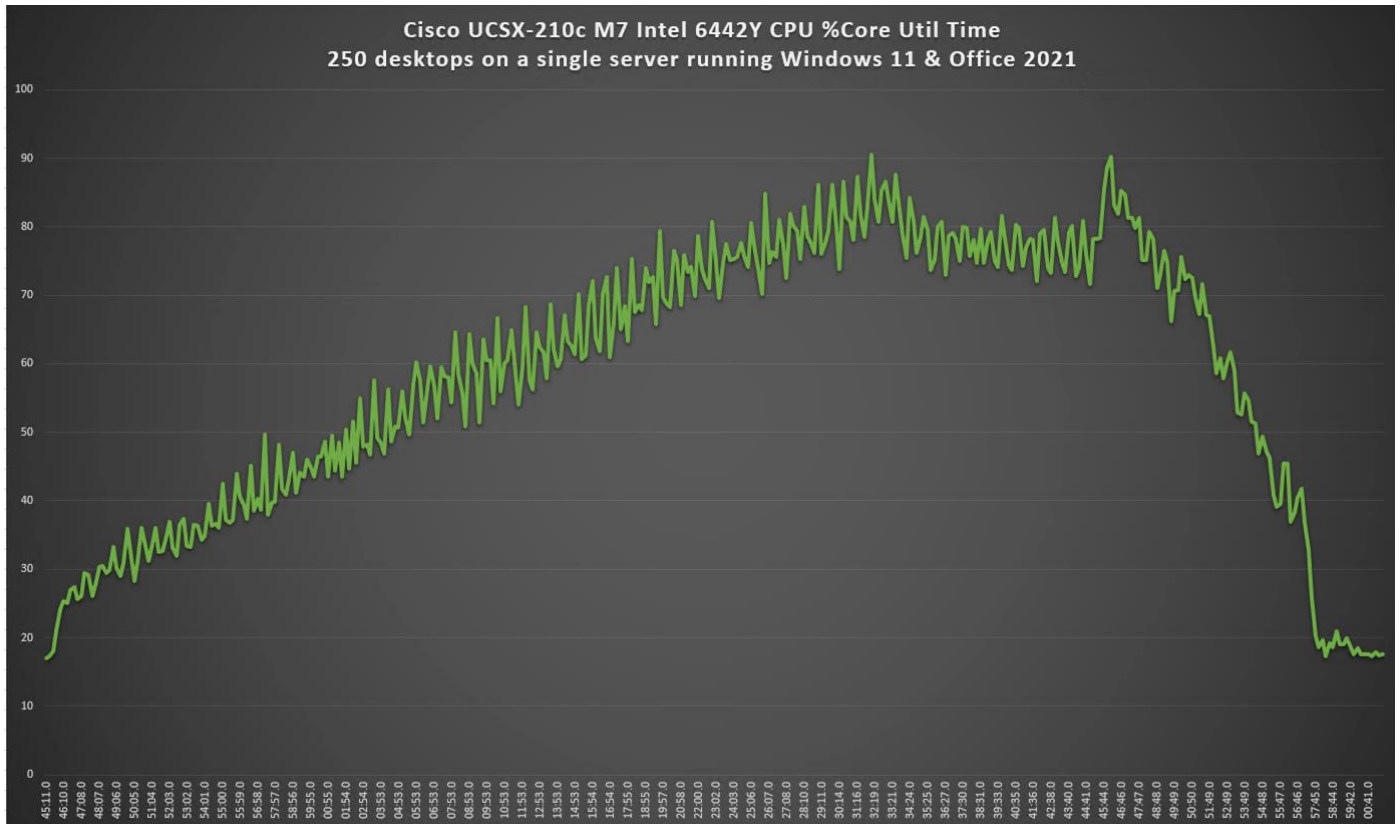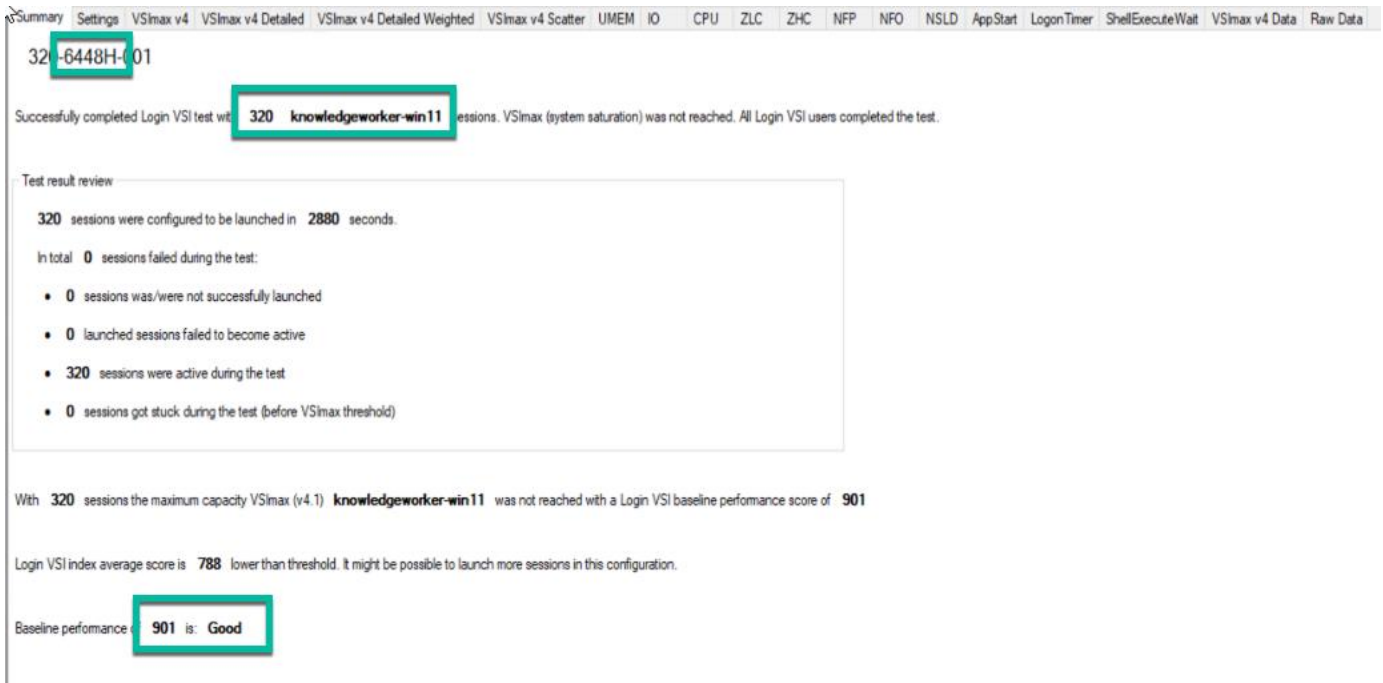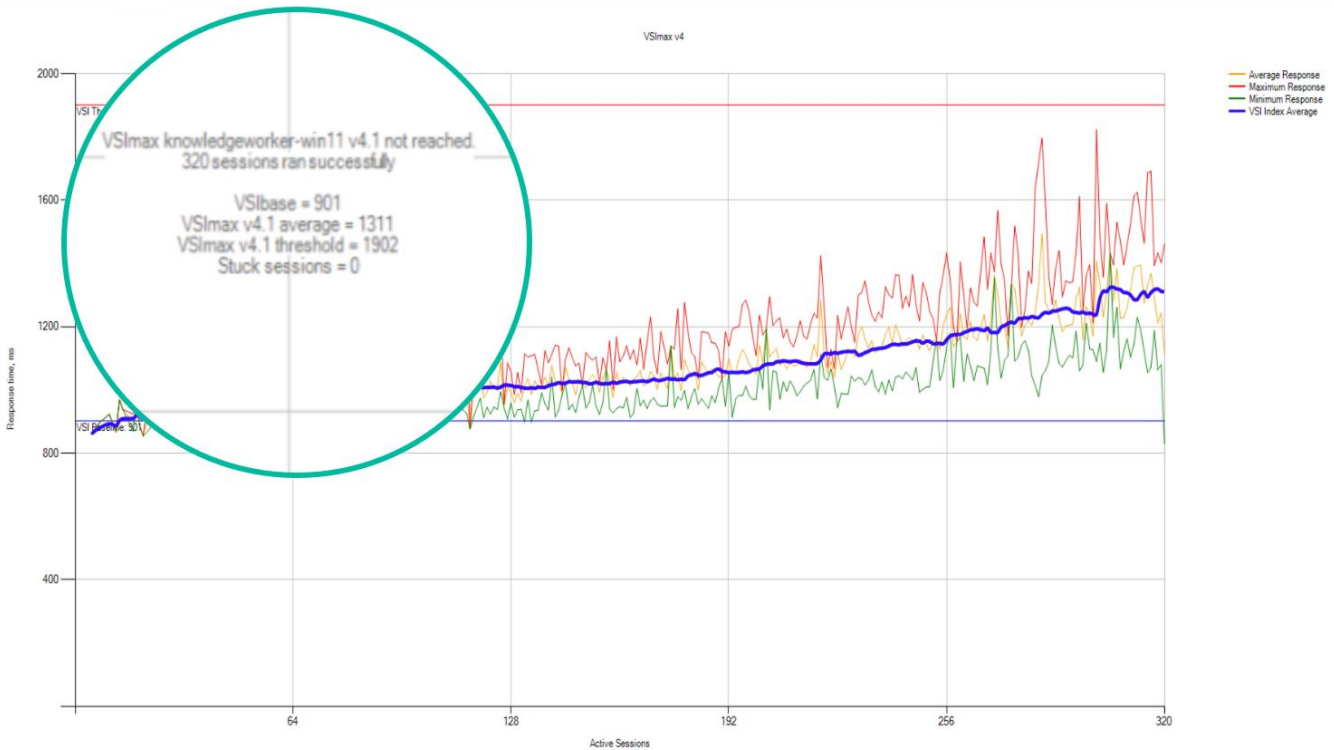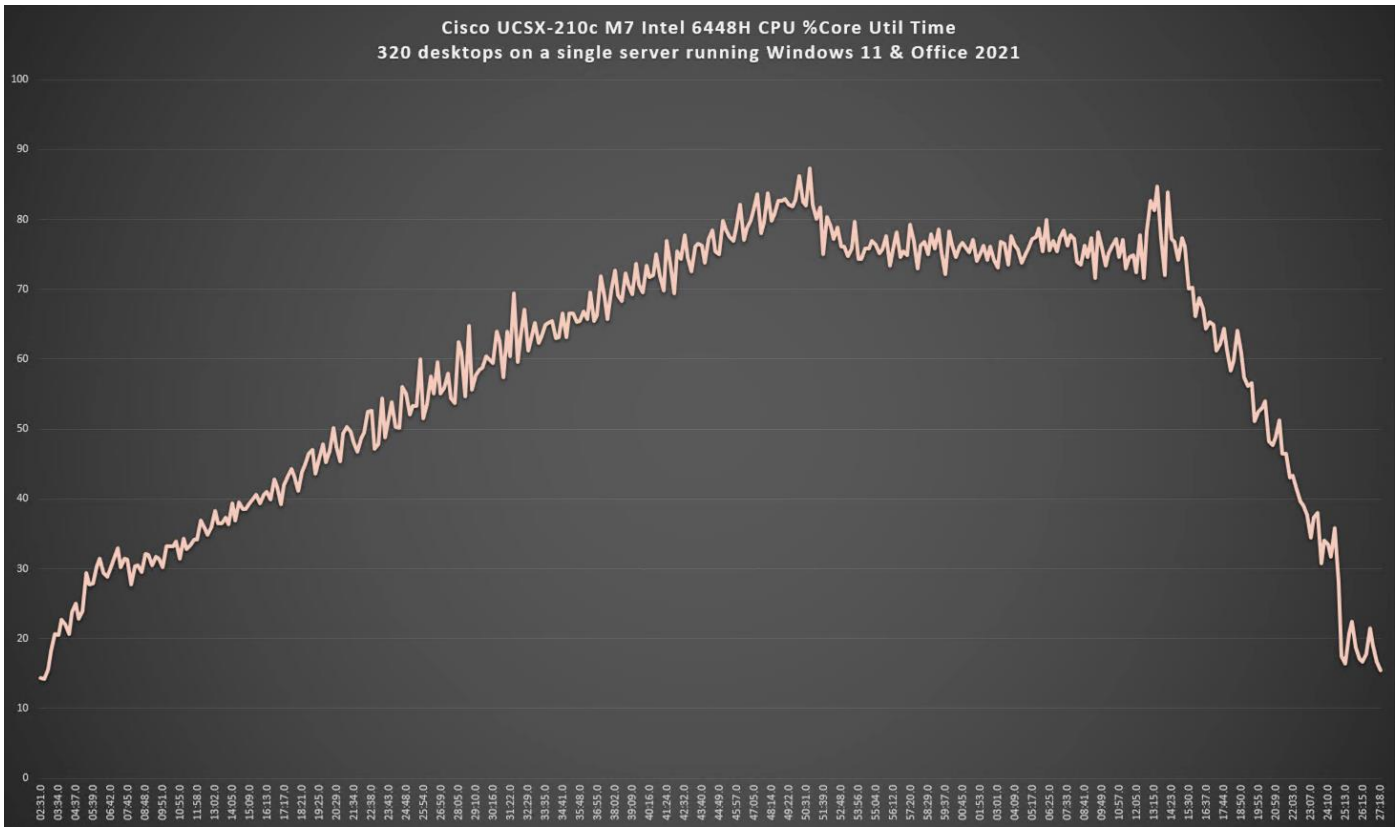Login VSI end-user experience performance chart for recommended number of users test

Figure 6.
VMware ESXi host CPU usage percentage during recommended number of users test

Microsoft Windows 11 and Citrix Virtual Apps and Desktops LTSR 2203 single-server synopsis: Intel Xeon Gold 6448H Processor

The test results are summarized here and in Figures 7−9.

- Operating system: Windows 11 64-bit with Microsoft Office 2021 and VMware optimizations

- 2 vCPUs; 4 GB of RAM

- Maximum number of users: 320 users running Login VSI Knowledge Worker workload with Windows 11

- Recommended number of users: 300 users running Login VSI Knowledge Worker workload with Windows 11

320-6448H-001

Successfully completed Login VSI test with **320** **knowledgeworker-win11** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

320 sessions were configured to be launched in **2880** seconds.

In total **0** sessions failed during the test:

- **0** sessions was/were not successfully launched

- **0** launched sessions failed to become active

- **320** sessions were active during the test

- **0** sessions got stuck during the test (before VSImax threshold)

With **320** sessions the maximum capacity VSImax (v4.1) **knowledgeworker-win11** was not reached with a Login VSI baseline performance score of **901**

Login VSI index average score is **788** lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **901** is: **Good**

Figure 7.
Login VSI maximum end-user experience summary
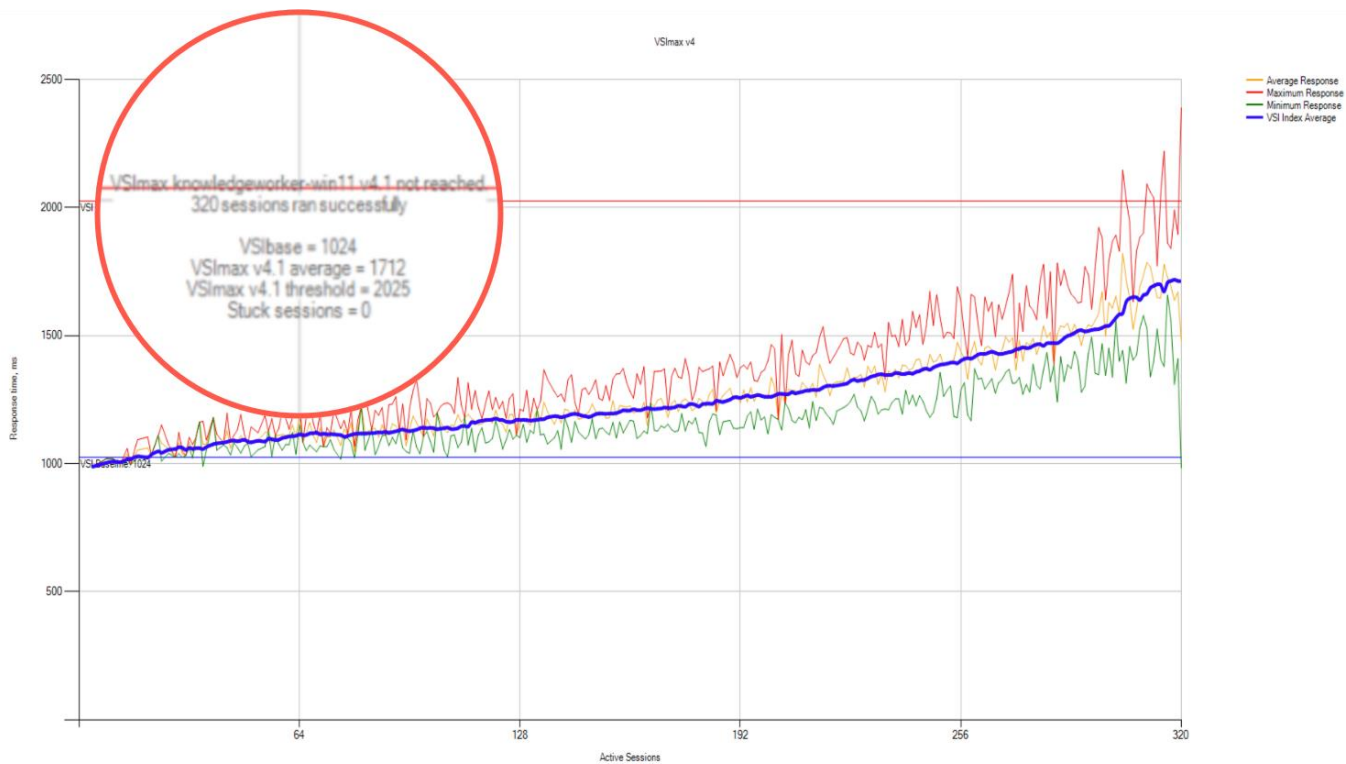

Figure 8.
Figure 8 Login VSI end-user experience performance chart for recommended number of users

Figure 9.
VMware ESXi host CPU usage percentage during recommended number of users test

Microsoft Windows 11 and Citrix Virtual Apps and Desktops LTSR 2203 single-server synopsis: Intel Xeon Gold 6454S Processor

The test results are summarized here and in Figures 10–12.

- Operating system: Windows 11 64-bit with Microsoft Office 2021 and VMware optimizations

- 2 vCPUs; 4 GB of RAM

- Maximum number of users: 320 users running Login VSI Knowledge Worker workload with Windows 11

- Recommended number of users: 300 users running Login VSI Knowledge Worker workload with Windows 11

Figure 10.
Login VSI maximum end-user experience summary

Figure 11.
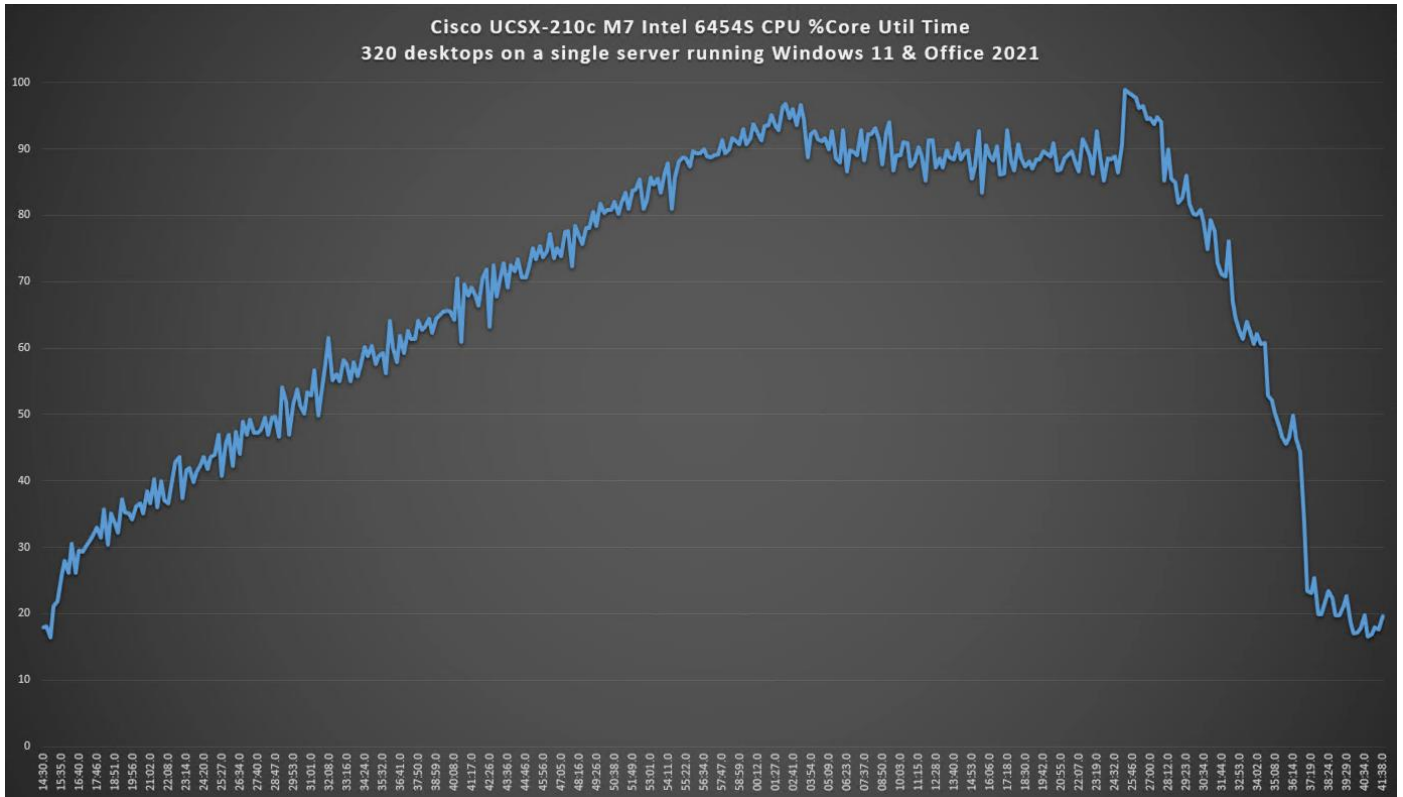Login VSI end-user experience performance chart for recommended number of users

Figure 12.
VMware ESXi host CPU usage percentage during recommended number of users test

Conclusion

By evaluating a variety of processor options in the 4th-generation Intel Xeon Scalable Family, we provided performance characteristics for Citrix Virtual Apps and Desktops LTSR 2203 and Microsoft Windows 11 virtual desktops sessions that should help Cisco customers more accurately architect VDI environments.

We identified the maximum recommended workload for each processor for Windows 11 virtual machines for knowledge workers. The maximum recommended workload is used to plan for maintenance and failure scenarios. During normal operations, fewer virtual machines would run on the clusters supporting your users. These numbers will vary based on the size of the VMware cluster hosting virtual machines to provide N+1 resiliency.

Table 4 summarizes our recommendations for the maximum recommended workloads.

Table 4. Microsoft Windows 11 and Citrix Virtual Apps and Desktops LTSR 2203 virtual desktops

Processor part number and quantity	Cores	Memory	Memory speed (Hz)	Knowledge worker
UCS-CPU-I6442Y x 2	48	2 TB (32 x 32 GB)	4400	250
UCS-CPU-I6448H x 2	64	2 TB (32 x 32 GB)	4400	320
UCS-CPU-I6454S x 2	64	2 TB (32 x 32 GB)	4400	320

Each customer's environment and workloads are different. The recommended densities shown in the table are starting points for your unique environment. They are not intended to be performance guarantees.

For graphics-intensive workloads and enhanced-experience Windows 11 workloads, you can use GPUs with additional processors that are suited for that purpose.

For more information

For additional information, refer to the following websites:

- Cisco UCS X-Series Rack Servers:
  - http://www.cisco.com/en/US/products/ps11265/
- Citrix Virtual Apps and Desktops LTSR 2203:
  - https://docs.citrix.com/en-us/citrix-virtual-apps-desktops/2203-ltsr.html
- VMware vSphere 8:
  - https://docs.vmware.com/en/VMware-vSphere/8.0/rn/vmware-vsphere-80-release-notes/index.html
- Login VSI
  - https://www.loginvsi.com/
  - https://www.loginvsi.com/products/login-vsi