Citrix Virtual Apps and Desktops (CVAD) for IBM Cloud Performance

August 2021

Author:
David Sandler

1. Overview

The Citrix Virtual Apps and Desktops (CVAD) for IBM Cloud offering provides virtualized solutions for accessing applications from anywhere. Citrix manages the core components on their cloud, while the rest of the customer-managed infrastructure hosting the core application are hosted on IBM Cloud.

The IBM Cloud solution provisions and configures a collection of Bare Metal and virtual infrastructure that is then passed on as a fully customer managed solution. The purpose of this white paper is to discuss the performance evaluation of this solution and subsequently the price-per-user that customers can roughly expect. The evaluation was done using Login VSI, a load testing tool used for scale and performance testing of virtual desktop infrastructure. The default workloads provided by the tool can simulate hundreds or thousands of users interacting with a CVAD resource location and provides metrics on stability and scaling performance.

2. CVAD Offering Architecture

The CVAD for IBM Cloud architecture is comprised of several virtual machines and bare metal servers to enable a full customer-managed virtual desktop resource location. Each resource location is comprised of 3 cloud connector VMs for communication with the Citrix Cloud, 1 Linux VM for managing DHCP, 1 Linux VM to function as a proxy to the internet, 1 Windows VM for Active Directory Management, and a variable number of bare metal servers running Citrix Hypervisor and using shared network-attached file storage. Citrix Cloud manages the core components and central control layer.

Once provisioned, it is the customer's responsibility to create the resource pools and the master VM image(s) which will run the Citrix Virtual Delivery Agent (VDA) and deliver various applications and desktops to end-users.

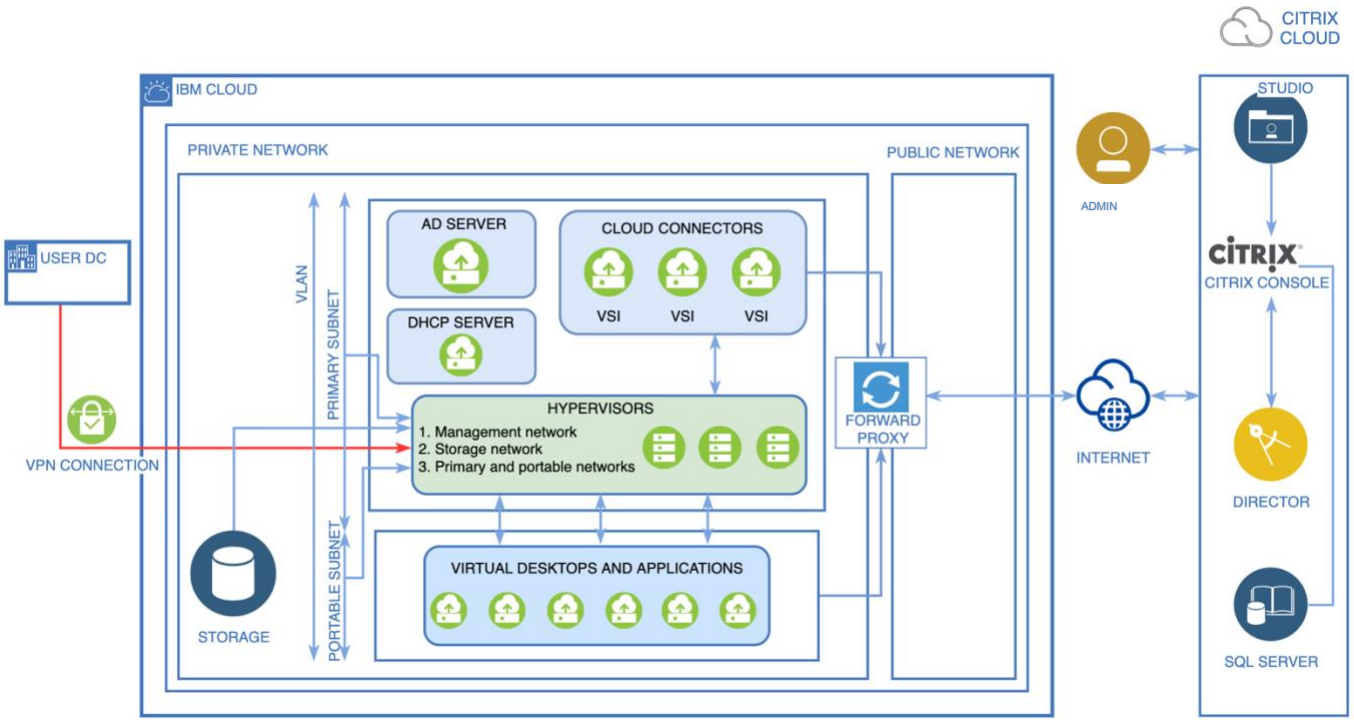

Figure 1: CVAD for IBM Cloud Architecture Diagram

3. Scalability Benchmarking Approach

In addition to the aforementioned CVAD components, in order to properly benchmark their performance, additional resources were provisioned to run the Login VSI benchmarking software. The Login VSI infrastructure was comprised of 1 Windows VM functioning as the VSIShare/test driver and a variable number (up to 100+) of Windows VMs functioning as launcher machines to simulate thousands of users and initiate sessions. Over a period of 48 minutes, hundreds to thousands of simulated-user sessions would launch and system performance would be monitored on the individual session level down to the bare metal host level.

Login VSI provides several built-in workloads, 3 of which we leveraged: Task, Knowledge, and Power Workers. These workloads simulate users at various degrees of usage, from light Microsoft Office and Email task workers up to CPU-heavy Java application power workers to maximize system load. The specific scalability numbers that were collected are then used in a resource sizing calculator that offers recommended bare metal server quantities and configurations to meet their end-user needs based on expected workload levels.

Metrics were collected using the built-in monitoring within Login VSI along with various Windows/Linux utilities such as dstat, perf, Performance Monitor, and the Citrix Hypervisor built in rrd tool. Login VSI provides 2 key high-level metrics for each test execution: The Baseline performance score and the VSImax. The Baseline performance is an indicator of system performance under little to no stress. It is comprised of a calculation of various measurements of CPU, memory, and IO performance, and can be viewed as a quality-of-service metric. The VSImax metric represents how many sessions a system can support before reaching saturation, i.e., when a certain delta threshold (1000ms) from the baseline is reached.

To evaluate system performance, we built and optimized Windows Server 2016 and Windows 10 Images and deployed them onto the hypervisors offered. By running the Login VSI tool, we could saturate the system and measure the tipping point and tune the system for optimal performance.

4. Scalability Testing Results and Observations

The graph below shows the results of a 275 Knowledge Worker run on a single Cascade Lake Bare Metal Server with dual 6248 processors. There were 10 VDAs in this resource location, each with 8 vCPUs and 32 GB of memory. The VMs boot on shared Network-attached file storage, and are running Windows Server 2016 Standard, with OS updates through November 2020. Additionally, the master image has been optimized with the Citrix Optimizer and the Virtual Desktop Optimization tool. In the case below, with a baseline of 657 ms, we tip over into saturation at 1658 ms once 264 users are active on the system.

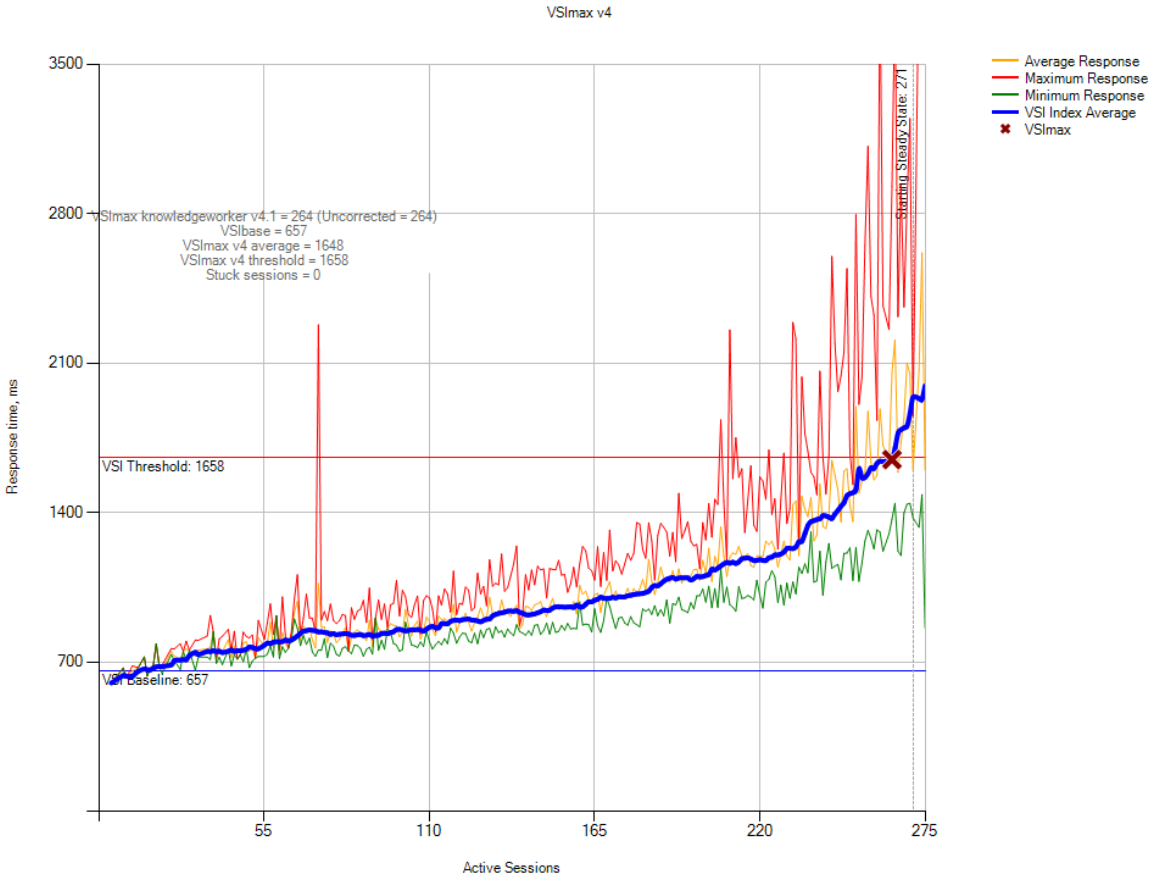

Figure 2. Login VSI Knowledge Worker on a Cascade Lake 6248 hosted-shared resource location.

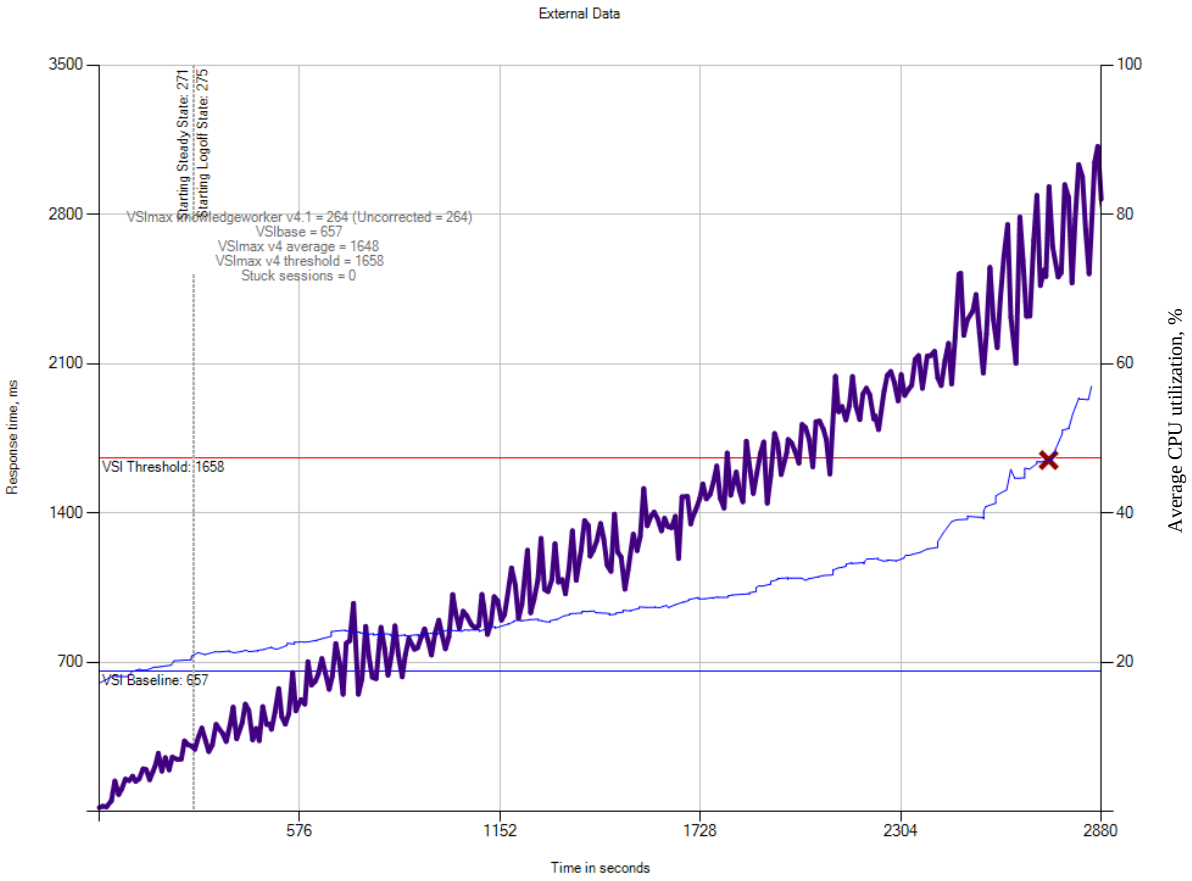

Figure 3. Average CPU utilization percentage

The primary bottleneck for single-host workloads is host CPU utilization. With this in mind, we consistently observed roughly 5-7x hosted-shared users per physical core on the Cascade Lake 6248 with dual processors (40 cores total), depending on the specific workload. The best performance/user density was achieved primarily by continuously optimizing the underlying master VM image and taking advantage of NUMA-placement available starting in Citrix Hypervisor 8.2.

Login VSI Workload	Cascade Lake 6248 + Citrix Hypervisor 8.2 (NUMA disabled)		Cascade Lake 6248 + Citrix Hypervisor 8.2 (NUMA enabled)	
	VSImax	Users per core	VSImax	Users per core
Task	173	4.8	275	6.9 (+35%)
Knowledge	161	4.5	264	6.6 (+29%)
Power	125	3.5	200	4.0 (+25%)

Table 1. Cascade Lake with NUMA

In December 2020, after installing the latest set of Windows 2016 Updates from November, we observed improved single-server scalability performance due to decreased durations for the Login VSI NFO metric, which measures the time to open the file open dialog in Notepad. Note the improvements below.

Login VSI Workload	Cascade Lake 6248 + Citrix Hypervisor 8.2 (NUMA enabled)		Cascade Lake 6248 + Citrix Hypervisor 8.2 (NUMA enabled) + November 2020 OS Updates	
	VSImax	Users per core	VSImax	Users per core
Task	275	5.1	304	7.6 (+11%)
Knowledge	264	5.1	312	7.8 (+18%)
Power	200	4.0	277	6.9 (+39%)

Table 2. November 2020 Windows Server 2016 OS Updates

Based on our evaluation of different server models, the best performance was seen on the platinum series 8260 Cascade Lake, allowing us to reach highs of 8-9x hosted-shared users per physical core. The increased user density can be seen in the price-performance chart below, with the 8260 processor the clear winner for large deployments ranging from several hundred to in the thousands. Additionally, note the inclusion of the silver series 4210 Cascade Lake. Based on a combination of factors including lower base price, lower user capacity and therefore memory requirements, it can be a favorable option for customers looking to support under 200 users.

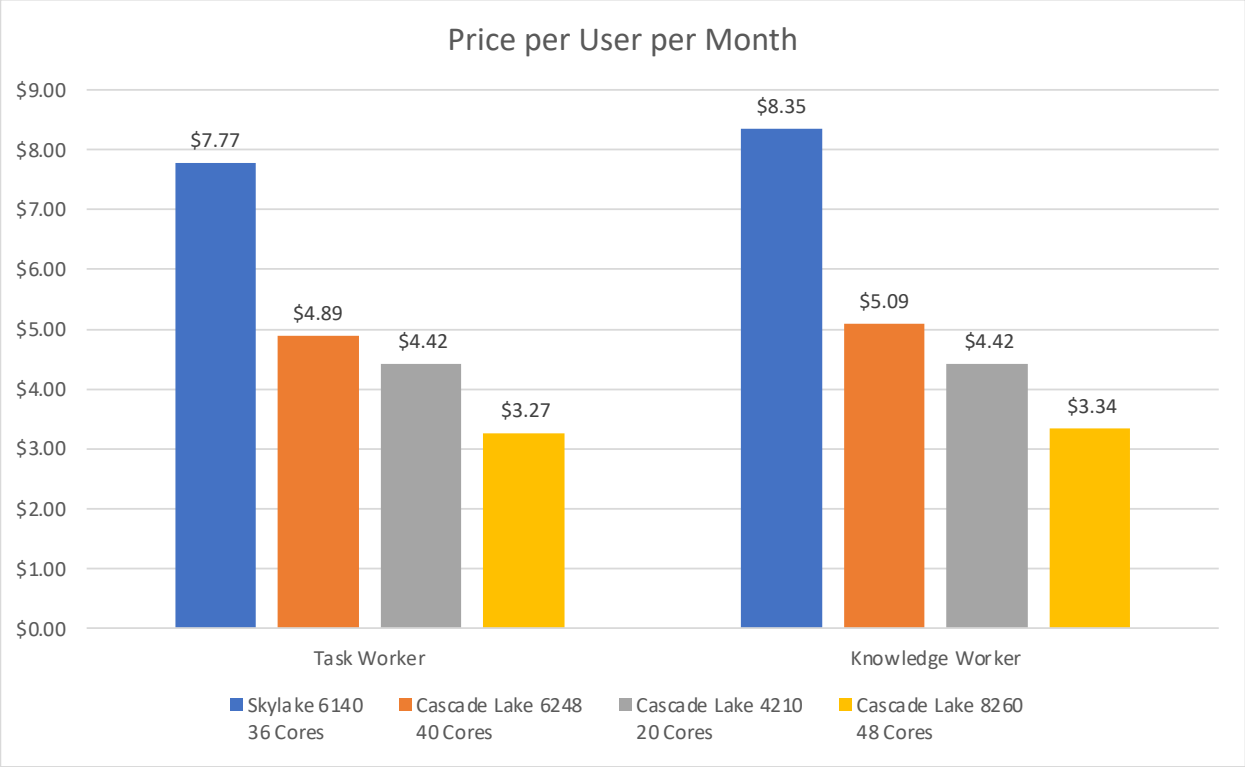

Figure 4. Price Per User per month comparison

In Table 3, we provide the raw VSImax numbers used for the price-performance calculations in Figure 4.

Workload	Processor	Number of System Cores	VSImax	Price Per User per month
Task Worker	Skylake 6140	36	173	$7.77
	Cascade Lake 4210	20	175	$4.42
	Cascade Lake 6248	40	275	$4.89
	Cascade Lake 8260	48	427	$3.27
Knowledge Worker	Skylake 6140	36	161	$8.35
	Cascade Lake 4210	20	175	$4.42
	Cascade Lake 6248	40	264	$5.09
	Cascade Lake 8260	48	418	$3.34

Table 3. VSImax and Price Per User per month breakdown

In addition to single-server scalability measurements, we also performed Login VSI benchmark runs using multiple servers. For large-scale enterprise environments with thousands of users, we observed linear scalability on the Cascade Lake 6248 processor when running 10 servers in a single pool. A test was set up to equally distribute user sessions across 10 hosts running 100 VDA virtual machines. As seen in Table 2, a single host had a Login VSI max of 312 users, while 10 hosts were able to support 3085

users (308.5 users per host) before reaching saturation. Similarly, the Login VSI baseline scores were comparable between the two datapoints.

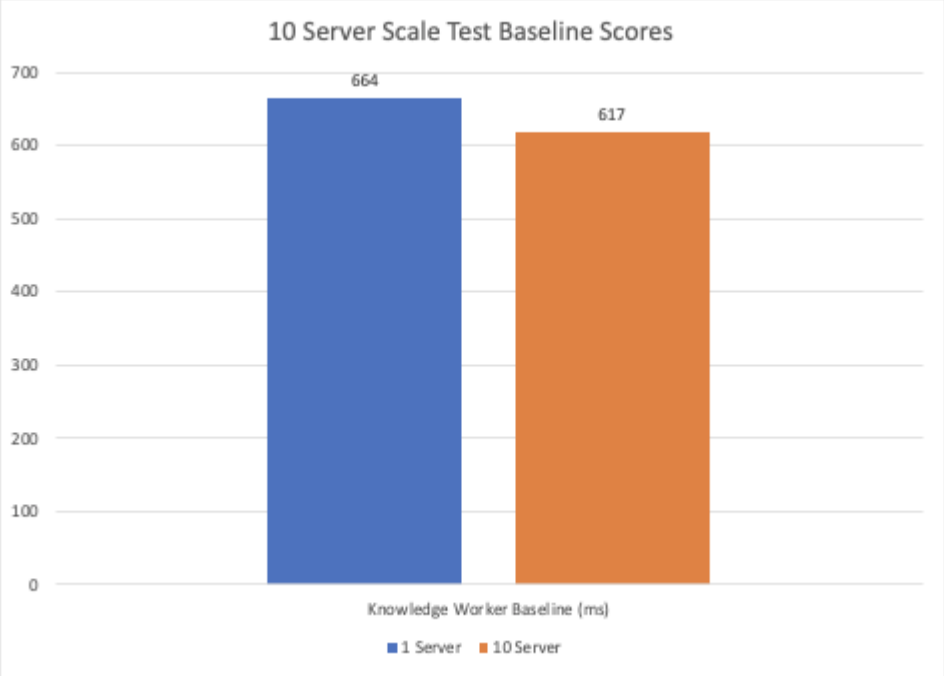

Figure 5. 10 Server Scale Test Baseline Scores

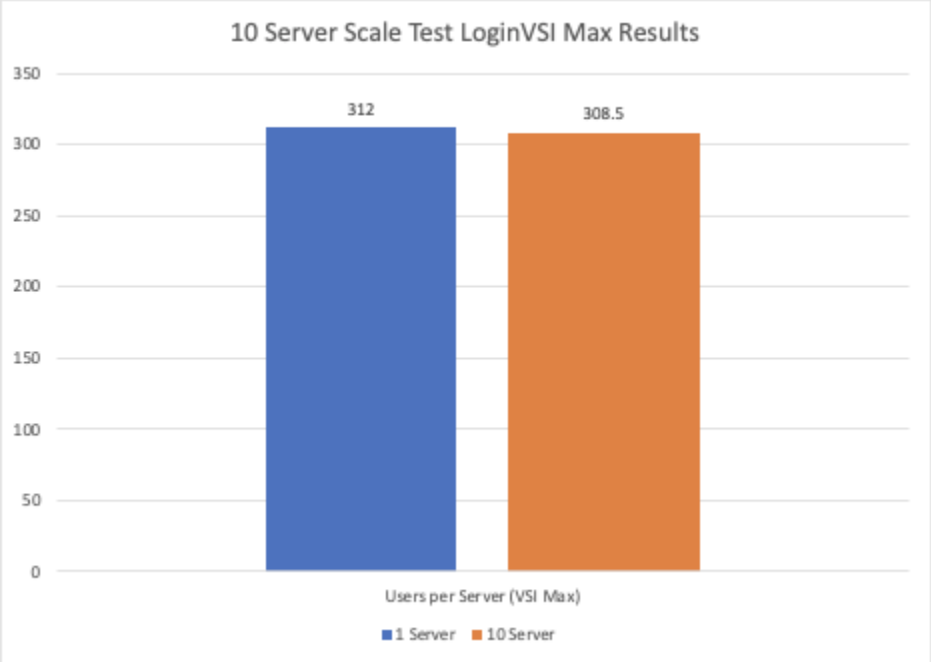

Figure 6. 10 Server Scale Test Login VSI Max Results

5. Performance Optimization and Tuning Recommendations

There are various Windows operating system optimizer tools available online designed to optimize VDI images. Because the same underlying OS image can potentially be deployed hundreds of times, removing unnecessary bloat by disabling system services can lead to gains in scalability and user experience. We recommend experimenting with the various OS optimizers that are available for free from The Virtual Desktop Team and Citrix.

https://docs.microsoft.com/en-us/windows-server/remote/remote-desktop-services/rds_vdi-recommendations-1909

https://support.citrix.com/article/CTX224676

Additionally, if your cluster is running Citrix Hypervisor 8.2, enabling NUMA (non-uniform memory allocation) placement on the host will keep each virtual CPU assigned to the same physical core.

Note that the tests were executed using the legacy TCP network protocol, instead of the default Citrix EDT protocol.

Since EDT is recommended and enabled as the default setting, customers are recommended to evaluate their specific infrastructure performance and tune this setting if needed. For more information on EDT, please refer to the Citrix documentation:

https://docs.citrix.com/en-us/citrix-virtual-apps-desktops/technical-overview/hdx/adaptive-transport.html

https://docs.citrix.com/en-us/citrix-gateway-service/hdx-edt-support-for-gateway-service.html

Conclusion

The purpose of this paper was to provide an overview of the performance of the Citrix Virtual Apps and Desktops for IBM Cloud offering. Given the price-per-user estimates that customers can achieve, the customer-managed solution provides a flexible, highly performant environment for delivering VDIs on the Cloud. For optimal performance, customers should take advantage of publicly available OS optimizers and Login VSI to tune their environment based on their needs. Additionally, decisions regarding NUMA-placement and EDT should also be considered.

References

Overview of Citrix Virtual Apps and Desktops: https://docs.citrix.com/en-us/citrix-virtual-apps-desktops-service.html

CVAD for IBM Cloud: https://cloud.ibm.com/docs/cvad?topic=cvad-about-citrix-virtual-apps-and-desktops

Login VSI: https://www.loginvsi.com/

Citrix OS Optimizer: https://support.citrix.com/article/CTX224676

Virtual Desktop Optimization Tool: https://github.com/The-Virtual-Desktop-Team/Virtual-Desktop-Optimization-Tool

Trademarks

IBM®, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.
SoftLayer® is a registered trademark of SoftLayer, Inc., an IBM Company.
Other company, product, or service names may be trademarks or service marks of others.
The information contained in this document has not been submitted to any formal IBM test and is distributed on an “as is” basis without any warranty either express or implied. The use of this information or the implementation of any of these techniques is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer’s operational environment. While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will result elsewhere. Customers attempting to adapt these techniques to their own environment do so at their own risk.