Citrix® VDI Handbook

Citrix Virtual Apps and Desktops 2402 LTSR

Version 1.0

01 July 2024

cloud™ SOFTWARE GROUP

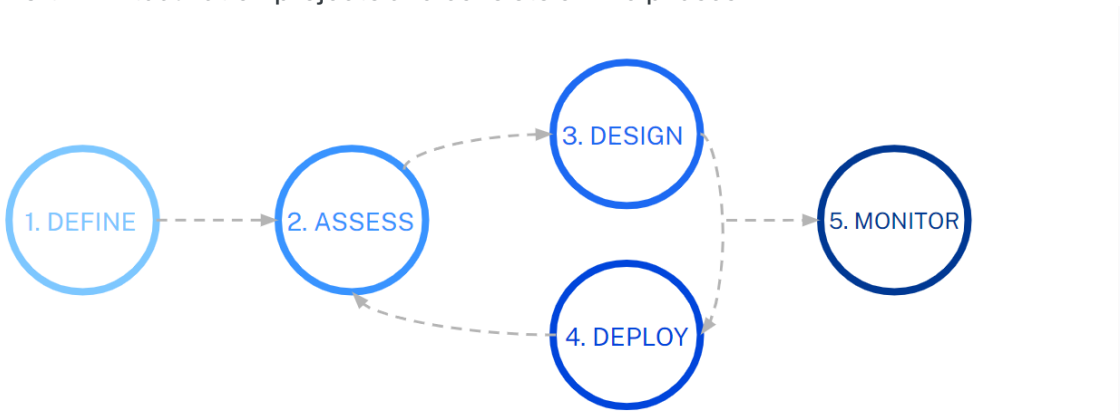
Contents

Overview

Welcome to the Citrix VDI Handbook. This guide is a collaborative effort between the Citrix Product Marketing, Product Management, and Professional Services teams. It is designed to give you an in-depth understanding of Citrix Infrastructure, a pivotal technology in modern enterprise IT environments. Citrix solutions empower organizations to deliver secure, virtualized desktops and applications to users anywhere, on any device, ensuring a seamless and productive user experience while simplifying IT management and enhancing security.

In this handbook, we will explore the fundamentals of Citrix, including its architecture, key features, deployment strategies, and best practices. Whether you are a seasoned IT professional looking to optimize your current VDI implementation or a newcomer aiming to understand the benefits and intricacies of Citrix solutions, this handbook is an essential resource to help you navigate the complexities of virtual desktop and application delivery.

The Citrix VDI Handbook offers detailed information and best practices for deploying and managing virtual desktop infrastructure (VDI) using Citrix technologies. It is a valuable resource for IT professionals, architects, and administrators responsible for designing, implementing, and maintaining Citrix-based environments. The Citrix VDI Handbook follows the Citrix Professional Services methodology, which has been successfully employed across thousands of Citrix virtualization projects and consists of five phases:



- Define: Build the business case for virtualization by creating a high-level project roadmap, prioritizing activities, and estimating storage and hardware requirements.
- Assess: Key business drivers are rated to prioritize work efforts accordingly. In addition, the current environment is reviewed for potential problems and to identify use cases for the project. This information will set the direction for Citrix deployment, upgrade, or expansion.
- Design: Define the architecture required to satisfy key business drivers and success criteria identified during the assessment phase. Topics such as environment scalability, redundancy, and high availability are addressed.
- Deploy: During the deployment phase, the infrastructure is installed and configured as described in the design phase. Before users can access the environment, all infrastructure components should be thoroughly unit and regression-tested.
- Monitor: Define architectural and operational processes required to maintain the production environment.

The Citrix VDI Handbook provides content on the Assess, Design, and Monitor phases. Deployment information can be found in Citrix Product Documentation and Citrix Tech Zone.

Assess

Creating an app and desktop delivery solution begins with a proper assessment. Architects who fail to assess the current environment properly find that they require the assessment information later, forcing them to backtrack, which could stall and put the project at risk. By gathering all of the information from the outset, the architect will gain an appreciation for the current environment and be able to work from the beginning on properly aligning business and user requirements with the overall solution.

The Assess phase is a four-step, simple-to-follow process:

Step 1: Define the Organization

The first step in your virtualization project should be understanding and prioritizing the organization's strategic imperatives. This enables the project management team to define success criteria and allows the design team to create a tailored and optimized architecture.

During this step, it is essential to identify specific business use cases for the Citrix environment. The associated business unit, user personas, workload, and application requirements should all be captured as a part of this process.

Requirements can be captured during meetings or by distributing questionnaires. Meetings are more time-consuming but allow for follow-up questions to be asked and help to simplify the prioritization process. This exercise must be completed jointly by business managers and IT decision-makers since they will have significantly different viewpoints. Most organizations do not focus on technology. They focus on the needs of the user and the organization. These needs can be met with technical solutions, but the team must understand the “Why” of the project.

This table identifies a few other priorities that many organizations often state:

Requestor	Requirement
Business Managers	**Better IT agility and responsiveness** – Flexible solution capable of accommodating periods of change such as rapid growth or downsizing. For example, it enables the business to rapidly set up project offices or temporary points of sale without long delays, hardware acquisitions, or IT notification periods.
	Bring your own device – Empower employees to choose their own devices to improve productivity, collaboration, and mobility.
	Remote, Mobile, and Hybrid Work – The business needs to support home workers to attract and retain top talent and/or traveling employees.

IT Decision Makers	**Cloud Migration —** Take advantage of the benefits of moving to the cloud, including increased flexibility and scalability, to meet business demands.
	Better desktop management — Simplify infrastructure management. IT is not as proactive as it would like and spends too much time "fighting fires."
	Increase security — Data theft, cyber-attacks, or the loss of devices containing sensitive data are big risks, and preventive measures are a top priority.
	Extend desktop hardware lifecycle — Replacing workstations every three to five years to keep up with the operating system or application requirements has been very costly.
	Reducing IT Management Scope: Improve IT efficiency by focusing only on business-critical aspects and offloading the remaining functions to third parties via cloud or service providers.
	Improving user experience - Increasing performance or enabling features that would otherwise not be possible with a geographically dispersed user population.

The prioritization process should be completed in collaboration with the project team, business managers, and IT managers so that all views are considered.

Step 2: Define the User Groups

Although there are multiple approaches to defining user groups, it is often easiest to align user groups with departments, as most users within the same department or organizational unit consume the same set of applications.

User Segmentation

Depending on the department's size, a subset of users with unique requirements might exist. Each defined user group should be evaluated against the following criteria to determine if the departmental user group needs to be further divided into more specialized user groups.

- **Primary datacenter** – Each user will have a primary datacenter or cloud resource location assigned to host their virtual desktop, data, and application servers. Identify the datacenter the user should be assigned to rather than the one they currently use. Users will be grouped based on their primary datacenter so that a unique design can be created for each one.
- **Personalization** – Personalization requirements are used to help determine the appropriate VDI model for each user group. For example, if a user group requires complete personalization, a personal desktop will be recommended as the optimal solution. There are three classifications available:

Personalization	Requirement
None	Users cannot modify user or application settings, such as in a kiosk.
Basic	Users can modify the user-level settings of desktops and applications.
Complete	The User can make changes, including installing applications.

- **Security** – Security requirements are used to help determine the appropriate desktop and policy (or policies) for each user group. For example, a hosted pooled desktop will be optimal if a user group requires high security. There are three classifications available:

Security Level	Description
Low	Users can transfer data in and out of the virtualized environment.
Medium	All authentication and session traffic should be secured; users cannot install or modify their virtualized environment.
High	Besides traffic encryption, no data should leave the data center (such as through printing or copy/paste); all user access to the environment should be audited. Regulatory compliance may be required for a portion or all of the data involved.

- **Mobility** – Mobility requirements are used to help determine the appropriate resource model for each user group. For example, if a user group faces intermittent network connectivity, any model requiring an active network connection is not applicable. There are four classifications available:

Mobility	Requirement
Local	Always uses the same device, connected to an internal, high-speed, and secured network.
Roaming Local	Connects from different locations on an internal, high-speed, secured network.
Remote	Sometimes, it connects from external variable-speed, unsecured networks.
Hybrid	It connects from an internal location on a secured network and external, unsecured networks.

- **Desktop Loss Criticality** – Desktop loss criticality is used to determine the level of high availability, load balancing, and required fault tolerance measures. For example, if there is a high risk to the business if the user's resource is unavailable, the user should not be allocated a laptop, as if that laptop fails, they will not be able to access their resources. Desktop loss critically should be used to inform recovery point objectives (RPO) and recovery time objectives (RTO). There are three classifications available:

Desktop Loss Criticality	Requirement
Low	No major risk to products, projects, or revenue.
Medium	Potential risk to products, projects, or revenue.
High	Severe risk to products, projects, or revenue.

- **Workload** –Types and number of applications accessed by the user impact overall density and the appropriate VDI model. There are three classifications available:

User Type	Characteristics
Task Worker	1-2 office productivity apps or kiosks.
Knowledge Worker	2-10 office productivity apps with light multimedia use.
Power User	Advanced applications, data processing, or application development.
Graphic-Intensive User	High-end graphics capabilities, 3D rendering, CAD, and other GPU-intensive tasks.

Note:

Performance thresholds are not identified based on processor, memory, or disk utilization because these characteristics will change dramatically following the application rationalization and desktop optimization process. In addition, the user's management tools and operating system will likely change during migration. Instead, workload is gauged based on the number and type of applications the user runs and the CPU and memory resources needed for the applications.

Assign VDI Models

As with physical desktops, it is impossible to meet every user requirement with a single type of VDI. Different types of users need different types of resources. Some users may require simplicity and standardization, while others may require high levels of performance and personalization. Implementing a single VDI model across an entire organization will inevitably lead to user frustration and reduced productivity.

Citrix offers a complete set of VDI technologies combined into a single integrated solution. Because each model has different strengths, it is important to choose the right model for each user group within the organization.

This list briefly explains each VDI model:

- **Hosted Apps** - The hosted apps model delivers only the application interface to the user. This approach provides a seamless way for organizations to deliver a centrally managed and hosted application to the user's local PC. The Hosted Apps model is often utilized when organizations must simplify the management of a few line-of-business applications. Hosted Apps includes a few variants:

    - **Windows Apps** – The Windows apps model utilizes a server-based Windows operating system, resulting in many users accessing a single VM model.
    - **VM-Hosted Apps** — The VM-hosted apps model utilizes a desktop-based Windows operating system, resulting in a single user accessing a single VM. This model is often used to overcome application compatibility challenges with a multi-user operating system, like Windows Server 2016, Windows Server 2019, or Windows Server 2022.
    - **Linux Apps** – The Linux apps model utilizes a server-based Windows operating system, resulting in many users accessing a single VM model.
    - **Browser Apps** – The browser apps model utilizes a server-based Windows operating system to deliver an app as a tab within the user's local, preferred browser. This approach provides a seamless way for organizations to overcome browser compatibility challenges when users want to use their preferred browser (Internet Explorer, Microsoft Edge, Google Chrome, Mozilla Firefox, etc.). Still, the web application is only compatible with a specific browser.
- **Shared Desktop** – The shared desktop model hosts multiple-user desktops from a single, server-based operating system (Windows Server 2016, 2019, 2022, Red Hat, SUSE, and Ubuntu). The shared desktop model provides a low-cost, high-density solution; however, applications must be compatible with a multi-user server-based operating system. In addition, because multiple users share a single operating system instance, users are restricted from performing actions that negatively impact other users, for example, installing applications, changing system settings, and restarting the operating system.
- **Pooled Desktop** — The pooled desktop model provides users a random, temporary desktop operating system (Windows 10, Windows 11, Red Hat, SUSE, and Ubuntu). Because each user receives their own instance of an operating system, overall hypervisor density is lower than in the shared desktop model. However, pooled desktops remove the requirement that applications must be multi-user aware and support server-based operating systems.
- **Pooled Desktop + User Personalization Layer (UPL)** – The pooled desktop + User Personalization Layer (UPL) provides all of the benefits of the pooled desktop model and allows you to personalize the individual user layer and preserve users' data and locally installed applications across sessions. There are some applications that do not work in this context.
- **Pooled Desktop + App Layering User Layer —** The pooled desktop + App Layering User Layer provides the pooled desktop benefits and persists each user's profile settings, data, and locally installed applications. Some applications are not supported on the user layer and should not be installed locally. An overview of these applications can be found here.
- **Persistent Desktop** – The personal desktop model provides users with a statically assigned, customizable, persistent desktop operating system (Windows 10, Windows 11, Red Hat, SUSE, and Ubuntu). Because each user receives their own instance of an operating system, overall hypervisor density is lower than in the shared desktop model. However, personal desktops remove the requirement that applications must be multi-user aware and support server-based operating systems. Persistent desktops can be used when end users require the ability to install applications and need setting changes to persist after reboot.

- **GPU Desktop** – The GPU desktop model provides each user with a hardware-based graphics processing unit (GPU), allowing for higher-definition graphical content.
- **vGPU —** The vGPU model allows multiple virtual machines to directly access the graphics processing power of a single physical GPU to render graphics without slowing down the server CPU.
- **Remote PC Access** – The remote PC access desktop model provides users with secure remote access to their statically assigned, physical office PC.
- **Web and SaaS application Access** - Citrix Secure Private Access with Citrix Enterprise Browser provides employees, contractors, and partners zero trust access to web/SaaS applications from anywhere while maintaining the security posture on unmanaged or managed devices. This enhances the user experience and allows for more flexibility. It also protects corporate data and networks, provides complete visibility and governance, and enforces last-mile security. Additionally, it reduces cost, provides a simplified and unified way to manage secure access and limits complexity.

Each user group should be compared against the following table to determine which VDI model best matches the overall user group requirements. In many environments, a single user might simultaneously utilize a desktop VDI model and an app VDI model.

Segment	Segmentation Characteristic	Hosted Apps	Hosted Shared Desktop	Hosted Pooled Desktop	Hosted Pooled Desktop + UPL	Hosted Pooled Desktop + App Layering User Layer	Persistent Desktop	Hosted GPU Desktop	vGPU IIRC	Remote PC Access
Workload	Light	✓	○	○	○	○	○	X	✓	○
	Medium	○	✓	✓	○	○	○	○	○	○
	Heavy	X	X	✓	○	○	○	✓	○	○
Personalization	None	✓	✓	✓	X	X	X	○	○	○
	Basic	✓	✓	✓	X	X	X	○	○	○
	Complete	X	X	X	○	○	✓	○	○	✓
Security	Low	○	○	○	○	○	○	○	○	○
	Medium	✓	✓	✓	○	○	○	○	○	○
	High	○	○	✓	X	X	X	○	○	X
Desktop Loss Criticality	Low	○	○	○	○	○	○	○	○	○
	Medium	✓	✓	✓	○	○	○	○	○	○
	High	✓	✓	✓	X	X	X	○	○	X

✓: Recommended / X: Not Recommended / ○: Viable

Citrix Tips for Success

Start with Windows apps, shared, and pooled desktops – As you can see in the VDI capability table above, the Windows apps, hosted shared, and pooled desktop models can be used in most situations. Reducing the number of VDI models and the required persistence VDI will help reduce deployment time, simplify management, and reduce costs (especially with public cloud resources).

Application Compatibility: Testing application compatibility with operating systems and other applications is essential. Some applications are not multi-session aware and, therefore, cannot be used on multi-session OS VDAs. Applications can also be sensitive to other applications if they have conflicting dependencies.

Perfect match – It may not be possible to select a VDI model that is a perfect match for the user group. For example, you can't provide users with a desktop that is highly secure and offers complete personalization at the same time. In these situations, select the VDI model that is the closest match to the organization's highest priorities for the user group.

Desktop loss criticality – Only three VDI models meet the needs of a high desktop loss criticality user group (backup desktops available) – none of which allow for complete personalization. If a high-desktop loss criticality user group also requires the ability to personalize their desktop, they could be provided with a pool of backup desktops (hosted, shared, pooled) and their primary desktop. Although these desktops would not include customizations made to their primary desktop, they would allow users to access core applications.

Consider Operations & Maintenance – The ongoing support of each VDI model should be factored in when deciding on a VDI model. For example, pooled desktops can be rebooted to a known good state, often leading to reduced maintenance versus a persistent desktop, where each desktop is unique. Additionally, solutions such as Microsoft SCCM or Intune must be budgeted for when deploying persistent desktops or Remote PC.

Review Application Compatibility — Reviewing application compatibility helps detect potential conflicts between new and existing applications and operating systems. Ensuring compatibility at this stage reduces the risk of unexpected deployment-related issues.

Step 3: Define the Applications

The next step is determining their required applications once the users have been divided into groups. This is a two-step process:

1. **Application rationalization** – Removing redundant applications from the inventory captured during the data capture will help simplify the application assessment.
2. **Link apps to users** — Use the data capture process results to map applications to user groups.

Application Rationalization

The number of applications identified during the inventory is often surprising, even for organizations that believe they have high application control. Consolidating the list of applications can help reduce the complexity and overall time required.

These guidelines will help ensure that your application list is consolidated appropriately:

- **Multiple versions** – Different versions of the same application may have been identified during the inventory. There are various reasons, including an inconsistent patching or upgrade process, decentralized application management, limited licenses, and situations where users require specific application versions for compatibility with other applications, macros, and document formats. Work with the application owners to reduce the number of versions required. The leading practice is standardizing on a single version of each application, typically the latest.
- **Non-business applications** – Applications not required by the business should be removed from the application inventory to reduce resource requirements and help simplify the overall project. Non-business-related applications are typically found in an application inventory where users can install their own applications, including games, communication clients, screensavers, peripheral software, and media players.
- **Legacy applications** – The inventory may identify legacy applications that have since been retired or are no longer required within the business. These applications may not have been removed from the desktops because there is no established process to do so or because there are always more high-priority activities to complete. These applications should be consolidated during the rationalization stage of the application assessment.
- **Management applications** — The antivirus, application delivery, monitoring, inventory, maintenance, and backup applications will be completely redesigned and consolidated across the organization during the desktop virtualization project.

Application Categorization

Each application included in the project should be categorized based on certain criteria, which will help determine the most appropriate way to host and integrate the app. Each application can be installed directly into the image, virtualized in an isolated container and streamed to the desktop (application packaging like App-V, MSIX, etc.), captured in a unique layer, and attached to the virtual machine (Citrix App Layering), installed on a multi-user Citrix Virtual Apps host and deployed as an application (Hosted Windows App), or run locally. Due to the uniqueness of every application, many large-scale deployments simultaneously utilize multiple approaches.

Applications can be categorized as follows:

- **Common –** Applications used by most users (greater than 75%).
- **Anomalous –** Applications used by a minority of users (less than 75%).
- **Resource Intensive –** Applications with high computing requirements like 1+ GB of RAM or more than 50% CPU resources.
- **Technically Challenging –** Applications that are complex to set up have extensive dependencies on other applications or have specialized configurations.

Application Characterization

These characteristics can be identified for each application so that the right application delivery model can be selected during the design phase of the project:

- **Complex** — An application should be classified as complex or technically challenging if it is difficult to set up, has extensive dependencies on other applications, or requires a specialized configuration, such as an Electronic Medical Records (EMR) application. Complex applications must be identified during the application assessment because they are not generally appropriate for installation in a pooled/personal desktop model or delivery by application streaming. Delivering complex applications as a hosted app often helps to reduce the complexity of the base desktop image.
- **Demanding** – Collecting application resource requirements allows the virtualization infrastructure to be sized and an appropriate application delivery model to be selected. For example, demanding or resource-intensive applications will not be delivered via a pooled/personal desktop model because there is limited control over how users share resources. There are two classifications available in the user assessment worksheet:
    - Resource Intensive – Application requires 2-4GB+ of RAM or averages 50%+ CPU utilization.
    - None – The application is not resource intensive.
- **Peripherals** — If applications require connectivity with peripheral devices, identify the interface required to make them available when running from a virtual session.
- **Restrictions** – Application access may need to be restricted due to insufficient licenses/resources and to protect sensitive data/tools. For example, applications with limited licenses should not be installed on a base image shared with unlicensed users. There are three restricted access categories in the application assessment workbook:
    - No — The application has no restrictions and can be accessed by any user within the organization.
    - User Group – The application may be installed on a multi-user operating system, but only a specific group of users should be provided with an icon.
    - Virtual Machine – The application should only be installed on a virtual machine accessible by authorized users, often because of licensing requirements.

Step 4: Define the Project Team

Desktop virtualization is a fundamental change that requires close collaboration between various business and technical teams to succeed. For example, the virtualization and desktop teams must work together to ensure that the virtual desktop image meets user needs while also being optimized for the data center or public cloud. Failure to build a cohesive project team with the right roles and skill sets can negatively impact performance, availability, user experience, and supportability while increasing costs and risk.

The following tables identify the business and technical roles required during an enterprise virtual desktop deployment. Although the list may seem large, many of these roles are only required briefly, and a single person may perform multiple roles. The project manager and Citrix architect are considered full-time roles, with other team members being brought in only when

required. The project manager role is key to ensuring that the right people are involved in the project at the right time.

Business Roles

This table is not meant to be prescriptive but to give examples of the various responsibilities and how they can be split between teams. Most organizations have the same people do multiple of these roles.

Role	Description	Example Responsibilities
Project Sponsor	The project sponsor is a senior company executive who recognizes the benefits that desktop virtualization will bring to the business. The chief technology officer (CTO) often performs the role of project sponsor.	All steps • Define key project milestones • Track high-level progress against the plan • Track expenditure against the budget • Manage scope changes • Brief steering committee on progress • Ensure project teams are in alignment on strategy
Project Manager	A project manager plans, executes, and oversees the implementation of projects throughout an organization. They ensure projects are completed on time and within budget, and they collaborate with cross-functional teams.	• Create and update the project plan • Maintain issue and risk register • Create weekly project reports and status meetings • Maintain project racking • Track expenditure against the budget • Organize project workshops and meetings • Ensure prerequisites are in place
Business Manager	Depending on company structure and size, business managers oversee planning and performance at a department, region, or company level. A business manager understands the requirements necessary for their employees to be successful.	Assess • Assist with application consolidation project. • Provide details on the connectivity requirements of the user group, including offline usage

		• Provide details on the risk tolerance of the user group • Identify requirements for peripherals Deploy • Promote the benefits of desktop virtualization • Assist with coordinating the rollout
Business Continuity Manager	The business continuity manager ensures that an organization can continue functioning after a disruptive event such as a natural disaster, crime, or human/computer error.	Assess • Provide Citrix architect with a detailed understanding of the current business continuity plan. Design • Update the business continuity plan to incorporate the new Citrix infrastructure. Deploy • Test business continuity plan
Test Manager	The test manager is responsible for ensuring that the test and user acceptance environments match the production environment as closely as possible. The test manager helps to reduce risk by ensuring that changes are fully tested before being implemented in production.	Assess • Provide Citrix architect with a detailed understanding of current testing infrastructure and processes. Design • Work with Citrix architect to design an appropriate testing infrastructure and test plan for the new Citrix environment. Deploy

		• Ensure that testing design is implemented correctly and new Citrix infrastructure is fully tested before rollout
Application owners	An application owner is a subject matter expert on specific applications deployed within the business. Application owners ensure that application problems are resolved and upgrades/updates are performed without issue. Application owners are also responsible for managing support agreements with the application vendors.	Assess • Assist with application consolidation project • Identify application licensing limitations • Provide details on security restrictions • Provide details on application dependencies • Provide location of backend resources Deploy • Provide installation prerequisites and install guide • Assist Citrix team with installing and testing applications in the VDI environment
Service desk manager	The service desk manager helps improve productivity and end-user satisfaction by ensuring that production issues are logged, escalated, and resolved in a timely manner. The service desk manager is also responsible for reporting on common issues, call volumes, and service desk performance. The service desk manager also ensures that support staff is trained in the technology.	Assess • Identify common issues with the existing environment • Provide details on support tools currently used • Determine the current skill set for support staff and end users Design • Assist Citrix architect with designing a delegated administration model • Participate in operations and support design workshops

		• Work with the training manager to identify training requirements • Create a training plan for support staff and end users Deploy • Monitor help desk calls for rollout-related issues • Implement a training plan for support staff and end users
Communications Manager	The communication manager is responsible for disseminating key information throughout the organization.	Design • Work with the project manager to create a communications plan Deploy • Relay benefits of desktop virtualization • Inform users of key migration dates • Ensure expectations are set accordingly

Technical Roles

This table is not meant to be prescriptive but to give examples of the various responsibilities and how they can be split between teams. Most organizations have the same people do multiple of these roles.

Role	Description	Example Responsibilities
Citrix Architect	The Citrix architect acts as the design authority for all Citrix products and liaises with other architects to ensure the Citrix infrastructure is successfully integrated into the organization.	Assess • Work with the project sponsor and key stakeholders to identify and prioritize key business drivers. • Oversee user segmentation and app assessment • Map VDI models to user groups • Perform capabilities assessment to determine the current state of readiness.

Role	Description	Example Responsibilities
		• Identify areas of risk and provide remedial actions Design • Create a Citrix design that includes hardware and storage estimates • Coordinate with other architects to integrate Citrix infrastructure into the organization • Work with monitoring architect to ensure that Citrix environment is monitored appropriately • Create operations and support design • Create implementation and rollout design • Create test plan Deploy • Ensure that the Citrix environment is implemented in accordance with the design • Verify that implementation passes the test plan • Creates change control requests • Ensure that the Citrix design is implemented correctly
Active Directory architect	Design authority on Microsoft Active Directory, including Organizational Units (OU) and Group Policy Objects (GPOs).	Assess • Provide Citrix architect with a detailed understanding of current Active Directory architecture. Design • Work with the Citrix architect to design OU structure, group policies, permissions, service accounts, etc. for the new Citrix environment • Update Active Directory infrastructure design to reflect the centralization of user data and accounts Deploy

Role	Description	Example Responsibilities
		• Ensure that the Active Directory design is implemented correctly
Cloud/Virtualization Architect	Design authority on the server and desktop virtualization using Citrix XenServer, Microsoft Hyper-V, Nutanix Acropolis, or VMware vSphere.	Assess • Provide Citrix architect with a detailed understanding of current virtualization architecture. Design • Work with Citrix architect to design hardware, networking, storage, high availability, etc. for server and desktop virtualization • Work with the monitoring architect to ensure that the virtualization environment is monitored appropriately Deploy • Ensure that the virtualization design is implemented correctly
Network Architect	Design authority on networking, including routing, VLANs, DHCP, DNS, VPN, and firewalls.	Assess • Provide Citrix architect with a detailed understanding of current networking architecture. Design • Work with Citrix architect to design a physical network, virtual networks, routing, firewalls, quality of service, remote access, network optimization, etc., for the new Citrix environment • Work with the monitoring architect to ensure that the network is monitored appropriately Deploy • Ensure that network design is implemented correctly
Desktop Architect	If there is no separate desktop team, design authority on Microsoft	Assess

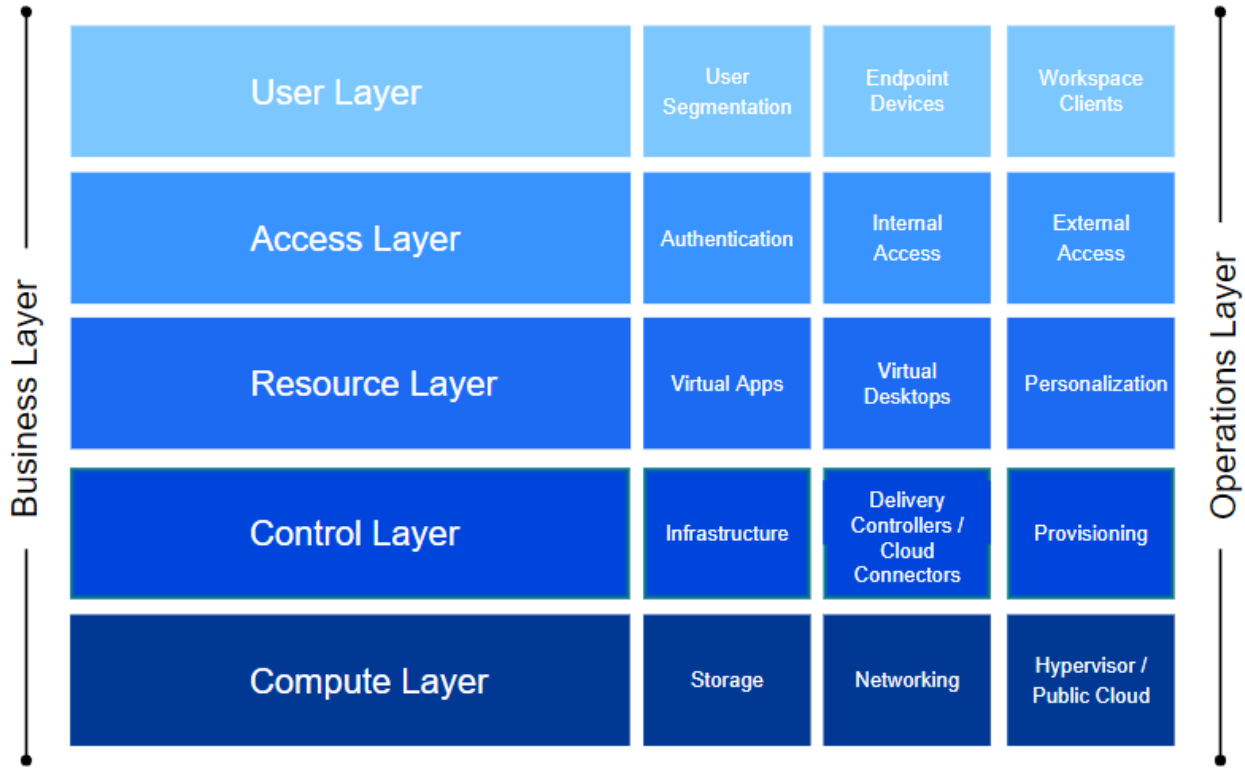
Role	Description	Example Responsibilities
	desktop operating systems often falls under the Citrix architect.	• Provide Citrix architect with a detailed understanding of the current desktop environment. Design • Work with Citrix architect to design core desktop virtual image, core applications, desktop optimizations, etc. for new Citrix environment • Work with monitoring architect to ensure that the virtual desktops are monitored appropriately Deploy • Ensure that desktop design is implemented correctly
Storage Architect	Design authority on storage solutions, including direct-attached storage, storage-attached networks, and network-attached storage.	Assess • Provide Citrix architect with a detailed understanding of the current shared storage environment. Design • Work with Citrix architect to design storage architecture, tiers, sizing, connectivity, etc., for the new Citrix environment • Work with the monitoring architect to ensure that storage is monitored appropriately Deploy • Ensure that storage design is implemented correctly
Backup Architect	Design authority on backup and recovery, including virtual machines, desktops, servers, user data, and databases.	Assess • Provide Citrix architect with a detailed understanding of current backup architecture and processes. Design • Work with Citrix architect and disaster recovery architect to design backup architecture, process, schedule, retention,

Role	Description	Example Responsibilities
		etc., for the new Citrix environment Deploy ● Ensure that the backup design is implemented correctly
Application packaging architect	Design authority on packaging applications for deployment via the systems management team.	Assess ● Provide Citrix architect with a detailed understanding of the current application packaging process and status Deploy ● Ensure that all required applications are packaged according to design
Monitoring architect	Design authority on monitoring, including hardware, network, servers, storage, and security appliances.	Assess ● Provide Citrix architect with a detailed understanding of current monitoring architecture and processes. Design ● Work with Citrix architect to design monitoring architecture, metrics, alerts, etc., for new Citrix environment and supporting infrastructure Deploy ● Ensure that the monitoring design is implemented correctly ● Provide regular reports on capacity and trends during the rollout
Systems management architect	Design authority on systems management, including server/desktop build process, patching, and automated application installation.	Assess ● Provide Citrix architect with a detailed understanding of the current systems management processes. Design ● Works with Citrix architect to define server/desktop build process, patching, and

Role	Description	Example Responsibilities
		application delivery strategy for new Citrix environment Deploy ● Ensure that the systems management design is implemented correctly
Security Architect	Design authority on security, including desktops, servers, networks, and VPNs.	Assess ● Provide Citrix architect with a detailed understanding of current security policy. Design ● Work with Citrix architects to design security standards for the new Citrix environment, including authentication, encryption, port numbers, firewall rules, etc. Deploy ● Ensure that security design is implemented correctly.

Design

Designing a virtualization solution follows a proven process and aligns technical decisions with organizational and user requirements. Architects randomly jump from topic to topic without a standardized and proven process, leading to confusion and mistakes. The recommended approach focuses on working through five distinct layers:



The top three layers, identified during the assessment phase's user segmentation section, are designed independently for each user group alongside the business layer. These layers define the users' resources and how they access them. After completing these three layers, the foundational layers (control, hardware, and operations) are designed for the entire solution.

This process guides design thinking in that decisions made higher up impact lower-level design decisions.

Conceptual Architecture

The conceptual architecture helps define the overarching strategies for the solution based on business objectives and organizational structure.

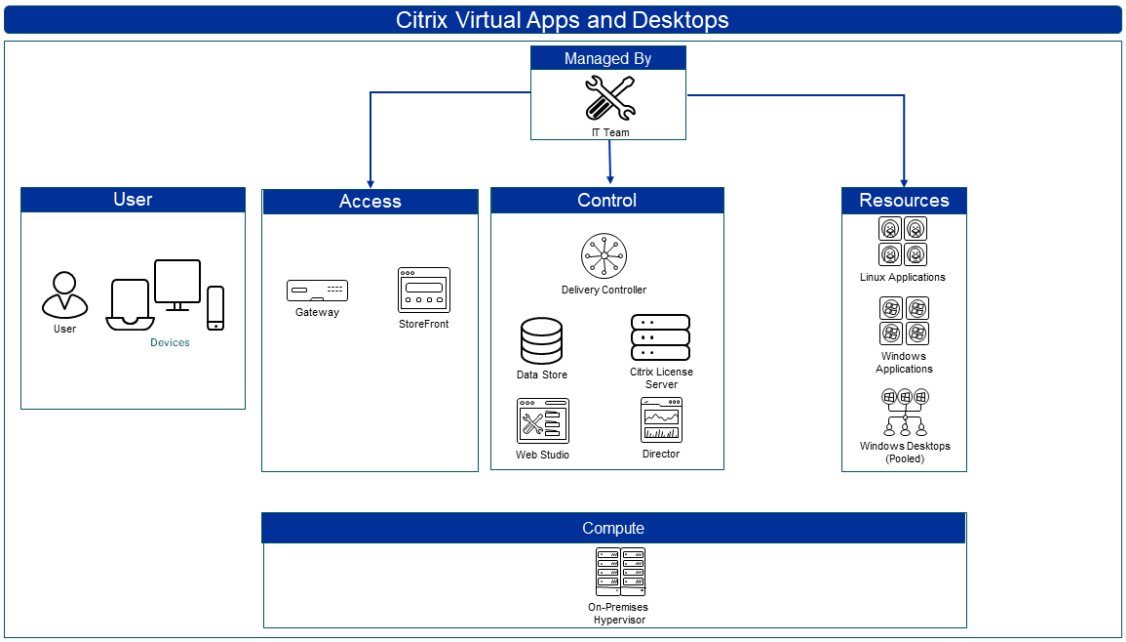
Although an organization's conceptual architecture will change over time, it is worthwhile to start the design phase by defining the long-term objectives around delivery models and the physical geographical distribution of the solution.

Decision: Delivery Model

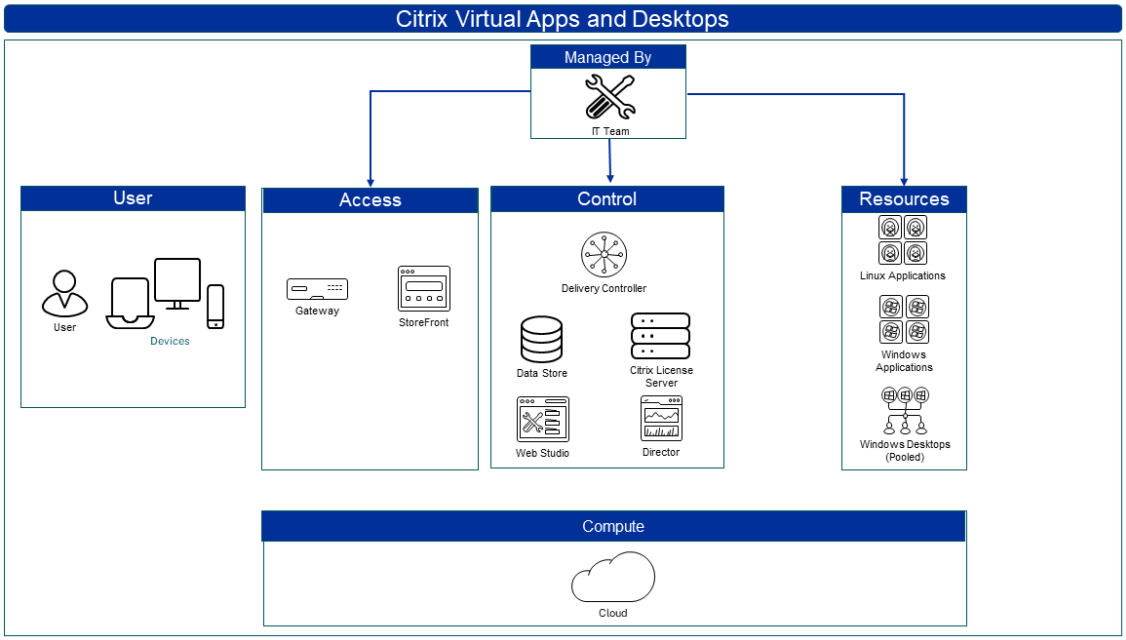
A Citrix solution can take on many delivery forms. The organization's business objectives help select the right approach.

The local IT team's management scope changes even though the infrastructure remains the same for all delivery models.

- **On-premises**: All components are hosted from the organization’s local data center. The on-premises model requires the local IT team to manage every aspect of the solution.



- **Public Cloud** - All components are hosted using Infrastructure as a Service (IaaS) from a public cloud infrastructure. The public cloud model eliminates hardware management from the local IT team‘s management scope.



- **Hybrid Cloud -** A single implementation executes from an on-premises data center and the public cloud. Even though components of the solution are using different hosting providers, the entire solution is a single solution using the same code and managed as a single entity. The local IT team continues to manage all aspects of the solution except for the hardware associated with the cloud-hosted resources.

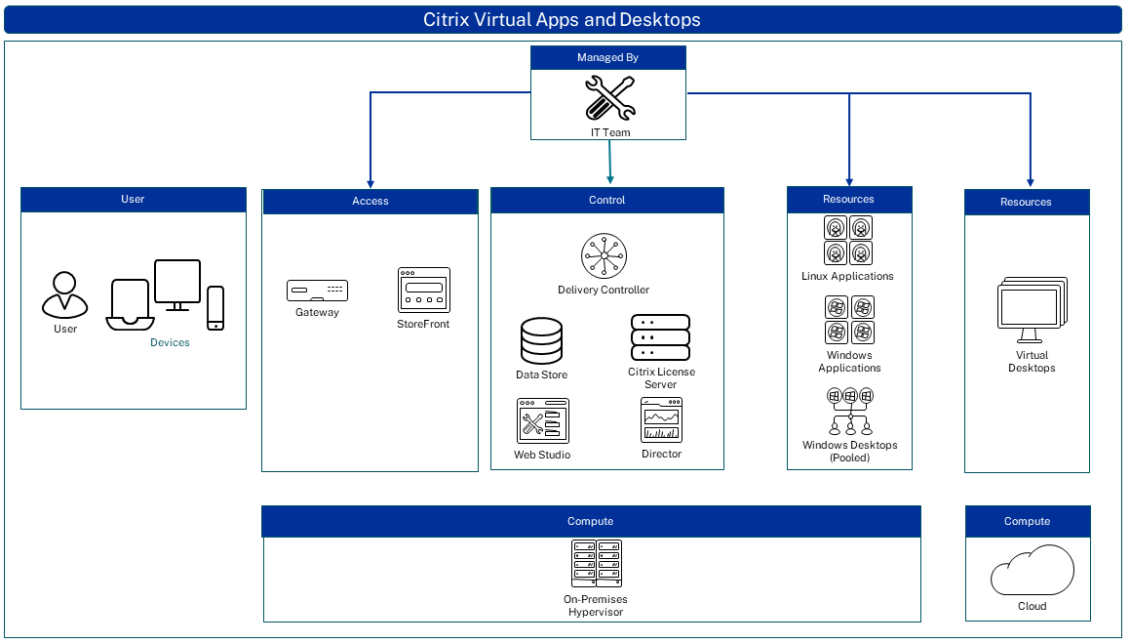
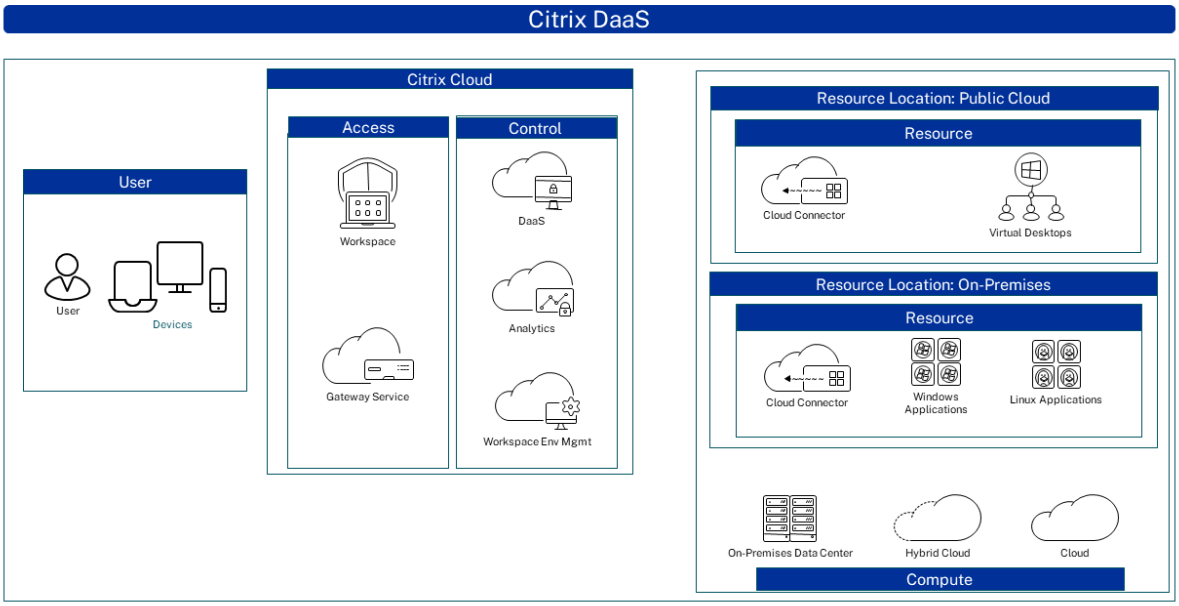


Figure 6: Citrix Virtual Apps and Desktops Hybrid Cloud Architecture

- **Citrix Cloud -** The Citrix DaaS offering from Citrix Cloud breaks a typical deployment into multiple management domains. Citrix hosts and manages the access and control

layer components in the Citrix Cloud. In contrast, the local IT team manages the resource layer components as an on-premises, public cloud, or hybrid cloud model. Citrix manages the hardware, sizing, and updates to the access and control components, while the local IT team manages the resources. In addition, if the public cloud hosts the resources, the local IT team does not have to manage the resource hardware.



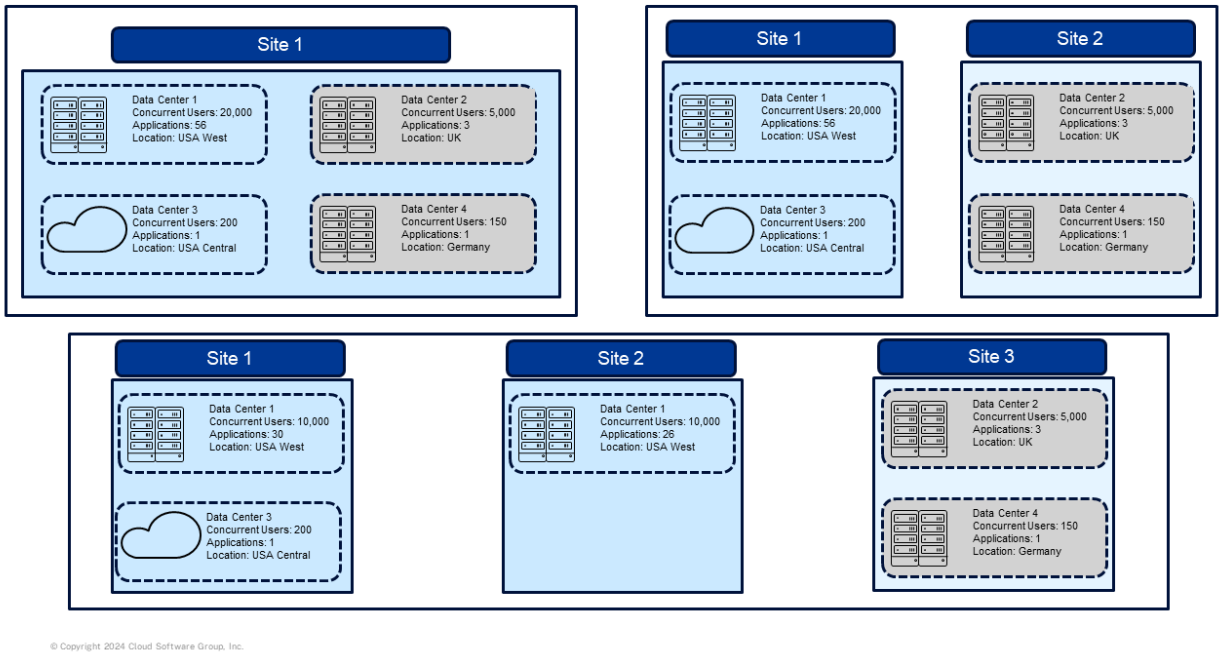
> Note:
>
> Citrix Universal Hybrid Multi-Cloud or the Citrix Platform License must use the Public and Hybrid Cloud options. Citrix for Private Cloud licenses only support Citrix environments deployed fully on-premises.

Decision: Site Topology

A Citrix Virtual Apps and Desktops or Citrix DaaS site groups desktops and applications together to form a single architectural and management entity. The site's persistent and dynamic data, including site configuration, desktop assignments, and session state, is stored in the site's database.

A Citrix DaaS site is a cloud service that provides app and desktop virtualization. It gives IT control of on-prem or cloud-hosted virtual machines, applications, and security while providing anywhere access for any device. End users can use applications and desktops independently of the device's operating system and interface. It is important to note that your cloud tenant is one site in Citrix DaaS.

A site can be contained within a single location, span across multiple locations, or be a partial location. Through rigorous testing, a single Citrix Virtual Apps and Desktops or Citrix DaaS site can support 40,000 concurrent sessions.



Zones (or Resource Locations in Citrix DaaS) subdivide a single site, often associated with geographical locations. There are several factors to consider when determining the overall topology of the Citrix Virtual Apps and Desktops and Citrix DaaS solution:

- **Risk Tolerance** – Create multiple sites to minimize the impact of a site-wide outage. For example, corruption of the site database could affect site-wide availability. For many organizations, the decreased risk from implementing multiple sites outweighs the additional management overhead and supporting infrastructure required. In Citrix DaaS, this would be done by having multiple Resource Locations (aka Zones).

Design Tip

For resiliency reasons, a single zone should be limited to 10,000 VDAs.

- **Security** – Although delegated administration is available, high-security organizations may require complete separation between environments to demonstrate compliance with specific service-level agreements.

Design Tip

If your organization requires complete separation for employees responsible for managing financial data, create two Citrix Virtual Apps and Desktops sites within the same datacenter - one for the financial employees and the second for all other employees.

- **Administrative Boundaries** – Due to billing/chargeback requirements or how IT is structured, multiple sites might be required to separate administrative boundaries.
- **Geographical Connectivity** — Although the zone implementation allows a single site to span geographical locations, there must be enough bandwidth between the satellite and

primary zones for the site database to capture the session information. Higher latency or more significant zones impact the user's response times.

CVAD 2402 LTSR			
Latency	45	90	160
Concurrent Requests	60	60	60
Avg Response Time (s)	7.3	16.2	28
Brokering requests per second	8.2	3.7	2.1
Time to launch 10k users	20m 19s	45m 04s	N/A

Session Count	Max concurrent session launches	Min site-to-site Bandwidth	Max site-to-site Round Trip Latency
Less than 50	20	1 Mbps	250ms
50 to 500	25	1.5 Mbps	100ms
500 to 1000	30	2 Mbps	50ms
1000 to 3000	60	8 Mbps	10ms
3000+	60	8 Mbps	5ms

Administrators should minimize the number of sites and zones to reduce architectural complexity and administrative effort.

Decision: Image Management Strategy

A Citrix Virtual Apps and Desktops solution requires creating and managing master image(s). Organizations must decide early what strategy to pursue for image management.

Installed Images

An installed image requires the administrator to install each image's operating system and applications. This approach is straightforward but duplicates effort as the admin may install the same operating system and core applications across multiple master images.

Maintaining master images also includes duplication of effort as the same operating system version and core applications are part of multiple images, each requiring the same update process.

Scripted Images

Administrators can automate many tasks associated with installed images with scripting or automation tools. Many operating systems and application installs support automated scripting, which mitigates the duplication of effort impact on the administrator's time with installed images.

Infrastructure as Code (IaC) deployments offer several advantages, such as improved scalability, reproducibility, and automation. Using code instead of manual processes to create and update images, administrators can reproduce images quickly at scale with reduced risk of human error. However, IaC deployments do require coding or development skill sets and also introduce an additional layer of complexity.

App Layering Layered Images

Each unique operating system (Windows 11, Windows Server 2022), platform (Citrix Virtual Apps and Desktops 2402 VDA, Citrix Virtual Apps and Desktops 2403 VDA), and application (Microsoft Office, Adobe Acrobat, Google Chrome, and Mozilla Firefox) is a layer. A master image merges one operating system layer, one platform layer, and potentially several application layers.

A layered image approach eliminates the duplication of effort challenges associated with installed and scripted images. Each unique layer is available to any master image. When an application requires an update, that layer receives the updates, and all master images utilizing the layer receive the update. If an organization requires ten unique Windows 11 images, each of the ten images shares the same Windows 11 OS layer. When the administrator needs to upgrade the VDA from version 2212 to 2402 across ten images, the administrator only updates a single platform layer.

Initially, the layered image approach does require more time to deploy because the administrator must build the organization's library of layers. However, once the layers are available, the time to maintain the images is drastically reduced. App Layering represents a significant change in how images are deployed and maintained compared to traditional image management methods. These changes must be coordinated with other teams, such as SCCM/Intune, Windows Server, etc., as the changes will be handled differently than in the rest of the org.

Decision: Disaster Recovery

High-level disaster recovery (DR) plans should be made before a Citrix environment design is completed. While plans and details can be subject to change, the business's DR needs will impact the overall site design.

The main drivers of DR design are Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO). These objectives govern how fast you need to be able to bring an environment back online in the event of an outage (RTO) and how much data the users can tolerate losing(RPO). These requirements often vary by use case, so it is important to assign tier classifications to the resources in the environment. Generally speaking, the shorter the RTO and RPO, the more expensive the DR solution.

Generally speaking, DR topology has two main options: active/active and active/passive. In an active/active deployment, users use both/all the datacenters actively, and if one site goes down, they failover to another datacenter. In an active/passive deployment, users only access their main datacenter for normal operations and failover to a (normally offline) datacenter in a DR event. For both options, there are considerations for the access, resource, control, and compute layers.

Layer 0: The Business Layer

The business layer describes the reasons for the Citrix environment. This includes business drivers, such as reducing total ownership cost, enabling hybrid work, consolidating vendors, protecting sensitive information such as Intellectual Property from leaving the corporate network, improving security for contractors, enabling you to bring your own device policies, etc. Technical objectives should also be discussed at this stage. Examples of technical objectives include centralizing image management and implementing a specific feature such as App Protection for contractors. Another key technical objective is to identify the pain points in the existing Citrix environment (if one exists) and what improvements need to be made.

In this layer, it is also important to identify future growth opportunities, such as mergers and acquisitions or natural business growth, as these can impact the environment's design and scalability.

All subsequent design decisions should incorporate these business and technical drivers.

Layer 1: The User Layer

The top layer of the design methodology is the user layer, which each unique user group defines.

The user layer appropriately sets the overall direction for each user group’s environment. This layer incorporates the assessment criteria for business priorities and user group requirements to define effective strategies for endpoints and the Citrix Workspace app. These design decisions affect each user group's flexibility and functionality.

Endpoint Devices

There are several endpoint devices that may be used with the Citrix environment, all with differing capabilities, including:

- Tablet-based
- Laptop
- Desktop PC
- Thin client
- Kiosks or shared workstations
- Smartphones

The Citrix environment and client apps must be designed with endpoints in mind. End users may be using multiple endpoint devices and different types, and endpoints often vary by user group and workload.

Decision: Endpoint Ownership

In many organizations, endpoint devices are corporate-owned and managed. However, many organizations have bring-your-own-device (BYOD) programs to improve employee satisfaction, reduce costs, and simplify device management. Even if BYOD is a business priority, it does not mean that every user should be allowed to use a personal device in the corporate environment.

Certain user requirements identified during user segmentation can greatly impact the suitability of personal devices:

- **Security** – Users requiring a high level of security might be unable to bring a personal device into the secured environment to risk data theft.
- **Unmanaged devices** -
    - **Bring your own devices (BYOD)** - Companies often allow employees to use their own devices to access work resources. This reduces the management overhead of procuring and managing the endpoints. However, it is recommended to issue minimum requirements when using a BYOD program to reduce the number of issues potentially encountered by end users (for example, minimum CWA version, OS versions, and/or browser versions for accessing the Citrix environment)
    - **Third-party users** — Often, users accessing the environment who are not directly employed by the company are not directly employed by the company. These users can include third-party contractors or outsourced employees using computers managed by their companies. This reduces the company's control over these users' devices. Similar to BYOD, it is also recommended that these use cases issue minimum requirements to reduce support tickets.

Decision: Endpoint Lifecycle

Organizations may repurpose devices to extend refresh cycles or provide overflow capacity for contract workers. Many endpoints can offer more capabilities, giving them longer useful lifespans. In many cases, these hardware capabilities vastly outstrip the needs of a typical user. When coupled with the ability to virtualize application and desktop workloads, this provides new options to administrators, such as repurposing existing workstations. These options go well beyond the simple three-year PC refresh cycle. However, the benefits of repurposing or reallocating a workstation should be balanced against the following considerations.

- **Minimum standards** – While cost factors of repurposing existing workstations may be compelling, certain minimum standards should be met to guarantee a good user experience. At a minimum, it is recommended that repurposed workstations have a 2GHz processor, 4 GB of RAM, and 30 GB of free disk space.
- **Business drivers** — Priorities underpin the success of any major project. Organizations that have prioritized reducing capital expenditure by prolonging the hardware refresh cycle can benefit from repurposing hardware. Conversely, if an organization's business drivers include reducing power consumption as part of an overall sustainability initiative, purchasing newer endpoints may be beneficial to take advantage of the latest power management capabilities available in the most modern devices.
- **Workload** – The type of work and VDI model for an end user can determine whether they are a good candidate for a repurposed endpoint or may be better served with a new device. Suppose the work performed by the individual involves locally installed applications. In that case, the individual may be best served by a new endpoint that offers the most powerful and recently updated processor and graphics architecture. However, if a user is largely performing tasks associated with virtualized applications that do not involve the latest multimedia capabilities, such as webcams, VoIP, and media redirection, then a repurposed workstation should be a viable alternative.

This planning matrix outlines the considerations when repurposing existing hardware:

Endpoint Provisioning Criteria	Repurpose Existing	Procure New
Capital restrained environment	X	
High number of virtualized applications	X	
Desire to prolong hardware refresh cycle	X	
High failure rate among existing desktops		X
Outmoded client-side feature set		X
Power consumption or sustainability initiative		X

Decision: Unified Endpoint Management (UEM)

VDI allows users to work on any device from any location while still getting access to their apps and data. With distributed users accessing the environment across multiple devices, including mobile devices, administrators need to be able to secure and support the mobile devices centrally.

Administrators need to:

- Selectively wipe a device if lost, stolen, or non-compliant.
- Require passcode security standards.
- Define WiFi parameters for office locations.
- Integrate certificates to secure communications.
- Configure conditional access based on security guidelines.

Managing the distributed endpoints is only part of the challenge. Administrators need to define levels of access. Administrators need to secure and control access to the apps and data. Security becomes more concerned when users access corporate virtual apps and desktop resources from personal devices. A few things to consider when delivering Citrix Virtual Apps and Desktop apps to mobile devices:

- What resources can a jailbroken device access?
- Can users copy/paste between personal apps and Citrix Virtual Apps and Desktops apps?
- Can a device with no configured passcode get access to corporate resources?
- Can users continue to use a native or untrusted third-party email client?
- Can mobile device users access Intranet sites with a browser optimized for mobile devices or a secure enterprise browser?

Unified Endpoint Management (UEM) solutions, like Citrix Endpoint Management and XenMobile, protect app data and let admins control app data sharing. UEM also allows for managing corporate data and resources separately from personal data.

Decision: Mobile Device Management (MDM)

VDI allows users to work on any device from any location while still getting access to their apps and data. With distributed users accessing the environment across multiple devices, including mobile devices, administrators need to be able to centrally secure and support the mobile devices, known as Mobile Device Management (MDM).

MDM solutions, like Citrix Endpoint Management and XenMobile, enable organizations to protect devices and data on devices at a system level. For example:

- Selectively wipe a device if the device is lost, stolen, or out of compliance.
- Require passcode security standards
- Define geo-location device restrictions
- Define Wi-Fi parameters for office locations
- Integrate certificates to secure communications

MDM is typically suitable for corporate-owned mobile devices because most users with personal devices do not want to give the IT team that much control over their personal devices.

Decision: Mobile Application Management (MAM)

With a distributed workforce accessing the Citrix Virtual Apps and Desktops environment across numerous devices, administrators must secure and control access to the apps and data. Security becomes a greater concern when users access corporate resources from personal devices.

A few things to consider when delivering Citrix Virtual Apps and Desktops apps to mobile devices:

- What resources can a jailbroken device access?
- Can users copy/paste between personal apps and Citrix Virtual Apps and Desktops apps?
- Can a device with no configured passcode get access to corporate resources?
- Can users continue to use a native or untrusted third-party email client?
- Can mobile device users access Intranet sites with a browser optimized for mobile devices or with a published desktop browser?

MAM solutions, like Citrix Endpoint Management and XenMobile, protect app data and let admins control app data sharing. MAM also allows for managing corporate data and resources separately from personal data.

MAM is often suitable for bring-your-own (BYO) devices because, although the device is unmanaged, corporate data remains protected. MAM can also be used for the following scenarios:

- Selectively wipe a device if the device is lost, stolen, or out of compliance without impacting the personal data
- Protect corporate app data with app data sharing controls.
- Enable intranet site access seamlessly with per-app VPN functionality (Citrix microVPN) to Citrix mobile productivity and in-house apps
- Block access to MAM-enabled apps if a device is rooted or jailbroken

Decision: Endpoint Form Factor

The endpoints' capabilities and efficiencies offered in thin client form factors have grown. Even mid-range thin clients now have graphics capabilities that utilize HDX features such as multi-monitor support and Microsoft Teams Optimization while offering management and power

efficiency benefits. It is important to note that thin clients support different workload levels (basic use cases vs graphics-intensive workloads). Be sure to discuss the requirements with your thin client vendor of choice.

Most organizations will likely deploy a mixture of fully featured and thin clients. However, as explained in this table, certain endpoint devices are more appropriate when combined with certain VDI models.

VDI Model	Thin Client	Desktop PC	Laptop	Tablet	Smartphone
Hosted Windows Apps	✓	✓	✓	✓	✓
Hosted Shared Desktop	✓	✓	✓	○	○
Hosted Pooled Desktop	✓	✓	✓	○	○
Hosted Static Desktop	✓	✓	✓	○	○
Hosted 3DPro Graphics Desktop	✓	✓	○	X	X
Remote PC Access	X	✓	✓	○	○

✓: Recommended / X: Not Recommended / ○: Viable

Citrix Workspace app Selection

The Citrix Workspace app is an easy-to-install software client that provides easy and secure access to applications and desktops from any device, including smartphones, tablets, PCs, Macs, Linux, and macOS.

This section provides a series of design decisions to consider when deploying the Citrix Workspace app.

Decision: Citrix Workspace app Type

Citrix Workspace app is available as both a Long Term Service Release (LTSR) or Current Release (CR) for Windows and CR for other platforms (including Mac, Linux, ChromeOS, HTML5, Android, and iOS). The release timelines for each platform vary and should be considered when deciding which version of the Citrix Workspace app to deploy. It is important to recognize that there are certain differences between editions, and each user group may require a different Workspace app depending on their need. For the latest feature matrix, please refer here.

Decision: Initial Deployment

Several deployment options are available to deliver the Citrix Workspace app to an endpoint. Although it is usually a best practice to have a full version of the Citrix Workspace app deployed to an endpoint to provide the greatest level of functionality, it is important to consider fallback options such as the Citrix Workspace app for HTML5 for those situations where the installation of Citrix Workspace app is not possible. Note that although the Citrix Workspace app for HTML5 can be used as a fallback option, it is not recommended for general use due to the limited feature set and common browser restrictions around unsecured WebSockets connections (see CTX134123 for more information).

These are some of the mechanisms commonly used to deploy and update the Citrix Workspace app:

- **StoreFront** – If Citrix StoreFront is available, administrators can deploy the Citrix Workspace app via a Receiver for Web site by enabling the "Client Detection" feature. When deployed, a Receiver for Web site enables users to access StoreFront stores through a web page. If the Receiver for Web site detects that a user does not have a compatible version of the Citrix Workspace app, the user is prompted to download and install the Citrix Workspace app. The Workspace app clients can be hosted on the StoreFront server. In this case, admins are responsible for keeping the versions on the StoreFront server up to date. Users can decline the upgrade. Alternatively, users can visit citrix.com/downloads for the latest Workspace app files.
- **Citrix Workspace** – If using Citrix Workspace, you can configure the service to prompt users to install the latest version of the Citrix Workspace app if a local client is not detected. Users can decline the upgrade.
- **Internal download site** – Users may be prevented from downloading software from the Internet, even with permission to install applications. Administrators can create an internal website for the supported Citrix Workspace app version or host the installation media on a common software distribution point for a more seamless user experience. This could be an alternative to enabling Client Detection on the StoreFront Receiver for Web site, which can result in an inconsistent user experience depending on the browser's ActiveX settings.
- **Markets and stores** – The Citrix Workspace app is available in Windows, Android, and iOS stores.
- **Enterprise software deployment** – Many organizations employ an enterprise software deployment (ESD) or Mobile Application Management (MAM) solution. ESD/MAM solutions can be used to deploy Citrix Workspace App to managed endpoint devices. Employee-owned devices can only be managed if the user successfully registers the device with the management tool.
- **Master image —** Most organizations deploy a group of master desktop images to each business-owned desktop, laptop, server, or virtual desktop. A common mechanism to ensure access to virtual desktops and applications is to include a supported version of the Citrix Workspace app in the master image. Subsequent Citrix Workspace app updates are handled manually or by enterprise software deployment tools.
- Group policy — For customers without a robust ESD solution, deploying and configuring the Citrix Workspace app via Microsoft Group Policy is possible. Sample start-up scripts for deploying the app are available in the product documentation. It should be noted that this method is limited to domain-joined or hybrid Azure AD-joined endpoints only.
- Manual install – All supported versions of the Citrix Workspace app are available from the Citrix Workspace app download site. Upon landing on this site, client detection is performed, and a platform and operating system-specific link is provided to allow users to download an appropriate edition of the Citrix Workspace app. It is important to note that no configuration will be accomplished via this method alone, so users will receive the first-time use prompt to enter a server URL or email address. This option is likely to be utilized with unmanaged devices.
- Citrix Global App Configuration service – Citrix Global App Configuration service (GACS) is a Citrix service used to configure, manage, and distribute Citrix Workspace app (CWA) or client app-specific end-user settings across all device OS. You can use the Citrix Workspace App Version setting to specify which Citrix Workspace app version your end users must use for optimal results. You can set up a rule that updates the app to the latest CR (Current Release) or LTSR (Long Term Service Release) version. You can also specify if the upgrade occurs automatically or if the end user can update the app

manually. Global App Configuration Service can be used on both managed and unmanaged devices.

This table describes these methods.

Deployment Options	Thin Clients	Desktop PC	Laptop	Tablet	Smartphone	Unmanaged
Base Image	✓	✓	✓	X	X	X
ESD	✓	✓	✓	X	X	X
MAM	X	✓	✓	✓	✓	X
Group Policy	X	✓	✓	X	X	X
Citrix Download site	X	✓	✓	X	X	✓
Internal Download site	X	✓	✓	X	X	X
App Store	X	✓	✓	✓	✓	✓
Global App Configuration Service	X	✓	✓	✓	✓	✓

✓: Recommended / X: Not Recommended

Decision: Initial Configuration

Citrix Workspace app must be configured to provide access to enterprise resources. The configuration method varies by Citrix Workspace app version, the form factor of the device, and lastly, the access method involved. Several methods may be viable for an organization. The method utilized is contingent on the resources (people, systems, time) available and larger organizational initiatives such as BYOD programs.

These methods can be used to configure the Citrix Workspace app:

- **Manually —** If allowed, it is usually possible to manually configure the Citrix Workspace app by entering the server URL. This method should be reserved for administrators or users with advanced knowledge.
- **Provisioning file –** For environments running StoreFront, providing users with a provisioning file containing store information is possible. Provisioning files are exported from the StoreFront console. The file is saved with a "*.cr" extension and can then be placed on a shared network resource, a Receiver for Web site, or another web-based resource or emailed to users. The file can then be launched from an endpoint, automatically configuring the Citrix Workspace app to use the store(s). If users browse the Receiver for Web site and select the "Activate" option under their username, this also automatically downloads this same ".cr" file and configures the Workspace app client for users.
- **Group policy –** Microsoft Group Policy can be used to configure the Citrix Workspace app. This can be done via start-up scripts used to deploy the Workspace app by ensuring there is a value for the SERVER_LOCATION=Server_URL parameter or by using the ADMX/ADML template files included with the installation of Citrix Workspace app to set the StoreFront Account List option in conjunction with another Workspace app

deployment method. Provide the URL of the server running Citrix StoreFront or Citrix Workspace service in the format https://baseurl/Citrix/storename/discovery. It should be noted that this method is limited to domain-joined or hybrid-Azure AD-joined endpoints only.

- **Email-based discovery** — The Citrix Workspace app can be configured by entering an email address. Based on Domain Name System (DNS) Service (SRV) records, the app determines the Citrix Gateway and StoreFront Server associated with the email address. The Citrix Workspace service can also use email-based discovery. The app then prompts users to log on to access virtual desktops and applications.
- **Global App Configuration service** - The Global App Configuration service provides a centralized setup for IT admins to easily configure Citrix Workspace app settings on Windows, Mac, Android, iOS, HTML5, and Chrome OS platforms. Each platform has different settings, such as Group policies, Registry, Default.ica, SF settings, and config files, and there are many ways to configure these settings. The Global App Configuration service allows you to push the settings to end-users from one centralized interface. Global App Configuration service can be used for StoreFront and Citrix Workspace URLs.

Decision: Updates

The Citrix Workspace app is in active development. As such, periodic updates are released to enhance functionality or address user issues. As with any actively developed product, the latest version of these products should be deployed to the endpoints so that users benefit from the latest functionality and maintain compliance with product support lifecycles. The Citrix Workspace app is released in two cadences: Current Release (CR) and Long Term Service Release (LTSR). LTSR releases are maintained for a longer time period but do not receive feature updates. CRs receive regular feature updates. Multiple methods are available to update the Citrix Workspace app and, if applicable, associated plug-ins.

- **Automatic Update** — The Citrix Workspace app includes an auto-update capability that automatically checks for newer versions. The auto-update service can be configured to allow users to defer updates and skip any updates that are not Long Term Service Release (LTSR) versions. Auto-update is unavailable for versions before Citrix Workspace app 2104 and Citrix Workspace app 1912 LTSR CU4.
- **Enterprise software deployment** — ESD tools give an organization direct control over the time/frequency of Workspace app updates to managed devices. However, additional thought must be given to updating unmanaged devices.
- **Manual updates** — Manual methods can be used to update the Citrix Workspace app when no automated solution is available. Whether deployed on Receiver for Web site, StoreFront, an internal Citrix Workspace app site, or an external site, these options require user involvement in updating the app. Due to the involved nature of manual updates and the opportunity for a user mistake, this option should only be considered a last resort.
- **StoreFront/Workspace deployment** - Administrators can manage Workspace app updates via either StoreFront or Citrix Workspace service.
    - **StoreFront** - If a locally deployed Citrix Workspace app cannot be detected, the user is prompted to download and install it. The default download location is the Citrix website, but you can also host the installers on the StoreFront server or elsewhere, as mentioned above.
    - **Workspace** — In the Workspace service, you can configure end users to be prompted with the latest release of the Workspace app if they do not already have it installed.

- **Global App Configuration Service —** Administrators can now manage the auto-update version for the organization's devices by setting the version in the maximumAllowedVersion property in the Global App Config Service. When the version is set, the Citrix Workspace app on the user's device automatically prompts the user to upgrade to the version specified in the maximumAllowedVersion property.

Layer 2: The Access Layer

The second layer of the design methodology is the access layer, which is defined for each user group.

Creating an appropriate design for the access layer is an important part of the virtualization process. This layer handles user validation through authentication and orchestrates access to all components necessary to establish a secure connection.

The access layer design decisions are based on each user group's mobility requirements and the endpoint devices used.

Authentication

Getting access to resources is based on the user's identity. Defining the authentication strategy considers the user's entry point into the environment and how the user will authenticate.

Decision: Authentication Provider

Previously, accessing Citrix Virtual Apps and Desktops required users to have an Active Directory username and password, typically managed within an on-premises Active Directory environment. This setup worked well for organizations with internal users.

However, with the increasing use of external contractors and affiliates, organizations seek more flexible authentication options. Many are turning to third-party identity providers (IdPs) like Microsoft EntraID, Google, or Okta to manage user access instead of maintaining their own user accounts.

Citrix has introduced the Federated Authentication Service to accommodate this shift, which allows integration with third-party IdPs. This simplifies the onboarding process and provides greater flexibility in managing user identities.

Decision: Authentication Point

Before a user connects to a virtual workload, they must successfully authenticate. The location of authentication is typically determined by the mobility requirements of the user groups, which are established during the user segmentation process. In Citrix Virtual Apps and Desktops, there are two primary authentication points:

- **Citrix StoreFront** — StoreFront provides authentication and delivers resources through the Citrix Workspace app, offering centralized enterprise stores for desktops, applications, and other resources. StoreFront authenticates users who can directly access the StoreFront URL, typically those on the corporate network.
- **NetScaler Gateway** – NetScaler Gateway is an appliance that facilitates secure application access and provides granular application-level policy controls. It enables users to access applications and data securely from anywhere, ensuring remote access for users outside the corporate network.

For Citrix DaaS, there is an additional authentication point option along with the two authentication points mentioned above:

- **Citrix Workspace –** Citrix Workspace aggregates and integrates Citrix Cloud services, enabling unified access to all the resources available to your end-users. Users authenticate to their resources using the primary identity provider configured for Workspace.

This table lists preferred authentication points according to user group mobility requirements.

User Group's Mobility Requirement (Network location)	Preferred Authentication Point
Local (internal)	Storefront*
Roaming or mobile users (internal)	Storefront*
Remote (external)	NetScaler Gateway

Note:

It is possible to route internal users through a NetScaler Gateway. The benefits of doing so include gaining HDX insights on internal user connections and HDX optimal routing.

Authentication for user groups requiring remote or mobile access may occur directly on StoreFront in certain scenarios. For instance, security policies in the DMZ might restrict access from the NetScaler Gateway to the domain, preventing support for Smart Card client certificate authentication. In such cases, authentication via StoreFront can be facilitated through a NetScaler SSL_BRIDGE virtual server, facilitating HTTPS traffic.

Typically, this virtual server would be deployed alongside a NetScaler Gateway on the same NetScaler appliance configured to provide HDX Proxy access to the virtual desktop environment. While this approach may sometimes be necessary, it's generally recommended that external users authenticate via NetScaler Gateway as a best practice.

Decision: Authentication Policy

Once the authentication point has been identified, the type of authentication must be determined. The primary methods available are:

- **StoreFront** supports several authentication methods, although not all are recommended depending on the user access method, security requirements, and network location. By default, StoreFront authenticates users directly with Active Directory. StoreFront can be optionally configured to delegate authentication to XML if required (such as if the StoreFront servers are in a domain that does not trust the user domains).
    - **Username and password** – Requires users to login directly to the site by entering a username and password.
    - **Domain pass-through** – Allows pass-through of domain credentials from users' devices. Users authenticate to their domain-joined Windows computers and are automatically logged on when they access their stores.
    - **NetScaler Gateway pass-through** — This option allows pass-through authentication from NetScaler Gateway. Users authenticate to NetScaler Gateway and are automatically logged on when they access their stores.
    - **Smart card** – Allows users to authenticate using smart cards and PINs through the Citrix Workspace app for Windows and NetScaler Gateway. To enable smart

card authentication, user accounts must be configured either within the Microsoft Active Directory domain containing the StoreFront servers or within a domain with a direct two-way trust relationship with the StoreFront server domain. Multi-forest deployments involving two-way trust are supported.

- **Anonymous** — Allow users to access applications and desktops without presenting credentials to StoreFront or the Citrix Workspace app. Local anonymous accounts are created on demand on the Server VDA when sessions are launched. This requires a StoreFront store configured for authenticated access, a Server OS-based VDA, and a Delivery Group configured for unauthenticated users.

> Note:
>
> Enabling anonymous access on the IIS hosting Citrix StoreFront is not advisable. With anonymous authentication, users can access resources without providing credentials, which risks exposing sensitive information to unauthorized individuals. This may result in unauthorized access, data breaches, and compromise of the Citrix environment.

- **Security Assertion Markup Language (SAML)** - Users can authenticate with their identity provider (IdP) credentials. SAML authentication provides seamless single sign-on (SSO) experiences for users, allowing them to access StoreFront stores and applications without entering separate credentials. This method is particularly beneficial for organizations that use identity federation with external partners or cloud-based services. To enable SAML authentication, StoreFront must be configured to trust the IdP's SAML assertions and properly map attributes to user accounts. The Citrix Workspace app and NetScaler Gateway can also be configured to support SAML authentication, ensuring consistent authentication experiences across different access points in the Citrix environment.

- **NetScaler Gateway** - The NetScaler Gateway supports several authentication methods. The list below includes those primarily used in virtual desktop environments. Each may be used individually but is often combined to provide multi-factor authentication.
  - **LDAPS** – The lightweight directory access protocol over SSL (LDAPS) is used to access directory information services such as Microsoft Active Directory. NetScaler Gateway uses LDAPS to authenticate users and extract their group membership information.
  - **RADIUS (token)** – Remote authentication dial-in user service (RADIUS) is a UDP-based network security protocol that provides authentication, authorization, and accounting. A network access server (NetScaler Gateway in this case) forwards credentials to a RADIUS server that can either check them locally or against a directory service. The RADIUS server could then accept the connection, reject the connection, or challenge and request a second form of authentication, such as a token.
  - **OAuth/OIDC** — OAuth is a standard for providing users access to resources without exposing credentials. OpenID Connect (OIDC) is an authentication standard built on top of OAuth. NetScaler supports OAuth/OIDC.
  - **SAML** - Security Assertion Markup Language (SAML) is an XML-based standard for exchanging authentication and authorization between Identity Providers (IdP) and Service Providers. NetScaler Gateway supports SAML authentication.
  - **TACACS+** - You can configure a TACACS+ server for authentication. Like RADIUS authentication, TACACS+ uses a secret key, an IP address, and a port number.
  - **Client certificate** – Users logging on to a NetScaler Gateway virtual server can also be authenticated based on a client certificate's attributes. Client

certificates are usually disseminated to users through smartcards or common access cards (CACs) that are read by a reader attached to each user's device.
    - **nFactor -** NetScaler Gateway supports nFactor authentication, which allows administrators to configure flexible authentication policies based on various factors, such as user group, device type, and location. nFactor authentication provides enhanced security by combining different authentication methods in a single policy, allowing organizations to implement multi-factor authentication.
- **Citrix Workspace** – Citrix Workspace supports several authentication methods, which include the following:
    - **Active Directory** – Citrix Cloud supports using on-premises Active Directory (AD) to authenticate end users. End users will use their AD credentials to authenticate.
    - **Active Directory plus token** — Citrix Cloud supports using a token as a second authentication factor for users signing in with AD. Users can generate tokens using any application that follows the time-based one-time password standard.
    - **Microsoft Entra ID** (formerly Azure Active Directory (AAD)) – Citrix Cloud supports using Microsoft Entra ID. Using Entra ID with Citrix Cloud, you can leverage your AD and security policies, configure MFA authentication, use a branded sign-in page, and use federation to an identity provider of your choice.

> Note:
>
> You need to use the Citrix Federated Authentication Service with Citrix Cloud to enable single sign-on and prevent a second login prompt at the VDA

    - **Citrix Gateway** — Citrix Cloud supports using an on-premises Citrix Gateway as an identity provider to authenticate subscribers. The authentication options for an on-premises Gateway are in the above section.
    - **Google Cloud Identity** – Citrix Cloud supports using Google Cloud Identity as an identity provider to authenticate subscribers signing in to their workspaces. By connecting your organization's Google account to Citrix Cloud, you can provide a unified sign-in experience for accessing Citrix Workspace and Google resources.
    - **Okta** – Citrix Cloud supports using Okta as an identity provider to authenticate subscribers signing in to their workspaces. By connecting your Okta organization to Citrix Cloud, you can provide a common sign-in experience for your subscribers to access resources in Citrix Workspace.
        - Subscribers have a different sign-in experience after enabling Okta authentication in Workspace Configuration. Selecting Okta authentication provides federated sign-in, not single sign-on. Subscribers sign in to workspaces from an Okta sign-in page but may have to authenticate again when opening an app or desktop from Citrix DaaS. You must use the Citrix Federated Authentication Service with Citrix Cloud to enable single sign-on and prevent a second login prompt.
    - **SAML 2.0** – Citrix Cloud supports using SAML (Security Assertion Markup Language) as an identity provider to authenticate Citrix Cloud administrators and subscribers signing in to their workspaces. You can use the SAML 2.0 provider of your choice with your on-premises Active Directory (AD).

> Note:
>
> You must use the Citrix Federated Authentication Service with Citrix Cloud to enable single sign-on and prevent a second login prompt at the VDA.

- **Adaptive Authentication service** – Adaptive Authentication is a Citrix Cloud service that enables advanced authentication (nFactor) for customers and users logging in to Citrix Workspace without needing an on-premises Gateway. The Adaptive Authentication service verifies the user identity and authorization levels based on location, device status, and end-user context. Using these factors, the Adaptive Authentication service intelligently chooses the appropriate authentication methods and enables access to authorized resources.

The authentication type for a user group is often determined based on security requirements and the authentication point used. Prioritizing the highest security standards is non-negotiable. Choosing a less secure option is risky and increases the environment's vulnerability.

Business Use Cases

- **Retail** — A small private retail company provides virtual desktop users access to non-sensitive data such as marketing catalogs and email. They are not required to adhere to security regulations such as Sarbanes Oxley. Therefore, LDAP authentication based on username and password has been implemented.
- **Financial** — A medium financial enterprise provides its virtual desktop users with access to confidential data, such as banking transaction records. These enterprises are governed by security regulations such as the Statement on Accounting Standards (SAS) 70 and must utilize multi-factor authentication for remote access users. LDAP authentication based on username and password has been implemented, along with SAML authentication.
- **Government** — A large federal institution provides virtual desktop users access to highly confidential data, such as private citizens' personal records. These users are subject to regulations by the Department of Defense (DOD) security standards. LDAP authentication has been implemented based on username and password, along with Client Certificate authentication using CAC cards.
- **Healthcare** - A hospital utilizes Citrix to deliver its healthcare Electronic Medical Record (EMR) application to users. Doctors and nurses use thin client devices stationed at workstations and mobile carts known as Workstations on Wheels (WOW) to access and manage patient data. Per HIPAA regulations, healthcare workers must authenticate their identity before accessing patient information. Imprivata offers a solution to streamline the authentication process and ensure compliance with data security regulations.

Decision: Multi-factor Authentication

Multi-factor authentication (MFA) adds an extra layer of security to your accounts. Instead of just requiring user credentials to gain access, MFA adds a second form of verification, like a phone push notification, one-time code, or biometric scan, thus making it harder for unauthorized users to access accounts.

MFA is strongly recommended if the environment involves publicly facing URLs like NetScaler Gateway or Citrix Workspace URLs. These sites are accessible online, so anyone can view the access point and try to force login. While internal-only sites are less vulnerable as they require the end user to be on the corporate network, MFA is still recommended due to the above mentioned benefits.

It is important to note that if a SAML-based MFA is used, the environment will require Citrix Federated Authentication Service.

StoreFront

Citrix StoreFront authenticates users to Citrix Virtual Apps and Desktop resources. StoreFront enumerates and aggregates available desktops and applications into a single interface that users access through the Citrix Workspace app or browser from their endpoint devices.

Decision: Store

The number of StoreFront Stores may vary depending on the use case. A single Store can be configured for both internal and external access. However, two Stores are recommended if there is a need to separate internal and external access. Additionally, the Stores can be separated based on the resources they enumerate, such as keyword inclusions/exclusions or managing Delivery Controllers (DDCs) or Cloud Connectors (CCs). It's essential to consider three key points when making this design decision:

- Segregating access traffic between Stores (internal vs external)
- Separate authentication use cases
- Default.ICA configurations for a specific use case may warrant a separate StoreFront Store

It's important to note that increasing the number of StoreFront stores can significantly impact storage requirements due to increased IIS-related logging.

Decision: High Availability

If the server hosting StoreFront is unavailable, users cannot launch new virtual desktops, published applications, or manage their subscriptions. Therefore, at least two StoreFront servers should be deployed to prevent this component from becoming a single point of failure. By implementing a load-balancing solution, users will not experience an interruption in their service. Options include:

- **Recommended: Hardware load balancing —** An intelligent appliance that can verify the availability of the StoreFront service and actively load balance user requests appropriately. Citrix + NetScaler is a great example of a hardware load balancer. It is ideal for this purpose and comes pre-configured with StoreFront health checks.
    - For the StoreFront load balancing VIP, it is recommended that the load balancing method be set to 'LeastConnection' and the Persistence setting to 'SourceIP.'
- **DNS round robin** – Provides rudimentary load balancing across multiple servers without performing any checks on availability. If a StoreFront server becomes unavailable, DNS round robin would still route users to the failed server. Because of this, Citrix does not recommend DNS round-robin.
- **Windows network load balancing** — A Windows service capable of performing rudimentary checks to verify the server's availability but cannot determine the status of individual services. This can cause users to be forwarded to StoreFront servers, which cannot process new requests. The user would then not be able to access applications or desktops.

Decision: Delivery Controller Reference

StoreFront must be configured with the IP address or DNS name of at least one Delivery Controller in each Citrix Virtual Apps and Desktops site to provide users with desktops and applications. For fault tolerance, multiple controllers should be entered for each site and/or farm specified. By default, StoreFront treats a list of servers in failover order (active/passive). Ensuring that the communication between Citrix StoreFront and Delivery Controller is securely transmitted using TLS 1.3 encryption over HTTPS (port 443) is recommended.

An active distribution of the user load (active/active) is recommended for large deployments or environments with a high logon load. This can be achieved using a load balancer with built-in XML monitors, such as Citrix + NetScaler, or by configuring StoreFront to load balance the list of Controllers instead of treating them as an ordered list. For the DDC load balancing VIP, it is recommended that the load balancing method be set to 'LeastConnection.'

Decision: Beacons

Citrix Workspace app uses beacons (websites) to identify whether a user is connected to an internal or external network. Internal users are connected directly to StoreFront for authentication, while external users are connected via Citrix NetScaler Gateway. It is possible to control what users see by restricting applications based on which beacon they can access.

The internal beacon should be a site that is not resolvable externally. By default, the internal beacon is the StoreFront base URL. This must be adjusted if the same external and internal URL is configured. The external beacon can be any external site that produces an HTTP response. Citrix Workspace app continuously monitors the status of network connections (for example, link up, link down, or change of the default gateway). When a status change is detected, the Citrix Workspace app first verifies that the internal beacon points can be accessed before checking the accessibility of external beacon points. StoreFront provides the Citrix Workspace app with the beacon points' HTTP(s) addresses during the initial connection/configuration download process and updates as necessary.

Specifying at least two highly available external beacons that can be resolved from public networks is necessary.

Decision: Resource Presentation

By default, StoreFront allows users to choose (subscribe) to the resources they want to use after they log on (favorites) regularly. This approach allows users to restrict the resources they see on their home screen to those they use regularly. The resources chosen by every user for each store are recorded by the subscription store service and stored locally on each StoreFront server (synced automatically between servers in the same server group) so that they can be displayed on the Citrix Workspace app home screen from any device that the user connects from. Although subscriptions are per store and server group by default, administrators can configure two stores within a server group to share a subscription database and/or sync subscriptions between two identically named stores in two separate server groups on a defined schedule if required. This is recommended to sustain the subscription across two server groups for a better user experience.

If favorites are not required or you do not wish to give users the ability to favorite resources, you can make the store mandatory, removing the option to favorite. This can reduce management overhead and the storage needed on the StoreFront servers.

Administrators should determine which applications should always be displayed to users on their home screen or the featured tab. These are generally common applications such as the Microsoft Office Suite and any other applications that every user in an environment may need. StoreFront can filter/present these resources using Keywords defined within the published application properties Description field.

This table explores the Keyword options:

Keyword	Description
Mandatory	Adds an application to the Home tab. Unlike favorites, users cannot remove mandatory applications from the Home tab. Has no effect if favorites are disabled for the store.
Auto	When users log on to the store, the application is automatically favorited and added to their Home tab. Users can unfavorite such applications. This has no effect if favorites are disabled for the store.
TreatAsApp	Apply to desktops to force StoreFront to treat them as apps. The desktop is displayed on the Apps tab rather than the Desktops tab. In addition, it is not automatically started when the user logs on to the store website and is not accessed with the Desktop Viewer, even if the site is configured to do this for other desktops.
prefer="pattern"	Where the application identifies a locally installed application. Applies only to the Citrix Workspace app on Windows. This indicates that the locally installed version of an application should be used in preference to the equivalent delivered instance if both are available.

Keyword	Description
Primary and Secondary	When using Multi-Site Aggregation, the one with the keyword primary specified will always be preferred over the one with the keyword secondary.

Decision: Remote Access

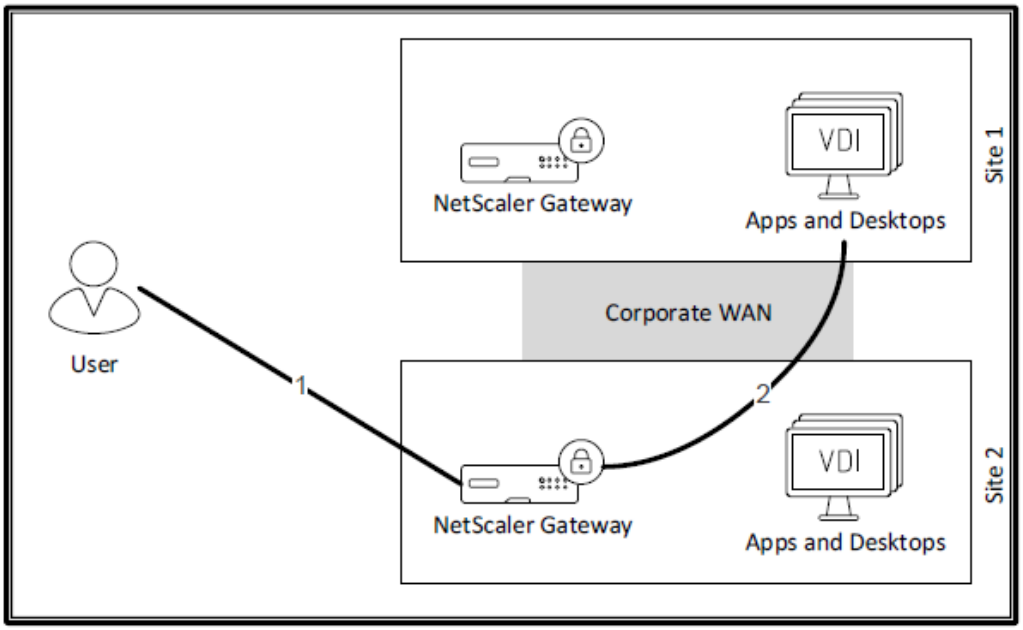
StoreFront stores can be accessible internally or internally *and* externally. Internal-only stores will only be accessible if the end-user device is connected to a corporate network that can resolve the StoreFront URL. If a store has remote access enabled, it can be accessible internally and through a Gateway. In environments where security requires the separation of internal and external traffic, internal and external users can use different stores.

When configuring remote access, you can configure a Callback URL. This is used to verify that requests received from NetScaler Gateway originate from that appliance. A callback URL is only needed if SmartAccess policies or password-less authentication methods (Smart Cards, SAML, and so on) are used. Additionally, it is recommended that you match the method used in the Gateway when configuring the Secure Ticketing Authorities (STAs) (HTTP vs. HTTPS, FQDN vs. IP).

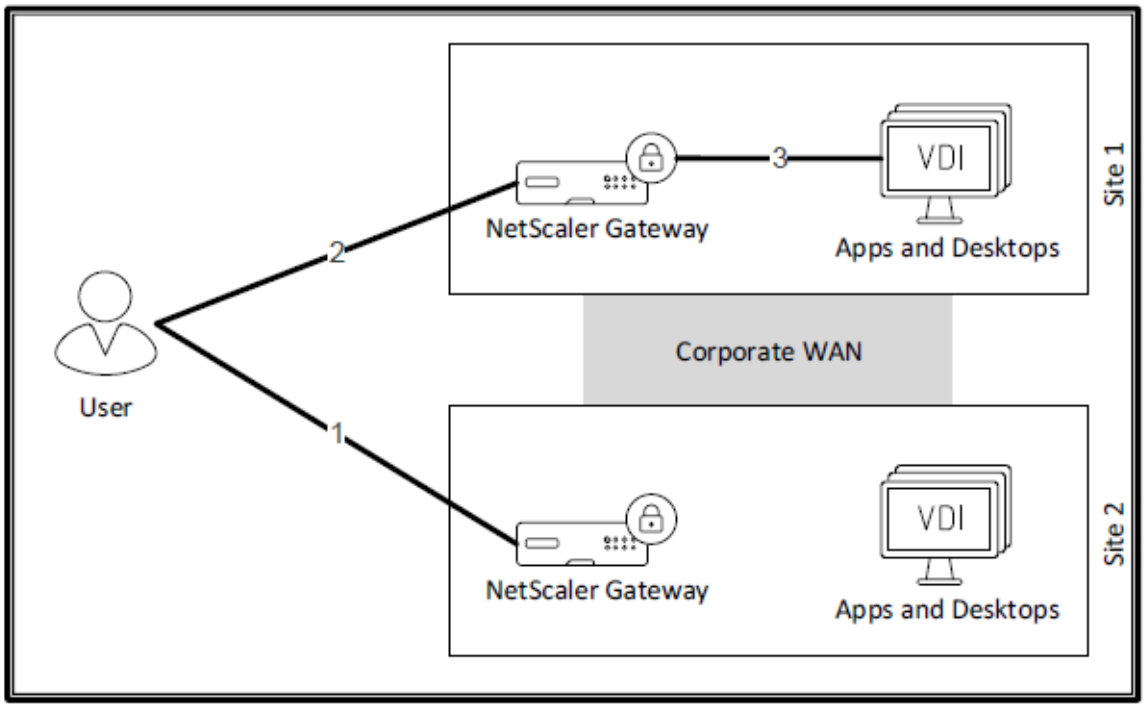
Decision: HDX Optimal Routing

In a multi-site Citrix Virtual Apps and Desktops solution, certain criteria route users to the optimal site, like the fastest response time or closest proximity. These algorithms do not consider the resources a user wants to access.

Improper routing of a user's session results in the following:



1. The user is routed to the most preferred site based on proximity or response time.
2. NetScaler Gateway proxies the HDX traffic to the correct resource, which could be across the corporate WAN. Ideally, optimized routing should resemble the following:

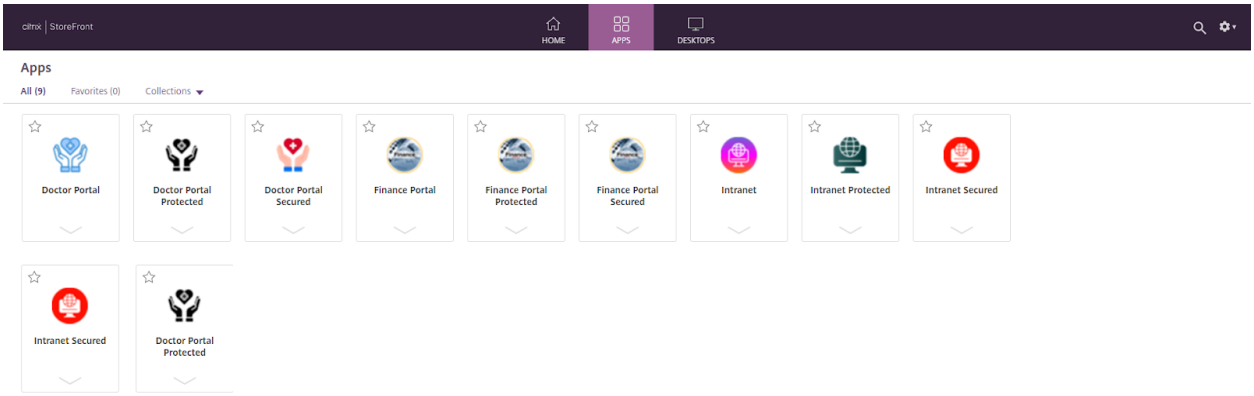


1. User routed to the most preferred site based on proximity or response time
2. Based on the selected resource, NetScaler Gateway reroutes the session to a NetScaler Gateway in the preferred site.
3. NetScaler Gateway proxies the HDX traffic to the correct resource, which stays on the local LAN.

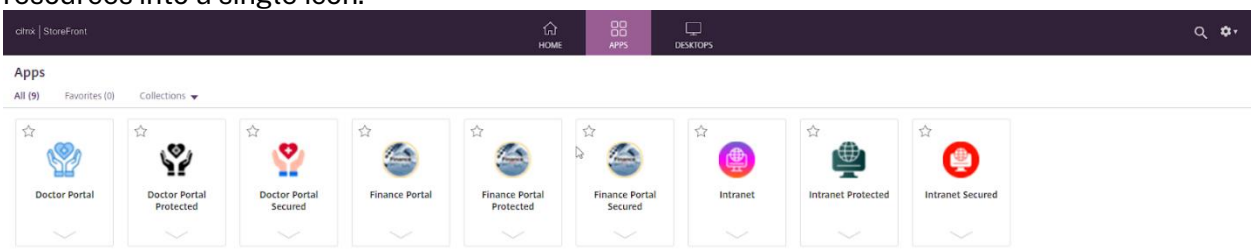
Using the optimized HDX routing option within StoreFront offloads traffic from the corporate WAN and places it on the public network, improving the user experience during their Citrix session.

Decision: Aggregation Groups

If the Citrix Virtual Apps and Desktops solution includes multiple delivery sites, StoreFront combines the available resources so the user has a single interface for all published resources. However, if multiple sites publish the same resources, the user might experience confusion as a single application appears multiple times.



StoreFront aggregation groups define how the resources in multiple sites merge to provide the user with a single, easy-to-understand view. StoreFront aggregates duplicate published resources into a single icon.



The administrator must determine how to load balance users across the different Citrix Virtual Apps and Desktops sites when the icon is an aggregation. The options are:

- Load Balancing – Used when duplicate sites are created based on capacity recommendations. StoreFront distributes user requests across all configured sites.
- Failover – Used when geographies need resources available during an outage or when migrating users from one site to another (like a Citrix migration project).

The “Map users to controllers” settings are commonly referred to as “User Farm Mapping” as they control which Sites a given group of users are allowed to enumerate against, whether those Sites are aggregated or not. This functionality has two primary use cases: limiting enumeration and assigning aggregation settings.

Limiting user mappings to stores with resources reduces the enumeration time by reducing the number of sites StoreFront has to contact. Aggregation settings can also be configured per user mapping, as it may be desired to assign different configurations to different sets of users—such as failover configurations or different combinations of sites.

You can override the specified deployment ordering for individual Citrix Virtual Apps and Desktops resources to define preferred deployments to which users are connected when they access a particular desktop or application. This enables you to, for example, specify that users are preferentially connected to a deployment specifically adapted to deliver a particular desktop or application but use other deployments for other resources. To do this, you can use keywords Primary to describe the desktop or application on the preferred deployment and Secondary to the resource on other deployments.

Documenting the users, stores, and aggregation methods during the design phase is advisable. For example:

User Group	Available Stores	Load Balancing Stores	Failover Stores
NA_FinanceUsers	NA_West_Store, NA_East_Store, EMEA Store	NA_West_Store, NA_East_Store	EMEA_Store
EMEA_SalesUsers	EMEA_Store, NA_East_Store	EMEA_Store	NA_East_Store

Decision: Scalability

The number of Citrix Workspace app users supported by a single StoreFront server depends on the resources assigned and user activity level. Receiver for Web users will consume more RAM on average than native Workspace app users, but a minimum of 8GB plus 30 MB for each store (assuming one website per store) is recommended. For each store with favorites enabled, an additional 5 MB plus 8 MB for each 1000 favorites is recommended. Additionally, more sites/farms enumerated per store will increase CPU utilization and server response time.

Depending on storage needs, a general starting point for StoreFront sizing is 4 vCPU with 8-16 GB of RAM. The CPU may need to be scaled up depending on the total number of connections and the load on the StoreFront servers. Each StoreFront server group should have at least 2 StoreFront servers but no more than 5. Tests have shown diminishing returns after a single StoreFront deployment grows beyond 3-4 StoreFront nodes with a maximum of 5 servers supported in a single server group. StoreFront server groups are not recommended to span data centers.

Citrix Workspace service

Citrix Workspace is the cloud-based access option for Citrix Cloud users. As a cloud service, it is accessible to all users via the Internet. The Workspace configuration has different design considerations.

Decision: Workspace URL

You can enable one URL within the *.cloud.com domain by default. The end user will visit this URL to authenticate, enumerate, and launch. It is also possible to use your own custom domain. You can also configure up to 10 Workspace URLs; however, a custom domain is currently

unsupported for multi-URLs. For multi-Workspace URLs, you can configure different branding, resource filtering, and authentication configurations for the separate URLs.

Decision: ICA proxy

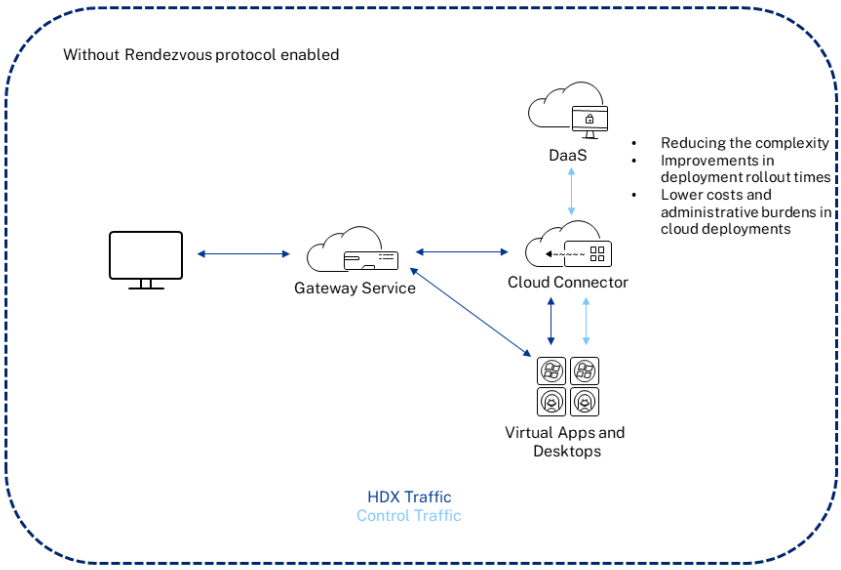
When using Citrix Workspace service, the default ICA proxy is Citrix Gateway Service. This is a Citrix Cloud service with Points of Presence (PoPs) around the globe that provides proxy sessions to end users. It is a turnkey service and does not require configuration. This service enables internal and external users to launch resources.

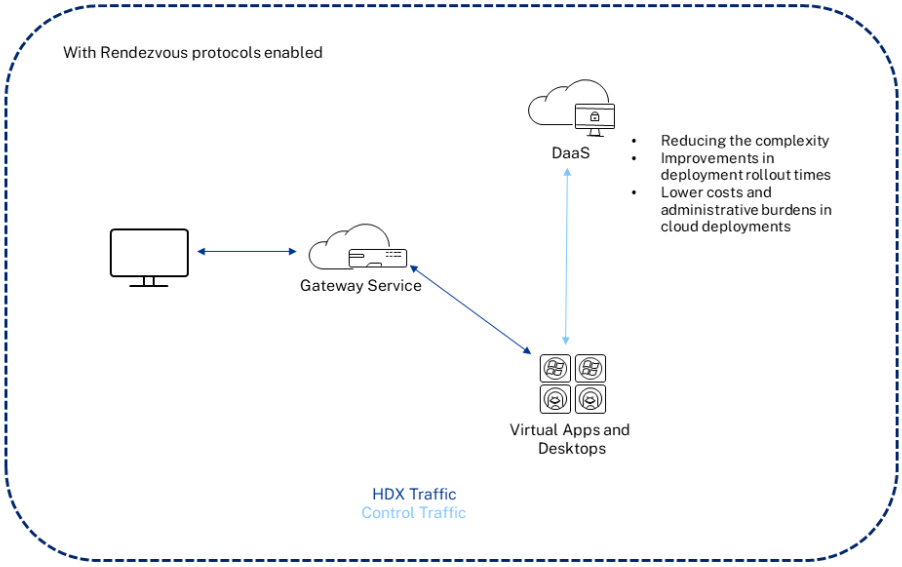
You can also configure Workspace to use an on-premises Gateway to route the ICA traffic if you want to control the routing of the session traffic, like through OGR. A caveat is that neither Service Continuity nor Local Host Cache is supported in this scenario.

Another option is Internal Only. This restricts resource launch only to scenarios where the user's endpoint can directly connect to the VDA, like from the corporate network. It is important to note that end users can still access their Workspace URLs from anywhere, but resource filtering can be used to hide the icons when connecting remotely.

Decision: Connection Routing

By default, all user sessions are routed through the Cloud Connectors via the Gateway service. This can increase the usage of the Cloud Connectors and affect their scaling. To reduce the amount of traffic handled by the Cloud Connectors, it is recommended that you implement the Rendezvous protocol.





If your environment uses VDA versions from before 2203, you can use Rendezvous V1. V1 supports bypassing the Citrix Cloud Connectors for HDX session traffic only. If you use a VDA version after 2203+, you can use Rendezvous V2. V2 supports bypassing the Citrix Cloud Connectors for both control and HDX session traffic, reducing the amount of traffic through the Cloud Connector.

With Direct Workload Connection in Citrix Cloud, you can optimize internal traffic to the resources to make HDX sessions faster. Ordinarily, users on both internal and external networks connect to VDAs through an external Gateway. This Gateway might be on-premises in your organization or provided as a service from Citrix. Direct Workload Connection allows internal users to bypass the Gateway and connect to the VDAs directly, reducing latency for internal network traffic. To set up Direct Workload Connection, you need network locations and public IP addresses corresponding to where clients launch resources in your environment.

Decision: Service Continuity

Service Continuity is a high availability (HA) service for customers using Citrix Workspace and Citrix Gateway service. Service continuity allows users to connect to their DaaS apps and desktops during outages, as long as the device maintains a network connection to a resource location.

Service continuity uses Workspace connection leases to allow users to access apps and desktops during outages. Workspace connection leases are long-lived authorization tokens. Workspace connection lease files are securely cached on the user's device. Workspace connection lease files are signed and encrypted and are associated with the user and the user's device.

Enabling Service Continuity is highly recommended; otherwise, users may be unable to connect during resilience events. When Service Continuity is enabled, a Workspace connection lease allows users to access apps and desktops for seven days by default. You can configure Workspace connection leases to allow access for up to 30 days. If you have users accessing via HTML5, they will need the browser plug-in to use Service Continuity.

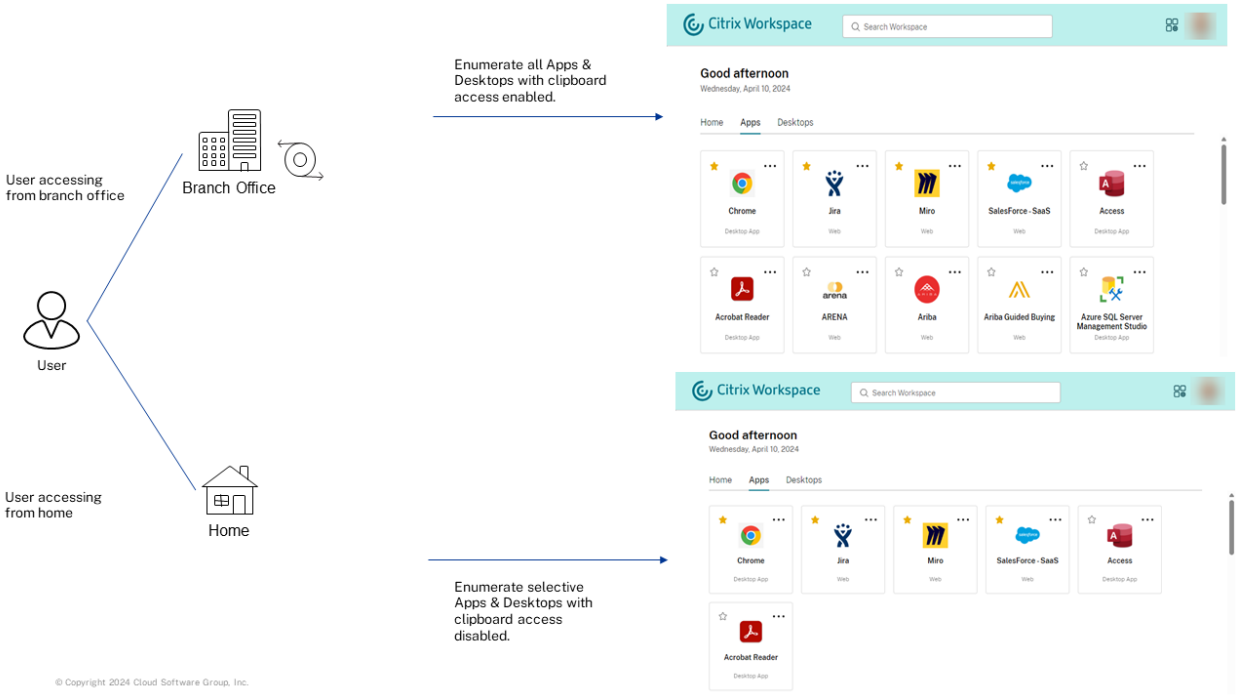
Decision: Resource filtering /network location service

It is possible to filter what resources appear to users based on their Workspace URL and network location.

With the new multiple Workspace URL feature, you can filter and deliver resources based on end users' Workspace URL. To configure an access policy based on Workspace URLs, you need to apply the following SmartAccess filters. The filter values are also sent as SmartAccess tags to the DaaS service.

- **Citrix.Workspace.UsingDomain** allows filtering of delivery group resources by the Workspace URL. The value is the fully qualified domain name of the Workspace URL.
- **Citrix-Via-Workspace** - Indicates that the end user is using the Workspace service rather than an on-premises StoreFront deployment.

The Citrix Workspace Adaptive Access feature uses advanced policy infrastructure to enable access to Citrix DaaS based on the user's network location. The location is defined using the IP address range or subnet addresses. To configure this, you must know the public IP ranges of the various access points.



Admins can define policies to enumerate or not enumerate virtual apps and desktops based on the user's network location. Admins can also control the user actions by enabling or disabling clipboard access, printers, client drive mapping, and so on, based on the user's network location. For example, admins can set up policies such that users accessing the resources from home have limited access to applications and users accessing the resources from branch offices have full access.

Decision: Site Aggregation

Only one Citrix DaaS tenant can be mapped to the Citrix Workspace service. You can also aggregate Citrix Virtual Apps and Desktop sites to display their icons and provide access via

Workspace. However, Service Continuity is not supported for Citrix Virtual Apps and Desktop sites.

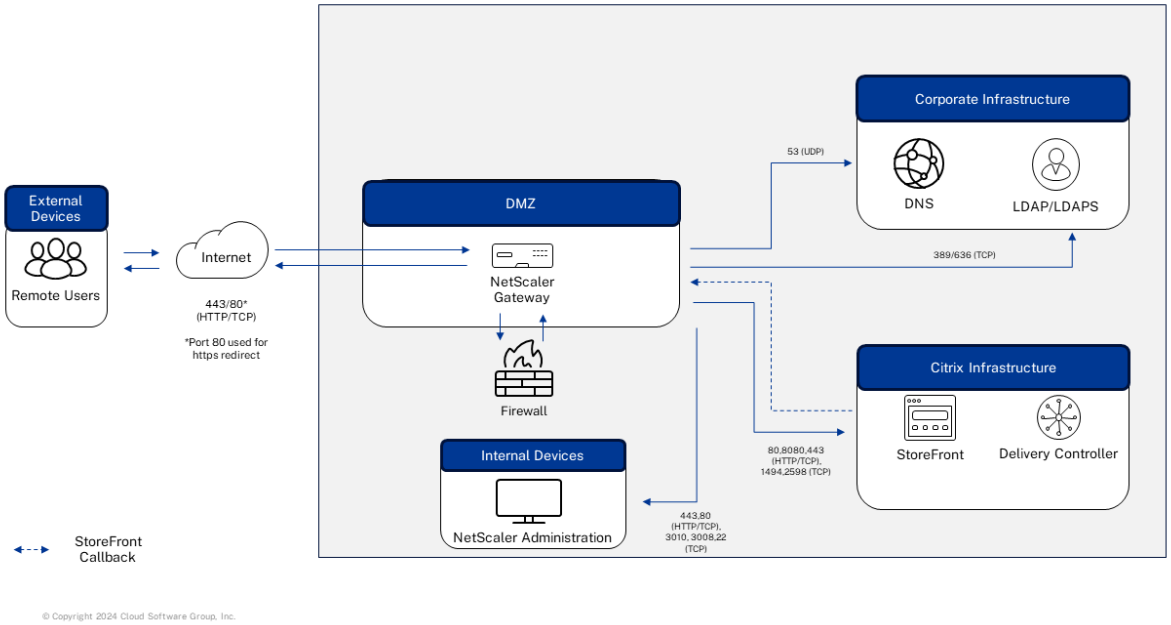
NetScaler Gateway

User groups utilizing NetScaler Gateway as their authentication point have additional design decisions to consider. These design decisions do not apply to non-NetScaler Gateway authentication points.

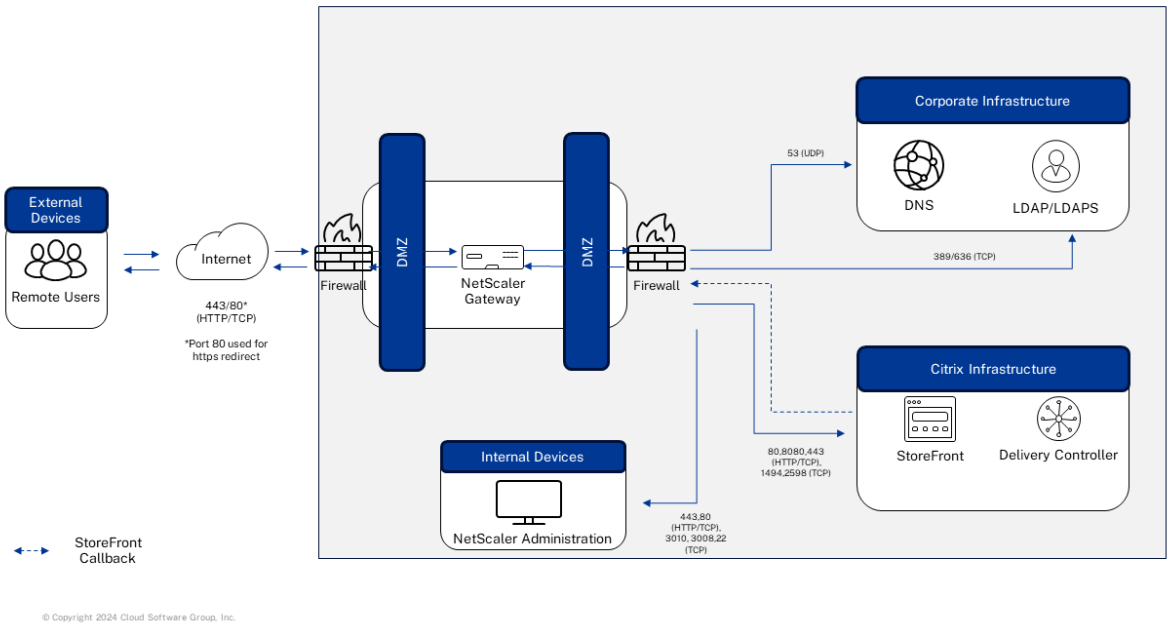
Decision: Topology

Selection of the network topology is central to planning the remote access architecture to ensure it can support the necessary functionality, performance, and security. The design of the remote access architecture should be completed in collaboration with the security team to ensure adherence to corporate security requirements. There are two primary topologies to consider each of which provides increasing levels of security:

- 1-Arm (normal security) — With a 1-arm topology, the NetScaler Gateway utilizes one physical or logical bonded interface, with an associated VLAN and IP subnet to transport both. If increased security is desirable, separate DMZ and intranet traffic VLANs can be configured with the appropriate firewall policies for each network.



- 2-Arm (high security) – With a 2-arm topology, the NetScaler Gateway utilizes two or more physically or logically bonded interfaces with associated VLANS and IP subnets. The frontend traffic for users is transported to one of these interfaces. The frontend traffic is isolated from backend traffic between the virtual desktop infrastructure servers and services, which is directed to a second interface. This allows using separate demilitarized zones (DMZs) to isolate frontend and backend traffic flows along with granular firewall control and monitoring.



Decision: High Availability

If the NetScaler Gateway is unavailable, remote users cannot access the environment. Therefore, at least two NetScaler Gateway hosts should be deployed per data center to prevent this component from becoming a single point of failure.

When configuring NetScaler Gateway in a high availability (active/passive) pair, the secondary NetScaler Gateway monitors the first appliance by sending periodic messages, also called heartbeat messages or health checks, to determine if the first appliance accepts connections. If a health check fails, the secondary NetScaler Gateway tries the connection again for a specified time until it determines that the primary appliance is not working. If the secondary appliance confirms the health check failure, the secondary NetScaler Gateway takes over for the primary NetScaler Gateway.

NetScaler Gateway can be deployed in a tiered design to provide high throughput, high availability, and scalability for HDX client traffic. Multi-tier NetScaler architectures enable expanding the number of supported HDX proxy users in a single data center while still using a single access URL. If done correctly, the user experience is no different from what you get using a single HA ADC pair.

NetScaler Gateway appliances or VMs operate in a cluster as a single system image to coordinate user sessions and manage traffic to network resources. The typical use case for using a cluster over a tiered architecture is simplifying management. With a cluster, a single management IP (CLIP) can be used to manage the entire cluster. With tiered, the top tier load balancer (used to distribute load across the Gateways) has a different configuration (and management IP) than the GW tier. However, this management difference can be automated with proper automation (Terraform, Ansible, etc.). A NetScaler Gateway cluster should be built with an odd number of nodes, e.g., 3, 5, or 7, with a max of 31 available. It is recommended to use odd-numbered cluster nodes to minimize issues with leader election for the cluster coordinator (CCO) node.

Decision: Platform

The key resource constraints must be identified to identify an appropriate NetScaler platform to meet project requirements. Since all remote access traffic will be secured using the secure sockets layer (SSL), transported by Hypertext Transfer Protocol (HTTP) in the form of HTTPs, two resource metrics should be targeted:

- **SSL throughput** — SSL throughput is the gigabits of SSL traffic that may be processed per second (Gbps).
- **SSL transactions per second (TPS)** – The TPS metric identifies how often an Application Delivery Controller (ADC) may execute an SSL transaction per second. The capacity varies primarily by the key length required. TPS capacity is primarily a consideration during the negotiation phase when SSL is first set up. It is less of a factor in the bulk encryption/decryption phase, which is most of the session life. While TPS is an important metric to monitor, field experience has shown that SSL throughput is the most significant factor in identifying the appropriate NetScaler Gateway.
- **User logon rate —** The default rate is 30 logins/sec, and a NetScaler can safely scale up to 60. If the peak user logon rate is higher than 60 logins/sec, the design needs to be re-architected using the abovementioned tiered architecture.

The bandwidth overhead for SSL depends on the packet size. The smaller the packet size, the higher the overhead. As ICA is typically a small packet protocol, an estimated 20% overhead for a typical interactive would be reasonable. If the app does bulk data transport, bandwidth overhead can be estimated at around 5%. In summary, this calculation depends on the type of app traffic. If it isn't possible to differentiate between highly interactive traffic and bulk data transport, then err at 20% overhead.

$$SSL\ Throughput = Maximum\ Concurrent\ Bandwidth * 1.20$$

For example, assuming 128Mbps maximum concurrent bandwidth, the appropriate NetScaler model can be determined as follows:

$$\sim 155\ Mbps = 128\ Mbps * 1.20$$

The SSL throughput value should be compared to the throughput capabilities of various NetScaler platforms to determine the most appropriate one for the environment. Three main platform groups are available, each providing broad scalability options.

- **VPX** – A NetScaler VPX device provides the same full functionality as hardware NetScaler. However, NetScaler VPXs can leverage 'off-the-shelf' servers for hosting. Typically, organizations create a baseline cap for the VPX instances at 500 users.
- **MPX** – A NetScaler MPX is the hardware version of the NetScaler devices. The MPX device is more powerful than the virtual NetScaler and can support network optimizations for larger-scale enterprise deployments, particularly when SSL offload will be configured, as this is done in software on the VPX versus dedicated SSL chips on the MPX.
- **SDX —** A NetScaler SDX blends virtual and physical NetScaler devices. An SDX machine is a physical device capable of hosting multiple virtual VPX NetScaler devices. This device consolidation aids in reducing the required shelf space. NetScaler SDXs are suitable for handling network communications for large enterprise deployments and/or multi-tenant hosting providers.
- **BLX** – A NetScaler BLX is a software form factor designed to run natively on bare metal. NetScaler BLX for bare metal runs as a Linux process on your hardware of choice. Because NetScaler BLX is a lightweight software package with no hypervisor or container overhead, you get extraordinarily fast performance. Certain use cases, such as

full disk at rest disk encryption, can only be done with BLX. BLX is also better suited for hyperscalers, as the hyperscalers tend to rate limit PPS, which causes varying performance randomly. BLX, since it occupies the entire host (bare metal instance), helps mitigate PPS rate-limiting issues with public clouds.

SSL throughput capabilities of the NetScaler platforms may be found in the Citrix NetScaler data sheet. However, actual scalability will depend on security requirements. NetScaler SSL throughput decreases with increasingly complex encryption algorithms and longer key lengths. Customers should generally prefer ECDHE over non-elliptic curve ciphers such as DHE. ECDHE is lighter on resources while providing the same level of cryptographic security. Also, this calculation represents a single primary NetScaler. At a minimum, N+1 redundancy is recommended, which would call for an additional NetScaler for the identical platform and model.

Note:

The Citrix NetScaler data sheet typically represents throughput capabilities under optimal performance conditions. However, performance is directly affected by security requirements. For example, if the RC4 encryption algorithm and a 1k key length are used, a VPX platform may be able to handle more than 500 HDX proxy connections. However, if a 3DES encryption algorithm and 2k key length are used (becoming more common), the throughput may be halved.

Decision: NetScaler Console

NetScaler Console and NetScaler Console service (formerly known as NetScaler ADM and ADM service) are solutions for managing all NetScaler deployments, including NetScaler MPX, NetScaler VPX, NetScaler SDX, NetScaler CPX, NetScaler BLX, and NetScaler Gateway that are deployed on-premises or in the cloud.

NetScaler Console provides all the capabilities required to quickly set up, deploy, and manage application delivery in NetScaler deployments, with rich analytics of application health, performance, and security. Key benefits involve deploying pooled licensing, HDX insights, SSL certificate tracking, automated upgrades, CVE tracking, and more.

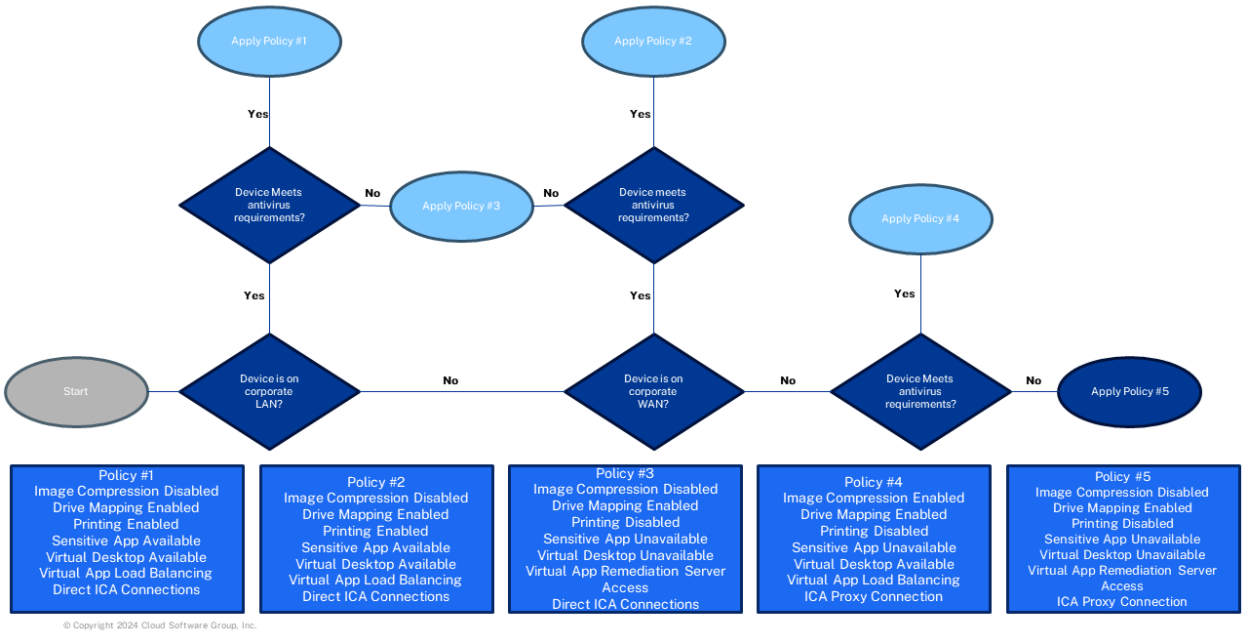
The NetScaler Console virtual appliance can be installed on Microsoft Hyper-V, VMware ESXi, Linux KVM, and XenServer platforms. The general requirements for the virtual appliance are a minimum of 32 GB RAM, 8 vCPUs, and 120 GB SSD. The NetScaler Console HA Deployment Guide can assist with the sizing requirements and overall deployment of the NetScaler Console virtual appliance.

Decision: Pre-Authentication Policy

An optional pre-authentication policy may be applied to user groups with NetScaler Gateway as their authentication point. Pre-authentication policies limit environmental access based on whether the endpoint meets certain criteria through Endpoint Analysis (EPA) Scans.

Pre-authentication access policies can be configured to test antivirus, firewall, operating system, or registry settings. These policies can prevent access entirely or by Citrix Virtual Apps and Desktops to control session features such as clipboard mapping, printer mapping, and even the availability of specific applications and desktops. For example, if a user device does not have antivirus installed, a filter can be set to hide sensitive applications.

This figure provides an overview of how multiple policies can be used to customize the features of a virtualization resource:



Design Tip

Use EPA scans to scan for updated antivirus definitions. Use Domain SID to verify that users are members of the enterprise domain or that the endpoint devices have a specific certificate installed before allowing access.

Decision: Session Policy

User groups with NetScaler Gateway as their authentication point must define corresponding session policies, which define the overall user experience after authentication.

Organizations create session policies based on the type of Citrix Workspace app used. For session policy assignment, devices are commonly grouped as either non-mobile (such as Windows, Mac, and Linux, ChromeOS-based) or mobile (such as iOS or Android). Therefore, a decision on whether to provide support for mobile devices, non-mobile devices, or both should be made based on client device requirements identified during the assess phase.

To identify device session policies, include expressions such as:

- Mobile devices - The expression is set to REQ.HTTP.HEADER User-Agent CONTAINS CitrixReceiver, which is given a higher priority than the non-mobile device policy, to ensure mobile devices are matched while non-mobile devices are not.
- Non-mobile devices – The expression is set to ns_true, which signifies that it should apply to all traffic sent to it.

An alternative use of session policies is to apply endpoint analysis expressions. These session policies are applied post-authentication yet mimic the previously mentioned pre-authentication

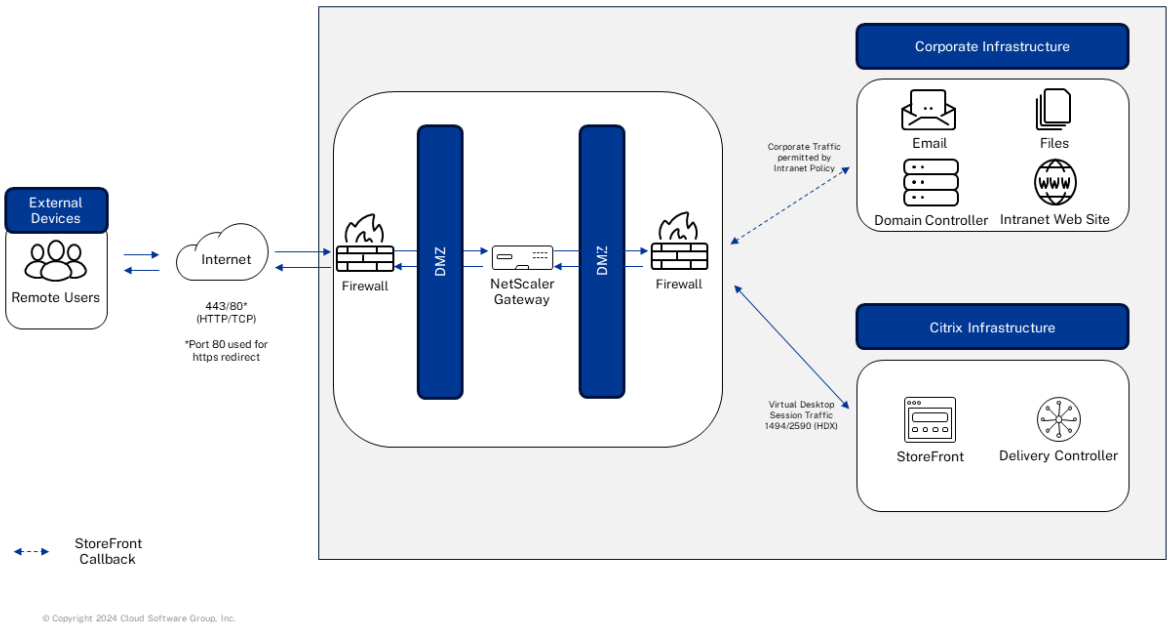
policies. The use of session policies is an option to provide a fallback scenario to endpoints that do not meet full security requirements, such as read-only access to specific applications.

Decision: Session Profile

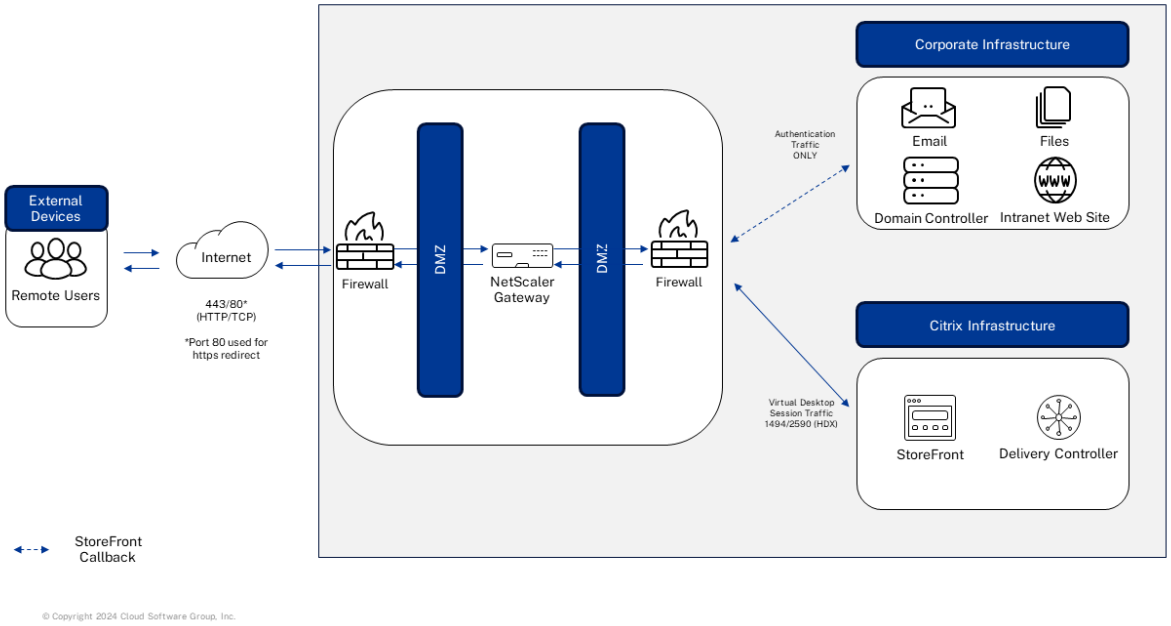
Each session policy must have a defined corresponding session profile. The session profile defines the details required for the user group to gain access to the environment. Two primary forms of session profiles determine the access method to the virtual desktop environment:

- SSLVPN – Users create a virtual private network and tunnel all traffic configured by IP addresses through the internal network. The user's client device can access permitted intranet resources as if it were on the internal network. This includes Citrix Virtual Apps and Desktops sites and any other internal traffic such as file shares or intranet websites. This is considered a potentially less secure access method since network ports and routes to services outside of the virtual desktop infrastructure may be opened, leaving the enterprise susceptible to risks that may come with full VPN access. Configuring traffic and authorization policies to limit this access can help mitigate this security issue, but this does introduce management overhead.
- These risks may include denial of service attacks, attempts at hacking internal servers, or any other form of malicious activity that may be launched from malware, Trojan horses, or other viruses via an Internet-based client against vulnerable enterprise services via routes and ports.

Another decision to consider when SSLVPN is required is whether to enable split tunneling for client network traffic. By enabling split tunneling, client network traffic directed to the intranet by the Citrix Workspace app may be limited to routes and ports associated with specific services. By disabling split tunneling, all client network traffic is directed to the intranet; therefore, traffic destined for internal services and external services (Internet) traverse the corporate network. The advantage of enabling split tunneling is that exposure of the corporate network is limited, and network bandwidth is conserved. The advantage of disabling split tunneling is that client traffic may be monitored or controlled through web filters or intrusion detection systems.



- HDX proxy—With HDX Proxy, users connect to their virtual desktops and applications through the NetScaler Gateway without exposing internal addresses externally. The NetScaler Gateway is a micro VPN in this configuration and only handles HDX traffic. Other types of traffic on the client's endpoint device, such as private mail or personal Internet traffic, do not use the NetScaler Gateway. A decision must be made whether this method is supported for each user group based on the endpoint and Citrix Workspace app used. HDX Proxy is a secure remote virtual desktop access method since only traffic specific to the desktop session can pass through to the corporate infrastructure.



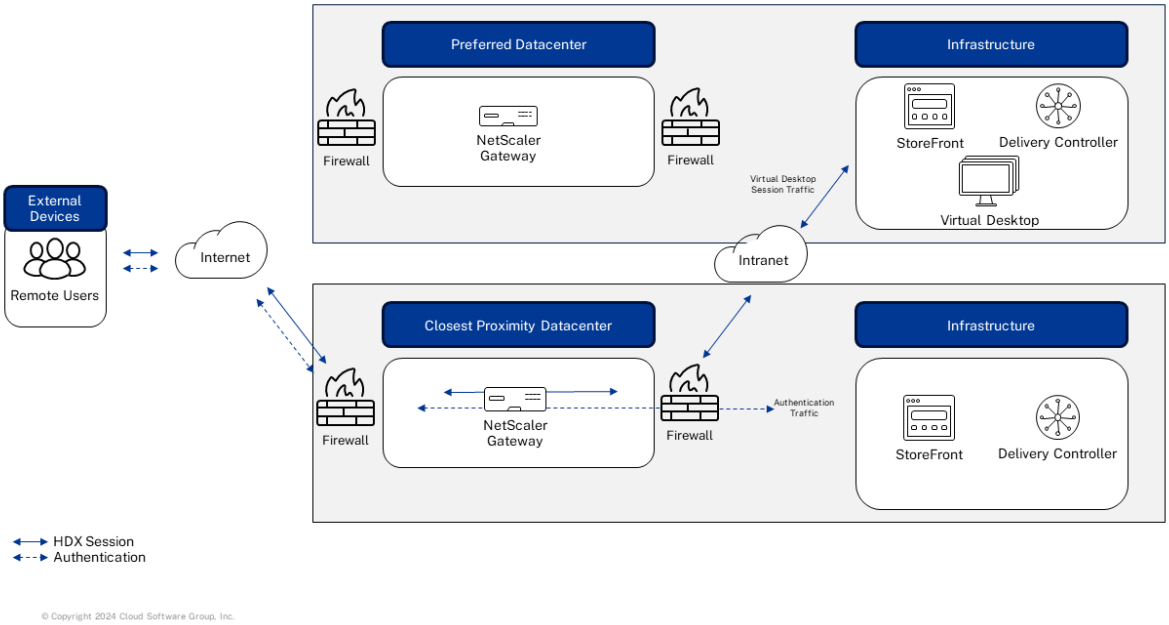
Decision: Preferred Datacenter

Enterprises often have multiple active datacenters that provide high availability for mission-critical applications. Some virtual desktops or applications may fall into that category, while others may only be accessed from a specific preferred datacenter. Therefore, the initial NetScaler Gateway a user authenticates to in a multi-active datacenter environment may not be within the preferred datacenter corresponding to the user's VDI resources. StoreFront can determine the location of the user's assigned resources and direct the HDX session to those resources; however, the resulting path may be sub-optional (WAN connection from the NetScaler Gateway to the virtual desktop/application resources as opposed to LAN connection).

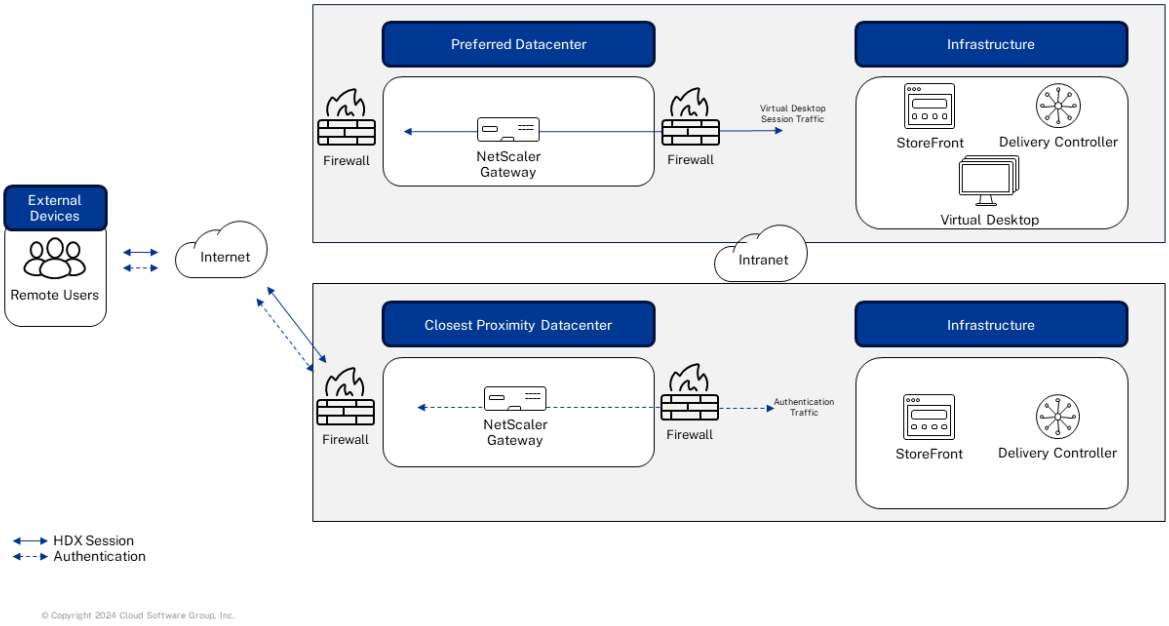
Static and dynamic methods are available to direct HDX sessions to their virtual desktop resources in their primary datacenter. The decision regarding which method to select should be based on the availability of technology to dynamically assign site links, such as Global Server Load Balancing (GSLB) along with the network assessment of intranet and Internet bandwidth as well as Quality of Service (QoS) capabilities.

> Note:
>
> For more information on configuring the static and dynamic methods of GSLB, please refer to NetScaler Product Documentation - Configuring GSLB for Proximity

- Static
    - Direct — The user may be given an FQDN mapped to an A record dedicated to the primary datacenter NetScaler Gateway(s), allowing them to access their virtual desktop directly wherever they are. This approach eliminates a layer of complexity added with dynamic allocation. However, it also eliminates fault tolerance options, such as the ability to access the virtual desktop through an alternative intranet path when a primary datacenter outage is limited to the access infrastructure.
- Dynamic
    - Intranet – For most dynamic environments, the initial datacenter selected for authentication is the one closest to the user. Protocols such as GSLB dynamic proximity calculate the least latency between the user's local DNS server and the NetScaler Gateway. Thereafter, by default, the HDX session is routed through the same NetScaler Gateway to whichever datacenter is hosting the user's virtual desktops and applications. The advantage of this approach is that most of the HDX sessions would traverse the corporate WAN, where quality of service may be used.



- Internet - Alternatively, the HDX session can be re-routed through an alternate NetScaler Gateway proximate to the backend VDI desktop / Citrix Virtual Apps server, resulting in most of the HDX session traveling over the Internet. For example, a user traveling in Europe with a dedicated desktop in the United States may be directed to a NetScaler Gateway hosted in a European datacenter based on proximity. However, when the user launches their desktop, an HDX connection will be established to the virtual desktop via a NetScaler Gateway hosted in the preferred datacenter in the United States. This conserves WAN network usage (at the cost of QoS) and is recommended in cases where the user's Internet connection may provide a more reliable experience than the corporate WAN.



Some customers will use a combination of these methods, such as geo-specific dynamic URLs, to provide fault tolerance within a geographic area (such as North America) without incurring the complexity of global GSLB.

Site-to-Site Connectivity

A Citrix Virtual Apps and Desktops site can span multiple locations. To successfully implement a multi-site solution, the design must consider the site-to-site links and Citrix Virtual Apps and Desktops session routing between locations to provide the best user experience.

Layer 3: The Resource Layer

The resource layer is the third layer of the design methodology, and the final layer focuses specifically on the user groups.

The decisions made within the resource layer define the solution's overall user acceptance. Profiles, printing, applications, and overall desktop image design are pivotal in aligning the desktop with the user group's requirements identified within the assess phase.

User Experience

Perception is everything when it comes to a good VDI experience. Users expect an experience similar to or better than a traditional physical desktop.

Codecs, transport protocols, and self-service capabilities affect the overall experience. Poor-quality graphics, lagging video, or extremely long login times can destroy the user experience. A proper user experience design can meet any network challenge.

Decision: Display Protocol

Selecting the correct display protocol determines the quality of static images, video, and text within the user's session and determines the impact on single-server scalability. Administrators have these options:

- **10-Bit High Dynamic Range (HDR)** — With 10-bit High Dynamic Range (HDR) virtual desktop sessions, you can use enhanced encoding and decoding capabilities to render high-quality images and videos with an extended range of colors and greater contrast and brightness. You can also customize the white luminance level, Extended Display Identification Data (EDID), and visual quality to improve the user experience.
- **H.264** - One of the most commonly used codecs, has great hardware and software decoding support. However, as screen resolutions increased and technology like HDR was introduced, H.264's limitations became apparent. It does come at the expense of CPU processing time, reducing single-server scalability.
- **H.265** - High-Efficiency Video Coding (HEVC), also known as H.265, is the successor to H.264. H.265 offers users the same quality at a lower bandwidth, improving user experience in bandwidth-restricted situations. However, H.265 requires a GPU for encoding and decoding. Performing this on a CPU is possible, but it influences overall performance and decreases scalability.
- **AV1** - Citrix added support for the AV1 video codec. As AV1 is a newer codec, it has newer graphics card requirements for both the VDA and endpoint. The benefit of AV1 is that it has superior image compression, better image quality, and lower bandwidth usage compared to H.264 and H.265.

The video codecs above can be automatically detected to optimize the VDA and endpoint communication. The decoding capabilities are evaluated when using Workspace App for Windows, and the VDA can select the best codec to use when connecting. Using a video codec, whether it is H.264, H.265, or AV1, provides a richer experience than adaptive JPEG at the expense of CPU to compress regions of fluid movement. Network bandwidth will generally be less with a video codec compared to Adaptive JPEG for multimedia workload, and newer codecs have better compression and, therefore, lower bandwidth usage. Running your own tests with your specific use case is highly recommended.

Decision: Video Codec Policy

Consider the workload of each user group when configuring Citrix policies to set the video codec and ensure the desired quality and user experience. When creating these policies, administrators have these options:

- **Thinwire** — Thinwire, part of the Citrix HDX technology stack, is the default display-remoting technology used in Citrix Virtual Apps and Desktops and Citrix DaaS. Display remoting technology allows graphics generated on one machine to be transmitted, typically across a network, to another machine for display. Graphics are generated from user input, such as keystrokes or mouse actions. Thinwire operates in two encoding modes: Thinwire full-screen H.264 or H.265 and Thinwire with selective H.264 or H.265.
- **HDX 3DPro** - By using HDX 3D Pro, you can deliver graphically intensive applications as part of hosted desktops or applications on Single-session OS machines. HDX 3D Pro supports physical host computers (including desktop, blade, and rack workstations) and, along with Citrix's entire graphics stack, supports GPU Passthrough and GPU virtualization technologies offered by XenServer, vSphere, and Hyper-V (passthrough only) hypervisors. You can create VMs with exclusive access to dedicated graphics processing hardware using GPU Passthrough. You can install multiple GPUs on the hypervisor and assign VMs to each of these GPUs on a one-to-one basis. Using GPU

virtualization, multiple virtual machines can directly access the graphics processing power of a single physical GPU.

Decision: Transport Protocol

There are three ways to transport the HDX protocol across the network:

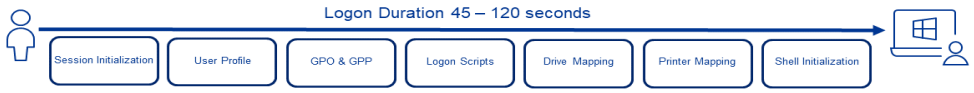
- **TCP** — Uses the industry-standard TCP transport protocol. It is a common transport protocol over LAN and low-latency WAN connections, but it suffers when connection distances increase, thus increasing latency and incurring more retransmissions.
- **EDT** uses a Citrix proprietary transport protocol called Enlightened Data Transport, based on UDP. It is most common on long-distance WAN links for high latency/packet loss networks. EDT provides a more interactive user experience without increasing the server's CPU load, but it consumes more network bandwidth than TCP.
- **Adaptive Transport** — Uses TCP and EDT transport protocols. EDT is used unless the network does not support transporting EDT over the network, in which case it automatically changes to TCP.

Most environments will use Adaptive Transport as the standard transport option unless the network does not have the appropriate firewall ports opened or NetScaler Gateway configured appropriately. NetScaler Gateway 14.1.12.30 or later is recommended. Alternatively, the Citrix Gateway Service with Rendezvous enabled can also be used. For the detailed requirements for using AV1, see the product documentation.

Decision: Logon Optimization

The logon process must be completed when a user logs onto a Citrix Virtual Apps and Desktops session. This includes session initialization, user profile loading, group policy preferences execution, drive mapping, printer mapping, logon script execution, and desktop initialization. Each process takes time and increases the duration of the login.

Most organizations include many mappings and complex logon scripts. When each of these items executes, the logon time drastically increases.



Minimizing GPOs and start-up scripts and removing unnecessary startup applications and tasks will help improve logon performance. In addition to these general recommendations, Citrix has features to help further improve logon performance.

Workspace Environment Management removes drive mappings, printer mappings, logon scripts, and roaming profiles from the standard login process. With asynchronous processing, Workspace Environment Management applies the mappings/scripts/profiles in the background after the session, and the desktop initializes. The user receives the same environment but receives their desktop interface faster. To use this asynchronous processing, most environments should consider implementing drive mappings, logon scripts, roaming profiles, and GPOs via Workspace Environment Manager.

The Citrix Optimizer is a tool that allows administrators to apply sets of recommended and optional configuration changes to both single and multi-session VDAs. It should be run before publishing an image to configure settings to optimize logon and run-time performance. The Optimizer can be run in three modes: Analyze, Optimize, and Rollback. Administrators should re-run the Optimizer Tool after image updates to re-apply the required settings should they be changed during the update process.

Citrix Director provides granular information on login times broken down by various phases, such as brokering, authentication, profile load, and interactive session. It also shows trends for the overall environment, comparing a particular user's average to their peers. Additional info can be found here.

User Profiles

A user's profile is critical to delivering a consistently positive experience within a virtual desktop or application scenario. A well-designed virtual desktop solution can fail if users are frustrated due to lengthy logon times or lost settings.

The user profile solution chosen must align with the personalization characteristics of the user group captured during the assess phase and the VDI model selected.

Decision: Profile Type

This section provides an overview of the available profile types and guides the optimal user profile for each VDI model.

- **Local profiles** – Local profiles are stored on each server or desktop operating system and are initially created based on the default user profile. Therefore, a user accessing these resources would create an independent profile on each system. Users can retain changes to their local profile on each individual system, but changes are only accessible for future sessions on that system. Local profiles require no configuration; if a user logging into a server or desktop operating system does not have a profile path administratively defined, a local profile is created by default.
- **Roaming profiles** – Roaming profiles are stored in a centralized network repository for each user. Roaming profiles differ from local profiles in that the information in the profile (whether it is a printer, a registry setting, or a file stored in the documents folder) can be made available to user sessions accessed from all systems in the environment. Configuring a user for a roaming profile requires an administrator to designate the user's profile path (for virtual desktops) or terminal server profile path to a particular network share. The first time the user logs on to a server or desktop operating system, the default user profile is used to create the user's roaming profile. During logoff, the profile is copied to the administrator-specified network location.
- **Mandatory profiles** – Mandatory profiles are typically stored in a central location for many users. However, the user's changes are not retained at logoff. Configuring a user for a mandatory profile requires an administrator to create a mandatory profile file (NTUSER.MAN) from an existing roaming or local profile and assign users with a

terminal services profile path. This can be achieved using Microsoft Group Policy, customizing the user properties in Active Directory, or Citrix Profile Management.

- **Citrix Profile Management (CPM)** – CPM combines a robust profile core (a mandatory or a local default profile) with user-specific registry keys or files merged during login. This technique lets administrators control which changes are retained tightly and keep the user profiles small. Furthermore, CPM addresses the last write wins issue using mature queuing techniques that automatically detect and prevent simultaneous writes that could potentially overwrite changes made in another session. Thus minimizing user frustration resulting from lost profile changes when accessing multiple servers or virtual desktops simultaneously. In addition, CPM captures and records only the changes within the profile rather than writing the entire profile at logoff.
- **Profile Containers** — Profile containers redirect the entire Windows user profile to a file, typically in a VHD or VHDX format, and store it in a storage repository, most often an SMB share. Profile containers are mounted when a user logs into a virtual session and can resolve issues with slow logins and improve the overall login experience. Citrix Profile Management provides a profile container in VHDX format.

This table compares the capabilities of each profile type:

Feature	Local	Roaming	Mandatory	CPM	CPM Container
Central management/roams with user	X	✓	○	✓	✓
User settings are stored persistently.	○	✓	X	✓	✓
Granular capture of user settings	X	X	X	✓	✓

✓: Functionality available / X: Functionality not available / ○: Optional

Understanding each user group's personalization requirements and the VDI model assigned is important for selecting the optimal profile type.

This table provides recommendations on selecting the appropriate user profile type based on VDI resources:

	Local	Roaming	Mandatory	CPM	CPM Container
User setting persistence required (personalization characteristic: basic/complete)					
Hosted Windows App	X	✓	X	✓	✓
Hosted Shared Desktop	X	✓	X	✓	✓
Hosted Pooled Desktop	X	✓	X	✓	✓
Hosted Static Desktop	○	✓	X	✓	✓
Hosted 3D Pro Graphics Desktop	○	✓	X	✓	✓
Remote PC Access	✓	○	X	○	○

	Local	Roaming	Mandatory	CPM	CPM Container
User setting persistence not required or not desired (personalization characteristic: none)					
Hosted Windows App	X	X	✓	X	X
Hosted Shared Desktop	X	X	✓	X	X
Hosted Pooled Desktop	✓	X	✓	X	X
Hosted Static Desktop	X	X	✓	X	X
Hosted 3D Pro Graphics Desktop	○	X	✓	X	X
Remote PC Access	○	X	✓	X	X

✓: Recommended / X: Not Recommended / ○: Viable

Decision: Folder Redirection

Redirecting special folders can supplement any of the described profile types. While redirecting profile folders, such as user documents and favorites, to a network share is a good practice to minimize profile size, architects need to be aware that applications may frequently read and write data to profile folders such as AppData, causing potential issues with file server utilization and responsiveness. It is important to thoroughly test profile redirection before implementation in production to avoid these issues. Therefore, it is important to research profile read/write activities and to perform a pilot before moving to production. Microsoft Outlook is an example of an application that regularly performs profile read activities, as the user signature is read from the user profile every time an email is created.

This table provides general recommendations to help identify the appropriate folders to redirect:

Folder	Local	Roaming	Mandatory	Hybrid File-Based	Profile Container
Application Data	X	○	X	○	○
Contacts	X	✓	X	○	○
Desktop	X	✓	X	○	○
Downloads	X	○	X	○	○
Favorites	○	✓	○	✓	✓
Links	X	✓	X	○	○
Documents	○	✓	○	✓	✓
Music	○	✓	○	○	○
Pictures	○	✓	○	○	○

Folder	Local	Roaming	Mandatory	Hybrid File-Based	Profile Container
Videos	○	✓	○	○	○
Saved Games	X	○	X	○	○
Start menu	X	X	X	X	X

✓: Recommended / X: Not Recommended / ○: Optional

Decision: Folder Exclusion

Excluding folders from being persistently stored as part of a roaming or hybrid profile can help reduce profile size and logon times. By default, Windows excludes the AppData\Local and AppData\LocalLow folders, including all subfolders, such as History, Temp, and Temporary Internet Files. In addition, the downloads and saved games folders should also be excluded. All redirected folders should also be excluded from the profile solution.

There are also recommended exclusions for Google Chrome and Firefox,

Decision: Profile Caching

Local caching of roaming or hybrid user profiles on a server or virtual desktop is the default Windows behavior and can reduce login times and file server utilization/network traffic. With profile caching, the system only has to download changes made to the profile. The downside of profile caching is that it can consume significant amounts of local disk storage on multi-user systems, such as hosted shared desktop hosts.

Local profile caches should not be deleted, even when non-persistent machines are restored to a pristine state on logoff, as deleting the cached profile consumes time and CPU resources. Administrators should consider deleting profile caches on multi-session desktop and application systems to reduce disk storage consumed by individual user profiles over time.

Configuring the *"Delay before deleting cached profiles"* Citrix policy sets an optional extension to the delay before locally cached profiles are deleted at logoff. Extending the delay is useful if a process keeps files or the user registry hive open during logoff. This can also reduce logoff times for large profiles.

Decision: Profile Permissions

For security reasons, user profile shares should always be configured to restrict permissions so others cannot view a given user's profile. Administrators, by default, cannot access user profiles but can take ownership of files and folders as an auditable action. If default administrator access is desired, the "Add the Administrators security group to roaming user profiles" policy setting can be configured. The configuration of this policy should be aligned with the security and privacy considerations of the user groups captured during the assess phase. For more information on the permissions required for the file share hosting user profiles and data, please refer to Microsoft's Deploy Roaming Profiles article.

Decision: Profile Path

Determining the network path for the user profiles is one of the most significant decisions during the user profile design process. In general, it is strongly recommended that you leverage a redundant and high-performance file server, NAS device, or cloud-based solution such as Azure Files.

Four topics must be considered for the profile share:

- **Performance** – File server performance will affect logon and logoff times, and depending on other decisions, such as redirected folders and profile streaming, can impact the user's experience within the session. A single file server cluster may not be sufficient for large virtual desktop infrastructures to handle peak activity periods. The file server address and share name must be adjusted to distribute the load across multiple file servers.
- **Location** – User profiles are transferred over the network using the SMB protocol, which performs poorly on high-latency network connections. Furthermore, WAN connections are typically bandwidth-constrained, which can add additional delay to the profile load process. Therefore, the file server should be located in close proximity to the servers and virtual desktops to minimize logon times.
- **Operating system platforms** — User profiles are tightly integrated with the underlying operating system, and reusing a single user profile for different operating systems is not recommended.
- **Availability** — The RTO and RPO goals will guide the availability strategy of the user profiles. It is generally recommended that profile data be backed up on a regular basis. It is critical that users have access to the data they need in a DR scenario.

Two methods can be used to address these challenges based on Windows' built-in technology:

- **User object** – An individual profile path can be specified for every user object in Active Directory containing the file server name and profile directory. Since only a single profile path can be specified per user object, ensuring a separate profile is loaded for each operating system platform is impossible.
- **Computer group policy or system variables** – The user profile path can also be configured through computer-specific group policies or system variables. This enables administrators to ensure a user profile is dedicated to the platform. Since computer-specific configurations affect all system users, all user profiles will be written to the same file server. Dedicated Citrix Virtual Apps and Desktops delivery groups must be created per file server to balance user profiles across multiple servers.

> Note:
>
> Microsoft does not support DFS-N combined with DFS-R for actively used user profiles. For more information, please refer to this Microsoft article.

When using Citrix Profile Management, a third option is available to address these challenges:

- **User object attributes and variables** – Citrix Profile Management enables the administrator to configure the profile path using a computer group policy and the attributes of the user object in Active Directory to specify the file server dynamically. To achieve this, three steps are required:
  - Create a DNS alias (for example, fileserver1) that refers to the actual file server.
  - Populate an empty LDAP attribute of the user object (for example, l or UID) with the DNS Alias.
  - Configure Citrix Profile Management using GPO to use a profile path that refers to the LDAP attribute (for example, if attribute UID is used, the profile path becomes \\#UID#\Profiles\profiledirectory)

In addition, Citrix Profile Management automatically populates variables to specify the profile path dynamically based on the operating system platform. Valid profile management variables are:

- **!CTX_PROFILEVER!** – Expands to the profile version depending on the profile version.

- **!CTX_OSBITNESS!** – Expands to x86 or x64, depending on the bit level of the operating system.
- **!CTX_OSNAME!** – Expands to the short name of the operating system, for example, Win11
- **!ctx_localsettings! –** Expands to AppData\Local
- **!ctx_roamingappdata! –** Expands to AppData\Roaming
- **!ctx_startmenu! –** Expands to AppData\Roaming\Microsoft\Windows\Start Menu
- **!ctx_internetcache! -** Expands to AppData\Roaming\Microsoft\Windows\Temporary Internet Cache
- **!ctx_localappdata! –**Expands to AppData\Local

By combining both capabilities of Citrix Profile Management, a fully dynamic user profile path can be created, which can be load-balanced across multiple file servers and ensure profiles of different operating system platforms are not mixed. An example of a fully dynamic user profile path is shown here:

\\#UID#\profiles$\%USERNAME%.%USERDOMAIN%\!CTX_OSNAME!!CTX_OSBITNESS!

Decision: Profile Streaming

> Note:
>
> The following design decision only applies to those environments that use Citrix Profile Management.

With user profile streaming, files and folders contained in a profile are fetched from the user store (file server) to the local computer when a user accesses them. Citrix Profile Management immediately reported that the profile load process had been completed during the login process, reducing profile load time.

Citrix recommends enabling profile streaming for all scenarios, and in Citrix Virtual Apps and Desktops 2402 LTSR, this is enabled by default. If a local cached copy of the user profile is desired for performance reasons, enabling the "Always Cache" setting and configuring a size of 0 is recommended. This ensures that the user profile is downloaded in the background and enables the system to use this cached copy going forward.

> Design Tip:
>
> Some poorly written applications might load faster if their AppData has already been streamed to the VDI resource. Enabling the "Always Cache" option for profile streaming can help improve performance when the AppData folder is not redirected.

Decision: Active Write Back

> Note:
>
> The following design decision only applies to those environments that use Citrix Profile Management.

By enabling the active write-back feature, Citrix Profile Manager detects when an application has written and closed a file and copies the file back to the network copy of the profile during idle periods. This feature can be tremendously beneficial when a single user leverages multiple virtual desktops or hosted shared desktops simultaneously. However, Citrix Profile Management does not copy any registry changes back to the network except during an ordered logoff. As such, there is a risk that the registry and files may get out of alignment on non-persistent systems, where locally cached profile information is wiped upon reboot. Therefore, it is recommended to disable active write-back functionality for non-persistent scenarios.

Decision: Configuration Approach

> Note:
>
> The following design decision only applies to those environments that use Citrix Profile Management.

Citrix Profile Management can be configured using a ".ini" file, Microsoft Group Policy, Citrix WEM, or Citrix Policy. Each option offers the same configuration settings, and selecting one over the other is based on factors such as ease of use. For example, Group Policy may be recommended because it allows administrators to perform Windows and Citrix profile configurations simultaneously, minimizing the tools necessary for profile management.

Decision: Replication

While having an active/active datacenter on a network level is easily accomplished with GSLB, replicating user data makes having a fully active/active deployment complex in most situations.

Administrators need to consider the implications for profile and other user data requirements to have an active/active configuration where users are not statically assigned to a specific datacenter. Profile solutions are not active-active, so user personalization will be limited or depend on intra-datacenter latency, which can impact the overall user experience.

User data, such as Windows profiles and documents, should be synchronized between datacenters for redundancy and failover purposes. Although it is recommended to replicate user data between datacenters, the replication would be an active/passive configuration. This means the data can only be actively consumed from a single datacenter. The reason for this limitation is the distributed file-locking method inside Windows that only allows a single user to write to a file actively. Therefore, active/active replication of user data is not supported. Any supported configuration consists of a one-way replication of data active in a single datacenter at any time.

Citrix Profile Management can be configured to replicate profile data between datacenters. On user login, CPM determines the datastore with the latest profile by default or may be configured to select either the earliest configured or best-performing store for access. Once users log off their session, the profile data is written back across all configured stores. This provides redundancy for the user profile but does not provide in-session failover capabilities.

User Data

To be effective, users must be able to access their data. For the user to have a good experience, the data must be in close proximity to the application. As the distance between the application and data increases, latency also increases, which slows down any file operation (opening, saving, modifying).

In a VDI-based environment, administrators must understand where users store their data and the impact of access.

Decision: User Data Location

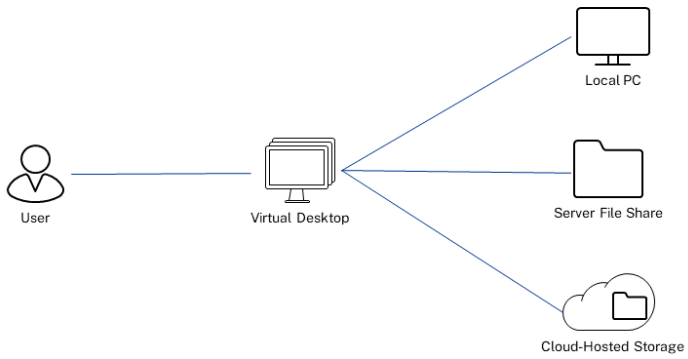
Users traditionally stored their data on their local device or on a network file server designated with a drive mapping. Due to IT storage space limitations or the inability to have the data follow the user to other mobile devices, users turned to free, cloud-based storage offerings like OneDrive, ShareFile, DropBox, and Box. To get access to the data, the user would install the storage vendor's agent on their traditional Windows PC, allowing them direct access to the cloud-hosted storage repository. One of the primary considerations for cloud-based storage offerings is caching on files upon connection, which can cause high network and I/O utilization and potentially fill up drive space. These solutions should utilize on-demand sync such as files are only downloaded upon access.

Administrators must design the solution to consider user storage using a Multi-Agent Strategy. In VDI, users require the admin to install and configure the agent for each storage provider, which assumes the storage agent supports the non-persistent VDI model. Each agent is a new application that the admin must manage and maintain.

Decision: User Data Access

A critical aspect of a successful VDI solution is for the user experience to remain the same as that of a traditional PC. If users open files from within the application or navigate with File Explorer to access a file, that functionality must continue functioning.

A user's data can exist on the local PC, network file share, or cloud.



With local PCs, on-premises network shares, and cloud-hosted storage options available to users, administrators must understand how users' data access affects the infrastructure and VDI experience.

- **Direct Data Access** – Users access a file on a remote server (on-premises Windows server or cloud-hosted storage provider). The distance between the application and the file directly affects the experience. Longer distances equate to higher latency. Each file operation (navigate, open, close, save, etc.) takes longer as the latency between application and file storage increases. Windows file servers are often located in the same data center as the user's VDI desktop, making direct data access feasible. Still, cloud-hosted solutions and local PC Access will experience poor response times if the connection between the VDI desktop and the storage repository has high latency.
- **Local Synchronization** – With a traditional PC, users are accustomed to having files local, which mitigates any slow application response times due to extremely low latency. Many cloud-hosted solutions synchronize data, enabling access speeds similar to a local storage model. Many cloud-hosted solutions provide full synchronization or user-configured partial synchronization of certain folders and files. With partial synchronization, only the synchronized files are visible and accessible on the device, causing user confusion. Full and partial synchronization increases VDI costs. For non-persistent VDI, each session is an entirely new desktop requiring synchronization of the user's folders/files, which takes time, network bandwidth, and VDI storage space. Every file synchronized to the VDI desktop must be stored within the organization's data center for the duration of the VDI session.
- **On-Demand Synchronization** – When navigating Explorer, users see a complete but virtual file/folder structure even though those files/folders do not physically exist on the desktop. Selecting a file begins an automatic synchronization to the VDI desktop for that single file. At this point, file access is local, creating a user experience like a traditional PC. When the user saves or closes the file, the file synchronizes back to the cloud. Only the files accessed synchronize, eliminating the waste incurred with the local

data access model. As only accessed files synchronize, the impact on the underlying storage infrastructure and associated storage costs is minimal.

	Direct Data Access	Local Synchronization	On-Demand Synchronization
Network File Server	✓	N/A	N/A
Cloud-hosted	X	X	✓
Local PC	X	X	✓

✓: Recommended / X: Not Recommended

Decision: Data Recovery

File corruption is an issue most users experience. Improperly shutting down the application or PC (hitting the power button instead of closing the application and shutting down the operating system gracefully) often causes many corruption issues.

A few options exist to provide users with data recovery options:

- **Multi-File** – With a traditional PC, users have few recovery options if the files are local. Users often manually create a new copy of the file each day to provide some level of recovery. This solution is hard to manage.
- **Backup/Restore** – Administrators can implement a backup and restore solution to help recover files. However, these solutions rarely work with local files, and for a network file share, the backup process usually only runs nightly or weekly. In addition, restoring a corrupted file requires the user to call support.
- **Versioning** — Cloud-hosted options include file versioning, which automatically creates new file versions as changes are saved. Versioning requires no user intervention and allows users to recover from corruption quickly and with little data loss.

Policies

Policies provide the basis for configuring and fine-tuning Citrix Virtual Apps and Desktop environments, allowing organizations to control connection, security, and bandwidth settings based on various combinations of users, devices, or connection types.

It is important to consider Microsoft and Citrix policies when making policy decisions to ensure that all user experience, security, and optimization settings are considered. Please refer to the Citrix Policy Settings Reference for a list of all Citrix-related policies.

Decision: Preferred Policy Engine

Organizations can configure Citrix policies via Citrix Studio or through Active Directory group policy using Citrix ADMX files, which extend group policy and provide advanced filtering mechanisms.

Using Active Directory group policy allows organizations to manage Windows and Citrix policies in the same location and minimizes the administrative tools required for policy management. Group policies are automatically replicated across domain controllers, protecting the information and simplifying policy application.

Citrix administrative consoles should be used if Citrix administrators cannot access Active Directory policies. Architects should select one of the above two methods appropriate for their organization's needs and use it consistently to avoid confusion with multiple Citrix policy locations.

It is important to understand how the aggregation of policies, known as policy precedence, flows to understand how a resultant set of policies is created. With Active Directory and Citrix policies, the precedence is as follows:

Policy Precedence	Policy Type
Processed first (lowest precedence)	Local server policies
Processed second	Citrix policies created using Citrix Studio
Processed third	Site-level AD policies
Processed fourth	Domain-level AD policies
Processed fifth	Highest level OU in domain
Processed sixth	Next-level OU in domain
Processed last (highest precedence)	Lowest level OU containing object

Policies from each level are aggregated into a final policy that is applied to the user or computer. In most enterprise deployments, Citrix administrators do not have the rights to change policies outside their specific OUs, which will typically be the highest level for precedence. In cases where exceptions are required, applying policy settings from higher up the OU tree can be managed using "block inheritance" and "no override" settings. Block inheritance stops settings from higher-level OUs (lower precedence) from being incorporated into the policy. It is preferred that block inheritance is applied at the Citrix level in the OU structure. However, the block inheritance setting will not be applied if a higher-level OU policy is configured with no override. Given this, care must be taken in policy planning, and available tools such as the "Active Directory Resultant Set of Policy" tool or the "Citrix Group Policy Modeling" wizard should be used to validate the observed outcomes with the expected outcomes.

> Note:
>
> Some Citrix policy settings, if used, must be configured through Active Directory group policy, such as Controllers and Controller registration port, as these settings are required for VDAs to register. These settings can also be configured via AD GPO.

Decision: Policy Integration

Organizations often require Active Directory and Citrix policies to create a completely configured environment when configuring policies. Using both policy sets, the resultant set of policies can become confusing to determine. In some cases, particularly with respect to Windows Remote Desktop Services (RDS) and Citrix policies, similar functionality can be configured in two different locations. For example, enabling client drive mapping in a Citrix

policy and disabling client drive mapping in an RDS policy is possible. The ability to use the desired feature may depend upon the combination of the RDS and Citrix policies. It is important to understand that Citrix policies build upon functionality available in Remote Desktop Services. If the required feature is explicitly disabled in the RDS policy, Citrix policy will not be able to affect a configuration as the underlying functionality has been disabled.

To avoid this confusion, it is recommended that RDS policies only be configured where required. There is no corresponding policy in the Citrix Virtual Apps and Desktops configuration, and the configuration is specifically needed for RDS use within the organization. Configuring policies at the highest common denominator will simplify understanding the resultant set of policies and troubleshooting policy configurations.

Decision: Policy Scope

Once policies have been created, they must be applied to groups of users and/or computers based on the required outcome. Policy filtering allows policies to be applied against the requisite user or computer groups. With Active Directory-based policies, a key decision is whether to apply a policy to computers or users within site, domain, or organizational unit (OU) objects. Active Directory policies are broken down into user configuration and computer configuration. By default, the settings within the user configuration apply to users who reside within the OU at logon, and settings within the computer configuration are applied to the computer at system startup and will affect all users who log in to the system. The challenges of policy association with Active Directory and Citrix deployments revolve around three core areas:

- **Citrix environment-specific computer policies** – Citrix servers and virtual desktops often have computer policies created and deployed specifically for the environment. Applying these policies is easily accomplished by creating separate OU structures for the servers and the virtual desktops. Specific policies can then be created and confidently applied to only the computers within the OU and below and nothing else. Depending on requirements, virtual desktops and servers may be subdivided within the OU structure based on server roles, geographical locations, or business units.
- **Citrix-specific user policies** – When creating policies for Citrix Virtual Apps and Desktops, several policies are specific to user experience and security that are applied based on the user's connection. However, the user's account could be located anywhere within the Active Directory structure, creating difficulty simply applying user configuration-based policies. Applying the Citrix-specific configurations at the domain level is undesirable as the settings would be applied to every system the user logs into. Simply applying the user configuration settings at the OU where the Citrix servers or virtual desktops are located will also not work, as the user accounts are not within that OU. The solution is to apply a loopback policy, which is a computer configuration policy that forces the computer to apply the assigned user configuration policy of the OU to any user who logs onto the server or virtual desktop, regardless of the user's location within Active Directory. Loopback processing can be applied by either merging or replacing settings. Replace is generally recommended. Using replace overwrites, the entire user GPO is governed by the policy from the Citrix server or virtual desktop OU. Merge will combine the user GPO with the GPO from the Citrix server or desktop OU. As the computer GPOs are processed after the user GPOs when merge is used, the Citrix-related OU settings will have precedence and be applied in the event of a conflict. Replacing settings is generally recommended for operational simplicity since other user-level policies may not be required or desired within the Citrix configuration. For more information, please refer to the Microsoft article.
- **Active Directory policy filtering** – In more advanced cases, a policy setting may be needed for a small subset of users, such as Citrix administrators. In this case, loopback processing will not work, as the policy should only be applied to a subset of users, not

all users who log in to the system. Active Directory policy filtering can be used to specify specific users or groups of users to which the policy is applied. A policy can be created for a specific function, and then a policy filter can be set to apply that policy only to a group of users, such as Citrix administrators. Policy filtering is accomplished using the security properties of each target policy.

Citrix policies created using Citrix Studio have specific filter settings available, which may be used to address policy-filtering situations that cannot be handled using group policy. Citrix policies may be applied using any combination of the following filters:

Filter Name	Filter Description	Scope
Access control	Applies a policy based on access control conditions through which a client connects. For example, users connecting through a Citrix NetScaler Gateway can have specific policies applied. This requires the Callback URL to be configured within StoreFront.	User settings
Client IP Address	This filter applies a policy based on the IPv4 or IPv6 address of the user device used to connect the session. If IPv4 address ranges are used, care must be taken with this filter to avoid unexpected results.	User Settings
Client Name	Applies a policy based on the client name, either an exact match or using a wildcard	User Settings
Delivery Group	Applies a policy based on the delivery group membership of the desktop running the session	User and Computer settings
Delivery Group type	Applies a policy based on the type of desktop or application	User and Computer settings
Organizational unit	Applies a policy based on the OU of the desktop or server running the session.	User and Computer settings
Tag	It applies a policy based on any tags to the session's desktop. Tags are strings that can be added to virtual desktops in Citrix Virtual Apps and Desktops environments to search for or limit access to desktops.	User and Computer settings
User or Group	Applies a policy based on the user's Active Directory group membership when connecting to the session.	User settings

Decision: Baseline Policy

A baseline policy should contain all common elements required to deliver a high-definition experience to most users within the organization. A baseline policy creates the foundation for user access and any exceptions that may need to be created to address specific access requirements for groups of users. It should be comprehensive to cover as many use cases as possible and should have the lowest priority, for example, 99 (a priority number of "1" is the highest priority), to create the simplest policy structure possible and avoid difficulties in determining the resultant set of policies. This way, exceptions to the baseline policy can be set at a higher priority. The unfiltered policy provided by Citrix as the default policy can be used to create the baseline policy; it should be set to the lowest priority and disabled.

A baseline policy configuration should also include Windows policies. Windows policies reflect user-specific settings that optimize the user experience and remove features that are not required or desired in a Citrix Virtual Apps and Desktops environment. For example, one common feature turned off in these environments is Windows update. In virtualized environments, particularly where desktops and servers may be streamed and non-persistent, Windows update creates processing and network overhead, and changes made by the update process will not persist a restart of the virtual desktop or application server. In many cases, organizations also use Windows software update service (WSUS) to control Windows updates. In these cases, updates are applied to the master disk and made available by the IT department on a scheduled basis.

In addition to the above considerations, an organization's final baseline policy may include settings specifically created to address security requirements and common network conditions or to manage user device or user profile requirements.

Decision: Peripherals

Citrix integrates with various peripheral devices, such as generic USB peripherals, client drives, specialty keyboards, etc. If your users do not require peripheral access, it is recommended that you restrict it. You can read more about peripheral policies and default settings in our product documentation.

Decision: Audio and Unified Communications

Citrix offers a wide array of policies to help improve audio performance during a virtual session. Our policy settings are in our product documentation.

One specific capability to call out is Browser Content Redirection (BCR). BCR prevents the rendering of web pages on the VDA side, which saves compute for your VDA resources. You can specify websites to be redirected (like YouTube, for example). BCR is supported on the Citrix Workspace app for Windows and Linux.

Citrix offers various optimization packs to improve softphone performance for different vendors in a Citrix environment. Using optimization packs and following the recommended settings in the product documentation will improve softphone performance. Citrix also provides specific guidance on optimizing Microsoft Teams for use in your VDI environment.

Printing

Citrix Virtual Apps and Desktops support a variety of printing solutions. Understanding the available technologies, their benefits, and their limitations is important for planning and successfully implementing the proper printing solution.

Decision: Printer Provisioning

The process of creating printers at the start of a Citrix Virtual Apps and Desktops session is called printer provisioning. There are multiple approaches available:

- **Auto Created** – Auto-creation is a form of dynamic provisioning that attempts to create some or all of the available printers on the client device at the start of a user session. This includes locally attached printers as well as network-based printers. Auto-creating all client printers can increase the session logon time as each printer is enumerated during the logon process.
- **Session-Based** – Session printers are a set of network-based printers assigned to users.
    - Proximity-based session printers are filtered by IP subnet. The network printers created under this policy may vary based on where the user's endpoint device is located. Proximity printing is recommended when Users roam between different locations using the same endpoint device (e.g., laptop, tablet) and where thin clients are used, which cannot connect directly to network-based printers.
    - Session printers may be assigned using the "Session Printer" policy or the "Printer Assignments" policy. The "Session Printer" policy is intended to set default printers for a farm, site, large group, or OU. The "Printer Assignments" policy assigns multiple printers to multiple users. If both policies are enabled and configured, the session printers will be merged into a single list.
- **Universal Printer** – The Citrix Universal Printer is a generic printer object that is auto-created at the start of a session and is not linked to a printing device. When using the Citrix Universal Printer, it is not required to enumerate the available client printers during login, which can greatly reduce resource usage and decrease user logon times. The Citrix Universal Printer will default print to the client's default printer. However, the behavior can be modified to allow users to select any compatible local or network-based printers. The Citrix Universal Printer is best suited for these scenarios:
    - The user requires access to multiple printers, both local and network-based, which may vary for each session.
    - The user's login performance is a priority, and the Citrix policy "Wait for printers to be created" must be enabled due to application compatibility.
    - The user is working from a Windows-based device or thin client.

> Note:
>
> Other options for provisioning printers, such as Active Directory group policy, "follow-me" centralized print queue solutions, and third-party print management solutions, can be used to provision printers into a Citrix session.

Decision: Printer Drivers

Managing printer drivers in a Citrix environment can be tedious, especially in large environments with hundreds of printers. Fortunately, several methods are available in Citrix Virtual Apps and Desktops to assist with print driver management.

- **User Installed** – When adding a printer within a Citrix Virtual Apps and Desktops session and the native print driver is unavailable, the user can install the drivers manually. Many different print drivers can potentially be installed on different resources, creating inconsistencies within the environment. Troubleshooting printing problems and maintenance of print drivers can become challenging since every hosted resource may install different sets of print drivers. User-installed drivers are not recommended to ensure consistency and simplify support and troubleshooting.
- **Administrator installed** – Administrators will pre-install the necessary print drivers in the VDA image. This ensures consistency and simplifies support and troubleshooting.

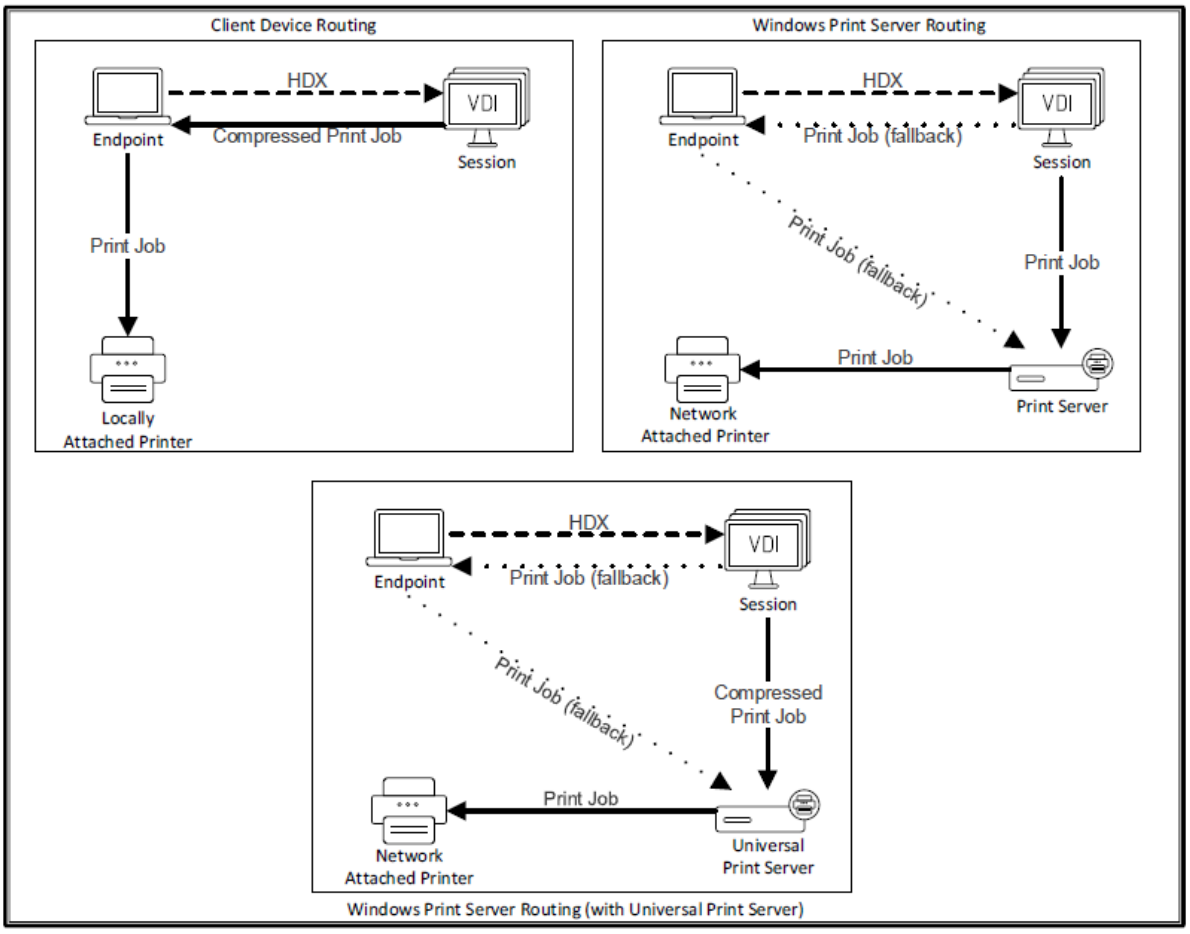
- **Automatic Installation** – When connecting a printer within a Citrix Virtual Apps and Desktops session, a check is made to see if the required print driver is already installed in the operating system. If the print driver is not already installed, the native print driver, if one exists, will be installed automatically. If users roam between multiple endpoints and locations, this can create inconsistencies across sessions since users may access a different hosted resource every time they connect. When this type of scenario occurs, troubleshooting printing problems and maintenance of print drivers can become very challenging since every hosted resource may have different sets of print drivers installed. Automatic installed drivers are not recommended to ensure consistency and simplify support and troubleshooting.
- **Universal Print Drive**r – The Citrix Universal Printer Driver (UPD) is a device-independent print driver designed to work with most printers. The Citrix Universal Printer Driver (UPD) simplifies administration by reducing the number of drivers required on the master image. For auto-created client printers, the driver records the application output and sends it, without any modification, to the end-point device. The endpoint uses local, device-specific drivers to finish printing the job to the printer. The UPD can be used with the Citrix Universal Print Server (UPServer) to extend this functionality to network printers.

Decision: Printer Routing

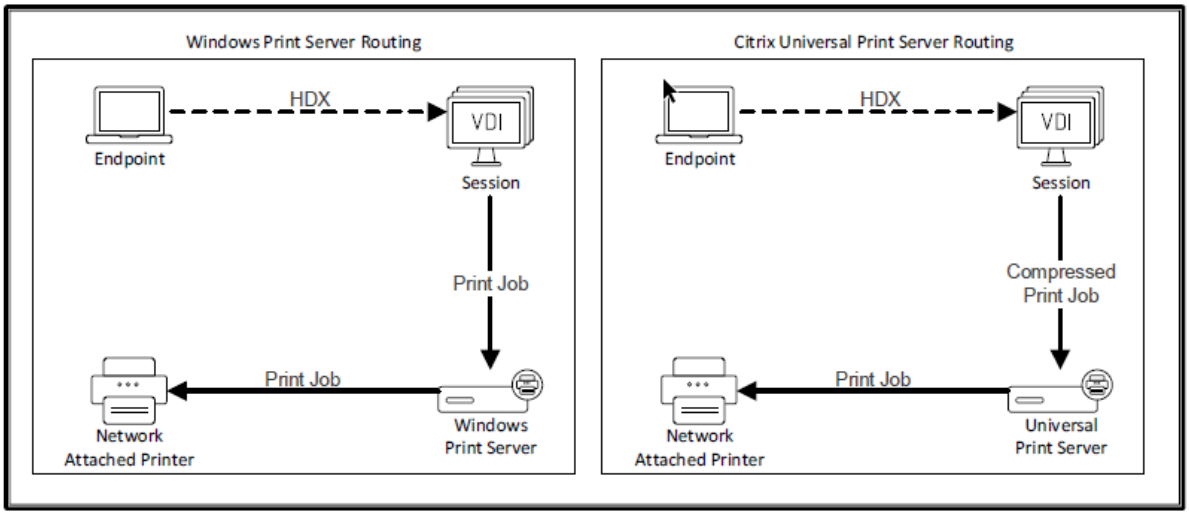
Print jobs can be routed along different paths: through a client device or through a print server.

- **Client Device Routing —** Client devices with locally attached printers (printers attached through USB, LPT, COM, TCP, etc.) route print jobs directly from the client device to the printer.
- **Windows Print Server Routing** – Print jobs sent to auto-created network-based printers will be routed from the user's session to the print server by default. However, the print job will take a fallback route through the client device when any of the following conditions are true:
    - The session cannot contact the print server
    - The print server is on a different domain without a trust established
    - The native print driver is not available within the user's session
- **Citrix Universal Print Server Routing** – Print job routing follows the same process as Windows Print Server Routing except that the Universal Print Driver is used between the user's session and the Citrix Universal Print Server.

The specifics of print job routing are based on the printer provisioning method. Auto-created and user-added printers can route print jobs based on the following diagrams:



However, if the printers are provisioned as session printers, the print job routing options change slightly. The jobs can no longer route through the user's endpoint device.



The recommended option is based on the network location of the endpoint device, the user's

session and the print server

- Client Device Routing
  - Use for locally attached printer implementations.
  - The use of a Windows endpoint device and printer are on the same high-speed, low-latency network as the Windows Print Server.
- Windows Print Server Routing
  - Use if the printer is on the same high-speed, low-latency network as the Windows Print Server and user session.
- Windows Print Server Routing (with Universal Print Server)
  - Use if non-Windows endpoint devices and printers are on the same high-speed, low-latency network as the Windows Print Server.

Decision: Printer Server Redundancy

Network-based printers managed with a Microsoft print server or the Citrix Universal Print Server should be configured with redundancy to eliminate a single point of failure. The Citrix Universal Print Server redundancy should be defined within a Citrix Policy.

Applications

Properly integrating an application requires understanding compatibility and how the user/business requirements influence the appropriate delivery method.

Decision: Compatibility

VDI typically requires significant changes to an organization's application delivery and management strategy. For example, many organizations will take the opportunity to upgrade their desktop operating system and simplify management by reducing the number of applications installed into the base image using techniques such as application streaming and application layering. These are significant changes that require comprehensive compatibility testing. Important compatibility requirements that may need to be verified include:

- **Operating System** - The application must be compatible with the preferred operating system.
- **Multi-User** — Some applications may be more appropriate for delivery via a hosted shared desktop or Windows App. In these situations, the application's compatibility must be verified against the multi-user capabilities of a server operating system like Windows Server 2022.
- **Interoperability** — Some applications may experience complications if they coexist on the same operating system. Possible causes include shared registry hives, dlls, INI files, and incompatible dependencies. Application interoperability issues should be identified so that appropriate remediation steps can be taken or an alternative delivery model can be selected.
- **Dependency** — Applications may need to interact with each other to provide users with a seamless experience. For example, applications that present information in PDF format require a suitable PDF viewer. The dependent (child) applications are often version-specific to the parent application.
- **Application virtualization** — Application virtualization techniques, like streaming and layering, help simplify image management by reducing the number of applications installed in the base image. However, not all applications are suitable for streaming and layering because they may install device drivers, use COM+, or form part of the operating system.

Application compatibility can be achieved by combining manual user testing, utilizing pre-verified lists maintained by the software vendor, or using an automated application compatibility solution, which runs through thousands of tests to verify compatibility.

Decision: Application Delivery Method

It is unlikely that a single delivery method will meet all requirements. Several application delivery methods can be considered based on the outcome of the application categorization assessment process and the overall image management strategy (installed images, scripted images, and layered images).

Choosing one of the appropriate application delivery methods helps improve scalability, management, and user experience.

- **Installed app** — The application is part of the base desktop image. The installation process involves copying dll, exe, and other files to the image drive and modifying the registry.
- **Streamed App (Microsoft App-V)**: The application is profiled and delivered to desktops across the network on demand. Application files and registry settings are placed in a virtual desktop container and isolated from the base operating system and each other, which helps to address compatibility issues. Microsoft App-V is scheduled for end-of-life in April 2026.
- **MSIX App Attach** — MSIX app attach enables you to attach applications from an application package to user sessions dynamically. Applications aren't installed locally on session hosts or images, making creating custom images for your session hosts easier and reducing operational overhead and costs for your organization. Delivering applications with MSIX app attach also gives you greater control over which applications your users can access in a remote session.
- **Layered App (Citrix App Layering)** – Each layer contains a single application, agent, or operating system. Layering simplifies ongoing maintenance, as an OS, agent, and application exist in a single layer; update the layer, and all deployed images containing that layer are updated. App Layering has two different delivery options:
    - Layered Image – An administrator can easily create new, deployable images by integrating one OS layer, one platform layer (Citrix Virtual Apps and Desktops VDA, Provisioning Services agent), and many application layers.
    - Elastic Layer — A Citrix Virtual Apps and Desktops user can dynamically receive a new app layer based on logon. On a Virtual App host, an elastic layer is session-aware, where an attached layer is only available to a user's session if the user is granted access to the layer.
- **Hosted Windows App** - An application installed on a multi-user Virtual Apps host and deployed as an application, not a desktop. A user accesses the hosted Windows app seamlessly from the VDI desktop or endpoint device, hiding the fact that the app executes remotely.
- **Local App** – An application deployed on the endpoint device. The application interface appears within the user's hosted VDI session even though it executes on the endpoint.

This table recommends the preferred approaches for integrating applications into the solution.

App Category	Installed App	Streamed App	MSIX App Attach	Layered App	Hosted Windows App	Local App
Common	✓	○	○	✓	○	X
Departmental	○	✓	✓	✓	✓	X
User	X	○	○	✓	○	✓
Management	✓	X	X	✓	○	X

✓: Recommended / X: Not Recommended / ○: Optional

Virtual Machines

Virtual resources require proper allocation of the processor, memory, and disk. These decisions directly impact the amount of hardware required and the user experience.

The key to successful resource allocation is ensuring that virtual desktops and applications offer similar performance levels to physical desktops. Otherwise, productivity and overall user satisfaction will be affected. However, allocating resources to virtual machines above their requirements is inefficient and expensive for the business.

The resources allocated should be based on the workload characteristics of each user group identified during the assessment phase.

Decision: Virtual Processor (vCPU)

For hosted desktop-based VDI models (hosted pooled desktops and hosted static desktops), the general recommendation is for two or more vCPUs per virtual machine to execute multiple threads simultaneously. Although a single vCPU could be assigned for extremely light workloads, users will likely experience session hangs.

Decision: CPU Optimization

In a shared and virtualized environment, a single user can monopolize CPU resources due to a runaway process or an intense data processing operation in Excel. If the processor is oversubscribed, it cannot fulfill other users' requests, resulting in a hung session.

Citrix Workspace Environment Management, a Citrix Virtual Apps and Desktops component, incorporates CPU optimization. When a process consumes a certain percentage of the CPU over a defined timeframe, the process priority lowers from normal to low or very low, giving all remaining processes a higher priority and overcoming the runaway process risk. CPU optimization will also remember processes that triggered CPU protection and automatically start the process at a lower priority on future launches.

Most environments should enable CPU optimization as a default configuration.

Decision: Virtual Memory (vRAM)

The amount of memory allocated to each resource is a function of the user's expected workload and application footprint. Assigning insufficient memory to the virtual machines will cause excessive paging to disk, resulting in a poor user experience; allocating too much RAM increases the overall cost of the solution.

This table provides guidance on the virtual RAM that should be assigned based on workload.

User Workload	Operating System	vRAM Configured for Scale
Light	Windows 10	4 GB
	Windows 11	4 GB
	Windows Server 2022	256 MB per user
Medium	Windows 10	8 GB
	Windows 11	8 GB
	Windows Server 2022	640 MB per user
Heavy	Windows 10	16 GB
	Windows 11	16 GB
	Windows Server 2022	1024 MB per user

Note:

Windows Server 2022 recommendations are based on the hosted Windows app and hosted shared desktop VDI model.

If used, the Machine Creation Services and Citrix Provisioning cache in RAM should be added to the virtual machine RAM specifications.

Decision: RAM Optimization

Even though users only work within a single application at a time, most have five or more applications running idle. When a process moves from active to idle, the application and operating system release a portion of the process's active working set of memory to free up system resources. However, this is only a small percentage of the applications working set. The rest remains locked for the application, severely limiting available system resources.

Using RAM Optimization within Citrix Workspace Environment Management, idle applications (have not been interacted with by a user) for a certain time are forced to release excess memory until they are no longer idle. When the application returns to an active state, the released memory is loaded back into the active working set.

Most environments should enable RAM optimization as a default configuration. If certain processes encounter issues with optimization, a RAM optimization exclusion list is available.

Decision: Disk Cache

The amount of storage that each VM requires will vary based on the workload and the image type. If creating a hosted personal desktop without leveraging an image management solution, each VM will require enough storage for the entire OS and locally installed applications.

Deploying machines through Machine Creation Services or Citrix Provisioning can substantially reduce the storage requirements for each virtual machine. Disk space requirements for the write cache, the temporary storage locations for write operations performed by the virtual machines, and difference disk, which captures any changes made to the virtual machine, such as user-installed applications or saved files, will depend on application usage and user behavior. However, the following table provides a starting point for estimating disk space requirements based on a machine sized with vCPU and vRAM per the earlier guidelines.

User Workload	Operating System	Storage Space (Differencing Disk/ Write Cache Disk)
Light	Windows 10	10 GB
	Windows 11	10 GB
	Windows Server 2022	20 GB
Medium	Windows 10	15 GB
	Windows 11	15 GB
	Windows Server 2022	60 GB
Heavy	Windows 10	20 GB
	Windows 11	20 GB
	Windows Server 2022	60 GB

Decision: RAM Cache

Provisioning Services and Machine Creation Services can utilize a portion of the virtual machine's RAM as a buffer for the storage cache. The RAM cache improves the performance of traditional storage by sharing the virtual machine's non-paged pool memory.

However, the following table provides a starting point for estimating RAM cache requirements based on a machine sized with vCPU and vRAM, per the earlier guidelines.

User Workload	Operating System	RAM Cache Configured for Scale	RAM Cache Configured for Experience
Light	Windows 10	128 MB	256 MB
	Windows 11	128 MB	256 MB
	Windows Server 2022	4 GB	4 GB
Medium	Windows 10	256 MB	512 MB
	Windows 11	256 MB	512 MB
	Windows Server 2022	8 GB	8 GB
Heavy	Windows 10	512 MB	1024 MB
	Windows 11	512 MB	1024 MB
	Windows Server 2022	10 GB	10 GB

Note:

The Machine Creation and Provisioning Services cache in RAM should be added to the virtual machine RAM specifications if used. Additionally, if additional RAM is available on the host, the RAM Cache amounts can be increased to provide even greater performance levels.

Decision: Storage IOPS

Storage performance is limited by the number of operations it can handle per second, referred to as IOPS. Under-allocating storage IOPS results in a VDI desktop with slow-loading apps, web pages, and data.

The following table provides starting guidance on the number of storage IOPS generated per user based on workload and operating system. Storage IO activity will be higher during user logon/logoff.

User Workload	Operating System	Storage IOPS (without RAM-Based Cache)	Storage IPS (with RAM-Based Cache)
Light	Windows 10	12 IOPS	1 IOPS
	Windows 11	12 IOPS	1 IOPS
	Windows Server 2022	4 IOPS	0.5 IOPS
Medium	Windows 10	20 IOPS	1 IOPS
	Windows 11	20 IOPS	1 IOPS
	Windows Server 2022	6 IOPS	1 IOPS
Heavy	Windows 10	35 IOPS	3 IOPS
	Windows 11	35 IOPS	3 IOPS
	Windows Server 2022	8 IOPS	1 IOPS

Decision: IO Prioritization

With shared environments, every user's IO process receives equal resources. A user running an IO-intensive task can affect mission-critical applications. Citrix Workspace Environment Management allows administrators to define IO priorities for processes.

If a process requires more IO resources or monopolizes IO resources, the process and process priority can be manually increased or decreased via the console.

Decision: Graphics (GPU)

The CPU renders graphical processing with software without a graphical processing unit (GPU). A graphical processing unit (GPU) can be leveraged to improve server scalability and user experience or enable graphically intensive applications. During the desktop design, deciding how the GPU (if used) will be mapped to the virtual machines is important. There are three methods available.

- **Pass-Through GPU –** Each physical GPU is passed through to a single virtual machine (hosted apps or desktops).
- **Hardware Virtualized GPU –** Using a hypervisor's vGPU technology (NVIDIA, Intel, AMD) the GPU is virtualized and shared between multiple machines. Each virtual machine has the full functionality of GPU drivers and direct access to the GPU.
- **Software Virtualized GPU —** The hypervisor manages the GPU and intercepts requests made by the VDI desktops. This process is used if a GPU is not installed within the host.

Hypervisor/Hyperscale	Pass-through GPU	Hardware Virtualized GPU (NVIDIA)	Hardware Virtualized GPU (Intel)	Hardware Virtualized GPU (AMD)	Software Virtualized GPU
XenServer 8	✓	✓	X	X	✓
Citrix Hypervisor 8.2 LTSR	✓	✓	✓	✓	✓
Microsoft Hyper-V	✓	X	X	X	✓
VMware ESX	✓	✓	✓	✓	✓
Nutanix AHV	✓	✓	X	X	X
Microsoft Azure	✓	✓	✓	✓	X
Amazon Web Services	✓	✓	X	✓	X
Google Cloud Platform	✓	X	X	X	X

✓: Available / X: Not Supported

User groups that heavily use graphical applications will often require a hardware-virtualized GPU.

Layer 4: The Control Layer

Active Directory

Decision: Forest Design

By default, multi-forest deployments do not have inter-domain trust relationships between the forests. An AD administrator can establish trusting relationships between multiple forests, allowing users and computers from one forest to authenticate and access resources from another.

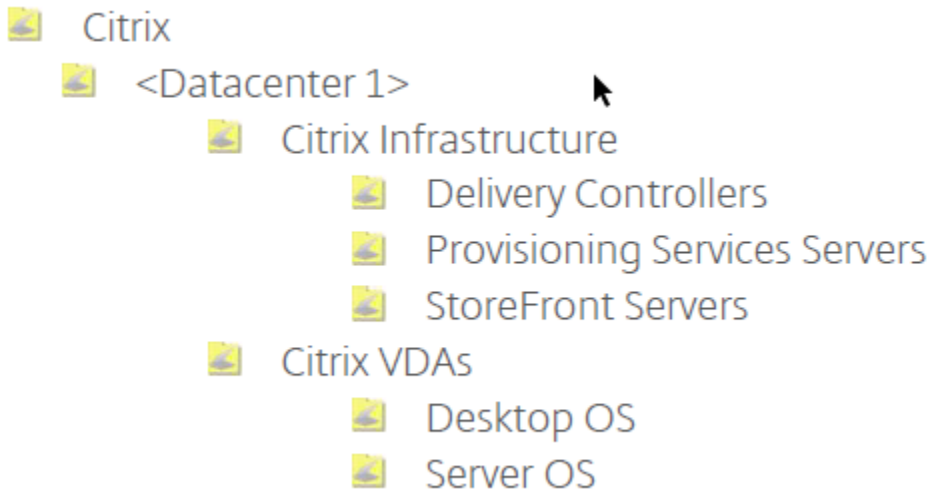
For forests that have inter-domain trusts, it is recommended that the appropriate settings be configured to allow the Delivery Controllers to communicate with both domains. Citrix Virtual Apps and Desktops sites for each forest must be configured and deployed when the appropriate trusts are not configured. This section outlines the requirements and calculates the size needed to successfully deploy Citrix Virtual Apps and Desktop environments.

For more information about deploying Citrix Virtual Apps and Desktops in a complex Active Directory environment, please refer to the Citrix Virtual Apps and Desktops product documentation.

Decision: Organization Unit Structure

The infrastructure components for a Citrix Virtual Apps and Desktops deployment should reside within their own dedicated organizational units (OUs), separating VDAs and infrastructure servers for management purposes. By having their own OUs and enabling block inheritance for the OU, the objects inside will have greater flexibility with their management while allowing Citrix administrators to be granted delegated control over the appropriate OU.

A sample Citrix OU structure can be seen here:



Decision: User Groups

Whenever possible, permissions and authorization should be assigned to user groups rather than individual users, eliminating the need to edit many resource permissions and user rights when creating, modifying, or deleting user accounts.

Permission application example:

- An application published to one group of 1,000 users requires the validation of only one object for all 1,000 users.
- The same application published to 1,000 individual user accounts requires the validation of all 1,000 objects.

User groups can be set along application users, business units, or use cases - as long as they can be identified easily.

Database

The majority of Citrix products discussed within this document require a database. Citrix Virtual Apps and Desktops requires three SQL server databases:

- **Site:** (also known as site configuration) stores the running site configuration, plus the current session state and connection information.
- **Logging:** This database (Configuration Logging) stores information about site configuration changes and administrative activities. It is used when the configuring logging feature is enabled (default = enabled).
- **Monitoring:** stores data used by Director, such as session and connection information.

Refer to the Citrix Virtual Apps and Desktops product documentation for additional information on the database requirements and how to provide high availability.

Decision: Edition

There are multiple editions of Microsoft SQL Server: Enterprise, Standard, Web, Developer, and Express. Based on the capabilities of the various SQL Server editions available, Standard or Enterprise editions of SQL are used to host the Citrix Virtual Apps and Desktops databases in production environments.

SQL Express should not be used for production databases due to its lack of supportability, scalability, and performance.

The Standard or Enterprise editions of SQL provide adequate features to meet most environments' needs. For more information on the databases supported by Citrix products, please refer to the Citrix Database Support Matrix. Different versions of Citrix products support different versions of the SQL server; therefore, it is important to check the support matrix to ensure the version of SQL server used is compatible with the Citrix product being deployed.

Decision: Database Server Sizing

The SQL server must be sized correctly to ensure an environment's performance and stability. Since every Citrix product uses the SQL server differently, no generic, all-encompassing sizing recommendations can be provided. Instead, per-product SQL server sizing recommendations are provided below.

Citrix Virtual Apps and Desktops Delivery Controllers use the database as a message bus for broker communications, storing configuration data, and monitoring and configuration log data. Databases are constantly being used, and the performance impact on the SQL server can be considered high.

Based on results from Citrix internal scalability testing, the following SQL server specifications for a server hosting all Citrix Virtual Apps and Desktops databases are recommended:

- 4 Cores / 4 GB RAM for environments up to 5,000 users
- 4 Cores / 8 GB RAM for environments up to 15,000 users
- 8 Cores / 16 GB RAM for environments with 15,000+ users

The database files and transaction logs should be hosted on separate hard disk subsystems to cope with many transactions. For example, registering 20,000 virtual desktops during a 15-minute boot storm causes ~500 transactions/second, and 20,000 users logging on during a 30-minute logon storm causes ~800 transactions / second on the Virtual Apps and Desktops Site database.

Citrix Provisioning — Besides static configuration data provisioning servers, database servers store runtime and auditing information. Depending on the boot and management pattern, the database's performance impact can be considered low to medium.

Based on this categorization, a SQL server specification of 4 Cores and 4 GB RAM is recommended as a good starting point. The SQL server should be carefully monitored during the testing and pilot phase to determine its optimal configuration.

Decision: Instance Sizing

When sizing a SQL database, two aspects are important:

- Database file – Contains the data and objects such as tables, indexes, stored procedures, and views stored in the database.
- Transaction log file – Contains a record of all transactions and database modifications made by each transaction. The transaction log is a critical component of the database, and if there is a system failure, it might be required to bring the database back to a consistent state. The usage of the transaction log varies depending on which database recovery model is used:
    - Simple recovery — No log backups are required. Log space is automatically reclaimed to keep space requirements small, essentially eliminating the need to manage the transaction log space. Changes to the database since the most recent backup are unprotected. In the event of a disaster, those changes must be redone.
    - Full recovery requires log backups. No work is lost due to a lost or damaged database data file. Data from any arbitrary point in time can be recovered (for example, before application or user error). Full recovery is required for database mirroring.
    - Bulk-logged – Requires log backups. This is an adjunct of the full recovery model that permits

For further information, refer to Microsoft SQL Server Recovery Models.

Understanding the disk space consumption for common database entries is important for estimating storage requirements. This section outlines the storage requirements on a per-product basis and provides sizing calculations.

Citrix Virtual Apps and Desktops uses three distinct databases:

- The **Site** database contains the static configuration and dynamic runtime data.
- The **Monitoring** database contains monitoring data, such as connection and session information, accessible via Director.
- The **Logging** database (accessible via Studio) records each administrative change and activity performed within the site.

Site Database

Since the database of a Citrix Virtual Apps and Desktops site contains static configuration data and dynamic runtime data, the size of the database file depends not only on the physical size of the environment but also on user patterns. The following factors all impact the size of the database file:

- The number of connected sessions
- The number of configured and registered VDAs
- The number of transactions occurring during logon
- VDA heartbeat transactions

Determining the size of the transaction log for the Site database is difficult due to factors that can influence the log, including:

- The SQL Database recovery model
- Launch rate at peak times
- The number of desktops being delivered

During Virtual Apps and Desktops scalability testing, Citrix observed the transaction log growth rate at 3.5MB an hour when the system is idle and a ~32 KB per-user-per-day growth rate. In a large environment, transaction log usage requires careful management and a regular backup to prevent excessive growth. This can be achieved using scheduled jobs or maintenance plans.

Monitoring Database

The Monitoring database is expected to be the largest of the three databases since it contains historical information about the site. Its size is dependent on many factors, including:

- Number of Users
- Number of sessions and connections
- Number of workers
- Retention period configuration
- Number of transactions per second. The monitoring service tends to execute updates in batches. It is rare for the number of transactions per second to go above 20.
- Regular consolidation calls from the Monitoring service cause background transactions.
- Overnight processing is carried out to remove data outside the configured retention period.

The transaction log size for the Monitoring Database is very hard to estimate, but scalability testing showed a growth rate of about 30.5 MB an hour when the system is idle and a per-user per day growth rate of ~9 KB.

Logging Database

The Logging Database is typically the smallest of the three databases. Its size and the size of the related transaction log depend on the daily administrative activities initiated by Studio, Director, or PowerShell scripts. Therefore, its size is difficult to estimate. The more configuration changes are performed, the larger the database will grow. Some factors that can affect the size of the database include:

- The number of actions performed in Studio, Director, and PowerShell.
- Minimal transactions occur on the database when no configuration changes occur.
- The transaction rate during updates. Updates are batched whenever possible.

- Data is manually removed from the database. Data within the Configuration Logging Database is not subject to any retention policy. Therefore, it is not removed unless done so manually by an administrator.
- Activities that impact sessions or users include session logoff and reset.
- The mechanism used for deploying desktops.

In environments not using Machine Creation Services (MCS), the database size tends to fall between 50 and 100 MB. For MCS environments, database size can easily exceed 200MB due to logging all VM build data.

Temporary Database

In addition to the Site, Monitoring, and Configuration Logging databases, SQL Server provides a system-wide temporary database (tempdb) to store Read-Committed Snapshot Isolation data. Citrix Virtual Apps and Desktops use this SQL Server feature to reduce database lock contention. **Citrix recommends that all databases use Read-Committed Snapshot Isolation**.

The size of the tempdb database will depend on the number of active transactions, but it is generally not expected to grow more than a few MBs. The performance of the tempdb database does not impact the performance of Citrix Virtual Apps and Desktops brokering, as any transactions that generate new data require tempdb space. Citrix Virtual Apps and Desktops tend to have short-lived transactions, which help keep the size of the tempdb small.

The tempdb is also used when queries generate large intermediate result sets. Guidance and sizing of the tempdb can be found in this Microsoft article.

Citrix Provisioning

The Citrix Provisioning farm database contains static configuration and configuration logging (audit trail) data. Please review the Database Sizing section of the Citrix Provisioning pre-install documentation for all the details required for proper database sizing.

During the Citrix Provisioning farm setup, a database with an initial file size of 20MB and a growth size of 10 MB is created. The database log's initial size is 10 MB and a growth size of 10%. Due to the nature of the data in the Provisioning farm database, the transaction log is not expected to grow very quickly unless a large amount of configuration is performed.

In contrast to Citrix Virtual Apps and Desktops, which also offer the ability to track administrative changes, the related information is not written in a dedicated database but directly in the Provisioning farm database. **It is recommended that the audit trail data be archived regularly to limit the size of the Provisioning database**.

Decision: Database Location

The Logging and Monitoring databases are in the Site Configuration database by default. Citrix recommends changing the location of these secondary databases as soon as the site configuration has been completed to simplify sizing, maintenance, and monitoring. All three databases can be hosted on the same server or on different servers. Ideally, these SQL databases are provided enough resources to ensure reliable performance and data integrity for the Citrix environment. For more information, please refer to Change Citrix database locations.

> Note:
>
> You cannot change the location of the configuration logging database when mandatory logging is enabled.

Decision: High-Availability

This table highlights the impact of Citrix Virtual Apps and Desktops and Citrix Provisioning during a database outage.

Component	Impact of Database Outage
Site database	Users cannot connect or reconnect to a virtual desktop unless Local Host Cache is enabled. Local Host Cache allows users to reconnect to their applications and desktops even when the site database is unavailable.
Monitoring database	Director will not display historical data, and Studio cannot be started. Brokering of incoming user requests and existing user sessions will not be affected.
Logging database	If allow changes are allowed when the database is disconnected and has been enabled within Citrix Virtual Apps and Desktops logging preferences, an outage of the configuration logging database will have no impact (other than configuration changes not being logged). Otherwise, administrators cannot make any changes to the site configuration. Users are not impacted.
Provisioning database	When offline database support is enabled and the database becomes unavailable, the stream process uses a local copy of the database to retrieve information about the provisioning server and the target devices supported by the server. This allows provisioning servers and the target devices to remain operational. However, when the database is offline, the console and the management functions become unavailable: Auto Add target devices, vDisk creation and updates, Active Directory password changes, Stream process startup, Image update service, and PowerShell and MCLI-based management. If offline database support is not enabled, all management functions become unavailable, and the boot and failover of target devices will fail, impacting end users.

Besides the built-in database redundancy options, Microsoft SQL Server and the underlying hypervisor (in virtual environments) offer many high-availability features. These enable administrators to ensure single server outages will have a minimal impact (if any) on the Citrix Virtual Apps and Desktops infrastructure. The following SQL/Hypervisor high-availability features are available:

- **AlwaysOn Failover Cluster Instances** – Failover clustering provides high-availability support for an entire instance of Microsoft SQL Server. A failover cluster combines two or more nodes or servers using a shared storage. A Microsoft SQL Server AlwaysOn Failover Cluster Instance appears on the network as a single computer but has functionality that provides failover from one node to another if the current node becomes unavailable. The transition from one node to the other node is seamless for the clients connected to the cluster. AlwaysOn Failover cluster Instances require a Windows Server Failover Clustering (WSFC) resource group. The number of nodes supported in the WSFC resource group will depend on the SQL Server edition.
- **AlwaysOn Availability Groups** – AlwaysOn Availability Groups is an enterprise-level, high-availability disaster recovery solution that enables administrators to maximize the availability of one or more user databases. AlwaysOn Availability Groups require the Microsoft SQL Server instances to reside on Windows Server failover clustering (WSFC)

nodes. Like failover clustering, a single virtual IP/network name is exposed to the database users. In contrast to failover clustering, shared storage is not required since the data is transferred using a network connection. Both synchronous and asynchronous replication to one or more secondary servers is supported. Compared to clustering, secondary servers can actively process incoming read-only requests, backups, or integrity checks. This feature can offload user resource enumeration requests to a secondary SQL server in Citrix environments to scale out an SQL server infrastructure. Since the data on active secondary servers can lag multiple seconds behind the primary server, the read-only routing feature cannot be used for other Citrix database requests at this point in time.

This table outlines the recommended high-availability features for Citrix databases.

Database Component	AlwaysOn Failover Cluster	AlwaysOn Availability Groups
Site	○	✓
Logging	○	✓
Monitoring	○	✓
Provisioning	○	X
Session Recording	○	○

✓: Recommended / X: Not Supported/ ○: Viable

Citrix Licensing

Citrix offers four platform subscription-based licensing options, with the best Citrix and NetScaler functionality tailored to fit your environment.

- Citrix Universal Hybrid Multi-Cloud — This subscription includes Citrix Virtual Apps and Desktops Premium, Citrix DaaS Premium, 1000 GB of NetScaler throughput with unlimited instances, and Citrix Endpoint Management. It also includes deploying sites and VDAs on the public cloud.
- Citrix Platform License — This subscription, available by invitation only, includes Citrix Virtual Apps and Desktops Premium, Citrix DaaS Premium, unlimited NetScaler instances and capacity, Citrix Secure Private Access, Citrix Endpoint Management, Citrix Analytics for Performance & Security, and uberAgent.
- Citrix for Private Cloud – This subscription is for fully on-premises environments only.
- NetScaler Fixed Capacity - This subscription is for stand-alone fixed-capacity NetScaler instances.

This table provides additional details about what is included in each Citrix subscription:

Citrix Feature	NetScaler Fixed Capacity	Citrix for Private Cloud	Citrix Universal Hybrid Multi-Cloud	Citrix Platform License
IT Managed CR/LTSR App and Desktop Control Plane	X	✓	✓	✓
HDX, Policy Control, Adaptive Authentication, StoreFront	X	✓	✓	✓
Hybrid Multi-Cloud Workload Support	X	X	✓	✓
Citrix Managed Cloud Control Plane	X	X	✓	✓
Mobile Device Management	X	X	✓	✓
NetScaler Application Delivery and Security	✓	X	✓	✓
NetScaler app and API Observability / NetScaler Console	X	X	✓	✓
Security and Performance Insights with WIEM Integration	X	X	X	✓
Secure Access to Web and SaaS Apps	X	X	X	✓
Enterprise-wide deployment	X	X	X	✓

✓: Included / X: Not Included

For more information on Citrix licensing, refer to our product terms. For more information on the features included in each subscription, please view our feature matrix.

Decision: Sizing

Internal scalability testing has shown that a single virtual license server with two cores and 8 GB of RAM can issue approximately 50-60 licenses per second. Citrix recommends you maintain 50,000 or fewer concurrent connections per License Server. If necessary, the specification of the license server can be scaled up to support a higher number of license requests per second.

Decision: High-Availability

For a typical environment, a single license server is sufficient. Should the license server become unavailable, dependent Citrix products will enter a 30-day grace period, which provides more than enough time to resolve connectivity issues and/or restore or rebuild the license server.

If the license server and the Citrix product do not communicate within 2 heartbeats (5-10 min), the Citrix product will enter a grace period and allow connections up to 30 days. Once communication with the license server is re-established, the license server will reconcile the temporary and actual licenses.

Citrix supports clustering for the license server if additional redundancy is required. Clustering the License Server lets users continue working during failure, and users cannot detect when one server in a cluster fails over another.

For more information, refer to Citrix Licensing Technical Overview.

Delivery Controllers

Decision: Server Sizing

Controller scalability is based on CPU utilization. The more processor cores available, the more a controller can support virtual desktops. Each VDA startup, registration, enumeration, and launch request impacts the controller's processor. As the storm increases in intensity, the controller's CPU utilization will increase. If the CPU reaches a critical threshold of roughly 80%, the site must either scale up or out.

Adding additional CPU cores to a Controller will lower the overall CPU utilization, thus allowing for greater numbers of desktops supported by a single controller. This is only feasible when dealing with virtualized controllers, as adding virtual CPUs is fairly straightforward. The other alternative is to add another controller to the site configuration. The controller would have the same configuration as other controllers, and the load would be evenly distributed across all controllers, thus helping to reduce the overall load on each single controller.

Testing has shown that a single Delivery Controller, with Local Host Cache enabled, can support more than 10,000 desktops using the following configuration.

Component	Specification
Processor	4 vCPU
Memory	8 GB RAM
Network	10 GBps networking
Host Storage	40GB Shared storage
Operating System	Windows Server 2022
Citrix Virtual Apps and Desktops	2402 LTSR

The following formula can calculate the number of Delivery Controllers required for a Citrix site.

$$\text{Number of Controllers} = (\text{Number of Active Sessions per Site}/10{,}000) + 1$$

Decision: High Availability

If the server hosting the Controller is unavailable, users cannot access their virtual desktops or published applications. Therefore, at least two Controllers (N+1 redundancy) should be deployed per zone on different physical servers to prevent this component from becoming a single point of failure. The others can manage connections and administer the site if one controller fails.

The locations of all Controllers are specified on the VDA, allowing it to automatically failover if communication with one controller is unavailable. The VDA checks the following locations in order, stopping at the first place it finds the Controller:

1. A persistent storage location is maintained for the auto-update feature. This location contains controller information when auto-update is enabled and after the VDA successfully registers for the first time after installation.
2. The VDA checks the following locations for its initial registration after installation or when auto-update is disabled.
3. Policy settings (Controllers, Controller SIDs).
4. The Controller information under the VDA ListofDDCs registry key. The VDA installer initially populates these values based on the information specified when installing the VDA.
5. OU-based discovery. This is a legacy method maintained for backward compatibility.
6. The Personality.ini file created by Machine Creation Services.

Citrix recommends utilizing the policy setting or managing the ListOfDDCs registry key to ensure VDAs connect to the correct Delivery Controllers. This feature will simplify the environment's management by keeping VDAs updated when adding and removing Controllers.

Decision: Local Host Cache

Even if the SQL database is highly available, there is the risk of not having access to the database if the network connection between the delivery controller and the SQL database fails, which is an important concern for sites that span geographical locations. To overcome this risk, the Delivery Controllers can utilize the Local Host Cache feature that creates a local copy of the SQL database, which can be used only if the Delivery Controller loses contact with the database.

The following must be considered when using Local Host Cache:

- Elections — When the zones lose contact with the SQL database, an election nominates a single Delivery Controller as master. All remaining controllers go into idle mode. The winner of the election is determined in alphabetical order.
- Sizing — When using LHC mode, a single Delivery Controller is responsible for all VDA registrations, enumerations, launches, and updates. The elected controller must have enough resources (CPU and RAM) to handle the entire load for the zone. A single controller can scale to 10,000 users, influencing the zone design.
    - RAM – The LHC services can consume 2+GB of RAM depending on the duration of the outage and the number of user launches during the outage.
    - CPU — LHC can use up to 4 cores in a single socket. Because of this, maximizing the number of cores per socket (e.g., multiplying 4 cores by 1 socket if using 4 vCPU) is recommended.

    - Storage – During LHC mode, storage space increased 1MB every 2-3 minutes, averaging 10 logons per second.
- Power Options – Powered-off virtual resources will not start when the Delivery Controller is in Local Host Cache mode. Pooled virtual desktops that reboot at the end of a session are placed into maintenance mode. To override this default behavior, the following PowerShell commands must be run:

  Site Wide:

  Set-BrokerSite -ReuseMachinesWithoutShutdownInOutageAllowed $true

  For each affected Delivery Group, run the following PowerShell command:

  Set-BrokerDesktopGroup -Name "name" -ReuseMachinesWithoutShutdownInOutage $true

  To enable the Delivery Group setting by default, run the following PowerShell command:

  Set-BrokerSite -DefaultReuseMachinesWithoutShutdownInOutage $true

- Consoles – Studio and PowerShell are unavailable when using LHC mode.

It is recommended to test LHC mode to ensure that everything is working correctly. You can do this by forcing LHC mode. Our Tech Paper has more information about Local Host Cache.

Decision: XML Service Encryption

In a typical session, the StoreFront server passes credentials to the Citrix XML Service on a Controller. The Citrix XML protocol exchanges all data using clear text, except for passwords transmitted using obfuscation.

If the traffic between the Storefront servers and the Controllers can be intercepted, it will be vulnerable to the following attacks:

- Attackers can intercept the XML traffic and steal resource set information and tickets.
- Attackers with the ability to crack the obfuscation can obtain user credentials.
- Attackers can impersonate the Delivery Controller and intercept authentication requests.

For most organizations, Citrix XML traffic will be isolated on a dedicated physical or virtual datacenter network, making interception unlikely. However, for security, it is recommended to use SSL encryption to send StoreFront data over a secure HTTPS connection.

Decision: Server OS Load Management

Default Load Management policies are applied to all Server OS delivery groups. The default settings specify the maximum number of sessions a server can host at 250 and do not consider CPU and Memory usage. Capping session count does not provide a true indication of load, which can lead to an overburdening of Server OS delivery groups, resulting in a degradation of performance or an underutilization of Server OS delivery groups, resulting in an inefficient usage of resources.

Citrix recommends creating unique “custom” Load Management policies for each Delivery Group based on performance and scalability testing. Depending on the different resource bottlenecks identified during testing, different rules and thresholds can be applied to each Delivery Group. For more information on the available load management policy configurations, click here. Refer to Load management policy settings.

If adequate testing cannot be performed before production, Citrix recommends implementing the following "custom" Load Management policy, which can be applied to all servers as a baseline:

- CPU Usage - Full Load: 80%
- CPU usage excluded process priority – Below Normal or Low
- Memory Usage - Full Load: 80%
- Memory Usage base load – Report zero load (MBs): 786
- Concurrent logon tolerance – 2
- Maximum number of sessions – X

The "Maximum number of sessions" policy is included for capping purposes – this is considered a best practice for resiliency. Organizations can choose an initial value of 250 (denoted by "X" above). It is highly recommended that this value and others be customized based on the results from scalability testing.

Cloud Connector

Citrix DaaS within Citrix Cloud utilizes a set of services contained within the Citrix Cloud Connector. These services allow communication between the VDAs, StoreFront (if applicable), and the cloud-based Delivery Controllers. Redundant Cloud Connector virtual machines must be placed in each data center/resource location containing VDA hosts.

Traffic from the Cloud Connectors to Citrix Cloud is encrypted by default. Similar to a delivery controller, it is recommended that traffic between the Cloud Connectors and the other infrastructure components be encrypted. Cloud Connectors also have the same Local Host Cache considerations as Delivery Controllers.

Decision: Server Sizing

Cloud Connector scalability is based on CPU utilization. The more processor cores available, the more virtual desktops a cloud connector can support. Each desktop startup, registration, enumeration, and launch request affects the cloud connector's processor. The cloud connector's CPU utilization will increase as the storm intensifies. If the CPU reaches a critical threshold of roughly 80%, the site must either scale up or out.

Citrix internal testing has shown that using the following configuration, with Local Host Cache enabled, a single resource location with one Cloud Connector can support up to 10,000 VDAs.

Component	Specification
Number of VMs (with N+1 Fault Tolerance)	3
Processors per VM	4 vCPU
Memory per VM	8 GB RAM

> Note:
>
> Citrix recommends three Cloud Connectors in each resource location to maintain a highly available connection to Citrix Cloud during Cloud Connector updates.

If you are using Citrix Workspace and Citrix Gateway services, it is recommended that you implement the Rendezvous protocol to reduce the traffic handled by the Cloud Connector. Rendezvous enables the VDAs to talk to Citrix Cloud directly, thus bypassing the Cloud Connectors. However, this does require the VDAs to have access to certain Citrix Cloud URLs via the Internet.

Cloud Connector Topology

Cloud Connectors are needed in each Resource Location in your site. Resource Locations often represent different data centers, hardware, or public cloud subscriptions. For example, if you had two on-premises data centers and one Azure subscription, you would need at least three Resource Locations.

There are limits to the number of domains, machines, sessions, and hosting connections that can be supported in a single Resource Location. If your Resource Location exceeds those limits, it will need to be split into multiple Resource Locations within Citrix Cloud.

Connector Appliance for Cloud Services

The Connector Appliance is also used to connect your environment to Citrix Cloud. The Connector Appliance is a Linux-based appliance that does not broker DaaS connections. Instead, it provides the following functions:

- Connecting Active Directory to Citrix Cloud, enables AD management, allowing the use of AD forests and domains within your resource locations. It removes the need to add any additional AD trusts.
- Image Portability Service -simplifies the management of images across platforms. This feature is useful for managing images between an on-premises resource location and one in a public cloud. The Citrix Virtual Apps and Desktops REST APIs can be used to automate the administration of resources within a Citrix Virtual Apps and Desktops site.
- Citrix Secure Private Access - enables administrators to provide a cohesive experience that integrates single sign-on, remote access, and content inspection into a single solution for end-to-end access control.

The Connector Appliance platform is part of Citrix Cloud Platform and Citrix Identity Platform and can process data, including the following information:

- IP addresses or FQDNs
- Device, user, and resource location identifiers
- Timestamps
- Event data
- User and group details from Active Directory (for example, used for authenticating and searching for users and groups)

Citrix Provisioning

Citrix Provisioning uses streaming technology to simplify the deployment of virtual and physical machines. Computers are provisioned and re-provisioned in real-time from a single shared disk image. In doing so, administrators can completely eliminate the need to manage and patch individual systems. Instead, all image management is performed on the master image.

Decision: Platform

Citrix Provisioning is supported on various platforms, including hypervisors and hyperscalers. For on-premises hypervisors managed by the customer, Citrix is committed to supporting VMware, Nutanix, Microsoft, and XenServer. This support article covers more details about the supported versions.

Citrix Provisioning is also supported on Azure and GCP. The next section will discuss public cloud considerations in greater detail.

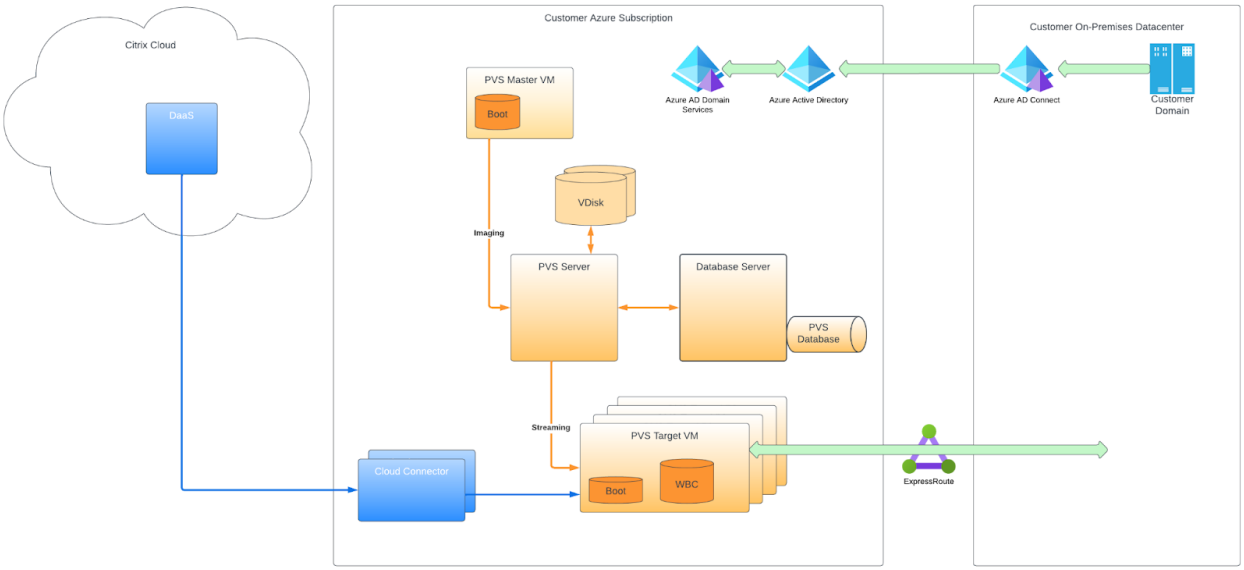
Decision: Public Cloud

When building out a PVS deployment in the public cloud, there are some additional considerations.

Microsoft Azure:

Knowing the subscription limits is important when building out deployments in Azure. A single subscription can only create 5,000 VMs, so a hub-and-spoke model will be required if your PVS farm exceeds that limit. Additionally, ensure that the subnet sizes are large enough to scale for the PVS farm. You can find more deployment guidance in our Reference Architecture.

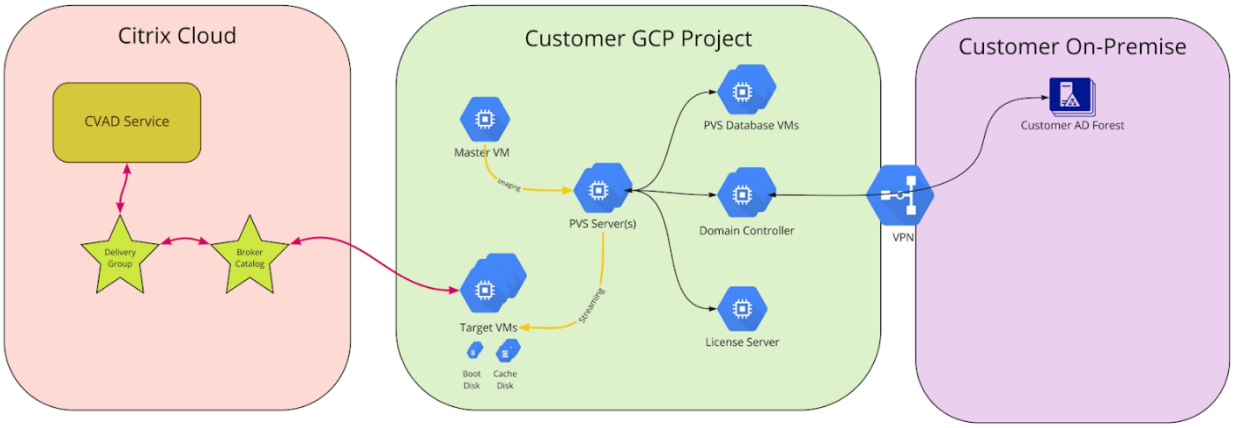
PVS In Azure Architecture with DaaS



There are some limitations to using PVS on Azure. For example, BDM boot is the only boot option that is supported. If you plan to use Azure File Services to provide storage for vDisks, you must create a Premium Storage Account. Additionally, the PVS API is not supported, and there are GUI and wizard limitations. From a VM perspective, you must create template VMs in each Azure region you wish to use. At this time, only standard SSD is supported. You can find other requirements in our product documentation.

Google Cloud Platform

When building out deployments in GCP, it's important to know the subscription limits. A single project can only create 3000 VMs, so a hub-and-spoke model will be required if your PVS farm exceeds that limit. Additionally, ensure that the subnet sizes are large enough to scale for the PVS farm.

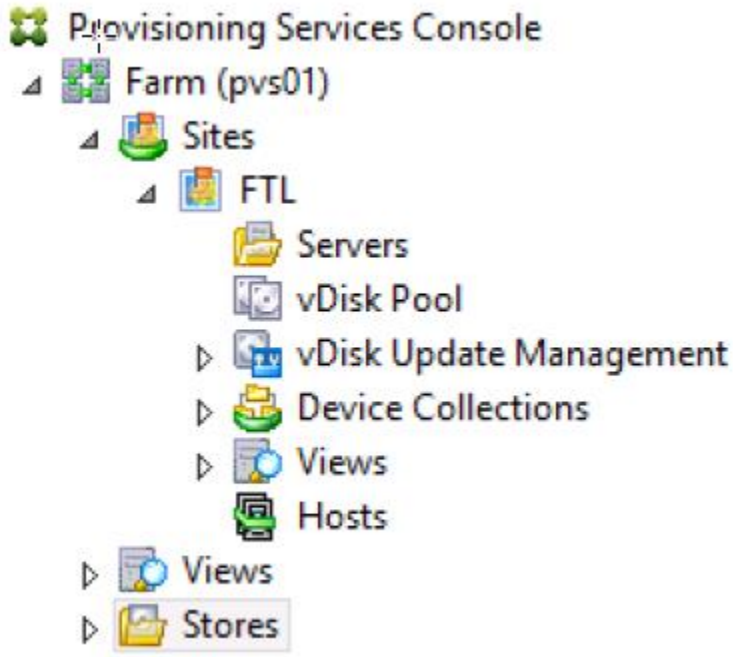


It is important to note that PVS in GCP supports only server operating systems. Windows 10, Windows 11, and sole tenant nodes are not supported. Additionally, power management of target devices cannot be done via the PVS console. You can find other requirements in our product documentation.

Decision: Topology

A Citrix Provisioning farm represents the top level of the Provisioning infrastructure, which can be further broken down into sites. All provisioning servers in a farm share the same SQL database and Citrix license server.

Each site is a logical entity containing provisioning servers, vDisk pools, and target device collections. Although all sites within a farm share the same database, target devices can only fail over to other provisioning servers within the same site.



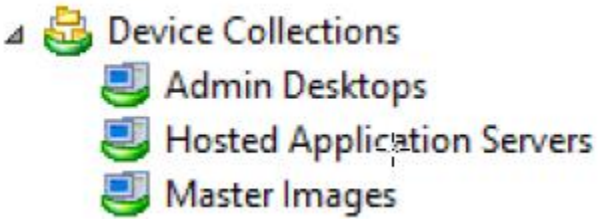
Some factors must be considered when determining the overall Citrix Provisioning topology:

- Network — Provisioning servers constantly communicate with the farm database to retrieve system configuration settings. Therefore, separate farms should be created for each physical location where target devices reside unless they are connected to the database server by a fast and robust connection.
- Administration — Organizations may need to separate departmental, regional, or countrywide administrative duties. Additional Citrix Provisioning farms will add some complexity to the environment's management. However, this overhead is typically limited to initial configuration, desktop creation, and image updates.
- Organization — A practical reason for building multiple sites is organizational changes. For example, two companies may have recently merged through acquisition but must keep resources separate while integration occurs. Configuring the organization to use separate sites is one way to keep the businesses separate but managed centrally through the Citrix Provisioning console.
- Platform — As discussed above, PVS is supported on various hypervisors, Azure, and GCP. Separate farms should be created for each platform unless they are connected to the database server by a fast and robust connection.

Only create additional sites if the business requirements warrant it. A single site per farm is easier to manage and requires no additional configuration.

Decision: Device Collections

Device collections provide the ability to create and manage logical groups of target devices. Creating device collections simplifies device management by allowing actions at the collection level rather than the target device level.



Device collections can represent physical locations, subnet ranges, chassis, or different organizational departments. They can also be used to separate production target devices from test and maintenance ones logically.

Consider creating device collections based on vDisk assignment so that the status of all target devices assigned to a particular vDisk can be quickly identified.

Decision: High Availability

Citrix Provisioning is a critical component of the virtual desktop infrastructure. These recommendations should be followed to eliminate single points of failure.

- **Provisioning Server –** At least two provisioning servers should always be implemented per site. Sufficient redundancy should be incorporated into the design so that a single server failure does not reduce the number of target devices supported per site. Citrix recommends an N+1 approach; no more than 2,000 VDA servers are handled per Citrix Provisioning server. When limited to 2,000 VDAs per server, if one server is unavailable, the other N servers pick up the total load for all VDAs.

- **DNS Alias FQDN** — Consider using a DNS Alias FQDN to locate the PVS servers for the initial logon process. Using this method, up to 32 servers can participate in the initial login process to assure high availability during boot up.
- vDisks and Storage — For vDisk stores hosted on local, Direct-Attached Storage (DAS), or Storage Area Network (SAN), replication should synchronize the vDisks. If using Network-Attached Storage (NAS), ensure the vDisks are hosted on a highly available network share.

- **Networking** — Provisioning Servers need a minimum of 1GB of network throughput; however, when streaming modern versions of Windows it is recommended to use 20 to 40GB NICs. Additionally, if the Provisioning Servers are virtual machine, it is recommended that hardware that supports SR-IOV is used to provide the best throughput.

> Design Tip:
>
> Consider putting all Citrix Provisioning servers and VDAs in a site using the same streaming subnet so no routing is required.

Trivial File Transfer Protocol (TFTP) is a communications protocol for transferring configuration or boot files between machines. Provisioning services can use TFTP to deliver the bootstrap file to target devices. There are several options available to make the TFTP service highly available. Some of the more commonly used options are:

- **DNS Round Robin** – A DNS entry is created for the TFTP service with multiple A records corresponding to the TFTP services running on the provisioning servers in the farm. This method is not recommended since the state of the TFTP service is not monitored. Clients could potentially be sent to a non-functioning server.
- **Hardware load balancer** — Use a hardware load balancer like NetScaler to create virtual IPs corresponding to the provisioning servers. The NetScaler can intelligently route traffic between the provisioning servers. If one of the servers becomes unavailable, NetScaler will automatically stop routing TFTP requests to that server. This is the best method for making TFTP highly available, but it takes additional time to set up.
- **Multiple DHCP Option 66 entries** – This method is easy to implement but requires a DHCP service that supports entering multiple entries in option 66. Microsoft DHCP server allows one option 66 entry, so this method would not be feasible in environments with Microsoft DHCP services. If using a non-Microsoft DHCP server or appliance, check with the manufacturer to verify that multiple option 66 entries are supported.

There are other options available that can achieve the same result without having to use TFTP:

- **Proxy DHCP** — The provisioning server's PXE service provides the bootstrap information. If one of the servers is down, the next available server in the farm can provide the bootstrap information. This method requires the provisioning servers to be in the same broadcast domain as the target devices. If other PXE services are running on the network (Altiris, SCCM, etc.), multiple VLANs may be required to keep the PXE services from interfering.
- **Boot Device Manager** – Use the Boot Device Manager to create a bootstrap file that is either placed on the local hard drive or used as a bootable ISO file. If the ISO file or bootable hard disk (BDM Boot Disk) is used, create an ISO file using BDM.exe, upload it to the hypervisor where the VDAs will run, and update the VM template to attach the ISO and change the boot order. Specify PXE boot as the method when running the

Virtual Apps and Desktops Setup Wizard. When either method is utilized, the TFTP service is not used at all.

Note:

Boot Device Manager is the only boot method available if you use PVS on Azure or GCP.

High availability should always be incorporated into the Provisioning Services design. Although high availability may require additional resources and increased costs, it will provide a highly stable environment so that users experience minimal impact due to service outages.

Decision: Bootstrap Delivery

A target device initiates the boot process by loading a bootstrap program, which initializes the streaming session between the target device and the provisioning server. There are three methods by which the target device can receive the bootstrap program:

Using DHCP Options –

1. When the target device boots, it sends a broadcast for its IP address and boot information. DHCP will process this request and provide an IP and scope option settings 66 (the name or IP address of the Provisioning Services TFTP server) and 67 (the name of the bootstrap file).

Note:

If a load balancer is being used for the TFTP service, the address of the load balancer is entered in option 66.

2. Using TFTP, the target device sends a request for the bootstrap file to the provisioning server, which downloads the boot file from the server.
3. The target device boots the assigned vDisk image.

Note:

To receive PXE broadcasts, the UDP/DHCP Helper must be configured when targets are not on the same subnet as the DHCP servers.

Using PXE Broadcasts –

1. When a target device boots from the network, it sends a broadcast for an IP address and boot information. DHCP processes this request and provides an IP address. In addition, all provisioning servers that receive the broadcast return the boot server and boot file name information. The target device merges the information received and starts the boot process.
2. Using TFTP, the target device sends a request for the bootstrap file to the provisioning server, which responds first. The target device then downloads the boot file from the provisioning server.

Note:

Ensure no other PXE services, such as the Altiris PXE service, are used on the same subnet or isolated using VLANs; otherwise, conflicts may occur with Provisioning Services.

When targets are not on the same subnet as the DHCP and PVS servers, the UDP/DHCP Helper must be configured to receive PXE broadcasts.

Using Boot Device Manager – The Boot Device Manager (BDM) creates a boot file that target devices obtain through a physical CD/DVD, a mounted ISO image, or a virtual hard disk assigned to the target device. A BDM partition can be upgraded in one of three ways:

- by collection
- by a group of highlighted devices
- by a single device

A summary of the advantages and disadvantages of each delivery method is listed in this table.

Delivery Method	Advantages	Disadvantages
DHCP Options	Easy to implement	This requires changes to the production DHCP service, which may only allow one option 66 entry. It also requires a UDP/DHCP helper for targets on different subnets.
PXE	Easy to implement	It can interfere with other running PXE services on the same subnet. A UDP/DHCP helper is required for targets on different subnets.
BDM ISO	Does not require PXE or TFTP services.	Booting physical target devices requires extra effort. BDM ISO is regarded as a single point of failure if a single file is used.
BDM Boot Disk	Using a DNS Alias FQDN with the BDM Boot Disk can expand to up to 32 individual PVS server addresses.	
BDM Partition	The BDM boot partition upgrade does not require PXE, TFTP, or TSB.	Booting physical target devices requires extra effort. An extra 8MB partition is created for each target device.

Note:

With UEFI, the choice of initial login server is random so the load is shared automatically.

Decision: vDisk Format

Citrix Provisioning supports the use of fixed-size or dynamic vDisks.

- **Fixed-size disk —** For vDisks in private mode, fixed-size disks prevent disk fragmentation and offer improved write performance over dynamic disks.

- **Dynamic disk** – Dynamic disks require less storage space than fixed-size disks but offer significantly lower write performance. Although vDisks in Shared mode do not perform writes to the vDisk, the time required to complete vDisk merge operations will increase with dynamic disks. This is not a common occurrence, as more environments choose to create new vDisks when updating.

Since most reads will be to the System Cache in RAM, there is no significant change in performance when utilizing fixed-size or dynamic disks. In addition, dynamic disks require significantly less storage space, so they are recommended.

Decision: vDisk Replication

vDisks hosted on a local, Direct Attached Storage or a SAN must be replicated between vDisk stores whenever a vDisk is created or changed. Citrix Provisioning supports the replication of vDisks from local stores to the provisioning server and across multiple sites that use shared storage. The replication of vDisks can be performed manually or automatically:

- **Manual** – Manual replication is simple but time-consuming, depending on the number of vDisks and vDisk stores. If an error occurs during the replication process, administrators can immediately catch it and take the appropriate steps to resolve it. The risk of manual replication is vDisk inconsistency across the provisioning servers, resulting in load balancing and failover not working properly. For example, if a vDisk is replicated across three servers and then one of the vDisks is updated, that vDisk is no longer identical and will not be considered if a server failover occurs. Even if the same update is made to the other two vDisks, the timestamps on each will differ; therefore, the vDisks are no longer identical.
- **Automated** – Automated replication is faster than the manual method due to the required number of vDisks and vDisk Stores for large environments. Some automated tools, such as Microsoft DFS-R, support bandwidth throttling and Cross File Remote Differential Compression (CF-RDC), which use heuristics to determine whether destination files are similar to the replicated file. If so, CF-RDC will use blocks from these files to minimize the data transferred over the network. The risk of automated replication is that administrators do not typically monitor replication events in real-time and do not respond quickly when errors occur unless the automation tool has an alerting feature. Some tools can be configured to automatically restart the copy process in case of failure. For example, Robocopy supports “resume copying” if the network connection is interrupted.

For medium and large projects, use a tool to automate vDisk replication. Select a tool that can resume from network interruptions, copy file attributes, and preserve the original timestamp.

> Note:
>
> Load balancing and high availability will not work unless the vDisks have identical timestamps.

Decision: Server Sizing

Generally, a Citrix Provisioning server is defined with the following specifications:

Component	Specification
Model	Virtual
Processor	4 to 8 vCPU
Memory	2GB + (# of vDisks * 2GB)
Network	10+ Gbps NIC / SR-IOV support if Citrix Provisioning server runs as a Virtual Machine
Host Storage	40 GB shared storage
vDisk Storage	Depending on the number of images/revisions
Operating System	Windows Server 2022

Model - Citrix Provisioning can be installed on virtual or physical servers:

- Virtual – Offers rapid server provisioning, snapshots for quick recovery or rollback scenarios, and the ability to adjust server resources on the fly. Virtual provisioning servers allow target devices to be distributed across more servers, helping to reduce the impact of server failure. Virtualization makes more efficient use of system resources.
- Physical – Offers higher levels of scalability per server than virtual servers. Physical provisioning servers mitigate the risks of virtual machines competing for underlying hypervisor resources.

Virtual provisioning servers are generally preferred when sufficient processor, memory, disk, and networking resources can be made available and guaranteed to be available.

> Note:
>
> For high availability, ensure that virtual Provisioning Servers are distributed across multiple virtualization hosts. Distributing the virtual servers across multiple hosts will eliminate a single point of failure and prevent the entire Provisioning Services farm from being brought down in case of a host failure.

CPU - Citrix Provisioning is not CPU intensive. However, underallocating the number of CPUs does impact the optimization of the network streams. The Streaming Service with default settings and the Citrix Provisioning server can stream up to 4,000 target devices. However, Citrix recommends limiting the streams to 2,000 target devices so that other Provisioning Servers can pick up load if one server is unavailable.

If the provisioning server does not have sufficient cores, the server will show a higher CPU utilization, and target devices waiting for requests to be processed will have a higher read latency.

- In small environments (up to approximately 500 virtual machines), 4 vCPUs are recommended.

- In larger environments, 8 vCPUs are recommended.

RAM - The Windows operating system hosting Citrix Provisioning partially caches the vDisks in memory (system cache), reducing the number of reads required from storage. Reading from storage is significantly slower than reading from memory. Therefore, Provisioning Servers should be allocated sufficient memory to maximize the benefit of this caching process.

The following formula can be used to determine the optimal amount of memory that should be allocated to a provisioning server:

*2GiB + (vDisk * 2GiB) + 15% (Buffer)*

Network — Unlike most Citrix Virtual Apps and Desktops components, Citrix Provisioning does not bottleneck the CPU. Its scalability is based on network throughput.

Determining how much time will be required to boot the target devices can be estimated using the following formula:

$$Seconds\ to\ Boot = \frac{(Number\ of\ Targets * MB\ Usage)}{Network\ Throughput}$$

> Tip:
>
> Firewalls can add latency and create bandwidth bottlenecks in Provisioning Services environments. If firewalls cannot be avoided, refer to the Citrix Tech Zone article – Communication Ports Used By Citrix Technologies, for the list of ports that should be enabled for full functionality.

Growth — As the farm grows, administrators must decide whether to add more resources to the provisioning servers or the farm itself.

Many environmental factors need to be considered when determining whether the Provisioning Servers should be scaled up or scaled out:

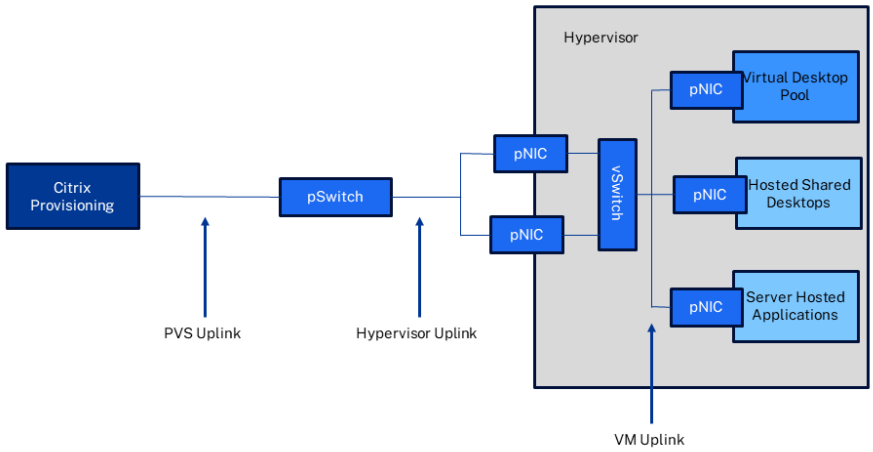
- Redundancy – Spreading user load across additional less-powerful servers helps reduce the number of users affected by a single provisioning server failure. If the business cannot accept the loss of a single high-specification server, consider scaling out.
- Failover times — The more target devices are connected to a single provisioning server, the longer it will take for them to failover if the server fails. Consider scaling out to reduce the time required for target devices to failover to another server.
- Data center capacity – The datacenter may have limited space, power, and/or cooling. In this situation, consider scaling up.
- Hardware costs — Scaling up may initially be more cost-effective. However, there will be a point where scaling out actually becomes more cost-effective. A cost analysis should be performed to make that determination.

Hosting costs – There may be hosting and/or maintenance costs based on the number of physical servers used. If so, consider scaling up to reduce the long-term cost of these overheads.

Decision: Network Configuration

As mentioned before, it is essential that the network is sized correctly to prevent network bottlenecks, which cause high disk access times and directly affect virtual desktop

performance. The following diagram outlines a common Provisioning Services network infrastructure:



The following network configuration is recommended for the network sections outlined within the diagram:

- PVS Uplink – All disk access from the target devices will be transferred via the PVS network uplink. This means hundreds or even thousands of devices will use this network connection. Therefore, it is vital that this connection is redundant and can failover without any downtime. Furthermore, Citrix recommends a minimum bandwidth of 1Gbps per 500 target devices. A respective QoS quota or a dedicated physical network uplink should be configured for virtual provisioning servers to ensure the best performance.
- Hypervisor Uplink — This uplink is used by all PVS target devices hosted on a particular hypervisor host. Therefore, redundancy with transparent failover is strongly recommended. Unless the target devices run a very I/O-intensive workload or perform I/O-intensive tasks (e.g., booting) simultaneously, a bandwidth of 1Gbps is sufficient for this uplink.
- VM Uplink – All network traffic for a virtual machine, including PVS streaming traffic, will traverse this virtual network connection. Unless the workload is extremely I/O intensive, a bandwidth of 100 Mbps can handle even peak loads during I/O intensive tasks, such as booting from vDisk. For example, a Windows 2022 Server will read approximately 232MB from the vDisk for a period of 90 seconds until the Windows Logon Screen is shown. During this period, an average data rate of 20.5 Mbps with peaks up to 90 Mbps can be observed.

The following switch settings are recommended for Citrix Provisioning. They are not required for PVS in Azure or GCP:

- Storm Control — Storm Control is a feature available on Cisco switches that allows a threshold to be set to suppress multicast, broadcast, or unicast traffic. Its purpose is to prevent malicious or erroneous senders from flooding a LAN and affecting network

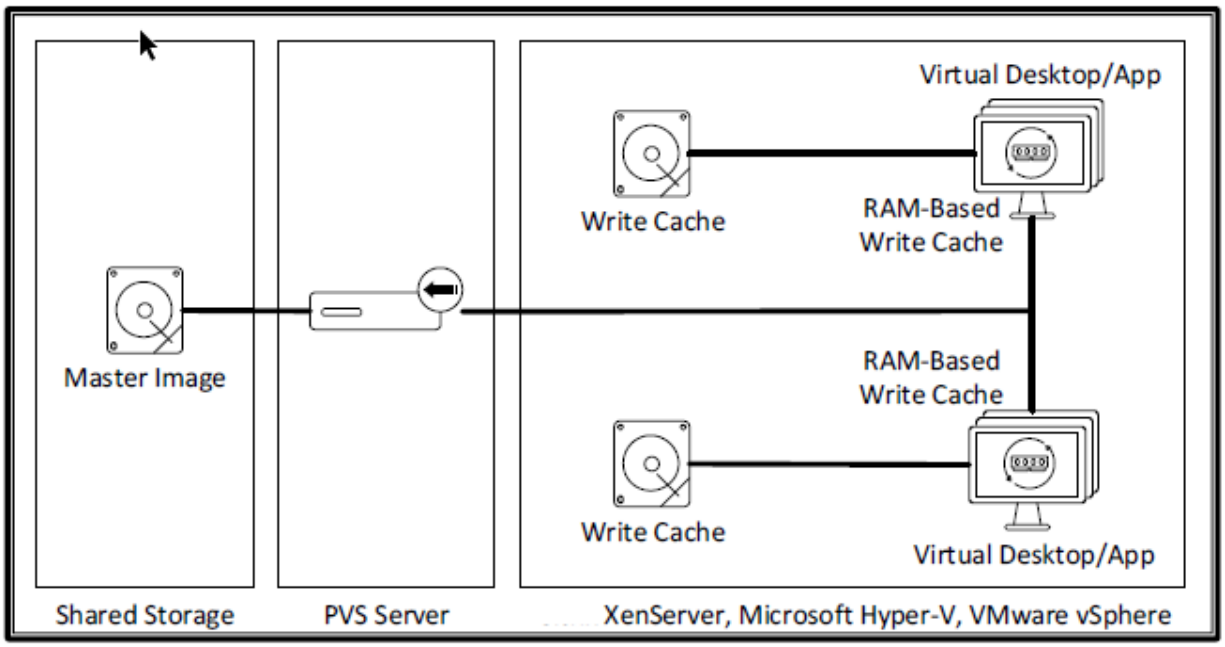
performance. Provisioning Servers may, by design, send large traffic that falls within a storm control threshold. Therefore, the feature should be configured accordingly.

- Broadcast Helper – The broadcast helper must direct broadcasts from clients to servers that would otherwise not be routed. In a Provisioning environment, it is necessary to forward PXE boot requests when clients are not on the same subnet as the servers. If possible, the recommended network design is to have Provisioning servers residing on the same subnet as the target devices. This mitigates the risk of any service degradation due to other networking infrastructure components.

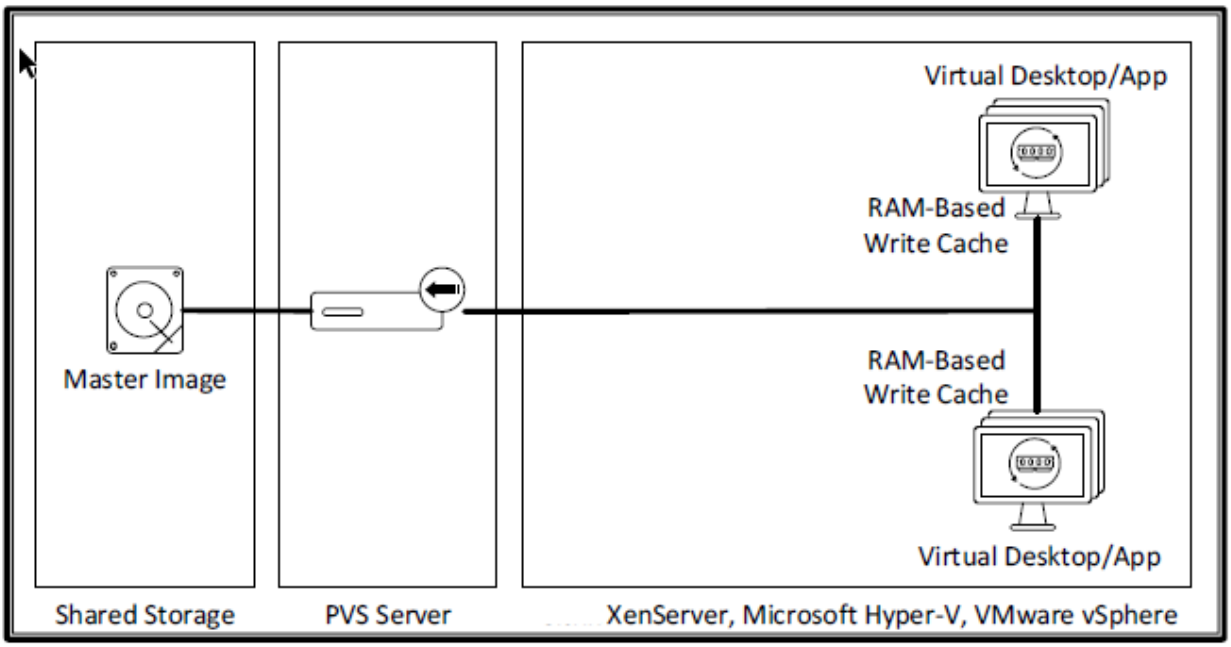
Decision: Write Cache

Because the master image is read-only, each virtual machine has a writable disk to store all changes. The administrator must decide where to store the write cache disk.

VM-Cache in RAM with Overflow to Disk - A combination of RAM and local storage is used for the write cache. First, writes are stored within the RAM cache, providing high performance. As the RAM cache is consumed, large blocks are removed from the RAM cache and placed onto the local storage write cache disk. This option is recommended as it provides high-performance levels with the low cost of local storage. This is the only option available in Azure.



VM — Cache in RAM — The RAM associated with the virtual machine holds the write cache drives for each target virtual machine. This option provides high performance due to the RAM's speed. However, the virtual machine will become unusable if the RAM cache runs out of space. To use this option, significant amounts of RAM must be allocated to each virtual machine, increasing the overall cost.



Decision: Antivirus

Most antivirus products scan all files and processes by default, significantly impacting Citrix Provisioning performance. For details on how antivirus software can be optimized for Provisioning Services, please refer to CTX124185 — Provisioning Services Antivirus Best Practices.

Antivirus software can cause file-locking issues on provisioning servers and VDAs being streamed. To prevent file contention issues, the vDisk Store and write cache should be excluded from antivirus scans. Regular vDisk updates must also be carried out to ensure Antivirus software and Windows updates are up to date and applied.

When a virtual disk runs in standard mode and must be restarted, it downloads the previously loaded virus definitions. This can cause performance degradation when restarting several target devices simultaneously, often causing network congestion while the operation persists. In extreme cases, the target device and provisioning server can become sluggish and consume more resources than necessary. If the antivirus software supports it, definition files should be redirected to the write cache drive to preserve them between reboots.

Machine Creation Services

Machine Creation Services (MCS) uses disk-cloning technology to simplify the deployment of virtual machines. Computers are provisioned and re-provisioned in real-time from a single shared disk image. In doing so, administrators can eliminate the need to manage and patch individual systems. Instead, administrators perform all image management on the master image.

Decision: Platform

MCS can be used on various on-premises hypervisors and several hyperscalers. When using different platforms, there are slightly different considerations.

On-premises hypervisors

Citrix supports XenServer, VMware vSphere, Nutanix Acropolis, and Microsoft SCVMM. A hypervisor administrator account with the appropriate permissions is required when connecting your site to the hypervisor via a hosting connection. The requirements for each vendor are in our product documentation. Additionally, if you are using Nutanix Acropolis, a plug-in is required on your Delivery Controllers or Cloud Connectors.

Microsoft Azure

Deploying a hosting connection to Azure requires a service principal with the appropriate permissions. Citrix can create the service principal to create the hosting connection, in which case the principal will have Contributor access. Alternatively, you can pre-create a service principal and designate the permissions. A list of minimum permissions can be found in the product documentation.

Resource groups are logical groupings of resources within an Azure subscription. To make them easier to track, it is recommended that you split your machine catalogs into separate delivery groups. You can pre-create resource groups with a desired naming schema, or Citrix can create the resource group while deploying the machine catalog.

Virtual machine golden images can be created via the Azure Resource Manager when deploying resources via Azure. Once created, these images can be managed via the Azure Compute Gallery, which provides versioning and grouping of resources and global replication across regions.

Machine profiles can be used within Azure to set various custom properties of VMs. You can view the specific properties that can be set via the machine profile here. A machine profile is required if you are using Entra ID.

Ephemeral disks offer a cost-effective storage option that reuses the local disk of the VMs to host the operating system disk. This functionality is useful for Azure environments that require a higher-performing SSD disk over a standard HDD disk. It is important to note that only certain Azure VMs support ephemeral disks. To learn more about ephemeral disks and their requirements and limitations, please visit our product documentation.

Google Cloud Platform (GCP)

Several APIs need to be enabled when deploying VDAs into Google Cloud. These APIs are the Compute Engine API, Cloud Resource Manager API, Identity and Access Management (IAM) API, and the Cloud Build API.

Deploying machines via MCS requires a hosting connection with appropriate Google Cloud Project resources rights. The permissions needed depend on when your hosting connection was originally created. View our product documentation for more information.

Similarly to Azure, when you create a catalog to provision machines using Machine Creation Services (MCS), you can use a machine profile to capture the hardware properties from a virtual machine and apply them to newly provisioned VMs in the catalog.

Amazon Web Services (AWS)

When creating a connection to AWS from the Full Configuration interface, you must provide the API and secret key values, as well as information about your AWS environment, such as region, VPC name, etc. You must also configure appropriate IAM permissions to connect AWS and DaaS.

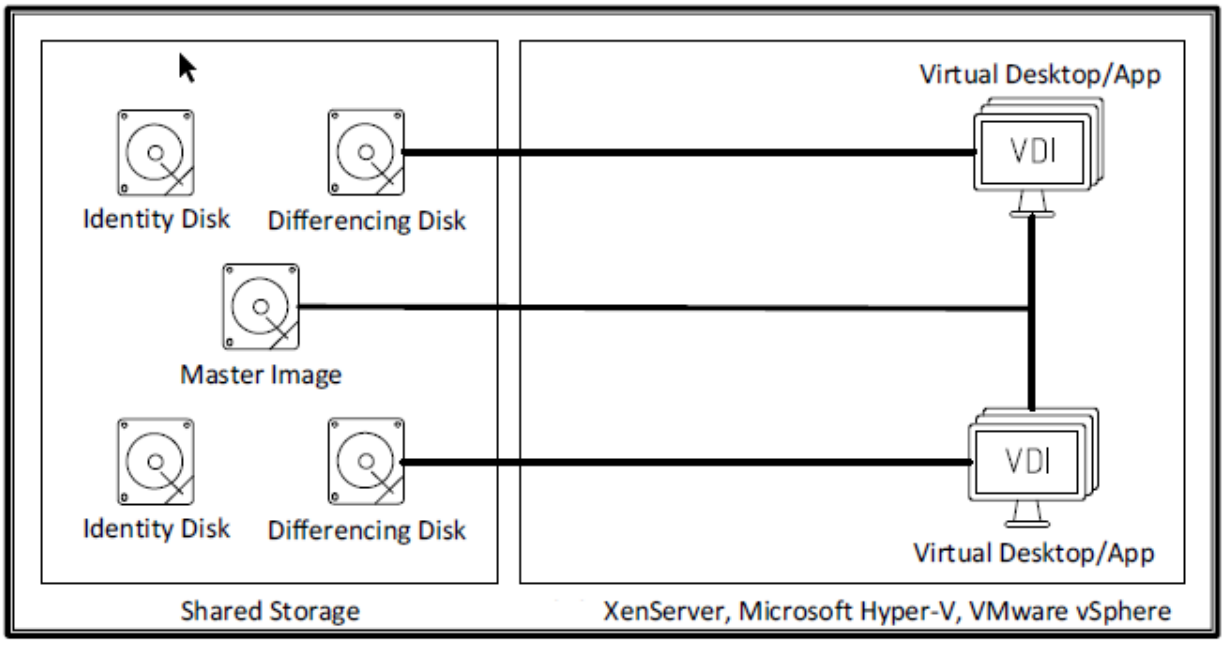
AWS provides the following options: shared tenancy (the default type) and dedicated tenancy. Shared tenancy means multiple Amazon EC2 instances from different customers might reside on the same physical hardware. Dedicated tenancy means that your EC2 instances run only on hardware with others you have deployed.

When you create a catalog to provision machines using Machine Creation Services (MCS) in AWS, you select an AMI to represent the master image of that catalog. From that AMI, MCS uses a snapshot of the disk.

Decision: Storage Location

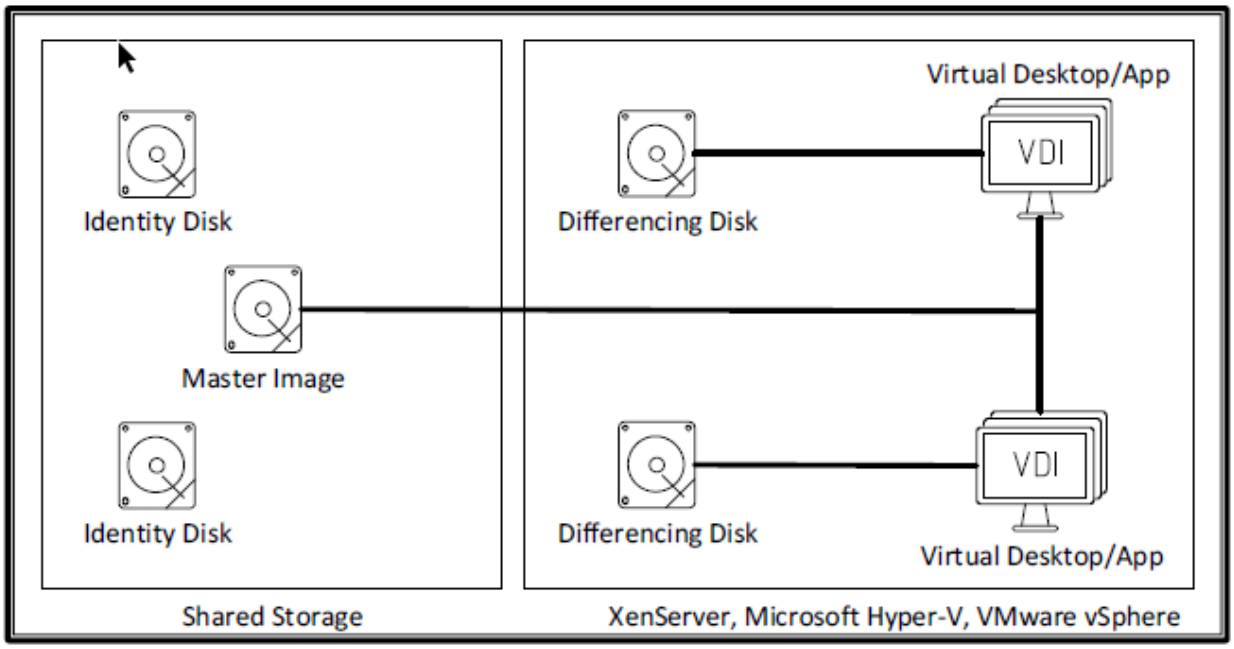
Machine Creation Services allows administrators to break up a virtual desktop into multiple components and store those pieces on different storage arrays.

Shared Storage - The first option utilizes shared storage for the operating system and the differencing disk.

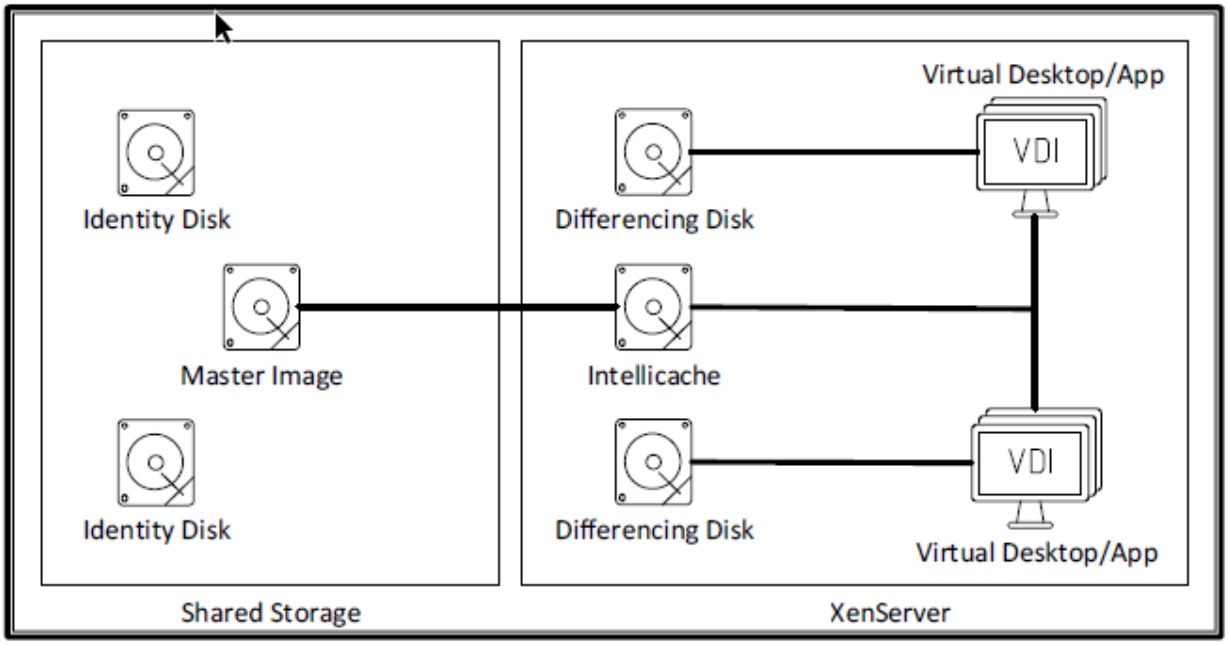


Although this option allows the sharing of the master image across multiple hypervisor hosts, it puts more strain on the storage array because it must also host the differencing disk, which is temporary data.

Hybrid Storage - The second option uses shared storage for the operating system disk and local hypervisor storage for the differencing disk.



XenServer IntelliCache Storage - The third option uses shared storage for the operating system disk, local hypervisor storage for the differencing disk, and local XenServer storage for a local cache of the operating system disk.



This is only an option for XenServer implementations. It provides the same value as the hybrid storage approach while reducing read IOPS from shared storage. IntelliCache can coexist with the XenServer RAM-based read cache if XenServer RAM is limited.

Public Cloud Storage — Public cloud vendors offer a variety of vendor-managed and customer-managed storage options. Please review our reference architectures for considerations regarding public cloud storage for Azure, AWS, and GCP.

Decision: Cloning Type

Machine Creation Services incorporates two types of cloning techniques.

- Thin - Every VM within the catalog utilizes a single, read-only virtual disk for all reads. A
- second virtual disk, unique for each VM, captures all write IO activity.
- Full - Every VM within the catalog receives a full copy of the master disk image. Each VM fully owns the disk, allowing for read/write activity. Full cloning technology is only available for personal virtual desktops, where a dedicated virtual machine saves all changes to a local disk.

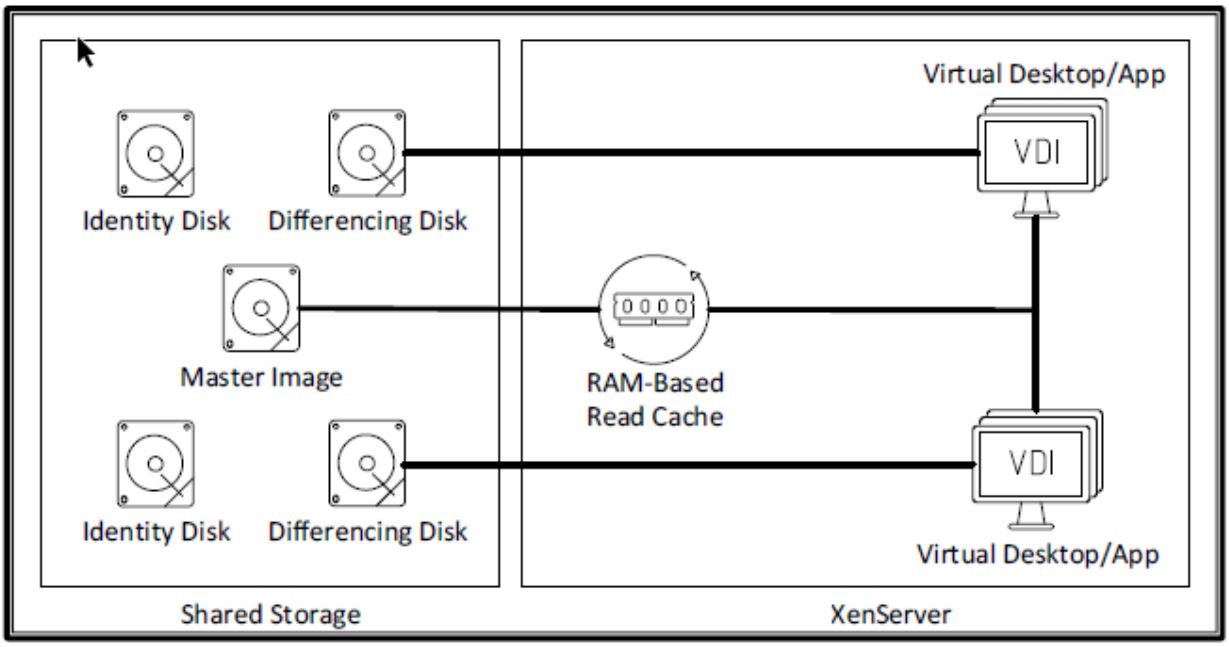
> Note:
>
> In the public cloud, there is only full clone mode.

Consider the following when deciding between thin and full cloning technologies:

	Thin Clone	Full Clone
Storage Space Requirements	Have the greatest storage space savings. A single master disk image is shared across multiple VMs. Only the differencing disk (writes) consumes space, which continues to grow until the VM reboots	High storage space requirements: Each VM receives a full copy of the master image. The size continues to grow as changes are made to the VM.
Backup/Restore	Difficult — Many third-party Backup/DR solutions do not support snapshot/delta disks, making thin provisioned VMs hard/impossible to backup or move to other storage arrays.	Easy - The VM exists within a single virtual disk, making it easy to backup and restore.
Provisioning Speed	Fast - Only requires a single disk image.	Slow (can be mitigated) - Each VM requires a full copy of the master image. Storage optimization technologies can help mitigate this.
Boot Storm	High Impact - In a boot storm, all differencing disks re-size to hold all writes from Windows startup, placing a high load on the storage as it happens simultaneously.	Low Impact

Decision: Read Cache

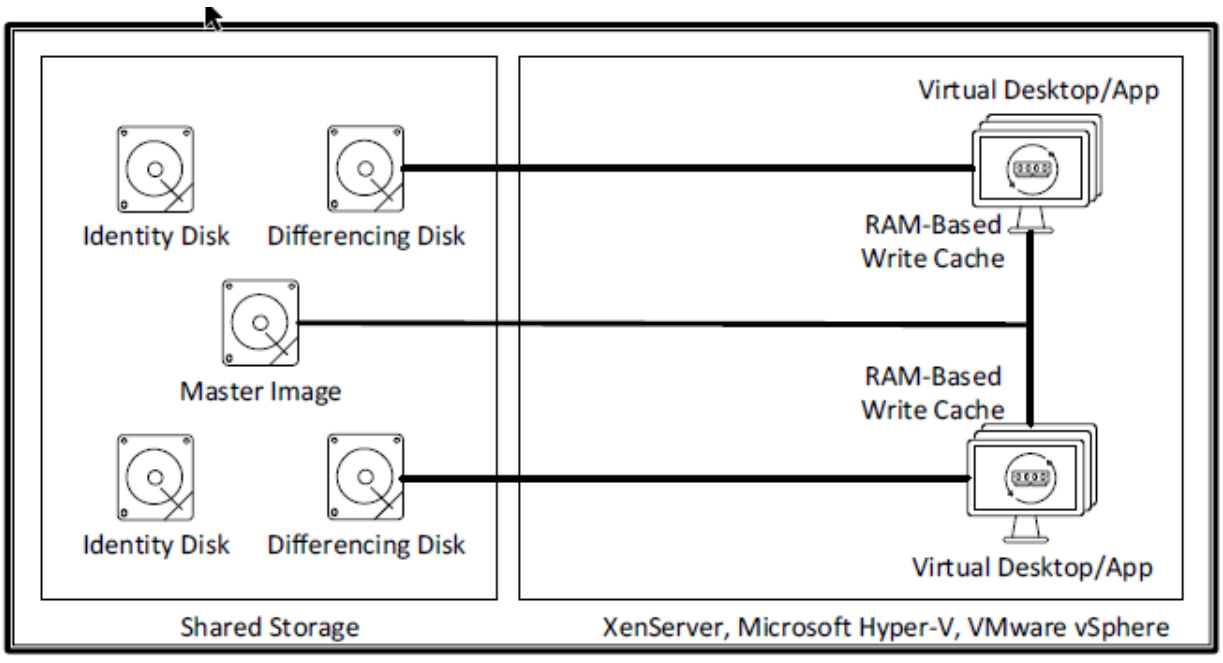
During boot and logon, virtual desktops incur high storage levels read IOPS, which can strain the underlying storage subsystem. When deployed on XenServer, Shared and Pooled VDI modes utilize a RAM-based read cache hosted on each XenServer.



Utilizing this integrated technology reduces read IOPS by 50-80%.

Decision: Write Cache

During steady state, virtual desktops incur high storage levels of write IOPS, which can strain the underlying storage subsystem. Shared and Pooled VDI modes can utilize a RAM-based write cache using non-paged pool RAM from the virtual machine operating system.



Utilizing this integrated technology reduces write IOPS by 95%.

Layer 5: The Compute Layer

Hardware Sizing

This section covers hardware sizing for virtual infrastructure servers, virtual desktops, and application hosts. Two methods are typically used to size these servers.

- The first and preferred way is to plan and purchase hardware based on the workload requirements.
- The second way is to use existing hardware in the best configuration to support the different workload requirements.

This section will discuss decisions related to both methods.

Decision: Workload Separation

When implementing a Citrix Virtual Apps and Desktops or Citrix DaaS deployment, the infrastructure, multi-session, or single-session workloads can be separated into dedicated resource clusters or mixed on the same physical hosts. Citrix recommends using resource clusters to separate the workloads, especially in an enterprise deployment. This allows for more targeted host sizing as each workload has unique requirements, such as overcommit ratios and memory usage.

In smaller environments where resource clusters are cost-prohibitive, the workloads may be mixed in a manner that still allows for a highly available environment. Citrix's leading practice is to separate the workloads; however, mixed workloads are a cost-based business decision.

Decision: Physical Process (pCPU)

The following tables guide the number of virtual desktops supported for light, medium, and heavy workloads per physical core.

In the first table, single-session desktops are represented and each desktop correlates to a single concurrent user, assuming that the operating system underwent optimization.

User Workload	Operating System	Desktops per Physical Core	VM Specs	VCPU to CPU overcommit
Light	Windows 10	6	2 vCPUs/4 GB RAM	12 vCPUs to 1 CPU
	Windows 11	5	2 vCPUs/4 GB RAM	10 vCPUs to 1 CPU
Medium	Windows 10	5	4 VCPUs /8 GB RAM	10 vCPUs to 1 CPU
	Windows 11	4	4 vCPUs/8 GB RAM	8 vCPUs to 1 CPU
Heavy	Windows 10	2	6 vCPUs/8 GB RAM	6 vCPUs to 1 CPU
	Windows 11	2	6 vCPUs/8 GB RAM	6 vCPUs to 1 CPU

The second table represents multi-session desktops and the number of users per physical core.

User Workload	Operating System	Users per Physical Core	VM Specs	vCPU to CPU overcommit
Light	Windows Server 2022	18	8 vCPUs/ 32 GB RAM	2 vCPUs to 1 CPU
Medium	Windows Server 2022	10	8 vCPUs/ 32 GB RAM	2 vCPUs to 1 CPU
Heavy	Windows Server 2022	6	8 vCPUs/ 32 GB RAM	2 vCPUs to 1 CPU

Note:

If using Windows Multisession, comparing Windows Server to Windows Multisession workload resulted in 19% fewer task workers and 32% fewer Knowledge workers. This performance decrease is expected because Windows Multisession is a full client version and is not optimized for server-based computing like Windows Server.

The “Users per Physical Core” estimate is a baseline number running Microsoft Office 365. The baseline number must be adjusted based on specific infrastructure requirements. As a general guideline, the following characteristics are baseline changes to server density.

Characteristic	Server Density Impact
Antivirus (not optimized)	25% decrease
Real-time Monitoring	15% decrease
Hyper-threading	20% decrease
Microsoft Office 365	25% decrease

To estimate the total number of physical cores required for the Citrix Virtual Apps and Desktops workload, use the following formula for each user group:

Total Virtual Desktops vCPU = ∑i Usersi/UsersperCorei(1+AV+ Mon + Off - HV))*

Total Virtual Apps vCPU = ∑i Usersi/UsersperCorei(1+Av+Mon+Off-HV))*

Σ represents the sum of all user group combinations “i.”
Usersi = Number of concurrent users per user group
UsersPerCorei = Number of users per physical core
AV = Antivirus impact (default = 0.25)
Mon = Monitoring tools impact (default = 0.15)
HT = Hyper-Threading impact (default = .2)
Off = Microsoft Office impact (default = .25)

If workloads will be separated, the formula should be calculated twice, once for all single-session users and the second for all multi-session users in order.

Decision: Physical Memory (pRAM)

The recommended method for sizing memory to a physical host is to size based on the total memory required to support the virtual machines and the host's CPU capacity. To calculate the total memory required for Citrix Virtual Apps and Desktops, simply multiply the number of virtual machines by the amount of memory allocated to them. The sum of the machine catalogs will be the total RAM required for Citrix Virtual Apps and Desktops hosts. This is shown in the formula below.

$$Total\ Virtual\ Desktops\ pRAM = \Sigma VMi * ivRAMi$$

$$Total\ Virtual\ Apps\ pRAM = \Sigma VMi * ivRAMi$$

Σ represents the sum of all user group combinations "i."
VMi = Number of concurrent users per user group
vRAMi = Amount of RAM assigned to each virtual machine

If workloads will be separated into different hosts (multi-session and single-session workloads), the formula should be calculated for each workload.

Decision: Physical Host (pHost)

In most situations, the number of physical hosts (pHost) to support the Citrix Virtual App and Desktop workloads will be limited by the number of processor cores available.

The following formula estimates the number of hosts required for the user workloads. The formula is based on the best practice of separating the Virtual App and Virtual Desktop workloads due to the different recommended CPU overcommit ratios for each.

$$Single\ SessionpHosts = (Total\ Desktop\ pCPU\ Cores\ per\ pHost + 1)$$

$$Multi\ SessionpHosts = (Total\ App\ pCPU\ Cores\ per\ pHost + 1)$$

The amount of RAM in each host is calculated once the number of physical hosts has been determined based on processor cores.

$$Single\ SessionpRAM\ per\ pHost = HypervisorRAM + (Total\ Desktop\ pRAM\ Desktop\ pHosts - 1)$$

$$Multi\ SessionpRAM\ per\ pHost = HypervisorRAM + (Total\ App\ pRAM\ App\ pHosts - 1)$$

Decision: GPU

Hosts that deliver graphical workloads require graphics processors to deliver a high-end user experience. Specific hardware hosts and graphics cards are required to support high-end graphics using HDX 3D Pro. An updated list of tested hardware is available in a knowledge base article. Sizing the desktop and application hosts of high-end graphics users should be based on the GPU requirements, ensuring that the host has adequate CPU and memory resources to support the workload.

NVIDIA, AMD, or Intel cards can be leveraged with vGPU profiles to support multiple users. This table provides the GPUs from each that are supported and have been validated, along with sizing guidelines for each.

Storage Sizing

Decision: Storage Architecture

The primary storage architectures are as follows:

- NAS - Provides file-level storage to computer systems through network file shares. The NAS operates as a file server, and NAS systems are networked appliances that contain one or more hard drives, often arranged into logical, redundant storage containers or RAID arrays. Access is typically provided using standard Ethernet and network file-sharing protocols such as NFS, SMB/CIFS, or AFP.

> Note:
>
> NAS can become a single point of failure. If the network share becomes unavailable, all target devices streamed from the disk will also be unavailable.

- SAN - Dedicated storage network that provides consolidated, block-level storage access. SANs allow computers to connect to different storage devices, so no server has ownership of the storage subsystem, enabling data to be shared among multiple computers. A SAN typically has its own dedicated network of storage devices that are generally not accessible through the network by standard means. A specialized adapter called the Host Bus Adapter (HBA) is required to connect a device to the SAN network. SANs are highly scalable with no noticeable change in performance as more storage and devices are connected. SANs can be a costly investment in capital and the time required to learn, deploy, and manage the technology
- Hyper-converged - Hyper-converged storage is a part of hyper-converged infrastructure (HCI) which integrates compute, storage, and networking resources in a single appliance. Storage resources are combined and managed through software rather than traditional hardware-based storage systems. These systems are easily scalable and managed alongside other compute and networking resources.

Monitor

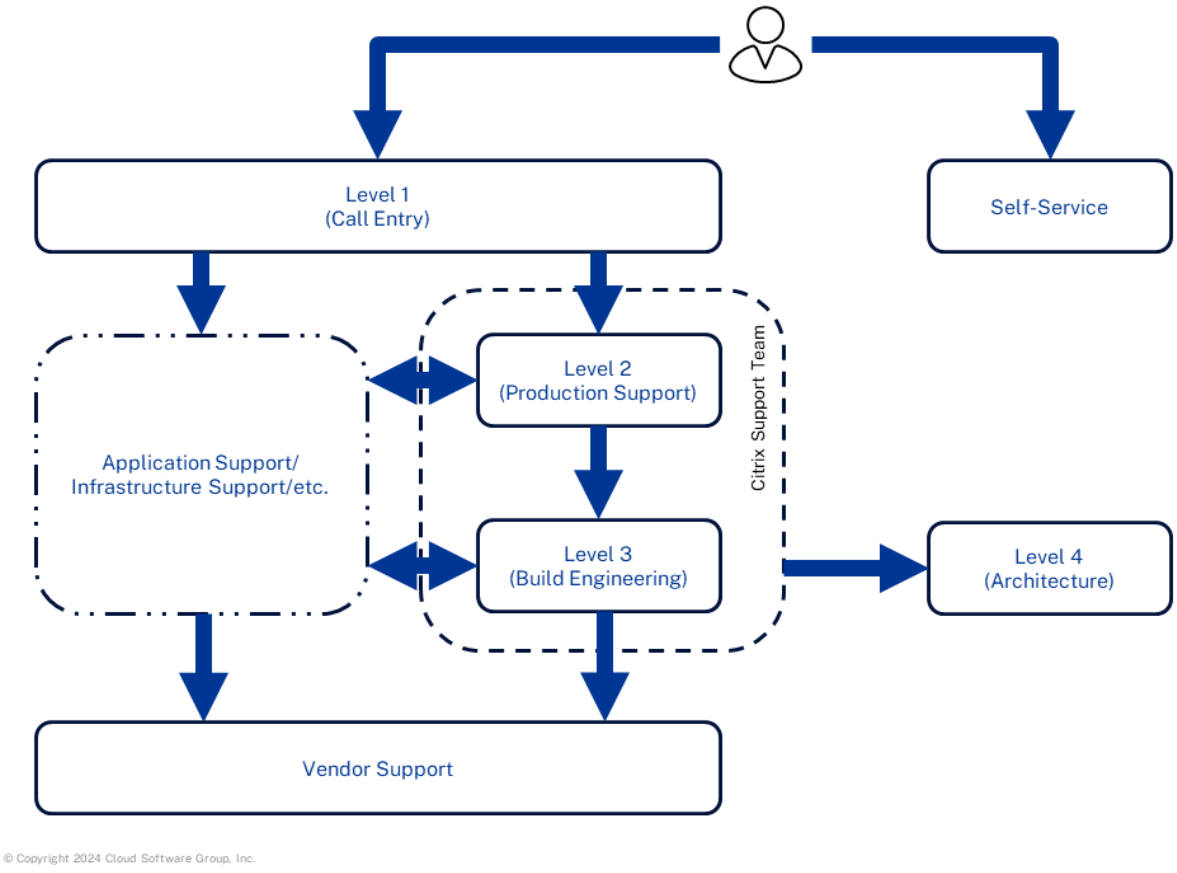
Like any integrated system, monitoring and maintenance are critical to the solution's health. Without proper support, operations, and health monitoring systems, the user experience will slowly degrade.

Process 1: Support

When problems arise, technical support is the first point of contact. This section addresses the proper staffing, organization, training, delegated administration, and tools for maintaining the Citrix deployment.

Decision: Support Structure

Multiple levels of support are the most effective ways of addressing support issues. Low criticality, low complexity, or frequently occurring issues should be managed and resolved at the lower support levels. High criticality and complex issues are escalated to more experienced architects or infrastructure owners. The diagram below outlines a common multi-level support structure.



If a user encounters an issue, Level-1 support (help desk) is the entry point to the support system. Level-1 should resolve 75% of all issues encountered, of which a majority will be routine problems requiring a limited knowledge of the Citrix environment. At this level, issues are quickly resolved, and some may be automated (self-service), such as password and profile resets.

Non-routine problems that exceed Level-1's abilities are escalated to Level-2 (Operators). This support level is generally composed of administrators supporting the production Citrix environment. Information on the end user's problem and attempted troubleshooting steps are documented at the first level, allowing Level-2 technicians to address the problem immediately. Level-2 technicians should handle only about 20% of the support tickets and be highly knowledgeable about the Citrix environment.

Complex issues that exceed Level-2's abilities should be escalated to Level-3 (Implementers). Level-2 and Level-3 support may often be members of the Citrix Support Team, with Level-3 comprising the senior staff maintaining the Citrix environment. Level-3 issues are complicated and often mission-critical, requiring expert knowledge of the virtual desktop and application environment. Level-3 support tickets should be at least 5% of all support issues.

The final level, Level-4 (Architects), is focused on strategic improvements for the solution, testing new technologies, planning migrations, and other high-level changes. Generally, Level-4 is not involved in active support of a production environment. The Citrix architect should regularly contact architects on other teams, such as security, Active Directory, etc.

Should support discover an issue related to an application or underlying infrastructure, the ticket is handed to the appropriate team for troubleshooting. The issue is re-escalated if a program bug is discovered and a ticket is established with the appropriate vendor.

Decision: Support Responsibilities and Skill Set

This table highlights the recommended characteristics of each support level.

Support Level	Description	Responsibilities	Skill Set
Level 1 (Help Desk)	Provide first-line support for reported issues. Initially, this involves servicing support messages and phone calls. This level must perform initial issue analysis, problem definition, ticket routing, and simple issue resolution. It often processes requests for application access or support with configuring plugins.	Perform issue definition, initial analysis, and basic issue resolution. Perform initial troubleshooting to determine the nature of the issue. Create a ticket, collect user information, and log all troubleshooting steps performed. Resolve basic Citrix-related, connectivity, and application-related issues using existing knowledge base articles. Escalate the issue to Level 2 if advanced skills or elevated permissions are required. Ability to isolate the issue as Citrix-related, Microsoft-related, or third-party Application-related. If it affects the production environment or is potentially causing a system-level outage, escalate directly to Level 3. Generate requests for additional issue resolution guides as necessary. Follow up with end users when a support ticket is closed to ensure the problem has been addressed.	• General Citrix knowledge • General Windows client OS/server OS knowledge • General Active Directory knowledge • General Networking Knowledge • General troubleshooting and security knowledge

Level 2 (Operators)	This role primarily supports the day-to-day operations of the Citrix environment, which may include proactive monitoring and management. It should also perform intermediate-level troubleshooting and utilize available monitoring or	• Perform intermediate issue analysis and resolution.
		• Identify the root cause of issues.
		• Respond to server alerts and system outages.
		• Create a weekly report on the number of issues, close rate, open issues, etc.
		• Review vendor knowledge base articles.
		• Respond to out-of-hours helpdesk calls.
			• Experience with Microsoft Windows Server including but not limited to:
			  ○ Configuring operating system options
			  ○ Understanding Remote Desktop Services policies and profiles
			  ○ Using Active Directory

Support Level	Description	Responsibilities	Skill Set
	troubleshooting tools. It will assist with resolving issues escalated by Level-1 support.	• Respond to critical monitoring alerts. • Generate internal knowledge base articles and issue resolution scripts and maintain Level-1 troubleshooting workflows. • Perform basic server maintenance and operational procedures. • Manage user profiles and data. • Escalate the ticket to Level 3 or the appropriate technology owner if advanced skills or elevated permissions are required. • Generate requests for additional issue resolution scripts and knowledge base articles as necessary. • Able to read built-in event logs for Windows and Citrix to do basic troubleshooting following public information via Google/Bing.	○ Creating users/managing permissions and administrator rights ○ Creating and modifying Active Directory group policies • Basic administration skills, including: ○ An understanding of protocols (TCP) o An understanding of firewall concepts ○ An understanding of email administration and account creation ○ An understanding of Remote Desktop Services policies and profiles ○ The ability to create shares and give access to shared folders/files • Experience performing the following: ○ Managing, maintaining, monitoring, and troubleshooting Citrix solutions ○ Backing up components in Citrix environments ○ Updating components in Citrix environments ○ Creating reports for trend analysis

Support Level	Description	Responsibilities	Skill Set
Level 3 (Implementer)	Central point for implementing, administering, and maintaining Citrix desktop and application virtualization infrastructure. This role focuses on deploying new use cases and leading lifecycle management initiatives. Generally, one Implementer could focus on one use case at a time. For example, three new concurrent use cases would require three Implementers. Escalates issues to software vendor-specific technical support and notifies Level-4 about this issue.	• Perform advanced issue analysis and resolution. Perform maintenance and environment upgrades. Addresses high-severity issues and service outages. Manage the Citrix environment. Oversee and lead administrative tasks performed by Level 2. Manage network and storage infrastructure related to the Citrix environment (depending on the company's size or Citrix environment). Review periodic server health reports, resource usage, user experience, and overall environment performance. Review vendor knowledge base articles and newly released updates. Perform policy-level changes and make Active Directory updates. Review change control requests that impact the Citrix environment. Perform advanced server and infrastructure maintenance. Review knowledge-base articles and issue-resolution scripts for accuracy, compliance, and feasibility Create knowledge-base articles and issue-resolution scripts to address Level-2 requests. Escalate the ticket to vendor-specific technical support when necessary and notify Level 4 of the issue.	• Knowledge of how the following Windows components integrate with Citrix technologies: ○ Active Directory Domain Services ○ Active Directory Certificate Services ○ Policies ○ Domain Name System (DNS) ○ Dynamic Host Configuration Protocol (DHCP) ○ Group Policy Objects (GPOs) ○ NTFS Permissions ○ Authentication and Authorization ○ Knowledge of IIS ○ Microsoft Windows Operating Systems ○ Roles and features of Windows Server • Knowledge of SQL Clustering and AlwaysOn Availability Groups. ○ General networking skills (i.e., routing, switching) ○ Knowledge of hypervisors ○ Knowledge of public clouds. ○ Knowledge of shared storage configuration and management.
Level 4 (Architect)	The Level-4 team has minimal exposure to administrative tasks but	• Provide technical leadership for upcoming projects.	• Advanced architectural assessment and design skills for:

Support Level	Description	Responsibilities	Skill Set
	focuses on scoping, planning, and executing Citrix-specific service and project requests. An architect translates business requirements into a technical design.	• Lead design updates and architecture revisions. • Address high-severity issues and service outages. • Review periodic server health reports, resource usage, user experience, and overall environment performance to determine the next steps and upgrade paths. • Initiate load testing to determine the capacity of the environment. • Review frequently recurring helpdesk issues. • Ensure technical specifications continue to meet business needs. • Update design documentation.	○ Citrix Virtual Apps and Desktops ○ XenServer, VMWare, Hyper-V ○ Azure, AWS, GCP ○ Citrix Provisioning ○ NetScaler ○ Citrix StoreFront ○ Active Directory ○ Storage solutions ○ Networking ○ Application delivery ○ Disaster recovery ○ Policies/policy structures and security restrictions ○ Licensing ○ Methodology • Intermediate knowledge of: ○ General networking skills ○ Change control process ○ Project management ○ Risk assessment
Vendor Support	If defects in a program are discovered, vendor assistance may be necessary. At this stage, Level-3 engineers must establish a support ticket with the appropriate vendor to find a solution.		
Self-Service	A self-service portal should be utilized for non-critical tasks such as application access, permissions, password resets, etc. The portal can range from a simple FAQ page to a fully automated process requiring no human interaction. The purpose of the self-service portal is to add an additional touch point for end users to address basic issues, preventing the creation of new support tickets.		

Decision: Certs and Training

The following table details each support level's recommended training, certifications, and experience.

Role	Recommended Training	Recommended Course(s)	Recommended Certifications	Relevant Experience
Help Desk (Level 1)	Level-1 support personnel should be provided with basic training on Citrix Virtual Apps and Desktops and/or Citrix DaaS. This can include internal training from subject matter experts or an authorized Citrix Learning Center. The training provided should focus on the following topics: High-level overview of the Citrix implementation Using Citrix Director to manage user sessions Troubleshooting Citrix sessions Troubleshooting Methodology	CWS-250: Citrix DaaS Deployment and Administration Course Description		1+ years (Entry level also acceptable)

Role	Recommended Training	Recommended Course(s)	Recommended Certifications	Relevant Experience
Operator (Level-2)	Level-2 personnel should conduct regular team training sessions to refine administrative skills and ensure a baseline knowledge level across the team. Formalized training is also essential when there are architectural updates to the environment, and the Level-2 team is working with unfamiliar technologies. All members of the Level-2 team should achieve the Citrix Certified Associate (CCA) certification. Advanced training on Windows concepts will also be essential for Level-2 team members who do not have desktop or server support experience. Finally, on-the-job training and close integration with Level-3 administrators is essential as the Level-2 roles are formalized and responsibilities are handed over from Level-3 to Level-2.	CWS-215: Citrix Virtual Apps and Desktops 7 Administration On-Premises and in Citrix Cloud Course Description	204 - Citrix Virtual Apps and Desktops 7 Administration Exam Preparation Guide	2-3 years

Role	Recommended Training	Recommended Course(s)	Recommended Certifications	Relevant Experience
Implementer (Level-3)	Level-3 support team members hold at least three years of enterprise experience implementing and supporting Citrix Virtual Apps and Desktops and/or Citrix DaaS, Citrix Provisioning, and Windows operating systems. Level-3 staff should also complete the Citrix Certified Professional (CCP) certification track, which will prepare them to proactively manage the user community and implement Citrix solutions according to Citrix's leading practices.	CWS-322: Citrix Virtual Apps and Desktops 7 Advanced Course Description	312 - Citrix Virtual Apps and Desktops 7 Advanced Administration Exam Preparation Guide	3-4 years
Architect (Level 4)	Experience is essential for Level-4 staff. A qualified Level-4 resource should have a minimum of five years of experience implementing, supporting, and serving in a technology architect role for a Citrix Virtual Apps and Desktops or Citrix DaaS environment, as well as additional administrative experience with integrated technologies such as application and profile management solutions. The ideal candidate will have served in such a capacity in two or more environments for purposes of product exposure and in at least one environment of over 1,200 concurrent users. A Citrix Certified Expert (CCE) certification or comparable training and experience should be a prerequisite for the role.	CWS-415: Citrix Virtual Apps and Desktops 7 Architect Design Solutions Course Description	403 - Citrix Virtual Apps and Desktops 7 Assessment, Design, and Advanced Configurations Exam Preparation Guide	5+ years

Decision: Support Staffing

The following table guides the recommended number of support staff.

Role	Small Environment Sites: 1 Users: <500 Images: 1-2	Mid-size Environment Sites: 1-2 Users: 1000-5000 Images: 3-5	Large Environment Sites: 2+ Users: >5000 Images: 5+	Enterprise Environment Sites: 2+ Users: >10K Images: 5+
Help Desk (L1)	3	5 -10	15 – 20	20+
Operator (L2)	1 - 2	2 - 3	4 - 5	5+
Implementer (L3)	1	1 -2	2 - 3	3+
Architect (L4)	1	1	1 - 2	2+

Note:

This table should only be used as a baseline. Support staffing decisions should be evaluated against an organization's defined requirements, projected workloads, and operational procedures. Multiple levels can be combined. For example, there may be insufficient design projects to have a dedicated architect role, or a more senior member of the Citrix team can act as an Operator and Implementer.

Decision: Job Aids

The following table details tools that should be made available to all support levels.

Tools	Details
Ticket Management System	Used to document customer information and issues. A typical ticket management system provides the following functionality: • Monitoring the queue of tickets. • Setting a limit on the number of open tickets. • Establish thresholds such as how long a certain ticket type should be answered. • Identifying a group of users or individuals who require higher-priority assistance. • Informing the user when their ticket is open, updated, or closed. • Provide an internal knowledge base for the support professionals to search for known resolved issues.
Help Desk Call Scripts	The first contact help desk personnel should have documented scripts to capture all relevant data while the user is on the phone. This practice also assists in proper triage and allows the next support level to perform research before customer contact. A sample call script is provided for reference in the appendix.
Remote Assistance Tools	Remote assistance tools are useful when troubleshooting user issues. Support technicians and administrators can remotely observe a user's actions.
Knowledge Base Articles	Documentation should be created and maintained in a knowledge base or library of known issues. Articles should be searchable for quick recovery. Knowledge bases help support staff quickly resolve known issues and reduce the need to perform time-consuming research.

Citrix Support Tools

The following provides recommendations on the Citrix support tools that should be available to each support level.

Tool: Activity Manager / Support Levels: Self Service

The Activity Manager is a Citrix Workspace and StoreFront feature that allows users to manage their resources by providing quick actions (Disconnect, Log out, Shut Down, Force Quit, Restart) on active applications and desktops from any device within the Citrix Workspace app.

Tool: Citrix Director / Support Levels: L1, L2, L3, L4

Citrix Director is a Citrix Virtual Apps and Desktops tool that overviews hosted desktops and application sessions. It enables support teams to perform basic maintenance tasks and to monitor and troubleshoot system issues.

Tool: Citrix Studio / Support Levels: L1, L2, L3, L4

Citrix Studio enables administrators to perform configuration and maintenance tasks for Citrix Virtual Apps and Desktops site-associated virtual desktops or hosted applications.

Tool: WEM Tool Hub / Support Levels: L2, L3, L4

The Citrix WEM Tool Hub is a collection of tools that aims to simplify the configuration experience for Workspace Environment Management (WEM). It includes the Windows Logon analysis tool, which can be used to view logon duration reports and get tips for optimizing login duration and troubleshooting.

Tool: Citrix Audio Diagnostic Tool: L2, L3, L4

The Citrix Audio Diagnostic Tool is a troubleshooting and monitoring tool for Citrix HDX Audio and can help identify root causes in the Citrix environment and surrounding infrastructure.

Tool: Citrix Connection Quality Indicator: L2, L3., L4

The Connection Quality Indicator (CQI) is a tool that notifies users of changes to user experience. It can assist users by pointing out issues that degrade the user experience, providing real-time data to find causes for lags in screen refreshes, and reducing the number of calls to help desks related to user experience issues.

Tool: Citrix Insight Services / Support Level: L3, L4

Citrix Insight Services (CIS) is an instrumentation, telemetry, and business insight generation platform. Its instrumentation and telemetry capabilities enable technical users (customers, partners, and engineers) to self-diagnose and fix problems and optimize their environments.

Tool: Citrix Analytics / Support Level: L3, L4

Citrix Analytics solutions allow organizations to detect and deflect potential threats and quickly address performance issues—long before security incidents occur or employees submit help desk tickets. Machine learning and artificial intelligence are used to provide real-time insights into user behavior and automate the process of preventing cybersecurity breaches, all while maintaining a reliable digital workspace experience for employees.

Tool: Citrix Provisioning Console / Support Levels: L3, L4

The Citrix Provisioning Console enables administrators to perform configuration and maintenance tasks for a Citrix Provisioning farm.

Citrix Insight Services

Administrators can utilize Citrix Insight Services to simplify the support and troubleshooting of the Citrix environment. Citrix Insight Services is run locally to collect environmental information. Online analysis capabilities analyze that information and provide administrators recommendations based on their Citrix environment and configuration. Additional information regarding Citrix Insight Services can be referenced in the Citrix Support article: CTX131233 - FAQ: Citrix Insight Services.

A full list of the tools Citrix Support provides for troubleshooting can be referenced in the Citrix Supportability Pack.

Citrix AlwaysOn Tracing

Citrix helps identify connection failures and reduces the need to reproduce a problem. As the name implies, it is “always on,” so traces are constantly captured. When issues occur, they are automatically captured. AlwaysOn Tracing has minimal impact on deployments, and the trace information is compressed as it is collected. The Citrix Telemetry Service retains a maximum of 10 MB of compressed recent trace information, with a maximum time limit of eight days.

Decision: Delegate Administration

Each support level must have sufficient rights to perform its role effectively. The following tables guide the recommended privileges per support level.

Citrix Virtual Apps and Desktops Delegated Rights

Admin Role	Description	Support Level
Help Desk Administrator	Can view Delivery Groups and manage the sessions and machines associated with those groups. Can see the Machine Catalog and host information for the Delivery Groups being monitored. Can also perform session and machine power management operations for the machines in those Delivery Groups.	L1
Full Administrator	Can perform all tasks and operations. A Full Administrator is always combined with the Entire scope.	L3, L4
Read-Only Administrator	It can see all objects in specified scopes in addition to global information but cannot change anything. For example, a Read Only Administrator with Scope=London can see all global objects (such as Configuration Logging) and any London-scoped objects (for example, London Delivery Groups). However, that administrator cannot see objects in the New York scope (assuming that the London and New York scopes do not overlap).	L1, L2
Machine Catalog Administrator	Can create and manage machine catalogs and provision the machines for them. Can build Machine Catalogs from the virtualization infrastructure, Provisioning Services, and physical machines. This role can manage base images and install software but cannot assign applications or desktops to users.	L2, L3, L4
Delivery Group Administrator	Can deliver applications, desktops, and machines; can also manage the associated sessions. Can also manage application and desktop configurations such as policies and power management settings.	L2, L3, L4
Host Administrator	Can manage host connections and their associated resource settings. Cannot deliver machines, applications, or desktops to users.	L2, L3, L4

Administrators can create or edit custom roles to enable only the necessary permissions for a support level and assign them appropriately. Visit the product documentation for more information on creating and managing custom roles.

Citrix Provisioning Delegated Rights

Admin Role	Description	Support Level
Farm Administrator	Farm administrators view and manage all objects within a farm, create sites, and manage role memberships throughout the entire farm.	L3, L4
Farm read-only Administrator	Farm administrators can view all objects within a farm.	L1
Site Administrator	Site administrators have full management access to all the objects within a site.	L2
Device Administrator	Device administrators manage device collections to which they have privileges. Management tasks include assigning and removing a virtual disk from a device, editing device properties, and viewing read-only virtual disk properties.	L2, L3, L4
Device operator	A device operator has administrator privileges to perform target device boot, reboot, and shut down operations.	L1, L2, L3, L4

For further information about delegated rights within a Provisioning Site, please refer to Citrix Provisioning Administrative roles.

Citrix StoreFront Delegated Rights

Admin Role	Support Level
N/A	L1 and L2
Local Administrator on StoreFront Server	L3
Full Administrator	L4

Users with local administrator rights can view and manage all objects within StoreFront or Web Interface. They can also create new sites and modify existing ones.

Citrix License Server Delegated Rights

Admin Role	Support Level
N/A	L1
N/A	L2
Administrator	L3 and L4

The account used during the license server installation becomes the console administrator by default. These accounts are often not intended for regular administration tasks. Please reference this documentation for the steps to change the default administrator.

Process 2: Operations

This section defines routine operations for the Citrix environment that help to improve stability and performance.

Decision: Administrative Tasks

The Citrix Support Team should perform regular operations and maintenance tasks to ensure a stable, scalable Citrix environment.

Each operation is categorized by the solution's associated component and the operation's frequency (ongoing, daily, weekly, and yearly). Tasks have been aligned to the roles described within Decision: Support Responsibilities and Skill Set.

If the administrators performing operations are the same the support team, then the designations are linked as follows:

- Level 2 Support = Operators
- Level 3 Support = Implementers
- Level 4 Support = Architect

Daily Periodic Tasks

The following table outlines the tasks that should be performed by the Citrix Support Team daily.

Component	Task	Description	Responsible
Generic	Review Citrix Director, Windows Performance Monitor, Event Log, and other SIEM alerts	Check for warnings or alerts within Citrix Director, event logs, or other monitoring software. If an alert occurs, investigate the root cause.	Operators

Component	Task	Description	Responsible
	Verify backups are completed successfully	Verify all scheduled backups have been completed successfully. This can include but is not limited to: • User data (user profiles/home folders) • Application data • Citrix databases • StoreFront configuration • Provisioning Services vDisks (virtual desktops and application servers) • XenServer VM/Pool metadata (or equivalent for other hypervisors) • Dedicated virtual desktops • License files	Operators
	Test environment access	Simulate an internal and external connection to ensure desktop and application resources are available before most users log on for the day. This should be tested throughout the day and may even be automated.	Operators
Citrix Virtual Apps and Desktops	Virtual machine power checking	Verify that the appropriate number of idle desktops and application servers are powered on and registered with the Delivery Controllers to ensure availability for user workloads.	Operators
	Perform incremental backup of Citrix-related databases	Perform incremental data backups of the following Citrix databases: • Site Database • Logging Database • Monitoring Database • WEM Database	Operators, Database team (if Citrix environment is using a shared SQL)
Citrix Provisioning	Check Citrix Provisioning Server utilization	If necessary, check the number of target devices connected to the Citrix Provisioning Servers and balance the load across the servers.	Operators

Component	Task	Description	Responsible
	Perform incremental backup of Citrix Provisioning database	Incremental backup of Citrix Provisioning Server database hosted on SQL Server infrastructure.	Operators, Database team (if Citrix environment is using a shared SQL)

Weekly Periodic Tasks

Component	Task	Description	Responsible
Generic	Review the latest hotfixes and patches	Review, test, and deploy the latest Citrix hotfixes and ascertain whether the Cirix Infrastructure and Server-Based OS / Desktop-Based OS virtual machines require them. **Note:** Any required hotfixes should be tested using the recommended testing process before implementation in production.	Operators, Implementers (review process)
	Create Citrix environmental baselines and a status report	Create a report on overall environment performance (server health, resource usage, user experience) and the number of Citrix issues (close rate, open issues, and so on).	Operators
	Periodically Review the status report	Review the Citrix status report to identify any trends or common issues.	Implementers, Architect
	Maintain internal support knowledge base	Create knowledge-based articles and issue resolution scripts to address Level-1 and Level-2 support requests. Review these articles and scripts for accuracy, compliance, and feasibility.	Operators (Level-2 requests), Implementers(Level-3 requests and review process)
Citrix Virtual Apps and Desktops	Check Configuration Logging reports	Confirm Citrix site-wide changes implemented during the previous week were approved through change control.	Auditors

Component	Task	Description	Responsible
	Perform a full backup of Citrix-related databases	Perform full-data backups of the following Citrix databases: • Site Database • Logging Database • Monitoring Database • WEM Database	Operators, Database team (if Citrix environment is using a shared SQL)
Citrix Provisioning	Check storage capacity (only before updating a vDisk)	Review storage utilization used and free storage space for the vDisk store and each vDisk. **Note:** Lack of space within the vDisk repository will be an issue only when the vDisks are updated using versioning or when a vDisk is placed in private mode during an update procedure. Storage utilization within vDisk should also be investigated. For example, a 20GB vDisk may only have 200MB of free storage. If the vDisk itself is limited for storage, it must be extended. Citrix does not support resizing a VHD file. Refer to the Microsoft link Resize-VHD for information on resizing a VHD file.	Operators
	Check auditing reports	Review the Citrix Provisioning Services auditing Logs. Note: Provisioning Server auditing is off by default and can be enabled to record configuration actions on components within the Provisioning Services farm. To enable auditing, refer to Enabling Auditing	Auditors
	Perform a full backup of the Citrix PVS database	Backup of Citrix Provisioning Server database hosted on SQL Server infrastructure.	Operators, Database team (if Citrix environment is using a shared SQL)

Monthly Periodic Tasks

The following table outlines the tasks that should be performed by the Citrix Support Team monthly.

Component	Task	Description	Responsible
Generic	Perform capacity assessment	Perform capacity assessment of the Citrix environment to determine environment utilization and scalability requirements.	Architect
	Review software upgrades	Review and assess the requirement for new Citrix software releases or versions. (Quarterly)	Architect

Yearly Periodic Tasks

Component	Task	Description	Responsible
Generic	Conduct Citrix policy assessment	Review Citrix policies and determine whether new policies are required and existing policies need to be updated.	Implementers
	Perform Business Continuity Plan (BCP)/ Disaster Recovery (DR) test	Conduct functional BCP/DR test to confirm DR readiness. This plan should include a yearly restore test to validate that the actual restore process from backup data is functioning correctly.	Implementers
	Perform application assessment	Review the usage of applications outside and within the Citrix environment. Assess the validity of adding additional applications to the Citrix site, removing no longer required applications, or upgrading the applications to the latest version.	Architect
Citrix Provisioning	Archive audit reports	Perform an archive of the Citrix Provisioning Server Audit Trail Information for compliance requirements.	Auditors

Decision: Backup Location

The location of backups directly affects the recovery time and reliability of the Citrix environment. Critical data backups should be stored both on-site and off-site. If off-site backups are not possible due to the costs associated with them or the sensitivity of the data, backups should be placed at separate physical locations within the same datacenter.

Each backup option is discussed further below.

- Onsite Backups — Onsite backups should be located on a storage device in the datacenter that will allow the data to be recovered quickly in the event of a failure. They are ideal for issues affecting only a small subset of datacenter hardware. Backups can also be stored on a cold storage solution such as tape. While this medium is slower to recover, it provides additional protection since it is only active during the backup process.
- Offsite Backups — Although the recovery time is much higher, offsite backups provide additional protection during a disaster. Offsite backups may require transferring data

over the Internet to a third-party provider or are created onsite and transported to a remote location on storage mediums such as tape. It is typical to put a limited number of backups offsite, for example, one backup a week or month.

Decision: Testing Process

Regular updates and maintenance are an everyday part of IT operations. Standard processes must be followed to ensure updates do not negatively impact the production environment. This includes maintaining a dedicated testing infrastructure to validate modifications before production implementation.

Since changes to Citrix infrastructure can impact thousands of virtual desktop and application users, multi-phase testing is critical for the reliability and performance of the environment. As such, the process for testing should resemble the following:

- Development — The development infrastructure exists outside of the production network. Typically, it consists of short-lived virtual machines whose configuration matches production as closely as possible. The purpose of the development phase is to provide change requestors with a non-production environment to perform proof of concepts, determine integration requirements, and perform iterative testing as part of a discovery phase. Proposed changes should be documented and applied in the test phase.
- Testing - The test environment is a standalone 1:1 copy of the production infrastructure and confirms that the proposed changes can be easily repeated before the pre-production staging environment. The changes made should follow documentation from the development stage. If testing fails within the testing stage, the architect must determine the severity of failure and determine whether minor updates to documentation is sufficient or a full development cycle is needed.
- Pre-production — The pre-production environment should mimic the current production environment. Staging aims to implement the proposed changes with little risk or uncertainty. Any changes to the staging infrastructure are expected to be tested and documented for repeatability. No iterations or adjustments should be required during this phase. During this phase and within this environment, User Acceptance Testing (UAT) should be performed.
- Production — The production environment is a fully redundant and scalable solution designed for normal use by end users. There should be minimal changes to the environment. All approved changes should be rolled out in stages to the production environment if possible. This process is known as a staged rollout and mitigates risk by

allowing changes to be rolled back, if necessary, without entirely using a normal environment.

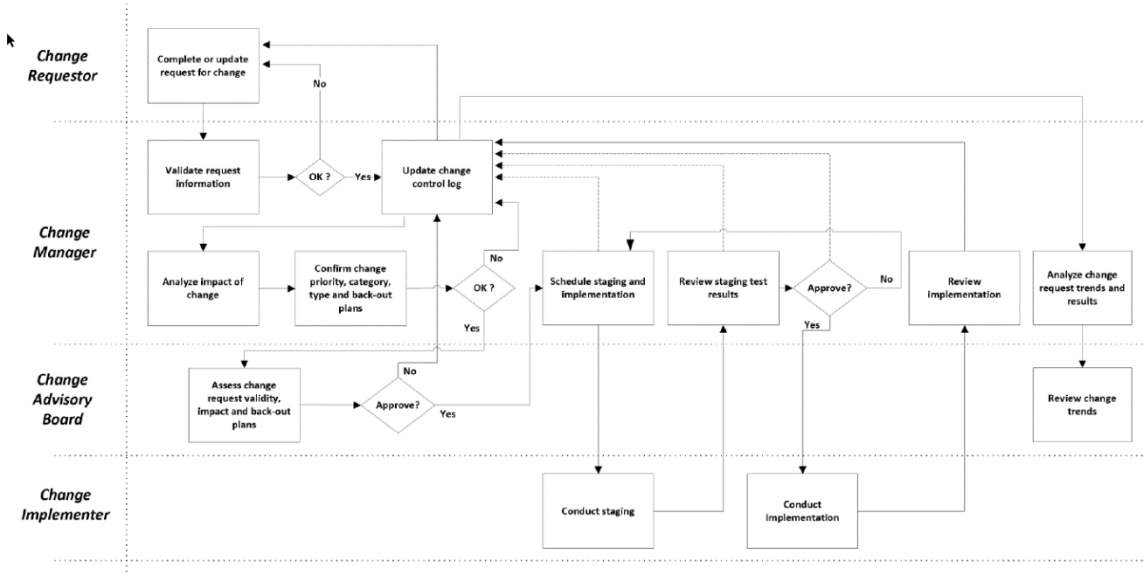
Decision: Change Control

Standardized processes that manage changes throughout a system's lifecycle are necessary to ensure consistent and accountable performance. The following change control leading practices should be considered.

- Use a change control window so that all applicable parties know when downtime might occur.
- Ensure all teams are represented in the Change Advisory Board (CAB).
- Every change should have a rollback plan.
- If a change fails, have a debrief to determine what went wrong.
- Always use an automated change control system, so support staff can quickly and easily identify changes.
- When available, ensure configuration logging is enabled to track any changes made to the Citrix environment.

The change control process should be closely followed, starting with a change request. A change request form should be filled out detailing the changes requested, reasons for the change, and intended timeframes for the action. This is then reviewed and edited if required by a Change Manager and advisory board. When the change request has gone through the entire change approval process, it is given to a change implementer who stages the change for testing and finally conducts the implementation in production.

A sample change control process, including detailed steps, is provided in the diagram below:



The process is as follows:

1. Any person requesting a change completes the Change Request (CR) form.
2. After appropriate manager approvals have been acquired, the CR is forwarded to the appropriate Change Manager(s).

3. The Change Manager validates the CR for completeness and logs the CR information into the Change Control Log for tracking. Incomplete change requests are returned to the requestor for update and re-submission.
4. The Change Manager assesses the change's impact in conjunction with subject matter experts and/or managers of the teams associated with/affected by it.
5. The Change Manager works with the associated/affected teams and the change requester to confirm the change's priority, category, and type, as well as the proposed rollback plan.
6. If the change is approved by the Change Manager, the CR is forwarded to the CAB for approval. If the change is rejected, the Change Control Log is updated with the current status and the reason for the rejection, and the CR is sent back to the requestor.
7. The CAB reviews and validates the change in detail and discusses and evaluates the purpose, reasons, impact, cost, and benefits. Each board member represents their department and provides guidance on the change requests. The CAB reviews multiple requests to coordinate implementations and "package" requests into a single release schedule.
8. Upon approval, the change is sent back to the change manager to schedule for implementation in the staging environment.
9. The change is implemented, and tests are conducted. The results are sent back to the Change Manager.
10. The change is scheduled for production implementation if the staging implementation and testing are successful. If the staging phase is unsuccessful, another staging iteration will be conducted.
11. The change is rolled out in stages to the production environment if possible. This process is known as a staged rollout and mitigates risk by allowing changes to be rolled back, if necessary, without impacting the entire environment. A rollback plan should be in place if there is an issue implementing a change in the production environment.
12. The Change Manager reviews the implementation and finally updates the Change Control Log.
13. The Change Manager periodically reviews the Change Control Log to identify trends in changes' type, frequency, and size and forwards the results to the CAB for review.

The processes may be expedited in an emergency. Should an issue be declared an emergency, a change request form is still filled out and delivered to the appropriate change management representative. The requested change is immediately implemented when approved, and the advisory board is notified.

Decision: Availability Testing

Availability testing ensures resources are still available during a failure. These tests are essential to ensure users always have access to business-critical resources. The testing should be conducted during non business hours or during a scheduled maintenance weekend when appropriate notice has been given to end users to inform them if any unforeseen issues arise.

The following is a list of the key components that should be tested regularly.

- **StoreFront**—StoreFront should be load-balanced and health-checked by a NetScaler or other load-balancing device. All but one of the StoreFront servers should be shut down to validate its configuration. This will ensure that the load-balancing device detects the failure and directs users to the functioning server.
- **SQL**—The SQL Server should be in a high-availability configuration. The primary SQL server should be taken offline to validate the configuration, and then the Citrix Studio console should be opened. Since Citrix Studio will not be accessible without a

functioning SQL server, it will validate that the SQL server failover mechanisms are functioning properly.

- **Controllers** — The deployed resources should be configured with a list of multiple Controllers. If one is unavailable, desktops and application hosts will automatically establish a connection to another server in the list. To validate this, shut down one of the controller hosts and determine if the resources initially connected to it are automatically registered to another server. This can be determined by viewing the registration status of the resources inside Citrix Studio.
- **Service Continuity -** If using Citrix Workspace, Service Continuity should be enabled. To validate Service Continuity is functioning properly, ensure lease files are downloaded, edit the local host file on your endpoint so that 127.0.01 points to your Workspace URL (<storename>.cloud.com), run ipconfig /flushdns, then exit and relaunch the Citrix Workspace app. Once validated, remove the entry in your local host file. For more information on Service Continuity configuration, see this Citrix Tech Brief.
- **Local Host Cache** - Local Host Cache is a combination of several services and components that come together to take over the brokering responsibilities until the connection to the Cloud Broker or SQL database can be reestablished. In DaaS, Local Host Cache engages when Cloud Connectors lose connectivity with Citrix Cloud for 60 seconds. For customer-managed Sites, Local Host Cache engages when Delivery Controllers lose connectivity to the Site database for 90 seconds. Review the information in this Citrix Tech Brief to test your Local Host Cache configuration.

Sample Testing Workflow: Citrix Provisioning

Prerequisites and configuration requirements:

- Hypervisor, Citrix Virtual Apps and Desktops services are up and running.
- At least two Citrix Provisioning servers are installed and configured to provide the streamed disk image.
- Resilient networking and storage infrastructure with multiple links to each server.
- Test users are active on the Citrix Virtual Apps and Desktops machines.

Steps	Expected Results
Citrix Provisioning Server Outage • Shutdown one of the Citrix Provisioning Servers. • Validate Citrix Provisioning continues to function. • Restart Citrix Provisioning Server. • Validate connections rebalance between Citrix Provisioning Servers. • Try the other(rest) Citrix Provisioning server(s) individually.	• Existing Citrix Virtual Apps and Desktops machines connect to another PVS server. • There is limited to no impact to the users utilizing that server. • New Citrix Virtual Apps and Desktops machines can be booted and start correctly. • SCOM reports that the PVS server is down / unavailable. • Live connections are rebalanced between both PVS servers once both PVS servers are made available again.
SQL Server Citrix Provisioning Database Outage • Admin reboots SQL Servers simultaneously • Validate Citrix Provisioning continues to function, but that administration is impossible. • Wait for the SQL server to come back online. • Validate Citrix Provisioning administrative functions are once again possible.	• Citrix Provisioning continues to function. • Citrix Provisioning administrative functions are no longer available. • Citrix Provisioning administrative functions are available once the SQL service has been restored.

Sample Testing Workflow: Citrix Virtual Apps and Desktops

Prerequisites and configuration requirements:

- Hypervisor, Citrix Virtual Apps and Desktops services are up and running.
- Network and storage services are available.
- Citrix Provisioning provides the streamed disk images.
- Citrix Workspace Environment Management (WEM) services are up and running.
- Test users are active on the virtual machines.
- SQL and Citrix Virtual Apps and Desktops servers are up and running.
- Ensure multiple StoreFront servers are running.
- NetScaler load balancing services.

Steps	Expected Results
Controller Citrix Broker Service Outage • Stop the Citrix Broker Service on one of the Controller servers. • Validate virtual desktops or applications can still be enumerated and launched. • Start the Citrix Broker Service on the Controller server. • Shutdown one of the Controllers. • Validate virtual desktops or applications can still be enumerated and launched. • When a desktop is launched, determine which Controller owns the host connection. Shut the Controller down and verify that another Controller takes over the session. **Note:** This should be done during the maintenance window. Once complete, the VDI resources should be rebooted to distribute the VDAs evenly across all controllers.	• StoreFront correctly identifies service as unavailable and redirects connections to the remaining Controller. • Desktops continue to be enumerated and launched successfully. • Launched desktop can be supported if a hosting Controller goes down.
SQL Service Outage: • Admin restarts servers in the SQL Always on Availability group. • Validate Citrix Virtual Apps and Desktops continue functioning, but that administration is impossible. • Wait for the SQL Service to come back online. • Validate administrative functions are once again possible.	• Existing sessions are not impacted • The local host cache allows access to recently used applications, hosted shared desktops, and assigned VDI. • Administrative functions are not possible • Administrative functions are possible once SQL service is available.

Sample Testing Workflow: Citrix License Server

Prerequisites and configuration requirements:

- Citrix Licensing Server up and running (with valid licenses installed).
- Hypervisor, Citrix Virtual Apps and Desktops, and StoreFront services are up and running.
- Users are active on the Server OS or Desktop OS machines.

Steps	Expected Results
Service continuity during the complete failure of the Citrix Licensing Server: • Shutdown the Citrix Licensing server. • Reboot an existing Server OS machine. • Log on to the Citrix StoreFront and launch a published application. • Reboot an existing Desktop OS machine. • Log on to the Citrix StoreFront and launch a virtual desktop.	• License Server connectivity error posted in Event Log. • Provisioned Server OS boots successfully. • Users can launch published applications. • Provisioned Desktop OS boots successfully. • Users can launch a virtual desktop. • Administrators will have 30 days' grace to recover the Citrix Licensing Server.

Process 3: Monitoring

By having an in-depth understanding of the Citrix environment and its components' current and expected behavior, administrators are better equipped to discover an issue before it affects the user community. Furthermore, the data tracked during normal operations benefits trending and capacity planning. This section defines the monitoring recommendations for a Citrix environment and some recommended tools.

Decision: Performance Monitor Metrics

Monitoring the overall environment's performance is crucial to ensuring that all components are available and performing effectively and that users have a high-quality experience.

Different components within the overall solution require monitoring of unique metrics with appropriately set thresholds. The metrics and thresholds presented are based on real-world experience but may not apply to all environments. Organizations must perform their own baseline, validity testing, and validation before implementing them within a production environment.

> Note:
>
> Some hypervisors, such as VMware vSphere and Hyper-V, provide specific performance counters for tracking CPU and Memory utilization within virtual machines (i.e., “VM Processor \ % Processor Time”). These performance counters should be used in addition to the general counters listed below.

General

These performance counters should be used to monitor the key performance metrics of the Citrix infrastructure, application servers, and virtual desktops.

Metric	Description	Warning (Yellow)	Critical (Red)	Troubleshooting/Remediation
Processor - % Processor Time	% Processor Time is the percentage of elapsed time the processor spends executing a non-Idle thread. It is calculated by measuring the duration of the idle thread is active in the sample interval and subtracting that time from interval duration. (Each processor has an idle thread that consumes cycles when no other threads are ready to run). This counter is the primary indicator of processor activity and displays	80% for 15 minutes	95% for 15 minutes	Identify the processes/services consuming processor time using Task Manager or Resource Monitor. If all processes/services work within normal parameters and the level of CPU consumption is expected, adding additional CPU resources to this system in the future should be considered. If a process or service that works outside normal parameters can be identified, it should be killed. Please note that killing a process can cause unsaved data to be lost.

Metric	Description	Warning (Yellow)	Critical (Red)	Troubleshooting/Remediation
	the average percentage of busy time observed during the sample interval. It is calculated by monitoring the service's inactive time and subtracting that value from 100%.			
System - Processor Queue Length	Processor queue length is the number of threads in the processor queue. Unlike the disk counters, this counter shows ready threads only, not threads that are running. There is a single queue for processor time, even on computers with multiple processors. Therefore, if a computer has multiple processors, you must divide this value by the number of processors servicing the workload. Depending on the workload, a sustained processor queue of less than ten threads per processor is normally acceptable.	5 (per core) for 5 minutes or 6 (per core) for 15 minutes	10 (per Core) for 10 minutes or 12 (per core) for 30 minutes	A long CPU queue is a clear symptom of a CPU bottleneck. Please follow the steps outlined for the counter “Processor - % Processor Time.”
Memory – Available Bytes	Available memory indicates the amount of memory that is left after nonpaged pool allocations, paged pool allocations, process’ working sets, and the file system cache have all taken their piece.	<30% of total RAM or 20% of physical memory over 6 minutes	<15% of total RAM or 5% of physical memory over 6 minutes	Identify the processes/services consuming memory using Task Manager or Resource Monitor. If all processes/services work within normal parameters and the level of memory consumption is an expected behavior, it should be considered to add additional

Metric	Description	Warning (Yellow)	Critical (Red)	Troubleshooting/Remediation
				memory to this system in the future. If a process/service can be identified that works outside normal parameters, it should be killed. Please note that killing a process can cause unsaved data to be lost.
Memory – Pages/sec	Pages/sec is the rate at which pages are read from or written to disk to resolve hard page faults.	>10	>20	A high value reported for this counter typically indicates a memory bottleneck, except if “Memory – Available Bytes” reports a high value simultaneously. In this case, most likely, an application is sequentially reading a file from memory.
Paging File - %Usage	This is the percentage amount of the Page File instance in use.	>80% over 60 minutes	>95% over 60 minutes	Review this value in conjunction with “Memory - Available Bytes” and “Memory - Pages/sec” to understand paging activity on the affected system.
LogicalDisk/PhysicalDisk - % Free Space	% Free Space is the percentage of total usable space on the selected logical disk drive that is free.	<20% of physical disk or 20% reported after 2 minutes	<10% of physical disk or 15% reported after 1 minute	Identify which files or folders consume disk space and delete obsolete files if possible. In case no files can be deleted, consider increasing the size of the affected partition or add additional disks.
LogicalDisk/PhysicalDisk - % Disk Time	% Disk Time marks how busy the disk is.	>70% over 15 minutes	>90% over 15 minutes	Identify the processes/services consuming disk time using Task Manager or Resource Monitor. If all processes/services work within normal parameters and the level of disk consumption is an expected behavior, it should be considered to move the affected partition to a

Metric	Description	Warning (Yellow)	Critical (Red)	Troubleshooting/Remediation
				more capable disk subsystem in the future. If a process/service that works outside normal parameters can be identified, the process should be killed. Please note that killing a process can cause unsaved data to be lost.
LogicalDisk/PhysicalDisk – Current Disk Queue Length	Current disk queue length provides a primary measure of disk congestion. It is an indication of the number of transactions that are waiting to be processed.	>= 3 over 15 minutes	>=10 over 30 minutes	A long disk queue length typically indicates a disk performance bottleneck. This can be caused by either processes/services causing a high number of I/Os or a shortage of physical memory. Please follow the steps outlined for counter “LogicalDisk/PhysicalDisk - % Disk Time” and counter “Memory – Available Bytes”
LogicalDisk/PhysicalDisk – Avg. Disk Sec/Read – Avg. Disk Sec/Write – Avg. Disk Sec/Transfer	The Average Disk Second counters show the average time of a read/write/transfer from or to a disk in seconds.	>=15ms consistently	>=20ms consistently	High disk read or write latency indicates a disk performance bottleneck. The affected systems will become slow and unresponsive, and applications or services may fail. Please follow the steps outlined for counter “LogicalDisk/PhysicalDisk - % Disk Time”
Network Interface – Bytes Total/sec	Bytes Total/sec shows the rate at which the network adaptor processes data bytes. This counter includes all application and file data, as well as protocol information, such as packet headers.	60% of NIC speed inbound and outbound traffic for 1 min.	70% of NIC speed inbound and outbound traffic for 1 min.	Identify the processes/services consuming the network using Task Manager or Resource Monitor. If all processes/services work within normal parameters and the level of bandwidth consumption is expected, it should be considered to move the respective process/service to a

Metric	Description	Warning (Yellow)	Critical (Red)	Troubleshooting/Remediation
				dedicated NIC (or team of NICs). If a process or service can be identified that works outside normal parameters, it should be killed. Please note that killing a process can cause unsaved data to be lost.

StoreFront

These performance counters are specific to the StoreFront servers.

Metric	Description	Warning (Yellow)	Critical (Red)	Troubleshooting/Remediation
ASP.NET – Request Queued	The number of requests waiting to be processed by ASP. To establish threshold values accurately, a baseline must be established in the environment.	Based on baseline values	Based on baseline values	Requests may be rejected if the queue length exceeds the critical limit. In this case, it should be considered that additional StoreFront or Web Interface servers be added to the load balancing team to distribute the load across more nodes.
ASP.NET – Requests Rejected	The number of requests rejected because the request queue was full.	None	>=1	When this limit is exceeded, requests will be rejected with a 503 status code and the message "Server is too busy." Please follow the steps outlined for counter “ASP.NET – Request Queued”
APP_POOL_WAS\ Current Application Pool State\Citrix Receiver for Web	A value of 5 indicates the Application Pool is stopped and Receiver for Web will display an error	-	5	Investigate the other performance counters, events, and configuration of the failing node on this counter.
APP_POOL_WAS\ Current Application Pool State\Citrix Delivery Services Authentication	A value of 5 indicates the Application Pool is stopped and StoreFront	-	5	

	Authentication will not be successful.			
APP_POOL_WAS\ Current Application Pool State\Citrix Delivery Services Resource	A value of 5 indicates the Application Pool is stopped and Gateway or PNAgent access will not be successful.	-	5	
Request response	Time taken for requests to be processed by StoreFront (Enumeration, Authentication app launch)	3 seconds	5 seconds	

Decision: Services Monitoring

Windows services critical to basic server functionality should be automatically monitored to ensure they are running properly. The following table lists the common Windows services that should be monitored. A warning (Yellow) or critical (Red) alert should be assigned when any of these services are restarted or stopped. The recommended recovery actions for the services listed below are as follows:

- First failure: Restart the Service
- Second Failure: Restart the Service
- Subsequent Failures: Put the server in maintenance mode and investigate the root cause.

Citrix Virtual Apps and Desktops

Service	Functionality	Administration Risk
Citrix AD Identity Service	Manages Active Directory computer accounts. Dependencies: WMI Service	Machine Creation Service relies on this service to create virtual machines. Administrators will be unable to create new or modify existing Machine Catalogs or establish new connections to Citrix Studio.
Citrix Broker Service	Manages connections to virtual machines and applications.	Citrix Virtual Apps and Desktops cannot communicate with the Configuration Logging Database if this service is stopped. Administrators cannot change the environment or establish new connections to Citrix Studio.

Service	Functionality	Administration Risk
Citrix Configuration Service	Stores service configuration information. Dependencies: • WMI Service	If this service is stopped, administrators cannot change the environment or establish new connections to Citrix Studio.
Citrix Delegated Administration Service	Manages configuration of delegated administration permissions.	If this service is stopped, Citrix cannot assign administrative permissions. Administrators cannot change the environment or establish new connections to Citrix Studio. Administrators will be unable to establish new connections to the Citrix Director, and existing sessions with Citrix Director will be interrupted.
Citrix Diagnostic Facility COM Server Service	Manages and controls Citrix diagnostic trace sessions on the system. Dependencies: RPC Service	This service does not impact the production environment. It generates CDF trace files, which aid in troubleshooting issues.
Citrix Environment Test Service	Manages tests for evaluating the state of a Citrix site.	If this service is stopped, administrators cannot establish new connections to Citrix Studio. Administrators will also be unable to check the status of the Citrix site configuration, machine catalogs, and delivery groups by running the tests under “Common Tasks” in the Citrix Studio administration console.,
Citrix Host Services	Manages host and hypervisor connection. Dependencies: WMI Service	Administrators cannot create new Machine Catalogs or control virtual machine power settings via Citrix Studio. Administrators will be unable to establish new connections to Citrix Studio. Users may experience issues connecting to virtual desktops when this service is unavailable. If this service is stopped, existing connections will not be affected.
Citrix Machine Creation Service	Creates new virtual machines. Dependencies: WMI Service	Administrators cannot create new or modify existing Machine Catalogs or establish new connections to Citrix Studio. Administrators will be unable to establish new connections to Citrix Studio.
Citrix Monitor Service	Monitors the system	If this service is stopped, Citrix will be unable to communicate with the Monitoring Database, Citrix Director will be unable to retrieve any data on the environment, and administrators will be unable to establish new connections to Citrix Studio.

Service	Functionality	Administration Risk
Citrix StoreFront Service	Manages deployment of StoreFront.	Administrators will be unable to establish new connections to Citrix Studio.

Citrix Cloud Connectors

Service Name	Description	Startup Type	Log On As	Dependencies
Citrix CDF Capture Service	Captures CDF traces from all configured products and components.	Automatic	Network Service	N/A
Citrix Cloud Connector Metrics Service	This service is responsible for collecting metrics from the NetScaler Gateway Service and generating Synthetics. It forwards all the data from the Cloud Connector to the Citrix Analytics Service platform.	Automatic (Delayed Start)	Network Service	N/A
Citrix Cloud Services AD Provider	Enables Citrix Cloud to facilitate the management of resources associated with the Active Directory domain accounts in which it is installed.	Automatic (Delayed Start)	Network Service	N/A
Citrix Cloud Services Agent Logger	Provides a support logging framework for the Citrix Cloud Connector services.	Automatic	Network Service	Citrix Cloud Agent System
Citrix Cloud Agent System	Handles the system calls necessary for the on-premises agents. Includes installation, reboots, and registry access. Can only be called by Citrix Cloud Services Agent WatchDog.	Automatic	NT AUTHORITY\SYSTEM	N/A
Citrix Cloud Services Agent WatchDog	Monitors and upgrades the on-premises agents (evergreen).	Automatic	Network Service	Citrix Cloud Agent System
Citrix Cloud Services Credential Provider	Handles storage and retrieval of encrypted data.	Automatic (Delayed Start)	Network Service	N/A

Service Name	Description	Startup Type	Log On As	Dependencies
Citrix Cloud Services WebRelay Provider	Enables HTTP Requests received from WebRelay Cloud service to be forwarded to On-Premises Web Servers.	Automatic (Delayed Start)	Network Service	N/A
Citrix ClxMtp Service	Citrix ClxMtp Service	Automatic (Delayed Start)	Network Service	N/A
Citrix Config Synchronizer Service	Copies brokering configuration locally for high availability mode.	Automatic	Network Service	N/A
Citrix High Availability Service	Provides continuity of service during an outage of the central site.	Automatic	Network Service	N/A
Citrix ITSM Adapter Provider	Automates provisioning and management of virtual apps and desktops.	Automatic	Local System	N/A
Citrix NetScaler Cloud Gateway	Provides Internet connectivity to on-premises desktops and applications without the need to open in-bound firewall rules or deploying components in the DMZ.	Automatic	Network Service	N/A
Citrix Remote Broker Provider	Enables communication to a remote Broker service from local VDAs and StoreFront servers.	Automatic	Network Service	N/A
Citrix RemoteHCLServer Service	Proxies communications between the Delivery Controller and the Hypervisor(s).	Automatic	Network Service	Workstation
Citrix WEM Cloud Authentication Service	Provides authentication service for Citrix WEM agents to connect to cloud infrastructure servers.	Automatic (Delayed Start)	Network Service	N/A
Citrix WEM Cloud Messaging Service	Provides service for Citrix WEM cloud service to receive messages from cloud infrastructure servers.	Automatic (Delayed Start)	Network Service	N/A

Citrix StoreFront

Service	Description	Impact on Failure
Citrix Credential Wallet	Provides a secure store of credentials.	Users will be unable to log in to access their desktops or applications. Users logged into StoreFront will also be unable to launch new applications or desktop sessions. Existing applications or desktop sessions are unaffected.
Citrix Default Domain Services	Provides authentication, password change, and other domain services.	Users will be unable to login to access their desktops or applications. Users currently logged in will not be affected.
Citrix Peer Resolution `	Resolves peer names within peer-to-peer meshes.	Both the Citrix Credential Wallet and Citrix Subscriptions store have stopped generating the risks associated with those services.
Citrix Subscription Store	Provides a store and replication of user subscriptions.	Citrix Workspace app cannot add, remove, or reposition applications within StoreFront. Users will need to re-add applications, and all changes made to their selection of applications within the StoreFront store will not be saved or replicated in other sessions. The original user configuration will be restored once the service is restarted.
World Wide Web Publishing Service	Provides web connectivity and administration through the Internet Information Services Manager.	Access to published applications or desktops will not be available through StoreFront. Users will be unable to resolve the Receiver for Web login page. Users logged into StoreFront will be unable to launch new applications or desktop sessions and will need to reenter credentials when the service is restarted. Existing applications or desktop sessions are unaffected.

Citrix Provisioning

Service	Functionality	Administration Risk
Citrix PXE Service	Provides the PXE Boot Server functionality. Note: Only applicable when PXE boot is used.	Target devices may not boot successfully if PXE booting is leveraged due to the failure of this service.
Citrix Stream Service	Streams contents of the vDisk to the target device on demand.	If this service is stopped, it will not be possible to stream vDisk images.
Citrix SOAP Service	Provides a framework for external or existing solutions	If this service fails, Provisioning Server to Provisioning Server communication and

	to interface with Provisioning services. Note: Only impacts console operations. User is unaffected	Provisioning Console to Provisioning Server communication is not possible.
Citrix TFTP Service	Provides the TFTP Server functionality. Note: Only applicable when TFTP is used.	If this service fails, target devices may be unable to boot if this server is used as a TFTP server for the bootstrap.
Citrix Two-Stage Boot Service	Provides the bootstrap functionality for booting devices using a BDM ISO file. Note: Only when BDM boot partitions are used.	If this service fails, target devices may be unable to boot if a BDM ISO file is used.

Citrix License Server

Service	Functionality	Administration Risk
Citrix Licensing Service	Provides licensing services for Citrix products.	Licensing mode changes to a grace period when service is stopped or the License Server cannot be contacted. If not monitored, the functionality of Citrix products will cease after the grace period expires.
Citrix Licensing Support Service	This account controls reading the license files and updating strings with license trailers (data dictionary functionality).	None
Citrix Licensing WMI	The Citrix License Management Console collects license data information using the WMI service.	None

Citrix Virtual Delivery Agent (VDA)

Service Name	Description	Startup Type	Log On As	Dependencies
Citrix Audio Redirection Service	Provides audio redirection between the endpoint device and the virtual desktop.	Automatic	Local Service	N/A
Citrix CDF Capture Service	Captures CDF traces from all configured products and components.	Manual	Network Service	N/A
Citrix CEIP Service for VDA	Citrix CEIP Service for VDA.	Automatic	Local Service	N/A

Service Name	Description	Startup Type	Log On As	Dependencies
Citrix Clipboard Service	Provides Clipboard virtual channel.	Automatic	Local Service	N/A
Citrix ClxMtp Service	Provides ClxMtp Protocol	Manual	Network Service	Remote Desktop Service
Citrix Control Channel Service	Provides Control Virtual Channel.	Automatic	Local Service	N/A
Citrix Desktop Service	The Citrix Desktop Service manages communication between the delivery controller and virtual desktops. It handles initial brokering of connections settings for connections and interaction with sessions.	Automatic	Network Service	Workstation
Citrix Device Redirector Service	Service to manage Citrix remote devices.	Automatic	Local Service	N/A
Citrix Diagnostic Facility COM Server	Manages and controls Citrix diagnostic trace sessions on the system.	Automatic	Network Service	Remote Procedure Call (RPC)
Citrix DND Service	Provides Drag and Drop command virtual channel.	Automatic	Local Service	N/A
Citrix Encryption Service	Secure ICA encryption.	Automatic	Local Service	Windows Management Instrumentation
Citrix End User Experience Monitoring Service	Service to collect and collate end-user experience measurements.	Automatic	Local Service	Citrix SMC Support Driver
Citrix Enterprise Browser Elevation Service (CitrixEnterpriseBrowserElevationService)	N/A	Manual	Local System	Remote Procedure Call (RPC)
Citrix GDT Service	Provides data transfer services to clipboard and drag and drop virtual channels	Automatic	Local Service	N/A
Citrix Group Policy Engine	Responsible for applying settings configured by Citrix administrators for the computer and users through the Group Policy component.	Automatic	Local System	Remote Procedure Call (RPC)
Citrix HDX Browser Redirection Service	Provides browser redirection between the endpoint device and the virtual desktop.	Automatic	NT AUTHORITY \SYSTEM	N/A

Service Name	Description	Startup Type	Log On As	Dependencies
Citrix HDX HTML5 Video Redirection Service	Provides HTML5 video redirection between the endpoint device and the virtual desktop.	Automatic	Local System	N/A
Citrix HDX MediaStream Service	Provides multimedia acceleration between the endpoint device and the virtual desktop.	Automatic	Local Service	N/A
Citrix HDX Port Forwarding Service	Provides port forwarding between the endpoint device and the virtual desktop.	Automatic	Local Service	N/A
Citrix HDX Teams Redirection Service	Provides teams redirection between the endpoint device and the virtual desktop.	Automatic	NT AUTHORITY \SYSTEM	N/A
Citrix Local User Service Manager	Citrix Local User Management Service.	Automatic	NT AUTHORITY \SYSTEM	N/A
Citrix Location and Sensor Virtual Channel Service	Enables a server side application to leverage Location and Sensor capabilities.	Automatic	Local Service	N/A
Citrix Mobile Receiver Virtual Channel Service	Enables a server side application to use mobile device capabilities.	Automatic	Local Service	N/A
Citrix MultiTouch Redirection Service	Provides MultiTouch redirection between the endpoint device and the virtual desktop.	Automatic	Local Service	N/A
Citrix NetScaler AppFlow Service	Provides AppFlow events to Netscaler.	Automatic	Local Service	N/A
Citrix Print Manager Service	This service supports the Citrix Advanced Universal Printing Architecture.	Automatic	Local Service	Print Spooler Remote Procedure Call (RPC)
Citrix Profile Management	Manages user personalization settings.	Automatic	Local System	N/A
Citrix Pvs for VMs agent	Pvs for VMs agent machine password update service.	Automatic	Local System	N/A
Citrix Services Manager	Citrix Services Manager.	Automatic	NT AUTHORITY \SYSTEM	Remote Desktop Service
Citrix Smart Card Service	Provides Smart Card redirection between the endpoint device and the virtual desktop.	Automatic	Local Service	N/A
Citrix Stack Control Service	RPC/Com Translation Service.	Automatic	Network Service	N/A

Service Name	Description	Startup Type	Log On As	Dependencies
Citrix Telemetry Service	Citrix Telemetry Service.	Automatic (Delayed Start)	Network Service	N/A
Citrix UWA Cache Service	The Citrix UWA Cache Service manages a private cache for Universal Windows Platform apps and their icons for Citrix use. The Broker Agent sends information in this cache to the Delivery Controller.	Automatic	Local System	N/A
Citrix WebAuthn Redirection Service	Provides Virtual Authentication FIDO2 Redirection Services.	Automatic	Local Service	N/A
Citrix WIA Service	Allows Redirection of WIA imaging devices over virtual channel.	Automatic	Local Service	N/A

Decision: Events Monitoring

Monitoring the Windows Event Log for unknown or critical events can help to discover issues and allow administrators to understand event patterns proactively:

- Licensing — Errors in the Event Log related to Remote Desktop licensing should be investigated. This might result from the installed Citrix product being unable to contact the Remote Desktop Licensing Server or the Citrix Licensing Server. If errors in the Event Log are not reviewed, users might eventually be denied access because they cannot acquire a valid license.
- Hardware Failure — Any event notification relating to a hardware failure should be looked at immediately. Any device that has failed will impact the system's performance. At a minimum, a hardware failure will remove the component's redundancy.
- Security Warnings — Customers should investigate security warnings or audit failure events regarding failed logins in the security log. This could indicate that someone is attempting to compromise the servers.
- Disk Capacity — When the drives of a Windows system reach 90% capacity, an event error message will be generated. To ensure continuous service, customers should poll these event errors. When the system runs out of hard disk space, it is put at severe risk. The server might not have enough space to service users' requests for temporary file storage.
- Application / Service errors - Any notification related to application or service errors should be investigated.
- Citrix errors - All Citrix software components will leverage the Windows Event Log for error logging. A list of the known Event Log warnings and errors issued by Citrix components can be found at the following links:

    - Citrix Event Logs
    - Citrix Provisioning
        - Citrix Workspace Environment Management

- Citrix App Layering

It is important to check the Event Viewer periodically for Citrix-related warnings or errors. Warnings or errors that repeatedly appear in the logs should be investigated immediately because it may indicate a problem that could severely impact the Citrix environment if not properly resolved.

Decision: Citrix Director Alerts

Baseline alerts within Director for Delivery Groups to alert Citrix administrators should be configured and adjusted as required.

These alerts can be configured in Citrix Director and should be monitored.

Alert Policy	Threshold and Alert Intervals	Description
CPU	All Delivery Groups: ● Warning: 80% (30 min) ● Critical: 90% (30 min)	Percentage of CPU usage. - Identify the processes or resources consuming CPU. - End the process if necessary. Ending the process causes unsaved data to be lost. - If all is working as expected, add additional CPU resources in the future.
Memory	All Delivery Groups: ● Warning: 80% (30 min) ● Critical: 90% (30 min)	Percentage of Memory usage. - Identify the processes or resources consuming memory. - End the process if necessary. Ending the process causes unsaved data to be lost. - If all is working as expected, add additional memory in the future.
Load Evaluator Index	All Delivery Groups: ● Warning: 80% (30 min) ● Critical: 90% (30 min)	Value of the Load Evaluator Index over the last 5 minutes. - Check Director for Server OS Machines that might have a peak load (Max load). - View both Dashboard (failures) and Trends Load Evaluator Index report.
Connection Failure Count	All Delivery Groups: ● Warning: 5% of total desktops ● Critical: 10%+ of total desktops	Number of connection failures over the last hour. - Check Director Connection Failures Trends view for events logged from the Configuration log. - Determine if applications or desktops are reachable.
Failed Machines (Server OS)	All Delivery Groups: ● Warning: 10 ● Critical: 20	Number of failed Server OS machines.

		Failures can occur for various reasons, as shown in the Director Dashboard and Filter views. Run Citrix Scout diagnostics to determine the root cause.
Average Logon Duration	• Warning: 60 seconds (30 min) • Critical: 120 seconds (30 min)	Average logon duration for logons that occurred over the last hour. - Check the Director Dashboard to get up-to-date metrics regarding the logon duration. Many users logging in during a short timeframe can increase the logon duration. - Check the baseline and breakdown of the logons to narrow down the cause.

Decision: Capacity Management

In addition to monitoring system-level metrics daily, performance metrics should be tracked from a historical perspective to help plan for future growth as more users access the environment.

A baseline of the environment's performance should be taken so that it can be compared against performance over time. For example, if a user complains of poor performance, this baseline can be used to identify whether the issues are related to the user load exceeding the environment's capacity.

Historical CPU, memory, and network utilization data on the controller and application servers or desktops would be an example of baseline performance metrics for capacity management.

Citrix Director

Administrators can utilize the Trends view within Citrix Director to track different parameters of the Citrix deployment over time. These parameters can be leveraged for capacity planning of the Citrix environment.

From the Trends view, administrators can see historical data that is broken up into several categories, including:

- **Sessions** - Provides concurrent session usage over time, enabling the ability to size the environment appropriately.
- **Connection Failures** - Gives an overview of the different types of connection failures that have occurred across different Delivery Groups.
- **Failed Desktop OS Machines** – Gives an overview of the different problems associated with failures in desktop machines.
- **Failed Server OS Machines** - Gives an overview of the different problems associated with failures in server machines.
- **Logon Performance** – This shows how long users can log on to their applications and desktops.
- **Load Evaluator Index** – Provides various performance counter-based metrics, including CPU, Memory, and Disk Usage for Server OS machines.
- **Capacity Management** – Shows utilization of published applications and desktops
- **Resource Utilization** – Provides information on CPU, Memory, and storage resource utilization
- **Custom Reports** – Allows administrators to create custom historical reports on numerous metrics captured by the system.

- **Hosted Application Usage** — This section Details all applications published on the site and can provide detailed usage information about each individual application (concurrent instances, launches, usage duration, and so on).

Citrix Analytics for Performance

The Citrix Analytics for Performance infrastructure dashboard gives administrators an overview of their environment's infrastructure health. The dashboard provides the VDA information across all sites. For multi-session OS VDAs, administrators can see which VDAs are unusable based on the load evaluator index. For single-session OS VDAs, administrators can see the number of VDAs in use and available.

From the Infrastructure Analytics page, administrators can gain insights into the key components of their Apps and Desktops sites to see:

- **Available Machines -** Shows the percentage of machines that were available in the last 15 minutes and the states they are in, either Ready or Active.
- **Unavailable Machines -** Shows the percentage of machines that were unavailable in the last 15 minutes, either in Unregistered, Failed, or Maintenance states. This information can be used to optimize machine utilization in the environment.
- **Machine Availability Trends -** Shows the Aggregate State of machines plotted across the selected period. The machine state is aggregated to consider the least favorable state from among Ready for use, Active, Maintenance, Unregistered, and Failed, in that order.
- **Machine Performance —** Available for Multi-session OS desktops only, this shows the number of machines in usable state categorized based on the load evaluator index, such as high, medium, and low, for the selected time period, Site, and Delivery Groups.

Acknowledgments

Creating and updating the handbook is time-consuming and requires real deployment experience across many scenarios. Citrix would like to thank all who have contributed, both current and past authors and subject matter experts who contributed to the Citrix VDI Handbook. This version of the handbook would not have been possible without the help from the following:

Name	Title
Steve Beals	Principal Product Marketing Manager
Emma Bland	Senior Product Marketing Manager
Uzair Ali	Professional Services Principal Architect
Rich Meesters	Professional Services Principal Architect
Michael McAlpine	Professional Services Lead Architect
Brendan Lin	Professional Services Senior Principal Architect
David Johnson	Professional Services Lead Architect
Michael Skowronski	Professional Services Lead Architect
Jason Delgado	Professional Services Lead Architect
Rob Zylowski	Professional Services Lead Architect
Sameer Sharma	Professional Services Lead Architect
Mike El-Safah	Professional Services Principal Architect
Jeff Qui	Principal Product Manager
Rainer Hasenzagl	Principal Product Marketing Manager
Eltjo van Gulik	Principal Product Manager
Adolfo Montoya	Lead Product Manager
Rody Kossen	Senior Principal Quality Engineer

Revision History

Revision	Description	Date
1.0	Citrix VDI Handbook 2402 LTSR	July 01, 2024