I appreciate those of you who provided questions for today’s webinar and since I received them before finalizing this presentation I was able to work the answers in for many of the posed questions. Some I will not be addressing directly because they did not fall into this topic, and the remaining ones I hope to cover during the Q&A at the end.
Why Is A Scalable Design Important?

- Return on investment (ROI)
  - A scalable design lowers the per desktop cost
  - An efficient design lowers management costs
  - A secure design limits liability

- User experience
  - Users ultimately dictate acceptability
  - Generate user demand for the technology

Since every implementation will have its own set of requirements I cannot tell you how to design your environment. What I can do is teach you the concepts that are required to design a scalable environment around your requirements. So today I plan to provide guidance around what things are important and what the best practices are recommended in a given situation.
What is your experience level with XenDesktop?

a. None (< 1 month)
b. Basic (1-6 months)
c. Moderate (6-12 months)
d. Experienced (12-24 months)
e. Expert (>24 months)
Agenda

- Quick architecture overview
- Key environment considerations
- Architecting the infrastructure
- Questions
Let’s start by looking at the end user view. A typical XD deployment caters very well for users access virtual desktops through the Intranet. In a desktop-replacement scenario, where the user’s desktop is entirely virtualized, Desktop Appliances are the most cost effective way of delivering access. DAs are available from a wide range of Citrix partners, and are access devices that require minimal management overhead and provide an excellent end user experience when accessing and using virtual desktops.

Of course our reference architecture works also well for remote and home users, who would typically connect through the Internet. The Citrix Access Gateway functionality is included in all XenDesktop editions and ensures that the communication with the virtual desktop is safe and secure – it works just like it always has when deployed in conjunction with XenApp.

In our reference architecture, we use PVS to stream the OS environment to the virtual machines, and VMs use shared disk images. In other words, you can set up and manage a single PVS disk image which is then used in parallel by any number of VMs. This reduces both storage cost and management overhead. The virtual disk image is best stored on a SAN, and each VM also has a write back cache which is implemented through a per-VM virtual disk that is also mapped through the hypervisor into the SAN. This set-up provides a good balance between network utilization and performance.

Since shared disk images are used, user personalization in our reference architecture is catered for through the use of roaming profiles, where user data and configuration is saved to a separate file server.

Finally, apps are delivered to the virtual desktop from XenApp.
So now you have a full-blown XenDesktop implementation under way, let’s look at what actually happens when end users connect.

First of all, the user authenticates, using the web service exposed by the DDC. XenDesktop supports a rich set of authentication methods, as you’re used to from XenApp.

The DDC determines the best virtual desktop for the user, given the current environment. Typically, this will be a VM that’s already up and running, in particular if the administrator has configure idle pool policies appropriately that enable the DDC to spin up VMs before demand occurs. However, in the case where the VM is currently suspended, the DDC communicates with the hypervisor to start it. This will cause the VM to PXE-boot from PVS, starting up the standard operating environment configured in the golden virtual disk you prepared earlier.

Once the VM is started, it will communicate with the DDC through a set of web service interfaces, and the DDC will relay the VM’s address to the desktop receiver on the endpoint device. This then results in an ICA connection to be made to the VM, which calls back into the DDC to acquire a license for the session. The DDC also determines policies that should apply to the session – XenDesktop supports the full set of ICA policies also available in XenApp, and also integrates with SmartAccess, in other words you can disable functionality based on information gathered on the end-user’s endpoint devices.

After a license has been acquired, the user is automatically logged into their virtual desktop, which also causes their profile to be applied using standard roaming profiles. Now the user can launch applications, whether they be installed locally, or delivered from XenApp (using the app receiver deployed into the standard operating environment).
Key Environment Considerations
Consider how the desktops will be accessed. Will all access come over the LAN with decent wirespeeds? Will the desktop via a small form-factor device such as an iPhone or iPad? Will that device be connecting over a 3G/4G connection? If the desktop access occurs over constrained pipes, consider implementing policies to control the ICA virtual channel behavior to disable unused channels. This saves not only the wire traffic, but it also reduces the CPU requirements (for packet processing) on both the server and the end-point device.

Windows XP
- Requires around 5000 IOPS each for startup
- Used for legacy applications
- Performed better on XenServer and ESX than Hyper-V

Windows 7
- Optimized for virtualized environment (Host Integration Services)
- More intelligent storage management
- Windows 7 performed better on Hyper-V
- Outperforms WinXP on same RAM (512/768)

Migration plans
- Starting with WinXP and moving to Win7 requires you plan for a Win7 environment (Profiles particularly)
Do you plan on migrating users from Windows XP to Windows 7 using VDI or XenDesktop?

a. Yes  
b. No  
c. Undecided, still researching  
d. Have not considered it
Key Environment Considerations

- Virtual machine availability
  - Idle desktop pool
  - Logon storms and storage capabilities
  - Hypervisor host goes down

- Security & monitoring software
  - Antivirus
  - Intrusion Detection/Prevention
  - EdgeSight
  - Systems Center Operations Manager

- Boot storm causes massive hit on storage
- Idle desktop pools are configured to maintain a specific number of idle workstations at a given time of day.
- Idle desktops are added automatically as new logons occur, which means a logon storm is accompanied by a boot storm.

Guidelines: Set workstation group peak early enough to settle down
  - Event log errors if not enough resources for new idle workstations
  - One option is to set idle count equal to workstation group size
  - Overcomes issues of new starting workstations
  - Lose power savings advantage

Security and monitoring software increase the required IOPS and reduce users per core. These items were designed for workstations that have plenty of free CPU cycles available. Now we are condensing 50-100 desktops to a single server and the availability of free cycles decreases dramatically.

Remember, the virtual desktops runs in a secure data center. The data center is normally the most secure zone and IDS/antivirus scanning is done at multiple levels both inside and outside the data center. In the case of a pooled desktop, the golden image is read-only and any virus infections are wiped out on a reboot. At the very least, disable drive scans for antivirus on pooled desktops. For persistent desktops, antivirus and IDS are probably still needed since a reboot will not reset the desktop to a golden image.

Reconsider what monitoring software you need for pooled desktops. If the monitoring software is there for troubleshooting machine issues is it necessary when a reboot fixes all the problems? In the case of asset management, that information is managed at the hypervisor host level. For patching/updates that can be handled at the vDisk/golden master creation level.

Scaling back the security and monitoring CPU cycles to just what is absolutely necessary for your environment will provide a more scalable solution with a faster response time.

Edgesight works better on XA than on XD and is more scalable. Hosting applications via terminal servers or XenApp where possible will allow troubleshooting where ES is not a scalability limiting factor. If ES on the desktop is necessary, it can be installed and the services disabled until needed.
Key Environment Considerations

• Application usage
  • Application types (Office applications, Video, VOIP)
  • Internet browsing level (heavy, medium, light)

• Application access
  • Hosted Applications (XenApp, Terminal Services)
  • Streamed Applications (App-V, Citrix App Streaming)
  • Local Applications

Application type

Applications that are CPU intensive will reduce the scalability. You cannot eliminate them, but you can be aware that your environment will probably have less scalability when these items are present. Microsoft Office is relatively light on storage. Internet browsing generates a significant amount of IOPS traffic as new pages are cached and written to disk.

Application Access

Hosted applications move processing and RAM requirements from the VM to the XA/TS server allowing for more desktops per host. Streaming applications will also be written to disk as the package is cached. Pre-cache application packages whenever possible to reduce the storage IOPS requirements. Streamed applications increase the IOPS required by the virtual desktop and decrease scalability of the storage tier.
Hardware Guidance

• Hypervisor desktop density
  • Users per core 4 → 10 (Power → Light)
  • Reserve one core for hypervisor

• RAM
  • Windows XP needs 512MB – 1GB
  • Windows 7 needs 1GB – 2GB
  • Overcommitting RAM resources will reduce performance

• Network
  • 1-2 Gbps per server is generally sufficient for desktop traffic

Calculate for active users per core
  • Light: 8-10 users/core
  • Normal: 6-8 users/core
  • Power: 4-6 users/core

Calculate users per server, reserve 1 core for hypervisor

• Influencers
  • XML response times
  • Session logon times
  • Hypervisor cluster / pool size
Storage Guidance

• Disk
  • User data requires SAN/NAS storage
  • Write-cache drives can be SAN/NAS or Local

• Measure user workload during a pilot
  • Analyze IOPS, size of I/O requests and data throughput (kbps)
  • Focus on peak average across guests not individual averages
  • Liquidware’s Stratusphere or Lakeside’s SysTrack can assist

• Consider new tools on the market that offload IOPS

Storage is the one thing you must get right!

Never make the mistake of purchasing storage based on the capacity you need instead of the performance you need. If you plan for performance you will always have enough capacity, but if you plan for capacity you may not have enough performance. Remember user experience is key. Most often the storage tier is cause of unexpected latency to the end user.

Consider that a spindle has a finite number of IOPS it is capable of generating. Controllers also have a finite number of IOPS they can handle. The transmission medium has a finite amount of bandwidth (fiber, iSCSI, ethernet) that is available. All these items combined generate a particular performance level maximum, a maximum you should be aware of when designing a system.

IOPS is a good start, here are some ball park figures:
  300 IOPS for booting
  100 IOPS for logging on
  25 IOPS for medium workload

Tools to offload storage like Atlantis that uses a virtual hard-drive system to offload the IOPS and reduce disk requirements.

Liquidware Fit has Peak Avg IOPS report.
Storage Guidance

- SAN best practices
  - Use fixed VHDs
  - Verify disk alignment
- Plan for the I/O profile
  - Pooled Desktops 10% Read / 90% Write
  - Assigned Desktops 40% Read / 60% Write
- Storage RAID level makes a difference
  - RAID5 4x write-penalty (vDisk storage)
  - RAID1/10 2x write-penalty (write-cache drives)
  - Controller cache performance improvement varies by vendor

Use fixed VHDs
  - Never use dynamic VHDs
  - Manually set/verify partition offset (disk alignment is key)
As VHD expands
  - Disk can become fragmented on physical media
  - Expansion algorithm occurs in 1 MB increments
  - Rapid expansion wreaks havoc on SAN such as first boot or page file creation
Alignment issues
  - Constructed with extra byte at end of file
  - Dynamic VHD always mis-aligns disk with storage
Use only fixed-size VHDs for write-cache drives and Provisioning services vDisk

Write-intensive operations
  - Pooled Desktops 10% Read / 90% Write
  - Standard Desktops 40% Read / 60% Write
Storage RAID level makes a difference
  - RAID5 4x write-penalty (vDisk storage)
  - RAID1/10 2x write-penalty (write-cache drives)
  - SAN Controller cache improves write through
Hardware Specifications

- **Desktop delivery controller**
  - Farm Master: 2 vCPU, 8GB RAM, 2Gbps
  - XML Controller: 2 vCPU, 4GB RAM, 2Gbps

- **Provisioning server**
  - Virtualized Server: 4 vCPU, 8GB RAM, 3Gbps (<1000)
  - Physical Server: 8 cores, 8GB RAM, 6Gbps (~3000)

- **Hypervisor hosts**
  - Dual quad-core Nehalem or Westmere processors are best performance
  - 4GB chips are best price point right now, though 8GB is fast approaching

- Farm Master recommended specs: 5,000 desktops
  - 2 cores (1.8GHz+)
  - 8 GB RAM
  - 2 Gbps NICs (Redundancy)

- XML Controllers recommended specs: 5,000+ desktops
  - 2 cores (1.8GHz+)
  - 4 GB RAM
  - 2 Gbps NICs (Redundancy)

**PVS Server Recommended Specs:** 1200-1500 streams
- 8 cores
- 2 GB + 2 GB RAM per vDisk (estimate)
- 3 Gbps NICs teamed

**Virtual Center recommended specs:**
- ESX 3.5: 1,700 virtual desktops
- vSphere: 2,500 virtual desktops
- 4 cores
- 4 GB RAM
- 2 Gbps NICs (redundancy)

**Host servers recommended specs**
- 8 cores
- RAM equal to workstation plus hypervisor needs - do not overcommit RAM
- 6 Gbps NICs - redundancy (2 user/infrastructure, 2 storage, 2 management)
Infrastructure Architecture
Which type of desktop pools are you most likely to use in your XenDesktop design?

a. Pre-Assigned (Persistent)
b. Assigned on first use (Persistent)
c. Pooled (Non-persistent)
d. All the above
General Guidelines

• Virtualizing components
  • Web Interface, License Server, SCVMM, vCenter - Yes
  • DDC, PVS, SQL Server – Depends

• Clustering
  • For infrastructure components
  • When using assigned desktops
  • Not recommended for pooled desktops
XenDesktop Farm Design Guidelines

- Desktop delivery controllers
  - Separate Roles (Master and XML/VDA)
  - Farm Master ~5,000 desktops (8GB / 2 cores)
  - XML/VDA Only ~5,000 desktops (4GB / 2 cores)
- Desktop group design
  - Assigned/Pre-Assigned/Pooled options
  - Keeping the number of desktop groups low reduces pool management overhead
  - Hypervisor selection limits maximum desktop group size

- Influencers
  - XML response times
  - Session logon times
  - Hypervisor cluster / pool size

- Farm Master recommended specs: 5,000 desktops
  - 2 cores (1.8GHz+)
  - 8 GB RAM
  - 2 Gbps NICs (Redundancy)

- XML Controllers recommended specs: 5,000+ desktops
  - 2 cores (1.8GHz+)
  - 4 GB RAM
  - 2 Gbps NICs (Redundancy)

Desktop group design:

Pool Management service handles each desktop group separately, so the more groups, the more connections into the hypervisor management system (Xen, SCVMM, vCenter). Also, since the groups are treated equally the more groups the more simultaneous startups you may have so if you have multiple desktop groups pointing to the same management system it may get overloaded.
Provisioning Services Guidelines

- Server RAM sizing (CTX125126): 2GB + (2GB * #vDisk)
- 1Gbps NIC = ~500 concurrent streams
- Failover ~200 streams/minute
- Key optimizations
  - Disable TCP large send offload on both server and target
  - Disable auto-negotiation on network ports where feasible
  - Windows File-caching ramifications

PVS Design Goal: Use system cache and not read from disk

\[ RAM = OS + OS_{\text{Cache}\_Req} + (\#_{\text{Active\_vDisks}} \times \text{Avg\_Data\_Read\_per\_vDisk}) \]

Recommended Specs: 1200-1500 streams
  - 8 cores
  - 2 GB + 2 GB RAM per vDisk (estimate)
  - 3 Gbps NICs teamed
Every 1Gbps NIC = 500 streams
More streams = longer failover process
  - 1,000 streams ≈ 5 minutes
  - 1,500 streams ≈ 8 minutes

Fixed vDisks: Fragmentation risk of dynamic vDisk during updates when size is increased
Disable TCP Large Send Offload
  - Adds latency as packets re-segmented
  - Must be set on server and target
No auto-negotiation
  - Causes latency in handshake
  - Can cause PXE timeout
  - Hard-set server/target NIC and switch ports appropriately

File Caching
  - If PVS server sees drive as local drive, OS will cache
  - If PVS server sees drive as network drive, OS will \textbf{not} cache
  - If connection between PVS and storage is CIFS, OS will \textbf{not} cache
Provisioning Services Guidelines

Windows XP Optimization Guide (CTX124239)
Windows 7 Optimization Guide Coming

http://support.citrix.com/article/CTX124239 (WinXP)
http://community.citrix.com/display/ocb/2010/01/15/Optimizing+Windows+7+for+FlexCast+Delivery (Win7)
http://community.citrix.com/display/ocb/2008/10/24/Windows+XP+Performance+Optimizations+for+XenDesktop+and+Provisioning+Server+vDisks (WinXP)

VDI Optimizer
Provisioning Services Write-Cache Guidelines

• Sizing the write-cache drive
  • Start with write-cache on PVS Server
  • Start with initial size of 2 GB + pagefile size

• Implementation recommendations
  • Client-side placement is recommended for maximum PVS server scalability
  • Pre-cache packages for streamed applications
  • Shared storage versus local disk placement
  • SSD considerations

- Start with write-cache on PVS Server to get some info on write-cache size. Write-caches are stored on the PVS server in a sub-directory under the vDisk store by machine MAC address. You can run a pilot for a week or so then view the maximum write-cache size (keep in mind if you are rebooting every night you will need to check for the maximum each day since it is recreated at boot time) during the pilot then add say 20-30% for a safety margin.
- Start with initial size of 2 GB + pagefile size

- Pagefile is written to target device partition where write-cache is located if “target device write-cache” is chosen, available space must be sufficient or OS may blue screen
- Pre-cache any streamed application packages (App-V or Citrix) so the packages do not need to be written down to the write-cache drive.
- Minimize network impact (limit number of hops)

- Shared storage is the more costly solution in many situations. If you can get away with local-storage for the write-cache drive, that is usually the better approach. Write-cache drives are volatile and get wiped on every reboot. They contain no real user data and just need to be available during the time when the VM is on. In a pooled desktop environment where users can just connect back into another desktop this architecture makes sense. Keep in mind though that using local drives prevents any type of VM migration while it is running. The local drive option also limits the IOPS so if you don’t have high IOPS requirements or can stack a server full of drives (like a DL360 with 8 drive bays) then this is feasible without using SSDs.

- SSDs have the high IOPS and can even remove the need for an idle desktop pool. They are a great option if you are willing to trust in their endurance which seems at this point to be the biggest drawback besides cost. However, preliminary estimates show their MTBF to be near that of a processor and when compared with the fully-loaded cost of a SAN (hardware, network, maintenance, mgmt, staff requirements) they are probably comparable.
Provisioning Services uses an optimized UDP-based protocol to communicate with the target servers.

Data is streamed to each target server only as requested by the OS and applications running on the target server. In most cases, less than 20% of any application is ever transferred. Network utilization is most significant when target servers are booting as the OS loads. After target servers start, there is minimal network utilization.

### Provisioning Services Network Usage

<table>
<thead>
<tr>
<th>Operating System</th>
<th>Client-Side Cache</th>
<th>Server-Side Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>XP SP2 32-bit</td>
<td>85 MB</td>
<td>90 MB</td>
</tr>
<tr>
<td>XP SP2 64-bit</td>
<td>110 MB</td>
<td>115 MB</td>
</tr>
<tr>
<td>Server 2003 SP1 32-bit</td>
<td>95 MB</td>
<td>100 MB</td>
</tr>
<tr>
<td>Server 2003 SP1 64-bit</td>
<td>120 MB</td>
<td>130 MB</td>
</tr>
<tr>
<td>Windows 7 32-bit</td>
<td>180 MB</td>
<td>190 MB</td>
</tr>
<tr>
<td>Windows 7 64-bit</td>
<td>220 MB</td>
<td>240 MB</td>
</tr>
</tbody>
</table>
ICA uses all available bandwidth
Compression algorithms applied as bandwidth is restricted

Performance Assessment and Bandwidth Analysis for Delivery XenDesktop to Branch Offices:
http://support.citrix.com/article/CTX124457

Optimizing HDX for High-Latency Connections:
http://support.citrix.com/article/CTX125027

**HDX/ICA Network *Estimates***

- ICA uses all available bandwidth
- XenDesktop Bandwidth Analysis: [CTX124457](http://support.citrix.com/article/CTX124457)
- Optimizing HDX for High-Latency: [CTX125027](http://support.citrix.com/article/CTX125027)

<table>
<thead>
<tr>
<th>Medium Traffic Workload</th>
<th>XenDesktop Native</th>
<th>XenDesktop w/ Branch Repeater</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft Office</td>
<td>43 kbps</td>
<td>31 kbps</td>
</tr>
<tr>
<td>Internet Browsing</td>
<td>85 kbps</td>
<td>38 kbps</td>
</tr>
<tr>
<td>Printing</td>
<td>553-593 kbps</td>
<td>155-180 kbps</td>
</tr>
<tr>
<td>Flash Video</td>
<td>174 kbps</td>
<td>128 kbps</td>
</tr>
<tr>
<td>Standard WMV video</td>
<td>464 kbps</td>
<td>148 kbps</td>
</tr>
<tr>
<td>High-definition WMV video</td>
<td>1812 kbps</td>
<td>206 kbps</td>
</tr>
</tbody>
</table>
Designing for High Availability

- Desktop Delivery Controller – Multiple controllers, heartbeat & failover via IMA, Web Interface
- Web Interface – Global Load Balancing
- Provisioning Services – Multiple servers, bootstrap list
- Desktops – Multiple desktops in pool, multiple pools
- Database – SQL Clustering
- File System – DFS
- For more information see [CTX123244](http://support.citrix.com/article/CTX123244)
Which hypervisor are you most likely to use for hosting the virtual desktops?

a. Citrix XenServer  
b. Microsoft Hyper-V  
c. VMware vSphere  
d. VMware ESX  
e. No preference
XenServer Design Challenges

- Multiple resource pools in a desktop group (CTX120077)
  - Steps to configure are not intuitive
  - Username/Password must be same across all pools
- Link bonding provides fault-tolerance not aggregation
- DOM0 memory modifications recommended (CTX124086)
  - Increase RAM allocated to DOM0
  - Increase heap size
- Recommended pool size: 1600 VMs (16 servers @ 100)

Aggregating multiple pools (if more than 16 XS hosts needed)
- XenDesktop Setup Wizard can only communicate to a single pool
- VM templates need to be copied to each pool
  - http://support.citrix.com/article/CTX120077

Add backup pool master for high availability
- XenServer Storage - IOPS, write-cache disks, HBAs, StorageLink, local RAID
- XenServer Networking - VLANs, NIC bonding, segregated networks (NFS/iSCSI)

Pool Size
- Recommended: 1,600 VMs (16 servers per pool * 100 VMs per server)
- Actual: Higher (20-40 servers per pool * 80-100 VMs per server)

DOM0
- Increase Dom0 RAM allocation to 2.94 GB (/boot/extlinux.conf)
- Increase Xen-heap setting to xenheap_megabytes=24 (/boot/extlinux.conf)
ESX/vSphere Design Challenges

• Pool Manager needs to throttle vCenter requests ([CTX123684])
• vCenter hosts clustered for fault-tolerance
• Overcommitting RAM reduces scalability
• Single desktop group cannot span vCenter hosts
• Recommended pool size:
  2500 VMs (vSphere 4)
  1700 VMs (ESX 3.x)

VMware DRS used a lower value should be set as DRS needs additional time to determine guest placement before powering it on. In our testing with DRS enabled the rate of 20 was used. Design impacted by logon/boot storms in environment.

In our testing we allowed DRS to do the initial VM placement through a full run, DRS was then disabled and this allowed the MaximumTransitionRate to be increased to 40 without VC becoming overloaded.

Edit Pool Management Service configuration file:
  C:\Program Files\Citrix\VMManagement\CdsPoolMgr.exe.config
  Modify the <appSetting> section
  <add key="MaximumTransitionRate" value="20"/>

Value:
  VMware DRS used: 20
  VMware DRS not used: 40
  Starting too many VMs at once through the API has been known to cause vCenter to crash. Throttle the outstanding machine count.

VPXD service
  Used by XenDesktop Pool Management to power on/shut down VMs
  Boot Storm: 2 cores fully consumed for 1,700 in 30-45 minutes

  Overcommit (ballooning) creates overhead that harms performance.

vCenter limited to 3,000 VMs
  VMware reference architecture: 1,000 VM blocks
  User habits dictate if 2000 is feasible

vCenter is critical to environment
  Can be a single point of failure, requires a clustered solution
  XD master uses API to do operational processes on VMs
  Numerous calls have caused API to crash, bringing down entire VC

Desktop Group
  A XD desktop group can only point to single vCenter
  Complicates design as more desktop groups required
Hyper-V Design Challenges

- Legacy network adapter required for PXE booting
- Supports only block-level storage (iSCSI/SAN)
- Desktop groups cannot span SCVMM servers
- CPU overcommit supported limit of 8:1
- Recommended pool size: 1000 VMs per SCVMM
- For more information see **CTX124687**

SCVMM templates
- Added a local write-cache disk template
- Disk dropped during create wizard
  - PowerShell scripts needed to copy/add write-cache VHD to all virtual machines after created
SCVMM server is able to run jobs asynchronously
- Running multiple instances of the XD Setup Wizard speeds up VM creation
- Setup Wizard divides the number of virtual machines equally across all the hosts on the SCVMM server

Hyper-V Design Guide:
- [http://support.citrix.com/article/ctx124687](http://support.citrix.com/article/ctx124687)
Questions