Delivering 3D Graphics from the Cloud with XenApp and XenDesktop

Derek Thorslund, Director of Product Management
Citrix Systems

March 2013
Rich Apps as a Service
HDX High Definition Experience

**HDX Broadcast**
ICA and RDP protocol support for access from any device, anywhere

**HDX MediaStream**
Video and audio playback, any format, on any device

**HDX RealTime**
Voice and video for real-time collaboration / unified communications

**HDX SmartAccess**
Simplified secure access

**HDX Plug-n-Play**
Access to local resources and peripherals such as printers, monitors and USB devices

**HDX RichGraphics**
2D and 3D graphics incl. Adaptive Display, HDX 3D, HDX 3D Pro and RemoteFX

**HDX WAN Optimization**
Performance and bandwidth optimizations via Citrix Branch Repeater

**HDX Adaptive Orchestration**
Best user experience based on server, network connection and user device
Milestones in 3D graphics remoting

2006
Project K2 delivers CATIA to Boeing Dreamliner designers

2009
GA of XenDesktop HDX 3D Pro with Deep Compression

2010
XenApp 6.0 high performance GPU Sharing for DirectX

2011
XenServer 6.0 hypervisor introduces GPU Passthrough

2012
Higher fps via NVIDIA VGX™ API plus improved compression
Business Drivers
for virtualizing 3D graphics apps & workstations
Global talent base
Secure IP
Work-from-home
Disaster recovery
Mobile device access
Improve time-to-market
Cost efficiency
Leverage worldwide talent pool
Centralize and secure design IP

- Engineering drawings
- Bills of Materials
- Cost info
- Supplier info
- Customer info
- Lifecycle data
- Product design decisions
Work-from-home & Disaster recovery
Leverage mobile devices
Improve efficiency & agility

Reduce operational costs
Global Product Development Teams - Real Example

- United States
- Brazil
- Germany
- India
- China
- Korea
- Australia
Global Development Effort – Real Example

- 30,000 CAD files or 70 GB of data to be synchronized every day
- Across 26 design centers (30,000+ users)
- Across 16 countries
- It took 2 weekends to sync all code updates!
- More challenging for 4,000+ suppliers and partners
Enhances IP control, collaboration and global agility

Data stays in data center
Access via LAN or WAN
Case studies
Case study and customer reference
Global CAD access with HDX 3D Pro
October 2011
Requirements from Business to IS

Extended Engineering Workbench in India

Engineers in India need to be able to work as if they were sitting in Switzerland
Challenges
Of course there were more than one…

3D CAD data is large
- Transferring our largest assemblies took 2.5 hours!

ABB’s corporate network
- Latency
- Bandwidth
Implementation at ABB
Citrix XenDesktop HDX 3D Pro
Learnings

**Service quality is a subjective matter**

Latency effect
(subjective scores)

![Graph showing latency effect and system quality percentage.](image)

- **China, Beijing**: 50%
- **India, Bangalore**: 75%
- **Canada, Montreal**: 80%
- **Poland, Cracow**: 90%
- **Switzerland, Turgi**: 100%

Latency [ms] - System Quality [%]

(These scores not incl. Repeater benefits)

Using Dassault SolidWorks, 5-6 hours per day; Designers can work from India as if in Switzerland!
Power and productivity for a better world™
HDX 3D Pro case study

Wind turbine manufacturer

- PTC Pro/E and Dassault SolidWorks from Europe to other continents
- HDX 3D Pro protects Vestas’ intellectual property, supports workforce globalization, eliminates inconsistencies in engineering design versioning and overcomes regulatory challenges
- Reduced cost per running hour by 30% from €416 (traditional CAD workstations) to €291 (data center blade workstations) via follow-the-sun utilization (Denmark, UK, US, India, China)
HDX 3D Pro case study
Daimler Digital Factory

• The Daimler Digitale Fabrik (Digital Factory) team can simulate an entire manufacturing plant in software

• HDX 3D Pro serves users across various Daimler plants and offices

• Siemens NX applications

• HP ws460c data center blade workstations

• NVIDIA Quadro Fermi Q2000 cards
HDX 3D Pro case study
Major European heavy vehicle manufacturer

• Access from Germany, Mexico and Brazil to Dassault CATIA apps hosted in Sweden

• At 220ms roundtrip latency, good performance working on models with 1500+ parts; bandwidth usage rarely reaches 2.5 Mbps

• Using 3D Space Mouse
HDX 3D Pro case study
John Deere

• In production since 2011
• IP security was a key requirement
• PTC Pro/E is primary application
• Follow-the-sun utilization
• NetScaler used to connect user to nearest data center
• HP WS460c workstation blades with NVIDIA Q2000M graphics cards
HDX 3D Pro case study
Aerospace and rail transportation manufacturer

- Dassault CATIA R18 access from India, Ireland, USA and Mexico to data center in Canada
- Dell workstations with NVIDIA Quadro GPUs
HDX 3D Pro case study
Large engineering, design and consultancy company

• Bentley, Navisworks, Revit, AutoCAD, and more
• “It’s faster than local!” (large 3D models no longer have to be transferred across the network)
• “It even works on 3G!”
• “We reduced the number of PCs per user from 1.6 to 1.05”
Product Overview
XenDesktop HDX 3D Pro and XenApp HDX 3D
HDX 3D Pro and 

- Heavy 3D graphics app usage
- Exceptional graphics performance and response optimizations
- Multiple users per workstation
- Best WAN performance
- OpenGL and DirectX acceleration
• Many users per server
• Works well over WAN at 3 Mbps
• Graphics performance and response optimizations
• DirectX and OpenGL acceleration
3D graphics acceleration options from Citrix

- **XenDesktop HDX 3D Pro**
  - High-end 3D professional graphics (OpenGL and DirectX)
- **XenDesktop w/ RemoteFX (DirectX only)**
- **XenApp HDX 3D**

**Bandwidth efficiency**
- 1 GPU/user:
  - 3-10 Mbps
  - 3D business graphics
- Shared GPUs:
  - ~3 Mbps
- Bandwidth at 1024 x 768:
  - ~1.5 Mbps

© 2013 Citrix
Segmenting the user population

Tier 1 (e.g. design engineers)
• Top rendering performance (dedicated GPU)
• Deep compression on WAN links
• 3D SpaceMouse

Tier 2 (viewing/editing of large 3D drawings)
• GPU sharing

Tier 3 (typical knowledge workers)
• Software rasterizer or highly shared GPU
Tier 1 and Tier 2 solutions for GPU-accelerated graphics

**XenDesktop HDX 3D Pro**
- GPU acceleration of DirectX and OpenGL
- One user per GPU (can also be used with GPU virtualization)
- H.264-based Deep Compression
- 3D SpaceMouse support

**TOP PERFORMANCE SOLUTION**

**XenApp HDX 3D**
- GPU acceleration of DirectX and OpenGL
- High performance GPU sharing
- Lower cost Microsoft licensing
- Apps must be compatible with RDSH (Terminal Services)

**MOST COST-EFFECTIVE SOLUTION**
XenDesktop HDX 3D Pro

- XenDesktop feature for high-end 3D professional graphics
- **GPU acceleration** for hardware rendering of large 3D models
- Multiple compression options including **deep compression** codec for access over narrow WAN links
XenDesktop: Powerful and flexible infrastructure

Universal client

High-Definition User Experience

Enterprise app store

Flexible Desktop and App delivery

PC
Mac
tablet
smartphone
thin client

Citrix Receiver

HDX

Citrix CloudGateway

© 2013 Citrix
HDX 3D Pro Architecture

- Receiver
- XenDesktop Controller
- Access Gateway
- Host (e.g. blade or rack) w/ multiple GPUs
- App Streaming
- User Profiles
- User Settings
- Desktop Provisioning
- OS
- Apps
HDX 3D Pro

- Feature of XenDesktop Enterprise and Platinum editions
- Broad app compatibility
  - OpenGL, DirectX (incl. WPF etc.)
- Blade/rack workstations are ideal, but any form factor can be used for the host
  - Multiple users per workstation using GPU Passthrough
- User device can be PC, Mac, tablet or thin client

© 2013 Citrix
HDX 3D Pro is not application-dependent 😊

Examples of applications tested with HDX 3D Pro:

<table>
<thead>
<tr>
<th>CAD</th>
<th>GIS</th>
<th>Multimedia</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATIA V5R19</td>
<td>Google Earth</td>
<td>HD Videos in YouTube</td>
</tr>
<tr>
<td>Autodesk Inventor 2010</td>
<td>ISRO – Bhuvan</td>
<td>Windows Media Player</td>
</tr>
<tr>
<td>Autodesk Revit 2010</td>
<td>ArcGIS Explorer</td>
<td>VLC Media Player etc.</td>
</tr>
<tr>
<td>Autodesk MAYA 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AutoCAD 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bentley</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D XML Player</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JT2Go</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SolidWorks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FurnPlan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adobe 3D</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Google Earth</td>
<td></td>
</tr>
</tbody>
</table>

**Medical**
- Fiat Lux

**Test Applications**
- NEHE Apps for OpenGL
- DirectX SDK
- CUDA SDK
- GLView

**Office Apps and Win 7 Gadgets**

**Multimedia**
- HD Videos in YouTube
- Windows Media Player
- VLC Media Player etc.

**Benchmark and Demo**
- Realtime HDR
- Turbine Demo
- SpecViewPerf
- CineBench

**WPF Applications**
- Mix Me
- Photo Shuru
- Visual 3D
HDX 3D Pro
Additional Features

• Host OS support: WinXP & Win7, 32-bit & 64-bit
  • Note: XenServer, however, does not support Windows XP 64-bit

• English-on-German/French/Spanish/Japanese/etc.

• Pixel-perfect lossless compression (e.g. for medical imaging)
  • Can combine with lossy compression during motion, for responsive remote access

• Seamless app delivery (see next slide)
Seamless Application Delivery

- Apps are merged seamlessly into the user’s local desktop
- Leverages the same architecture as desktop delivery
Deep Compression codec technology
Customer-reported bandwidth utilization on long-haul connections

• First user requires 1.5 to 2 Mbps minimum
• Navistar:
  Branch with 12 concurrent users requires 700-800 Kbps per user
• CCI Valves:
  20 Mbps WAN link serves branch with 17 users (1.2 Mbps/user)
• Bandwidth requirement does not scale linearly 😊
User control… and Admin control

Adjust image quality settings with HDX image quality configuration tool
Lossless Compression (pixel-perfect)
GPU Options

• NVIDIA cards deliver highest frame rate with VGX™ API ("Monterey") and support full-screen applications

• CPU-based deep compression is default, but if server CPU is limited then deep compression can be offloaded to the NVIDIA GPU
HDX 3D Pro
Desktop Virtualization for High-end Graphics Users

Client options

Optimized with special codecs:

Std. ICA codec:
HDX 3D Pro on a tablet

“Performance is great! Nearly like my Windows 7 PC!”
HDX 3D Pro on thin clients

- HDX Ready Premium thin clients supporting Deep Compression decoding
- More to come, including lower cost HDX SoC devices

Photos not to scale
Host requirements

- Windows 7 (32/64-bit), Windows XP (32/64-bit)
- XenServer 6.x or vSphere 5.1 or physical machine
- Quad core CPU at 2.3 GHz or higher, or four vCPUs
- 4 GB of RAM minimum
- XenDesktop 5.6 Feature Pack 1 (select “HDX 3D Pro” during installation)
- GPU card supported by ISV (recommend NVIDIA Quadro or GRID)
Dual monitor support

• Citrix Receiver for Windows or Linux
• Efficient use of bandwidth
3D mouse support

USB redirection for 3D Space Mouse and similar devices

Virtual Channel can be prioritized to maximize responsiveness
Citrix CloudBridge™ Branch Repeater
Ideal for low bandwidth and high latency connections

• Improves responsiveness of apps delivered via HDX 3D Pro over high latency connections

• Further reduces bandwidth consumption due to local caching, enabling more users to share a given size of pipe (e.g. ABB reports 3:1 compression at just 5 users)
GPU Passthrough introduced in XenServer 6

Reduced cost per user

- XenDesktop HDX 3D Pro and XenApp HDX 3D
- Windows 7 (32/64-bit), Windows XP (32-bit), Windows Server 2008 R2 (64-bit)

- Multiple GPUs per host
  - Servers with 4 GPUs currently on HCL; testing underway with 8
  - One graphics-accelerated VM (single-user or multi-user) per GPU
  - Depending on CPU power, same host may also support non-graphics-accelerated users (i.e. regular office workers)

- Leverages NVIDIA high-performance Fermi/Kepler GPUs
  - e.g. Quadro 1000M/2000/3000M/4000/5000/6000, Tesla M2070Q, GRID K1/K2
XenDesktop HDX 3D Pro

Option 1:
1 Server : 1 GPU : 1 User

Option 2 (GPU Passthrough):
1 Server : 4 GPU: 4 Users
1 Server : 2 GPU : 2 Users
GPU Passthrough

XenDesktop
Windows VMs

<table>
<thead>
<tr>
<th></th>
<th>non-3D VM</th>
<th>non-3D VM</th>
<th>non-3D VM</th>
<th>non-3D VM</th>
<th>non-3D VM</th>
<th>non-3D VM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D Pro VM</td>
<td>3D Pro VM</td>
<td>3D Pro VM</td>
<td>3D Pro VM</td>
<td>3D Pro VM</td>
<td>non-3D VM</td>
<td>non-3D VM</td>
</tr>
</tbody>
</table>

Hypervisor

Hardware platform

GPU  GPU  GPU  GPU
XenApp HDX 3D: GPU sharing on Windows Server
Up to 4x more users

Option 1:
• 1 Server : 1 GPU : ~10 Users

Option 2 (GPU Passthrough):
• 1 Server : 4 GPU* : ~40 Users
• 1 Server : 2 GPU* : ~20 Users

Note - * Server & GPU dependent
XenServer GPU Passthrough with XenApp

- XenApp Windows Server VMs
  - XenApp VM
  - XenApp VM
  - XenApp VM
  - XenApp VM
  - XenApp VM
- Hypervisor
- Hardware platform
  - GPU
  - GPU
  - GPU
  - GPU

© 2013 Citrix
GPU sharing scalability

Reports from the field

• With two NVIDIA Quadro 4000 cards we ran 18 users using a test app that works with ESRI ArcGIS, and we still had space for more

• Running Dassault SolidWorks, Ansys Workbench and Fluent, scalability was 6 to 10 users per Quadro 4000

• The Quadro 6000 was able to support 30 users running Dassault 3DVIA Composer Player with only minor slowdown; and this test was harder on the graphics card than the real world is!

New NVIDIA GRID K2 promises even higher user densities!
# Multi-GPU Hardware Platforms

<table>
<thead>
<tr>
<th>Hardware Platform</th>
<th>Certification Status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HP</strong> ws460c G6 blade (Q4000/5000/6000)</td>
<td>XenServer certified</td>
</tr>
<tr>
<td>Servers: DL980 G7, DL370 G6, SL390S G7 and others</td>
<td></td>
</tr>
<tr>
<td><strong>Dell</strong> R720 rack workstation (two dual-slot cards, e.g. M2070Q or GRID)</td>
<td>tested</td>
</tr>
<tr>
<td>R5500 rack workstation (3x Q4000 or 4x Q2000)</td>
<td>XenServer certified</td>
</tr>
<tr>
<td>R5400 rack workstation (2 PCIe x16 GPUs, e.g. Q4000)</td>
<td></td>
</tr>
<tr>
<td>PowerEdge T620 (4x Q2000)</td>
<td></td>
</tr>
<tr>
<td>PowerEdge M610x with M2070Q plus C410x PCIe expansion chassis (4 GPU devices per server, 16 total GPUs per chassis)</td>
<td>tested w/ XenApp</td>
</tr>
<tr>
<td><strong>IBM</strong> HS22 with GPU expansion blade</td>
<td>tested with 2x M2070Q</td>
</tr>
<tr>
<td>System x iDataPlex dx360 M3/M4 (supports Q4000/Q5000/M2070Q)</td>
<td>tested</td>
</tr>
<tr>
<td>System x iDataPlex x3650 M3/M4 (2 dual-slot cards, e.g. GRID)</td>
<td>certified</td>
</tr>
<tr>
<td><strong>Others</strong> Cisco C240 M3; Fujitsu TX300; SGI Asterism ID112</td>
<td>XenServer certified</td>
</tr>
</tbody>
</table>
Designer Engineer Virtual Workstation solution

Citrix ready design engineer ecosystem around HDX 3D Pro
GPU Sharing on XenApp
Multiple concurrent users per GPU

• XenApp HDX 3D GPU Sharing has been available since XenApp 6.0
• New: XenApp 6.5 OpenGL GPU Sharing – NOW AVAILABLE on citrix.com
• Lots of great comments from our Tech Preview participants: http://blogs.citrix.com/2013/02/22/how-many-users-can-share-a-gpu/
• Works with Fermi-generation NVIDIA Quadro cards and with the latest Kepler-architecture GRID K2 (higher user density)
• Directly leverages the GPU video driver (unlike API Intercept vGPU)
• Supports all versions of DirectX and OpenGL
• Ideal for second tier users of 3D professional graphics

© 2013 Citrix
XenDesktop/XenApp on vSphere 5.1

GPU passthrough and GPU sharing on vSphere 5.1

• XenDesktop/XenApp is hypervisor agnostic
• vSphere 5.1 has introduced “virtual Direct Graphics Acceleration” (vDGA) and “virtual Shared Graphics Acceleration” (vSGA)
• Testing for XenDesktop HDX 3D Pro compatibility with vDGA has been successfully completed ([http://support.citrix.com/article/CTX136291](http://support.citrix.com/article/CTX136291))
• Testing with vSGA successful (single monitor) for “tier 3” use cases
• vSGA uses API Intercept method of sharing the GPU so performance may be an issue with large 3D models, and graphics API compatibility is limited to DirectX 9c and OpenGL 2.1
Looking ahead…
3D graphics enhancements in Excalibur (first release from project Avalon)

XenDesktop HDX 3D Pro
• Multi-monitor support (i.e. more than two)
• Dynamic resizing without disconnect
• Auto-config / auto-adapt (no config tool / user intervention required)

XenApp HDX 3D becomes XenApp “HDX 3D Pro”
• Deep Compression

Citrix Receiver (enhancements targeted for Excalibur timeframe)
• Deep Compression support on Mac, iPad, Android
• HDX SoC support for Deep Compression decoding on low-cost Linux thin clients
Looking ahead…
Hardware vGPU

At the 2012 GPU Technology Conference, NVIDIA and Citrix announced our collaboration on **GPU Hardware Virtualization**

- Enhancements to the VGX™ API and XenServer along with new GRID graphics cards will extend Citrix’s high performance GPU sharing capability from Windows Server RDSH to Windows XP/7/8 VDI
- Higher performance with large 3D models compared to API Intercept method used by VMware vSphere/ESX and Microsoft Hyper-V
- Ideal for second tier users of 3D professional graphics
- Beta trials with selected partners starting Q2 2013 (see demo in Exhibit Hall)
GPU Virtualization

XenDesktop
Windows VMs

1
3D Pro VM

2
3D Pro VM

3
3D Pro VM

4
3D Pro VM

... N
3D Pro VM

XenServer hypervisor

... vGPU

Hardware platform

... GPU
Summary: Citrix solution for 3D graphics

- Best WAN performance on the market
- First to market with NVIDIA VGX API support
- Lowest cost per user
- Any device
And don’t miss…

Citrix booth in the Exhibit Hall
• XenApp high-performance GPU sharing (OpenGL and DirectX)
• GPU sharing on XenDesktop
  (pre-beta of GRID VGX Software for GPU hardware virtualization)

Thomas Poppelgaard’s seminar S3540
• Successfully Delivering 3D Graphics Solutions for Your Business
• Best practices from real-world implementations, plus cool videos!
• Wednesday at 9:30am, room 230C
Work better. Live better.