S4783: Virtual is Better than Physical

Delivering a Delightful User Experience from a Virtual Desktop
About the speaker

Kenneth Fingerlos - @KFingerlos
Solutions Architect – Lewan Technology
Focus on Application, Desktop, Datacenter Virtualization

VMware Premier, Citrix Platinum, Microsoft Gold
Liquidware, AppSense, Atlantis Computing
Dell, HP, Cisco, NetApp, Nimble
Planning for desktop virtualization
Design choices and how they impact user experience
About the Any’s (Any Device, Anytime, Anywhere)
“Those who fail to plan, plan to fail”

-Michael Baisden
Who are your user groups?
What applications do they (really) use?
How and When do they work?
How will they use a virtual desktop (or applications)?
How will you deliver applications (or desktops) to them?
What are the user’s current expectations?
Planning

Why are we talking about virtualization?
What are we trying to accomplish?
What is our plan to get from here to there?
How will we know when we get there?

You need to be able to articulate the answers!
“One size does not fit all.”

-Frank Zappa
Break users down into homogenous groups
- Similar application needs
- Similar performance needs
- Similar usage patterns
- Similar expectations

Then design a solution around each group
Consider a formal assessment

- Discover application usage
- Discover resource usage
- Discover user login/logoff patterns

…and lots more you probably didn’t know…
<table>
<thead>
<tr>
<th>Application Name</th>
<th>Machine Install Count</th>
<th>Machine Used Count</th>
<th>Approx. Launches Per User</th>
<th>Avg. Launch Delay</th>
<th>CPU Time Used</th>
<th>ANR per Sec</th>
<th>Peak Avg. Mem Used</th>
<th>Peak Avg. IOPs</th>
<th>Total Install Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microsoft .NET Framework 4 Extended (v4.0.30319)</td>
<td>23</td>
<td>1</td>
<td>2</td>
<td>00:00:04.547s</td>
<td>00:00:00.1s</td>
<td>3030.3</td>
<td>17 MB</td>
<td>0</td>
<td>33 GB</td>
</tr>
<tr>
<td>ShoreTel Communicator (v16.43.1410)</td>
<td>23</td>
<td>20</td>
<td>18</td>
<td>00:00:11.881s</td>
<td>00:00:07.364s</td>
<td>1011.6</td>
<td>30 MB</td>
<td>3.03</td>
<td>2 GB</td>
</tr>
<tr>
<td>Google Earth (v7.1.2)</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td>00:00:01.614s</td>
<td>00:00:01.641s</td>
<td>78.6</td>
<td>660 MB</td>
<td>9.07</td>
<td>4 GB</td>
</tr>
<tr>
<td>persistence Module (v8.15.10)</td>
<td>20</td>
<td>19</td>
<td>6</td>
<td>00:00:00.33s</td>
<td>00:00:00.000s</td>
<td>12610000.0</td>
<td>3 MB</td>
<td>0</td>
<td>8 MB</td>
</tr>
<tr>
<td>hrcmd Module (v8.15.10)</td>
<td>19</td>
<td>19</td>
<td>5</td>
<td>00:00:00.29s</td>
<td>00:00:00.000s</td>
<td>0.0</td>
<td>2 MB</td>
<td>0</td>
<td>7 MB</td>
</tr>
<tr>
<td>Internet Explorer (v11.00.9600)</td>
<td>19</td>
<td>16</td>
<td>33</td>
<td>00:00:00.73s</td>
<td>00:00:43.825s</td>
<td>42.9</td>
<td>82 MB</td>
<td>5.02</td>
<td>15 MB</td>
</tr>
<tr>
<td>Java(TM) Platform SE binary (v6.0.220)</td>
<td>18</td>
<td>17</td>
<td>8</td>
<td>00:00:23.770s</td>
<td>00:00:31.407s</td>
<td>73.4</td>
<td>124 MB</td>
<td>1.02</td>
<td>2 MB</td>
</tr>
<tr>
<td>Google Chrome (v32.0.1700)</td>
<td>17</td>
<td>17</td>
<td>58</td>
<td>00:00:40.415s</td>
<td>00:02:52.272s</td>
<td>49.5</td>
<td>58 MB</td>
<td>90.26</td>
<td>14 MB</td>
</tr>
<tr>
<td>Microsoft Office Single Image 2010 (v14.0.7015)</td>
<td>17</td>
<td>17</td>
<td>25</td>
<td>00:00:01.418s</td>
<td>00:00:49.843s</td>
<td>100.3</td>
<td>67 MB</td>
<td>11.28</td>
<td>35 GB</td>
</tr>
<tr>
<td>Roxio Burn (v1.8)</td>
<td>15</td>
<td>13</td>
<td>4</td>
<td>n/a</td>
<td>00:00:00.000s</td>
<td>3533333.3</td>
<td>5 MB</td>
<td>0</td>
<td>484 MB</td>
</tr>
<tr>
<td>Trusted Drive Manager (v4.1.1)</td>
<td>13</td>
<td>13</td>
<td>6</td>
<td>00:00:00.29s</td>
<td>00:00:00.000s</td>
<td>0.0</td>
<td>2 MB</td>
<td>0</td>
<td>363 MB</td>
</tr>
<tr>
<td>FF_Protection MFC Application (v1, 0, 0, 0)</td>
<td>11</td>
<td>10</td>
<td>5</td>
<td>00:00:00.31s</td>
<td>00:00:00.000s</td>
<td>14158333.3</td>
<td>2 MB</td>
<td>0</td>
<td>7 MB</td>
</tr>
<tr>
<td>Microsoft .NET Framework 4.5.1 (v4.5.50938)</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>00:00:01.193s</td>
<td>00:00:00.000s</td>
<td>0.0</td>
<td>2 MB</td>
<td>0</td>
<td>2 GB</td>
</tr>
<tr>
<td>Intel(R) Management and Security Status (v7.1.20)</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>00:00:00.408s</td>
<td>00:00:00.13s</td>
<td>141136.4</td>
<td>13 MB</td>
<td>0.01</td>
<td>18 MB</td>
</tr>
<tr>
<td>Microsoft Office Professional Plus 2010 (v14.0.7015)</td>
<td>10</td>
<td>9</td>
<td>24</td>
<td>00:00:01.788s</td>
<td>00:00:11.905s</td>
<td>295.6</td>
<td>24 MB</td>
<td>2.82</td>
<td>16 GB</td>
</tr>
<tr>
<td>Hour of Day</td>
<td>User Count</td>
<td>Machine Count</td>
<td>Total Avg. IOPS</td>
<td>Total Avg. Peak IOPS</td>
<td>Total Avg. Peak Reads</td>
<td>Total Avg. Peak Writes</td>
<td>Total Avg. Peak Read Percent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>---------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:00a</td>
<td>267</td>
<td>331</td>
<td>7962.39 IOPS</td>
<td>31222.57 IOPS</td>
<td>21162.34 IOPS</td>
<td>10076.51 IOPS</td>
<td>68.04 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:00a</td>
<td>268</td>
<td>332</td>
<td>8898.85 IOPS</td>
<td>36900.53 IOPS</td>
<td>26074.88 IOPS</td>
<td>10850.64 IOPS</td>
<td>70.66 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:00a</td>
<td>267</td>
<td>331</td>
<td>9811.15 IOPS</td>
<td>42893.44 IOPS</td>
<td>31052.31 IOPS</td>
<td>11891.73 IOPS</td>
<td>72.33 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00p</td>
<td>264</td>
<td>329</td>
<td>8014.73 IOPS</td>
<td>32616.65 IOPS</td>
<td>21842.03 IOPS</td>
<td>10879.89 IOPS</td>
<td>66.92 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:00p</td>
<td>264</td>
<td>330</td>
<td>8007.05 IOPS</td>
<td>35309.77 IOPS</td>
<td>25063.33 IOPS</td>
<td>10371.71 IOPS</td>
<td>70.75 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:00p</td>
<td>264</td>
<td>330</td>
<td>6376 IOPS</td>
<td>25896.63 IOPS</td>
<td>19216.9 IOPS</td>
<td>6700.74 IOPS</td>
<td>74.42 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:00p</td>
<td>265</td>
<td>331</td>
<td>7770.43 IOPS</td>
<td>33773.63 IOPS</td>
<td>24596.59 IOPS</td>
<td>9442.38 IOPS</td>
<td>72.41 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:00p</td>
<td>260</td>
<td>329</td>
<td>6047.71 IOPS</td>
<td>25194.77 IOPS</td>
<td>16996.11 IOPS</td>
<td>8223.83 IOPS</td>
<td>67.42 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:00p</td>
<td>245</td>
<td>323</td>
<td>6474.03 IOPS</td>
<td>25046.35 IOPS</td>
<td>15105.41 IOPS</td>
<td>9947.83 IOPS</td>
<td>60.36 %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metrics</td>
<td>Allocated</td>
<td>Average Used</td>
<td>Peak Average Used</td>
<td>Peak Used</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------</td>
<td>--------------</td>
<td>-------------------</td>
<td>------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPU MHz</td>
<td>1458.75 GHz</td>
<td>88.47 GHz</td>
<td>196.64 GHz</td>
<td>599.9 GHz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAM</td>
<td>1013.81 GB</td>
<td>532.93 GB</td>
<td>573.12 GB</td>
<td>727.79 GB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Page/Swap File</td>
<td>1.23 TB</td>
<td>128.6 GB</td>
<td>136.45 GB</td>
<td>236.5 GB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage Capacity</td>
<td>68.78 TB</td>
<td>13.74 TB</td>
<td>13.75 TB</td>
<td>16.66 TB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read IOPs</td>
<td></td>
<td>3270.51 IOPs</td>
<td>15572.53 IOPs</td>
<td>89330.2 IOPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write IOPs</td>
<td></td>
<td>2384.91 IOPs</td>
<td>7391.52 IOPs</td>
<td>84254.97 IOPs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read Rate</td>
<td></td>
<td>14.39 GBps</td>
<td>34.76 GBps</td>
<td>253.9 GBps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Write Rate</td>
<td></td>
<td>13.29 GBps</td>
<td>30.65 GBps</td>
<td>164.06 GBps</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network B/w Capacity</td>
<td>52.62 GB/s</td>
<td>634.97 MB/s</td>
<td></td>
<td>2.18 GB/s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Send Rate</td>
<td></td>
<td>17.69 MB/s</td>
<td>48.14 MB/s</td>
<td>318.38 MB/s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network Receive Rate</td>
<td></td>
<td>19.99 MB/s</td>
<td>54.79 MB/s</td>
<td>363.51 MB/s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“It’s the art of design”
- anonymous
Know that industry averages are … average…
Best practices and reference architectures are …stuff that works well most of the time…

As you begin your design, know if you are average, and if best practices apply to your specific situation!
## Design

<table>
<thead>
<tr>
<th>CPU - vCPU/Core</th>
<th>vCPU</th>
<th>vCPU</th>
<th>vCPU</th>
<th>vCPU/PVD</th>
<th>vCPU/PVD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>old</td>
<td>Dual/2S</td>
<td>Quad/4S</td>
<td>Dual/2S</td>
<td>Quad/4S</td>
</tr>
<tr>
<td>User Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>8-10</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Normal</td>
<td>6-8</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Heavy</td>
<td>2-4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RAM - GB/VM</th>
<th>Old</th>
<th>Pooled</th>
<th>PVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>1-1.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Normal</td>
<td>1.5-2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Heavy</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IOPS - IOPS/VM</th>
<th>Old</th>
<th>MCS</th>
<th>PVS</th>
<th>MCS/PVD</th>
<th>PVS/PVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light</td>
<td>4-6</td>
<td>8</td>
<td>5</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Normal</td>
<td>8-12</td>
<td>16</td>
<td>10</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Heavy</td>
<td>25-50</td>
<td>32+</td>
<td>20+</td>
<td>29+</td>
<td>18+</td>
</tr>
</tbody>
</table>

Guidelines updated from [http://virtualfeller.com/2013/05/02/virtual-desktop-resources-then-and-now/](http://virtualfeller.com/2013/05/02/virtual-desktop-resources-then-and-now/) - by Daniel Feller, Lead Architect

"old" guidelines from 2011, new guidelines updated 2013. Principal change - Faster processors
Consider the customer assessment shared earlier:

• 268 users
• Peak 600Ghz of CPU, 728GB of RAM, 90,000 IOPS
• Peak Averages of 197Ghz, 574GB RAM, 23,000 IOPS

Observed:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHz per user</td>
<td>0.73-2.3 Ghz</td>
</tr>
<tr>
<td>GB RAM/user</td>
<td>2.14GB-2.71GB</td>
</tr>
<tr>
<td>IOPS per user</td>
<td>86-335</td>
</tr>
</tbody>
</table>

Guideline (Normal-Heavy):

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHz per user</td>
<td>0.31-0.63 Ghz</td>
</tr>
<tr>
<td>GB RAM per user</td>
<td>2-4 GB</td>
</tr>
<tr>
<td>IOPS per user</td>
<td>16-32+</td>
</tr>
</tbody>
</table>

KNOW YOUR ENVIRONMENT!
VDI vs. SBC
Design

VDI – Windows Desktop OS running in a VM – 1:1
SBC – Windows Server OS running RDS – 1:n

VDI – Virtual Desktop Infrastructure
SBC – Server Based Computing
a.k.a. Terminal Services
VDI Advantages

- Every user is in their own VM, providing per user isolation
- Users can be granted admin rights within their session
- Software is running on a desktop OS
- Device Pass-through is more robust
- GPU shared at the virtual hardware level
Design

SBC Advantages

• Lower resource requirements since a single instance supports multiple users

• Lower per-user license costs (MS RDS CAL vs. VDA subscription)

• Fewer instances required to support a total number of users

• GPU can be shared at the OS level
SBC and VDI Disadvantages

• Some software requires admin privs - VDI
• Some software requires desktop OS – VDI
• More protocol options for VDI – VDI
• VDI resources can be 10x SBC – SBC
• MS VDA is $100/year, RDS CAL is $92. – SBC
SBC vs VDI Conclusion

- If cost is of greater concern - SBC
- If flexibility is of greater concern – VDI

Mix and Match in the same environment
- XenApp + XenDesktop!
- View + MS RDS!
- MS RDS supports both!
Persistent vs. Non-persistent
Non-Persistent – VDI or SBC where changes/customization are lost after a reboot

Persistent – VDI or SBC where changes are preserved and users are assigned to specific instances
Persistent Advantages

- Physical desktops are inherently persistent
- Do what you’ve always done

Why change what works?
Persistent Considerations

- Backups
- Availability
- Users must be returned to the same instance
- Management infrastructure (patches, software dist)
- Provisioning
- Cost
Non-Persistent Advantages

- Gold Image based
- User connects to first available instance
- Potential lower total resource needs
- Cost (instance count = max concurrent users)
- Always new
- Inherent backup/availability solution
Non-Persistent Advantages (cont.)

- Option for local storage (vs shared)
- User changes can be intentionally lost - lab/demo/kiosk
- Simplified update process – no SCCM, LANDesk, etc.
but I need something in the middle...
Persistent – Non-Persistent Hybrid options

- Roaming Profiles, Profile managers, UEV
- Folder Redirection
- Layering – PVD, Mirage, Liquidware

Adds some complexity; rules; adds some flexibility
Looking to merge benefits of both NP and P desktops
Folder Redirection

Take the common folders from the desktop – Desktop, Documents, Downloads, etc. and redirect them from the user profile to a network location

This alone will make the contents of the folders persist and float between desktops
Profile Managers and Roaming Profiles

Take the user profile (c:\users\username) and copy it to a network location at logoff, copy it back at logon.

Provides persistence for the contents of the profile which includes most user preferences and data.
MS Roaming profiles are built into the OS
Citrix and VMware provide profile managers with additional capabilities (deeper control, policy)

3rd parties such as Liquidware Labs Profile Unity provide yet more control and the ability to map settings between v1 and v2 profiles, which is useful to mix XP, Win7, Win8 or support migration
User Environment Virtualization (UEV)

• More sophisticated approach than typical profile management.

• Wrap a bubble around the user environment and track changes made, then replay those changes to a new session upon application launch

• Can get very granular about what is persisted
Also introduce a significant policy engine to allow for managing the user environment

- Map printers
- Map drives
- Logon and logoff scripts
- Policy actions within the environment
Layering

- Break the desktop into “layers” for OS, Applications, and user data
- User desktop is constructed by assembling the layers into a composite image
- Layers can then be updated/removed/added independently of each other
Citrix Personal vDisk (PVD) and VMware user data disk (UDD) provide a “user layer” on top of the system image. This allows the system image to be updated without impacting data in the user layer.

These technologies are included with their respective products and provide persistence while allowing administrators to utilize image based management.
VMware Mirage provides a more sophisticated layering technology which can be used with VDI and Physical desktops.

Liquidware Labs Profile Unity supports a user layer attached dynamically to non-persistent desktops and also supports admin assigned departmental layers.
Persistent vs Non-Persistent

Computer Labs, Training Rooms – NP
Application Developers – P
Everybody else … It depends

Most customers end up with a mix of Persistent, Hybrid, and Non-Persistent with the mix changing over time.
“And now here’s something we hope you’ll really like”
-Rocky the Flying Squirrel
Design

CPU
Memory
Disk

Network
Graphics
Desktop virtualization is a high demand workload
Don’t mix desktop workloads with server workloads; when resources are in contention desktops win or UX suffers
Do not build your desktop virtualization project on last-year’s hardware or the castoffs from the server folks
If your servers are 5ms slower, users won’t care
If their desktops are 5ms slower they will notice
Windows Experience Index

- Benchmark built into Windows OS in Windows Vista, Windows 7 and Windows 8 (removed in 8.1)
- Requires at least some 3D support to run
- Rates subsystems separately and the lowest overall score determines the system’s score
- Useful for comparing a system’s overall ability to deliver a user experience
Rate and improve your computer’s performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

New hardware detected
Your Windows Experience Index needs to be refreshed

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor:</td>
<td>Calculations per second</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td>Memory (RAM):</td>
<td>Memory operations per second</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Graphics:</td>
<td>Desktop performance for Windows Aero</td>
<td>4.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Gaming graphics:</td>
<td>3D business and gaming graphics performance</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>Primary hard disk:</td>
<td>Disk data transfer rate</td>
<td>5.7</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?
Tips for improving your computer’s performance.

Learn more about scores and software online.

Last update: 8/14/2012 5:28:33 PM
Rate and improve your computer’s performance

The Windows Experience index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Calculations per second</td>
<td>6.1</td>
<td>5.9</td>
</tr>
<tr>
<td>Memory (RAM)</td>
<td>Memory operations per second</td>
<td>6.1</td>
<td>5.9</td>
</tr>
<tr>
<td>Graphics</td>
<td>Desktop performance for Windows Aero</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Gaming graphics</td>
<td>3D business and gaming graphics performance</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>Primary hard disk</td>
<td>Disk data transfer rate</td>
<td>5.9</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?

Tips for improving your computer’s performance.

Learn more about scores and software online.

Your scores are current
Last update 9/23/2012 10:18:45 PM
Typical Laptop – 4.4-6.0 base score
• Processor – 5.6-7.8
• Memory – 5.9-7.9
• Business Graphics – 4.6-6.2
• Gaming Graphics – 6.1-6.6
• Disk – 5.6-6.3

So we want our virtual desktop in these ranges

Rate and improve your computer's performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Calculations per second</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>Memory (RAM)</td>
<td>Memory operations per second</td>
<td>7.9</td>
<td></td>
</tr>
<tr>
<td>Graphics</td>
<td>Desktop performance for Windows Aero</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Gaming graphics</td>
<td>3D business and gaming graphics performance</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>Primary hard disk</td>
<td>Disk data transfer rate</td>
<td>6.1</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?

Tips for improving your computer's performance.

Learn more about scores and software online.

Your scores are current
Last update: 10/12/2013 10:54:31 AM
Rate and improve your computer’s performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Calculations per second</td>
<td>7.2</td>
<td>7.4</td>
</tr>
<tr>
<td>Memory (RAM)</td>
<td>Memory operations per second</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>Graphics</td>
<td>Desktop performance for Windows Aero</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>Gaming graphics</td>
<td>3D business and gaming graphics performance</td>
<td>7.9</td>
<td>7.9</td>
</tr>
<tr>
<td>Primary hard disk</td>
<td>Disk data transfer rate</td>
<td>7.4</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?
Tips for improving your computer’s performance.
Learn more about scores and software online.

Your scores are current
Last update: 10/14/2013 9:13:55 PM
Compute
CPU is virtualized on a time-slicing basis with a scheduler and priority mechanism to address contention.

Not enough and user sessions run slow, hang, disconnect.

Generally VDI configs are memory bound rather than CPU. SBC configs are more likely to be CPU bound.

Generally faster is better, yielding more responsive desktops.
CPU - AMD vs INTEL

• Generally AMD offers more cores, but Intel offers faster cores.
• Benchmarks and load testing favor Intel systems hosting greater densities (>60%), but at a higher price tag per host.
• Generally Intel’s density offsets the higher cost
• Generally Intel’s faster core gives a better UX
CPU –

- 12 cores at 2.5 Ghz = 30Ghz
- 0.3Ghz/desktop means 100 desktops
- 0.6Ghz/desktop means 50 desktops
- 2.5Ghz/desktop means 12 desktops

Plan CPU capacity according to user workloads, allowing for peak demands to be met.
CPU – Hyperthreading

• Can provide some additional logical capacity however it’s very dependent upon the workloads being executed

• If workloads are largely homogenous the benefit decreases

• Plan environment based on physical core capacity

• Know that HT will provide some additional headroom
Multiple virtual CPUs (vCPUs)

- General guidance is to only use what’s needed, but
- User experience generally improves with 2 or more
- Moving from 1 vCPU to 2 will directly reduce VM density
Memory
Design

Memory

- Swapping degrades UX rapidly
- Physical memory bus is a shared resource
- 100 2GB VM’s means 200GB of real memory
- Hypervisor will use some RAM itself
Hyper-v and XenServer allow variable memory assignments (1-3GB for instance)
vSphere requires a hard assignment (2GB) but allows memory overcommit and will swap at the hypervisor if needed … but swapping remains bad for UX
Because of the per-VM swap, vSphere creates a swap file for each VM on storage when the VM powers on
This swapping (and the file) can be disabled via memory reservations
Storage
Design

• The number one performance issue encountered with desktop virtualization is undersized storage performance

• Storage performance and capacity are independent

• There is more than one way to skin this cat

• Remember to think about UX when making storage decisions
Design

IOPS – IO Operations Per Second
Local Storage, DAS – Storage internal or connected to a single physical host
Shared, SAN – Storage physically connected to multiple hosts such that all have equal access

IO Latency – Measure of how long a drive takes to respond to a command (perform an IO operation) generally measured in milliseconds (ms)
Design

• 7,200 RPM drives generally perform 80 IOPS
• 10,000 RPM drives generally perform 120 IOPS
• 15,000 RPM drives generally perform 180 IOPS
• SSD drives generally perform 2000+ IOPS
SAN Advantages

- Conventional approach for most datacenters
- Provides high availability
- Provides high performance
- Relatively easy to manage
- Facilitates VM mobility between hosts
- Simplified DR through SAN replication
- Good choice for persistent deployments
SAN Downsides

- Relatively high cost
- Stair-step scalability
Local Advantages

- Low cost
- Simple
- Easy to scale
- Easy to configure
- Good choice for non-persistent
Local downsides

- Complicates live migration and HA
- Complicates provisioning
- Complicates persistent scenarios
Design

Local vs SAN

• Non-Persistent desktops – Local storage
• Persistent desktops without HA – Local storage
• Persistent desktops requiring HA – SAN storage
RAM based IO accelerators

- Virtual machine using RAM to either optimize or eliminate physical disk IO
- Products from Atlantis Computing support both persistent and non-persistent desktops
- Basic concept – 256GB of RAM costs $3900 and 200GB enterprise SSD costs $5900
- IO Accelerator can eliminate the need for high performance disk
Design

• Accelerator uses RAM from host – usually 20-64GB to optimize and compress VM IO.
• Only optimized writes go to physical disk
• Can result in a 100x reduction in physical disk IO
• For non-persistent solutions physical disk IO can be completely eliminated
• IO from RAM measured in nano-seconds as compared to disk IO measured in milli-seconds
Design

- Accelerator performance scales as additional hosts are added
- End result is a lower cost solution for providing very fast IO to virtual desktops
- HA is handled at the accelerator appliance level – the appliance and dependent desktops must migrate together
- Compatible with all existing storage – local and SAN
Storage Sizing – Capacity – Full Clones/Persistent

• (size of the image+size of RAM)*number of images

• (40GB Image+2GB RAM)*200 desktops = 8,400 GB

Storage deduplication can reduce this footprint significantly by only tracking unique blocks for each VM
Storage Sizing – Capacity – Linked Clones Non-P

- Varies with placement of replica files, number of pools and technology.
- VMware Composer allows placement of replicas on separate storage from the VM’s themselves.
- Size of Image + (VM Delta + RAM)*VMs
- $40\text{GB} + (5\text{GB}+2\text{GB})*200 = 1800\text{GB}$
Design

- Citrix Machine Creation Services (MCS) requires placement of replicas on the storage with the VM’s themselves; with max of about 64 VM’s in a group
- With 200 VM’s you’d have 4 groups of 50 each
- \( 4 \times (\text{size of image} + (\text{size of RAM} + \text{VM Delta}) \times 50) \)
- \( 4 \times (40\text{GB} + (2\text{GB} + 5\text{GB}) \times 50) = 1560\text{GB} \)
• Citrix Provisioning Services (PVS) requires placement of images on the PVS server
• PVS Server Disk Size = Size of Image(s)
• Number of VMs*(Size of RAM + VM Delta)
• 200*(2GB+5GB) = 1400GB

Note that the “size of RAM” figure is vSphere specific and is zero for Hyper-v and XenServer.
Storage Sizing – Performance

- Understand the IO profile of the desktops – how many IOPS to size for and the read/write mix
- RAID5 requires 4 IOPS for every write IO
- RAID1 and RAID10 require 2 IOPS for every write IO

These are referred to as the “raid penalty” associated with performing writes in a RAID configuration
General VDI IO mix is accepted as being 80-90% writes

If we assume 200 desktops @ 20 IOPS each, (4000 raw IOPS) and 90% writes...

• RAID10 effective IOPS are 7,600 IOPS
• RAID5 effective IOPS are 18,400 IOPS
Physical disks to support this IO load would be:

- RAID10 – 96 7.2K, 64 10K, 44 15K, or 4 SSD’s
- RAID5 – 230 7.2K, 154 10K, 103 15K, or 8 SSD’s

Citrix PVS and VMware Composer would allow you to size separate disk solutions for reads and writes.
Size performance first, *then* capacity
Example we need 4000 IOPS @ 90% write and 1400GB usable capacity

Solutions then are:

- RAID10 – 96 29GB 7.2K, 64 44GB 10K, 44 64GB 15K, 4 700GB SSD
- RAID5 – 230 6GB 7.2K, 154 9GB 10K, 103 14GB 15K, 8 200GB SSD
- RAM Accelerators + 1400GB of slow-medium speed disk
Design

• Use drive sizes and quantities to price out storage options
• All Flash arrays will often include 8-16 SSD drives in a single chassis
• Enterprise arrays from major vendors can incorporate SSD drives into their configurations
• IO Accelerators will reduce or eliminate disk performance requirements
SSD drive use flash memory to store data. The physical nature of flash memory is such that repeated overwrites of the same memory cell eventually wear out the cell.

For this reason use of SSD to store data which is frequently overwritten may result in premature failure of the SSD device.
Design

Technologies such as TRIM and “wear leveling” attempt to minimize these failures but do not eliminate the issue.

For this reason it is very important to understand how vendors use and warranty SSD storage.
Perspective – SAN costs for 2000 IOPS (persistent)

- Ram costs $3900 for 256GB + disk backing (~$2500)
- Enterprise SSD costs – $11,800 (per 200GB usable)
- Enterprise 15K costs - $21,450
- Enterprise 10K costs - $40,640
- Enterprise 7.2K costs - $46,752
Design

Perspective – Local costs 16-disks – Non-persistent
• Ram costs $3900 for 256GB @ 20,000 IOPS
• Local SLC SSD costs - $2760 @ 2400 IOPS – 200GB Usable
• Local 15K costs - $6192 @ 1400 IOPS – 2.4TB Usable
• Local 10K costs - $6192 @ 900 IOPS – 4.8TB Usable
• Local 7.2K costs - $3920 @ 600 IOPS – 4.0 TB Usable

Its all about user density and UX here
Perspective – RAM IO Accelerators

- RAM is inherently faster than disk
- Persistent RAM accelerators will greatly reduce costs for SAN configs
- Non-persistent SAN represents an expensive option, but less so with RAM accelerators
- Non-persistent RAM accelerators will reduce cost or be cost neutral but will improve performance in local configs
Network
How much network bandwidth do I need?
It depends on a number of factors; what is the user actually doing? What display protocol is in use? and how much compression is happening.

Office applications can be as little as 384Kb/s/user, but video, animation, and graphics can be 10Mb/s or higher.
Network bandwidth is not constant. It will vary from mouse click to mouse click.

Display protocol choice (RDP, HDX, PCoIP) will also impact the bandwidth use; as can the tuning of these protocols.

Easiest to evaluate actual bandwidth during a pilot.
Insufficient bandwidth will result in slow screen paints as data trickles in; and may even result in dropped or hung sessions.

Excessive network latency can give the impression that the session is running slow and even render it unusable, and can cause dropped sessions.
Design

- Can generally tolerate network latency up to 120-150ms
- Some cases can be usable up to 300ms
- Tuning can reduce bandwidth requirements

It is important to understand the network conditions you expect to deliver desktops over
Graphics
Consider that

- nearly every PC (or computer) shipped in the last 20 years has had some form of hardware graphics accelerator
- Applications are becoming increasingly graphics intensive
- Traditional virtual machine have no graphics hardware at all
Design

- Traditional virtual desktops render all of their 2D graphics on the system CPU
- Traditional virtual desktops have limited (if any) 3D rendering capability
- Traditional virtual desktops are therefore heavily optimized to minimize graphic content

But this creates problems…
Also consider that

- Windows 7 is optimized for graphics hardware
- Users expect a rich graphical experience
- Applications expect to be able to render 3D graphics via OpenGL and DirectX

How do we solve this?
Software Rendering
• Provides basic 3D rendering support using CPU
• Suitable for basic applications and static images
• Compatible with virtually any hardware

• Slow and resource intensive
• Limited capability for recent DirectX and OpenGL versions
Rate and improve your computer’s performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Calculations per second</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>Memory (RAM)</td>
<td>Memory operations per second</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>Graphics</td>
<td>Desktop performance for Windows Aero</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Gaming graphics</td>
<td>3D business and gaming graphics</td>
<td>2.5</td>
<td>Determined by lowest subscore</td>
</tr>
<tr>
<td>Primary hard disk</td>
<td>Disk data transfer rate</td>
<td>7.9</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?
- View and print detailed performance and system information
- Tips for improving your computer's performance.
- Learn more about scores and software online

Your scores are current
Last update: 10/26/2013 3:35:31 PM
GPU Pass-Through

• Physical pass-through of a GPU to a virtual machine
• Virtual machine runs the GPU’s native drivers
• Provides full support of latest OpenGL and DirectX
• Provides full performance of the GPU
• Suitable for Professional Graphics needs

• Limited to 1 VM per GPU
Rate and improve your computer’s performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor:</td>
<td>Calculations per second</td>
<td></td>
<td>7.2</td>
</tr>
<tr>
<td>Memory (RAM):</td>
<td>Memory operations per second</td>
<td></td>
<td>7.9</td>
</tr>
<tr>
<td>Graphics:</td>
<td>Desktop performance for Windows Aero</td>
<td></td>
<td>7.9</td>
</tr>
<tr>
<td>Gaming graphics:</td>
<td>3D business and gaming graphics performance</td>
<td></td>
<td>7.9</td>
</tr>
<tr>
<td>Primary hard disk:</td>
<td>Disk data transfer rate</td>
<td></td>
<td>7.4</td>
</tr>
</tbody>
</table>

Determined by lowest subscore

What do these numbers mean?
What do these numbers mean?
Tips for improving your computer’s performance.

Learn more about scores and software online

Your scores are current
Last update 10/14/2013 9:13:55 PM
GPU Virtualization – VMware vSGA

- API intercept and GPU virtualization using VMware SVGA driver
- Leverages GPU for faster graphics rendering
- Allows fall-back to software rendering

- Limited OpenGL and DirectX version support
- Variable performance based on system load
Rate and improve your computer’s performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor:</td>
<td>Calculations per second</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>Memory (RAM):</td>
<td>Memory operations per second</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>Graphics:</td>
<td>Desktop performance for Windows Aero</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Gaming graphics:</td>
<td>3D business and gaming graphics performance</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>Primary hard disk:</td>
<td>Disk data transfer rate</td>
<td>7.9</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?

Tips for improving your computer’s performance.

Learn more about scores and software online.

Your scores are current
Last update: 10/25/2013 11:53:27 PM
Design

GPU Virtualization – NVIDIA vGPU
• Allocate a subset of a GPU’s physical resources
• Virtual machine runs the GPU’s native drivers
• Supports latest OpenGL and DirectX
• Performance based on allocated resources

• XenServer only
• Max users dependent on resource allocation
Rate and improve your computer’s performance

The Windows Experience Index assesses key system components on a scale of 1.0 to 7.9.

<table>
<thead>
<tr>
<th>Component</th>
<th>What is rated</th>
<th>Subscore</th>
<th>Base score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor:</td>
<td>Calculations per second</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>Memory (RAM):</td>
<td>Memory operations per second</td>
<td>7.2</td>
<td></td>
</tr>
<tr>
<td>Graphics:</td>
<td>Desktop performance for Windows Aero</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>Gaming graphics:</td>
<td>3D business and gaming graphics performance</td>
<td>7.1</td>
<td></td>
</tr>
<tr>
<td>Primary hard disk:</td>
<td>Disk data transfer rate</td>
<td>6.6</td>
<td></td>
</tr>
</tbody>
</table>

What do these numbers mean?

Tips for improving your computer’s performance.

Learn more about scores and software online

Your scores are current
GPU Sharing – SBC

• Pass-Through GPU to a SBC desktop
• SBC runs native GPU Drivers
• Supports latest OpenGL and DirectX
• High performance based on allocated resources
• Allows GPU to be shared by multiple sessions on the VM
• Performance varies with user load
Graphics Performance (best to worst)
- GPU Pass-Through/GPU Sharing
- NVIDIA vGPU
- VMware vSGA
- Software Rendering
- Traditional VDI (no 3D)

Balance cost and performance with need
Endpoints Matter
The users endpoint is the device charged with presenting the desktop experience to the user

The endpoint can’t improve (much) on what the datacenter does, but it can certainly detract from it

Rendering 60FPS of HD video in the desktop is useless if the endpoint can only render 5fps
Endpoints

Thin Client types –

- Zero Clients – firmware only devices, often have hardware assists to decode display protocols
- Linux Embedded – customized Linux distribution and client software
- Windows Embedded – Windows XP/7/8 and client software
Client software does matter.

CPU in the endpoint does matter, as do hardware decoding capabilities.

A thin client which drives a 17” or 19” monitor ok, may not be able to handle the additional CPU load of a 20” or 24” display.

You get what you pay for; an inexpensive thin client may be fine for a task worker, but be insufficient for professional graphics.
The Any’s

Anytime, Anywhere, Anything
Endpoints Matter!

Networking Matters!

Time of day?
- Yes! Backups, Maintenance, After-School Surfing!
Virtual is better than Physical

Delivering a delightful experience
Consider Windows Experience Scores (7.9 max)
- Typical laptop provides a maximum of a 5.9; limited by disk IO
- Typical laptop scores 4.6-6.2 for business graphics
- Using IO Accelerators and GPU Pass-Through you can build a VDI desktop scoring 7.9 (maximum)
- Using IO Accelerators and vSGA you can build a desktop scoring 6.0 graphics and 7.9 on disk.
Virtual is Better

And you get the benefits of being virtual
- Access from your iPad
- Access from home
- Access from Starbucks
- Security
- Control
- Flexibility
- ...and so many more
@KFingerlos #GTC14

S4783  Virtual is Better than Physical