Document: Stratusphere Spot Check #'s for VMware VDI
Last Updated: 01/15/2017
Document Purpose: Define metric and thresholds for a Spot Check as it relates to User Experience in a VMware VDI environment utilizing Liquidware Labs Stratusphere UX.

This document is designed to bring together the recommendations from many experts in the industry about the metrics that need to be monitored and the thresholds that are deemed acceptable.

This document does not make recommendations on changes needed due to the many industry, usage, costing, and application variables that are in play.

What is a "Spot Check"

The thresholds represented below are at a 1 hour level of granularity unless otherwise specified in the description. Spot Checks are basically point in time health checks. Review of data from multiple dates and times is critical before making changes or recommendations. Normal and High Usage dates and times should be examined based on the industry and user requirements.

CRITICAL NOTES:

A) Know your company!

Knowledge of the industry/company/department work habits, loads and applications are critical for data interpretation and threshold evaluation.

Example 1: Moderate/High Storage Latency may be acceptable during shift changes with lots of people logging in and out. BUT, Not Acceptable during work hours as this impedes productivity.

Example 2: Law firms and Hospitals generally demand sub 10 second login times whereas most other industries are satisfied with under 30 second.

B) Know your data!

There are many monitoring and diagnostic systems on the market. Each of the solutions available collect data in different ways and with different levels of granularity. They all also render the data in different ways and granularity with role ups that can vastly change the data and perspective for the viewer. For this reason, the metric values represented in this document are only for Liquidware Labs Stratusphere and may not apply well to other products.

Examples:

Depending on the view you could be looking at average, peaks or peak averages.
Did the data come from Broker, Hypervisor, "In-Guest", "In Band" or "Out of Band"?
How much impact did the "In-Guest" Agent put on the OS?
How much impact and time lag is on the "Out of Band" Agent, Broker and Hypervisor?

C) Good Blogs

Grey Matter is Required – Automated Solutions don't work
Monitoring vs. Diagnostics

Documentation needed for analysis, conclusion and recommendations:

Multiple Spot Check Dates:

MM/DD/YYYY (Monday), MM/DD/YYYY (Wednesday), MM/DD/YYYY (Friday)

Multiple Time Frames Examined on each date:

(Time frames for review are based on business requirements)

9-10AM, 10-11AM, 2-3PM, 4-5PM

The systems will be examined on each date and time for the following information based on the max values shown below:

Critical Sections for Review:

1. ESX Host Criteria: (VMware Best Practices)

CPU % - (Max 75% Average over 60 minutes)

CPU % Ready (Max 3% Average over 60 minutes)

Memory Utilization (Max 85% at any point in time)

Memory Swapping and Ballooning (Should always be ZERO)

Datastore Latency (Should be under 10 Milliseconds – Max 15 Milliseconds)

2. Virtual Machines Criteria: (VDI)

Login Delay (Industry Average is under 30 Seconds – This is a company preference)

Application Load Time (Industry Average is under 3 Seconds – Company Preference)

CPU Utilization (Max 80%) – Higher than 50% generally is bad over 60 Minutes

- This generally denotes a stuck or run away process on the machine

CPU Queue (Should not be more than 2 per vCPU assigned to VM)

https://technet.microsoft.com/en-us/library/Cc940375.aspx

Memory Usage (Should be less than 75%)

VMware Best Practice – to reduce Windows Paging

Page File Usage (Should be as close to zero as possible)

You cannot stop Windows from paging.

Soft Page Faults are in memory – Hard Page Faults are to disk.

Do Not turn off the paging file in windows. Set to Minimum Size.

Windows Paging causes extra CPU and Disk Overhead and should be reduced whenever possible.

Disk Queue (Should be ZERO for 99% of Users)

Graphics Intensity will be noted has high when over 100 for more than 1/3 of users.

-This is not a concern but must be examined to see if graphics off load processor would help.

https://blogs.msdn.microsoft.com/dsui_team/2013/04/23/debugging-a-gdi-resource-leak/

Applications Non Responsive – (1 per Day/Per Machine)

3. PCoIP: (Remote Display Protocol)

Image Quality:

This is a good "Base" metric to gage/monitor the user connection quality. PCoIP will lower the image quality if it has packet loss, high latency or a low bandwidth connection to the end user.

- VMware View Version 5 Default Image Quality is 90%
- VMware View Version 6 Default Image Quality is 80%

Adjusting image quality higher than the default is only needed for clients that may be viewing broken bones and need to see very small fractures.

Image quality directly affects the number of frames per second sent from the VM to the end client. This can dramatically affect the ESX Host CPU and network bandwidth required per user.

Session Latency:

- General Max Observations:
  - NY to California – 30-50 Milliseconds
  - USA to India – 150-200 Milliseconds
  - Inter Office Same City – 10 Milliseconds
  - Inter Office Same Building – 5 Milliseconds
- HUGE Latency #s in Stratusphere shows users dropping on and off the network. (Huge Denotes 800+ ms)

Protocol: (Good and Bad… This is Just info for you)

PCoIP is UDP based.

UDP Packets have a lower priority than TCP Packets on most networks.

UDP is dynamic and bursty by nature of the protocol.

UDP is faster than TCP because there is no error-checking for packets

UDP is lightweight. There is no ordering of messages, no tracking connections, etc.

UPD can do error checking if turned on but there is no recovery. Packets must be resent and with no ordering it is up to PCoIP to request large blocks for retransition.

Packet Loss:

Packet Loss with PCoIP is bad and can cause users with poor experiences. Mouse Drag, Artifacts on the screen.

General Recommendations:

1. QOS (Quality of Services) should be implemented on all routers.
   a. PCoIP should be right under VOIP (Voice Over IP) if in use.
2. Lower the Maximum Image Quality using the PCoIP GPO settings that best fit your use case.
3. USB and Audio Channels.
   a. Disable and lower the priority of these channels as it suites your business requirements. Disabling USB or lowering Audio quality can dramatically lower VM/Host CPU and Network requirements.
4. There are many options for PCoIP Tuning. Play with them all, consult the best practice guides from VMware and tune/monitor your users for best user experience in your environment.