Wide Open World – Tech Tip
How performance craters when defective power supplies, power throttling and BIOS settings intersect!

Copyright © 2017 SAS Institute Inc. All Rights Reserved.

Version: 201701131500gcmA - * RFC / Final *****
By Gregory Mobley, gregory.mobley@sas.com

BLUF: Lenovo/IBM issued an Engineering Change Announcement, ECA-003, to replace defective power supply units (PSU). Server models in ECA-003 should be inspected. If defective PSUs are installed, then follow the instructions to replace them. Be prepared to provide the make, model, serial and IMM logs with screen captures of the PSU serial numbers to identify the defective PSUs.

ECA quantity notice: If more than forty (40) PSUs need replacing, customers should open a CRIT/CMT with their sales representative and provide the number of PSUs to be replaced.

ECA-003 --> https://support.lenovo.com/us/en/documents/ht116322

ECA-003 is for Lenovo / IBM systems. Other server manufacturers may have defective power supplies and/or throttling options similar to those discussed below. Servers not listed may still be power-throttling workloads if they are using incorrect IMM Power Management settings!

Required actions:

1. Inspect, verify and replace all defective PSUs, even when the PSU appears to operating normally.
2. ECA-003 recommends a full system power down because the defective PSUs may not maintain the workload when hot-swapped.
3. Flash/update all FW/BIOS to current versions for system, IMM, PCIe cards, etc. Then remove all power for one minute and perform a full cold restart.
4. Verify system “Power Policies” are set appropriately. Review the screen captures and sections below for:
    1. **Power Module Redundancy without throttling**.
    2. **No Power Limiting**
5. Restart the system after completing the above steps.
6. Return all defective PSUs within 30 days or the customer will be billed for each PSU.

Why? These defective PSUs may **not** present the orange **"!"** server warning or insert entries into the IMM logs. In fact, the PSUs may appear to be operating normally when internally the PSUs may have throttled the system workloads. Usually, throttled workloads will be identified by the IMM log showing the machine being “powered throttled.” When a defective PSU is present, the pCPUs may be cycling between normal and degraded. Also, the

system workloads may show increasing latencies when examined using Foglight’s or VMware’s historical performance services.

Indications? For our x3750 M4 systems, at least one of the PSU had failed. We were **not** alerted by the IMM nor were ANY IMM log entries generated until the IMM was restarted. There were also no warnings on VMware's Hardware alerting tab. Overall, the system appeared "100% Normal".

It was from experience that we restarted the IMM. Once the IMM restarted, we immediately began receiving repeating log entries that the processors were cycling between normal and degraded states. The IMM also notified us that the workload had been power throttled.

Multiple Impacts? We experienced ***severe*** performance degradations once the defective PSUs had "failed." Foglight's performance / GUI transitioned from snappy and responsive to very sluggish. We opened service requests with Foglight support but no software or severe tuning problems were identified.

Upon further analysis, we saw steep latency increases reported by Foglight. We then confirmed similar increases using VMware. Since these latency measurements began increasing remarkably at the same time, that coincidence became a major red-flag for us to continue digging.

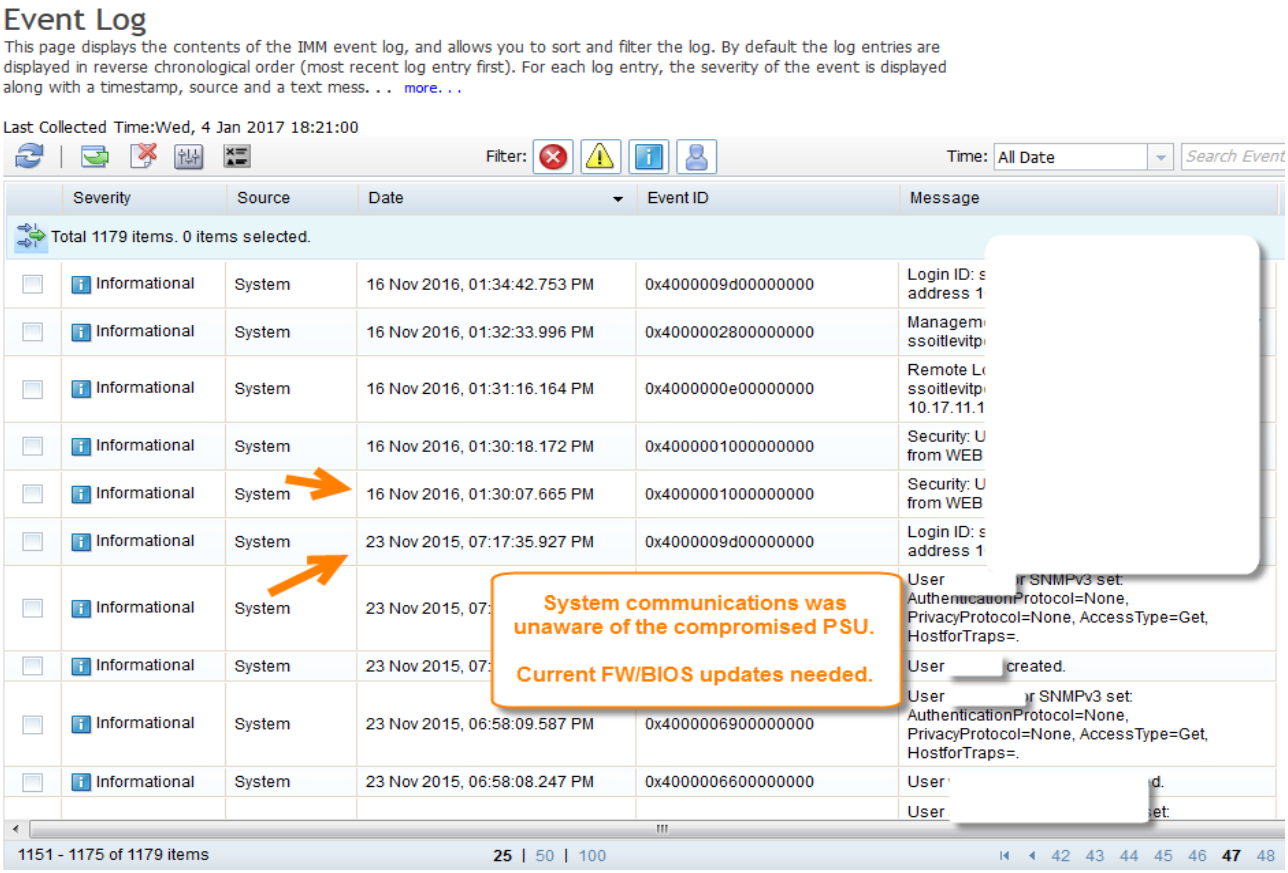
Our customer impacts were **dramatic**. We saw:

1. Foglight GUI responses increase from sub-second to multi-second.
2. Our Foglight Federator accumulate and amplify delays from **each** kid.
3. Increases in Foglight database latencies
    1. Before → sustained 140-400ms +peaks
    2. After → 200us – 3ms
4. Dramatic physical host increases for VM ready, latencies, and co-stop.
5. Higher than expected VMware Core Utilization
    1. Before → Peaking 60-82%
    2. After → Peaking 15-30%
6. Increases in MYSQL database backups (240+%) and multiple backup errors
    1. Before → ~ 12 hours (19:30 – 07:30)
    2. After → ~ 4 hours, 15 mins (19:30 – 23:45)

Note: The observed impacts from power throttling may be similar to the results seen when the BIOS is allowed to default into "Balanced Power" mode.

Supporting materials

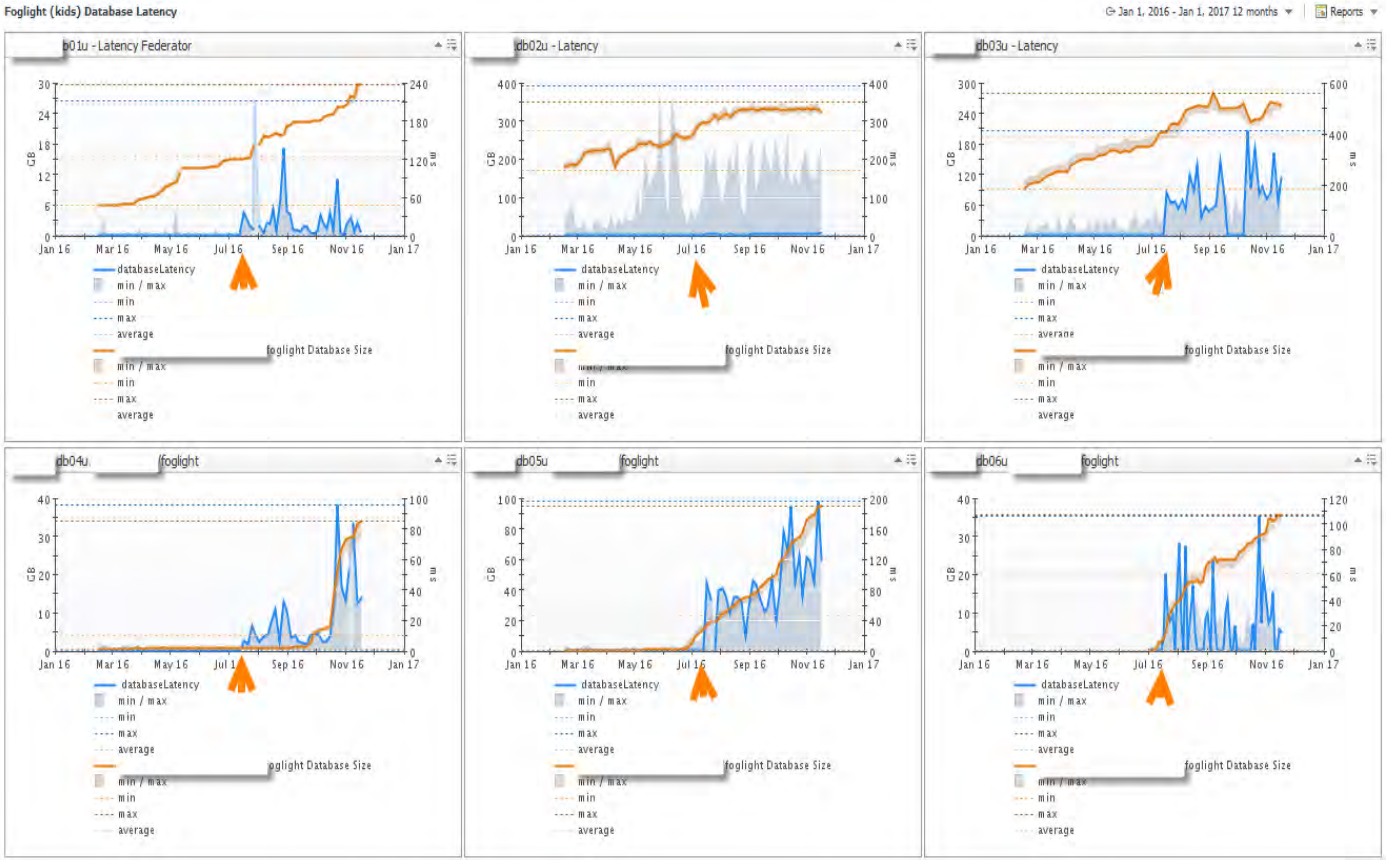
IMM log showing no events prior to restarting the IMM.



IMM log indicating “Power Throttling” and each pCPU bouncing from normal to degraded states. (Note: these entries were presented only AFTER the IMM was restarted.)

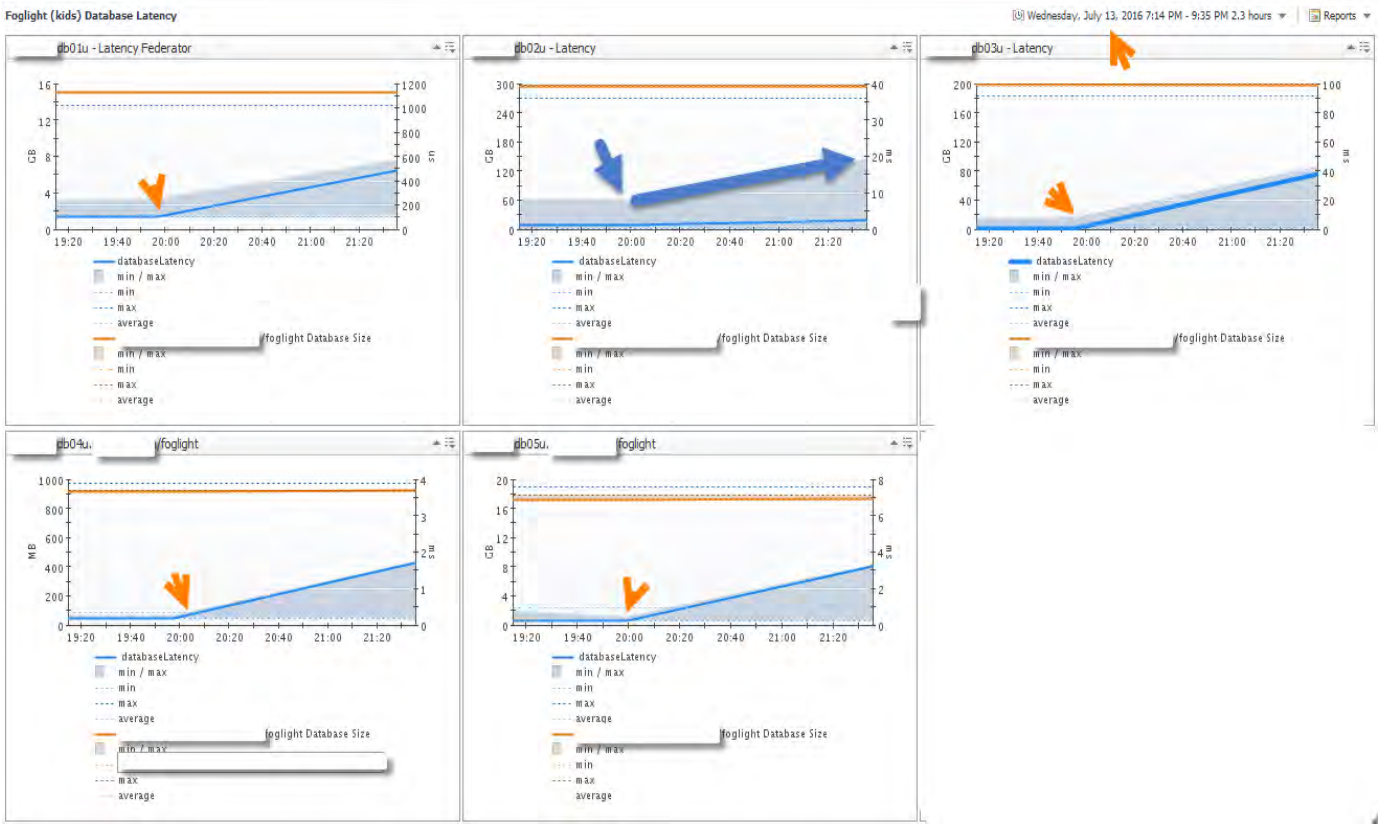
Informational	System	19 Nov 2016, 04:52:15.233 PM	0x4000001000000000	failures from WEB client at IP address
Informational	System	19 Nov 2016, 04:50:03.099 PM	0x4000007f00000000	Power cap throttling occurred.
Informational	Processors	19 Nov 2016, 04:37:54.346 PM	0x816f0a070304ffff	The Processor processor 4 is no longer operating in a Degraded State.
Informational	Processors	19 Nov 2016, 04:37:54.339 PM	0x816f0a070303ffff	The Processor processor 3 is no longer operating in a Degraded State.
Informational	Processors	19 Nov 2016, 04:37:54.308 PM	0x816f0a070302ffff	The Processor processor 2 is no longer operating in a Degraded State.
Informational	Processors	19 Nov 2016, 04:37:54.305 PM	0x816f0a070301ffff	The Processor processor 1 is no longer operating in a Degraded State.
Warning	Processors	19 Nov 2016, 04:37:44.291 PM	0x806f0a070304ffff	processor 4 is operating in a Degraded State.
Warning	Processors	19 Nov 2016, 04:37:44.289 PM	0x806f0a070303ffff	processor 3 is operating in a Degraded State.
Warning	Processors	19 Nov 2016, 04:37:44.097 PM	0x806f0a070302ffff	processor 2 is operating in a Degraded State.
Warning	Processors	19 Nov 2016, 04:37:44.094 PM	0x806f0a070301ffff	processor 1 is operating in a Degraded State.

Foglight showing steep increases in multiple VM database latencies



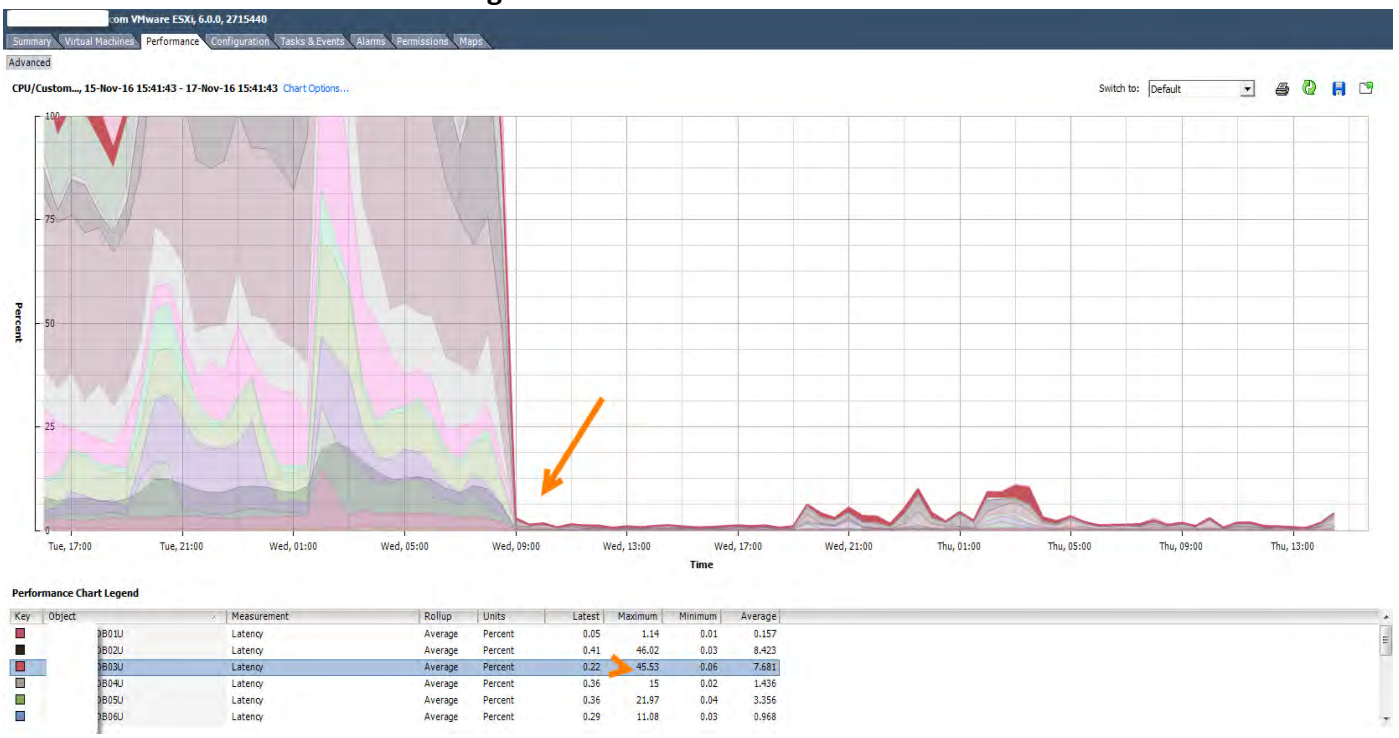
Foglight showing a drill-in on VM DB03 database latency 20:00 13 Jun 2016

Foglight showing multiple VM database latencies for ~ 20:00 13 Jul 2016

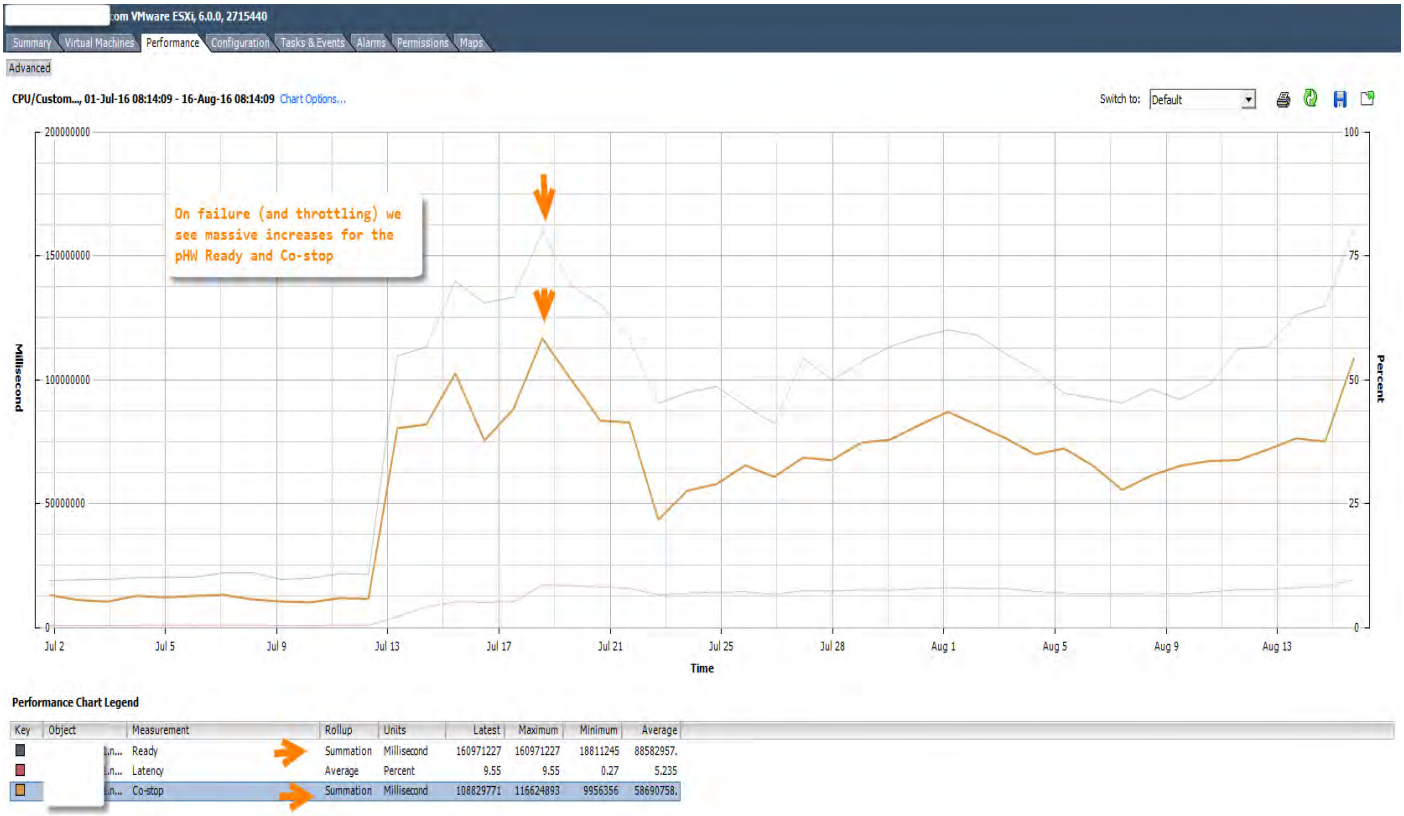


VMware showing VM DB03 database latency + co-stop – increasing dramatically

VMware showing multiple VM database latencies while throttled vs. un-throttled with HOP BIOS settings



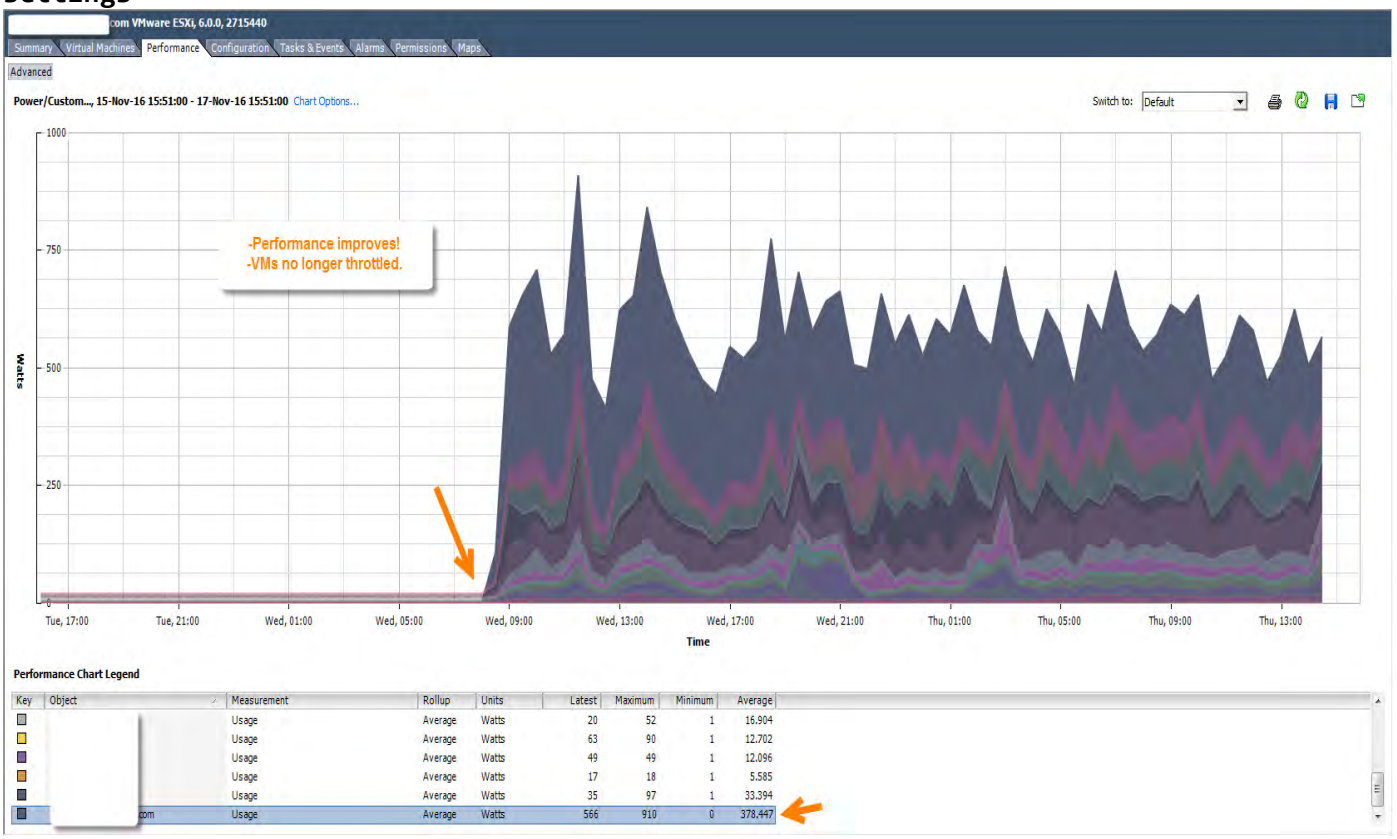
VMware showing Ready, Latency and Co-stop increases during throttling



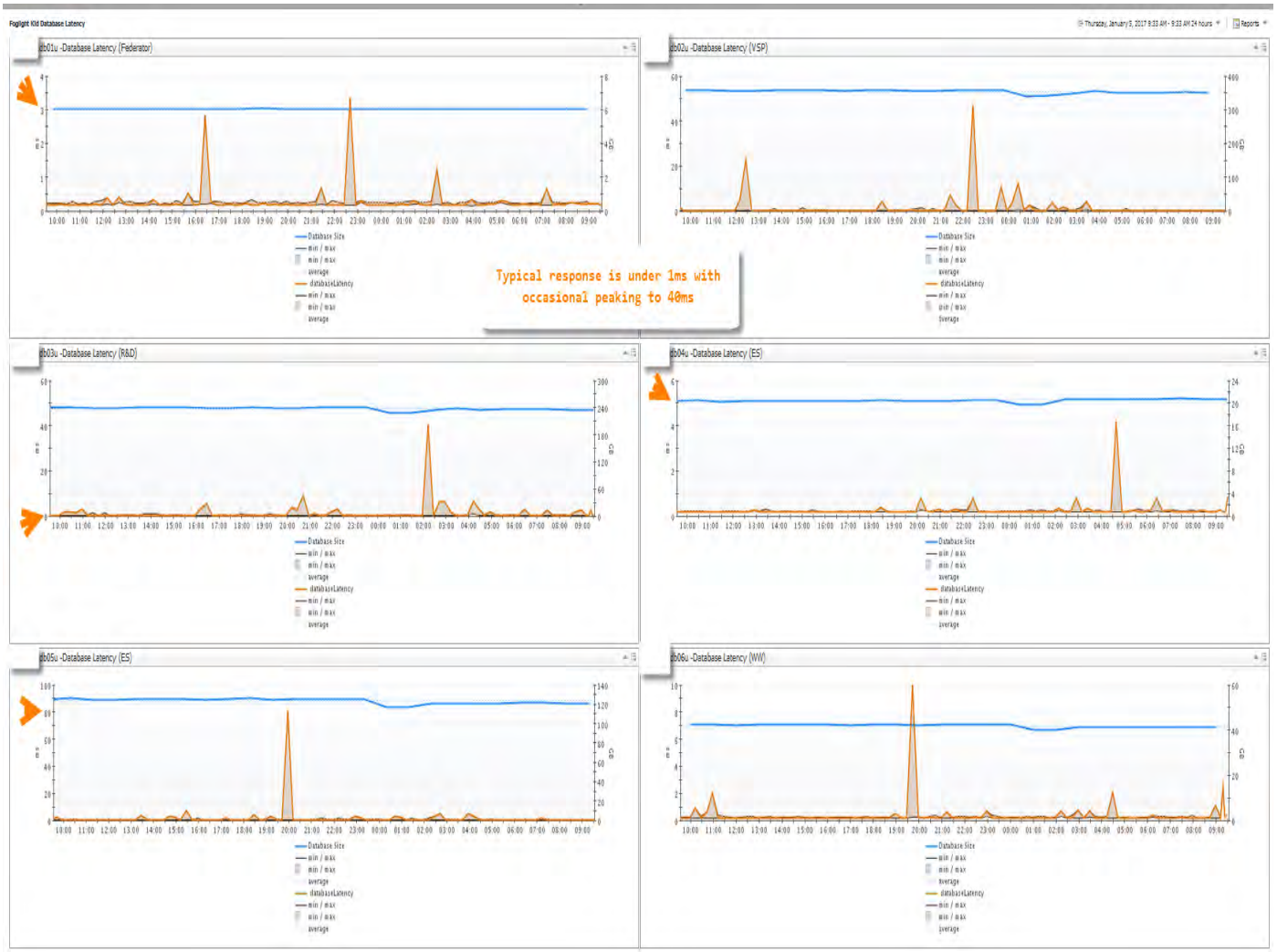
VMware showing pHOST "Core Utilization" with 82% peaks during throttling vs. 30% peaks when not throttled and using HOP BIOS settings

The GUI sluggishness, as reported by customers, is to be expected on over-allocated pHosts. Given the left-most graphing below, engineers may conclude the pHW pCPUs are being challenged by the active workloads and that ordering new pHW is imperative to restore performance. However, when viewing the right-most graphing, those same engineers would conclude the pHOST is modestly loaded.

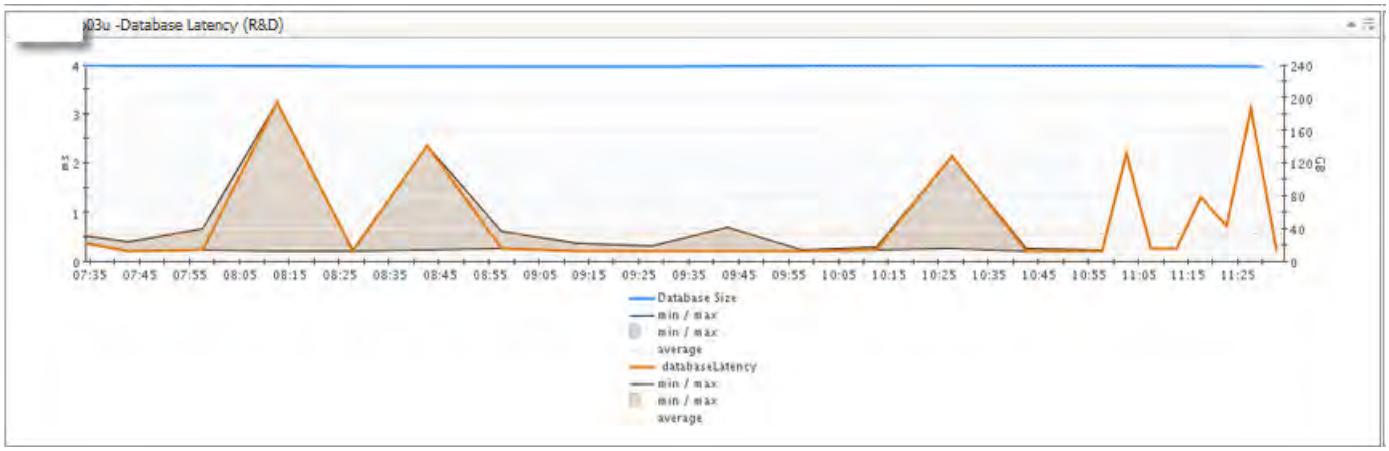
VMware showing VMs power-throttled vs. VMs not power-throttled with HOP BIOS settings



Foglight showing multiple VM database latencies with no throttling and HOP BIOS settings fully engaged.

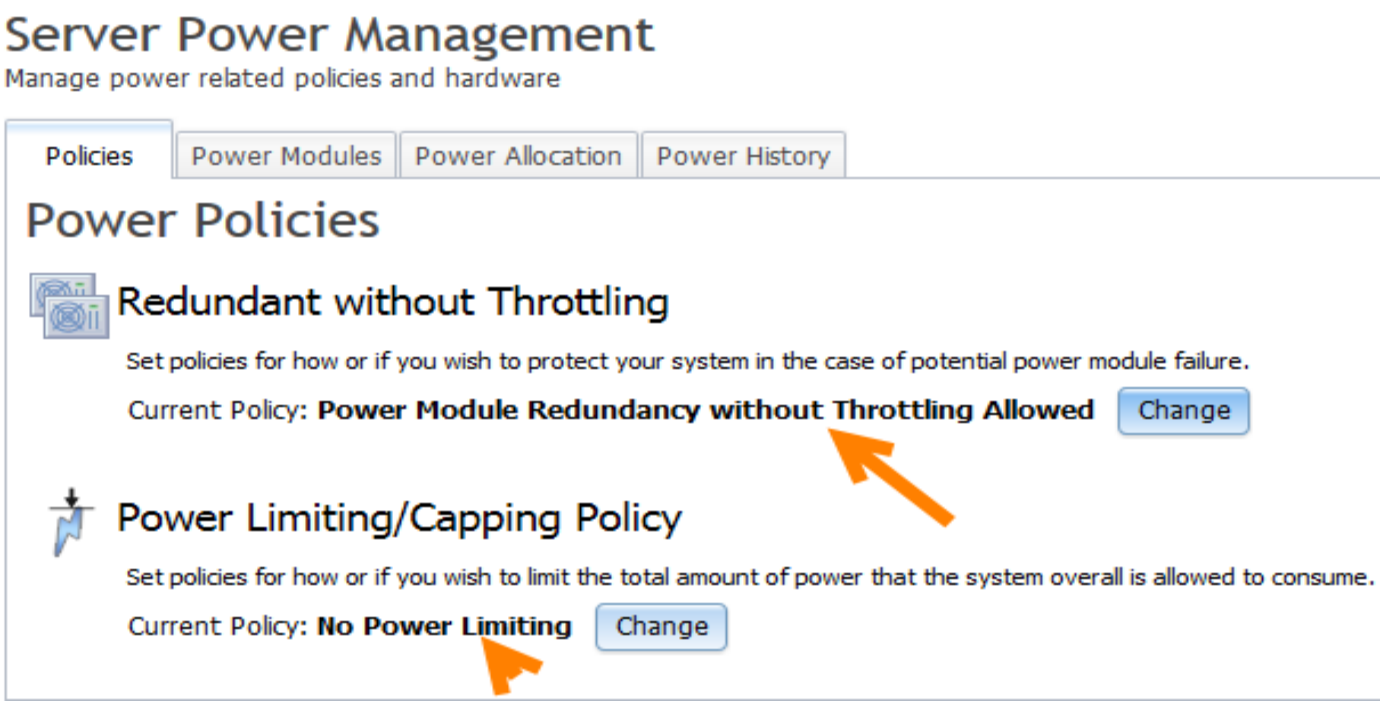


Today our Foglight database latencies range from 400us - 3ms (4hr view)

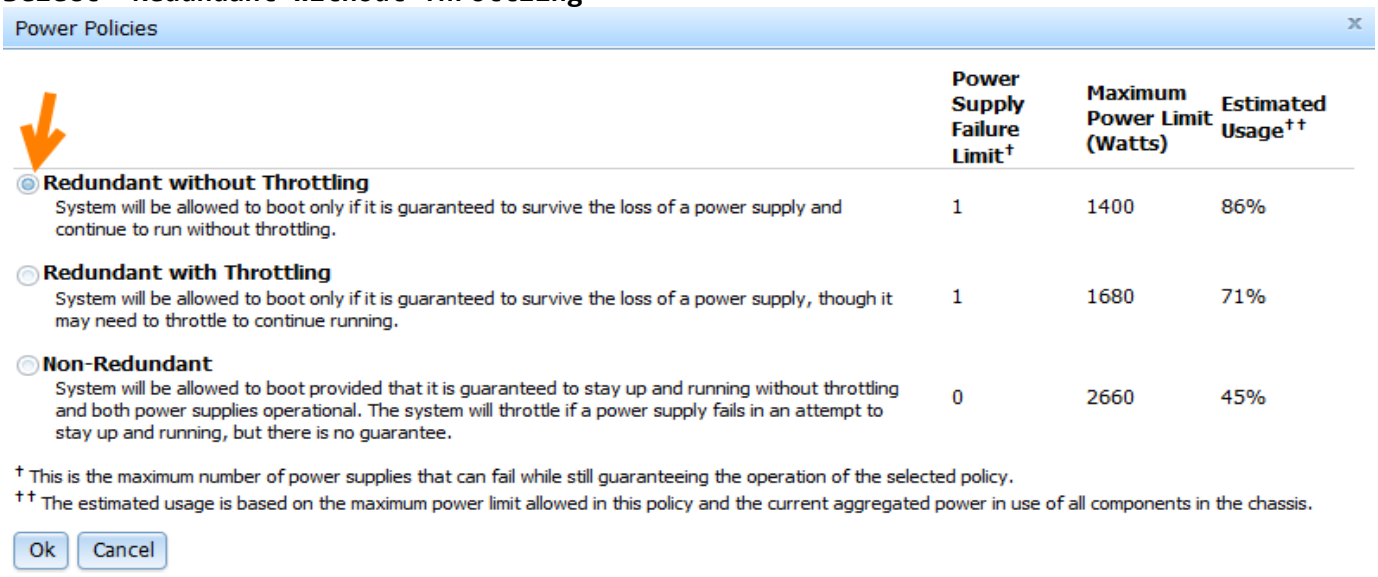


Lenovo Power Policies

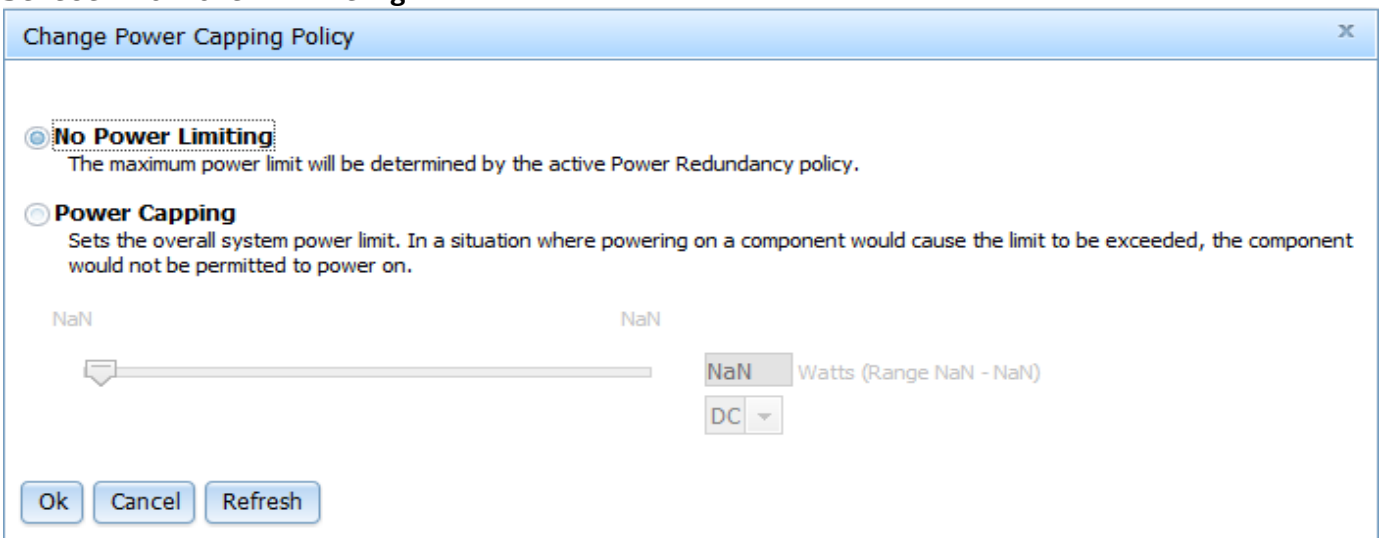
"Power Policies" are features which may be available from many hardware manufacturers. Often, the Power Policy defaults to allow power throttling. These policies should be changed in the IMM Power Management section. The policies and defaults may not update when redundant power supplies are installed. Do not assume, verify your system settings!



Select "Redundant without Throttling"



Select “No Power Limiting”



<EOF>