Private Cloud Management

Transforming VMware SDDC into Your IaaS



PART 2 OF 4

October 2026

Iwan "e1" Rahabok
VMware Cloud Foundation | Broadcom
e1@broadcom.com

5 Edition

Back of cover page. Delete if you do not plan to print.
Thank you Mai Mai for the design. Daddy proud of you!

PART 2

Consumption

Table of Contents

How do you implement the concepts in Part 1 so they are consumable by various roles?

That's the goal of PART 2.

Chapters:

Proactive Operations	This implements the concept covered in Part 1 Chapter 1
The Art of Dashboard	Best practices that balance usability and functionality.
Inventory	Dashboards for inventory management
Performance	Dashboards for inventory management
Capacity	As above, but for capacity management
Configuration	Dashboards for configuration management
Availability	Dashboards for availability management
Role-based Content	Dashboards for specific roles in the management of IT operations.
Log Analysis	
Reports	Report best practices

Proactive Operations

Part 2 Chapter 1

Proactive Operations is a complex concept. How does it work in real life?

System Engagement

As a product, how does VCF Operations "communicates" to its users?



In terms of urgency, live screen is the most pressing, followed by alert. Report is suitable for long term review. Use dashboard as much as possible as it offers the richest functionality and interaction. That's why the diagram above shows it with bigger arrow than the other 3 ways.

If you apply the above thinking, you will vrealize several benefits:

Less reports	Encourage users to login as they get richer experience. Compliment with login-less dashboards, made available at user-friendly intranet website
Higher engagement	Higher engagement with IT Leadership Team. Take time to educate and demonstrate how dashboard meets their needs better. Use a login-less dashboard, displayed prominently near their office to encourage involvement.
Better telemetry	It's not possible to track if users actually read your reports. With dashboard, you can even track how they interact with VCF Operations, using 3rd party website tracker software.
Flexible threshold	You can tailor the threshold (green – yellow – orange – red) accordingly, as each of the 4 way of communication have different level of urgency.

Synergy

The following table details how the 3 ways of engaging are complementary. I've excluded screen as it has narrow usage.

	Alerts	Dashboard	Report
Nature	Reactive	Proactive	Passive
Good For	Exception	Exception. Analysis	Exception.
Depth	Detail	Detail and Summary	Summary
Use Case	Troubleshooting	Troubleshooting. Monitoring. Optimizing	Export to spreadsheet or PowerPoint
Time	Minutes	Daily – Monthly.	Weekly onwards

	Focus is “now”	Focus is days	Focus is weeks
Roles	L1	L1 onwards	IT Leadership
		Architect Team	Finance Team
		Capacity Team (Dashboard + Report work together)	
		Audit Team (Dashboard + Report work together)	
		Tenant (Dashboard + Report work together)	

Timeline

In VCF Operations, the following are implemented as dashboard:

- NOC screen
- Daily health check
- Longer cadenced

However, they have diferent timeline. As a result, they have different purpose.

	Live NOC Screens	Daily Preventive Check	Longer Cadence
Purpose	Real time visibility into the overall environment, so live fire is attended right away.	Prevent alerts of the day. Tomorrow is another day as it will be checked tomorrow morning.	Larger optimization and issue avoidance. Deep and broad analyzis of the whole environment.
	Urgent (hopefully not important)	Important (hopefully not urgent)	Important (definitely not urgent)
Timeframe	Focus: last 5 minutes	Focus: next 12 hours	Focus: Weekly – Quarterly
	Data: 1 hour	Data: 1 week	Data: 1 year
Usage and Interaction	Live. Always on and projected on the big screen.	2x a day. Once at start of day, the other at end of day.	Less standard. Multiple usage throughout the week/month, depending on the persona and need.
	Time spent to analyze: 10 seconds	Screentime: 10 minutes	
User	L1 (first liner)	L2 (expert)	Auditor, Capacity Planner, Performance Specialist, Cost Specialist
Primary Content	Availability. Performance. Security.	Same area of concern, but the Daily Check is broader and deeper as it has much longer horizon.	Capacity (utilization, reclamation, optimization). Sustainability.

How are their differences translating in the product UI?

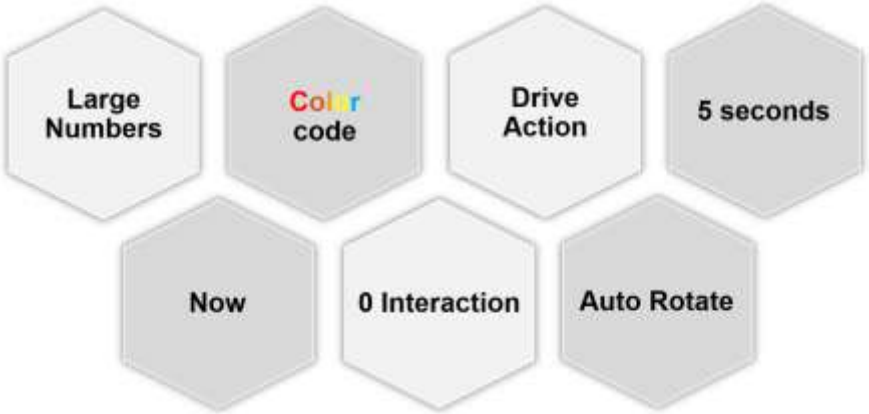
Live NOC Screens	Daily Preventive Check	Longer Cadence
Dashboard size matches the projector/TV. No scrolling.	Dashboard size is 1 -2 screen deep.	Dashboard size is >2 screen deep.
Table not suitable. Hard to conclude by seeing many numbers. Easier to see trends + present number	Scoreboard is suitable, as L2 will click to do interactive filtering and analysis	Table for export to spreadsheet for further analysis
No interaction as there is no keyboard/mouse. See from a distance. It auto-refreshes and auto-rotate	Standard interaction process	Rich interaction. Flexible flow, depending on the need at that time

Live Screen

The real-time nature of life inside the Network Operations Center room calls for a different type of visibility.

Collaborate with NOC Room team, as they are ultimately the consumer. What information do they need? What do you need from them? What's the Exit Criteria and Entry Criteria?

The following diagram summarizes the key requirements and constraints. As a result, the dashboard needs to be designed accordingly.



0 interaction	Think of TV, not monitor. NOC screen does not have keyboard and mouse for user to click and type. So the information is auto-refresh and auto-rotate. If your TV has a touch screen, that's a bonus. This should not replace the basic need of being useful from afar.
	Toolbars & buttons on screen should not be used. They add confusion, plus there is no way to use them from afar.
	Most widgets are not suitable as a result.

Large Number	As the information is seen from a distance, the font used is much larger than the standard font used in laptop or desktop monitor. This also means the screen real estate is relatively less. When designing, use a 14" screen instead of 24" screen.
Now	It is placed strategically because it displays a *time sensitive* information. As such, many operational use cases are actually not suitable for these. For examples, Capacity, Cost, and Sustainability management are not highly time sensitive. They tend to stay static for hours, rendering them unsuitable for live screen. It's showing the latest 5-minute data. The past is less relevant. Ideally, project the next 1 hour. The information auto-refresh frequently. Think 1 minute, not 10 minutes. If the display is static for 300 seconds, users will not be drawn into it. The information presented is more urgent in nature than alerts (otherwise you simply use alerts!) and is used to complement alerts.
Color	Heavy use of color. Color is easier to digest than text, as you don't even need to read. Lots of text can confuse viewers. Text can be hard to read from afar. Use key colors (green, yellow, amber, and red) classify the severity of the issue. By default, all should be green. If you display something that is red most of the time, after a while the viewer will ignore it. This defeats the very purpose of displaying on the big screen. Use alerts for something less urgent. Ideally, all the numbers are in %, with 0 being bad and 100 being perfect.
5 second	KISS (Keep It Simple Show). Remember the 5-second test. The screen should be easy to interpret, user friendly and do not require an explanation.

The above principle applies to the part of the dashboard that is projected on the big screen. You can dual-purpose the same dashboard, to cater for the operators in the NOC room. These help desk administrators should have the same dashboard on their desktop or laptop. They can then use keyboard and mouse to interact with the dashboard, enabling them to drill down and find out more information.

I'm aware that other persona, especially IT leaders, may drop by to NOC room to check things out. This does not mean we design a NOC dashboard for them. For example, say CIO drops by every single day for 10 minutes. In this case, if it makes life easier, prepare a dashboard to facilitate the "check". Only bring up during the visit. There is no need to have the dashboard on the wall for the remaining 23:50 hours.

Action

Action, not information. NOC show data for your action, not for your information.

Focuses on immediate remediation. Remediation action has to be immediate, as soon as possible if it's red.

If something can't be fixed within the same day, why show it live? Remediation that takes >1 day should not be shown, as the dashboard will be red for hours. Use alerts for longer remediation window.

Problems that don't require immediate attention should be avoided, as they are distraction. Your NOC Screen is not your To Do List.

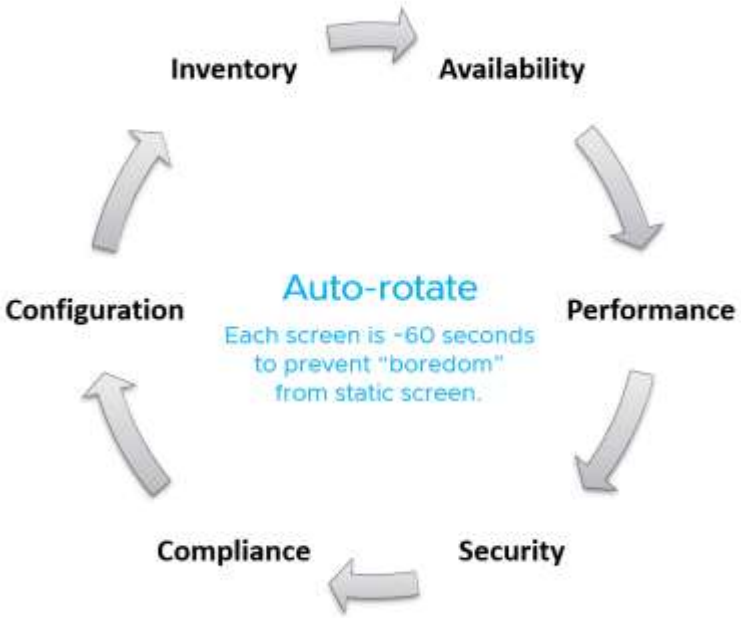
Examples of suitable actions: stop provisioning of new VM, take action on VMs that abuse the shared infrastructure.

Examples of not suitable actions: Increase supply of infrastructure, such as adding hardware.

Auto Rotate

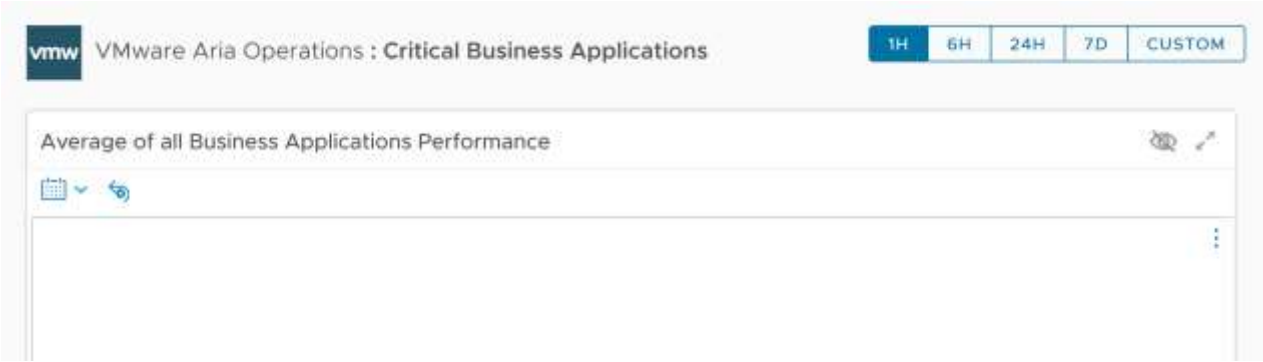
It's not enough to auto-refresh due to the sheer amount of information. Auto-rotate enables you to show more content without scrolling the dashboard, and prevent human's boredom from seeing the same screen for hours.

Configure the dashboards to auto-rotate in a logical flow. Explain the flow to the users as it can get overwhelming.



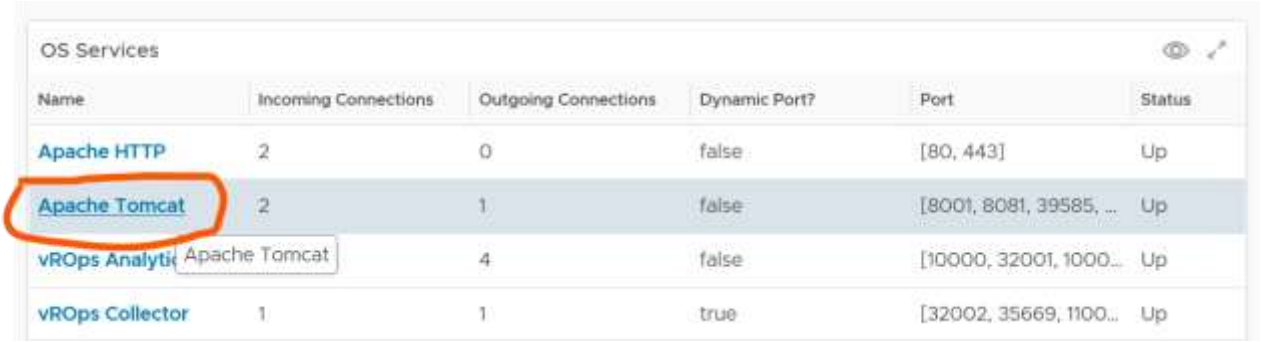
Implementation wise, you need to disable Guest OS screen saver, Aria Operations user time out and Guest OS going to sleep.

Full Screen



For improved focus, hide the menu and navigation by using a login-less dashboard. This will also maximise the screen real estate. Complement this by enabling the browser full screen, so viewers do not even see it's a web page. The preceeding screenshot is the result.

Take note that viewers cannot leave the page and navigate out into other pages in vR Ops. The object link below does not work. Yes I agree the product should have removed the URL.



Daily Health Check

This daily, proactive health check forms the 2nd leg of your proactive management.

Start each day with a proactive daily health check. Initially, this is done by both Level 2 and Level 1 team. As Level 1 gain confidence, they may even do it at the start of their shift. As there are 2-3 shifts in 24/7 operations, the dashboard gets used multiple times a day. The dashboard aims to minimize alerts by analysing underlying issues, hence it requires expertise of how your private cloud performs. It also requires someone who knows the environment well, especially what is happening on that day that can impact performance and security.

Morning Routine

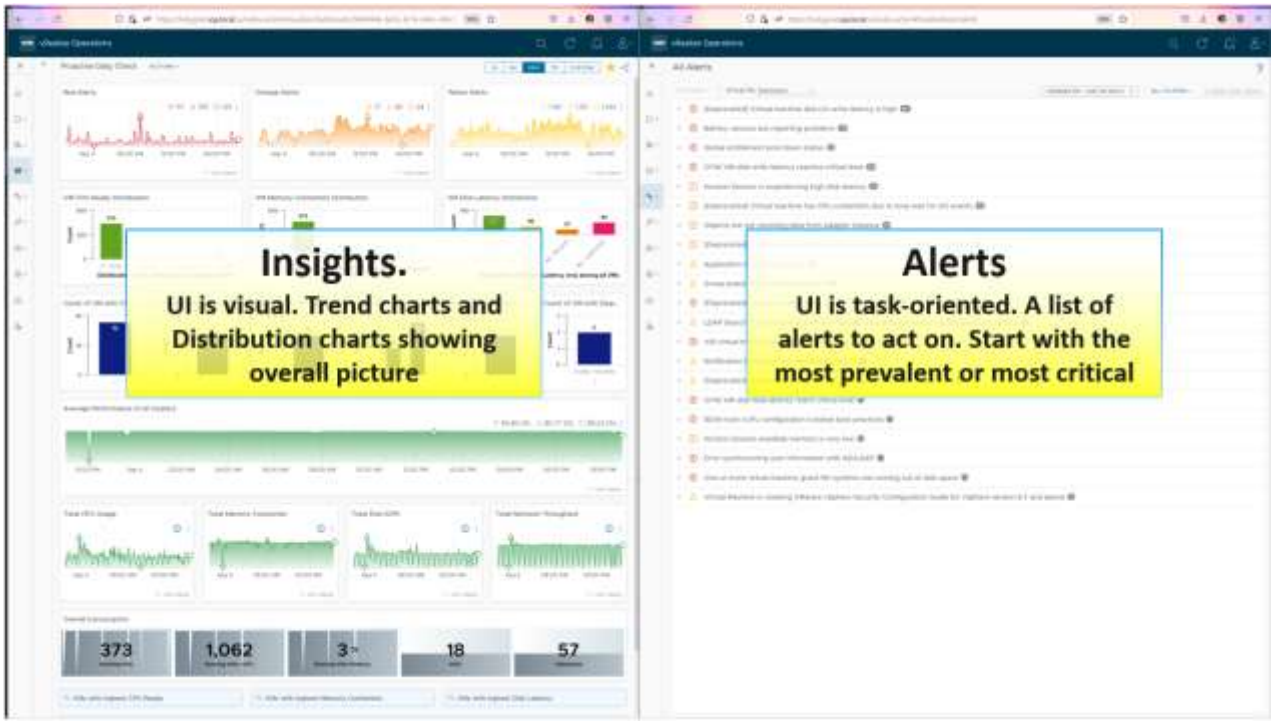
As the person responsible for the health of the overall environment, what do you check every morning, with your first cup of coffee?

If the answer is a list of alerts or complaints, how do you know the big picture?

By and large, your environment is good, otherwise you're not reading this specific chapter 😊. If it is not good, you have some knowledge on the problems as you need to explain it to your management. On this assumption, this dashboard focuses on things that need attention.

It complements alerts by showing insights (situation that has not triggered an alert) and show the overall picture. The overall picture is also useful as before you troubleshoot something specific you should ensure the issue is not with the overall environment.

As insights complements alert, open the 2 screens side by side. Simply open 2 browser windows, as shown in the following screenshot.



There are 2 dashboards, one from VCF Operations and one from Log Insight.

Overall Daily Check

As part of our daily check, you're after changes and unusual events. There are 4 types of changes:

- Configuration changes. Focus on undesirable changes as they have negative impact to the overall environment.
- Consumption changes. Both unexpected increase and drop are not desirable.
- Supply changes. Focus is on unexpected drop, which could be due to a variety of reasons (e.g. maintenance taking longer than expected).
- Dynamics: VM state change, VM location change, VM inventory change

The first thing to check is the trend of alerts. Since you're checking this daily or even more, the number gives you a good indicator of what's ahead of you.

Alarm can give insight when reviewed as a set and over the last 24 hours.

It cuts across objects, so we must be able to see ripple and correlation.

Check the patterns of alarms in the last 24 hours. Is that within expectation?



Check both the absolute amount and the pattern. Compare it with your expectation on that day. Is today a special date, where high demand is expected? Any relevant changes approved during last night change window that will impact supply and demand?

Is the trend unusual? Form an expectation of what's normal for your environment as it develops pattern. Compare it with what's happening (e.g. DC upgrade, public holiday).

List of Clusters



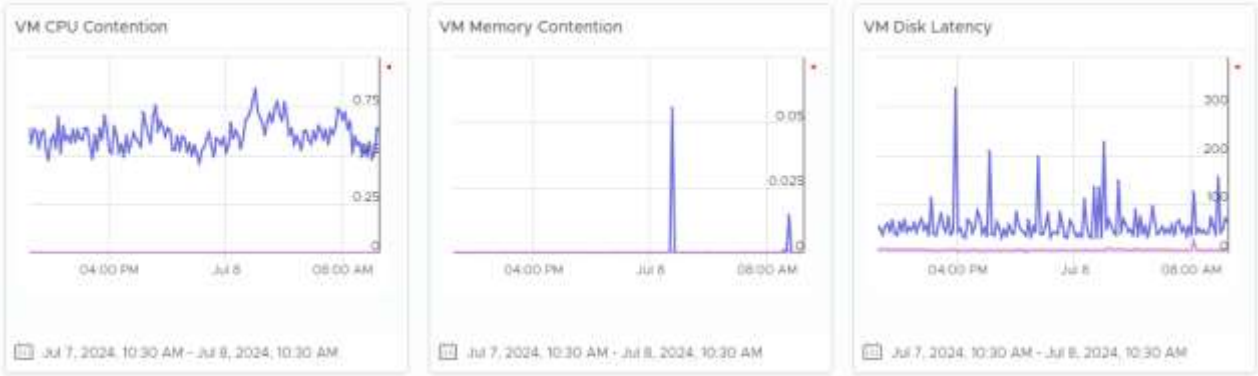
vSphere Cluster is a logical group where "similar" VMs live together.

The table is sorted by the overall performance, where the least performing cluster is shown at the top.

The columns sport leading indicators, as it uses worst() not average. It checks 9 key performance metrics. Adjust the threshold to your comfort level. 4 utilization metrics are shown to catch abnormal peak.

Performance Trends

What's the overall IaaS performance among all the VMs? Is it a CPU, memory, disk problem? Expect the numbers to be within your expectations of that day.



What's the overall IaaS performance among all the VMs? Is it a CPU, memory, disk problem? To do that, dive into a specific cluster. This section automatically shows its performance in the last 24 hours.

The 3 table covers CPU, memory and disk. Expect CPU contention to be higher and more volatile than memory contention. Add network drop packets if the issue exists in your environment.

Each table shows 2 lines (worst and average). Are they in line with your expectation, both the absolute amount and their pattern over time? Make sure the average is well below your comfort level.

If the worse is far higher, you have an isolated incident. If they are close and the average is not low, you have a widespread issue.

Affected VMs

The distribution charts place all the VMs in 4 levels of performance (green, yellow, orange, red). Attend to the red category first.

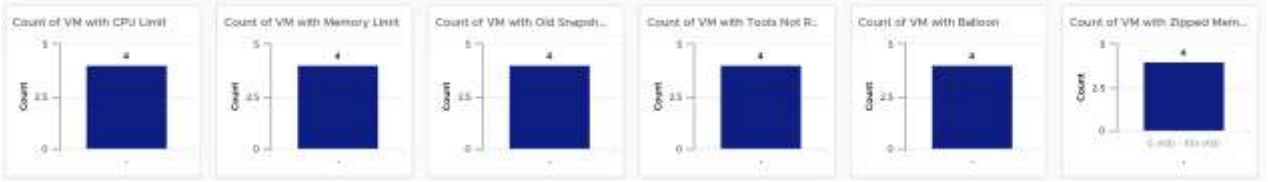
They are based on 20-second peak metric, not 5-minute average. This gives you a leading indicator. Each value is based on the highest in the last 24 hours, to enable daily comparison.

Add network drop packets if the issue exists in your environment.



Configuration

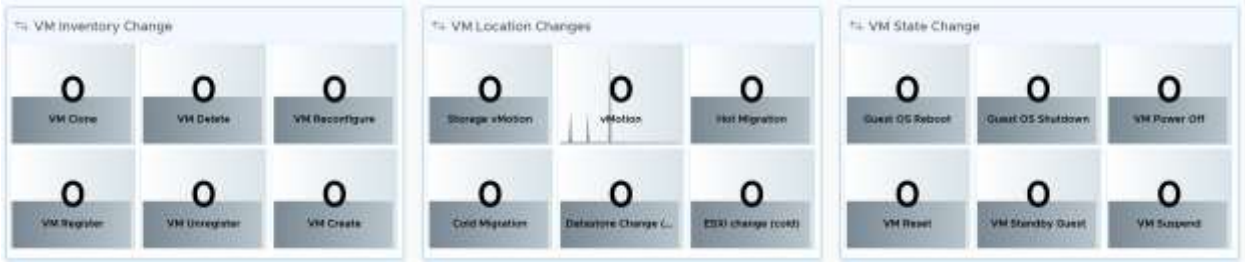
Next is configuration. What are typical misconfiguration that cause performance problems?



Ideally the above returns 0 VMs.

VM Changes

What are the important changes on VMs? Is the environment more volatile than expected? Ensure they match the expectation of the day.



High volatility increases the chances of alerts.

There are 20 types of VM changes that can happen to a VM. They are grouped into 3 logical sets. For each set, the changes are shown in order of urgency, where the least desired changes are shown first.

Each of the 20 changes are unique. Ensure they match the approved change requests and expected changes of that day.

- State = VMs power state change. Abnormal changes such as reset, suspend and power off are shown first.

- Location = VMs are moved to another host or datastore. This can be hot or cold migration. Hot migration is shown first as it might impact performance.
- Inventory = VMs are added or removed, hence impacting count of inventory.

Utilization

To complement the performance, we need to check utilization. Are they higher than expected? On the other hand, a big drop means the amount of work completed is less, which could indicate availability or performance issues.

We check the 4 elements of infrastructure. Since this is the overall utilization, it should be stable or cyclical overtime. Make sure both the pattern and the amount are in line with your expectation of that day.



The CPU widget uses Usage metric as it is aware of CPU clock speed. The memory widget uses allocation as memory load is much more stable, reflecting its nature as cache of disk space.

If you have enough screen real estate, add the Memory Balloon and Compresed + Swapped.

Customization Tips

If you operate a large environment, you can add a table of data centers so you can zoom into specific DC. You can even add vSphere Clusters in the table, just take note the table will be longer.

One of the table rows has to be vSphere World object. This lets you see the overall picture.

vSphere Daily Log Check

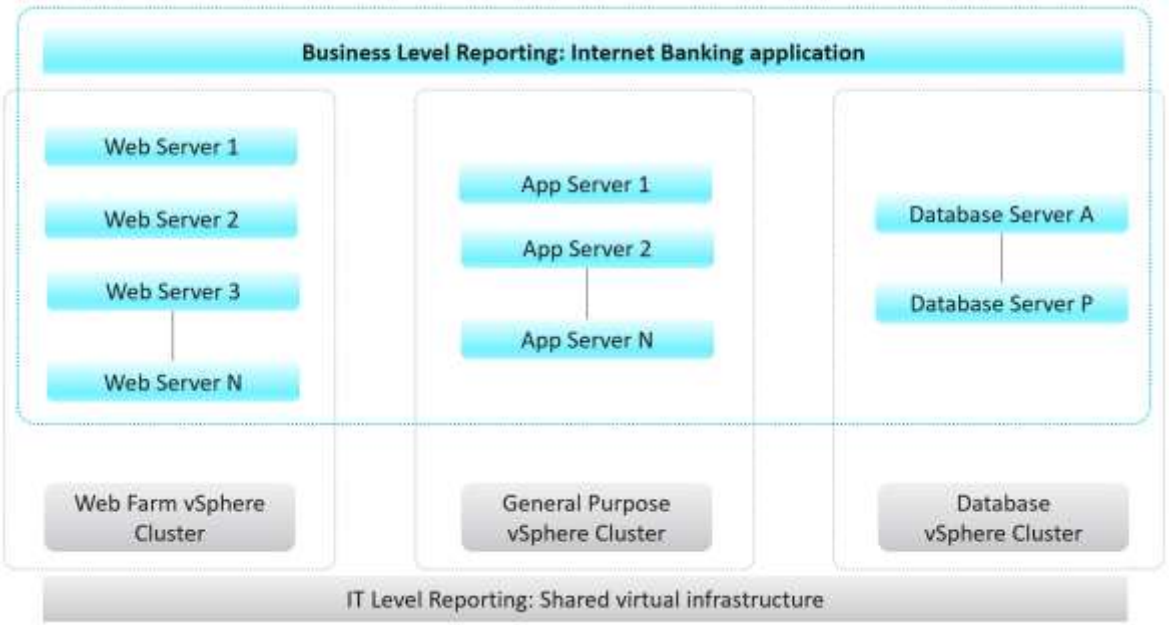
A lot of information is only available in the form of logs. Use a daily Log Insight dashboard to perform basic hygiene, such as:

- Undesirable VMs events. Are they in line with your expectation? That means the correct amount and right timing. Harmless events such as VM relocation, snapshot, powered on were filtered out from the dashboard.
- Undesirable VM changes & Snapshot. Are they in line with your expectation? Harmless changes filtered out. Are changes and snapshot taken during office hours?
- Issue. vCenter Alarms, vCenter tasks that ended with error, and ESXi potential problems.
- Security. Any malicious changes? Anyone issuing shell commands at ESXi console? Login by root?

- vMotion. This is a leading indicator of a larger performance issue. Is the bandwidth lower than expected? Is the downtime longer than expected? Some clusters did a lot more vMotion than usual?



Business Applications



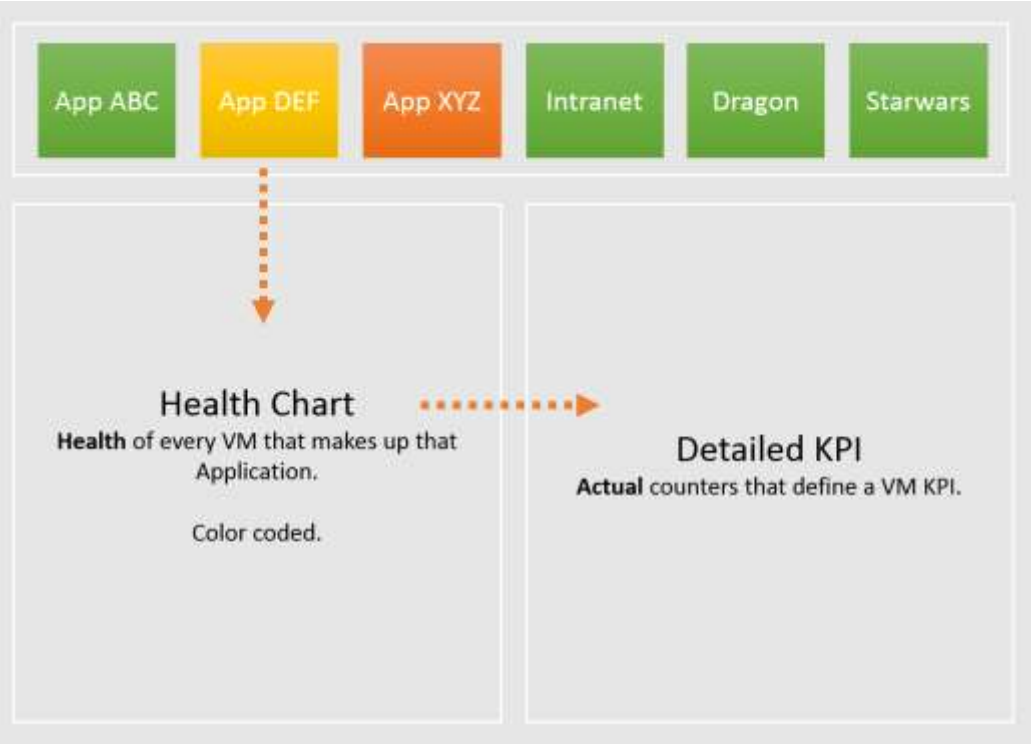
There is a need for more in-depth business-oriented reporting focused on logical business constructs such as services, applications, departments, divisions, teams, groups, or any other type of logical business structure. Quantifying performance, capacity, and cost at the logical business unit level makes more relevant information to business leaders. This is because services, applications, and business units can span multiple infrastructures from private, through hybrid, to public. Not only is business-oriented reporting more intuitive for a business to consume, it also makes the IT more transparent and aligned with actual business outcomes.

Frequently, business unit stakeholders will ask IT to provide reporting around their workload's performance, capacity, configuration, and cost. For example, a business owner may ask the following questions:

- What is the infrastructure cost for an application, service, or department?
- What can we do to reduce the infrastructure costs of our service/application?
- How much compute and storage are our application/service consuming?
- What are the poorly performing VMs in an Application Stack and why?

These questions are mostly about the underlying infrastructure, but instead of being posed at a vSphere level, they are positioned around the business applications. Traditionally, it has been very difficult for IT to answer these types of questions because typical Element Managers used by IT, represent various objects in a very rigid hierarchical fashion that does not translate well to the elastic business structure.

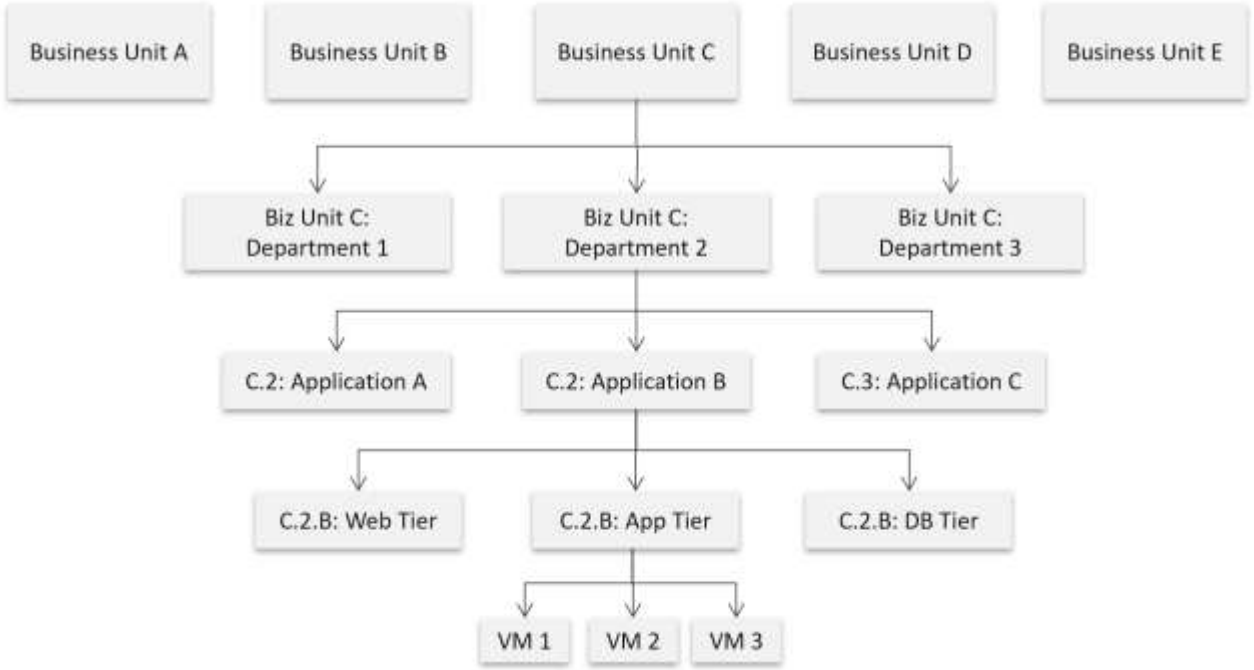
The Crown Jewel



vSphere Folder Structure

To implement the above, the first step is to design a vSphere folder structure that reflects the organisation structure. These folders and their relationships are then used by VCF Operations groups and business applications.

The following structure shows Business Unit as the top folder. Each business unit can have 1 or more departments (Business Unit C spans 3 departments in the diagram below). Each department owns multiple business applications. A business application typically consists of multiple tiers (e.g. web tier, application tier, database tier). A tier is a group of VMs performing the same function, running the same set of software and have identical hardware configuration. You expect the VMs are either scale-out (farm) or active/passive.



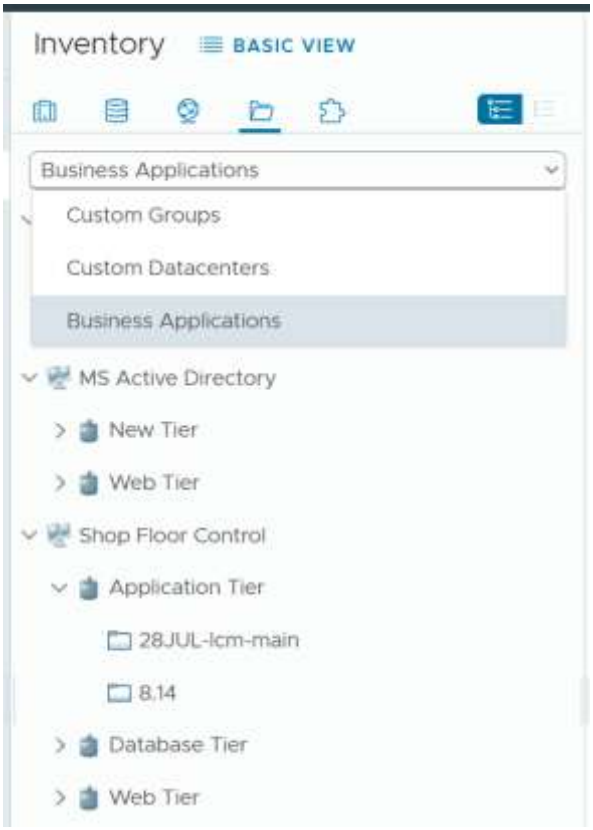
The limitation of the above is reorganisation. You will need to rename, move folders to the new parent folders, and delete folders that are no longer relevant. For example, if Business Unit B merges with Business Unit C and the combined entity has a new name, then you need to rename one of them, and delete the other. This is why I only have 2 levels of organisation in the above. Your goal is to have some level of context, not to replicate the entire organisation chart as that's the job of HR department 😊

The application name should be unique. If not, prefix with the department name.

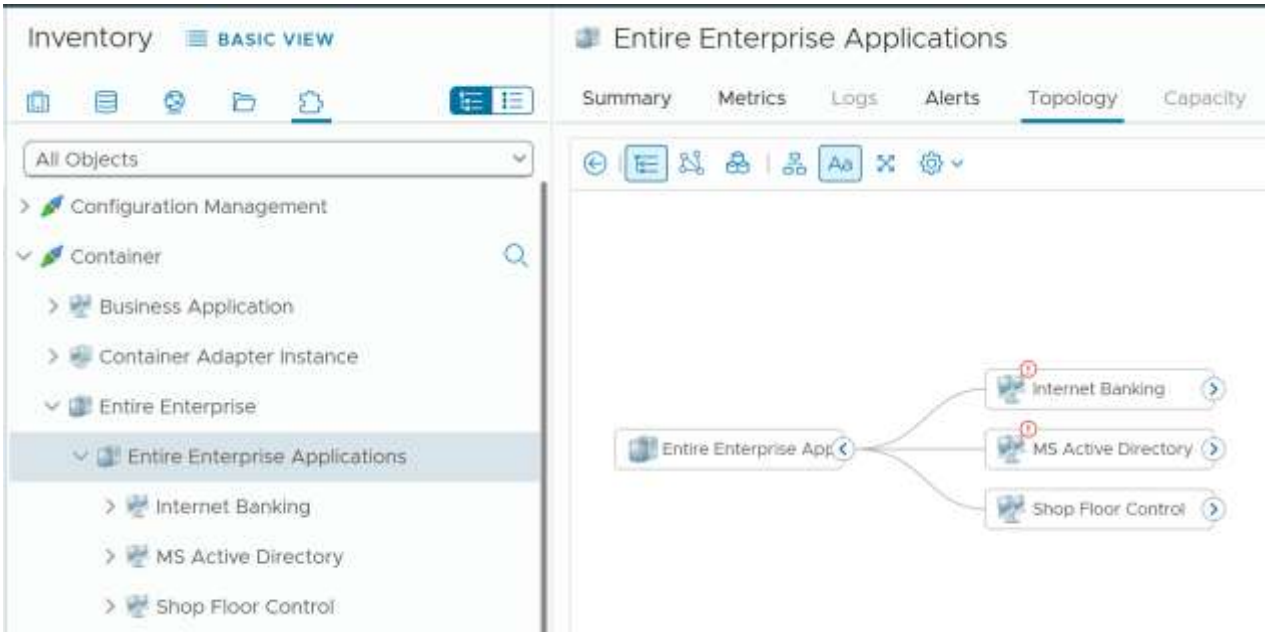
The folder name needs to be unique, even across vCenter Servers. So, it's paramount to have the application name as prefix.

If you have multiple vCenter servers, the name should be consistent across all of them.

Objects Design



Business application is not an object that belongs to vSphere or VMware. It can potentially contain objects outside VMware objects. So the parent object is different. It's called Entire Enterprise Applications, as shown on the following screenshot.



In the screenshot, you can see the structure that it's a parent of 3 business applications

Limitation: The tiers are manually added. It has no dynamic membership the way Custom Groups has.

Dashboard Design

The dashboard focuses on Performance. To some extent, if the Guest OS is down, the dashboard below will detect it. If you enhance it to include availability and compliance, let me know!

The top part of the dashboard shows the overall trend.

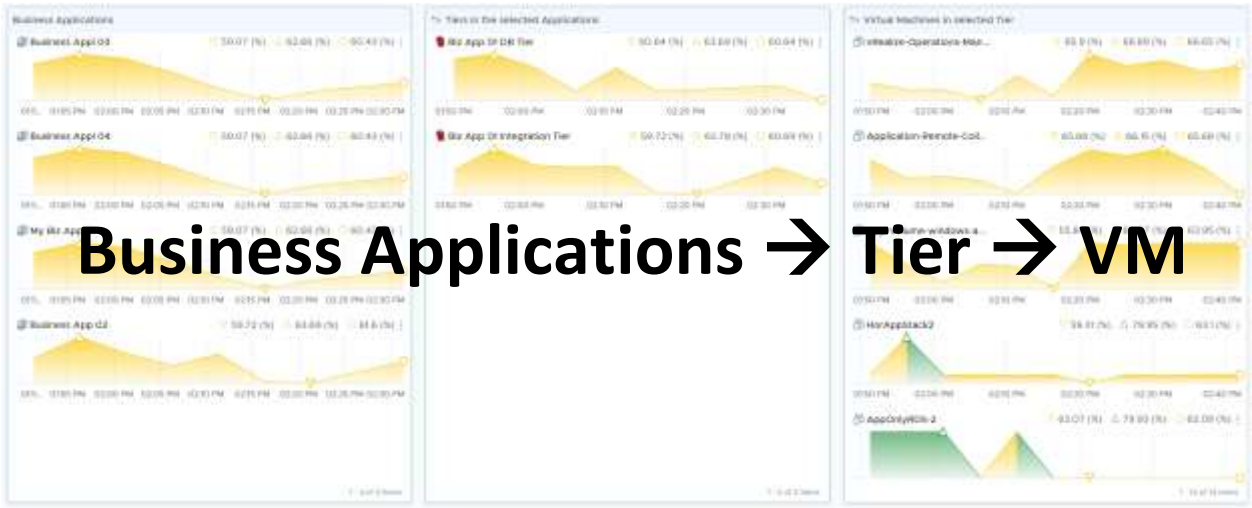


Assuming you only list your mission critical applications, you should expect the value to be good and match business reality.

The problem with average is it might mask a problem. For example, you have 2 mission critical applications. 1 has is value goes up (for whatever reason) while the other goes down. The average will fail to show it.

This is why the next row on the dashboard complements the overall picture. It has 3 columns

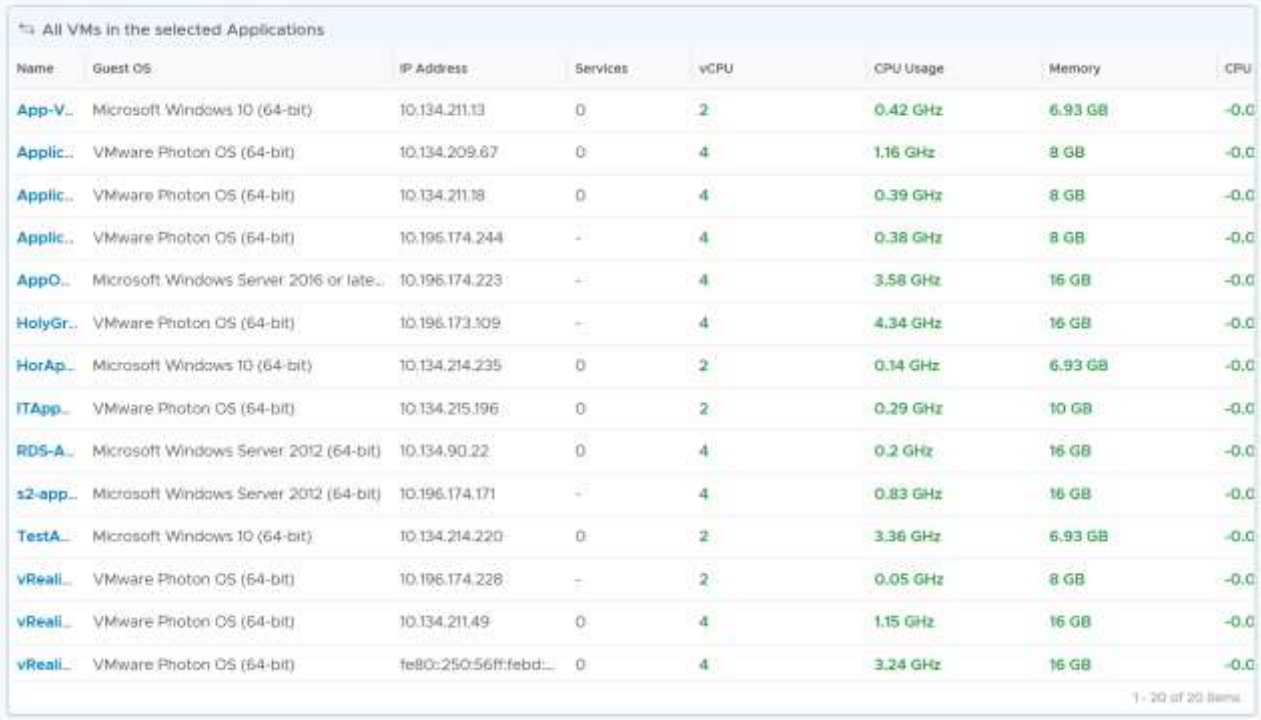
- The 1st column shows all the business applications. It's sorted by the worst performing app, judged from the most recent value.
- The 2nd column shows the tiers within the selected business application.
- The 3rd column shows the VMs within the tiers.



The columns also support drill down into VM Performance dashboard and VM Capacity dashboard.

The last column, because it's a list of virtual machines, also sports a drill down into the guest operating systems. Take note this requires Telegraf agent.

The 3 columns also drive the table underneath them. The table lists all the VMs for ease of analyzis.



All VMs in the selected Applications

Name	Guest OS	IP Address	Services	vCPU	CPU Usage	Memory	CPU
App-V..	Microsoft Windows 10 (64-bit)	10.134.211.13	0	2	0.42 GHz	6.93 GB	-0.0
Applic..	VMware Photon OS (64-bit)	10.134.209.67	0	4	1.16 GHz	8 GB	-0.0
Applic..	VMware Photon OS (64-bit)	10.134.211.18	0	4	0.39 GHz	8 GB	-0.0
Applic..	VMware Photon OS (64-bit)	10.196.174.244	-	4	0.38 GHz	8 GB	-0.0
AppO..	Microsoft Windows Server 2016 or late..	10.196.174.223	-	4	3.58 GHz	16 GB	-0.0
HolyGr..	VMware Photon OS (64-bit)	10.196.173.109	-	4	4.34 GHz	16 GB	-0.0
HorAp..	Microsoft Windows 10 (64-bit)	10.134.214.235	0	2	0.14 GHz	6.93 GB	-0.0
ITApp..	VMware Photon OS (64-bit)	10.134.215.196	0	2	0.29 GHz	10 GB	-0.0
RDS-A..	Microsoft Windows Server 2012 (64-bit)	10.134.90.22	0	4	0.2 GHz	16 GB	-0.0
s2-app..	Microsoft Windows Server 2012 (64-bit)	10.196.174.171	-	4	0.83 GHz	16 GB	-0.0
TestA..	Microsoft Windows 10 (64-bit)	10.134.214.220	0	2	3.36 GHz	6.93 GB	-0.0
vReali..	VMware Photon OS (64-bit)	10.196.174.228	-	2	0.05 GHz	8 GB	-0.0
vReali..	VMware Photon OS (64-bit)	10.134.211.49	0	4	1.15 GHz	16 GB	-0.0
vReali..	VMware Photon OS (64-bit)	fe80::250:56ff:febd:...	0	4	3.24 GHz	16 GB	-0.0

1 - 20 of 20 items

Customization

You can enhance the visibility by corelation. For example, you can compare the application-level metric with infrastructure metric. CPU is a good general candidate. For example, you compare the number of web server session with the CPU Usage in GHz. If application developer releases an improved code that enable the web server to serve more users with the same resource, but the reality is the opposite, then they have something to troubleshoot.

You can enhance this by adding process-level metric. This requires you to install Telegraf agent.

A common request among VMware Admin is to give their customers a self-service access to their own VMs. The VM Owners should be given a simple portal, where they can easily see all their VMs and its performance. You can use the dashboard sharing feature of VCF Operations for a login-less access.

But what if your tenants are from different companies? They are not allowed to see one another VMs.

You don’t want to create a dashboard for each of them one by one. The challenge here is how to use the same dashboard for multiple applications teams. This requires a security mapping. Each tenant needs to have a login ID, which must be mapped to their VM.

Role	The first step is to create a role and give it limited access. All tenants user accounts will be mapped to this role. This role should not be able to browse the inventory. Its only access is to the group of tenants.
Dashboard	Create a common dashboard and map to the role. This role can’t see any other dashboards
Tenant	For each tenant, you need to create a user ID. This ID is then mapped to a group. The group has the tenant VMs. In this way, the tenant ID will not be able to see other VMs.

Capacity

In order to calculate the capacity demand, you need to know whether it is active-passive or active-active set up.

CPU Consumption =

```
(
    If Active-Active then Average (VM Member CPU Usage) Else Max (VM Member CPU Net Run)
)
/
Min (VM Member Total Capacity) * 100%
```

The above gives you a number in percentage, making it easier to see across many tiers. Let me know if you find a use case for absolute numbers (in GHz for CPU and GB for memory)

For total capacity, we take the lowest value among the members VM.

Alert

Managing alerts is not the same as minimizing alerts. Managing is dealing with alerts that are already triggered. Minimizing takes us towards preventing alerts to begin with. Use insight to minimize alert definition, as alert should be reserved for urgent and important issue.

Your goal is to minimize alerts storm while providing the broadest coverage. This calls for a careful planning on the alert definition, symptom and threshold.

Alerts Planning

For each alert, ask yourself: what action must be taken, today, by the person seeing the alert?

- If the answer is nothing, then why disturb the person today? He will be bombarded with early warnings. Dashboard gives better picture as it shows issues that are yet to breach the threshold.
- If the answer is something, can that action be automated? If the answer is yes, then why do you want to be alerted? You should automate the response. Be careful of simplified logic as computer has no common sense. An automation that is not fool proof can result in a disaster.

If the answer is escalating to the next level (e.g. Level 2 support) after adding some context and initial analyzis, then it depends on the urgency. Generally speaking, L2 has longer timelines and need to see bigger picture. So the L1 frontliner may group the alerts into 1 instead of escalating each as isolated cases. If this is the case, help the L1 to see the big picture. A dashboard is more effective.

Target Persona: The Who

It is essential to design the alert for each persona. What is relevant for one may be irrelevant to others.

There are at least 3 main personas, with additional sub-personas:

- Application team
- Infrastructure team. This is the main persona, but it has at least the following sub-personas:
  - Network Team
  - Storage Team
- Security team. They look at both, but purely from security perspective.

Class of Service: The What

Not all workloads are equal. The mission critical applications are certainly more important than its development version.

Classify your workload into 3:

- Mission critical applications
- Standard production
- Test and Development

Having more than 3 can complicate operations.

Tailor the alerts according to the class of service. The higher the criticality, the lower the threshold. For example:

- Mission Critical: VM Disk Latency > 30 ms for 5 minutes.
- Standard Production: VM Disk Latency > 60 ms for 10 minutes.
- Test and Development: VM Disk Latency > 120 ms for 10 minutes.

What to Alert

What areas do you want to monitor with alerts? What can you do to reduce the noise?

Define the types of problem for each pillar of operations. For each type, classify if an alert is needed or not.

Availability	Proactive hardware replacement before reaching the manufacturing limit. For example, the SSD disk has limit of number of writes as each write introduces wear and tear. The manufacturer has a number in mind for endurance. If you have thousands of disks, you create a weekly dashboard just to track this limit and schedule proactive replacement for those disks nearing their limit.
	Check the hardware temperature as heat is the #1 enemy of electronics. Okay, if you have water leakage you have a different problem
	Ensure software stack compatibility as versions incompatibility is known to cause outage.

	Have redundancy, so unplanned downtime triggers non-critical alerts so long it's within the allowance set in the design. If you design N+1 availability, then a single downtime is not a red alert
	Trap soft errors. They are non-critical events that act as early warning. Treat them as insights.
Performance	Minimize reactive operations not by using lower threshold as that will result in noise. Instead, track early warning metrics via insight.
	For consumption type of alerts, only define red alerts. Orange level is best covered by daily dashboard.
	For contention, suppress for low utilization, as likely the impact is not felt at the application level. Use CPU as the proxy for application.
	Daily check for sustained and high utilization, especially those trending upwards over time.
	Separate formal SLA (alert) with internal KPI (insight).
Security	Security attack. If the attack is live, the alert screen needs to be accompanied by NOC screen.
	Check for undesirable events
Configuration	Only high-risk settings that cause urgent problems such as slow performance, outage, or security hacking. Non-urgent config issues are best managed via daily health check.
Compliance	Only the urgent one, where you cannot wait until tomorrow
Capacity	Only define red alert as it's a guard rail.

Alert should not replace operations best practices. For examples, you should ensure VMware Tools are up to date. If you set an alert for outdated VMware Tools, you become reactive and may get excessive alert. Since it's not urgent, it's better to have a daily SOP where you have a list of VMs with outdated VMware Tools. Make sure the list is empty.

Alert should be triggered on all affected objects, both the victims and the villain. It is common that a failure at the lowest ripples through the upper layers. For example, a physical disk failure in a vSAN cluster. This in turn can bring down an entire vSAN disk group. If the ESXi host has only 1 disk group, it can impact the host as it will have no storage.

What not to Alert

Do not set alert on issue that you can review the next day. You're better off with a daily dashboard, as it lets you see the big picture.

Availability	Backup failure. For example, you do nightly back up of 1000 VM and 1 fails. Since you still have the previous night backup, you don't need to troubleshoot at night. The exception here is there are many VMs affected.
Cost	Nothing urgent.
Inventory	Since there is already a NOC screen tracking inventory volatility, you don't need an alert here as inventory it's merely an account of what you have. There is no good or bad.

Capacity	Reclamation and downsizing is best done via daily dashboard. This is often mistaken as performance problem, as high utilization is a common alert.
Configuration	Non-urgent issues, such as outdated VMware Tools.
Compliance	Non-urgent issues, such as non-security compliance to corporate standard.

Examples of important but not urgent:

- "Host is in maintenance mode for at least 72 hours" is not urgent as it's been down for 3 full days. It's important as that's not normal.
- "A recoverable memory error has occurred on the host" is not urgent as the issue has been recovered. You likely need to replace hardware. As the problem can spread, it's an important one. You need to schedule a hardware scan, which may require a schedule downtime.
- You notice vMotion has been taking 300% longer in the last 12 hours, going up from the usual 0.2s to 0.6x second in 10 large clusters, where a total of 10K production VMs are running. But it's a low level issue that none of the app owners have not noticed. Not urgent, but important as it can potentially turn into a big problem.
- You notice a performance problem that has happens randomly. It went away by itself within 5 minutes, so the owner kind of forgive your IaaS. But the problem hit VMs randomly, and it's been going on a few times a day for a few days. Different owners are impacted so each does not really complain as each is rarely hit. You feel it's risky to add more VMs.

Design

Design the alerts as a set.

- Start with the 4 elements of infrastructure (CPU, memory, disk, and network). The reason is each has their own behaviour.
- For each element, look at the entire set of objects being monitored, and design top down. Do both consumer objects and provider objects at the same time.
- For each object, cover the 2 types of metrics: contention and consumption.

Once you have a list of alerts, validate with

- Consistency check. Run through each object type, and review all the 4 elements of infrastructure to make sure they are correct.
- Bottom-up scenarios. Ensure all the possibilities are captured.

Once you do the above, you will end up with an alert count for each.

The following shows the overall count of alerts for the top 5 objects in vSphere. The first 2 are consumer objects, and the last 3 are the primary provider objects.

Some objects have 2 alerts as we need to account for different problem (e.g. read vs write latency).

		Guest OS	VM	Res Pool	ESXi	Cluster
Contention	CPU	1	1	1	1	1
	Memory	Not yet	1	1	1	1
	Disk	1	2	N/A	Not yet	Not yet

	Network	Not yet	1	N/A	2	Not yet
Consumption	CPU	Not yet	1	N/A	1	1
	Memory	Not yet	N/A	N/A	1	1
	Disk	1	(Guest OS)	N/A	N/A	N/A
	Network	Not yet	Not yet	N/A	Not yet	N/A

Not yet = I've not implemented it yet. Some of the Guest OS alerts requires Telegraf agent.

N/A = alert is not applicable for this object as daily check or live NOC screen is better.

Let's now add the last 4 objects in vSphere. They provide storage and network resources. Data Center is not included as it's just a group from alerts design perspective.

Type of Alerts		Datastore	Datastore Cluster	vSwitch	Port Group
Contention	Disk	0	2	N/A	N/A
	Network	N/A	N/A	N/A	1
Consumption	Disk	1	Not yet	N/A	N/A
	Network	N/A	N/A	N/A	N/A

Alert Definition

There are 2 main approaches to define an alert.

	Generic Alert	Specific Alert
What it is	It covers specific symptoms. The cause could be many. It's symptom oriented.	It covers specific cause. It's cause oriented.
Examples	VM did not get the CPU time it asked	VM had CPU contention due to CPU Co-stop caused by snapshots
		VM had CPU contention due to CPU Co-stop caused by too many vCPUs
Strength	Alert name will be simpler	Alert name is more specific
	Overall less alert definition	Action is clearer
Weakness	Can't tell the actual problem. For example, is it read or write latency? But do we need to know at this stage, since alert is just the tip of the iceberg? Can we pass this to Insight, where we can see more?	Very complex to design. It does not cover unknown cause. So we still need the catch all as the last resort. That means designing the alert is complex as the catch all typically contains certain traits.
	Recommendation becomes a long list of possibilities. Can be solved by having a condition in recommendation itself, so it's not a long list of irrelevant recommendation. The recommendation should contain live, clickable information instead of just plain text.	Most causes are not known, so this has limited use case. For example, we can't prove that a snapshot is causing CPU contention or even disk latency. We only know probable cause, so we're unable to prove the actual. We pray that we make some changes, and the problem goes away magically.

		Many known causes are due to misconfiguration. Why wait for an alert to address the incorrect config? Regular health checks should be provided.
		Mostly only for Infrastructure Team. Not so suitable for VM Owner as she does not and should not care about underlying infrastructure problem.

Threshold Design

Follow industry standard. For compliance, adhere to the relevant industry regulation.

Align with your promised SLA. If you do not have SLA with your customers, use your internal target.

Review the complaints (formal via ticket, or informal via your management chain).

Have many symptoms instead of a single and simple one.

- This sharpens the alert as it becomes narrower.

For contention, make sure the issue is both acute and sustained

- Acute is sharp, lasting only 20 second but severe.
- Sustained is average over 300 seconds.
- Set the acute issue to be 4x higher over the sustained issue.

Make sure the contention alert does not occur at the point of very low utilisation, as there may not be meaningful business workload.

Design a consistent scaling within the context, so the relative urgency of red vs orange is maintained.

For example, Red is always 4x of orange. You can make red 2x as bad by modifying any 2 of the following:

- 2x the threshold. If orange uses 10 ms latency, then red is 20 ms.
- 2x the population. If orange impacts 10% of the VM, then red impacts 20%.
- 2x the time. If orange lasts for 5 minutes, then red is 10 minutes. I'm not in favour of this as I think it 0.5 hour makes more sense to operations team.

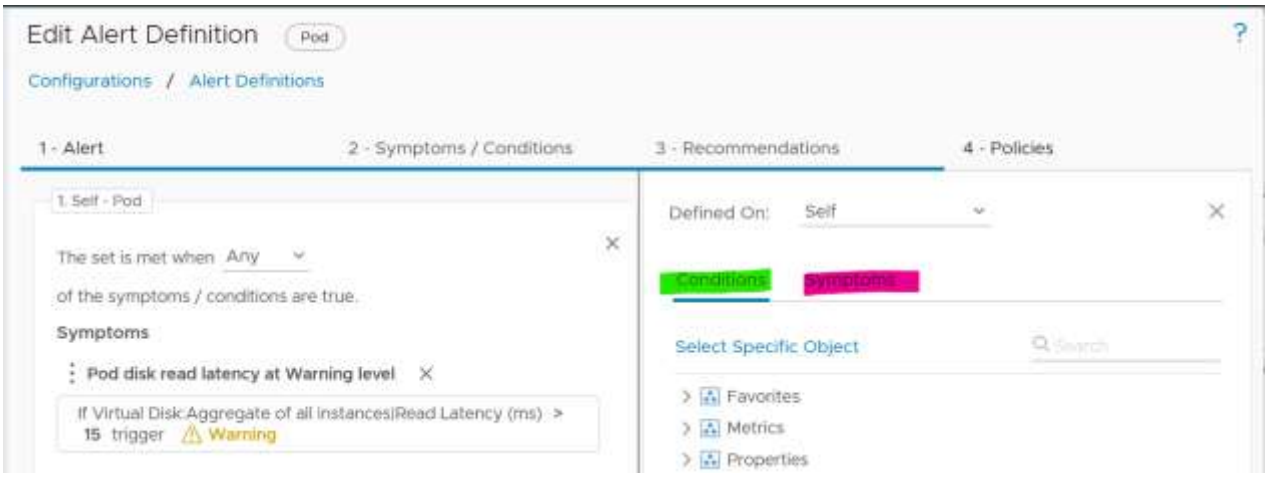
Wait Cycle

For contention, use 5 minutes. In cases where the metrics tend to have false positive, increase to 10 minutes.

For consumption, use 15 minutes.

Conditions

Do not use Symptoms. Instead, use Conditions. Symptoms means lighter or small problems that do not warrant an alert. Think of it as you got a light body ache which would go away after resting.



Recommendation | Remediation

Recommendation is just a piece of advice. It does not have ability to change. Remediation, on the other hand, will execute change in your environment. Recommendation is safer as it's not automated.

It's hard to develop a great remediation, or complete-yet-concise recommendation. The reason is the root cause is typically far from the alert triggered. There are too many combinations. We need to stop thinking that IT products have good observability. IT systems actually have poor visibility into what happened. Can you troubleshoot when your web browser is slow even though your high-end laptop is fast and overall utilization is low?

No, you can't. If you can't troubleshoot a single PC, why do you expect we can troubleshoot something far more complex? Observability is always a second-class citizen in product development. Most products only have few meaningful metrics. Even if they do have the metrics, they don't provide API. Even if they have, the metrics could be unreliable or buggy.

A good recommendation is clear. It needs to list its assumptions and explain the reason behind the recommendations.

In the ideal world, your alerts are fully automated. The problem is RCA today is largely an exercise of elimination. It is a process that requires human expertise on the architecture. For example, to troubleshoot Tanzu, you need to know K8 + vSphere + Tanzu specific implementation.

The reason why we have to resort to elimination exercise is the lower layer or stack does not carry information from upper stack. For example, ESXi SCSI commands can't be traced to specific VM.

Maybe one day we can have AI-assisted troubleshooting. I envisage the human expert "records" the steps. This becomes the training dataset.

Application Team

We start with Guest OS as this is the closest layer to the application. This alert is for someone responsible for the application, running on top of Windows or Linux. It is not relevant to the infrastructure team.

Guest OS Alerts

Since performance problem can happen within Guest OS, independently of virtualization, you need to monitor within this layer. The goal is to proactively manage performance before business is affected. The challenge happens when there is no application level or user transaction level monitoring. Another word, the application team simply rely on complaint. The problem with complaint is it's highly subjective. The correlation to technical counters is both weak and unpredictable.

CPU Contention

Goal	The CPU Run Queue is relatively less known due to the focus on CPU Utilization. This alert is likely independent of the underlying virtualization layer due to the nature of some applications which spawn excessive threads. Demonstrate to application team or VM owner that the IaaS team has visibility into Windows or Linux.
Red Condition	All these conditions must be met for 15 minutes: • Guest OS Peak CPU Queue > 20 • VM Net Run > 75%
Orange Condition	This is not provided to reduce noise. If you need to set, follow these: • Guest OS Peak CPU Queue > 10 • VM Net Run > 75%
Design Consideration	As the CPU Queue counter is known to have false positives, we increase the wait cycle to 15 minutes, and add Net Run to ensure it only happens when there is sufficiently high workload.
	The alert works together with the VM CPU Utilization alert and VM CPU Contention alert. Use these 3 alerts together.
Remediation[1]	If the alert of high VM CPU contention is also triggered, then follow the remediation for this alert.
	If the alert of high VM CPU consumption is also triggered, then follow the remediation for this alert.
	If the above 2 are false, look at the application on why it's creating many threads. Compare the values with other software or code that are part of the larger business applications. Also, compare the value with the same software in other business applications. If the software is a commercial software from IT vendor, ask the vendor for their recommendation as the maker of that product. If you do not get the answer, exclude this application from this alert.

[1] Thanks Darren Farrer for the collaboration.

Disk Contention Alert

Red Condition	All these conditions must be met for 5 minutes: • 20-second peak disk queue > 10000 • CPU Usage > 1 GHz
Orange Condition	All these conditions must be met for 5 minutes: • 20-second peak disk queue > 5000 • CPU Usage > 1 GHz
Design Consideration	There is no official guidance from either operating system. This alert is set at relatively high number, so adjust as you deem appropriate
	If you have excessive false positive, disable the orange condition.

Other Contention Alerts

What other performance problems can exist only within the Guest OS, which is not caused by underlying virtualization? There are many possibilities. I'm not creating alerts as their threshold vary wildly by applications.

You should create your own alerts for the following:

- Check if CPU Context Switch is consistently high. Inform developers that this could cause application performance problem. If the VM CPU Net Run is low, and run queue is low, then reduced the vCPU size.
- Heavy memory paging, coupled with high amount of page fault.
- High number of outstanding disk operations.
- Network dropped packet. Note this requires Telegraf agent.

Consumption

There is no alert defined, as it could result in false positive.

There are, however, proactive check you can do:

- Check CPU Usage Disparity. If it's consistently high, discuss with VM Owner on why application unable to use all vCPU equally. Reduce size if not required, especially if CPU Context Switch is high and run queue is low.
- Check free memory. If it's consistently near 0 MB, coupled with moderate or high amount of paging, discuss with VM Owner if this impacts performance. Do application-level benchmarking before and after as too much memory can negate performance.
- Low disk space. Not yet implemented as certain partitions such as swap is always near 100%. This cannot be implemented at VM virtual disk layer.

VM Alerts

In overcommit environment, a VM can experience contention. How do you balance between application team catching this alert, and infrastructure team to address the issue first?

The balancing will minimize blamestorming. Some techniques:

- Avoid using the orange alert. So it's only red. You keep the orange for the infrastructure team, so they have a heads up.
- Use a higher threshold. For example, you set CPU Ready > 8% for application team and > 6% for infrastructure team.

Alert at infrastructure level, such as ESXi, is not relevant.

CPU Alerts

Design the CPU alerts as a set. That means cover all objects, and for each object, cover both contention and consumption. The contention alert and utilization alert should be used together. Especially for consumer, where capacity can be increased at a moment notice, increasing capacity when there is contention will result in inferior performance.

Contention Alerts

VM and IaaS have different perspective. For the shared infrastructure, the alert should focus on population issue, as opposed to just a small subset of VMs. We use 25% of VM population as the overall threshold. As this can mean hundreds of VM, we choose a lower threshold per VM.

This means ESXi, Resource Pool and Cluster. Data Center is not added as it can consists of clusters of different purposes.

VM Contention

Red Condition	All these conditions must be met: • Peak vCPU Ready > 32% • CPU Ready > 8% • CPU Contention > 8% • CPU Usage > 1 GHz
Orange Condition	All these conditions must be met: • Peak vCPU Ready > 32% • CPU Ready > 4% • CPU Contention > 8% • CPU Usage > 1 GHz
Design Consideration	An absolute unit is used for CPU utilization as opposed to relative (%) as a small percentage of a monster VM could still be sizable
Remediation	This is complex. See the diagram at the start of CPU Alert section.

Infrastructure Contention

Red Condition	All these conditions must be met: • Peak vCPU Ready > 24% • CPU Contention > 6%
Orange Condition	All these conditions must be met: • Peak vCPU Ready > 12% • CPU Contention > 3%
Design Consideration	For cluster, we measure directly at VM level so it's closer to what we care, which is the application. Aggregating from ESXi is less accurate.
	For resource pool, the symptom does not work with cascading RP. Stacking RP on top of RP complicates operations.
Remediation	This is complex. See the diagram at the start of CPU Alert section.

Consumption Alerts

Utilization alert is tricky to define as high utilization could mean good or bad.

VM Consumption

Goal	High utilization could cause business performance, especially for online workload where end users expect a "real time" response time.
Red Condition	All these conditions must be met for 15 minutes: • VM CPU Net Run > 95% • Guest OS CPU Queue > 4
Orange Condition	There is no orange alert as it will create more false positives.
Design Consideration	15 minutes is used instead of 5 to reduce false positive. Take note this means you will lose 15 minutes.
	Added Guest OS CPU Queue to ensure that there is actual bottleneck, as opposed to running optimally. This should result in a more meaningful alert.
Remediation	Find out the nature of the application. If it's batch job, the CPU Net Run could be sustaining at >95% for the entire duration of the batch run. So long the job is completed within the expected window, this is a healthy behaviour. Exclude this application from this alert.
	Find out from application team what metrics at application-layer or business transaction layer are proving that there is queue and latency. If the pattern matches the number at the infrastructure layer, then add vCPU. In the event you decide to add CPU, make sure it's NUMA compliant.

Infrastructure Consumption

Red Condition	All these conditions must be met for 10 minutes: • CPU Thread Utilization > 40% • CPU Core Utilization > 80%
Orange Condition	There is no orange alert as it will create more false positives.
Design Consideration	As this is a provider, the focus is on capacity management, not performance. Space, not speed. That's why counter such as Usage is not used.
	The threshold is not that high as we need to consider HA. Ideally, we should add the HA buffer into the threshold dynamically. If this results in excessive noise, increase the threshold to 50% and 100%.
	It is not set for Resource Pool as its use case varies.
	For cluster, it's simply based on > 25% of ESXi members of the cluster

Memory Alerts

Design the memory alerts as a set. That means cover all objects, and for each object, cover both contention and consumption.

Contention Alerts

Memory performance is hard to trace. The memory contention can be come from different factors. We can't distinguish how much each part contributes. For example, VM 001 has 51% memory contention. This could be due to:

- 5 MB was limited Limit.
- 1 MB was swapped.
- 2 MB was compressed.
- Low shares setting played a part.

VM Contention

Red Condition	All these conditions must be met: • Peak Memory Contention > 24% • Memory Contention > 6% • CPU Usage > 1 GHz
Orange Condition	All these conditions must be met: • Peak Memory Contention > 12% • Memory Contention > 3% • CPU Usage > 1 GHz

Design Consideration	An absolute unit is used for CPU utilization as opposed to relative (%) as a small percentage of a monster VM could still be sizable

Infrastructure Contention

Red Condition	All these conditions must be met: • Peak Memory Contention > 16% • Memory Contention > 4%
Orange Condition	All these conditions must be met: • Peak Memory Contention > 8% • Memory Contention > 2%
Design Consideration	For cluster, we measure directly at VM level so it's closer to what we care, which is the application. Aggregating from ESXi is less accurate.
	For resource pool, the symptom does not work with cascading RP. Stacking RP on top of RP complicates operations.

Consumption Alerts

There is no alert set at VM level as VM is just a motherboard.

There is no alert set at Guest OS level as it varies per application.

Infrastructure Consumption

Red Condition	All these conditions must be met: • Balloon > 10 MB AND • Consumed > 95%
Orange Condition	There is no orange alert as it will create more false positives.
Design Consideration	10 minutes is used instead of 5 to reduce false positive. Take note this means you get notified later.
	Balloon is set in an absolute term (10 MB) and not relative to the ESXi memory consumed so the impact is standardized.
	Swap and Compressed are not included so that alerts are triggered earlier, as remediation action, such as reducing demand, can take hours or days to complete.
	It's set at 50% of host members as clusters are mostly <20 hosts, and memory functions as cache (so more does not mean it's slow)

Storage Alerts

Design the storage alerts as a set. That means cover all objects, and for each object, cover both contention and consumption.

Contention Alerts

Read latency and write latency are separated as their causes are typically different. The remediation actions likely differ as a result.

VM Contention

Red Condition	For read latency, all these conditions must be met: • Average Read Latency > 80 ms • Peak Read-or-Write Latency > 320 ms • CPU Usage > 1 GHz • Disk Read IOPS > 100
	For write latency, all these conditions must be met: • Average Write Latency > 80 ms • Peak Read-or-Write Latency > 320 ms • CPU Usage > 1 GHz • Disk Write IOPS > 100
Orange Condition	For read latency, all these conditions must be met: • Average Read Latency > 40 ms • Peak Read-or-Write Latency > 160 ms • CPU Usage > 1 GHz • Disk Read IOPS > 100
	For write latency, all these conditions must be met: • Average Write Latency > 40 ms • Peak Read-or-Write Latency > 160 ms • CPU Usage > 1 GHz • Disk Write IOPS > 100
Design Consideration	Due to difference between vendor best practice and reality in customer environment, we set the wait cycle to 10 minutes and increase the threshold. In theory I agree we should pick a lower number, but in reality, there is hardly any complaint. Try to convince your boss to invest when there is no complaint 😉
	The 20-second peak cover both read and write. Yes, this is a known limitation.
	CPU Usage and Disk IOPS are added to reduce the alerts during low activity. The idea is business impact could be negligible during low disk activity.

Infrastructure Contention

Not yet implemented. The reason is the many:many relationship between VM and datastore makes it impossible for us to set condition at VM level.

If you create, use these as a guide:

>25% of VM population has latency > 10 ms AND highest VM latency > 40 ms AND Overall Datastore Latency > 5 ms.

Consumption Alerts

Storage consumption covers both capacity (disk space) and performance (IOPS and throughput). Performance is hard to quantify due to the lack of "100%". The ceiling is hard to define.

Object	Threshold for Red level
Datastore	Used space > 90%.
Datastore Cluster	Not yet implemented. It needs to trigger when a member datastore has > 90% used. The remediation is to trigger storage DRS.

Network Alerts

Consumption alert is hard to quantify due to the lack of "100%". The ceiling is hard to define.

Contention Alerts

Object	Threshold for Red level
VM	Transmitted packet drop > 4% and
	In future the threshold should be relative to packets transmitted. 4% is a high threshold as that means 1 in 40 packets were dropped. However, since it happens within a 20-second period, it is set high.
	Orange alert is not set to minimize alert storm. Dropped packets are typically retransmitted, so users do not experience problem. So the ratio is always 100%.
ESXi	Transmitted packet drop > 50 packets for virtual network
	Transmitted packet drop > 50 packets for physical network
Distributed Port Group	Red alert: Dropped packet > 10% Orange alert is not implemented due to false positive. In future I will change to transmit only.
Distributed Switch	Likely not applicable, as it's just a group of ESXi and port groups.

The Art of Dashboard

Part 2 Chapter 2

There are many considerations required in designing a suite of dashboards that tell a story. The last few releases of VCF Operations introduces many new capabilities that change the way your dashboards can look. Your dashboards are no longer confined to be standalone dashboard.

Design Consideration

To a few of us who love visualizing information with VCF Operations, we see the dashboard as a canvas. Granted, the widgets have limitation but that's part of the art 😊

Creating dashboards is an art as you need to balance many conflicting requirements, such as:



Can the dashboard be understood within 5 seconds? If yes, you buy yourself a few more minutes. The user understands what the dashboard does, and is willing to spend more time mastering it. To pass this test, think of which information, object, metric can you take away from the dashboard? See the KISS Principle, because an uber-dashboard that tries to please everyone and cover all scenarios will end up not being used.

Begin with the end in mind. The purpose dictates the dashboard design. 2 dashboards can have an identical target role, purpose, and use case (e.g. performance), but if the size of the environment differs, the 2 dashboards will be different. An environment with 50,000 VMs is managed differently with an environment of 500 VMs.

A good dashboard answers a set of questions. So jot down the questions. Be specific, including the metric and how it needs to be visualized.

A small environment with 100 VMs in just 8 hosts in 1 cluster (hence 1 data center, 1 vCenter) needs less dashboards than an environment with 10,000 VM spread over 800 ESXi hosts, 100 clusters, 10 data centers and 3 vCenter Servers.

Product wise, the dashboard feature in VCF Operations complement its built-in pages. Each page is hand coded HTML with precision. For customers who need to customize and personalize the information, we've provided out of the box dashboards that complement the existing page.

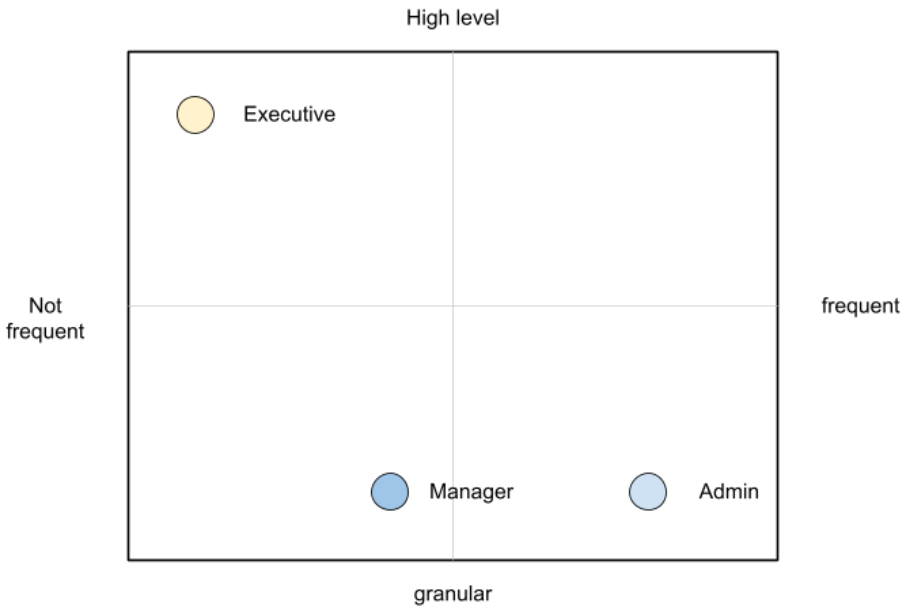
Design Methodology

This section introduces us to a few considerations surrounding dashboard design. We will walk through a method where we are able to understand the user, the activity, the workflow and the other aspects needed to understand the context before we start.

When we create dashboards that are used by others in our team we have to consider a few aspects that are helpful in making information useful, usable and delightful.

Who	Begin with the end in mind. Who are actually using the dashboards you create? What specific role and persona is going to use it? Start by creating a list of all users who would be accessing and using information in the dashboard Against this list mention the kind of information they would find useful for the specific persona.
What	What is the type of activity (granular vs. high level) they will perform? List next to the task if the information needs to be granular or high level. Sometimes users are looking for just a KPI number, sometimes they are looking for trends. Sometimes users might want to interact with the information, create drill downs etc. The complexity of the activity will help you identify if you need a complex widget or a simple one.
How	How (workflow) will they use the dashboard? List down what happens after they view the information, do they need to use another application, another visualization. Plan accordingly and place navigation to enable this to happen.
When	How often (hourly, daily, weekly, monthly, quarterly) will they use the dashboard? Think about the usage of the dashboard, is it a frequent use case needed for daily use to perform work, or is it for monitoring on an irregular basis.
Why	Why (outcome) do they need the dashboard? Also list down outcomes from the dashboard usage. What is the usefulness of using the dashboard and what goals or outcomes the user is likely to achieve.

Once we get this information we can better design dashboards based on the needs of the users. One way could be to plot it on a grid to understand the impact better. Let's start with a grid like this one.



An executive might use the dashboard less frequently and the type of tasks they would perform could be high level. If you make a dashboard that is very granular and requires a lot of effort the executive might not have the time to use the dashboard.

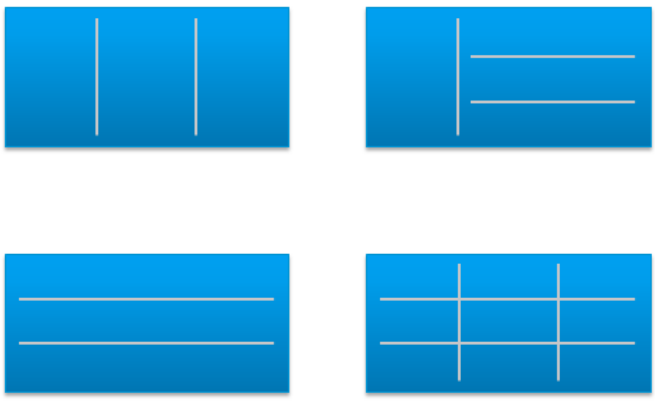
A manager might use the dashboard once a week, or on a monthly timeline, might need information on both granular or high-level.

Similarly an admin agent might use it on a daily basis and might need information at times that needs a lot of effort and at other times just a metric number would do.

Once you have reached a visualisation like this, you can better judge. If it needs to be a single dashboard or multiple based on the persona. What items should you prioritise and what kind of complexity is necessary based on the timeframe and time available with the stakeholder

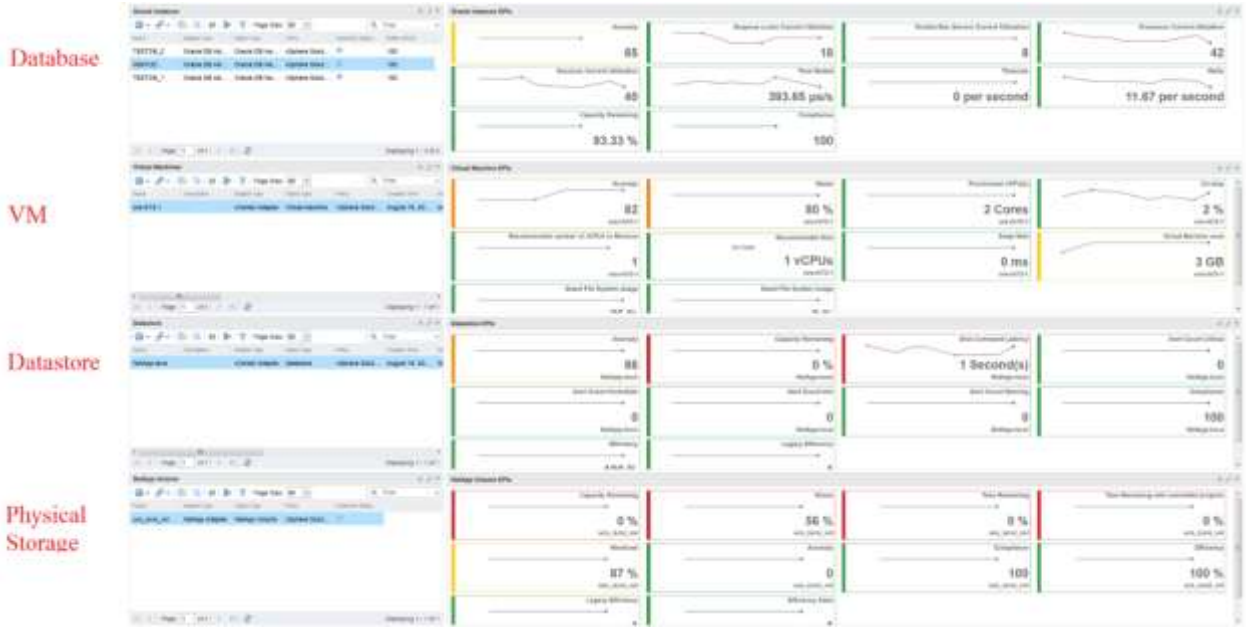
Clean Layout

Divide the screen into sections visually. This makes the dashboard easier to read. Here are some examples of how you can divide the screen.



Example 1

Here is a good example of layout from Brock Peterson. Notice how simple it is. It is clear that it has 4 layers as layout is consistent among them.



Example 2

You should also play with Dark Mode, and see how that looks. The following shows a clean layout in dark mode. Notice how Dale Hassinger has arranged this dashboard well.



Visually, he has designed the layout “top down”, making it easier to see.

Another technique is to create 2 sections: summary and detail.

- The summary section is typically placed at the top of the dashboard. It gives the big picture.

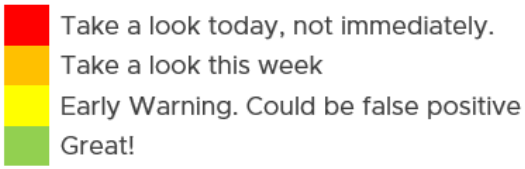
- The detail section is placed below the summary section. It lets you drill down into a specific object. For example, if it's a VM performance, you can get the detail performance of a specific VM.

This detail section is also designed with quick context switch, as you may want to check the performance of multiple objects during performance troubleshooting. Take for example VM performance. The dashboard gives you all the VM-specific information and allows you to see the KPIs without changing screens. You can move from one VM to another and view the details without opening multiple windows.

Color as Meaning

It's easier to see color than read numbers, if the color has meaning. There are many occasions in operations that you just need to know if things are good or bad first, before diving into the actual raw value. Color is also easier at a glance, especially if you have read hundreds of numbers inside a large table.

Here is the color I recommend:

Green –> Yellow –> Orange –> Red	Green means good, and gradually getting worse as it moves toward red. Typically used in performance monitoring and compliance. 
Dark Grey –> Green … Red	We use dark grey in Capacity as wastage (unused) is a bigger issue than over utilization. It's important to show wastage as it can also impact performance. For performance dashboard, you should consider using red to convey that oversized is bad for performance.
Blue	Neutral color. Used when it does not have any meaning.
Grey	System error. The data is missing due to collection error. Typically happens in heat map. We use a color instead of white as white is hard to read.

Threshold

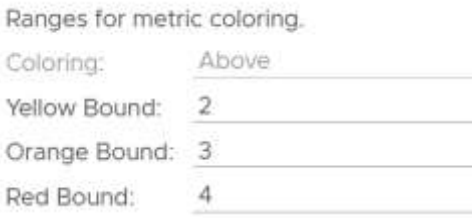
At which point do you change the color? This is where threshold comes in.

Try to make the scale easy to remember by adopting a consistent scale

- Yellow = 2x Green
- Orange = 2x Yellow
- Red = 2x Orange

The above also results in red being 8x green.

The above does not work when the scale is small. For example, in Horizon, a user typically has 1 – 4 desktops. It's rare for users to go beyond that. So in this case, your threshold is as follow:



Gradient

Gradient can be handy in widget where *relative* comparison among many objects is important. In the following example, the widget is answering the question about the severity of imbalance among vSAN disk groups. What's your conclusion if you solely rely on the color?



Oh no, they are not balanced. Let's initiate rebalancing, which can potentially move Terabytes of data!

But wait! We only care about unbalanced when they are significant. Minor unbalance is expected. In fact, if they are perfectly identical, it would be strange.

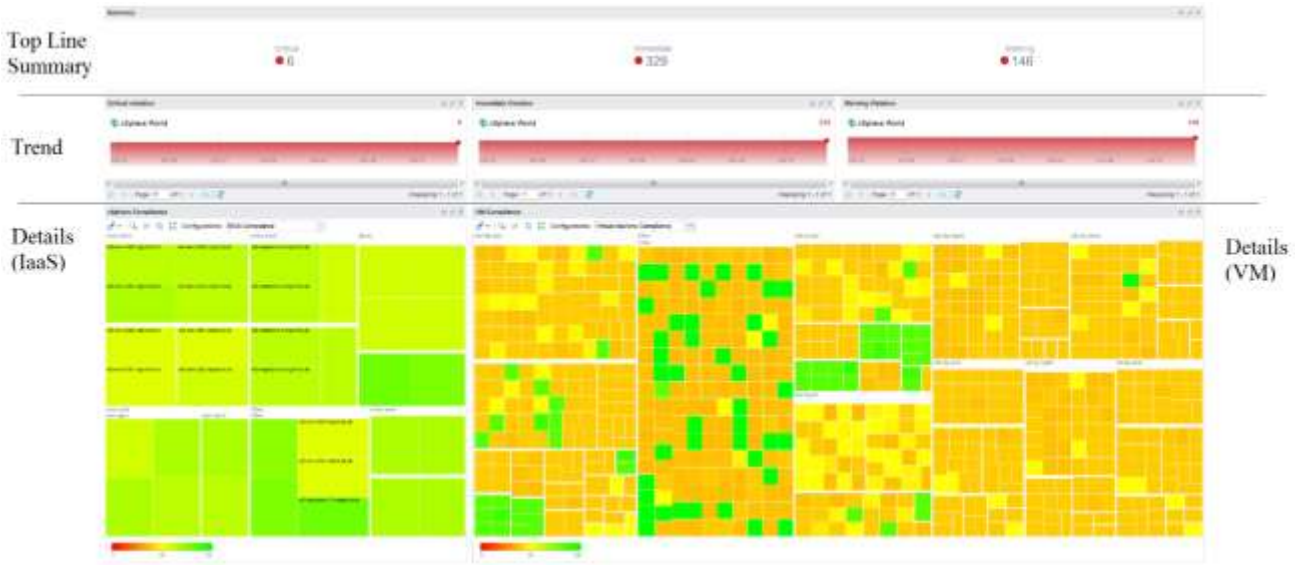
The problem with the above widget is utilization should only turn red when it reaches 75%.

After fixing the 2 preceding problems, we have a meaningful heat map.



Example Implementation

In the Compliance dashboard below, color is used to quickly show the various level of compliance. If all you see is Green, there is no need to look at the numbers & texts!



Last but not least, think of users who may not be able to distinguish all the colors. Provide alternative way for them so the functionality is not lost.

Past vs Present

Think of what timeline you need to show, as there are use case for both the present and the past. The past useful when the present data does not tell the full picture as it depends on the time you login and see it. If you login at 8 am in the morning to see what's happening over night, then the value at 8:00 am is not answering your question.

The past is harder to visualize as it has >1 data points. Ideally, you show the data as a *line chart* so you can see the trend. Showing a single data such as the maximum or average can miss critical information such as "is it trending upwards or downwards" and "how long did the peak last?". The problem with line chart is it takes up screen real estate.

If you need to show a lot of objects or metrics, then you need to summarize. As covered in Leading Indicator, you may need to show 2 numbers, especially if one of them is an average.

Summary → List → Detail

In large environment, you manage a large number of items (e.g. K8 pods, AWS EC2). This means you need a way to summarize the information, so you can work with a subset of the group. Within this subset, you need to know which particular item to work on.

The flow Summary → List → Detail is one common technique to achieve the above.

Summary

The summary shows the overall picture, not a subset.

There are 2 techniques of visualizing:

- Number oriented. Useful when the absolute value matters.
- Non number oriented. Useful when relative values of the numbers matter, as the numbers form a set.

The number-oriented uses a scoreboard. I find scoreboard is a good way to start, especially if the numbers are suitable for color coding.



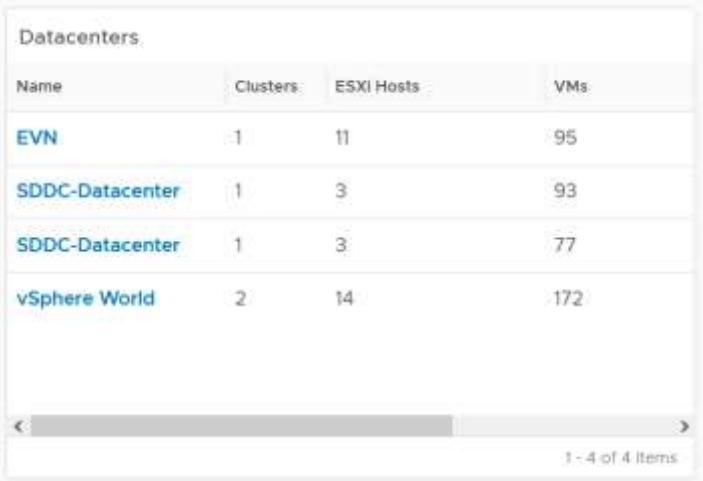
Based on the scoreboard, you may want to drill down to a subset of the environment. Take note that the drill down is based on the object owning the metric.

The non number oriented uses a distribution chart, such as bar chart or pie chart.



List

The View List widget is a table that lists all the objects. Use a filter if the number of items is large.



Having a filter also helps the dashboards scalability as showing tens of thousands of objects will impact the dashboard performance. It also improves usability as you can use progressive disclosure, and draw user attention to the big picture first.

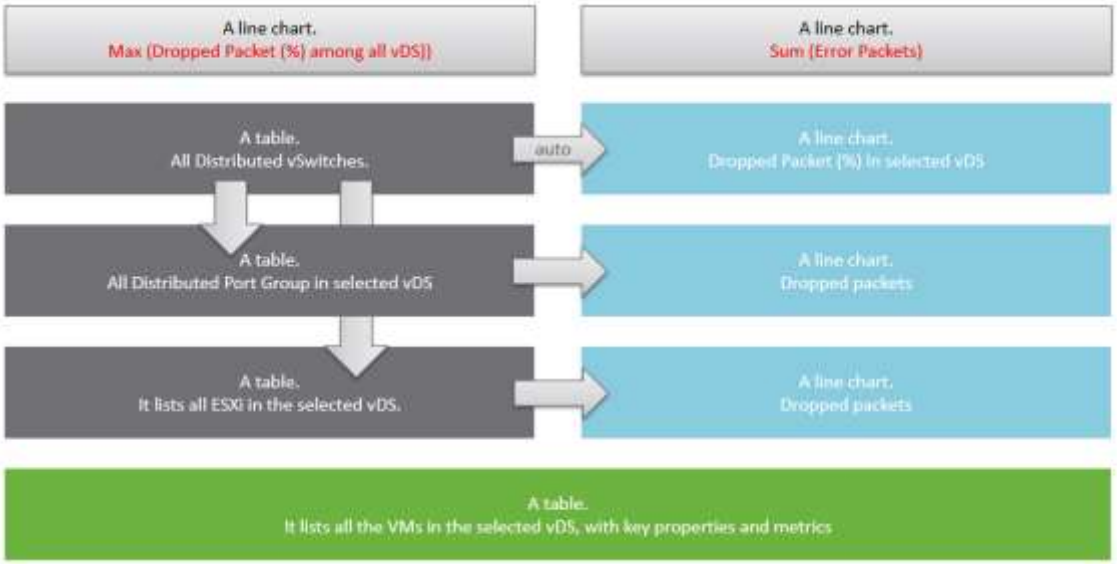
Detail

At the detail level, you deal with a single item, such as a Horizon session. If the item has large numbers of metrics to show, you need a way to see them before plotting the trend chart. I found the Scoreboard the most suitable as it can be color coded and supports drills down to see the trend chart.

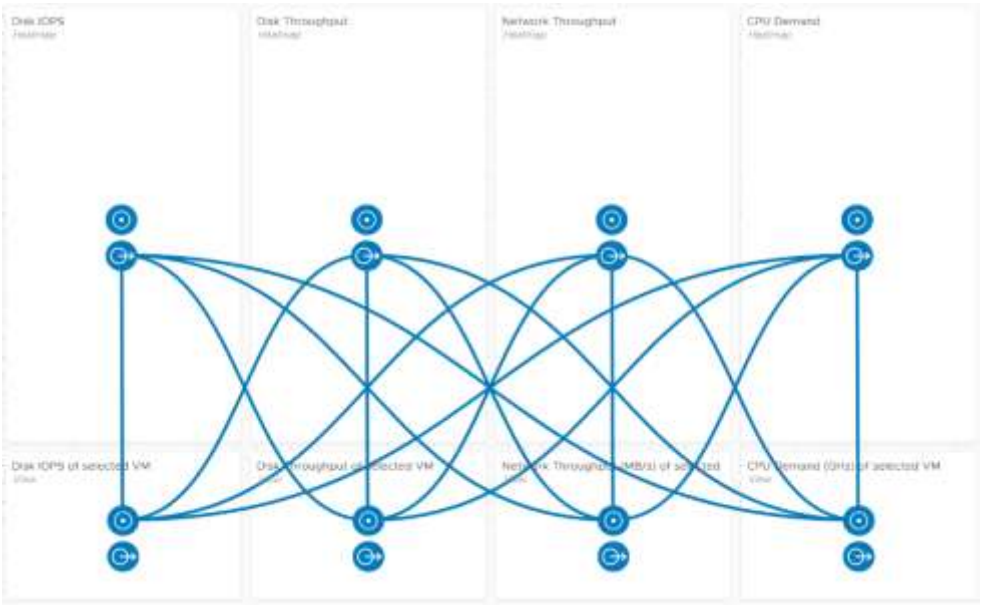
Interaction

Interaction allows you to build interaction within the widgets. So it is intra-dashboard, while navigation is inter-dashboard. Take advantage of the interaction & navigation, but keep them consistent so they are not too complex.

Logically design the dashboard first using any drawing tools so you can see the flow of information. I typically use PowerPoint, as shown in the following example.



A rich interaction will increase the functionality of the dashboard. Aim for a symmetrical interaction as it's easier for the dashboard consumer to understand. The following shows 4 heat maps driving 4 line charts in a 4x4 combination. It looks complicated in the screenshot below but feels natural when actually used.



Use progressive disclosure to minimize information overload and ensure the webpage loads fast. On the other hand, so long your browser session remains, it remembers your last selection.

Widgets

Keep the variants of widgets minimal, as it improves usability, especially when you have multiple dashboards or a dashboard with many widgets. You notice that I mostly use the same set of widgets in all the dashboards I have created.

The following lists the widgets I use:

Scoreboard	Headline or Banner or Summary. To get attention to the key point, I use large text, space out the content and apply a gradient to make it stand out. I may even use color. Object Property. To show relevant configuration of an object, I use small font. I keep the visual plain and simple. Many metrics per object. Being more compact than health chart, it's handy if you have more metrics than screen real estate.
Bar Chart	Absolute distribution. The actual value of each bucket is important to you. I find manual bucket size useful when the bucket range values is important.
Pie Chart	Relative distribution. The relative size of each slice is important to you. Look for patterns in the result, such as the number of slices shown matches your design or desired outcome.
List	Object List is good for one-time, disposable dashboards as it's convenient. I use View List for permanent dashboard as it has more features, especially the data transformation and parent values.
Heat Map	Use it if there are many objects, and you only need to see the present value. So it's suitable for capacity, configuration, live NOC, and not suitable for performance (where you need to see trend over time and data older than 5 minutes)
Health Chart	Where the color is more important than the value.
Trend Chart	I use View Trend Chart as my main line chart. I rarely or do not use other forms of line chart, such as Metric Chart. I rarely use the threshold for visual reason. The thresholds make the chart hard to read.
Top N	A table listing for top N numbers within the specified period.
Alert List	Use to show relevant alert. For example, for performance, I only show performance alerts as the full alert is available in the object summary page.
Relationship	I use the Advance version of the widget as it's visually more polished and meets all my requirements. I do not use the simpler version.
Text Display	I use it as a "Read Me" to either explain the whole dashboard or certain part of it.

Practitioners like Dale Hassinger use a lot of Text Widget, because he creates custom web pages. It certainly makes a functioning and cool dashboard!

I rarely use the following

- Metric Picker. I'd use it if as a designer I'm unsure of what metrics users will use.
- Metric Chart. I'd use it if I need to have multiple line charts on a single widget. Using less widget helps simplify the overall dashboard look.
- Object List. Use View List instead. Functionally it's a superset.
- Scoreboard Health. Use heat map instead.

Scoreboard

At the top of the dashboard, it's useful to provide a banner or headline, giving the big picture or summary. For example, this is what I use for Inventory summary.



The above is not color coded as each number does not have a good or bad meaning. It's visually less noisy also to have a neutral color.

The limitation is the number you see is the present number.

You can convey meaning by using color. There is also a line chart in the background of each box.

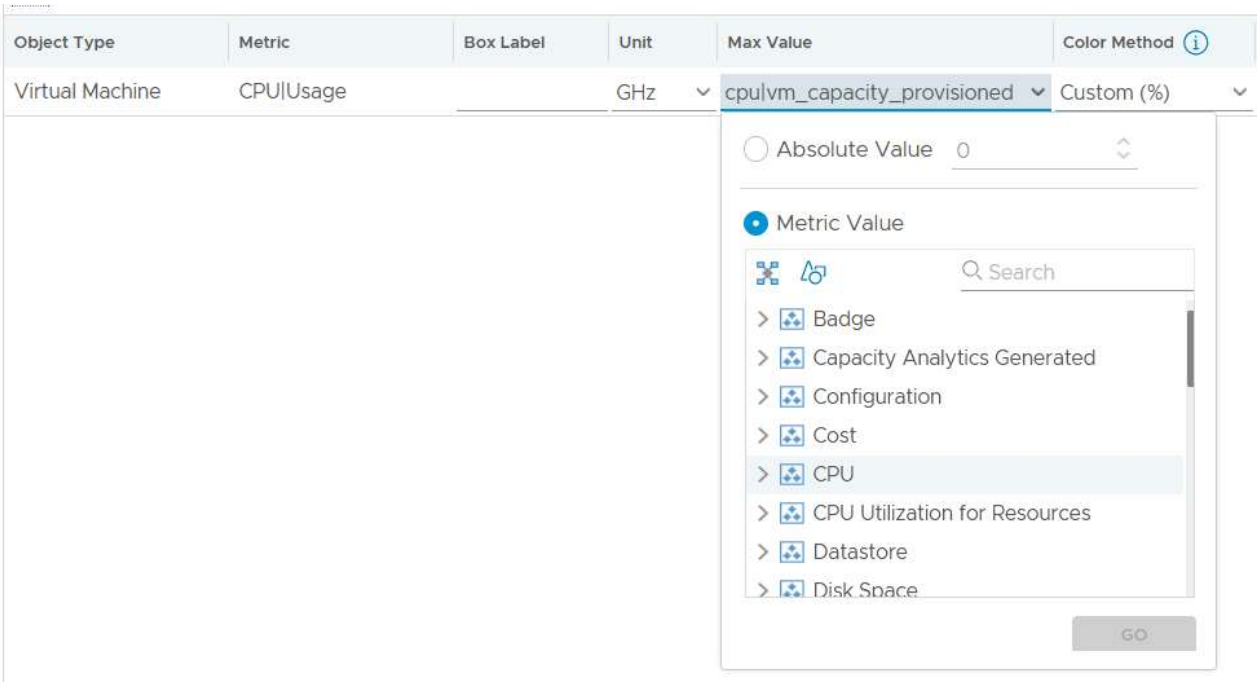


Take note that the background color is also based on this last value, not the average or peak of the monitoring period. Another limitation is the threshold

To see a trend chart, double click on it.

Gauge

This is a variant of scoreboard. It can show relative comparison. You can use another metric as the "total". This means you do not have to create a super metric.

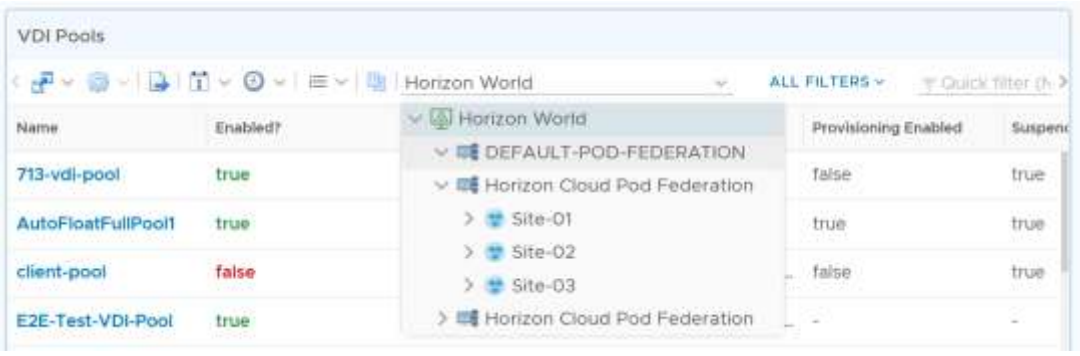


You can also change the color coding of the threshold based on percentage.

Table

A table is simply a list, where each row represents an object, and each column shows a *single value*. This enables us to list hundreds of rows, with ability to filter and sort. Each cell value can also be color coded.

View List also supports filter. The following screenshot shows Omnissa Horizon hierarchy, so you can drill down into specific area in the world of Horizon.



Aggregation

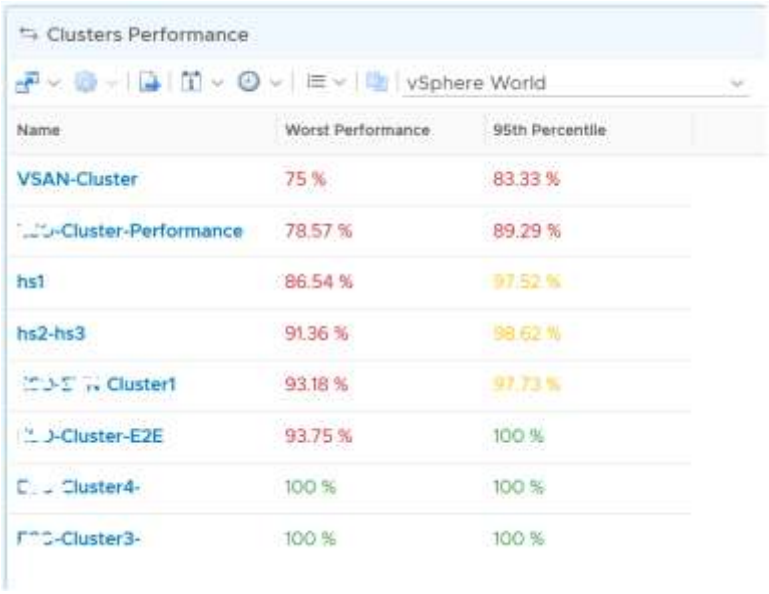
Table is good for details. However, as a *summary*, its main problem is how to give an insight over time as each cell can only hold *1 value*. How to give an insight into what happens in the past? For example, how to see the performance in the last 1 week? There are 2016 datapoints in the last 7 days, which one do you pick to represent?

There are a few possible options for the values:

Current	This could the present value or the last known value. It's useful to show the present situation. However, this does not tell what happened 5 minutes ago. Its useful in capacity, compliance, configuration. It's rather limited in performance.
Average	The average of the period. Average is a Lagging Indicator. By the time the average is bad, roughly 50% of the number is unlikely to be good. It is not suitable for proactive monitoring.
Worst	The worst of the period. This is suitable for *daily average*, as there are only 288 data points in 24 hours. If you find that results in outlier, then replace it with 99th percentile. As it only takes 1 peak to set this value, your chance of outlier is 7x higher in a week. It's great for peak detection, but needs to be complemented with worst 5th percentile when looking at weekly or monthly period
Percentile	The worst 5th percentile number. This is a good midpoint between Average and Worst. For performance monitoring, the worst 5th percentile is a better summary than average. BTW, we use the org.apache.commons.math3.stat.StatUtils library. The number differ slightly to Microsoft Excel or Google Sheet in general as they use a different formula.

From the above, we should choose worst and worst 5th percentile.

The following table implements the above concept.

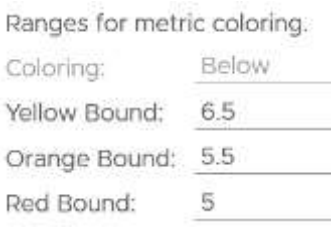


Coloring

BTW, I vrealize that the threshold coloring works on string, so long it contains number. The following table shows that any version higher than 6.7 is green. The rest follows the logic shows next to this table.

Name	Present Compliance	Lowest Compliance	Version
evn1-...	0 %	0 %	6.5
evn1-...	0 %	0 %	6.5
bos1-...	45 %	45 %	7.0.2.0
bos1-...	45 %	45 %	7.0.2.0
10.186...	45 %	45 %	6.7.3
10.186...	45 %	45 %	6.7.3

To do that, set the following:



It also works on text.

Allow Promiscuous	Forged Transmits	MAC Address Changes
-	-	-
true	true	true
true	true	true
true	true	true
true	true	true
false	false	false
false	false	false
false	false	false

Yes, you do not have to specify every color!

Coloring:	Below
Yellow Bound:	Yellow Bound
Orange Bound:	Orange Bound
Red Bound:	true

You cannot specify the green color. What is not red is automatically shown as green.

Calendar Summary

Name	Calendar Month	Count of Total	Count of Running VMs
vSphere World	March 2024	1,324	1,132
vSphere World	April 2024	1,283	1,116
vSphere World	May 2024	1,383	1,170
vSphere World	June 2024	1,423	1,207
vSphere World	July 2024	1,388	1,165
vSphere World	August 2024	1,428	1,201

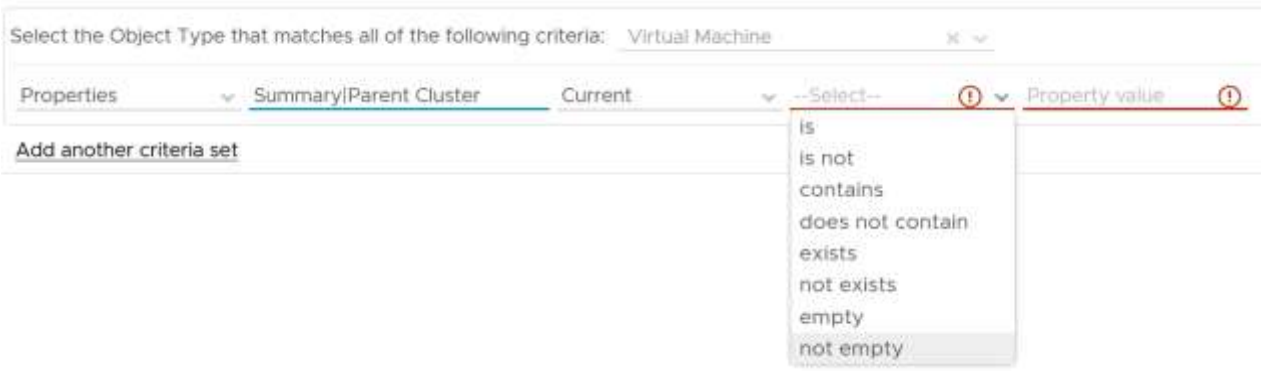
30 Mar 2024, 22:19 - 30 Sept 2024, 22:19 1 - 7 of 7 items

By the way, a little trick from Robert Mesropyan. There are times where you need to hide the object name. To do that, export the view, and change the value below to true.

```
<Property name="hideObjectNameColumn" value="true"/>
```

Empty or Exist

In configuration check, we can encounter a need to check for exception. This is where the Exist or Not Exist check and Empty and Not Empty check can come in handy.



Distribution Chart

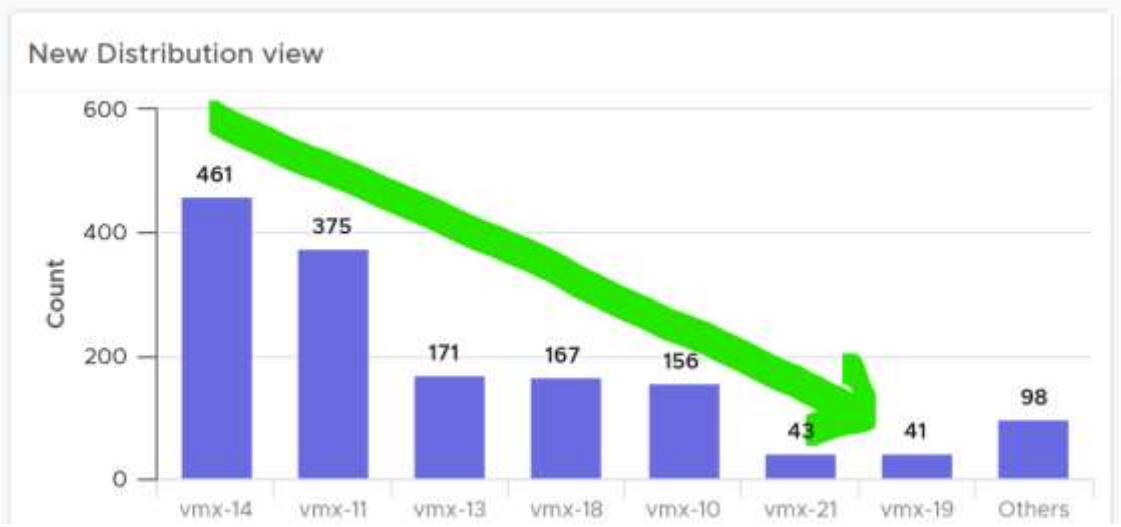
Distribution Charts can be used to give insight to a large dataset, as table will show too many rows.

There are 2 types of distribution charts:

- Absolute distribution: bar chart.
  It's called absolute as you care about the actual value of each bar.
  You also call about the order of how they are sequenced. The brackets are linear and follow a step up.
- Relative distribution: pie chart and donut chart.
  It's called relative as the number you care is in percentage or ratio. You care about how the numbers are relative to one another. You also do not care about the order they are displayed (the brackets are not in order)

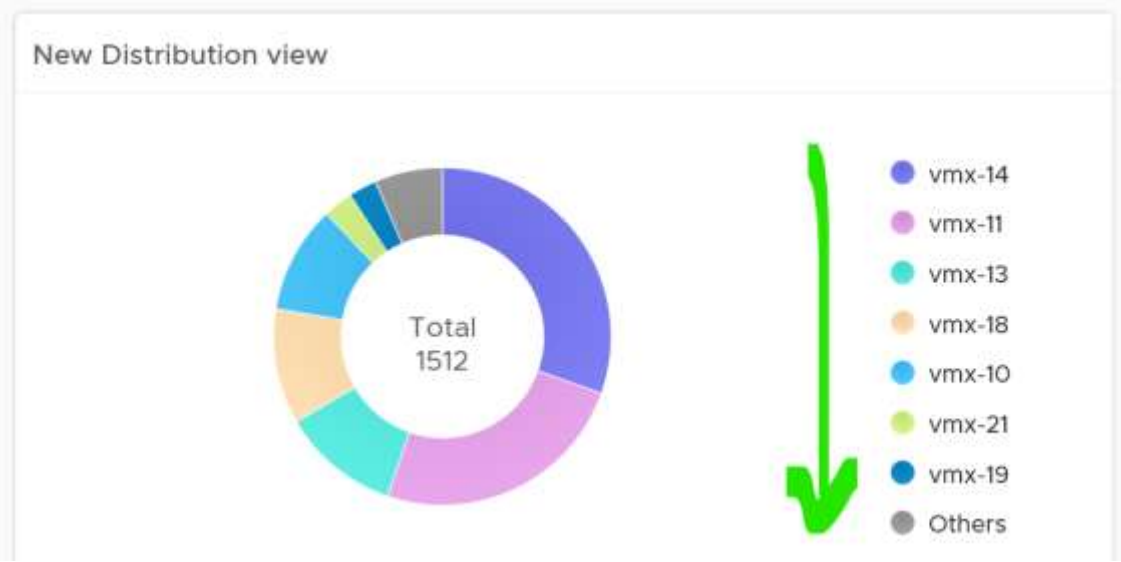
The number of buckets on the chart should be balanced between the available screen estate, ease of use and functionality. Modify the buckets to either reflect your current situation or your desired ideal state.

The chart automatically starts with the largest slice. It ends with "Others", which aggregates all the remaining buckets.



How does it work with a pie chart or donut chart?

Well, it starts at 12 o'clock, goes clockwise and end with "Others".



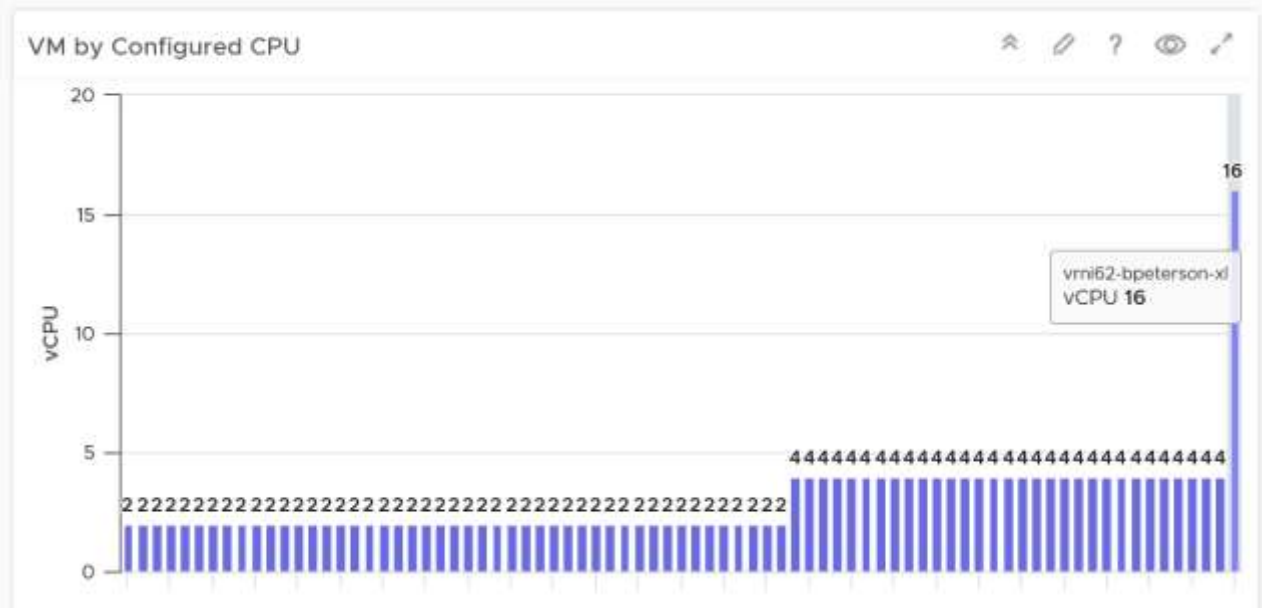
Bar Chart

When the order matters, use bar chart to make sure each bucket is meaningful. Take the following as an example. It shows vSphere shared datastores by their capacity remaining. They are categorised into 5 buckets, from the lowest capacity remaining to the highest. Each bucket is given a color to convey a meaning. Can you guess why >80% is represented by grey, as opposed to even more green?

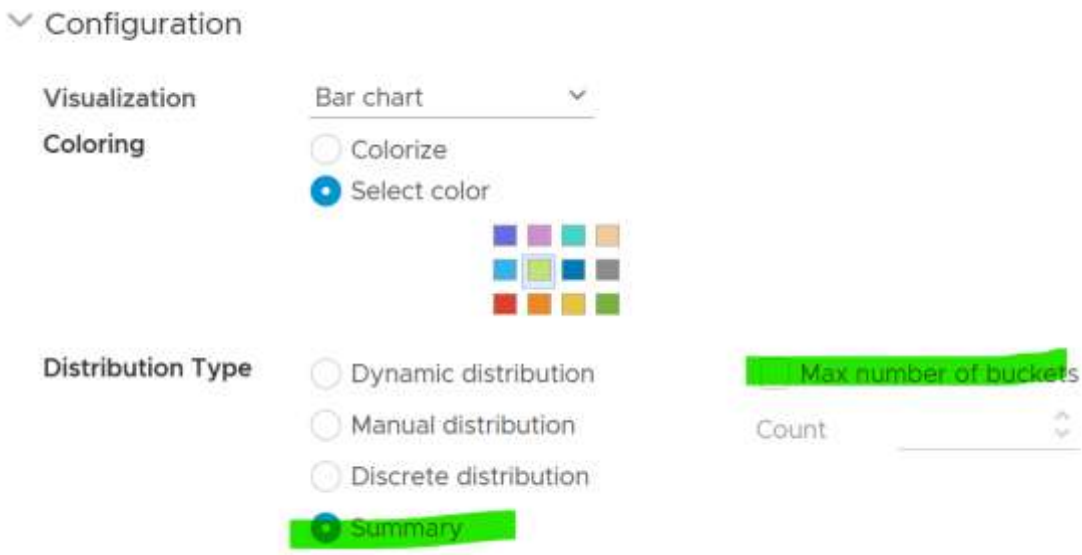


The reason has to do with the reason you bought the capacity in the first place. It's to be consumed. If the capacity is not well used after months or even years, it's a wastage. You overbought capacity.

The bar chart is also better at showing the outlier. The following cluster has 79 VM. If you want to know their distribution by size, you can plot all VMs in order of their vCPU size.

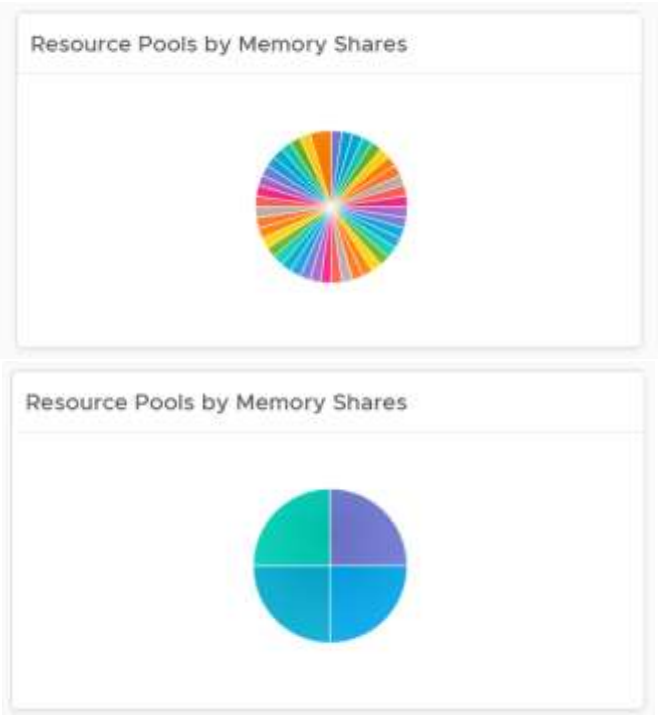


The distribution type is "Summary", and leave the "Max number of buckets" blank.



Pie Chart

When the number of variants matters, use pie chart as it does not have space constraint. It's useful when you have an expectation of what is good or bad. See the following 2 widgets. They show the number of resource pools in a cluster. What does each pie chart tell you?



The first cluster has way too many resource pools. This alone will make performance management difficult.

The second cluster has 4 resource pools, but they are of equal share. What's the point of having resource pool when they are the same share? It looks like a common mistake of using resource pool as folder.

Summary

This widget can do 2-level of summary

- First, the data is summarized over time.
- Second, the data is summarized over members of the group.

The first summarization is optional. In the following example, it simply gets the VM configured size for each VM.

The widget then aggregates 3 numbers: the average size of all VMs, the largest size, and the smallest size.

Average VM

	Average	Largest	Smallest
vCPU	6.49	16	1
Memory	35.7 GB	128 GB	0.16 GB
Disk Space	0.34 TB	1.05 TB	0.0098 TB
vDisk	2.46	17	1

You can also add standard deviation to see the size of the distribution.

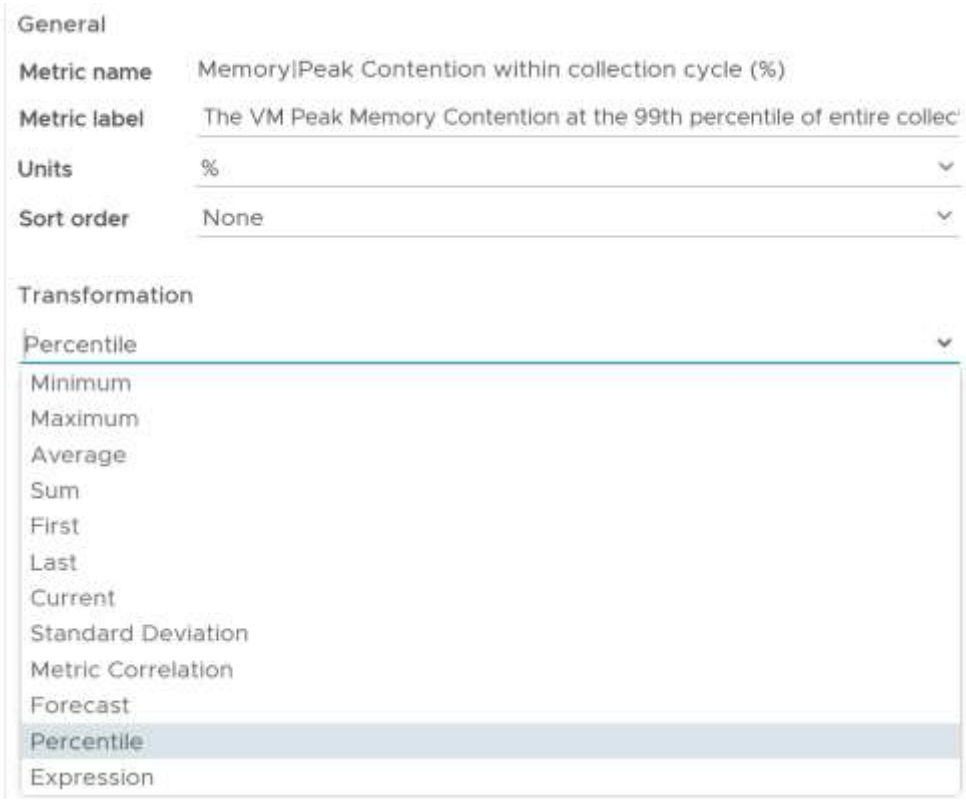
Advance Example

You can use the widget to perform advance analysis, such as baseline performance profiling.

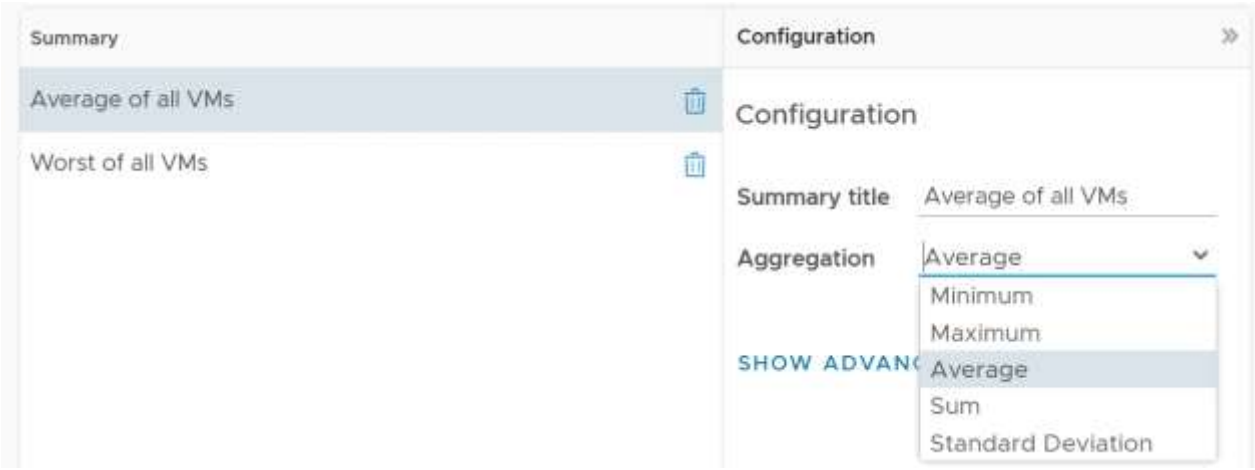
In the example below, we first take the 95th percentile of each VM contention. This is then used as the input to get the overall average, and the worst, among all the VMs.



The 99th percentile is set on each row.



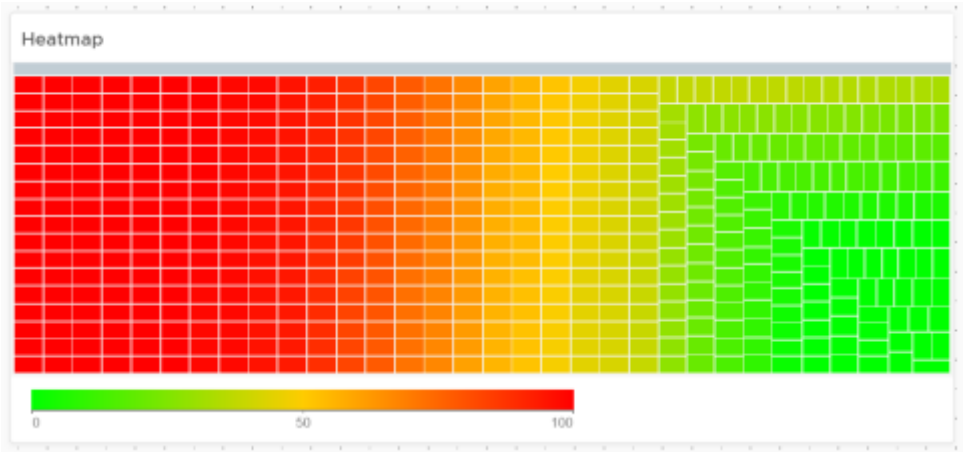
The 2 columns (average and worst) are actually summary



Limitation: there is no function such as count and percentile.

Heat Map

Heat Map lets you vary both the color and the size of each object. Fixing the size can be powerful as it shows a different meaning. From the following example, you can see that there are more red than orange + yellow combined. It also makes the overall layout neater, which is always a nice bonus.

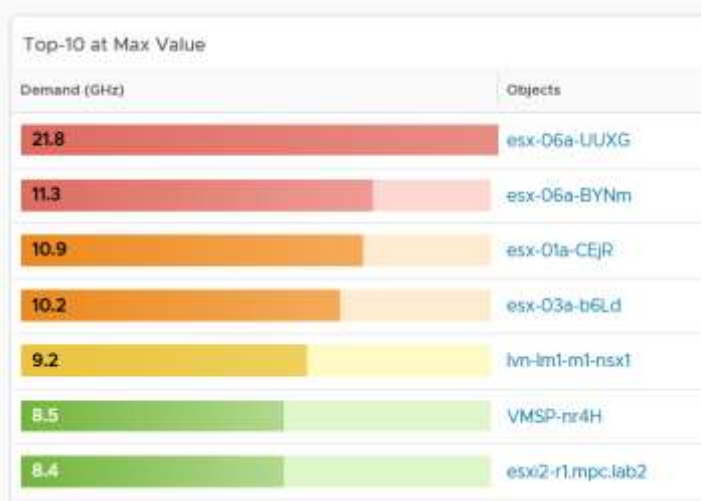


Fixing the size also enable you to use multiple heat map widgets together. The reason is the exact location and shape of each box is identical across widgets. In the following example, the red dot comes from the same ESXi host. This means you can see while this server across the 4 different metrics.



Top N

Top N is a simple list with visual. It is ideal when you need to convey the value before the object name. As you can see below, the presentation is more eye catching than View List table.



Relative to View List, Top-N has limitation on data transformation. By default, it will average the value. Since average is a lagging indicator, use the percentile function. As a bonus, the percentile is not locked in the design mode. Any user can change the percentile on the fly!

Limitations

Distribution charts (pie chart, donut chart, bar chart) have the following limitations:

- "`No data to display`" does not imply that there is something wrong with VCF Operations data collection process. It might signify that none of the objects meets the filtering criteria of the widget, hence there is really nothing to display.
- To see the content of a slice in a pie-chart or a bucket in a bar-chart, simply click on it. Note that the list cannot be exported.
- The pie-chart and bar-chart cannot drive other widgets. For example, you cannot select one of the pie-slices or buckets, and expect it to act as a filter to a list or a table. What you can do, is to select one of the object in the slice. I use this technique to allow drill down. You can also click on the object name, which will take you to the object summary page. The page provides key configuration information, alongside other summary information.
- You can apply a specific color in a pie chart or distribution chart for a specific numeric value, but not string value. For example, you can't apply a red color to the value "Not Installed".

The following widget can only show the present data:

- Heat Map
- Scoreboard

Log Insight. Unable to specify color in the bar chart. The color is automatically chosen. It can result in "red" in the legend being shown in green.

Implementation

Ok, enough of theory 😊 . How do you apply all those in real world operations management when there are many systems to manage and many roles are involved?

Architecting a suite of dashboards spanning a wide area is complex. To address this, begin with the end in mind. Different team in a large organisation are interested in different things. Once you have the standards, you can personalize as individuals within a team have specific ways of doing his or her job.

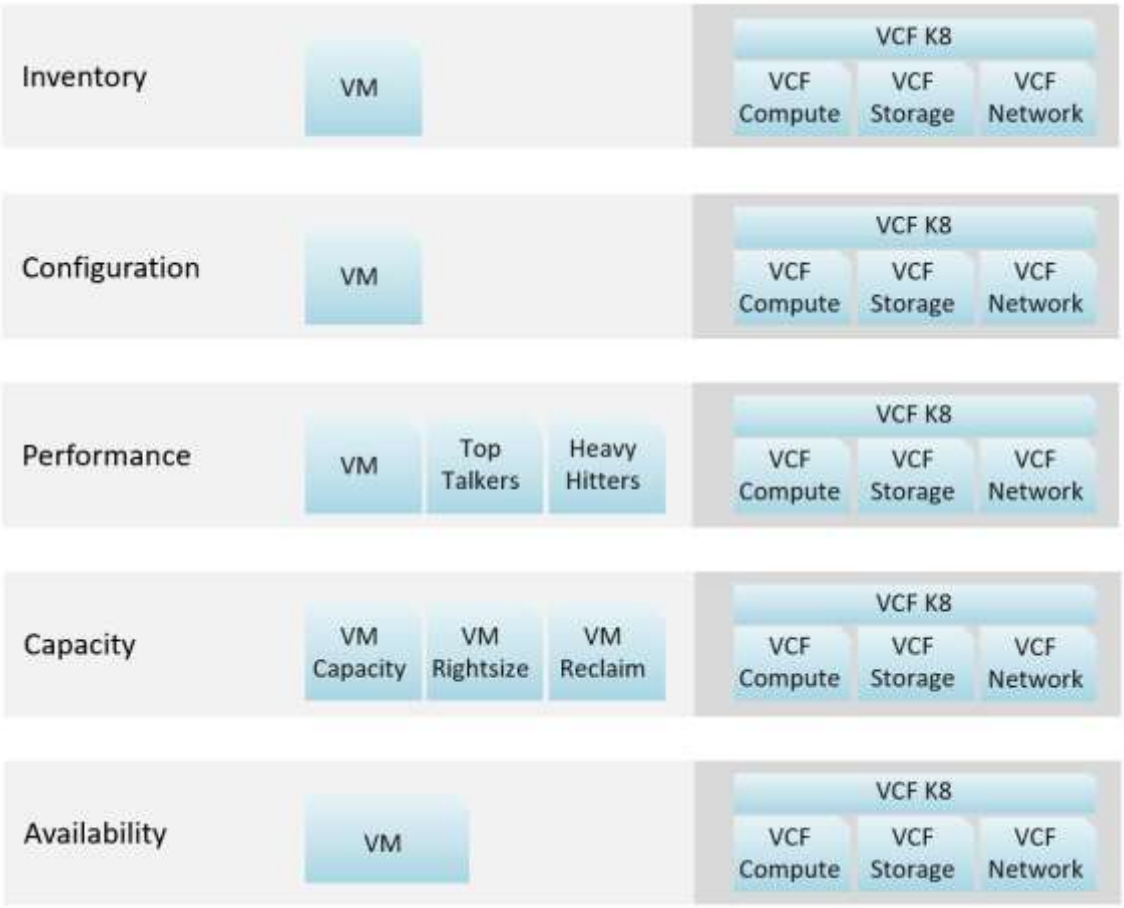
Each operation is like a fingerprint. While there are commonalities, each customer runs their operations a little differently. Hence it's not possible to design one dashboard that meet every customer's operational needs. A configuration that is important for one may not even be relevant for another. Tailor this dashboard to your unique environment. If needed, widgets can be collapsed or expanded allowing more relevant data to be displayed.

In VCF Operations, the primary role we target in our default dashboards is the VMware team. They are often called the Platform Team, especially if they run vSAN and own the Guest OS that runs on top of vSphere. The team supports the Level 1 Team and performs troubleshooting. That's why the default dashboards span both monitoring and troubleshooting. They are interested in inventory, performance, availability and configuration.

Role-based	Platform Team Capacity team: Capacity and Reclamation Compliance team: Compliance and Configuration IT Finance team: Cost and Price Level 1 team: Live NOC, Alerts, Availability, Security IT Leadership: Summary, long term, big picture.
Pillar-based	Develop the following set of dashboards, each cover a pillar of operations • Inventory. What do I have where? Is the inventory movement & volatility matching expectation of the day? • Configuration. Any important settings that are incorrect, inconsistent, outdated, or suboptimal? • Availability This is the foundation. If the system is down, you don't have performance nor security problem. • Performance. Any VMs that is not being served as promised by our shared IaaS platform? What's causing the issue? • Capacity Are we utilizing our capacity well? How much can we reclaim? When must we buy additional capacity?
System-based	Each system has their own architecture; hence their observability varies. Popular systems are: • VMware vSphere • VMware vSAN • VMware NSX • Microsoft Windows

Overall Design

Once you consider all the above, you will come up with a suite of dashboards that work like one.



Explanation:

- VCF Network = NSX + Distributed Switch
- VCF Storage = vSAN + VMFS + RDM. We keep SRM separate as it's separate SKU.
- VCF Compute = vSphere Cluster + ESXi + Resource Pool.
- VCF K8 = vSphere Supervisor.

How about Kubernetes workload?

My take is they need their own set. Just like VDI needs their own set, as these 2 technologies have many object types.

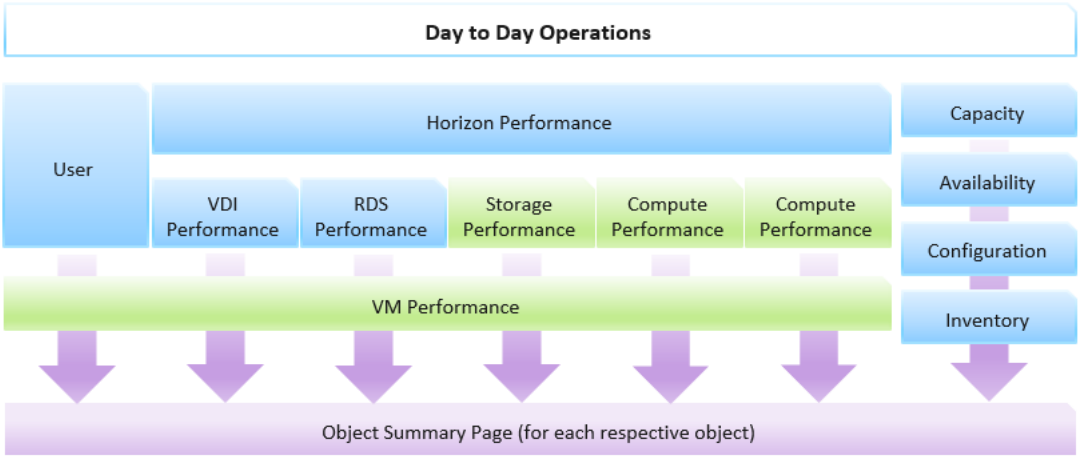
You need to have some consistency for dashboards that covers, so users need not learn repeatedly. It will be confusing if each dashboard looks totally different from one another, considering they have the same objective. Apply some design standards for different pillars of operations. After all, performance is performance, whether it’s Kubernetes or Amazon Web Services is a matter of objects and metrics.

Multi-Dashboard

Dashboard to dashboard (D2D) navigation has to be part of your overall approach. You can drill down in the object hierarchy from one dashboard to another, or move laterally. Note that you cannot go up the parent hierarchy.

D2D capability also helps in avoiding a deep dashboard with many widgets that requires multiple pages of scrolling. They are harder to understand and may suffer from loading time.

Create a set of dashboards that act like one system. This is the principle we use when creating the out of the box dashboards. The following shows an example for Horizon, a VDI solution. It integrates into existing vSphere dashboard by enabling drill down.



Focus on the pillars of operations, not objects.

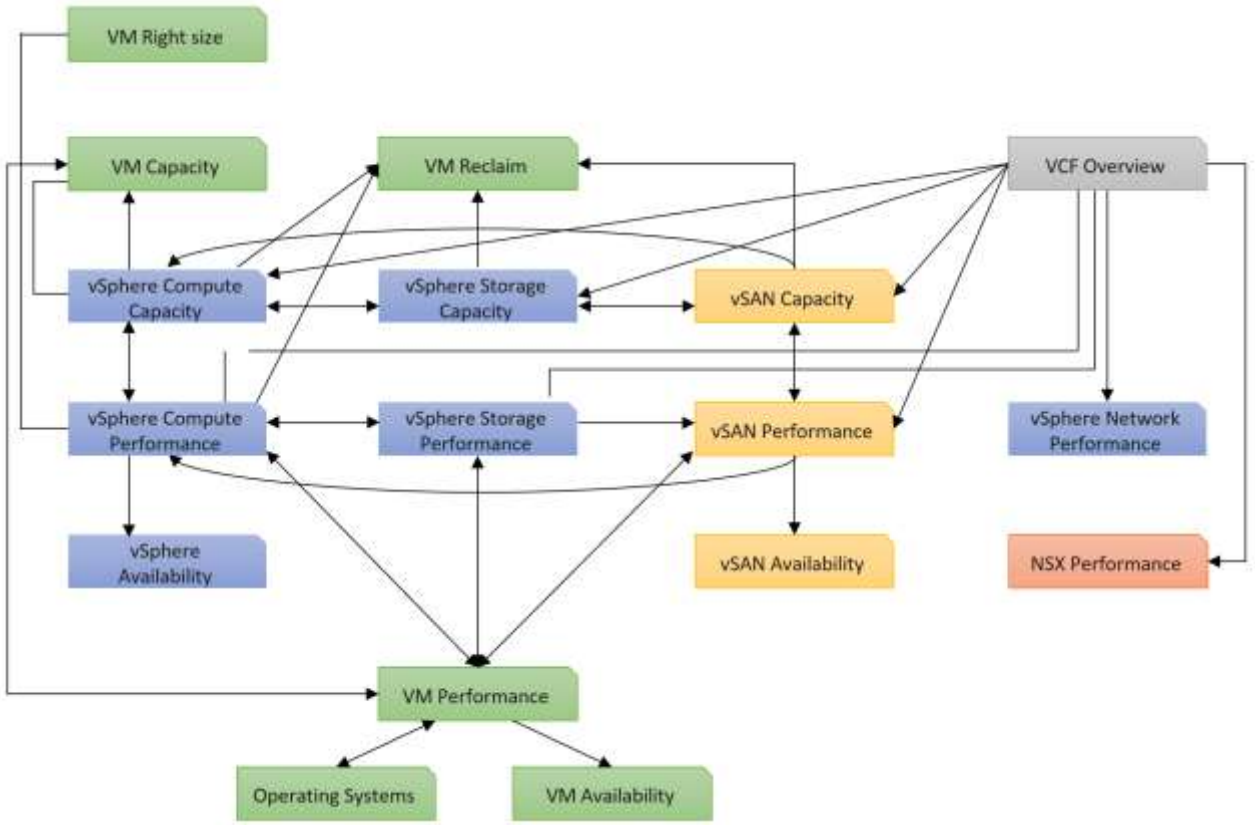
Let's say you're designing for Capacity management and Performance management, as they often go together. Look at the systems being managed and the major tasks. For the case of vSphere and vSAN, we can group the components into Consumer and Provider. For each, we need to keep the lights on (Operate) and improve (Optimize). Operate covers the capacity and performance management, while Optimize covers reclamation, rightsizing and upgrade. Optimize also applies to infrastructure. You adjust cluster size, upgrade ESXi, migrate from traditional array to vSAN. All these increase efficiencies, reduce cost, increase capacity, etc.

You may end up with something like this as your initial overall design.

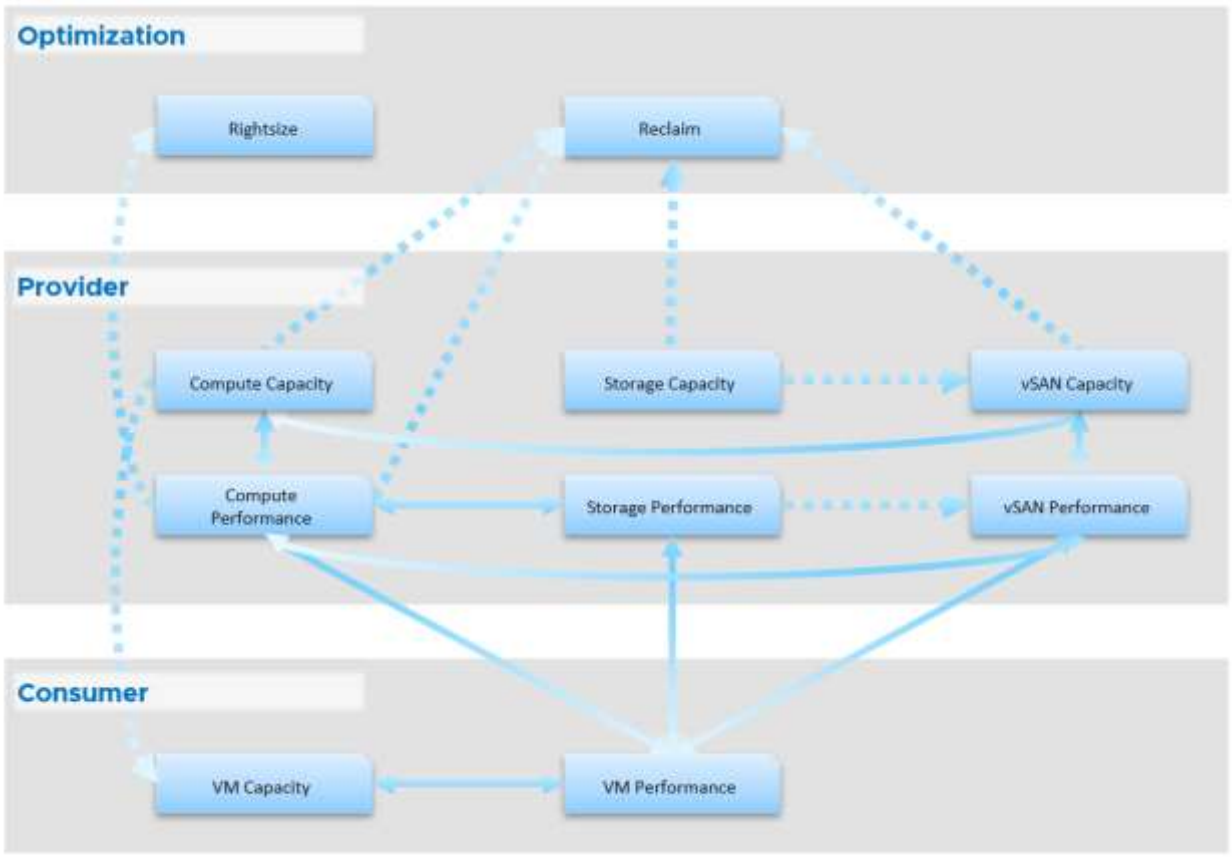


Once you have an overall approach, you can plan your dashboards. Because vSAN has many metrics and components that need further check than a simple datastore, I need to give it its own set of dashboards. There is a drill down from Storage to vSAN as not all datastores are vSAN. By keeping them separate, you can also add physical array dashboards, so your Storage dashboard can drill down to both vSAN and array.

Take time to draw your multi-dashboard solution. I found it helps clarifying my thought. The following example shows an attempt to connect all products, not just vSphere and vSAN. It will be interesting to connect VCF, VMware Clouds on AWS, NSX, Horizon, vSAN and vSphere together!



You can and should also organise by function, not just by VMware products. In the following example, I group them into 2 (Provider and Consumer). I added Optimization as the lines became too complex. By working at this logical level, it also helps me spot if there is inconsistency. Having gone multiple rounds, I ended up with something like this.



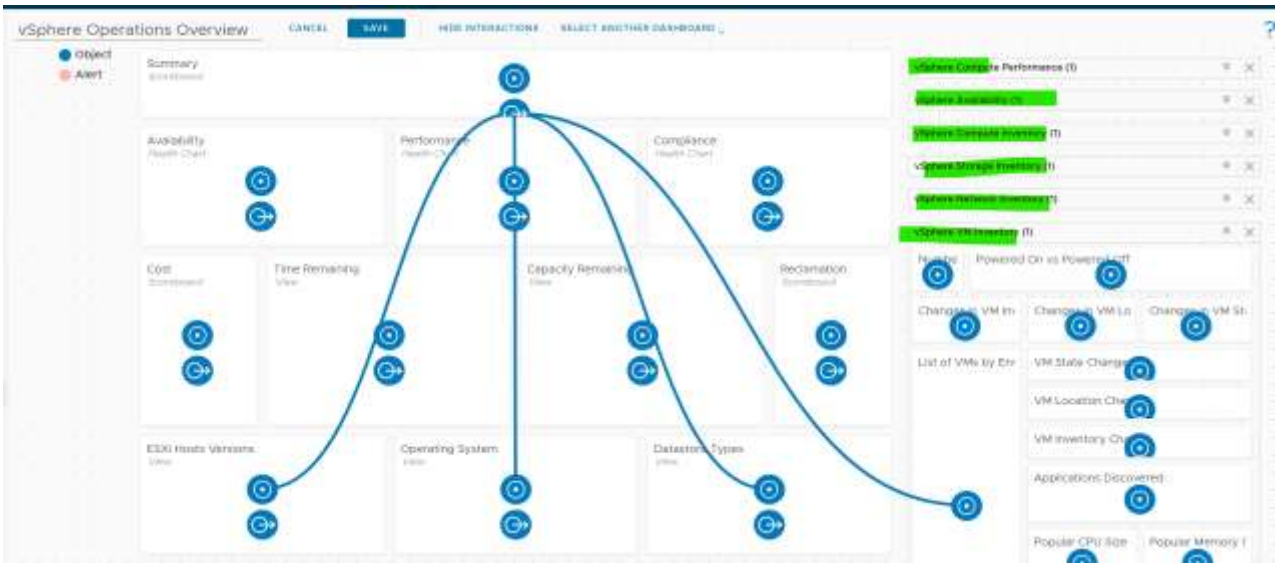
Notice something missing?

I do not have Network. I cover in Part 2 Chapter 5 Network Metric that the nature of network means its capacity model is largely a non issue. You focus more on performance as you do not tolerate network dropped in your data center. I plan to include Network Performance after I consider integrating NSX. If you have an idea on how to enhance the current dashboard, let's collaborate!

From the above, why are ESXi and Resource Pool missing?

The reason is they are part of cluster. Instead of separate dashboard, I feel adding it into the cluster dashboard will increase usability. The drawback is it does not cover standalone ESXi.

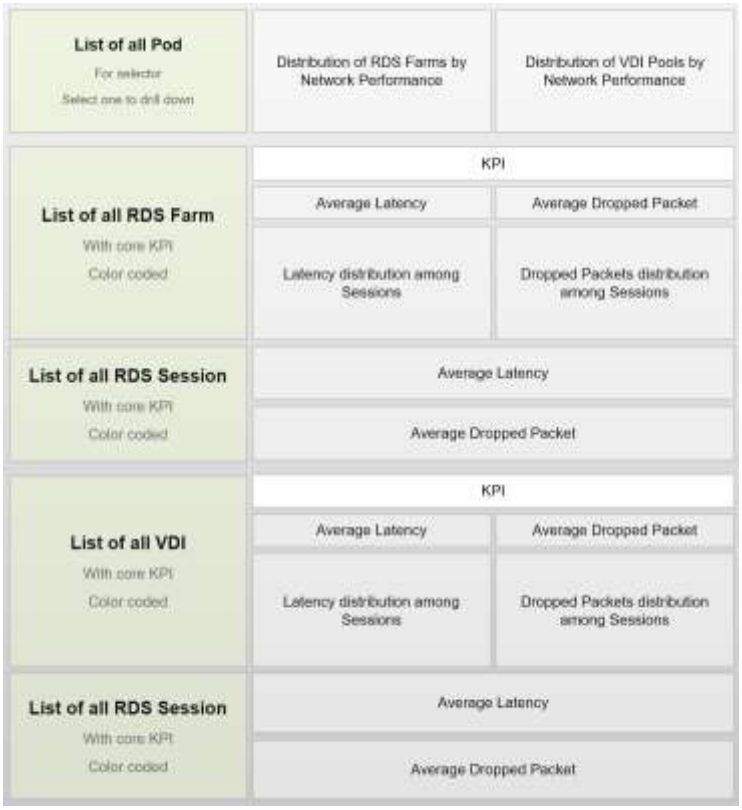
Once you design the overall flow, you implement it on each dashboard. Here is an example, where this dashboard drills down into 6 other dashboards.



The screenshot shows the drilldown 1 dashboard at a time, hence I chose the last dashboard else the first 5 could not be captured.

Logical Design

I find it useful to design my dashboards in Microsoft PowerPoint before I implement them, especially if I need to design a suite of dashboards that work together with back-and-forth relationship. I find it helps me see the big picture, as I can see multiple dashboards on one screen and spot any inconsistency among them. It's also faster to make changes.



Large Environment

The number of dashboards you will have depends on the size of the environment and the number of people managing it. An environment with 100 VMs in just 5 hosts and 1 cluster will need far fewer dashboards than a global environment with 100,000 VMs spread over 5,000 ESXi, 500 clusters, 20 data centers, and 15 vCenter Servers.

In a large environment, where you have many physical data centers and even more vSphere clusters, you will likely need to display the information per physical data center. There are several reasons for this:

- Aggregating data at a global level, which spans many physical data centers, will hide too much information. Presenting data at such a level means you are getting an average of thousands of objects. If your environment is generally healthy (and it should be), the average will logically fall within a healthy range.
- In most cases, the performance in a given physical data center is independent from that of other data centers. For example, your Singapore data center typically does not impact the performance of your London data center. An exception to this case is when you link your data center at the network (stretched L2) and storage layers (synchronous replication). From experience in troubleshooting such a scenario, we recommend you keep the physical layer independent from each other. Assuming your data centers are independent, it makes more sense to display the chart on a per data center basis.
- VMs typically do not move from one physical data center to another (unless they are paired with storage replication and your network is stretched), so an imbalance among multiple data centers does not translate into a realistic rebalancing action.

Inventory

Part 2 Chapter 3

The inventory dashboards aims to implement the inventory pages of VCF Operations.

Overall Design

The suite of dashboard works together as one integrated set. They also have similar design.

Goals	Check if the inventory matches your expectation. Not just the count but also the change.
	Cover all objects in VCF where inventory matters.
Questions	What VMs do I have? Where are they located? What's the changes I need to know?
Assumptions	To be used with configuration dashboards
Target Users	Platform Team
Usage frequency	Mostly ad hoc, on a need basis Daily for VM inventory changes.
Features	Visual relationship of objects.
	Ability to traverse the inventory hierarchy quickly
	Show the hierarchy and count of objects in the hierarchy. Ideally, it shows the movement of the inventory.

VM Inventory

VM is more volatile than infrastructure, so it's important to track their changes.

Template is not counted as VM. From capacity point of view, it does not consume CPU and memory resources. From cost point of view, it does not have a price as it's considered as part of IT asset.

The dashboard aims to answers questions such as:

- What do we have in the environment? How does it change over time?
- What's the ratio between running and not-running VMs in the environment?
- What's the ratio between Windows and Linux in the environment? How does it map to your licensing and ELA?
- Do you have to many variants of Windows OS and Linux OS? Too many different versions, editions, and updates can complicate operations.
- The dynamic nature of private cloud means change is constant. Is the volatility matching your expectation and business cycles? For VM, there are 3 types of changes (location, state and inventory). They are tracked in 3 separate health charts. Tailor the threshold accordingly for ease of report analysis.
- How many VMs were deleted? How many were added?

- Are you having many large VMs? Are the relative size of CPU and memory balance? A CPU intensive VM should have high CPU:Memory ratio, while a memory intensive VM should have lots of memory relative to its CPU.
- What's the average VM size? How does it compare to the largest size? What are the popular CPU size and memory size?
- What are the largest VMs in your environment? Why do these VMs need that many resources? Is the list matching your expectations?

Summary section

The summary section consists of 2 rows of widgets. It's designed to help you see the big picture.

Check that these numbers are within your expectation.



If the number of powered off VM is relatively high, navigate to the VM Reclamation dashboard to see if they can be deleted and taken off the systems.

Review if the numbers over time match your expectation. For example, if there was an activity last night that impacted the number of powered off VM, yet the actual number differs, perhaps the activity was not executed as per plan.



The 2nd row of widget consists of 3 charts. They summarize the overall movement of all the VMs in your environment. Make sure the amount and the pattern both match your expectation.

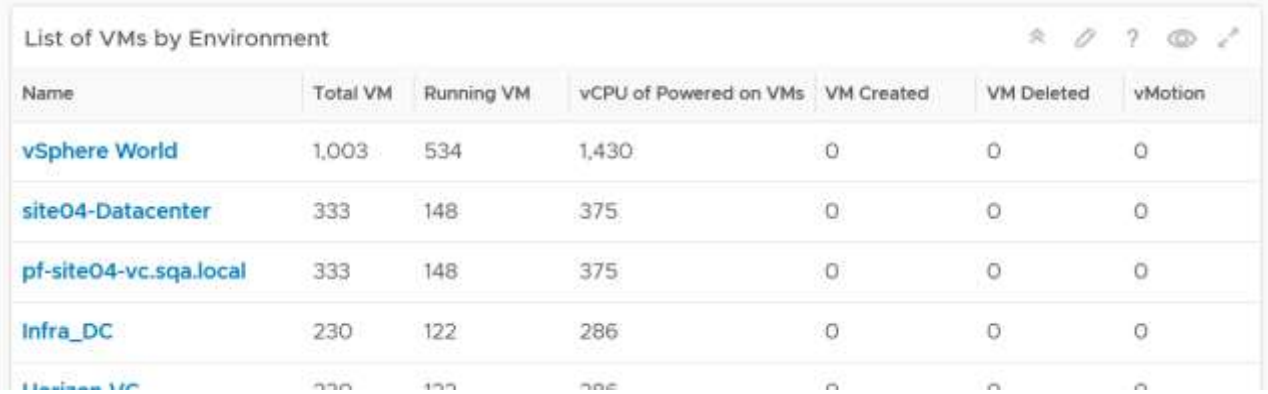


You should customize the thresholds, so it's easier to see if the values exceed what's acceptable in your day to day operations. The value is every 5 minutes, so if 100 changes of states is not something you can ignore, set the threshold of yellow, orange and red accordingly.

Detail section

The detail section has interaction where you can zoom into specific data center or clusters.

The first row is always the total. vSphere World object covers all the clusters, so its numbers are the summation of all clusters.



Name	Total VM	Running VM	vCPU of Powered on VMs	VM Created	VM Deleted	vMotion
vSphere World	1,003	534	1,430	0	0	0
site04-Datacenter	333	148	375	0	0	0
pf-site04-vc.sqa.local	333	148	375	0	0	0
Infra_DC	230	122	286	0	0	0

To see a trend, click on any of the row. The 3 sets of scoreboards will automatically show the details.

The first scoreboard is about the availability state. Try to customize the threshold so it's easier to see at a glance. I've set the threshold for the VM Reset as that's an abnormal operations.

When you customize, consider the steady clusters. They will have a lower threshold as there are relatively less changes or movement. If you set a threshold based on your busy customers, these small clusters will tend to show green.

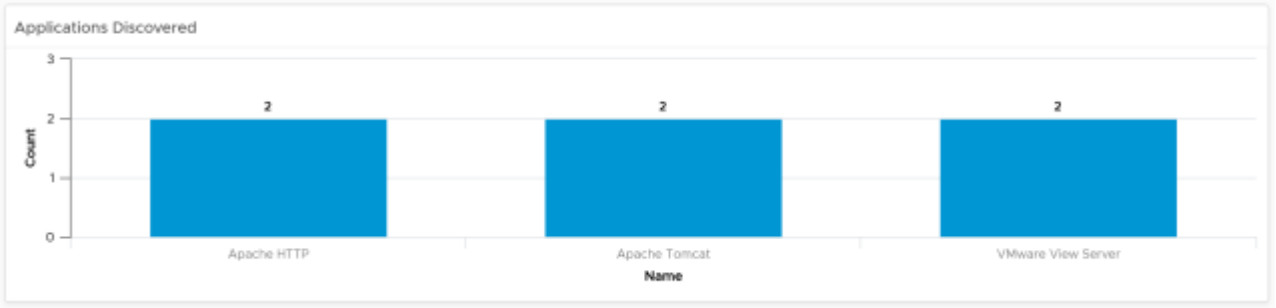


The next 2 scoreboards cover the location change and inventory change. Again, customize accordingly.



The dashboard lets you see the applications running inside the VM on any part of your environment. This can be handy if you have policy that certain applications should only run on certain clusters, for either cost or compliance reason.

Note: This requires Service Discovery adapter.

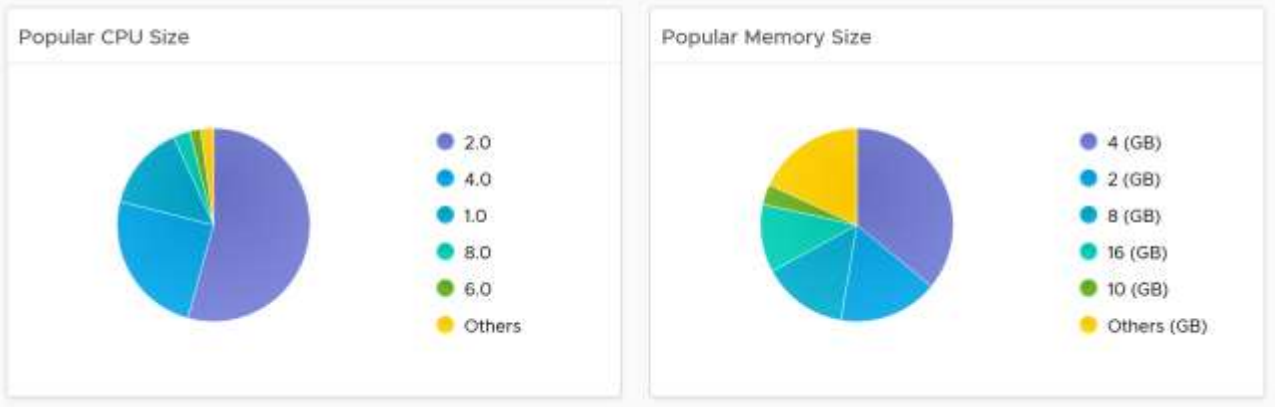


Popular VM Size

Since this is about inventory and not configuration, the following pie charts focus on the *popular* size (most common). If this was a configuration dashboard, we would have ordered them based on the configured size.

Make sure there is no odd sizes. They need to match your underlying CPU and memory.

Customize the pie charts if you need to see more than 5 sizes. I find 5 is a good balance since I can see the details in the next table below the pie charts).



Popular OS Distribution

The next pie chart shows the OS distribution, again sorted by the most popular.

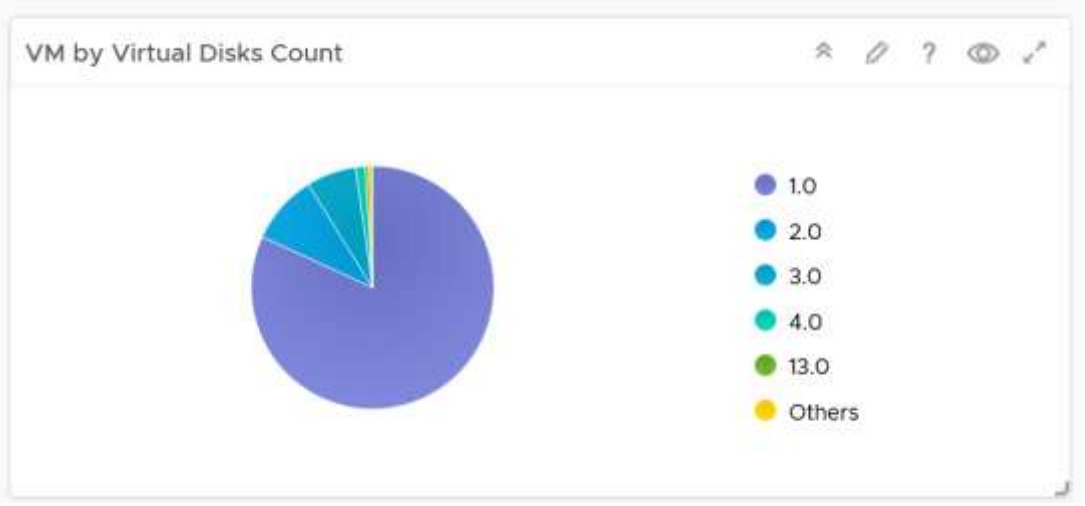
They should match your expectation. For example, if you separate Windows and Red Hat into different clusters, then you expect not to see one of them when you select a particular cluster.



If you see the value "none" under Operating System, that means the VM does not have Tools. I chose the value from Tools instead of vCenter as that's more accurate.

Virtual Disk

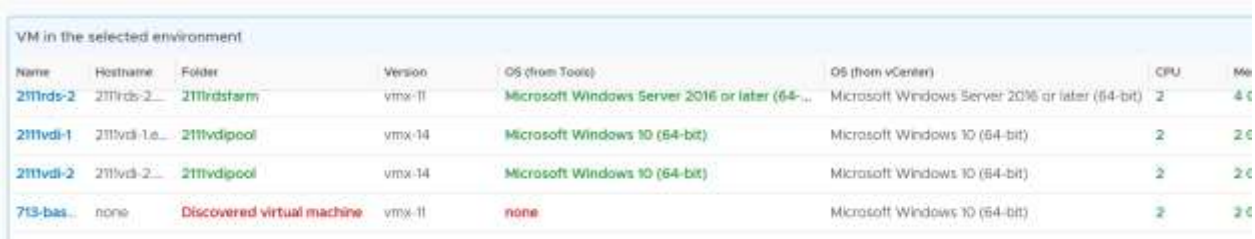
The last pie chart shows by the number of disks.



Since a VM needs to have at least 1 disk, I've excluded VM with just 1 virtual disk so those with many are more visible. You're welcome.

Individual VM Section

The last section shows the individual VMs, and you can drill down to see its details. This is why I keep the pie charts to just show 5 as this table lets me sort in any columns.



Select a VM, and you can drill down to see its capacity and performance.

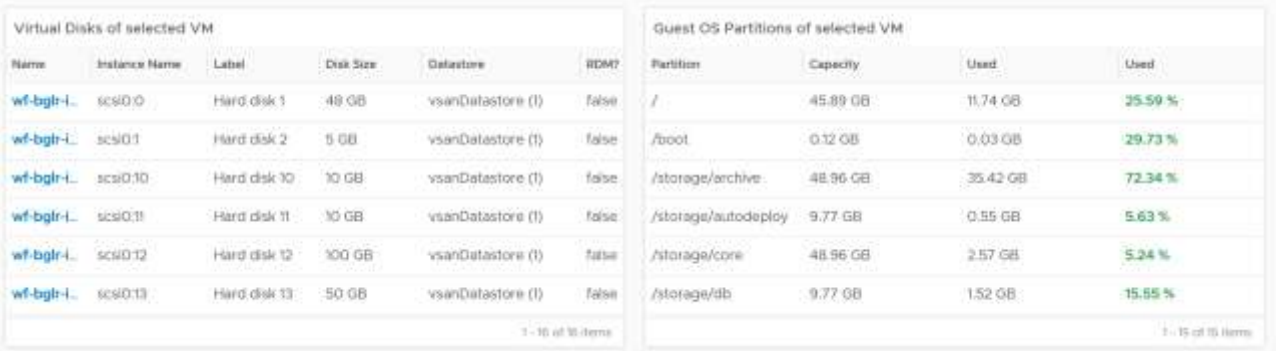
Add or remove columns to suit your environment.

Customize the table further by color-coding the values. I've color coded the following:

- VMs that are not properly placed in a folder. They will appear under the "Discovered virtual machine" folder.
- VMs that are not reporting its operating system. They will report the value "none".
- VMs with large CPU, memory or disk. Tailor this to your environment.

Select a VM, and its details of virtual disk, operating system partitions and network cards are automatically shown.

Limitation: It is not possible to see the mapping between virtual disk and partition. A partition or drive can spans multiple virtual disk, and a virtual disk can contain multiple partitions. This is why it's best to keep the relationship simple (1:1).



vSphere Inventory

The dashboard aims to answers questions such as:

- How are the VMs distributed across clusters?
- What are the largest clusters in your environment? Which clusters have the most VM?
- What are the largest datastore in your environment?

There are 3 dashboards

- Compute
- Storage
- Network

They have a fairly similar design. The compute has more details because it has more things to manage.

The design has 4 sections, as shown in the following screenshot



If you think their design copies the Horizon Inventory dashboard, you're right! That dashboard was basically the 2.0 of the vSphere inventory dashboards, so now you're seeing vSphere delivers the 3.0.

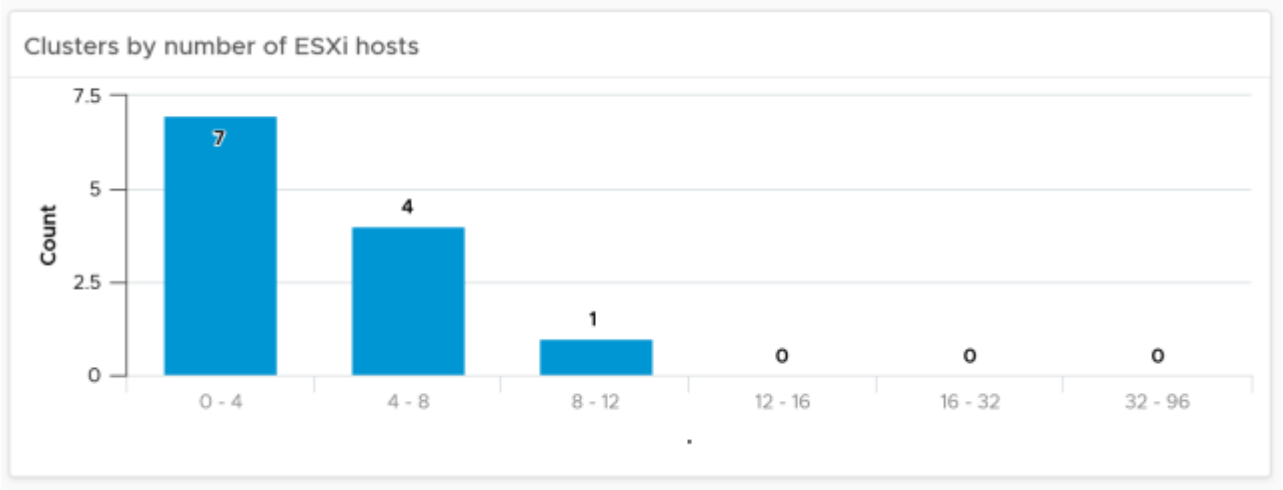
They can also navigate to one another. The relationship widget below sports the navigation. Pick an object and then choose the dashboard navigation.



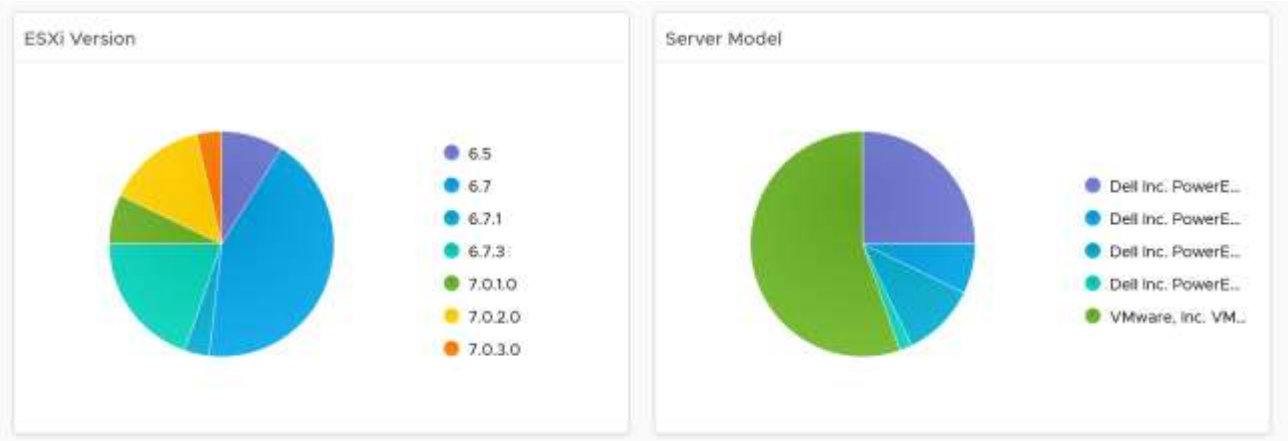
Compute Inventory

Let's use the Compute Inventory dashboard to represent the 3 dashboards.

The next row is a set of distribution chart. Due to limited space, I've provided only 2 charts, focusing on the cluster object. Since this inventory and not configuration, we should focus on the count of members.



The next row is a set of pie charts. Let me know if you have improvements on what the top 3 things you want to be shown as pie charts.



If you have larger screen real estate, add more pie chart for better insight into the type of items you have in your inventory.

Detail Section

The last row is always the lowest level of details, hence a table is chosen. For the compute inventory, we need to cover cluster, ESXi host and resource pools. That's why you notice 3 tables, one for cluster, one for ESXi and one for resource pools.



I added a navigation from the cluster table into the cluster performance and cluster capacity dashboard.

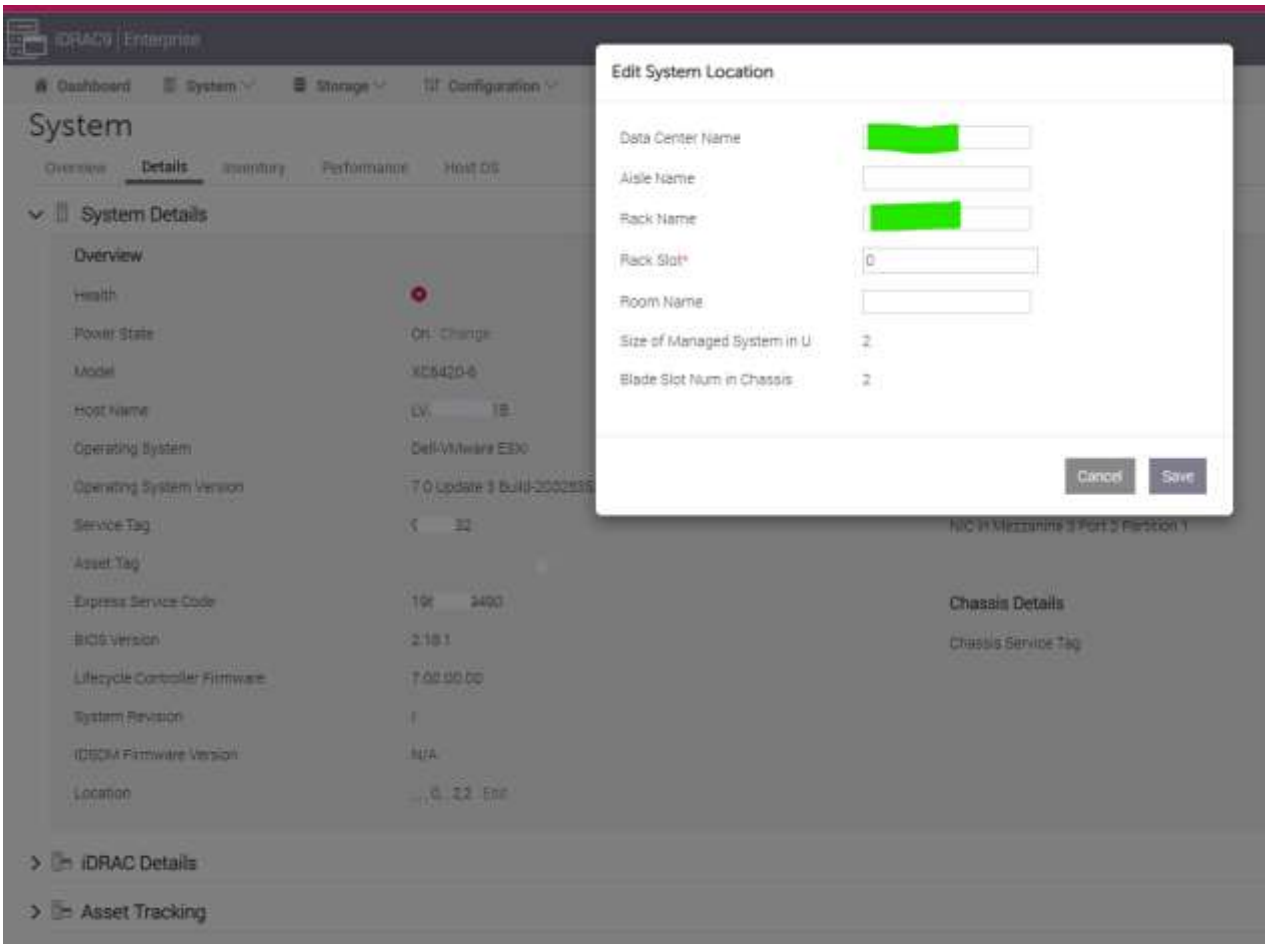
Customization tip

- A heat map showing clusters in the report scope. Size by the number of ESXi hosts, color by the number of VM. Alternatively, you fix the size and use the VM:Host ratio as the color.

Hardware Integration

In large environment with hundreds of ESXi, it is useful to be able to locate the physical location. To do that, integrate the hardware management with VCF Operations.

Ensure the rack name is unique, as you will use it to create the rack object in VCF Operations.


Edit System Location
Data Center Name
Aisle Name
Rack Name
Rack Slot*
Room Name
Size of Managed System in U
Blade Slot Num in Chassis
Cancel
Save
System
Details
System Details
Overview
Chassis Details
iDRAC Details
Asset Tracking

Performance

Part 2 Chapter 4

The performance dashboards aims to implement the concept covered in Performance Management chapter.

Overall Design

The suite of dashboards works together as one integrated set. They also have similar design.

Goals	Monitor performance
	Troubleshoot performance
Questions	Examples of questions answered by the dashboards:
	Are the VMs performing well? If not, which pods are affected by what problems (CPU, Disk, RAM, Network)?
	Is the VM performance caused by IaaS not serving it, or by contention within the Guest OS?
	Are the VMs running high utilization? If yes, which VMs, how high, and what resource (CPU, RAM, Disk, Network)?
	Are they really high relative to the underlying IaaS capacity? That can cause strain in the shared infrastructure.
Assumptions	They are designed for both day-to-day operations (proactive monitoring), and ad-hoc reactive troubleshooting. As a result, it's an advanced dashboard, not designed for Level 1 Support.
Target Users	VMware Administrators or Architects
Usage frequency	Daily for monitoring. This is to encourage daily usage, as what happens beyond 24 hours ago may be practically irrelevant from performance troubleshooting viewpoint. Hence the views are set to show data in the last 8 - 24 hours.
	Ad hoc for troubleshooting. Designed to be used within the day of the performance issue, or at most the next day.
Features	Color coded. Each of the color is according to the best practice of that counter. If you are unsure of what suitable numbers to set for your environment, profile the metrics. The Guest OS Performance Profiling dashboard provides an example of how to profile metrics. We also document the steps in Baseline Profiling section of the book.
	Design to see trend over time, and go back to any point in time. This is important as by the time you have the chance to look at the problem, 5 minutes have passed, or the problem is no longer happening

There are 2 types of performance troubleshooting

- Consumer. You focus on a single VM, container or application. Other consumers do not have a problem, or they are independent.
- Provider. You focus on the shared infrastructure as the problems impact many consumers. It could be hitting them at random, meaning different VMs/Applications got hit at different time.

At the infrastructure layer, we care whether it serves everyone well. Make sure that there is no contention for resource among all the VMs in the platform. Only when the infrastructure is clear from contention can we troubleshoot a particular VM. If the infrastructure is having a hard time serving majority of the VMs, there is no point troubleshooting a particular VM. Notice all the previous sentences are about the VM. Yes, the infrastructure metrics are *not* that relevant.

Logically split the dashboard into 3 parts:

Contention	This should be the first part, and most visible.
Consumption	Once you determine there is contention, use the **Utilization** portion to see if the contention is caused by *very* high utilization. When utilization exceeds 100%, performance can be negatively impacted especially when queue developed inside the Windows or Linux. By default, VCF Operations has a 5-minute collection interval. For 5 minutes, there may be 300 seconds worth of data points. If a spike is experienced for a few seconds, it may not be visible if the remaining of the 300 seconds is low utilization.
Configuration	Only include the relevant settings that can impact performance

This separation keeps each dashboard simple, while emphasizing the concept of contention as the primary counter for performance. You will notice in the dashboard design that contention is color coded, while utilization is not.

Take note that health chart is not ideal when the metrics do not have a "ceiling", meaning we don't know what "high value" is, as 100% is hard to define. Example of such metrics are disk IOPS and CPU context switch. If your operations team has a standard that utilization should not exceed a certain threshold, you can add that threshold value into the line chart. The threshold line will help less technical teams as they can see how the real value compares with the threshold.

For large vSphere environment, group the VM by clusters of the same class of service (e.g. Gold), so you can see the profile for each environment. I find the Data Center as good boundary. In general, storage, network and compute do not extend beyond a vSphere Data Center object. Performance problems tend to be isolated in a single physical environment, unless the WAN link connecting the 2 data centers are causing the problem. A performance problem in country A typically does not cause performance problem in country B, especially if they share little in common.

The dashboards share the same design principles, hence they are intentionally designed to be similar. It will be confusing if each dashboard looks totally different from one another, considering they have the same objective. To avoid repeating the explanation, read the VM Performance dashboard before reviewing others.

IaaS Performance Profiling

Why does this chapter begin with this dashboard? Why not start with VM performance dashboard?

Because you want to be proactive. Optimize your environment then it becomes easier to manage.

While performance is about what the VM experience, you can see it from VM perspective or the IaaS perspective.

VM Perspective

We covered how to baseline performance in PART 1 Chapter 2. So in this section, we will go straight into the dashboard.

Metrics to Use

Now that you know how to profile, what metrics do you choose?

CPU	We chose CPU Ready. It's good enough to represent the 4 types of CPU contention metrics as we're taking 20-second peak. I don't think it's worth the extra effort to include Co-stop, Overlap and Other Wait. You need to wait for 3 months as you need to create the super metric first.
Memory	We chose VM Memory Latency as that's the only true measurement of performance.
Disk	We chose VM Disk Latency. For profiling purpose, there is no need to split Read and Write. Overall is good enough as the improvement will be across the board.

There is no need to do network as you should not have dropped packets in the data center.

When you have time, profile the following metrics

- CPU Overlap. I expect the value to be <1%, with >90% of them below 0.25%.
- CPU Other Wait. I expect the value to be <1%, with >90% of them below 0.25%. Read the documentation of this metric as there is a false positive.

Metrics not to Use

You do not need to profile utilization metrics. For contention metrics, you do not have to profile less important metrics.

CPU Run Queue	These are Windows or Linux level metrics. They are application dependent, meaning not something the IaaS platform controls.
CPU Context Switch	
CPU Usage Disparity	This is also application dependent.
CPU Co-stop	This typically happens when the VM is oversized.
Disk Outstanding IO	This is impacted by IOPS, which is not something the IaaS platform control
Disk Queue Length	This depends on the storage driver. I recommend you use PVSCSI
Free Memory	It depends on the application. Certain applications such as JVM and DB manage their own memory, not something Windows or Linux can control.
Page-in rate	
Page-out rate	
VM Balloon	Not a VM performance counter, but an ESXi capacity counter.
VM Compressed	

VM Swapped	

Summary

How to summarize from millions of datapoints?

Here is the technique I use:



After that, we take the average and worst value among these VMs.

The following screenshot provides an example.

Summary of VM Performance SLA

	Average of all VMs	Worst of all VMs
The VM Peak Memory Contention at the 99th percentile of entire collection period	0 %	0 %
The VM Peak Disk Latency at the 99th percentile of entire collection period	7.53 ms	57 ms
The VM Peak CPU Ready at the 99th percentile of entire collection period	0.47 %	1.95 %

The 2 summary numbers become your **Before Optimization** numbers. Once you optimize the environment, revisit this table and take the new value.

The average columns should be fairly good. The worst of all them should not be too bad.

Here is my guidance:

		Average of all VMs	Worst of all VMs
CPU	Low:	< 2%	< 5%
	Normal:	2 – 6%	5 – 15%
	High:	> 6%	> 15%
Memory	Low:	< 1.5%	< 4%
	Normal:	1.5 – 4.5%	4 – 12%

	High:	> 4.5%	> 12%
Storage	Low:	< 10 ms	< 50 ms
	Normal:	10 – 30 ms	50 – 150 ms
	High:	> 30 ms	> 150 ms

The guidance is set with High = 3x Low.

The normal range is the widest range, and most environmental will fall here. Naturally, your gold cluster should be closer to the low range, while your bronze cluster will be closer to the high range. If your bronze cluster is showing very good performance, you can put more VMs to lower the overall cost per VM.

Likely, you see that the average is good but the worst is bad.

That means majority of the VMs are being served well, but there is a small percentage that are not getting served. You want to investigate which segment of your IaaS is not performing. It could be due to some configuration, or lack of capacity.

Details

If the summary is good, then there is no need to see the details.

What if the numbers are worse than expected?

One way is to use a bar chart. Design the buckets to represent the ranges from good to bad.

Generally speaking, for environment where performance is number 1, you want the value to be in the green to yellow range since you're taking the 99th percentile. For environment where cost is number 1, you likely have to lower expectation of your customers.

The following table shows the recommended buckets for CPU and memory.

Green	Yellow	Orange	Red
0 – 4%	4 – 8%	8 – 16%	> 16%

Depending on the result, adjust the bucket accordingly. For example, if your environment performance is worse than the above, yet nobody complains, you can adjust the above buckets to something like this:

Green	Yellow	Orange	Red
0 – 6%	6 – 12%	12 – 24%	> 24%

For disk latency, I recommend the following:

Green	Yellow	Orange	Red
0 – 25 ms	25 – 50 ms	50 – 100 ms	> 100 ms

Depending on the result, adjust the bucket accordingly. For example, if your environment performance is worse than the above, yet nobody complains, you can adjust the above buckets to something like this:

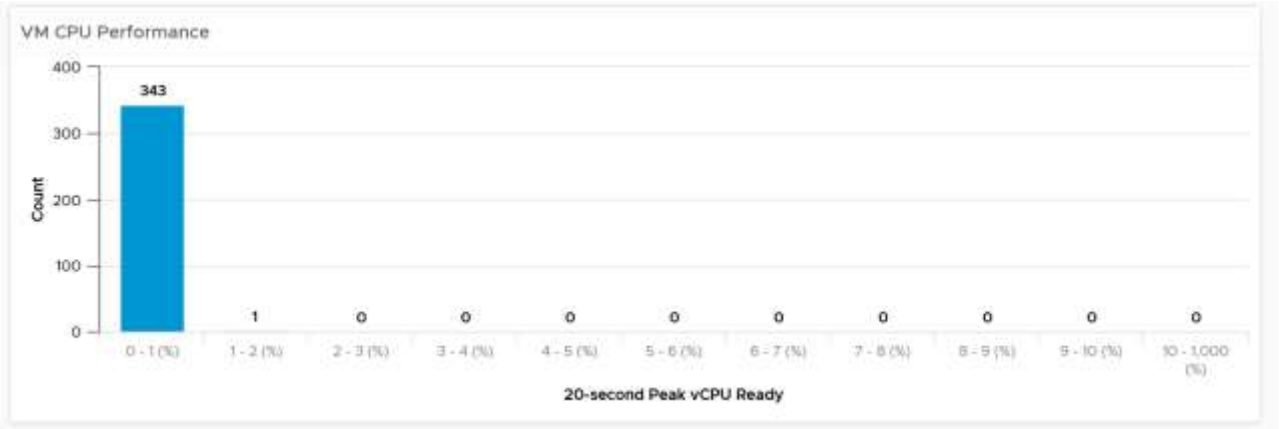
Green	Yellow	Orange	Red
0 – 60 ms	60 – 120 ms	120 – 240 ms	> 240 ms

If you have more screen real estate, do a split within each range for more granular visibility. Using the preceding table as example, you may split this way:

Green	Yellow	Orange	Red
0 – 30 ms	60 – 90 ms	120 – 180 ms	> 240 ms
30 – 60 ms	90 – 120 ms	180 – 240 ms	

Compute Profiling

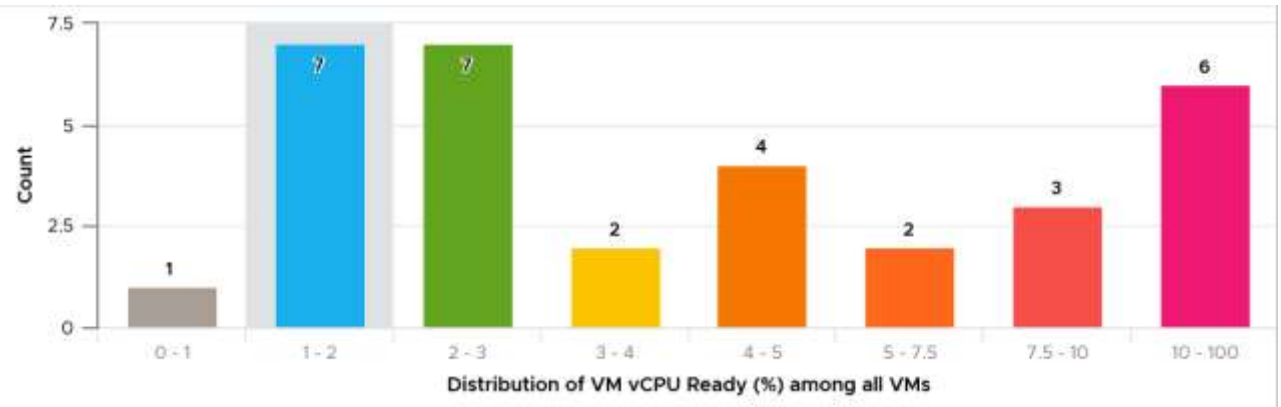
The following screenshot shows a sample profiling. What insight do you get from it?



The above shows an example of an ideal environment. All the 343 VMs experienced very low CPU ready time. It is possible that either utilization is too low or no overcommit. This means the relative cost is high.

Let's take a different example. Obviously, it's from another environment.

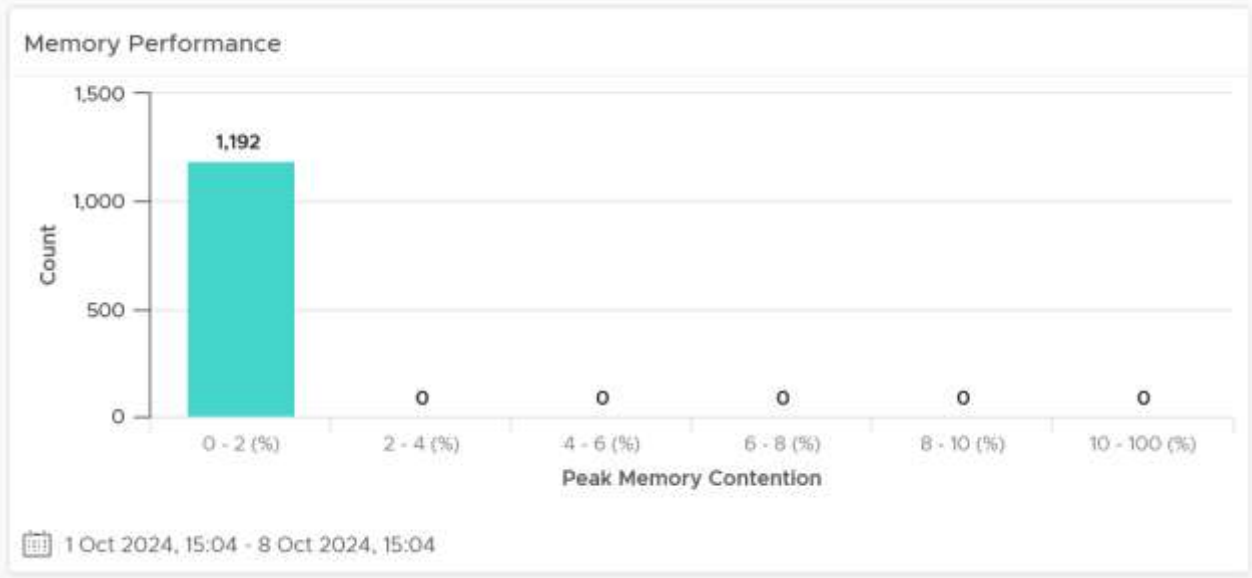
What's your conclusion from seeing the following chart?



This environment has worse performance. The distribution is evenly spread, meaning many VMs experienced CPU Ready above 5%.

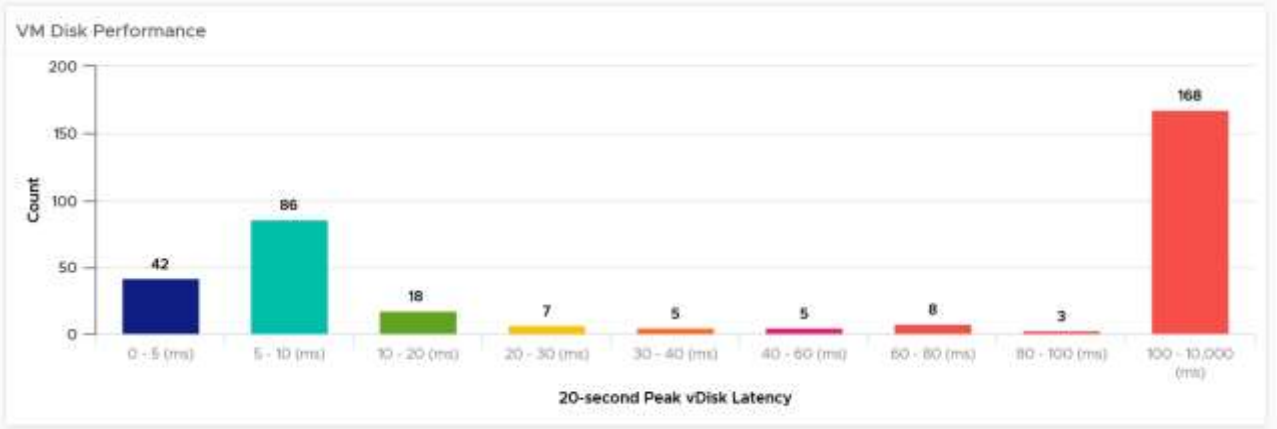
As a general guidance, investigate all those values that are really high, as they are not normal.

For memory, most customers do not overcommit memory. IMHO, this is too conservative for non-production environments. The following is what you can expect is you do not overcommit.



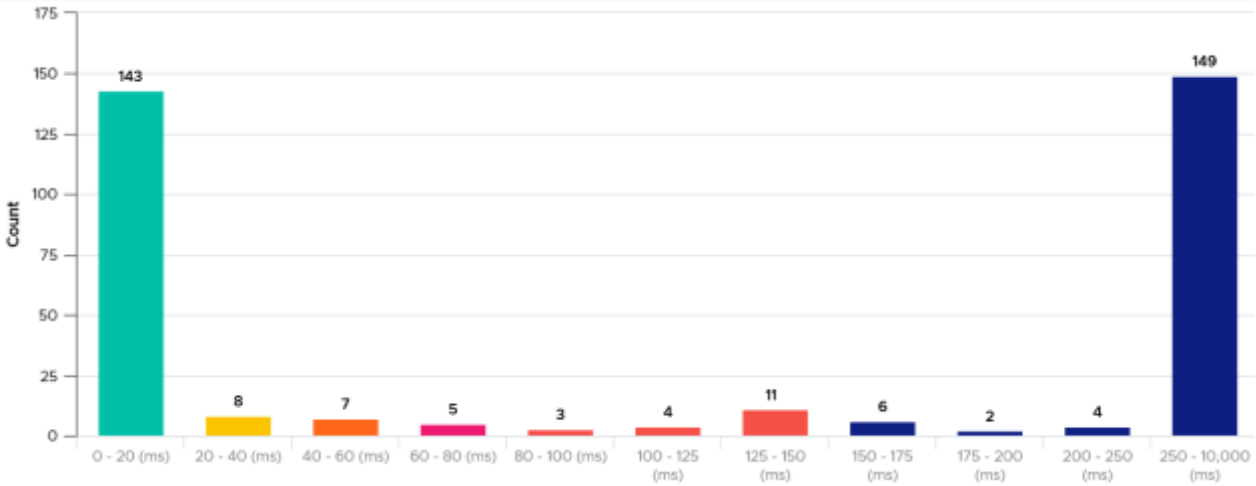
Storage Profiling

The threshold that you set can provide interesting insight. For example, the following looks normal at a glance. There are some bad latency numbers, as shown by the tall red bar at the end.



Since the pattern is a bit odd, I decided to make the “shift the buckets” by applying higher threshold. I did it by applying a higher latency bucket. T

The following is what I got!



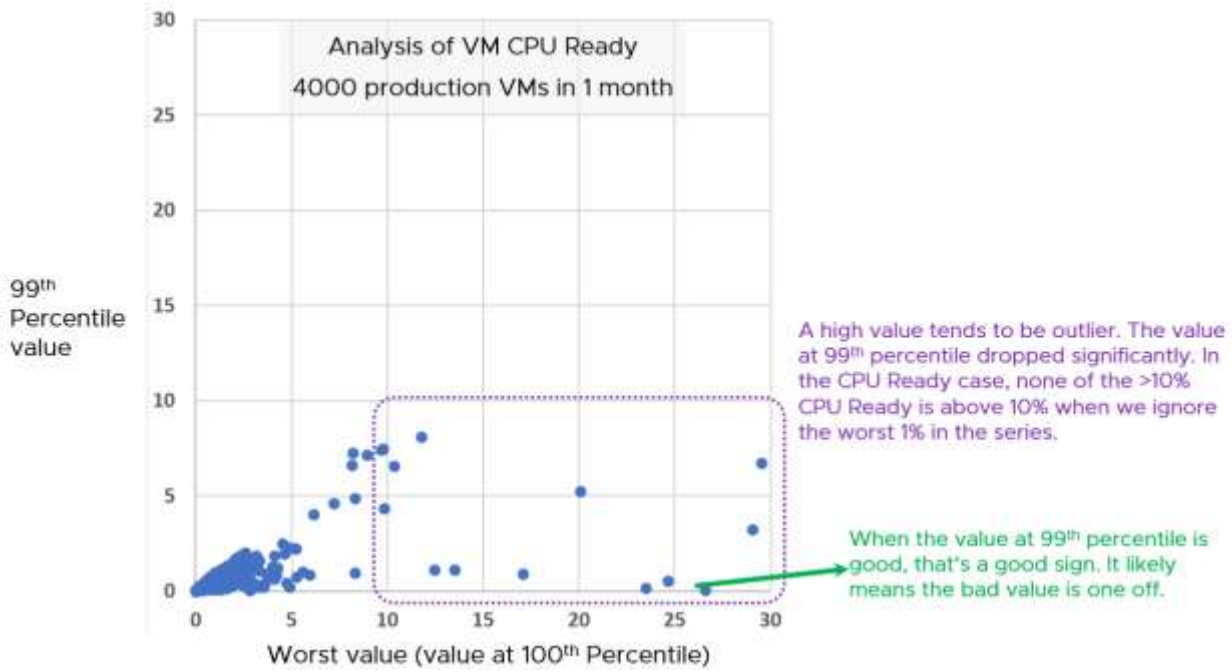
What do you think of the result?

It's interesting to see the polarizing result! In this case, the lab has 2 different classes of storage. One is SSD the other is magnetic.

Extra Analyzis

If you want to confirm that the worst value is indeed an outlier, compare the value (which is 99th percentile) against the worst value. If the gap is sizable, that means it's a one-off occurrence.

Export the table into a spreadsheet, and create a scatter chat. The following proves in my environment that CPU Ready tend to be short-lived.



Notice the high values of 99th percentile value are around 5 – 8%, while the high of the worst values are around 15 – 30%.

Infrastructure Perspective

Now that you know the overall picture affecting all the VMs, the next step is to see the performance over time. You also want to quickly compare across clusters or datastores.

In a large environment with many clusters and datastores, you may have a few clusters or datastore that fail to deliver the expected performance.

For each cluster, you want to measure both the depth and the breadth. That means tracking 4 metrics

- the worst CPU Ready experienced by any VM
- percentage of VMs experiencing CPU Ready
- the worst Memory Contention experienced by any VM
- percentage of VMs experiencing Memory Contention

For datastore, you take VM disk latency, not outstanding IO.

Having the depth and breadth give you better insight.

Just like the case in VM, take 3 months data.

A cluster with 1000 VM will have a metric that represent 1000 VM in any given 5-minute interval. Since there are 8765 instances of 5 minutes in a month, that means you analyze 8,765,000 data points in that one cluster. Taking the worst among millions will likely return you with an outlier.

This is why we use the average of worsts. So it's the average of all data points, where each datapoint is the Worst VM CPU performance data point.

To implement the above comparison for compute, create 1 view that lists the metrics. Color code the table so you can focus on the red ones.

Performance of all vSphere Clusters

Name	Worst VM CPU Ready	Worst VM Memory Contention	% VMs facing CPU Ready (%)
Management_Cluster	2.57 %	0 %	16.18 %
Client_cluster	2.97 %	0 %	16.07 %
Horizon 713 Cluster	2.77 %	0.17 %	8.52 %
Infra_Cluster	0.16 %	0 %	0.01 %
Horizon8_Cluster	0.17 %	0 %	0.0087 %
Site-04-Cluster	0.19 %	0 %	0.0019 %

Create a similar table for datastores. I added the percentage of VM facing disk latency just to get some insight into how widespread the problem is.

Performance of all Shared Datastores

Name	Average of Worst VM Disk Latency	Average of Percentage of VMs facing disk latency
w1_hs1_680af_ds02	33.42 ms	2.75 %
w1_hs1_680af_ds01	12.04 ms	0.78 %
FC-5TB-new	2.68 ms	0.14 %
w1_hs1_unity680f_...	2.3 ms	0.01 %
FC_LUN2_2TB-new	1.48 ms	0.0026 %
500GB-FC-02	1.23 ms	0 %

As in the case with VM, compare this with the default threshold used in VCF Operations, as shown previously.

The above covers "how bad the problem is". We need to cover "how widespread the problem is". How many percent of the VM population is experiencing the problem? The range used is

- Green: 0 – 2.5% of the population
- Yellow: 2.5 – 5%
- Orange: 5 – 10%
- Red: more than 10%

Analyzis

Use both the breadth and depth as input to your decision.

Depth	Breadth	Analyzis Conclusion
Good	Good	Safe to add more load, assuming the utilization is low and overcommit is below your plan
Good	Bad	Performance is good but do not add more load as many VMs are experiencing contention already.
Bad	Good	Since it's not widespread, check for common configurations among the impacted VM
Bad	Bad	If there has been no complaint for months and business are not impacted, it is possible that you do not have to do anything. If you have good relationship with the application team, this is the time to set their expectation lower. Otherwise, quietly and proactively plan for improvement as future workload may be more sensitive to latency. As a principle, IT should be ahead of business.

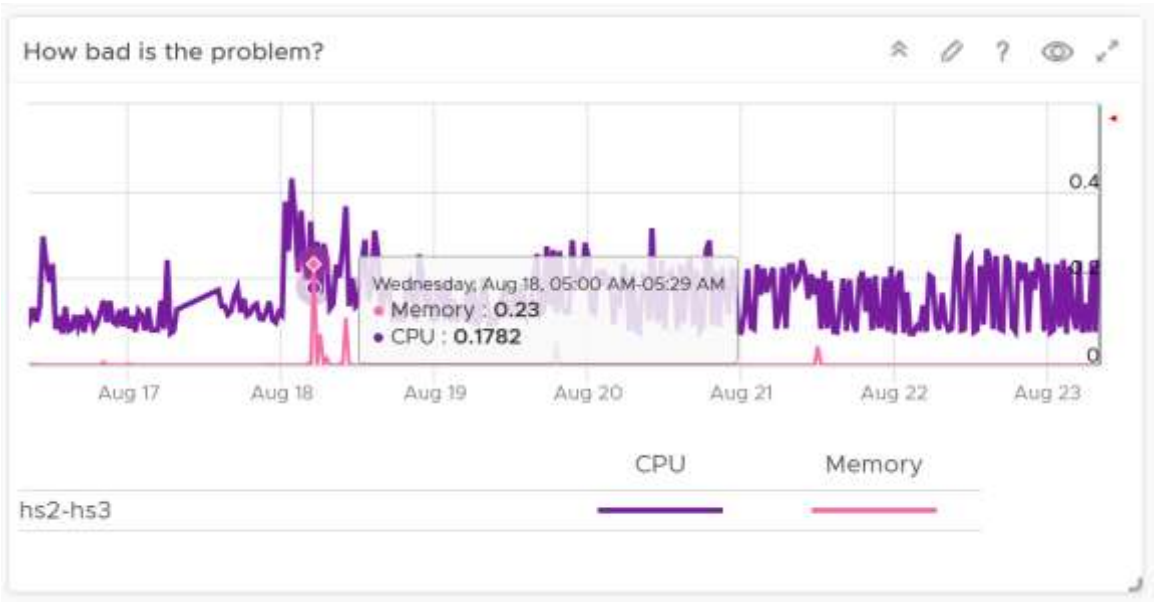
Using the above as example, you can see that all the clusters can serve their VMs well, as the highest values are 3.6%. If this is what you want to maintain, then use 3.5% as the threshold. The first cluster is already full. It can't handle anymore VM as it's already exceeding 3.5%. The second cluster is near full.

For memory, all the clusters are doing well. If this is what you want to maintain, then set 0.1% as the threshold.

If you want to be more aggressive, then you set 5% for CPU Ready and 1% for Memory Contention.

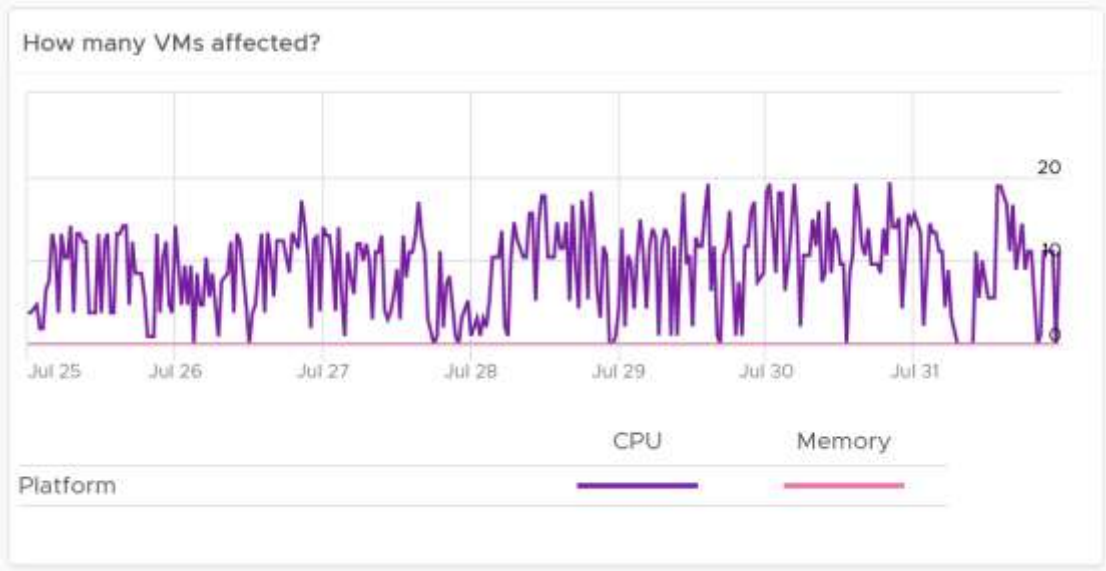
For better analyzis, create a health chart or line chart showing the breadth and depth metric for a selected cluster. This lets you see trend over time. Perhaps the time of performance problem happens during Sunday night where there is hardly any users and you're doing full back up. In that case, you can adjust your decision.

The following is an example where both CPU and memory are well within the green range.



Next, look at the breadth dimension.

In the following example, what problem do you see?



A high percentage of the VM population is experiencing CPU contention. On the other hand, memory is perfect. If it turns out that there is no issue with both the cluster and VM settings, you should consider buying more CPU as your workload turns out to be CPU heavy.

Network Profiling

Network is different due to its nature as an interconnect. We cover this in-depth in vSphere Metrics book. As a result, we need to investigate separately. Instead of investigating per VM, we investigate per network. This means per distributed port group.

We should investigate both the VMkernel network and VM network, as both can impact VM performance.

What metric should we choose?

There is no network latency counter. The proxy for it (network dropped packet) tends to have false positive, which makes profiling inaccurate. Luckily the false positive rarely happens. The false positive can be overcome by using 99th percentile metric, as it removes the outlier. The false positive happens more frequently on received side than transmit side. When profiling, we focus on the transmit side.

As there is fewer network than VM, it's more effective that we use table instead of distribution chart. List all the networks, and show the received dropped packets at 99.9th percentile. 99th percentile is less suitable as your tolerance for drop packet is higher, and they rarely happen.

Sort the list in descending order so the problematic network appears at the top.

Guest OS Performance Profiling

I'm covering this dashboard next as they reveal performance issue that application team should consider.

There are metrics that *directly* impact the performance of Windows or Linux, the Operating Systems running inside a VM. These KPIs are outside the control of the hypervisor, meaning ESXi VMkernel is not able to control the increase or decrease of their values. Visibility into these KPIs also requires an agent, such as VMware Tools. As a result, they are typically excluded in performance monitoring.

Because these KPIs are closer to the applications, it is critical to know their values and establish an acceptable range, which will vary in your environment. By profiling the actual performance over time and from of all VMs, you can establish a threshold that is supported by facts. Since there are 8766 instances of 5 minutes in a month, profiling 1000 VM over a month means you are analyzing 8.8 million datapoints, more than enough to draw a conclusion.

Use this dashboard together with application team to determine the acceptable level of these metrics. Once you determine that, add thresholds to the table so you can easily see the VMs that exceed a threshold.

Take note that these guest OS metrics do not appear unless vSphere prerequisites have been met.

How to Use

Select a data center from the data centers list.

- The three tables listing CPU, memory and Disk will automatically show the VMs in the selected data center or vSphere World.
- Each table shows the highest value in the last 1 week (2016 datapoints based on 5-minute collection cycles), hence their columns are prefixed with Max.

About the CPU table widget



CPU

Name	Max CPU Queue	Max Context Switch/sec	Configured CPU
[redacted]	13.35	5,215.87	4
[redacted]	12.47	9,527.73	2
[redacted]	7.17	9,824.6	4
[redacted]	6.75	5,856.07	4

The CPU queue is the sum of all vCPU. A larger VM can tolerate higher queue as it simply has more processors. If you want to compare VMs of different size, create a super metric that calculates the queue per vCPU.

The Max CPU Queue column shows the highest number of processes in the queue during the given period. Best practices indicate you should stay below 3 for each queue. For a VM with 8 CPUs (8 queues), you want to be below 24.

Show the corresponding CPU Run at the peak of CPU Queue. High queue not supported with high run indicates application creating excessive threads.

CPU Context Switch. There is cost associated with context switch. The problem is there is no guidance for this number, and it varies widely. This is the very purpose of this dashboard!

Review the Memory list widget below. Why are the Guest OS Paging metrics chosen, as opposed to In Use, Cache or Free memory?



Memory

Name	Max Page-In/sec	Max Page-Out/sec	Page Size	OS Type
evn-uwsim-drivevm	298,431.19	538.47	0 GB	Ubuntu Linux (64-bit)
vROps HCI Mesh main	36,442	34,010.13	0 GB	VMware Photon OS (64-bit)
BA latest	35,022.2	24,089.73	0 GB	VMware Photon OS (64-bit)
demo_1713496	31,448.33	19,210.33	0 GB	VMware Photon OS (64-bit)

Because they are measuring rate of change (performance), while the other measures a static disk space consumption (capacity).

Compare the page-in vs page-out. Is that what the application team expect?

Review the Disk list widget below. Why are these 3 metrics chosen?

Disk

Name	Max Guest OS Disk Queue	Max VM Outstanding IO	Max VM IOPS
ContainerMP-vRealize-Op...	739.71	290.88 OIOs	163.4
HCI-Mesh-togomain-versi...	445.57	173.6 OIOs	273.6
vROps HCI Mesh main	11.71	1.88 OIOs	287.8

The Guest OS Disk Queue shows potential issue inside Windows or Linux.

The 2 metrics at VM level provides additional contexts. If queue is high inside the Guest, and OIO and IOPS are low at VM level, the problem is at Guest OS layer.

Select any VM from any tables above. The 3 line charts will appear automatically and will show data from the same VM to facilitate correlation.



VM Performance

Now that you've profiled your environment, you're ready to monitor specific VM performance.

There are 2 team who have an interest. As a result, we need to cover both Guest OS and VM metrics. In addition, we also need to cover both contention and consumption are they are intertwined. As a result, there are 2 x 2 sets of metrics. The following table shows the matrix.

	Performance (it is about "speed" or contention)	Capacity (it is about "space" or consumption)
Guest OS (unaware being virtualized)	Contention inside the Guest OS directly impact the application performance. It is not visible to the hypervisor.	Guest OS unaware of Hyper Threading. The ideal utilization is 100%, so long there is no queue.
VM (it is a motherboard)	This is about VM not getting the CPU resource. It results in Guest OS experiencing time freeze.	This is about the footprint of the VM, which includes work outside Guest OS such as IO processing.

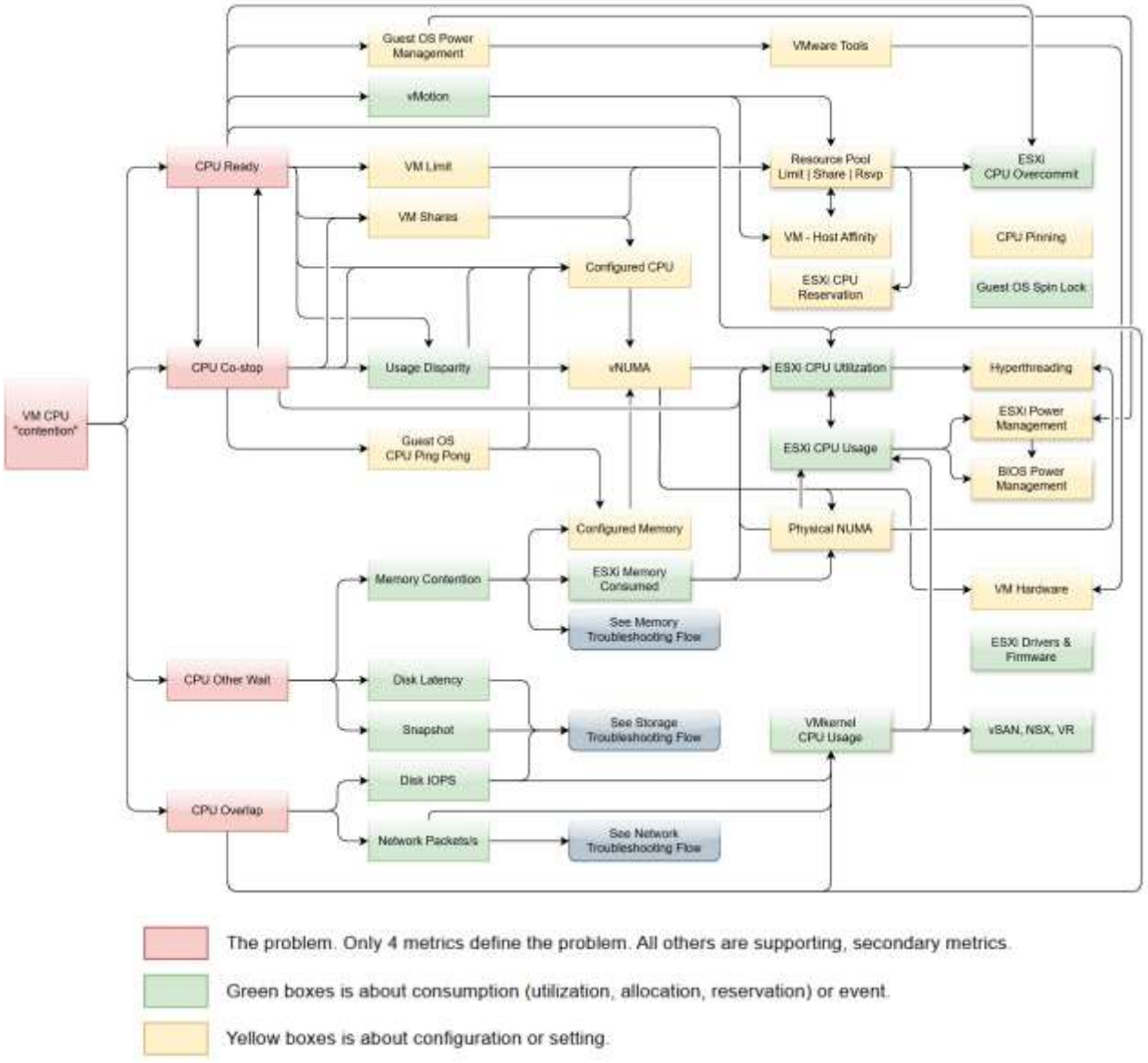
App Team

Infra Team

Keep the value low, ideally 0.
Both Guest OS and VM metrics matter.

Keep the value high for Guest OS, with buffer for spike.
VM metrics are not relevant from application point of view.

The following diagram is only showing CPU contention. A VM can have CPU, memory, disk, network or system problem.



Why doesn't the preceeding diagram include CPU Latency?

It includes a reduction of CPU efficiency due to hyperthreading. That's not a performance problem.

There are 4 primary metrics proving that a VM does not get the CPU it demands. If these are low, the VM does not have a CPU performance problem, regardless of what other metrics show.

Primary metrics are explained by secondary metrics. For a single VM troubleshooting, check the VM itself, its parent ESXi, and Cluster-setting such as Resource Pools and VM-Host Affinity. Factors outside the VM is shown with shadow.

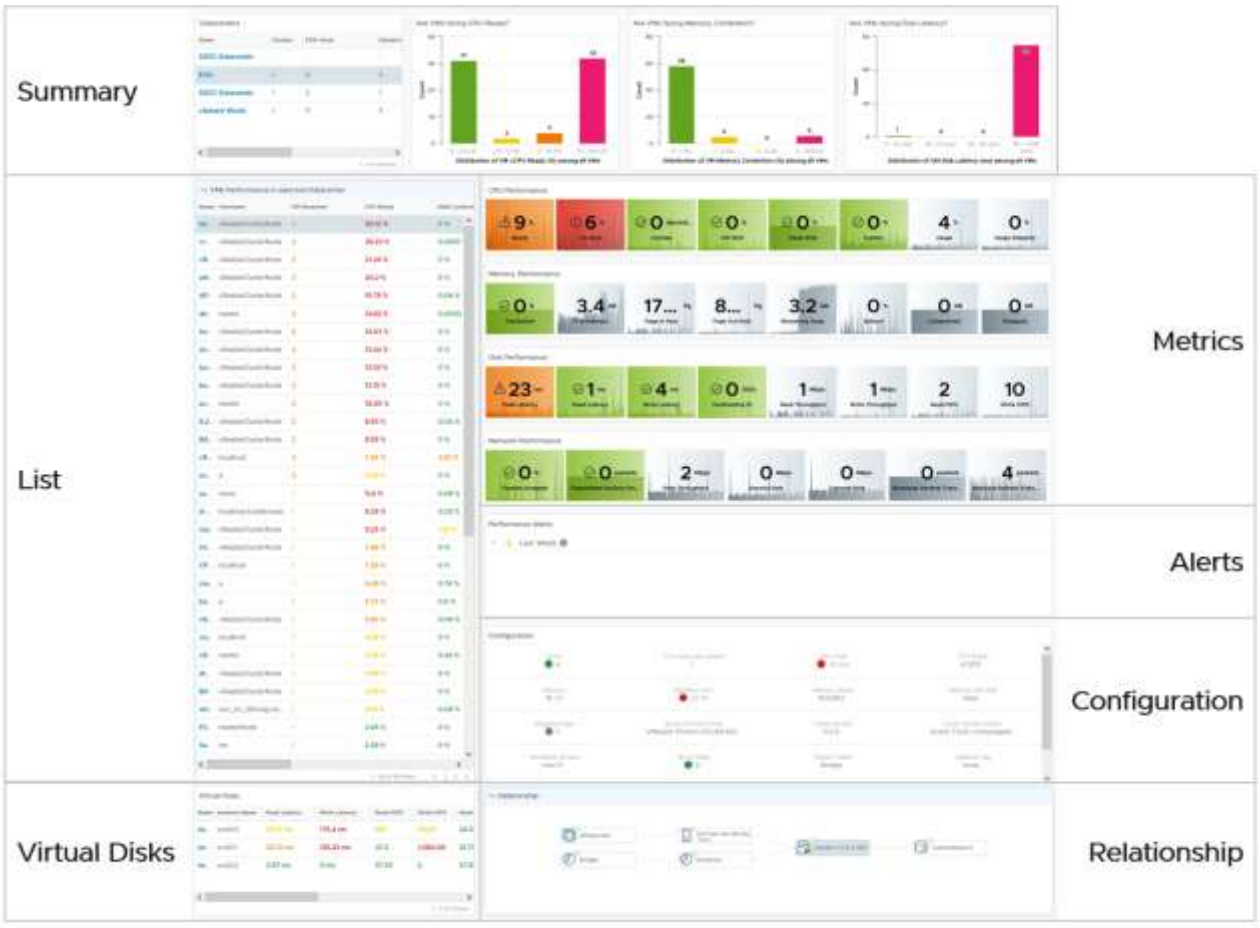
This diagram only covers VM. Guest OS has its own troubleshooting flow. Ensure BIOS is optimized for virtualization. Logically, cluster level setting such as DRS plays a part. See the Cluster diagram.

Dashboard

The troubleshooting flow is complex, hence the dashboard is complex.

The dashboard is laid out logically into 3 sections to provide a top-down flow of performance analysis.

Summary	Quickly see the big picture. The first thing to check when a VM has a performance problem is if other VMs have the same problem. If the problem is widespread the root cause is not with the VM. Hence it's important to see the big picture, before diving into a specific VM
List	A table listing all the VMs. It's convenient to analyze many VMs. The table supports sorting, filtering and export for further analyzis.
Detail	This section is made of multiple widgets
	Metrics. Only metrics impacting performance are shown. They are grouped logically by category of problems.
	Alerts. Only alerts impacting performance are shown.
	Configuration. Only settings impacting performance are shown.
	Relationship. Navigate into objects related to the selected VM.
	Virtual Disk. Drill down into the individual virtual disks.
	Virtual Network. Drill down into the individual virtual NIC. This is not yet added out of the box.



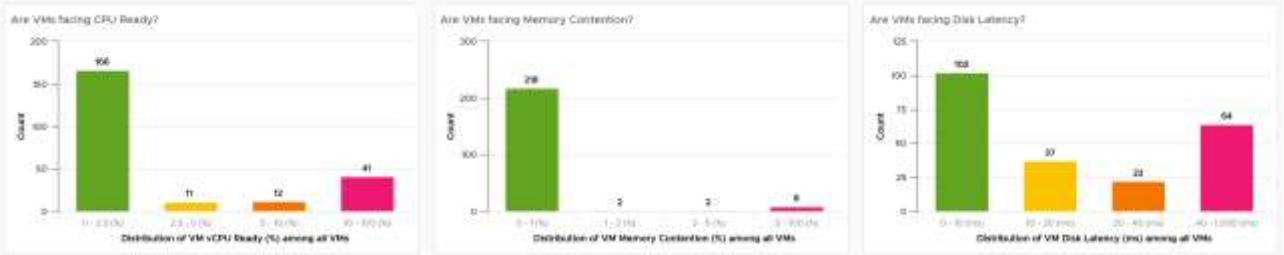
Summary

The dashboard uses a list of Data Center as a selector/filter. Why don't we use vSphere Cluster as the selector?

Because this is about VM performance, not cluster performance. Multiple clusters can share the same datastore, and they often share the same network. These shared infrastructure can impact the performance of VMs on it.

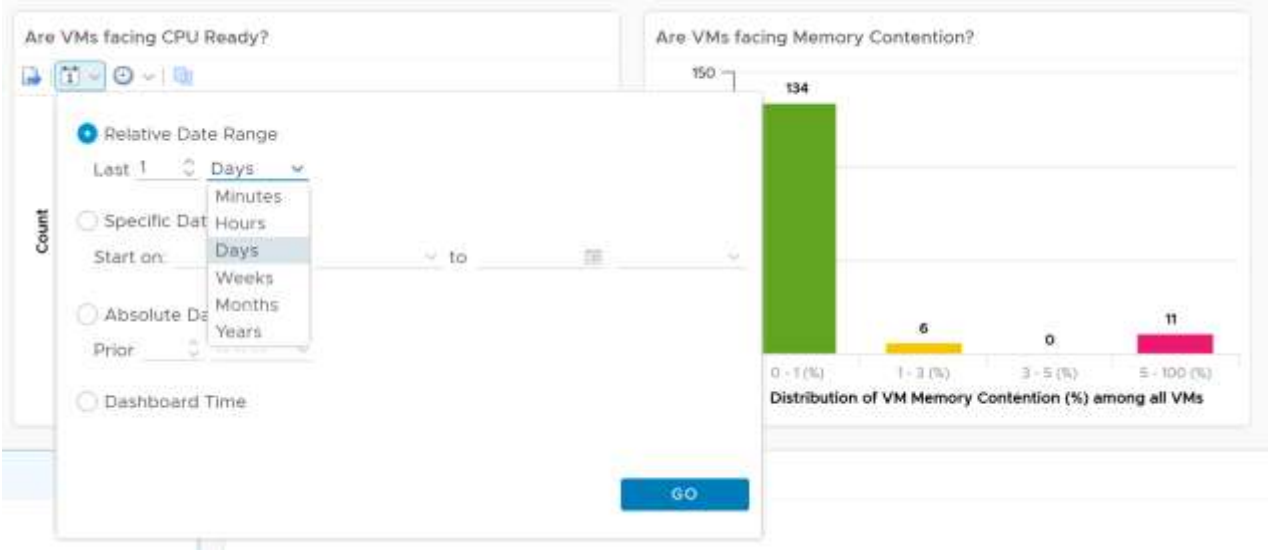
Select one of the data centers from the table. The following bar charts will be automatically shown.

What does the following 3 bar charts tell you about the environment? They work in tandem to reveal if there is any performance problem, and if yes, what problems and how bad.



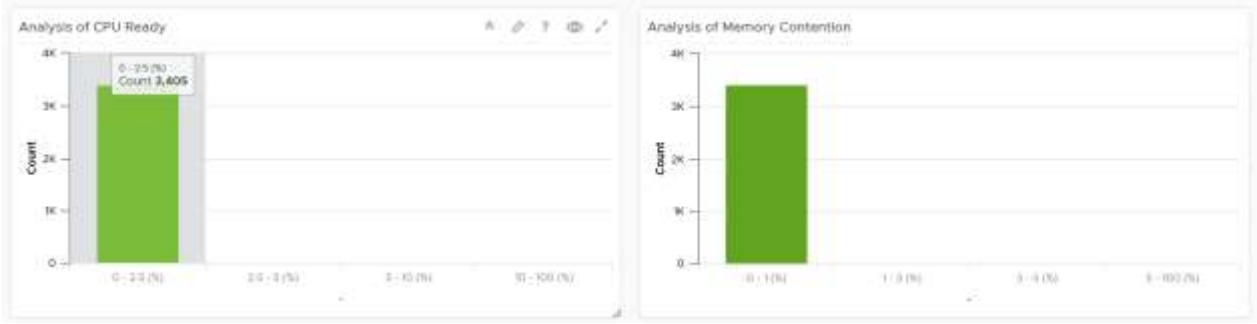
Each chart analyzes how the VMs are served by the cluster. For each VM, it picks the *worst* metric in the last 24 hours. By default, VCF Operations collects every 5 minutes, so this is the highest value among 288 datapoints. Once it has the value from each VM, the bar charts put each VM in the respective performance buckets. The threshold in the buckets consider best practice, hence they are color coded.

You can change the time period to the period of your interest. The *maximum* number will be reflected accordingly.



The value is actually the worst 20-second within the 5-minute. Yes, it's using the peak metrics, covered here in the Troubleshooting Metrics section. Why do I choose the 20-second average instead of the 5-minute average, considering your SLA should be based on 5-minute average?

It's to give you a heads up. Leading indicators. Otherwise, you get something like this:



Yes, the above shows 3405 VM. And not a single one of them has CPU Ready > 2.5% and memory contention > 1% in the last 24 hours. It's not so useful in giving you early warning.

For your mission-critical environment, you should expect that all the VMs are being served well by the IaaS. So expect to see green on both distribution chart. If they are, there is no need to analyze further.

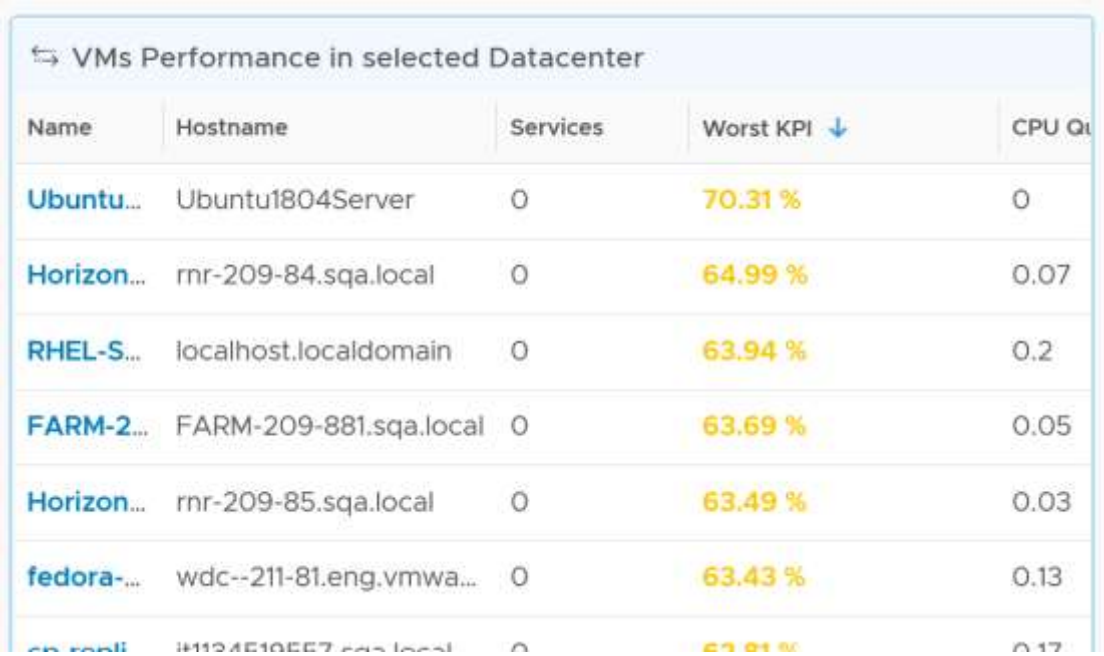
For development, you may tolerate a small amount of contention in both CPU and Memory as you need to balance cost.

If it makes more sense for you, change the filter from data center to cluster. Once you are listing cluster, you can then add the cluster performance (%) metric and sort them in an ascending order. This way the cluster that needs immediate attention is on top.

You can click on any of the bar to see the list of VMs under that performance bucket. From there, you can select a VM and have its KPI automatically shown on the lower section of the dashboard.

Multi-VM Analyzis

When you select a data center from the table, the table listing all the VMs in the data center is automatically shown. A portion of the table is shown below.



VMs Performance in selected Datacenter

Name	Hostname	Services	Worst KPI	CPU Qu
Ubuntu...	Ubuntu1804Server	0	70.31 %	0
Horizon...	rnr-209-84.sqa.local	0	64.99 %	0.07
RHEL-S...	localhost.localdomain	0	63.94 %	0.2
FARM-2...	FARM-209-881.sqa.local	0	63.69 %	0.05
Horizon...	rnr-209-85.sqa.local	0	63.49 %	0.03
fedora-...	wdc--211-81.eng.vmwa...	0	63.43 %	0.13

The table shows the hostname as known by Windows or Linux. This is the name that application team or VM owner know, as they may not be familiar with the VM name.

The table is sorted by KPI column, directing your attention to the VMs that are not performing. The column is based on the super metric VM KPI (%), which we covered in Part 2.

BTW, if the KPI is blank, that means at least one of the metrics is blank. Use the dashboard to check which metrics are blank. If not all the Guest OS metric appear, then the VMware Tools is likely outdated.

The rest of the columns show contention metrics. Notice there is no utilization metrics at all. You should know why by now 😊

Because the goal is proactive monitoring, as opposed to reactive troubleshooting, the metrics show the *worst* value instead of the average of the monitoring period. You can see from the following screenshot that the Transformation is set to Maximum, instead of Current.



Data	Transformation
Hostname	Current
KPI Breached	Maximum
CPU Ready	Maximum
RAM Contention	Maximum
Disk Latency	Maximum
Transmit Packet Dropped	Maximum

For example, the CPU Ready counter shows the highest CPU ready within the period you specify. The default period is 24 hours, as the dashboard is designed to be part of the daily SOP. The reason 1 week is not chosen is the performance > 24 hours are likely to be less relevant.

Unlike the bar chart, this shows average of 5 minutes. The reason is we're showing VM, and the VM SLA should be based on 5-minute average.

If you have screen real estate, group the VMs by cluster. In this way, you can quickly see if the problem is in particular cluster. Avoid grouping by ESXi as VMs may relocate.

Per-VM Analyzis

Once you're down on a specific VM, you need to check both at the Guest OS layer and the VM layer. The following are Windows or Linux metrics to check.

Metric	Optimize or Fix
Runaway process	The process goes into an infinite loop. Remedy is to kill it. In a large VM with many vCPU, this can be impossible to detect if the process is only consuming 1 CPU.
CPU Run Queue	Ideally, measure both the queue per CPU and the total queue. A 64 vCPU having 1 queue on each CPU means 64 processes were waiting. Performance could be affected. If utilization is high, then add vCPU. If utilization is not high, check if there is imbalance among the vCPU. Some applications may need to be configured to use all vCPU.
CPU Context Switch	IO causes context switch. If utilization is not high, reduce vCPU to minimize ping pong. More on CPU Context Switch is covered here.
Network latency	Change driver. More CPU to process network packets
Disk Queue Length	Problem could be inside the Guest or outside. Check if VM outstanding IO and VM disk latency is high. If not, then the issue is with Guest OS, not underlying infrastructure. Check if Guest OS latency is high. If yes, change SCSI driver to PVSCSI is one good remediation. If not, leave for now.
Driver Queue	Storage driver such as LSI Logic and PVSCSI has their own queue. If you know how to get the value, let me know.
Network Buffer	The ring buffer in the virtual NIC card. This can get filled up.
Memory Page Fault	Look inside the Guest OS and Process to see if it's memory settings aligns with configured RAM. It's common for JVM or DB to have memory settings that don't match Guest OS configuration. Need Telegraf agent as it is not tracked by Tools.

At the vSphere VM level, these are the metrics to check:

Metric	Optimize or Fix
CPU Ready	Check config: Limit, Shares (also relative to other VM, not just its absolute amount) Check if it's caused by resource pool Check ESXi utilization. A sign of ESXi struggling is other VMs are affected too. Check vMotion stun time (requires Log Insight)
CPU Co-Stop	As per CPU Ready, but note that VM with many CPU has higher chance of experiencing co-stop. If VM utilization is not high, reduce vCPU.

CPU Overlap	VM has high IO (disk or network), causing its own CPU to be interrupted for IO by VMkernel. Check if ESXi cores utilization is balanced.
RAM Contention	High ESXi memory utilization. Check ESXi Balloon, ESXi reservation. Check vMotion stun time (this requires Log Insight)
CPU IO Wait	High disk latency causes high CPU. Check for snapshot as snapshot requires double IO processing.
Network Dropped Packet	After verifying this is not a false positive, check for packet retransmit and underlying infrastructure. If packets are broadcast packets it might be dropped by the network.
Network retransmit	When a packet is loss, it needs to be retransmit (unless it's UDP protocol).
Network latency	Latency caused by hops. Optimize the traffic. Note this needs Aria Network Insight
Disk Latency	Check VM outstanding disk IO Check if IOPS Limit has been placed.
Outstanding Disk IO	Check underlying ESXi or datastore
vMotion	Check DRS automation setting If the actual vMotion is too long, check if the active memory. You can use Log Insight to see the throughput. For stunned time, you need Log Insight as it's not shown in the UI.

VM KPI (%)

Choose a VM of your interest from the table.



All the CPU, Memory, Disk and Network performance charts are automatically shown, each widget showing the KPIs of that VM.

CPU Metrics

The CPU panel sports the following 8 metrics. Review them. Do you know why we pick these 8, and show them in this specific order?



The metrics are shown in the order of importance or urgency. The most important one for Guest OS is CPU Run Queue, while the most important for VM is CPU Ready.

Peak CPU Run Queue. Ideally we divide by the number of vCPU. Create a super metric for that. Good practice 😉

Peak vCPU Ready (%) tracks if any of the virtual CPU is experiencing high CPU Ready. It takes the highest among the vCPU. This can be useful in large VMs with many vCPU.

Why are the last 3 metrics are not color coded?

- CPU Context Switch actually impacts performance. I'm just unsure what to put as it varies among customers. For your environment, set it as appropriate after profiling.
- The CPU Usage (%) is shown ***grey*** as it actually has *negative* corelation with performance. Grey is chosen as it also conveys wastage. Resources that are hardly utilized may not mean performance is at peak. In fact, it could be the opposite. If a VM just need 1+ vCPU, configuring it with 2 CPU will result in better performance than configuring it with 20 CPU.
- The CPU Usage Disparity (%) metric. Color code it based on your expectation

The following screenshot shows the threshold used:

Metric	Box Label	Unit	Color Method	Yellow	Orange	Red
Guest|Peak vCPU Queue within collection cycle	Queue		Custom	4	12	36
CPU|Peak vCPU Ready within collection cycle	Ready	Auto	Custom	3	6	12
CPU|Peak vCPU Co-Stop within collection cycle	Co-Stop	Auto	Custom	1	2	4
CPU|Peak vCPU Overlap within collection cycle		%	Custom	1	2	4
CPU|Peak Other Wait within collection cycle	Other Wait	%	Custom	1	3	9
Guest|Peak CPU Context Switch Rate within collection cycle	Context Switch		None			
CPU|Highest vCPU Usage of all instances	Usage	Auto	None			
CPU|vCPU Usage Disparity	Usage Disparity	Auto	None			

The Guest OS CPU Queue has a high threshold due to false positive. For details, see the metric documentation in this book.

The Other Wait metric has a high threshold due to false positive. For details, see the metric documentation in this book.

Yes, they are based on the 20-second peak. Read more behind the 20-second metrics here.

The chart also displays the present value, so you know if the problem is still happening or not.

Why none of the metrics have decimal point?

I round all the counter because the decimal is not significant and clutters the dashboard with no real value. There is really no difference between 0.1% and 0.9% in all these metrics.

Why CPU Latency or CPU Contention and CPU System metrics are not shown? Refer to the metric chapter for the reason.

Why is CPU Swap Wait not shown? Its value is a subset of Memory Contention.

Memory Metrics

Now that you know how the widget is designed, let's move content from CPU to Memory. Review the following screenshot.



Why is contention the only color coded metric? Balloon, page in rate, page out rate, etc are not. See this example of where severe ballooning did not result in much contention at all.

Because memory is a form of storage. Most metrics measure usage of the disk space, not latency of access. Consider the disk space occupied. A utilization at 90% of the space is not slower than 10% utilization. It's a capacity issue, not performance.

Why are the memory paging shown in color?

Because they are leading indicator and possible cause of memory performance within Windows or Linux.

Metric	Box Label	Unit	Color Method	Yellow	Orange	Red
Memory|Peak Contention within collection cycle	Contention	Auto	Custom	1	2	4
Guest|Free Memory	Free Memory	GB	None			
Guest|Page In Rate per second	Page In Rate	Page/s	Custom	10000	50000	200000
Guest|Page Out Rate per second	Page Out Rate	Page/s	Custom	4000	16000	64000
Guest|Remaining Swap Space	Remaining Swap	GB	None			
Memory|Balloon	Balloon	%	None			
Memory|Compressed	Compressed	GB	None			
Memory|Swapped	Swapped	GB	None			

Why is Balloon placed towards the end?

Because that's a VM level counter. For memory, you want to measure at Windows or Linux level. The following shows the metrics we use. Guest means it's coming from the Guest OS, not VM.

Why is ballooned, compressed and swapped not shown in color?

Because there presence do not mean the VM has performance problem. They are not metrics for VM performance. That's a counter for the underlying ESXi and Cluster. As usual, refer to the metric chapter for details 😊

Storage Metrics

What do you expect to see for disk? What are the contention metrics, and what are the utilization metrics?



The most important one for Guest OS is Disk Queue, while the most important for VM is the Latency. The latency is based on the 20-second metric. It's also the highest of read or write.

Peak Virtual Disk Read Latency (ms) and Peak Virtual Disk Write Latency (ms) track whether any of the virtual disks (either VMDK or RDM) is experiencing latency. This can be useful in large VMs with many virtual disks.

I put IOPS before throughput metrics as it's more popular. Use Disk IOPS and Throughput together, especially for applications that use large block size. An IO with 250x block size (e.g. 1 MB instead of 4 KB) will generate equal throughput at 250x less IOPS, all else being equal.

For throughput, I use byte or bit as it's the amount of disk space written, not the amount of data travelling in network line.

I use the following threshold. They are based on the profiling result documented in Part 2.

Metric	Box Label	Unit	Color Method	Yellow	Orange	Red
Guest|Peak Disk Queue within collection cycle	Queue		Custom	10	20	40
Virtual Disk|Peak Latency within collection cycle	Peak Latency	ms	Custom	20	40	80
Virtual Disk|Highest Read Latency of all instances	Read Latency	ms	Custom	10	20	40
Virtual Disk|Highest Write Latency of all instances	Write Latency	ms	Custom	10	20	40
Virtual Disk:Aggregate of all instances|Write IOPS			None			
Virtual Disk:Aggregate of all instances|Read IOPS			None			
Virtual Disk|Write Throughput		MBps	None			
Virtual Disk|Read Throughput		MBps	None			

I round all the metrics to the nearest whole number. By now you know why 😊

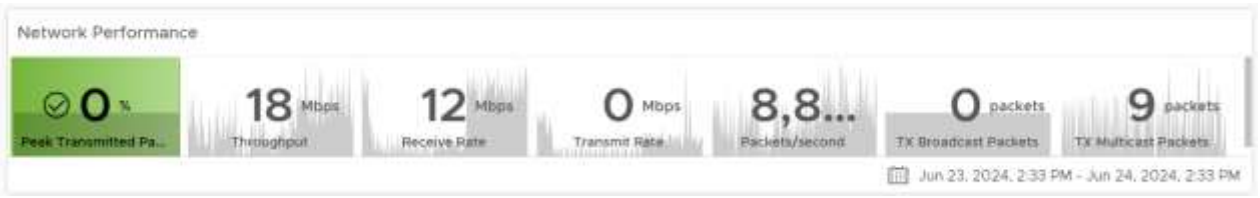
Why is Outstanding IO metric excluded?

Read the description of the metric in Part 2 😉

Network Metrics

The last component is Network. What metrics do you expect?

These are the metrics I use:



The Packet Dropped (%) formula is (dropped / (dropped + transferred Packets)) * 100%. In a rare case where there is no packet at all, the result shows undefined (blank). This is because the maths of 0 / 0 = undefined.

The following is the threshold I used:

Metric	Box Label	Unit	Color Method	Yellow	Orange	Red
Network|Aggregate of all instances|Packets Dropped		Auto	Custom	1	2	4
Network|Total Transmitted Packets Dropped	TX Packets Dropped	packets	Custom	25	100	400
Network|Peak Usage Rate within collection cycle	Peak Throughput	Mbps	None			
Network|Data Receive Rate	Receive Rate	Mbps	None			
Network|Data Transmit Rate	Transmit Rate	Mbps	None			
Network|Total Broadcast Packets Transmitted	TX Broadcast Packets	packets	None			
Network|Total Multicast Packets Transmitted	TX Multicast Packets	packets	None			

Why did I set the packet drop higher than I set in the super metric? It is in fact 2x higher.

The reason is this includes RX and TX. The RX tends to be several times higher, so setting it at 2x is actually conservative, relatively speaking.

Notice the missing counter?

There is no received packets dropped. The reason is false positive.

If it's relevant to you, add the packets per second counter.

Broadcast packets and multicast packets metrics are added as most VMs should be unicast. If the number if high, investigate with the application team why the applications are behaving that way.

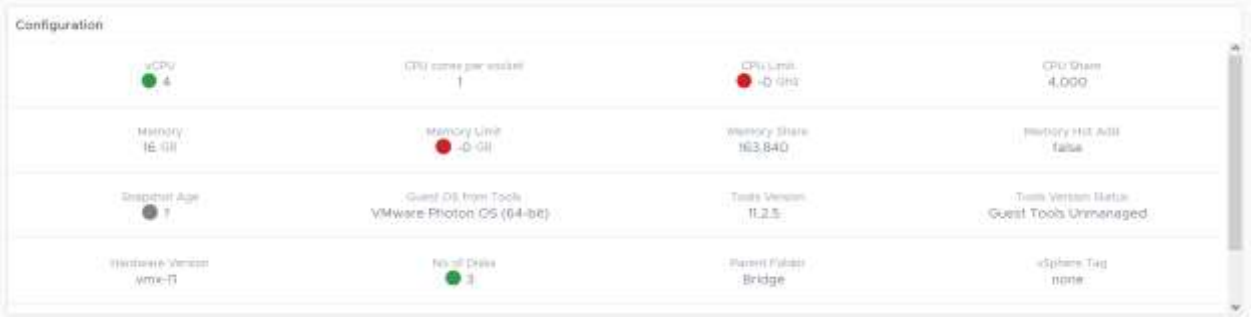
Alerts

The relevant alerts are also automatically shown. You can see the settings by editing the widget, and adjust them accordingly to fit your operational needs.



Configuration

This shows the relevant settings of selected VM. Customize as you deem appropriate.



Virtual Disks

A VM can have many disks, and it's possible that these disks have different performance. The following table lists the individual virtual disks and their contention and utilization metrics. Both read and write are shown.

Virtual Disks

Name	Instance Name	Read Latency	Write Latency	Read IOPS	Write IOPS	Read Throughput	Write Throughput	Label
ma...	scsi0:0	10.87 ms	19.47 ms	127.8	64.07	12.64 Mbps	15.55 Mbps	Hard disk 1
ma...	scsi0:1	15.73 ms	67 ms	39.73	39.8	22.39 Mbps	3.45 Mbps	Hard disk 2
ma...	scsi0:2	0 ms	0 ms	0	0	0 Mbps	0 Mbps	Hard disk 3

To see a trend over time, click on the VM object, and choose All Metrics.

Relationship

From the VM, you want to navigate to the parent cluster or datastore. Use the relationship widget to navigate and auto select the associated cluster or datastore.



Heavy Hitters VM

IaaS provides four services, CPU, Memory, Disk, and Network.

- CPU, Memory, and Disk are bound. A VM with 4 vCPU and 16 GB memory cannot consume more than this amount, the same applies to disk space. A VM configured with 100 GB disk space cannot consume more than that.
- Network and Disk utilization are not bound. An active VM can consume all your network bandwidth, packet per second capacity and storage IOPS capacity.

The **Storage Heavy Hitters** dashboard forms a pair with the **Network Top Talkers** dashboard. To understand the IO demands in your environment, use them concurrently. If you are using ethernet based storage, storage traffic will run over the same physical network your network traffic is travelling.

Why not combine into 1 dashboard?

- The users are different. Network dashboard may be required by Network Team, while storage dashboard by Storage team.
- The remediation actions are different.
- The customization is different. You extend each dashboard, going into physical array or NSX.

The dashboards are designed to help you analyze the impact of these VMs on your IaaS. It classifies the workload into two categories: short bursts and sustained hits.

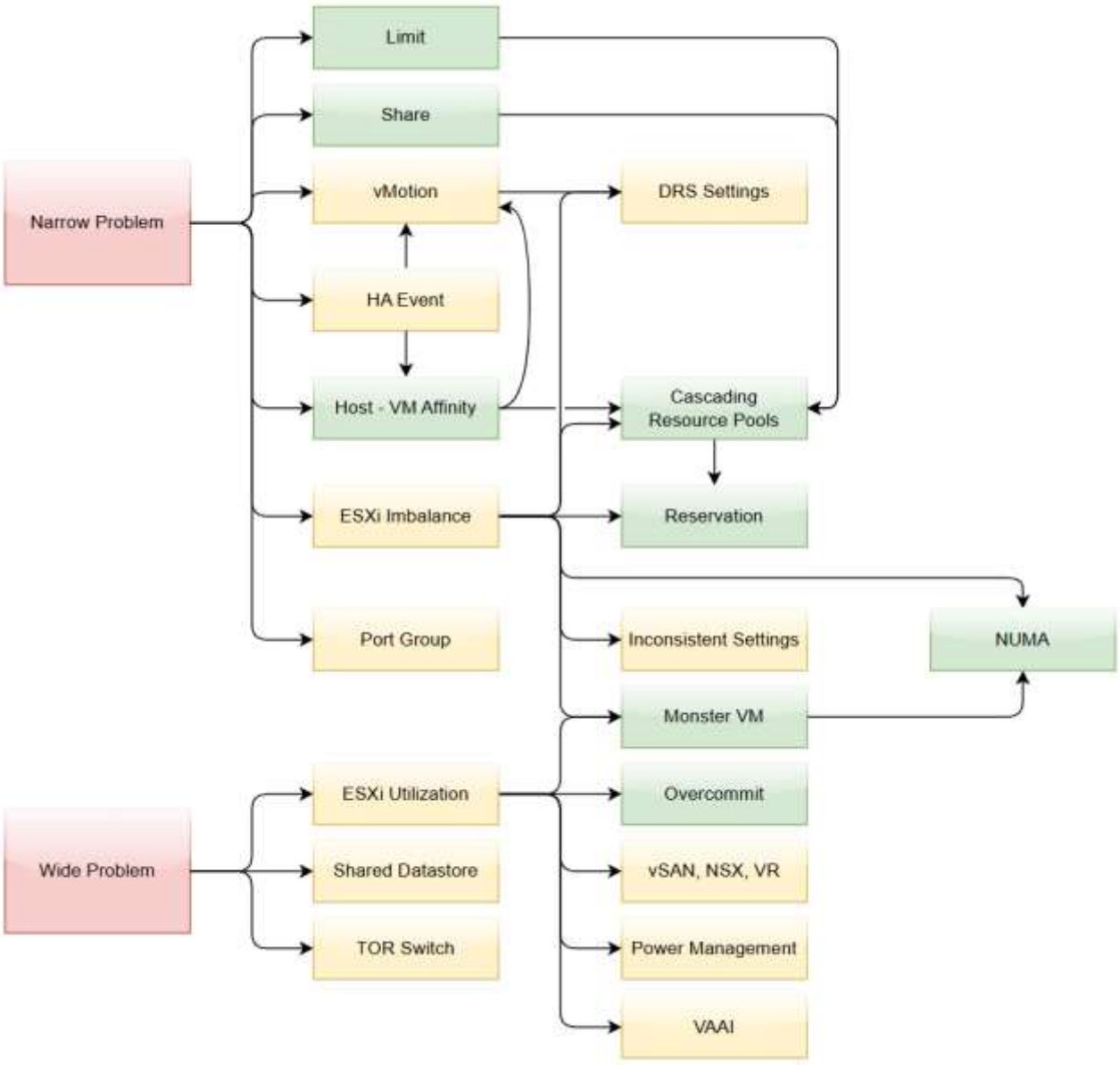
Short burst last for a few minutes, while sustained hits can last much longer. A sustained hit that lasts for an hour can cause serious problems.

A heat map would enable us to visualize the data easier. However, it can't show the past data, hence it's not used. If you want see in heat map form, see the Live! Heavy Hitters dashboard.

Points to Note

- add a line chart to automatically show the selected VM usage. Add the CPU usage also to see correlation between disk and CPU.
- Group the VM by clusters of the same class of service (e.g. Gold), so you can see the profile for each environment.
- For smaller environment, change the table from listing data centers to listing clusters.
- For larger environment, add a distribution chart so you can see the spread of the metrics.

Compute Performance



Slowness in VM population is actually a range. Wide problem means a large percentage, while narrow means a small percentage. They have different root cause, hence different worflow.

Green boxes is typically caused by consumer. This means VM, or how you slice the clusters (Resource Pools, VM - Host Affinity).

Yellow boxes is typically caused by provider. This means ESXi, cluster, hardware, and infrastructure outside the clusters (TOR Switch, Shared Datastores, vSAN)

This dashboard covers vSphere Cluster, ESXi host and resource pools, hence the generic name compute is adopted.

Design wise, while we can further add analyzis, I feel it's not worth the cost of complexity. I understand the very large customers may want to add more analyzis. If you are one of those, drop me a note.

Overall Analyzis

2 color-coded trend charts are provided.

The first one gives the overall performance. Expect this to be steady, especially in a large environment.

The second one complements it by showing the count of clusters that are no longer in the green zone. If your problem is isolated, the first counter may miss.

Average Performance

Look at the "Average Cluster Performance (%)" health chart at the top of the dashboard. In a high performing environment, where all the clusters are doing well, you will see something like this.



There was only one occurrence where the color is not green. At that time, the actual value is also relatively good.

In this case, all is good, and you may not need to look further.

On the other hand, if the clusters are unable to serve the VMs well, you will see something like this.



The average among all the clusters is no longer green, with a few occurrences of reds. The good part is the value does not trend downwards. Check if this is normal in your environment.

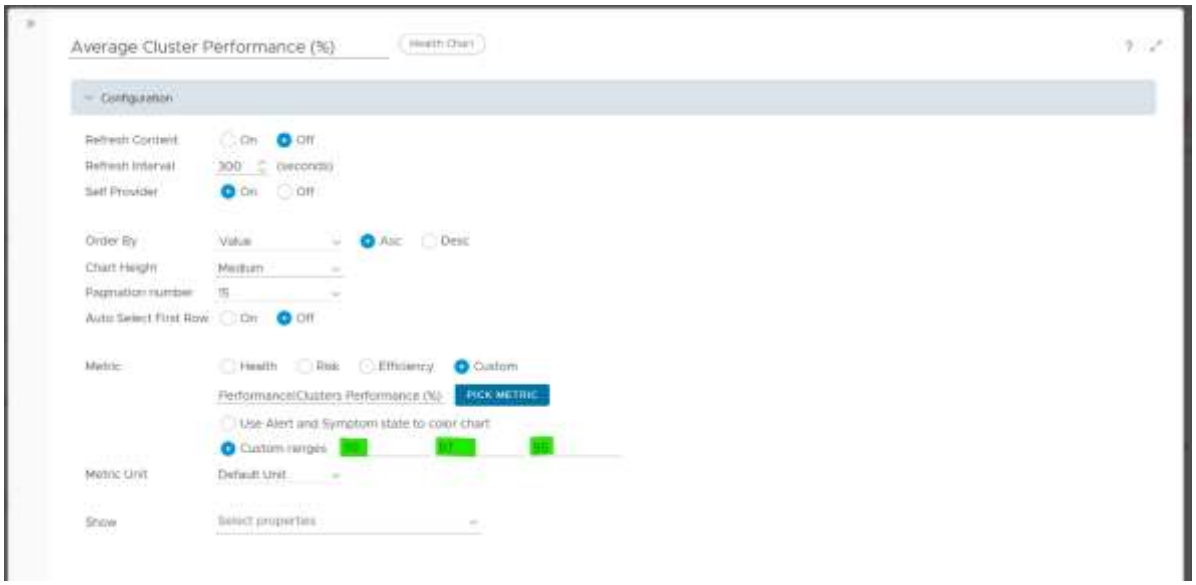
As this KPI considers *every single running VMs* in your environment, the number should be steady, especially in a large environment. The analogy in real life is the stock market index. While individual stocks can be volatile on a 5-minute by 5-minute basis, the overall index should be relatively steady. A big drop is called a market crash and that's not something you want in your environment. So a drop like the following warrants an immediate investigation.



The relative movement of the metric is as important as the absolute value of the metric. Your absolute number may not be as high you wish it to be, but if there have been no complaints for a long time, then perhaps there is no urgent business justification to improve it.

As the chart shows all the clusters, it uses the vSphere World object. This object is the parent of vCenter object, so it will show all clusters from all vCenter, making it suitable when you want to show everything.

The actual metric used is Performance \ Clusters Performance (%), as shown in the following dialog box. This is the primary KPI for your entire IaaS. It plots how your IaaS is performing every 5 minutes, giving you the trend view of overall performance.



What is this metric based on?

It is simply the average of Cluster KPI \ Performance (%) metric. This performance metric in turn averages the VM Performance \ Number of KPIs Breached metric from all running VMs in the cluster. Hence a value of 100% indicates that every single running VM in the cluster is served well.

More details on the formula were covered earlier here as it's an important foundation of IaaS performance KPI.

Non-Green Clusters

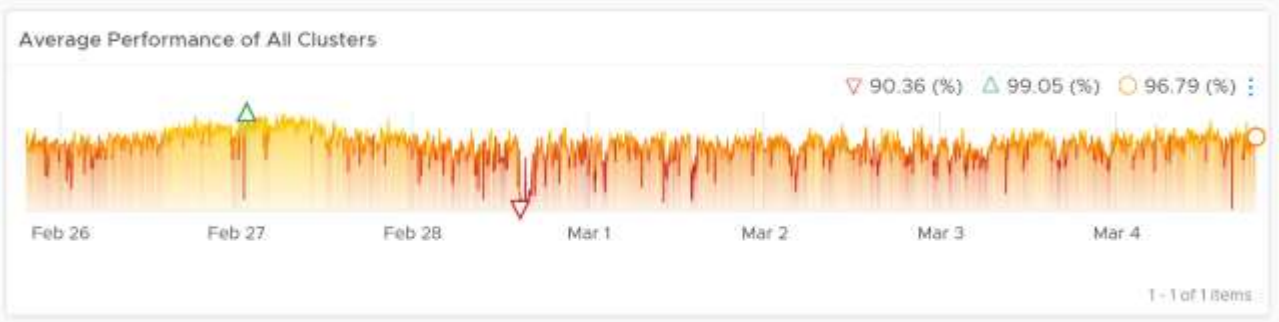
This trend charts shows the number of clusters that are no longer in the green zone. Another word, the value of the Cluster KPI \ Performance (%) metric has gone down below 75%.

Review the following chart. What trend do you spot in the last 1 week?



That's correct, the number of clusters falling below green zone has gone up from 0 to 4.

Earlier, I wrote that this counter complements the first counter. The following chart is what you get from the 1st counter. Notice it's hard to deduce that 3-4 clusters have fallen below the green zone. In this case, some clusters actually went up as we powered off their VMs, so the overall numbers remain similar.



Multi-Cluster Analyzis

If the chart is showing green, then all is good. If not, you want to know which clusters are not performing. This is where the Clusters Performance table comes in.

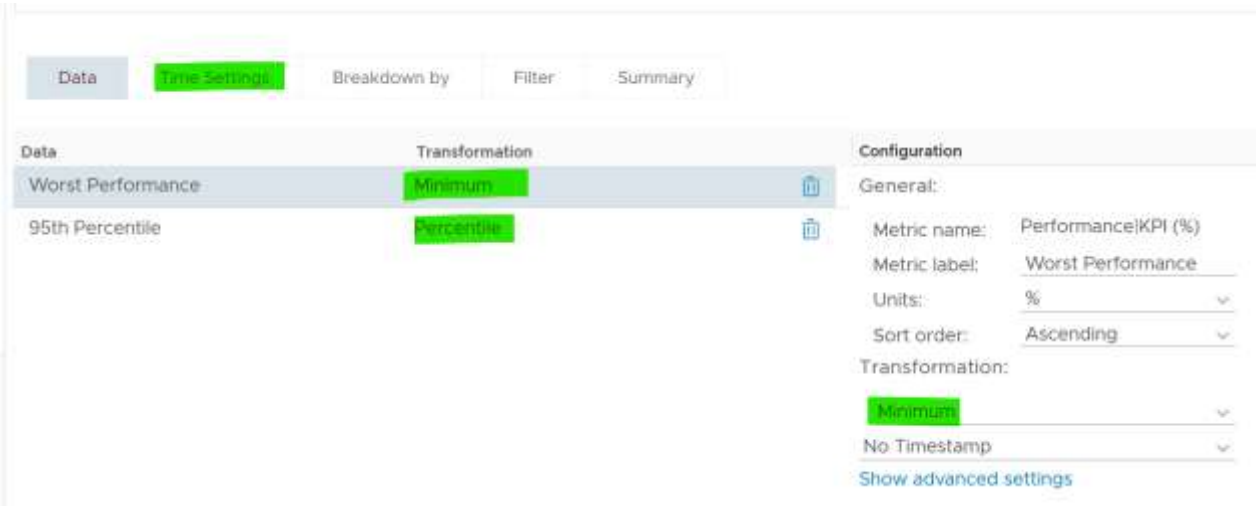
The table lists all the clusters, starting with the lowest performance. By default, it's showing data from the last 24 hours as this dashboard is designed to be part of your daily SOP.

vSphere Clusters

Name	Worst KPI	Average KPI	Running VMs	ESXi
ER HaaS Cluster	30.08 %	49.41 %	136	4
Management_Cluster	44.63 %	82.02 %	15	4
Site-04-Cluster	66.76 %	78.75 %	138	3
Infra_Cluster	67.66 %	79.9 %	59	12
Client_cluster	75.55 %	77.94 %	21	5
Horizon8_Cluster	87.24 %	98.13 %	27	5

1 - 10 of 10 items

The Worst Performance shows the lowest number in the time period. As VCF Operations collects every 5 minutes, there are 12 x 24 = 288 data points in a day. This column shows the worst point among 288 datapoints.



For a very large environment with many clusters, add a grouping to make the list more manageable. Group it by class of service, so you can focus on the more critical clusters. You can then adjust the threshold accordingly. For example, add the column worst 1st percentile to complement the worst and the worst 5th percentile. This is useful if you have clusters with more than 1000 VMs, as 5th can be too late.

You may not be comfortable not seeing utilization metrics. Old habits die hard 😊. Modify the table to add key utilization metrics. The following table shows an example of metrics to add. Notice that I do not color code the disk IOPS and Network Usage.

Clusters Utilization

Name	CPU	Memory	Balloon	IOPS	Network Usage	Running VMs
Clust...	74.92 %	77.15 %	0 GB	8,762.67	0.23 Gbps	46
2-...	50.61 %	146.22 %	12.3 GB	30,627.27	4.52 Gbps	194
1	32.4 %	107.38 %	104.47 GB	13,345.8	3.26 Gbps	159
[redacted]	26.84 %	67.31 %	0 GB	80.13	0.0046 Gbps	6

Per-Cluster Analyzis

The table is accompanied by the health chart. It allows for quick toggling among clusters.

Select a cluster to see the trend over time, and see the Performance (%) health chart.



If the cluster has a daily or weekly pattern, increase the dashboard time duration to at least 1 week. Here is an example where the performance problem is clearly shown. You can see the cluster performance has regular drop in the last 7 days.



Once you determine the cluster to investigate, review the scoreboards. There are 5 of them:

- CPU
- Memory
- Disk
- Network
- Others

They are shown separately, because you can have one problem and not the other. You may also drill down into a different path (e.g. into vSAN dashboard), and talk to different team (e.g. Network team).

CPU problems tend to be more common than Memory problems, due to lower overcommit ratio in memory in practice among customers. It is common to see customers do 4:1 CPU overcommit and only 2:1 memory overcommit. This conservative practice was due to the inherent higher value of memory counter. vSphere cluster shows high memory value as memory, giving the impression the actual utilization is high. And the reason for high value is modern Operating Systems like ESXi VMkernel uses memory as disk cache.

For performance, it's important to show both the depth and breadth of the performance problem, as explained here. A problem that impacts 1-2 VMs requires a different troubleshooting process than a problem that impacts all VMs in the cluster.

The depth is shown by reporting the worst among any VM counter. So the highest value of VM CPU Ready, VM Memory contention, VM Disk Latency among all the running VMs are shown. If the worst number is good, then you do not need to look at the rest of the VM.

A large cluster with thousands of VM can have a single VM experiencing poor performance while >99.9% of the VM population is fine. The depth counter will not be able to report that most VMs are fine. It only reports the worst. This is there the breadth metrics comes in.

The breadth metrics report the percentage of the VM population that is experiencing performance problems. The threshold is set to be stringent, as the goal is to provide early warning and enable proactive operations.

CPU Metrics

This is the first of a series of performance scoreboards.



There are 3 metrics measuring the depth of the problem:

- Worst CPU Ready among all the VMs in the cluster.
- Worst CPU Co-stop among all the VMs in the cluster.
- Worst CPU Ready among all the VMs in the cluster.

CPU Ready is the primary counter. Do you why CPU Contention is not used? See CPU Contention metric in Part 2 of the book.

CPU Co-stop is included, but placed at lower priority than Ready because high Co-stop does not mean the ESXi is struggling to serve the VMs. Co-stop can be reduced by right-sizing an oversized VM, so the remediation action is not always on the ESXi Host. CPU Overlap is included as it can happen when there are many active VMs in the cluster.

CPU Overlap is included as it indicates interrupts. A running VM was interrupted because VMkernel needs the physical core to run something else. A high and frequent number of interruptions is not healthy. This can also impact the VM performance. It's placed third as the value tends to be small.

Here are thresholds used.

Box Label	Unit	Color Method	Yellow	Orange	Red
Worst VM Ready	%	Custom	2.5	5	10
Worst VM Co-Stop	%	Custom	2.5	5	10
Worst VM Overlap	ms	Custom	500	1000	1500
VMs with Ready	%	Custom	2.5	5	10
VMs with Co-Stop	%	Custom	2.5	5	10

If you find that the CPU Co-Stop values in your environment is much better, adjust it accordingly.

Take note that the Workload metric can exceed 100% because it's demand / usable capacity * 100. So this could happen if you have 4 hosts in a cluster with each host running at 100% demand and admission control is set to 50%.

If your cluster has many resource pools and Host to VM Affinity, create a super metrics to track the gap between highest and lowest CPU usage among ESXi hosts in the cluster. The reason is imbalance is pretty common, especially in large clusters with many hosts or VMs. There are many settings that can contribute to it (e.g. DRS settings, VM Reservation, VM – Host Affinity, Resource Pool, Stretched Cluster, Large VM).

Memory Metrics



Expect memory utilization to be higher than CPU, as it's a form of cache. The Memory Consumed counter is used, as it's more appropriate than the Memory Active counter. If active is low, no need to upgrade RAM as Consumed contains disk cache. For me, it's fine for Consumed to be 99% so long RAM Contention is 0.

The high utilization metrics are explicitly shown. Balloon, Compressed, Swapped. Notice they can exist even though utilization is not even 90%, indicating there was high pressure in the past. If you look at only utilization, you'd think you are safe!

I use the following settings for the threshold. What do you notice?

Box Label	Unit	Color Method	Yellow	Orange	Red
Worst VM Contention	%	Custom	1	2	5
VMs with Contention	%	Custom	2.5	5	10
Contention	MB	Custom	256	512	1024
Compressed	GB	Custom	0.5	2.5	10
Swap In Rate	MBps	Custom	0.5	2.5	10
Swap Out Rate	MBps	Custom	0.5	2.5	10
Swap Out	GB	Custom	0.5	2.5	10
Ballooned	GB	None			
Highest Host Usage	%	None			

The threshold for contention is much lower than the threshold for other metrics. This is because contention is the only performance metric. The rest is secondary, playing supporting role. You can have 10 GB compressed, so long they are not used, the VM does not experience delay.

Storage Metrics



The disk IOPS is split into read & write to gain insight into the behaviour. Look at the values for that cluster. Do they match your expectation?

For throughput, why are the unit set at Gbps? Why not GBps or Mbps?

Because we're referring to IO commands sent down the cable (your DC is not wireless, yet!). Your ethernet cable or FC cable use bit, such as 16 Gbps or 32 Gbps.

Mbps is too small for a cluster. A cluster doing useful work should easily exceed 1 Gbps as it's just 128 MB/s. If you have distributed storage like vSAN, you can very well exceed 10 Gbps.

If you use VMFS datastores, add the disk contention metrics such as Bus Reset and Aborted Commands. You should expect they return 0 at all times.

Network Metrics

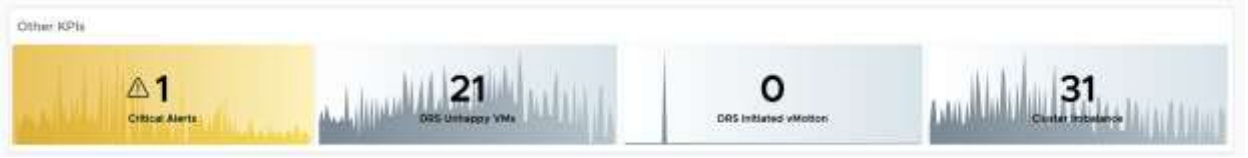


For network dropped packets, there are 2 types: physical (ESXi) and virtual (VM). The physical counter is used as that better represents infrastructure problems. The VM counter may suffer from a false positive. I recommend you customize this dashboard and add it anyway, for completeness.

Expect the network error 1% and dropped packet to be 0 most of the times, if not always. If it's not, analyze to see if there is any patterns across all ESXi Hosts, and bring it up to your network team.

The network throughput is split into sent (transmit) and received to gain insight into the behaviour. Plus, the total usage can be misleading because it sums send and received traffic. In reality the network pipe is 1x for each direction (due to the full duplex nature of ethernet), not 2x shared by both.

Other KPI Metrics



The vMotion line chart is added as a high number of vMotion can indicate the cluster load is volatile, assuming the DRS Automation level is not set to the most sensitive setting.

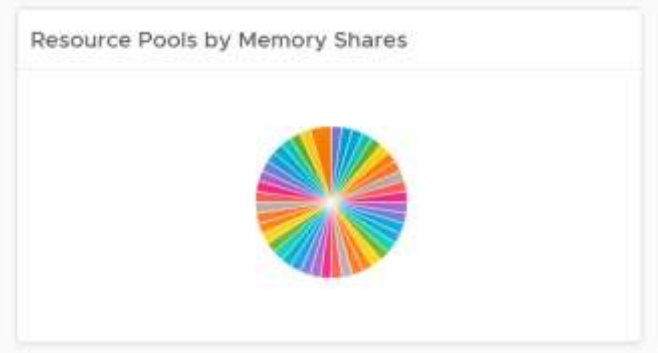
VM Shares

One common root cause for uneven performance problem is uneven shares. This is easy to make a mistake, as shares are relative. So when you see many slices of the pie, ask yourself why you need that many. Each slice should correspond to a Class of Service. So the entire cluster is serving 1 class, then you should see a simple circle with no pizza slice.

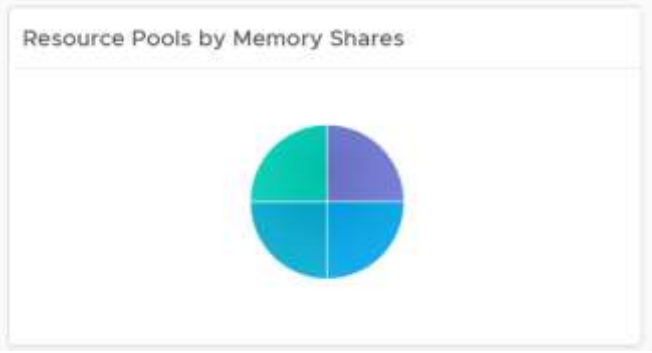


Resource Pool Analyzis

Resource Pool is another common reason behind uneven VM performance. They can also get complicated as you can cascade resource pool. The following example shows a cluster with way too many resource pools. This alone will make performance management difficult.



The second cluster has 4 resource pools. They are sized by their respective share. A pool with 2x share will have 2x the size. With that in mind, can you spot a problem?



Yes, they are of equal share. What's the point of having resource pool when they are the same share? It looks like a common mistake of using resource pool as folder.

The dashboard also provides a table listing all the resource pool.



You can navigate to the VM performance dashboard. It will show the VMs in the selected resource pool.

ESXi Analyzis

At the end of the day, a cluster is just a collection of ESXi. The performance can be caused by uneven performance among the member hosts. You can drill down from cluster to ESXi. The following table lists all the ESXi hosts in the cluster, sorted by the worst performance in the last 24 hours.

ESXi Hosts in the Selected Cluster

Name	Worst Performance	95th Percentile	Max CPU Ready	Max RAM Contention
ev...	60.71 %	92.86 %	17.14 %	22.9 %
ev...	64.29 %	92.86 %	16.23 %	1.03 %
ev...	68.75 %	100 %	45.39 %	1.76 %
ev...	77.78 %	97.22 %	37.68 %	0.87 %
ev...	79.17 %	87.5 %	4.78 %	0.11 %
ev...	82.14 %	96.43 %	27.76 %	0.02 %
ev...	85 %	97.5 %	32.03 %	0.1 %
ev...	85.71 %	92.86 %	2.41 %	2.88 %
ev...	95 %	100 %	6.16 %	0.13 %

If the table is all showing green, then there is no need to analyze further. The reason 24 hours is chosen instead of 1 week is the performance > 24 hours ago are likely to be irrelevant.

You can change the time period to the period of your interest. The maximum number will be reflected accordingly.

The table helps you quickly compare the performance of each ESXi. You can also see the performance over time, to see a trend. For example, here is the performance of one of the hosts.



Another ESXi on the same cluster is showing a different experience. What does it tell you?



Yes, there is imbalance. A cluster is meant to be balanced, if you enable DRS. The above proves that something has constrained DRS ability.

Just like the cluster KPI, you can also see the KPI of a host. They are similar to the cluster, so to avoid duplication, I'm only covering the difference.

CPU Metrics



An ESXi typically has a few dozen physical cores, which take turn to run a few dozen VMs. This could create imbalance among the cores. The metric Highest CPU Core Usage tracks the highest utilization among the cores. If this number is consistently near 100%, that indicates at least 1 core is always busy. If this number if much higher than the Usage (%), that indicates imbalance.

Memory Metrics



For metrics where the cluster is the sum of its ESXi, the threshold used for ESXi is lower. For example, the red threshold used for cluster for compressed is 10 GB, while the threshold for individual ESXi is 2 GB.

Storage Metrics

It's likely that the ESXi hosts in a cluster will have different disk IOPS pattern. They could be running VMs with different purpose. Even if all the VMs are the same (e.g. they are all from the same VDI pool), the way the users use the applications are different.

For ESXi, we show both the datastore level and adapter level. Why is that so?

Review the Storage Metrics chapter for the detail. In short, an ESXi can mount datastore and non datastore (e.g. RDM) via its storage adapter.



Network Metrics



Configuration

Certain settings such as power management and hyper threading can impact performance. The configuration widget shows the relevant property of a selected ESXi Host.

VM Analyzis

Selecting a cluster or ESXi will automatically list the running VMs. Use this table to check if the cluster or host performance problem were caused by VM configuration and usage. Note that it is possible that the VM was not on the same host at the time of problem, due to vMotion.

To drill down into a particular VM, select it and click the double arrow before the widget title. Choose the VM Performance dashboard.

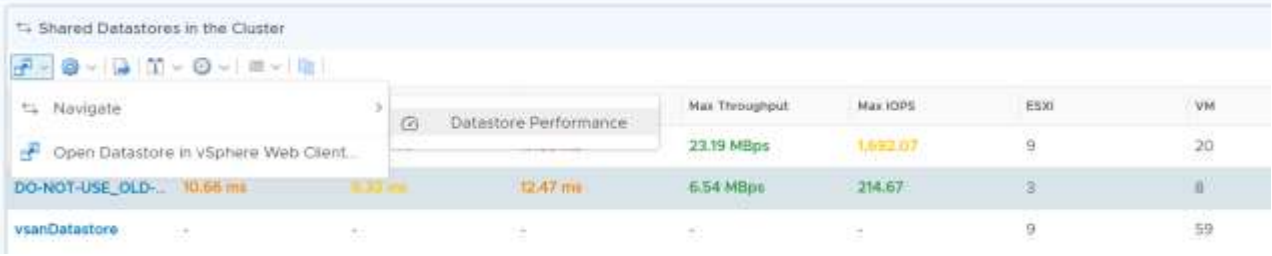


The table is color coded for ease of analyzis, as you can have >1K VM in a cluster. Tailor the threshold to your environment.

Each of the column is showing the maximum value of the time period, so you can spot even if there is a short burst.

Datastore Analyzis

The list of shared datastores accessible by hosts in the cluster is shown. Take note that if your design has M:N relationship between clusters and datastore, then the numbers at datastore level is a number from all clusters.



You can also drill down into the selected datastore.

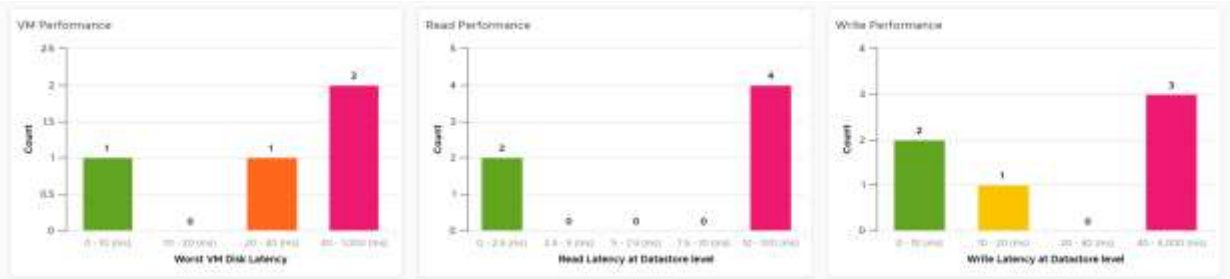
Storage Performance

Use the Datastore Performance dashboard to view performance problems related to storage such as high latency, high outstanding IO, and low utilization. This dashboard is designed for both VMware administrator and Storage administrator, with the goal of fostering closer collaboration between the 2 team.

Local datastores are treated separately, as they have their own use case.

Overall Analyzis

The three bar charts provide the overall analyzis of the datastore performance in a given vCenter DC or World. They work together to give a better insight. Just like other performance charts, the value shown in the worst value during the time period.



The first one shows how many VMs experience what kind of latency. This is your primary chart as it's measuring at VM level. The next two charts are measuring at datastore level, which means they are the normalized average of all VMs in that datastore. Expect the first chart to be higher than the last two.

Read and Write latency are shown separately for a better insight. The nature of read and write problem may not be the same so it's useful to see the difference.

The dashboard does not have datastore clusters. If your environment use it, add a View List right after the list of data center, and have it also drives the Datastore Performance view list.

Datastore Analyzis

A table listing all the shared datastores in the data center or vSphere World will also be automatically shown.

Datastores Performance

Name	Worst Write Latency	95th percentile Write Latency	Worst Read Latency	95th percentile Read Latency	VM
Datastore-ISCSI-	380.71 ms	48.18 ms	183.51 ms	31.41 ms	105
Datastore-ISCSI	127.16 ms	30.04 ms	52.77 ms	14.93 ms	87
DC1	107.19 ms	64.94 ms	57.23 ms	22.53 ms	64
vsanDatastore	100.78 ms	6.52 ms	2.43 ms	0.95 ms	25
DC2	86.51 ms	53.81 ms	49.29 ms	19.63 ms	79
vsanDatastore	6.79 ms	6.35 ms	1 ms	0.15 ms	15
nfs0-1	0 ms	0 ms	0 ms	0 ms	1

Since it's a table, we can afford to show multiple numbers. Both the worst (peak) performance and 95thpercentile are shown. If the latter is close to the peak and it's also high, then it's a sustained problem. If the latter is low, then it's a short duration.

The table is color coded. If you require a different threshold, adjust accordingly.

Select a datastore you want to troubleshoot. The relevant metrics and configuration will be shown automatically.



Its latency and outstanding IO is automatically shown.

Note the latency is the normalized average of all VMs in the datastore.



Its IOPS and throughput are also shown. These line charts are not color coded as it varies per customer. Edit the widget and add your expected threshold. It will make it easier for the operations team.

The list of running VMs in the datastore is automatically shown, with the relevant contention and utilization metrics.

VM in the selected datastore

Name	Worst Write Latency	Worst Read Latency	IOPS	Throughput
ale...	82.1 ms	9.11 ms	29	2.69 Mbps
ma...	81.38 ms	8.58 ms	21.33	2.45 Mbps
lat...	80.93 ms	24.39 ms	50.73	11.42 Mbps
loc...	77.8 ms	50.33 ms	25.8	2.66 Mbps
he...	74.6 ms	16.67 ms	32	3.15 Mbps
Al...	55.42 ms	25.22 ms	53.93	12.34 Mbps
dev	26.6 ms	0.93 ms	0.13	0.02 Mbps

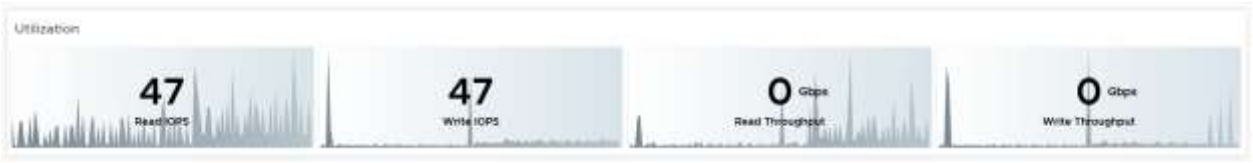
VM Analyzis

Select a VM you want to troubleshoot. Its contention and utilization are automatically shown.



Note that this number is at the VM level. If you suspect one of the virtual disk has high latency, use the counter Peak Virtual Disk Read Latency (ms) and Peak Virtual Disk Write Latency (ms).

For utilization, here are the key metrics.

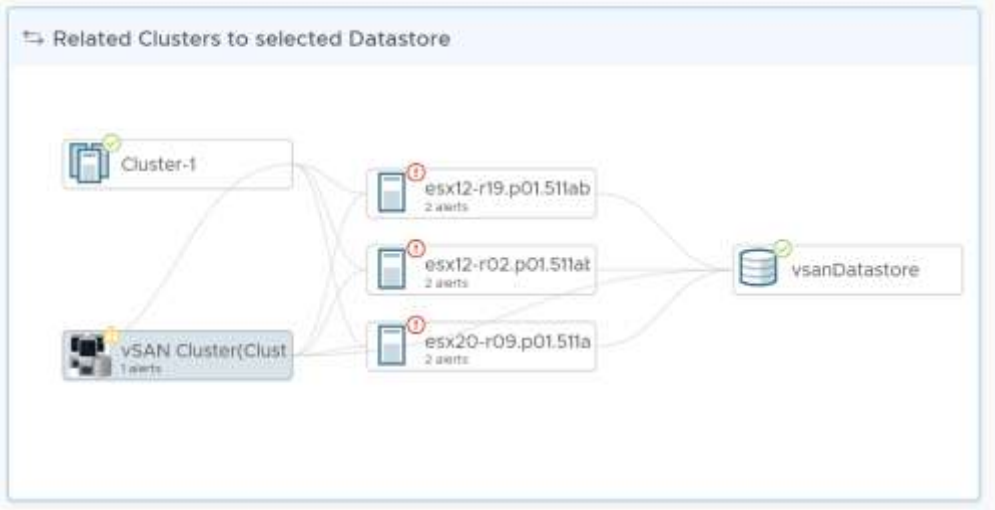


If you have many VMs with virtual disks on multiple datastores, add a View List widget to list the individual virtual disks. Use this list to plot the latency & utilization of individual virtual disk.

Relationship

From the datastore, what if you want to check which clusters or hosts are causing the problems? If it's a vSAN datastore, what other relevant KPIs should you check?

From the relationship widget, select either an ESXi, a vSphere cluster or a vSAN cluster. It's relevant contention and utilization metrics will be automatically shown.



The following table shows the contention metrics for the 3 object types:

Object Type	Metric	Box Label	Unit	Color Method	Yellow	Orange	Red
Host System	Datastore|Highest Write Latency of all instances	Write Latency of all Datastores	ms	Custom	10	20	40
Host System	Datastore|Highest Read Latency of all instances	Read Latency of all Datastores	ms	Custom	10	20	40
Host System	Datastore|Outstanding IO requests	Total Outstanding IO	OIOs	Custom	10	20	40
Host System	Storage Adapter|Highest Read Latency of all instances	Read Latency of all Adapters	ms	Custom	10	20	40
Host System	Storage Adapter|Highest Write Latency of all instances	Write Latency of all Adapters	ms	Custom	10	20	40
Cluster Compute Resource	Datastore|Outstanding IO requests	Total Outstanding IO	OIOs	Custom	10	20	40
Cluster Compute Resource	Datastore|Worst VM Disk Latency	Worst VM Disk Latency	ms	Custom	10	20	40
Cluster Compute Resource	Disk|Read Latency	Read Latency	ms	Custom	5	10	20
Cluster Compute Resource	Disk|Write Latency	Write Latency	ms	Custom	5	10	20
Cluster Compute Resource	Disk|Total Latency	Latency	ms	Custom	5	10	20
vSAN Cluster	KPI|Max Disk Group Congestion	Disk Group Congestion		Custom	60	100	150
vSAN Cluster	KPI|Max Disk Group Read Cache/Write Buffer Latency	Cache Latency	ms	Custom	2.5	5	10
vSAN Cluster	KPI|Min Disk Group Read Cache Hit Rate	Cache Hit Rate	Auto	Custom	95	90	85
vSAN Cluster	KPI|Max Capacity Disk Latency	Capacity Disk Latency	ms	Custom	5	10	20
vSAN Cluster	KPI|Sum vSAN Portgroup Packets Dropped	Packet Dropped	%	Custom	0.25	0.5	1

Just like all the contention metrics, they are color coded for ease of analyzis.

The following table shows the utilization metrics.

Object Type	Metric	Box Label	Unit	Color Method
Host System	Datastore|Read IOPS			None
Host System	Datastore|Write IOPS			None
Host System	Datastore|Read Throughput		Gbps	None
Host System	Datastore|Write Throughput		Gbps	None
Cluster Compute Re...	Datastore|Read IOPS			None
Cluster Compute Re...	Datastore|Write IOPS			None
Cluster Compute Re...	Datastore|Read Throughput		Gbps	None
Cluster Compute Re...	Datastore|Write Throughput		Gbps	None
vSAN Cluster	KPI|Max Capacity Disk IOPS			None
vSAN Cluster	vSAN|Performance|Write IOPS			None
vSAN Cluster	vSAN|Performance|Read IOPS			None
vSAN Cluster	vSAN|Performance|Total Throughput		Gbps	None

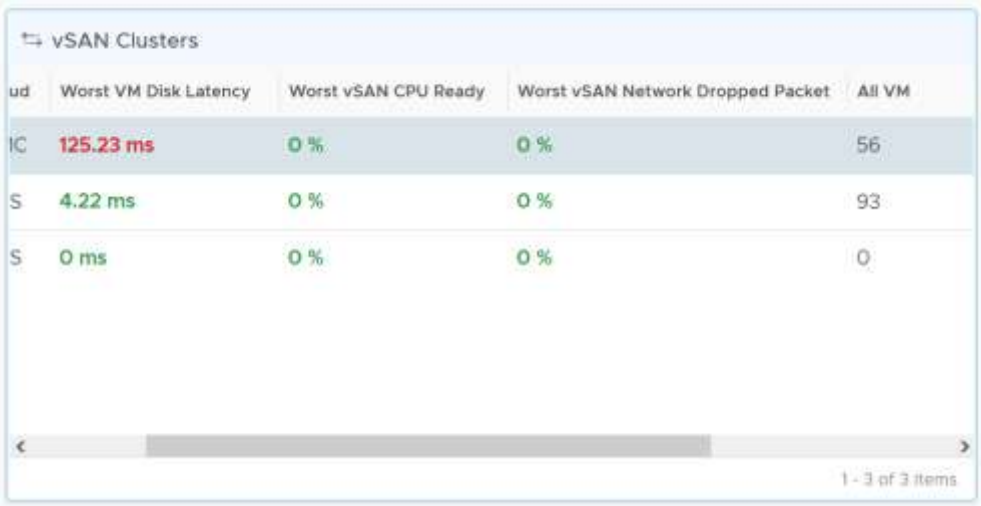
Just like all the utilization metrics, they are not color coded so not to give the wrong conclusion.

vSAN Performance

This dashboard is designed to complement the vSphere Cluster Capacity dashboard. It focuses on the storage and vSAN specific metrics, and does not repeat what's already covered. It also does not list non vSAN cluster. That's a long winded way to say you gotta use both dashboards to manage vSAN performance 😊

Cluster Analyzis

The table lists all the vSAN clusters, starting with cluster with the highest VM disk latency. By default, it's showing data from the last 24 hours as this dashboard is designed to be part of your daily SOP.



All Flash and Hybrid have different performance. Since there is only 1 benchmark, the all flash will perform better than hybrid, and that's what you expect to see.

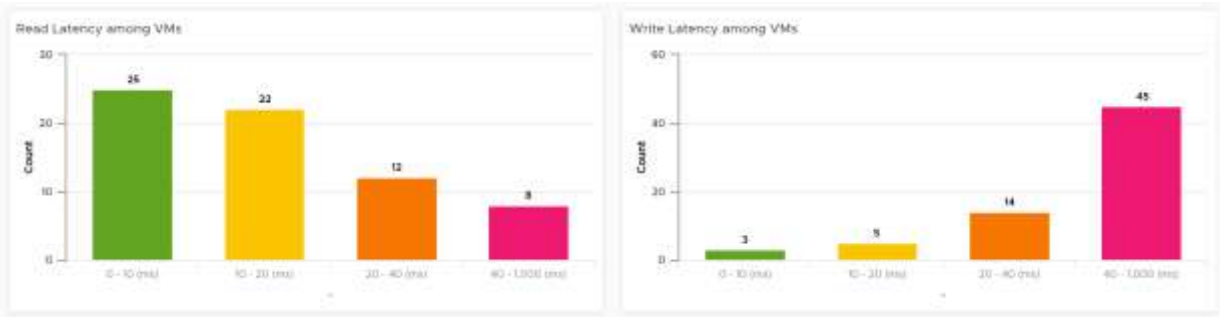
The first column shows if the distribution of disk latency experienced by all the VMs in the cluster. You should expect majority of the VMs to experience latency that matches your expectation. For example, in an all flash systems, the VMs should not be having >10 ms disk latency. If your vSAN environment is all flash, you may need to adjust the distribution bucket to a more stringent set.

The second column shows if any of the vSAN kernel module has to wait for CPU. Expect this number to be near 0% and below 1%, as vSAN should not be waiting for CPU time. vSAN gets higher priority than VM World as it lives in the kernel space.

The third column shows if any of the vSAN cluster is dropping packet in the vSAN network (not the VM network). vSAN relies on network to keep the cluster in-sync. This number should be near 0% and less than 1%.

Select a cluster to investigate further. Its VM latency distribution will be automatically shown.

What can you conclude from the following 2 bar charts?



You can tell that most VMs in the selected cluster experience write latency, but not read latency. This helps you choose the right path of investigation.

The latency counter above is not at VM level. It's at virtual disk level, meaning if a VM has 10 virtual disks, it takes the highest among any of them.

Contention

You can see various disk-related contention metrics of the cluster, as shown below.



All these metrics are at vSAN cluster level, so there are some roll up involved. Why are the Read Latency and Write Latency not shown as the first metrics?

They are average. By the time the group average is bad, you could have half the population in worse situation. One way to overcome is to use a tighter threshold, as shown in the following table.

Metric	Box Label	Unit	Color Method	Yellow	Orange	Red
Summary|Worst VM Disk Latency	VM Disk Latency	ms	Custom	10	20	40
KPI|Max Capacity Disk Latency	Capacity Disk Latency	ms	Custom	5	10	20
vSAN|Performance|Read Latency	Read Latency	Auto	Custom	5	10	20
vSAN|Performance|Write Latency	Write Latency	Auto	Custom	5	10	20
vSAN|Performance|Outstanding I/O	Outstanding I/O		Custom	10	20	40
vSAN|Performance|Congestion	Congestion		Custom	60	100	150

Read and Write are split as they tend to have different patterns.

Utilization

As expected, we complement the contention metrics with utilization metrics.



Large block size can result in high throughput in relatively low IOPS. If you are seeing large block size when You are not expecting it, investigate which applications are the using it.

Max IOPS among capacity disk is shown as a disk has a limit, especially magnetic disk. A typical magnetic disk delivers ~200 IOPS, and this can be easily saturated.

Disk Groups

You can drill down to the disk group level. All these metrics are the worst value among the disk groups.



Same thing with the read cache. All the values shown below are the worst value among the disk groups's read cache.



Other KPIs

The performance problem could be caused by non-storage.



vSAN Resync is a type of utilization metric, but its presence can impact performance. vSAN has two scenarios that trigger rebalance (the reason for resync).

- Proactive: A large variance among devices disk space utilization.
- Reactive: When devices reach a critical capacity threshold (typically ~80%). This adjustment can be in the form of object components moving to other disks, disk groups, or hosts, or even splitting of existing large components into smaller components to achieve the desired result.



Disk Group Analyzis

You can drill down into individual disk groups.



Disk Groups

Name	Read Latency	Write Latency	IOPS Congestion	Write Buffer Free
Disk group (5201d39e-50e4-6474-f376-36ac9d81176f)	86.91 ms	3.22 ms	0	69 %
Disk group (52665924-d400-c52c-bc39-3dd3f5897d08)	18.57 ms	0.94 ms	0	69 %
Disk group (528ae737-3361-37ee-99b2-f7e5d55e9b1a)	13.87 ms	0.96 ms	0	69 %
Disk group (528e86da-ca26-31d7-a6a8-f96a704d0883)	9.21 ms	0.59 ms	0	69 %

Make sure they are fairly balanced. If not, you have a hot spot.

Select a disk group you want to analyze. Both the contention and utilization metrics are automatically shown.



There are more metrics for the utilization. To avoid the name being truncated, I keep them short. To see the full name, simply mouse-hover over the counter name, as shown below.



For the throughput metrics, why do I choose bytes instead of bit? I use MBps, not Gbps.

Revisit the vSphere Metrics book for the answer.

Cache Disks

You can drill down into individual disk groups. As this is SSD, you should expect their values to be below 5 ms.



Cache Disk

Name	Device Latency	Guest Latency	Write IOPS	Read IOPS	Type	Vendor	Size	Queue Depth
evn-lab-esx-78.eng.vmware.com_NAA.55CD2E404C2068CD	0.67 ms	0.69 ms	6,314	2,022	SSD	ATA	200 GB	32
evn-lab-esx-79.eng.vmware.com_NAA.55CD2E404C2068F1	1.57 ms	1.58 ms	6,936	1,725	SSD	ATA	200 GB	32
evn-lab-esx-80.eng.vmware.com_NAA.55CD2E404C205DB8	1.79 ms	1.81 ms	6,433	1,832	SSD	ATA	200 GB	32

The configuration is also shown. Make sure they are consistent.

Select a cache disk you want to analyze. Both the contention and utilization metrics are automatically shown.



vSAN has various layers, which is covered in Part 2 Chapter 4 Storage Metric.

Congestion is a special derived metric in vSAN.



Why is the cache disk space utilization not considered contention?

Because 90% used does not mean it's faster or slower than 80%. We can't color code it.

Capacity Disks

You can drill down into individual disk groups. The latency value will vary between magnetic and SSD.



The configuration is also shown. Make sure they are consistent.

Select a capacity disk you want to analyze. Both the contention and utilization metrics are automatically shown.



At this level, what you can do is basically observing if there is a hot spot.



Network Performance

Use the Network Performance dashboard to view performance problems related to network such as high latency, unusual traffic, and many dropped packets. This dashboard is designed for both VMware administrator and Network administrator, with the goal of fostering closer collaboration between the 2 team.

The dashboard enables drill down from distributed switch to the ESXi host and port groups in the switch, and then to the VM.

To prove that Network is performing well:

- No Errors.
  - Not a single ESXi host is experiencing packet drops in any of its NICs (vmnics). If there are, show the ESXi names.
  - Not a single VM is experiencing packet drops.
- Latency
  - Good round trip latency. No network trombone across physical data centers if NSX is used.
- Utilisation
  - Not a single VM is hitting its limit, be it 1 GE or 10 GE or higher.
  - Not a single ESXi vmnic is hitting its limit.
  - Total bandwidth hitting the physical switches is below capacity.
- Special network
  - The broadcast network is minimal. For both ESXi and VM.

How to Use

Review the **Distributed Switches** table

- It lists all the switches, sorted by the highest packet dropped. The table splits the incoming traffic and outgoing traffic for better analyzis.
- As the focus is performance and not capacity, the throughput metrics are not shown.

Select a switch from the table

- The health chart will automatically show the dropped packet trend over time.



- However, it will not narrow down the list of port groups automatically, as the list of port groups are always showing all the port groups in your environment.
- If necessary, expand the 2 collapsed widgets. They are showing the network throughput and broadcast packets. Utilization is also shown so you can correlate if the dropped packets are due to higher utilization

Review the port groups and ESXi hosts in the selected switch

- They are automatically listed when you selected a switch from the table above.
- Just like the distributed switch, you can also see their relevant counters. The following shows the same chart, except for a single port group.



If your environment has unused network switches, you can filter them out from this list, as this dashboard focuses on performance.

Points to Note

Network latency within a data center should be below one millisecond. Use Aria Network Insight to study the latency or the retransmitting problems, caused by moving into the lateral traffic.

Add a physical network using the appropriate management pack, such as True Visibility Suite.

Most packets are unicast, between a pair of sender and receiver. If your environment has many VMs sending broadcast packets to everyone and multicast packets to many targets, add a Top-N widget to find out which VMs are sending these packets.

Enhance to include analysis of distributed port group. Does any of them hit their limit?

Capacity

Part 2 Chapter 5

The capacity dashboards aims to implement the capacity concept covered in Capacity Management chapter, and complement the existing product pages.

Overall Design

You may notice that the performance dashboards and the capacity dashboards share similar layout. The reason is there is commonality in both pillars of operations.

The suite of dashboards work together as one integrated set. They also have similar design.

Goals	Determine overall capacity. Cover compute, storage and network. Network covers ESXi physical NIC, NSX Edge and NSX Manager.
	Optimize capacity. Reclaim, rightsize, and redistribute.
Questions	Examples of questions answered by the dashboards:
	What's the overall capacity? Do we have enough CPU, memory, disk and network to meet present demand + demand in the near future?
	Any hot spot? Do we need to rebalance?
	How severe is the wastage? How much can we easily claim within a month? How much will take much longer to reclaim?
Assumptions	Users will use the built-in pages. These dashboards are designed to complement, not replace.
Target Users	**Capacity Team.** This is the main target. The team responsible for overall capacity life cycle, from planning to upgrade. Not interested in day to day, such as a host goes down or put into maintenance. Keen on long term and top down view, as they are planning future expansion and ageing hardware technological refresh. Their primary focus is the Provider
	Operations Team. The team running the day to day, live operations. Not interested in long term, both the past and the future. Their primary focus is the Consumer, especially oversized VM.
	Storage Team. Specialist team. Collaboration become critical if the architecture is "thin on thin".
Usage frequency	Weekly
	Quarterly
Features	The dashboard is designed "top down". It has 2 sections: summary and detail. The summary lets you see the big picture. The detail section is placed below the summary section. It lets you drill down into a specific object. For example, if it's a VM capacity, you can get the detail capacity of a specific VM.
	Quick context switch. You can cycle through objects quickly without changing screens or opening multiple browser windows.

	UI wise, the dashboard uses progressive disclosure to minimize information overload and ensure the webpage loads fast. On the other hand, so long your browser session remains, it remembers your last selection
	Color coded. A high capacity remaining could indicate wastage. For ease of identifying, wastage is shown by a new color. Dark grey indicates wastage as capacity is not used. In fact, there can be performance problem is the low utilization was caused by bottleneck somewhere else
	Complement the out of the box pages by visualizing information differently and giving more choice of customization. For examples, the reclamation size is grouped into buckets so you can focus on the largest reclamation opportunities first, and trend charts are provided so you can quickly see the growth over time, without changing context (e.g. open a new screen).

There are 2 types of capacity management

- Consumer. You focus on a single VM, container or application. You want to right size them and reclaim unused portion.
- Provider. You focus on the shared infrastructure as the problems impact many consumers.

For cluster, special types of cluster alter the capacity model. One example is stretched cluster. It needs its own capacity model and visualization. You will need to have an object or custom group for each site, and then displays them side by side.

Compute Capacity

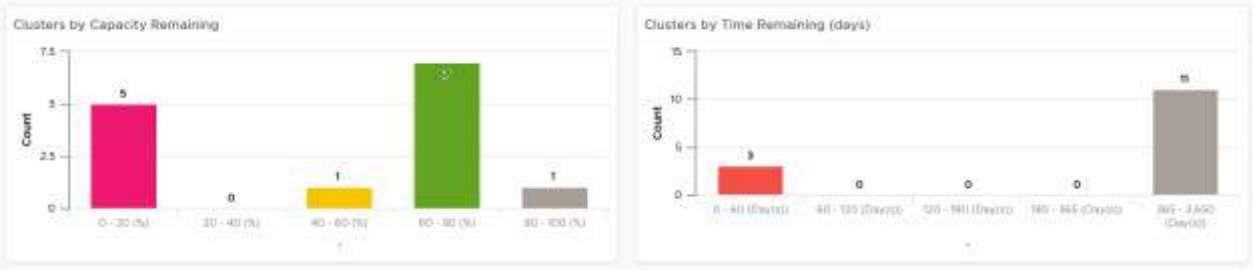
The compute capacity dashboard covers vSphere clusters with their associated ESXi host and resource pools, as they impact the cluster capacity.

The dashboard is designed for cluster, and not standalone ESXi host. The layout has these sections:

- Multi Clusters. It should have a table listing all clusters.
- 1 Cluster. What's the overall utilization. It should show both CPU, memory, and network.
- Subcluster. Break into the ESXi hosts. This is only needed if there are imbalances.

Overall Analyzis

The 3 bar charts show all the clusters, summarizing the overall situation. The first 2 are shown below.



The first two bar charts work *together*. Just because you are running low on capacity does not mean you are running out of time. Cluster with a cyclical load that hits high utilization but never trends toward 100% will have low capacity remaining, but plenty of time remaining.

Generally speaking, the ideal situation is low Capacity Remaining and high Time Remaining. This means your resources are cost effective and working as expected. For clusters where you intentionally need to run at low utilization (e.g. stretched cluster), increase its buffer accordingly so the capacity metrics will reflect that.

The 3rd bar chart is VM Remaining. It gives more complete contexts, as different clusters can have different VM size.

The bar charts can't show the parent data centers. It also can't show all the clusters, individually. For a large environment, a heat map comes in handy.



The three heat maps are Time Remaining, Capacity Remaining, and VM Remaining.

The color indicates usage. Low utilization is marked as grey, not green, as it represents waste.

Why is the box size made identical?

For ease of use and better focus on the action to be taken. Otherwise the small clusters will be dwarfed by the large ones.

If your cluster sizes are not standardized, create another heat map, and use the number of ESXi hosts to show the size difference.

A table provides the next level of details.

Clusters Capacity

Name	ESXi Hosts	Time Remaining	Capacity Remaining	VM Remaining	Running VMs	Total CPU	Total Memory
Clus...	-	-	-	-	-	-	-
vRO...	9	> 1 Year	45.91 %	77	64	302.4 GHz	1.69 TB
Clus...	3	> 1 Year	65.51 %	312	82	247.82 GHz	1.68 TB

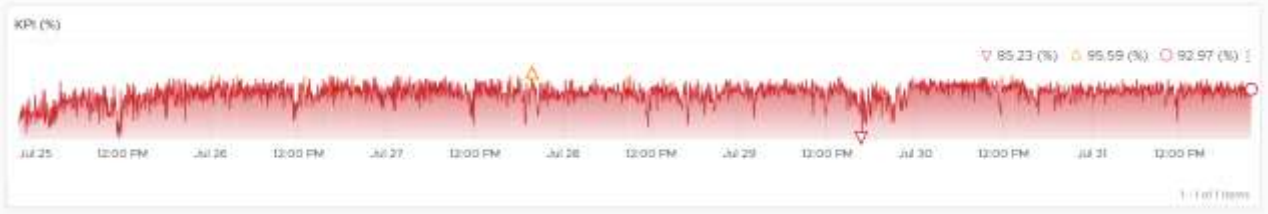
The number of ESXi hosts are color coded as smaller clusters have relatively higher overhead.

Cluster Analyzis

Select a cluster from the table. Its detail capacity will be automatically shown.

Performance

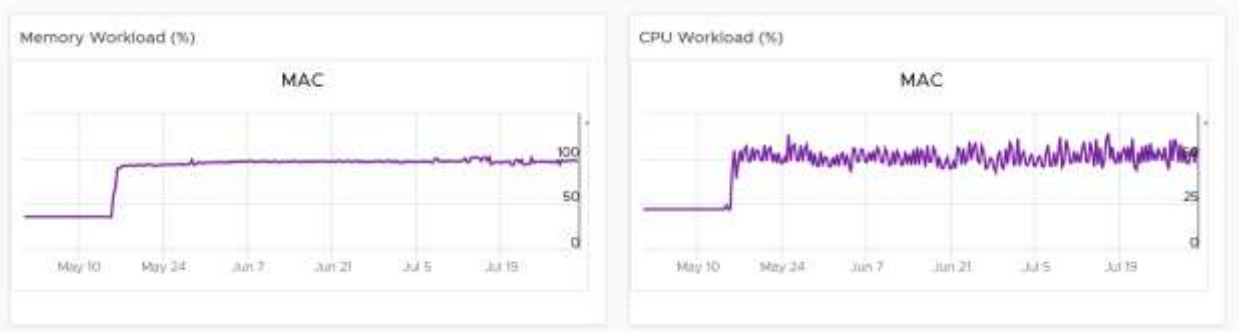
The first thing it shows is the cluster performance. Make sure this is within your expectation.



The counter is covered in Performance chapter and Performance dashboard.

Utilization

The next chart is utilization. It's showing in %, as it's relative to your usable capacity.



Utilization displayed for three months and not one week. The daily average is displayed and not the hourly average, so you can focus on the overall trend.

For memory, the focus is on consumed memory and not active memory.

If it's useful to you, add peak to complement average utilization. The peak is defined as the highest among any ESXi hosts. If the peak is higher than the cluster-wide average, then the cluster is imbalance. Find out the cause of imbalance and optimize it.

If you have the screen real estate, add the absolute utilization.

Allocation

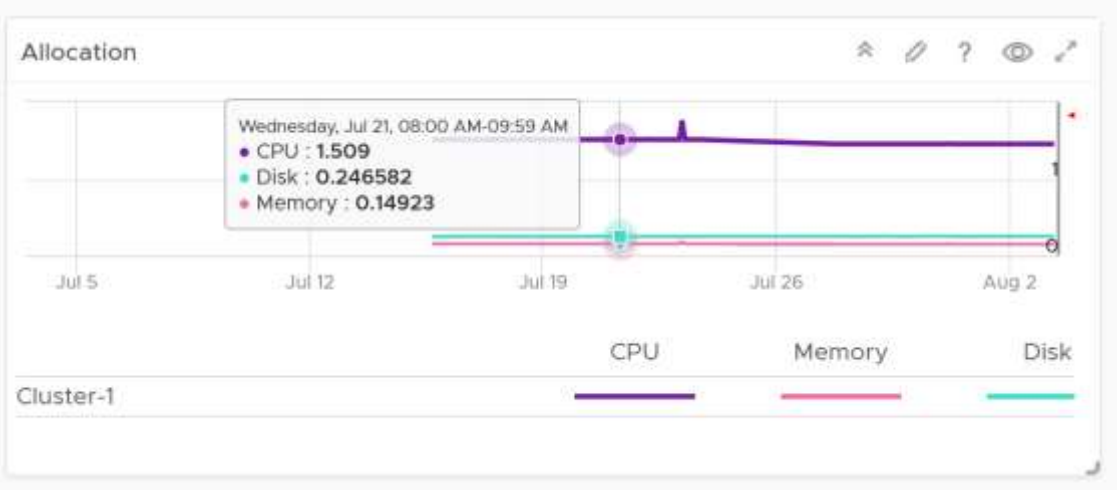
Next is Allocation, as not all demand is real. You need to use both models.

Workload can be low, but is overcommit ratio high? Newly provisioned VMs tend to be idle for weeks, and suddenly grow.

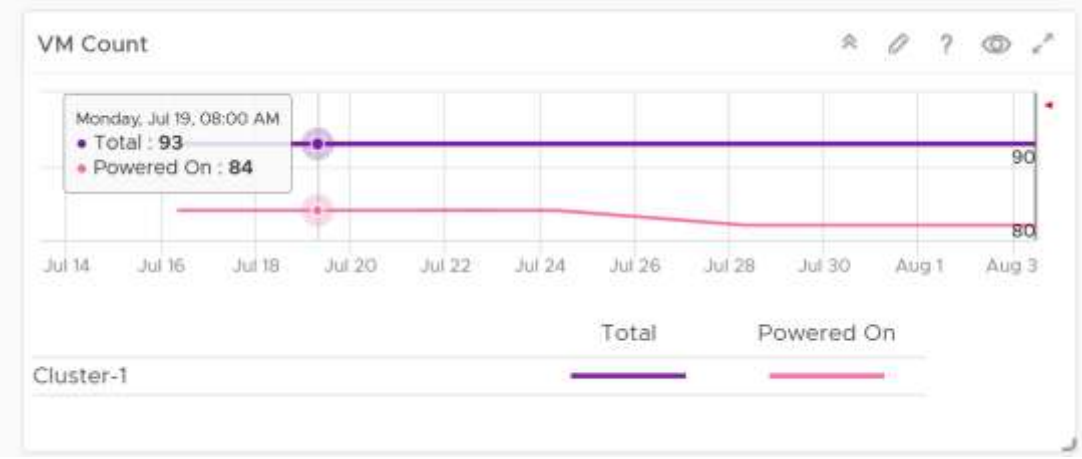


You can see the trend of the 3 components together on a chart if that is relevant to you. In general, your CPU overcommit should be the highest, followed by disk (due to thin provision). Memory overcommit tends to be near 1 due to its nature as cache.

Use the line chart to see the trend. As usual, the data is averaged hourly so you can focus on the big picture.



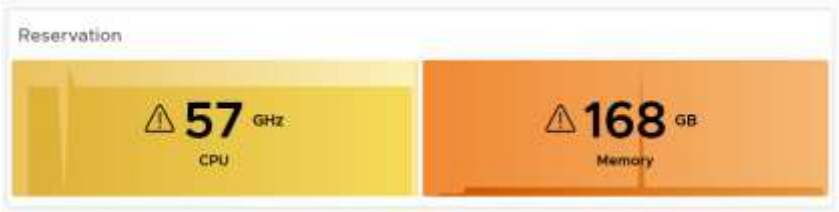
Next is the VM Count. A trend line of number of VM over time is important to spot if there are many newly provisioned VM. If you see VM growing but demand remains low, that's a sign of potential demand coming up in the future.



BTW, the overall VM:ESXi ratio is a common number cited to senior management. It’s often measured as cost efficiency proof.

Reservation

Reservation can impact the efficiency of your cluster. Why is the cluster low on capacity? Is it because of real workload, or just reservation?



If your cluster size varies, complement the reservation number by showing relative value. Once you have a standardized number, you can visualize them on a heat map! This enhancement requires super metric, good practice! 😊

Reclamation

As covered in the Capacity Management chapter, there are 6 types of reclamation. Some of them are shown below.

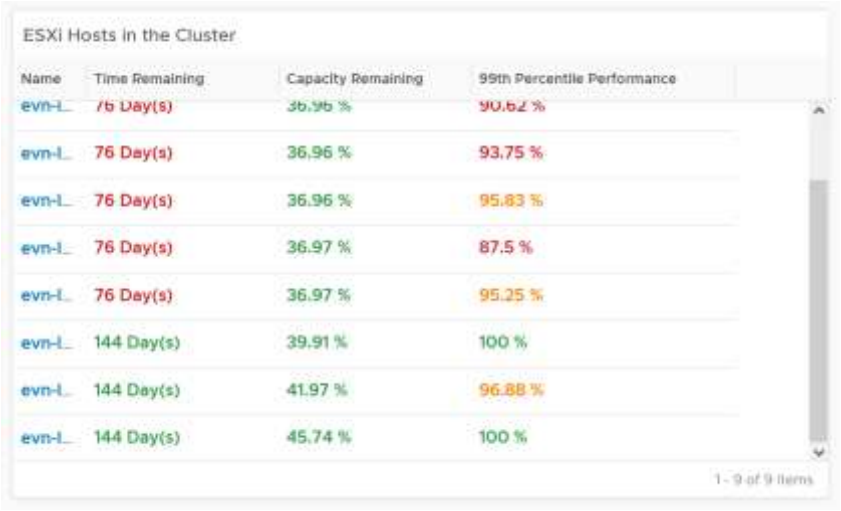


If it's relevant to your environment, add the number of undersized CPU and undersized memory.

ESXi Analyzis

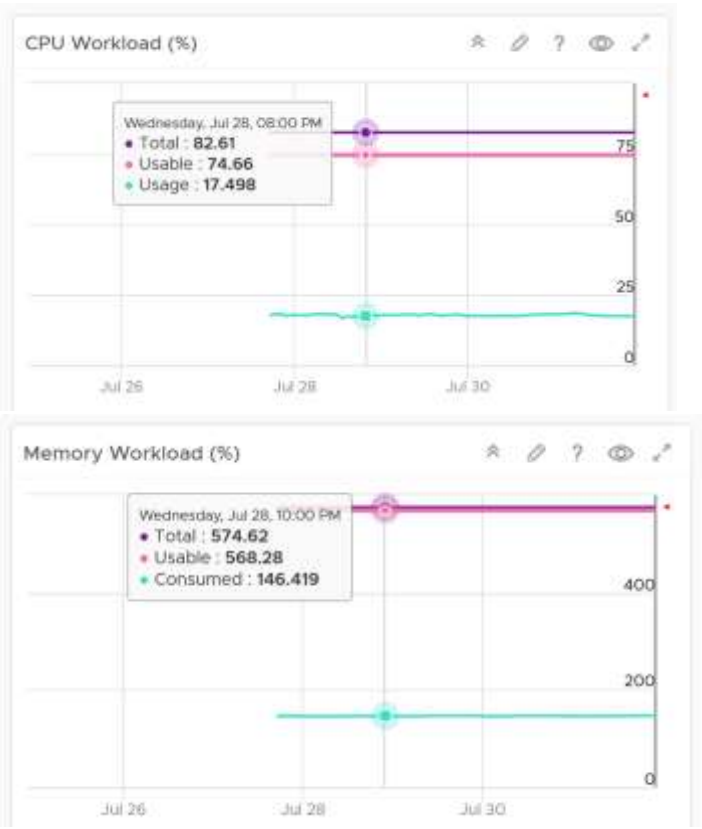
Just because the cluster capacity is good, does not mean there is no issue at ESXi level. Imbalance is common issue, especially in large cluster and stretched cluster.

The following table shows all the member ESXi. You can see the imbalance clearly, thanks to the color code.



The 99th percentile Performance column takes the 99th percentile value of the ESXi Performance (%) metric. The reason we're not taking the worst performance (which is equivalent to 100thpercentile) is to rule out outlier. In addition, the performance threshold has been set to be stringent.

Select one of the ESXi. All its detail will be automatically shown.



Both CPU and memory trend line charts show if it's steady demand, cyclical demand, rising demand, or declining demand. The trend is as important as the present value. To see trend, you need to see over longer time. Utilization is displayed for three months and not one week. The daily average is displayed and not the hourly average and the focus is on memory consumed and not memory active. Note that memory consumed includes the total memory consumed, so it includes memory consumed by VMkernel.

Both total and usable utilization in terms of memory and CPU are displayed. This gives you the absolute amount of capacity.

A technology refresh is often used to address capacity shortage. The relevant configuration shows the hardware model and specification to help you determine the age of the hardware.

VM Analyzis

What's causing the low capacity remaining? Which VMs are impacting what infrastructure resource (CPU, memory, disk space)?

Use the following table to analyze. It lists either the VMs in the cluster or host.

VMs in the selected Cluster or Host

VM Name	Guest OS	vCPU	Memory	Disk	# Disks	Creation Date
AVS-Clo...	Other 3.x Linux (64-bit)	4	12 GB	80 GB	2	February 8, 2021 3:5...
CentOS-...	CentOS 4/5 or later (64-bit)	2	1 GB	8 GB	1	February 8, 2021 9:2...
CentOS-...	CentOS 4/5 or later (64-bit)	4	2 GB	8 GB	1	February 8, 2021 8:2...

You can select one of the VM, and additional relevant information will be shown.

You can enhance this by using 2 heat maps, one for CPU and one for memory. Shows all the VM, sized by configured and color by capacity remaining. If you see many large VMs running low on capacity, that means you should stop provisioning until you upsize existing VMs first.

Datastore Capacity

This dashboard complements out of the box capacity pages and dashboard. It focuses on storage, provides overall picture, and highlights the datastores that need attention.

Overall Analyzis

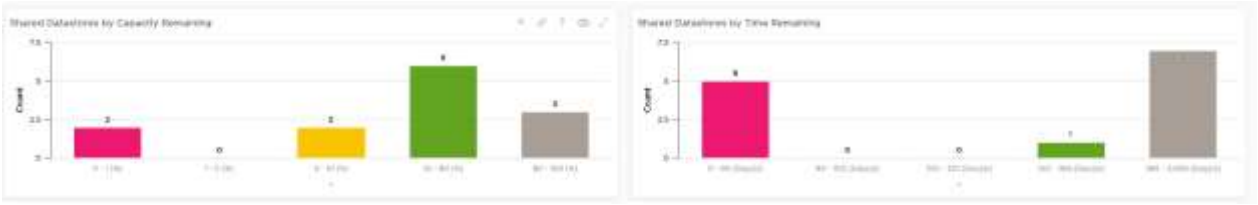
"What do I have in terms of capacity?"

The summary banner answer basic questions like this.



If required, customize it. I keep it simple as you don't actually have 1 contiguous storage pool. Just because you have 100 TB available in South Pole DC does not mean your servers in North Pole DC can use it, unless you tunnel dark fibre over earth core.

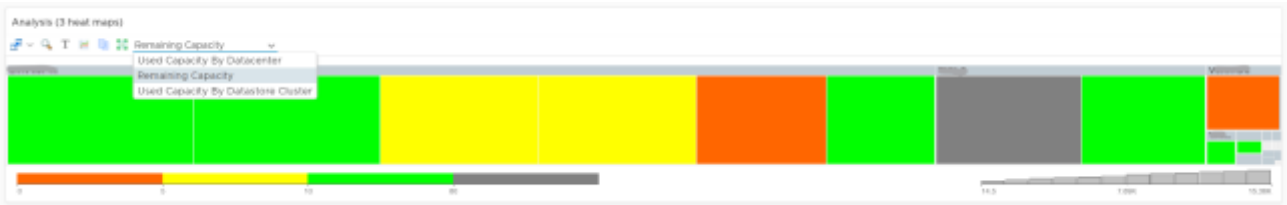
Next is the distribution chart. Do they look familiar?



It's intentionally designed to be consistent with the Cluster Capacity dashboard.

As you can expect by now, next in the dashboard is the heat map.

There are three heat maps, the primary being Remaining Capacity heat map. The 2 other heat maps cover Used Capacity. One of them is designed for environment that use Datastore Clusters.



Each box represents a datastore. If you have many datastores, the heat map will group them. You can drill down to see its members. The larger the datastore, the larger its box is. If you have many small datastores, consolidation can make operations easier.

Following the design used in Capacity dashboard, next is the table listing all the **Shared Datastores**.

Shared Datastores

Name	Capacity Remaining	Total	Remaining	Time Remaining	VM	Orphaned Disk	Snapshot
vsa...	0 %	19.65 TB	0 TB	0 Day(s)	56	4.39 TB	0 GB
vsa...	58.96 %	41.92 TB	18.54 TB	> 1 Year	93	87.93 GB	1.01 GB
ev...	88.56 %	10 TB	8.86 TB	> 1 Year	19	7.96 GB	0 GB
DO...	96.91 %	11.75 TB	11.38 TB	> 1 Year	8	99.34 GB	0 GB
To...	-	83.31 TB	38.78 TB	-	176	4.58 TB	1.01 GB

The table provides a summary, showing all datastores a glance. They are grouped by Data center. If you use Datastore Cluster as your standard and it suits you better, replace the grouping with it.

By default, the table is sorted by the least capacity remaining.

There are 3 reclamation opportunities: powered off VM, snapshot, and orphaned VMDK.

Why is Idle VM not included?

Because you should power off first, and let it go through powered off period before deletion.

Datastore Analyzis

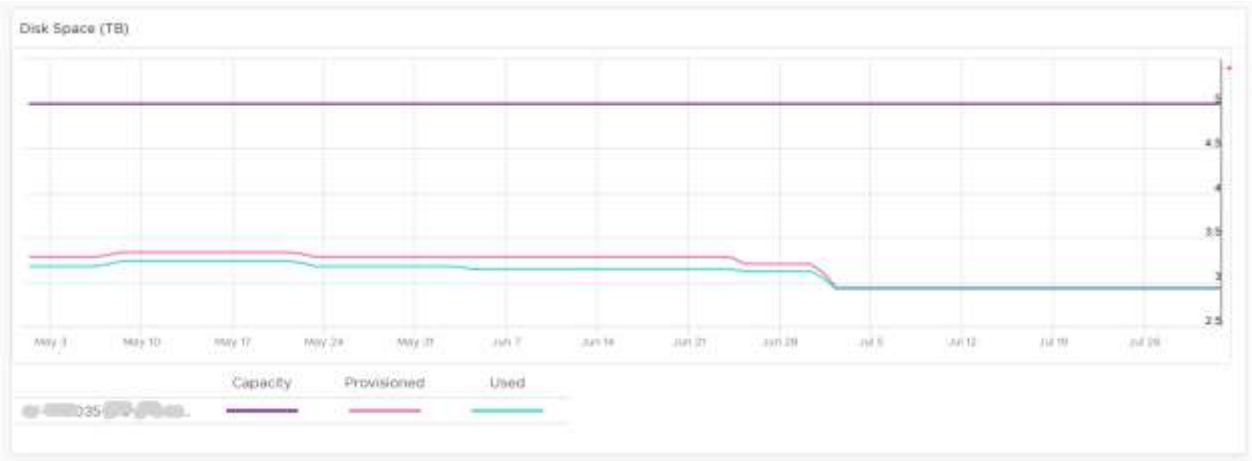
Select a datastore from the table. Its detail capacity will be automatically shown.



The snapshot should be 0 GB. If it is not 0, then it should be temporary. A snapshot lasting beyond a few days should be investigated.

Orphaned Disks are VMDK files that are not associated to any VM. Expect it to be 0.

For disk space, the total capacity, allocated and actual used are shown. What does it mean when the provisioned disk space metric grow but the actual used does not?



That means the VMs are yet to use it. Watch out, you can run out of space sooner than expected.

The value is daily averaged to show the big picture. Having 5-minute spike can visually make the chart harder to read.

If you use allocation model, add a second line chart showing the allocation over time. Add a threshold so you know how close you are.

VM Analyzis

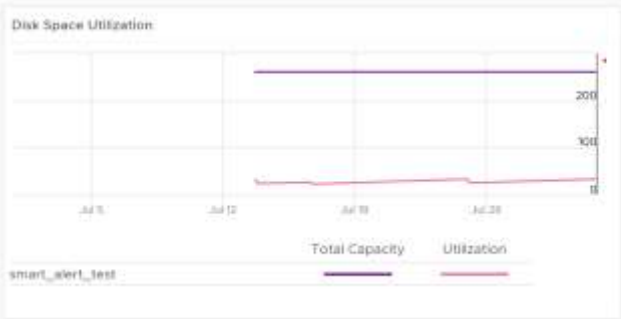
If you need to analyze at VM level, review the table listing all the VMs in the datastore.

VMs in the Datastore

Name	Used	Used	Provisioned Space	Snapshot	Powered Off?	Idle?	Age
ale...	7.67 %	21.01 GB	274 GB	0 GB	0	0	1 Day(s)
ma...	9.29 %	25.45 GB	274 GB	0 GB	0	0	2 Day(s)
he...	12.75 %	34.94 GB	274 GB	0 GB	0	0	3 Day(s)
lat...	12.99 %	35.6 GB	274 GB	0 GB	0	0	3 Day(s)
Al...	14.18 %	38.86 GB	274 GB	0 GB	0	0	5 Day(s)
loc...	16.79 %	46.09 GB	274.54 GB	0 GB	0	0	10 Day(s)
Su...	202.68 %	375.05 GB	2,204.1 GB	0 GB	-	-	-

If you have hundreds of VMs, create a heat map showing all the VMs. Color it by the used space so you know if the VM is reaching its full size. Size it by the allocated size so you can see who the big VMs are. If the big VMs have plenty of space to grow, watch out.

Click on the VM you want to investigate further. You get to see its usage over time.





Local Datastores

They are shown separately as a table on its own, at the end of this dashboard. Avoid running VMs on local datastores, unless its storage requirements can be met with a local disk and it does not need vmotion.

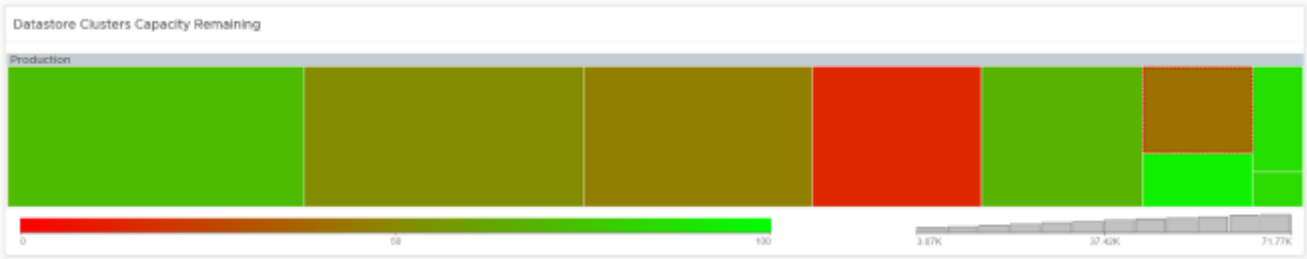
Local Datastores Capacity

Name	Total	Remaining	VM	Datacenter	ESX Hosts	Type	VMFS	vSphere Tag	vSphere Folder
Datastore-98	6.53 TB	5.92 TB	2	EVN	1	VMFS	6.82	none	
Datastore-99	6.53 TB	5.76 TB	8	EVN	1	VMFS	6.82	none	
Total	13.06 TB	11.68 TB	10	-	-	-	-	-	-

Points to Note

If you are using “thin on thin”, meaning the underlying LUN is also thin provisioned, add visibility into the physical array.

The dashboard does not have datastore clusters. If your environment use it, modify this dashboard or create a new one. In a large environment with many datastores and datastore clusters, add a View List to list the datastore clusters so you get summary information. From this list, drives the datastore view list. Alternatively, create a heat map, listing the datastore clusters.



vSAN Capacity

The **vSAN Capacity** dashboard complements the **vSphere Cluster Capacity** dashboard by showing vSAN related capacity. It focuses on the storage and vSAN specific metrics, and does not repeat what’s already covered. That’s a long winded way to say you gotta use both dashboards to manage vSAN capacity 😊

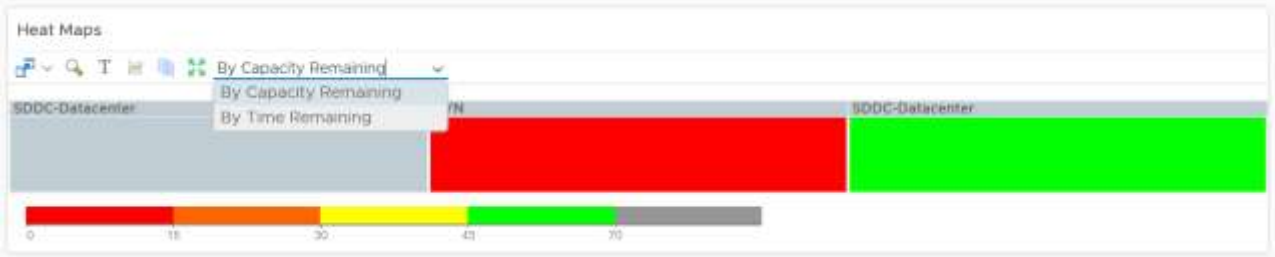
Because the metrics are vSAN specific, this dashboard does not list non vSAN cluster.

Overall Analyzis

The dashboard starts with the familiar bar-chart combo that Cluster Capacity dashboard has. The difference is this focuses on the vSAN disk space, not compute and network.



Similar to Cluster Capacity dashboard, next is a heat map. There are 2 heat maps provided.



Why is the box size made identical?

Head to the Cluster Capacity dashboard for the explanation.

As you can expect, a table provides the next level of details. The difference is you get vSAN specific metrics.

vSAN Clusters

Name	Capacity Remaining	Time Remaining	Configured	Hosts	Disk Groups	Capacity Disks
vSAN Clust...	-	-	-	3	6	18
vSAN Clust...	0 %	0 Day(s)	19.65 TB	9	9	18
vSAN Clust...	58.41 %	> 1 Year	41.92 TB	3	6	24

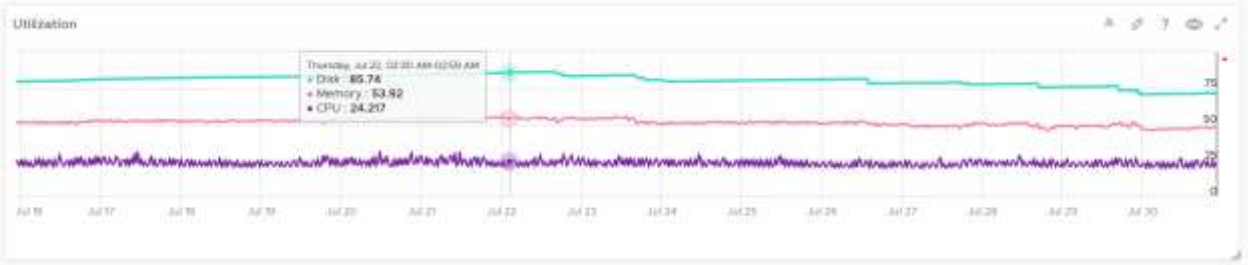
If it's useful, add metrics such as total disk capacity and used.

Cluster Analyzis

Select a vSAN cluster from the table. Its detail capacity will be automatically shown.

Utilization

It's showing the utilization for all 3 elements, as you need to consider all three. Network is not shown as typically it's not an issue and it's complex to model.



Just like physical array, there can be hot spot and imbalance. The following heat map shows individual disk groups.



Notice they are all green but not identical. It's normal for them to have minor variance.

Reclaimable

It's a key component of proactive capacity management. The dashboard shows you both the VMs and non VMs portion.



Dedupe and Compressed

The scoreboard shows details on this area.

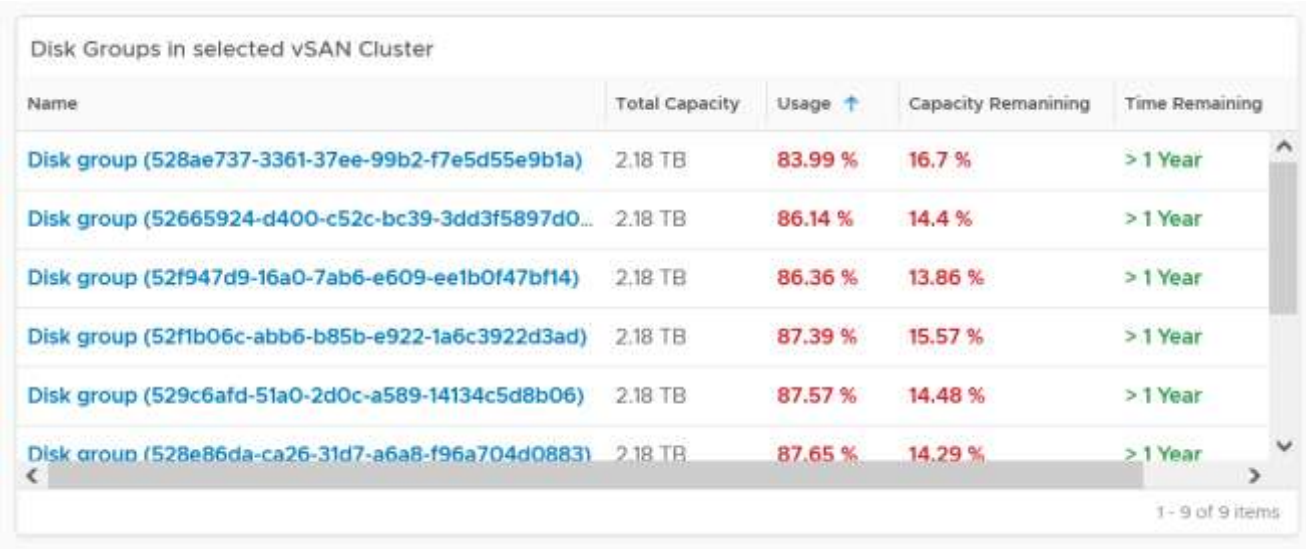


If you are concerned with the CPU usage from the dedupe and compress feature, add a line chart of vSAN CPU usage.

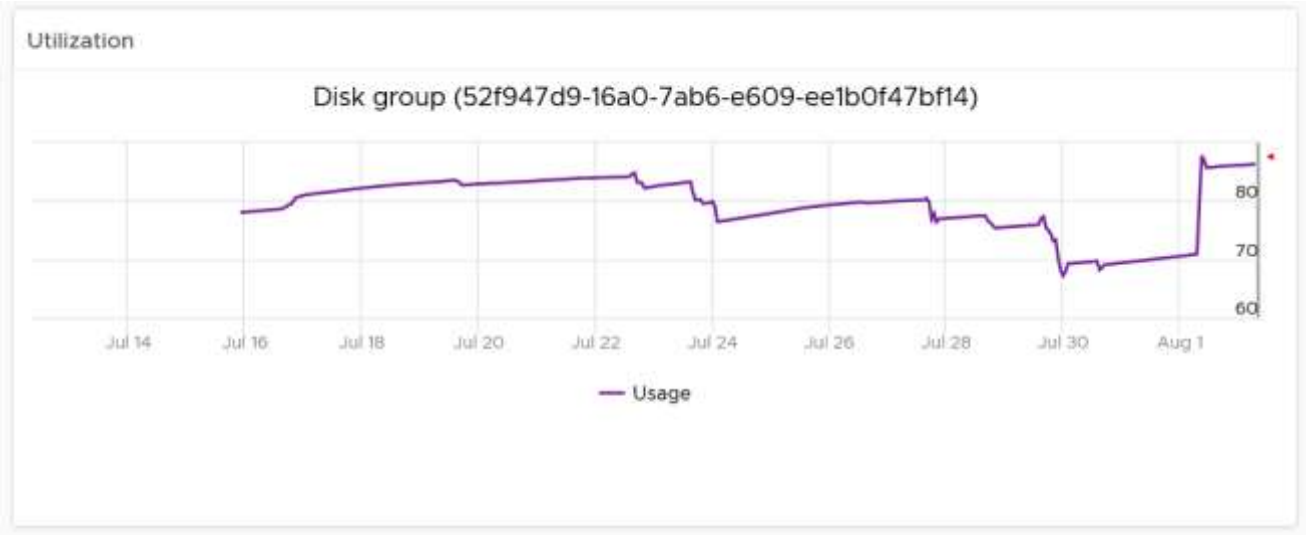
Disk Group Analyzis

If there is imbalance, you can drill down into each disk group. Otherwise, skip this section.

The following table shows all the disk groups in the cluster. Their usage may not be similar but they should not deviate drastically.



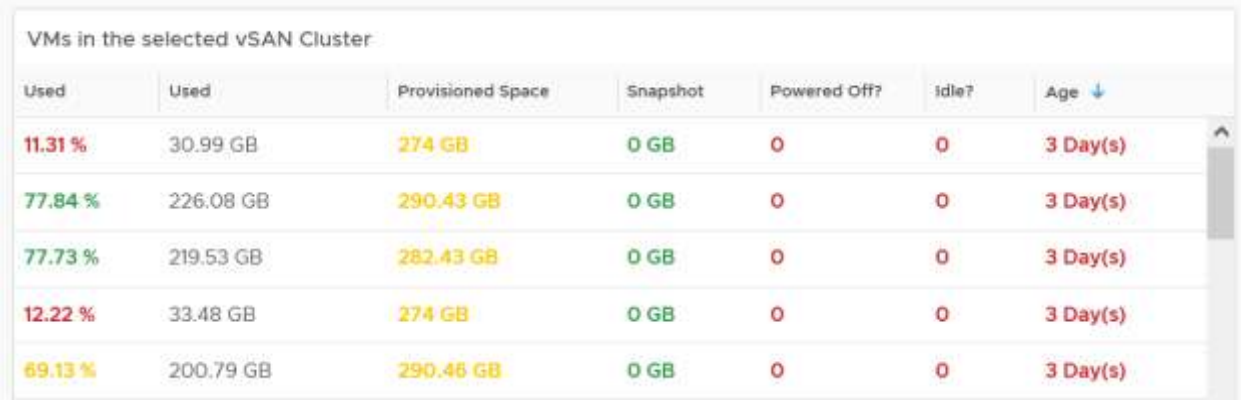
To see any of the disk group usage trend, click on it.



In addition a heat map is provided to see the usage among the capacity disks in the selected disk group. Expect them to be uniformed at this level as it's RAID striping.

VM Analyzis

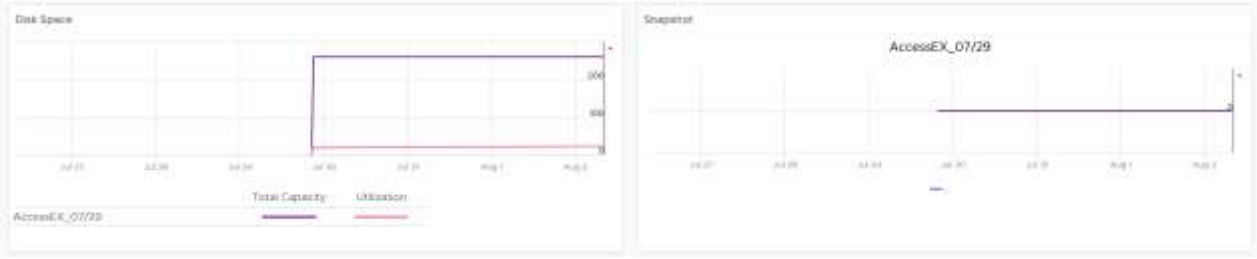
You can get down to individual VMs in the selected cluster, and check their usage and snapshot.



VMs in the selected vSAN Cluster

Used	Used	Provisioned Space	Snapshot	Powered Off?	Idle?	Age
11.31 %	30.99 GB	274 GB	0 GB	0	0	3 Day(s)
77.84 %	226.08 GB	290.43 GB	0 GB	0	0	3 Day(s)
77.73 %	219.53 GB	282.43 GB	0 GB	0	0	3 Day(s)
12.22 %	33.48 GB	274 GB	0 GB	0	0	3 Day(s)
69.13 %	200.79 GB	290.46 GB	0 GB	0	0	3 Day(s)

To see the usage trend, select the VM.



In addition, the relevant configuration of the selected VM is also shown.

VM Capacity

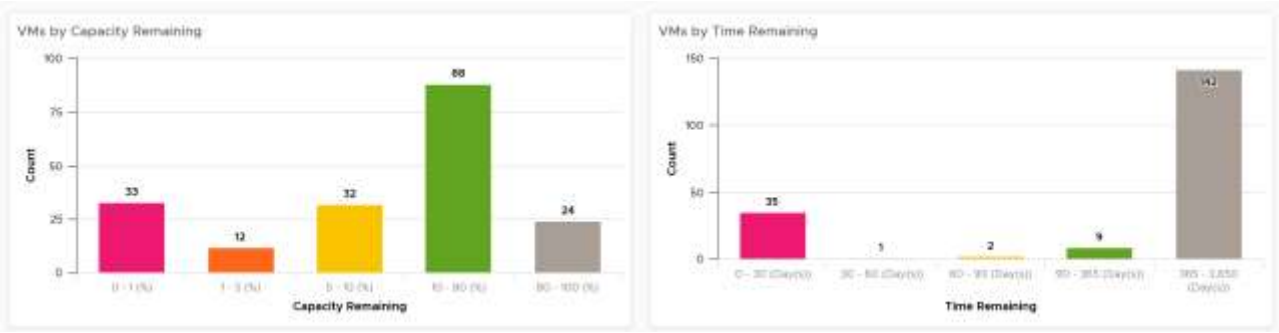
The dashboard helps you analyze the capacity of all the VMs, with ability to drill down into each VM.

Overall Analyzis

The dashboard begins with a table listing all your data centers.

vSphere World is also included so you can see all the VMs from all data centers. Unlike infrastructure objects, there are potentially tens of thousands of VM. So take note that the charts will take longer due to refresh if you select vSphere World.

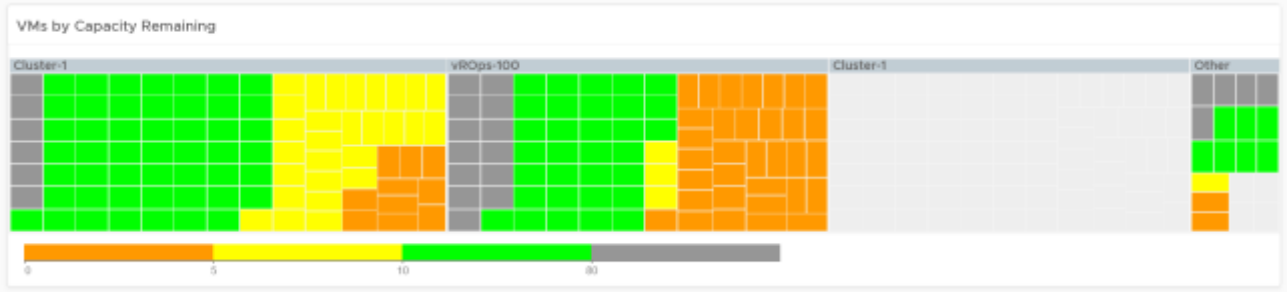
How do the following 2 charts provide insight into the capacity of all the VMs in your environment? What's the ideal distribution?



The first chart groups them by capacity remaining, while the second one by time remaining. Ideally, you want all of them to be low on capacity remaining (meaning the given capacity is actually fully used) but high on time remaining (meaning they do not need additional).

The bucket size has been designed to map the default settings in VM capacity policy. If you change it, also change the value in policy so it's consistent.

The heat map provides additional view by grouping them by cluster. It helps you spot which cluster is at risk (majority of its VMs need more capacity) and which cluster can provide extra resource (majority of its VMs are not using their capacity).



Review the heat map. It provides the next level of details by grouping the VMs by clusters, so you can see which clusters need attention.

Note the VM size has been standardized for better visualization. If it suits your capacity team better, add the size. Note that by doing that you have to pick CPU or Memory, so you may have to create 2 heat maps.

Review the table listing all VMs in the selected data center.

VMs Capacity in the Selected Datacenter

Name	Time Remaining	Capacity Remaining (CR)	CPU CR	Memory CR	vCPU	Memory	Creation Date
Filter...	0 Day(s)	0 %	4 GHz	0 GB	2	8 GB	-
7/30	0 Day(s)	0 %	2.47 GHz	0 GB	2	8 GB	3:13:08 PM GMT+08:00
Expo...	0 Day(s)	0 %	7.73 GHz	0 GB	4	16 GB	July 20, 2021 7:09:21 PM GMT+08:00
lates...	0 Day(s)	0 %	3.2 GHz	0 GB	2	8 GB	July 28, 2021 8:25:49 PM GMT+08:00
Shop...	0 Day(s)	0 %	2.81 GHz	0.56 GB	2	1 GB	November 10, 2020 5:40:14 PM GMT+08:00

The list is sorted by the VM with the least capacity remaining. If sorting by Time Remaining suits your operations better, modify the dashboard.

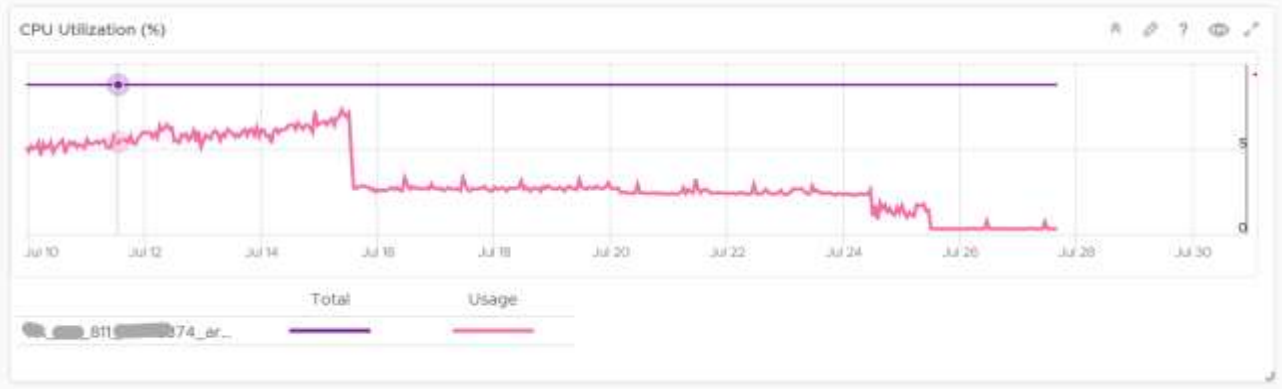
The table is also color coded. Take note that the threshold is unable to show the grey (wastage) color.

The creation date is added for the same reason in idle VM case.

VM Analyzis

Select a VM from the table. Its detail capacity will be automatically shown.

You get both the CPU and memory trend over time. 3 months data is shown, and they are averaged to hourly so you can see the overall trend instead of spiky chart.



For memory, it uses Guest OS, for the same reason explained in VM Rightsizing dashboard.



For the disk, it uses the Guest OS partition. Do not use VM virtual disk as there may not be 1:1 mapping to the actual partition.

Disk

Partition	Capacity	Used	Used
/	15.56 GB	4.94 GB	31.76 %
/boot	0.12 GB	0.04 GB	34.94 %
/storage/core	19.56 GB	1.04 GB	5.33 %
/storage/db	205.7 GB	11.75 GB	5.71 %
/storage/log	19.56 GB	2.63 GB	13.43 %

Right-sizing recommendation is also shown for both CPU and Memory. Unlike physical server, it's important to right-size VM for the benefits listed here.

For CPU, the CPU Usage counter is used instead of Demand. Use the knowledge you learned here to figure out why.

For disk, it's showing at the Guest OS partition level. There is no overall capacity at VM level because different partitions have different capacity.

Relevant Configuration

The relevant configuration is automatically shown to give context to the VM.

Information such as VM Owner and business units can be useful in the analyzis.

Reclamation

The Reclamation dashboard helps you managing various types of reclamation that can be done on VMs and datastore. It is designed for both the Capacity team and the Operations team.

Overall Analyzis

The scoreboard provides a summary of the total reclamation. Guest what infrastructure resource is missing from the scoreboard?



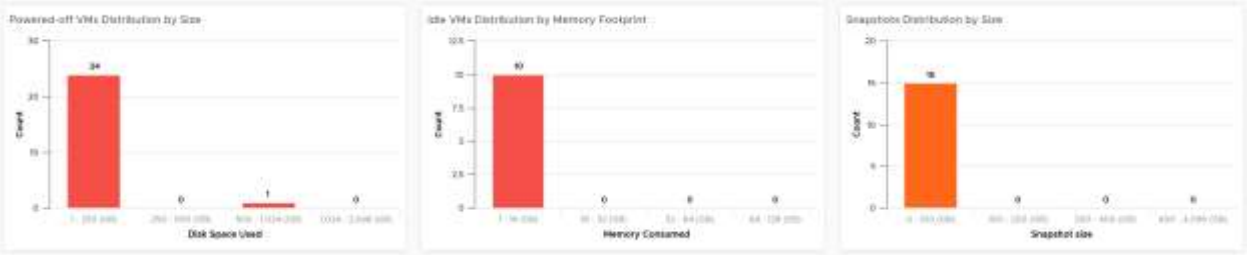
CPU is missing. Because you're reclaiming blank air.

You can drill down at either Data Center level, cluster level or datastore level. Datastore level is required as orphaned disk does not have VM association, hence it's not related to any cluster.



The table above can drive all the charts underneath them, shown below, giving you a flexible way to slice the information. Take note that the datastore table *only* drives the snapshot table. The reason is traversal spec. The view widget can only use 1 traversal spec.

The summary information will be automatically shown. To show from all clusters, select vSphere World. This object covers all clusters. Take note that the charts will take longer due to refresh due to higher amount of data.



If necessary, adjust the bucket size in the charts to suit your operational requirements.

The reclamation potentials are presented as 3 bar charts, each corresponds to an area you can reclaim:

- Powered off VMs that are no longer needed contribute to wasted disk usage. Consider deleting them to free up space or moving them to archival storage.
- Idle VMs are running, but not being used actively. These VMs consume memory that may be used by active VMs. Consider removing these VMs to reduce memory contention. Guess why is VM-level metric used instead of Guest OS metric?
  Memory reclamation is based on the memory footprint at the parent ESXi. The value inside the Guest is not what is being reclaimed, and so it is irrelevant
- Snapshots are meant to be temporary and can cause performance issues and waste disk space if not deleted after a few days.

Focus on snapshot first, as it does not involve changing VM.

Next is powered off VM. The longer it has been powered off, the lower the risk of deletion. Ideally, you want a confirmation from the owner, and have a back up outside vSphere. This is why it's important to have meta tag such as VM Owner email address. Note if the VM has snapshot, it's included in the calculation, so there can be *double counting* since the VM also appears in the snapshot table.

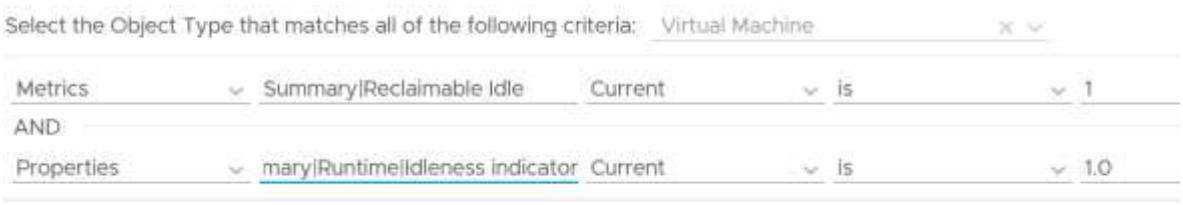
Idle VM

As the VM is still running, it's relatively harder to reclaim as shutting it down may require approval from the VM Owner.

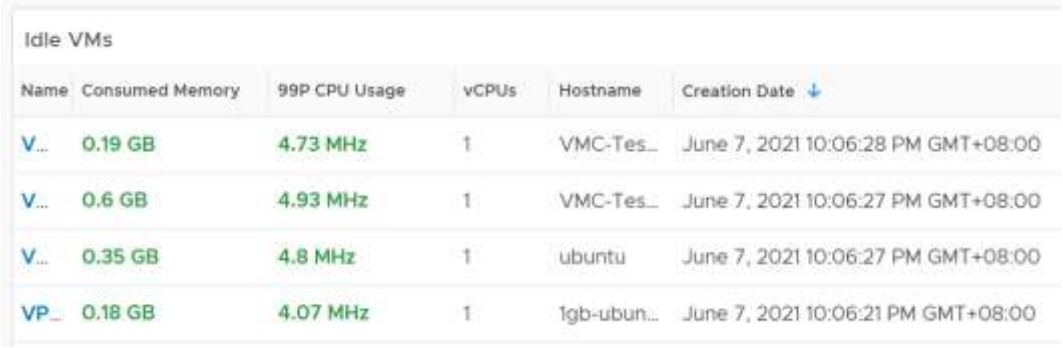
BTW, if there is nothing to reclaim, you get a blank table. The following example shows no VM meets the criteria for either powered off or idle. If you suspect the table is wrong, review your criteria.



We discussed in Capacity chapter that both metrics need to be true to ensure the list does not contain VMs that has recently become active.



Let's drill down into idle VM as that's the most complex part. Guess why we show the 99th percentile CPU usage?



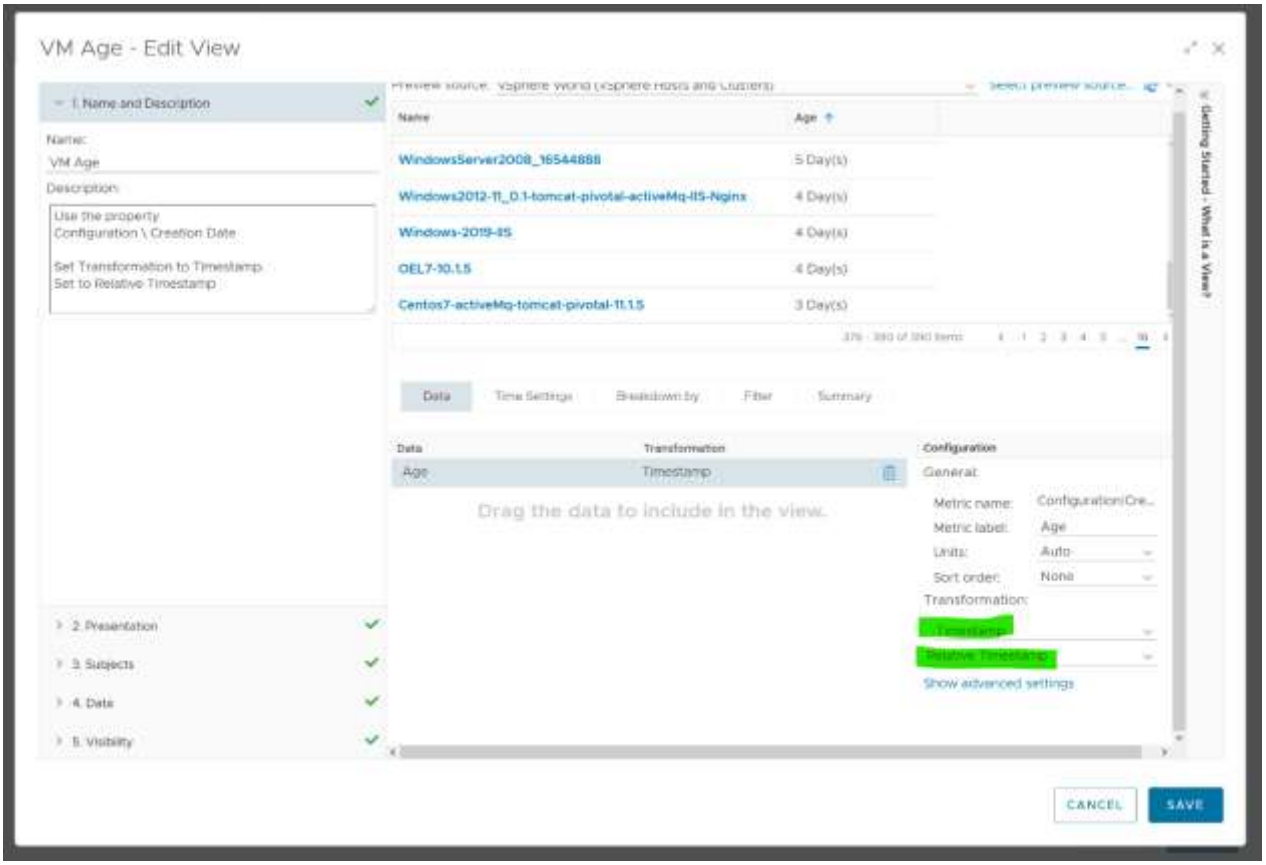
Idle VMs

Name	Consumed Memory	99P CPU Usage	vCPUs	Hostname	Creation Date
V...	0.19 GB	4.73 MHz	1	VMC-Tes...	June 7, 2021 10:06:28 PM GMT+08:00
V...	0.6 GB	4.93 MHz	1	VMC-Tes...	June 7, 2021 10:06:27 PM GMT+08:00
V...	0.35 GB	4.8 MHz	1	ubuntu	June 7, 2021 10:06:27 PM GMT+08:00
VP...	0.18 GB	4.07 MHz	1	1gb-ubun...	June 7, 2021 10:06:21 PM GMT+08:00

The 99P CPU Usage shows the CPU usage at 99th percentile during the time period. It's a handy way to check if it's indeed idle. Essentially, if it the CPU utilization is low for 99% of the time, the VM could indeed be idle.

Why is the Creation Date especially important for Idle VM case?

The answer is newly provisioned VM may remain unused for months. The following show how to set it.



Choose VM property Configuration \ Creation Date, then choose the Transformation: Timestamp and then Relative Timestamp.

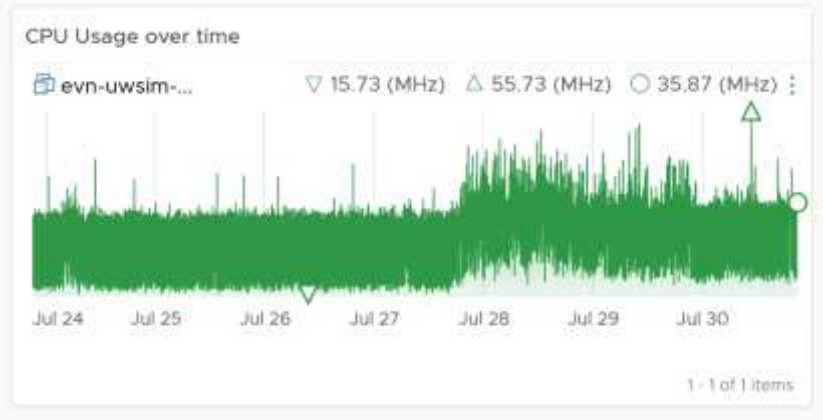
VM Analyzis

To analyze VMs for reclamation opportunities, select a VM from one of the three tables (Powered Off VM, Idle VMs or VM Snapshots). Its detail value over time will be automatically shown. Notice the VM name is the same in all the 3 charts.



Powered off over time shows VM status (on/off) over time. 1 = true, meaning the machine is powered on.

CPU Usage over time provides insights into the aggregate CPU usage, including peak usage periods. This way you can validate that an idle VM has not had any brief usage. What you want to see is green over a long period of time, as shown in the following example.



If the snapshot is expanding rapidly, ensure that the VM disk is small (relative to size of the underlying datastore) as it can fill up the datastore.

Lastly, the relevant configuration of the VM is also shown to give context.



Points to Note

If your environment is large, change the dashboard filter to a functional filter. Group by the class of services such as gold, silver, and bronze and default the selection to the least critical environment. In this way, you can be more active in reclamation.

If reclaiming is a long drawn manual process in your organization, add a filter by department or VM owners. To organize your reclaim efforts it is helpful to create custom groups to make it easier to filter by department or VM owner. This can make it easier to seek approvals and communicate with anyone who may be impacted.

You should enhance this to include Trim and Unmap. Happy to collaborate and make this into the product. We need to check on only the thin provisioned disk. We should also check at the array level, using the TVS adapter.

Rightsizing

The VM Rightsizing dashboard helps you in adjusting the VM size for best long term performance at lowest cost. It covers both undersized and oversized scenarios. It's not designed for short term, 5-minute performance burst. The reason is high utilization is actually good for performance. As a result, this dashboard is categorized under Capacity and not performance.

It is designed primarily for Capacity team, not day to day operations team.

Before vs After

How do you minimize the chance of rolling back? This means the downsize has to be undone and the VM was increased further because of performance problem. If you do that, do you notice that sometimes it did not solve the slowness, or make it even worse?

To address this, you need a proof that the downsize did not cause the slowness. Something else did. In theory, a reduction should result in:

Contention	These metrics should not increase. In fact, they should decrease, proving the theory that smaller size results is more efficient.
	It should be measured at 20 seconds, not 5 minute.
	This should be measured at component level, as measuring at aggregate level can mask out problem.
Consumption	There are 2 types of consumption metrics. • Absolute. This should remain the same. • Relative. This logically will go up.

	This should be measured at the aggregate level.
	What if the consumption metrics go down? That can impact performance, as the amount of work done goes down.

What is the complete list of metrics?

CPU Consumption	Thread Count	Do we split active threads, etc?
	Thread Migration	
	VM Usage (GHz)	If not, increase the share.
	VM Run Time (ms)	Run Time should be the same
	OS System Time	Can reveal if performance degradation is due to changes in workload
	OS User Time	
CPU Contention	OS Queue	
	VM Ready	The highest among any vCPU, measured at 20-second peak
	VM Co Stop	
	OS Context Switch	May go higher due to less CPU, but lower due to less ping-pong
	VM NUMA Migration	Does not exist yet in VCF Ops. Need esxtop.
Memory Consumption	OS Used Memory	This may go down
	OS Free Memory	This is likely to go down. So long it's
Memory Contention	OS Page-in + Page-out	This may go up
	VM Balloon	This is likely to go down. If it's 0, it should remain 0.
	VM Zipped & Swapped	
	VM Latency	
	VM % Remote Memory	Does not exist yet in VCF Ops

In theory, reducing CPU and memory should not impact disk and network. Only check them to troubleshoot the cause of the slowness.

Disk Consumption	VM Disk IOPS	
	VM Disk Throughput	
Disk Contention	VM Read Latency	The highest Read or Write latency at any vDisk, measured at 20-second.
Network Consumption	VM Packets/seconds	Spilt into Transmit and Received
	VM Throughput	Sum of Transmit and Received, measured at VM level.
Network Contention	OS Packet Dropped	
	VM Packet Dropped	

Overall Analyzis

The scoreboard provides a summary of the total undersized and oversized CPU and memory.



If the number is acceptable to you, that's it. No need to further analyze.

The reason why I'm not adding the number of VMs that are undersize or oversized is screen real estate. It needs 4 metrics to account for all the combination.

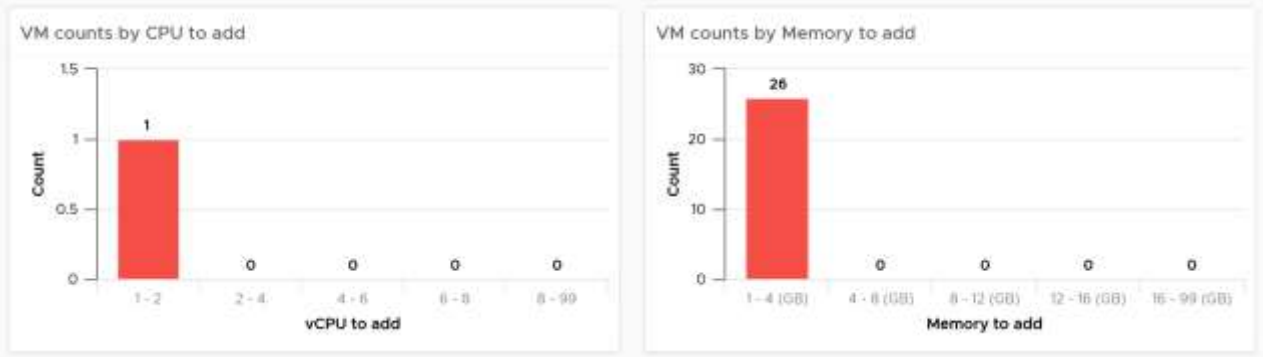
You can drill down at either Data Center level or cluster level. In most cases, rightsizing analyzis should be done at Cluster level as VMs typically do not move inter-cluster. At cluster level, the table looks like the following. Guess why are the cluster capacity metrics shown?

Clusters in Selected DC

Name	Time Remaining	Capacity Remaining	VM Remaining	ESXi Hosts	Running VMs
Cluster-1	-	-	-	-	-
vROps-100	> 1 Year	45.84 %	78	9	64
Cluster-1	> 1 Year	65.48 %	313	3	82
Total	-	-	391	12	146

The metrics are shown to give better context. Focus on reclaiming on cluster that is low on capacity remaining. For upsizing VMs, ensure the parent clusters have good capacity remaining.

The distribution charts showing the rightsizing is automatically shown. Let's dive into the undersized first as it's more common. Why are the vCPU to add focused on the first 8 vCPU, and memory to add focused on the first 16 GB?



The main reason is you want to avoid adding large amount without strong justification.

For CPU, the primary justification is the CPU run queue counter, not the CPU usage. Even if the CPU Usage is flat out 100%, if the CPU run queue is below 2 per vCPU, you don't need to add many vCPU.

For memory, Guest OS does not provide deep visibility into the depth of memory shortage. Adding memory may also requires changing application setting so it can take advantage of it.

Do the same capping for oversized VM. To make a drastic reduction, discuss with the VM Owner. Reducing memory may also requires changing application setting, especially application that manage its own memory such as JVM and DB.

Other than the bar charts, you also get the table listing the actual VM. You have 2 tables, one for undersized and one for oversized. Why can't we just use 1 table?


Undersized VM

Name	CPU to Add	Configured CPU	Memory to Add	Configured Memory	Creation Date
Al...	0	4	2 GB	16 GB	July 29, 2021 4:54:50 PM GMT+08:00
BA...	0	2	2 GB	8 GB	July 27, 2021 4:06:36 PM GMT+08:00
VS...	0	2	2 GB	8 GB	July 29, 2021 2:39:49 PM GMT+08:00
7/...	0	2	3 GB	8 GB	3:13:08 PM GMT+08:00
vR...	0	4	3 GB	16 GB	July 2, 2021 2:04:48 PM GMT+08:00
To...	2	91	37 GB	361 GB	-

21 - 25 of 26 items

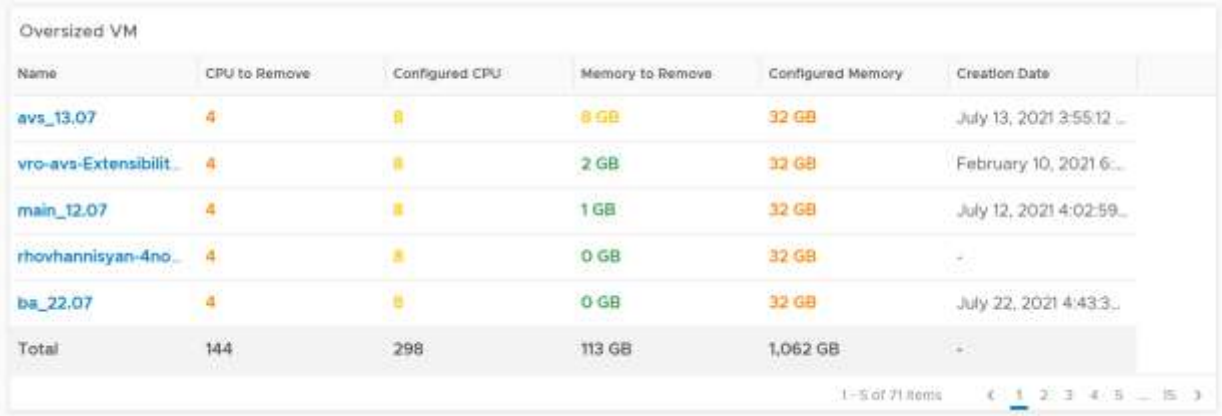
The reason for separate table is the business processes for oversized and undersized VMs are different, as one requires the affected VM to be shut down and the VM Owner to give back resources. For upsize, you want to add incrementally or even automate this process. For downsize, you want to remove in one change window as the effort to reduce is the same and there will be only one downtime.

CPU/Memory to remove is color coded to help you prioritize. Red color means they are excessively oversized.

Configured CPU/Memory is color coded to help you focus on the large VM. They are shown in red.

Take note that a VM can appear in both tables. It can be undersized for CPU, oversized for memory, and vice versa.

Here is the table for oversized.


Oversized VM

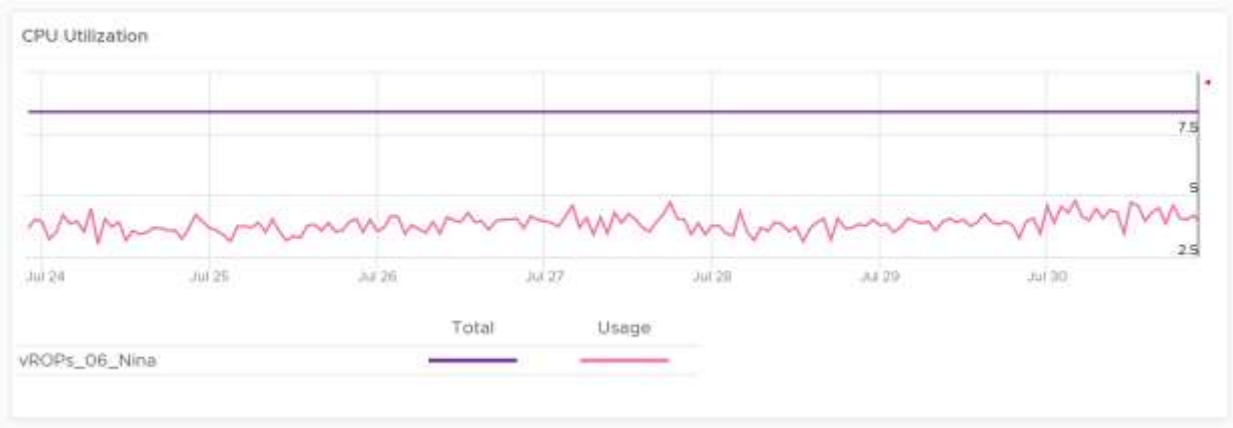
Name	CPU to Remove	Configured CPU	Memory to Remove	Configured Memory	Creation Date
avs_13.07	4	8	0 GB	32 GB	July 13, 2021 3:55:12 ...
vro-avs-Extensibilit...	4	8	2 GB	32 GB	February 10, 2021 6:...
main_12.07	4	8	1 GB	32 GB	July 12, 2021 4:02:59...
rhovhannisyan-4no...	4	8	0 GB	32 GB	-
ba_22.07	4	8	0 GB	32 GB	July 22, 2021 4:43:3...
Total	144	298	113 GB	1,062 GB	-

1 - 5 of 71 items

The metrics used are Summary \ Oversized \ Virtual CPUs and Summary \ Undersized \ Virtual CPUs. It stores the capacity engine calculation on recommended number of vCPUs that must be removed or added.

VM Analyzis

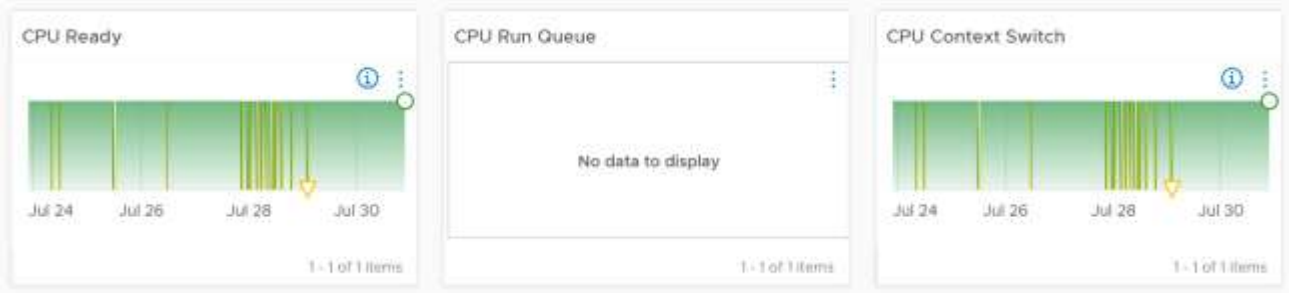
Select a VM to investigate further. You get its utilization automatically shown. Naturally, you expect the utilization to go higher if the VM is reduced.



Take note since the purpose here is capacity and not performance, you should not dive into individual vCPU utilization, nor increase collection granularity to shorter than 5 minute (e.g. to 20 second).

Having said that, rightsizing a VM can help improve its performance, or maintain its current performance. How do you prove that?

You show the performance bottlenecks, such as CPU ready and CPU run queue. The following is what you get out of the box.



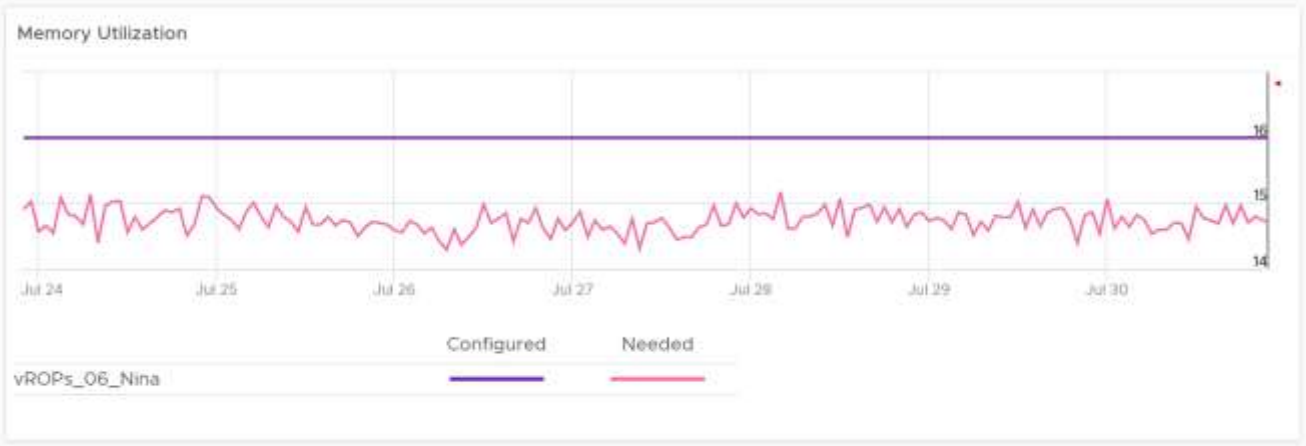
You should expect the VM contention metrics such as CPU Ready and Co-stop to either drop or remain the same.

You should expect the Guest OS CPU Run Queue to remain the same. It might go up a bit, so long it is lower than 3 per vCPU, no need to increase CPU size despite it's running high. If you want to be safer, modify the widget to use the 20-second counter.

You should expect CPU Context Switch to be less, as there are less CPU to switch. All else being equal, this is a positive change.

If you have the need, add more metrics such as Co-stop.

Memory utilization is collected from the Guest OS. If guest OS metrics are not available, then you will only see the **Configured** counter. Guess why isn't there a fall back to VM level memory?



It does not fall back to VM level counter as it's not a good replacement. See Part 2 Metrics Chapter 3 Memory Metrics.

Rightsizing memory improves performance by reducing memory ballooning and contention. For example, VMs with overprovisioned memory are more likely to experience ballooning

If you want to be conservative, add the free memory counter. If this never touch 0.5 GB, you do not need to add RAM despite its utilization.



If you have the need, add more metrics such as Page-Out rate.

Relevant configuration for the purpose of right sizing



Points to Note

If your environment is large, change the dashboard filter to a functional filter. Group by the class of services such as gold, silver, and bronze and default the selection to the least critical environment. In this way, you can be more active in reducing the oversized VMs.

For another example, see Dale Hassinger at code.vmware.com, where he has more data showing the proof that performance can improve with reduced size.

Monster VMs

You can enhance the dashboard if you need to convey the overall situation to senior management. You plot every single powered-on VM in the environment, sized by their vCPU. In this way, the monster VMs will be highlighted. A 64 vCPU VM will appear 64x larger than a single vCPU VM. This is good as the focus should be on the large VM, as discussed here.

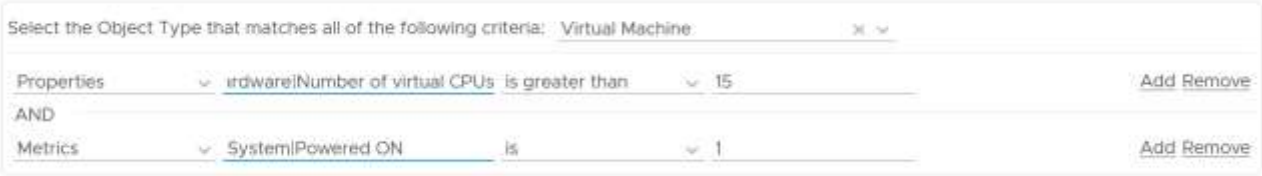
The heat map colors the VM by its capacity remaining. A VM with high capacity remaining means it has plenty of wastage resources.

What would it look like on an environment with many monster VMs that are oversized? You get something like this. The grey boxes dominate the space of the heat map. One of the boxes consist of 48 VMs with total > 1000 GHz. All of them have 87.79% capacity remaining. Another word, they are oversized.

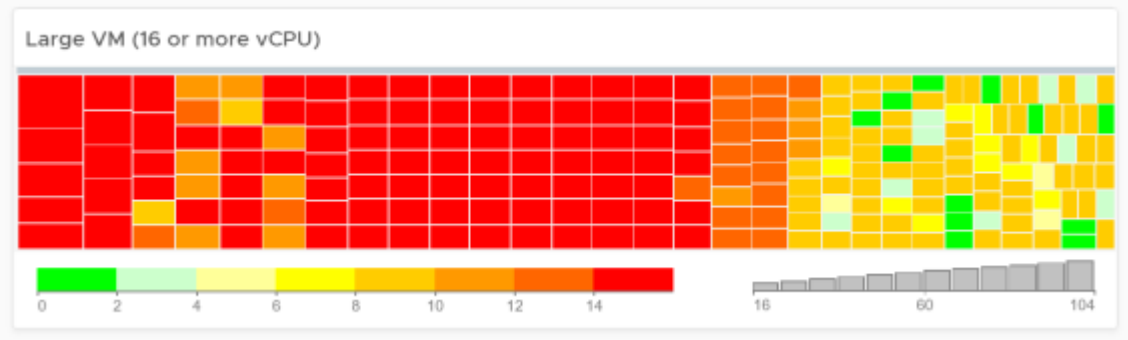


You may want to do the same for memory.

You can also focus on the large VM by specifying a filter. In the following example, I set to 16 vCPU or more, and the VM has to be powered on at present.



Since we only have the large VM, we can further refine the heat map. The size remains by vCPU size. However, the color changes to the number of oversized vCPU. The more I can reclaim the more red the color shows, and anything above 14 vCPU is red.



Configuration

Part 2 Chapter 6

The configuration dashboards aims to implement the concept covered in Configuration Management chapter.

Overall Design

Use the Configuration dashboards to view the key configuration in your VCF environment, especially for areas that need attention.

The suite of dashboards works together as one integrated set. They also have similar design.

Goals	Address configuration issues before they cause impact
	In the event of issues, check configuration to see if it plays a part.
Questions	Are the configuration matching best practices?
	Are the settings consistent when they have and up to date as per our plan?
	Do we have undesirable drifts?
Assumptions	You have the enterprise standard and policy documented.
Target Users	Platform Team
Usage frequency	Daily, for urgent settings
	Weekly, for non-urgent settings.
Features	Show the configuration that needs attention first, before showing overall configuration. The idea is to drive action towards optimizing configuration.
	Balance ease of use, performance (loading time of the dashboard page) and completeness of configuration check. As a result, you may not be able to show all settings of desired objects. Lack of screen real estate is another consideration behind the design

The current dashboards are:

VM Configuration	These 2 dashboards work together. The 2nd dashboard is created as the dashboard was simply too large, impacting usability.
VM Storage Configuration	
ESXi Configuration	These 2 dashboards work together. The 2nd one is prefixed with vSphere as the word cluster is a generic English word. Ideally, there is a drill down from cluster dashboard to ESXi. However, that will increase complexity in the user interface.
vSphere Cluster Configuration	
vSphere Resource Management	This needs to be a separate dashboard as the configuration spans VM, resource pool and clusters. The flexibility and complexity warrant a dedicated dashboard to help you optimize.
vSphere Network Configuration	These are basic dashboards. In future, I'd enhance them.
vSAN Configuration	

In some dashboards, there are simply too many configuration items to check than the screen real estate provides. If you have a larger screen, add the additional check as you deem fit, or add legends to the pie-charts.

In a large environment, create a filter for this dashboard. Group by the class of services such as, Gold, silver, and bronze. Default the selection to Gold. In this way, your monitoring is not cluttered with less critical workloads.

Resource Management

[e1: this section is draft]

Consumer Dashboards

The VM Configuration dashboard and the VM Storage Configuration dashboard were designed to highlight the settings that you need to fix or improve. To balance usability, it does not show every single setting.

How to Use

Select a data center from the **Data center** table

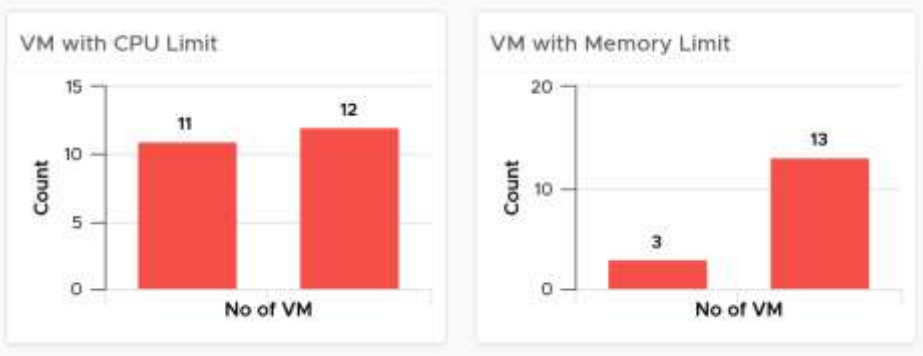
- In a large environment, loading thousands of VMs increase the web page loading time. As a result, the VM is grouped by data center. In addition, it may make sense to review the VM configuration per data center.
- For a small environment, the vSphere World is provided so you can see all the VMs in the environment.

The dashboard is organized into 3 sections for ease of use.

- All 3 sections will automatically display the VM configuration in the selected data center

The first section covers limit, share and reservation

- Limit should not be used, as explained previously here.



- Their share and reservation values can easily become inconsistent among VMs, especially in an environment with multiple vCenter Servers. The following shows an environment with far too many variations in shares.



- Shares should be mapped to a service level, to provide a larger proportion of shared resources to those VMs who pay more. This means that you should only have as many shares as your service levels. If your IaaS provides Gold, silver, and bronze, then you should have only three types of shares.
- The value of share is relative. If you move a VM from one cluster to another (be it in the same or different vCenter Server), you may have to adjust the shares.
- Reservation impacts your capacity. Memory reservation works differently from CPU reservation, it's more permanent.

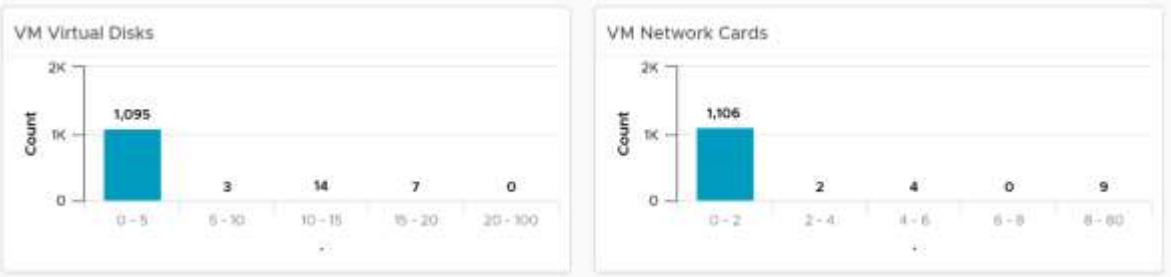


The second section covers VMware Tools

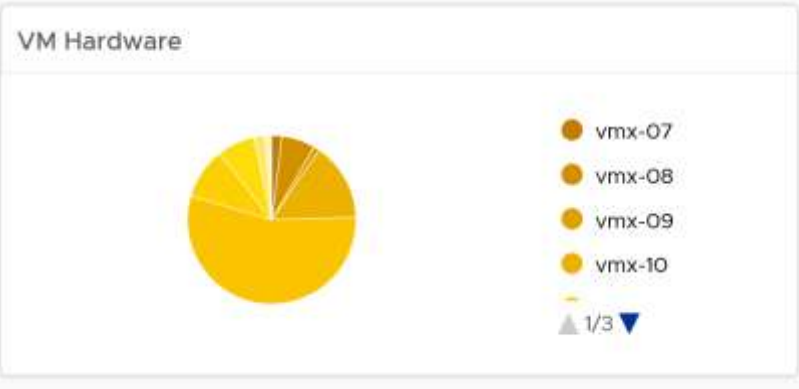
- Tools is a key component of any VM, and should be kept running and up to date.
- The distribution chart shows the various versions. You should keep them minimal

The third section covers other VM key configuration

- Keep the configuration consistent by minimizing variants. This helps to reduce complexity.
- Pay attention to VMs with many virtual disks or many virtual network cards.



- Keep the number of VM hardware versions minimal, and keep them current.



- VM Network Card widget. It you suspect your environment may have VM with no NIC card, consider adding it as dedicated bucket.

The last part of the dashboard is collapsed by default.

- It shows all the VMs with their key configuration.
- You can sort the columns and export the result into spreadsheet for further analyzis.

ESXi Configuration

The dashboard is organized into sections for ease of use.

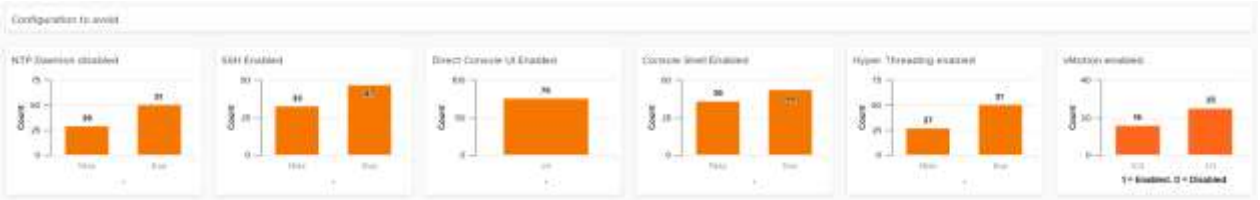
The upper part of the dashboard displays basic ESXi configurations that should be standardized for ease of operations.

- There are six pie-charts that are displayed as one set because there is a relationship in their values. There should be a correlation between them. Ideally, the ESXi version, the ESXi build, and the BIOS should be identical across all ESXi hosts in a cluster". Keep the variations of hardware model, NIC speed, and storage path minimal. The more complex the pie chart, the more variants you have. This results in complex operations, potentially resulting in increased OPEX.
- The configurations should reflect your current architecture standard. Each pie-chart counts the occurrence of a particular value. A large slice signifies that the value is the most common value, and if that is not your current standard, then you must address it.



The second section of the dashboard displays configurations that are potentially suboptimal.

- The three bar-charts display various size dimensions of the ESXi hosts. The bar-charts are designed to be seen as one set. Ensure minimal number of variations to reduce complexity.
- Smaller ESXi hosts have a relatively higher overhead, and are limited in running larger VMs. If they have a low core count, they could be using outdated CPU. Small ESXi hosts are more expensive on a per core, per GB, per rack unit basis than larger ones if they occupy the same space. On the other hand, a 4-CPU socket ESXi host is likely to be too large, resulting in a concentration risk (too many VMs in a single ESXi host). Maintain a good balance that balance your budget and risk constraints . You should adjust the distribution chart bucket size to fit your environment.



The third section of the dashboard displays configurations that you may want to avoid.

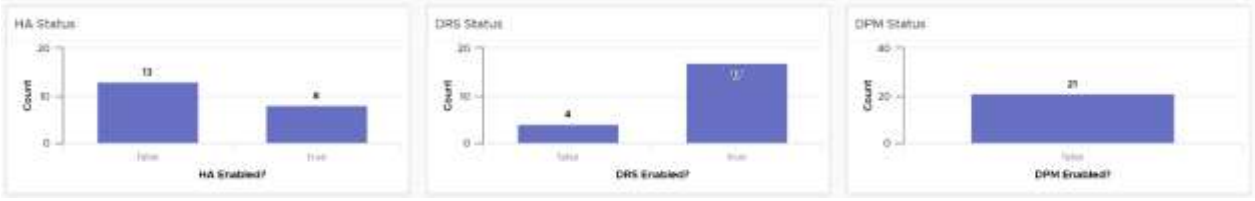
- The six bar-charts focus on security, availability, and capacity settings that you can set as a standard. For example, you should consider enabling the NTP daemon for consistent time, which is critical for logging and troubleshooting.
- The three tables list the actual ESXi hosts that are in a non-productive state. They can be in maintenance mode, powered off, or in a disconnected state.
- BTW, I've modified the last one to show information, to show an example if that makes sense for your operations. I do not do it for the out of the box version as visually it will look awkward as the first 5 charts do not need it 😊

The last part displays all the ESXi hosts in your environment.

- You can sort the columns and export the result into spreadsheet for further analyzis.
- Some of the columns are color coded to facilitate quick reviews. Adjust their threshold to either reflect your current situation or your desired ideal state

Cluster Configuration

The dashboard is organized into sections for ease of use.

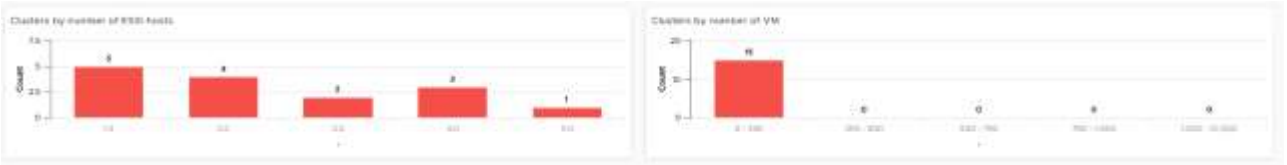


The first section shows 3 bar charts. They correspond to the 3 main features of vSphere clusters, namely High Availability HA, Dynamic Resource Scheduler DRS and Distributed Power Management DPM.

- **HA**: The best practice is to enable HA admission control. You can specify the Admission Control Policy in vCenter and the threshold for failover shares.
- **DRS**: The best practice is to have DRS enabled. Think of a vSphere cluster as a single logical computer that balance within itself.
- **DPM**: The best practice is to enable DPM in an environment where environmental concern is the top priority or the high peak rarely occurs (most of the time You are running very low utilization).

The second section of the dashboard shows 8 pie charts. They show the relative distribution of key configurations.

- 2 bar charts cover Admission Control. You should enable admission control. The pie chart displays the policy code instead of the policy name, as it is based on the property Cluster `Configuration | Das Configuration | Active Admission Control Policy`.
- 2 bar charts cover the HA Failover Share, one for CPU and one for memory.
- 2 bar charts cover DRS setting. Generally speaking, you want to have DRS fully automated, meaning no operator intervention is required for both initial VM placement and subsequent load balancing, but with a moderate migration threshold (value = 3.0). The value range from 1.0 to 5.0.
- There are 2 pie charts showing reservation, one for CPU and one for memory. Minimize the total reservation value as it prevents overcommit of resources and hence results in a less optimal utilization. Memory reservation can remain and occupy the memory space of the ESXi host even though the VM does not use the memory anymore. Consider the analogy of unused files that you have not opened for months in your laptop c:\ drive. They still take up space of the hard disk. Keep the number of distinct shares below three (or at a minimum), matching the distinct classes of service.



The third section of the dashboard shows 2 bar charts. They show the absolute distribution of cluster.

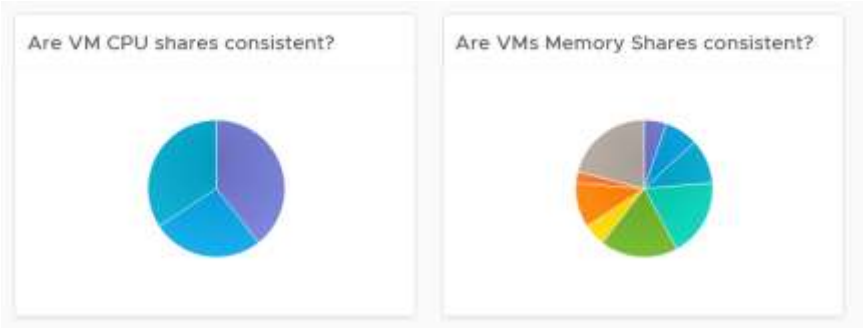
- The first shows the cluster grouped by the number of ESXi Hosts. Ensure this matches your plan and cluster sizing standards.
- Small clusters (defined as having less ESXi hosts) have higher overhead while large clusters have a higher risk in case of cluster-wide outages. For large cluster, have a disaster recovery plan an unexpected cluster-wide outage can impact many VMs.
- Performance risk is lower in large clusters partly because there are more nodes that DRS can tap on, but if there is an actual problem troubleshooting can be harder (because there are more nodes to analyze).
- In large environment, small clusters can result in cluster sprawl.

Clusters key configuration

Name	Hosts	Datastores	Admission Control Enabled?	HA Enabled?	DRS Enabled?	DRS Automation Level
test-v...	3	4	false	true	true	fullyAutomated
Cluste...	4	1	true	true	true	fullyAutomated
Produ...	4	5	true	false	true	manual
[redacted]	4	5	true	false	true	fullyAutomated
VSAN...	5	12	true	false	true	fullyAutomated
Total	36	62	-	-	-	-

The fourth section of the dashboard lets you drill down into individual cluster.

- It begins with a table listing all the clusters with their key configuration. You can export this list as a spreadsheet for further analyzis or reporting.
- Select a cluster. The list of ESXi Hosts under the cluster, along with shares and resource pools information, is automatically filled up. Ensure as all its ESXi Hosts have identical configuration.
- Keep the number of distinct shares below three (or at minimum), matching the distinct classes of service. You should avoid giving different service level to individual VMs as that increase the complexity of the cluster performance.
- Verify that Shares are used consistently throughout your entire vSphere environment. Do note that they are relative numbers, not absolute. Ideally, avoid using it altogether as it's easy to overlook. When you move VM to another cluster or vCenter, you may forget to set the new share appropriately.



- Keep the number of resource pools minimal. Resource Pools can impact performance, if the number of VMs in the pool do not match its intended shares. The new **Scalable Shares** feature in vSphere 7 automates this adjustment, which has to be done in older version. More about it here.
- Some of the columns are color coded to facilitate quick reviews. Adjust their threshold to either reflect your current situation or your desired ideal state

Points to Note

- You might notice that the **DRS Migration Threshold** widget is using the property **DRS vMotion Rate**. This is the correct property. We will correct the property name so it's less confusing.
- The **HA Memory Failover Shares** widget should be named **HA Memory Failover percentage.** This is a known typo.

- The **HA CPU Failover Shares** widget should be named **HA CPU Failover percentage.** This is a known typo.

Network Configuration

For a more complete visibility, consider adding physical network device monitoring by using the appropriate management pack. More info here.

The dashboard is organized into 2 sections for ease of use.



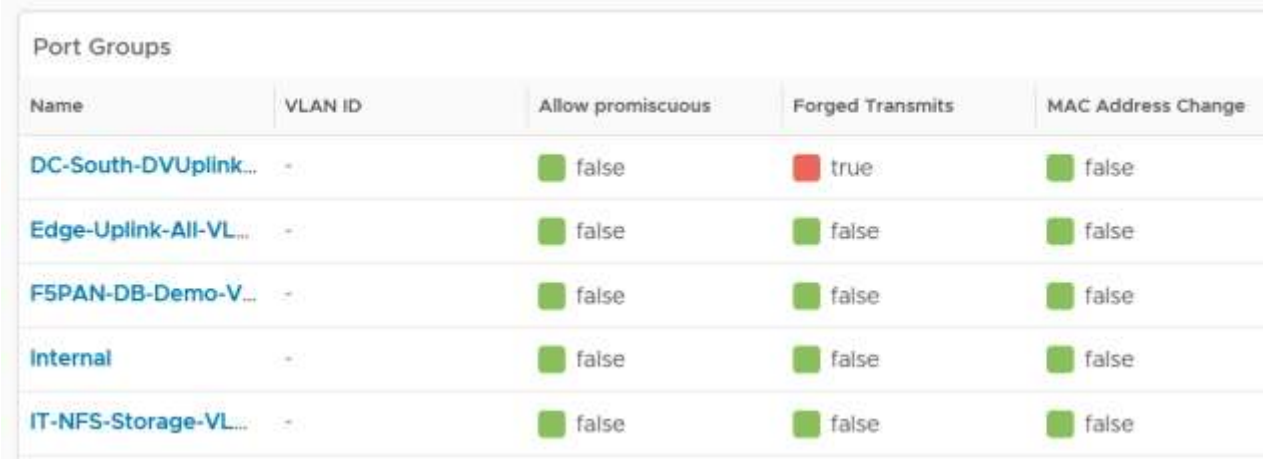
The first section displays network configurations that needs your attention

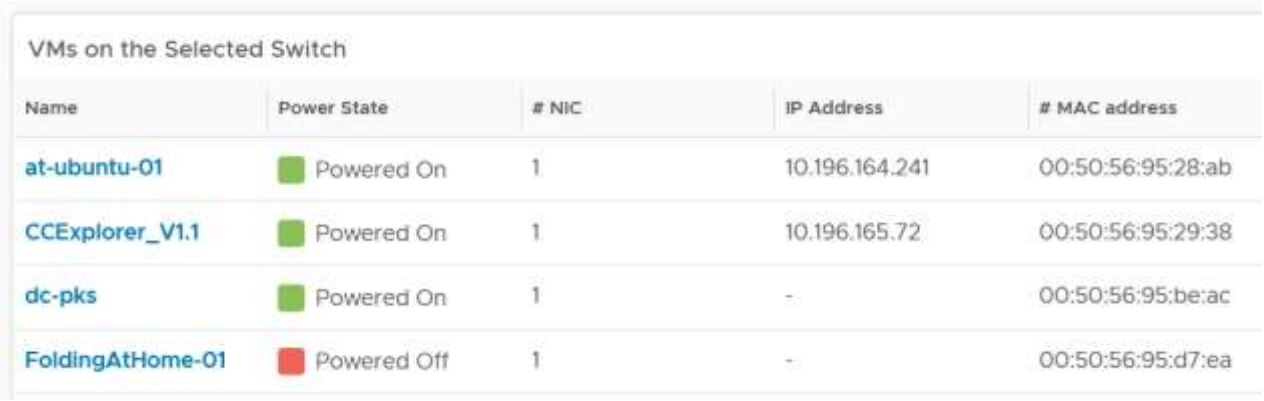
- There are 5 bar charts that focuses on critical security settings.
- The last bar chat shows the version of the vSphere Distribution Switch. Aim to keep the version current, or matching your vSphere version.



The second section provides overall configuration, with ability to drill down into a particular switch

- Start by selecting a switch from the list.
- The ESXi Hosts, port groups and VMs on the switch will automatically be shown.





VMs on the Selected Switch

Name	Power State	# NIC	IP Address	# MAC address
at-ubuntu-01	Powered On	1	10.196.164.241	00:50:56:95:28:ab
CCExplorer_V1.1	Powered On	1	10.196.165.72	00:50:56:95:29:38
dc-pks	Powered On	1	-	00:50:56:95:be:ac
FoldingAtHome-01	Powered Off	1	-	00:50:56:95:d7:ea

Review each of the tables. For the ESXi Host table, ensure their settings are consistent.

Some of the columns are color coded to facilitate quick reviews. Adjust their threshold to either reflect your current situation or your desired ideal state.

You can sort the columns and export the result into spreadsheet for further analyzis.

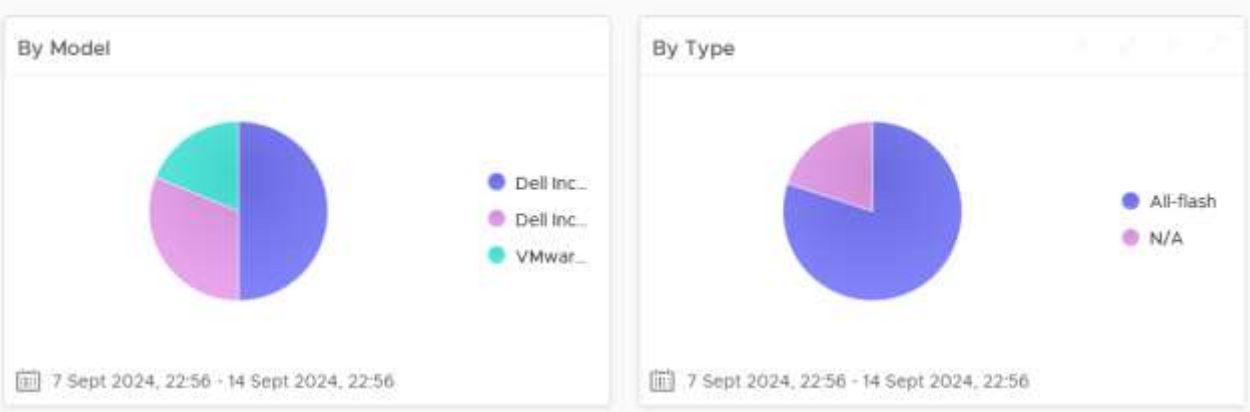
Storage Configuration

The dashboard covers both vSAN and non vSAN.

The dashboard is organized into 3 sections for ease of use.

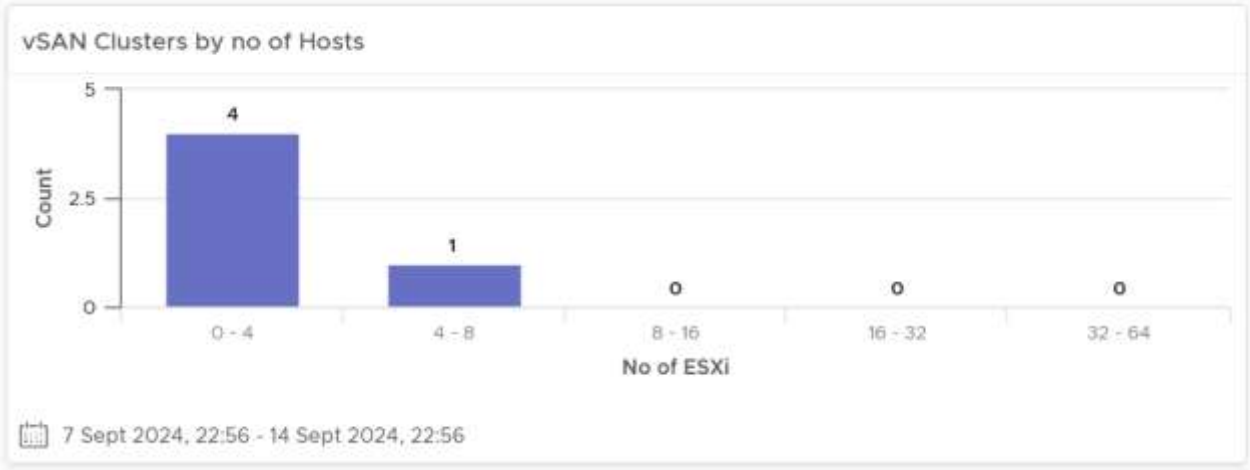
The first section displays 6 pie charts

- There are 5 bar charts that focuses on critical security settings. Their values should match your security policy.
- The last bar chat shows the version of the vSphere Distribution Switch. Aim to keep the version current, or matching your vSphere version.



The second section displays 3 bar charts

- Together, they provide good overview of the vSAN key capacity configuration. By seeing the distribution, you can see if you have capacity configuration that is outside your expectation.



The last part of the dashboard shows all the vSAN clusters with their key configuration.

- Some of the columns are color coded to facilitate quick reviews. Adjust their threshold to either reflect your current situation or your desired ideal state
- You can sort the columns and export the result into spreadsheet for further analyzis.

Availability

Part 2 Chapter 7

The availability dashboards aims to implement the capacity concept covered in Availability Management chapter. They are rather limited, hence if you have something let's collaborate!

VM Availability

Use the VM Availability dashboard to calculate the availability of the Guest OS. The availability of the Guest OS is calculated because the Guest OS might not be running even when the VM is powered on. There are two layers of Availability, that is, the Consumer layer and the Provider layer. This dashboard covers the Consumer layer. You can view VMs in the selected data center, uptime trend for a selected cluster, and so on.

Questions asked:

- Overall availability over time.
- What are causes of the availability? How many VMs affected by HA? How many were reset?
- Tools availability. Which VMs have Tools not running? While technically “tools not installed” is a configuration issue, it impacts availability.

Design Consideration

The dashboard is designed to help check the availability (uptime in percentage) of VMs, as availability is typically a part of services provided by the IaaS provider.

This dashboard does not check the application up time. It is possible that the application (e.g. a database, web server) is down while the underlying Windows or Linus is up. Generally, the service provided by IaaS team is only until Windows or Linux. To check application level, use network ping or application specific agent (e.g. Telegraf).

How to Use

Start in **Data centers** widget by selecting one of the data centers listed.

- In small environment, or if you want to see overall, you click the **vSphere World** object.
- The above action will update other widgets automatically.
- Think of creating a filter for this table that reflect your class of service. Group by the class of services such as Gold, silver, and bronze and default the selection to Gold. In this way, the monitoring is not cluttered with less critical workloads, and you can focus on the important VMs. One way to achieve this is by creating a VCF Operations custom group for each class of service

About the VMs by Uptime in the last 30 days bar chart

- It displays the average uptime of VMs grouped by their availability. The bucket distribution is designed to cater for a wide array of environment. If You are monitoring only production VMs, where uptime is expected to be near 100% all the time, edit the bucket to meet your operational need.

About the VMs in the Selected Data center table

- It lists all VMs currently deployed to the data center. Average Uptime is displayed for the last 1 month of data. Expect this number to be 100% or near there for production VM.
- Note that the **Services** column will be blank unless Service Discovery is enabled and services/processes were discovered on a specific VM.
- The column **VMs** includes all VMs including powered off VMs.

Select a VM from the above table.

- The remaining widgets will automatically show the detail of the selected VM.
- Selected VM Uptime Trend displays the selected VM's Guest Tool Uptime (%) across the last 30 days.

Expand the 2 collapsed widgets

- If Guest OS services or processes are discovered inside a VM, their availability is analyzed. Service 'state' over time is displayed in Guest OS: Services.
- The dashboard displays the process or services running inside the Guest OS. This requires the Service Discovery Management Pack.
- The ESXi Host where the VM has run widget can show historical migration of the VM. This can be useful in determining the cause of a VM downtime.

Points to Note

- The metric is only tracking the availability of VMware Tools, not the entire Guest OS. If Tools is not up, it assumes the Guest OS is down. To help you check that this is not a false negative, add a few line charts that shows sign of life. A good counter is IO metrics such as Disk IOPS, Disk Throughput and Network Transmit Throughput, because IO requires CPU processing. CPU Usage is not a reliable counter as work by VMkernel on the VM is charged to the CPU metrics.
- VCF Operations 8.2 sports a new ping adapter. This means you can enhance the accuracy of the uptime measurement by creating a super metric that adds the ping information or checking the process (needs an agent, such as Telegraf).
- Add a property widget that lists the selected VM properties to give you more context about the VM. In large environment, it is possible that the VM name alone may not be providing enough context.

vSphere Availability

There are two layers of Availability, that is, the Consumer layer and the Provider layer. The vSphere Availability dashboard covers the Provider layer. This dashboard includes a cluster and not an ESXi host because the cluster is operationally a single compute provider. This dashboard considers the N+1 design, where the cluster can withstand one host failure. Logically, a cluster with fewer hosts has a higher risk.

The dashboard is designed to help you analyze and report the uptime, as availability is typically part of official business SLA. It's also often required in the monthly operational summary report.

The dashboard is not designed for live monitoring of the uptime. An NOC style of dashboard is better suited for that use case. Tools such as Aria Log Insight should also be leveraged as fault is typically preceded with soft errors.

How to Use

The first information shown is the average availability of all vSphere clusters in the environment. It's a simple average, regardless of the clusters function.



Ideally what you want is something like this:



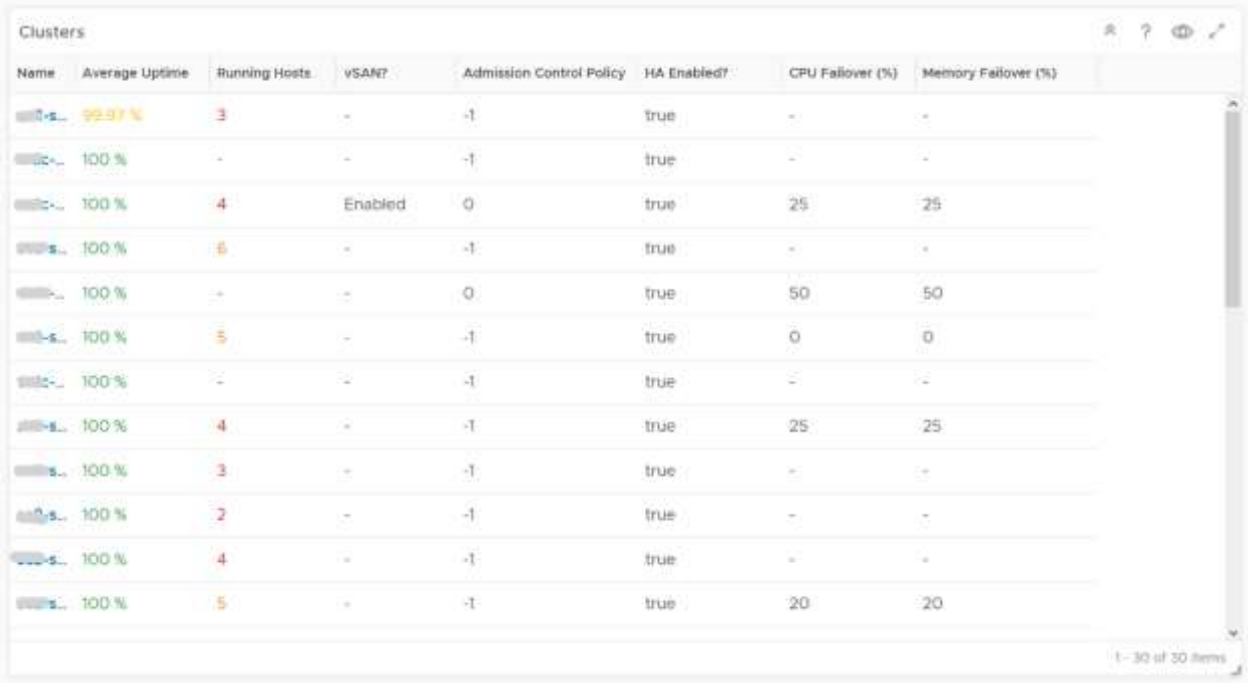
The next health chart shows the count of ESXi hosts that are not available.



Adjust the threshold accordingly. By default, it's using 1 for Yellow, 2 for Orange and 4 for Red.

Next is a table showing all vSphere Clusters.

In a very large environment, creating a filter for the list of clusters can make it more manageable. One is to group the clusters by their class of services. Group your clusters into Gold, silver, and bronze and default the selection to Gold. In this way, you can see your Gold clusters more easily.



It is sorted by the lowest uptime, so your attention is drawn to the cluster with the lowest uptime in the last 1 month. The column is based on the *average* of the last 1 month as availability SLA is should be calculated per month (but reported much more frequently)

The metric used is Summary \ Cluster Availability (%), which assumes N+1 design. That means if there is 1 node failure, the metric is still showing 100%.

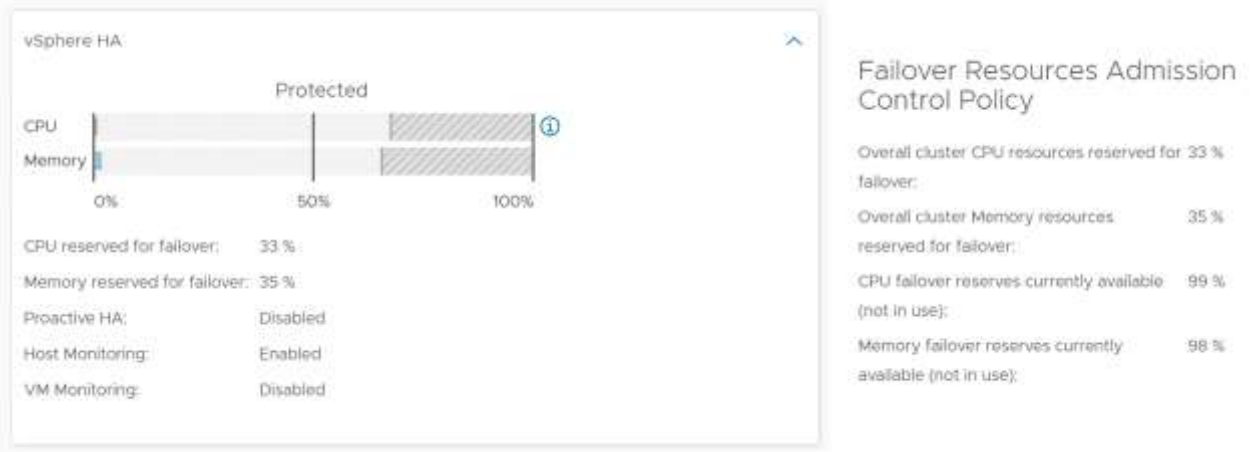
The column **Running Hosts** are color coded as logically a smaller cluster has higher risk. A single host failure results in relatively higher capacity degradation.

The column **vSAN?** Is added as hyper-converged means you need to consider both the compute part and the storage part.

The Admission Control Policy is based on the property `Cluster Configuration \ DAS Configuration \ Active`. The mapping between code to name is

- -1 = Disabled
- 0 = Cluster Resource percentage
- 1 = Slot Policy (Powered-on VMs)
- 2 = Dedicated Failover Hosts

The cluster failover percentage columns map to the following values in vCenter client UI.



Select a cluster from the above

- The cluster uptime will be automatically plotted. It's using 25%, 50%, and 75% as the threshold for red, orange and yellow respectively. The reason for low threshold is the 5 minute window. A complete 5 minute downtime is only 0.0116% when measured against a 30-day SLA. If the uptime was 100%, it will only go down to 99.9884%
- The ESXi in the selected cluster table will be automatically filled up. For more context, you can add a property widget that lists the selected ESXi Host properties.
- The 'Connected to vCenter' and 'Maintenance State' columns are not the average values, as both are string. However, they display the last state in the selected period. This allows you to go back to a specific point in time and view availability at that point.

About the "Datastore not available" list



- It filters to only those datastores with status powered off. This covers both local and shared datastore. To add context, consider adding extra column such as the data center where it resides, and the datastore type (e.g. NFS, VMFS)

About the Port Group availability list

- This lists port groups which at present has uptime of less than 100%.
- To add context, consider adding extra column such as the data center where it resides, used number of ports and maximum number of ports

Points to Note

Consider adding vCenter Server availability and NSX components availability. This requires the VMware SDDC Health monitoring solution.

Role-based Contents

Part 2 Chapter 8

It is imperative to get other departments to use VCF Operations. Different persona have their own agenda and goals, so take time to sit down with the individuals and understand the why, not just the what.

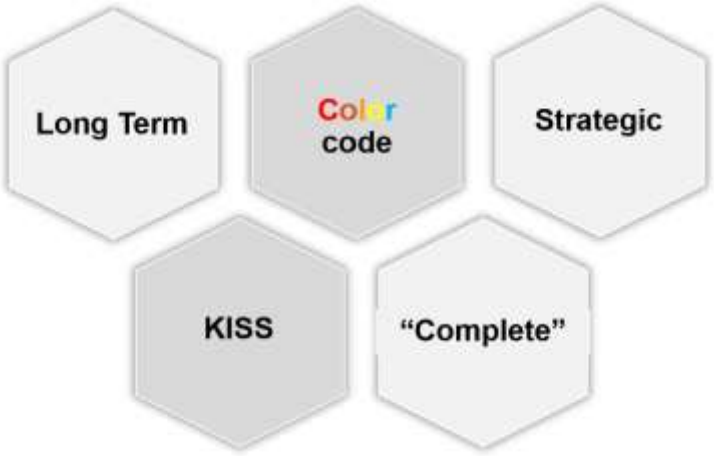
IT Leadership Team

The most important persona or role is your boss. In large organisation, there can be multiple layers between you and your CIO. Each layer has their own needs. Understand the need of each individual. The nuance matters. Different managers are likely having slightly different requirements. Take time to understand the real needs behind the ask. Learn what questions need to be answered, what answers are expected, and what actions are taken based on the answer.

You also need to think of what you want from them, as they hold decision making authority. Show them problems that you can get help, which is budget and resource. Do this by showing data. By giving visibility into live environment to senior management, you prove that you do need additional hardware. If there is wastage to be reclaim, you also prove on where and how large the wastage is.

Think of what approval you need to do so you can do your work. How do you present the supporting information without burdening them with too many details?

General Requirements



Summary	Generally, they want summary and the big picture, not details. Or at the very least, the big picture is presented first, and then drill down is provided.
	As part of showing the big picture, trend should be included. Show the situation in the last few months, ideally coupled with projection. The data should be averaged out, so that a 5-minute spike should not show up.
	Exception. Things that they need their attention. Complete means nothing else can be taken out. Which information, object, metric can you take out while preserving the point?
Offline	They prefer information to be emailed so they can access from anywhere easily. They may not want to login to VCF Operations.
	They appreciate offline access as they tend to be mobile. Some may even ask for the information to fit on a mobile phone screen, or just a single screenshot. For off-line access, verify the screen size. A small screen of 7" can only display a small subset than say a 14" display.
KISS	KISS. Keep it simple solution. Keep the interaction, clicking, zooming, sorting, etc. minimal. Use larger fonts, round numbers (law of significant figure).
	UI that is easy to understand. So keep each dashboard to a specific question. Make sure the dashboard is easy to use. So keep the interaction, clicking, zooming, sorting, etc. minimal.
	Can the dashboard be understood within 5 seconds? If yes, you buy yourself a few more minutes. Your IT Management has understood what the dashboard does, and is willing to spend more time appreciating its full capabilities.
	No technical info. Ideally, present in business terms, not IT jargons. Terminology such as datastore, distributed switch may need to be replaced with something suitable in your organization.
	Take note of the "size" of your dashboard. A dashboard that has many widgets and scrolls deep is harder to understand.

If they ask for a self-service portal, then make it easy to access. They may not want to login to VCF Operations. If they do, they may forget their password, so the portal should not require a password.

Sample Requirements Analysis

The following table provides an example analyzis of 2 different IT leadership layers. Notice the 2 groups have a different set of requirements, resulting in 2 different approaches.

	Middle IT Management	Senior IT Leadership Team
Users	John and Jill	Sam, Andy, and Julia
	They will read on their laptop. They have access to VCF Operations but only login occasionally in a week.	They expect a 15-minute presentation (live environment), with a copy of the PDF on their inbox for future reference.
Familiarity	High. They know the environment and context well	Low. They need explanation.
Focus	Keeping the lights on.	Last month review, especially compared against plan and SLA. Finalize next month plan.
	Availability, Performance, Security.	Capacity and Cost. How to lower cost while increasing capacity.
Authority	Medium-level changes. Budget up to ___.	Major IT decision that is strategic and long term. Represent IT to Business Leaders

While you are clear on what they want from you, what do you want from them?

Having analyzed the requirements from each person, and from you to them, you develop the following approach:

	Middle IT Management	Senior IT Leadership Team
On Mobile	Daily. First thing in the morning. Forward looking, not about the past. A simple summary. Fits into 1 screen.	Weekly. On Monday. 1 week ahead outlook. A simple summary. Fits into 1 screen.
On Desktop	Weekly. 2 pages. They will read all sections.	Monthly. 4 pages. They do not read all sections, as it will be presented to them. Done as part of overall operations review, which includes non VCF
Others	On demand login to dashboard. They can login and interact to get richer information.	You place a screen on the wall near their office. No interaction. It's for at a glance look.

Business Applications

This is likely the most important, both from their viewpoint and your viewpoint. It proves that your infrastructure is playing its part well.

The report should provide a summary for all the mission critical business applications. As they are of the same level of importance, the list should be grouped by the business units.

Information	Description
Name	The business application name as per what the business or senior IT leaders call it. As business applications span multiple tiers, VM, and K8 Pod, you need to build the structure first in VCF Operations
Availability	Average uptime during the reporting period. This shows Actual Availability (%), not Operational Availability (%) as the end customers (which are typically external or public) do not care about your scheduled downtime. Majority of websites has 24 x 7 x 365 days availability expectation, even if your customers are not global.
Performance	It shows their average performance over the reporting period. As average can mask out poor performance at specific times, they may ask for worst number also. If there are many business applications, they may ask for the average of all of them. They may ask for the information to be presented over time, like stock market performance.
Compliance	This is configuration compliant to both security and master template.

The above is one example. Another common example is infrastructure-centric report. As usual, understand the reason behind the ask. For example, they may not be familiar with the environment and inventory, and hence they ask for what you think is basic.

Operations Overview

This report summarizes what happened to VCF operations management.

- Was it a busy week or month? Were the operations hectic and environment volatile with a lot of debates?
- Was the environment well managed? Was the availability, performance, and compliance meeting the SLA?

Information	Description
Inventory	Summary of the key inventory items. This can be absolute or relative.
	Changes in the inventory. Was there a lot of items added or removed?
	Dynamics of the inventory. Was there a lot of settings changed?
Configuration	Summary or important configuration that either they care, or you want to bring up to their attention
Capacity	Is there enough capacity? If not, when should action be taken? What was the change in capacity?
Performance	Ideally this is compared with SLA.

Sample Dashboards

VCF Operations provide two example dashboards to get you started. They are designed for you to present live information to your senior management. They are not designed as self-service.

This dashboard is used by the Ops team to provide the overall picture to the IT Management team. This dashboard works together with the set of inventory dashboards. The inventory provides details on available resources and what is running on these resources.

It also works with the capacity summary dashboard, which provides details on the resources remaining capacity and time remaining to act.

Capacity Summary

This dashboard is used by the Ops team to explain capacity to the IT Management team. It works together with the **Inventory Summary** dashboard. The inventory provides details on available resources and what is running on these resources. The capacity provides details on the remaining capacity and time.

See the Executive Summary Dashboards page for common design consideration among all the dashboards for IT senior management.

How to Use

The dashboard has 2 sections.

- The top section provides summary at the vSphere World level.
- The bottom section enables drill down into individual compute or storage capacity



The weekly average of VM growth is displayed to provide holistic visibility of overall growth across all Data centers for both running and powered off workloads. If the increase in VM count is not accompanied by corresponding increase in utilization, these newly provisioned VMs are likely not yet used.

Overcommit Ratio highlights the efficiency gained by vSphere virtualization running multiple workloads on shared infrastructure. It is important to note that overcommitment needs to be further reviewed in conjunction with elevated resource contention (refer to contention dashboards) to understand the performance impact when running VMs competing for resources. In general, Overcommit is required to be financially more economical than the public cloud. As a reference, AWS typically overcommits CPU 2:1 by counting the hyper-threading and does not overcommit memory. Note: VCF Operations uses Physical CPU Cores not Logical Cores (Hyper-threading) for all CPU-based capacity calculations

The bottom part of the dashboard is split into 2 columns:

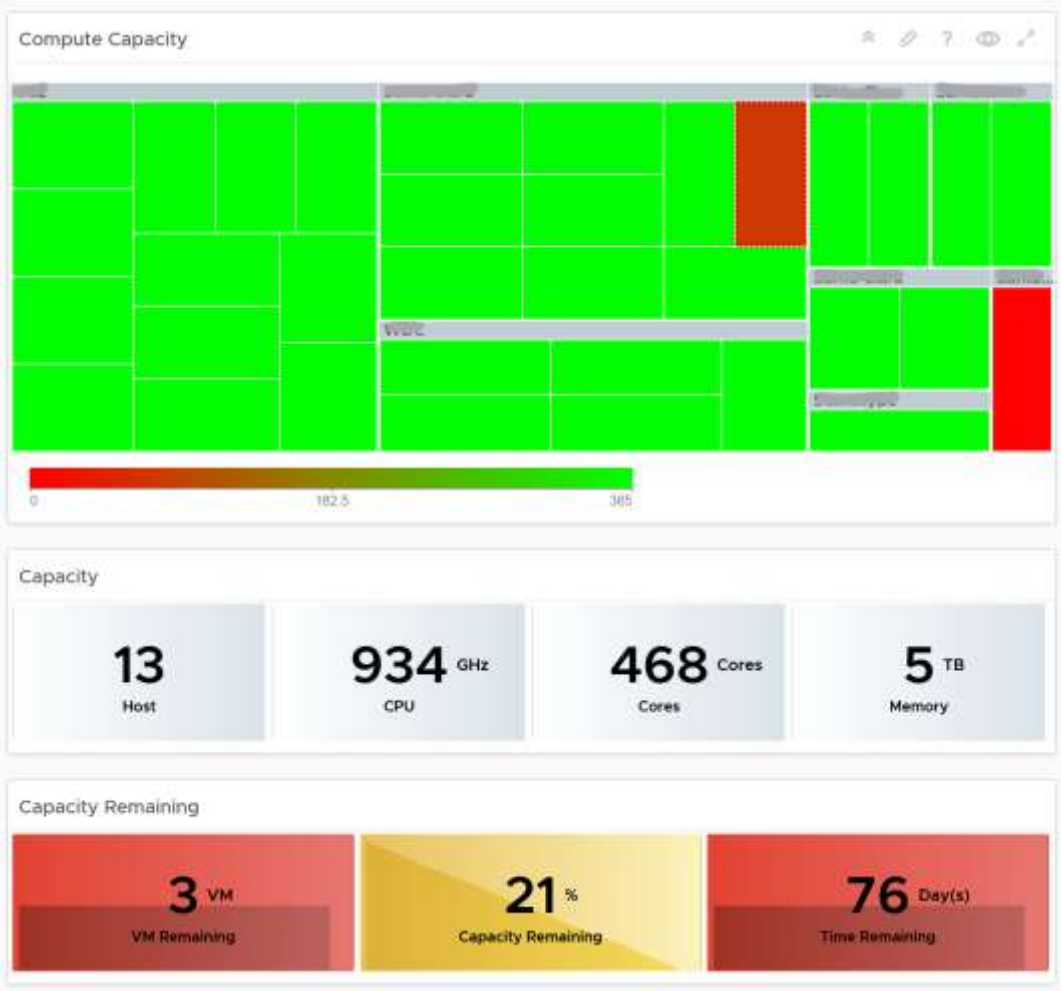
- compute
- storage.

Network is not added due to its nature. It's an interconnect, not nodes, so capacity is much harder to compute.

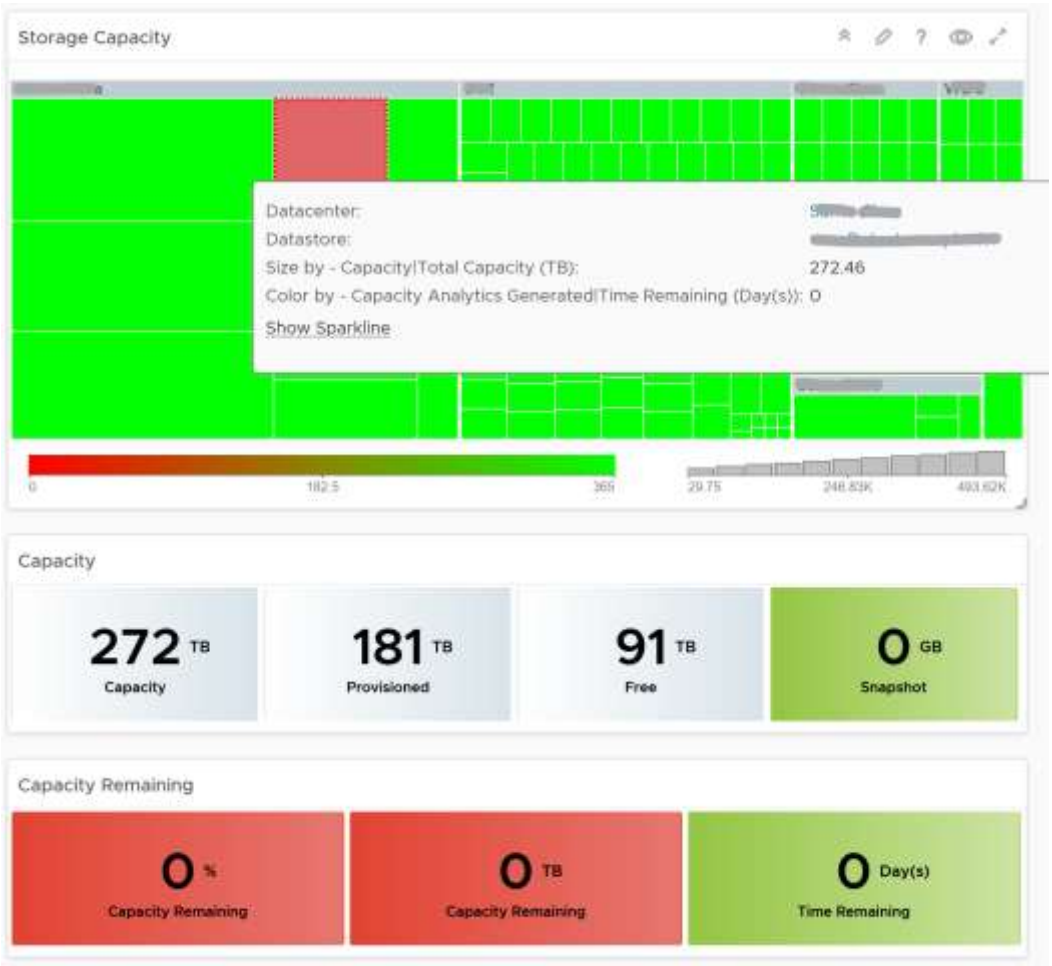
The dashboard uses the term compute (and not vSphere Cluster) and storage (and not vSphere datastore) to keep the visual simple.

The two columns have identical design. The heat map displays capacity by size and colored by time remaining. However, for compute the size of each box in the heat map is fixed as there is no single metric to represent the size. A cluster size can be measured in 4 different ways (No of ESXi Hosts, Total CPU GHz, Total CPU cores, and Total Memory).

Here is what the compute portion looks like. The heat map is interactive. By selecting one of the cluster, you can further drill-down to clearly understand remaining capacity and time (in days).



Here is what the storage portion looks like. They are designed to be similar



Points to Note

- Capacity Remaining is not shown at the world level as it could be misleading, especially in global or large infrastructure. Clusters also tend to serve different purpose, and they are not interchangeable.
- If you are using both on-prem and external cloud, for example, VMware on AWS, consider splitting the dashboard into 2 columns. You would need extra screen real estate though.

Level 1 Team

There are 2 dashboards for VM and 2 dashboards for infrastructure, covering the most common scenarios.

What information do you show? This is a tricky answer. For example, when something is down, it should be shown. But what if it's part of planned downtime? A planned downtime can last hours, making the NOC screen displaying red for too long. This means you need to exclude them. Create a group called a planned down time, and move the object as part of the process.

Live! vSphere VM Changes

The dashboard starts with a set of summary numbers at the top.



Make sure the numbers match your expectation. Set a threshold for each so it's easier to see if their values change (drop or increase) to a level that is not within your comfort level.

The changes above is supported with details. There are 3 rows of detail information covering the 20 types of changes. This covers all the possible changes that could happen to a VM, hence providing a real time visibility into the movement or volatility.

Each of the 20 changes are unique. Ensure they match the approved change requests and expected changes of that day.

Inventory Changes	VMs are added or remove, hence impacting count of inventory. Are VM are added or removed as per your expectation? If yes, are they done the right way (cloning vs free style creation vs template based deployment)? Are the changes matching our change window policy? Why do we have high count of VMs being unregistered or deleted?
Location Changes	VMs are moved to another host or datastore. This can be hot or cold migration. Hot migration is shown first as they likely impact performance as it changes both the compute and storage. A change of both parents at the same time is shown first as that's a relatively more major change. Pay attention to the storage migration as they take the longest amount of time. If it takes longer than expected, something could be wrong. On the other hand, cold migration is typically an activity that should match change request as the VMs are shutdown.
State Changes	VMs power state change. Reset is the least desired, hence it's shown first. It is used as last resort. Suspend and Power Off are least preferred, as these action should be done from within the Guest OS. Expect these numbers to be low. A high number of powered on VMs could lead to high demand. Typically, how many days passes before your New VMs becomes active? It can take days or weeks before the VMs usage grow to its "full size".

For each of the 3 types, the dashboard shows 2 widgets:

- The first is a trend line. Make sure both the absolute amount and the pattern match your expectation.
- The second is a scoreboard showing the present number of the each of change. The change is sorted from the least desired (most impactful to operations) to the least important.

Threshold Adjustment

The changes are color coded. Since your environment is likely different, adjust the threshold accordingly. Set a different threshold for each while keeping the variations minimal. Different numbers are needed because they are different events. Minimal variants are needed as they make it easier for the NOC room team.



Live! Heavy Hitter VMs

This dashboard helps you analyze the misuse of the shared infrastructure. This dashboard displays details of VMs misusing the shared infrastructure and if that has caused performance problems to the other VMs. The cause for excessive load might be due to security attacks, for example, denial of service, process runaway, or mass activation of agents.

Design Consideration

In shared environment, it is possible to have victim-villain problem. In the heat map, the villain VM is the one with the largest box size, while the victim is the one with red box. If a handful of VMs are dominating the shared infrastructure, their collective size will be highly visible on the dashboard.

There are 4 areas where a monster VM can impact its neighbours.

- CPU.
- Disk IOPS. Any VM with PVSCSI can generate very high IOPS. Since this hits the underlying physical storage, the view is grouped by Data Center, not Datastore or Cluster.
- Disk Throughput. Applications with large block size (e.g. 4 MB) can consume bandwidth without generating high IOPS
- Network. Network packet/second not shown as dashboard can't support 5 columns. If you rearrange, it will look odd. It's added in the trend chart

Memory is not needed as it's rare for Guest OS to actively read/write from DIMM. It's mostly cache. Think of it like disk space (passive).

Disk space is not shown for the same reason as memory.

How to Use

The dashboard starts with a set of summary numbers at the top.



Make sure the numbers match your expectation. Set a threshold for each so it's easier to see if their values increase to a level that is not within your comfort level.

There are 4 heat maps, showing the 4 different loads that can be excessive. The heat map displays the relative value and not the absolute value. A VM does not generate a high load in the absolute term just because it has large configuration.

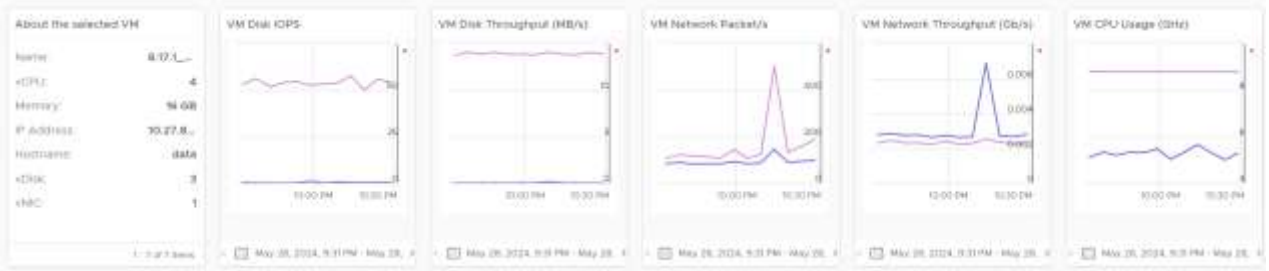


Each heat map has their own color threshold, reflecting the nature of the contention metrics used in each of them.

Remediation Action you should take: Check why the VMs are generating excessive load.

Interactive Section

For NOC Operator, who have access to mouse and keyboard on their desk, they can drill down by selecting one of the VM on the heat map. All the information below is automatically shown, giving user more detail about the VM.



5 types of "hit" on the shared infrastructure are shown. For each, breakdown between read and write, transmit and receive, etc. are provided

Points to Note

Memory is not shown as it's a form of storage. The memory metrics are space utilization, not speed. Think of disk space instead of IOPS. It can cause capacity problem on the shared ESXi host, but not performance problem to other VMs.

In a large environment, it might be difficult to view a small victim VM. Consider having multiple dashboards and rotate among them.

Cluster Dashboards

There are 2 dashboards working together:

Live! Cluster Performance	This the primary dashboard for performance. It provides live information on whether the requests of the VMs are met by their underlying ESXi host, which is grouped by the compute clusters.
Live! Cluster Utilization	This secondary dashboard complements it by showing if the performance problem (read: contention) was caused by high utilization.

The primary dashboard answers the question "Is our IaaS performing?", while the secondary dashboard answers the question "Is our IaaS working hard?".

The focus of performance is on population, not a single VM. This is not a single VM troubleshooting dashboard. We are looking at infrastructure problem, not a single VM problem. As the infrastructure counter is mathematically an aggregation of VM metrics, we need to pick the right roll up strategy. As the goal is to provide early warning, we're not using average as the roll up technique, as it is too late as early warning. We use percentage of population exceeding a threshold. The threshold is set to be stringent so we can get early warning.

The 2 dashboards show a set of 4 heat maps side by side. They complement each other and should be used together. The location of each cluster and ESXi hosts within those clusters is identical in all heat maps. This fixed positioning enables viewers to compare if the problem is caused by memory contention, CPU ready or CPU Co-stop.

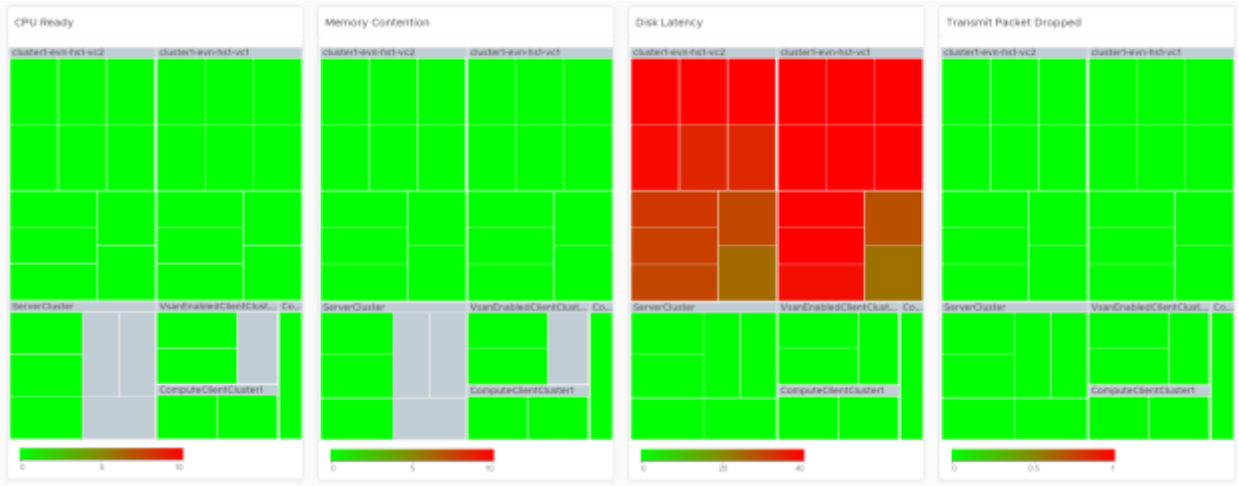
The size of each cluster and ESXi hosts has to be constant. Variable sizing creates a distraction, as the focus here is not capacity. Variable sizing can potentially result in small boxes, making it hard to read from a far.

Live! Cluster Performance

The dashboard starts with color coded trend chart.



The number is the average of all clusters. If you have 10K VM, this is the average of 40K metrics as each VM has 4 (CPU, memory, disk, network). As a result, it takes a serious degradation to bring this number down. This number should match your business cycle if your environment is mostly virtualized.



Look at the 4 heat maps and see if there is any color other than green. Since you have HA in the cluster, where at least 1 host can go down without impacting performance, that means no hosts should be running hot when there is no HA.

- While each box is an ESXi, the counter is coming from all the VM in the host. It's not taking ESXi level counter at all. The metrics used are % VMs facing CPU Ready, % VMs facing CPU Co-stop, % VMs facing RAM Contention.
- The color is by percentage of VMs not being served well. We are not using Max among the VM as it's too extreme, placing too much focus on a single VM. On the other hand, we're not doing ESXi wide average, as that will be too late.
- Green indicates that almost 100% of the VMs are getting the CPU and memory they are asking. The threshold is set such that if 10% of the VM population is not getting the resources they are asking, the heat map will turn full red.
- Red indicates an early warning. Stringent thresholds are used to enable proactive attention & remediation operations. Because high standard is applied, it is possible that the heat map is showing red, but there is no complaint from VM owner yet.
- Light grey likely means there is no VM running on the host, hence the metric is not computing.

Check if there is imbalance.

- There are 2 types of imbalances: cluster imbalance and resource type imbalance
- The ESXi hosts are grouped together by the cluster, so imbalance within a cluster can be seen easily. Cluster imbalance is a real possibility that is best monitored and not just assumed.
- If the 4 heat maps are quite different, there is resource imbalance. For example, if the memory contention is mostly red, but the 2 CPU heat maps are green, that means you have imbalance between memory and CPU.
- If a single ESXi host is displaying a different color across 3 heat maps, it indicates imbalance in the host.

Interactive Section

For NOC Operator, drill down by selecting one of the ESXi on the heat map.

Memory Consumed is used, and not Guest OS metrics, as it better represents the footprint on the VM.

CPU Usage is shown, and not CPU Net Run, as it better represents the footprint on the VM.



The “Trends of selected ESXi Host” will automatically show the performance metrics

It’s showing from all 3 heat maps, so you can correlate. To hide any metric, simply click on its name on the legend.



Live! Cluster Utilization

Use it to view the clusters and ESXi hosts within those clusters that are working excessively and are close to their physical limit. This dashboard displays ESXi hosts which have CPU or Memory saturation, and which can lead to performance issues for the VMs running on the host.

2 metrics for Memory are used as Consumed metric alone is not enough. It could be mostly cache or historical data. Balloon also persist (past data), but at least it gives an insight of ESXi under memory pressure.

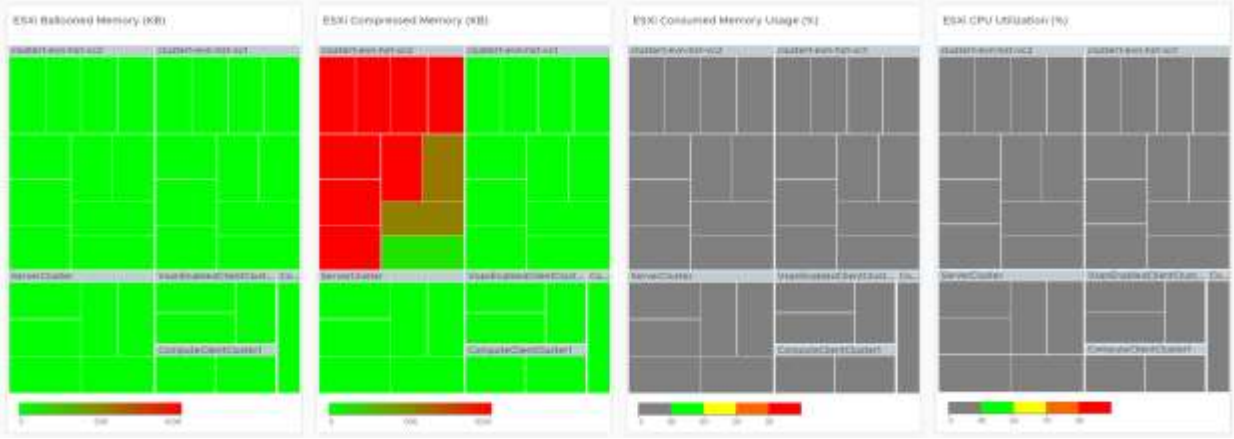
The dashboard help answer the following questions:

- Is the IaaS running very high utilization right now? If yes, which cluster? Within the cluster, which ESXi?
- Do we size ESXi correctly, meaning balanced usage of CPU and Memory?
- Are the clusters balanced? Are load distributed equally among member ESXi hosts?
- Is all ESXi contributing? Light grey box indicates that the host is part of the cluster but there is no utilization. It is likely the host is in maintenance mode or is powered off.

How to Use



Unlike the heat maps in the **Live! Cluster Performance** dashboard, the 4 heat maps in this dashboard has a different scale, reflecting the different nature of the metrics.



Logically, memory is a form of storage. It acts as a cache to disk as it's much faster. Hence a high utilization is better, as that means more data is being cached. The ideal situation is when ESXi Consumed metric is red but ESXi Ballooned metric is green. When Ballooned is red and Consumed is grey, it means it was likely high in the past but not anymore. The reason the ballooned stays red because the ballooned pages were never requested back.

Ballooned memory counter was chosen over the swapped or compressed memory metrics as it's a better leading indicator. Because all 3 can co-exist at the same time, all 3 are shown in the line chart. Ballooned is shown in absolute amount and not percentage, as the higher the size the higher the chance it might impact a VM. If you feel using percentage is easier for your operations, create a super metric to translates the value.

Interactive Section



Quiz Time: why is the heat map showing grey and light grey?

The heat map is showing dark grey and light grey to convey wastage and data error, respectively. While the use case is performance, there is a need to educate that oversized VM cause performance issue.

Points to Note

If you have the screen real estate:

- Add a Heat Map for Disk Latency. Use the counter "Percentage of Consumers facing Disk Latency (%)". It is part of datastore object, not cluster, as a VM in a cluster can have disks across multiple datastores. Organize this storage performance by data center and not by cluster.
- Add trend chart of Disk Throughput and Network Packet/second.
- If your network is highly utilized, add a Heat Map to check the ESXi physical NIC Network Throughput.

Storage Team

Ask any Storage Team and Platform Team whether the collaboration between them can be improved by a mile, and you are likely to get a nod. One reason for this issue is there is lack of common visibility. You need to see the same thing if you want to collaborate. Storage Team do not get always get access vSphere. Even if they do, vCenter UI is not designed for Storage team. It is designed for VMware Admin.

VCF Operations and Log Insight can bridge that providing a set of read-only, purpose-built dashboards, that answer common questions such as:

- When a VM Owner complains, can we agree if it's a storage issue within 1 minute? This will help reducing the ping pong game between VM Owner, vSphere Admin, and Storage Admin.
- Is the Storage serving all the VMs well? If not, who are affected, when and how bad? Read or Write? The answer has to be tier based, as Tier 1 VM expects lower latency than Tier 3.
- What's the total demand hitting the array? Are they growing fast and becoming a risk? Who are the heavy hitters among the VMs?
- When & where are we running out of capacity? How much disk space can be reclaimed? From which VMs?
- What have we got? Are they consistently configured?

The questions above cover the main areas of SDDC Operations, such as performance, capacity, configuration and availability. They enable joint troubleshooting, capacity planning, performance monitoring. For better collaboration, add physical array monitoring into VCF Operations and Log Insight, so you can analyze physical arrays and fabrics, and then correlate back with vSphere.

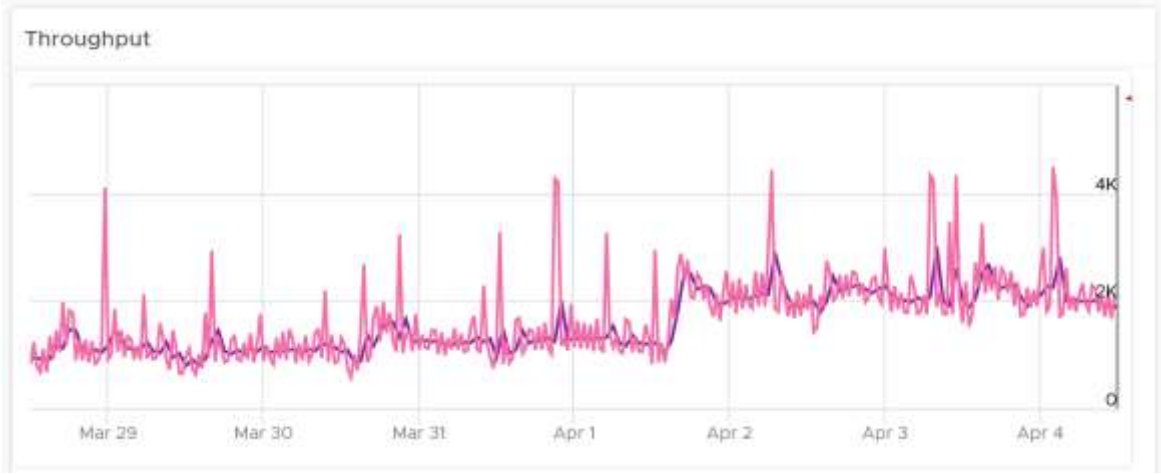
The dashboards should provide overall visibility to Storage team. They give insight into the SDDC by showing relevant objects by:

- quickly showing the summary of key information.
- showing VM, datastore, datastore clusters, compute cluster, and data center. It shows their relationship, which you can interact and drill down.
- showing all the VMs, where they are located, how much space they are allocated, and how they are using it.
- Showing physical arrays inventory and how they map into vSphere.

Storage Heavy Hitters

Interpreting IOPs and throughput metrics depends on your underlying physical storage capability. For visibility into this hardware layer, add physical storage arrays and physical network switches metrics to the dashboard.

See the **Network Top Talkers** dashboard as they basically have the same design. The main difference is we show IOPS and throughput. They are related, so use both to gain insight, they should display a similar pattern. If not, that indicates varying block sizes. For example, a throughput spike without an accompanying IOPs spike indicates large block sizes.



About the “Which VMs hit storage the hardest” table

Which VMs hit storage the hardest?

Name	IOPS (5 minute Max)	IOPS (1 hour Max)	Throughput (5 minute Max)	Throughput (1 hour Max)	Datacenter
[redacted]	22,418.93	8,357.15	627.58 MBps	378.35 MBps	[redacted]
[redacted]-stage-[redacted]-r3	13,517.8	5,535.46	528.15 MBps	318.2 MBps	[redacted]
[redacted]-cstg-[redacted]	7,884.73	2,005.47	495.1 MBps	126.14 MBps	[redacted]
[redacted]-auth1a	6,026.73	4,959.43	1,221.67 MBps	1,098.39 MBps	[redacted]
[redacted]	4,755.8	1,885.83	345.77 MBps	147.02 MBps	[redacted]
[redacted]	1,696.53	347.05	123.11 MBps	10.57 MBps	[redacted]

The table shows the most demanding VM. You can identify the villain VM and compare their demands with the capabilities of the underlying IaaS. Knowing the infrastructure capability is important, because different class of SSD have different IOPS and throughput capabilities.

After identifying the villain VM, talk to the VM owners if the numbers are excessive during peak hours and identify the reasons behind the excessive usage. You must ensure that they do not create a hot spot, for example, vSAN cluster with >100 disk can handle numerous IOPS but if the VM objects are only on a few disks, those disks can become a hot spot.

Network Team

Similar to the problem face between Storage Team and Platform Team, VMware Admin needs to reach out to Network Team. A set of purpose-built dashboards will enable both teams to look at issue from the same point of view.

The network covers both the physical and virtual environment

The dashboards must answer the following basic questions for Network Team:

- What have I got?
    - This
    - What is the virtual network configuration? What are the networks, and how big are they?
    - We have NSX, Distributed Virtual Switches, Distributed Port Group, Data center, Cluster, ESXi, etc. How are they related? Distributed Switch does not span beyond vSphere Data Center. So data center is a logical choice to start analysing the relationship.
    - Who are the consumers of my network? Where are they located?
- Are they healthy?
    - Do we have any errors in our networks? Which port groups see packets dropped? If there is problem, which VMs or ESXi, are affected?
    - Do we have too many special packets? Broadcast, multicast and unknown packets. Who generates them and when?
    - The two primary metrics are bandwidth and latency. Bruce Davie explains both in this book[2], specifically this page, that the two metrics work together. The reason is some applications are latency sensitive, while others are bandwidth hungry.
- Are they optimized?
    - Just because something is healthy does not mean they are optimized. Look for opportunity to right size.

Once Network Admin know what they are facing, they are in better position to analyze:

- Performance
    - When VM Owner thinks Network is the culprit, can both Network Team and Platform verify that quickly?
- Configuration
    - Is the configuration consistent across objects of the same kind? Do they follow best practice?
    - What are the virtual networks, and how do they map into the physical top-of-rack switches?
- Capacity
    - Is any VM or ESXi near its peak in network? Which VXLAN is the busiest? Does it match their expectation?
    - Who are the top consumers for each physical data center? What's their workload pattern?
    - How is the workload distributed in this shared environment?

[2] The book is freely available, and that's part of the reason why I decided to make mine free.

- Which networks are not being used? Why? Could we decommission them as they could become security risk?

Network Top Talkers

Use the Network Top Talkers dashboard to monitor network demand in your IaaS. In a shared environment, a few VMs generating excessive activity can impact the entire data center. While a single VM might not cause a serious problem, a few of them can. Understanding high demand helps you monitor IaaS and allows you to plan capacity.

Network throughput, disk throughput, and disk IOPS can spike as their physical limit is generally very high per VM. This means that IaaS has enough capacity for all workloads and performs well until VMs start consuming abnormally high amounts of network and disk bandwidth.

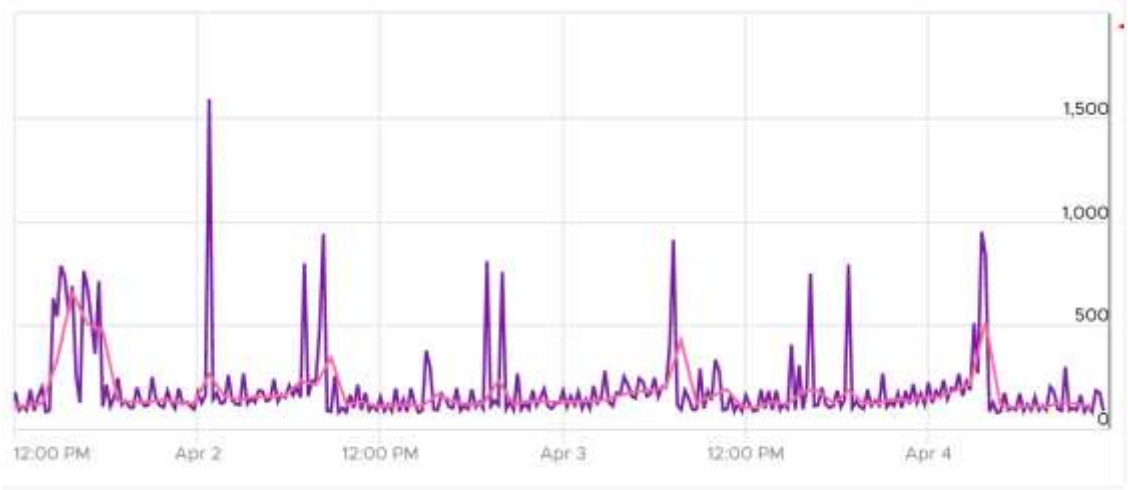
The dashboard shows the big picture, while allowing you to see the individual VM. It is important to see the VM utilization in the larger context.

The dashboard begins by showing the current workload. This is the total network load (received and transmit) from all vSphere environments monitored by VCF Operations. The idea is to give you an indicator on how hard the overall load is.

Select a data center from the data centers list.

- The columns show the number of clusters, ESXi Hosts and VM for each Data center. The VM count includes powered off VM. If you need to show only the running VM count, edit the widget.
- If you want to see from all Data center, select the **vSphere World** row.
- Upon selection, the Total Demand line chart and the Top Talkers table will be automatically filled up.

About the Total Demand line chart:



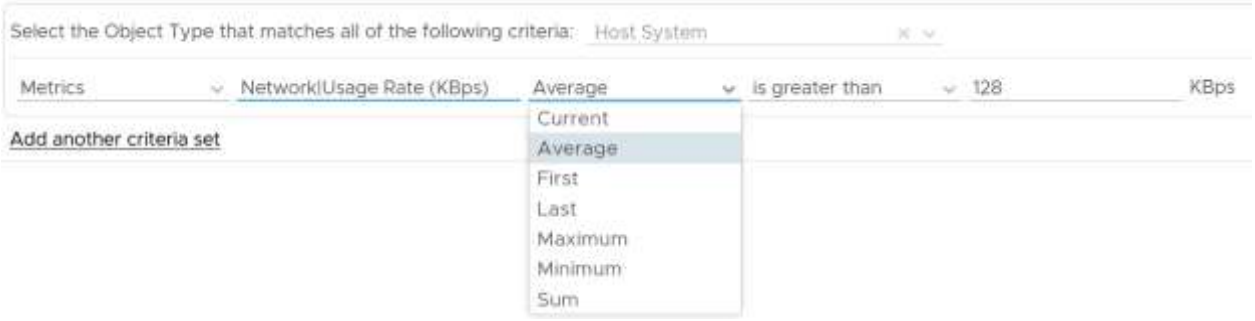
This shows the total throughput (received and transmit) in the selected data center.

It shows both the 5-minute peak and the hourly average into 1 line chart. You would expect that the 5-minute peak is much higher than the hourly average, indicating it is just a short burst. You can click on the metric name to hide the corresponding line chart.

About the Top Talkers table:

Name	5 minute Max	1 hour Max	No of NIC	ESXi	Datacenter
[redacted]1	14.06 Gbps	8.02 Gbps	2	[redacted].com	[redacted]
[redacted]2	13.97 Gbps	7.91 Gbps	2	[redacted].com	[redacted]
[redacted]3	13.83 Gbps	7.71 Gbps	2	[redacted].com	[redacted]
[redacted]12	13.66 Gbps	7.8 Gbps	2	[redacted].com	[redacted]
[redacted]3	13.63 Gbps	8.03 Gbps	2	[redacted].com	[redacted]
[redacted]5	13.6 Gbps	8.15 Gbps	2	[redacted].com	[redacted]

The table shows the most demanding VM. To help you focus on the VMs that hit the network hard, VMs that are not hitting the network hard are filtered out. The threshold used is 1 Megabyte/second sustained in the *last* collection cycle. You can change the filter to average of all time if that makes more sense to you.



You can identify the villain VM and compare their demands with the capabilities of the underlying IaaS. Knowing the infrastructure capability is important so you know the absolute limit.

After identifying the villain VM, talk to the VM owners if the numbers are excessive during peak hours and identify the reasons behind the excessive usage. You must ensure that they do not create a hot spot, for example, vSAN cluster with >100 disk can handle numerous IOPS but if the VM objects are only on a few disks, those disks can become a hot spot.

Migration Team

Workload migration from one infrastructure to another need to be tracked and monitored.

Migration Sign Off	Purpose is to get a sign off from VM Owner.
Migration Tracking	Purpose is to show overall progress of a long migration project to senior IT leaders

Migration Sign Off

Since we do not know which VMs are being migrated, we need to use VCF Operations group. Since the dashboard needs to know the group ID, I need to create the group in advance. You have to import the group, and not create your own. If you do, then you need to modify the dashboard to use yours.

Migration Groups

The purpose of the groups is to get a sign off from VM Owner, preferably a bulk sign off if you're migrating a large number of VMs in one migration window.

Create 1 group for each batch. Batch means VMs that will be migrated together. Ideally they are from the same owner or they have high dependency.

Create the group as early as possible. In this way, you get the longest historical data. This can be useful if there are many VMs with monthly or quarterly workload.

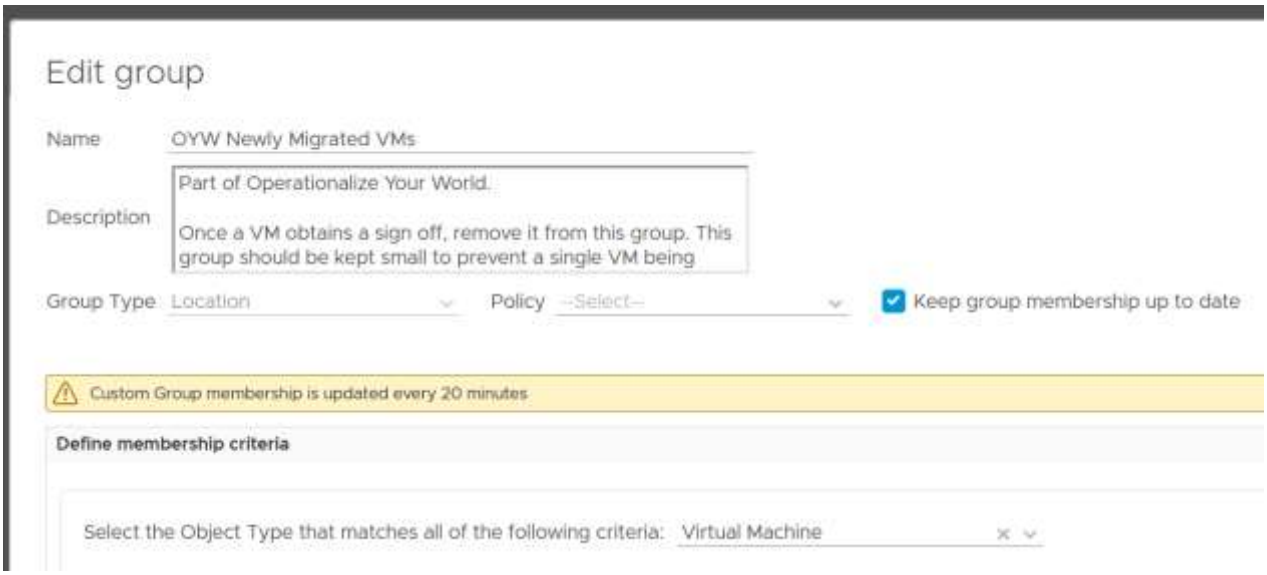
Make sure the group members do not change when comparing Before and After, as it makes comparison invalid. This includes any configuration change at both application and infrastructure level. For example, if after migration the workload increases due to whatever reason, it's unfair to compare with before migration as it's no longer apple to apple.

Have a deadline to minimize risk of changes. I recommend 1 week as your migration schedule is probably over the weekend. This enables you to move on and focus on the next batch.

Avoid very large group. Keep it below 100 VM. A large group can mask details. For example, a group of 1000 VM can have 50 VM performing worse. If the remaining VMs are performing better, the average number may not reveal that.

The group KPI is simply the average of its VM KPIs. Expect this to remain good or improve post migration.

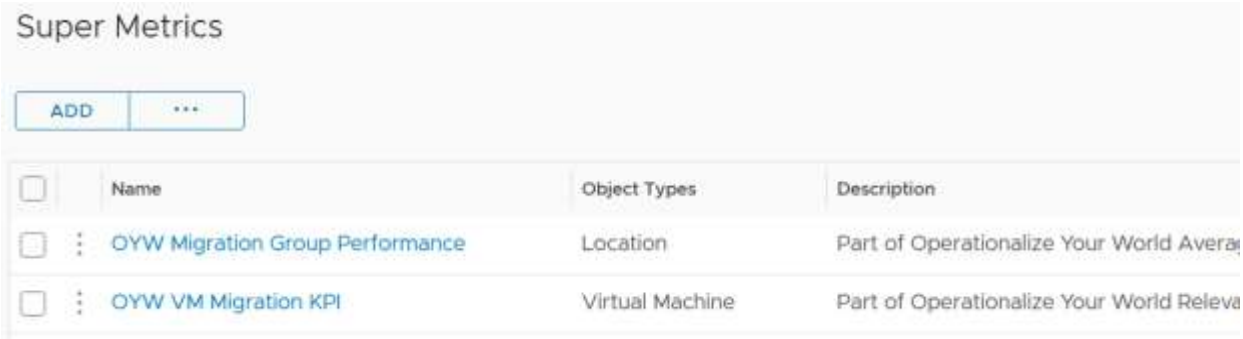
The following screenshot shows the group. Notice the Group Type is Location. That's the object type used in the super metric.



The group needs to have super metrics to summarize the overall performance.

The following screenshot shows the 2 super metrics used.

- The super metric "Migration Group Performance" is simply the average of the "VM Migration KPI". Average is a suitable function as it represents all the members VM in the group.
- The super metric "VM Migration KPI" takes a subset of the VM KPI metrics. It excludes the utilization metrics as they can create confusion when comparing before and after.



Migration KPI (%)

How do you compare the performance of a VM since there are many metrics? One way is to plot multiple line charts per VM.

This is time consuming since there are at least 9 metrics to check:

- Guest OS CPU Run Queue
- Guest OS Disk Queue Length
- Guest OS RAM page-out rate
- VM CPU Co-Stop
- VM CPU Ready
- VM CPU Overlap
- VM Memory Contention
- VM Disk Latency
- VM Network Dropped Transmitted Packet

Notice the above focuses on contention. No consumption metrics are used as their values are expected to go up.

The group KPI is simply the average of its VM KPIs. Expect this to remain good or improve post migration.

The solution is to create a metric. It's basically the same KPI used to calculate VM Performance (%) minus the consumption metrics.

To be Migrated VMs

This is an optional group as it does not involve VM Owner.

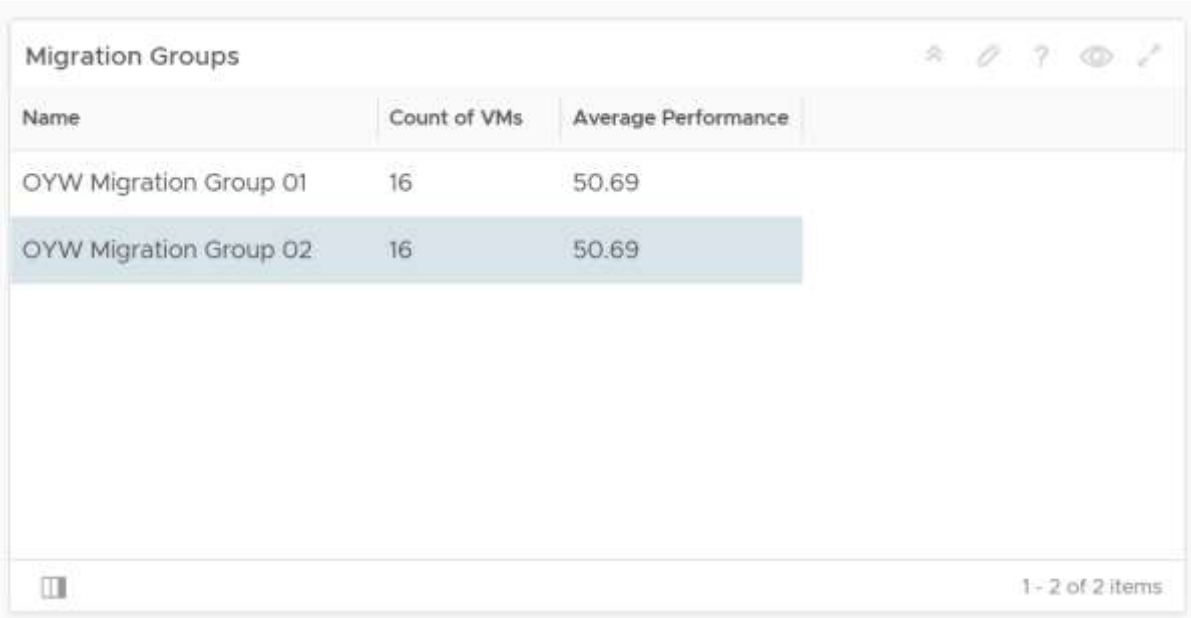
This group only contains VMs that are scheduled for the next migration window. Do not put VMs that are not going in the next schedule, as their load changes the group value.

If the VMs in the group are going to multiple destination, then you need to create 1 group per destination.

Once a VM is migrated, remove it from this group.

Summary Section

The dashboard section starts with a list of all the migration VM groups.



To see a group details, select the row. All the other widgets will be automatically refreshed.

The first thing to check is the health chart. As we need something to show over time that the migration is successful, we pick a color coded chart. The chart shows the average performance of all VMs in the group. You should expect the value to be same or better after migration.



In migration, you want to have visibility into the application of services. I added a bar chart to show the services discovered by the Service Discovery adapter.

Ideally, install Telegraf agent so you have process-level metrics.

Detail Section

We need to use 2 separate tables due to limitation. You need to set the 2 dates manually as different migration groups have different migration dates.

The tables list all the VMs in the selected migration group, along with their relevant contention metrics, consumption metrics and business context.



CPU Ready, CPU Co-Stop and Memory Contention should go down, unless the numbers are already very good to begin with.

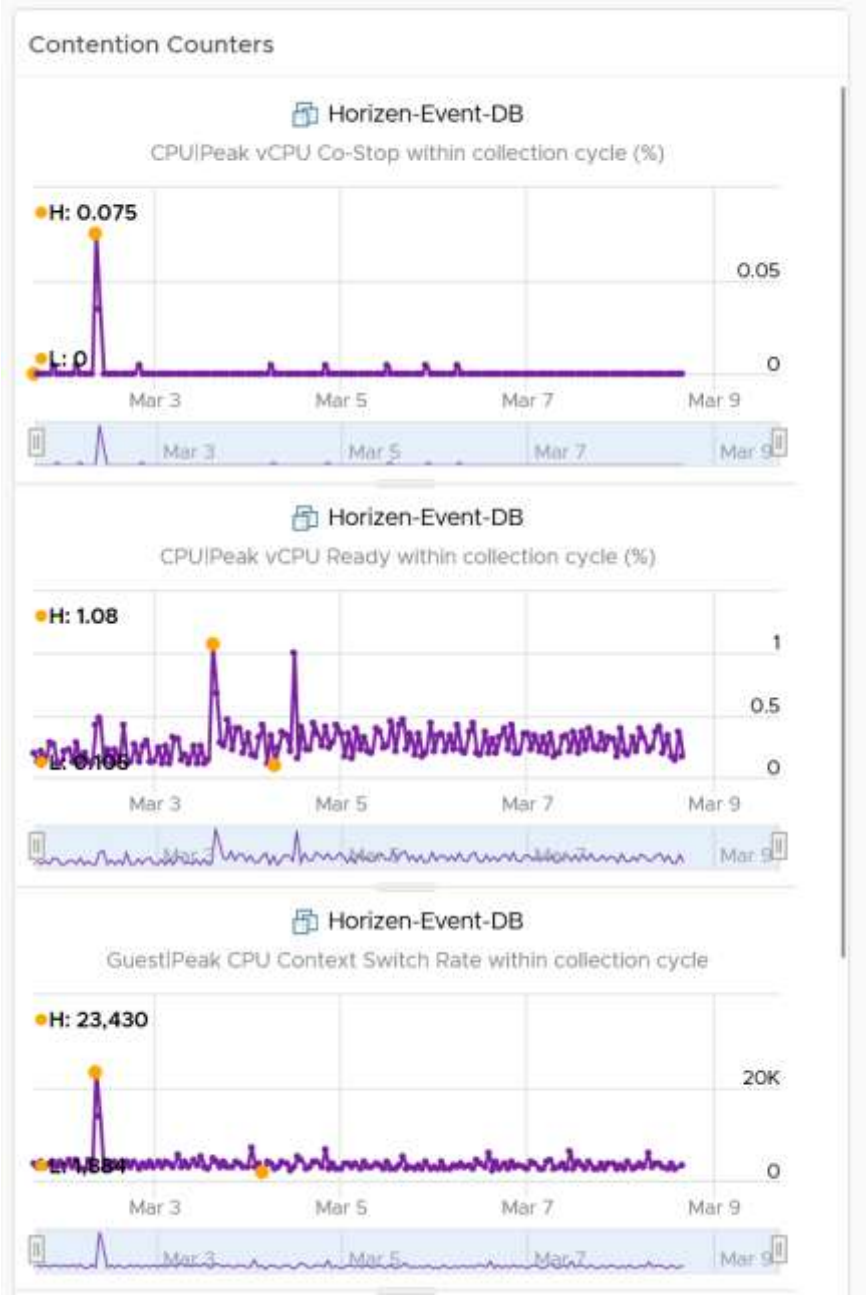
The numbers are taken at 20-second average, so they report much higher number than the 5-minute average.

Any numbers less than 2% is not something to worry about.

Disk Latency and Disk Queue should remain low even with more IOPS, assuming the new DC has much better storage subsystem.

CPU Context Switch could go up if the CPU Usage goes up, as there are more threads or execution to be done.

CPU Queue could go up if the CPU Usage goes up. So long the value is less than 3 per vCPU, there is no need to add more vCPU



I added a set of consumption metrics as we expect the consumption to go up.

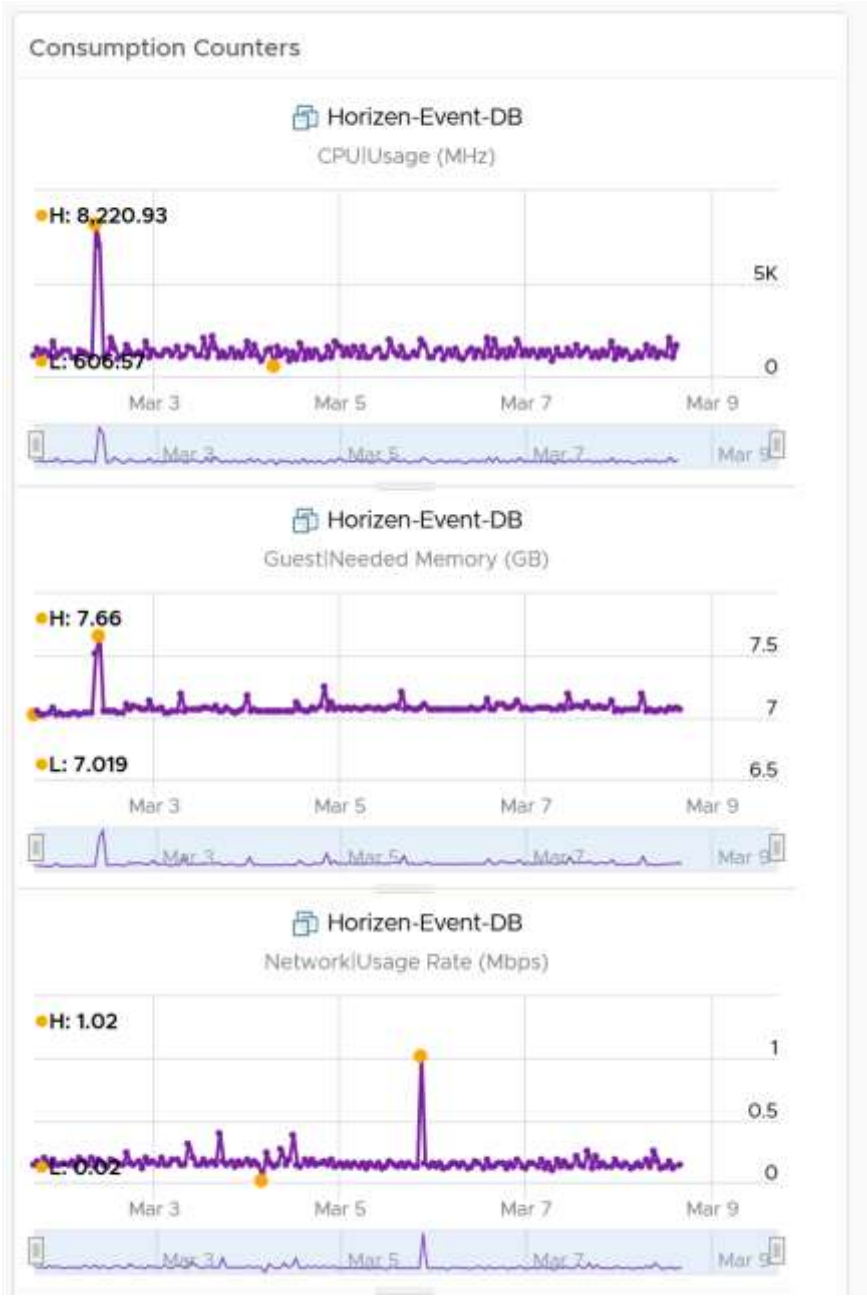
For CPU, we need to use GHz and not percentage as the 2 data centers could use a different CPU speed.

CPU Usage could be tricky to compare if the new data center uses a different CPU architecture, meaning 1 GHz in the old DC does not equate to same amount of CPU processing with 1 GHz in the new DC.

For memory, we have to use the one inside Guest OS. The counter at VM level is not relevant. For details, refer to the memory chapter.

For disk, we use both IOPS and throughput.

For network, we use network throughput. You can add packets per second if you think the behaviour changes.



Migration Tracking

The dashboard should compare both data centers. The layout should be 2 equal columns (Old DC and New DC). This is relevant in the case of 1:1 migration and a huge chunk of the VMs in the old DC are being migrated. If the old DC will continue having new VMs, then the comparison will be invalid.

The first row shows a summary

- Number of running VM. This shows you the overall progress. You expect the number on the old DC to go down over time. I exclude powered off VMs as you may not be keen on migrating them.
- Overall CPU Utilization. You should expect the number in the new DC to go up while the number in the old DC to go down
- Overall Memory Utilization
- Overall Disk IOPS
- Overall Disk Throughput
- Overall Network Throughput
- Overall Network Packets/second.

The second row shows the list of clusters

- If you select a cluster, you can drill down into its performance and capacity.

The third row shows the list of VMs

- If you select a VM, you can drill down into its performance and capacity.

The fourth row shows the list of datastores

- If you select a datastore, you can drill down into its performance and capacity.

The following table shows a sample design of a comparison dashboard. It has two identical columns, allowing you to do show before vs after comparison.

Overall Monitoring	
List of on-prem clusters to be migrated No of ESXi, No of core, RAM, No of running VM	Same info, but on VMC
Line Chart Selected Cluster Performance (%) Shown over 1 month, 5 minute average.	Same info, but on selected VMC cluster
Line Chart Selected Cluster Worst Disk Latency (ms) Shown over 1 month, 5 minute average.	Same info, but on selected VMC cluster
Line Charts (CPU, RAM) Selected Cluster Utilization (%) Shown over 1 month, hourly average	Same info, but on selected VMC cluster
Line Chart (Disk IOPS) Selected Cluster Disk IOPS. Shown over 1 month, hourly average	Same info, but on selected VMC cluster

Logs Analysis

Part 2 Chapter 9

Those of us who have experienced troubleshooting VMware Cloud Foundation (or any enterprise infrastructure) know the importance of logs. In fact, it is the first thing that VMware Support and many other vendors support ask you. The reason is there are many useful information that only exists in the logs. They don't surface as alerts, metrics or events. The information is not limited to just security and compliance. They cover availability and performance too.

Having said that, >99% of the logs are not useful, especially after a couple of days. So how do you minimize the cost while maximizing the benefit?

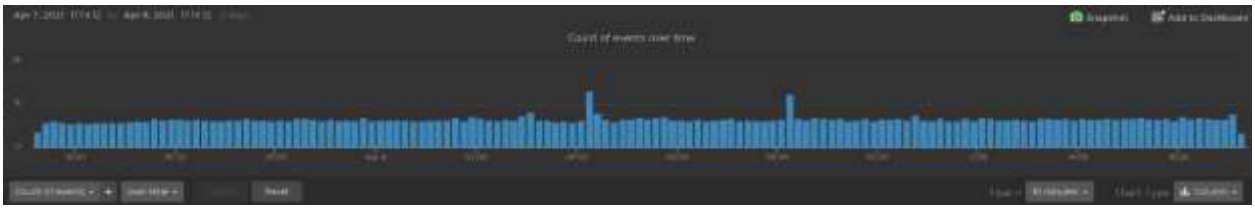
vSphere

Log Insight provides the ability to slice and dice vCenter events, tasks and alarms. This can be handy in audit investigation.

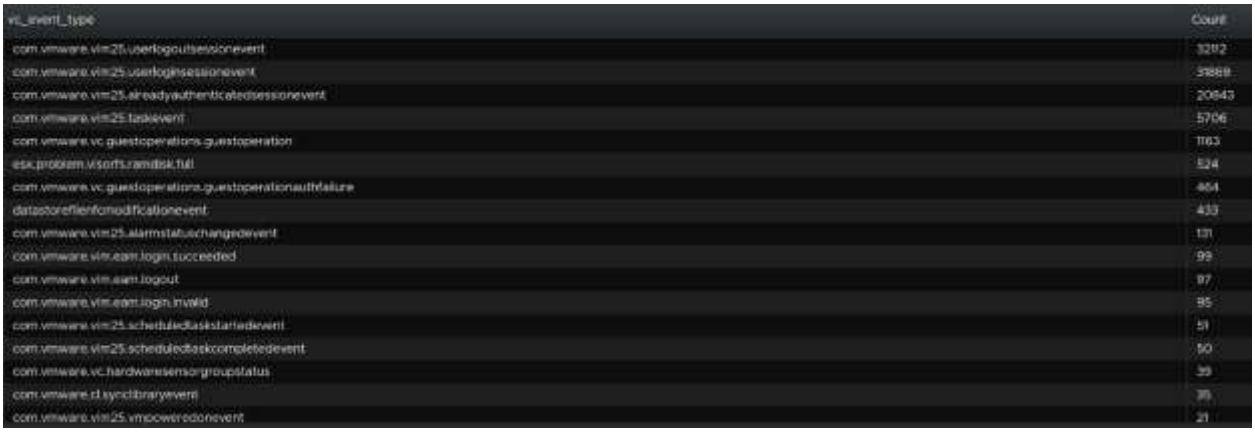
vCenter Events Analyzis

To see all the vCenter events, all we need is to select a built-in variable called VC Event Type. A log entry that has this field exist will appear.

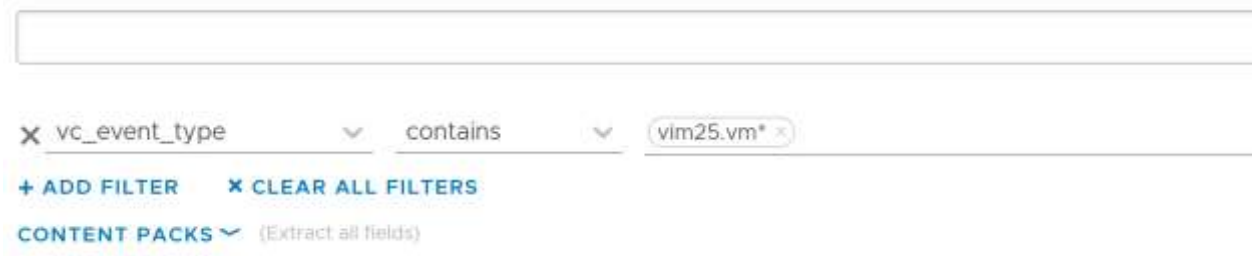
From the following chart, we can see there are steady stream of events every 10 minutes. You can change the data granularity.



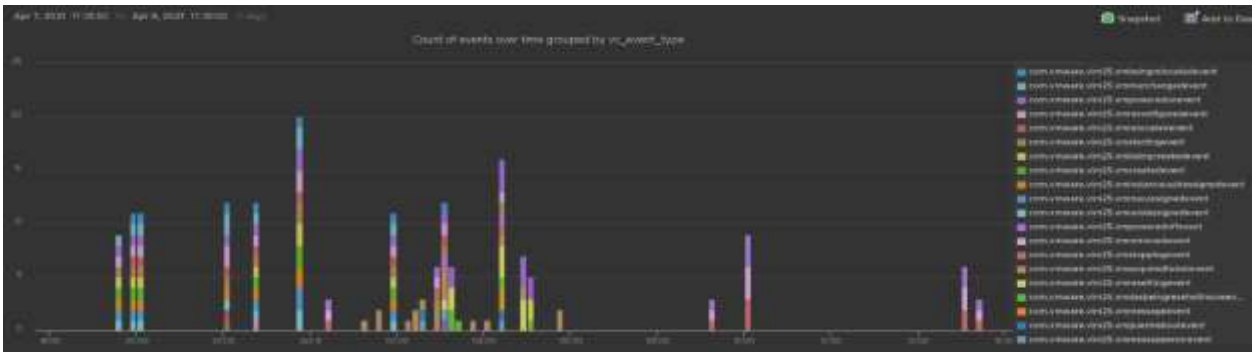
We can see actual event names by changing into a table. We can sort it to show the top events if required.



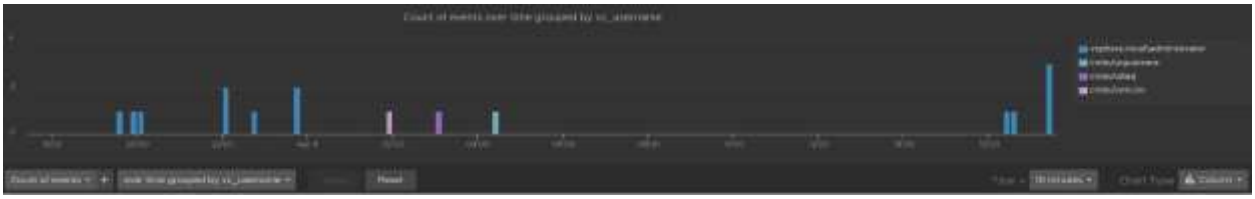
We are interested in events impacting our consumer (VMs), so let's filter it out. The filter is vim25.vm* as that is what vCenter shows in its log as we can see from the above table. I did not know that vCenter uses **vim25.vm***, but looking at the table above I could make an educated guess.



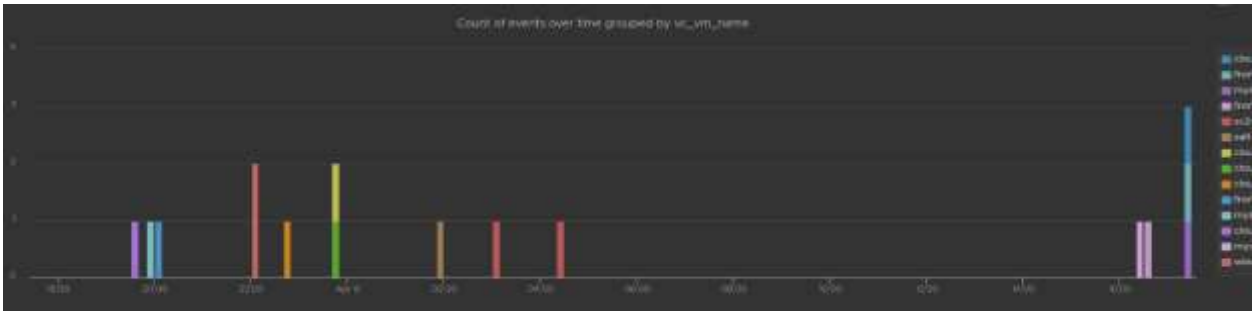
Once the above filter is set, I rerun the search and get all events impacting VM.



Let's zoom into one of the events. Let's say we're interested in VM configuration event and want to know what exactly was changed. So let's zoom into that event and group it by the user who changed it. We get the following chart, showing the user in the legend.



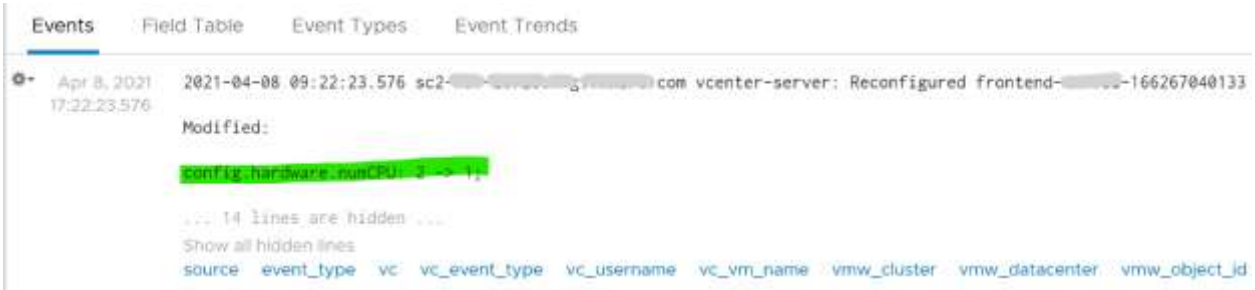
You can group the data by VM name to see which VM were changed when. I've cropped the VM name in the legend.



You can see the result in tabular format. You have the VM name, user name and additional context such as the parent ESXi host at the time of the change.

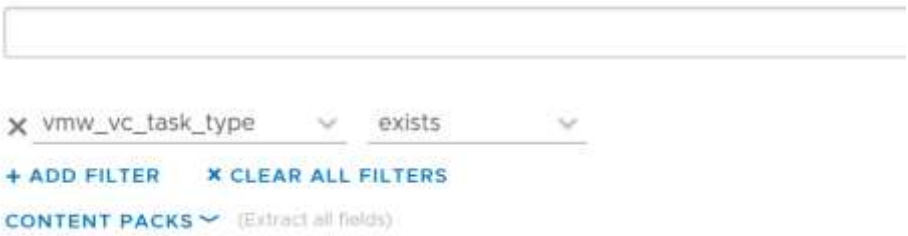


Last but not least, you can see the actual change, as highlighted in green.

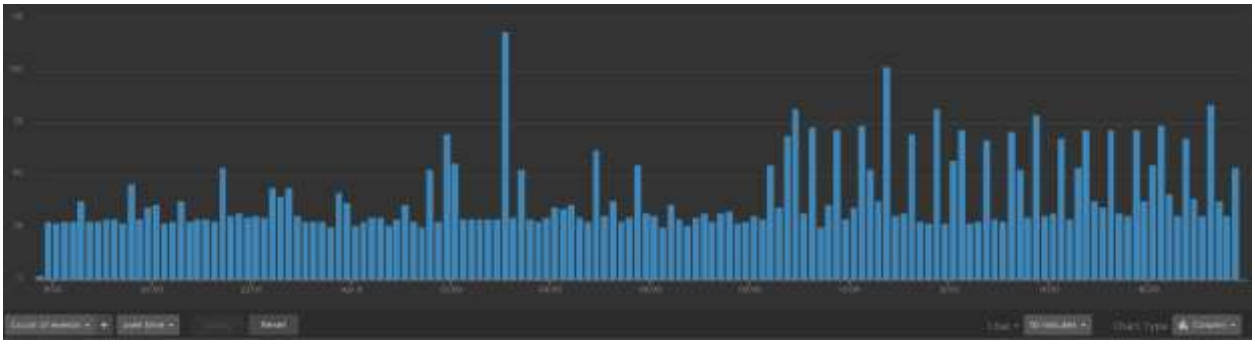


vCenter Tasks Analyzis

To see all the vCenter Tasks, all we need is to select a built-in variable called VC Task Type. A log entry that has this field exist will appear.



We can see that there is a regular stream of events throughout the day. The pattern looks normal.



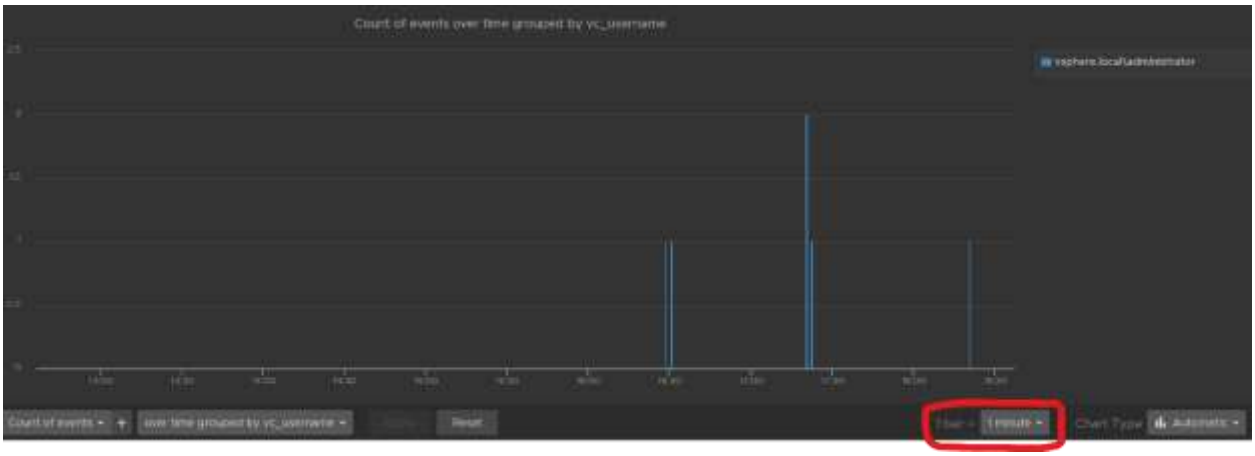
Let's show the top tasks by showing the result in table format.

vmw_vc_task_type	Count
delete a virtual storage object	2039
delete container volume	1982
update container volume	832
fetch content of a library item	520
query	179
check new notifications	48
sync library	36
power on virtual machine	15
deploy ovf template	13
reconfigure virtual machine	13
call drs for cross vmotion placement recommendations	12
deploy ovf package from content library to resource pool	12
transfer file(s)	12
reset virtual machine	11

To zoom into any of the tasks, we specify the task name. I only specify one below, but it can take multiple.



Using the above, and limiting the result to a narrower time window, we can zoom into the nearest minute.

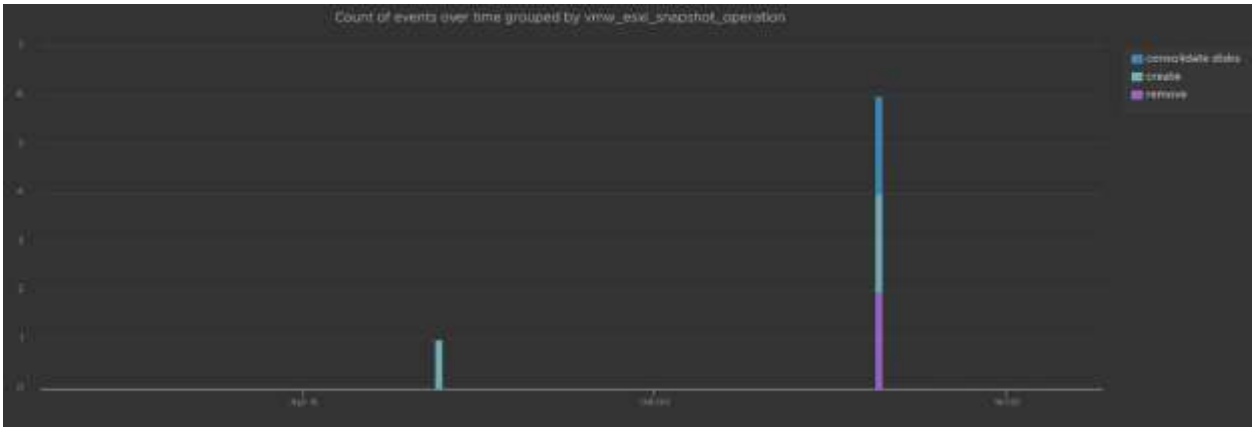


Snapshot Analyzis

You can visually see all the snapshot operations with a single filter **vmw_esxi_snapshot_operation**. Just use the **exist** operator.



Group the data by the operations and you will get something like this. I can see there are 3 snapshots created but only two were removed. So one of the VM still has a snapshot.



The above shows the time too. In production, you should not take snapshot during busy hours, especially on mission critical VM. So if you run the query in the last 1 week, you should expect no data during the busy hours, and all the daily back up should appear within the backup window.

You can see the details such as the VM name and other context, so see which VM did not have its snapshot removed.



You can check the snapshot name (partially masked out in grey) and whether the snapshot include memory.

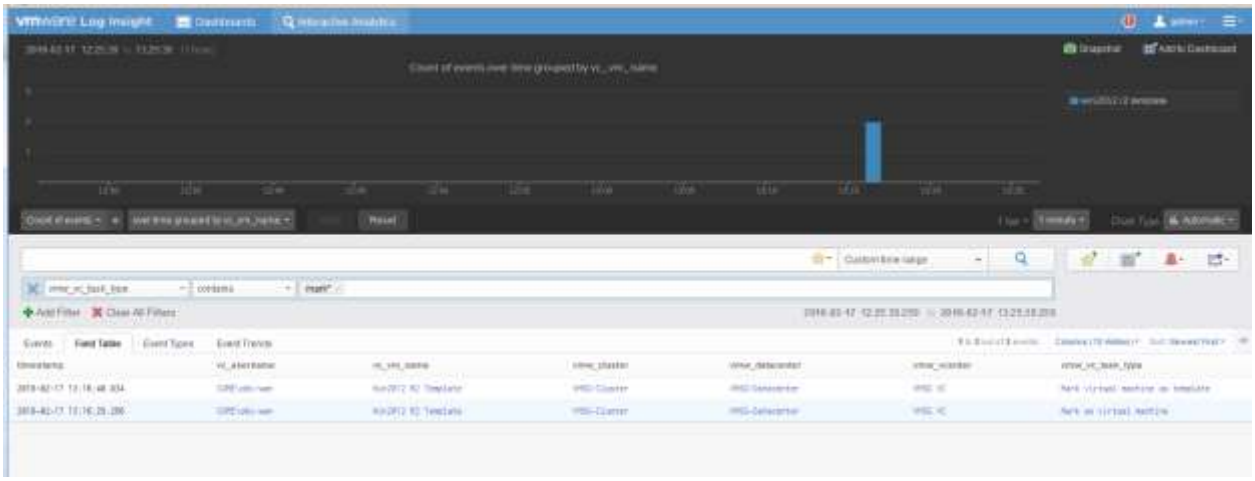


Template Analyzis

How do you prove to auditor that your VM templates have not been modified by unauthorised person. If a template has been modified, you want to know who did it.

The good thing is there are only a few things you can change to a template. You can rename the template, change the permission, and convert it into a VM. All other changes require the template to be converted into a VM first. That means we can focus on this conversion.

The vCenter logs the entry as "mark virtual machine as template" when you convert a VM into a template. When it is converted back to, it writes "mark as virtual machine". So it's a matter of tracking these 2 entries.



vSphere Health

In general, you know that you did a good job with your vSphere IaaS as the VM owners are happy with the performance of their VMs. The business is powered by the VMware infrastructure that you design and operate. However, there is a chance that the vSphere, NSX, vSAN logs bear evidence of hidden issues which are not visible from the UI.

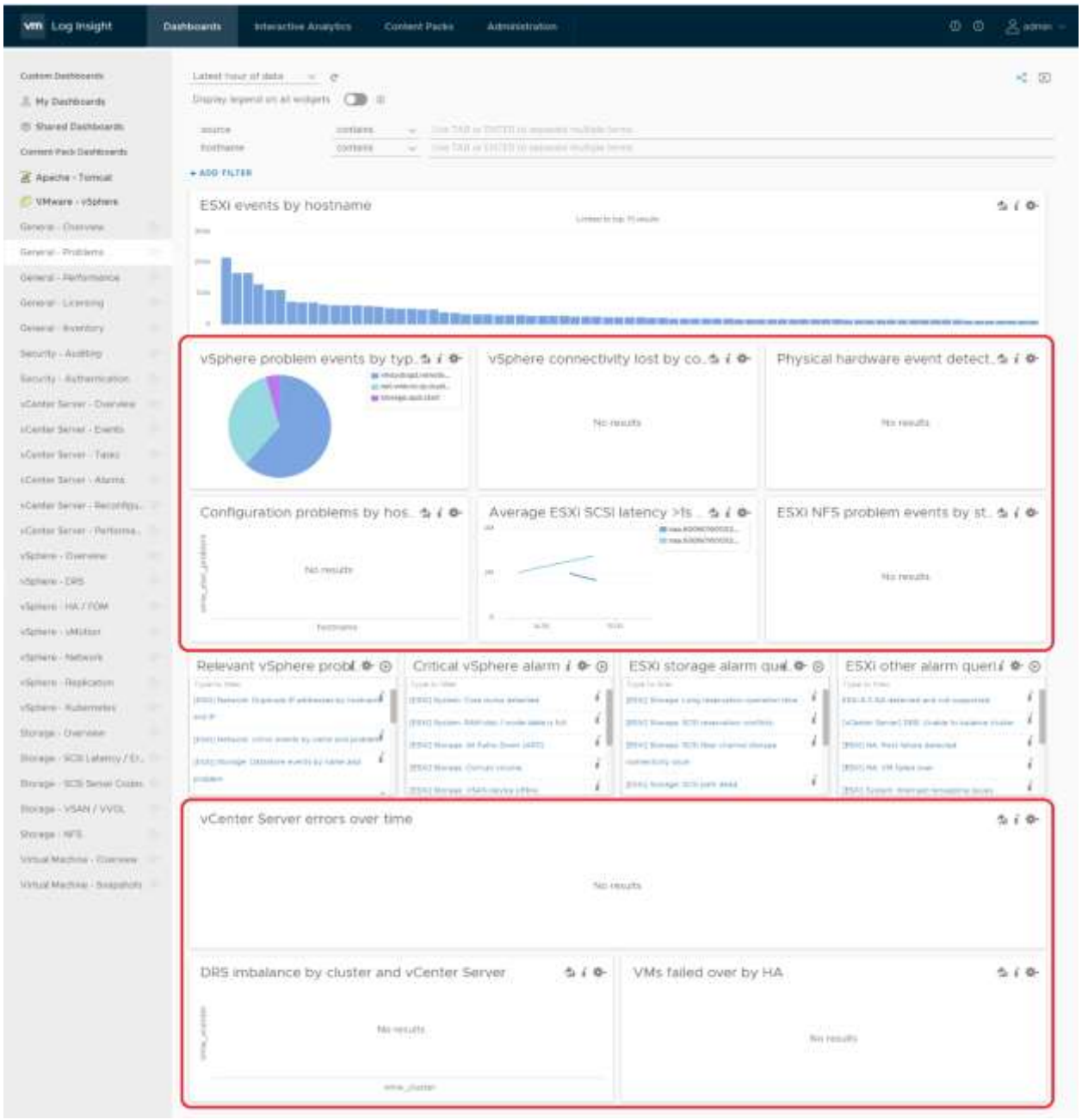
As VMware professionals, we know vSphere well and probably have years of experience working with vSphere. We can architect, design, implement, upgrade, and troubleshoot it.

The same thing cannot be said about the logs. Generally speaking, deep knowledge of vSphere logs belongs to VMware GSS engineers, as they read logs on a daily basis, and perform all kinds of troubleshooting activities. That knowledge has been transitioned to Log Insight release after release in the form of a vSphere content pack.

One common question I get from customers is how to prove that there is not hidden warning lurking around in the log files. As you know, vSphere produces a lot of logs.

Your first stop should be the General Problems dashboard in Log Insight. This dashboard checks the health of your vSphere using 8 queries. You expect a flying color, meaning it should be blank like this. That means vSphere has not logged any issues.

BTW, by the default the detail log files of individual VM is not collected. The file **vmware.log** contains detailed VM activity messages including reconfiguration events, vmotions, VMware tools messages, memory state, power on/off events, features enabled, API requests, etc. It can be used to troubleshoot events leading up to core dumps or kernel panics. For more information, read this blog by Julie Roman.

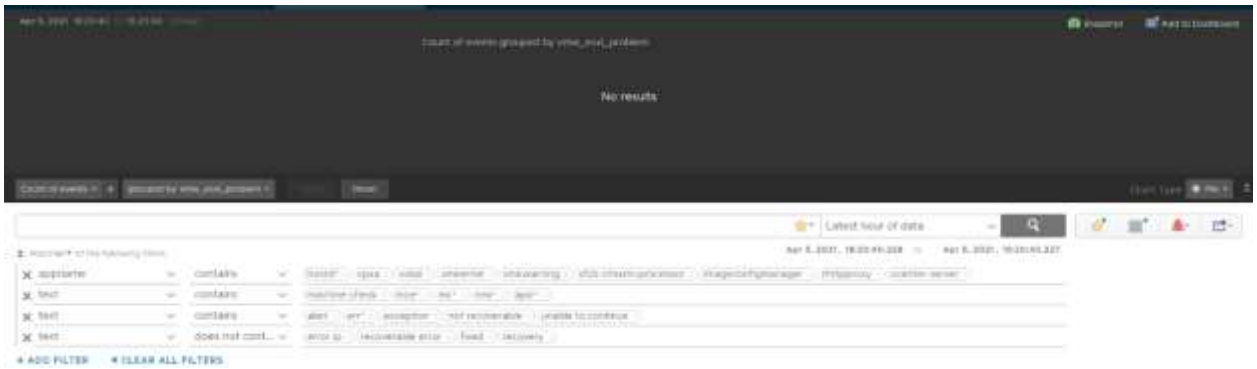


Let's look at some of the queries that Log Insight runs. The SCSI latency is based on 1 second, which is 1,000,000 microseconds. Here is what the query looks like:



1 second is on the high side; you can change it to a lower number. Do note that this is from VMkernel viewpoint and it's taking 1 SCSI operation (1 read or 1 write), so the number will be much higher than vCenter average. I've seen 12 ms value in vCenter (from the real time chart, so it is a 20 second average) became 600 ms.

The above query is simple, as it's looking for a specific item. Here is a much broader health check.



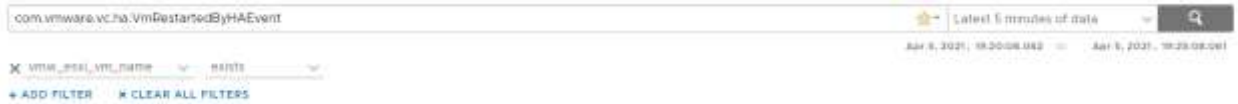
The example below checks for any errors in the vCenter which have not yet been reported as an alarm.



This query below checks for cluster imbalance.



And this query tracks for VM which were rebooted due to HA.



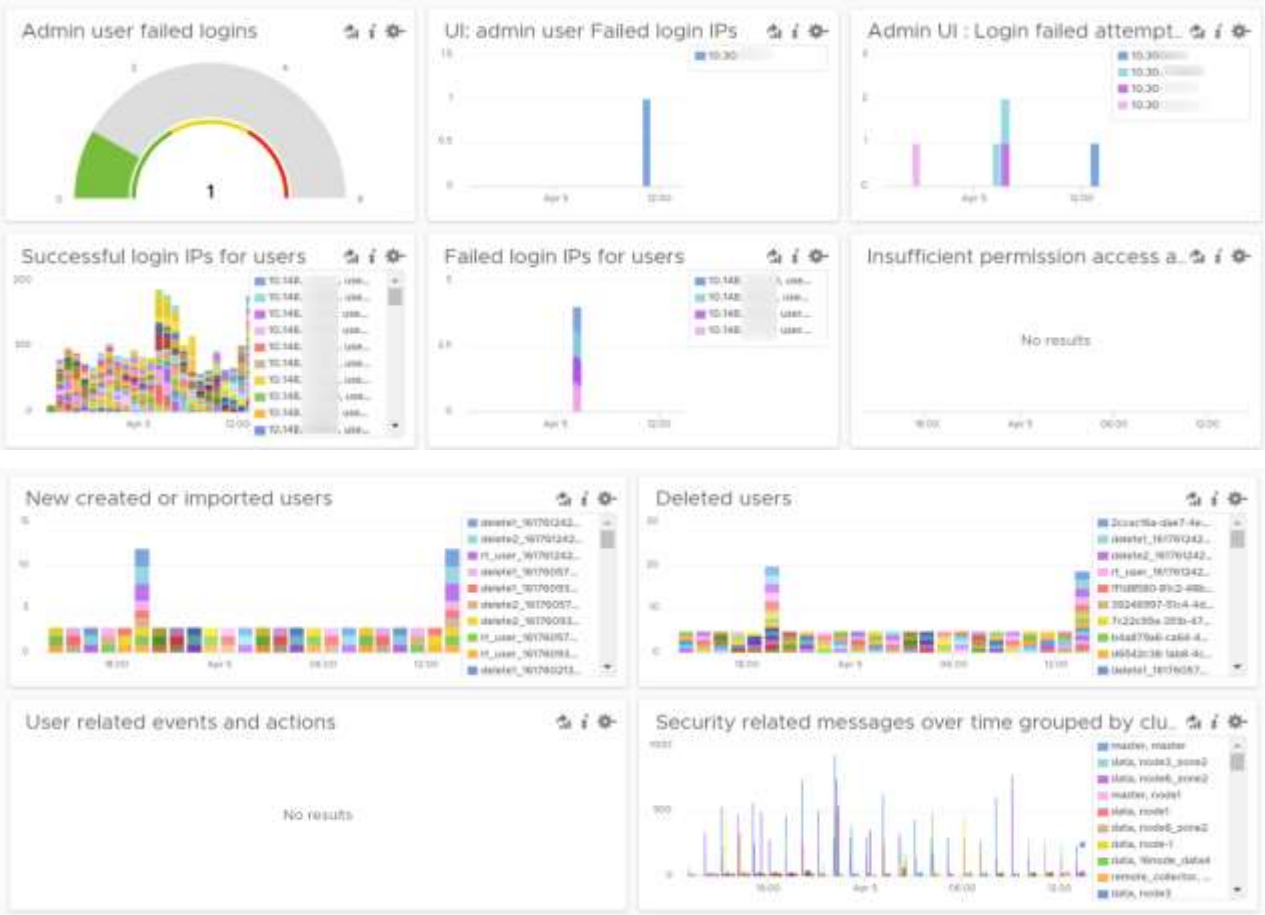
All the above widgets are what you would check out first. You might also want to ensure that there are no errors across major vSphere components.



Aria Operations

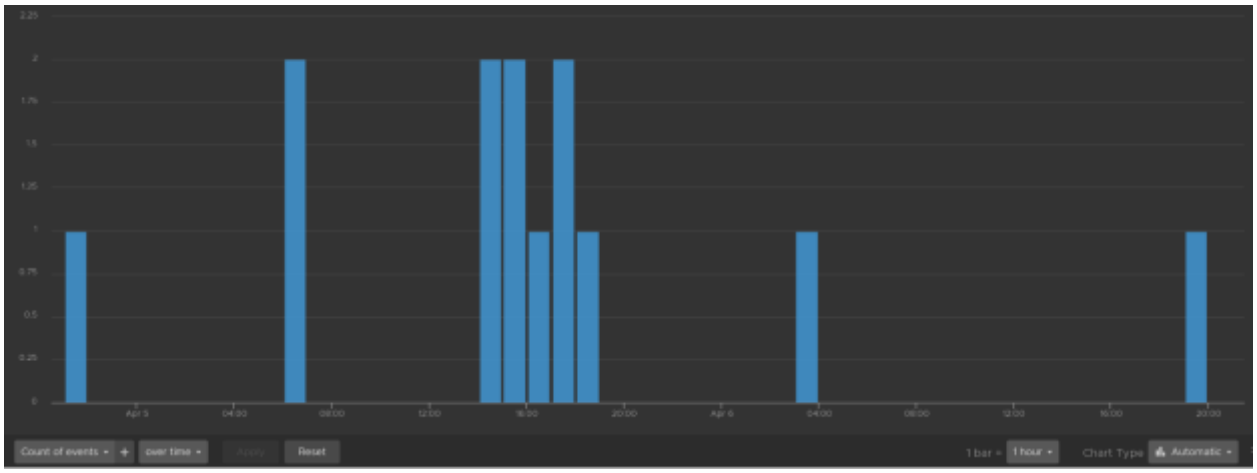
Log Insight sports out of the box dashboards that visualizes security-related activities in your Aria Operations. It also has a dashboard to help you watch and troubleshoot upgrade.

Authentication Analysis



Let's go through some of its contents in-depth.

There are two privileged login IDs that customers have access that audit requires compliance reporting: root and admin. Other privileged account such as MaintenanceAdmin is not accessible. How many times do admin ID login with the wrong passwords in a specific period? The following chart counts each time a login failure happens.



The count is possible because Log Insights create a field out of the log entries. The field is named vmw_vr_ops_admin_attempt and that's what plotted over time. I'm showing two examples of actual log entries, with value added context by Log Insight.



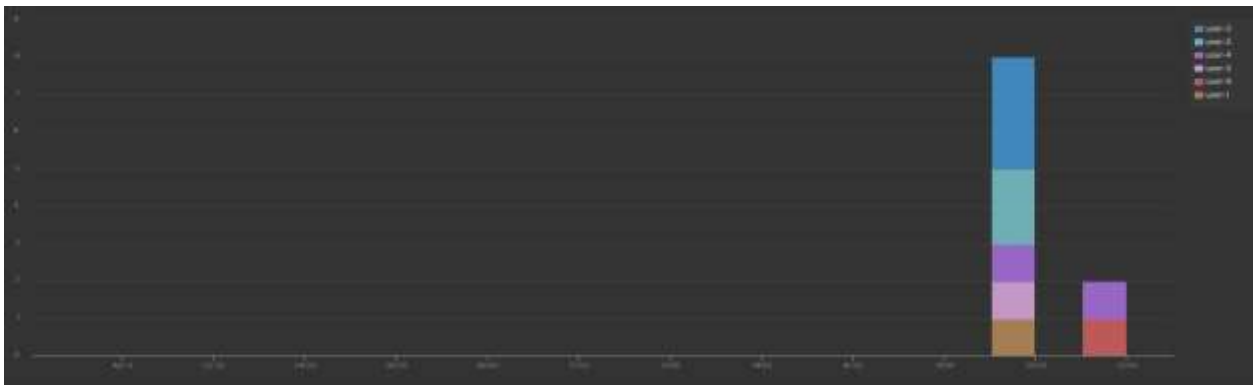
The log entries themselves are in turn filtered using the following query. Log Insight has awareness of Aria Operations via the variable vmw_vr_ops_appname.



Aria Operations has 2 UI that are accessed via separate URLs. The Admin UI is for platform administration such as upgrading Aria Operations, so it's important to track login activities. How many failed login attempts made in the Admin UI is also provided out of the box.

We've covered admin. How about root? As this is a Linux account as opposed to Aria Operations account, check the Linux content pack. The concept is the same.

What IP address do they login from? This helps you trace the location of the user who used the account. The following chart shows 6 different users and when they log in. We can drill down to any of them to see the exact time and the IP address used.

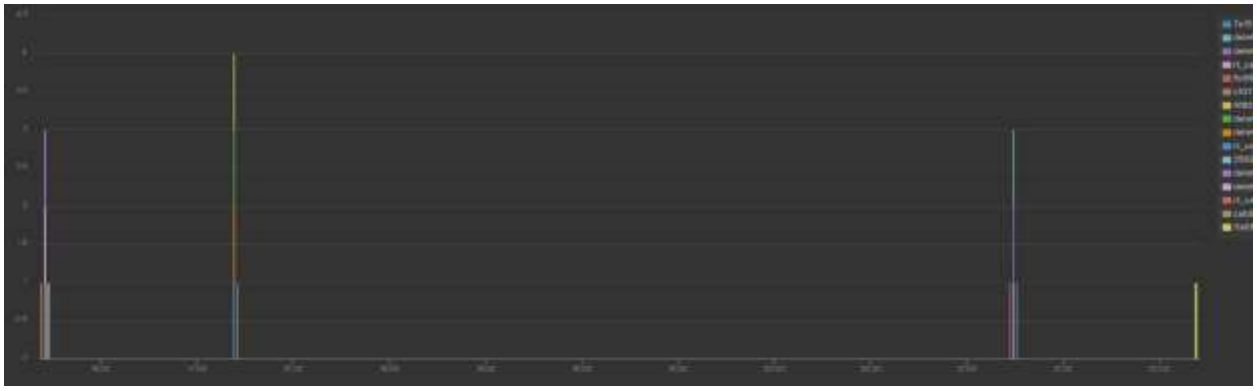


Alternatively, you can present in table format to show the IP address information. If you want the time stamp, use the **Field Table** feature.



Users addition, especially unexpected ones, can be a cause of security concern. Audit team may ask for the lists of users added in certain period. Users deletion is typically not an audit concern, but could be useful in troubleshooting. You can figure who deleted the user account and when.

The widgets in the third row of the dashboard covers users creation, import and deletion. By now you can guess that we can plot this event over time too. The following shows the list of user accounts that got deleted and when. I cut the full name as that's part of actual product validation.



The query that produces the above chart is the following. To some extent, it's actually human-readable!



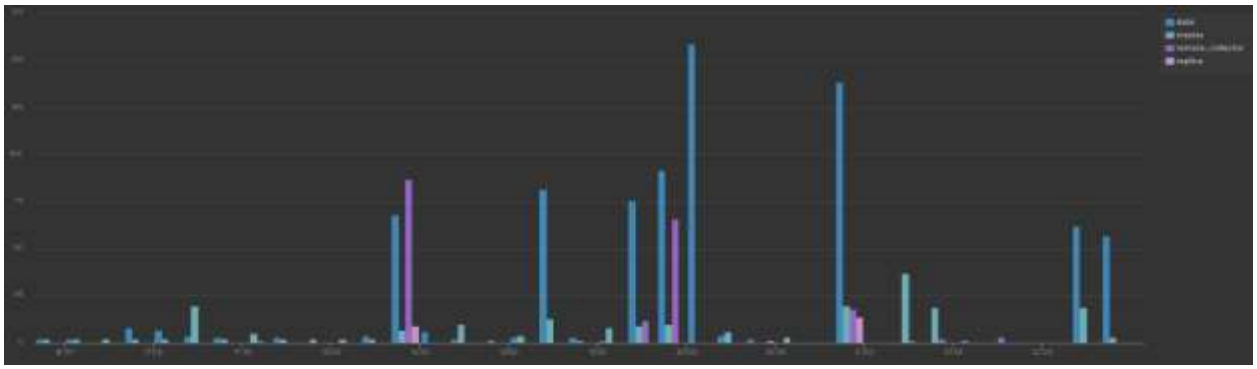
Insufficient permission access logs events where users tried performing activities that they do not have sufficient privilege. The following shows some example where users were denied access.



If user has the access, you will see something like this



The widget "Security related message" is my personal favourite, as it demonstrate the adhoc troubleshooting ability of Log Insight. The following shows the many security related messages. They are grouped by the Aria Operations nodes, so you can exclude certain types or zoom into a particular type.



How was the above achieved? The following shows the actual query. Log Insight automatically group log entries of similar type and give them a unique **event_type**. This means you can filter out the types that are not relevant, like what I have done below.



Activity Audit



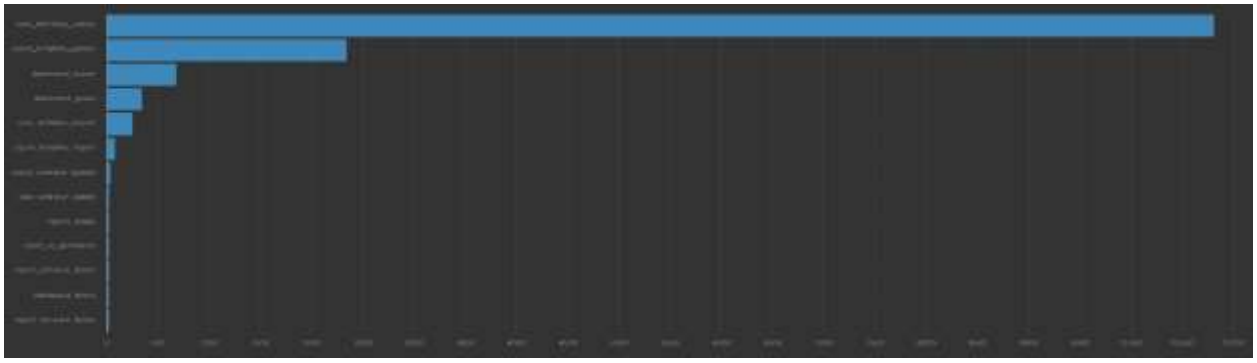
Use this dashboard to select a user and audit that user's actions. The type of actions you can audit are

- Dashboards, Views and Reports. This covers creation, update, deletion, import, schedule (report) and generate (report).
- Alerts. This covers alert definition, symptom definition, notification rules and recommendation.
- Environment. This covers Application, Custom Data Center and Custom Group activities

- Inventory. This covers resource creation, resource deletion, resource changes, collection start, collection stop, maintenance mode start, maintenance mode end.
- Configuration. This covers changes in global setting, cost settings, credential, collector groups, etc.

The widgets are separated as in the menu of Aria Operations Manager UI, so you can easily to orientate.

Let's go through some of its contents in-depth.



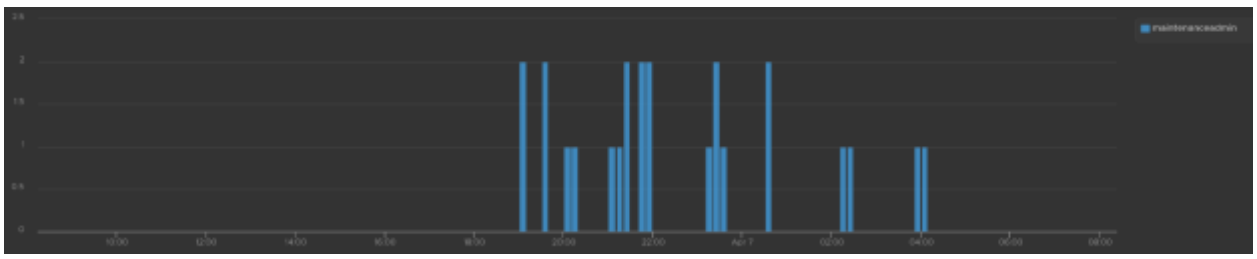
The query that produces the above chart is the following. I've excluded **dashboard_add_tab** from the category as it dominates the chart.



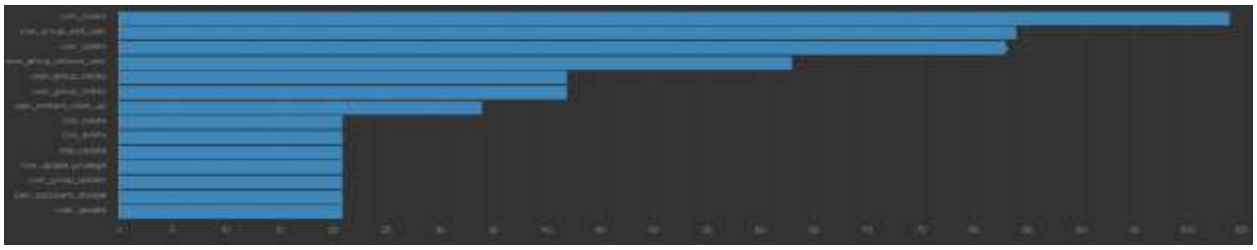
Let's drill down into a specific task. Let's say we have some views deleted and we need to know who deleted them and when. For that, we select the view_definition_delete and add it. Log Insight will automatically add the field name and operator. No need to manually type!



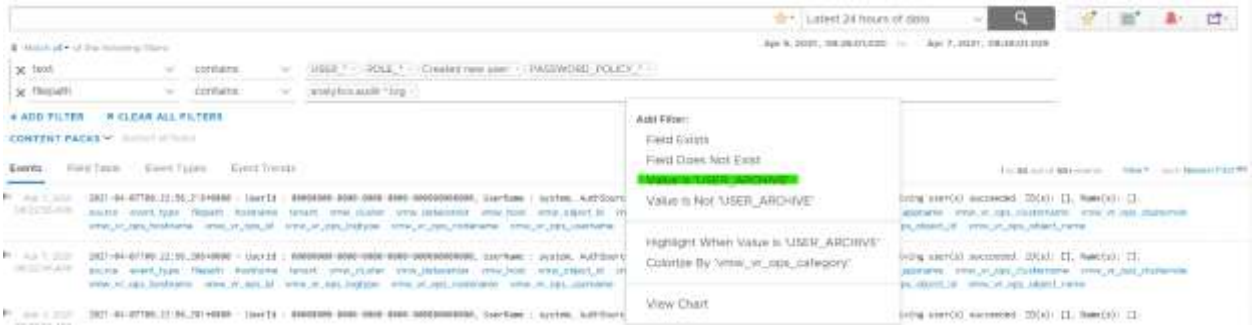
Since we're down into a single activity, we can now plot the chart over time, grouped by the user. We can see here that the user is the system account maintenanceadmin.



Let's take another example from the dashboard. This time we will take user management, where you can track things like user deletion, role creation, user password change and many others. The following shows some of those activities.



I've filtered out user_archive event as its value dominates the chart. As mentioned earlier, no need to manually type. Simply click on one of the log entries, choose a filter from the pop-up menu and that's it!

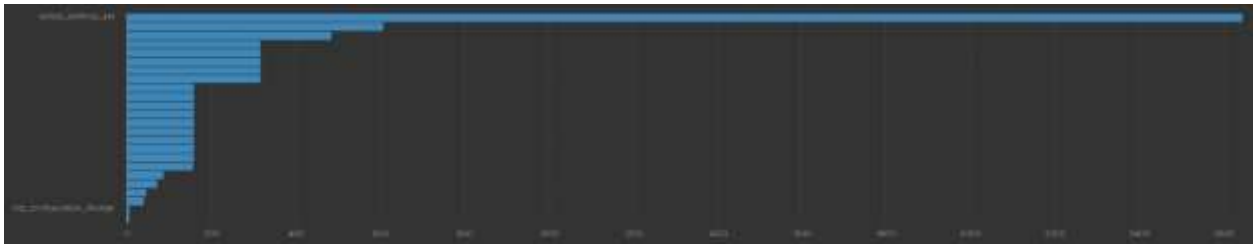


As usual, you can have a table of who did what when.

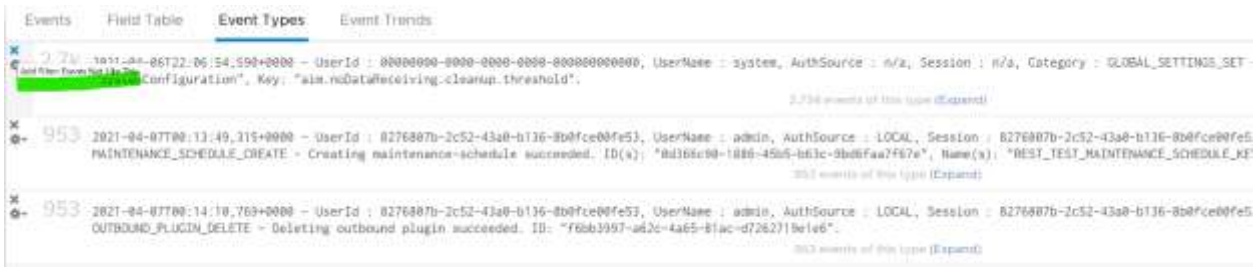
Events | Field Table | Event Types | Event Trends

timestamp	vmw_vr_ops_username	vmw_vr_ops_category	vmw_vr_ops_useraction
Apr 7, 2021, 8:13:48 AM	maintenanceAdmin	USER_DELETE	Deleting user(s) succeeded
Apr 7, 2021, 8:13:41 AM	maintenanceAdmin	USER_CREATE	Creating user(s) succeeded
Apr 7, 2021, 8:12:51 AM	admin	USER_DELETE	Deleting user(s) succeeded

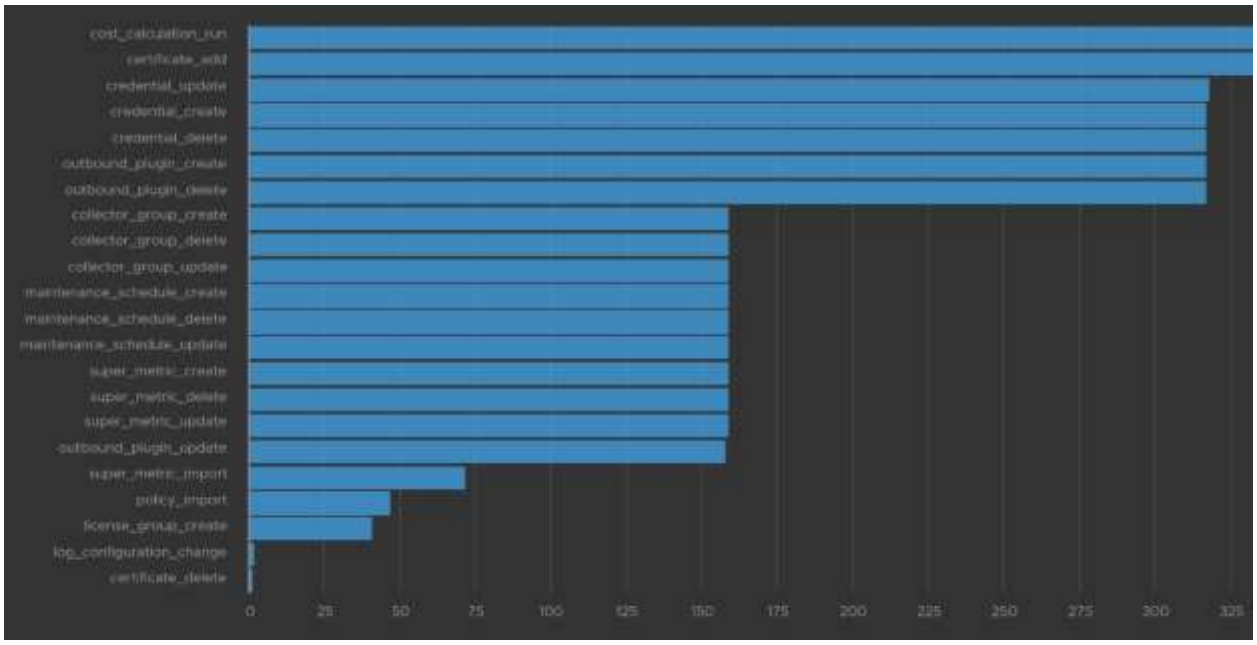
Let's take one last example. I will take the configuration activity as it shows a range of interesting events. As usual, I started with a bar chart as it lets me see the activity name. We can see that changes in global settings dominate the result.



We already know how to filter it, so I'll show you another way. Go to the Event Types. You'll see the log entries grouped by type of events. To filter out, simply click on the **X** icon.



Once filtered out, the following is what I get. Let us know if there are events that you need to be logged that's not trapped.

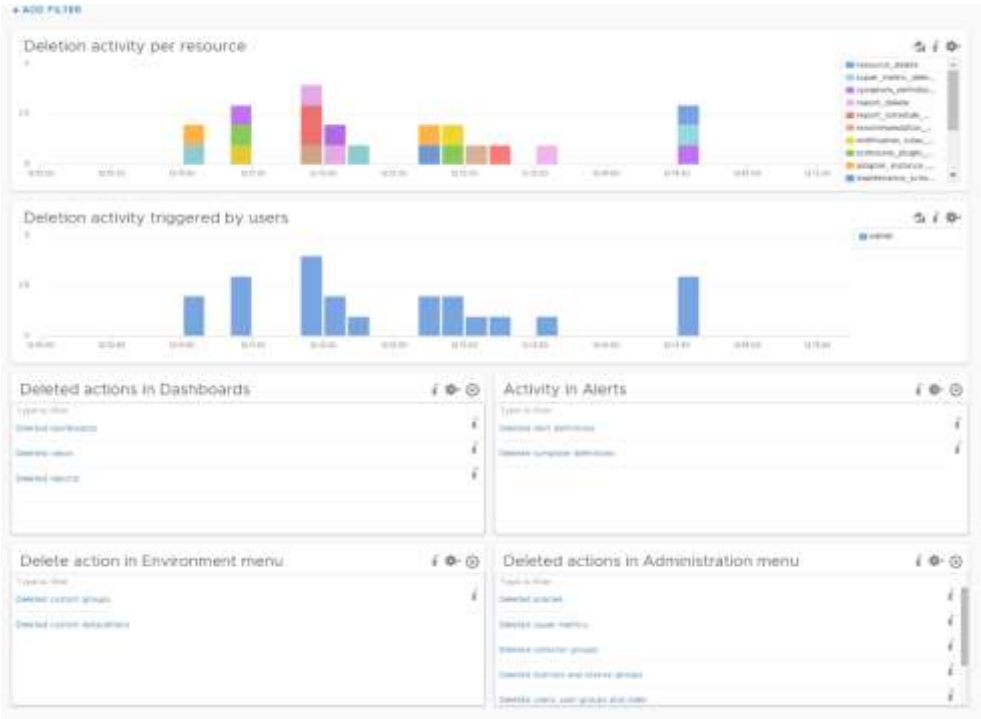


The query to get the above is complex

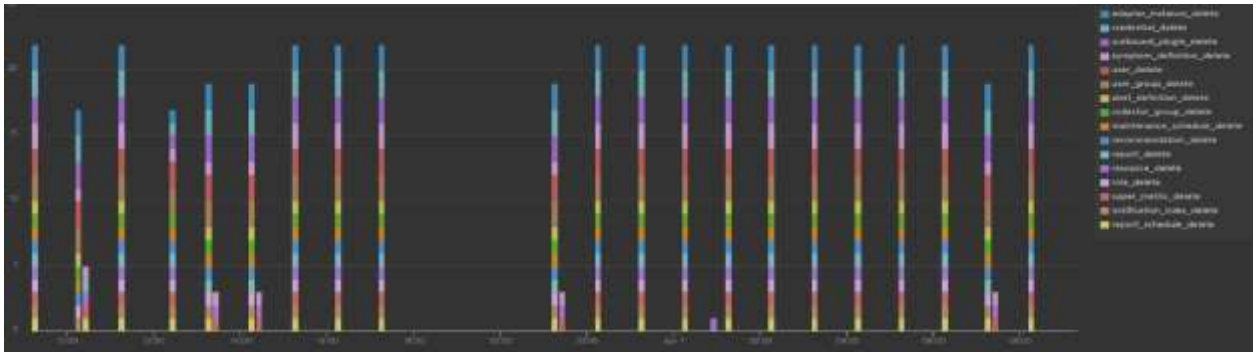


.*POLICY.*|super_metric_.*|disk_rebalance|collector_group.*|global_settings_set|GLOBAL_SETTINGS_OVERRIDE|dynamic_threshold_.*|license.*|DESCRIBE|CREDENTIAL_.*|CERTIFICATE_.*|MAINTENANCE_SCHEDULE_.*|WLP_.*|SCHEDULE_.*|RIGHTSIZING_.*|RECLAIM.*|OUTBOUND_.*|LOG_CONFIGURATION_.*|COST_.*|REFLIB_UPDATE

Delete Activity Audit



Let's dive into "Deletion activity per resource" widget by opening it in **Interactive Analytics** page. You get the same information shown on the widget, but this time you can adjust it. I've made each time block to be 10 minute instead of 1 hour so I can see the changes better.



Note that not all activities show the actual resource name being added/modified/deleted. In the following screenshot, I've highlighted in green where the affected resource name being shown, and in orange where it is not captured.



All the columns above are Log Insight field, a type of variable. Each extracted field has its own rules for extraction. Log Insight scans events and extracts fields whenever predefined patterns get matched. Let's take vmw_vr_ops_username as an example, and shows its extraction formula.





All the VMware extracted fields are prefixed with vmw_ followed by the product name.

We're now ready to evaluate the filters used in the preceding chart. It requires three filters working together, meaning they all must be true. It's an AND operator, not an OR operator.



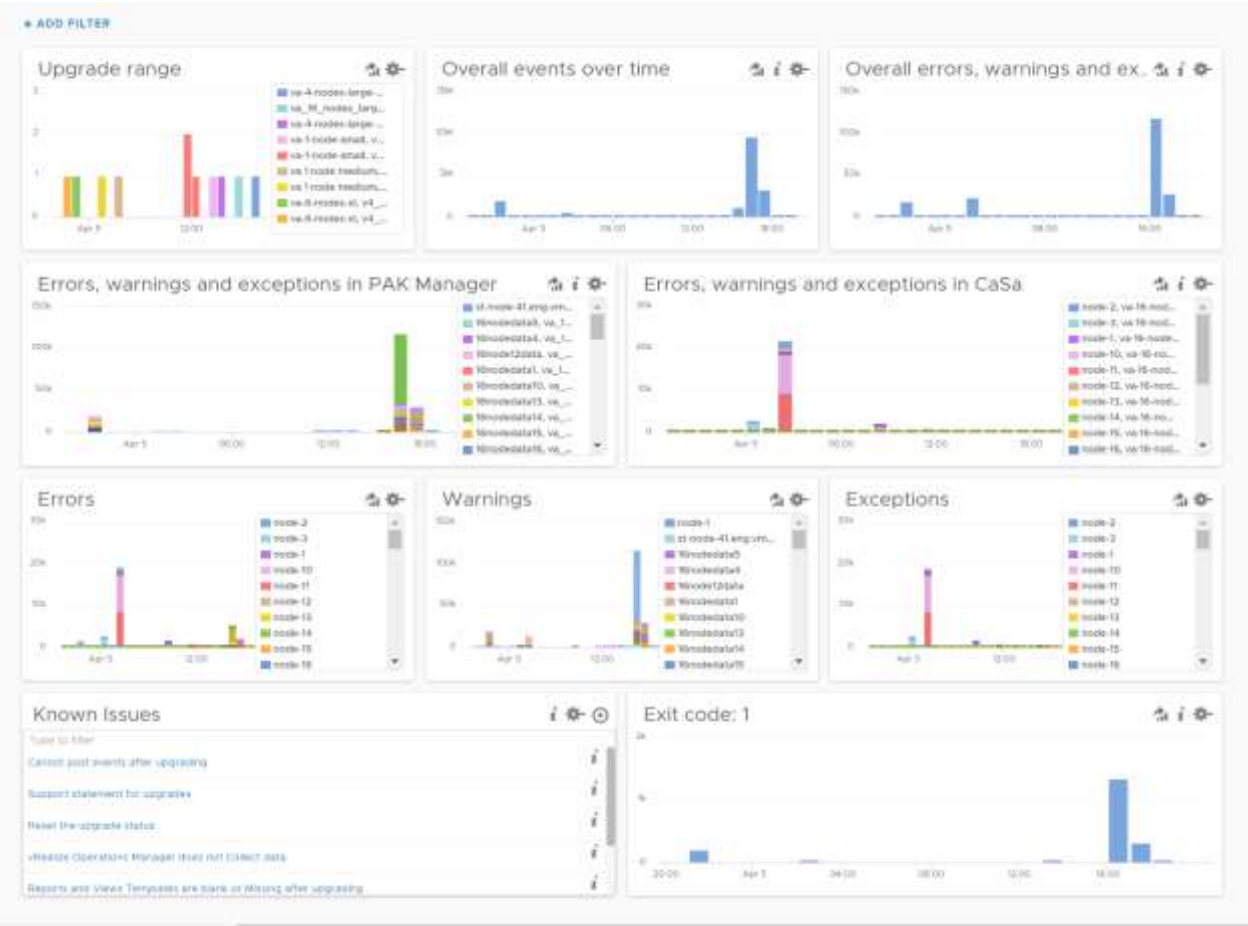
The first filter uses vmw_vr_ops_category field to filter out the events which were generated as a result to some kind of deletion.

The second filter uses filepath filter, a special system wide filter, to track the name and the path of the file from where events are collected. In Aria Operations, all audit related events are collected from analytics.audit log files.

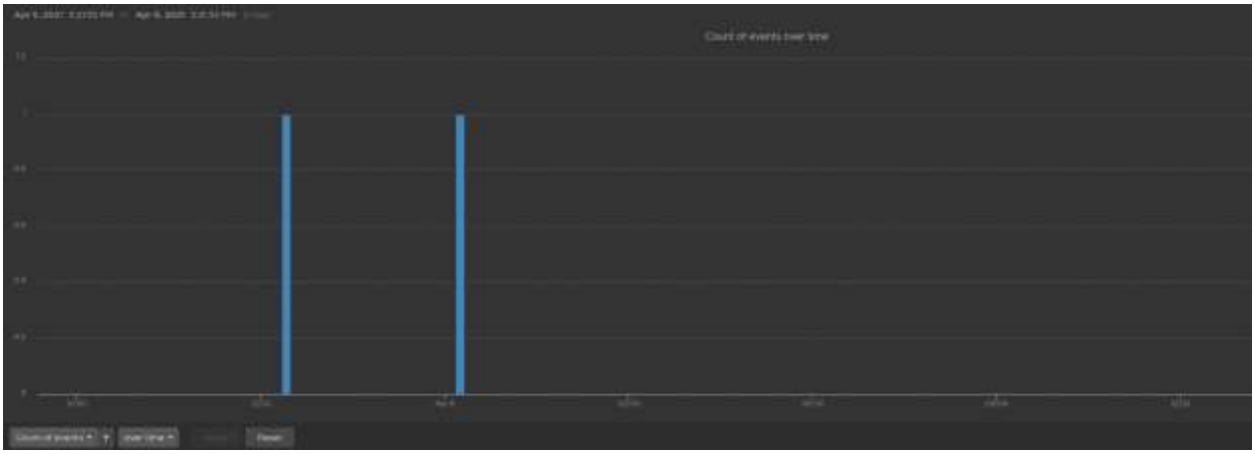
The last filter uses vmw_vr_ops_username field to exclude logs generated by service and system users. We have to exclude automation admin, maintenance admin, migration admin and system as they are not accessible by users.

Upgrade Monitoring

You can monitor the above stages as they progress via Log Insight. Yup, pretty much like watching a live streaming, as the logs are streamed into the Log Insight dashboard. The dashboard sports 9 widgets arranged in 4 rows.

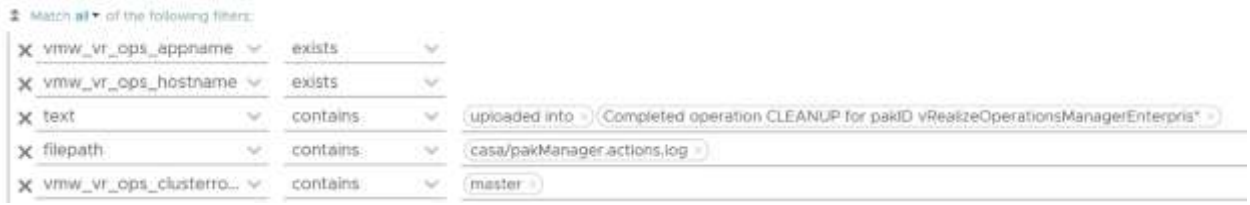


The "Upgrade Range" widget shows when the upgrade started and when it completed. It covers the time range of the upgrade process. If the process was successful, you'll see two columns, one marking the start and one marking the end, as shown in the following.



This widget is actually capable of monitoring multiple upgrades running in parallel. You can filter to only show the environment or nodes you are going to monitor by specifying their values in the fields vmw_vr_ops_clustername, vmw_vr_ops_hostname, and vmw_vr_ops_nodename.

The actual query to produce the chart is this.



When you see “uploaded into reserved” that means the upgrade process has started. When you see “Completed operation CLEANUP for pakID” that means the upgrade process for that node has been completed successfully. The following shows examples of actual messages you should expect to see.



The “Overall events over time” widget is showing the proportion of all logs generated during the upgrade.

The “Overall errors, warnings and exceptions” widget is showing the proportion of all logs with errors, warnings and exceptions generated during the upgrade.

The second row of the dashboard shows 2 widgets. They cover the main two services responsible for upgrade: PAK Manager and CaSa. The widget monitoring errors, warnings and exceptions generated from that services.

The query is as the following:



The third row of the dashboard covers errors, warning and exceptions separately. Ensure that the spikes here are not unusual high.

An example of the errors during the upgrade:



Some examples of warnings during the upgrade:



An example of the exceptions:



Note that the errors, warnings and exceptions may be not critical in this case and upgrade process may not be affected.

The last row contains the list of queries which are based on known Knowledge Base articles. Use it to check if a particular KB article is relevant in your environment. The last widget covers all log entries with **exit code: 1**. This is typically the reason for the failure of the upgrade process.

Reports

Part 2 Chapter 10

How can report, which is static, complement dashboard?

Report | Dashboard

The nature of report means it is better than dashboards in certain situation or use case, such as:

Offline	A typical example is the user is travelling, such as in a plane. If the user is not offline, but has limited network bandwidth, a text-based AI-powered chat could be better than report.
	Typical audience: Senior IT Leaders.
No access	The users don't have network access or security clearance to login to VCF Operations.
	Typical audience: external users who cannot reach VCF Operations
Casual User	Both PDF files and CSV files require less training to users than a system. There is no login required, and usage is generally much simpler.
Export	Export into spreadsheet for further analyzis & reporting. Typically, the Finance team or Audit Team want the data as part of their spreadsheet report or PowerPoint presentation. If it makes sense, educate and empower users on using a dashboard instead. They will need to know how to filter a table and do data transformation.
Legal requirement	Legal requirement may dictate that the information is pushed to the user. Politically, in some situations you can say you've done your job when the report is delivered to the user inbox, regardless whether the user actually reads it or not.
	Typical audience: different legal entity. Paying customer.
Printed document	In some meetings, stack of printed documents can complement laptop as it's easier to pass around.
	In some situations, government policy may require the report to be signed.
	Limitation: you need to manually design the widget size to fit the paper.
Historical copies	You need to keep an external copy of the data. This is applicable for functions such as capacity management and compliance management. If you want to compare at fixed and regular intervals, such as every week from Monday 00:00 hours to Sunday 00:00 hours, then it's more convenient via report. Comparing a specific date (e.g. each month of the year) is also easier on a report. This is best done via a simple CSV file as files can be compared in a spreadsheet if they have identical layout. Since many historical data is stored, be specific on the information as changing the format makes comparison difficult.
Attachment	A file can easily be attached to email or other system.

Suitability

Report is not suitable for:

- Daily consumption.
  If you need the information daily, you're better off using interactive dashboard. You get richer experience this way. Even if you login once a week, consider using personalized dashboard with rich interactivity. It's way more adaptive to your needs, which can have variations over time.
- Details.
  Vast amount of data that requires analysis is better served via a set of interactive dashboards. The exception is you need free style analysis on spreadsheet.
- Dynamic time period.
  Unlike dashboard, in which users can change the date and time being covered, report is fixed. To change the coverage from say 1 day to 1 week, you need to have design access and change of each widget one by one.

Because of the limitations above, report is not where you do management. However, it's great for planning and overall review, especially long-term planning where you don't want to be distracted with details.

Unlike dashboard, report has 2 models of consumption:

- To be read.
- To be explained.

If a report is to be read, then it must be self-explanatory.

If a report is to be presented and explained, keep the content for strategic conversations. The inability to drill down help your goal to focus on the big picture. If what you need is details, you're better off using dashboard. In this way, you can drill down as needed to answer questions from your senior management.

Inventory	Focus on the big picture.
Availability	The goal is just to capture overall availability. Focus on the big picture. As a result, individual ESXi host availability is not suitable for reports.
Performance	Not for monitoring or troubleshooting. The goal is just to report (proof) that performance was good. So capture the overall performance. Focus on the big picture. As a result, individual VM level performance is not suitable for reports.
Compliance	Not for remediation. The goal is simply reporting for compliance purpose.
Capacity	The goal is for planning, not managing the capacity. Focus on the big picture. Look ahead, preferably 1 quarter ahead.
Configuration	Use it as a backdrop on why you have configuration issues, and what can be done to address them.
Security	Not for remediation. The goal is simply reporting.

Target Audience

Report should be purpose-built, designed specifically for a persona and use case. A persona may need multiple reports, with different frequency and contents. Avoid lumping these different reports into a single report. Find out the actual, end persona.

In the product, it's impossible to cater for everyone in all customers, so we're providing the building block approach. You may need to combine the reports, and perhaps even customize the input and output. As a convenience, we provide some examples out of the box.

Once you know the actual human being reading the report, schedule a meet to understand the need well.

Purpose	Is it for action or just FYI? Senior management may just want to receive, but not actually open the report on time and everytime. The report is just there for "just in case" additional details are needed.
Next Step	If it's for action, it must not be time sensitive (within the day). For daily information, the users are better off served by dashboard. Exception is they need to export to spreadsheet for further integration
Timeline	How frequent do they need it? The longer the timeline, the less fresh the information becomes, so the focus needs to shift towards the bigger picture. The longer the timeline, the higher the chance the recipient forgets the previous value, so you might have to repeat and compare current and previous.
	Past focus or future focus? This requires brainstorming as users typically ask for historical data, while their intent is for future.
	The main limitation in report is the time period needs to be hardcoded in the widget. Unlike dashboard, report does not have date control at the report level.

The most common persona for report is IT senior management. As their requirements are comprehensive, they are documented separately in the Role-based Contents chapter.

Auditor

Majority of their needs are better addressed by dashboard.

This role is interested in knowing that the environment is kept secured. They also want to know if there are unauthorised changes, especially by privileged users (read: system administrators, yes, that means you and your team).

Report Frequency: Weekly – Monthly.

Capacity Planning Team

Majority of their needs are better addressed by dashboard.

There are several use cases where report is better than dashboard. They may need to keep historical data for years, to enable more accurate long-term planning. Having the record of each data center, cluster and datastore can be handy. This should be provided as simple CSV files.

Report Frequency: Weekly – Monthly.

Finance Department

They typically prefer information in a spreadsheet, so they can use it as part of their financial analyzis and reporting.

As they care about asset, only information about hardware is relevant. For example, a list of ESXi host with their purchase date, end of warranty, and cost.

Report Frequency: Weekly – Quarterly.

Tenant

In some organisations, it can be effective to send regular emails on reclamation to the respective owner. While the owner has access to a self-service portal, they may react differently as the email is cc-ed to their senior leadership and Office of the CFO.

It's also useful and make actionable request highly visible to the broader team. The email can consist of Top 10 VMs, with the VM owner name, with the largest cost savings potential.

Report Frequency: Weekly – Monthly.

Consultant

This is a special kind of report for 2 reasons:

- The nature is comprehensive, covering all aspects of operations management.
- Typically, the consultant conducts an assessment and need to produce report. This assessment is done rarely, such as once a year.

Technical Considerations

Once the target audience and purpose are defined, the report design consideration comes into play:

Scope	Is it the whole environment or a subset? The amount of data can impact your choice of visualization. For example, while a table can handle 100 rows well, 10K rows results in endless pages. In a large environment with thousands of vSphere Clusters and datastores, the data at world level may be too high. The scope impacts the type of information shown. One way to address is to focus on the exception or important. For example, when listing VM to downsize, only focus on large VMs.
	The above translates into what objects to use. In general, we use vSphere World, vCenter and vCenter DC as the scope. This means the content is from lower-level objects.
Layout	For consumption on a laptop or desktop, a landscape will suit user better. This is why the default report is all landscape.
	For consumption on a mobile phone, a portrait fits how a user holds her phone. For consumption on a tablet, it depends on the size of the tablet.
Color	A report that is color coded will be easier to read.
Data Recency	Present value vs past value. There are use cases for both.
	The present data point is relevant when you need to know the current or latest situation.
	The historical data points are more relevant when you need to analyse over a period. Since there are more than 1 data points, you need to pick 1 to represent them. Typically you take the maximum, average, or percentile. If you need to present a pattern or trend, you plot a line chart.

Output

There are 2 types of outputs:

- PDF
- CSV

CSV	Only the view list is suitable. While Summary List can produce CSV output, the content lacks details for further analysis. So use it in PDF output, as you likely need another widget to show detail
	The page orientation is also not relevant. You can go as deep and as wide as you want.
	1 report per object. CSV file cannot support tab. 1 file only has 1 table or list.
	Do not color code as it will be ignored by the plain CSV format.

	Do not group the output, as it's better to use a spreadsheet to group.
	While you can add a summary row at the end, it's more flexible to use spreadsheets as you can have different formula for each column.
	There is a limit of 50 columns. For object with many metrics and properties, logically separate them.
	For utilization and contention, show at 95th percentile. Maximum will likely show outlier while average is "too late" as the goal is to serve everyone almost all the time.
	For event, sum the numbers. For example, the number of Guest OS rebooted.
PDF	Design the report to use landscape orientation. This makes it suitable for laptop, tablet and printed paper.
	Just like mobile phone friendly app is best designed separately, the same goes for report. The small screen requires a purpose-built design.

Context

Put a context. It helps when the CSV file contains thousands of rows. For example, for VM, add the following columns. I put them at the end as the first few columns should be more important information and used for sorting.

Column 1	The parent vCenter Data Center. This is the first column, so the VM is "grouped" by DC.
Column 2	The parent vSphere Cluster. This is the second column, so the VM is "grouped" by cluster.
Column 3	The parent folder. This is the third column, so the VM is "grouped" by cluster

Widgets

Only a handful of widgets are suitable for PDF printed report, because reports have no interaction and fixed height.

A report should flow from summary to details.

- Start with headline numbers. This shows large, eye-catching number.
- Break down the headline numbers by groups. The focus is on the groups, especially their relative size or value.
- Show trend over time, if the pattern is important.
- Last is show the individual members. To focus on the important ones, limit to say top 10.

The following shows the various section and the widgets suitable for it.

By being aware of the limitations of each widget in PDF layout format, you avoid redoing your report. Measure twice, cut once.

Purpose	Description
Headline	Large font, eye catching. 3 – 5 numbers shown on portrait paper visualize the summary well. If you put too many it will look crammed. If you put 1 – 2 numbers it will look incomplete and sparse.
	Widget: Score Board. Limitation: score board can only show present value, hence not suitable for performance.
Summary	A simple table. Use when you need to summarize values of many members at the parent level. It complements the score board widget as you can do computation.
	Widget: Summary View. Limitation: The filter applies to all rows in the table, despite the fact each row has different metrics. There is no count function. Only sum, maximum, minimum, average and standard deviation
Grouping	Show objects by group. Focus is on the group, not the individual members.
	Use Pie Chart when you don't know how many slices as it's dynamic. Value wise, pie chart is good for relative distribution. Its focus on relative means it's less suitable for performance management. Limitation: Legend gets truncated easily unless it's short.
	Use Bar Chart when you need to control the range of each slice, or focus on the quantity of each slice. Value wise, bar chart is good for absolute distribution. Limitation: Need to manually set the number of buckets for best visualization
	Use Heat Map when you need to show relative size and intensity. For the color, heat map is the only widget capable of showing continuous gradient of values. Limitation: The group name is hard to read if there are too many groups
Trend	Shows trend line. Focus on pattern over time.
	Health Chart. Use when you can color code the value. Limitation: The shortest height is too small. Use the medium or tall option instead. It is less suitable if the number of objects is not fixed. Visually, there is no control for the y-axis minimum and maximum value.
	View Trend. When values can't be color coded, and you need multiple lines on the same y-axis for comparison. Limitation: Quite tall, taking up sizeable page. Since your report is landscape, try to put 2 side by side.
Individual	Individual means listing each object on the report, with their names.
	Top-N. This is a simplified table. Focus on the value of 1 metric. Best when the relative values among them matter and you can color code. Use the Percentile function as the default roll up is average Limitation: Set the number of rows, so it's only suitable for items when this number can be capped without losing meaning
	View List. This is a multi-column table. Focus on the object name. View can show before vs after. The limitation is the time setting is hardcoded in the view. Users without designer access cannot modify it.

Functional Considerations

Let's summarise the applicability of present value vs past value.

	Present Data	Past Data
Inventory	Yes	Yes
Availability	No	Yes
Performance	No	Yes
Compliance	Yes	Yes
Capacity	Yes	Yes
Configuration	Yes	Yes
Cost	Yes	Yes

For historical data, you need to hardcode the time period on each widget.

Since Performance and Availability requires data across time, not all widgets can be used. Specifically, the following widgets are not suitable:

- Scoreboard
- Heat Map

Synergy

The reports were designed together as one suite. As a result, none of them is independent as we minimize overlapping content. Run the reports as a set.

Let's take an example on how they synergise. Inventory, Configuration, and Capacity synergize to provide complete coverage.

- Configuration is about settings of that object.
- Capacity is about size. It also tends to use line chart, so it can show changes or pattern over time.
- Inventory is about count. It focuses on the type of sizes, not the actual size. It's showing the most common ones. The goal is to know the variation.

Let's take VM configured size for example.

Inventory	It reports how many sizes you have, and which ones are the most popular size. Are the common sizes larger or smaller than your expectation? It's using a pie chart as it's about relative distribution and the number of the variations.
Configuration	It reports the sizes in order, so you focus on the monster VMs. It's using bar chart, with buckets filtering out the small ones, focusing on the large configuration.
Capacity	It reports the undersized and oversized VMs.

	It's using bar chart, but the bucket shows the range of undersized – rightsized – oversized. For details, you can use a table to show the top 10 VMs to be acted upon.

Let's take ESXi host for example:

Inventory	Relative distribution of ESXi by vendors, server type (blade, rack mount), server generation.
Configuration	Absolute distribution of ESXi by version, by CPU model and generation.
Capacity	Absolute distribution of ESXi by CPU Size, by memory size.

Let's take vSphere Cluster for example:

Inventory	Bar chart of clusters, grouped by number of VM. You determine the value of each bar.
Configuration	Bar chart of clusters, grouped by number of ESXi host The number of VM is not part of a cluster configuration.
Capacity	Bar chart of clusters, grouped by size. For example, you group them into 5 buckets of size based on total memory capacity.

Overall Design

Executive Summary	1 summary report. Designed to be presented by someone familiar with the environment, not self-reading by the IT leaders. The content focuses on the big picture. It is to facilitate discussion on major initiative. It does not contain detail.
Operations Management	1 set of reports, covering inventory, capacity, performance and configuration. Run them as a set.
Spreadsheet Export	1 CSV export per vSphere object type. Run as needed.

Executive Report

They may not be familiar with the environment, so the report starts with key inventory.

It avoids the use of technical jargon, unless the generic term causes confusion.

It does not have details. As a result, it does not cover individual items. This means it does not list the names of any VM, clusters, datastore, etc.

Customisation tips:

- If the context allows, put items you want their approval at the top. For example, you need to get budget for technology refresh on ageing hardware.

Operations Management

For the Operations Management, the 4 reports have similar table of content.

- Consumer: VM
- Provider: Compute
- Provider: Storage
- Provider: Network

Depending on the number of pages, the above maybe further split or combined.

Based on the above, we end up with the following "section" across 4 reports. Each section, which is technically a dashboard, may span multiple pages.

Report	Consumer	Provider
Inventory	VM Inventory	vSphere Inventory
Performance	VM Performance	vSphere Compute Performance
		vSphere Storage Performance
Capacity	VM Capacity	vSphere Compute Capacity
	Reclaimable Capacity	vSphere Storage Capacity
Configuration	Guest OS Configuration	ESXi Hardware Configuration
	VM Configuration	ESXi Software Configuration
	VM Size Distribution	vSphere Cluster Configuration
		vSphere Network Configuration

END OF PART 2

Part 3 covers metrics. It has been made as standalone book as Microsoft Word became slow at 1K page.