How To Guide

NetApp Cloud Insights: A New Way to Monitor Your Cloud Infrastructure

Innovate faster with data insights across your application infrastructure stack

Richard Treadway, NetApp
February 2019 | WP-7291

Abstract

This How To Guide provides an introduction into NetApp® Cloud Insights. In this document you will learn how to set up and begin using Cloud Insights to monitor, troubleshoot and optimize your entire infrastructure.

NetApp®

TABLE OF CONTENTS

1 The Cloud Operations Imperative

Your infrastructure is increasingly complex, and you are being asked to do more with fewer resources. The cloud has made it easier to deploy quickly but harder to control costs and optimize use. You are responsible for ensuring performance and preventing failures. You are the first in the line of fire if something goes wrong.

Increasingly, your application development and deployment teams are taking advantage of new innovative cloud technologies that speed time to market and enable higher levels of customer intimacy. But corresponding advancements in infrastructure monitoring and optimization tools have not kept pace. Technologies such as microservices and distributed tracing require new monitoring tools. This gap between implementation technologies and monitoring tools is leaving the Ops side of your DevOps projects exposed to increased risk of failures and runaway costs.

As a site reliability engineer, you are part of a DevOps team that's chartered to build new cloud applications. You are the single point of arbitration between Developer and Operations teams. Your goal is to improve operational efficiency and performance. You identify and manage risks in the development and deployment of these new applications. Failure of these applications will adversely affect your company's business.

You need real-time service-level indicators (SLIs) so that you can be assured that your systems are meeting your service-level objectives (SLOs) and service-level agreements (SLAs). Your tools need to be able to monitor both cloud and on-premises infrastructure. When you do experience infrastructure outages, you need to find the problem fast. The bottom line is that everyone points to you if something goes wrong.

You need a simple, easy-to-use, cloud-based infrastructure monitoring tool that can reduce troubleshooting time, accurately predict performance needs, and help control costs. We designed NetApp® Cloud Insights to meet those needs.

Cloud Insights is a SaaS monitoring tool that gives you actionable knowledge of your infrastructure. Because it's hosted in the cloud, Cloud Insights is easy to use. You'll be up and running fast, receiving real-time data visualization of the availability, performance, and usage of your entire IT infrastructure.

This *How To Guide* explores the challenges of monitoring cloud infrastructure and how Cloud Insights can help you save time and money.

We'll show step by step how to:

- Collect your infrastructure data
- Monitor and watch for problems
- Troubleshoot to find and resolve issues
- Optimize costs and run more efficiently

2 Key Features of Cloud Monitoring Tools

Cloud applications are written to consider infrastructure as code, which essentially means that infrastructure is provisioned and deprovisioned dynamically through APIs. It also means that applications must know about their state as they run to be able to make real-time adjustments. Most cloud applications make extensive use of cloud-based provisioning and control services like Puppet, Chef, containers, and Kubernetes, which means that they can expand and contract the infrastructure they use at the speed and scale of the cloud.

This speed and scale require that cloud monitoring tools can capture data in milliseconds, not minutes or hours, and they need to understand not just the state of each component but the relationships between components.

Monitoring is the forensic tool that can spot a transient spike in latency or a fleeting failure that is visible for only an instant. Problems left uncorrected in a self-healing dynamic infrastructure lead to overprovisioning, higher costs, and eventually customer impacts that affect your business.

3 NetApp Cloud Insights

NetApp Cloud Insights is designed specifically for today's cloud-based infrastructure and deployment technologies, offering advanced analytics on the connections among resources in the environment. You'll have real-time data visualization of the topology, availability, performance, and usage of your entire infrastructure, including both cloud and on-premises multivendor resources, as well as support for NetApp Cloud Volumes, NetApp HCI, and AFF.

Cloud Insights makes it possible to connect traditional service and software-defined infrastructure layers, giving you visibility into both your traditional and modern application architectures. Cloud Insights quickly inventories your resources, figures out their interdependencies, and assembles a topology of your environment, giving you end-to-end visibility into what resources are supporting which applications.

Unlike today's monitoring tools, Cloud Insights is designed to handle the transient nature of modern cloud infrastructure and its connectivity to sets of services, giving you a complete understanding of demand, latency, errors, and saturation points of all your services.

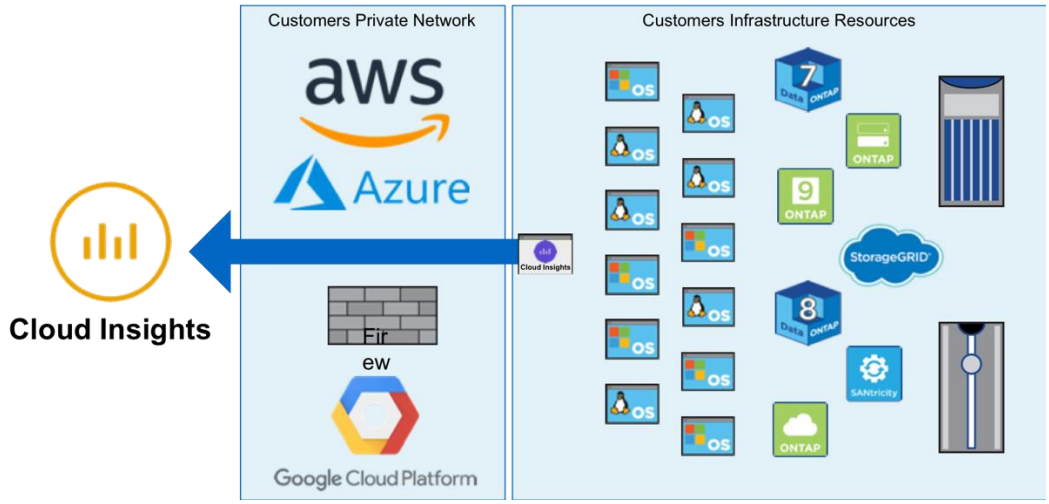
With Cloud Insights, you'll be able to:

- **Improve customer satisfaction** by preventing up to 80% of cloud infrastructure issues before they impact end users. You'll be better equipped to meet customers' demands by proactive monitoring of your complete environment. You can visualize your topology with automated discovery to see end-to-end service paths. You'll know exactly how your systems are performing and how they're being used. When a performance-level violation is detected, you get the necessary data to quickly determine the root cause of the violation. With that analysis, you can be confident that you're keeping up with customer demand.
- **Proactively prevent failures** and reduce mean time to resolution (MTTR) by up to 90%. With advanced analytics, you can identify which resources are greedy and degraded. You can use correlation analysis to correlate services to modern transient infrastructure to help identify the root cause of a problem faster. You can also set up advanced conditional alerts, which save you time tracking down false positives. Finally, predictive analytics based on machine-learning technology alert you to potential issues before they become major problems.
- **Optimize and reduce cloud infrastructure costs** by an average of 33%. With resources in play from your on-premises data centers to multiple public clouds, it's hard to know what's really in use and what can be freed up. You need to be able to identify unused or abandoned resources. Knowing the performance requirements of your applications lets you identify when they might be overprovisioned. With that knowledge, you can re-provision applications to less costly infrastructure.

4 Step 1: Collect Your Data

You can access all NetApp Cloud Data Services from NetApp Cloud Central. Once you've logged in and launched Cloud Insights, the first step is to discover the infrastructure that you want to monitor. To do this you need to install an acquisition unit. Figure 1 shows the role of the acquisition unit in your infrastructure topology.

Figure 1) Acquisition unit and infrastructure topology.

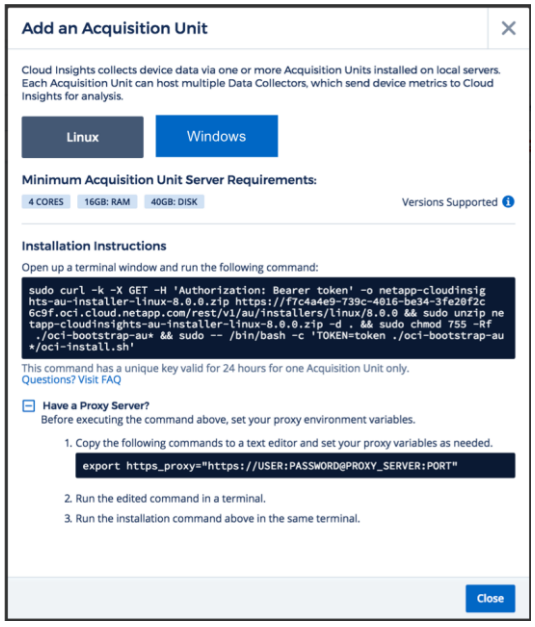


The acquisition unit is deployed in your infrastructure, and therefore it sits behind your firewall, virtual private cloud, or VNet. The acquisition unit is installed on a virtual machine with access through the firewall, and it has the ability to push data out and back into the cloud.

Cloud Insights supports acquisition units for both Windows and Linux. The acquisition unit requires a very lightweight virtual machine with 4 cores, 16GB of RAM, and 4GB of disk.

For the Linux version, all you need to do is copy the supplied text to the console, as shown in Figure 2. If you have a proxy server, you need to cut and paste the additional line shown to set the proxy variables.

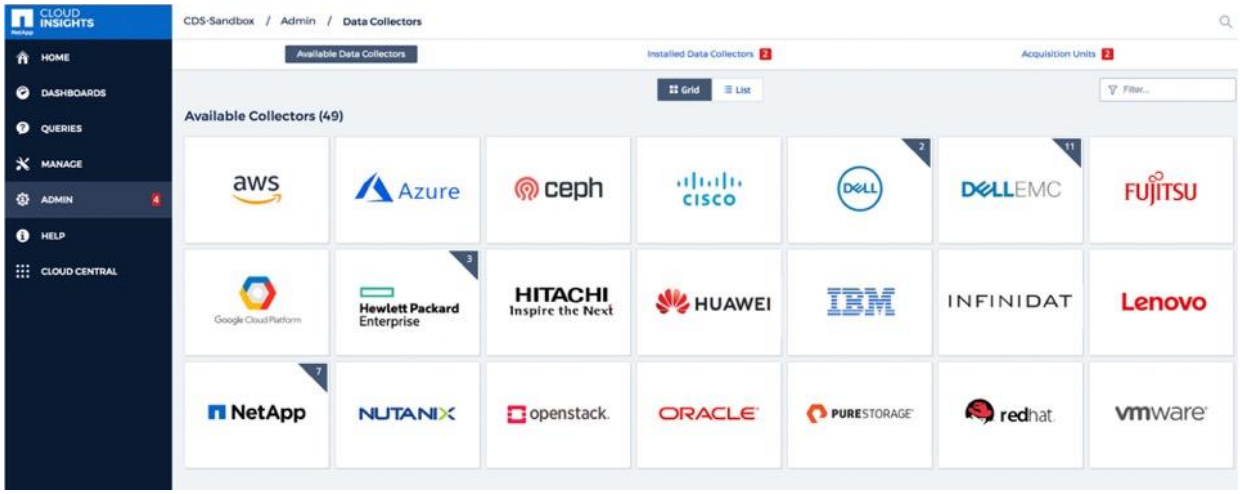
Figure 2) Setting up the acquisition unit.



Windows follows a similar process in which you download the Microsoft Installer and run it to set up the acquisition unit.

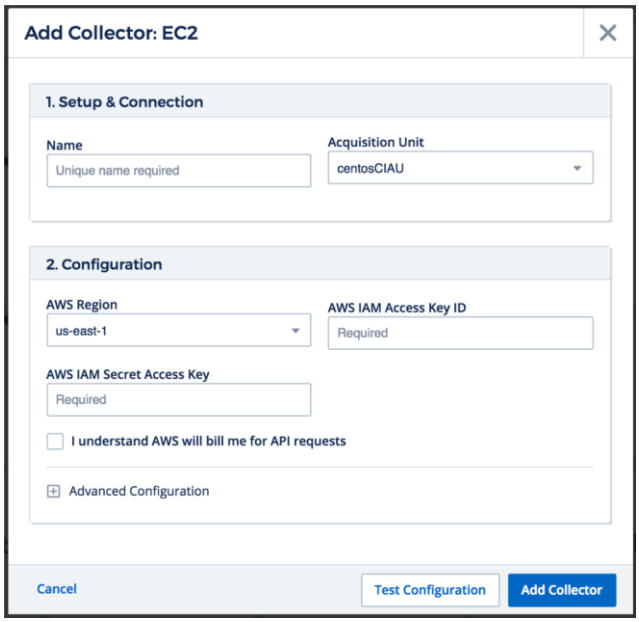
Cloud Insights is not just for monitoring NetApp systems. It provides full coverage of public cloud providers, AWS, Azure, and Google Cloud Platform (GPC), as well as third-party devices from Dell/EMC, Fujitsu, IBM, Hitachi, and others. Figure 3 shows the breadth of options supported.

Figure 3) Available collectors.



It's easy to set up a collector. You simply give the collector a name in the environment and choose the acquisition unit that you just deployed. You then supply additional information specific to that collector. Figure 4 shows the specifications for an AWS collector. In this case you enter the region, the access key ID, and the secret access key. The acquisition unit is a read-only account for Cloud Insights to pull data from the resource. That data includes configuration details, how the resource is connected to other resources, vendor, capacity, memory utilization, and so on. Cloud Insights uses that data to perform the critical analysis necessary to monitor, troubleshoot, and optimize the complete environment.

Figure 4) AWS Collector dialog box.



5 Step 2: Monitor and Watch for Issues

Once data is flowing into the system, you can begin watching for issues. Cloud Insights comes with a gallery of dashboards to get you started. The gallery you see initially is based on the collectors you have enabled. For example, if you want to discover your AWS environment, you are presented with a list of reports that you can select for your personalized dashboard. Your dashboards give you visibility into your data so you can start looking for issues.

When monitoring, you are looking for four key SLI signals.

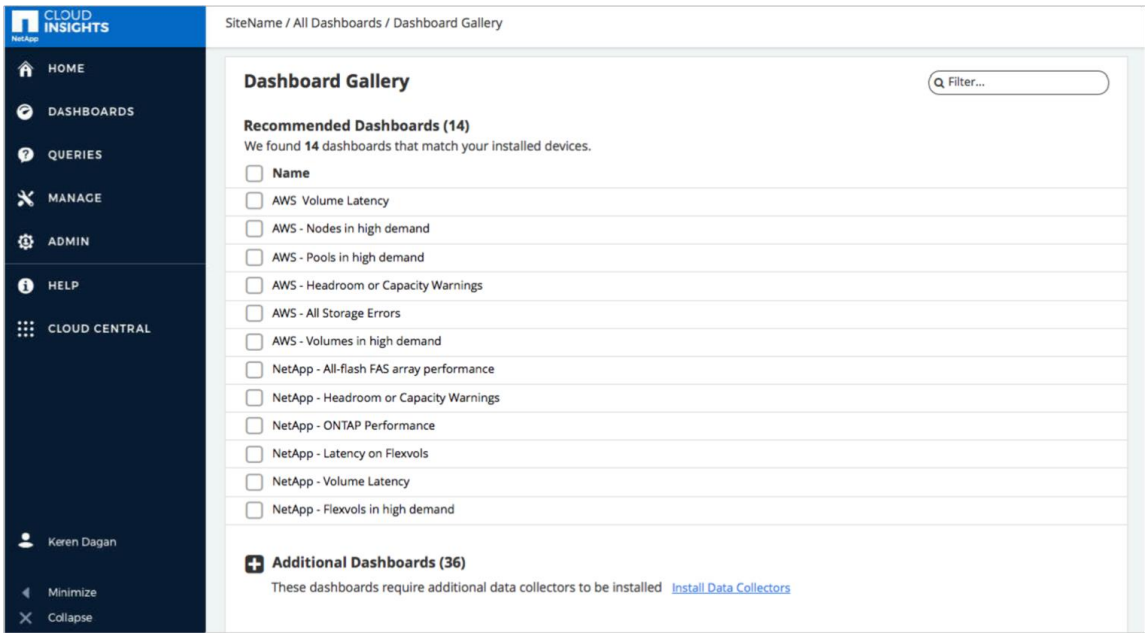
- **Latency.** When latency spikes, users are affected and they start submitting trouble tickets, which leads to remediation. You don't want that. Instead, you want to discover and resolve latency issues before they impact end users.
- **Saturation.** Latency issues are often caused by saturation of devices supporting the workload. Are resources exceeding their performance capabilities? Memory utilization? CPU utilization? Disk capacity? That's when the saturation starts to affect latency. Understanding what is causing saturation is a key step in discovering the root cause.
- **Traffic.** Saturation is usually triggered by an increase in traffic. An increase in I/O or megabytes per second causes specific resources to exceed saturation thresholds. As an operations person, you need to understand latency, saturation, and traffic, because the sooner you know that SLI thresholds are exceeded the sooner you can take remedial steps to prevent trouble tickets and user complaints.
- **Errors.** Finally, you need to know if there are errors in the environment, and you need to be able to discover the root cause and begin remediation as soon as possible.

Dashboards

Cloud Insights creates a default set of dashboards based on the data collectors you have enabled. Each dashboard is designed to answer your specific questions about monitoring your infrastructure.

- What systems are experiencing high latency?
- Where have SLOs been exceeded and are causing errors?
- What VMs are inactive or powered off?

Figure 5) Example dashboard gallery.

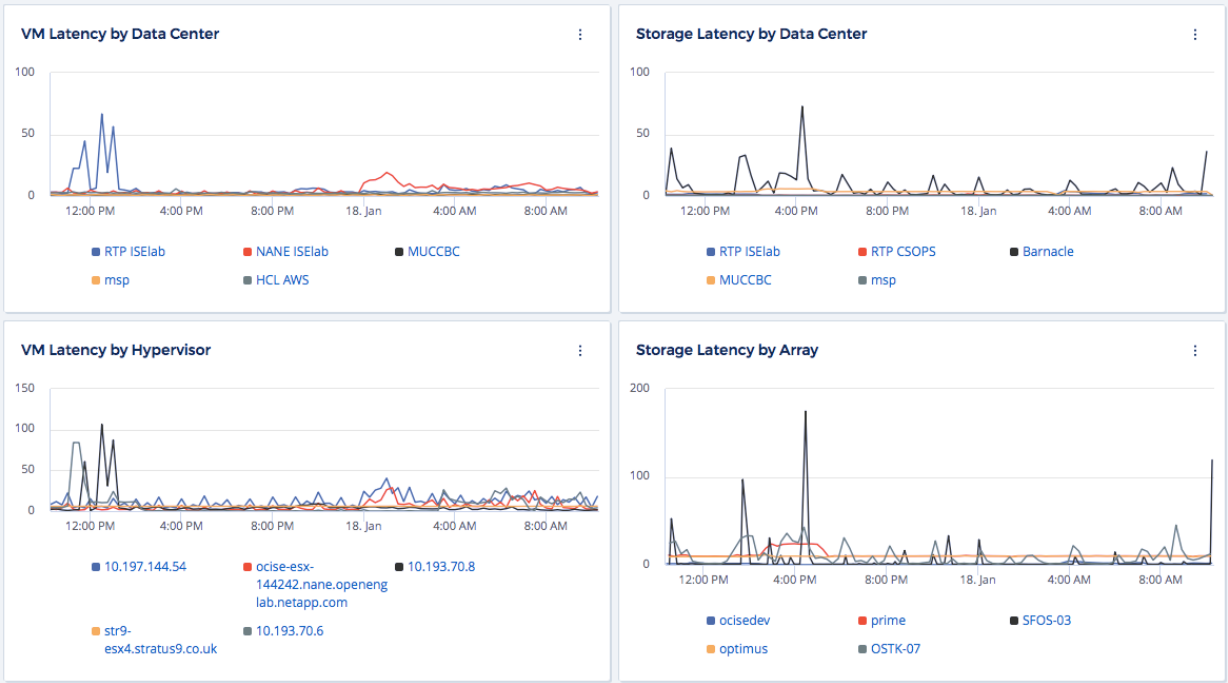


Let's examine a typical monitoring scenario to query which systems are experiencing high latency that, if not corrected, might lead to degraded performance and end-user impact. Figure 6 shows a dashboard that reports on all systems across all infrastructure, sorted by latency. In this case, there are multiple data centers with resources in AWS, Azure, and GCP, as well as on-premises systems.

From this dashboard you can easily click down to details in a data center to a specific resource to see what is creating the high latency. The Cloud Insights correlation engine shows which resources are connected to a spike in latency and the probability that they are affected.

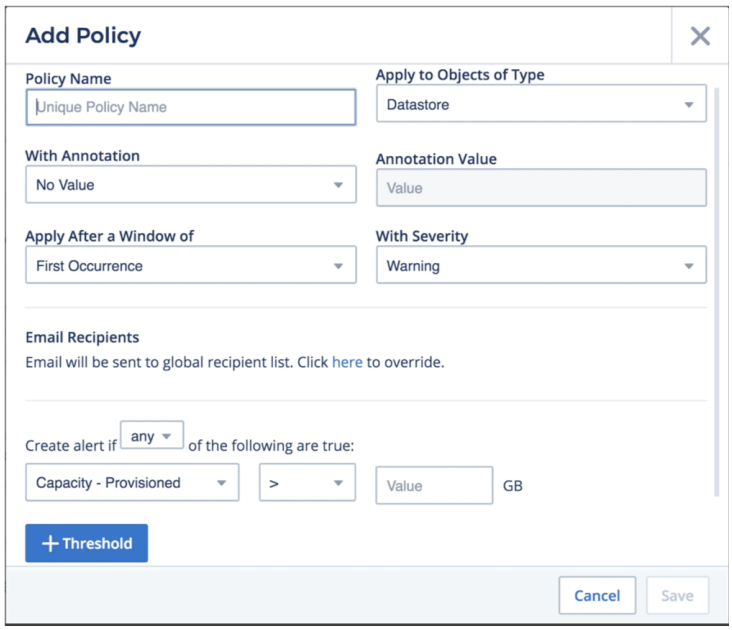
When the affected resource is found, the topology view shows how it is connected to other resources.

Figure 6) Latency by data center dashboard.



With Cloud Insights you can set policies to create alerts if a resource exceeds a specific SLI. You can proactively monitor the environment to uncover problems before they impact your operation SLAs. As Figure 7 shows, it's easy to create an alert. In the Add Policy dialog box, you specify the alert name and what you want to monitor. Options are numerous and include data stores, VMDKs, hypervisors, volumes, and virtual machines. You also specify the type of alert and the set of resources to be monitored. Policies specify the resource, the metric, and the associated threshold, giving you the flexibility to target the exact condition that you want to trigger an alert.

Figure 7) Add Policy dialog box.



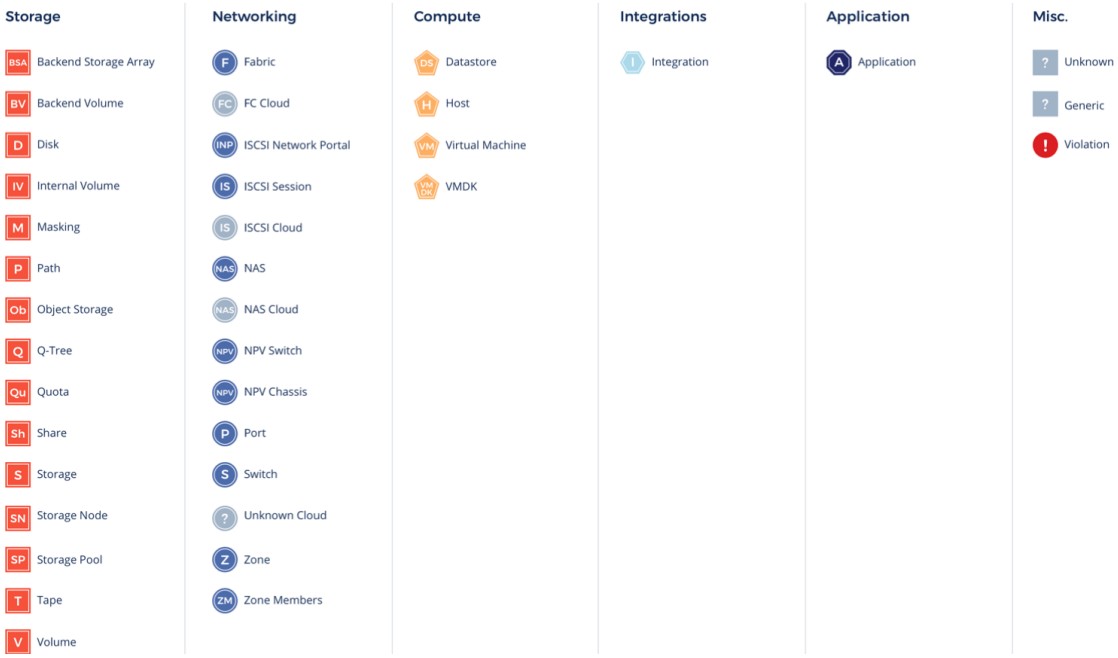
When an alert is triggered you receive an email with a link that takes you to the list of policies that have exceeded their thresholds. Figure 8 shows an example list of policy violations for FlexVol® volumes latency that have exceeded 5 minutes. From this list you can further filter violations and drill down on specific violations to investigate their root cause.

Figure 8) Detected alerts.

6 Step 3 Troubleshoot to Find Failures Quickly

To troubleshoot effectively, you need to understand how your resources are related and how they interact. Figure 9 shows the list of icons that represent elements of your infrastructure.

Figure 9) Topology icons.

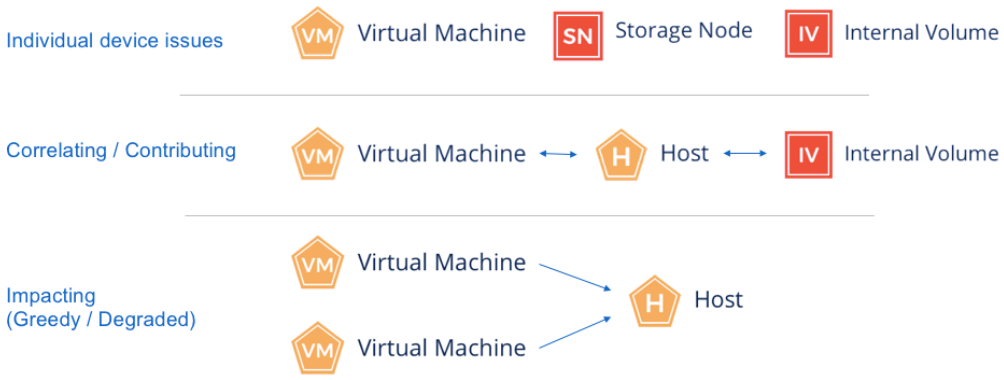


In troubleshooting your infrastructure, you look for specific scenarios that lead you to understand the root cause of the problem. Figure 10 illustrates common issues.

- **Individual device issues.** The problem could be related to a single virtual machine, storage node, or volume. In these cases, other related resources are not affected. For example, a single VM might experience latency violations not caused by other resources but by coding errors, such as a memory leak.
- **Correlating or contributing.** In this case, a problem in one resource affects other related resources. Cloud Insights maps all the resource relationships so that you can see if a resource failure is affecting other connected resources.
- **Greedy resources.** The third case in Figure 10 shows a VM that is having issues that affect another VM somewhere else in the stack. Cloud Insights understands these correlations and provides probabilities for which resources are most likely being affected by a greedy or degraded resource.

Correlating and probability detection are unique to Cloud Insights. No other tool on the market today has this capability.

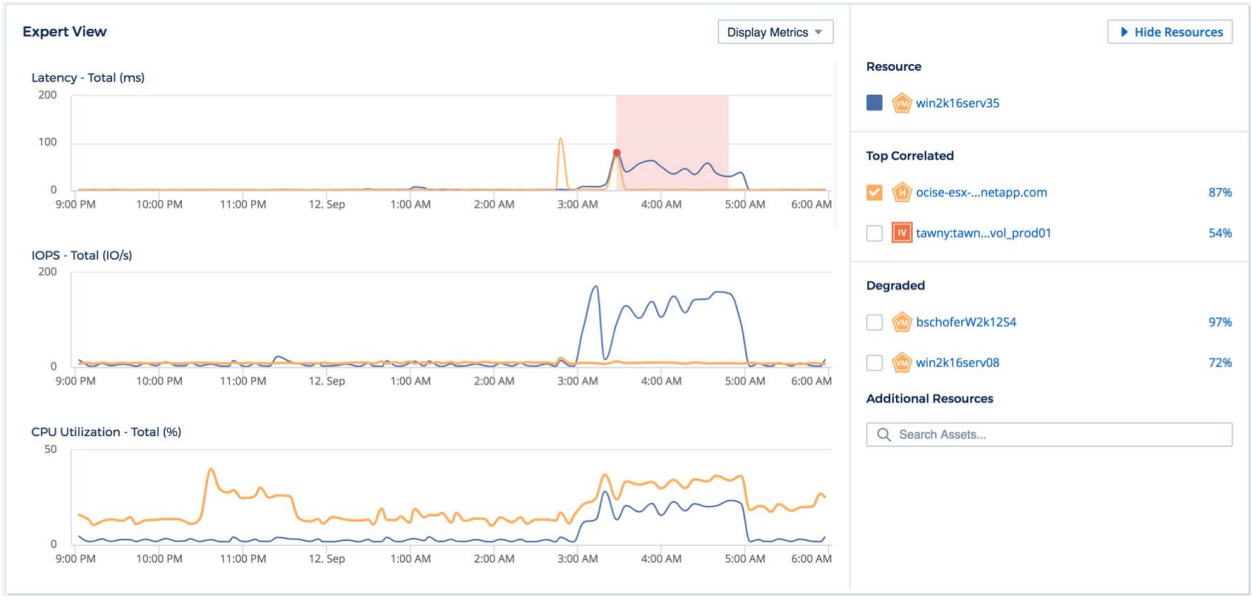
Figure 10) Possible issues in the infrastructure stack.



When an issue is detected, you follow the alerting email link and drill down to the performance violation. Figure 11 shows an example violation in which you can see that a VM's latency spiked at 4 a.m. You can also see the other correlated resources. You can display the related resources' metrics by selecting their checkboxes. In this case there is an 87% probability the host ocise-esx is affecting the VM bschoferW2k245.

Understanding these correlations helps you quickly identify exactly what is going wrong and how it's affecting other resources in your infrastructure.

Figure 11) Finding the root cause.



7 Step 4 Optimize Cost and Run Efficiently

As you monitor and troubleshoot your environment, you also want to make sure that you're not overprovisioning and spending more that you need to. Especially in the cloud, it's easy to provision resources and forget about them.

Cloud Insights offers some easy wins to quickly reduce your costs. Because Cloud Insights knows the state of all the virtual machines in the environment, it can quickly identify those that are idle and the capacity associated with them. You can apply a cost to each one to see the savings you can realize by deprovisioning them.

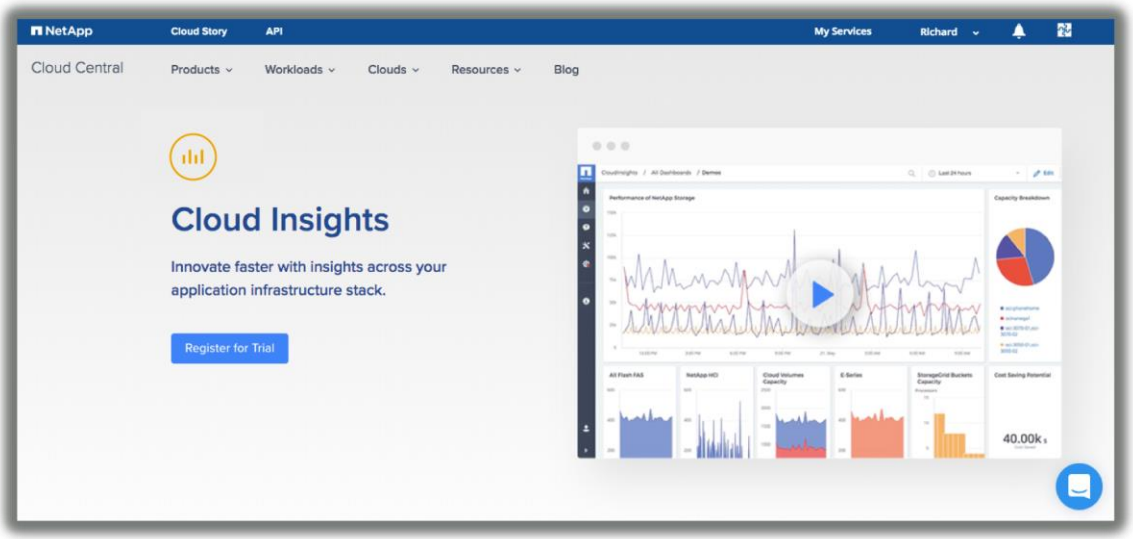
Figure 12 shows a report of all powered-off and suspended VMs and the costs they are incurring. You can drill down to each one to see who they are allocated to and how they are connected to other resources in the infrastructure. You can use this information to determine which resources can be released.

Figure 12) Finding savings.

8 Summary and Next Steps

Cloud Insights can help you monitor, troubleshoot, and optimize your infrastructure.

To experience Cloud Insights for yourself, register for a 14-day free trial. Go to NetApp Cloud Central to learn more about [NetApp Cloud Insights](#) and start your free trial.



Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.