

White Paper

# NetApp Cloud Insights

**Optimize Your IT Infrastructure** 

Joshua Moore, NetApp January 2020 | WP-7319

#### Abstract

Your IT environment is more complex and more dynamic than ever before. This makes tracking and controlling costs even more challenging. NetApp<sup>®</sup> Cloud Insights can help you achieve greater efficiency and optimization both in the cloud and on-premises. This white paper shows you how to configure Cloud Insights to identify areas of waste to reduce costs and optimize hybrid and multicloud IT operations.



#### TABLE OF CONTENTS

1	About This White Paper Series	. 3
2	The Complexity of Understanding and Controlling Infrastructure Costs	. 3
	2.1 Cloud Costs	3
	2.2 On-Premises Costs	3
	2.3 Keep It Simple: Focus on Identifying and Reducing Waste	4
3	Using Cloud Insights Queries	
	3.1 About Cloud Insights	5
4	Using Cloud Insights Dashboards	. 6
5	Incorporating Cost Information in Cloud Insights	. 7
6	Shameback Reporting	. 9
7	Summary and Next Steps	10
Wł	ere to Find Additional Information	10
Ve	rsion History	10

#### LIST OF FIGURES

Figure 1) On-premises costs depend on management, networking, and facilities costs in addition to upfront capital costs.	
Figure 2) NetApp Cloud Insights enables you to quickly identify wasted resources to optimize costs.	
Figure 3) Example ad hoc query to find large AWS compute instances with low utilization.	5
Figure 4) Example Cloud Insights dashboard showing waste in AWS.	6
Figure 5) Dashboard Gallery showing available options based on installed data collectors	7
Figure 6) Custom Azure dashboard showing cost metrics.	8
Figure 7) A widget to show Azure spending by instance type.	8
Figure 8) Example AWS dashboard showing usage by user	9

## **1** About This White Paper Series

This white paper is number four in a series about NetApp Cloud Insights, an innovative software-asservice (SaaS)-based monitoring tool that spans on-premises and cloud environments. Cloud Insights helps you monitor, troubleshoot, and optimize your hybrid cloud resources — from legacy hardware to container environments. The papers in this series cover the following topics:

- <u>A New Way to Monitor Your Cloud Infrastructure | WP-7291</u>
- Enhance Monitoring by Understanding the Relationships Between Resources | WP-7302
- <u>Accelerate Troubleshooting Across Your Hybrid Cloud | WP-7308</u>
- Optimize Your IT Infrastructure | WP-7319

This series focuses on both on-premises and cloud environments.

# 2 The Complexity of Understanding and Controlling Infrastructure Costs

Your IT infrastructure today is more complex than ever before, and this makes determining and controlling costs—both in the cloud and on-premises—more challenging.

## 2.1 Cloud Costs

If you've provisioned resources in Microsoft Azure, Google Cloud, or Amazon Web Services (AWS), you already know that there are many factors that affect cost. For a compute instance, for example, the cost might depend on the number and type of CPUs, number and type of GPUs, amount of memory, type of storage, which region it's located in, whether it's on-demand or reserved, and so on. The situation is similarly complex for storage, and these are just two of the possible cloud building blocks. Across the three public clouds, there are more than 500 services available, making it extremely difficult to figure out the cost for any single resource or service, or compare similar services across clouds in an apples-to-apples fashion.

## 2.2 On-Premises Costs

We tend to assume it's easier to determine the cost of our on-premises infrastructure. However, at least cloud costs are written down. The true cost of on-premises infrastructure includes not only the costs for servers and storage resources used (capital expenditure (capex)), it also includes management costs, associated network costs, and facilities costs (operational expenditure (opex)). (See Figure 1.) Just as in the cloud, these costs vary with location.

Figure 1) On-premises costs depend on management, networking, and facilities costs in addition to upfront capital costs.



This can make it extremely difficult in practice to know exact costs or to arrive at a per-virtual machine (VM) or per-gigabyte (GB) cost that can be compared to public cloud.

## 2.3 Keep It Simple: Focus on Identifying and Reducing Waste

A practical way to address the uncertainties in determining and optimizing costs both in the cloud and onpremises is to keep things simple. Regardless of how much you're spending on a given resource, if you use less of that resource, it will cost less. This is particularly true in the cloud where you might consume many resources on demand (Figure 2).

Figure 2) NetApp Cloud Insights enables you to quickly identify wasted resources to optimize costs.



NetApp Cloud Insights can help you achieve greater efficiency and optimization both in the cloud and onpremises. This white paper explores the optimization capabilities of Cloud Insights in more depth, showing you how to configure Cloud Insights to identify areas of waste—and the individuals responsible—and thereby reduce costs. While the examples in this paper focus on public cloud, the same approaches apply to reduce the waste of on-premises resources—and potentially deferring the need for additional data center investments.

## **3 Using Cloud Insights Queries**

Cloud Insights includes an easy-to-use but powerful query mechanism that provides a set of filters and finders that allow you to quickly zero in on sets of resources. Queries are explained in detail in an earlier white paper in this series: <u>WP-7302</u>.

For example, suppose you want to identify all the large compute instances (meaning, the most expensive ones) in AWS that are underutilized. Cloud Insights makes it simple to create a query to do this, as illustrated in Figure 3.

	NetAp	op CDS Sa / All Queri	es / New Qu	uery							① Last 24 Hours		<ul> <li>Save</li> </ul>	
MONITOR & OPTIMIZE		Virtual Machine *												
HOME	Filt	ter By Name Any	× Processors	>= 16 × CP	U Utilization - Total (%) <= 5	× Hype	rvisor OS Amazo	n 🗙 🕂						
DASHBOARDS	Qu	ery Results (43)											Actions *	1
QUERIES	0	Name 1	Capacity - Total (GB)	Capacity - Used (CB)	os	Memory (GB)	Processor	Hypervisor Name		or	Application	IOPS - Total (IO/s)	Latency - Tota (ms)	4
REPORTS										Stat e				
		c01-node-customer-1	150.00	N/A	RHEL-7.5_HVM_GA- 20180322-x86_64-1- Hourly2-GP2 x86_64	64.0	16	us-east-1a	1	On		10.06	1.39	
ADMIN 23		c01-node-customer-1	150.00	N/A	RHEL-7.5_HVM_GA- 20180322-x86_64-1- Hourly2-GP2_x86_64	64.0	16	us-east-1b	I.	On		12.43	1.15	
LOUD SECURE		c01-node-customer-1c	150.00	N/A	RHEL-7.5_HVM_GA- 20180322-x86_64-1- Hourly2-GP2 x86_64	64.0	16	us-east-1c	L	On		11.59	1.96	
HELP		c01-node-kañka-1a-01	1,150.00	N/A	RHEL-7.5_HVM_GA- 20180322-x86_64-1- Hourly2-GP2 x86_64	64.0	16	us-east-1a	L	On		15.81	2.26	
		c01-node-kafka-1b-01	1,150.00	N/A	RHEL-7.5_HVM_GA- 20180322-x86_64-1- Hourly2-GP2 x86_64	64.0	16	us-east-1b	1.02	On		16.37	2.72	

Figure 3) Example ad hoc query to find large AWS compute instances with low utilization.

To create a query like that shown in Figure 3, you simply add elements to the query from the available menu items. In this case, the query specifies: resources of type Virtual Machine; with more than 16 processors; with CPU utilization less than 5%; running on the AWS hypervisor. For the NetApp lab environment, this results in a manageable set of results. The results can be narrowed even further by adding elements to the query. The same general approach can be used to zero in on waste in any resource. For example, you might create a query to identify storage volumes on expensive, high-performance storage types that have low IOPS.

Once you've identified a list of wasted resources, there are several actions you can take:

- Export the list to a file and send it to the bill payer for follow-up.
- Drill down on each item in the list, then identify and contact the owner to find out if it can be reclaimed, right sized, and so on.
- Tag each resource and monitor it over time to see if it remains underutilized.

#### 3.1 About Cloud Insights

NetApp makes it simple to get started with Cloud Insights. However, to get the most from your Cloud Insights environment, you should take some additional steps, including configuring annotations, queries, alerts, and dashboards specific to your environment:

- Annotations. Some information can't be obtained through discovery alone. By using annotations, you can add custom metadata that is specific to your IT and business needs. Cloud Insights provides a set of default annotations that you can use, plus you can create your own annotations to slice monitoring data in other ways. Annotation rules can be created that assign the appropriate annotations to new resources as they come online automatically.
- Queries. Cloud Insights provides a powerful visual search engine with filters and finders that enable you to monitor and troubleshoot a hybrid cloud environment easily. You can search assets at a granular level based on various criteria, including annotations and performance metrics. You can also define rules to annotate resources based on these query results to automatically tag new assets without any manual intervention.
- Alerts. You can create performance policies to generate alerts when a resource exceeds a specific SLI. This feature enables you to proactively monitor your environment to detect problems before they affect your operations. You can easily create targeted alerts to detect problems while reducing the noise in your monitoring environment to avoid alert fatigue.

• **Dashboards**. Cloud Insights comes with a default set of dashboards based on the data collectors that you enable. Each dashboard is designed to help you answer specific questions about your environment. Your team can also create highly tailored dashboard views to meet diverse monitoring and troubleshooting needs. You can start from an existing dashboard and customize it, or you can create entirely new dashboards from scratch using powerful tools to visualize available data.

These capabilities are described in detail in the White Paper <u>Enhance Monitoring by Understanding the</u> <u>Relationships Between Resources | WP-7302</u>, including tips on configuring and using each feature.

# 4 Using Cloud Insights Dashboards

You can also use Cloud Insights to create custom dashboards for the express purpose of identifying and tracking wasted resources. The example shown in Figure 4 is designed to identify and track wasted compute instances (EC2 instances) and block storage (EBS storage) capacity in AWS.

EC2 With < 10% CPU % of Over-Specced Instances	Powered off EC2	EBS Powered Off Capacity	Orphaned EBS Capacity	Idle EBS Cost
9		EBS Powered Off Capacity	Orphaned EBS Capacity	Idle EBS Cost
9		EBS Powered Off Capacity	Orphaned EBS Capacity	Idle EBS Cost
567 76.2	FC			
<b>50/ 76 2</b>		0.2	096	9 44 051
	56	9.3 тв	<b>98.6</b> TB	11.05k
Instances	Instances	Capacity - Total	Capacity - Total	\$/Month
EC2 Worst Offenders		unt		
stagro-g	0.15	•		
CPU Utilization - Total (%)	0 2	<b>8</b> 4 6 8	10 12 Processors	14 16 18
map-041ead50033395b82	Powered Off EC2 Inventory 70 items found in 21 groups			
snap-0fe6e36aa7314bb8	Instance Type	Name	Capacity - Total ↓ Memory (GB) (GB)	Processors IOPS - Total (IO/s)
snap-0216196576793cd6 snap-046c3044aacb06832	+ m5.xlarge (7) + r5.4xlarge (4)			28 N/A
	Ec2 Worst Offenders many man	Largest Orphaned EBS Vols  support Constant 215005564  support Constant 21500556  support Constant 21500556  support Constant 21500556  support Constant 2150055  support Constant 21500556  support Constant 2150055  support Constant 215005  su	Largest Orphaned EBS Vols regol 1005/2017/11-100505 regol 2005/2017/11-100505 regol 2005/2017/11-1005 regol 2005/2015 regol 2005/2015 rego	Linstances Capacity - Total Capacity - Total Capacity - Total Capacity - Total Capacity - Total Cap

Figure 4) Example Cloud Insights dashboard showing waste in AWS.

This dashboard shows high-level EC2 and EBS statistics across the top, including idle and powered off compute instances and storage. The four detail panes provide more detail on wasted resources. From a dashboard like this, you can drill down to find out more about a resource by clicking on it. This type of dashboard allows you to quickly identify the biggest opportunities for reclaiming wasted resources, and take action as described in the previous section.

To create your own custom dashboards to track resource waste, you have two options. When you add a data collector (as described in earlier papers in this series), you are presented with a gallery of relevant dashboards to choose from. An example is shown in Figure 5. As you can see, many of the recommended dashboards focus on reclaiming waste, right sizing compute instances, and re-tiering storage volumes.

Coul Insights x Coul Insights     Coul Insights     F → C O ■ ps1325.c01.cloudinsights.netapp.co	x Coud insights x Coud insights x Coud insights x Coud insights x R
	NetApp CDS Sa / All Dashboards / Import from Gallery
MONITOR & OPTIMIZE	Dashboard Gallery
🏠 номе	Recommended Dashboards (13) We found (13) dashboards that match your installed devices.
C DASHBOARDS	Name
	AWS Admin - Where do I have EC2 Latency?
	AWS Admin - Where can I reclaim waste?
X MANAGE	AWS Admin - Which EC2 are in high demand?
ADMIN 25	AWS Admin - EC2 Instance Performance by Region
	AWS Admin - EBS Storage that can be moved to the lowest tier
CLOUD SECURE	Azure Admin - Where can I potentially reclaim waste?
	NetApp Admin - How are my all-flash FAS arrays performing?
T HELP	NetApp Admin - Where do I have latency on my FlexVols?
	Storage Admin - Which nodes are in high demand?
	Storage Admin - Which pools are in high demand?
	VMware Admin - Where are opportunities to right size?
	VMware Admin - Where can I potentially reclaim waste?
	VMware Admin - Where do I have VM Latency?

Figure 5) Dashboard Gallery showing available options based on installed data collectors.

You also have the option to create your own custom dashboards, either by modifying an existing dashboard to meet your needs or creating entirely new dashboards from scratch. More details on creating custom dashboards are discussed in <u>WP-7302</u>.

## **5** Incorporating Cost Information in Cloud Insights

You might have noticed that the dashboard shown in Figure 4 incorporates AWS cost information. Figure 6 shows an example of an Azure dashboard with several additional cost views.

Courregets × Courregets     F = C = pr132.c01.doudreight.retapp.com	X Courtregets X Overlangtes X Overlangtes X Overlangtes X	99. X [ +			×	
	NetApp CDS Sa / All Dashboards / zz Azure Spend (webinar)		① Last 24 Hours	▼ Add Variable ▼ Add Widget ▼	🗈 Save 🔽 🏭	
MONITOR & OPTIMIZE	the dottom right then gives them a phonty list to target and linestigate more. The tor	unest towards the bottom right corner are the viv	is with the lowest cirio and highest o	22.		
🏦 номе	Estimated Monthly Azure VM Spend :	Cold Instances	: Spend on C	Cold Instances		
DASHBOARDS						
	72 0 11	120		11 111	**************************************	
	73.84k	128		11.44K		
🔅 ADMIN 🛛 🔁	\$/Month	Instances	5	\$/Month	5	
CLOUD SECURE	Instance type spend	: Low CPU Usage vs				
• неср		CPU Utilization - Tota 12 6 4 2	(X)			
💄 Joshua Moore	Standard_A2 Standard_DS5_v2 Standard_B2ms Sta	ndard_DS2_v2 ndard_D2s_v3	100 200	300 400	•	
4 Minimize	Standard_DS3_v2 Standard_DS1_v2	5	100	Pay as you go	Checklist (0/3) +	

Figure 6) Custom Azure dashboard showing cost metrics.

The panel in the lower left of the dashboard breaks down spending by instance type. By clicking on the three vertical dots in the top right of the panel and selecting Edit, you can see how this is accomplished, as shown in Figure 7.

Figure 7) A widget to show Azure spending by instance type.

nstanc	e type	sper	nd						
🚖 Virtu	ial Machin	e	¥						
Pay as yo	ou go 📎	*	Filter By Hypervisor OS	azure × +					
Roll Up	Sum	• t	y Instance Type 🔻 🗙	Show Top 🔻	10	Include others	Time aggregate by	Avg 👻	Less Options
	Pie Char		y Instance Type 👻 🗙	Snow Top *	10	Include others	Time aggregate by	Avg 👻	Less Opt



Note the similarity between this process and that of creating a query as shown in Figure 3. In this case, we are filtering resources of type Virtual Machine, that have the Pay As You Go tag. We then roll the data up by Instance Type in the resulting pie chart.

Cloud Insights includes a flexible metadata engine with customizable tags or annotations. Metadata associated with the Pay As You Go tag includes the on-demand costs for each Azure instance type, providing accurate cost information that can be manipulated and reported. You can create simple rules to tag instances with these costs based on instance type or use the Cloud Insights API for greater flexibility.

## 6 Shameback Reporting

Using Cloud Insights metadata, you can associate each individual resource with a creator. This enables reporting at varying levels of granularity all the way down to individual users, as shown in the example dashboard in Figure 8.



Figure 8) Example AWS dashboard showing usage by user.

Many of the reporting tools available—if they identify potential resource waste at all—provide a very top down view of costs and usage. This can be useful for bill payers, but often the administrators, engineers and developers incurring these costs in the first place don't use them. Typical showback and chargeback reports simply provide visibility of costs but not what cost is associated with waste. This Cloud Insights capability can empower you to control your own usage and limit waste.

With a dashboard like Figure 8, you can immediately see the information you need to clean up after yourself. These views can be integrated into the daily workflows of front-line engineers, along with performance and capacity metrics, to proactively control waste. Many users are conscientious and want to use resources efficiently, they just need the tools to help them identify the resources they might have forgotten. Even less conscientious users probably don't want their names visible at the top of the list for long. That is why reports like this are sometimes referred to as shameback rather than showback reports.

## 7 Summary and Next Steps

NetApp Cloud Insights can help you monitor, troubleshoot, and optimize your infrastructure. This paper examines the optimization capabilities of Cloud Insights, explaining how you can quickly identify wasted resources—like idle VMs and orphaned storage volumes—in the cloud and on-premises. Cloud Insights goes beyond typical chargeback and billing reports, showing usage down to the granularity of individual users.

To learn more, go to NetApp Cloud Central at <u>https://cloud.netapp.com/cloud-insights</u>, where you can experience Cloud Insights with a 30 day free trial.

## Where to Find Additional Information

To learn more about the information that is described in this document, review the following documents and/or websites:

- NetApp Cloud Central
   <u>https://cloud.netapp.com/cloud-insights</u>
- NetApp Cloud Insight Documentation Center <u>https://docs.netapp.com/us-en/cloudinsights/</u>
- NetApp Product Documentation
   <u>https://docs.netapp.com</u>

## **Version History**

Version	Date	Document Version History
Version 1.0	February 2020	Initial release.

Refer to the <u>Interoperability Matrix Tool (IMT)</u> 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.

#### **Copyright Information**

Copyright © 2020 NetApp, Inc. All Rights Reserved. Printed in the U.S. No part of this document covered by copyright may be reproduced in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or storage in an electronic retrieval system—without prior written permission of the copyright owner.

Software derived from copyrighted NetApp material is subject to the following license and disclaimer:

THIS SOFTWARE IS PROVIDED BY NETAPP "AS IS" AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE, WHICH ARE HEREBY DISCLAIMED. IN NO EVENT SHALL NETAPP BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

NetApp reserves the right to change any products described herein at any time, and without notice. NetApp assumes no responsibility or liability arising from the use of products described herein, except as expressly agreed to in writing by NetApp. The use or purchase of this product does not convey a license under any patent rights, trademark rights, or any other intellectual property rights of NetApp.

The product described in this manual may be protected by one or more U.S. patents, foreign patents, or pending applications.

Data contained herein pertains to a commercial item (as defined in FAR 2.101) and is proprietary to NetApp, Inc. The U.S. Government has a non-exclusive, non-transferrable, non-sublicensable, worldwide, limited irrevocable license to use the Data only in connection with and in support of the U.S. Government contract under which the Data was delivered. Except as provided herein, the Data may not be used, disclosed, reproduced, modified, performed, or displayed without the prior written approval of NetApp, Inc. United States Government license rights for the Department of Defense are limited to those rights identified in DFARS clause 252.227-7015(b).

#### **Trademark Information**

NETAPP, the NETAPP logo, and the marks listed at <u>http://www.netapp.com/TM</u> are trademarks of NetApp, Inc. Other company and product names may be trademarks of their respective owners.

WP-7319-0220

