FinOps
Operational Best Practices

January, 2023

DRAFT

Federal Technology Investment Management (FTIM)
Community of Practice (CoP)

&

Cloud & Infrastructure (C&I) CoP

General Services Administration

Office of Government-wide Policy

Overview

The Federal Technology Investment Management (FTIM) Community of Practice (CoP) and the Cloud and Infrastructure (C&I) CoP established a FinOps Pilot project to better understand cloud cost management and optimization in alignment with the best practices of the FinOps Foundation.

Unmanaged and ungoverned Infrastructure as a Service (IaaS) services can lead to significant cost overruns. It's important to implement guardrails to ensure costs are accounted for, managed, and allocated to the consumer of these resources. While agencies can leverage native IaaS cloud spending tools to get an overview of their expenditures, it's recommended to take a more structured approach to govern cloud spend. This can be accomplished through the FinOps model and framework.

FinOps is the practice of bringing a financial accountability cultural change to the variable spend model of cloud, enabling distributed engineering and business teams to make trade-offs between speed, cost, and quality in their cloud architecture and investment decisions. Through this pilot, three federal executive agencies at various points in their cloud journey shared their implementation of cloud resources and cloud financial management practices with the General Services Administration Technology Business Management Project Management Office (GSA TBM PMO). This data was then analyzed and feedback provided to pilot participants on how they may optimize their cloud resources using the FinOps framework.

This document describes best practices in use by the pilot agencies, as well as best practices defined by the FinOps Foundation and the Cloud Service Providers (CSPs)

Note: The federal agencies involved in the pilot described above used Amazon Web Services (AWS) and Azure Cloud Service Providers (CSPs), but the best practices within this document can also be applied to the Google Cloud Platform (GCP).

Account Structure

CSPs provide mechanisms to relate accounts to each other and align them hierarchically to how they will be consumed and funded. Agencies can take advantage of the AWS (account/sub account) or Azure (subscriptions) construct to align costs to consumers by function or business unit. This functionality provides a hierarchical arrangement to compartmentalize cloud spend to the appropriate consumers and separate shared services from client-specific services. It also ensures untagged or untaggable resources are still accounted for in the proper account or subscription. One equitable approach to allocation is a mechanism that uses proportionalized costs from shared services to other sub-accounts or business units based on spend or consumption.

Additionally, as part of the cost allocation model, there will be administrative, operational, and overhead costs associated with cloud spend and should be allocated similarly to cloud costs.

Optimization Best Practices

The cloud is complex and there are many facets in how organizations use it and how it matures over time. Determine the highest priority opportunities based on the most value for the time and money spent on cloud and address those first.

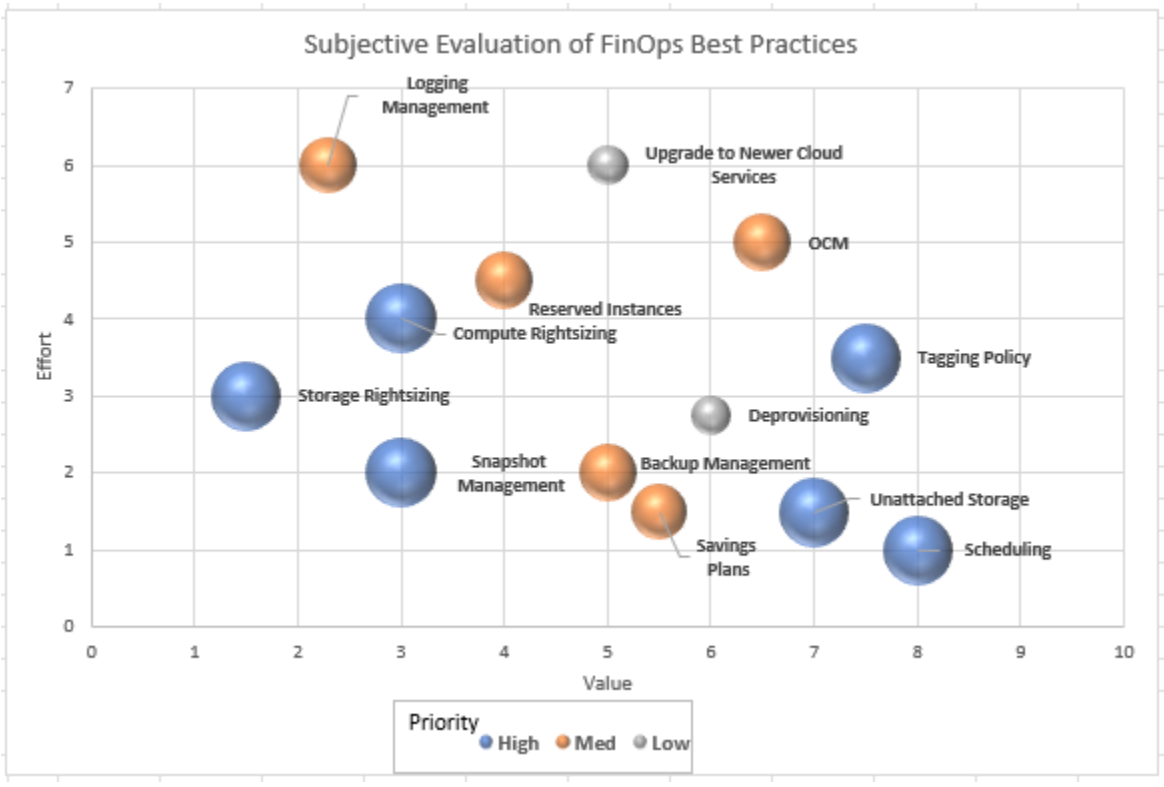
Depending on the needs for high compute load, storage and retention, security standards, risk tolerance, etc., no two organizations' cloud usage profiles are the same. Therefore, the optimization opportunities are different as well. However, these best practices are considered to be common ground that can be incorporated into the existing processes of finance, cloud, and engineering teams. Managing cloud resources and collecting relevant data will need to be performed on an hourly, daily, or at least weekly cadence. Establishing metrics and monthly reporting will benefit cloud operations and should be integrated with the overall IT financial management program. Anomaly detection, cost allocation, and trending over time are common reports that will provide insight and inform data-based decision making. Chargeback or showback will provide transparency of cloud costs directly to the business consumers and will improve overtime as you incorporate more FinOps standards.

The table and chart below are subjective lists of some of the most common best practices for finding optimization opportunities and their general prioritization. Again, depending on your organization, your effort to value ratio may change and you may prioritize them differently or modify the metrics. The ones listed in the below table link directly to the details for that metric.

- "Value" in the table below should be considered as "the value to the organization." It could be a combination of performance, operational efficiency, and saving money.
- "Effort" should consider the level of effort to implement and maintain a particular metric.
- "Priority" would be based on the importance the metric provides to the organization in alignment with the maturity of your cloud program. In the below example, an organization that is just beginning to migrate to the cloud would not have much opportunity for deprovisioning so the priority would be "low."

Optimization Metric	Value (1-10)	Effort (1-6)	Priority
Compute Rightsizing	3	4	High
Scheduling	8	1	High
Snapshot Management	3	2	High
Storage Rightsizing	1.5	3	High
Tagging Policy	7.5	3.5	High

Unattached Storage	7	1.5	High
Backup Management	5	2	Med
Organizational Change Mgmt (OCM)	6.5	5	Med
Logging Management	2.3	6	Med
Reserved Instances	4	4.5	Med
Savings Plans	5.5	1.5	Med
Tagging Compliance and Maintenance	9	5	Med
Change Management & Governance	9	5	Med
Deprovisioning	6	2.75	Low
Upgrade to Newer Cloud Services	5	6	Low



Each element of the table and accompanying graph are described below in order of priority

High Priority

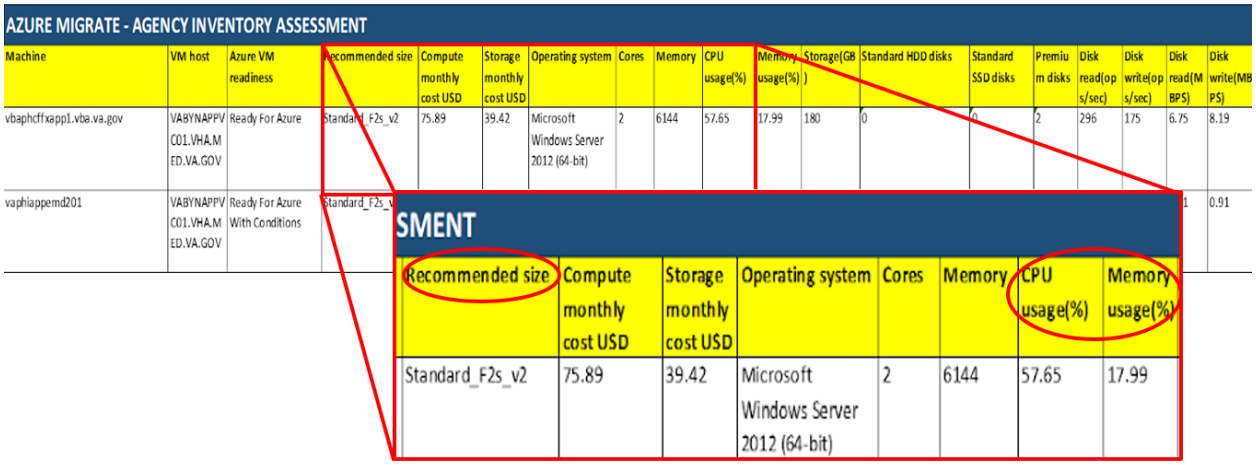
Compute Rightsizing

Cloud compute resources have standard prices referred to as “On Demand” (or Pay-as-you-go Pricing for Azure) and it is a best practice to rightsize your purchased instances to avoid procuring unnecessary extra resources. Because a one-to-one relationship does not exist for on-prem resources (CPU, Memory) being migrated to the cloud, a lift and shift methodology may create overallocated or underutilized resources that require monitoring and fine tuning once in the cloud. Therefore, during post-migration, close attention needs to be given to your compute utilization so it can be adjusted based on performance to the right instance size and type. Once you understand your cloud compute workload, saving plans or Reserved Instances (RIs) should then be purchased.

This discussion pertains to rightsizing compute, such as Instances or virtual machines (VMs). The first step is to monitor and analyze your current use of services to gain insight into Instance or VM performance and usage patterns. Generally, one month will provide sufficient data to observe performance and capture the workload and business peak. However, a monthly or quarterly review will determine if there are other anomalies or cyclical peaks.

There are many tools and methods that can be used to monitor and assess Instance and VMs usage performance. One of the available discovery and assessment tools is Azure Migrate. Azure Migrate can be used for both AWS and Azure environments. You can also use AWS Discovery as well, but for only AWS.

Azure Migrate generates performance assessments based on CPU and RAM-utilization data. See the example assessment below.



AZURE MIGRATE - AGENCY INVENTORY ASSESSMENT

Machine	VM host	Azure VM readiness	Recommended size	Compute monthly cost USD	Storage monthly cost USD	Operating system	Cores	Memory	CPU usage(%)	Memory usage(%)	Storage(GB	Standard HDD disks	Standard SSD disks	Premium disks	Disk read(ops/sec)	Disk write(ops/sec)	Disk read(MBPS)	Disk write(MBPS)
vbaphcffxapp1.vba.va.gov	VABYNAPPVC01.VHA.MED.VA.GOV	Ready For Azure	Standard_F2s_v2	75.89	39.42	Microsoft Windows Server 2012 (64-bit)	2	6144	57.65	17.99	180	0	0	2	296	175	6.75	8.19
vaphiappemd201	VABYNAPPVC01.VHA.MED.VA.GOV	Ready For Azure With Conditions	Standard_F2s_v															0.91

SMENT

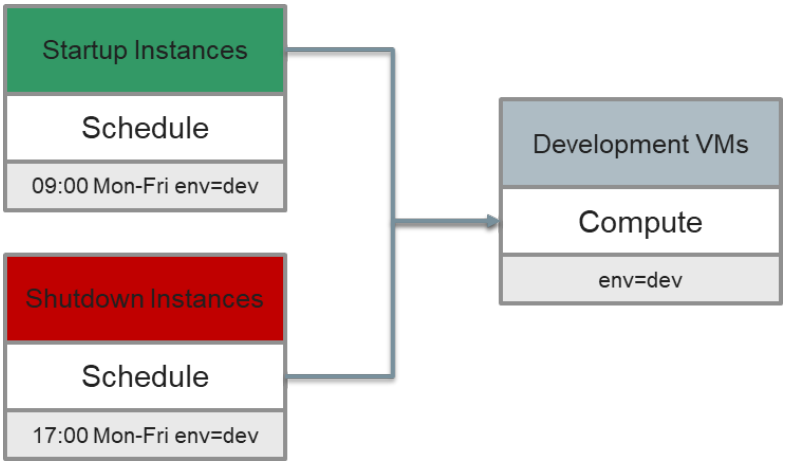
Recommended size	Compute monthly cost USD	Storage monthly cost USD	Operating system	Cores	Memory	CPU usage(%)	Memory usage(%)
Standard_F2s_v2	75.89	39.42	Microsoft Windows Server 2012 (64-bit)	2	6144	57.65	17.99

Once usage performance data has been analyzed, rightsizing aims to align the sizes and types of Instances and VMs to your specific workload capacity and performance requirements at the lowest

possible cost. This process can be long and challenging. Instances and VMs run various workloads and applications and the owners of these applications (i.e., developers, business owners, etc.) need to be consulted when making recommendations to downsize.

Scheduling

Scheduling provides the ability to turn off VMs when they are not in use and turn them back on when they are needed again. The most common use case for this is applied to dev and test servers that are not typically required to run 24/7.



Most CSPs have Instance Scheduling features that can be used to start and stop instances. This can be done manually through their native console or scripted by using server names or tags to determine which machines are impacted by a particular schedule.

By following this solution, cloud architects can automate the start and stop of services per their usage pattern to save more than 60 percent (108 [night & weekend hours]/168 [hours per week]) of the costs on their nonproduction environments.

Snapshot Management

Snapshots are a mechanism for creating a backup of a system's current state. Snapshots are often put in place and forgotten. New snapshots do not overwrite older ones, so they will continue to grow in quantity. A policy should be put in place that determines the best fit for the snapshot lifecycle. Then, in accordance with that policy, snapshots can be managed on a daily or weekly basis so older snapshots are deleted. In many instances, this process can be scripted and automated.

Snapshots can be managed through the AWS or Azure Management Console, Command Line Interface (CLI), or Software Development Kits (SDKs).

For example, Amazon Snapshots integrate with Amazon Data Lifecycle Manager (DLM), which allows you to define policies that help you automate snapshot lifecycle management. This allows you to create policies to automate multiple tasks, including creating, deleting, retaining, and sharing snapshots. This reduces the operational burden of snapshot management and helps reduce storage costs as DLM automatically deletes outdated backups based on your specified policies. Azure has a similar Lifecycle Manager. One important consideration is that the Lifecycle Manager uses resource tags to identify the snapshots. Hence, another reason to tag resources.

Azure Storage lifecycle management offers a rule-based policy you can use to transition BLOB (Binary Large Object) data to the appropriate access tiers or to expire data at the end of the data lifecycle. A lifecycle policy acts on a base BLOB, and optionally on the BLOB's versions or snapshots.

Disk Volume Storage Rightsizing

Disk volume storage tends to be configured and forgotten and, in the cloud, this can cause costs to spiral out of control quickly. Disk volume storage is also often over provisioned, unattached, or underutilized, incurring costs unnecessarily.

Disk volumes should be rightsized with the application's appropriate capacity, IOPS (input/output operations per second), and throughput. You should monitor the read-write access for all disk volumes periodically. If the throughput is low, downgrade the disk volume type in question to reduce disk volume storage costs.

In AWS, you can use this command to identify existing disk volumes.

AWS CLI (Command Line Interface) command:
aws ec2 describe-regions --query "Regions[].RegionName" --output text

You can then use AWS CloudWatch to monitor and identify disk volumes that have low throughput. Using this data, you can determine the average number of IOPS provided and the average throughput of your volume. The average number of IOPS that a disk volume provides is then calculated by adding the CloudWatch metrics for VolumeReadOps and VolumeWriteOps.

AWS Cloudwatch Steps

1. Open the CloudWatch console.
2. Under Metrics in the sidebar, select All Metrics.
3. Select EBS, and then select Per-Volume Metrics.
4. Select the metrics you want to graph:
   To graph the actual average, select IOPS, VolumeReadOps, VolumeWriteOps, and VolumeIdleTime.
   To graph the actual average throughput, select VolumeReadBytes, VolumeWriteBytes, and VolumeIdleTime.
5. Select the Graphed Metrics tab.
6. In the Statistic dropdown list, select Sum.
7. Select the period of time you want to view from the Period dropdown list.
   Note: The Period in the previously mentioned formulas represents a given time in

CloudWatch. The specified Period of the CloudWatch graph equals the volume's collection period.

8. In the Add Math dropdown list, select Start with an empty expression. Then, enter the following expression: (m1 + m2)/(PERIOD(m3)-m3)
9. Select Apply.

Once you identify a disk volume type to downgrade, you can switch the volume type.

For example, if you can, switch from an io1 volume type to gp2. This saves on IOPS per month and has a lower hourly storage rate. If your volume is 500GB or larger, convert to sc1 or a cold HDD to dramatically lower your storage rate. If your volumes start getting more traffic, you can easily return to the IOPS volume.

In Azure, you can use this command to identify existing disk volumes

```
Azure CLI
az disk list [--resource-group]
```

You can gain insights on your disk volume performance with Azure Monitor. The disk volume performance metrics in Azure Monitor will identify disk I/O (Input/Output) or throughput. These metrics describe the utilization of the disk volume IOPS or throughput on a scale of 0 to 100 percent, where 0 percent utilization means that resource is not being used at all and 100 percent utilization means that the resource is being run at its limit and is a bottleneck.

Tagging Policy

Tagging occurs twice in the above graphic because it is vital to have a tagging policy and to continuously govern that policy. The ongoing maintenance and compliance with the tagging policy may require more effort long term and is of equal value, but it is listed as a lower priority because the policy should come first.

Tags are metadata labels associated with IT resources. They are necessary to track individual IT resources throughout their lifecycle and their relationships to other resources and dependencies. There are a variety of mechanisms CSPs provide to enable tagging as well as a number of different categories that each resource may need to tag.

A robust cloud tagging strategy must integrate with the on-prem tagging of resources. This provides consistency, enabling the calculation of the application's or system's Total Cost of Ownership (TCO) and insights on when items can be decommissioned. The tagging strategy and account structure ensure resources are accounted for in the appropriate sub-account/subscription. Not every resource that appears on an invoice is taggable. The remaining untaggable resources will still be associated with the

appropriate sub-account or subscription so the costs will get allocated to the business unit associated with the sub-account. However, some granularity is lost and these costs will have to be spread to the lower level systems, applications, or solutions within that sub-account or subscription.

Tags are typically categorized in key/value pairs and generally follow the guidelines below:

- The name of the application, system, or service with a consistent naming convention.
- Abbreviations must be consistent across all resource types (i.e., compute, storage, etc.), and also across cloud providers and on-premise resources.
- A tag should serve a single purpose.
    - Use separate tags for the environment (i.e., dev, test, prod) vs. the system.
- Tags should be human readable and machine readable to comply with the mechanisms for automation.
    - Use delimiters purposefully and consistently.
- Do not use generic names, such as "log server" or "file server," or hostnames.

CSPs vary in the degree of flexibility, tag length, number of tags per resource, etc., so consistency and integrity of tagging is the responsibility of the cloud team.

GSA has created a complete Cloud Tagging Strategy Guide for developing your organization's cloud policy.

Unattached and Zero Throughput Disk Volume Storage

Storage volumes should be monitored regularly. At times, a server will be decommissioned and the storage becomes unattached and inaccessible. However, unless it is deleted, it is still accruing costs. Unattached storage volumes should be evaluated to determine whether the data is still required or can be deleted or moved to a lower cost archive storage. This process can be automated through scripts.

Managing disk volume storage costs can be complicated. You have to ensure you have the right kind of storage for your workload and that you're using your storage effectively. You also have to periodically look for unused and underused volumes. The right tools can help you find elusive disk volumes and manage storage costs.

Choosing the Right Disk Volume

There are many types of Solid State (SSD) backed or Hard Disk (HDD) backed disk volumes designed for various workloads. Generally, the first major decision to make is whether your priority is IOPS (the count of disk operations per second) or throughput (the volume of data being transferred per second). For throughput, you need to determine if you are doing many small, random I/O, which means you want SSD, or if you are performing large, sequential reads, in which case you want HDD. These decisions impact cost and performance.

Always Tag Disk Volumes

Tagging is the first step to effective management of disk volume. Manage disk volume tagging the same way Elastic Compute Cloud (EC2) instances and other resources are tagged. Use tags to assign a simple key/value pair to your instances, disk volumes, and snapshots so you can group and manage those resources together.

After disk volumes are created, it can be tricky to know how they are being used, which makes them difficult to tag. Graffiti Monkey is a good tool for associating disk volumes with their instances. You can also use Lambda Graffiti Monkey to run Graffiti Monkey serverless.

Track Disk Volumes and Control Costs

1: Remove Unattached Disk Volumes

Disk volumes persist after the instances stop. This is good for data retention, but bad if you're paying for storage no longer needed. Unused disk volumes are called unattached or orphaned volumes, and they're marked as "available" if you check their status. These volumes are not being used so removing them will reduce costs.

Below are a few recommendations for AWS that will help identify unattached disk volumes and delete them to optimize cloud costs and prevent underutilized resources. In addition to identifying zero throughput disk volumes, there are similar commands and scripts for Azure.

Step 1

To find all volumes, go over all available regions.

Aws CLI command:

```
aws ec2 describe-regions --query "Regions[].RegionName" --output text
```

Step 2

Thoroughly review all volumes for every available region and check the current status. If the current status is **available,** this volume is not attached to any instances.

AWS CLI command:

```
aws ec2 describe-volumes --region "$region" --filters Name=status,Values=available
--query 'Volumes[].[VolumeId]' --output text
```

AWS CLI has a pagination mechanism for large amounts of data in the output, such as when there are many volumes in a region.

```
for

   region in $(aws ec2 describe-regions --query "Regions[].RegionName" --output
text);

 do for

   volumeId in $(aws ec2 describe-volumes --region "$region" --filters
Name=status,Values=available

   --query 'Volumes[].[VolumeId]' --output text);

 do echo "Region: $region VolumeId $volumeId";

 done;

done;
```

- The scripts above can be automated and you can execute the script twice with a one-day delay to find volumes still not attached after a day.
- It is recommended that your account has cloud trail logs enabled. With AWS, you can try to find the last attachment date by following these instructions.

https://aws.amazon.com/ru/premiumsupport/knowledge-center/list-attachments-history-ebs-volume/

2: Identify Zero Throughput or Zero IOPS Volumes

Once you have removed unattached volumes, look for attached volumes not in use. These often show up when the associated instances have been turned off and the disk volumes were forgotten. To discover these volumes, look at the volume network throughput and IOPS. If there hasn't been any throughput or disk operations in the last 10 days, the volume is probably not in use.

AWS

Use CloudWatch metrics to identify possible zero throughput volumes. CloudWatch monitors the IOPS (op/s) and throughput (byte/s) for all AWS disk volume types by collecting samples every minute.

1. Check the VolumeIdleTime metric. This metric indicates the total number of seconds when no read or write operations are submitted in a specified period. If VolumeIdleTime is high, that means the volume remained idle for most of the collection period.

With the VolumeIdleTime metric for throughput, there are VolumeReadBytes and VolumeWriteBytes metrics.

2. Use the following formula to calculate the average throughput that is reached when the volume is active:

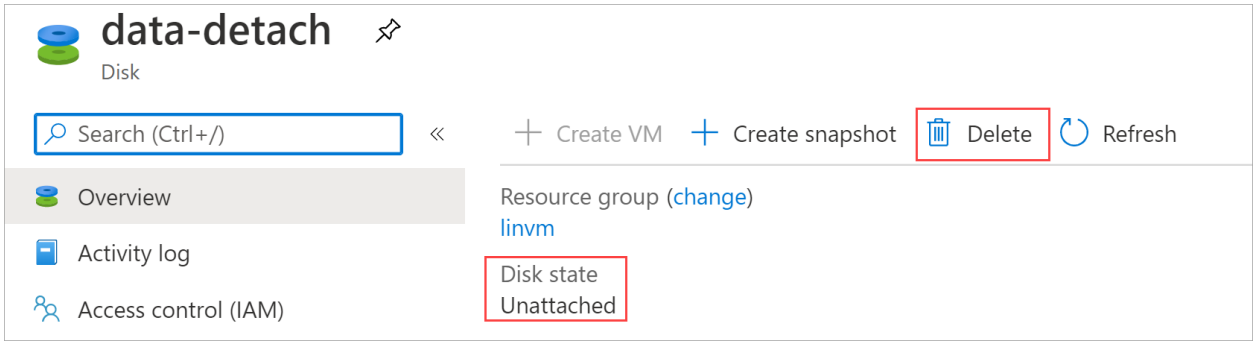
> Actual Average Throughput in Bytes/s = (Sum(VolumeReadBytes) + Sum(VolumeWriteBytes)) / (Period - Sum(VolumeIdleTime)).

Azure

For Azure disk management, find and delete unattached disk volumes following the process within the Azure portal explained below:

1. Sign in to the Azure portal.
2. Search for and select Disks.
3. On the Disks blade, you are presented with a list of all your disks.
4. Select the disk you'd like to delete, this brings you to the individual disk's blade.
5. On the individual disk's blade, confirm the disk state is unattached, then select Delete.

The following screenshot is of an individual managed disk blade. This blade will show 'unattached' in the disk state if it is unattached. You can delete this disk if you do not need to preserve its data any longer.



3: Rightsizing and Managing Disk Volumes

PIOPS volumes are costly but easy to change, so ensure they are managed appropriately. Simply go through your list of disk volumes and reduce the number of PIOPS volumes if necessary.

Example: If you can, switch from an io1 volume type to gp2. This saves on IOPS/month and has a lower hourly storage rate. If your volume is 500GB or larger, then convert to sc1, a cold HDD, to

dramatically lower your storage rate. If your volumes start to increase, you can easily go back to the PIOPS volume.

Medium Priority

Backup Management

Consideration has to be taken into the type of data that needs to be backed up, its retrieval likelihood, and why. The backup mechanism and the selected storage service are critical decisions with cost implications.

If there is a compliance mandate for seven years of records retention and a low likelihood the data will actually be accessed, then the storage solution should match the need where the cost to store the data would be very low but accessing the data is more expensive.

On the other end of the spectrum, if backup data is frequently used then you want to optimize for a storage service design and price for frequent I/O cycles.

Log Management

Log management is the continuous process of centrally collecting, parsing, storing, analyzing, and disposing of data to provide actionable insights to support troubleshooting, application performance, and security monitoring. This process can generate hundreds of gigabytes of log data per day. It is important to have a log management strategy that can analyze the logs, report on them, and then delete the log data once it is no longer needed, preventing excessive data that continues to accrue cost.

It is important to understand which workloads are the biggest contributors to log data to determine ways to reduce volume, such as restructuring how log entries are written. There also may be opportunities to configure log exclusions or to archive (on less expensive storage) logs only required for audit purposes.

Federal log management falls under memo M21-31, which details what has to be logged. There are requirements for log retention for each category of logging, typically 12 months of active storage followed by 18 months of cold storage.

Reserved Instances (RIs)

Balance is required between On Demand (list price) and discount commitments, such as RIs. You need to determine the scope of an RI so that it applies to either a specific availability zone or an entire region. Regionally-scoped RIs offer more flexibility because they can apply to any matching instance type within a region. However, unlike those scoped to an availability zone, they do not guarantee a capacity reservation.

The level of discount you can expect to achieve also depends on the attributes you specify during your purchase. These include:

- **Instance type:** The compute resource you want to reserve
- **Platform:** The operating system (OS)
- **Region:** The AWS region in which your RI will apply
- **Tenancy:** A standard EC2 instance or single-tenant hardware
- **Payment terms:**
    - **Length of term:** Can be either one or three years
    - **Upfront payment:** Pay for the entire commitment at its start
    - **Partial upfront:** Pay a lump sum at its start and the remainder monthly
    - **No upfront:** The commitment amount is divided by the term to determine monthly payments

NOTE: Further detail on RIs is available in the following document:
Cloud Rate Optimization through Discounts

The fundamental difference between Standard and Convertible RIs is the level of flexibility. With Standard RIs, what you reserve is what you get. By contrast, at any time during the term, you can exchange a Convertible RI for another Convertible RI with different specifications. This can include a different instance family, scope, platform, or tenancy.

Standard, regionally-scoped Linux reservations are more adaptable than those hosting Windows or enterprise Linux distributions. This is because AWS applies your RIs to any instance size in the same family as the Linux instance you specify. AWS then uses a normalization factor to apportion your credits. They can spread your RIs across a combination of smaller instance sizes or use them to reduce the cost of running a larger instance.

Savings Plans

Savings Plans offer significant savings over On Demand or Pay-As-You-Go pricing in exchange for a commitment to use a specific dollar amount of computing power. This rate is measured by dollars-per-hour over a one-year or three-year commitment period.

Savings plans can provide savings of up to 72 percent on your compute usage regardless of instance family, size, OS type, tenancy, or region. Savings plans are much simpler to manage and provide additional flexibility over RIs. They can apply across instance types and regions and generally do not apply to database usage.

NOTE: More information on savings plans is available in the following document:
Cloud Rate Optimization through Discounts

Tagging Compliance and Maintenance

Having a tagging policy is just the beginning of effective tagging. Operationally, tagging governance, enforcement, compliance, and maintenance must be ongoing. Tagging should be part of configuration management and cloud resources provisioning policies.

In large organizations, there may be a need for a dedicated employee who is accountable for tagging compliance. This person does not actually do the tagging but is responsible for ensuring all taggable resources are tagged and that tags comply with policy. Regular monthly reporting on tagging should be pushed out to those responsible for implementing or correcting tags. These reports should be generated automatically from cloud source data. The types of tagging reports should include but not be limited by:

- Count of untagged resources organized by cloud service and business system
- Trends of untagged resources month after month
- Details of service and resource names that are not tagged
- Tag exception reports that detail resources where tags have changed or been deleted

Organizational Change Management & Governance

Effective organizational change management and governance are crucial components of a successful FinOps initiative in federal government agencies. By implementing OCM processes, agencies can establish support, develop specialized skills, enhance communication and collaboration, and create a culture of continuous improvement. These processes help ensure the success of the FinOps organization by establishing governance frameworks, optimizing costs, improving financial visibility, and ensuring financial accountability. With proper change management and governance in place, agencies can achieve their goals and maximize the benefits of their FinOps initiatives.

For more information and resources around Organizational Change Management (OCM) check out our Organizational Change Management Guide. While many of the resources within the guide refer to TBM the templates are generic enough to be repurposed to create a FinOps OCM strategy.

The following lists a few areas OCM can help federal agencies in their pursuit of standing up a successful FinOps team:

- **Cultural Shift**: Implementing a FinOps organization requires a significant cultural shift in how an agency manages its financial operations as well as system engineering . Change management is

essential to ensure that all stakeholders understand the goals of the FinOps initiative and are willing to adopt new processes and procedures. For example, agencies can use change management to encourage developers to take responsibility for their cloud resource usage and work more closely with financial teams to optimize costs.

    - Use the following template to Determine your OCM Strategy which provides assistance in translating other industry frameworks and change management terminology.

- **Governance Framework**: A well-defined governance framework is essential to the success of a FinOps organization. Change management can help agencies establish the governance framework by involving stakeholders, or change agents,early on in the design process and building buy-in and support. For example, agencies can use change management to ensure that the FinOps team has a clear understanding of the agency's cloud policies and procedures and works closely with security teams to ensure compliance.
    - Use the following template to Engage Change Agents which provides best practices and related tools for gaining sponsorship and fully utilizing your FinOps stakeholders.
    - Use the following template to Charter the Change Agenda which helps to craft a change agenda and perform a gap analysis.
    - Use the following template to Align to Enterprise Governance and Reviews which provides additional guidance to align and integrate FinOps with agency enterprise governance.
- **Skill Development**: A FinOps organization requires specialized skills in areas such as cloud technology, finance, and analytics. Change management can help agencies identify the necessary skills and provide training and support to develop those skills within the workforce. For example, agencies can use change management to provide training for financial analysts to understand cloud billing and cost allocation, and training for developers to optimize resource usage and reduce costs.
    - Use the following template to Measure for Change Success for best practices and example progress and performance measures to evaluate achievements, assess skill gaps, and determine next steps.
- **Communication and Collaboration**: Effective communication and collaboration are continuously needed for a successful FinOps initiative. Change management can help agencies establish clear lines of communication and collaboration between different departments and stakeholders to ensure that everyone is aligned with the goals of the FinOps initiative. For example, agencies can use change management to encourage regular meetings between the FinOps team, developers, and financial teams to discuss cost optimization strategies and identify opportunities for improvement. With everyone on the same page marketing the value proposition of FinOps becomes less of an uphill battle.
    - Use the following template to Market the Value Proposition which provides tips on evaluating competing values and facilitating value conversations across core leadership and agency functions to drive change implementation.
- **Monitor for Continuous Improvement**: Since cloud is a subscription service costs are incurred monthly and continuously. This poses a conflict in the way an engineering or operations team is accustomed to working. For on-prem resources the process was much more of a stand it up and

let it run approach whereas the cloud requires more continuous monitoring and tweaking to maximize performance and minimize cost. And for high demand or high throughput systems this can be necessary to do multiple times a day. This shows that a FinOps organization is an ongoing process of continuous improvement. Change management can help agencies establish a culture of continuous improvement by encouraging feedback, data-driven decision-making, and continuous learning.

- Use the following template to Foster a Sustaining Change Environment which provides guidance and sample strategies to develop a sustainability plan and evaluate lessons learned to continuously monitor, communicate, and improve a FinOps program.

Lower Priority

Lower priority items are also important, but if you are in the beginning of cloud migration and implementation then these are not as relevant until your cloud environment matures.

Deprovisioning and Dependencies

Cloud and enterprise architects may need to rely on multiple provisioning tools to customize how they use cloud resources. Applications and workloads in the cloud often tap into basic cloud infrastructure resources, such as compute, networking and storage, creating many interrelated dependencies. Many agencies also deploy workloads on more than one cloud platform, which makes it even more challenging to know what should be running.

Costs continue to accrue monthly so when an application is no longer needed or reaches its end of life, it needs to be deprovisioned or decommissioned. Some services may carry dependencies that need to be clarified, which can lead to unexpected overuse and surprise costs. Adherence to tagging strategies and policies is key to knowing all associated dependent resources that can be turned off. Monitoring CPU, memory, and storage at regular intervals can also alert the team to unused or underutilized resources that may not have been tagged sufficiently.

Upgrade to Newer Cloud Services

CSPs continue to add services, features, and capabilities and provide better performance sometimes at a better rate. Consider migrating certain workloads to those newer feature sets. If the current implementation uses RIs, moving such workloads may only be feasible once the RI period ends. Generally, an RI is tied to a particular underlying technology. However, it is key to understand the newest features and how they can benefit your organization through better system performance and availability. Then, determine a migration program to take advantage of such benefits.

Cloud Workforce Considerations

Having a workforce that is capable of supporting cloud technology should be a key planning consideration and part of your overall agency's cloud adoption strategy. The CIO organization should work with HR and other senior agency officials to identify skill gaps and develop a hiring, retention, and training strategy. One approach is to map the new skills required to manage, broker, and operate cloud services to legacy job series and categories. IT is rapidly evolving and CIOs should plan for ongoing and continuous upskilling and education. All IT disciplines are impacted by the cloud regardless of title and will need some foundational training. It is not uncommon to have personnel dedicated to tagging governance and cloud financial management through FinOps best practices.

Recruitment may be a challenge with a large pay disparity between government and the private sector. To hire talent, plan to emphasize other benefits outside of pay, such as remote work, mission, and work-life balance. Also, consider hiring contract staff to fill skills gaps, but ensure you understand the types of skill sets needed to carry out management activities in complex environments such as IaaS CSPs. Agencies can use GSA contract vehicles to acquire complete solutions that include CSP services and the requisite contract workforce to support them.

Conclusion

Cloud best practices are a continuous effort, not just for steady state operations but for longer term optimization, planning, and forecasting. This document is just one of a series of assets required to build a FinOps practice and improve the management of your cloud environment.

Resources and Credits

FinOps Foundation
https://www.finops.org/project/reducing-waste/

AWS Best Practices
https://aws.amazon.com/blogs/aws-cloud-financial-management/category/post-types/best-practices/

AWS console (information only). The actual console requires login from an active account.
https://aws.amazon.com/console/

Azure Best Practices
https://azure.microsoft.com/en-us/solutions/cost-optimization/#tools

Azure Migrate (information only). The actual Azure migrate hub requires login from an active account.
https://azure.microsoft.com/en-us/products/azure-migrate/

Federal log management memo M21-31

Open Source tools available at GitHub

Graffiti Monkey
Lambda Graffiti Monkey

Glossary

IaaS	Infrastructure as a Service
AWS	Amazon Web Services
CLI	Command Line Interface - A mechanism to interact with the cloud environment.
BLOB	A binary large object (BLOB or blob) is a collection of binary data stored as a single entity; typically images, audio, or other multimedia objects.
EC2	Amazon Elastic Compute Cloud (Amazon EC2) is a web service that provides secure, resizable compute capacity in the cloud.
EBS	Elastic Block Storage is an AWS storage offering.
IOPS	Input/output Operations Per Second - A measure of performance for storage systems. IOPS is a count of the read/write operations per second, but throughput is the actual measurement of read/write bits per second that are transferred over a network.
PIOPS	Provisioned IOPS SSD (io1) EBS volume types - A special type of volume created to fulfill the needs of very intensive I/O workloads that require very high throughput; useful for cases that are latency-sensitive, such as large database workloads.
CSP	Cloud Service Provider (Ex - AWS, Azure, GCP)
OS	Operating System
RDS	Relational Database Service
VM	Virtual Machine
GCP	Google Cloud Platform
HDD	Hard Disk Drive - A storage device with rotating disks.
SSD	Solid State Drive - A storage device using integrated circuits with no moving parts.

io1 volume type	io1 - A SSD storage volume type for provisioned IOPS, optimized for high throughput high performance transactional workloads involving frequent read/write operations.
gp2 volume type	gp2 - A SSD general purpose storage volume type also optimized for transactional workloads, however, at a lower performance level than the io series
Lambda	AWS Lambda - A serverless, event-driven compute service that lets you run code for virtually any type of application or backend service without provisioning or managing servers.