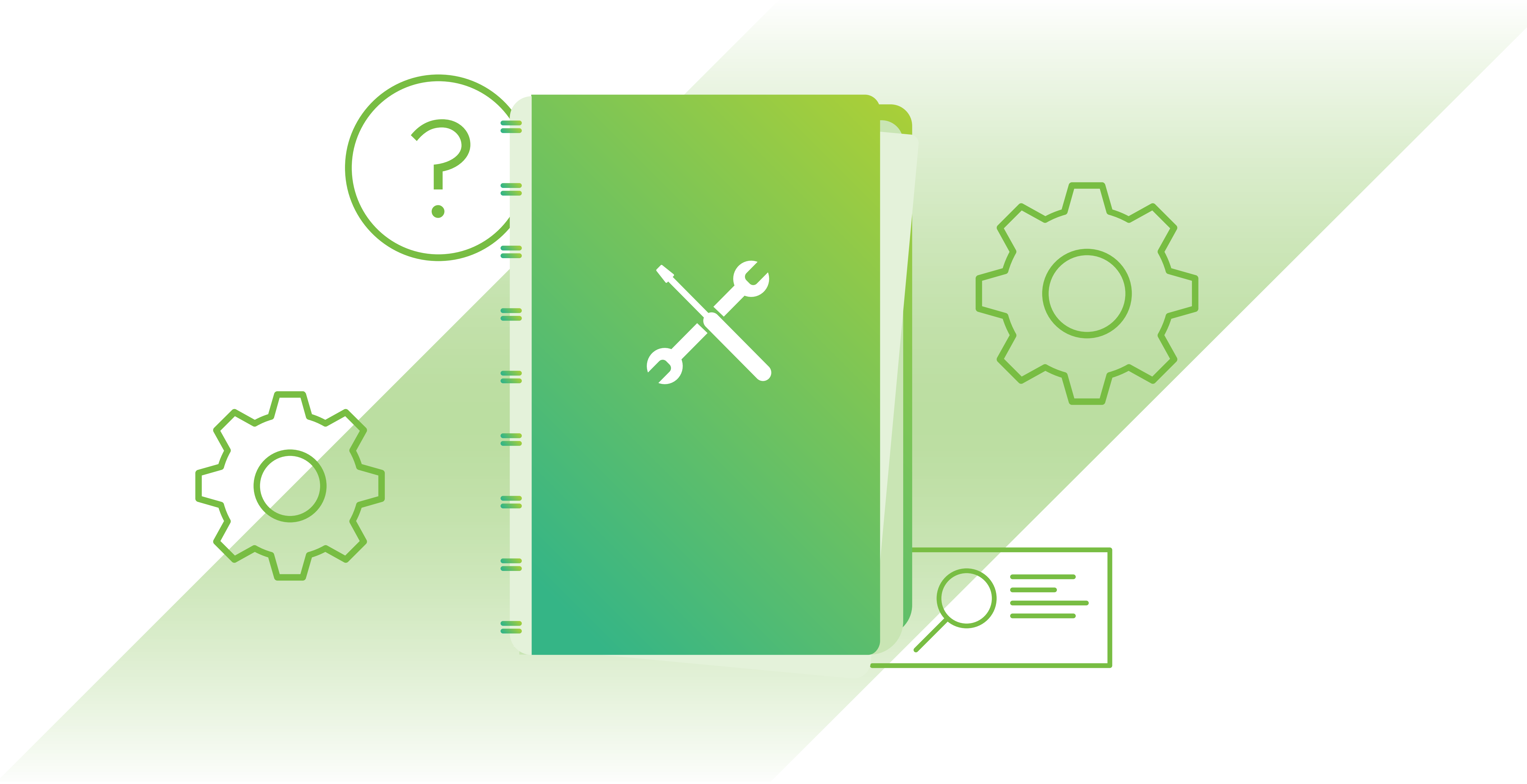
6 Policy Types for Multi-Cloud Governance

› Get Started

Why does governance matter?

If your organization has a large cloud environment, managing your assets and monitoring the rapid rate of change can be an extremely time-consuming task. The best way to govern these dynamic environments at scale is to implement automated policies that allow you to manage your environment in a relatively hands-off manner. These policies consist of a set of customizable rules that give you a simple and effective way to eliminate noise, gain consistency and control, and reclaim time that can be spent on more strategic projects.

This ebook outlines several types of policies and examples that you must put in place to centralize governance across your cloud environments.



1. Asset and configuration management

In most on-premises environments, IT teams have a solid process for understanding asset and configuration management. Using tools such as configuration management databases (CMDBs) and frameworks such as the IT Information Library (ITIL), organizations can tightly control deployments and ensure standardization. In the cloud world where almost any user can provision infrastructure in a few clicks with a credit card, these previously developed frameworks fall apart. To bring asset and configuration management back under control, advanced IT shops realized they needed to manage their environments by exception: by setting up rules for non-approved configurations and assets, and then closely monitoring them.

Tag compliance

Tagging is an essential way to accurately categorize assets to their appropriate business groups. By using tags, you can assign labels for categories such as department (e.g., engineering), product, environment (e.g., production), application (e.g., human resources information system [HRIS]), customer, or application role (e.g., Cassandra). Once a resource is tagged, usage associated with this resource is reported by this tag, allowing you to more easily associate costs to different business groups. Setting policies to identify assets that do not conform with your organization's internal tagging standards can help you stay on top of tag compliance. This includes untagged assets, mistagged assets, or misspelled tags.

Sample tag compliance policies

If any asset is missing the tag "Environment," send a notification and execute an AWS Lambda function to tag the asset.

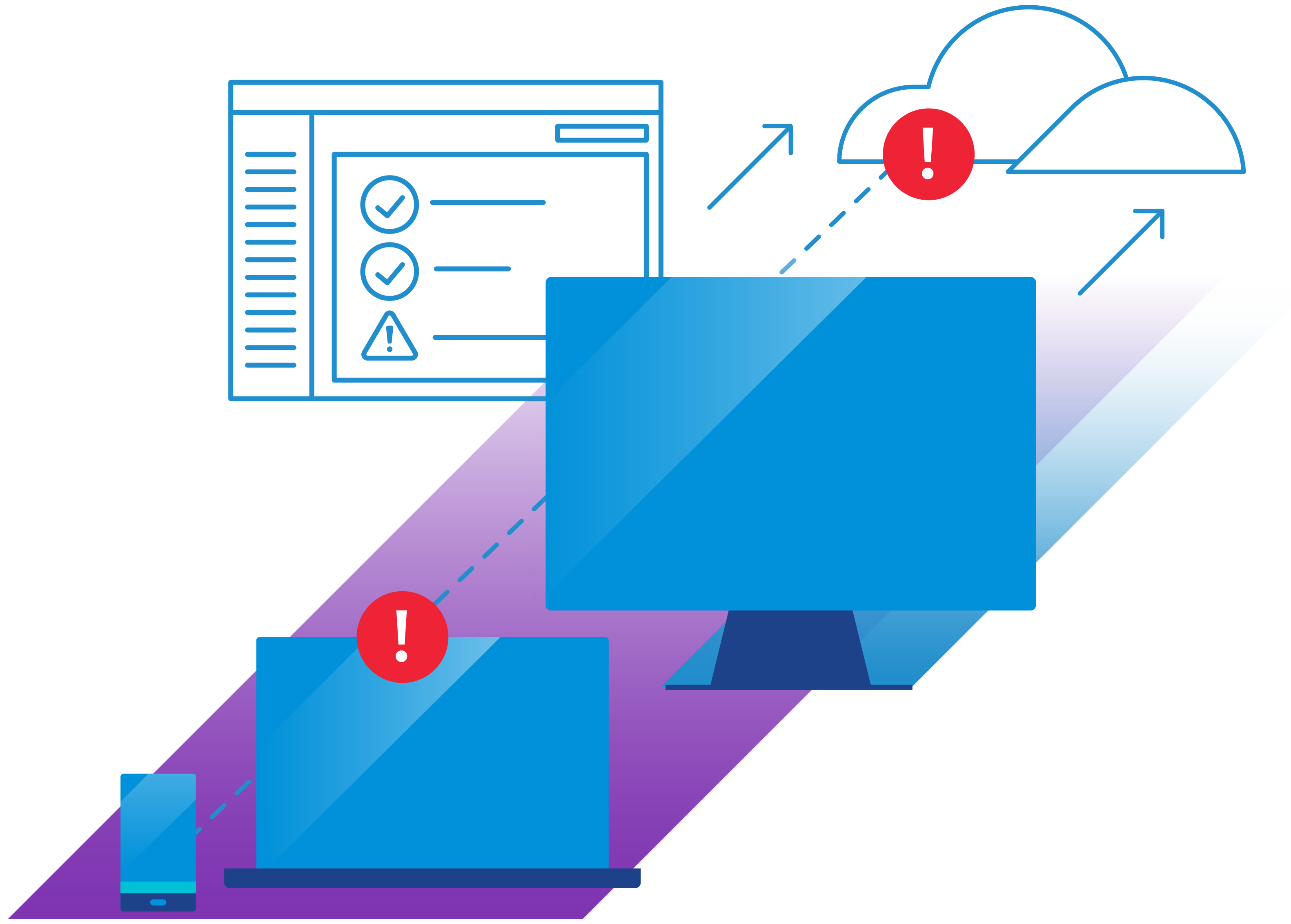
If any Amazon Elastic Compute Cloud (EC2) instance is untagged, alert its owner via email and stop the instance.

Identify non-conforming assets

In any organization, there are asset types and configurations that are not preferred or not allowed. Organizations have many reasons to set a policy preventing certain assets from running: the organization might receive special discounts on certain instance types or have decided that certain instance types are too costly. An organization might want to prevent users from launching instances in certain regions, such as China, for example, for security reasons. Launching an older AMI type or OS could lead to security or interoperability challenges. Whether it is certain instance types, regions, AMI types, operating systems, or network types, it's critical that you can quickly identify these and take action to correct them.

Sample non-conforming asset policies

If any X1 instance type is launched, send through the approval workflow first. If any instance is launched in a non-conforming region, send through the approval workflow first.



2. Financial management

To keep costs under control, the best practice is to implement financial management policies that will help identify which lines of business, cost centers, or projects are accountable for driving up costs, and will alert you when costs unexpectedly spike. Financial management policies primarily focus on budget and cost trend monitoring.

Budget

Creating a budget is easy. Staying within that budget, not so much. To help departments and lines of business stay within the allocated budget, set a policy to alert budget owners when projected spend is greater than their set budget. You may also want to set additional policies that send alerts when overall spend is nearing budget limits. These policies can help departments track their actual spend compared to allocated budget and avoid unpleasant surprises at the end of the quarter or year.

Cost trend

Closely related to budget policies, cost trend policies look for unexpected cost increases. You can have greater control over your costs by benchmarking the cost of each asset type month over month and identifying variances by business group. You can get extremely granular with this policy type. For example, get an alert when costs grow by more than 10 percent in a given month for your production assets. Or you can take a broader approach, such as getting an alert when the total cost of any department increases by 20 percent.

Sample budget policies

If projected month to date cloud spend is greater than 100 percent of the budget, then send an email notification to the budget owner.

If total spend reaches 85 percent of the budget for a given month, then send an email notification to the budget owner.

Sample cost trend policy

If total Amazon Simple Storage Service (S3) costs increase by more than 10 percent in 1 day, send an alert to the owner.

3. Cost optimization

While financial management policies are critical for keeping pace with budgets and trends, they don't help you optimize and reduce costs on their own. To achieve this, you need to create policies that will help you proactively reduce and optimize costs in your cloud environment.

In AWS, one of the most effective ways to reduce costs is to purchase Reserved Instances (RIs). Therefore, many of the best practice policies for cost optimization focus on managing and automating the full lifecycle of RIs.

Sample RI opportunity policy

If an instance is running On-Demand for more than 550 hours over the course of a month, send an email alert (potential RI purchase).

Identify commitment-based discounts

Every leading cloud provider offers flexible pricing and discount structures that help organizations optimize rates while using their services. Purchasing these commitment-based discounts is one of the most effective ways to optimize your cloud spend. In the simplest of terms, if you commit to using a certain amount of a cloud provider's services or to spending a certain amount of money on services, the cloud provider will give you a discounted rate on those services.

AWS Reserved Instances

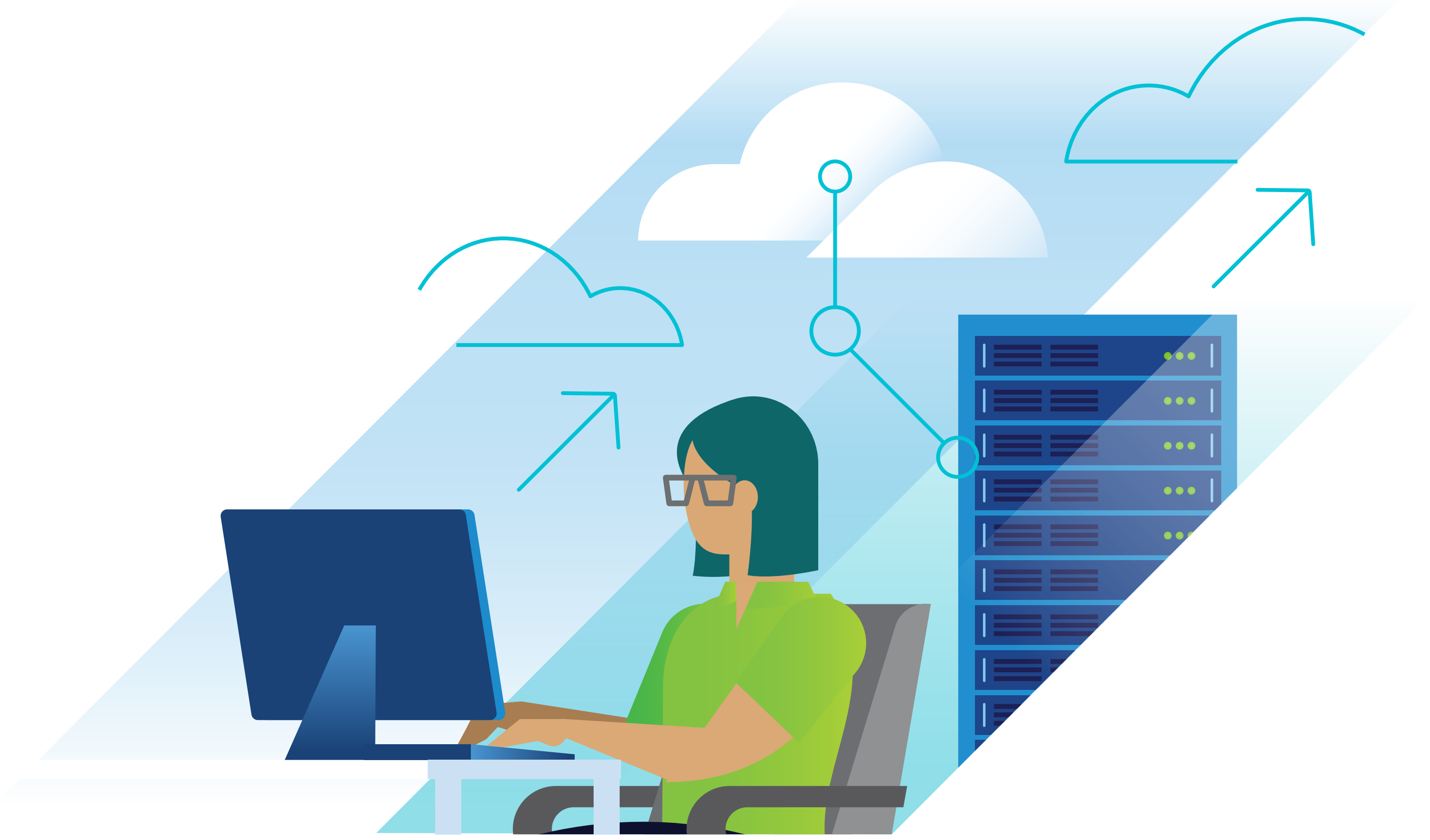
Amazon Web Services (AWS) customers can purchase reservations for a variety of services, including Amazon Relational Database Service (RDS), Amazon ElastiCache, Amazon OpenSearch Service, Amazon Redshift, and Amazon DynamoDB. Flexible payment terms are also available, with greater discounts on three-year commitments (compared to one-year commitments) and when more money is paid upfront. AWS offers all upfront, partial upfront, and no upfront payment options.

AWS Savings Plans

AWS Savings Plans provide more flexibility while offering the same cost savings as RIs. Currently, AWS Savings Plans are only available for compute services (including EC2, AWS Fargate, and AWS Lambda), they can't be sold on a marketplace if unused, and they provide no capacity guarantees. So, while Savings Plans have become extremely popular, RIs aren't going away any time soon.

Azure Reservations

Microsoft Azure also offers commitment-based discounts called Azure Reservations. Azure Reserved Virtual Machine (VM) Instances have an estimated cost savings of up to 72 percent (80 percent when combined with the Azure Hybrid Benefit). Azure RIs are available with a Microsoft Enterprise Agreement or pay as you go. They can be paid for in full upfront or in monthly installments. Azure Reservations can cover a single resource group, a single subscription, or shared. Azure Reservations can be modified, and you can return reservations at any time during the term for an adjusted refund. Microsoft also offers reservations for Azure Storage, Azure Cosmos DB, Azure SQL Database, MySQL servers, PostgreSQL, and Redis cache. VMware Tanzu CloudHealth® can provide asset reporting on all reservable services for Azure as well as optimizations.



Google Cloud Platform committed and sustained use discounts

Google Cloud Platform (GCP) committed use discounts (CUDs) and sustained use discounts (SUDs) offer similar deals to those who not only commit to a certain level of spend or usage in advance, but also those who are able to continuously use GCP services. CUDs can be gained through commitments to a specific number of resources or money spent. Resource-based CUDs are only available for Google Compute Engine (GCE) and can earn discounts of up to 70 percent for memory-optimized machine types. Spend-based CUDs generally increase with longer time commitments and can be purchased for several services, including Google Cloud SQL, Google Cloud VMware Engine, Google Cloud Run, and Google Kubernetes Engine. CUDs are not to be confused with SUDs, which are automatic discounts calculated by Google and received on incremental usage after continuously running GCE resources for a significant portion of a billing month. While these only apply if you have already achieved a required level of sustained use, they can also apply to overages on your CUDs.

Exchange and optimize purchases

It's not enough to simply utilize commitment-based discounts, you must also keep them optimized. For example, AWS allows customers to modify Standard RIs in the following ways:

- Switching between regional and Availability Zone scopes
- Switching between Availability Zones for reservations scoped to a specific zone within the same region
- Switching between Classic EC2 and Virtual Private Cloud
- Altering the instance size within the same family

You can make modifications through the AWS console, directly through the API, or automatically with a cloud service management solution. Because modifications are free, mature organizations continuously look for modifications to maximize their ROI from RI purchases.

Sample RI modification policy

If potential RI reallocation savings exceed $10, then modify the Reserved Instance.

4. Performance management

Understanding and monitoring performance in your environment is critical but not always easy. It's important to consider core utilization metrics, such as CPU, memory, disk, and network in/out, which can be gathered from performance monitoring tools. Using these trended metrics over time, you can gain information on whether instances and volumes are sized properly and performing as expected. It is a best practice to have predefined thresholds for what constitutes normal behavior for your infrastructure. For example, if CPU is less than 20 percent, then you deem that asset as underutilized. Underutilized assets should be downgraded for cost efficiency, while overutilized assets should be upgraded to avoid performance headaches.

Rightsizing underutilized assets

It's common for developers to spin up new instances that are substantially larger than necessary. This may be intentional to give themselves extra headroom during production or accidental because they don't know the performance requirements of the new workload yet. Over-provisioning can lead to exponentially higher costs, so it's critical to set up policies that will notify you when an asset is over-provisioned.

For example, the critical factors to consider with Amazon Elastic Block Store (EBS) volumes are capacity, IOPS and throughput. AWS offers several types of EBS volumes, from Cold HDDs to Provisioned IOPS SSDs, each with their own set of pricing and performance. You can find candidates for downgrading by analyzing the read/writes on all volumes. If a volume has hardly any read/writes, it is either attached to a zombie instance or the volume is unnecessary. Many organizations find they've deployed General Purpose SSD or Provisioned IOPS SSD volumes that have hardly any read/writes for a long period of time. They can be downgraded to Throughput Optimized HDD or even Cold HDD volumes to reduce costs.

Sample underutilized instance policies

If any Amazon RDS instance has an average read throughput, write throughout, and swap usage of less than 35 percent for more than two weeks, then send an email notification (potential downgrade).

If any io1 volume type average reads are less than 10,000 and average writes are less than 10,000 for one week, then send an email notification (potential downgrade).

Rightsizing overutilized assets

Rightsizing is important for identifying cost savings in underutilized assets and critical for identifying overutilized assets that can be impacting performance and causing a poor experience for the user. Overutilization policies will look similar to the underutilization policies, except with different thresholds and more inclusive clauses: instead of “and” clauses, use “or” to find any area of resource constraint.

Sample overutilized instance policies

If any gp2 volume type throughput averages more than 150 MiB/s for one week, then send an email notification (potential upgrade).

If any S3 Infrequent Access Object was retrieved more than three times in the past 30 days, send an email alert (potential migration to S3).

5. Operational governance

Automating basic operational tasks is one of the best ways to reclaim time to focus on more strategic business projects. Whether automating the detection and elimination of zombies or unused infrastructure, flagging older instance types, or scheduling environments to turn off and on again, these policies can yield significant time savings.

Identify and terminate zombie infrastructure

Zombie assets are infrastructure components running in your cloud that aren't in use. For example, an EC2 instance was used for a project that has ended, but the instance was never turned off. Zombie assets can also come in the form of an unattached elastic IP, an empty Amazon Elastic Load Balancer (ELB), or an idle RDS instance. No matter the case, AWS will charge for them as long as these assets are in a running state. They must be isolated, evaluated and immediately terminated if deemed nonessential. It is recommended that you take a snapshot, or a point-in-time copy, of the asset before terminating or stopping it to ensure you can recover it, if needed.

Instance scheduling

The most cost efficient environments dynamically stop and start instances based on a set schedule. Each cluster of instances can be treated a different way. These types of lights-on/lights-off policies can often be even more cost effective than commitment-based discounts, so it's crucial to analyze where this type of policy can be implemented.

Sample zombie termination policies

If an EBS volume is unattached for one week, then trigger a snapshot, delete the volume, and send an email notification.

If a snapshot is older than two months, send an email notification and delete.

If an elastic IP is unattached for two weeks, then send an email notification and release the elastic IP.

Sample instance scheduling policy

Stop Development EC2 instances at 7 PM on Friday, start Development at 6 AM on Monday.

6. Security and incident management

In a rapidly evolving cloud environment, it is important to keep up with changes that might impact your security posture. The best way to do this is with automated security policies, which can monitor for issues and flag them before they become catastrophic.

Access control

Cloud security starts with users and access controls. Without proper access controls and identity management, users can intentionally or unintentionally create security flaws with serious implications. Set policies to validate you have properly and securely configured access to your cloud and to help you stay ahead of breaches by monitoring for leading indicators, such as:

- Misconfigured users (i.e., users not in a group)
- Users with too broad a span of control (i.e., root accounts enabled for API access, too broad privileges, etc.)
- Users with vulnerable accounts (i.e., not compliant with password policies, identity and access management [IAM] user access keys in need of rotation, multifactor authentication disabled, etc.)
- Inactive users (i.e., IAM user with access keys not being used, etc.)

While it's always best to proactively catch security vulnerabilities before they are exploited, it's prudent to also monitor for events that can turn into security incidents, or lagging indicators, such as:

- Suspicious activity (i.e., a large volume of instances launched outside normal usage patterns, a new IP address for logging in to IAM user accounts, etc.)
- Changes to security groups or users (i.e., a new IAM group or user recently created or changed, root account recently used, etc.)

Sample access control policy

If any accounts have root account API access, then send an email notification and execute a Lambda function to revoke user access.

Conclusion

It's important to remember that these best practices are meant to be ongoing processes, not one-time activities. Because of the dynamic and ever-changing nature of the cloud, governance policies should ideally be automated so they can take place continuously. It's also critical to periodically revisit policies to ensure they still make sense for your organization.

Learn more about how Tanzu CloudHealth can help you implement and automate governance policies across your environment by visiting tanzu.vmware.com/cloudhealth.

Get Started Today

Find out how to manage your environment with Tanzu CloudHealth.

LEARN MORE

Join us online: