OverWatch SOLUTIONS

Spot by NetApp

Guide

# Azure cost optimization guide

Microsoft Azure

# Table of contents

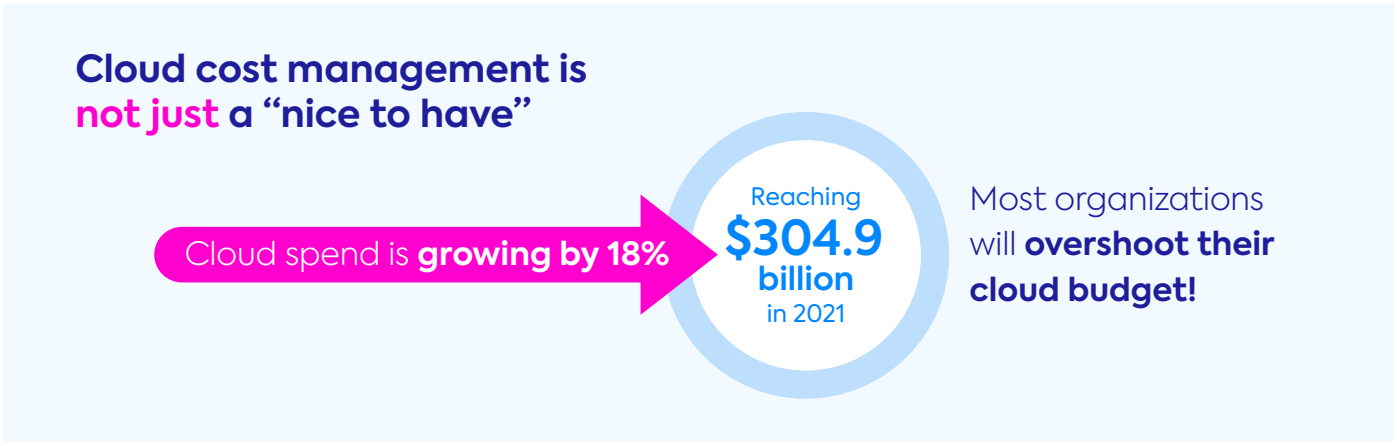
# Introduction

## Why Azure cloud cost optimization is important

Although it is no surprise that Gartner is forecasting overall public cloud consumption growing by 18% and hitting $304.9 billion in 2021, Gartner's research also indicates that the vast majority of organizations overshoot their cloud infrastructure budgets due to a lack of cost optimization approaches.[1]

With the global business climate frequently subject to change, companies that are reliant on public cloud infrastructure might consider that keeping cloud spend within budget is more than just a "nice to have" and could even be mission-critical for some.

Based on our experiences working with over 1,500 organizations, this guide focuses both on best practices for setting up Microsoft Azure cloud governance as well as hands-on cloud cost management, so you can fully control your Azure infrastructure costs, both now and into the future.




1. "How to Identify Solutions for Managing Costs in Public Cloud IaaS", 19 August 2019, Gartner, Inc.

# Cloud governance

## Starting with the right framework

To create financial accountability and efficiency in the cloud, ideally the entire company should be involved. However, all the following can be implemented within individual business units and teams and with success, expanded company wide.

## Here are the key steps to follow:

1

### Understanding your organization's needs

Before getting started, it's imperative to get a clear view of your organizational landscape with perspective on the goals of individual business units and teams. The following activities are a good place to begin:

- Interviewing engineering teams to understand their current, near-term and future projects to anticipate their needs for cloud resources
- Reviewing engineering's projects with your DevOps, SREs, and TechOps teams to verify requirements for cloud architecture, associated infrastructure and budget.
- Verifying with the Finance team that the requested cloud spend is in line with available budget.

Along the way, these interviews will give you familiarity with the various projects and organizational hierarchy which will be essential for tagging all cloud resources, which itself is a prerequisite for accurate spend visibility and cost allocation.

## 2 Creating cross-organizational alignment

As cloud cost management will impact multiple teams, all relevant stakeholders must understand and commit to the process. With your initial round of interviews (see previous section) uncovering what is needed in terms of cloud resources, a potential next step is to correlate your business KPIs (e.g., desired cost per customer, product, etc.) to projected cloud spend, and then confirm that these are well aligned. This sort of analysis will make it easier to explain cost management activities and gain cross-team and executive buy-in which is so critical for implementing strong cloud cost governance.

## 3 Mandatory tagging and cloud spend monitoring

Once you know which tags should be applied to deployed resources, a mandatory tagging strategy combined with real governance should be established. Ensuring adherence can be owned by any relevant persons or teams such as DevOps, FinOps, Engineering, etc.

Open-source tools such as Cloud Custodian offer a comprehensive library of scripts for reporting on and enforcing compliance with fundamental aspects of Azure cloud management such as new resource labeling or tagging.

Once proper tagging is in place, the next step is to gain comprehensive and unified reporting of your cloud spend - all based on the business logic of your own company. This can be done by leveraging Azure's native, cost management tool as well as third-party solutions such as Spot by NetApp's Cloud Analyzer which provides not only comprehensive cost analysis, but also cost optimization recommendations that can be implemented in just a few clicks.

## 4 Assessing and prioritizing savings opportunities

Mapping your cloud infrastructure footprint can leave you with an overwhelming amount of information. To identify the best places to reduce costs, it is important to balance the potential savings against the complexity and resources required to realize those savings. A systematic approach that assesses compute, storage, and network infrastructure is crucial. Modern tools that calculate potential savings, not just summarize current costs, are important to leverage in this process.

For example, Spot by NetApp's Cloud Analyzer offers a complete analysis of your cloud spend as well as actionable recommendations for cloud cost reduction such as right-sizing cloud resources and guidance on when and where to use reserved or spot instances for different workloads.

## 5 Establishing ongoing processes that reduce costs today and in the future

One-time cost reductions are often the first step taken to save money, however even more important is ensuring that cloud infrastructure costs are kept under control in the future as well. Without that, inefficiency is almost certain to increase again until the next cost reduction project.

Creating cost alerts to identify anomalous spend as well as frequent reviews of cloud spend, and optimization recommendations provided by cost management tools are examples of cloud cost management best practices that can be easily implemented.

**Elastigroup saves you up to 90% with zero overhead**

Cloud cost optimization tools such as Spot by NetApp's Elastigroup allow you to continuously and reliably run production and mission-critical workloads on **Azure's highly affordable spot VMs.**

# Taking action

With comprehensive governance and monitoring of Azure cloud spend in place, as soon as cloud cost overages are identified, the following sections, which outline practical methods for significant cloud cost reduction, can be applied, whether separately or all together.

## Here are key ways to quickly reduce Azure costs:

### 1 Eliminating cloud waste

The ease of provisioning Azure resources is a double-edged sword, especially in large teams without comprehensive governance controls. Not only is it all too easy to provision resources and then leave them running even when no longer needed, errors in deployment templates can also generate unneeded resources or leave them orphaned after post-job cleanup. Here are some areas to look out for:

- **Idle and underutilized resources**

  Once you have a defined threshold for what is considered idle or underutilized behavior, keep an eye out on Azure VMs (virtual machines), Azure SQL (relational databases), Azure SQL Data Warehouse and Azure Scale Sets (scaling groups) to make sure they are sufficiently active. If they are not, they should be considered for either right-sizing or termination.

- **Orphaned resources**

  Orphaned resources are the result of terminating a VM but leaving previously attached resources running. Make sure that unattached Azure Virtual Disks (volumes), unassociated Static Public IPs and orphaned Azure Block Blobs (object storage) are terminated or reassigned.

To identify unused resources, Azure and 3rd party vendors offer monitoring solutions that look at usage telemetry such as network traffic, CPU load, memory and similar data points to identify resources that are no longer active. Even better are tools that deploy advanced analytics and automation to monitor resources continuously, and automatically shut down those that are no longer in use.

## 2 Optimizing use of Azure's pricing models

Utilizing Azure's discounted pricing models for compute instances can potentially deliver the most significant cost savings. However, familiarity with all the various options is essential for successful cost optimization. The following sections provide comprehensive details on Azure's three main pricing models:

1. **Pay-as-you-go**

   This option allows you to spin VMs up and down at will, only paying for what you actually use, by the second, without any commitment. However, for longer-term usage, this non-discounted pricing model can be prohibitively expensive.

2. **Reserved Virtual Machines (RIs)**

   This pricing model delivers up to 72% cost discounts compared to pay-as-you-go pricing in exchange for either a 1- or 3-year commitment. While the discount is attractive, a financial commitment of this length might not be ideal for dynamic environments. For those with longer-term projects and resource requirements, see the "Must-know Azure RI concepts" section below for informed Azure RI planning and management.

3. **Azure Spot VMs**

   All public cloud providers need to have some excess compute capacity, commonly referred to as spot instances, to service any surges in demand. To offset the financial loss of idle resources, Azure offers access to this excess compute capacity at greatly reduced pricing — sometimes up to 90% less than pay-as-you-go rates. However, pricing for spot instances fluctuates based on supply and demand with the result being that workloads can be evicted any time, and in the case of Azure Spot VMs, with just a 30 second notification.

   While interruptions of this nature pose a challenge for workloads requiring high availability, companies like Spot by NetApp allow customers to enjoy the dramatic discounts while ensuring SLA-backed availability for even mission-critical and production workloads. Spot by NetApp is able to predict Azure's evictions, gracefully drain the workload and move it to another spot or pay-as-you-go VM, thereby avoiding any interruption or performance impact. Read more here.

## Must-know Azure RI concepts

### Azure account and subscription

For those familiar with Amazon Web Services, an Azure account is similar to AWS's master payer account and Azure subscriptions are similar to AWS's sub-accounts. The Azure account is used for overall billing purposes, but each subscription generates its own set of billing data that is helpful for cost allocation purposes. Each subscription can also be used for access control and isolating environments.

### Prioritized capacity

For reserved instances that have a "single" scope you can select "Capacity priority" which as the name implies, prioritizes data center capacity for your deployments. This option is similar to AWS's zonal reserved instances.

### Services covered

Azure provides a very detailed list of where reservations can be applied. Aside from Azure VMs, some of the services covered include Blob storage capacity, the compute component of Azure Database for MariaDB, MySQL, and PostgreSQL, as well as quite a few other Azure services.

### Flexibility

When buying an Azure RI you can choose "instance size flexibility" ensuring that the RI will be applied to any size VM in the same instance size flexibility group. For example, if you bought an RI for a Standard_DS4_v2 with a footprint of 8 but you actually ran two Standard_DS2_v2 sized VMs with a footprint of 2 each and a Standard_DS3_v2 sized VM with a footprint of 4, your initial RI purchase would fully cover what you actually used.

### Exchanging or cancelling reservations

In the event that you no longer need the reserved instance you purchased you may exchange it for a different instance family, region, etc. You currently can also cancel up to $50,000 worth of reservations without penalty.

### Scope

Reserved instances can be either "single" and applied to VMs (or other services) in a specific subscription, or they can be "shared" and applied to reservations in any other subscriptions that you are the owner of and that share the same billing context as the reservation.

## 3 B-Series VMs for burstable workloads

Many cloud workloads such as web servers, small databases, and development and test environments, often experience bursty CPU performance. These workloads will run for a long time using a small fraction of available CPU and then when traffic or processing increases, will quickly begin using all the CPU power in the VM. While most of the time these VMs are greatly underutilized (and hence not cost-effective), they cannot be easily downsized as there are times when full CPU capacity is required.

To address this, Azure offers their B-Series burstable VMs. When not being used at a defined baseline of usage, these VMs build up credit so when full capacity is required, the cost is discounted, anywhere between 15-55% in comparison to equivalent VMs. However, if the VM becomes too CPU-intensive, it'll be throttled down to the defined baseline performance until enough credits are available. Therefore, some analysis is required to determine if your workloads are sufficiently low-use to benefit from Azure's B-Series.

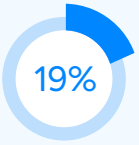
## 4 Azure region selection

While some VM workloads need to run in a very specific region to properly service customers, other workloads have more flexibility in terms of location. For this latter group, certain regions, even within the same geographic area, may offer significant cost savings.

### Azure region pricing tip

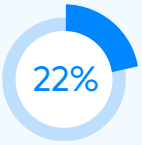
**Significantly reduce cost** by running workloads in a different Azure region within the same geographic area!

| | Instance | Azure region | Pay-as-you-go |
|---|---|---|---|
| **US West** | **DS2 v2** | West US | $0.14/hour |
| | | West US 2 | $0.114/hour |



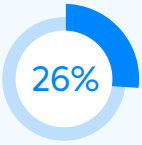

Regional cost savings

| | Instance | Azure region | Pay-as-you-go |
|---|---|---|---|
| **US East** | **DS2 v2** | East US | $0.146/hour |
| | | East US 2 | $0.114/hour |




Regional cost savings

| | Instance | Azure region | Pay-as-you-go |
|---|---|---|---|
| **Asia Pacific** | **DS2 v2** | East Asia | $0.214/hour |
| | | Southeast Asia | $0.158/hour |




Regional cost savings

**Note:** For additional examples of regional pricing differences see Azure's Linux Virtual Machine Pricing

## 5 Storage optimization

Azure offers several types of storage services including:

- Azure Blob for unstructured data storage
- Azure Files  for managed file sharing
- Azure Queues for reliable messaging between application components
- Azure Tables for schema-less storage of structured data with NoSQL
- Azure Disks (block-level storage volumes) for Azure VMs

**Here are some cost reduction best practices for each type of storage:**

### 1. Azure Blob and Azure Files

For both of these storage types, tiering can be employed to significantly reduce costs for data that doesn't need regular access.

Here is an example of cost, accessibility and performance by tiers for Azure Blob:

| Azure Blob | | | |
|---|---|---|---|
| **Storage tier** | **Price** | **Retrieval performance** | **Minimum storage period** |
| Hot | $0.0184 per GB | Milliseconds with 99.9% availability | NA |
| Cool | $0.01 per GB | Milliseconds with 99% availability | 30 days |
| Archive | $0.00099 per GB | Up to 15 hours | 180 days |

**Note:** Price listed is just for first 50 terabytes of data per month. Rates may vary by region.

While storage costs for the Archive tier is significantly lower, reading data can be a costly activity at $5 for every 10,000 read operations.

In situations where your data access patterns are unclear, best practice is to start with the hot storage tier, monitor for a while, and then move the data to cool storage or archive storage tier based on access frequency.

Azure Storage Reserved Capacity can further lower costs for all three tiers. For example, 100 terabytes of the Cool storage tier costs $1,000 per month. If you know you will need it, committing to a one-year reservation reduces cost by 16% down to $840 per month.

Similar to Azure Blob, using the Cool tier for Azure Files can reduce cost by up to 40% in comparison to the Hot tier. Here too monitoring how often data is being accessed can guide your decision of which tier to use.

2. **Azure Queues and Azure Tables**

   Both Azure Queues and Azure Tables offer different levels of storage redundancy. Options range from locally redundant storage (LRS) where your data is replicated three times within a single physical location in a single region, with at least 99.999999999% (11 nines) durability, all the way to Geo-zone-redundant storage (GZRS) where your data is copied across three availability zones in the primary region and is also replicated to a secondary geographic region with at least 99.99999999999999% (16 nines) durability of objects over a given year.

   Using reduced redundancy levels for non-critical data can immediately bring down storage costs. For example, unstructured and semi-structured data storage in Azure Tables with basic redundancy can be 65% less expensive than the highest level of redundancy.

3. **Azure Disks**

   In addition to seeking out and terminating orphaned disks which are no longer attached to a VM, monitoring the utilization of your disks can help determine if you can use smaller disk sizes with lower IOPS and throughput. Additionally, carefully consider whether your workloads require HDD, SSD and the various premium disk options that Azure offers. All of these will have significant impact on cost.

4. **Moving from SQL Server instances to the Azure SQL service**

   Moving from Database VMs to Elastic Databases Running SQL Server or other database servers on Azure can get expensive quickly. VMs themselves are expensive and you will often find that database instances are under-utilized, and it is not straightforward to distribute loads between instances. In many cases, transitioning to a PaaS model, for example moving from SQL Server instances to the Azure SQL service, will result in a dramatic reduction of cost, because you are paying only for actual database resources used.

## 6 Networking optimization

Cloud computing platforms vary significantly in their pricing for data transfer, however costs associated with data transfer in the cloud can quickly become significant. Without needing to rearchitect applications, these steps can help bring down those costs.

### Reduce traffic across zones and regions

Network traffic between datacenters can be an overlooked but significant contributor to cloud costs, not only for traffic between availability zones but even more for traffic between regions. Although the need to ensure redundancy and resiliency is one reason that creates cross-datacenter traffic, the ease of provisioning services in multiple availability zones can also make it easy to inadvertently create significant amounts of unnecessary traffic between regions or between availability zones. The data transfer costs that result can quickly add up, especially for "chatty" workloads as well as for applications that consist of large numbers of distributed services.

**Starting July 1, 2021** Azure will be charging for data transfer, ingress and egress, from a VNet resource deployed in an availability zone to another resource in different availability zone in the same VNET.

However, data transfer between VNet resources located in same availability zone as well as between a VNet resource and a Public IP address in the same Azure region will not be charged. Likewise, data transfer between VNet resources located in peered VNets across availability zones will not be charged. However, this data transfer will be charged as per VNet peering rates.

As such, rebalancing services across zones to minimize data transfers across regions and zones provides one way to reduce these costs. Done thoughtfully, rebalancing can significantly reduce cross-datacenter communication without compromising resiliency and can even help improve performance. Network tracing and logging tools can also help, bringing to light misconfigurations that are creating unnecessary cross-datacenter traffic that can be quickly reduced.

### Optimize network configurations

Not only does the amount of network data transfer impact costs, so does the way in which networks are configured. Depending on how network traffic is routed, data transfer costs can vary widely and grow quickly with usage.

Modifying network configuration can have a significant impact on data transfer costs without affecting available throughput. For example, choosing private IP addresses rather than public or elastic IP addresses where possible can have a big impact on data transfer costs.

### Deploy distribution and caching solutions

Many applications and application components make requests for the same data from other cloud services or from remote repositories. Particularly for media objects and for transfers of large data sets for data processing and analytics, the repeated data transfers created by these requests can drive up costs.

Deploying content distribution networks and caching services can help reduce these costs by reducing repeated transfers from remote services and locations. These solutions do come with their own costs, making it important to assess the potential savings from deploying them.

## 7 Special pricing and cost management tools

Azure offers additional pricing models for both enterprises and startups that can help with lowering costs.

- Azure Credits provide new users with $200 free credit and various free services for up to 12 months.
- Enterprise Agreement is intended for larger organizations (500 users or more) and provides pricing discounts and other benefits.
- Azure Dev/Test pricing provides subscribers of Visual Studio as well as Enterprise Agreement customers with monthly credits and lowers rates for various Azure services and VMs.
- Hybrid Benefit provides users with existing Windows Server and SQL Server on-premises licenses, up to 85% savings over the standard pay-as-you-go rate when bringing these licenses to Azure.

# Continuous Azure optimization

## Reducing Azure costs with Spot by NetApp

Taking steps to reduce and control cloud infrastructure costs requires ongoing effort and diligence that could all too easily consume significant amounts of time for already overstretched CloudOps teams.

To help address that challenge, Spot has created a suite of products built on unique machine learning and analytics, going beyond cloud visibility and guidance and actually automating Azure workloads on an optimal blend of spot VMs, reserved instances and if needed, pay-as-you-go pricing. All of this is provided without compromising availability, performance and flexibility, ensuring optimized infrastructure for containers, Kubernetes, autoscaling applications and more.

**Learn more about Spot by NetApp's products and solutions for Azure at spot.io/solutions/msft-azure/**