solliance

COST OPTIMIZATION CONSIDERATIONS FOR LINUX ON AZURE

White Paper

Sponsored by Microsoft Azure and AMD

Written by Reid Patrick

Microsoft Azure AMD

About Solliance

Solliance delivers expert led end-to-end technology solutions—from strategic planning to architecture design and implementation—for a diverse range of customers, including Fortune 500 companies and start-ups. Our global team of more than 300 top tier industry experts spans 18 countries, features 30 Microsoft MVPs and averages 20 years of experience in Cloud, AI & Data, Infrastructure & DevOps, Security, Applications, Microservices, Advanced Computing, and Training.

Recognized as one of Microsoft's top global partners, Solliance was awarded the exclusive Advanced Specialization in AI and Machine Learning on Microsoft Azure. Our team also regularly leads advanced 300 & 400 level trainings for the Microsoft field.

Our fast growing AI & Data practice brings unparalleled expertise in Artificial Intelligence, Cloud Analytics, Machine Learning, Natural Language Processing and Large Language Models. Years before the excitement around OpenAI, Solliance was using generative AI technologies to build real-world, innovative applications.

About the Author

Reid Patrick is a 15-year veteran at Microsoft and a nine-time Microsoft Azure MVP for Networking. He has extensive DevOps experience using Terraform, Bicep, PowerShell, and Bash, and is highly skilled in Cloud Architecture, Microservices, DevOps, Automation, Hybrid Networking, Infrastructure Modernization, Migrations, and Disaster Recovery.

With a career spanning over three decades, Reid has worked with at-scale internet deployments using Linux since 1995, building some of the largest websites in the world. He is also highly experienced in leveraging Kubernetes for deployments on Azure, solidifying his expertise in modern cloud-native practices. Currently, Reid is focused on DevOps for AI solutions using Azure OpenAI and FoundationaLLM (https://foundationallm.ai).

Reid is the Chief Infrastructure Architect at Solliance. He has a distinguished background in IT Infrastructure and Operations, having architected and led teams supporting some of the largest service providers in North America, managing environments with as many as 15,000 Windows Servers and 120 million endpoints.

As a published author, Reid contributed to the Networking, Azure Active Directory, and Containers sections of the 70-533 Exam Reference for Microsoft Press. Additionally, he has written nearly 250 pages of the Cloud Adoption Framework for Microsoft Azure, further solidifying his thought leadership in the field.

CONTENTS

Executive Summary

In recent years, Linux has emerged as a critical operating system for cloud deployments due to its flexibility, open-source nature, and widespread community support. Microsoft's cloud platform Azure has recognized this trend and offers robust support for various Linux distributions running on Infrastructure as a Service (IaaS). However, while Azure provides robust infrastructure and tools for Linux environments, managing and optimizing costs is top of mind and an enduring priority for most customers.

This whitepaper focuses on three key categories that can help optimize cloud consumption costs: **Azure Services, Pricing Models and Offers, and Cost Analysis Tools.** The intent is to equip IT professionals, system administrators, and decision-makers with the knowledge to effectively manage the cloud consumption costs of Linux workloads running in Azure.

Azure Services: Microsoft offers various virtual machines (VM) and storage services tailored to different Linux distributions and workloads. This section explores the various VM and storage options, detailing their features, performance capabilities, and best use cases. By understanding these services, readers can make informed decisions about the most suitable VMs and storage for their specific requirements, optimizing performance and cost. Additionally, there are advanced techniques that administrators can take to ensure costs remain at an expected level for deployments.

Pricing Models and Offers: Azure provides several pricing models and offers that provide flexibility and cost savings. This section examines the different pricing options, including different subscription types. It also highlights Azure's unique offers, such as Azure Hybrid Benefit, Azure savings plan for compute, and Reserved Instances, which can further reduce costs. Understanding these models and programs will help readers design cost-effective architectures and make strategic decisions to minimize expenses.

Cost Analysis Tools: Effective cost management requires robust monitoring and analysis tools. Azure offers a variety of cost analysis tools, such as the pricing calculator and Microsoft Cost Management, that provide real-time insights into spending patterns and potential savings opportunities. This section delves into these tools, explaining how to set up cost alerts, generate detailed reports, and use advanced analytics to identify and mitigate unnecessary expenses. By leveraging these tools, readers can maintain continuous control over their expenditures and ensure their Linux environments on Azure remain cost-efficient.

By exploring these three categories, this whitepaper provides a comprehensive guide to managing Linux environments on Azure, focusing on optimizing costs while maintaining high performance and security.

Azure Services

When considering workloads for running Linux on Azure IaaS, it is essential to understand the various compute options available. Azure offers virtual machines (VMs) for various Linux workloads, Azure Virtual Machine Scale Sets for automatic scaling to meet demand, and Azure Spot Virtual Machines to run interruptible workloads cost-effectively. Several VM services provide various options to save, depending on the nature of the workload. These can include dynamically autoscaling VMs according to demand or provisioning spare Azure compute capacity at up to 90% discount versus pay-as-you-go rates[1].

These compute options provide flexibility in managing costs and resources, ensuring efficient handling of different workload requirements.

Azure also provides cost-effective storage solutions to support various workloads. Azure Disk Storage offers simplified management with high availability and durability, Azure Blob Storage provides scalable object storage for unstructured data, and Azure Files delivers fully managed file shares accessible via SMB and NFS protocols. Additionally, Azure Files for NetApp enables seamless integration with enterprise applications, offering high-performance and low-latency access to files while benefiting from the scalability and flexibility of the cloud. These storage options ensure robust, scalable, and flexible data management in the cloud.

Azure Virtual Machines

Azure offers a diverse range of VM families tailored to meet the unique demands of various workloads. Each VM family has specific features and performance characteristics, making it suitable for different workloads, from entry-level applications to high-performance computing and confidential data processing. Table 1 – Azure VM Families by workload, highlights key features and example workloads to select the best VM family for specific workload requirements.

Type	Description
Burstable	Ideal for workloads that do not need the full performance of the CPU continuously, like web servers, proof of concepts, small databases, and development build environments. These workloads typically have burstable performance requirements. These VMs are sized in line with the General-Purpose families.
General purpose	Balanced CPU-to-memory ratio. Ideal for testing and development, burstable workloads, small to medium databases, and low to medium traffic web servers
Compute optimized	High CPU-to-memory ratio. Suitable for medium-traffic web servers, network appliances, batch processes, and application servers.
Memory optimized	High memory-to-CPU ratio. Great for relational database servers, medium to large caches, and in-memory analytics.
Storage optimized	High disk throughput and IO ideal for Big Data, SQL, NoSQL databases, data warehousing and large transactional databases.
GPU	Specialized virtual machines targeted for heavy graphic rendering and video editing, as well as model training and inferencing (ND) with deep learning. Available with single or multiple GPUs.
High performance computing	Fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA).

Table 1 – Azure VM Families by workload

[1]Actual discounts may vary based on region, VM type, and Azure compute capacity available when the workload is deployed

Cost Optimization When Selecting VMs

Selecting the appropriate VM family based on the workload's requirements while maintaining optimal performance can significantly reduce costs. Utilizing the Azure Virtual Machine selector can further enhance this process, ensuring that selected VMs align perfectly with the organization's needs and budget constraints.

Optimizing costs when selecting VMs on Azure involves aligning the VM type with the specific needs of the workload. For example, burstable VMs are ideal for applications with intermittent performance needs, allowing cost savings by purchasing a VM size with a baseline utilization rate. These VMs build up credits when utilization is below the baseline, which can be used during periods of increased demand. General-purpose VMs offer a good balance of CPU and memory, making them suitable for a wide range of applications, including development and testing environments, web servers, and smaller databases. Compute-optimized VMs are best suited for high CPU performance needs, ensuring efficient resource usage and cost management.

For more specific requirements, memory-optimized VMs, with their high memory-to-CPU ratio, are designed for workloads like database servers and in-memory analytics, offering cost efficiency for large datasets. Storage-optimized VMs, with their high disk throughput, are tailored for big data and large transactional databases, enabling cost-effective data processing. GPU-enabled VMs provide the necessary power for specialized tasks such as graphic rendering, deep learning, and AI, optimizing costs by delivering targeted performance. High-performance computing VMs cater to the most demanding applications, offering the fastest CPUs and high-throughput network interfaces.

AMD EPYC™ VM options

Using the latest Azure virtual machines based on the latest AMD processors can help optimize the price-to-performance ratio for Linux workloads. The new v6 VMs based on 4th Gen AMD EPYC™ CPUs (Genoa) offer better price performance than the 3rd Gen AMD EPYC™ CPUs (Milan)-based v5 VMs designed to meet demanding performance requirements while balancing costs. On average, workloads running the new Dasv6 and Easv6 VMs can expect 20% CPU performance improvement over the Dasv5 and Easv5 VMs and better price performance. The new Falsv6, Fasv6, and Famsv6 VMs offer the fastest x86 CPU performance amongst same-segment VMs from major cloud providers. This VM series has up to 2x CPU performance improvement over the 3rd Gen AMD EPYC™ CPUs (Milan)-based v5 VMs. Table 2 – Azure VMs capture these options.

VM Family	vCPU	Memory (GiB)	VM Category	Workload Types
Dalsv6/Daldsv6 and Dasv6/Dadsv6-series	2-96	4-192 (Dalsv6/Dasldsv6) 8-384 (Dasv6/Dadsv6)	General Purpose	General applications including web servers, gaming, video encoding, AI/ML, and batch processing.
Easv6 and Eadsv6-series	2-96	16-672	Memory-intensive	Enterprise applications, data warehousing, business intelligence, in-memory analytics, and financial transactions.
Falsv6, Fasv6, and Famsv6 -series	2-64	4-128 (Falsv6) 8-256 (Fasv6) 16-512 (Famsv6	Compute Optimized	Enterprise and scientific applications requiring the highest CPU performance.

Table 2 – Azure VMs based on 4th Gen AMD EPYC™ CPUs (Genoa)

General Purpose: The Dasv6 and Dadsv6 series virtual machines, powered by AMD's 4th Generation EPYC™ 9004 processor in a multi-threaded configuration, offer up to 320 MB of L3 cache and are ideal for general-purpose workloads. These VMs balance CPU and memory, making them suitable for a wide range of applications. For workloads that don't require as much memory per vCPU, the Dalsv6 and Daldsv6 series offer a more affordable option, with 2 GiB of RAM per vCPU. These VM series are optimized for non-memory-intensive applications, providing significant cost savings while maintaining performance.

Memory-intensive workloads: The Easv6 and Eadsv6-series virtual machines harness the power of AMD's 4th Generation EPYC™ 9004 processors, achieving a boosted maximum frequency of 3.7GHz. These VMs provide configurations with up to 96 vCPUs and 672 GiB of RAM. The Eadsv6-series offers a balanced mix of vCPU, memory, and fast local NVMe temporary storage, making them ideal for memory-intensive enterprise applications. Furthermore, Eadsv6-series virtual machines support various disk types, including Standard SSD, Standard HDD, and Premium SSD.

Compute Optimized: The Falsv6, Fasv6, and Famsv6-series virtual machines leverage AMD's 4th Generation EPYC™ 9004 processor, which can reach a boosted maximum frequency of 3.7GHz and features up to 320 MB of L3 cache. These VM series are designed without Simultaneous Multithreading (SMT), ensuring that each virtual CPU (vCPU) is mapped to an entire physical core. This configuration allows software processes to run on dedicated and uncontested resources, making these full-core VMs ideal for workloads requiring the highest CPU performance.

Azure Spot Virtual Machines

Azure Spot Virtual Machines offer compute capacity at significantly reduced costs by utilizing currently unused compute resources. While workload eviction is possible, the compute capacity is available at a significantly reduced price, often up to 90% lower than standard rates. Spot Virtual Machines are ideal for interruptible and non-time-sensitive workloads, such as machine learning model training, financial modeling, or CI/CD processes.

Incorporating Spot Virtual Machines can be a crucial component of a cost-savings strategy. Azure provides significant pricing incentives to use any available spare capacity. Evaluating the potential use of Spot Virtual Machines for suitable Linux workloads can maximize cost savings. Understanding how Spot Virtual Machines work and assessing their suitability for specific workloads can help organizations leverage these cost-effective resources effectively.

Pricing and Management of Spot Virtual Machines: Spot Virtual Machines are priced according to demand. This flexible pricing model allows setting a price limit for the Spot Virtual Machines in use. If the market causes the price for Spot Virtual Machines to exceed a predetermined limit, workloads can be paused and resumed when demand decreases. When high utilization rates are anticipated in a particular region at specific times of day or month, selecting another region or planning for higher price limits may be advantageous. If the workload timing is flexible, setting a low-price limit ensures that workloads run only during periods when Spot capacity is the cheapest, minimizing the cost of Spot Virtual Machines.

When using Spot Virtual Machines with price limits, it is also essential to consider the different eviction types and policies. These options determine what happens to Spot Virtual Machines when a pay-as-you-go customer reclaims them. To maximize cost savings, the delete eviction policy is recommended first. This allows for faster redeployment of VMs, reducing downtime while waiting for Spot capacity, and eliminates the need to pay for disk storage. However, if a workload is region or size-specific and requires some level of persistent data in the event of an eviction, the deallocate policy is a better option.

Admins can further explore detailed best practices for building applications with Spot Virtual Machines to understand available options better.

Deployment Scenarios: There are various cases where Spot Virtual Machines can be highly advantageous. Workloads that are stateless, scalable, or time, location, and hardware-flexible are a good fit for Spot Virtual Machines. While Spot Virtual Machines can offer significant cost savings, they are unsuitable for workloads that require high availability, consistent performance, or long-running tasks. Examples of ideal workloads for Spot Virtual Machines include continuous Integration and continuous delivery (CI/CD), Financial Models, and Media Rendering:

- **CI/CD:** Pipelines are among the most accessible places to implement Spot Virtual Machines. The temporary nature of many development and test environments makes them well-suited for Spot Virtual Machines. Since the time difference between a few minutes to a few hours in testing an application is often not business-critical, deploying CI/CD workloads and build environments with Spot Virtual Machines can significantly reduce the cost of operating a CI/CD pipeline.
- **Financial Modeling:** Creating financial models is compute resource-intensive but often transient. Researchers frequently face challenges in testing all their hypotheses with rigid infrastructure. Spot Virtual Machines allow adding extra computing resources during periods of high demand without committing to a higher amount of dedicated VM resources, enabling the faster creation of more and better models.
- **Media Rendering:** Rendering jobs, such as video encoding and 3D modeling, require substantial compute resources but may not need these resources consistently throughout the day. These workloads are typically computationally similar, independent, and do not require immediate responses. These characteristics make media rendering another ideal use case for Spot Virtual Machines. For rendering infrastructure often at capacity, Spot Virtual Machines can add extra compute resources during high-demand periods without committing to more dedicated VM resources, thus lowering the total cost of ownership (TCO) of running a render farm.

Azure Virtual Machine Scale Sets

Azure Virtual Machine Scale Sets enable the management and deployment of groups of VMs at scale, offering a variety of load balancing, resource autoscaling, and resiliency features. While many of these features can indirectly save costs by simplifying deployments or enhancing high availability, specific features like autoscaling and Spot Mix contribute directly to cost reduction.

Autoscaling: Autoscaling is a critical feature within Virtual Machine Scale Sets, allowing for the dynamic increase or decrease of the number of virtual machines running within the scale set. This capability enables the infrastructure to scale out to meet demand when necessary and scale in when compute demand decreases, reducing unnecessary costs from extra VMs running without need.

VMs can be autoscaled according to self-defined rules based on various metrics. Host-based metrics, such as CPU usage and memory demand, or application-level metrics, like session counts and page load performance, trigger these rules. This flexibility allows for precise scaling requirements, helping to control infrastructure scaling to match compute demand optimally without extra overhead.

Scaling can be scheduled to anticipate cyclical changes in VM demand throughout certain times of the day, month, or year. For example, workloads can automatically scale out at the beginning of the workday when application usage increases and scale in overnight when usage decreases, minimizing resource costs. Similarly, scaling can be adjusted for specific events, such as holiday sales or marketing launches. Additionally, for more complex workloads, Virtual Machine Scale Sets offer the option to leverage machine learning to predictively autoscale workloads based on historical CPU usage patterns.

These autoscaling policies facilitate adapting infrastructure usage to various variables. Leveraging autoscale rules to fit application demand is essential for reducing costs.

Spot Mix in Virtual Machine Scale Sets: Spot Mix in Virtual Machine Scale Sets allows for the configuration of scaling policies to specify a ratio of standard to Spot Virtual Machines as VMs increase or decrease. For instance, with a 50% ratio, every ten new VMs added by the scale-out policy would consist of five standard VMs and five Spot Virtual Machines. A lower ratio of standard to Spot Virtual Machines can be set to maximize cost savings, deploying more Spot Virtual Machines instead of standard VMs as the scale set grows. This approach suits workloads that do not require much-guaranteed capacity at larger scales. Conversely, increasing the ratio ensures adequate baseline standard capacity for workloads needing more resiliency at scale.

Administrators can find additional information on choosing appropriate VM families and sizes using the VM selector and the Spot Advisor.

Azure Compute Fleet

Azure Compute Fleet is a foundational service designed to accelerate access to Azure's capacity within a region. Compute Fleet can efficiently launch a combination of virtual machines (VMs) at the lowest price and highest capacity, catering to various use cases, including stateless web services, significant data clusters, and continuous integration pipelines. Compute Fleet is particularly beneficial for workloads such as financial risk analysis, log processing, and image rendering, which require the capability to run hundreds of concurrent core instances.

Azure Compute Fleet allows the deployment of up to 10,000 VMs with a single API, utilizing both Spot Virtual Machines and Standard VM types. It enables users to attain superior price-performance ratios by leveraging a mix of pricing models, including Reserved Instances, Azure savings plan for compute, Spot Virtual Machines, and pay-as-you-go (PAYG) options. Rapid provisioning from a customized SKU list tailored to specific preferences expedites access to Azure capacity. Using Standard and Spot Virtual Machines, personalized allocation strategies can optimize for cost, capacity, or a combination of both.

The “Fire and Forget-it” model automates VM deployment, management, and monitoring processes, eliminating the need for complex code frameworks and simplifying the initial setup process. This streamlines resource management by removing concerns about scripting complexity related to determining optimal VM pricing, available capacity, managing Spot evictions, and SKU availability. Azure Compute Fleet also maintains Spot target capacity in the event of VM evictions due to price or capacity constraints. Notably, there are no additional charges for using Compute Fleet; users are only billed hourly for the VMs launched by the Compute Fleet. This integration significantly enhances the ability to manage and optimize large-scale compute deployments in Azure, ensuring cost efficiency and operational agility.

Azure Storage

Azure Storage provides a range of services and features that enable businesses to optimize storage costs while maintaining the necessary performance and reliability for various workloads. From tiered storage models to advanced disk options and elastic storage solutions, Azure offers flexible, scalable options tailored to different business needs. Organizations can balance performance requirements with budget constraints by leveraging these cost-saving opportunities, ensuring efficient and cost-effective cloud storage management.

Azure Disk Storage

Azure Disk Storage offers several options for cost optimization, making it an attractive choice for managing storage in the cloud. It comes in five performance tiers: Ultra Disk, Premium SSDv2, Premium SSD, Standard SSD, and Standard HDD. Each tier is designed to meet different performance and cost requirements.

Ultra Disks provide the highest performance with customizable IOPS and throughput. They suit data-intensive applications such as SAP HANA, high-performance computing (HPC), and other demanding workloads. Although the most expensive, Ultra Disks offer the best performance for the most demanding applications.

Premium SSD v2 offers higher performance than Premium SSDs while generally less costly. Performance can be tweaked (capacity, throughput, and IOPS) of Premium SSD v2 disks at any time, making workloads cost-efficient while meeting shifting performance needs. Azure Premium SSD v2 is designed for IO-intense enterprise workloads that require sub-millisecond disk latencies and high IOPS and throughput at a low cost. Premium SSD v2 is suited for a broad range of workloads such as SQL server, Oracle, MariaDB, SAP, Cassandra, Mongo DB, big data/analytics, gaming, on virtual machines, or stateful containers.

Premium SSDs best suit production workloads requiring consistent low latency, high throughput, and high-performance I/O. They are ideal for database applications, high-transactional workloads, and mission-critical systems.

Standard SSD offers a balance between cost and performance. It suits web servers, lightly used enterprise applications, and development/test environments. Standard SSDs provide better performance than Standard HDDs at a moderate cost increase.

Standard HDD: This is the most cost-effective option, suitable for backup, non-critical, and infrequently accessed data. It provides a lower cost per GB and is ideal for workloads where performance is not a primary concern.

Businesses can optimize costs by choosing the appropriate disk type based on workload requirements while ensuring adequate performance.

Azure Elastic SAN

Azure Elastic SAN is a cloud-based storage solution that provides scalable, high-performance block storage for virtual machines and containers. It offers the flexibility to configure and manage storage resources like traditional SAN systems but with the elasticity and scale of the cloud. Azure Elastic SAN allows seamless integration with Azure VMs and AKS, delivering simplified management, rapid provisioning, and high availability for mission-critical workloads. This service is ideal for large-scale storage applications with consistent performance, such as databases, ERP systems, and transactional applications.

Azure Blob Storage

Azure Blob Storage offers several options for cost optimization through its tiered storage model. Blob Storage is ideal for unstructured data storage, providing a scalable and cost-effective solution.

Hot Tier is designed for data that is accessed frequently. It has higher storage costs but lower access and transaction costs, making it suitable for active data and workloads requiring low-latency access.

Cool Tier is intended to store data that needs to be accessed infrequently for at least 30 days. It offers lower storage costs than the Hot tier but higher access and transaction costs. This tier is ideal for backup data, disaster recovery, and long-term data storage that does not require frequent access.

Cold Tier is a cost-effective storage solution designed for infrequently accessed data, balancing low storage costs and higher retrieval costs. It ensures that organizations can retain large amounts of data at minimal expense. It is ideal for archiving, backup, and long-term data storage, which is rarely needed but must be kept for regulatory or compliance reasons. It integrates seamlessly with Azure's broader storage options, allowing easy data movement between tiers based on access patterns while maintaining the durability and security of data within the Azure ecosystem.

Archive Tier is the most cost-effective storage option for data that is rarely accessed and stored for at least 180 days. It has the lowest storage cost but the highest access and transaction costs. This tier is perfect for archival data, compliance data, and any data that can tolerate high access latency.

By leveraging these tiers, businesses can optimize their storage costs by aligning the data storage tier with the data access patterns and retention requirements.

Azure Data Lake Storage

Azure Data Lake Storage does not have traditional storage tiers like Azure Blob Storage, which includes hot, cool, and archive tiers. However, Azure Data Lake Storage (ADLS) Gen2 is built on top of Azure Blob Storage, so the same tiering options can be leveraged by using Blob Storage tiering. ADLS Gen2 allows data to be stored in a way that optimizes big data analytics. However, blob access tiers must be set within the same account at the container or blob level to manage costs using different storage tiers. For ADLS Gen2, cost management is typically handled through lifecycle policies, which can move data to lower-cost storage options like the archive tier after specific periods.

Azure Files

Azure Files provides fully managed file shares in the cloud, accessible via SMB and NFS protocols. There are several options for cost optimization within Azure Files, including standard, premium, transaction optimized, and cold.

Standard File Shares: These are the most cost-effective file shares, suitable for general-purpose file storage and infrequently accessed data. They balance performance and cost and are ideal for development/test environments, user home directories, and lightly used applications.

Premium File Shares: Premium file shares offer high-performance storage with low latency and high throughput. They are designed for I/O-intensive workloads such as databases, high-performance computing, and enterprise applications. Although they are more expensive than Standard file shares, they provide the performance needed for demanding applications.

Transaction Optimized File Shares: This option balances standard and premium file shares, optimizing for transaction-heavy workloads with moderate performance needs. They suit scenarios with frequent transactions and moderate performance requirements, such as shared application data and user profile storage.

Cold File Shares: Cold file shares are designed for infrequently accessed files and provide the lowest cost per GB. They are ideal for backup, archival, and long-term data storage where access is rare, but data durability is critical.

By selecting the appropriate file-share option based on workload requirements and access patterns, businesses can effectively manage their storage costs while meeting performance needs.

Azure Lifecycle Storage Policies

Incorporating Azure storage lifecycle policies is a crucial strategy to optimize storage costs in cloud environments. By leveraging features such as lifecycle management, tiered storage, and data retention policies, organizations can automatically move data to cost-effective storage tiers based on usage patterns. For example, frequently accessed data can be stored in the Hot tier, while infrequently accessed or archival data can be transitioned to the Cool or Archive tiers, reducing storage costs.

Azure also allows data to be automatically deleted, which is no longer needed after a specific retention period, further lowering storage expenses. These policies enable organizations to efficiently manage data growth and control costs without compromising performance or availability. Another strategy would be creating a rule that moves blobs to the archive tier if that blob hasn't been modified in 90 days for applications that don't need real-time access to the data.

Advanced Techniques for Cost Control

Adopting advanced techniques and optimization strategies allows organizations to deploy Linux solutions on Azure cost-effectively and efficiently while maintaining high performance and reliability.

By following these advanced techniques and optimization strategies, organizations can deploy Linux solutions on Azure cost-effectively and efficiently while maintaining high performance and reliability.

Cost Savings Through Automation and Scripting: Automating routine tasks and optimizing resource usage through scripting can lead to substantial cost savings:

- Azure Automation: Azure Automation schedules and manages workloads, reduces manual intervention, and ensures efficient resource utilization.
- Infrastructure as Code (IaC): Implement IaC using tools like Azure Resource Manager (ARM) templates, Bicep, or Terraform to automate the deployment and management of infrastructure, ensuring consistency and reducing human errors.

Additional Optimization Strategies: To further optimize costs and performance within an Azure environment, consider implementing the following strategies:

- **Marketplace Images:** Utilizing marketplace images instead of building and maintaining custom images reduces maintenance overhead and streamlines updates.
- **Schedule VM Start and Stop:** Automating the scheduling of VM start and stop times to align with usage patterns minimizes costs during off-peak hours when resources are not in use.
- **DDoS Protection:** Implementing Azure DDoS Protection for internet-facing workloads safeguards against attacks and prevents additional costs associated with mitigating such incidents.
- **Shared Azure Resources:** Sharing Azure resources like Azure Firewall, Azure Application Gateway, or Azure Front Door among multiple workloads or landing zones optimizes resource utilization and reduces costs.
- **Correct Azure Purchase Method:** Adopting appropriate purchase methods for different Azure services, for instance, leveraging trial or no-license-required options for development environments and reserving license purchases for production use only.

Integrating these strategies can achieve substantial cost savings and enhanced performance, ensuring a more efficient and secure Azure infrastructure.

Pricing Models and Offers

Azure provides several pricing models and offers relevant to Linux deployments, each catering to different needs and usage scenarios. Understanding these models is crucial for optimizing costs and ensuring deployment aligns with budget and operational requirements.

Organizations must also understand and evaluate the subscription location. Azure offers plans in countries worldwide, each with different pricing and tax implications.

Azure Plan and Pay-As-You-Go

The Azure Plan and Pay-As-You-Go models offer maximum flexibility, allowing users to pay for Azure services based on actual usage. This model is ideal for dynamic workloads that experience varying demand levels, eliminating the need for long-term commitments and upfront costs. Resources can be scaled up or down as needed, with payment required only for what is used, making it a cost-effective option for unpredictable workloads.

Cloud Solution Provider (CSP)

The Cloud Solution Provider (CSP) program allows organizations to purchase Azure services through a third-party provider. CSP partners can offer customized solutions and support, often bundling Azure services with additional value-added services. This model benefits organizations that prefer a managed services approach or require extra support and customization beyond what is available directly from Azure. CSP pricing can also include flexible billing options and potential discounts the provider negotiates.

Dev/Test

The Dev/Test subscription is designed for individual developers or small teams who need access to Azure services for development and testing purposes. This model provides a cost-effective way to explore Azure capabilities without incurring significant expenses. Developer subscriptions often come with free credits, access to premium developer tools, and discounted rates on specific services, making it an attractive option for those looking to innovate and experiment in a cost-controlled environment.

Linux distribution pricing

The choice of Linux distribution (distro) can significantly impact pricing. Different distributions come with varying costs and support options. Additionally, the scalability and flexibility required in Linux deployments can influence overall expenses, as some distributions may be more cost-effective under specific conditions.

When comparing the costs of Linux VMs in Azure, licensing fees and support may or may not be included. This is determined by the image deployed to a customer's subscription. Options are available based on the distro and Microsoft's support for the OS.

It is essential to note the distinction between unpaid and paid Linux options in Azure. Some distributions are free, relying on the community for support, while others come with licensing fees and professional support options. Examples of paid distributions include Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES), and Ubuntu Pro, which offer enterprise-grade support and additional features. On the other hand, distributions like Ubuntu LTS, AlmaLinux, Debian, and Rocky Linux can be used free of charge, though commercial support may be available as an optional add-on. Depending on the need for stability, support, and cost, organizations can choose between these paid and unpaid options to best suit the deployment requirements.

All Linux distributions are welcome in Azure. Several different sources of Linux VM images are available for Azure, such as marketplace, platform, custom, and community gallery. Each source provides a different expectation for quality, utility, and support. Endorsed Linux distributions on Azure are those that have been tested and validated by Microsoft to run effectively on the Azure platform. Their publishers regularly update these distributions in the Azure Marketplace to include the latest patches and security fixes.

Optimize Azure compute spend with Savings Plan vs. Reserved Instances

Azure provides two pricing offers to help maximize an organization's compute investment: Azure savings plan for compute and Reserved Instances (RIs).

This section will examine the key features and benefits of Azure Savings Plans for Compute and Reserved Instances to distinguish between them and determine the best suited for specific workloads. It will also explore how both can be leveraged to achieve optimal results. The differences between the two are highlighted in Figure 1 – Azure savings plan vs reserved instances.

Azure savings plan vs. reserved instances

Azure savings plan for compute give customers savings with built-in flexibility and automatic optimization. **This plan works best when customers:**

Reserved instances offer the greatest savings. **They work best when customers:**

Azure savings plan	Reserved instances
Have planned or unplanned changes	Run workloads continuously within one instance family and in the same region
Run different compute services or VM instances families	Have no planned changes
Utilize different regions at different times of the day (For example, 24-hour global help desk)	Have no expected changes to machine series or region (For example, those that have highly stable workloads, like domain controllers)

Figure 1 – Azure savings plan vs reserved instances

Azure Savings Plans and Reserved Instances enable optimization of Azure compute usage by committing to a specific amount of spend or usage over a defined period. However, each option has distinct characteristics and advantages designed to suit different scenarios and requirements.

Azure Savings Plan for Compute

Azure savings plan for compute enables organizations to reduce eligible compute usage costs by up to 65% (off list pay-as-you-go rates) by making an hourly spend commitment for 1 or 3 years. Unlike Azure reservations, which target stable and predictable workloads, Azure savings plans target dynamic and evolving workloads. To learn more, visit Decide between a savings plan and a reservation. Savings plans are a billing discount – it doesn't affect the runtime state of Azure resources.

Azure savings plans are available to organizations with either an Enterprise Agreement (EA), Microsoft Customer Agreement (MCA), or Microsoft Partner Agreement (MPA) agreements. Enterprise Agreement customers must have an offer type of MS-AZR-0017P (EA) or MS-AZR-0148P (DevTest) to purchase Azure savings plans. To learn more, visit Buy an Azure savings plan.

Reserved Instances

Reserved Instances (RIs) in Azure provide a cost-effective solution for businesses running Linux VMs by offering significant savings compared to pay-as-you-go pricing. By committing to a one-year or three-year term, organizations can receive up to 72% discount on VM costs, making it an attractive option for predictable, steady-state workloads. Figure 2 – Reserved Instances shows the difference in the cost model for Unpredicted vs. Predicted Capacity. Before the cloud, companies would always have to purchase hardware for the busiest day possible; in Azure, that's no longer the case, allowing workloads with different profiles to be budgeted more intelligently.

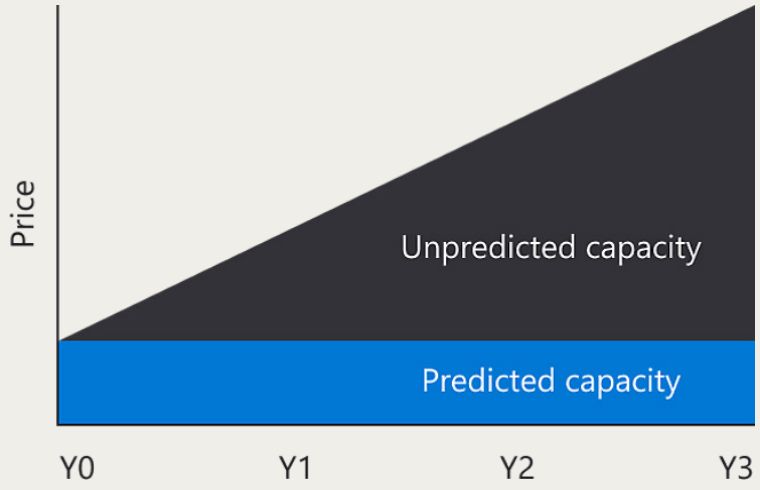

Figure 2 – Reserved Instances

Calculate break-even point: Calculating the break-even point is essential to determining the financial viability of purchasing a reserved instance compared to using the pay-as-you-go (PAYG) pricing model. The break-even point represents the time required for the reserved instance to pay for itself relative to the PAYG price. Understanding this timeframe helps assess whether the reserved instance is a prudent investment for the workload and budget.

The formula in Figure 3 – Break-even point calculation, can be used to understand the point at which a reserved instance makes sense for a Linux deployment in Azure:

$$\text{Break-even in months} = \frac{\text{Reserved instance price (yearly)}}{\text{PAYG Price (monthly)}}$$

Figure 3 – Break-even point calculation

For example, consider purchasing a reserved instance for a D4av4 VM in the East US region, which includes 4 vCPUs and 16 GiB of memory. The PAYG price for this VM is $140.16 per month or $1,681.92 yearly. The reserved instance price for a one-year term is $82.42 monthly or $989.00. This results in a cost savings of $692.92, which represents a 41% discount. Using this formula, the break-even point is calculated in Figure 4 — Example RI calculation:

$$\text{Break-even in months} = \frac{989.00}{140.16} = 7.05$$

Figure 4 – Example RI calculation

This calculation indicates that the reserved instance needs to be used for at least 7.05 months, or approximately seven months, to break even with the PAYG price. Usage beyond this period results in cost savings, whereas using it for a shorter duration leads to a financial loss.

Understanding the break-even point is crucial for making informed decisions about investing in reserved instances, ensuring that compute resources are cost-effective and aligned with budgetary constraints.

Azure Hybrid Benefit for Linux: Azure Hybrid Benefit for Linux can provide additional cost savings for organizations with active Linux subscriptions, including Red Hat Enterprise Linux or SUSE Linux Enterprise Server running in Azure. It allows eligible customers to use their existing on-premises licenses on Azure, reducing the need for new licenses and lowering overall costs.

For Red Hat and SUSE Linux implementations, Azure Hybrid Benefit for Linux offers significant financial advantages by leveraging existing investments in software licenses. Enterprises that have invested substantially in Red Hat or SUSE Linux for their on-premises environments. Organizations can optimize their cloud expenditure by applying these licenses to Azure deployments while maintaining consistency and compliance with their existing infrastructure and software standards. This approach not only maximizes the value of current licenses but also facilitates a smoother transition to the cloud by ensuring compatibility and support for critical applications and workloads.

Combined with Reserved Instances, Azure Hybrid Benefit for Linux can provide substantial cost savings. In some cases, up to 76%[2] with Azure Hybrid Benefit for Linux and reserved instances.

[2]The 76% savings is based on one standard D2s v3 Azure VM with RHEL or SLES subscription in the East US region running at a pay-as-you-go rate vs a reduced rate for a 3-year Reserved Instance. Based on Azure pricing as of October 2022. Prices subject to change. Actual savings may vary based on location, instance type, or usage.

Cost Analysis Tools

Effective cost management is crucial for organizations leveraging Azure for their Linux environments. Azure provides a suite of tools and services designed to help estimate, manage, and optimize costs. This section overviews these tools and their capabilities, focusing on how they enable precise cost tracking, allocation, and optimization. These tools allow organizations to gain detailed insights into their spending patterns, identify cost-saving opportunities, and ensure financial transparency and accountability across different departments and projects. The tools highlighted in this section include a pricing calculator, TCO calculator, Azure Migrate, Microsoft Cost Management, Azure Advisor, and tags. Each tool offers unique features contributing to effective cost management, enabling organizations to control their cloud expenditures and optimize their Azure investments.

Pricing on Azure

Getting the best value for a company's investment in Linux on Azure requires a better understanding of the deployment implementations. Microsoft understands that moving to a new pricing and purchasing model is just as complex as the technology. As part of its commitment, Microsoft provides tools to help cloud administrators and leaders plan for their new budgets. Two essential tools are the pricing calculator and the TCO calculator.

Azure Pricing Calculator

The pricing calculator is an essential tool designed to help users estimate the costs associated with deploying and running services on Azure. This comprehensive tool, shown in Figure 5 – Azure Pricing Calculator, lets administrators input their configurations, including virtual machines, storage options, networking, and other resources, to generate a detailed cost breakdown. By offering a granular view of potential expenses, the calculator helps users plan and budget more effectively, ensuring they understand the financial implications before committing to deployment.

Additionally, the Azure pricing calculator allows users to adjust various parameters and immediately see the impact on costs, which is particularly useful for comparing different scenarios and making informed decisions. This tool is invaluable for initial cost estimation, ongoing cost management, and optimization as workloads and requirements evolve. Calculations can be saved, shared, or exported to Excel.

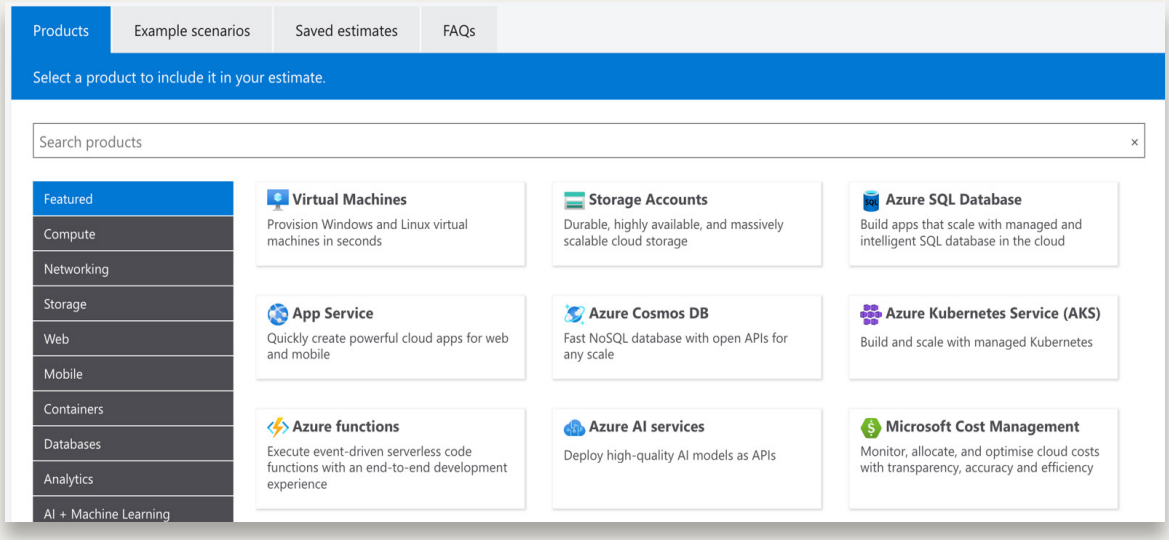

Figure 5 – Azure Pricing Calculator

TCO calculator

In addition to the pricing calculator, Azure offers a Total Cost of Ownership (TCO) calculator. This tool helps organizations compare the costs of running workloads on Azure versus on-premises or other cloud platforms, providing a comprehensive view of potential savings and cost implications.

Business leaders can evaluate the potential cost savings of migrating to Azure by calculating and comparing the TCO for Azure with that of a comparable on-premises deployment.

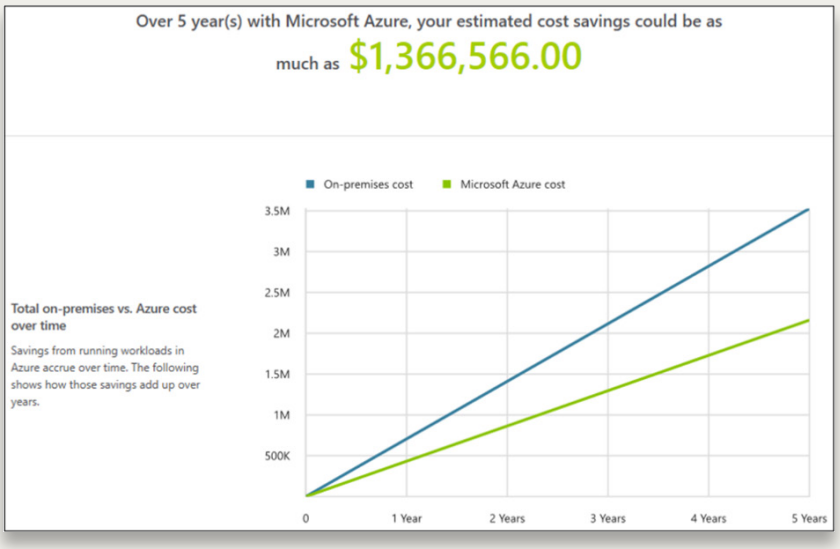

Figure 6 – Azure TCO Calculator

Azure Migrate

Azure Migrate facilitates the planning and execution of migration and modernization projects to Azure. It provides a centralized hub to track the discovery, assessment, and migration of on-premises infrastructure, applications, and data to Azure. This hub includes Azure tools for assessment and migration and offerings from third-party independent software vendors (ISVs).

Azure Migrate provides price recommendations derived from comprehensive assessments to optimize cloud spending. By analyzing current on-premises workloads and usage patterns, Azure Migrate generates tailored pricing suggestions that align with an organization's applications' specific needs and performance requirements. These recommendations help identify the most cost-effective Azure services and configurations, enabling informed resource allocation and budget planning decisions. By leveraging these insights, organizations can ensure they utilize Azure resources efficiently, avoiding over-provisioning and minimizing unnecessary expenses.

The Business Case capability shown in Figure 7 - Building a business case with Azure Migrate assists in developing a comprehensive business proposal to understand the value Azure can bring. It highlights critical factors such as the total cost of ownership (TCO) comparison between on-premises and Azure environments, year-on-year cash flow analysis, and resource utilization insights to identify servers and workloads ideal candidates for cloud migration. It also identifies quick wins for migration and modernization, including end-of-support Windows OS and SQL versions. It outlines long-term cost savings by transitioning from a capital expenditure model to an operating expenditure model, paying only for what is used.

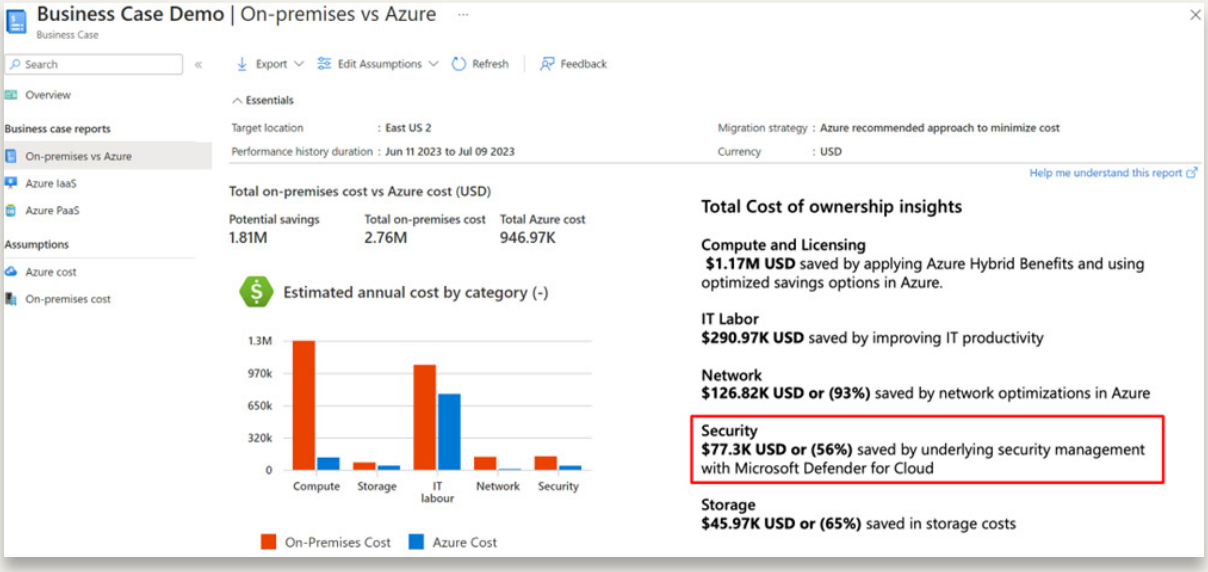

Figure 7 – Building a business case with Azure Migrate

This capability eliminates the guesswork in the cost-planning process by providing data-driven insights and calculations. It can be generated in just a few clicks after performing discovery using Azure Migrate. Additionally, the feature is automatically enabled for existing Azure Migrate projects, streamlining the process of building a business case for migration to Azure.

Microsoft Cost Management

Microsoft Cost Management offers comprehensive capabilities for tracking, managing, and optimizing cloud expenses. These tools provide detailed insights into cost and usage patterns, enabling organizations to identify trends and anomalies. Key features include cost analysis, budget creation, and cost alerts, which help control expenditures and prevent unexpected overruns. The billing tools also support various billing scenarios and provide detailed invoices, making allocating costs accurately across different departments or projects easier. Two critical tools are Cost Analysis and Budgets.

Cost Analysis

Cost analysis is a critical component of Microsoft Cost Management, providing detailed insights into cloud expenditure patterns. It allows organizations to break down costs by various dimensions, such as services, resources, and time periods, offering a granular view of where money is spent. This analysis helps identify cost drivers, detect anomalies, and uncover opportunities for optimization. By utilizing cost analysis, organizations can make informed decisions to control and reduce their cloud spending, ensuring that their Azure investments align with their financial and operational goals.

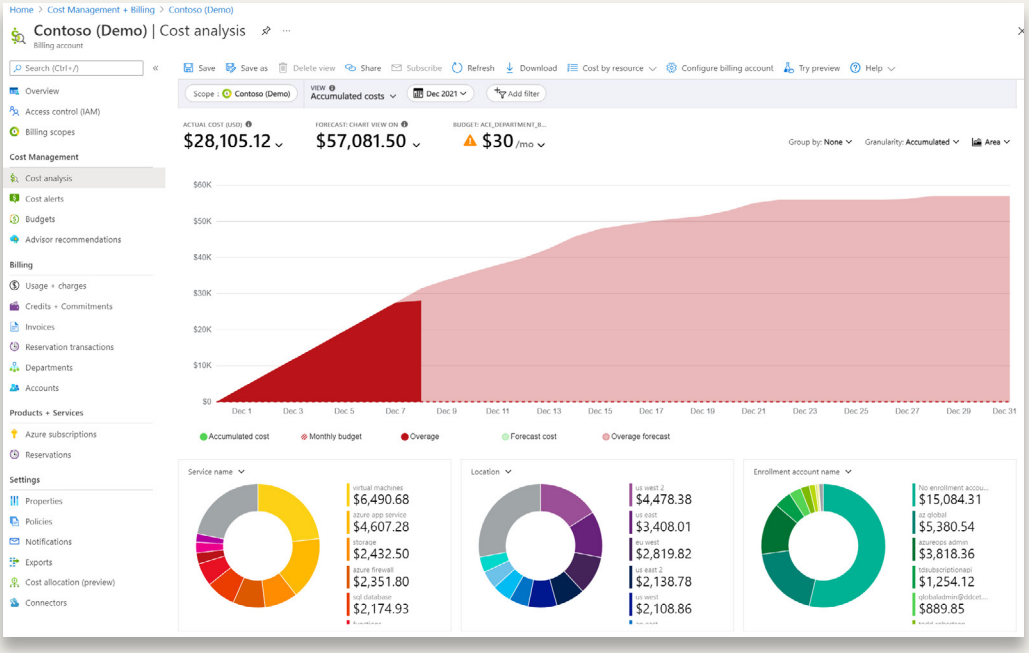

Figure 8 – Cost Analysis in the Azure Portal

Available for any resource in the Azure Portal, cost analysis is easy to use and helps administrators:

- Visualize and analyze organizational costs
- Share cost views with others using custom alerts
- View aggregated costs by organization to understand where costs occur over time and identify spending trends
- View accumulated costs over time to estimate monthly, quarterly, or even yearly cost trends against a budget
- Create budgets to provide adherence to financial constraints
- Use budgets to view daily or monthly costs and help isolate spending irregularities

Budgets

Budgets in Microsoft Cost Management allow for setting and managing budgets, enabling proactive cost control. Users can create budgets at various scopes, such as subscriptions, resource groups, or individual resources, and set up alerts to notify stakeholders when spending approaches or exceeds budget limits.

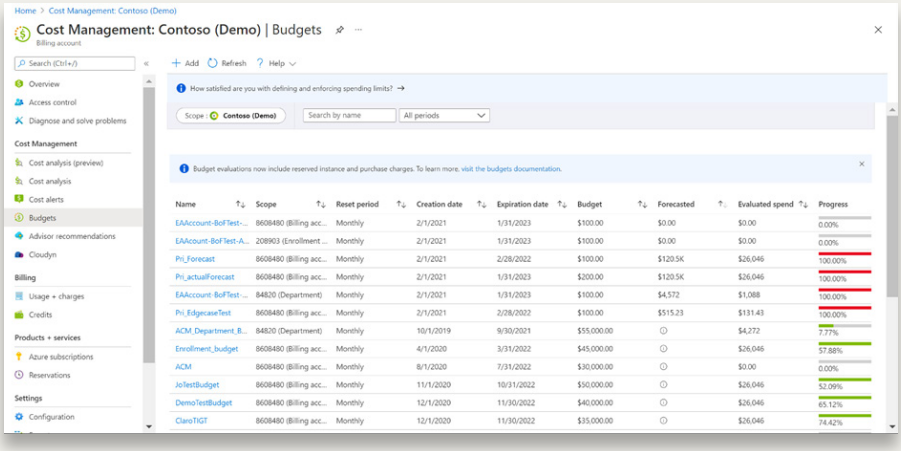

Figure 9 – Budgets in the Azure Portal

Azure Advisor

Azure Advisor is a personalized cloud consultant that provides best practice recommendations to optimize Azure deployments. It includes cost optimization suggestions that can help reduce unnecessary spending. The tool refreshes regularly, giving specific recommendations on Linux VMs deployed in the customer subscription. Here are some of the benefits of this tool:

- Get proactive, actionable, and personalized best practices recommendations
- Improve the performance, security, and reliability while identifying opportunities to reduce overall Azure spend
- Get recommendations with proposed actions in line

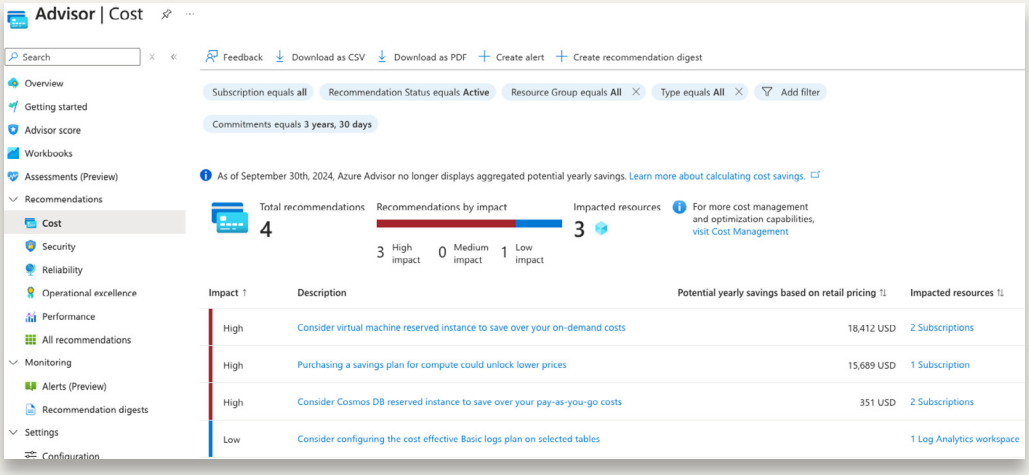

Figure 10 – Azure Advisor Cost Recommendations

Tags

Tags are a powerful tool for cost management and chargeback within an organization. By assigning metadata tags to Azure resources, businesses can categorize, and track expenses based on departments, projects, or cost centers. This granular level of detail enables precise cost allocation and ensures that each department or project is accountable for resource consumption.

Tags facilitate detailed reporting and analysis, helping identify cost-saving opportunities and trends. Moreover, they support automated workflows and governance policies, streamlining the cost management process and enhancing financial transparency across the organization. This systematic approach to tagging not only aids in effective budget control but also simplifies the chargeback process, ensuring equitable distribution of costs.

Administrators can use Azure Policy to enforce tagging rules and conventions. Through the creation of a policy, the risk of deploying resources without the required organizational tags within a subscription is mitigated. By applying tags to all deployed resources, cost tracking becomes feasible across the subscription. For effective cost tracking, it is recommended to use Azure tags such as “CostCenter,” “Department,” “Project,” “Environment,” and “Owner” to categorize and monitor resource expenditures accurately.

Managing and Monitoring Costs in Real-Time

Real-time cost monitoring and management are essential for maintaining budget adherence and avoiding unexpected charges. Azure provides several strategies and tools to facilitate this, including setting up alerts on budgets and leveraging Azure automation.

Cost Alerts

Azure allows configuring cost alerts to notify when spending exceeds predefined thresholds. These alerts can be set at different levels, such as resource groups, subscriptions, or specific services, ensuring timely notifications to stakeholders.

Automation

Automation: Using tools like Azure Automation and Azure Logic Apps, organizations can automate routine tasks, such as scaling resources up or down based on usage patterns. This helps optimize costs and improve operational efficiency.

Using Azure Log Analytics for Cost and Usage Data Analysis

Azure Log Analytics plays a critical role in cost and usage data analysis. It enables collecting, querying, and analyzing log data from various sources. Organizations can gain actionable insights into spending patterns and resource utilization by creating custom queries and visualizations. This information is crucial for identifying cost-saving opportunities and ensuring efficient use of cloud resources.

Azure Orphaned Resources Workbook

Orphaned objects in Azure, such as unused virtual machines, disks, and IP addresses, can accumulate unnoticed and continue to incur costs. Implementing the Azure Orphaned Resources Workbook helps centralize and identify these orphaned resources within Azure environments. As shown in Figure 11 – Azure Orphaned Resources workbook, this resource makes it simple to find resources that are no longer needed but are still being billed to the subscription.

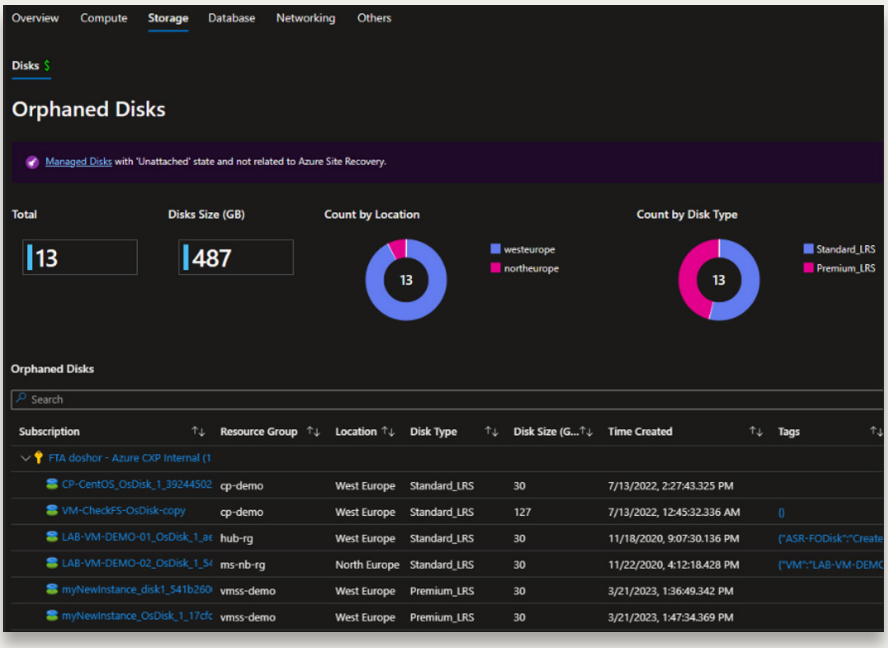

Figure 11 – Azure Orphaned Resources workbook

This workbook provides an overview of orphaned resources, enhancing efficiency by saving money, preventing misconfigurations, and simplifying operational management. Organizations can optimize their Azure spending and improve overall cloud resource management by regularly reviewing and addressing orphaned resources.

Leveraging FinOps on Azure

Adopting FinOps practices in Azure enables organizations to achieve greater financial accountability and efficiency in cloud operations. FinOps, or Financial Operations, combines financial management with operational and technical best practices to optimize cloud spending. By implementing FinOps, organizations can gain visibility into their Azure consumption, track and forecast expenditures, and align spending with business objectives. This approach involves continuous monitoring and analysis of cloud costs, enforcing budget controls, and identifying opportunities for cost savings. Utilizing Azure's native tools and dashboards, such as Microsoft Cost Management, enhances FinOps capabilities, allowing teams to make data-driven decisions and ensure that cloud investments deliver maximum value. To learn more, download the FinOps with Azure e-book.

Conclusion

This whitepaper provides a comprehensive overview of managing and optimizing Linux environments on Azure, focusing on Azure services, pricing models and offers, and cost analysis tools. By leveraging Azure's diverse VM and storage options, cost-saving pricing models, and robust cost analysis tools, organizations can efficiently manage their Linux deployments, ensuring optimal performance and cost efficiency.

Azure services: Azure's range of VM and storage services are tailored to different workloads, allowing organizations to select the most suitable VMs for their needs. Techniques such as autoscaling and utilizing Spot Virtual Machines can significantly reduce costs, while reserved instances provide long-term savings for predictable workloads. Selecting the correct storage tiers for workload requirements will provide significant long-term savings.

Pricing models and offers: Understanding Azure's various pricing models and offers, including pay-as-you-go, reserved instances, and Azure Hybrid Benefit, enables organizations to design cost-effective architectures.

Cost analysis tools: Azure offers several tools that provide real-time insights into spending patterns. Tools like Microsoft Cost Management, Tags, and Azure Advisor help monitor and control expenditures, ensuring financial transparency and accountability.

By adopting these strategies and tools, organizations can manage their Linux environments on Azure more effectively, achieving significant cost savings while maintaining high performance and security. This whitepaper guides IT professionals, system administrators, and decision-makers to navigate the complexities of cloud cost management and make informed decisions that align with their operational and financial goals.