Azure e-book series

Microsoft

Enterprise Cloud Strategy

Infrastructure as a Service

Barry Briggs / Third edition

Published by

Microsoft Corporation
One Microsoft Way
Redmond, Washington 98052-6399

Author: Barry Briggs

07 /
Accelerate Productivity with Containers and Orchestration

08 /
Create, Manage, and Maintain Your IaaS Ecosystem

09 /
Governance, Risk, Compliance, and Security

Summary and Conclusions

For Further Reading

Introduction

Since its introduction a decade ago, the cloud has become an increasingly important aspect of corporate and IT computing. Today, nearly every organization uses cloud computing in some way.

This book describes one of the most popular and foundational approaches to cloud computing, called infrastructure as a service (IaaS), in which the cloud provides the basic underlying components of computing—facilities, servers, networks, storage—allowing organizations to decrease or even eliminate their reliance on private datacenters. Infrastructure as a service enables companies to rapidly move applications away from expensive datacenters in order to take advantage of the cost efficiencies of the cloud.

In this book, we define what IaaS is and place it in context with other leading cloud computing models, platform as a service (PaaS) and software as a service (SaaS). We then describe a number of common uses of IaaS, ranging from offloading development and testing from the datacenter to web applications, and for hosting your enterprise applications.

You might be asking, "How much will my organization have to pay?" We show how you can understand the costs of cloud computing and compare them to your costs on-premises in your datacenter, then help you to avoid common pitfalls, such as overprovisioning, so that you get the most value for your cloud investment.

Then we go on a "deep dive" of Azure IaaS, where we walk through the major technologies and capabilities offered by Azure, and how you can use them. We discuss the three foundational aspects of IaaS in Microsoft Azure: compute, storage, and networking, and your options in each.

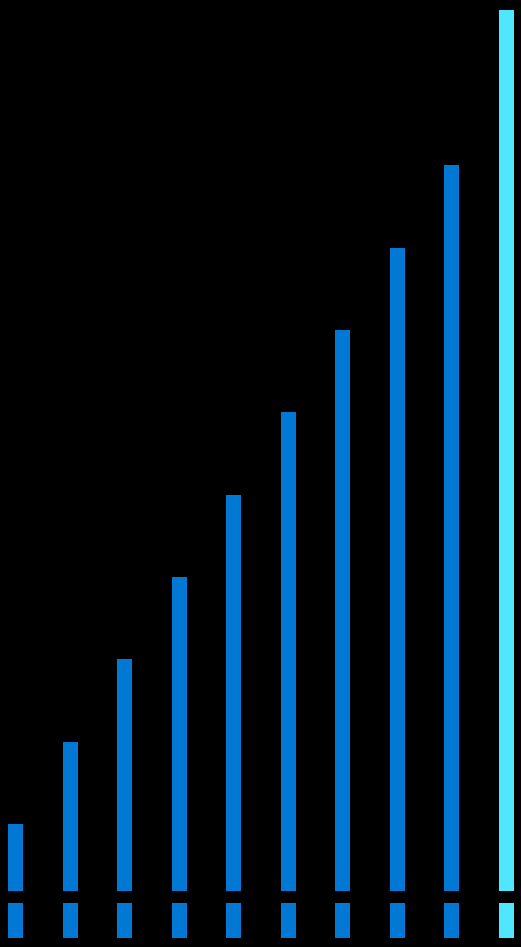
We cover more ways you can get value from the cloud, even if your applications remain on-premises: using the cloud's inexpensive data backup features; its support for business continuity and disaster recovery (BC/DR); and others.

From there we discuss a relatively new set of capabilities, containerization and orchestration, respectively, and how these technologies can both save you money and provide you with new levels of scale and resilience. And of course, we show how you manage your growing IT ecosystem in the cloud, describing the various tools available to you to monitor and manage your applications.

Finally, we cover two critically important areas of cloud computing: security and compliance, which are naturally top of mind for every enterprise IT executive. We talk about what capabilities and support Azure offers you, and what your responsibilities are in these spaces.

We hope you enjoy your tour of Azure infrastructure as a service. After reading this book, you and your teams should be prepared to embark on your cloud journey!

01 / Why the Cloud?



These days, there are fewer reasons to have your own datacenter.

Data centers are expensive

There's no denying that datacenters are hugely expensive. Studies have shown that building out an existing property can cost $200 per square foot, and for a new facility, upward of $1,000 per square foot. To reach your datacenter with your network you would need to lay fiber, which can cost anywhere from $10,000 to $25,000 per mile. For a datacenter of any reasonable size, overall costs begin in the millions of dollars.

And, of course, datacenters incur significant expenses in operations: for everything from facilities maintenance and electricity (which could be tens of thousands of dollars per megawatt, or more) to cooling, physical security, network security, the servers themselves, software licenses, storage networking, human resource costs, and so on. None of these expenses directly return value to the business.

Once functioning, the day-to-day processes of a datacenter can be complex, expensive, and time-consuming. A new application requires hardware (server, network,

storage) to run, and all these must be procured and provisioned, which can take days, weeks, or longer—delaying time for the application to actually generate the value toward its original goal.

Needed: Computing on demand

We require instantly available computing technology that can be used as needed and then discarded when not. Like electricity and the telephone, computing should be available "as a service" and not as a massive capital expense, and that is the promise of the cloud—a promise that is being realized today by hundreds of thousands of companies around the world. Cloud computing is available to you on a pay-as-you-go basis: you pay for the resources you use, and when you no longer need them, you no longer pay.

As we show, with cloud technology, and specifically a form of it called infrastructure as a service (IaaS—pronounced "eye-as"), you can outsource your datacenter to a public cloud provider such as Microsoft Azure. Using the IaaS capabilities of Azure, you can migrate your applications to the cloud and over time decrease, and perhaps eliminate altogether, your dependence on traditional datacenters.

By using Azure as a "virtual datacenter," you can

- eliminate expenses associated with the physical aspects of computing: servers, storage, and networking;
- migrate your applications at your own pace, closing and consolidating datacenters while moving functions to the cloud;
- use other Azure IaaS features to complement your remaining on-premises applications and replace expensive products, such as backup/restore, failover, and other essential IT capabilities;
- adjust your usage to get the most optimal balance between cost and compute power, adding more servers during peak usage times and removing unnecessary ones during quiet times; and
- take full advantage of massive economies of scale, since Azure has millions of servers distributed all across the globe.

Considerations when choosing a cloud provider

You have choices when it comes to cloud providers. When considering which one to use, think about the following criteria:

Global coverage

Is your cloud provider global? Does it have datacenters in every area where you may consider doing business? Even if yours is not a global company today, it may be tomorrow; therefore, it behooves you to be ready.

Microsoft Azure, as shown in the chart below, has more than **54** different regions all over the world, with more coming online all the time (check here for the latest list).

Capacity

Only a few cloud providers can operate at the scale needed to assure you that your applications—today and tomorrow—will be able to consume all the capacity they need. According to the website, Azure is available in 140 countries with up to 1.6 Pbps of bandwidth in a region and counting.

SLAs and uptime

Your cloud provider should provide you with contractual service-level agreements that are well documented and broken down by service. In the case of Microsoft Azure, the overall SLA, or guaranteed uptime, is 99.9 percent, with some individual services having higher SLAs.

Security and compliance

Look at how your cloud provider can help you ensure compliance with country and local regulations as well as with industry standards. Cloud providers like Microsoft work closely with governments and industry to ensure Azure is strictly compliant.

In addition, examine how your cloud provider helps you maintain the security of your applications and data. In Microsoft Azure, your data *is your data* and is only accessible to you. Determining the level of physical security of cloud datacenters, involving technologies like encryption and strong authentication, among others, is key to helping you evaluate a cloud provider.

We spend more time discussing governance, risk, compliance, and security later on.

Experience

Think about choosing a cloud provider that has extensive experience not just in the cloud but in all aspects of corporate computing, and one with an extensive ecosystem of partner solutions that can accelerate the realization of your business objectives.

Your business road map

Finally, consider your long-term business road map and the goals you want to achieve. You want a partner who will be “in it” with you for the long term, who can supply state-of-the-art technology when you need it, and who is constantly innovating like you.

02 / Cloud Computing: Terms to Know

As you're undoubtedly aware, you have many approaches at your disposal to use cloud computing. Over time, a taxonomy has developed that has gained wide currency.

Let's take a closer look:

Infrastructure as a service

With the infrastructure as a service (IaaS) model, you are renting only the server hardware and a small amount of software (the hypervisor) to host your application's virtual machine (VM), where the VM consists of the operating system, associated system software, and the application itself. IaaS means that VMs are simply *moved* from on-premises to the cloud. Figure 2-1 illustrates that many operating systems and applications can coexist on a cloud server. A thin piece of code called a hypervisor ensures that each one runs in a timely and efficient fashion.

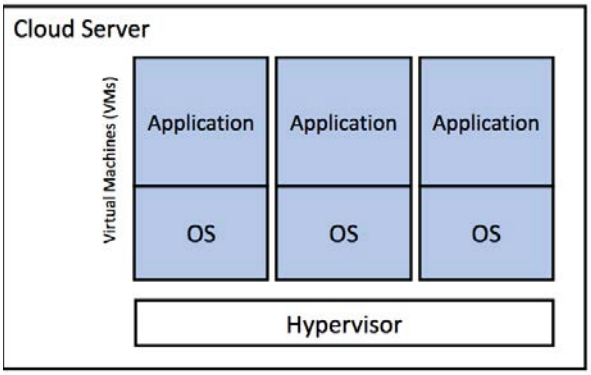


Figure 2-1 Infrastructure as a service

In other words, you supply—and maintain—the pieces highlighted in blue in Figure 2-1.

This is the easiest and fastest migration strategy; it offers many benefits, including cost savings. However, it still means that your operations staff will need to perform such tasks as patch management, updates, and upgrades.

Nevertheless, IaaS is one of the most common cloud deployment patterns to date because it reduces the time between purchasing and deployment to almost nothing. Additionally, because it is the most similar to how IT operates today, it provides an easy onboarding ramp for your current IT culture and processes. As we shall see, the bulk of migration, especially in the early phases of cloud adoption, is to IaaS.

Platform as a service

In platform as a service (PaaS), the cloud provider maintains all system software, removing the burden of upgrades and patches from the IT department. In a PaaS deployment model (Figure 2-2), all that the enterprise needs to focus on is deploying its code on the PaaS machines; the cloud provider ensures that operating systems, database software, integration software, and other features are maintained, kept up to date, and achieve a high SLA.

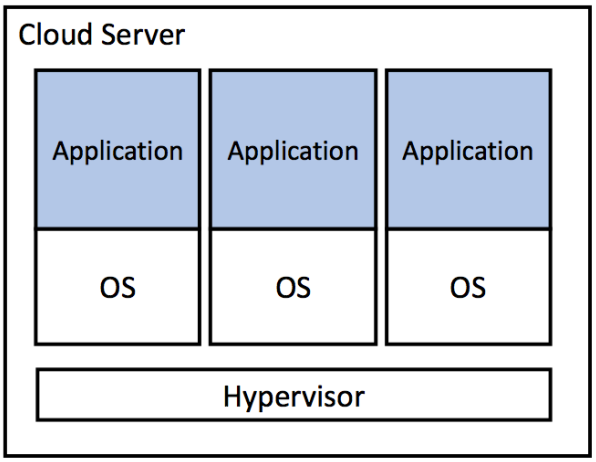


Figure 2-2 Platform as a service

PaaS provides IT departments with important benefits, most important among them being the cost savings associated with reduced or eliminated maintenance of system software and other rote functions. However, PaaS usually implies some redesign of the application in order to best take advantage of the model.

Software as a service

In software as a service (SaaS—pronounced "sass"), you simply rent an application from a vendor, such as Microsoft Office 365 for email and productivity. This is by far the most cost-effective of all the options because typically the only work involved for the IT department is provisioning users and data and, perhaps, integrating the application with single sign-on.

Typically, SaaS applications are used for functions that are not considered business-differentiating, for which custom or customized applications encode the competitively differentiating business models and rules.

Overview of capabilities

A good way to visualize the differences is with a chart of the capabilities provided by the respective models:

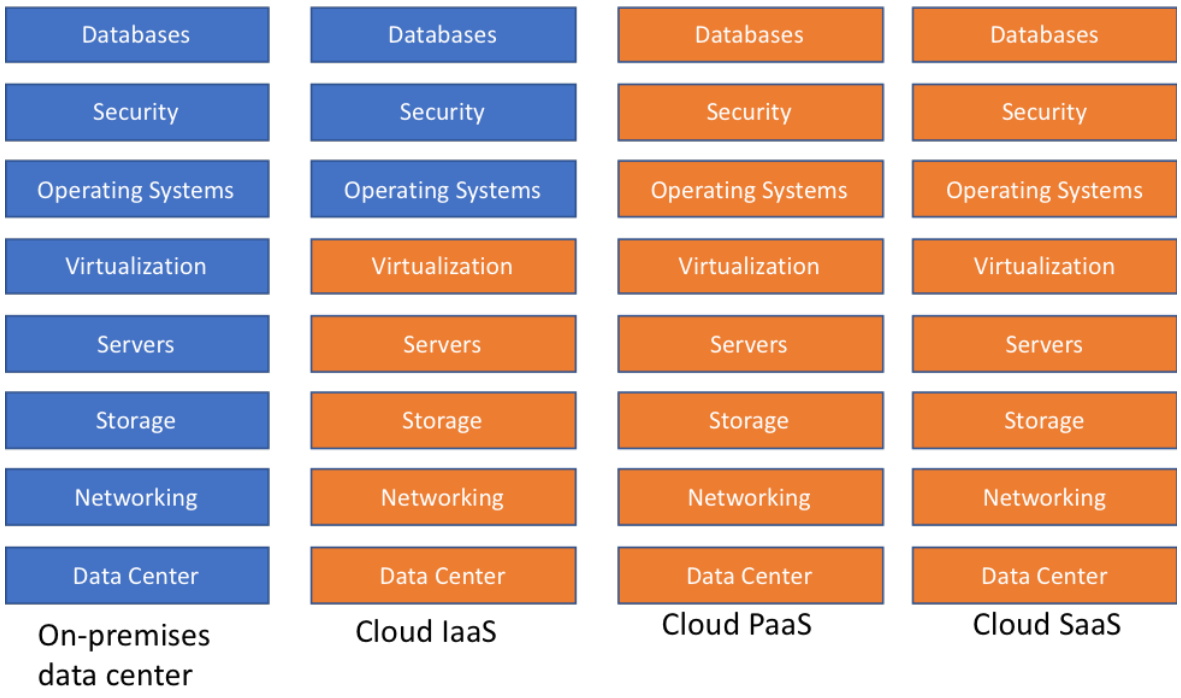


Figure 2-3 "As a service" compared

In this diagram, the services in blue are provided by the customer (you), and the services in orange, from the cloud. When IT is run from a traditional on-premises datacenter, the customer is fully responsible for the entire stack: from the physical datacenter all the way through to the application.

In an IaaS model, the lower components—physical datacenter, networking storage, servers, storage, and virtualization services—are offered by the cloud. Typically, the customer will then supply one or more virtual machines—perhaps created with the Microsoft Hyper-V or VMware tools—comprising the application and its required system software (operating system, database, and so on).

Platform as a service brings more to the table: the customer only focuses on the application software, as all system software—operating system and so forth—

is provided by the cloud. And in software as a service, a vendor makes an entire application available in the cloud for you to use; examples include the Microsoft Office 365 and Dynamics 365 suites.

Advantages and disadvantages

It would appear that IaaS offers perhaps the least value. In fact, in many ways, the opposite is true: IaaS usually offers the quickest "on-ramp" to the cloud. Since most corporate applications run as virtual machines in datacenters today, it is a relatively simple matter to move them from the on-premises datacenter to the cloud datacenter intact.

With IaaS, the customer supplies the operating system and other systems software, which can be a potential disadvantage. Just as in the on-premises datacenter, you need to ensure that patches and upgrades are applied in a timely fashion, and you need to ensure that systems software is appropriately licensed (but as we discuss below, you can transfer on-premises licenses for Windows Server and SQL Server to instances in Azure).

PaaS applications by contrast, while being more "cloud native" in many respects, usually require some rearchitecture to take advantage of native-cloud services, which can be time-consuming and expensive.

Finally, SaaS applications can be the most cost-effective, and for commodity functions—email, collaboration—SaaS is the best bet.

The lines are blurring

In the past, the distinctions between these different cloud architectures were sharp. Over time, however, the lines have begun to blur. Where in the past, only PaaS applications could take advantage of the cloud's ability to scale elastically, today IaaS applications can scale up in response to high demand and scale down as the load decreases. IaaS applications can tap into all the cloud resources that PaaS can.

Hosting models

In the last few sections, we discussed the "how" of cloud computing. Below, we discuss the "where."

Private clouds

It's possible, of course, to build a "cloud" in your datacenter. By purchasing a large number of servers and corresponding amounts of storage, you can create an environment in which you can make computing available to your users in an "on-demand" fashion. If you have extremely sensitive data, or if for regulatory compliance reasons you cannot move your applications and data out of your datacenter, your own cloud in your datacenter might be a good—if expensive—option. Using Azure Stack, you can even create an on-premises environment that mirrors that of the cloud.

The hyperscale public cloud

Of course, the public cloud—such as Microsoft Azure—is where the action is today. Azure, with millions of servers worldwide, supports hundreds of thousands of customers and millions of applications.

In the public cloud, computing can operate at *hyperscale*, meaning that computing resources scale with the demand placed on them. Hyperscale computing provides the ready availability of whatever computing capabilities you need, whenever you need them. Thus, if you need 10,000 servers for an overnight big data analytics job, but only for a few hours, you'll have them, and then you can release them when finished. Hyperscale also implies the notion of configurability (and reconfigurability) at scale. Today, a given server might be allocated to a particular real-time application with a very high service-level agreement (SLA); tomorrow, it might be assigned a background task with a very different SLA, all at the request of the consumer of cloud functions.

Hybrid cloud

For many companies, a hybrid cloud—in which some of its applications are in the cloud and others remain on-premises—is essential, for a variety of reasons:

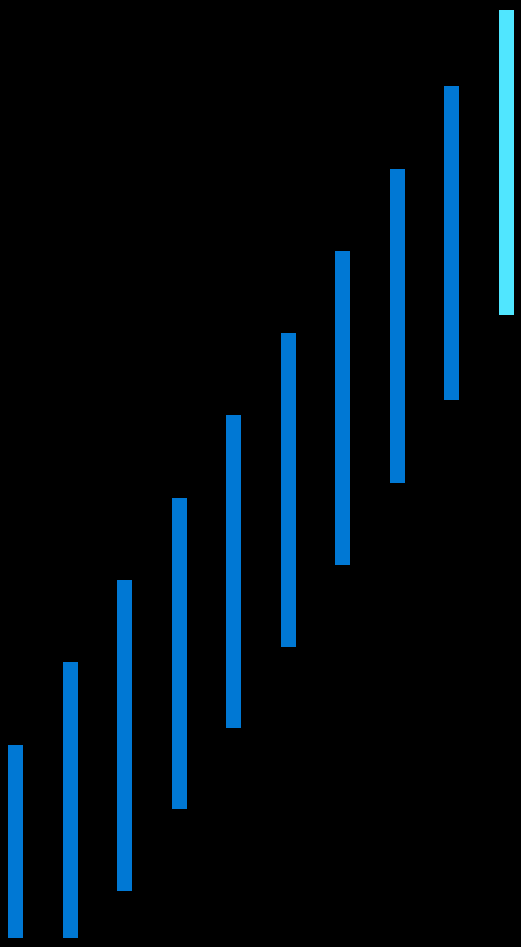
- In some cases, legacy systems (for example, mainframes) cannot be easily migrated.
- Occasionally, for policy reasons, some applications cannot be moved.
- Simply, moving all the applications to the cloud takes time.

We talk more about the mechanics of a hybrid cloud later.

Government cloud

Governments often have unique requirements; for example, such that only citizens of their country may work in the cloud datacenter or for specific handling and treatment of classified data. For this reason, Azure offers dedicated instances of the cloud for the US and other governments supporting all the same features as the public cloud but limited tgovernment access and use.

03 / Common Scenarios for Infrastructure as a Service



In this chapter, we'll describe some of the most common use cases for the IaaS model of cloud computing.

DevTest

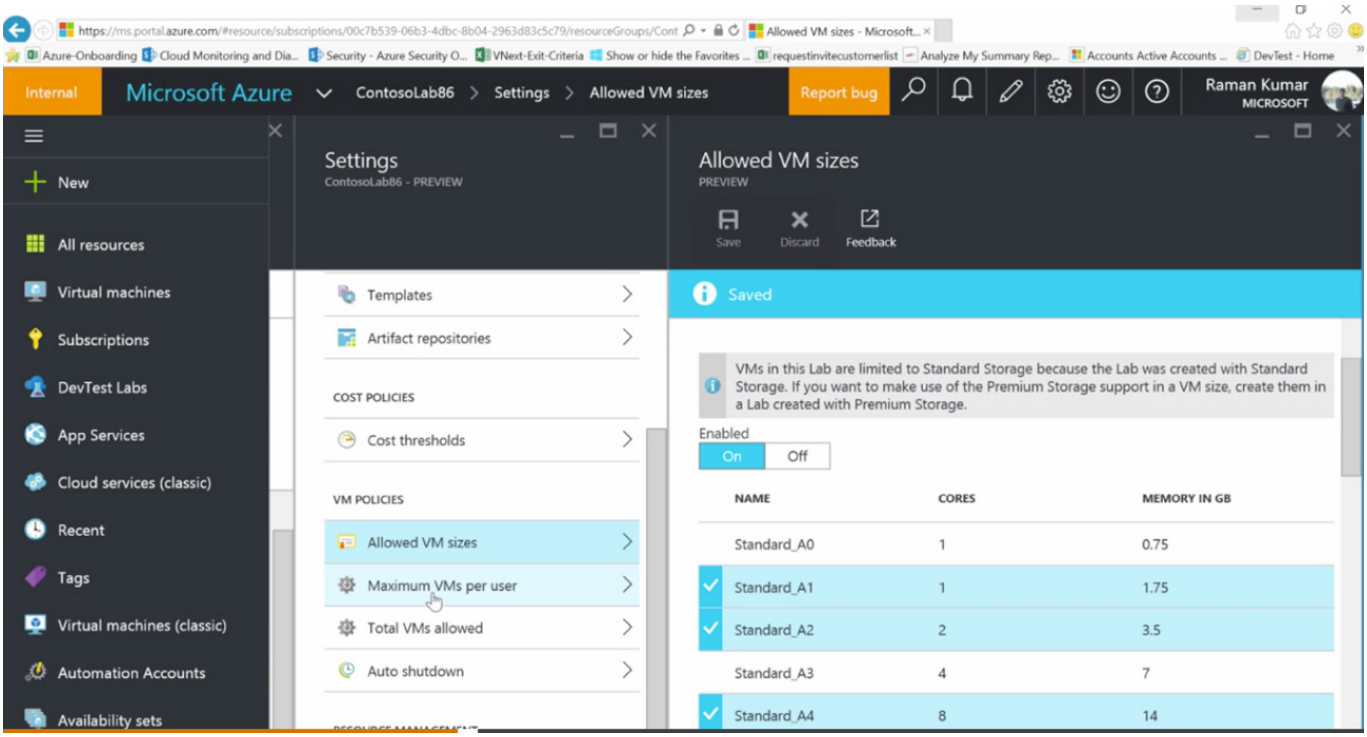
One of the best places to start with cloud deployments at scale is in development and testing. You can quickly show the cost benefits of cloud computing as the multiple environments (development, testing, user acceptance testing, and so on) can be expensive and do not provide direct business value. Additionally, developers are typically more tolerant of problems than production users, so if you do encounter problems, developers and IT can learn from their mistakes.

Teams can quickly set up and dismantle development and test environments, bringing new applications to market faster. IaaS makes it quick and economical to scale dev-test environments up and down, and Azure tooling ensures that teams do not exceed usage quotas.

Try it for free! Click here to get an instance of Azure DevTest Labs.

In fact, with Azure DevTest Labs, your teams can develop and test in the cloud in a self-service, controlled fashion. DevTest Labs allows you to allocate servers to do development. A separate set of servers can be spun up—under configuration control—at a certain hour of the day (say, at night) to run tests, and then deallocated as the tests complete or at a particular time. As with an on-premises lab, policies can be created that regulate what kind of test machines are used, how many can be allocated for each user, and when the project ends (an expiration date) for the lab.

Figure 3-1 Setting allowed VM sizes for a development/test lab



Want to try it? Click here for a test spin of Azure App Service for a limited time without a subscription, free of charge and comment.

Website hosting and web applications

IaaS provides all the infrastructure to support web apps, including storage, web and application servers, and networking resources to support websites and web applications. Organizations can quickly deploy web apps on IaaS and easily scale infrastructure up and down when demand for the apps is unpredictable.

Migration

As we've mentioned, IaaS VMs are an easy and straightforward destination for applications that you're moving out of your datacenter. Because, in all likelihood, your on-premises applications are running as virtual machines, it's relatively simple to move them to the cloud: it's a matter of sizing your cloud VMs appropriately and allocating the correct number of cloud servers.

We discuss the details of this shortly.

Storage, backup, and recovery

Organizations can avoid the capital outlay for storage and complexity of storage management, which typically requires a skilled staff to manage data and meet legal and compliance requirements. IaaS is useful for handling unpredictable demand and steadily growing storage needs. It can also simplify planning and management of backup and recovery systems.

Read more about Azure Storage and recovery, and success stories from customers using it, here.

High-performance computing

High-performance computing (HPC) on supercomputers, computer grids, or computer clusters helps solve complex problems involving millions of variables or calculations. Examples include earthquake and protein-folding simulations, climate and weather predictions, financial modeling, and evaluating product designs.

Distributed HPC workloads often require extremely fast interconnects, and for those, Azure uses InfiniBand and remote direct memory access (RDMA). Azure also offers Cray hardware in the cloud for supercomputing workloads.

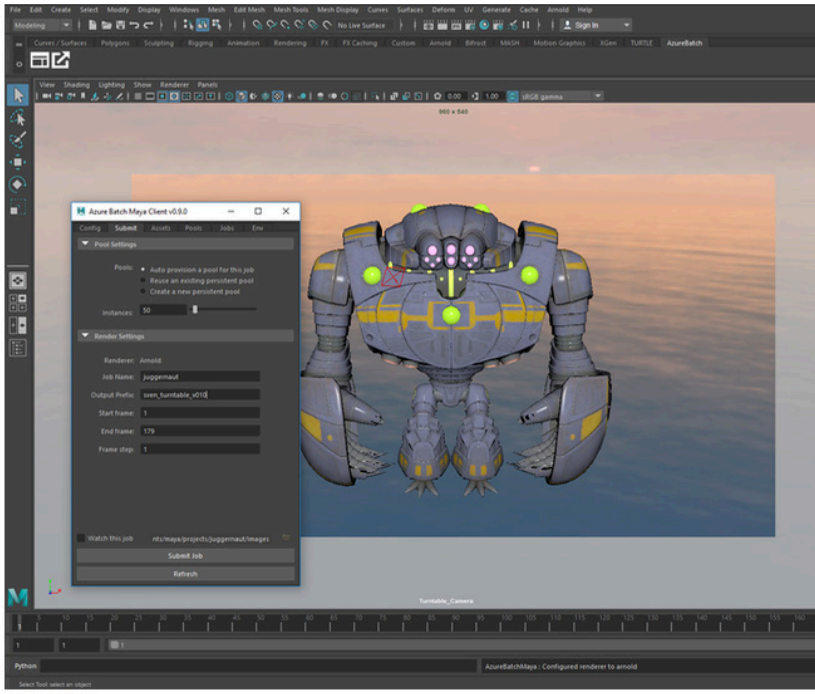
Big data management and analysis

Big data is a popular term for massive data sets that contain potentially valuable patterns, trends, and associations. Mining data sets to locate or tease out these hidden patterns requires a huge amount of processing power, which IaaS economically provides.

GPU-centric applications

Many diverse types of modern applications can now take advantage of the graphics processing unit, or GPU.

Using Azure Batch Rendering, customers can automate and scale up their graphics development pipelines, using as much graphics horsepower as they need, with plug-ins available for many popular design tools (see above).

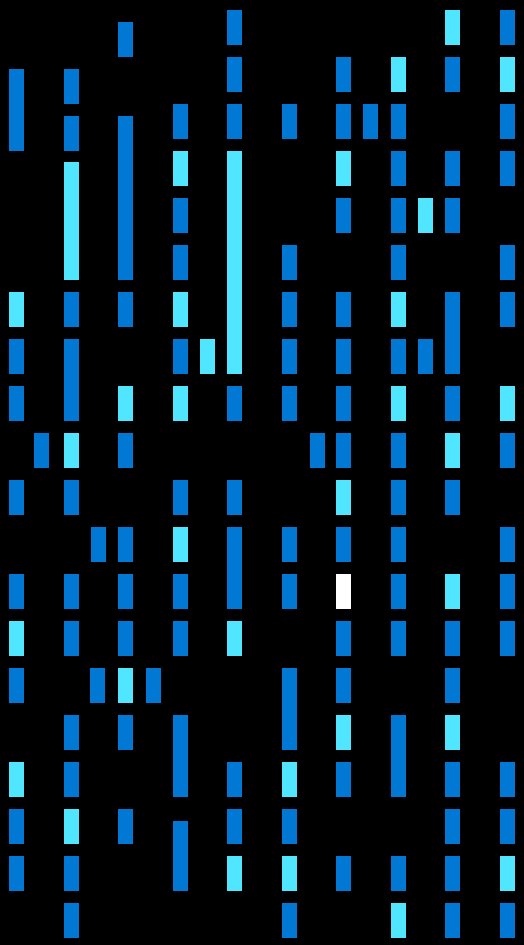


In addition, artificial intelligence and machine learning applications—including CNTK, TensorFlow, and Caffe—can be significantly accelerated using GPUs. For these, Azure offers virtual machines with NVIDIA Tesla GPUs, specifically designed for AI workloads.

Specific workloads

The SAP HANA platform is widely used for enterprise resource planning and in-memory data analysis at scale. Azure can support HANA in a VM with up to 4 terabytes of memory, and on a "bare-metal" processor (a server supporting only one workload) up to 20 terabytes of RAM.

04 / The Economics of Infrastructure as a Service



In this chapter, we'll look at how you're charged for IaaS services and how you can show significant return on your IaaS investment. Please note that specifics may change over time.

Comparing on-premises application costs to IaaS

Many people will want to get a sense of the potential cost differences between an on-premises datacenter and the cloud.

It's beyond the scope of this e-book to go into a full analysis, and moreover, it will vary from customer to customer. But here's a methodology and some considerations you can use when performing your own analysis.

Consider creating a metric called "cost per operating system instance" (we say this instead of cost/server or cost/application in order to allow for applications not running as VMs and applications that span multiple servers, and so on).

How much will your Azure IaaS deployment cost? Try our cost estimator for free here.

For your datacenter, cost per operating system instance should consider

- facilities (building, real estate, among others);
- electricity and cooling charges;
- network, including both the fiber running to the datacenter as well as the internal networking inside the datacenter;
- hardware (servers, racks, routers, SSD and disk storage, and so on);
- licenses for both system and applications software; and
- operations staff.

In the cloud, you can perform the same analysis, which, broadly speaking, is similar but without the first four bullet points. (Microsoft provides a free online total cost of ownership calculator here, which can help you compare the costs of running on-premises to running in Azure.)

However, there are three very important cost-saving tools you have in the cloud:

1. As we've mentioned, with Azure you can save up to 72 percent of your costs with reserved instances.
2. You can also use existing on-premises licenses for Windows Server and SQL Server to save on Azure with the Azure Hybrid Benefit, which is described here.
3. Typically, on-premises IT will have the *maximum* number of servers provisioned for an application, in order to support spiky usage. In the cloud, you can often provision a much smaller number, and only purchase more capacity when you need it. We talk about how to do this later.

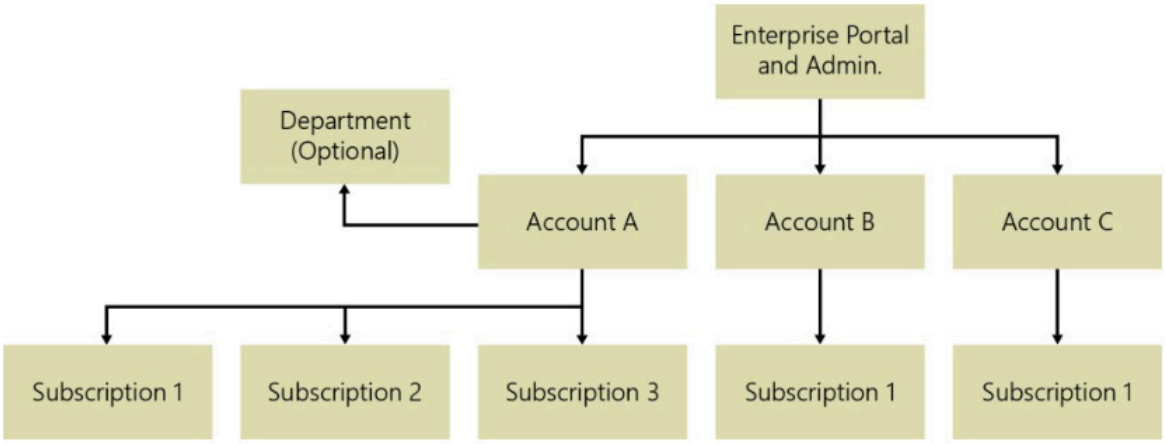


Figure 4-1 Subscription management governance hierarchy

Subscriptions

As you begin to go live with applications in the cloud, you should consider how to manage subscriptions. It's tempting at first to say all your enterprise is on one cloud subscription, but that model results in substantial inefficiency: it is difficult to account for by application, cost center, or department in a single-subscription model. In addition, managing a large number of applications—some in production, some in test, and so on—can be cumbersome; and finally, the administrator of the single subscription can become overwhelmed with new requests for VMs and other resources.

It's usually more effective to assign subscriptions to individual cost centers or even to applications, or application groups (for example, sales apps). This facilitates better visibility into costs by function, and it provides CIOs with a way to assign each group cost targets that the groups then can manage independently.

In a large organization, for better visibility and accountability, you might want to set up a cloud governance hierarchy, as shown below.

In this model, there is a single enterprise-wide portal from which all costs across the enterprise can be viewed. Department-level accounts can contain one or more subscriptions, perhaps for cost centers or for individual solution areas.

Microsoft Power BI gives you a convenient way to visualize your subscriptions and their usage:

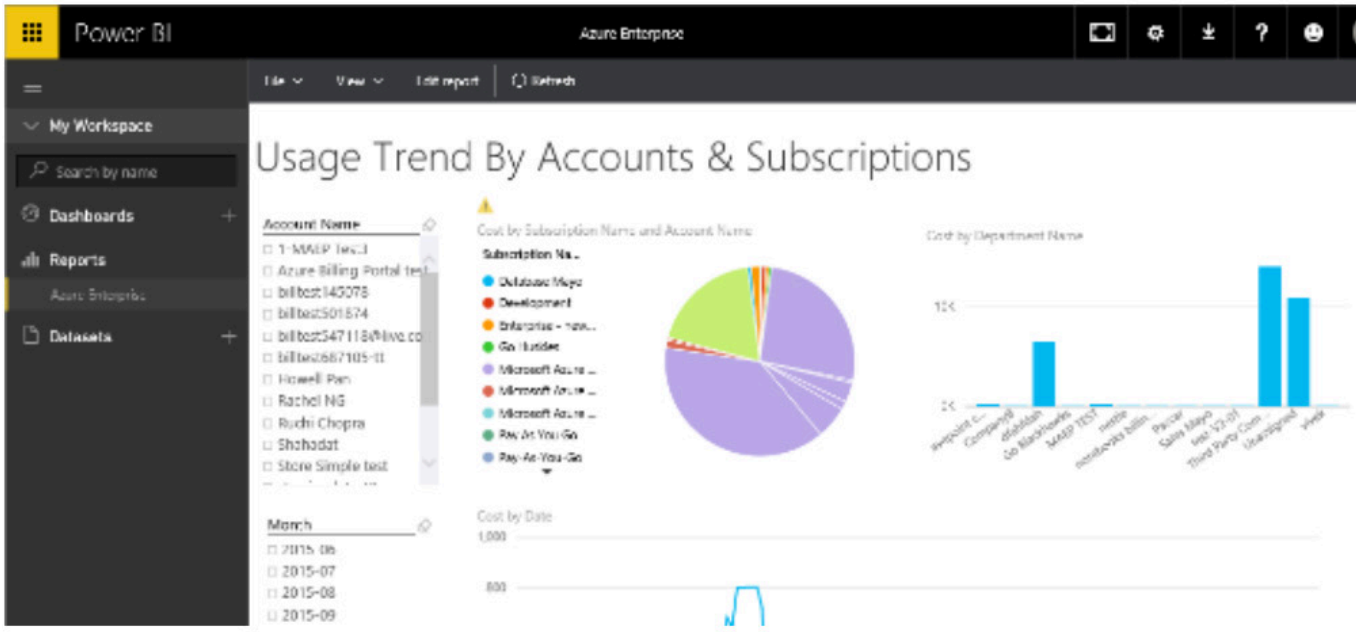


Figure 4-2 Azure usage and cost reporting

How you're charged for Azure IaaS

When planning your IaaS deployment, you have two options for how you're charged:

- **Pay-as-you-go**. In this model you pay for the IaaS capacity you use (by the second). You have no long-term commitment or up-front payments, and you can increase or decrease capacity on demand. The pay-as-you-go model is useful if you're not sure of how much capacity you need, or if you have particularly seasonal or "spiky" (or otherwise unpredictable) workloads.

- **Reserved instances**. An Azure reserved virtual machine instance is an advanced purchase of a virtual machine for one or three years in a specified region. The commitment is made up front, and in return, you get up to 72 percent

price savings compared to pay-as-you-go pricing. Reserved virtual machine instances are flexible and can be easily exchanged or returned.

The reserved-instance model is useful for customers who are looking for budget predictability and (perhaps significantly) reduced cost, and who can commit to long-term usage.

As examples of the cost, a small VM (at this writing in September 2018) in pay-as-you-go mode with 1 GB of RAM and one virtual CPU starts at $0.005 per hour. A much larger machine, with eight cores and 64 GB of RAM is charged at $0.438/hour. (These prices change frequently, so check back here for updated information.)

Return on investment

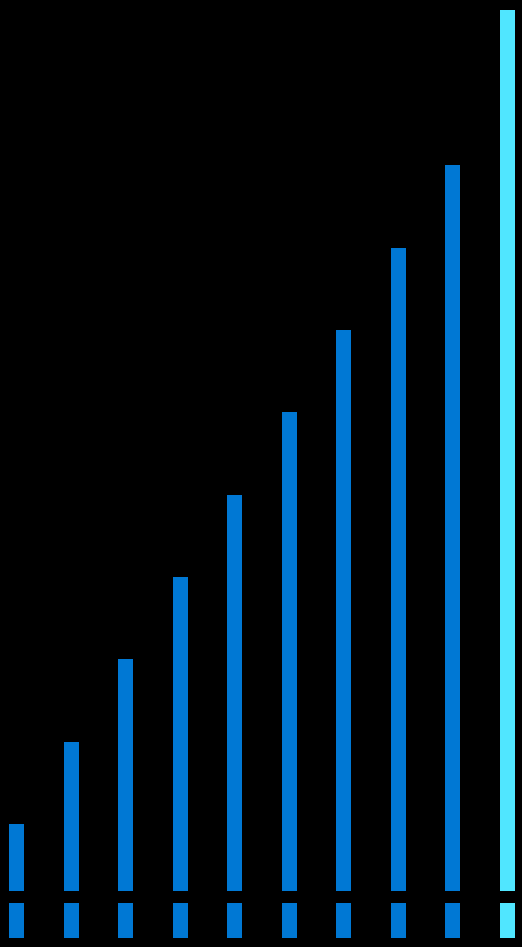
In June 2017, Microsoft commissioned Forrester, a leading analysis firm, to study the financial benefits of Azure IaaS. Their conclusions, stated in a Total Economic Impact report, showed the huge benefits Azure IaaS cloud computing can bring:

- A datacenter reduction of 50–73 percent, leading to cost savings of $12 million. Servers were deployed on-site or hosted with a partner, but still managed by organization resources. Migrating these workloads to Azure reduced internal office space or hosting costs. These benefits add up to a five-year present value (PV) of about $12 million.
- An annual reduction between 33 percent and 83 percent in IT outsourcing needs, leading to $9.8 million in cost savings. Migrating to Azure IaaS enabled significant cost savings. The five-year PV is about $9.8 million.

- Re-engineering that delivers an 85 percent improvement for a key process, leading to $2.8 million in cost savings. Azure IaaS enables process improvement opportunities; a textile manufacturer now provides virtual samples instead of many prototypes, leading to cost savings and faster production times. The five-year PV is about $2.8 million.
- New sales leading to $1.7 million in new income from improved processes, global reach, and better customer service. With better processes and improved global performance, there was an improvement of $1.7 million (five-year PV) in enterprise sales with Azure IaaS.
- Website scale and performance improvements, increasing annual customer sales between 48 percent and 63 percent and increasing transaction size between 20 percent and 27 percent, adding up to $1.2 million in new income. Significant improvements in global scale (such as faster page load times in China) reduce purchase barriers, meaning more and larger sales of about $1.2 million (five-year PV).

The full report is available here.

05 / Infrastructure as a Service: A Deep Dive



In this section, we'll describe the major components that make up infrastructure as a service and how they work together to create a virtual datacenter.

Overview

Think about the devices and software you have in your datacenter today: all these have counterparts in the cloud, and the same amount of design flexibility—in terms of network configuration, server types, storage, and so on—is present in the cloud.

Compute

For your applications, you can select from a wide variety of server capacities in Microsoft Azure. You can choose from machines as small as 1 CPU and 1 GB of memory through very high-performance machines with 128 virtual CPUs and 438 GB of RAM.

In addition, you can select from offerings that are tailored for specific workloads. For example, for database applications, you may wish for a machine with a moderate number of CPU cores but a large amount of available memory.

See the chart below for examples of available options for Linux VMs. For the latest information on server size options, see here (Linux) and here (Windows).

Type	Sizes	Description
General purpose	B, Dsv3, Dv3, DSv2, Dv2, Av2	Balanced CPU-to-memory ratio. Ideal for testing and development, small to medium databases, and low to medium traffic web servers.
Compute optimized	Fsv2, Fs, F	High CPU-to-memory ratio. Good for medium traffic web servers, network appliances, batch processes, and application servers.
Memory optimized	Esv3, Ev3, M, GS, G, DSv2, Dv2	High memory-to-CPU ratio. Great for relational database servers, medium to large caches, and in-memory analytics.
Storage optimized	Ls	High disk throughput and IO. Ideal for big data, SQL, and NoSQL databases.
GPU	NV, NVv2, NC, NCv2, NCv3, ND	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 compute	H	Our fastest and most powerful CPU virtual machines with optional high-throughput network interfaces (RDMA) and 100 Gb/s InfiniBand

Want to create an Azure virtual machine? Here's how.

Storage

Storage and persistence for applications generally fall into one of two categories: simple storage for files and the like, and databases for record-based access with varying amounts of so-called ACID (atomic, consistent, isolated, and durable) capabilities. Azure provides applications a rich variety of functionality in both.

Basic storage options

You have an equally wide spectrum of offerings in the storage space for your Azure IaaS applications. Consider which makes the most sense for you, based on the needs of your application, its performance requirements, and the types of data it uses.

For example, for an application that is being migrated or ported from on-premises you may find it convenient to use Azure Files, thus requiring few or no code changes in file management. For applications with demanding I/O requirements, think about the high-performing SSD options available with an Azure Disk. And for simple, unstructured data—useful for webpage assets, for example—use Azure Blob Storage; or for infrequently accessed data, the cost-effective Cool and Archive storage options.

These forms of physical storage are summarized below:

- **Files.** Azure Files make it easy to migrate your existing file-based applications to the cloud. Azure Files support the industry-standard SMB 3.0 (Server Message Block) protocol, meaning that applications written in standard file APIs will work unchanged in the cloud.
- **Disks.** Azure Managed Disks allow you to create a virtual disk on the medium of your choice. Azure Disks are replicated to three separate replicas for maximum tolerance against failures.

Azure Disk Type	Workload	Size	IOPS
Ultra SSD	Demanding, data-intensive	4 GB–64 TB	100,000–160,000
Premium SSD	High-performance database	32 GB–16 TB	Up to 80,000
Standard SSD	Cost-effective option for less demanding workloads	128 GB–32 TB	Up to 2,000
Standard HDD	Low-cost disks	Up to 32 GB	Up to 2,000

- **Blobs.** Azure Blob Storage is useful for unstructured data, like text files and images, or streaming media used on websites. Azure Blob Storage is massively scalable and can be configured for very fast access times ("hot") or not so fast ("cool" or "archive").
- **Data Lakes.** Use an Azure Data Lake as your storage mechanism for vast quantities of data intended to be processed using tools like Hadoop and Apache Spark.
- **Archives.** For data that is infrequently accessed, inexpensive archive data is an excellent choice, with straightforward "tiering" between hot, cool, and cold layers.

Other basic Azure storage options include tables for easy key-value lookup, and queues, for sending data between applications.

Database options

As with the basic storage choices described above, Azure provides many options for database management. If you are using SQL Server or other relational databases in your datacenter, you can easily migrate them to the cloud. One common approach is to simply migrate the database server as a VM. For example, you can back up your on-premises SQL Server instance (either on Windows or Linux), spin up a virtual machine in Azure, and restore the backup into an instance of SQL Server running in Azure, as described here.

Alternatively, you can deploy to Azure SQL Database, an Azure cloud version of SQL running as a PaaS. Azure SQL Database is designed so that you can migrate without changing your applications, and it features built-in intelligence that learns your unique database patterns and tunes it for improved performance and protection. It also provides enhanced security features and a pay-as-you-go model (optionally, you can "bring your own license").

Azure also provides many other options for bulk data storage, including a number of NoSQL data stores like Cosmos DB, a globally distributed database service, and numerous partner relational database and nonrelational offerings.

Networking

Of course, as you move applications to the cloud, you want them to stay connected to each other and to systems remaining on-premises. To do this, you set up a private network in the Azure cloud. This is a straightforward task in the Azure Portal. If you need higher bandwidth for your datacenter, or if it is required that your data stay off the public internet, you can add a dedicated physical line connecting your datacenter to the Azure datacenter.

VPN options

IT departments can connect VPNs either using software only (called point-to-site) or by using a hardware VPN device (site-to-site). In point-to-site, only one local computer is connected to cloud resources, and it is generally only useful when connecting from home or from a conference, or for testing purposes.

In site-to-site configurations, a specialized hardware VPN device creates an encrypted (using IPsec, with Internet Key Exchange) tunnel between the datacenter and the cloud. IP addresses are configured in the device such that cloud resources appear to be on the local network.

VPNs of this sort can be set up across multiple on-premises datacenters, as shown below:

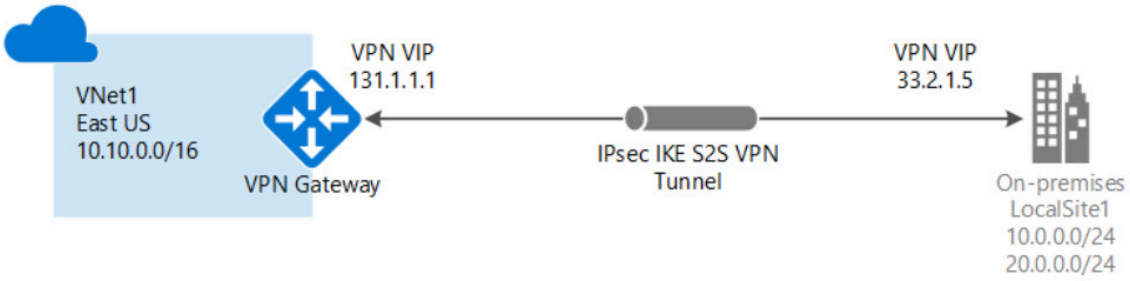


Figure 5-1 Hardware VPN

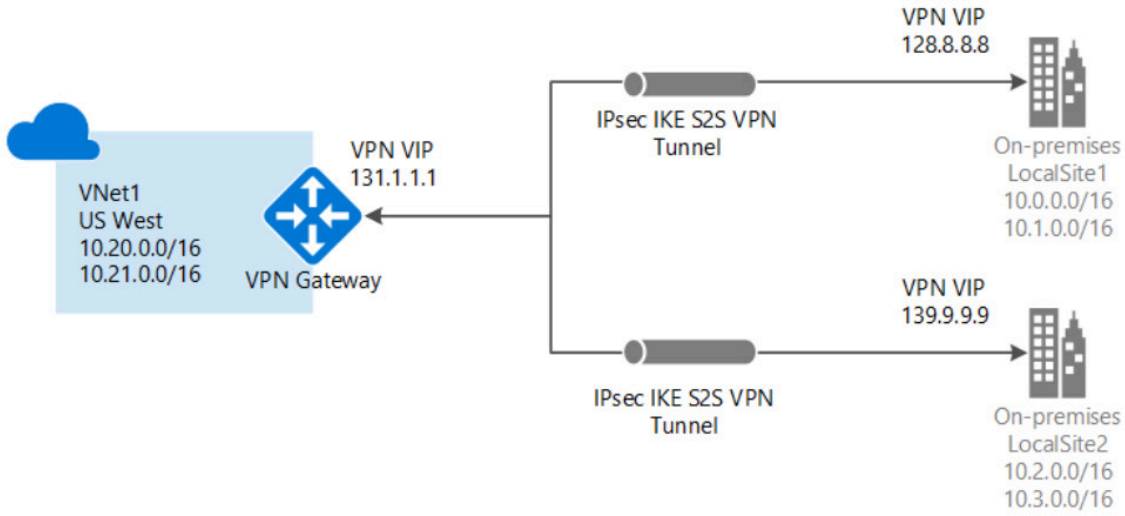


Figure 5-2 Multisite VPN connections

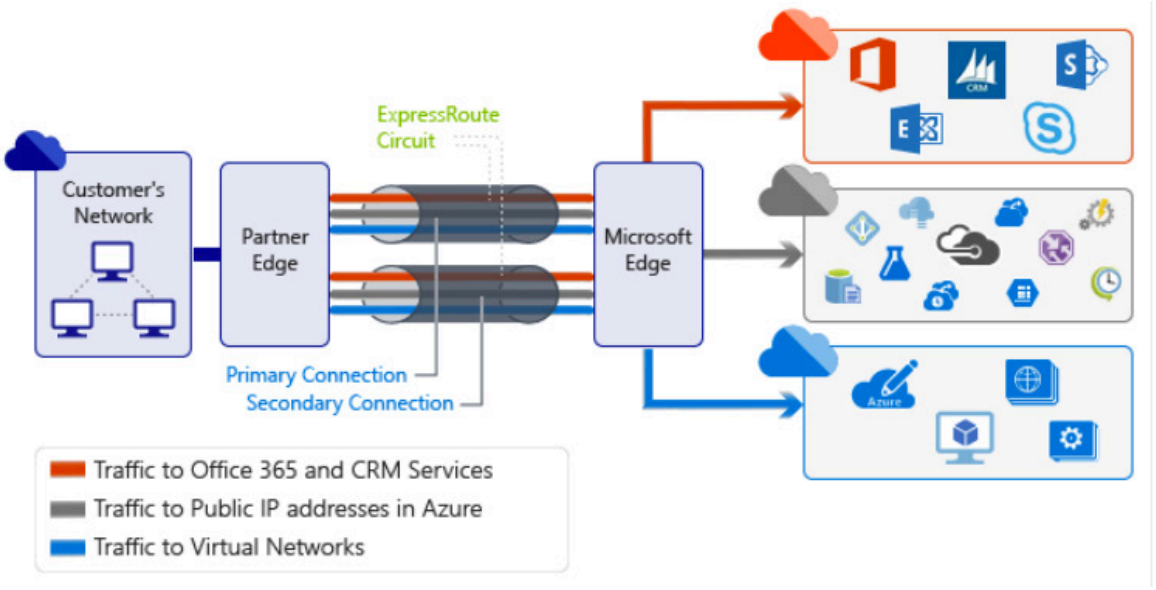


Figure 5-3 ExpressRoute

ExpressRoute

With a dedicated line, such as Microsoft Azure ExpressRoute, enterprises can connect directly from their site to the cloud. However, dedicated lines must be purchased from the local telecommunications provider, and you will need the appropriate edge router and other hardware installed at your site.

Such dedicated lines have the following advantages:

- You can typically purchase guaranteed bandwidth from your telco.
- ExpressRoute can be used to connect to any Microsoft cloud service.
- Messages do not go over the public internet, for an added layer of security.

However, dedicated lines such as ExpressRoute will incur additional costs, depending on bandwidth desired, rates defined by the selected telco provider, and so on.

ExpressRoute Global Reach

For global cloud services with demanding data transfer requirements, Azure sports a feature called ExpressRoute Global Reach, which allows applications to transmit data over the Azure fiber network, from an Azure datacenter in one region to an Azure datacenter in another.

App Service Hybrid Connections

An Azure App Service Hybrid Connection allows you to access a specific resource on another network from your Azure application, such as a mainframe or on-premises database. Each Hybrid Connection correlates to a single TCP host and port combination. This means that the Hybrid Connection endpoint can be on any operating system and any application, provided you are accessing a TCP listening port.

Migrating your applications

Microsoft Azure provides a number of tools to help you migrate applications to the cloud, including Azure Migrate, Azure Site Recovery, Azure Data Migration Assistant, and Azure Data Migration Service.

For much more detail, see *Cloud Migration Essentials*, and this series of articles that describes how a (fictitious) company named Contoso managed a large migration project.

Optimizing your spend

Here is an outcome from moving your applications to the cloud that you may not expect: you may discover that you are actually *spending more* in the cloud than on-premises.

What happened? After all, we have spent much of this book advancing the idea that the cloud will save your company both time and money!

One very common reason for this undesired situation is that when applications are initially moved to the cloud, their configurations are more or less replicated exactly. That is, if you had eight servers devoted to the application in your on-premises datacenter, it's likely that in the initial move eight IaaS cloud servers were allocated.

Of course, the reason you had eight servers allocated to the application in the first place was that you needed these to handle peak capacity loads: most of the time their CPUs operate at single-digit utilization.

Your datacenter servers also likely ran 24/7—yet usage was confined to working hours! That made little difference on-premises, but spinning down unneeded resources in off-hours can save a lot of money.

Here is where the cloud and DevOps can show their value. By providing ongoing monitoring through tools such as Application Insights described above, you can understand day by day (or for that matter, minute by minute, if you're so inclined) what your applications are doing.

It's not at all uncommon to find that a significant percentage of your servers are running in single-digit utilization—and that can lead to important savings.

If, for example, you see that you have eight servers all running at 6 percent CPU utilization, you can consolidate that load on to two servers and return the rest to the pool—and now you will only be charged for the two. When applying this technique broadly through your application portfolio in the cloud, you should see considerable savings.

In the following example, a particular IaaS application was monitored and CPU utilization was measured according to the industry standard P95 algorithm. Running on a relatively large server, its monthly costs came to around $1,400.

Here is a chart of the utilization:

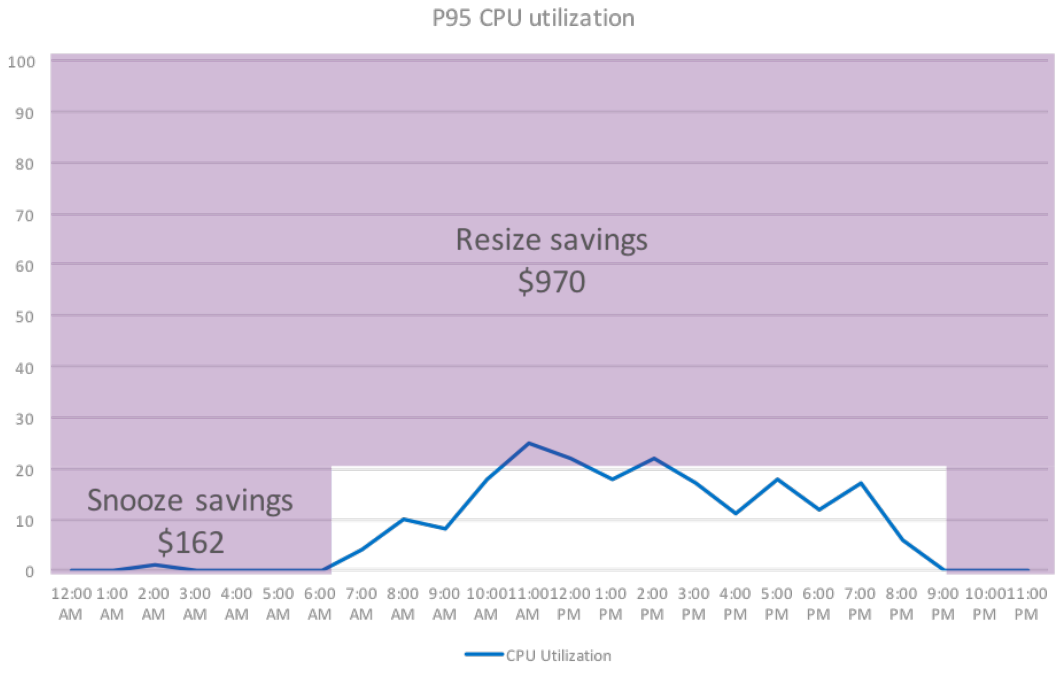


Figure 5-4 CPU utilization of IaaS app

You can see that the application is only active between the hours of 6:00 AM and 9:00 PM, and that its maximum CPU utilization—the *most it ever used* of the processor—was around 25 percent! As a result, the application was put on a "snooze" schedule to take it offline during off-hours—during this time the cloud server resources were freed. They also moved the application to a smaller server more suited to the light load.

Overall, these two simple actions saved more than a thousand dollars per month.

With Azure Cost Management, you can achieve the same sort of results. Azure Cost Management, which can also help you track expenditures on other cloud providers, can generate reports (an example shown below) for cost allocation and showbacks/chargebacks as well. And, as mentioned above, Azure Cost Management can help you optimize your cloud spending by identifying underutilized resources, which you can then manage and adjust.

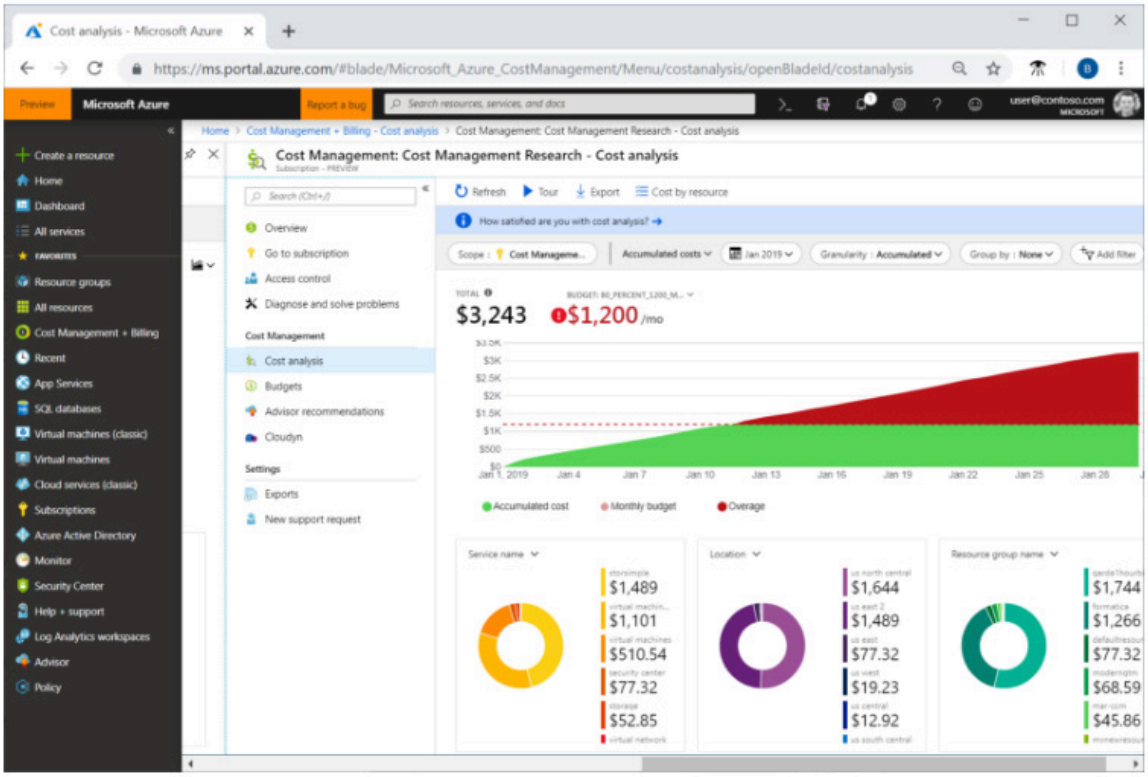


Figure 5-5 Azure Cost Management

Achieving resilience and scale

The cloud offers many capabilities to ensure you achieve your resilience and scale goals.

Availability sets and availability zones

An *availability* set is a logical grouping capability for isolating VM resources from each other when they're deployed. Azure ensures that the VMs you place within an availability set run across multiple physical servers, compute racks, storage units, and network switches. If a hardware or Azure software failure occurs, only a subset of your VMs are affected, and your overall application stays up and continues to be available to your customers.

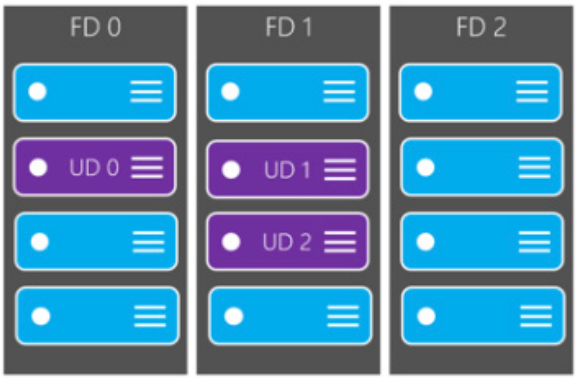


Figure 5-6 Availability sets

Availability sets (and availability zones, below) employ the concepts of *fault domains* and *update domains*. In an availability set, a fault domain represents a group of virtual machines that share a power source and a network switch (a rack or racks). *Update domains* represent sets of machines that can be rebooted at the same time without causing an application to fail, and to ensure resiliency, update domains are distributed across fault domains.

Availability zones extend the concept of availability sets and are unique physical locations within a single Azure region. An availability zone is composed of one or more cloud datacenters. Like availability sets, availability zones have the concept of isolation: in this case meaning separate power, cooling, and networking so that applications are protected against datacenter failures.

There is no extra charge for using availability zones.

Want to get the resiliency that Azure Availability Zones offer? Try it here.

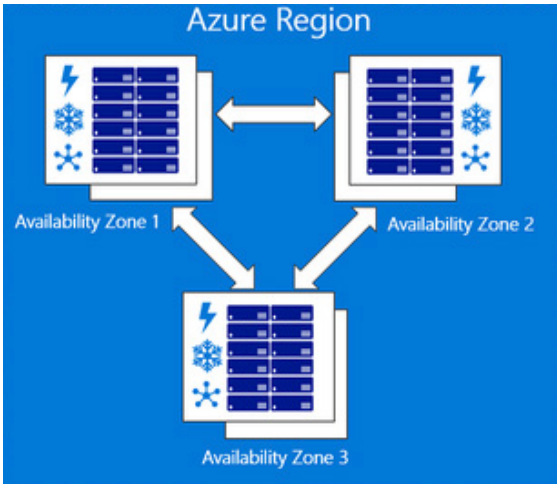


Figure 5-7 Availability zones

Redundancy options for your data

With cloud datacenters distributed around the world, you can take advantage of different regions to achieve data and compute redundancy.

Azure offers four different options for data redundancy:

1. Locally redundant storage (LRS)
2. Zone-redundant storage (ZRS)
3. Geo-redundant storage (GRS)
4. Read-access geo-redundant storage (RA-GRS)

Locally redundant storage protects your data against failure of an individual node within a datacenter. Zone-redundant storage protects against the failure of a datacenter, and geo-redundant, against an outage spanning an entire region. When you enable read-only access to your data in a secondary region (RA-GRS mode), your data is available on a secondary endpoint as well as on the primary endpoint for your storage account.

Handling peak load with VM Scale Sets

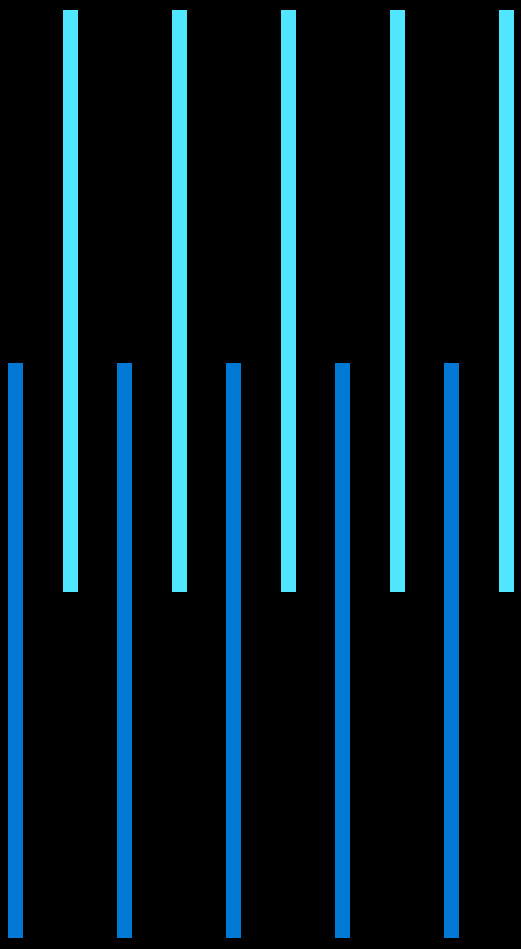
Azure Virtual Machine Scale Sets let you create and manage a group of identical, load-balanced, and autoscaling VMs. The number of VM instances can automatically increase or decrease in response to demand or a defined schedule. Scale sets provide high availability to your applications, and allow you to centrally manage, configure, and update a large number of VMs. With virtual machine scale sets, you can build large-scale services for areas such as compute, big data, and container workloads.

Global scale with Azure Traffic Manager

If your application reaches a global audience, you want your users to have the best possible experience no matter where they're located. For example, if you have an instance of your website in Europe, you want your European customers routed to that instance, while your North American customers are directed to an instance in the US or Canada. With Azure Traffic Manager, you can ensure that requests are routed to the nearest geographical datacenter or, alternatively, to the endpoint with the lowest latency.

Azure Traffic Manager instances can themselves be "nested," that is, progressively routed to the nearest datacenter. For example, a user in Germany may first be routed through a traffic manager instance to western Europe, and then a traffic manager in Europe can route the request to a cloud datacenter in Germany.

06 /
To the Cloud and Back Again



While much of this book concerns moving entire applications and ecosystems to the cloud, it's worth pointing out that such migrations usually take time. It's also often the case that for various reasons some applications or data must stay on-premises.

We call a mixed on-premises/cloud IT ecosystem a hybrid cloud, and in this section, we show how there is much about the cloud that makes on-premises computing more efficient and cost-effective. Many of the more mundane tasks of IT, such as backup and restore, can be performed to and from the cloud inexpensively and securely.

Further, with messaging buses and integration brokers, enterprises can quickly connect to B2B sites. And by extending corporate directories to the cloud, they can propagate secure identity management to cloud resources.

In addition, as you move more and more applications into the cloud, it often becomes useful to mirror cloud computing paradigms on-premises using appliances. For environments where network latency or the volume of traffic must be controlled, or for scenarios in which data cannot leave the datacenter, Azure Stack provides Azure-style IaaS and container programming models in your datacenter. We discuss Azure Stack later.

Backup and restore

One of the most important, if unheralded, functions of an IT department is to ensure that corporate data is never lost, in spite of server crashes, power outages, accidental erasure, and the like. In the past, backup was typically handled by copying the contents of disks to an offline media, like tape, often in the middle of the night, and then transporting that tape to some offsite location.

The cloud offers a new approach to backup, both for on-premises and cloud applications. It's easy to see why: with enormous capacities of cheap storage, with built-in security, and cloud datacenters all over the world, the cloud matches or surpasses the capability of traditional backup solutions.

When thinking about a backup strategy, there are two metrics that will help you formulate your plans:

1. Recovery time objective (RTO): How fast do you need to get your data back?
2. Recovery point objective (RPO): How current must the data be when restored? (In other words, how frequently must you back up? Daily? Hourly?)

There are many cloud solutions for backup and restore, each targeted at a specific workload or scenario. Azure Backup Services, for example, as the name suggests, backs up data to storage in the cloud. The data is encrypted (using AES-256), with as many as six separate copies in two separate datacenter regions (if you choose the geo-redundant option, the datacenters are at least a hundred miles apart). Like everything else in the cloud, Azure Backup is a pay-as-you-go service: you pay for what you use.

You could also consider the backup method. Modern backup technologies, including Azure Backup, allow you to select either a full backup, in which you copy the entire source data; a differential backup, which only stores data blocks that have changes since the initial full backup, or an incremental backup, which copies data blocks that have changed since the previous backup. The most efficient of these, of course, is to do a full backup initially, followed by periodic incremental backups.

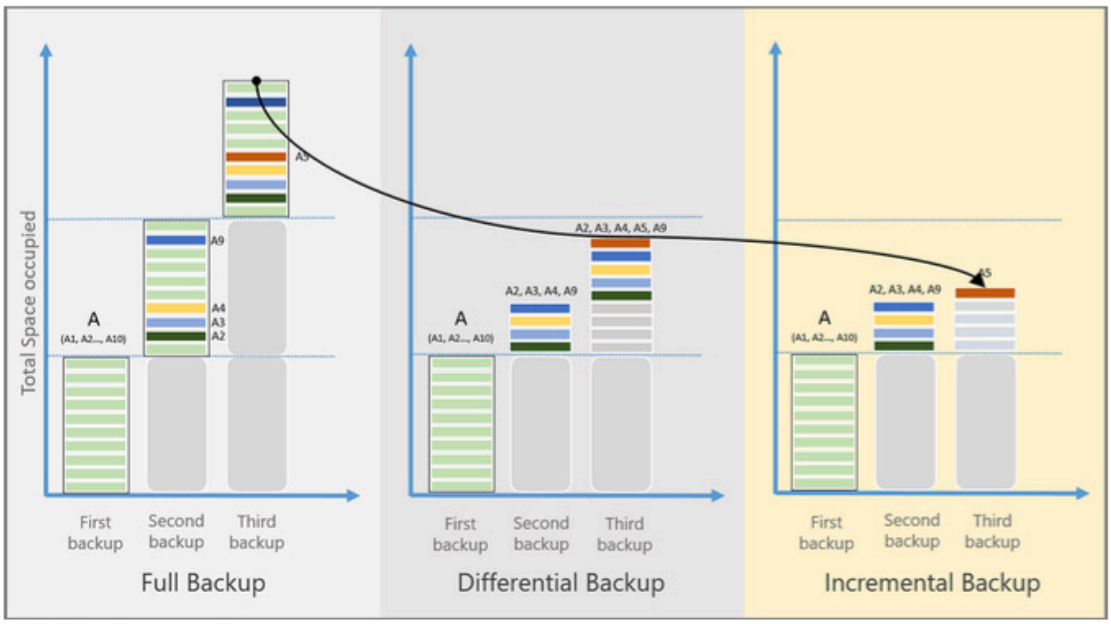


Figure 6-1: Backup modes

Obviously, you will have to select the frequency and type of backup that best suits your needs and meets the RTO and RPO objectives you set. Azure Backup Services ensure that application data is always consistent so that whatever RPO you set, the data will be in a useful state.

If you already have the Microsoft systems management tools installed in your enterprise, Microsoft System Center, you can extend its Data Protection Manager function to back up to the cloud. DPM is a full-featured backup solution that can back up to tape or other media, as well as to the cloud, to the same "data vaults" that Azure Backup Services use. Similarly, Microsoft SQL Server can be configured using Transact-SQL to back data up to the cloud.

Extending on-premises storage to the cloud

IT executives are often faced with regulatory requirements to retain vast quantities of historical data. This data is typically rarely accessed (usually in response to a legal demand), but by law it must be available.

In such cases, a storage appliance that lives on-premises but has knowledge of the cloud can be very useful. Specifically, as storage space begins to run low, such an appliance can offload infrequently or rarely used data to the cloud. Thus data that is needed is still available locally, but the device maintains knowledge of where all the data is, so in response to a regulatory request or other need, it can quickly restore it. The Microsoft StorSimple appliance is such a device.

Business continuity and disaster recovery

One CIO once told the story of a datacenter his company built that, unknown to them, had an ungrounded metal antenna on the roof. Sometime later, lightning struck the datacenter, on that very antenna, causing a catastrophic failure of all the systems inside. IT executives work hard to avoid such disasters, but they do happen, and IT needs to be prepared for them.

The best BC/DR solution is one that seamlessly fails over from the disaster-struck site to another replica, running the same software with up-to-date data. Now, as with simple backup, the concepts of RTO and RPO apply to BC/DR as well, and IT leaders should determine their targets for these metrics as part of an overall BC/DR strategy. You also want to test your BC/DR failover solution periodically—monthly or quarterly—and your BC/DR solution should permit that without interruption to daily operations.

Finally, when the failed site recovers, you need to control the order in which applications are brought back online, as it's not uncommon that applications depend on one another.

With the Azure Site Recovery Service, you can implement a full BC/DR solution in the cloud, ensuring full data consistency, testing without disruption, and customized recovery plans.

Integration

As we've mentioned in previous sections, even if you plan to move your entire application portfolio from on-premises into the cloud, there will be a period of time in which some of your applications remain in your datacenter while others have been migrated. Alternatively, and actually the more likely scenario, you will choose to leave some applications in the on-premises datacenter for the foreseeable future: a hybrid cloud.

In both cases, enterprises will want to have their application portfolio integrated in such a way that all applications continue to run as before, as if they were all on the same network, and with little or no change to user experiences. In the next few sections, we outline a few approaches to ensure this integration.

Application networking from your datacenter to the cloud

To provide integration between applications, the cloud offers a number of approaches. Of course, applications can use standard, REST-based APIs to communicate between on-premises and the cloud.

For high-speed, real-time communications, consider using WebSocket with .NET's library implementation called SignalR. SignalR is useful for highly interactive and responsive web applications, and it is available as a managed service.

The Azure Relay Service allows you to securely expose services within the corporate enterprise network to cloud applications without opening a firewall connection or making other intrusive changes. The relay service supports traditional one-way, request/response, and peer-to-peer traffic. It also supports event distribution at internet scope to enable publish/subscribe scenarios and bidirectional socket communication for increased point-to-point efficiency.

Serverless application integration: Logic Apps

At the highest level of application integration are brokers that implement B2B protocols directly and can also be used to create custom enterprise workflows.

The easiest to use of these are brokers of integration platform as a service (iPaaS), of which Microsoft Azure Logic Apps is a leading example.

Logic Apps permit enterprise developers to connect applications using industry protocols—with no code; they are "serverless," a concept we discuss in more detail in the next chapter.

Logic Apps connectors include EDI X12, HL7 FHIR, XML, SMS, SAP, and literally hundreds of others. Since Logic Apps require no code, they make application integration fast and reliable.

Extending directory services into the cloud

Three key goals in identity management in the enterprise are that

1. users have a "single sign-on" experience to applications both in the datacenter and in the cloud;
2. users should be able to authenticate to applications from outside the corporate network (for example, to work from home); and
3. for certain applications, authentication via external internet authorities (for example, Microsoft accounts, Facebook, or Google sign-in credentials) might be allowed, perhaps with limited privileges.

To achieve these goals, enterprises should consider extending their directory services function into the cloud, for example, with Azure Active Directory.

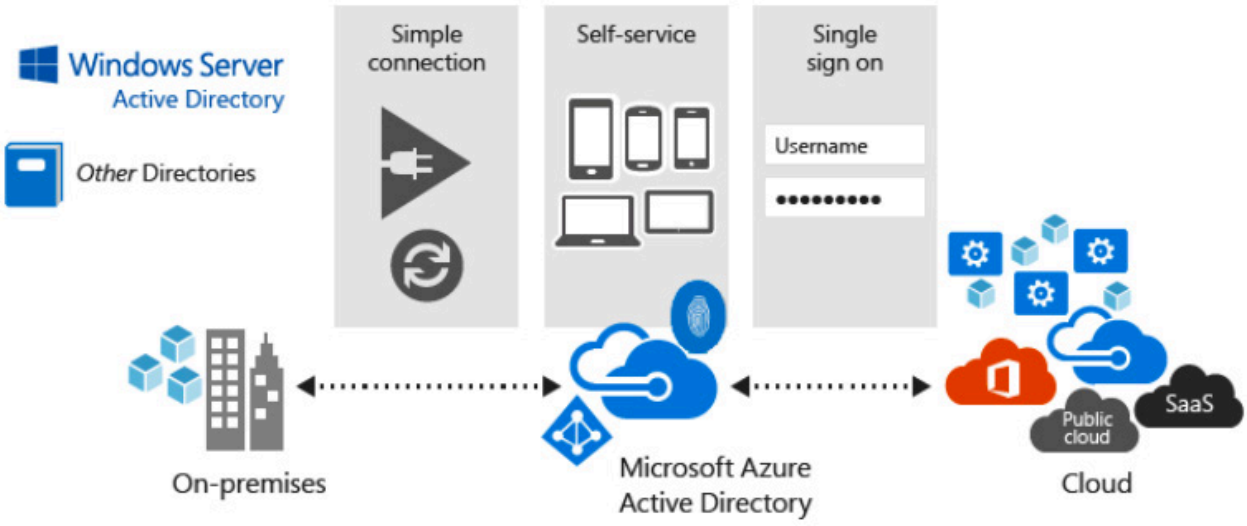


Figure 6-2 Azure Active Directory

Azure Active Directory synchronizes with on-premises directories such as Windows Server Active Directory and others. This enables users to easily sign on once and have access to applications locally in the datacenter and those residing in the cloud. Additionally, users can sign on from outside the datacenter, and AAD will manage the authentication process, coordinating with the on-premises directory. In addition, AAD can manage internet authentication sources such as Facebook and Microsoft accounts.

One of the most important aspects of AAD is its connection to leading SaaS applications; users need only sign on once to access corporate applications and others like Office 365, Salesforce.com, Dropbox, Concur, and many others.

An added feature of AAD enables consumer authentication at scale, for example, for an e-commerce site that wants authentication of its customers.

Cloud computing in your datacenter

As you migrate applications to the cloud, and in some cases perhaps rearchitect them, there may come a day when your staff is more fluent in cloud technologies than in traditional on-premises models.

Alternatively, you may have scenarios in which your cloud applications must have absolutely deterministic latency: that is, certain applications cannot tolerate the variable response time inherent in going over the open internet (for example, manufacturing devices on an assembly line). Or there may be situations in which connectivity to the cloud cannot be guaranteed (for example, on board a ship).

For these types of applications that are, admittedly, uncommon, consider bringing in a cloud “appliance”; that is, server equipment that runs cloud software, such as Azure Stack. Azure Stack consists of packaged cloud software that can be run on selected server platforms.

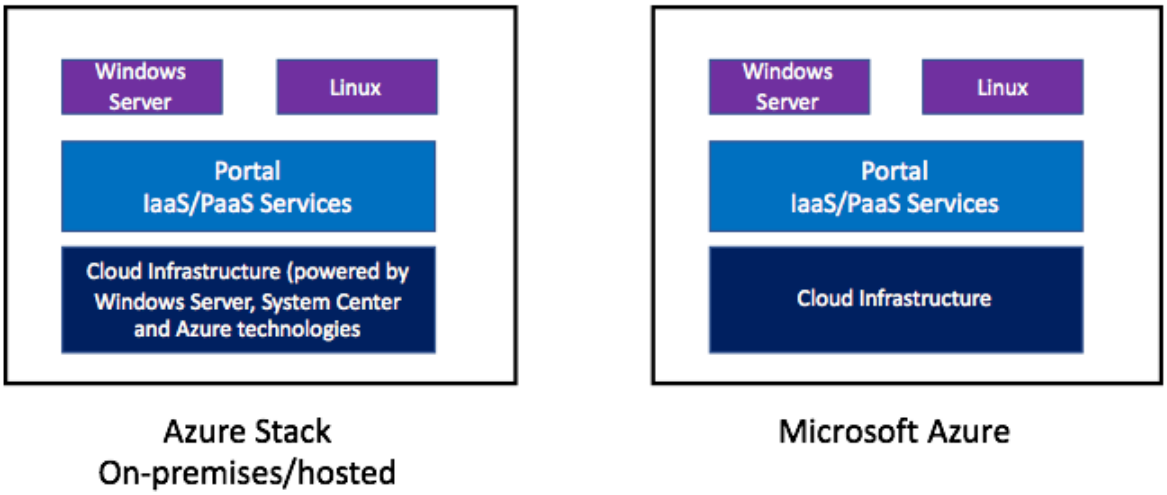


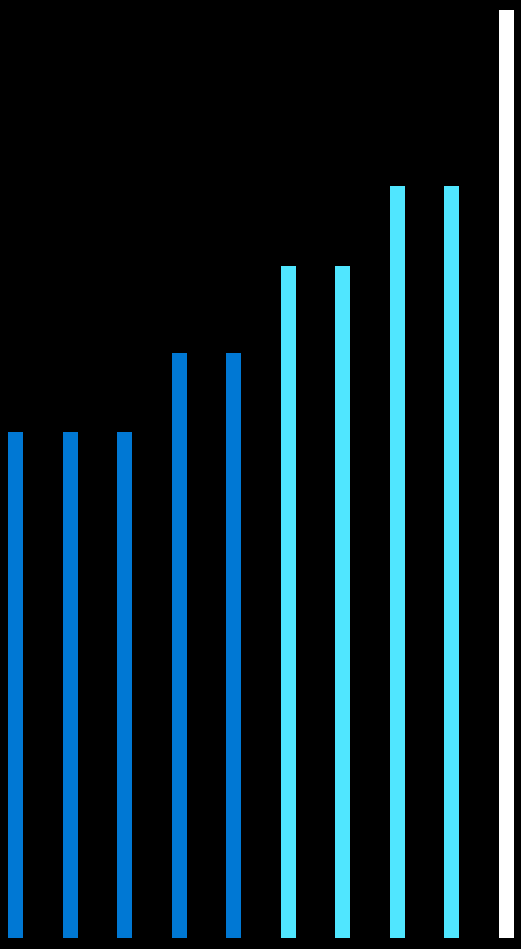
Figure 6-3 Azure Stack

By running cloud services in your datacenter, you can guarantee network latency at local network levels; if for example, you have manufacturing equipment that requires responsiveness within some narrow margin of time, you can use Azure Stack on-premises to eliminate variations in latency caused by the open internet.

Alternatively, if connectivity to the internet cannot be guaranteed, you can use Azure Stack to ensure the availability of your services even if your link to the internet is down or unavailable—a problem faced, as it happens, by a large passenger cruise ship company that used Azure Stack to solve it.

Azure Stack allows you to develop once for the cloud—say, a PaaS application, or serverless function, and deploy either to the cloud or on-premises, thus providing your development teams a single programming model.

07 / Accelerate Productivity with Containers and Orchestration



So far, we've talked about migrating your IT ecosystem to the cloud in terms of moving your virtual machines (VMs). As we've noted, you can save considerable costs by creating such a "virtual datacenter" in the cloud.

Recently, two new technologies have emerged that promise to make your use of cloud resources even more efficient and your teams more productive.

The first of these are containers. *Containers*, as the name suggests, allow you to package applications.

What are containers?

Consider a server hosting three virtual machines. In this case the server holds three, and possibly four (depending on the style of virtualization) full operating systems, with attendant libraries and support code.

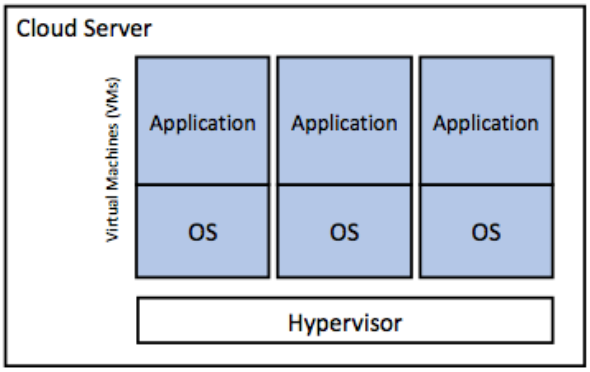


Figure 7-1 Virtual machines in IaaS

By contrast, a server hosting three containerized applications holds one copy of the operating system. The OS guarantees each container full isolation, enabling each container's security.

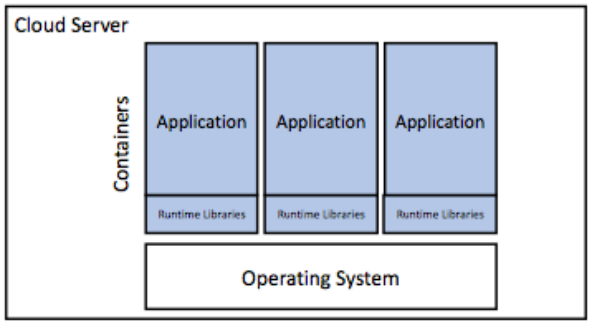


Figure 7-2 Containers in IaaS

Further, each container "believes" it owns a full machine: each has its own IP address, and while all containers share the single operating system's kernel resources, they all function independently.

Thus containerized applications make better use of server resources than VMs. Containers load much faster than VMs, they do not require as much maintenance (patching) as VMs, and they can run faster than applications running in VMs.

It's no wonder then that containers have rapidly become the de facto method of building and deploying cloud applications.

Microservices

Containers lend themselves well to a concept called microservices. Microservices are small units of code that generally perform a single function, such as web serving, search, shopping cart, database management, and so on. Microservices allow developers to break down large application monoliths into parts.

Microservice-based applications have the following advantages:

- Separate teams can develop microservices independently of one another (as long as APIs remain constant).

- Microservices can be deployed and scaled independently. For example, before a holiday season an e-commerce site may see a lot of activity in search, as customers browse for ideas. Closer to the holiday, the search activity may taper off in favor of the transaction microservice, and so on.

Managing large numbers of containers: Orchestration

Of course, most modern cloud applications consist of much more than a single container running on a single server. As we've said, modern applications require the ability to scale up and down in response to changing load, and must be able to recover in the event of hardware or software anomalies or failures.

Here is where a technology called *orchestration* comes in. An orchestration engine allows you to treat large groups of hardware and software as a single resource: you can deploy, start, stop, and manage an entire "mega-application" as a unit.

Properly configured, an orchestration server can, in response to demand, reach into a registry (repository) of container images and instantiate new containers as needed. Orchestration servers can also "watch" for system failures and re-instantiate running containers on new servers when necessary.

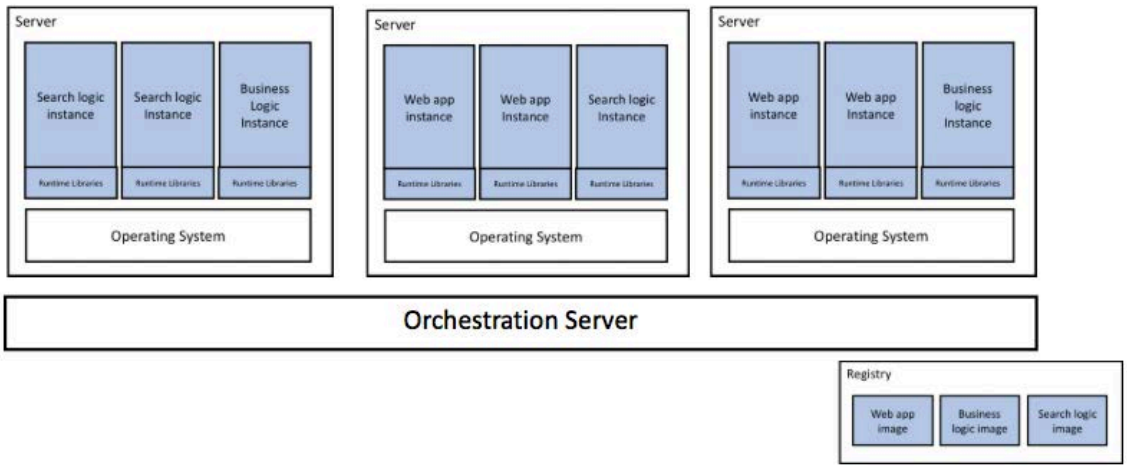


Figure 7-3 Containers in IaaS

Azure cloud orchestration engines

Microsoft Azure offers you a choice of orchestration services.

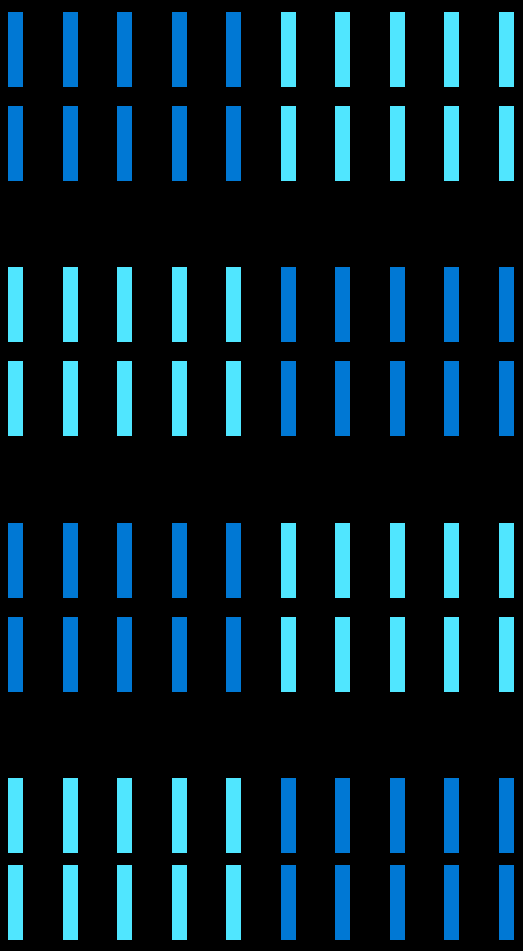
Of course, you can build and deploy an orchestrated application using the open-source Kubernetes engine, Mesos and DC/OS, or Docker Swarm (and others) yourself.

However, Azure offers several services that can make your life much easier:

- Azure Kubernetes Service (AKS) is a fully managed implementation of the Kubernetes orchestration technology. AKS provides the full benefits of Azure to Kubernetes applications, including management through the Azure Portal, deployment globally to datacenters worldwide, integration with Active Directory and other tools for identity management and security, and a wide variety of other capabilities.
- Azure Service Fabric is a battle-tested orchestration engine that originated from technologies used to support products like Microsoft Bing, Skype for Business, and other core Azure services. Service Fabric supports a variety of application models including containers, actors, and full executables, providing a wide variety of services to make your applications scalable and resilient.
- Azure Container Instances (ACI) is a "serverless" version of AKS, meaning that developers no longer need to concern themselves with physical clusters and servers. With ACI, as new instances of containers are created, resources can be allocated from a pool of "warm" servers automatically.

AKS and Service Fabric are both open source and support both .NET and Linux.

08 / Create, Manage, and Maintain Your IaaS Ecosystem



You have many options and resources available to you for creating and managing your IaaS ecosystem. These range from the Azure Portal—where you create, start, stop, and scale your IaaS VMs—to technologies where you can declaratively create and manage templated solutions consisting of multiple individual components and applications working together.

Azure Portal

The Azure Portal is your first stop in creating and managing your cloud environment, including web apps, databases, virtual machines, virtual networks, storage, and Visual Studio Team projects. The highly customizable portal gives you fine-grained access control, allowing you to select who can manage what, by subscription, service, and operation levels.

From the portal you can choose from more than 3,000 services from Microsoft and its partners—many at no charge. You can set your applications to scale automatically, or scale them manually through the portal user interface. And through the portal you have visibility into current and projected costs.

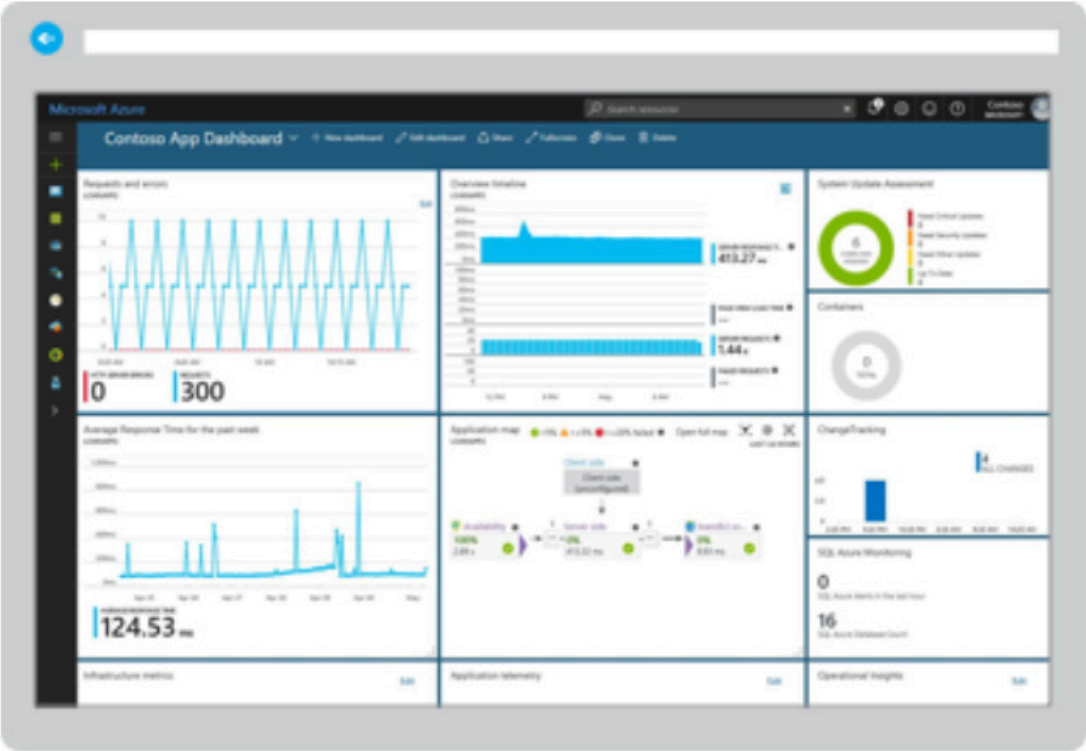


Figure 8-1 The Azure Portal

The portal can be run from within a browser on your desktop, or as a mobile application on iOS or Android.

A command-line interface, the Azure Cloud Shell, gives a fully authenticated shell experience in either Bash or Windows PowerShell.

For more details on the Azure Portal, see here, or get familiar with it through Azure trainings.

Azure Resource Manager

Most corporate and enterprise applications actually consist of multiple components: for example, a web app, a database server, a user feedback service, and analytics services. The Azure Resource Manager allows you to group all these components together so you can effectively deploy and manage them as a unit—improving the consistency and reliability of your deployments.

The Azure Resource Manager relies upon a file that declares the resources you wish to use; here's a simple snippet for example (giving a name to a Linux VM):

```
{
 "type": "Microsoft.Compute/
virtualMachines",
 "name": "demoLinuxVM",
 ...
}
```

Of course, most ARM files are much more complex. Fortunately, there are numerous prebuilt templates available for you to choose from, which you can either use straightaway or customize for your particular application. You can find many prebuilt ARM templates with these Azure quick-start templates.

DevOps and "infrastructure as code"

The Azure Resource Manager is one piece of the overall Azure DevOps story. DevOps, taking its cue from lean manufacturing methodologies, strives to make the code-to-deployment process faster and more efficient. Defining infrastructure in advance as ARM templates—as code—is one way to do that.

In addition, by using features in Microsoft Visual Studio, the integrated development environment, or other tools, developers can set up release "pipelines." Here's the idea: as a developer checks a code change, the pipeline automatically starts a series of automated tests and, if appropriate, automatically deploys the code directly to the cloud service. This is called continuous integration / continuous deployment, and by using these methodologies, developers can make many reliable, tested code changes per day, therefore enabling rapid responses to new market conditions, simultaneously running numerous live "A/B" experiments, and so on.

Azure management and monitoring

Azure provides a wealth of tools to help you monitor the performance and health of your services, applications, and solutions. Services such as App Insights and Log Analytics can help you understand the health of your applications.

Application monitoring with App Insights and Log Analytics

With Azure Application Insights, you can get rich real-time performance monitoring, powerful alerting, and easy-to-consume dashboards to help ensure your applications are available and performing as you expect. With rich analytics that can be fed into machine learning applications you can detect and predict anomalies, failure counts, performance changes, and Azure Cloud Services behaviors.

As the name suggests, you can also use Log Analytics to monitor log data from both cloud and on-premises environments. You can query, correlate, and analyze data from a variety of different sources, and use various analytic tools including Application Insights.

Log Analytics requires minimal configuration and is already integrated with other Azure services. You just need to create a workspace to enable collection. You can then install agents on virtual machines to include them in the workspace and enable management solutions that include logic to provide additional insights into different applications. Behind the scenes, data types are either predefined or automatically created as data is collected.

Log Analytics also powers other monitoring solutions, including

- Container Monitoring, which helps you view and manage your container hosts, and
- Azure SQL Analytics, which collects and visualizes performance metrics for Azure SQL databases.

Core monitoring with Azure Monitor

With a diverse set of applications running in different datacenters and some in the public cloud, management of all this can become a challenge. An application that can manage all of them—a "single pane of glass" as it is sometimes called—can traverse the boundaries of operating systems, datacenters, and clouds to provide administrators with a consolidated view of their ecosystem.

Azure Monitor collects base-level infrastructure metrics, statistics, and logs for most services in Azure, accessible from the Azure Portal. It collects Activity Logs (which track all operations performed on Azure resources), service health issues, and metrics, and allows you to configure alert rules.

Network monitoring

Numerous tools help you monitor your cloud networks. These include:

- Network Watcher provides scenario-based monitoring and diagnostics for different network scenarios in Azure. It stores data in Azure metrics and diagnostics for further analysis. It works with the following solutions for monitoring various aspects of your network.
- Network Performance Monitor (NPM) is a cloud-based network solution that monitors connectivity across public clouds, datacenters, and on-premises environments.
- ExpressRoute Monitor is an NPM capability that monitors the end-to-end connectivity and performance over Azure ExpressRoute circuits.
- DNS Analytics is a solution that provides security, performance, and operations-related insights, based on your DNS servers.
- Service Endpoint Monitor tests the reachability of applications and detects performance bottlenecks across on-premises, carrier networks, and cloud/private datacenters.

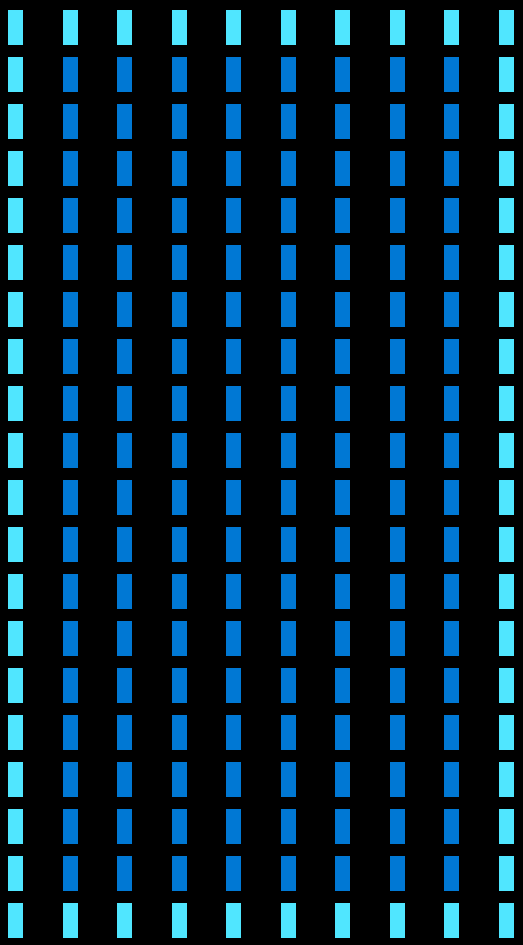
Operations Management Suite

Previous technical documentation related to Azure management includes Operations Management Suite, which is a bundling of the following Azure management services:

- Azure Automation
- Azure Backup
- Log Analytics
- Site Recovery

These services continue to be available individually.

09 / Governance, Risk, Compliance, and Security



Perhaps the most responsible and important question any IT executive can ask is if applications in the cloud are secure and compliant with relevant regulations and standards. The answer can be an unqualified yes—if the appropriate technologies and controls are applied.

As we shall see in this chapter, moving applications to the cloud does not negate many of the "traditional" roles of IT. Security and ensuring that enterprise data is properly managed, costs are controlled, and change is managed appropriately remain key areas of responsibility—but how enterprises govern themselves changes when in the cloud.

Governance, Risk, and Compliance (GRC)

Governance, risk management, and compliance ensure that a business

- follows proper procedures;
- avoids unnecessary uncertainty and risk in its operations; and
- complies with relevant laws, regulations, standards, and conventions.

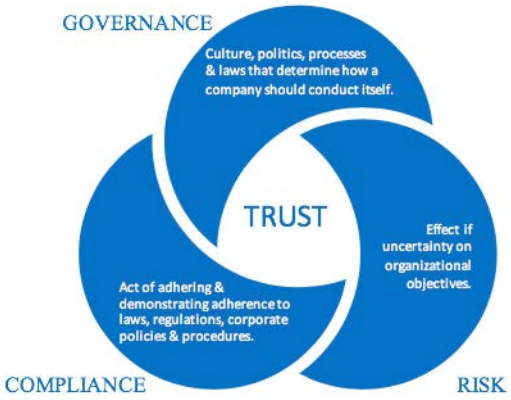


Figure 9-1 Governance, compliance, and risk framework

GRC aims to synchronize information and activity across governance, risk management, and compliance in order for an organization to operate more efficiently, enable effective information sharing, report activities more effectively, and avoid wasteful overlap.

The goals, then, of any GRC program must include

- keeping risk at acceptable levels,
- maintaining availability to systems and services,
- complying with relevant laws and regulation, and
- protecting customer data.

Many organizations will establish a "cloud governance" team to adapt existing processes, create new ones where needed, and ensure their usage. For the cloud, governance implies, among other things, when, how, and how often applications are deployed to the cloud; how teams develop cloud applications and which applications; how updates, patches, and revisions are built, tested, and deployed; that application data is properly secured; and so on.

Regulatory compliance

Managing regulatory compliance can be a complex task under any circumstance, and for multinational organizations and regulated industries such as healthcare and financial services, it can be even more challenging. Standards and regulations abound, of course, and they change frequently, making it difficult for businesses to keep abreast of all the international electronic data–handling laws.

Businesses should understand the division of responsibilities regarding regulatory compliance in the cloud. Cloud providers like Microsoft make every effort to ensure that their platforms and services are compliant, but companies should also ensure that their applications, or those from third parties which they use, are compliant.

Similarly, applications in regulated industries that use cloud services may require certification from the cloud provider. For example, a healthcare application that processes protected health information (PHI) is subject to the Health Insurance Portability and Accountability Act (HIPAA)'s Privacy Rule and the HIPAA Security Rule, and thus requires that a healthcare business receive written assurances from the cloud provider that it will safeguard any PHI received or created.

Another important regulation is the Payment Card Industry Data Security Standard (PCI DSS), a proprietary information security standard for organizations that handle branded credit cards from the major card schemes including Visa, MasterCard, American Express, Discover, and JCB. The PCI Standard is mandated by the card brands and administered by the Payment Card Industry Security Standards Council. The standard was created to increase controls around cardholder data to reduce credit card fraud. Validation of compliance is performed annually, either by an external qualified security assessor or by a firm-specific internal security assessor that creates a report on compliance for organizations handling large volumes of transactions, or by a self-assessment questionnaire for companies.

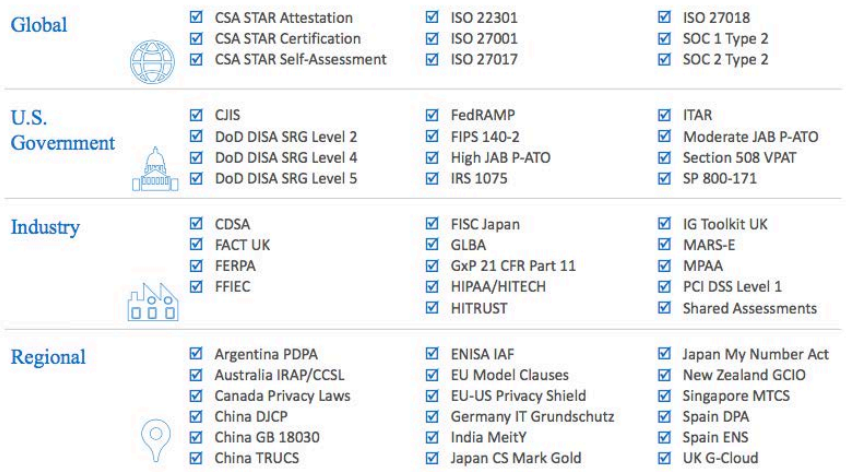


Figure 9-2 Microsoft Azure compliance attestations

Microsoft has received at this writing more than 50 compliance attestations, shown in the chart above. Check the Microsoft Trust Center to see updates as well as important prescriptive guidance regarding compliance.

The General Data Protection Regulation (GDPR)

On May 25, 2018, a sweeping set of regulations went into effect in the European Union, valid in 31 countries of the EU and some non-EU countries (Iceland, Liechtenstein, and Norway). The GDPR gives individuals greater control of their personal data and imposes new obligations on organizations that collect, handle, or analyze personal data. The GDPR also mandates strict monetary penalties on organizations found not to be compliant.

Under the GDPR, individuals have the right to

- access their personal data,
- correct errors in their personal data,
- erase their personal data,
- object to processing of their personal data, and
- export personal data.

These rights have a set of corresponding requirements for organizations that collect personal data, which include

- protecting personal data with appropriate security,
- notifying authorities of possible breaches,
- obtaining consents for processing data, and
- keeping records describing the processing of personal data.

Microsoft has made many resources available to organizations to help in understanding the GDPR, including this site containing best practices.

Cloud security

Companies worldwide are challenged by the ongoing volume of evolving security threats and with retaining qualified security talent to respond to these threats. In fact, the average large organization gets 17,000 security alerts each week, which results in an average of 99 days to discover security breaches. That contrasts with the less than 48 hours it takes for security breaches to grow from one compromised system into significantly broader issues.

As you look for solutions to address these challenges, Azure can help strengthen your security posture, while reducing cost and complexity. Azure provides value in three key areas—a secure foundation that is provided by Microsoft, built-in security controls to help you quickly configure security across the full stack, and unique intelligence at cloud scale to help you safeguard data and respond to threats in real time.

Secure foundation of Azure

In this section we cover how to strengthen the security of your cloud workloads.

Physical security

Every security story starts with physical security, that is, the safety of the physical facilities in which the cloud runs—the cloud datacenters. Microsoft takes a layered approach to physical security. Datacenters managed by Microsoft have extensive layers of protection: access approval, at the facility's perimeter, at the building's perimeter, inside the building, and on the datacenter floor. Employees at cloud datacenters must undergo rigorous background checks. Admission to the server areas requires multiple forms of authentication, including a biometric check. All activity is monitored and audited.

Infrastructure security

The secure network infrastructure of Azure has built-in protections against distributed denial of service to safeguard your resources against volumetric or protocol layer attacks. Azure DDoS Protection has the operational capacity to scale protection to the largest of workloads and the experience protecting Microsoft services such as Xbox and Office 365.

Security controls are integrated into the firmware and hardware of Azure to ensure data is secure by default and continues to be secure throughout its lifetime.

Of course, security is ever-evolving. At its datacenters, Microsoft manages the basics such as guaranteeing the servers that run Azure are patched. It actively works to identify vulnerabilities through continuous testing and monitoring by running exercises such as red team versus blue team cyber-penetration testing.

Operational security

Microsoft's developers follow the Security Development Lifecycle to ensure they are meeting core security principles throughout development, resolving security issues before their code is deployed, and adhering to the security standards used by all software developed for the Azure platform.

In addition to securing your code, Azure operations and security professionals also work to protect your data from unauthorized access. This includes implementing controls that restrict unauthorized access from Microsoft personnel and contractors and requiring Azure infrastructure and security operators are required to use a secure access workstation when accessing the Azure infrastructure.

Microsoft Azure has a global security incident management team that detects and responds to a wide array of security threats 24/7/365. Microsoft has more than 3,500 cybersecurity experts that act as human intelligence while working with sophisticated, automated processes to detect, respond, and remediate threats.

Built-in security controls

Even with the secure foundation that Azure provides, security is ultimately a joint responsibility between you and Microsoft. When you put your workloads and data on Azure, we recommend you use built-in security controls across identity, network, data, and tools to help you with security management and threat protection.

Manage identity and access

Consider multifactor authentication for secure user sign-ins. MFA requires that a second form of identity beyond username and password be supplied to gain access to corporate resources. Various forms of MFA are available including biometric models, phone calls, and text messages. For example, a sign-in might trigger a phone call to a mobile phone that has thumbprint identification capability; the user is prevented from signing in until the phone returns a valid entry.

Another form of MFA changes a random number on a mobile device every few seconds according to a predetermined algorithm; the user must type in the number displayed by the phone in order to gain access.

Secure your network and data and manage keys and secrets

Applications should use encryption wherever possible. For hybrid cloud connections, that is, connections between the on-premises datacenter and the cloud, VPNs and ExpressRoute use IPSec with Internet Key Exchange as the underlying transport. Consider using Transport Layer Security, which is the security technology behind secure HTTP (HTTPS) for client access to cloud websites.

Data at rest in Azure storage or in databases should also be encrypted wherever possible. Azure SQL Database, for example, offers transparent data encryption for real-time data encryption and decryption, using a server certificate. Replicas in different geographic regions have different certificates, which are rotated every 90 days (considered a norm).

A best practice in security is to separate encryption keys from the application, and with a vault, such as Azure Key Vault, this is possible. With this capability, an administrator first creates a key vault for the application and then places the key(s) into it. Azure Key Vault then supplies the developer with URLs to the keys, which the application can use at runtime to decrypt arbitrary data, such as in Azure Storage or elsewhere.

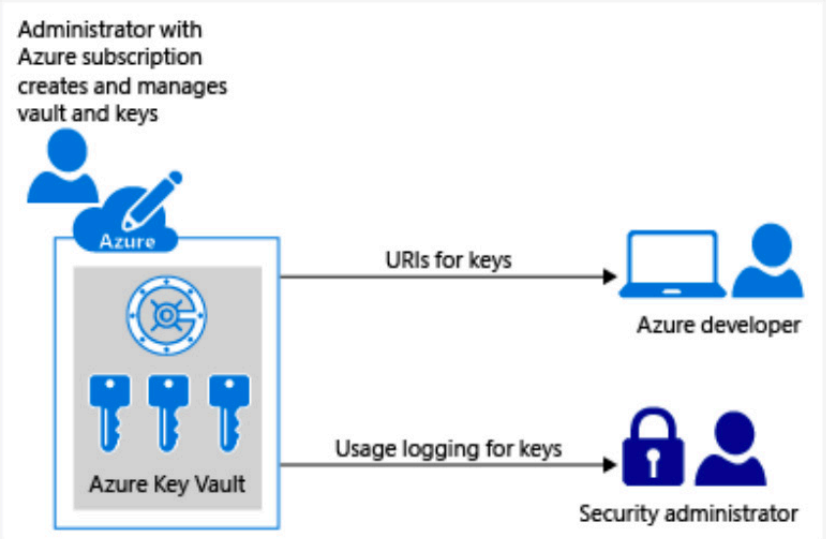


Figure 9-3 Azure Key Vault

For added protection, the keys or secrets can be stored in a hardware security module (HSM), a physical appliance that can both store as well as generate keys. HSMs can also offload cryptographic processing (normally a CPU-intensive activity), performing encryption and decryption on board.

For customers in more highly regulated industries, you can also use Azure Dedicated HSM to maintain full administrative and cryptographic control over your HSMs. Each HSM device comes validated against FIPS 140-2 Level 3 and eIDAS Common Criteria EAL4+, ensuring tamper resistance. This enables you to meet a wide variety of security and compliance requirements.

Threat protection and security management

Azure Security Center provides security professionals in your organization with a wide array of capabilities to help them strengthen their security posture and protect against threats. These capabilities include providing recommendations (such as applying patches or updating antivirus software) to improve your security score, applying security policies to meet compliance requirements, protecting IaaS resources and Linux and Windows servers from threats, and responding to security alerts (such as an RDP brute-force or SQL injection attack).

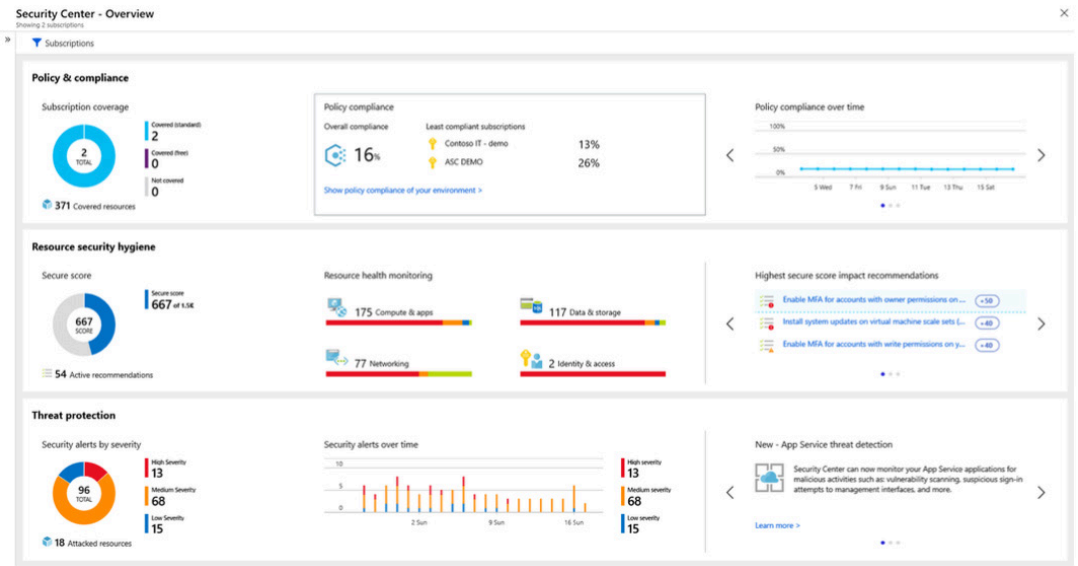
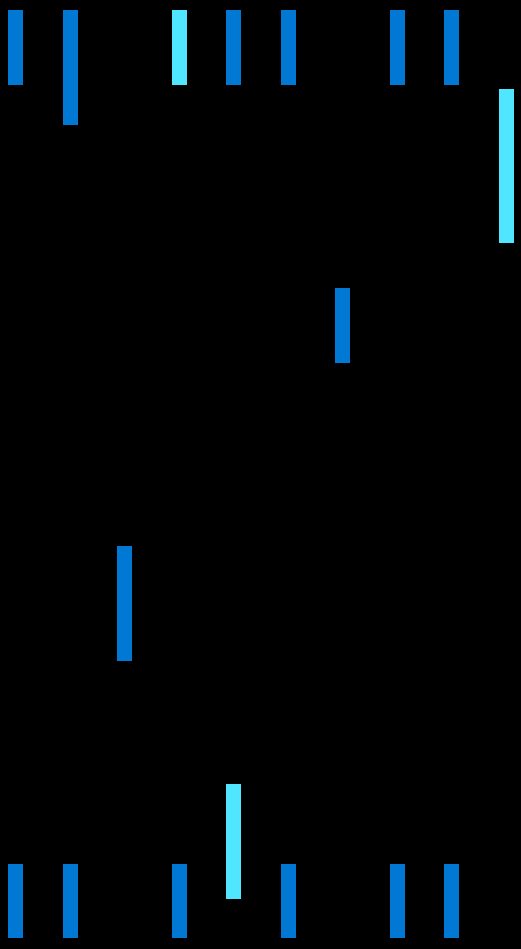


Figure 9-4 Azure Security Center

Summary and Conclusions



In this book, we've detailed Microsoft Azure infrastructure as a service (IaaS) and put IaaS in context with the other predominant cloud computing models, PaaS and SaaS. IaaS has become a popular model for cloud computing because of the ease with which applications running in a private datacenter can be migrated to the cloud.

In Chapter 2, we described some key terms to know when talking about the cloud: IaaS, PaaS, and SaaS, and the pros and cons of private, government, hybrid, and public clouds.

In Chapter 3, we outlined some common use cases for infrastructure as a service, focusing on the DevTest scenario as one of the first you should consider.

In Chapter 4, we discussed the charging model behind IaaS in Microsoft Azure. We also showed how you can optimize your deployments so that you achieve the maximum efficiency and cost savings.

Then in Chapter 5, we took a deep dive into the services in the Microsoft IaaS offering, including compute, storage, and networking. We also provided a description of the mechanics of assessing your applications' readiness to migrate and how to perform a migration.

In Chapter 6, we talked about the hybrid cloud: how you can continue to use your on-premises applications while also taking advantage of cloud services to improve your operations and make them more cost-effective.

Chapter 7 discusses one of the more recent advances in cloud application architecture: containers, and how you can take advantage of capabilities in Azure to accelerate your container deployment.

In Chapter 8, we provided an overview of the various management and monitoring tools available in Azure, and in Chapter 9, we discussed some of the critical issues that IT managers face when moving data to the cloud: governance, risk management, compliance, and security.

We hope you enjoyed your tour of Azure IaaS, and feel well prepared to start your IaaS journey. Happy clouding!

For Further Reading

General

Briggs, Barry, James Farhat, and Eduardo Kassner. *Designed to Disrupt: Reimagine Your Apps and Transform Your Industry*. (Microsoft Press, 2018). https://azure.microsoft.com/resources/designed-to-disrupt-reimagine-your-apps-and-transform-your-industry/.

Azure Friday (YouTube series). https://www.youtube.com/playlist?list=PLLasX02E8BPDT2Z2pdCHNCkENpcQWy5n6.

Sources

Chapter 1: Why the Cloud?

Datacenter costs
https://www.datacenterknowledge.com/archives/2015/02/11/data-center-building-vs-outsourcing-whats-best-business

Microsoft Azure regions
https://azure.microsoft.com/global-infrastructure/regions/

Microsoft Azure SLAs
https://azure.microsoft.com/support/legal/sla/summary/

Chapter 2: Terms to Know

What is IaaS?
https://azure.microsoft.com/overview/what-is-iaas/

What is PaaS?
https://azure.microsoft.com/overview/what-is-paas/

What is SaaS?
https://azure.microsoft.com/overview/what-is-saas/

Chapter 3: Common Scenarios for IaaS

Common scenarios
https://azure.microsoft.com/overview/what-is-iaas/

SAP HANA on Azure IaaS
https://azure.microsoft.com/blog/why-you-should-bet-on-azure-for-your-infrastructure-needs-today-and-in-the-future/

GPU-centric applications on Azure IaaS
https://azure.microsoft.com/blog/massive-scale-cloud-rendering-with-autodesk-on-azure/

Chapter 4: The Economics of IaaS

Forrester Total Economic Impact report on Azure IaaS
https://azure.microsoft.com/mediahandler/files/resourcefiles/b97e0ab9-3aff-49b4-af44-eaf1bef33085/Azure%20IaaS%20Total%20Economic%20Impact%20Report%20(TEI)%202017%20by%20Forrester.pdf

Chapter 5: IaaS In-Depth

Windows VM sizes
https://docs.microsoft.com/azure/virtual-machines/windows/sizes

Linux VM sizes
https://docs.microsoft.com/azure/virtual-machines/linux/sizes?toc=%2fazure%2fvirtual-machines%2flinux%2ftoc.json

Availability sets
https://docs.microsoft.com/azure/virtual-machines/windows/manage-availability

Availability zones
https://docs.microsoft.com/azure/virtual-machines/windows/manage-availability

Availability zone—how-to
https://docs.microsoft.com/azure/virtual-machine-scale-sets/virtual-machine-scale-sets-use-availability-zones

Traffic manager
https://docs.microsoft.com/azure/traffic-manager/

Networking for IaaS
https://docs.microsoft.com/office365/enterprise/designing-networking-for-microsoft-azure-iaas

Azure Virtual Network
https://docs.microsoft.com/azure/virtual-network/virtual-networks-overview

Azure Storage
https://azure.microsoft.com/services/storage/

Chapter 6: To the Cloud and Back Again

Azure Backup
https://azure.microsoft.com/services/backup/

StorSimple
https://azure.microsoft.com/services/storsimple/

Azure Site Recovery
https://azure.microsoft.com/services/site-recovery/

Chapter 7: Containers and Orchestration

Containers on Azure Overview
https://azure.microsoft.com/overview/containers/

Azure Kubernetes Service (AKS)
https://azure.microsoft.com/services/kubernetes-service/

Chapter 8: Create, Manage, and Maintain Your IaaS Ecosystem

Azure Portal
https://portal.azure.com

Monitoring overview
https://docs.microsoft.com/azure/monitoring/monitoring-overview

Chapter 9: Governance, Risk, Compliance, and Security

The Three Ways Azure Improves Your Security
https://azure.microsoft.com/blog/the-3-ways-azure-improves-your-security/

Azure Trust Center
https://azure.microsoft.com/overview/trusted-cloud/

Azure Compliance Overview
https://www.microsoft.com/trustcenter/compliance/default.aspx

Privacy
https://www.microsoft.com/TrustCenter/Privacy/default.aspx

GDPR Webcast
https://info.microsoft.com/ww-ondemand-getting-ready-for-GDPR.html


Microsoft Azure