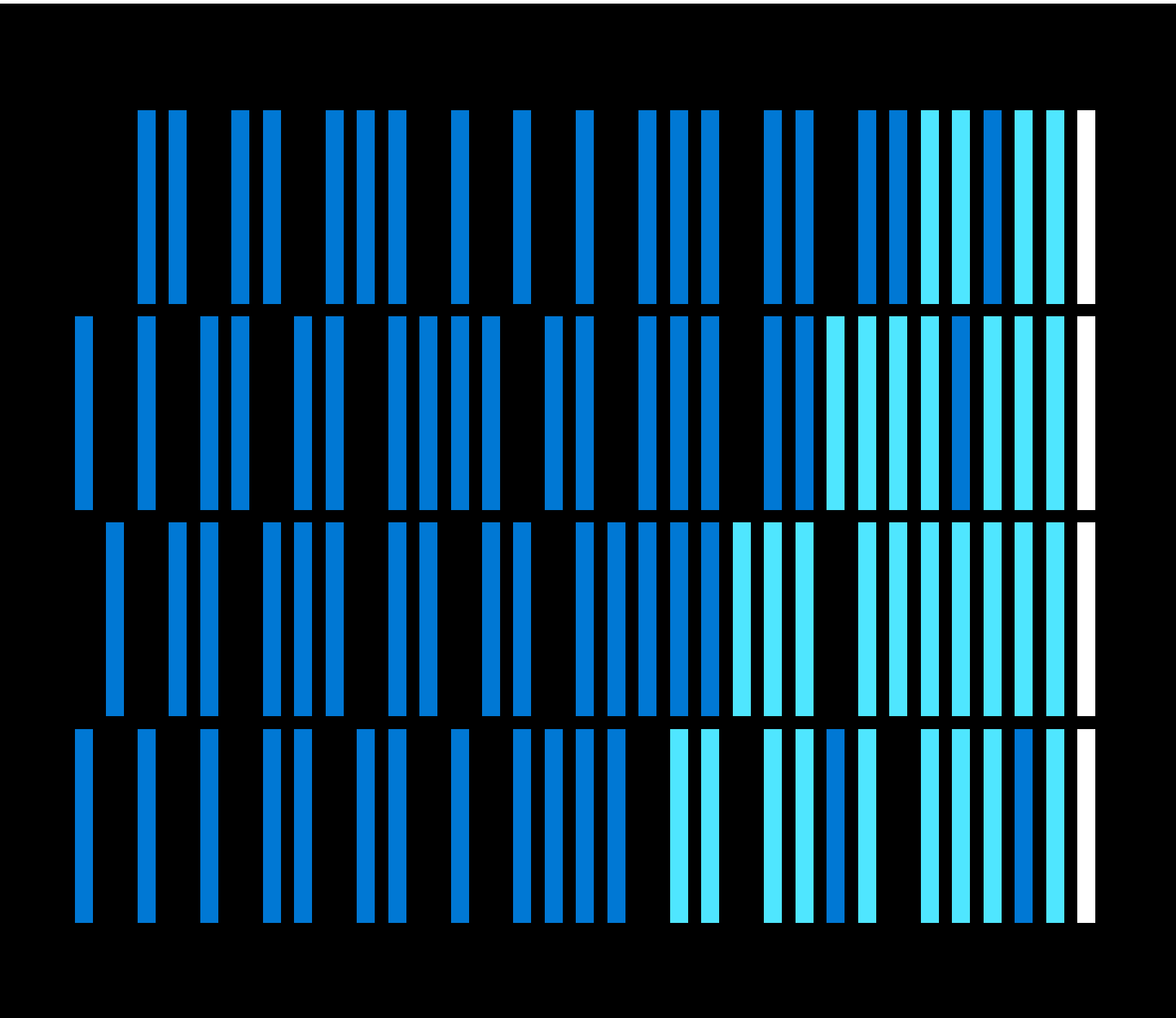
Azure e-book series

Microsoft Azure

Enterprise Cloud Strategy
2nd Edition



PUBLISHED BY
Microsoft Press
A division of Microsoft Corporation
One Microsoft Way
Redmond, Washington 98052-6399

Microsoft Press books are available through booksellers and distributors worldwide. If you need support related to this book, email Microsoft Press Support at mspinput@microsoft.com. Please tell us what you think of this book at http://aka.ms/tellpress.

This book is provided "as-is" and expresses the author's views and opinions. The views, opinions and information expressed in this book, including URL and other Internet website references, may change without notice.

Some examples depicted herein are provided for illustration only and are fictitious. No real association or connection is intended or should be inferred.

Acquisitions Editor: Janine Patrick
Developmental Editor: Bob Russell, Octal Publishing, Inc.
Editorial Production: Dianne Russell, Octal Publishing, Inc.
Copyeditor: Bob Russell

Defining correct enterprise cloud strategy is a complex task that is easy to understand but hard to master. That is why the *Enterprise Cloud Strategy* book represents one of the best publications for cloud experts who want to learn how to shape the efficient enterprise cloud transformation. This book provides a clear yet deep level of detail on all aspects of cloud strategy, offering a large collection of case studies and repeatable patterns, and even providing cloud architectural blueprints to accelerate specific cloud use cases. It is a perfect companion for all advisors, strategists and enterprise planners that are shaping complex cloud journey roadmaps – and is mandatory reading for all our technology strategy professionals. It is important that the book covers non-technological aspects of cloud strategy, but it is even more important that it guides cloud strategy implementations. And above all, it is very pleasant to read and won't deter any cloud novices.

– Miha Kralj, managing director of cloud strategy, Accenture

The move to the cloud has opened many opportunities. With it comes the need for best practices and guidance on how to adopt cloud platforms with enterprise grade rigour and governance. This book fills this troublesome gap in a clear, concise and practical way. It is an easy read, too. It is based on the best practices of Microsoft IT and many global companies, therefore I recommend that you review the approaches outlined. I wish you the best of luck on your journey.

– Gavriella Schuster, corporate vice president, One Commercial Partner, Microsoft Corporation, July 2017

The *Enterprise Cloud Strategy* book provides a strategic and, more importantly, end-to-end perspective on the multitudes of concerns, challenges and, most crucially, opportunities for enterprises in the age of cloud computing. From strategic reasoning, to practical guidance in assessing needs and exploring proven practices, this book shows the enterprise architect or executive what cloud computing is all about and why any modern enterprise needs to care now.

– Ulrich Homann, distinguished architect, Cloud + Enterprise Engineering, Microsoft

Contents

Introduction to the second edition

In the two years since we first began working on *Enterprise Cloud Strategy,* much has changed. Cloud technology has evolved from being an "if" to a "when," and across almost all enterprises, the cloud has now become an integral part of IT strategy. Further, there is now a growing realisation that cloud computing not only represents a set of technical opportunities for efficiencies and cost savings, but also provides the potential to significantly transform the scope of enterprise computing. In fact, many enterprises are finding that cloud computing offers entirely new business models, revenue streams and vehicles for customer intimacy.

Few technological changes have the potential to so dramatically change the way we do business. The last time we experienced such a tectonic shift was the with arrival of the Internet itself.

Yet, as we noted in the first edition of this book, when briefing CIOs and senior IT executives at Microsoft, we are often told that migrating IT workloads to the cloud ranks among their highest priorities. And quite frequently this is followed by questions such as “How do I start?”, “How should I build a plan for cloud migration for my entire portfolio?” and “How will my organisation be affected by this change?”

Today, a new question has been added: "How can I use cloud computing to become a true partner to the business?"

This book, based on real-world cloud experiences by enterprise IT teams, seeks to provide answers to these questions. Here, you’ll see what makes the cloud so compelling to enterprises; which applications you should consider as you start your cloud journey; how your organisation will change and how skill sets will evolve; how to measure progress; how to think about security, compliance and business buy-in; and how to exploit the ever-growing number of features the cloud offers in order to gain strategic and competitive advantage.

Acknowledgments

For the second edition, we would like to thank Miha Kralj, managing director of cloud strategy at Accenture, and Brian Cawelti of Avanade for their insights. In addition, many thanks to Brad Wright and Rick Ochs of Microsoft IT. Frank Simorjay, Ranger Due, Rodrigo Souza, David Cervigon Luna, Pete Apple, Andre de Beer, Brian Harrison and William Bories of Microsoft all provided highly useful input as well.

We further wish to express our deep gratitude to the following individuals for their support, guidance and willingness to freely share their expertise: Scott Woodgate, Javier Nino, Tom Schinder, Venkat Gattamneni, Martin Vliem, Ulrich Homann, Robert Hanegraaff, John Devadoss, Brenda Carter, Michael Washam, Zoiner Tejeda, Nadia Matthews, Rob Beddard, Jeff Fryling, Kevin Gee, Colin Nurse, Raman Johar, Walter Myers, Uwe Hoffman, Ashish Sharma, Ashutosh Maheshware, Rich Nickerson, Michel Declercq, Arlindo Alves, Dennis Mulder, George Moore, Richard Ochs and Christopher Bennage.

Rob Boucher and Monica Rush created the graphic representations of the blueprints in Appendix A.

Free eBooks from Microsoft Press

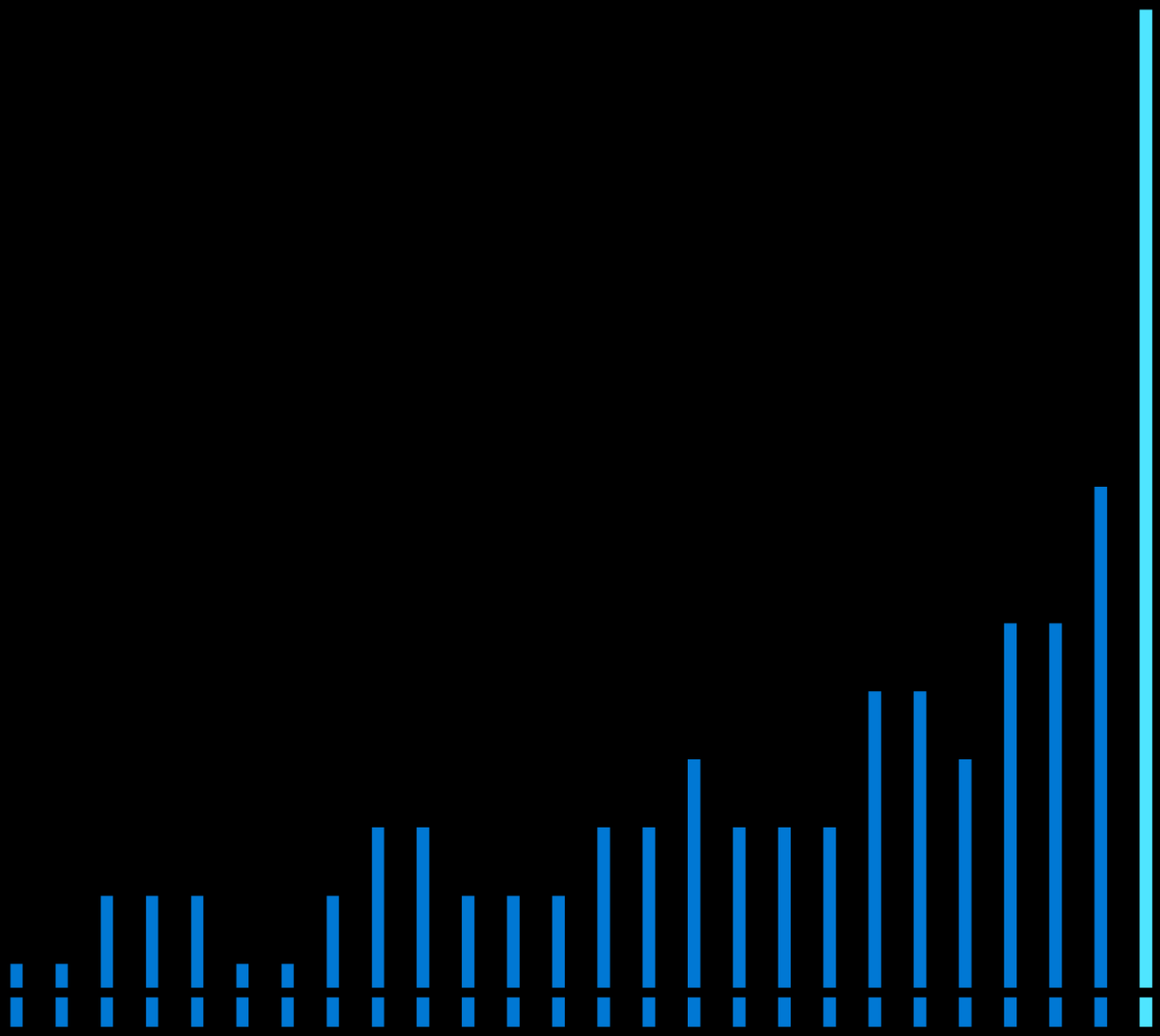
From technical overviews to in-depth information on special topics, the free eBooks from Microsoft Press cover a wide range of topics. These eBooks are available in PDF, EPUB and Mobi for Kindle formats, ready for you to download at:

http://aka.ms/mspressfree

Or check out the full collection at https://www.microsoftpressstore.com/

Part I

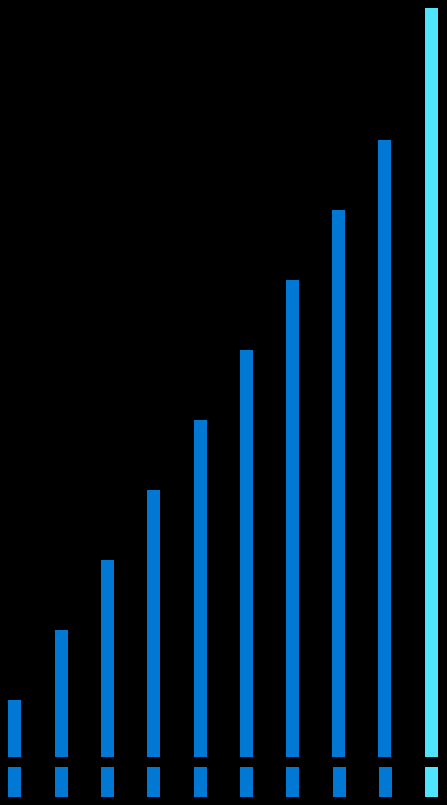
Beginnings



Microsoft

Chapter 1

The cloud, efficiency and innovation



Most people now agree that the cloud has become a core element of any enterprise's technology strategy. Indeed, in the past few years we have seen the conversation around cloud adoption move from "if" to "when" and "how."

Nevertheless, it remains one of the most disruptive changes in computing in years, and it is worth reviewing what makes the cloud so compelling to enterprise IT. Its value proposition is many-faceted, ranging from significant cost savings over a traditional datacentre approach, to the ability to quickly build robust, resilient applications that can scale-up as traffic spikes, and scale-down as it recedes.

Enterprise computing before the cloud

For nearly a half century, the economics of enterprise computing remained relatively constant. Enterprises purchased computing equipment and software from vendors and housed them in their own datacentres. Computers were like any other capital expense: a (usually large) one-time purchase followed by several years of depreciation.

As enterprises grew, so did the number of datacentres, for various reasons. Often as new facilities or plants were constructed, a new computing centre would be built nearby. As they grew into other countries, a datacentre in that location would be required for both technical reasons (to reduce networking costs) and perhaps also to comply with local regulations. And, finally, as computing became mission-critical for the operations of the business, new datacentres were built solely to support business continuity and disaster recovery requirements.

For the CIO, all of this expansion meant an IT organisation that perhaps spanned the globe, but also one which required large numbers of skilled individuals to maintain all of the systems. It was not uncommon that a third of the IT staff was dedicated to "operations" – that is, maintaining the datacentres; procuring new hardware; deploying new servers, software, and retiring depreciated hardware; managing networks; ensuring that system software patches were applied in a timely fashion; debugging router loops; and other such arcane issues.

Moreover, most CIOs intuitively understood that, then as now, demand on enterprise applications is, by and large, seasonal. Enterprise Resource Planning (ERP) systems that manage the corporate ledger are under the heaviest usage toward the end of the quarter and the end of the fiscal year. Performance management systems for employee reviews are most heavily used during the review period but are practically idle for the rest of the year. Many IT managers had "rules of thumb" to purchase three or four times the amount of hardware expected for the load – to ensure that applications never failed during peak usage.

Of course, the consequence was that average CPU usage in the datacentre was, surprisingly, sometimes in single digits. Virtualisation – putting multiple workloads on a single server – went some distance in improving usage, but overall it remained low, which suggested that money was being wasted on IT assets that still were not being fully used.

Between operations staff, capital equipment management and software maintenance, an IT department could easily spend 80 % or more of its budget, with only a small amount left over for innovation. No wonder, then, that CEOs and CFOs constantly searched for ways to trim the IT budget, given that any money disbursed to IT was typically money lost to growing the business.

Something had to change.

Economics of the cloud

Shortly after the turn of the century, several technology vendors began offering computing services, in effect for rent – the birth of the cloud. It soon became evident that this model yielded important advantages for enterprise customers.

In cloud computing, enterprises pay for what they use, much as they would pay a telecom provider. If demand decreases and you no longer need capacity, you can turn off systems and you are not charged. This simple model stands in stark contrast to the *traditional, capital-intensive* model of enterprise computing just described.

The cloud, being subscription-based, is an *operating expense* model. In the cloud, computing becomes a service for which customers are billed a monthly charge. Like other such services, it is metered by usage. The more computing, network and storage resources you use, the higher your bill. Of course, the reverse is also true: the less you use, the less you are charged. Indeed, most IT organisations find wide variations in system usage: some applications (for example, retail shopping) are seasonal; other applications (for example, training applications) run for a short period of time before being shut down; others are simply unpredictable. The cloud addresses this variability (shown in Figure 1-1) perfectly via its “pay for what you use” model.

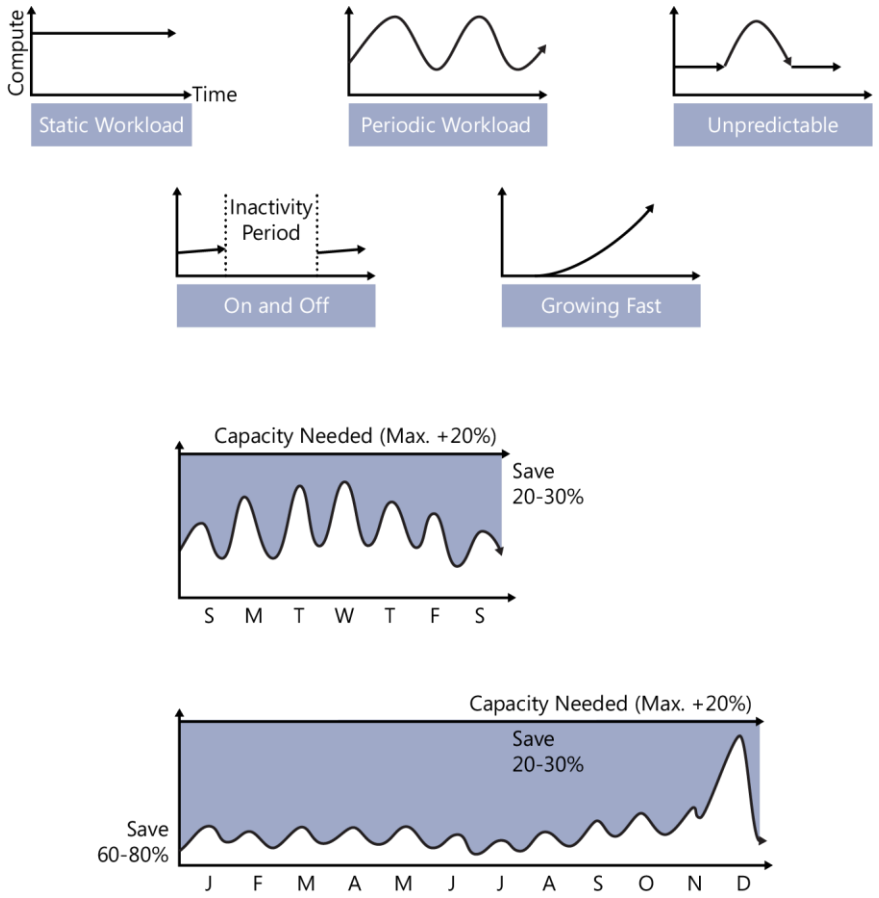


Figure 1-1: Common application utilisation models

(It is worth mentioning that in the on-premises datacentre, the *maximum* usage must be planned for and provisioned, which is financially far more inefficient than in the cloud.)

But there is more to it. Operating in the cloud frees enterprises of the mundane tasks of system backup, network maintenance, patches and software upgrades, because the cloud provider can handle these in their entirety. The cloud provider in turn is heavily incentivised to use and, in many cases, pioneer best practices for system maintenance; the benefits are then passed on to the customer.

Moreover, cloud providers such as Microsoft can achieve economies of scale by buying hardware in massive bulk, tens of thousands of servers at a time, for example. Very large datacentres hosting public clouds can also achieve economies in purchasing other resources; cloud datacentres pay only a quarter of the average cost of electricity in the United States. In many cases, cloud datacentres take advantage of local renewable energy; for example, Microsoft's datacentre in Quincy, Washington, is located near a hydroelectric facility and other datacentres use wind-generated electricity as well as other green sources.

Figure 1-2 shows how overall total cost of ownership (TCO) per server declines dramatically at scale.

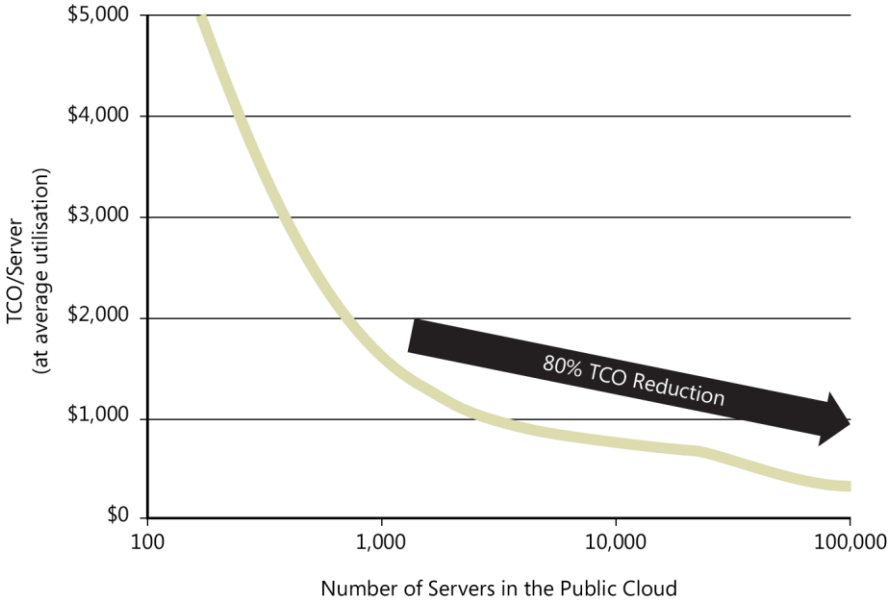


Figure 1-2: Economies of scale in the cloud

These savings can, and are, passed on to cloud service customers.

Later, we will discuss how IT departments can quantify the savings they can expect to achieve by adopting cloud computing.

Perhaps most importantly, the cloud is not an "either/or" proposition. It is certainly possible, and indeed in many cases desirable, to leave some applications running in a local, traditional datacentre while others are moved to the cloud. Providers such as Microsoft have made huge investments in this *hybrid cloud* model that securely connects applications in the cloud to those remaining in a customer's datacentre. As we shall see, the hybrid model makes it possible for companies to move their applications to the cloud *at their own pace*.

After there is an on-demand computing service available, all sorts of other efficiencies become possible. For example, systems devoted to development and application testing often constitute a large cost area for IT departments, yet, when all is said and done, they do not actually provide any direct value to end users. With the cloud, developers and testers can quickly allocate cloud-based resources, use them for their work and then free them up when they've finished. Similarly, with the vast, capacious amounts of cheap storage available in the cloud, data backup to the cloud, and across multiple geographies if desired, becomes a straightforward and inexpensive function. We will cover more of these in the course of the book.

After TCO: the journey continues

Many companies we talk to have agreed that migration to the cloud will help them save money and operate more efficiently (and we agree with them). In fact, we talk later (in Chapter 7) about how, after companies move to the cloud, they can optimise their use of the cloud on a day-to-day basis, adjusting consumption and usage to achieve their cost goals.

But that is only half the story. As many companies are discovering, the drive for lower costs is really only the first step in a journey.

The cloud opens up all sorts of possibilities for innovation, which does not only make IT better, but also provides direct benefit to the business, making the CIO not just a cost centre, but a real partner in driving value and growth for the business.

In 2016, Microsoft commissioned Forrester Consulting to conduct an independent study[1] on the return on investment (ROI) of using the cloud. In this case, the study focused on platform-as-a-service (PaaS) usage (more on this in Chapter 2), but the results were striking: an ROI of 466 %, with a reduction in the amount of IT time spent on maintenance of 80%, among other benefits:

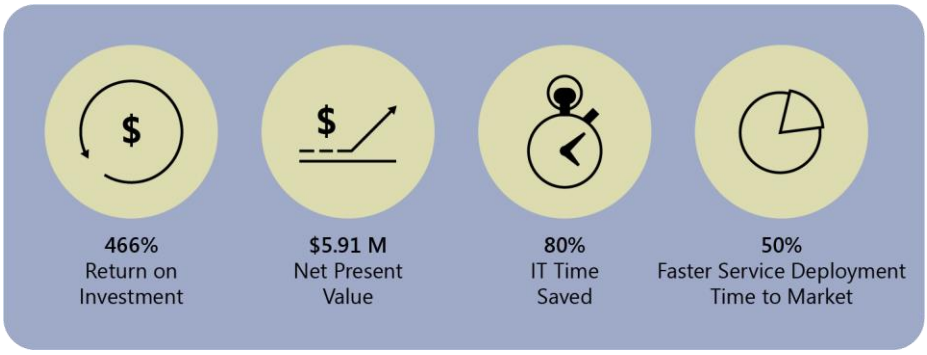


Figure 1-3: Cloud platform-as-a-service benefits

Of course, your mileage may vary, and we have considerably more to say about measuring cloud costs and cost savings in Part II of this book. However, what is salient at this point is that because of this substantial reduction in non-value-added tasks, such as maintaining servers and managing patches, enterprises were more able to focus their IT resources on *business innovation*.

Innovation

At the end of the day, the goal of any enterprise strategy is to create competitive differentiation and advantage, and little doubt remains that IT has become a key element in modern strategy. IT now drives transformative innovation, making it possible for enterprises to compete more effectively by establishing processes that deliver ongoing competitive advantage.

As we will see, the emergence of a global computing cloud heralds the arrival of entirely new classes of innovation across applications and markets. Indeed, such new forms of innovation can actually *transform* an organisation and a business.

Transformational innovation drives a different culture and mind-set than most organisations currently have. Affecting both IT and the leadership of the enterprise as a whole, this culture requires a close alignment between IT and business leadership.

In the next few pages, we will examine a number of case studies from various global companies, all of which have reaped rewards by their use of the cloud.

Accuweather

AccuWeather, a leading provider of weather forecasts worldwide, needed a better solution for handling more than four billion daily data requests. Accuweather uses the cloud for development and for proofs of concept – a straightforward task given that, by using the cloud, it does not need to procure and provision hardware.

[1] "The Total Economic Impact of Microsoft Azure PaaS," July 2016

It also has increased on-demand scalability, improved access to real-time weather data and cut IT costs by up to 40 %.

Scale was particularly important: "As more connected devices came on the market worldwide, we went from two million to more than four billion requests a day within five years," says Chris Patti, Vice President of Technology at AccuWeather. "Scale became a challenge" and within a few short years, that quadrupled to 17 billion requests every single day.

Weather, of course, is all about data. The company is using analytics and artificial intelligence capabilities in the cloud (Microsoft Cortana Intelligence Suite) to integrate sales data with weather information. In a recent project with Starbucks, AccuWeather helped the coffee giant solve seasonal problems like running out of ice and cups in hot weather. And, in another example, AccuWeather helped a global candy manufacturer identify which products sold best, and if the sales spikes were weather-related.

In short, by taking advantage of the cloud, Accuweather discovered what many enterprises have discovered, or soon will: the cloud can save you money and open up new markets.

GEICO

GEICO, a direct auto insurer since 1936 and now the second-largest private-passenger auto insurer in the United States, is enhancing its digital presence to better connect with customers through multiple digital venues. Referring to the rise of the mobile Internet and the explosion in social media participation, Fikri Larguet, Director of Cloud Services at GEICO, notes:

> *In the last five to eight years, the customer appetite for digital engagement has grown enormously. Customers are engaging with us much more frequently and in new and interesting ways. We want to be ahead of the curve when it comes to where the next digital engagement opportunities will occur.*

But what does digital engagement mean? It means 24/7 availability, on every kind of device the customer might have, from anywhere.

Like Accuweather, Geico discovered that by moving to a cloud model, it could easily reach all of its customers, at any time and at any scale. Moreover, it found many of its IT costs declined: development teams accelerated as a result of their adoption of a DevOps (Chapter 7) model in conjunction with cloud development. Because of the cloud's ability to run multiple copies of applications in different datacentres, redundancy and business continuity/disaster recovery (BC/DR) operations are greatly simplified – again, these are benefits that any enterprise can reap.

Rolls-Royce

Rolls-Royce has more than 13,000 engines for commercial aircraft in service around the world, and for the past 20 years, it has offered customers comprehensive engine maintenance services that help keep aircraft available and efficient. As the rapidly increasing volume of data coming from many different types of aircraft equipment overtakes the airlines' ability to analyse and gain insight from it, Rolls-Royce is using the Microsoft Azure platform to fundamentally transform how it uses data to better serve its customers.

Rolls-Royce uses the scalable, on-demand nature of analytics (Chapter 12) in Azure, along with its artificial intelligence (AI) capabilities (Chapter 13), to perform data modelling and analytics at scale to accurately detect operational anomalies and help customers plan relevant responses. Says Nick Farrant, Senior Vice President of Rolls-Royce:

There are terabytes of data coming from large aircraft fleets, with gigabytes per hour – rather than kilobytes – to process and analyse. Microsoft Cortana Intelligence capabilities are helping us filter the signal from the noise across large datasets so that we can focus on finding the real value in the data. Our vision of future digital capability will need to aggregate many sources of data and provide a platform for collaboration with customers.

We believe, because of the remarkable technologies that exist today, which make it possible for enterprises to capture huge amounts of data about what their customers, their partners and their machines are doing, that every enterprise will become a data-driven one. CIOs and IT decision makers should include data, analytics and AI in their cloud plans because of the benefits that will accrue to their businesses.

Brainshark

Brainshark is a cloud-based sales training and readiness platform that helps sales people achieve mastery in the presentation of sales materials to clients, slashing the costs and resources needed for training and maximising the effectiveness of sales engagements.

With half of the Fortune 100 as its clients, Brainshark is a clear worldwide leader in its space. And continuous innovation and improvement have kept it a business leader for 17 years, and poised the company for continued dominance.

Brainshark began its use of the cloud by placing all of its video training materials there. According to Brainshark's Vice President of Engineering, Michael Ferioli:

By moving video to Azure we've virtually eliminated the management and cost of maintenance we used to incur. We actually spend less with Microsoft than we thought we would on an ongoing basis. Actually, I have not bought a piece of hardware in more than two years.

And what did the company do with its savings? It began innovating new ways of immersive sales training. For example, by using Microsoft's advanced augmented reality HoloLens device, Brainshark created much more realistic training scenarios. Sales trainees can experience a simulated client engagement through Microsoft HoloLens, complete with presentation capabilities and life-like avatars representing clients. In contrast to virtual reality technologies, HoloLens combines real spaces with virtual elements, letting trainees practise in places they're familiar with.

By taking costs out of non-value-added functions associated with an on-premises datacentre, Brainshark was able to truly innovate and differentiate in remarkable new ways.

Disaster Relief: Oso, Washington, 2014, and Nepal, 2015

Because the cloud gives IT the ability to create applications and make them operational very quickly, disaster recovery teams around the world rely on it to rapidly bring aid to people in need.

On 22nd March, 2014, a hillside saturated by heavy rains collapsed on the small Northwest town of Oso, Washington, flattening homes and killing 43 people. In the aftermath, nearly 200 government and aid agencies, including the Red Cross, the Federal Emergency Management Agency, the Washington National Guard and the US Navy's search and rescue team, as well as thousands of representatives of the media, descended upon Oso.

The local government's record-keeping and co-ordination systems were quickly overwhelmed, so Microsoft Services Disaster Response, with help from the Azure product team, migrated Oso's records

to the cloud. With its nearly limitless capacity, the cloud made it possible for everyone who needed access to the records to retrieve and search them quickly and efficiently. Using Microsoft Office 365, the team also quickly deployed an Incident Command Collaboration System that provided a way for incident commanders and emergency liaisons from the various agencies to connect with one another.

A year later, a massive earthquake levelled some 600,000 buildings and killed thousands of people in Nepal, leaving the remote, mountainous country faced with the massive task of rebuilding. "Disaster relief is always overwhelming," Dan Strode, project manager for the United Nations Development Programme (UNDP), said at the time. "There's too much to do, too many people that need help and never enough time or resources."

The daunting task of rebuilding began with mapping where the original structures had stood. In the past, such records were maintained on paper. However, to expedite reconstruction, the Microsoft Innovation Center in Nepal built a mobile phone application (Figure 1-4) that used a device's GPS to help workers record the outline of a damaged home and store it in the cloud before clearing the debris. And to help restart the economy, the app also managed daily cash payments to the workers. Cloud applications like Office 365 and the Microsoft Power BI data visualisation tool helped them to co-ordinate and track progress.

Figure 1-4: Nepal's debris management application

Lessons Learnt

What have we learnt? These examples demonstrate the potential that the cloud offers. We explored how customers are able to do the following:

- Build and rapidly deploy applications with reach and scale that would have been impossible from their own datacentres
- Communicate with Internet-connected devices all over the world
- Tap into Big Data and analytics services for personalisation, better products and more efficient processes
- Enjoy unprecedented development, test experimentation and innovation cycles

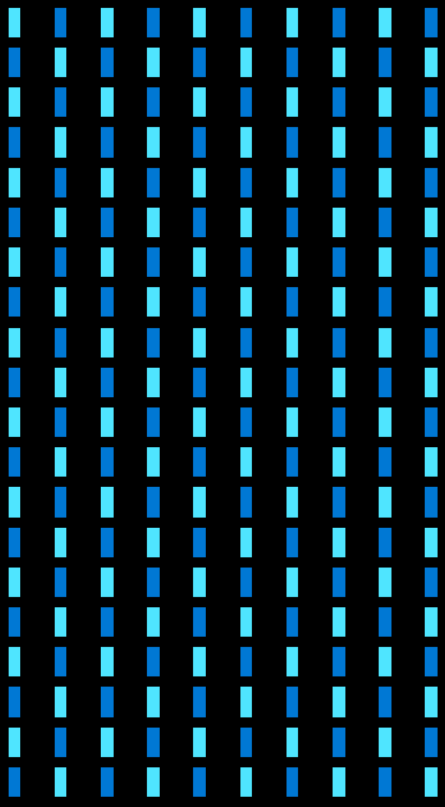
Every IT department is charged with safeguarding its company's information assets, reducing costs and "keeping the lights on." These functions are, and always will be, critical components of any IT organisation. Yet IT must also facilitate and foster innovation, both to make existing processes faster and cheaper as well as to support new and emerging business models.

With the cloud, the balance between maintenance and innovation shifts. As we shall see, operating in the cloud provides many cost advantages, allowing IT departments to focus more on innovation. Running in the cloud can reduce the need for rote operations such as system software upgrades and patching, thus permitting IT to redirect staff towards revenue-centric activities. And new capabilities in the cloud make new kinds of powerful applications possible. As we have seen in the preceding examples, more and more companies now see the cloud as a way to accelerate business innovation and competitive differentiation.

But, as with any great technological change, this kind of transformation cycle involves much more than pure technology. It also requires a shift in corporate culture, enterprise and IT processes, individual roles, governance and (for that matter) engineering. How an enterprise achieves this transformation is the subject of the remainder of this book.

Chapter 2

The cloud: what is it?



Like any new technology, the cloud comes with a whole new set of terms, acronyms and abbreviations. Nevertheless, it's important to understand the different forms of cloud computing in order to make the right decisions about how to use it. In this chapter, we examine the ways in which the cloud manifests itself and how you can employ each in enterprise computing.

So, what is the cloud?

At its core, the cloud physically consists of millions of servers distributed across multiple, very large datacentres strategically located all over the world. All cloud providers use custom-designed server hardware that is focused on reducing cost, improving environment footprint and, of course, providing the greatest computing capability.

The datacentres in which the servers are contained are themselves designed for maximum efficiency and minimal environmental impact; considerable research goes into making datacentres as green as possible. For example, Microsoft's datacentre in Quincy, Washington, is located next to a hydroelectric facility on the Columbia River; this is emblematic of how cloud providers take advantage of local opportunities to reduce their carbon footprints. Elsewhere, datacentres in cooler climates use ambient air rather than air conditioning systems to reduce electric consumption. Some providers use wind power, and others use cheaper non-potable water in the air conditioning systems where necessary.

A key measure of datacentre efficiency is called *Power Usage Effectiveness* (PUE), which measures how power coming into the datacentre is used. A perfect PUE score is 1.0, meaning that all of the power goes to the computing equipment (formally, PUE is defined as total facility energy used divided by that consumed by IT equipment). Traditional enterprise datacentres typically realise a PUE of 2.0, meaning that half of the incoming energy is used by non-computing equipment such as air conditioning, lighting and so forth. Cloud datacentres are now achieving PUE ratings of 1.1 and sometimes even less. This is the result of significant investment and innovation on the part of the cloud providers.

Public, Private and Hybrid Clouds

In the following subsections, we define and examine the three major cloud models.

Private Cloud

The first set of definitions we'll discuss is the distinction between private and public clouds.

The term *private cloud* is often misused; some will say it is the same as a traditional on-premises datacentre. In fact, they are very different. In the traditional on-premises model, IT departments purchase hardware as applications need it, and often this year's servers will look and behave very differently from last year's. Moreover, IT departments traditionally maintain a mix of hardware and software, ranging from mainframe to PC server, with a variety of operating systems, databases and other system software. All of this effectively prevents the notion of on-demand computing, which is the essence of the cloud.

In a private cloud, technologies specific to the cloud model are hosted in an on-premises datacentre, with large numbers of commodity hardware running identical system software: in other words, a "cloud" that belongs to you. Private clouds can be useful because they can implement a technology stack that is consistent with the public cloud. This might be necessary in scenarios for which certain applications or data cannot be moved off premises (we discuss reasons for not moving to the public cloud in Chapter 6).

However, private clouds are of very limited use. They do not provide the cost savings and efficiencies that the public cloud can, because private clouds require a significant capital expense budget and an operations staff, so they remain on your company's balance sheet. Moreover, individual companies cannot achieve the aforementioned economies of scale of a public cloud provider, so their costs are proportionately higher.

Public Cloud

A *public cloud*, which is the primary focus of this book, is built, managed and maintained by a large technology vendor that makes computing, storage and software available for hire. The leading public cloud vendors have datacentres all over the world with literally millions of servers available for use. Customers (enterprises) can either take advantage of applications that already exist in the cloud or they can upload their own proprietary applications, and, as we shall see, there are a number of ways in which applications can physically exist in the cloud but appear to be private to the enterprise corporate network.

Hybrid Cloud

Often, an enterprise will want to keep some of its applications on-premises while moving others to the public cloud. Of course, it is desirable that all of these applications continue to run as they did previously – that is, as if they were all still local and on the same network. When some applications are in the cloud and some are on-premises, this is termed a *hybrid cloud*. Every enterprise will have a hybrid cloud at some time, even if they plan to eventually move all of their applications off-premises, there will be a time during the transition when some applications have moved and others have not: a hybrid model.

To securely connect the two environments, multiple solutions exist. You can set up a Virtual Private Network (VPN), which makes cloud applications appear to be on the same internal network as the enterprise. You can set up VPNs on a per-application basis or, with a hardware device, for the entire corporate ecosystem.

Alternatively, enterprises can purchase a dedicated line through their telecom provider, linking the corporate datacentre with the cloud; bandwidth can be purchased as needed. This solution is preferable when you want to keep all traffic off the public Internet or when substantially higher bandwidth is required. However, it entails additional cost, of course.

Hyperscale makes computing an on-demand service

With the cloud, computing can operate at *hyperscale,* meaning that computing resources scale with the demand placed on them. Hyperscale computing requires 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 back 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 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.

Hyperscale also means that computing capability can be accessed from anywhere in the world with similar latency, which in turn means that cloud providers must build enormous datacentres all over the planet (which they have). The global scale of the public cloud, in turn, provides any number of new capabilities, such as the ability to do geo-distribution of data and to do cross-region failover, to name just two.

The potential, then, of hyperscale computing – its features and its economics – far exceeds that of any enterprise datacentre.

Because of this incredible global scale, computing can be provided *as a service,* meaning that the cloud offers a set of capabilities that enterprises can rent and use for a period of time, add on to as more capability is needed, and then discontinue when no longer needed. Of course, as we've noted, this model is analogous to other commonly used services such as telecom, electricity and so on: you pay for what you use and no more.

"As a service"

As we've said, in the cloud, computing is made available as a service, and there are three predominant application models for cloud computing. Let's take a closer look at each of them.

Infrastructure as a service

With the infrastructure as a service (IaaS, pronounced "eye-as") 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 co-exist on a cloud server. A thin piece of code called a *hypervisor* ensures that each one runs in a timely and efficient fashion.

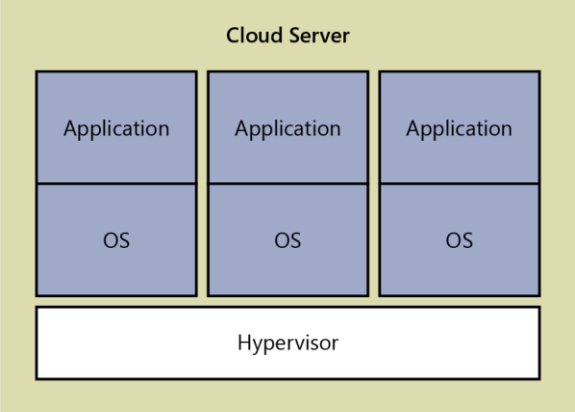


Figure 2-1: Infrastructure as a service

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. But, it still means that your operations staff will need to perform tasks such 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, pronounced "pahz"), 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.

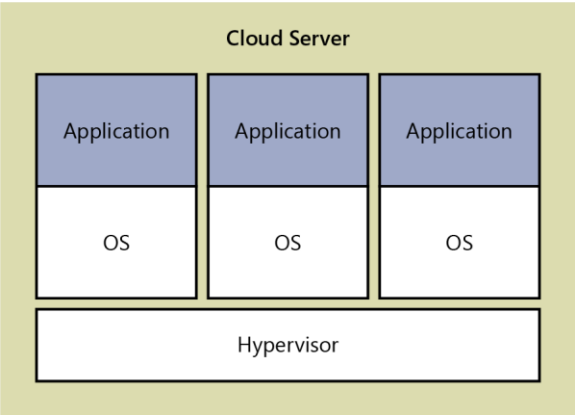


Figure 2-2: Platform as a service

Note that in Figure 2-2 the pieces in blue – the parts that the user must supply and maintain – consist *only* of the application.

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 requires some redesign of the application in order to best take advantage of the model.

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

As we discuss further in Chapter 6, when choosing how to move functionality to the cloud, you should always be on the lookout for opportunities to use SaaS-based applications. Usually, they will provide you with the highest return on investment.

Containers

Containers, which lie somewhere between IaaS and PaaS on the "as-a-service" spectrum, are a means by which applications can share a single instance of an operating system, as illustrated in Figure 2-3. This provides the appropriate isolation and security guarantees preventing applications from "stepping" on one another. Because starting a containerised application typically does not involve loading and initialising an entire VM with an operating system, container start-up can be very fast, so scale-up and scale-down can be very high performing.

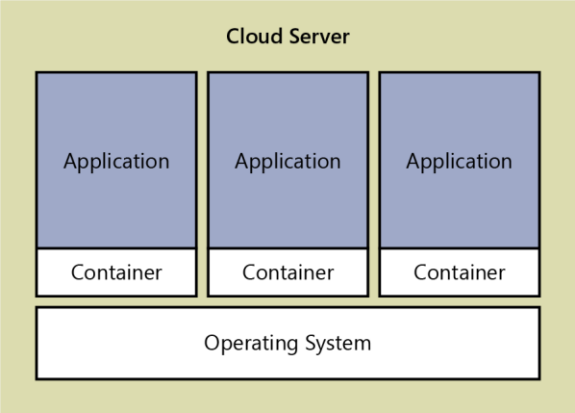


Figure 2-3: Container architecture

Containers have many advantages. Often it is possible to package an application with few or minor changes to run within a container. Having created containers, it's often useful to deploy multiple copies for scale or resiliency reasons. A related technology, *orchestration,* can help automate the process of deploying many copies of many different applications or components to a *cluster* of servers. We discuss all of this, including tradeoffs, in more detail in Chapter 10.

“As-a-service,” compared

Figure 2-4 compares the various "as-a-service" technologies with on-premises computing. The items in blue represent components or software that the enterprise (you) are responsible for maintaining; the items in orange are the responsibility of the cloud provider.

As you can see, for an on-premises datacentre, the enterprise is fully responsible for everything, from the datacentre's operation, the facilities, electricity and air conditioning, all the way to the application. As the migration to the cloud progresses, more and more of these expenses are borne by the cloud provider.

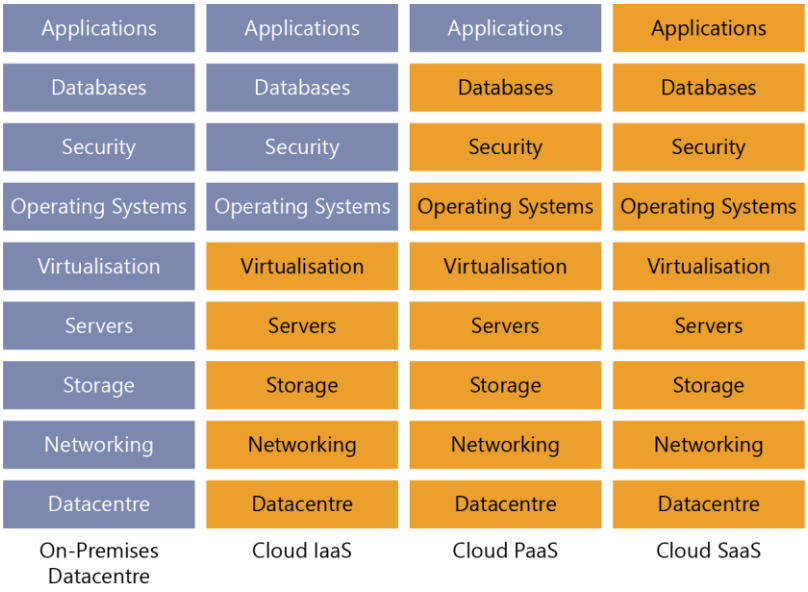
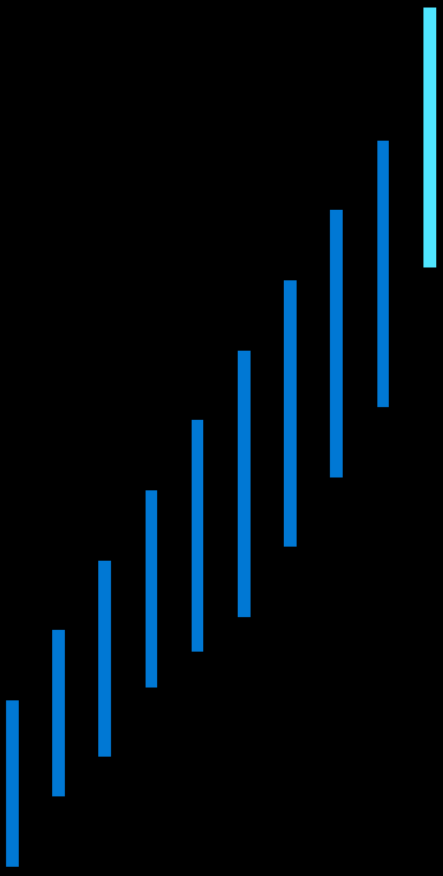


Figure 2-4: "As-a-service" compared

Chapter 3

Journey to the cloud: the roadmap



What if you were able to achieve both efficiency and innovation in all of the business domains and applications across your entire portfolio? What if you could take advantage of the cloud and all of its resources and features to get a "the whole is greater than the sum of its parts" effect? With a good roadmap to lead the way, you can. This chapter covers what it means to move your enterprise to the cloud.

In any transformative change, it's important to understand what the destination is and what the waypoints along the journey will be. There are multiple potential destinations for any application, and IT cloud deployments will be a mixture of them.

Don't miss the opportunity to modernise

Before we go on, it's worth noting that the cloud provides an opportunity to consider the IT ecosystem as a whole and how you can *modernise* it. As you shall see, cloud migration at scale involves looking at each application and determining how it should be thought of in this new environment. Is further investment in certain applications justified? Should they be retired?

Many enterprises have held on to their applications for far too long without assigning them a maintenance or retirement schedule. Therefore, due to fear of complexity, lack of documentation, resources, source code or other reasons, applications remain untouched.

Even for applications that remain on-premises, modernisation can save time and money. An internal Microsoft IT study several years ago demonstrated that the number of problem reports ("tickets") and the time to resolve them increased with the age of the application and system software. (This analysis led to a focused effort to ensure that all applications were on the latest version of the operating system and other systems software such as databases.)

The opportunities provided by the cloud to IT represent a seminal event to re-evaluate your entire ecosystem, in particular an opportunity to evaluate and modernise applications. This activity in and of itself can provide great returns on investment and positively affect the top-line revenue.

Evolution of the five Rs of modernisation

To focus our efforts on guidance for existing applications, let's proceed with the most convenient way to think about modernisation, which is commonly called "the five Rs":[2] retire, replace, retain and wrap, re-host and re-envision. This ontology was originally formulated by Gartner in 2011, and we've expanded it over the years based on our experiences.

It's likely that no single approach will be appropriate for all of an enterprise's legacy applications, and a mix of differing approaches might be warranted, as illustrated in Figure 3-1, based on the value that an application delivers versus the cost of any given approach. Because these approaches depend greatly on the situation, application and types of cost involved, there is no one-size-fits-all solution.

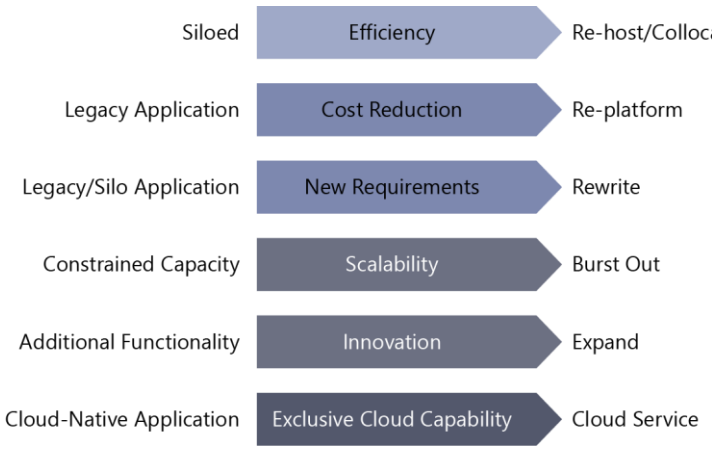


Figure 3-1: Types of modernisation initiatives

- **Retire** Of course, if a legacy application is providing little value compared to its costs, the enterprise should consider it a candidate for retirement. When few people are using an application relative to its cost impact, the enterprise needs to run a cost-benefit analysis to determine if it is worth the expense. Additionally, some functionality provided by legacy systems can be rolled into a consolidated modern application running in the cloud, allowing some applications to be retired while others are replaced and modernised.
- **Replace** Often, a legacy application is providing some value, but an off-the-shelf replacement with a lower total cost of ownership (TCO) is available. Many legacy applications were originally built because there was no alternative at that time. A modern, readily available application that is better suited to running in the cloud – most cost-effectively of all, a SaaS application – may now exist that you can use to replace the older one. Also, when a legacy application is replaced with a more comprehensive modern solution, there might be a chance to consolidate functionality from several older applications, thereby replacing multiple applications with a single system.

[2]Based on "Gartner Identifies Five Ways to Migrate Applications to the Cloud," Gartner Inc, 2011. http://www.gartner.com/newsroom/id/1684114

- **Retain, wrap and expand** If a legacy application is providing good value and not incurring a high TCO, the best approach might be to retain it but put a modern "wrapper" around it in order to gain additional value and benefits. Examples of the "retain and wrap" approach include the following:
    - Using API management tools, such as Microsoft Azure API Management, to add an API so that authorised external applications can use the application functions.
    - Extend a legacy application with third-party tools; for example, by using a C# wrapper around older applications, or by making their data available through Extract, Transform and Load (ETL) or other approaches, to connect them to other software components, such as analytics applications, machine learning or mobile access.
- **Re-host** If a legacy application is providing good value, but is expensive to run, it might be a candidate for re-hosting. Re-hosting involves keeping the same basic functionality, but moving it to the cloud where it is easier to manage and less expensive to run. This is also called "lift and shift". In a re-hosting situation, the legacy application might be currently located either on a local virtual machine (VM) or on local hardware. Some VMs might be eligible to move with a simple migration. Those on local hardware might be able to be converted with a physical-to-virtual migration and then hosting the VM in the cloud. Some VMs, especially older ones, might not relocate to the cloud easily without some significant work. In those cases, you might want to consider re-envisioning and building the application in the cloud.
- **Re-envision** If a legacy application is providing good value, but cannot be easily moved, the best solution might be to re-envision it and build it again in the cloud. Re-envisioning is a process of rebuilding the application in the cloud using modern technology, a new architecture and best practices; it normally also involves adding more business value to core functionality, such as improving market differentiation. Re-envisioning an application might require rewriting the main logic by using a modern development language and tools and making it service orientated. Re-envisioning an application can be facilitated by starting with VMs in the cloud, which can be installed in a matter of minutes.

  There are many ways to take advantage of the cloud when re-envisioning, as we shall see. For example, it can be useful to think about "bursting" approaches, wherein, as load on the on-premises application increases, new instances are created in the cloud to handle the temporary overage. Keeping frequently used ("hot") data locally while aging-out infrequently accessed ("cold") data to far cheaper cloud storage is another common pattern. We cover more of these strategies later in the book.

There are a number of ways to think about your strategies for legacy applications. One way is to consider them by workload, as depicted in Figure 3-2.

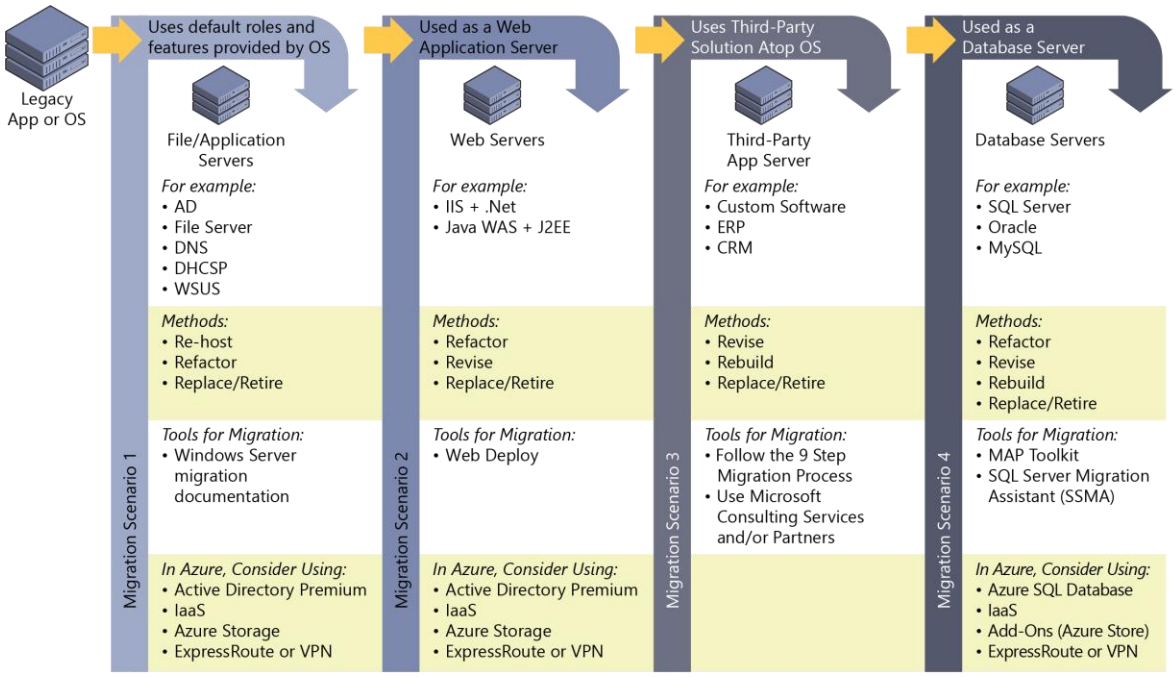


Figure 3-2: Legacy application strategies

In the figure, you can see that you can upgrade legacy software by rehosting it on more modern platforms, or you can move it to the cloud and gain more advantages. We have much more to say in succeeding chapters about the individual technologies, how to migrate to them and how to use them.

Cloud migration: three stages

When planning migration to the cloud, there are many ways to think about a roadmap. From our experience, however, we've seen three basic stages: *experimentation, migration,* and *transformation*. A note, however, before we begin our discussion: in almost every case we've seen, these three stages do not take place in order; often they occur all at the same time. The reasons why might not be apparent at this moment, but (briefly) what often happens is that one group in the enterprise will be experimenting with certain applications in the cloud while others have already moved on to, for instance, a SaaS application. In other words, you don't need to wait for the experimentation phase to complete before trying something transformative, and so on.

Experimentation

In the essential experimentation phase, two processes take place. In the first, the engineers and others create the IT department's first cloud applications, with the objective of learning what the cloud is all about: how to develop for it, how to test, how to deploy and how to monitor and maintain a cloud application. Concurrently, businesses and IT departments envision the art of the possible; design new solutions to demonstrate how to advance the status quo; and envision a newer, expanded, more agile and better application or service.

Migration

In the migration phase, which in many ways is the most demanding of the phases, the bulk of the IT portfolio is moved to the cloud in one form or another. This requires co-operation and collaboration across a number of different enterprise functions, including the technical staff and the operations staff, as well as the executive team, business sponsors, security professionals, regulatory compliance staff, legal department and HR. We spend a significant amount of time in this book covering migration in all its aspects.

Transformation

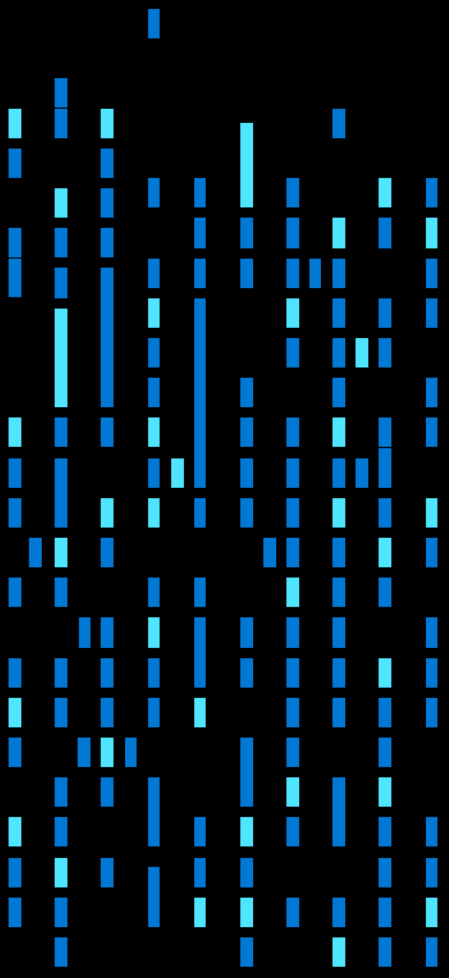
In the transformation phase (which will often coincide with the migration phase), selected applications are redesigned to take maximum advantage of the cloud – using the PaaS model – affording greater scale, greater integration with other cloud services and numerous other advantages.

Moving forward, the now-cloud-native applications can take advantage of cloud services such as machine learning, big data, streaming analytics and many others – making them much, much richer in functions and features than before.

The following chapters cover each phase in detail.

Chapter 4

Experimentation



There is always a first cloud application. In every IT organisation, some brave soul will either move an existing application to the cloud or create a new one there. In so doing, this person will gain an understanding – beyond all the hype – of what developing, testing, deploying and maintaining a cloud application is all about.

Microsoft IT's first cloud application

Microsoft IT developed its first cloud application in 2010. It was an employee auction application, used once a year as part of the Microsoft charitable giving campaign (see Figure 4-1). With it, employees donate items (ranging from mentoring sessions, to cooking classes, to software, to perks such as the use of an executive's car for a day!) and others buy them, with all the proceeds going to charity. The auction, typically held in October, runs for a month.

Why did we pick this as our first cloud application? A number of factors led to this decision. Firstly, it was *not* a business-critical application. Therefore, news of any application problems would not cause damage to the company's finances or reputation or appear on the front page of any newspaper.

Secondly, we could see the scalability features of Microsoft Azure in action. As the end of October approached, traffic on the application continually rose, reaching a peak in the last few days of the auction.

Finally, it was a relatively simple application whose deployment in the cloud did not require other applications to be updated in concert.

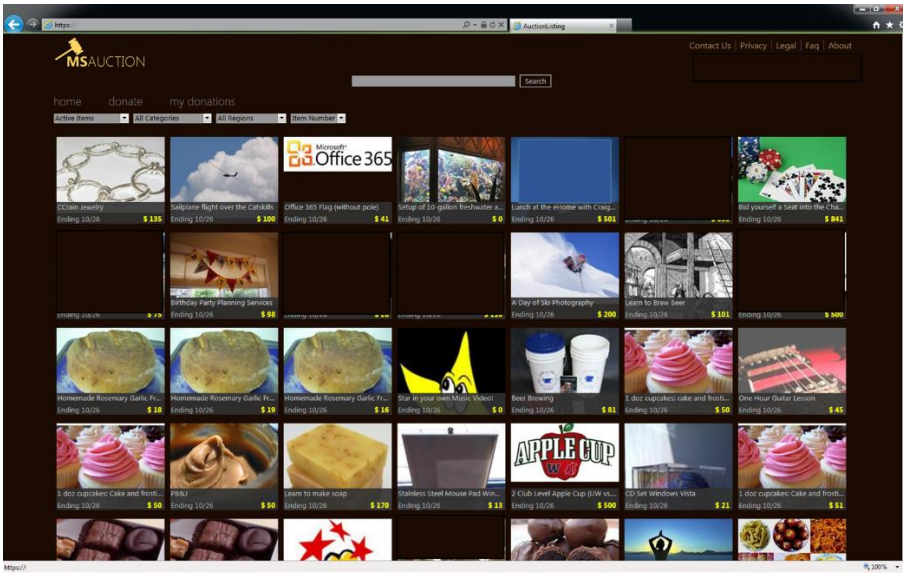


Figure 4-1: Microsoft internal auction application, circa 2010[3]

In the end, the application was very successful and the auction met its goals (incidentally, over the years, Microsoft's employees have raised more than one billion dollars for charity). Microsoft IT many lessons on cloud development and deployment, which we used in subsequent stages of our own journey. We saw the application easily scale to meet the increased demand over the month. At the end of the auction, we could shut it down and no longer pay for the resources required to run it (as we would have – for servers, operations staff and so on – had we run the application in our own datacentre). By every measure then, this first experiment was a success.

There were many other early experiments in this period, trying out new approaches, testing new features and so on; we learned that developing a "culture of experimentation" was useful in that we could be continuously trying new things and innovating.

If you are just beginning your cloud journey, consider using your first few applications as ways to thoroughly explore the possibilities of the cloud: use telemetry to monitor their operation; collect the data and analyse it using any of the big data and analytics capabilities in the cloud; build a dashboard or use machine learning to predict user behaviour. These are all excellent ways to familiarise your staff with cloud services. These individuals then will become the core of your cloud centre of expertise, passing along the lessons they have learnt to others on your team.

Shadow IT and the culture of experimentation

IT departments often live in a world of contradiction. On the one hand, they must "keep the lights on," by keeping servers and networks up, by delivering reports on time, and by ensuring that systems, data and the processes they support meet regulatory obligations such as Sarbanes-Oxley, the Health Information Portability and Accountability Act (HIPAA), the financial Payment Card Industry Data Security Standard (PCI DSS) and numerous other forms of compliance. These requirements are nothing if not rigorous – and essential. (We cover compliance in more detail in Chapter 9.)

[3] Mentoring, tutoring and personal one-on-ones with executives are always among the items offered for auction. In the interest of privacy, we have removed the faces of the individuals offering these sessions from the screenshot; hence, some squares are blank.

On the other hand, IT and its business partners recognise the importance – indeed the absolute *necessity* – of innovation: new programmes and new applications to support both new and evolving business opportunities, to better serve their customers, and so forth. Yet the costs of IT operations – sometimes 70–80 % of the overall budget – reduces the ability of IT to spend on new programs and innovation.

In many cases (in fact, in every enterprise we know), there are occasionally applications created and deployed outside of the IT department in response to critical business needs. These unofficial applications are often referred to as "shadow IT." So, instead of going through the usual budget, requirements analysis, design and deployment phases typical in the creation of a new IT application, a marketing department publicising a new campaign might simply create a new website on its own.

Because it eliminates the capital-expense investment component (i.e. servers, storage and network) of application development, the cloud makes this sort of rapid innovation much, much easier. In effect, all that is needed are a few coders to write the application – and a credit card.[4]

IT executives should realise that this sort of innovation and experimentation is inevitable, and in many cases actually desirable. As the business climate rapidly evolves, it is critical for both businesses and IT organisations to foster rapid experimentation and innovation.

It will be important to educate businesses on the importance and consequences of regulatory issues and noncompliance, of course. IT departments can actually help them by providing controlled, managed access to critical data, such as customer information, rather than letting those businesses gather and manipulate the data on their own.

As soon as a company starts this process of envisioning and creating the culture of experimentation, it learns a disruptive truth: in the cloud era, you must experiment, fail fast and learn fast. It is important to experiment in order to learn quickly from both successes and failures. Learning from how you succeed and what makes you fail provides the basis for delivering the disruptive innovation and value from the cloud.

Principles of a culture of experimentation

The culture of experimentation might seem jarring to the traditional IT shop, which, as often as not, focuses on carefully controlled development and risk reduction. Fostering experimentation, however, will greatly enhance the cloud adoption process.

The principles we've used are *go fast*, *push the boundaries*, make *data-driven decisions*, *simplify* and, finally, *communicate* to succeed. Table 4-1 provides an overview of these principles, followed by detailed descriptions of each.

[4] The elimination of such capital expenses has greatly accelerated the pace of start-ups, as well.

Table 4-1: Cloud migration principles

Go Fast	Push the boundaries	Make data-driven decisions	Simplify	Communicate to succeed
Fail fast, learn fast Try many, use best	Design new applications and capabilities for PaaS/SaaS Refactor legacy apps for PaaS/SaaS Build your plan-of-record to take advantage of cloud capabilities Think "Experience"	Manage your costs Use telemetry to gain insight into operational efficiency Understand your blockers Manage your plan-of-record	Retire, retire, retire legacy applications wherever possible Aggressively right-size Review frozen and cold servers weekly Clean up Configuration Management Database (CMDB) data	Communicate customer and stakeholder impacts – transparency is key Share what has been learnt and best practices

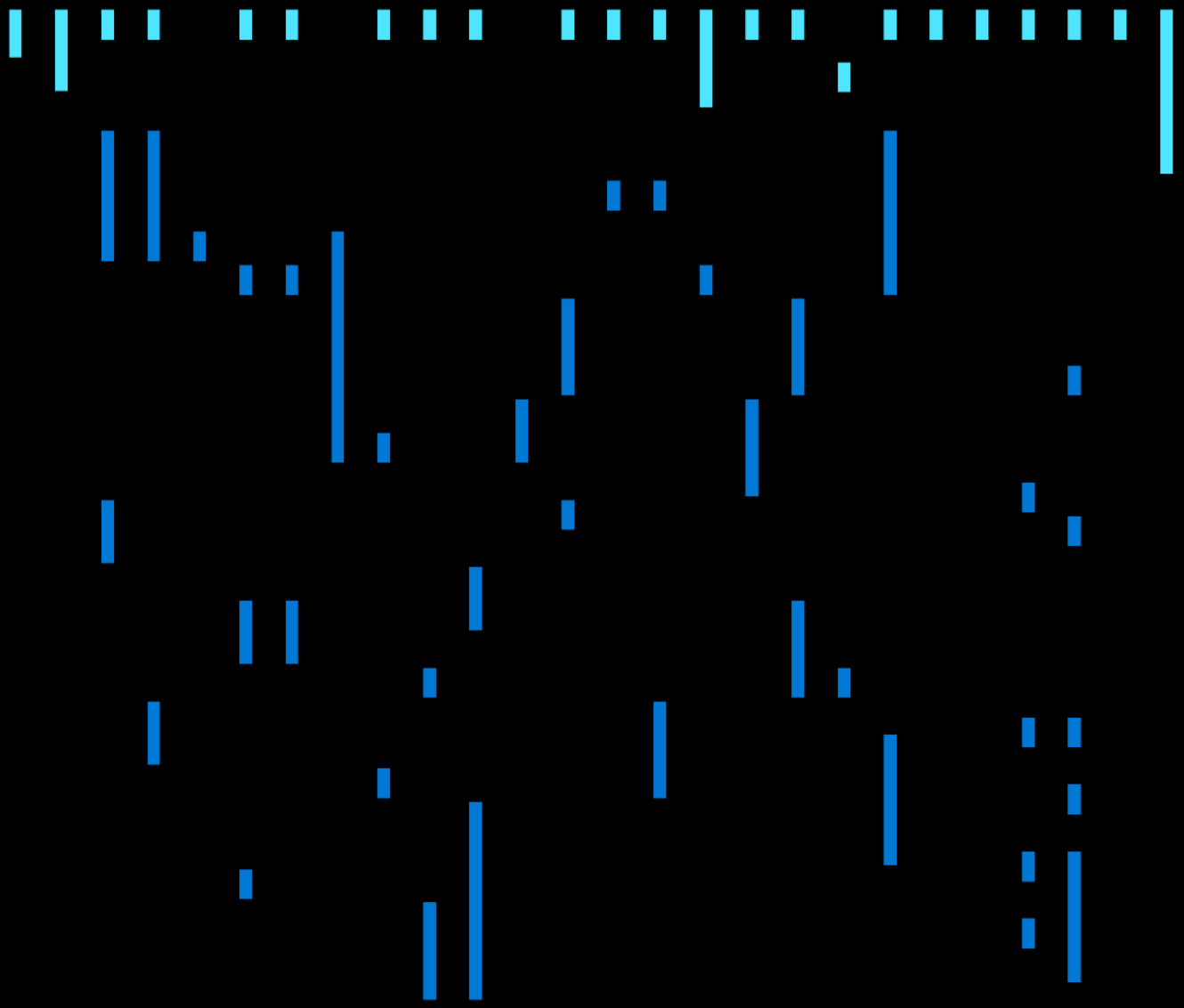
- **Go fast** This exemplifies the spirit of the experimentation phase. For some, it might represent a new way of thinking for IT because, with the cloud, you can "spin up" new projects quickly with a few clicks rather than having to plan, allotting datacentre space, procuring equipment and so on. We call this the *try many, use best* approach because the cloud uniquely facilitates the ability of IT departments to choose the best of many solutions.
- **Push the boundaries** This principle suggests that wherever possible, IT should not simply adapt to the new paradigm of the cloud, but embrace it and adopt new architectures and processes quickly to best exploit the new opportunities.
- **Make data-driven decisions** This proposes that you carefully track and measure the numbers, including the cost effectiveness of the cloud for financial reasons, system telemetry for technical efficiency reasons and so on. Following the data carefully will make it possible for you to make informed decisions about which applications are generating the most return, about which you should prioritise, about which are performing well in the cloud and where potential problem areas exist.
- **Simplify** This focuses on retiring, right-sizing and consolidating as many services and applications as possible. Applications that are infrequently or rarely used often generate significant costs for an IT organisation, with little return. Retiring them and consolidating them with applications that perform similar functions can, conversely, generate savings in a number of areas such as hardware, system software licences and maintenance. Consider generating metrics around "hot" and "cold" applications based on CPU, network and database usage; for example, an application that averages 2 % of CPU and has few authenticated users might be just such a "cold" application.
- **Communicate to succeed** This principle is the single most important mechanism that guarantees continued success, not just the migration of a single application or a service. Establish a clear and continuous communication channel for stakeholders to visualise success and impact as well as to understand failures and the lessons learnt from them. Key stakeholders remain engaged and continue to invest when they feel their participation in the joint effort is required to make this a continuous journey, not just a single trip.

A thoughtful approach to experimentation can yield great rewards. Experimentation by no means implies all controls are removed; rather, IT executives should set in place "sandboxes" where this can happen. As examples of sensible constraints, one might posit that experiments should be conducted on non-critical business applications or processes; that they should not access sensitive data such as Personally Identifiable Information (PII); and so forth.

Much will be learned from experimentation. These lessons set us up for the migration phase, which we cover in Part II.

Part II

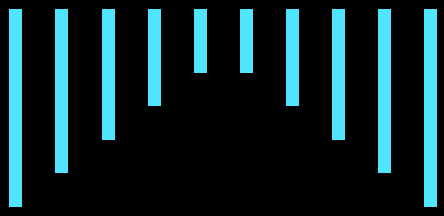
Moving IT to the cloud

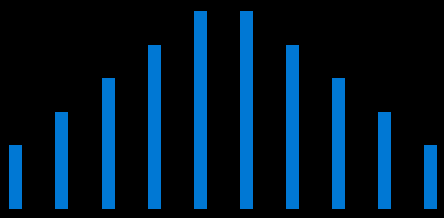


Microsoft

Chapter 5

Building the capability





Sooner or later, it becomes obvious that running a large portion of the IT portfolio – perhaps even the majority of it – in the cloud makes sense from a variety of perspectives. As we discussed in Part I, running in the cloud provides a number of advantages, including cost savings, agility and innovation opportunities. The cloud is very compelling, yet, the migration phase typically involves many more applications and many more people, and potentially affects more of IT's customers than any other phase – by far.

It can be daunting when a large enterprise IT department manages hundreds or thousands of applications running on perhaps tens of thousands of virtual machines (VMs). Which ones to move first? How to prioritise? How does operating in the cloud affect regulatory compliance, data security and enterprise processes? What does it mean for organisational roles, training and change management? And, last but certainly not least, how to do all this while continuing to serve the business?

Where to begin?

In Part II, we describe how to establish strategy and goals for a cloud migration activity; what roles the various organisations in the enterprise play; how to prioritise application migration; and how to extend IT governance to cover the cloud.

Establish strategy and goals

Every journey must have a sense of its destination, its route and when it will arrive. A migration journey to the cloud is no different. It is time well spent to engage senior members of IT – and, indeed, business leaders from around the enterprise – to understand all aspects of the cloud and which of the many options and approaches to take.

In Microsoft IT, as in many enterprises, the journey began with the creation of a Cloud Strategy Team, driven (in our case) by the CTO and consisting of members of the enterprise architecture team, IT finance, the most senior technologists from the various IT applications groups (HR, finance and so on), and leaders from the infrastructure, security and networking teams. Figure 5-1 shows the structure of the Cloud Strategy Team.[5]

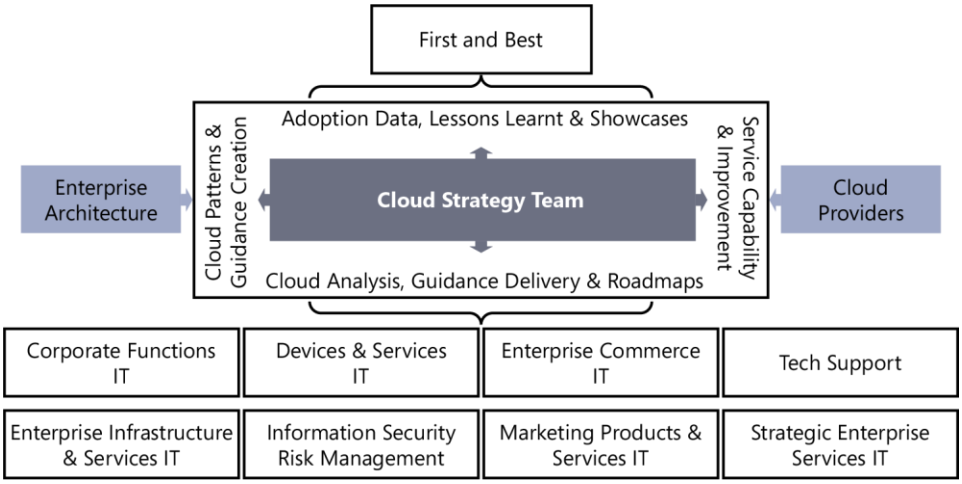


Figure 5-1: The Cloud Strategy Team at Microsoft IT

The Cloud Strategy Team was chartered to lead the cloud analysis and experimentation phase previously described (see Figure 5-2). In addition, it built (or facilitated the building of) the architectures, patterns and guidance for deployment of the re-envisioned applications or services to finally manage the communications to key stake holders and promote the success and what has been learnt from the programme.

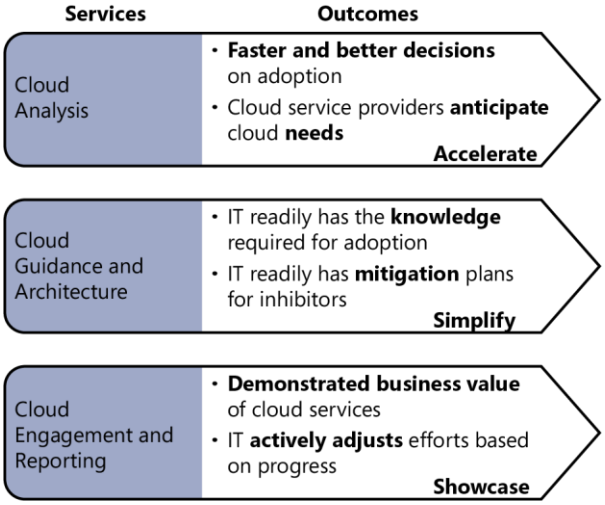


Figure 5-2: Cloud Strategy Team charters

[5] Microsoft IT's "First and Best" team ensures that Microsoft IT is Microsoft's "First and Best" customer by testing all of the company's products in IT prior to their general release to the public, a practice often referred to as "dogfooding."

The creation of this team was one of the key forces promoting long-term commitment to the journey. It established a practice for continuously evaluating and experimenting to help determine the appropriate platform and destination for each application – namely, what is moved to the most appropriate platform, such as the following:

- If moved to the cloud, should it be left as a simple VM for infrastructure as a service (IaaS) or redesigned for platform as a service (PaaS)?
- Could a pre-existing software as a service (SaaS) model replace it, saving costs?
- Or should it remain on-premises?

Among its first tasks, the team educated itself, ensuring that all participants were on a level playing field. For better or worse, cloud technology comes with its own set of acronyms, as is discussed in Chapter 2; learning to speak a common language early accelerated future conversations. The team also spent time familiarising itself with the offerings from platform, tools and cloud applications providers.

When the team began to draft out the strategy, members understood that not all services or applications would end up in the public cloud, for various reasons. Microsoft IT's strategy, therefore, was based on the notion of a *hybrid* cloud (see Figure 5-3). This meant that certain applications would remain on-premises for some time.

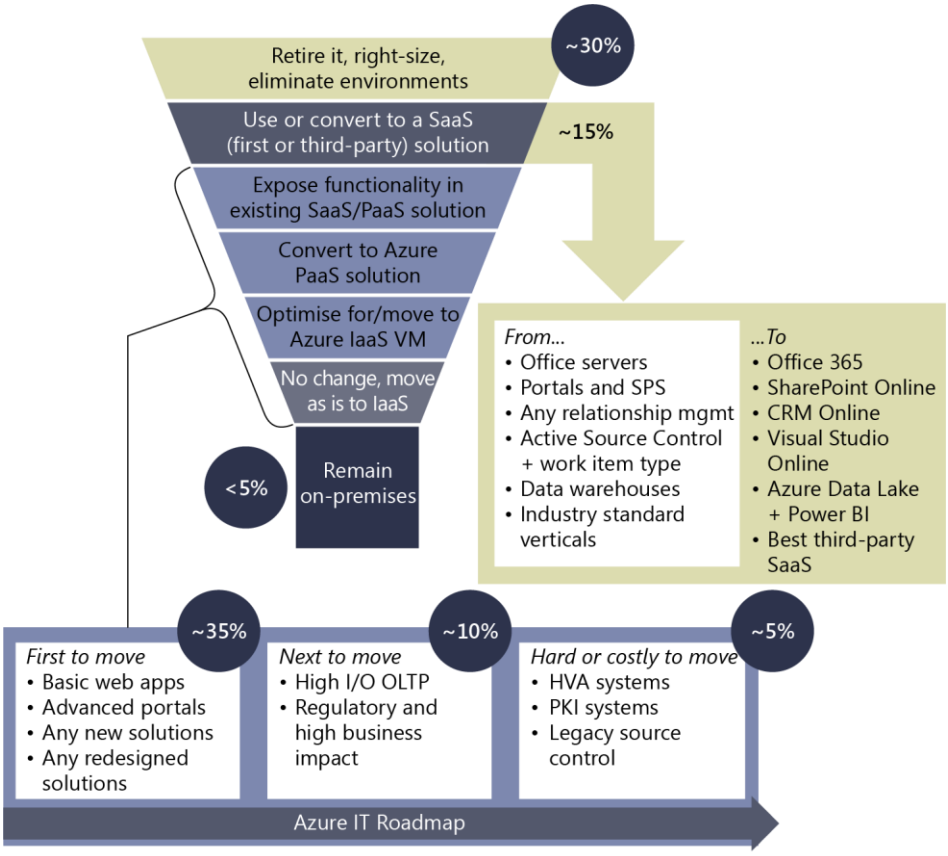


Figure 5-3: Hybrid cloud strategy

On the other hand, we clearly realised that the optimum strategy from an efficiency and cost point of view was, first, to see if we could retire the application by consolidating it with another one with similar functionality, or, if it had very little usage, to remove it altogether.

Next, we looked for applications that could be moved to a SaaS model; one in which Microsoft IT did not need to maintain either hardware or software. Then, if we had already invested in a SaaS application but weren't using all of its functionality, we looked to see if we could use more of it to replace applications. Certain customer applications – those undergoing significant new development – were converted to Microsoft Azure PaaS applications.

Other applications were migrated to a cloud IaaS environment, meaning they were hosted in the cloud but still required operating system and database maintenance from us. Lastly, a very small number of applications remained on-premises for various reasons such as legacy code.

In Chapter 6, we show, in considerable detail, the mechanics of this analysis.

The Cloud Strategy Team's deliverable was a document describing the goals of the migration, proposed timeframes, recommended technical strategy (that is, technical platform and tools) and expected results and benefits. For example, the recommendations included statements like the following:

- The majority of existing applications will be initially moved to IaaS VMs as a relatively quick way to move them to the cloud (no code changes required).
- To take advantage of scalability and other features, new applications and major releases will be (re)designed as PaaS applications.
- During the transition, on-premises applications will communicate with cloud applications via a dedicated connection (typically MPLS or WAN) line – in our case, we used Microsoft ExpressRoute.
- Applications that provide little competitive differentiation (applications that can be commoditised) will be transitioned to external SaaS providers (for example, Microsoft's Office365 for mail and productivity applications).
- Expected cost savings will be *x* % after the first year and *y* % after the second.
- Certain applications will remain on-premises for the near term (due to legacy architectures, complexity of integration and similar issues).
- Security will be provided through combinations of encryption, cloud identity federated with on-premises identity providers (such as Active Directory) and other controls.
- Operations teams will be trained in cloud deployment and systems management in the cloud and will evolve to a DevOps model (discussed later).

Documents of this nature can include different models and options to facilitate discussion and informed choice.

Organisational responsibilities in creating the strategy

As we conducted our initial investigation, it became evident that the cloud affected many organisations inside IT and a number outside, as well. That being the case, it was important to ensure that they participated in the decision-making process. In this section, we describe how each of the following organisations contribute to the cloud strategy:

- Enterprise architecture
- Information security and risk management
- Data classification
- Enterprise risk management
- Finance
- Operations
- Human resources
- Applications teams
- Business units

Enterprise architecture

The enterprise architecture (EA) organisation can play a key leadership role in cloud migration. The goal of any EA team is to ensure that the highest business value is received for most efficient use of technology resources – as such, EA provides the essential bridge between business and IT.

Typically, EA maintains the list of IT capabilities and processes, facilitates the creation and implementation of IT strategies, works with businesses and executives to understand the long-term goals of the company in order to plan for the future and drives various enterprise-wide governance activities, such as architecture review. For such reasons, the EA team is an ideal choice to lead the Cloud Strategy Team.

The EA team overseeing the IT ecosystem as a whole is in a position to provide the appropriate analyses of system capabilities and application impacts of any large-scale changes to the ecosystem. Often, it is EA that creates and maintains the portfolio management system (the catalogue of applications) from which the prioritisation of applications to be moved to the cloud can be drawn (we will have much more to say about this process later). Enterprise architects should examine what is known about the portfolio and where additional information is needed – for example, whether an application is virtualised – the EA team should add this and other attributes to the knowledge base and engage with other parts of IT to collect the data. Other examples of such metadata will be described shortly.

Cloud migration offers the enterprise architect many opportunities. By using modelling techniques such as business capability analysis[6] and capability maturity models, it might be possible, as the prioritisation process for applications takes place, to *simplify* IT by consolidating applications with similar functions. Consolidation will have clear financial benefits both by reducing the compute, data and network requirements, and by simplifying the operations and maintenance functions.

The enterprise architect, and in particular the enterprise information architect, can also use the opportunity afforded by cloud migration to analyse the data models used by applications and update them to enterprise-wide canonical models. Such an effort will streamline application integration and reduce semantic mismatches between disparate data models, which often require manual adjustment in a complex on-premises environment.

In addition, it is the EA team's core responsibility to create and maintain as-is and to-be roadmaps of the overall IT ecosystem. The EA team should easily be able to communicate the various stages of the migration, summarising the current thinking of the Cloud Strategy Team.

Finally, the EA team should direct the investigation into the use of new cloud technologies to either augment existing capabilities and/or provide entirely new functionality to IT applications, and as these are validated, to add these to the existing roadmaps. Enterprise architects need to experiment with new technologies as well as understand and communicate their business value to IT management and business stakeholders. Successful investigations should lead to the development and publishing of reference architectures that applications teams can reuse.

Information security and risk management

Every major change in the way you conduct business entails some amount of risk; few aspects of the cloud have generated more discussion and controversy than those regarding its security and risk. In this time of breaches, nation-state hacking and growing and profound concern with individual privacy on the Internet, cybersecurity has become a board-level concern, and rightly so. Governments and organisations have created regulations and requirements to rein in the risks.

Begin by understanding the security postures of the cloud platform providers. Issues to examine include the availability of antimalware software for cloud-hosted applications; the presence of intrusion-detection software and tools; sophisticated and secure identity management; at-rest and in-motion encryption options; networking options for on-premises and off-premises communications; the ability to do carry out penetration testing; and so on. The requirement to implement "defence in depth" remains; you will need to determine how you can collaborate with your cloud provider to implement and enhance it.

You should also understand the physical security practices of the cloud provider. Are employee background checks required? Does access to the cloud datacentre require biometric authentication?

[6] A modelling technique that analyses an enterprise in terms of its business capabilities, independent of organisation or technology, pioneered by Gartner. See https://www.gartner.com/doc/1415831/use-business-capability-modeling-explore. Capability models are just one possible enterprise architecture modelling methodology, others, such as the famous Zachman Framework pioneered by John Zachman or Business Process Model and Notation (BMPN), can be used either with or instead of capability modelling.

Next, because the cloud potentially makes it possible to access corporate computing devices from anywhere in the world, the information security team should address what requirements should be levied on these devices to grant them such access. For example, it might require all client devices to have encrypted local storage by using such technologies as Microsoft BitLocker. Similarly, because typing usernames and passwords on mobile devices can be tedious, the team should consider the merits of alternate forms of authentication, such as biometrics. Or, it might choose to implement "multifactor authentication," requiring both a username/password as well as some other form of identity (such as a smart card or secondary authentication using a smartphone).

A related capability in the cloud is its ability to accept authentication credentials from a multitude of sources by using the Open Authorisation (OAuth) protocol. Information security professionals should decide which, if any, applications may accept (for example) Facebook or Google credentials. E-commerce sites might benefit from usage of these credentials but internal applications likely would not.

Third, verify key regulatory compliance certifications (for example, HIPAA, the Health Insurance Portability and Accountability Act; FedRAMP, the Federal Risk and Authorization Management Program; and the GDPR, the European General Data Protection Directive). Different industries and different geographies will be governed by different regulations and standards. Learn how to detect a suspected breach and how to report it to the provider, and what the response time Service-Level Agreement (SLA) is expected to be. The Azure Trust Center provides details on all of these as they relate to its offering. The Cloud Security Alliance is an excellent independent resource bringing together experts from across the industry to develop recommendations for best practices for secure computing in the cloud.[7]

You can find a more detailed discussion of cloud security and governance in Chapter 9.

Data classification

Think about the data your applications can store in the cloud and how they might influence security and risk. Many companies classify their data according to its sensitivity: a marketing document has a very different security requirement than, say, a draft of a 10-K filing prior to earnings release.

One possible schema is to divide data into several categories, based upon the impact to the business in the event of an unauthorised release. For example, the first category would be public, which is intended for release and poses no risk to the business. The next category is low business impact (LBI), which might include data or information that does not contain Personally Identifiable Information (PII) or cover sensitive topics, but would generally not be intended for public release. Medium business impact (MBI) data might include information about the company that might not be sensitive in and of itself, but when combined or analysed could provide competitive insights, or some PII that is not of a sensitive nature, but that should not be released for privacy protection. Finally, high business impact (HBI) data is anything covered by any regulatory constraints, anything that involves reputational matters for the company or individuals, anything that could be used to provide competitive advantage, anything that has financial value that could be stolen, or anything that could violate sensitive privacy concerns.

[7] Azure Trust Centre: http://azure.microsoft.com/en-us/support/trust-center/
Cloud Security Alliance: https://cloudsecurityalliance.org

Next, you should set policy requirements for each category of risk. For example, LBI might require no encryption. MBI might require encryption in transit. HBI, in addition to encryption in transit, would require encryption at rest. You should also consider creating audit requirements, access control and other security guidelines based on these categories. The cloud strategy team working with the information security group might, in fact, choose to prioritise applications that manage low-security data (LBI) to migrate to the cloud first because they represent the least risk. High-risk data (HBI) such as customer PII might require a security review before being migrated, whereas LBI applications might not.

Enterprise Risk Management

If you have an Enterprise Risk Management (ERM) team, work closely with it to determine how the cloud affects its risk models. Most ERM teams have a detailed, documented list of enterprise risks along with the likelihood of them happening and the impact if they do. To address these risks, ERM teams will implement controls and establish teams to either remediate or monitor the risk, depending upon its severity. The cloud, as with any significant change, will introduce changes and new risks to the existing risk model, and it is important that these be examined and discussed. For example, in the extremely unlikely case of a cloud datacentre failure, IT departments should consider geo-replicating data to mitigate the risk of data loss.

Finance

It is imperative to involve your CFO and your enterprise's finance department in developing your cloud migration plan. You will need to work with them on developing cost models that compare IT operations on-premises (in the datacentre) against those in the cloud. You'll also need to build models showing how purchasing and procurement of new hardware draws down over time. You might also even build models showing when and how datacentres can close.

Develop some key measurements to quantify the savings more particularly. For example, one measurement we used in Microsoft is called "cost per operating system instance (Cost/OSI)". (We used this so as to include both applications and operating systems running on bare-metal servers as well as those running in VMs as a single metric.) Cost/OSI includes hardware, licensing, facilities, network, operations staff and all the costs of running an operating system and its applications in an on-premises datacentre. You can segment systems if this is useful: we used "t-shirt sizing" and had a metric for small, medium, large and extra-large deployments.

With this metric, you can now compare the cost of running an on-premises system against one in the cloud. Of course, the parameters for Cost/OSI in the cloud are different and include size of the application, number of cores required, amount of storage and estimated network traffic. And, unlike the on-premises case, you can spin-down servers in the cloud when they're not needed or not used, and thus reduce or even eliminate charges.

You should determine your Cost/OSI currently as a baseline. Then, you can forecast costs for various operations in the cloud. Most cloud service providers, including Azure, provide cost estimation tools to help you determine what your Cost/OSI will be under various configurations and requirements.

You need to work with your finance department to develop several scenarios for your cloud migration, including aggressive, moderate and slow migration plans, as shown in Figure 5-4. An aggressive plan might involve moving 50 % of your workloads to the cloud in the first year, whereas a moderate plan might be 30 % and a slower plan might be 10 %. Aggressive plans will potentially save you more, but this must be weighed against greater risk and higher migration costs.

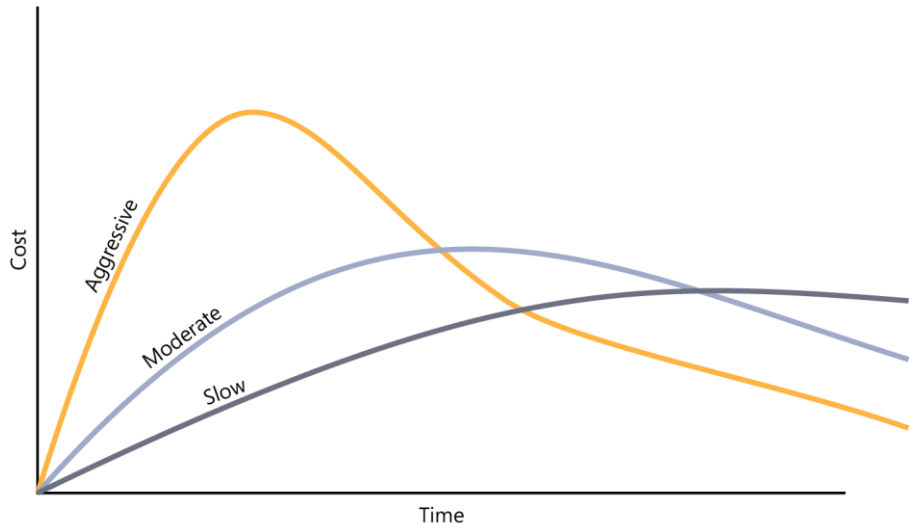


Figure 5-4: Adoption rates and costs

Of course, finance leaders need to understand that the journey to the cloud is about more than just cost savings. They need to view the enterprise's data as a valuable asset that can be made to have greater value based on what we can do with it. Using new types of data; analysing the data to discover insights on your products, customers and processes; frequent experimentation to determine how to maximise the impact from these insights; and scaling these innovations will add significant value to your data. In turn, these actions will provide increased control and reduce the risk to a company's operation; something about which all CFOs care deeply. The more you can quantify how much the value of data will increase and what costs will be saved by moving to the cloud, the easier it will be to get more of the highest-level decision makers to support the move.

Operations

Cloud migration has a very significant impact on daily operations in an IT department. Although *functionally* the requirements of this team remain intact, the *mechanics* of how many of these functions are performed changes in some important ways. Consider some of the following operations tasks and how they will change in the cloud-centric world:

Task	On-premises function	Cloud function
Health monitoring	Using various tools, such as Microsoft System Center, to monitor applications and provide Root Cause Analysis (RCA) of failures	Embed with developers to monitor the applications in real time and rapidly understand impact of (perhaps daily or even hourly) updates (such as DevOps)
Security operations (SecOps)	Use Security Information and Event Management (SIEM) tools to analyse events and ensure event logs are regularly audited	Use products like Azure Security Center to prevent, detect and respond to threats

Data backup	Use on-premises tools such as Microsoft System Center Data Protection Manager (DPM) to create disk- or tape-based data backups	Use DPM for IaaS VMs or Azure Backup Services for PaaS to create online (optionally geo-replicated) backups
Scalability	Add and provision additional hardware instances (servers) in the datacentre; ensure proper operation and network connectivity	*Configure* scale up/out options to automatically respond to spikes by enabling scale, reliability and resiliency
Business continuity/disaster recovery testing	Use custom scripts to failover to alternate datacentres	Turn on tools such as Azure Site Recovery to perform script-driven orderly failover and recovery of applications and storage
Network configuration and optimisation	Use various tools to analyse and optimise network performance, discover router loops and so on.	Ensure hybrid network connections such as V-Nets and MPLS routers ("ExpressRoute") are appropriately tuned and load balanced
Identity provisioning and de-provisioning	Maintain user directory (for example, Active Directory), ensure appropriate user access to resources and enable/enforce single sign-on (SSO)	Extend directory to cloud and possibly use alternate forms of authentication for specific applications and resources

This list is neither exhaustive nor conclusive; rather, it is illustrative of the types of issues the operations staff will want to address.

The operations staff, in addition, typically maintains a Configuration Management Database (CMDB) for all of its hardware assets. There is much in the CMDB that is relevant for the cloud migration process. As we will discuss later, the CMDB can provide information such as the size of servers required for a given application, the typical number of VM instances, what storage is being used and so on. This information in combination with the portfolio management system will provide the raw data used to prioritise application migration.

Human resources and the evolution of roles

Migration to the cloud will force the roles and responsibilities of IT professionals to evolve. Much has been written about how the cloud will eliminate IT jobs. Our experience is that this is not the case; instead, IT roles change (see Figure 5-5), and become less about rote IT functions and more about high-value contributions to the business of the enterprise.

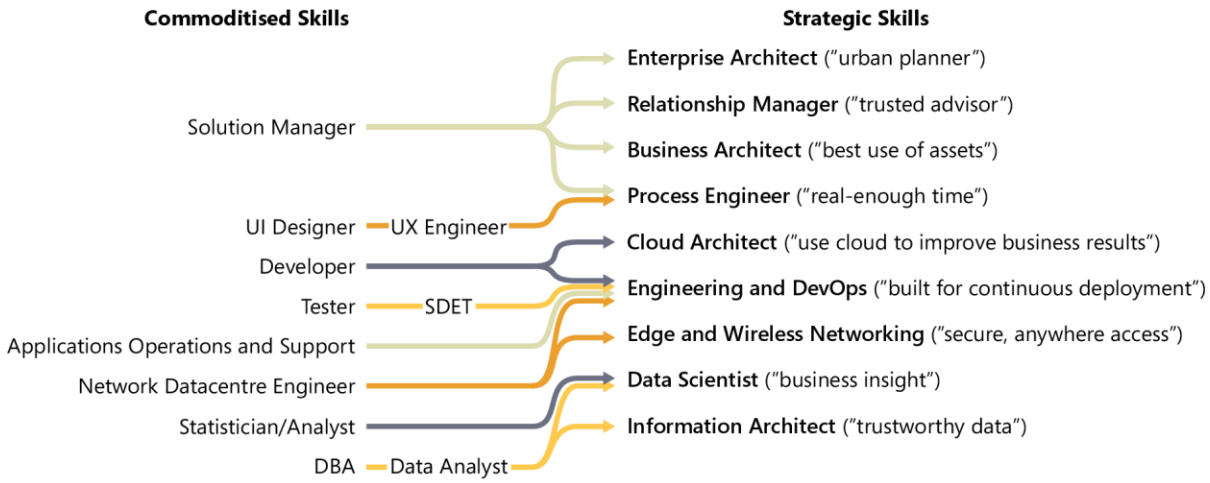


Figure 5-5: The evolution of IT roles in the cloud era

Existing IT skills will remain, but become of less value than the newer, cloud-centric skills. Enterprise architects, evolving from senior technologists, solution architects and, in some cases, relationship managers, will maintain the portfolio as a whole, understanding how to extract the most business value from large collections of applications and people. In a sense, they are the urban planners of the organisation. Business architects, using quantitative models and working closely with their partners in the actual business units, examine technical assets and business processes in various business domains and plan their evolution into the future. Process engineers optimise business processes such that they run in real time wherever appropriate and "real-enough" time (where appropriate) elsewhere. Six Sigma as well as other quality methodology skills are useful here.

With the cloud comes greater reach and with greater reach comes the essential requirement to create applications that are both productive and pleasing for the user. User interface (UI) design has evolved from simply creating menus and dialog boxes to ensuring that the entire experience of performing a task online, from end to end, is efficient and, in this era of Facebook, YouTube and Twitter, enjoyable.

Cloud architects focus on envisioning and enhancing an application or set of applications focused on a particular domain, such as finance, and work closely with their counterparts in business architecture (BA) and EA to build the optimal application in the cloud, taking advantage of its capabilities. The solution architects provide oversight and direction to the development of new features and capabilities within the applications in their space. They typically are very technical individuals.

Perhaps one of the most interesting and talked about evolutions in cloud migration is the merging of two communities, development and operations, that were previously separate. This is now called the *DevOps* movement. As applications move to the cloud and the ability to deploy applications quickly and repeatedly (sometimes adding new features each week, or even more often, using agile methodologies) is recognised, the traditional boundaries between developers, testers and operations staff begin to blur. Developers will test their applications in staging areas in the cloud. Testers will necessarily be as conversant in cloud technologies as others and often write cloud-based automation scripts or applications in the cloud, making them cloud developers, as well. And, operations personnel will less and less manage hardware assets such as servers and networks, and more and more handle creating automated configuration, deployment scripts, insight portals, monitoring scripts and orchestration flows, or using those provided by the cloud or tools vendor. (We cover DevOps in more detail in Chapter 7.)

Lastly, the information architect will ensure both the consistency of data models across the enterprise and their lifecycle. A well-designed, documented and maintained set of models–for example, for "customer" and "product" data entities–ensure ease of system integration and consistency of reporting, among other benefits.

The human resources team should work with the relevant leaders to build readiness and training plans for the affected individuals. Nearly all roles in IT will evolve. Many will require specialised training; for example, in new tools or new processes.

Skills development

Think about building cloud readiness in phases, with increasing levels of education and rigour. One framework that you can use encompasses three phases of training, which you can see in Figure 5-6: beginning with educating individuals and then building a cloud practice, i.e., teams of educated engineers, and then using them to educate the entire organisation.

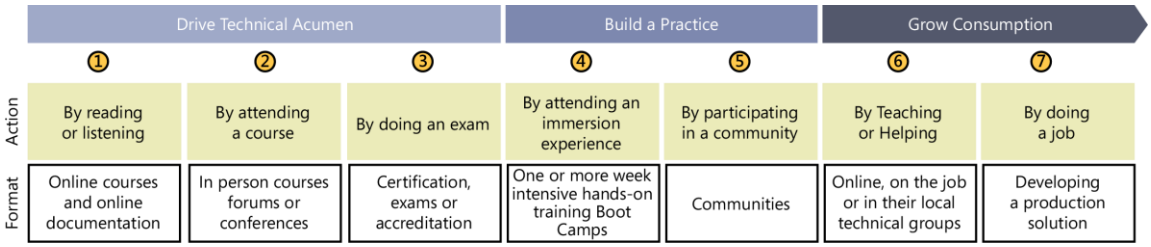


Figure 5-6: Building skills

In the first phase (Figure 5-7), drive technical acumen for individuals by having them attend classes either online or in person.

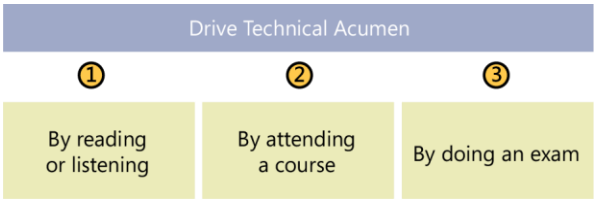


Figure 5-7: Driving technical acumen

Numerous free or inexpensive online courses for Azure exist, both on the Azure website (https://azure.microsoft.com/en-us/community/training/) and through the Microsoft Virtual Academy. You can also find a comprehensive set of training resources available for Microsoft partners in the Azure Skills Initiative (https://blogs.partner.microsoft.com/mpn/new-cloud-trainings-for-next-generation-tech-professionals/), and many other organisations provide similar types of training.

These courses help individuals learn the basic skills for the cloud. Other forms of training, such as those shown in Figure 5-8, help your teams develop common sets of skills and the beginnings of your cloud practice.

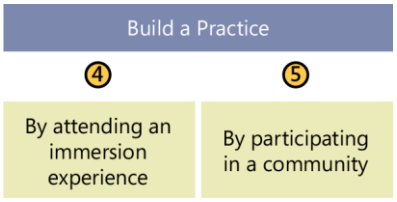


Figure 5-8: Building your cloud practice

Immersion experiences such as boot camps bring teams together for a full-day immersion in the cloud as it applies to their organisation, and participating in any of the large numbers of communities can further deepen your team's skills.

Of course, there's nothing like doing, as they say, either by leading or by helping a team build and deploy cloud applications (Figure 5-9), or by simply being in a cloud-centric role.

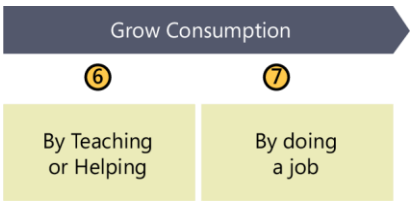


Figure 5-9 Growing the capability of your team

What kinds of training should you have your teams participate in? Here is a sample catalogue of courses provided by Microsoft; this list, incidentally, will help you build the specific cloud roles we mentioned in Figure 5-10.

Cloud Administration	Data & Analytics*	Cloud Development*	DevOps
• Microsoft Azure Fundamentals • Microsoft Azure for AWS Experts • Microsoft Azure Virtual Machines • Microsoft Azure Virtual Networks • Microsoft Azure Identity • Microsoft Azure Storage • Microsoft Azure Security and Compliance • Managing Azure Workloads • Microsoft Azure App Services • Databases in Azure • Migrating Workloads to Azure • Application Deployment and Management	• Processing Big Data with Hadoop in Azure HDInsight • Implementing Real-Time Analysis with Hadoop in Azure HDInsight • Implementing Predictive Solutions with Spark in Azure HDInsight • Processing Big Data with Azure Data Lake Analytics • Processing Real-Time Data Streams in Azure • Orchestrating Big Data with Azure Data Factory • Delivering a Data Warehouse in the Cloud • Developing NoSQL Solutions in Azure • Developing Big Data Solutions with Azure Machine Learning • Provisioning Databases in Azure and SQL Server • Recovering Data in Azure and SQL Server • Securing Data in Azure and SQL Server • Managing Organisational Data Sources with Azure Data Catalogue	• Developing IoT Solutions with Azure IoT Hub • Azure App Configuration • Creating an Angular Web App on Azure • Developing Azure App Service Components	• DevOps on Azure Paas • Continuous Integration and Continuous Deployment • DevOps Testing • Infrastructure as Code* • Configuration* • Mobile DevOps* • Application Monitoring and Feedback Loops*

*Course(s) coming soon

Figure 5-10 Types of training available for Azure

By taking a thoughtful approach to skills development, you not only can provide the requisite training to individuals in your organisation, you also can improve the overall effectiveness of entire teams.

Applications teams

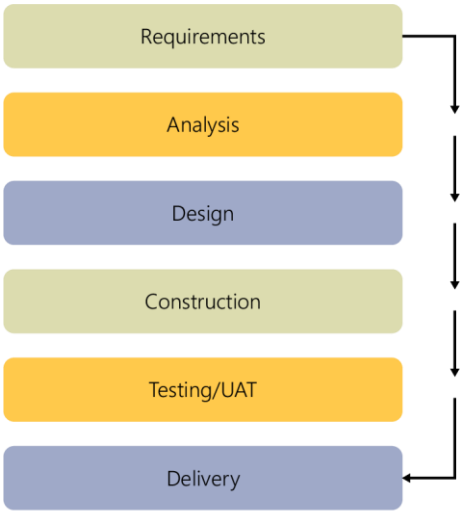
Applications teams must be consulted on a range of topics. Typically, it is these teams that will provide the required information for the application catalogue or portfolio management system (discussed later) that will help prioritise application migration.

In addition, discuss with them the technical implications of running their applications in the cloud. If an application is "chatty" in the datacentre (meaning it sends and receives a lot of messages to accomplish a task), it's possible that the latency inherent in moving to an off-premises cloud datacentre will amplify delays. To ameliorate this, application teams might either want to update

the application or recommend using a high-speed dedicated line to provide additional bandwidth. If you're using a cloud database, it might impose certain size restrictions, but this can be addressed by using specific approaches such as database *sharding* (a database shard is a partition of data in a database; each shard is commonly hosted on a separate database server instance).

Applications teams should know the longer-term possibilities of a cloud-centric application. For example, redesigning an application to be PaaS or to be a collection of *microservices* (discussed in more detail in Chapter 11) will require awareness and training.

From a methodology perspective, applications teams should consider if using a traditional *waterfall* approach (as shown in Figure 5-11) is appropriate, or if an *agile* methodology, incorporating many short development sprints with feedback and potential course correction can be used. For certain types of applications (e.g. financial accounting, for which strict regulatory requirements can essentially dictate the functional specification) you might use a waterfall approach. Waterfall-based projects usually include a detailed, comprehensive requirements document that project managers can validate.



*Linear Approach

Figure 5-11: Traditional waterfall approach to software development

However, fewer applications today require this amount of rigour and most actually benefit from short amounts of development followed by user testing and feedback, which is a hallmark of Agile development (see Figure 5-12). In this way, users can get a sense of the application, request new features, suggest others be removed and so on. In many cases, the Agile methodology leads to a solution that meets users' needs far better than waterfall's linear approach.

Having this discussion is important because the cloud accommodates much faster development/ deployment cycles and thus lends itself very well to Agile.

Figure 5-12: Cloud software development

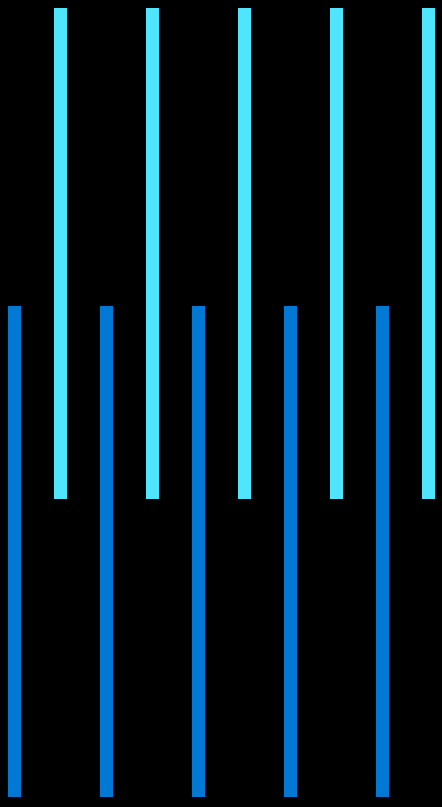
Business units

Business units should also be consulted. Some will embrace and champion the change; others might initially resist because such a change introduces risk, as we discussed earlier. Providing an understanding in nontechnical terms of how you will address those risks will go a long way toward easing their fears. Of course, describing the expected benefits in terms of cost savings, increased reach and quicker deployment times will, hopefully, whet their appetites. Partnering with your champions early and building real applications that demonstrate the benefits will sway the others.

Take note of their business calendars. Often business leaders have important times of the year when their systems must be available, such as at critical financial reporting periods or, for e-commerce functions, around holidays such as so-called “Black Friday” in the United States. You can then plan around these times.

Chapter 6

Portfolio analysis



How exactly do we prioritise the migration of applications to the cloud? To understand what applications you should move, when and how, it's important to create a well-attributed catalogue of applications managed by IT. Then, the relative importance of each attribute (say, business criticality or amount of system integration) can be weighted and you can build the prioritised list.

Building the catalogue

There might be many attributes ranging from document classification types to server counts to protocols, and so forth. It is often useful to roll these up into management sets of overall attributes, such as that shown in Figure 6-1. As the figure depicts, the top-level criteria include performance, architecture, financial, risk, operations, and security and compliance.

Many enterprises already have a portfolio management system in which such a list is maintained, and usually this can be used or extended for cloud purposes. Others might need to use an ad hoc tool such as a spreadsheet. Either can be effective.

It can be useful to think about application characteristics, or attributes, from two perspectives: the business ("top-down") and technical ("bottom-up") models. This is because the data comes from different constituencies. The top-down approach asks where each application or workload *should* go; the bottom-up approach describes where each *can* go. The sections that follow explain each along with the attributes they capture.

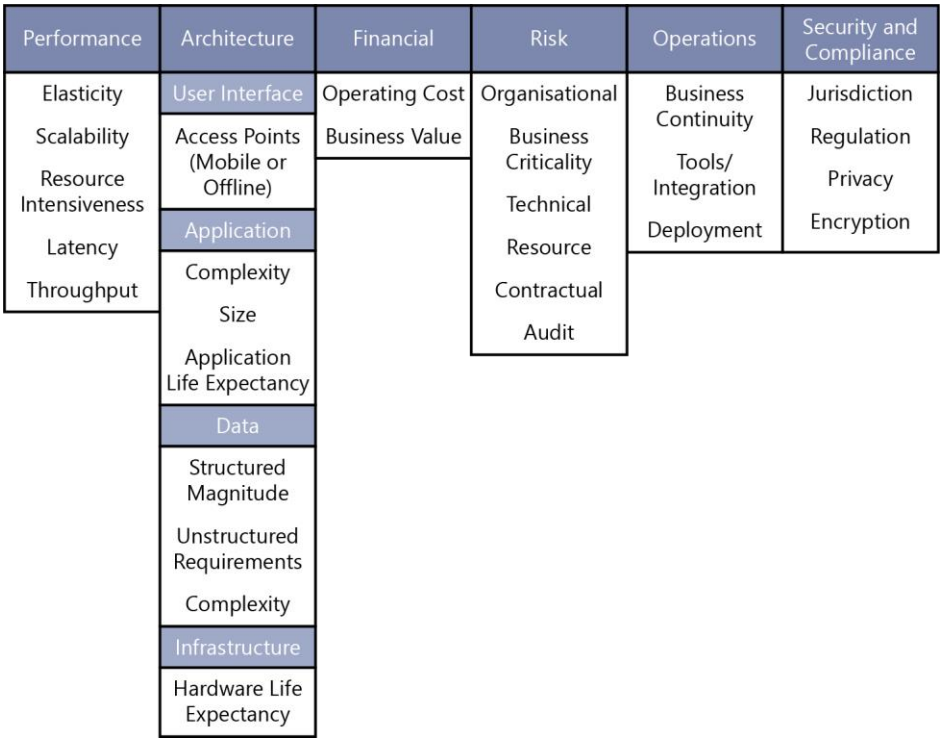


Figure 6-1: Evaluation criteria

Top-down portfolio analysis

So far, we have discussed the migration process as a systematic approach, examining objective and subjective metadata to determine where applications or workloads should go. This is a top-down assessment method, which provides a strategic approach, driven by planning and your detailed analysis and modernisation needs.

Figure 6-2 demonstrates how the top-down assessment first evaluates the security aspects previously mentioned, such as the categorisation of data (high, medium or low business impact), compliance, sovereignty and security risk requirements. Then, it assesses the current complexity interface, authentication, data structure, latency requirements and coupling and application life expectancy of the architecture. Next, top-down assessment measures the operational requirements of the application, such as service levels, integration, maintenance windows, monitoring and insight among others. When all of those aspects have been analysed and taken into consideration, the result is a score that reflects the relative difficulty of migrating this application to each of the cloud platforms – infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).

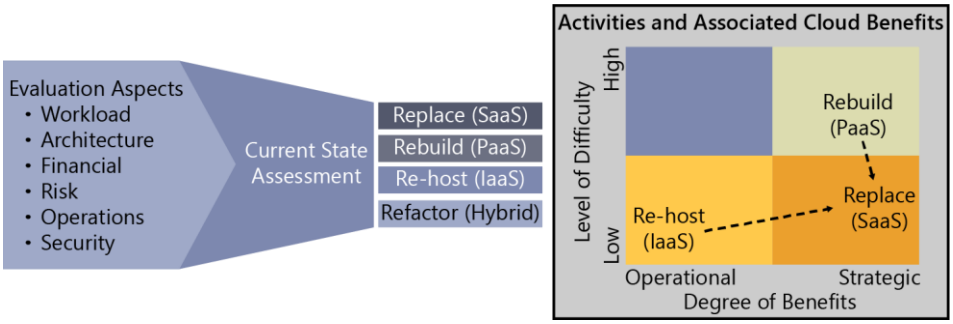


Figure 6-2: Top-down current state application assessment

Second, the top-down assessment evaluates the financial benefits of the application such as operational efficiencies, total cost of ownership (TCO), return on investment or any other appropriate financial metrics. In addition, the assessment also examines the seasonality of the application (are there times of the year when demand spikes) and overall compute load. Also, it looks at the types of users it supports (casual/expert, always logged on/occasionally logged on, etc.) as well as the consequent required scalability and elasticity. Finally, the assessment concludes by examining the business continuity and resiliency requirements that the application might have as well as dependencies to run the application if a disruption of service occurs.

The two parts of the process result in an application valuation score that reflects the balance resulting from the difficulty to migrate to each platform versus the potential benefit gained by it. You can see the entire process in Figure 6-3.

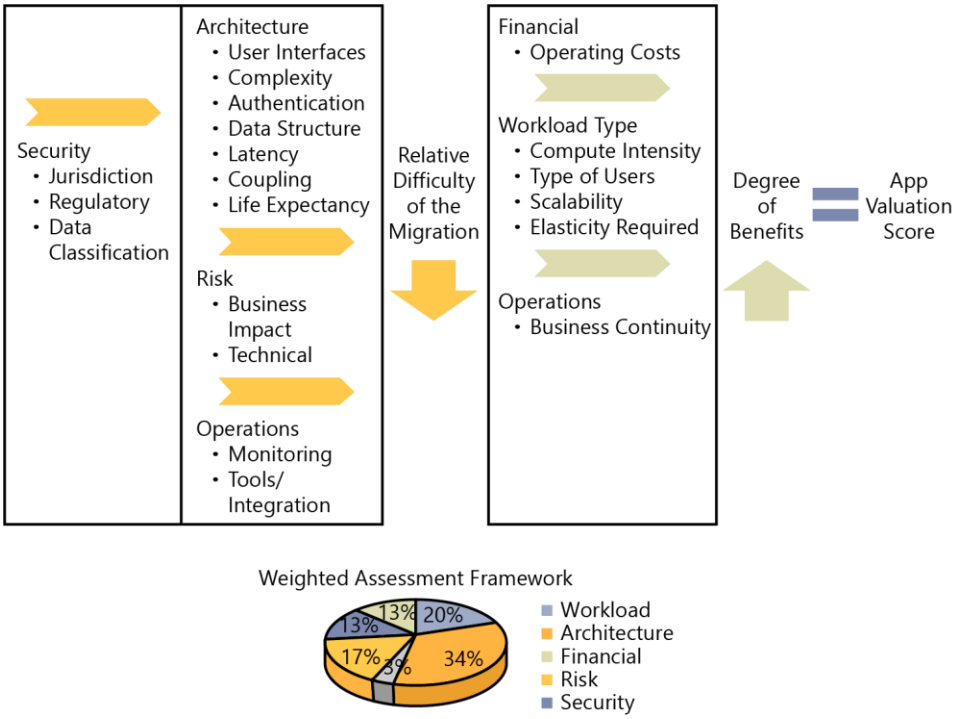


Figure 6-3: Top-down assessment process

With the results from the top-down assessment method, you can then map out which applications have the highest potential value and are better suited for migration, and start there. You might even be able to combine that list with the quick wins which are based on lower potential value applications that are also better suited for migration. After your organisation has gained the appropriate experience, built the right set of tools and processes, and gained confidence in its methods, it is then time to move to the applications that are more difficult to move but have a high potential value, leaving those that are more difficult to move but have a low potential value until last. We can visualise this more easily in Figure 6-4.

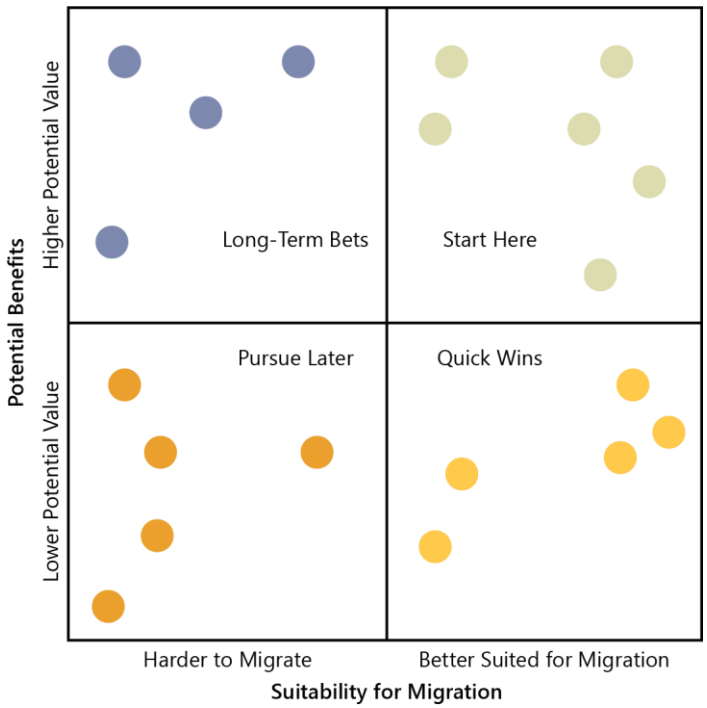


Figure 6-4: Application migration suitability versus potential benefit

Bottom-up portfolio analysis

There is a complementary approach that is more tactical and technical and is focused more on requirements. As we mentioned, the top-down approach analyses where an application *should* go; here we are asking where a particular workload *can* go, based on its purely technical requirements.

The bottom-up approach (Figure 6-5) occurs simultaneously with your top-down planning and is aimed at providing a view into the eligibility of an application to migrate at a technical level. We can typically pull much of this information from a Configuration Management Database (CMDB). Enterprises use this method to provide additional insight to the top-down approach.

The type of requirements evaluated by the bottom-up assessment cover the application or service required: maximum memory, maximum number of processors (CPU cores), maximum operating system storage space, maximum data drives, network interface cards (NICs), IPv6, network load balancing, clustering, version of the operating system, version of the database (if required), domains supported and third-party components or software packages, among others.

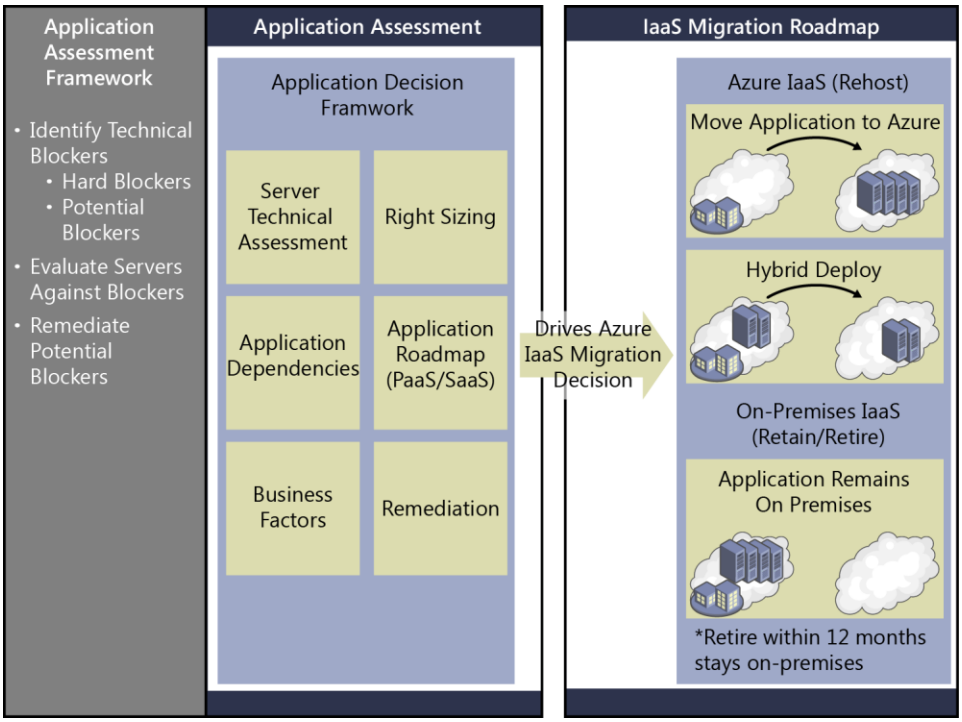


Figure 6-5: Example of bottom-up assessment for IaaS

Each day, it seems, cloud platforms are becoming more and more capable of handling different application profiles. Servers equipped with 16 and 32 cores have become commonplace, and massive amounts of memory and storage are available. Still, you might find applications that for one technical reason or another you cannot move at this time[8] or should wait until cloud capabilities are further extended.

As part of the bottom-up planning, catalogue the technical aspects of your application, including its operating system type, version, number of processors required, memory required, disk space and number of drives needed and so on. Knowing the size of an application's database, and its data types, will help inform a decision as to whether to use, for example, a cloud-centric relational database such as Microsoft Azure SQL Database, Microsoft SQL Server or Oracle in a virtual machine (VM), or perhaps even a NoSQL database.

Your existing integration systems will be affected by cloud migrations, at least temporarily, so you should spend time documenting the potential impact on these systems. You will need a clear understanding of which applications connect to which: Is there an order of precedence for integration operations? How much data is moved and how frequently? What is the architecture of your Extract, Transform and Load (ETL) tools? The complexity of your integration operations should be an important factor in prioritising cloud migration goals.

Moreover, many cloud vendors now implement options for business continuity and disaster recovery, including failover to alternate datacentres, redundant data storage and online backup. Understand their offerings and their capabilities and how they map to the needs of your applications.

[8] For example, applications that depend on deprecated features in operating systems, run on operating systems not supported in the cloud, or rely upon poor programming practices such as hardcoded IP addresses.

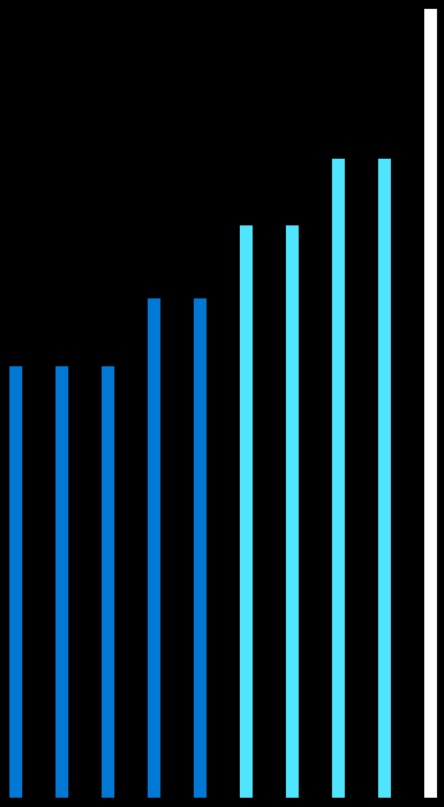
A large number of tools that can help you assess the current state of your applications exist. As mentioned, a portfolio management system will be of great utility in listing and attributing your applications. The Microsoft System Center suite includes a CMDB as well as a host of monitoring and health management capabilities that will help you to ascertain the state of your systems.[9] The Microsoft Assessment and Planning Toolkit is a solution accelerator that provides a large feature set for assessing existing IT environments. MAP facilitates the automated inventory and assessment of applications to determine basic suitability and VM sizing requirements.[10] You can also use a number of third-party applications and utilities to perform similar functions.

The type of information you collect with these instruments will address objective factors concerning hardware/VM eligibility and application/workload eligibility for cloud migration. These tools are useful for automated metadata collection about your applications and operating system instances.

[9] http://www.microsoft.com/en-gb/server-cloud/products/system-center-2012-r2/
[10] http://technet.microsoft.com/en-gb/solutionaccelerators/gg581074

Chapter 7

Building and executing the plan



With every great endeavour, there must be a plan. With a well-understood portfolio, prioritised according to criteria, it's now time to finally begin the actual migration of applications to the cloud. In this chapter, we look at building the actual plan – and then making it happen.

Consider beginning with dev/test

Before you consider the problem of moving production applications, it's worth considering moving your dev/test environments first. There are a number of good reasons to begin here:

- **Dev/test environments are expensive** You can gain cost savings – sometimes considerable – by moving them. Often, there are three separate environments per application, completely distinct from production: the development environment; the test environment, used by QA to run unit and automated tests; and the user acceptance test (UAT) environment. That can be a lot of servers and virtual machines (VMs) that are no longer needed in the datacentre.
- **It's a great way for teams to get familiar with migration tools** Teams can learn migration tools without impacting the production environment, so, when it is time to move production, that migration benefits from the experience gained with dev/test.
- **Developers are generally more tolerant of problems than users** In other words, if something goes wrong, production users are not affected, and developers and IT staff can learn from the mistakes.

- **Order matters less** As we will discuss, moving production applications should follow the prioritisation guidelines we created in Chapter 6. This is less important with dev/test applications, with the caveat that the compliance rules and regulations around protected data such as Personally Identifiable Information (PII) apply whether in dev/test or production (which is why anonymised data is often used in development).

The cloud migration plan

Returning to production applications, your mapping exercise leading from the current state to the desired state is the root of your cloud migration plan. The migration plan takes the map and adds specifics such as priorities and sequencing.

You should set priorities within your plan, based upon a combination of business factors, hardware/software factors, and other technical factors. Your business liaison team should work with the operations team and the business units involved to help establish a priority listing that is widely agreed upon. Figure 7-1 illustrates principles that you might use in establishing prioritisation guidelines.

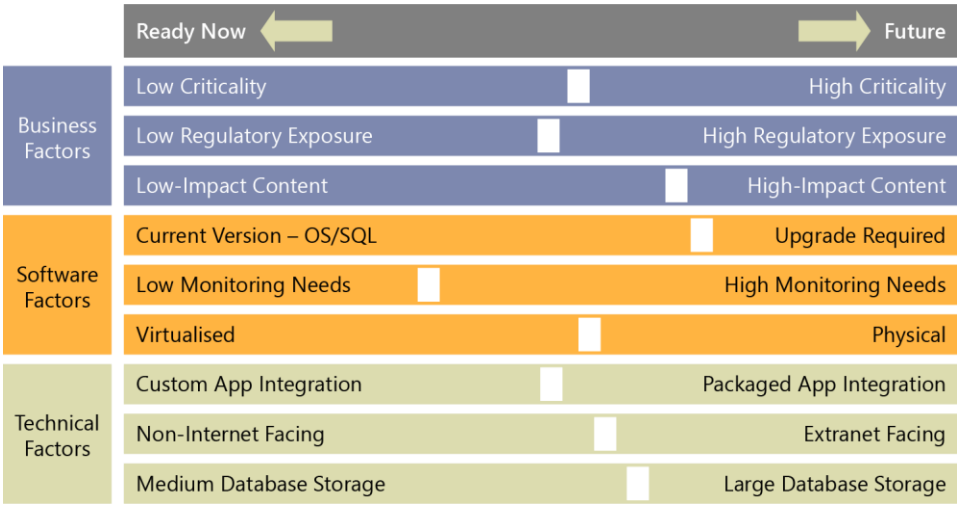


Figure 7-1: Migration priority strategy example

For sequencing the migration of your workloads, you should begin with less-complex projects and gradually increase the complexity after the less-complex projects have been moved. As with running a pilot project, you will gain valuable experience while moving applications with lower complexity and lower business risk, which can help prepare you for the more complex and more business-critical migrations.

Your cloud migration plan will be more of a process than a static plan document. In its essentials, your plan will actually be a compilation of a number of smaller plans that deal with the migration of each departmental workload, based upon the sequence you establish. The particulars of each migration will generally follow this pattern:

1. **Analysis** This process will help you to identify the gaps between what you currently have and what it will take to move that workload to the cloud. Those gaps might involve changes to the architecture of the workload or might require a complete rewrite of the program. (See the section "Evolution of the five Rs of modernisation" in Chapter 3) Additionally, many legacy programs will require significant work to improve their performance and make them more scalable, and you should identify this work during your analysis of the workload.

As part of the analysis, you'll want to fully understand the security and compliance implications of the application in question, making sure to call out any PII and other sensitive information. Where this exists, you should work with your information security and risk management teams to ensure the proper controls are in place (we cover this in more detail in Chapter 9).

2. **Application migration** When you determine that a particular workload should be moved to the cloud*, it is a best practice to create a version of the workload with a minimal amount of data in order to get the application working on the cloud or to build a new version of the application there*. If the application is already running on a VM, it might be possible to simply move the VM to the cloud without further changes. In general, many on-premises applications can run on Microsoft Azure with minimal or no changes, but this does not mean that the application will be optimised for performance, scalability and security. So, you might need to redesign and rebuild the application, to some degree, by using modern service-orientated principles.

3. **Networking** There are many ways, as we discuss in detail in Chapter 10, of connecting on-premises applications securely to the cloud. These range from a simple Virtual Private Network (VPN) to a dedicated line (i.e. Azure ExpressRoute), with different performance and price points.

4. **Data migration** This is somewhat similar to the application migration in that you can move the data structure as-is to either a relational (Azure SQL Database, SQL Server in Azure VM) or non-relational (blob, table, queue, Azure CosmosDB, and so on) location on the cloud. Several of these kinds of migration are extremely easy and you can conduct them with the help of a wizard such as the SQL Server Azure Migration Wizard.

   However, you might want to consider rebuilding the data model as a new Azure SQL Database to gain performance, scalability, resiliency and security improvements. If you need to synchronise data between on-premises and SQL Database or between different SQL Database servers, set up and configure the SQL Data Sync service. In addition, *it is a best practice that you set up and configure a data recovery plan in case of user errors or natural disasters*.

5. **Optimisation and testing** After you move your application and data to Azure, you need to perform functional and performance tests. At this phase, test your application in the cloud and confirm that it works as expected. Then, compare performance results between on-premises and Azure. After that, resolve any feature, functionality, performance or scalability issues in your cloud application.

6. **Operation and management** After the testing and optimisation phase, set up and implement application monitoring and tracing with Azure Application Insights, which enables you to collect and analyse telemetry from your application. You can use this data for debugging and troubleshooting, measuring performance, monitoring resource usage, traffic analysis and capacity planning and auditing.

   You can use the Microsoft Operations Management Suite to manage applications running both on-premises and in the cloud. Operations Management Suite provides a single view of all your applications, regardless of where they are hosted.

These six phases of migration will be conducted for each workload that you want to move. However, there is also an iterative process that is greater than any one migration, by which you can begin moving applications that meet your initial minimum standards, based on priority and sequence. When the initial group is moved, you can begin to work on making more applications and hardware eligible by upgrading operating system/SQL versions, getting current with all security patches, moving applications from physical machines to VMs, addressing issues caused by multiple IP addresses and so on.

Tools

You can use a number of readily available tools from assorted vendors to move VMs to IaaS. Azure's native Business Continuity and Disaster Recovery (BC/DR) service, Azure Site Recovery, includes a suite of tools for moving VMs (including VMware VMs) from an on-premises datacentre to the cloud, and from one cloud region to another. For websites, Azure's Websites Migration Assistant can quickly move an Internet Information Server (IIS) site to the cloud. The Data Migration Assistant (Figure 7-2) can help plan the migration of an on-premises SQL Server database to Azure SQL Database.

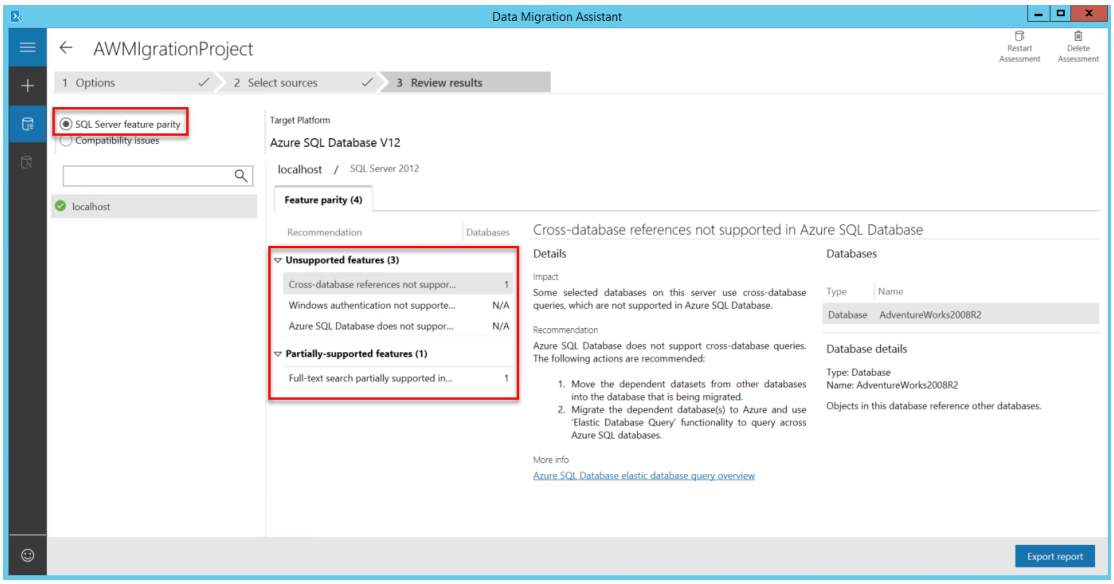


Figure 7-2: Azure Data Migration Assistant screenshot

Subscription management

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 of your enterprise is on one cloud subscription, but that model results in substantial inefficiency: it is difficult to account for by application and/or by cost centre or department in a single-subscription model. In addition, managing a large number of applications – some in production, some in testing 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 centres or even to applications or application groups (e.g. 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 can then manage independently.

In a large organisation, for better visibility and accountability, you might want to set up a cloud governance hierarchy, such as the one illustrated in Figure 7-3.

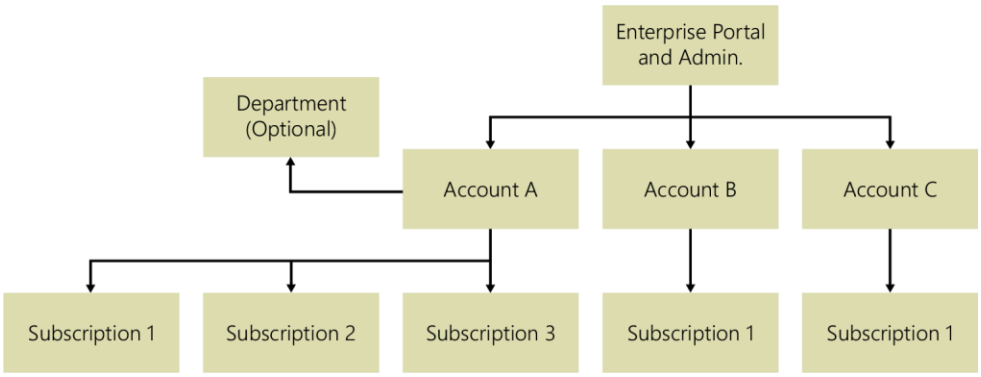


Figure 7-3: Subscription management governance hierarchy

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 centres or for individual solution areas.

Microsoft Power BI gives you a convenient way to visualise your subscriptions and their usage, as depicted in Figure 7-4.

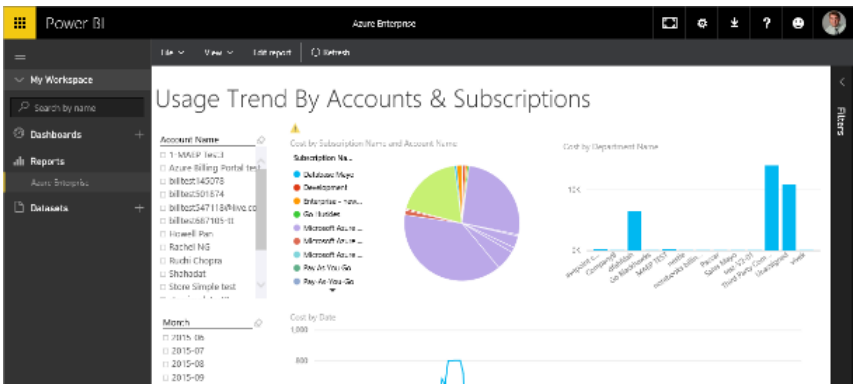


Figure 7-4: Azure usage and cost reporting, as viewed in Power BI

Microsoft IT's experience

When Microsoft IT began its cloud migration journey in 2009, it followed a similar process. First, it catalogued its operating system instances and application workloads. This assessment included both quantitative data that was mostly retrievable by tools as well as qualitative data that was partially retrievable by tools, and also required examination by both the operations team and the business liaison team. This latter category of metadata included relationships, dependencies and integration points.

Microsoft IT then identified the initial prioritised eligible operating system instances and initial prioritised eligible workload/applications. These initial migration candidates were then reduced by removing any business-critical systems, which would be moved after more experience had been gained. Next, this initial list of candidates was prioritised and sequenced with less-complex applications placed before more-complex applications, and applications running on updated VMs prioritised over those running on physical machines or legacy VMs. Some applications were identified as ineligible for various reasons (most of which no longer apply in 2017) and these were moved to an optimised on-premises datacentre.

After the initial set of migrations was completed, Microsoft IT completed work to make less-eligible operating system instances (OSIs) and workloads more eligible. For example, OSIs with older operating systems or database versions were updated, more applications on physical machines were moved to VMs and more mission-critical applications were deemed eligible. Applications and workloads that were identified as requiring a major overhaul were rebuilt as services on Azure.

Here's the process, which you can see illustrated in Figure 7-5:

1. Identify eligible hardware (OSIs) per Azure compute, storage and RAM limits.
2. Identify eligible applications, remove applications with sensitive data for the time being, sequence critical and complex apps for later, and right-size to include more apps.
3. Increase eligible hardware and applications by doing the following:
    - Virtualising more servers
    - Expanding to more regions
    - Including externally-facing apps
    - Including applications with sensitive data ("high business impact," or HBI as shown in the diagram)
    - Getting current (OS, SQL)
    - Increasing Azure VM limits
4. Build new applications as services for SaaS IT.

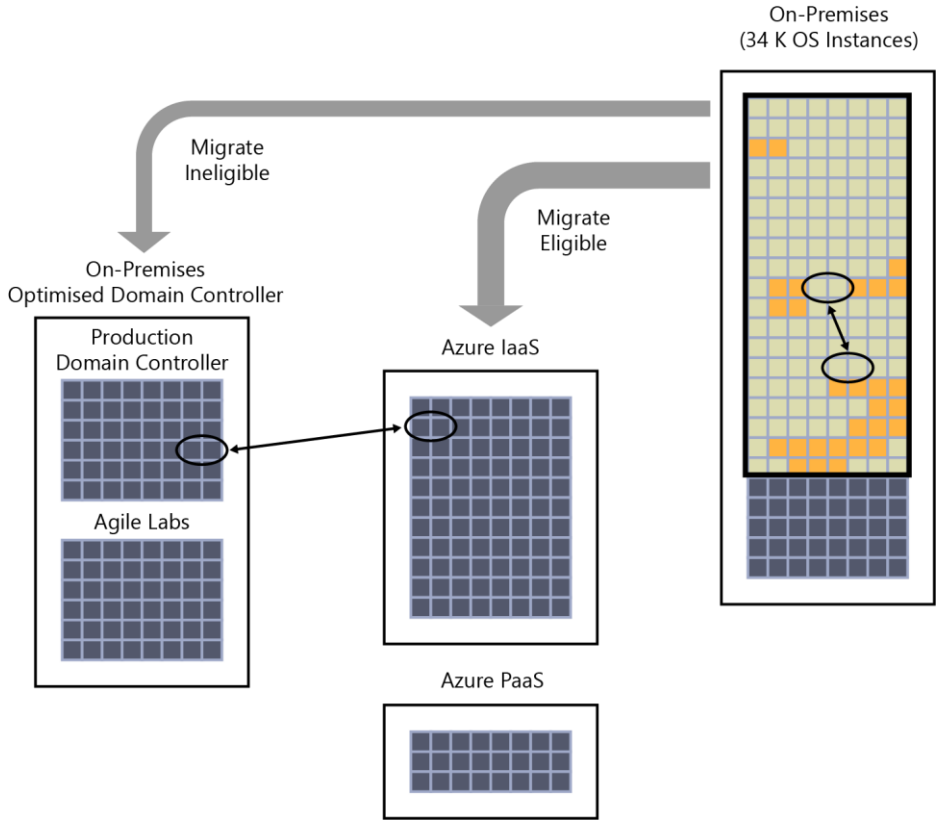
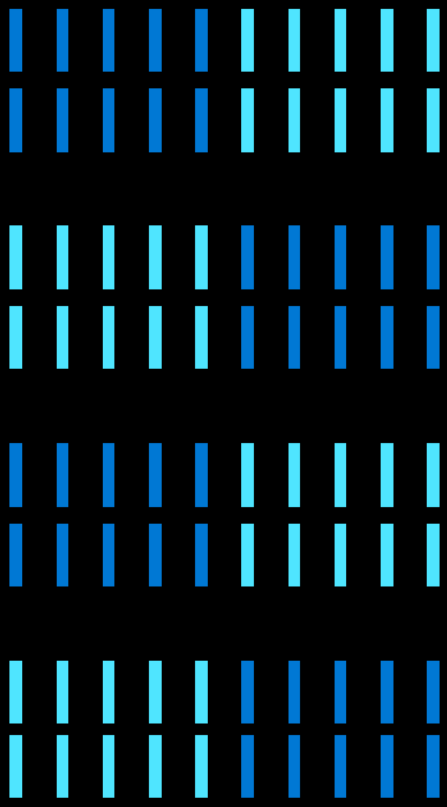


Figure 7-5: Implementing the pla

Chapter 8

DevOps makes teams more productive



Developing, configuring, deploying, managing and updating applications in the cloud creates many new opportunities to make teams more productive and to reduce costs. Previously separate teams, IT software developers and operations staff are coming together to make the processes of putting applications in the cloud seamless, fast and efficient. In this chapter, we look at how DevOps teams can "spin-up" cloud resources for development and testing and then release them when they're no longer needed. We also talk about how DevOps treats "infrastructure as code" to make deployments and updates fast and safe.

Using the cloud to do development and testing

Not so very long ago, one of the biggest inhibitors to IT productivity was testing. Before going live with an application, it needs to be tested, first by the developers and testers under very controlled conditions, and then live with real users in a phase usually called User Acceptance Testing (UAT). Compliance testing would normally be performed at around the same time.

There were many issues with this model of development and testing. Bugs often appeared in production that had never been seen before in development or testing. Why? Because as often as not the production configuration varied in some way from the dev/test environment. Perhaps the hardware configuration was different. Perhaps "real" data was qualitatively different in some way from test data. Perhaps the load was unexpectedly high. The net result was that in some way the dev/test environment was different from the real-world production case, and, unsurprisingly in hindsight, bad things happened.

When all development was done on-premises, creating an environment that somehow matched the production mode could be very difficult. Procuring sufficient hardware, for example, was a costly capital expense. Simulating user load could yield inaccurate results if hardware environments or software configurations differed; and so on.

Using the cloud for creating dev/test environments presents a number of opportunities; of course, the essence of cloud computing is that capacity can be checked out, used and then returned when it's no longer needed – precisely the model of development and testing.

With Microsoft Azure DevTest Labs (Figure 8-1), it becomes possible to perform development and testing in the cloud in a self-service, controlled fashion. Using Dev Test Labs, you can allocate servers to do development. A separate set of servers can be spun-up – under configuration control – at a certain hour of the day (e.g. at night) to run tests, and then deallocated as the tests are completed 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 expiry date) for the lab.

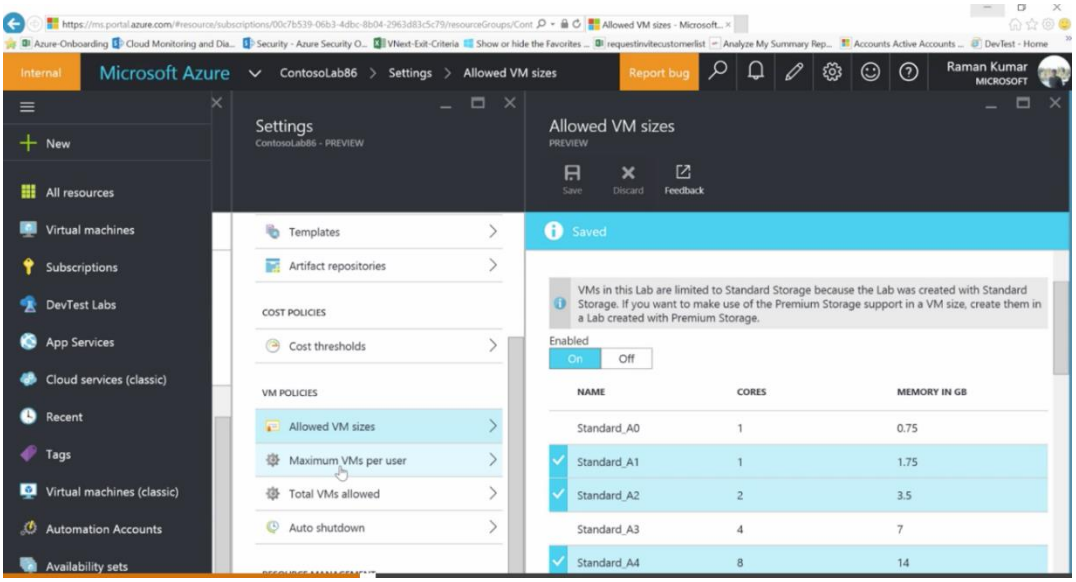


Figure 8-1: Setting allowed VM sizes for DevTest Lab

To ensure that the right environment is tested each time, you can create virtual machine images in advance that contain the required software, reducing or eliminating common errors that occur from having different environments or configurations. In addition, you can include commonly used tools such as Telerik's Fiddler (network traffic analysis tool), PowerShell scripts, logging functions or other tools in the test environment. These can be very helpful in diagnosing problems as they occur. Finally, the configuration of the environment can be updated easily so that you can test differing configurations quickly.

The DevOps revolution

For decades, enterprise computing focused on cost control and risk management. Every piece of hardware and software was carefully accounted for, and significant purchases were made only after many layers of review by various groups across the enterprise, all to ensure money was being spent wisely.

The development process was equally rigid. "Account managers" worked with their business partners to create requirements documents. A functional specification was written and presented to the development team, which responded with a design or technical specification. When the documentation was signed off after multiple reviews, development began. At various milestones code was handed over to the test team, which ran feature tests and filed bug reports. Eventually the code was put into UAT, and after that was signed off – and only after that – the software was put into production.

Although it was easy to understand and straightforward to track in a project-management application, this "waterfall" approach to software development was, obviously, laden with waste and inefficiencies. New releases often took months or years to develop and deploy. Applications failed on deployment because of configuration differences with test environments, or because the amount and the nature of traffic in production was not what was expected. Because of the large number of groups involved – development, testing, operations – resolving critical failures took unacceptably long periods of time. Many IT departments had to deal with the criticism that they were "too expensive and too slow."

Moreover, as Internet-style computing became more the norm, the focus on cost and risk control began to give way to a new emphasis on *speed*, i.e. the demand for (much) more frequent releases. E- commerce websites, for example, constantly promoting new products and creating new incentives, found themselves needing updates many times, even hundreds of times *each day*.

Something had to give, and it did.

Continuous Integration and Continuous Deployment

Taking a cue from lean manufacturing methodologies, many in the field began to experiment with more agile development strategies, minimising the tedious documentation stages or skipping them altogether and using short coding "sprints" to implement some features and test them with real users, getting feedback and incorporating it into the next sprint. Instead of thinking of software development as a series of stages, these pioneers conceived of it more as a continuous pipeline of develop, test, deploy, repeat.

Automated testing has replaced much of the manual feature testing that used to occur (note the change in title from *Tester* to *Software Development Engineer in Test* – SDET). The act of checking in a new piece of code now kicks off a suite of automated tests to verify that nothing has broken (or if it had, the developer was quickly notified and the check-in rejected). This became known as *Continuous Integration* (CI).

Moreover, it soon became obvious that the scripts and configurations that control deployment *were themselves code* in a way and could be treated as such: with version control, bug tracking and so forth. Companies began to realise the now-famous saying that they should treat "infrastructure as code." In Azure, for example, infrastructure is described by using Azure's Resource Manager in JavaScript Object Notation (JSON). Following is a small snippet of ARM providing a name to a Linux VM:

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

Eventually, this radical change – as much cultural as technological – became known as DevOps, as the traditionally separate roles of development, testing and operations began to blur together. As new code was checked in, automated tests were run, and if these were passed, the code could be automatically deployed to production. This is known as Continuous Deployment (CD). This is how some applications can now be deployed as many as 200 times per day!

Monitoring and instrumentation

In DevOps, developers also build instrumentation into the code so that telemetry is being received and application health can be monitored at all times. Azure Application Insights provides not only the APIs, but a dashboard on which you can monitor telemetry, including alerts. This data can also be fed into a repository such as Hadoop or its Azure-managed service version, HDInsight (covered in more detail in Chapter 12), where you can analyse usage patterns and other trends. For example, Figure 8-2 presents an Application Insights analysis showing average user dwell time on a site.

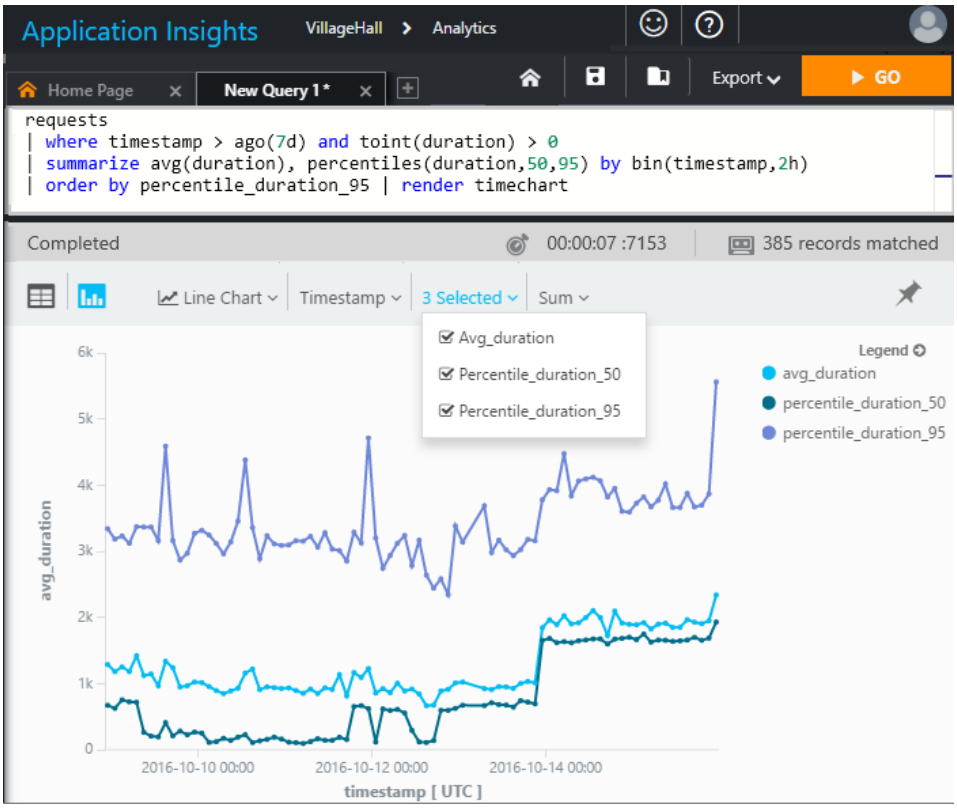


Figure 8-2: Application Insights showing average user dwell time

You can then see the new development cycle: coding, user testing, automated testing, automatic deployment, followed by runtime feedback from telemetry and tuning made to the deployment parameters where necessary.

Want to see a unified view of all the relevant metrics for your application? You can use the newly announced Azure Monitor, shown in Figure 8-3, which brings together Activity Logs (which track all operations performed on Azure resources), metrics and diagnostics logs, and provides tools with which you can configure alert rules.

Figure 8-3: Azure Monitor

Lastly, Azure Log Analytics collects and aggregates data from many different monitoring sources, though with a delay of 10 to 15 minutes. Part of the Operations Management Suite, Log Analytics provides a holistic IT management solution for Azure, on-premises and third-party cloud-based infrastructure. It provides richer tools to analyse data across more sources, facilitates complex queries across all logs, and can proactively alert on specified conditions. You can even collect custom data into its central repository so that you can query and visualise it.

Use DevOps to optimise your infrastructure

Here is an outcome of moving your applications to the cloud that you might not expect: you might 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 enterprise 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 replicated more or less exactly. That is, if you had (for example) eight servers devoted to the application in your on-premises datacentre, it's likely that in the initial move, eight infrastructure as a service (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 use.

Here is where the cloud, and DevOps, can show their value. By providing ongoing monitoring through tools such as the aforementioned Application Insights, 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 use, and that can lead to important savings.

If, for example, you see that you have eight servers all running at 6 % CPU usage, you can consolidate that load onto two servers, for instance, and return the rest to the pool – now you will only be charged for two. Applied broadly throughout your application portfolio in the cloud, you should see considerable savings.

In the example presented in Figure 8-4, a particular IaaS application running in Microsoft IT was monitored and CPU usage was measured according to the industry standard P95 algorithm. Running on a relatively large server, its monthly costs came to around $1,400.

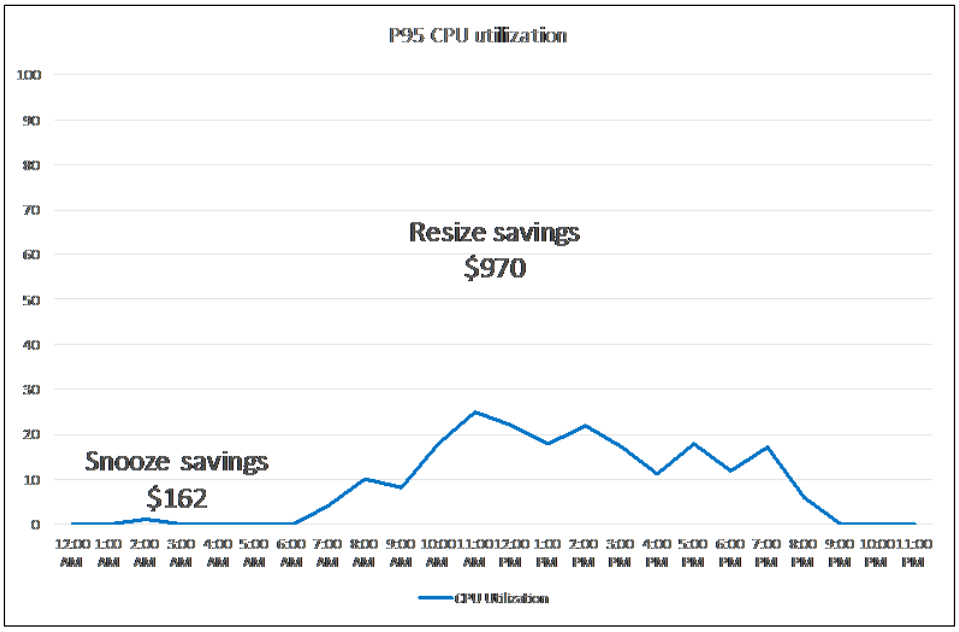


Figure 8-4: CPU usage of IaaS app

You can see that the application is active only between the hours of 6 AM and 9 PM, and that its maximum CPU usage – the most *it ever used* – was around 25 %!

Microsoft IT put the application on a "snooze" schedule to take it offline during off hours, during which 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!

Now compare these costs to running the application on-premises: in your datacentre, you are always paying for the servers; you cannot free them back to the "pool," nor is it easy to resize them. Here is a clear example of how using DevOps capabilities with the cloud can return significant savings.

Changing the conversation

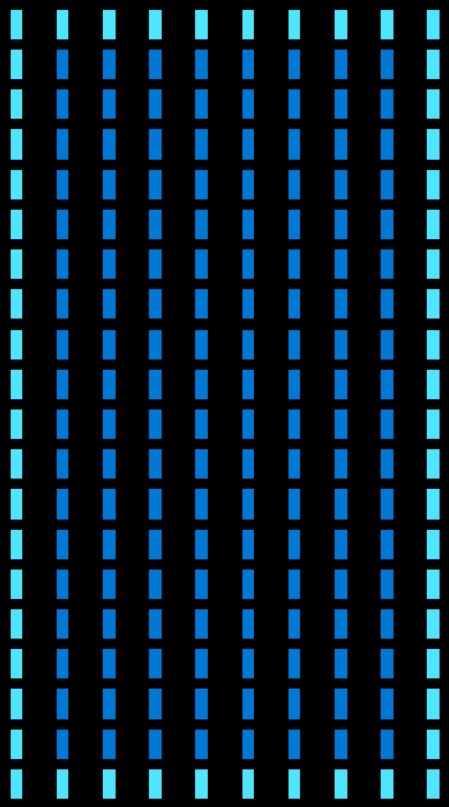
With continuous optimisation, you are always on the lookout for overprovisioned servers, for underused databases and servers, and for applications that are only used lightly or not at all in off hours. For a large IT ecosystem in the cloud, this can save literally millions of dollars.

But optimisation can also change the conversation that IT has with its business partners. One solution owner recognised that creating a particular report on worldwide sales could benefit from more CPU horsepower. This then became a business conversation: the business could have their report on an hourly basis, *if they were willing to pay for the extra servers.* But if a daily report was sufficient, fewer servers would be needed and the cost would be lower.

Moreover, if during certain periods, for instance, toward the end of a quarter, hourly reports were temporarily required (at extra cost), IT could accommodate this request. This agility is unprecedented and would have been impossible in a pre-cloud world.

Chapter 9

Cloud security and governance



One of the first questions every responsible IT executive asks is whether applications and data in the cloud are secure. The answer is a definite yes – so long as the appropriate technologies and controls are applied. Ensuring security is one area in which both technology and governance are applied.

Indeed, moving applications to the cloud does not negate many of the "traditional" roles of IT; security, ensuring that enterprise data is properly managed, that costs are controlled and that change is managed appropriately remain key areas of responsibility. But, the *ways* in which enterprises govern themselves change when in the cloud. In this chapter, we look at cloud security and governance and how IT executives should plan for them.

Cloud security

Nearly every IT executive we talk to admits to a bit of initial discomfort when thinking about moving their IT ecosystems to the cloud. After all, when all the applications and data reside in the on-premises datacentre, enterprise IT is in control. We can liken it to the concept of a bank 150 years ago. At that time, people kept their cash under their mattresses. But eventually everyone came to the realisation that their money was far safer in a bank where it could be protected by professionals.

But cloud security differs from banks in an important way: as application and data owner, you must be an active participant in your security. In the next few paragraphs we'll outline some of the areas to which you should devote attention and resources.

Physical security

Every security story begins with physical security, i.e. the safety of the physical facilities in which the cloud runs – the cloud datacentres. Cloud providers make significant investments in physical security; all of them feature 24/7 video surveillance. Employees at cloud datacentres 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.

Software Updates

Recall that if you have deployed your applications to the cloud as infrastructure as a service (IaaS) virtual machines (VMs), your staff retains responsibility for ensuring that system software updates and patches are applied in a timely manner. If you are using a platform as a service (PaaS) model, your cloud provider will maintain the system software for you.

Encryption everywhere

It is recommended that applications use encryption wherever possible. For hybrid cloud connections (connections between the on-premises datacentre and the cloud), VPNs and Microsoft Azure ExpressRoute use IPSec with Internet Key Exchange as the underlying transport.

Consider using Transport Level Security (TLS), which is the security technology behind secure HTTP (HTTPS) for client access to cloud websites.

You also should encrypt data at rest in Microsoft Azure Storage or in databases 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).

Key vaults and hardware security modules

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 places the keys into it (Figure 9-1). 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 data in Azure Storage or elsewhere.

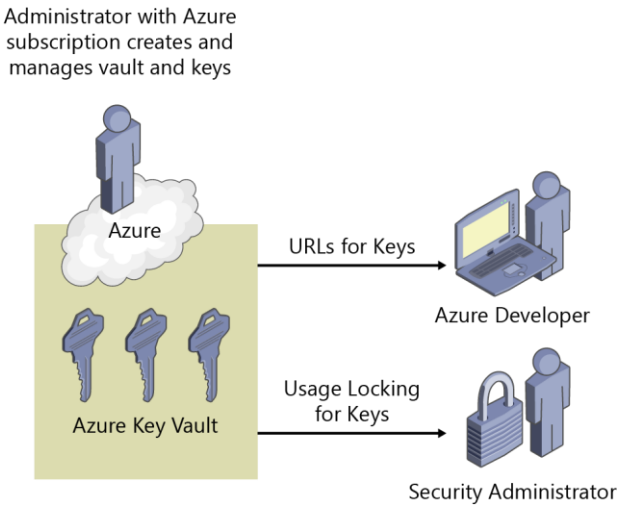


Figure 9-1: Azure Key Vault

For added protection, the keys can be stored in a Hardware Security Module (HSM), which is a physical appliance that can both store and generate keys. HSMs also can offload cryptographic processing (normally a CPU-intensive activity), performing encryption and decryption onboard.

Antivirus software

Nothing can hurt an application more, whether on-premises or in the cloud, than the discovery that it is spreading a virus – knowingly or not. Applications (particularly in IaaS) should take advantage of the antimalware software supplied either by the cloud provider (such as Microsoft Antimalware) or by a partner in the cloud marketplace. Any events that the antimalware software detects are logged. Cloud administrators should periodically examine these logs to determine if any action should be taken.

Multi-factor authentication

For additional security, consider using multi-factor authentication (MFA) when users log in. 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 user logging in might trigger a phone call to a mobile phone that has thumbprint identification capability; the user is prevented from logging 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 the number displayed by the phone in order to gain access.

Secure development life cycle

Even though the cloud provides many security advantages, hosting an application in the cloud does not entirely relieve application writers and security professionals of their responsibilities. We strongly recommend that developers and testers adhere to the Security Development Lifecycle (https://www.microsoft.com/sdl/default.aspx), which provides a set of steps for anticipating and mitigating threats. Antivirus and antimalware options should be included in your deployments.

Monitoring for security breaches

IT executives should remain vigilant for security breaches in cloud applications just as they do for on-premises applications. Fortunately, cloud providers also have trained security professionals at their disposal who monitor cloud activities 24/7.

You might want to deploy a Security Intrusion and Event Management (SIEM) application for additional security. SIEM systems scan applications for vulnerabilities, provide intrusion detection and monitor user behaviour for signs of malicious actions.

In addition, the Azure Security Centre (Figure 9-2) provides security professionals in your organisation with a wide array of capabilities, including making recommendations (such as applying patches or updating antivirus software), security alerts (such as your application communicating with known malicious IP addresses) and setting security policies for your applications.

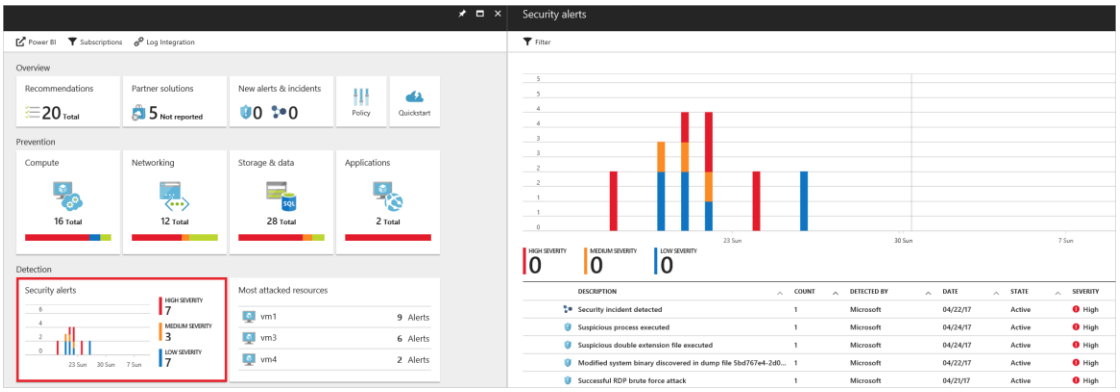


Figure 9-2: Azure Security Centre

Penetration testing

Sometimes it is only possible to find vulnerabilities by actually attempting to hack an application. Many enterprises employ teams of computer security professionals to perform so-called penetration testing, and this is a best practice.

However, you should work with your cloud provider to schedule such testing because it can be difficult for the cloud provider to distinguish between a test and a real attack without advance warning.

Understand cloud security controls

Figure 9-3 delineates the distribution of security responsibilities by application model – on-premises, IaaS, PaaS and software as a service (SaaS).

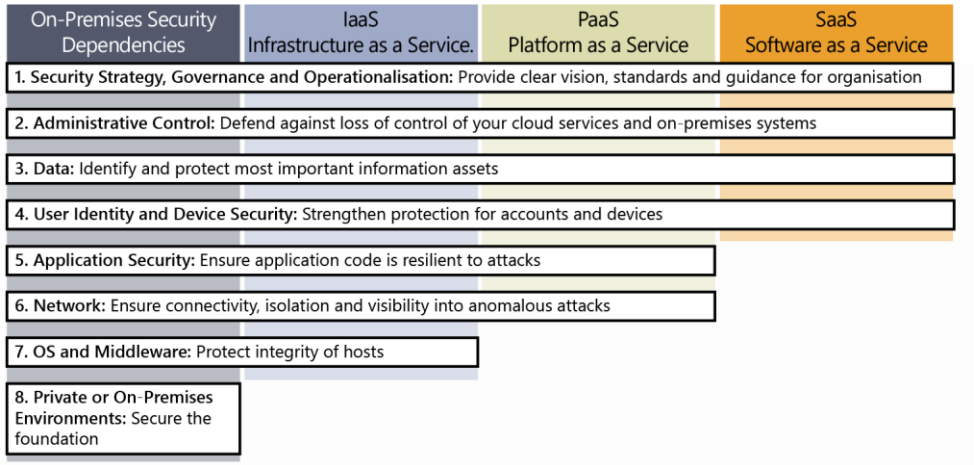


Figure 9-3: Understanding responsibility for security controls

Governance, compliance and risk

Organisations have been managing governance, compliance and risk management since the dawn of business. Every organisation has its own approach to the aspects of governance, risk management and compliance, from the ad hoc and disorganised to the mature and aligned.

Governance, risk management and compliance (GRC) are three facets that help to ensure that an organisation meets its objectives, as illustrated in Figure 9-4.

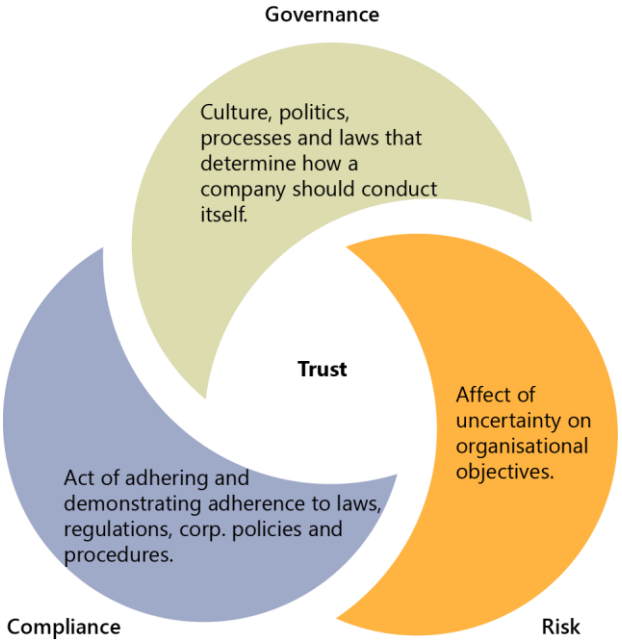


Figure 9-4: Governance, compliance and risk framework

- Governance is the combination of processes that the directors (or the board of directors) establish and execute to achieve their goals.
- Risk management is predicting, understanding and managing risks that could otherwise hinder or prevent the organisation from achieving its objectives.
- Compliance refers to adhering to the policies and procedures as well as laws and regulations.

GRC is a discipline that aims to synchronise information and activity across governance, risk management and compliance in order to operate more efficiently, facilitate effective information sharing, more effectively report activities and avoid wasteful overlap.

The goals, therefore, of any GRC programme must include the following:

- Keeping risk at acceptable levels
- Maintaining availability to systems and services
- Complying with relevant laws and regulation
- Protecting customer data

In general, GRC is not normally a "net-new" function for the cloud, instead, it extends existing activities. GRC professionals should therefore fully understand the implications the cloud has for their areas and extend existing practices.

Ensuring regulatory compliance

Managing regulatory compliance can be a complex task, and for multinational organisations, particularly those in heavily 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.

As with security controls, businesses should understand the division of responsibilities regarding regulatory compliance in the cloud. Cloud providers such as Microsoft make every effort to ensure that their platforms and services are compliant, but companies also need to ensure that their applications, or those supplied by third parties, are compliant.

Similarly, applications in regulated industries that use cloud services might require certification from the cloud provider. For example, a healthcare application that processes patient health information (PHI) is subject to the Privacy Rule and the Security Rule encompassed in the Health Information Portability and Accountability Act (HIPAA), and thus requires that a healthcare business receives 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 organisations 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 (QSA) or by a firm-specific Internal Security Assessor (ISA) that creates a Report on Compliance (ROC) for organisations handling large volumes of transactions, or by Self-Assessment Questionnaire (SAQ) for companies.

Microsoft has received more than fifty compliance attestations at the time of writing this, which you can view in Figure 9-5. Check the Azure Trust Centre from time to time to see updates as well as important prescriptive guidance regarding compliance.

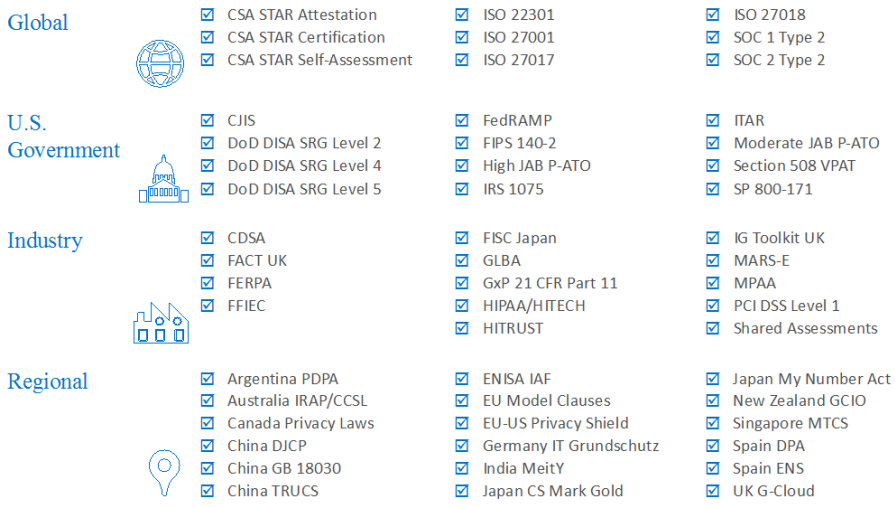


Figure 9-5: Microsoft Azure compliance attestations

Many emerging laws, particularly those dealing with privacy and individual Personally Identifiable Information (PII), require that businesses themselves comply and report on compliance and any breaches that might occur.

One of the most important developments in this area is the recent enactment by the European Commission of the General Data Protection Regulation (GDPR), which is designed to strengthen the protection of data for individuals within the European Union. The GDPR requires that data about individuals, "a name, a home address, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer's IP address,"[11] be maintained on servers within the EU and not transferred out of it. It also requires that companies notify individuals of any data breaches, and mandates that companies have a Data Protection Officer. Other countries have, or are developing, similar types of regulation.

Azure has a number of services to help with GDPR and similar initiatives (Figure 9-6). Azure Information Protection provides document tracking and revocation capabilities, which make it possible for you to both monitor the flow of data through the organisation and to revoke access. With Microsoft Office 365 Advanced Data Governance you can assign classifications to corporate data.

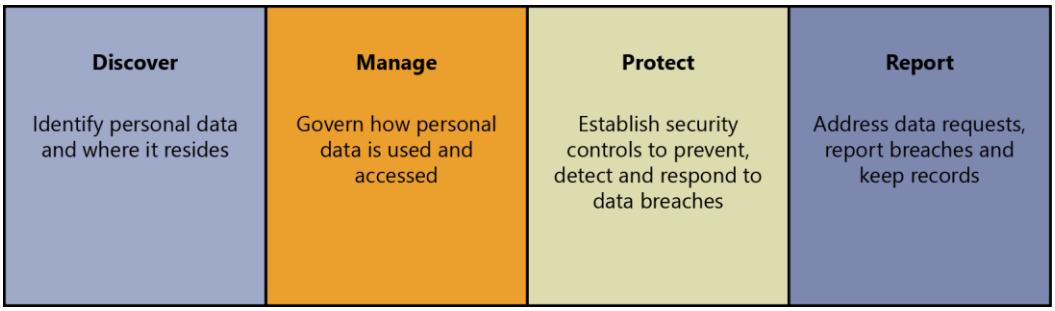


Figure 9-6: Complying with GDPR and other initiatives

Data governance

To ensure compliance with many of the regulations and standards, a data governance function is essential. Since long before the cloud, data governance in IT has been a critical function. Creating and ensuring adherence to common data models, providing extensibility where needed, managing changes, ensuring regular and controlled taxonomy updates, specifying use of master and reference data, implementing data classification, instituting formal processes around data retention and destruction: all of these activities have been a part of the IT governance function for decades. Figure 9-7 depicts the process involved in data governance.

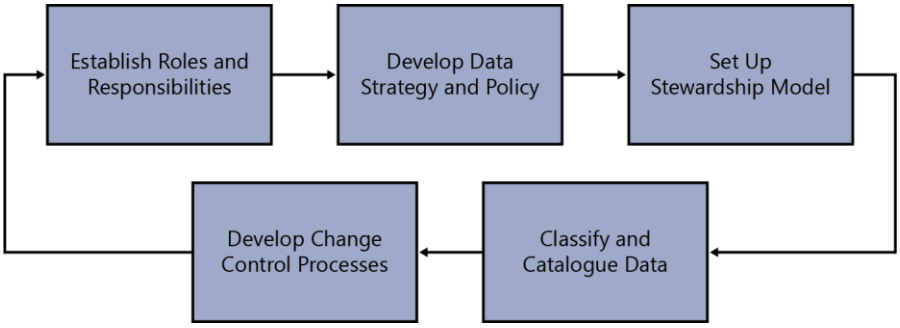


Figure 9-7: Data governance process

[11] http://europa.eu/rapid/press-release_IP-12-46_en.htm?locale=en

For your part, you need to know what data your applications are keeping in the cloud as well as what the laws of your country or region are with regards to data sovereignty and cross-border data movement. Some potential measures you might want to implement include the following:

- Not placing any individual or customer data in the cloud
- Encrypting key PII such as email addresses or physical addresses prior to moving data to the cloud
- Disabling geo-replication to other geographies

Financial governance

We have already described the fairly significant changes to IT finance that are part and parcel of the cloud, for example, the change from a capital expense model to an operational expense or subscription model. Financial governance ensures that the financial changes are managed in a methodical and predictable fashion, including the following:

- Actual cloud costs are in line with predicted cloud costs
- Capital expenses are declining in line with expectations
- Subscriptions are managed in a reasonable way (e.g. by cost centre or application area) and no "rogue" credit card accounts are allowed
- Appropriate chargeback mechanisms are created or extended to support cloud computing
- Quarterly or annual budgeting shows the appropriate changes
- Reporting systems accurately reflect current spend on IT

Change management

Most IT organisations have a programme management office (PMO) of one form or another. The PMO's function is to ensure that changes entail minimal risk and disruption to the IT function. In moving to the cloud, the PMO will need to handle new change management functions. These include the following:

- Operational readiness, to ensure the operations or DevOps function is ready to manage a cloud-resident application
- User readiness, in the case of functional changes to applications
- Organisational readiness, to ensure (for example) that dependent applications continue to function and that all security, compliance and financial requirements are completed
- Application and ecosystem readiness, to ensure applications moving to the cloud and applications that are remaining but integrated with the cloud applications are fully tested and ready, and that all issues are known in advance

There are other aspects of governance (e.g. supplier management). However, it should now be clear that governance in the cloud, by and large, extends existing functions, and professionals in each of these areas should consider the impact cloud applications will have on their space.

ITIL and the Cloud

As Chapter 8 points out, by switching to a DevOps model of Continuous Integration and Continuous Deployment, enterprises can often realise much faster implementation of new features in their cloud applications. However, there are many applications for which change must be strictly controlled, such as applications that manage a company's finances, the most obvious among these being the core Enterprise Resource Planning (ERP) system. The need for these traditional frameworks to control change – and the consequent risk – remains very valid.

Many IT organisations rely on the ITIL framework for service management and operations. Over the years the Information Technology Infrastructure Library[12] has proven a useful set of practices for IT Service Management (ITSM) and for aligning IT investments and operations with business goals. Advocates and practitioners of ITIL point to increased reliability, uptime and predictable costs among its benefits.

ITIL fundamentally concerns itself with IT *services*, that is, the functions and processes that the IT organisation provides to the business. A service is something–an application, a set of applications, information, people–that a business user consumes in order to perform a business function. In general, the cloud as a technology does not change the goals of ITIL; however, the cloud can dramatically change how services are delivered, as we have shown.

ITIL consists of five key strategic areas:

- **IT Service Strategy** ITIL's Service Strategy provides a set of frameworks for determining what services are delivered, how their value is measured, how to measure cost and provide a measure of return on investment (ROI) and how to manage the IT relationship with its business partners. Earlier in this chapter we described how to set up a strategy effort that defines the overall goals – technical, financial and organisational – of the cloud migration effort.
- **IT Service Design** This area covers the design of processes and how they relate to one another, Service-Level Agreements (SLA), capacity and availability management, business continuity management, security and supplier management. We discussed some of these topics earlier in this chapter; Appendix B provides patterns for backup and business continuity.

  IT Service Design also notes the need for a service catalogue, which the portfolio management and configuration management systems are key parts of.
- **IT Service Transition** Service Transition governs how services are delivered and deployed. Such areas as change management, release and deployment management and service evaluation are typically part of the transition phase. The goal, of course, is that new services and changes to existing services are deployed with minimal impact to the overall IT ecosystem.

  Although the structure of Service Transition remains the same, the actual tasks when deploying a service to the cloud change significantly, as we have described. In particular, the emergence of DevOps and its associated methodologies means that the processes and tools associated with deployment are new and different. In addition, IT departments might want to think about areas such as SLA measurement differently. For example, there might be additional latency incurred for network traffic to the cloud over the open Internet.

  Similarly, IT departments should set up a test cloud environment that mirrors the production environment in order to allow User Acceptance Testing (UAT), load and penetration testing, and integration testing with other applications prior to full production deployment.

[12] ITIL® is a registered trademark of the UK Cabinet Office.

- **IT Service Operation** Service Management covers the management and monitoring of services, and how issues are managed and resolved. Key to the Service Management component is the notion of a Service Desk, the primary point of contact for service incidents and events. The service desk, call centre and help desk, if separate, will need to be trained to support cloud-based services.
- **IT Continual Service Improvement** In Continual Service Improvement (CSI), IT personnel and business teams work together to ensure services can quickly meet new and emerging business requirements. CSI is heavily data driven and relies upon operational statistics as well as business insights to determine where focus should be placed.

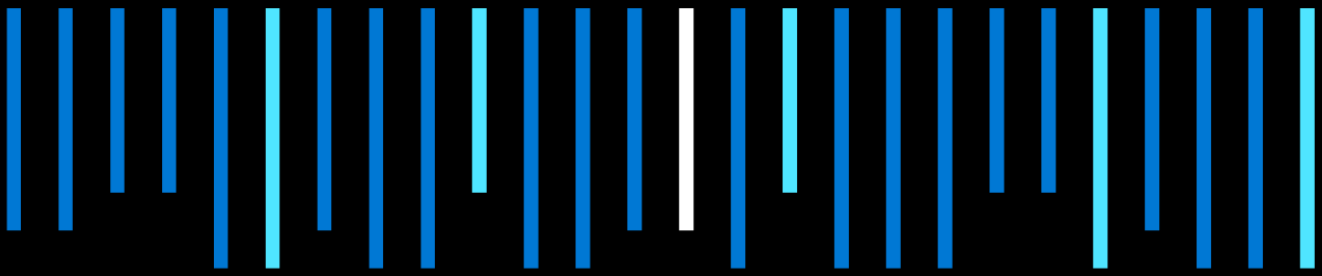
In general, cloud migration will force organisations to change some of the mechanisms and processes by which they implement ITIL, although the basic structure of ITIL is generally technology-independent. However, organisations also should consider how to extend their own processes to be more agile than ITIL might suggest; given that experimentation and prototyping are quick, think about how to do them as part of the strategy and design phases of ITIL.

Moving applications to the cloud is an important and significant activity, requiring changes to how both businesses and IT operate. In Part II, we have described how to form and use a Cloud Strategy Team to drive the migration; how to involve the many organisational stakeholders; how to prioritise application migration; and how to extend existing governance activities.

Of equal importance are the transformational aspects of the cloud, which should be examined and performed concurrently with migration. In Part III, we outline what we mean by transformative innovation and the opportunities afforded by the cloud.

Part III

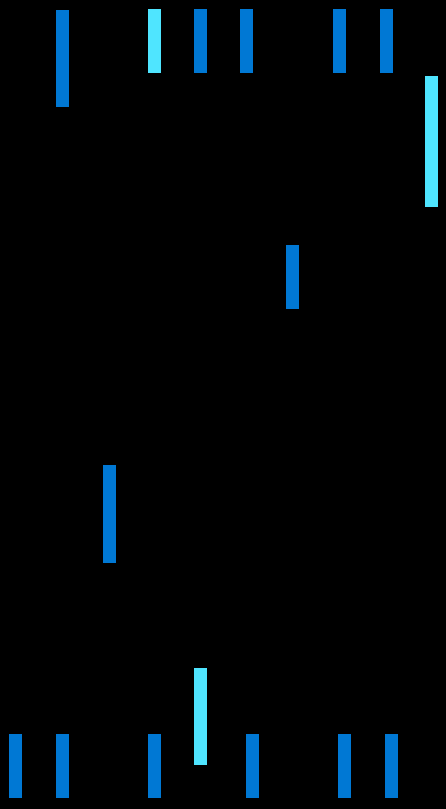
A new age of IT



Microsoft

Chapter 10

To the cloud, and back again



Although much of this book focuses on moving entire applications and ecosystems to the cloud, there is much about the cloud that makes on-premises computing more efficient and cost effective. Many of the more mundane IT tasks, such as backup and restore, can be performed to and from the cloud inexpensively and securely. With messaging buses and integration brokers, enterprises can quickly connect to business-to-business (B2B) sites and, by extending corporate directories to the cloud, they can propagate secure identity management to cloud resources. And, over time, it can become useful to mirror cloud computing paradigms back into the datacentre.

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 drives to an offline media (e.g. tape) in the middle of the night and then transporting that tape to some offsite location.

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

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

- **Recovery Time Objective (RTO):** How fast do you need to get your data back?
- **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. For example, Microsoft Azure Backup, as its 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 datacentre regions (if you choose the geo-redundant option, the datacentres are at least 100 miles apart). Like everything else in the cloud, Azure Backup is a pay-as-you-go service: you pay for what you use.

You also should consider the *backup method*. As Figure 10-1 illustrates, modern backup technologies, including Azure Backup, provide you with the ability to select either a *full* backup, in which you copy the entire source data; a *differential* backup, which stores only data blocks that have changed since the *initial full* backup; or *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.

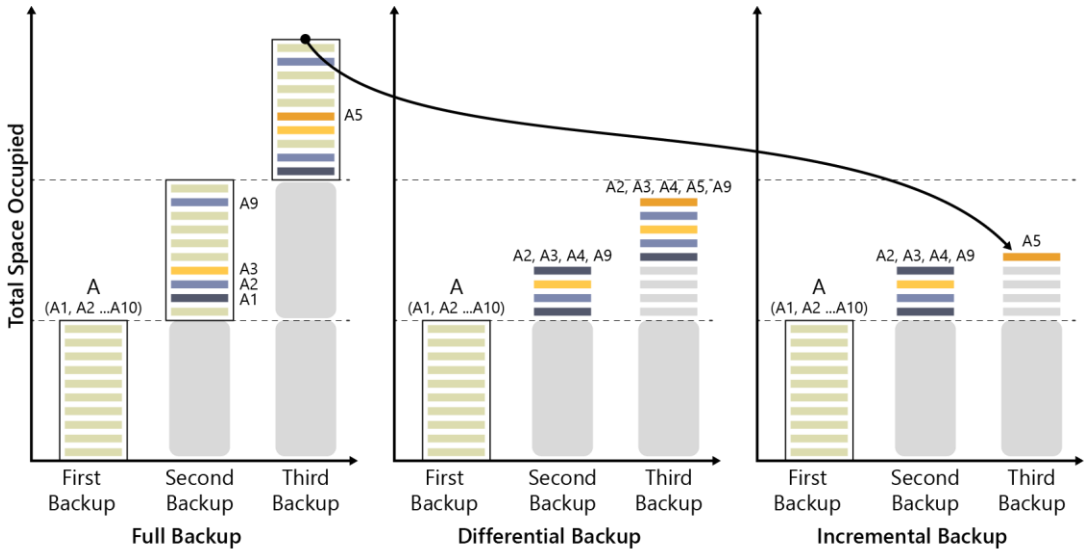


Figure 10-1: Backup modes

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

If you already have Microsoft's systems management tools installed in your enterprise, Microsoft System Center, you can extend its Data Protection Manager function to back up to the cloud. Data Protection Manager 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 uses. Similarly, you can use Transact-SQL to configure Microsoft SQL Server to back data up to the cloud.

Extending on-premises storage to the cloud

IT executives are often faced with the regulatory requirement to retain what often turns out to be vast quantities of historical data. This data is typically very 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. This means data that is needed is still available locally, but the device maintains knowledge of where the entire body of data is, so in response to a regulatory request or other need, it can quickly restore the information. Microsoft offers the StorSimple appliance, which is just such a device.

Business continuity and disaster recovery

A CIO once told me a story about a datacentre his company built that, unknown to them, had an ungrounded metal aerial on the roof. Some time later, lightning struck the datacentre, on that very aerial, 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 Business Continuity and Disaster Recovery (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'll 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'll need to control the *order* in which applications are brought back online, because it's not uncommon that applications depend on one another.

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

Integration

As we've mentioned previously, 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 datacentre, whereas others have been moved. Alternatively, and actually the more likely scenario, you will choose to leave some applications in the on-premises datacentre for the foreseeable future: this is 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.

Networking

First, you'll want to ensure that cloud-based applications are visible to the corporate network; that is, on the appropriate subnet. You can accomplish this by using a Virtual Private Network (VPN), or by implementing a dedicated physical line that connects the enterprise datacentre to the cloud datacentre.

VPN options

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

In site-to-site configurations, a specialised hardware VPN device creates a tunnel, encrypted using IPSec, with Internet Key Exchange [IKE], between the datacentre and the cloud. IP addresses are configured in the device such that cloud resources appear to be on the local network, as depicted in Figure 10-2.

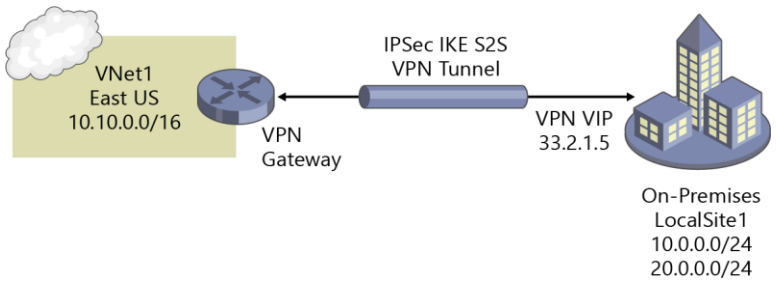


Figure 10-2: Hardware VPN

You can set up VPNs of this sort across multiple on-premises datacentres, as shown in Figure 10-3:

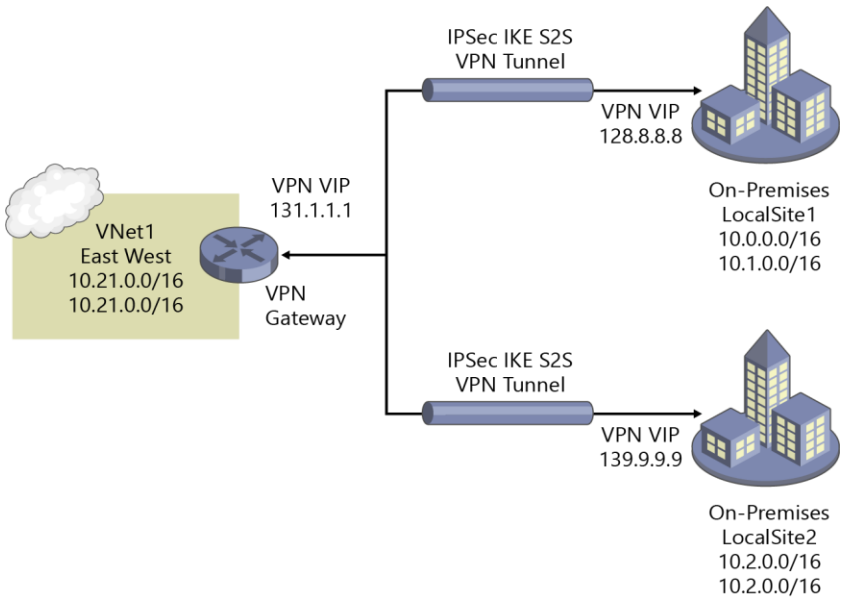


Figure 10-3: Multisite VPN connections

Azure ExpressRoute

With a dedicated line, such as Azure ExpressRoute, enterprises can connect directly from their site to the cloud. However, you must purchase dedicated lines from the local telecommunications provider, and you need to install the appropriate edge router and other hardware at your site. Figure 10-4 presents an overview of an ExpressRoute configuration.

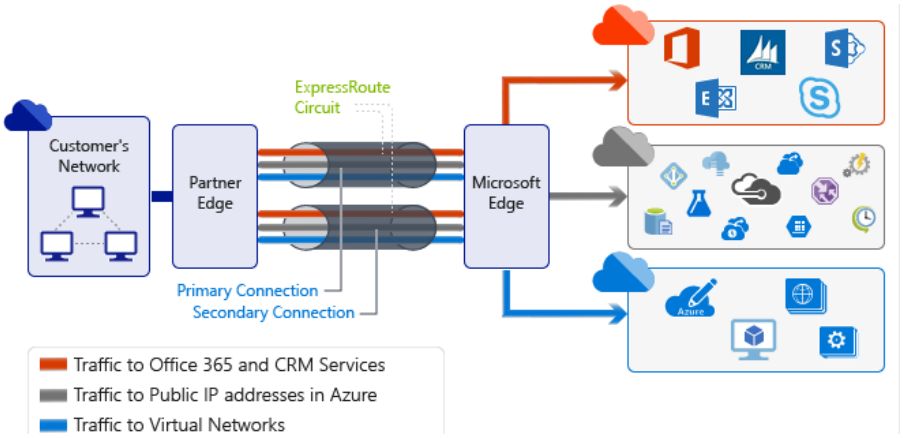


Figure 10-4: ExpressRoute

Such dedicated lines have the following advantages:

- You can typically purchase guaranteed bandwidth from your telecommunications provider.
- You can use ExpressRoute 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 your desired bandwidth, rates defined by the selected telecommunications provider and so on.

Messaging: Service Bus

To provide integration between applications, the cloud offers a number of approaches. For application-to-application messaging, Azure Service Bus, which can connect applications only in the cloud, or applications on-premises with cloud applications, provides a number of different architectural options. Similar in concept to a physical post office, Service Bus is a reliable information delivery service.

Different messaging paradigms supported include:

- **Queues** For first-in, first-out messaging
- **Topics and subscriptions** Applications can declare certain messages to be of a certain type; other applications can then subscribe to them.

Azure Service Bus, a general-purpose message broker, is highly secure and it ensures that messages are reliably delivered. Its actions are *transactional,* meaning that if a given action (i.e. delivery) cannot be completed, its state is rolled back to a known consistent state.

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 way to use these is as so-called integration-platform-as-a-service (iPaaS) brokers, 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 Chapter 11. Logic Apps connectors include EDI X.12, HL7 FHIR, XML, SMS, SAP and literally hundreds of others. Because Logic Apps require no code, they make application integration fast and reliable.

Extending directory services to the cloud

There are three key goals in identity management in the enterprise:

- Users have a "single sign-on" (SSO) experience to applications, both in the datacentre and in the cloud.
- Users should be able to log in to applications from outside the corporate network (e.g. to work from home).
- For certain applications, authentication via external Internet authorities (e.g. Microsoft account, 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 to the cloud – for example, with Azure Active Directory (Azure AD), as illustrated in Figure 10-5.

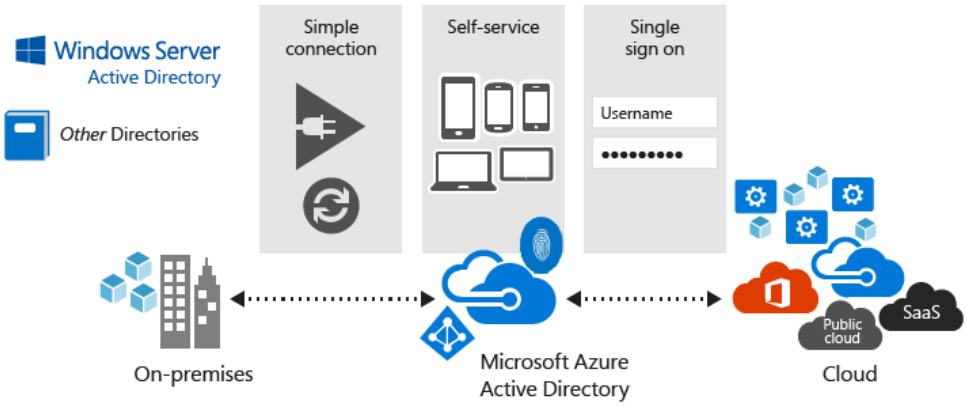


Figure 10-5: Azure Active Directory

Azure AD synchronises with on-premises directories such as Windows Server Active Directory and others. This makes it possible for users to easily log in once and have access to applications locally in the datacentre and those residing in the cloud. Additionally, users can log in from outside the datacentre, and Azure AD will manage the authentication process, co-ordinating with the on-premises directory. In addition, Azure AD can manage Internet authentication sources such as Facebook and Microsoft account.

One of the most important aspects of Azure AD is its connectors to leading software as a service applications; users only need to log in once to access not only corporate applications but others like Microsoft Office 365, SalesForce.com, DropBox, Concur and many others.

An added feature of Azure AD provides the tools to set up consumer authentication at scale, for example, for an e-commerce site that needs to be able to authenticate its customers.

Cloud computing in your datacentre

As you move applications to the cloud, and in some cases perhaps redesign them, there might come a day when your staff is more fluent in cloud technologies than in traditional on-premises models. Alternatively, you might 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 (manufacturing devices on an assembly line is one example). Or, there might be situations in which connectivity to the cloud cannot be guaranteed.

For these types of application, which are, admittedly, uncommon, consider bringing in a cloud "appliance," that is, server equipment that runs cloud software. An example of this is Azure Stack. Azure Stack consists of packaged cloud software that you can run on selected server platforms.

By running cloud services in your datacentre, 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. Figure 10-6 presents an overview of Azure Stack.

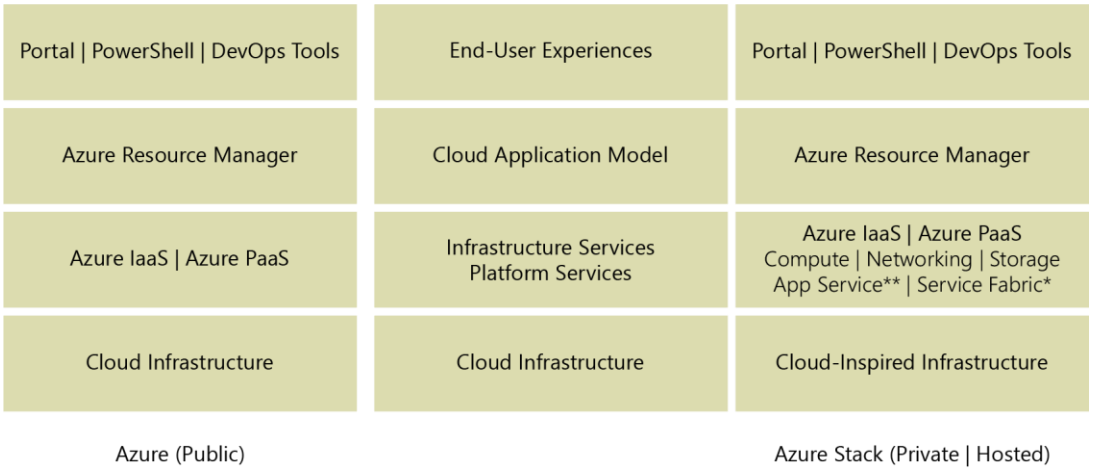


Figure 10-6: Azure Stack

Alternatively, if you cannot guarantee connectivity to the Internet, you can use Azure Stack to guarantee the availability of your services even if your link to the Internet is down or not available – a problem faced, as it happens, by a large passenger cruise ship company, who used Azure Stack to solve it.

With Azure Stack, you can develop a PaaS application or serverless function, and deploy either to the cloud or keep them on-premises, thus providing your development teams with a single programming model.

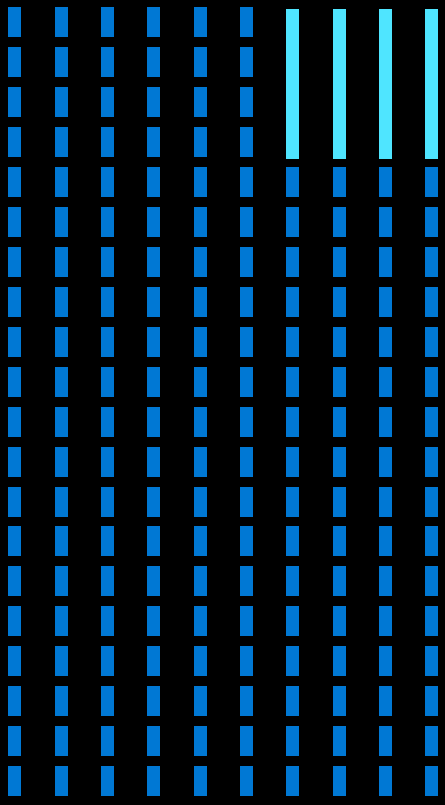
Hybrid cloud management

With a diverse set of applications running in different datacentres and some in the public cloud, management of all of 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, datacentres and clouds to provide administrators with a consolidated view of their ecosystem.

Microsoft's Operations Management Suite provides these capabilities. It can continuously monitor the health of critical workloads such as Active Directory and SQL Server, ensure that all systems (on or off-premises) are properly protected with up-to-date antimalware programs and signatures; and can analyse petabytes of data from both datacentre and cloud to provide a consolidated view of trends, working with both Windows and Linux virtual machines on-premises, in Azure and in Amazon Web few) have similar features.

Chapter 11

New application models



Conceptually, relocating applications to the cloud in the infrastructure as a service (IaaS) model is simple, and, as we have discussed, it comes with a number of advantages. However, both IT and business stakeholders can realise truly transformational value by taking advantage of capabilities native to the cloud: new application models and cloud-native services. In the next few chapters, we discuss how enterprises can take advantage of the unique features of the cloud to bring these benefits to their businesses. We begin in this chapter with a discussion of new application architectures: what they are and the benefits they bring.

Cloud computing has revolutionised the way we develop, test and deploy applications. Because of the easy availability of cloud resources, it has become faster to make new applications available and to quickly update them. And because of this, new models have arisen to support this rapid application development and deployment model.

What does it mean to transform?

Recently, we had a conversation with a start-up that's creating an Internet of Things (IoT) application – in this case, capturing output from a home medical device over the Internet. Figure 11-1 demonstrates how the architecture of the application was quite simple.

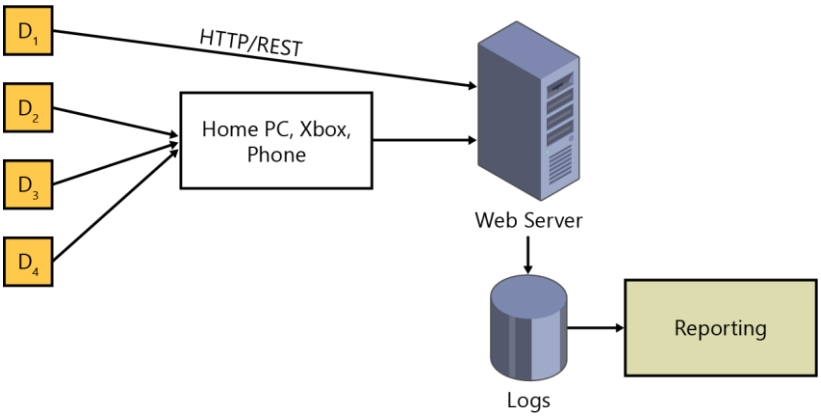


Figure 11-1: Simple IoT architecture

Here, the devices sent data over the REST protocol either directly to a web server or through a router to the server. The server, in turn, did some processing and stored the data in a database where it was subsequently displayed in a home-grown reporting application.

We asked some simple questions:

- What happens when the company becomes wildly successful and must support tens of thousands of devices online, at any given time?
- How would the company support resiliency?
- How could it carry out preventative or predictive maintenance?
- How could it discover its most unreliable suppliers for its devices?
- How could the company add new reports quickly?

We suggested that the company integrate with capabilities in the cloud. In the proposed architecture, shown in Figure 11-2, the start-up simply needs to connect its devices to various services.

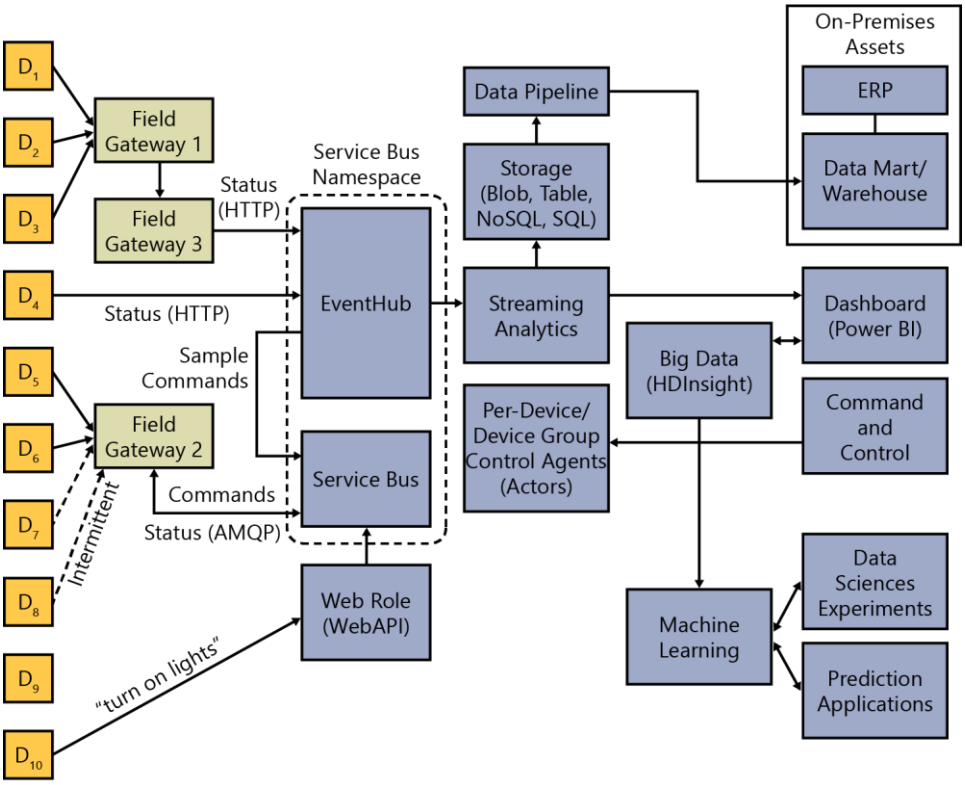


Figure 11-2: A rich end-to-end IoT application

Using Microsoft Azure Event Hubs, which support massive throughput in event ingestion, the start-up's application can easily scale to as many devices as it needs. The data received is stored in Azure Tables which automatically keeps two replicas, ensuring that no data is ever lost. The data can be analysed by large scale MapReduce programs in Azure HDInsight, and actors can provide real-time programmatic command and control. Machine learning applications can be written to predict upward trends in sales or part failures, and all of the data can be visualised in an intuitive and visually pleasing dashboard – all with a *minimum of coding*.

In short, what was once a fairly limited application very rapidly became one that was innovative, insight-rich and transformational.

The point here, of course, is that, by taking advantage of cloud-native capabilities, applications have at their disposal a wealth of possibilities for enrichment. In the next few sections we'll examine these new models in some depth.

Platform as a service

As we have discussed, moving applications to the cloud in IaaS is one approach, and perhaps the most simplistic. Of course, IaaS carries with it a number of advantages, such as passing responsibility for the datacentre to the cloud provider. To really *transform* to a cloud-centric model, designing applications specifically for the cloud is the next step.

IaaS has certain limitations: you are still responsible for maintaining the system software, operating system, and database for your application including items such as periodic patches and software upgrades. In fact, we can say that IaaS is only the *first step* to taking full advantage of the cloud.

Figure 11-3 demonstrates how in platform as a service (PaaS) models you need to maintain only *your application* (in blue in the illustration), whereas the system software is provided by and maintained by the cloud provider. In addition, PaaS offerings typically add seamless scalability and resilience by providing scale-out and data replication, and PaaS can interact with cloud services such as Microsoft Azure Active Directory (Azure AD) for robust identity management.

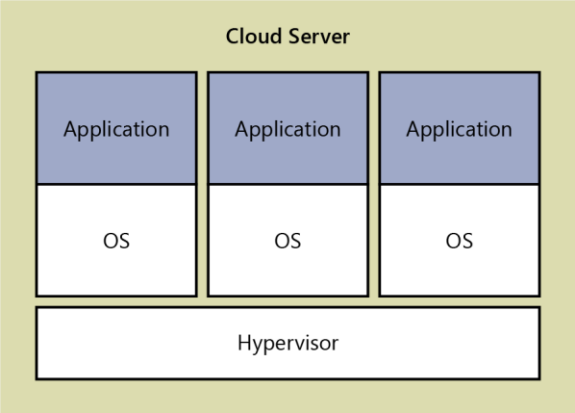


Figure 11-3: Platform as a service

Azure App Service Web Apps, for example, provides a way to rapidly provision a scalable website in the cloud with a minimum of effort. Microsoft provides the underlying web infrastructure (operating system, networking stack, storage, language support and scalability features) that remove much of the systems overhead of managing a large-scale web application. It is straightforward to configure scalability, backup and monitoring capabilities into a Web Apps application. Web Apps also connects to all the other services offered by the cloud for rich applications (more on this later).

Azure Cloud Services are a cloud analogue to the "three-tier" line-of-business (LoB) applications of a decade ago. In Cloud Services, an application consists of three components: a *web role*, effectively a web front end, scalable independently from other parts of the application; a *worker role*, providing background computation and processing (analogous to the business logic layer in the three-tier model); and *persistent storage* using an Azure-enabled version of SQL Server (Azure SQL Database). Although it requires some redesign to take an existing application to Cloud Services, this will be relatively straightforward because the model is intentionally similar to three-tier.

Containers and orchestration

PaaS applications have a tradeoff, which is that while the cloud vendor provides and maintains the operating system environment, the OS and application still operate at runtime as a virtual machine, which means that the start-up time is the same as that of a full operating system.

One of the most important new trends in application architecture is the so-called *container model*, which takes its name from the ubiquitous shipping containers we see every day on ships and lorries. Like these physical containers, software containers are standardised packages of software that are highly portable and that you can deploy quickly.

To understand the container model, it's useful to contrast it with IaaS and PaaS. In both of these, a *hypervisor* manages multiple operating system instances on a server (in IaaS, you provide the OS, in PaaS, the cloud provider does). On any given server then, one might find multiple (very large) operating systems running concurrently and in parallel – with exactly the same functionality (each OS has a file manager, a network subsystem etc.).

In the container model (Figure 11-4), applications *share* a single instance of the operating system. Both Microsoft Windows and the various Linux distributions have been enhanced to support the isolation required to ensure that each application appears to "own" the OS. Applications are packaged to be deployed on container-capable systems.

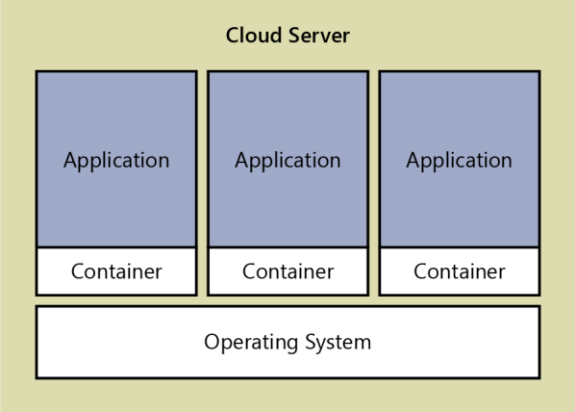


Figure 11-4: Container architecture

One result is that application start-up is considerably faster because the overhead of loading an entire operating system for each app is avoided. Another is that you can create standard, portable packages or *images*, such as an image for a web server or for a database, and you can deploy these without complex installation.

Another result is that containers achieve much more efficient use of the hardware because on a given server the number of actual operating system instances is limited (and there might only be one).

In a typical container environment, a number of servers – a *cluster* – run instances of containers, which often include web servers, back-end logic, search, real-time analytics and so forth. Software to deploy the desired number of instances of each, to update them in a controlled way, to handle failures and to manage scale-out is called *orchestration* (Figure 11-5). Most orchestration services provide tools so that administrators can create rules to, for example, prevent one container type from being on the same server as another, or to handle failover and recovery in a controlled and logical fashion. Some well-known orchestration products include Kubernetes, Mesosphere, Docker Swarm, and Deis, as well as Microsoft's Service Fabric, which we cover in more detail in a moment.

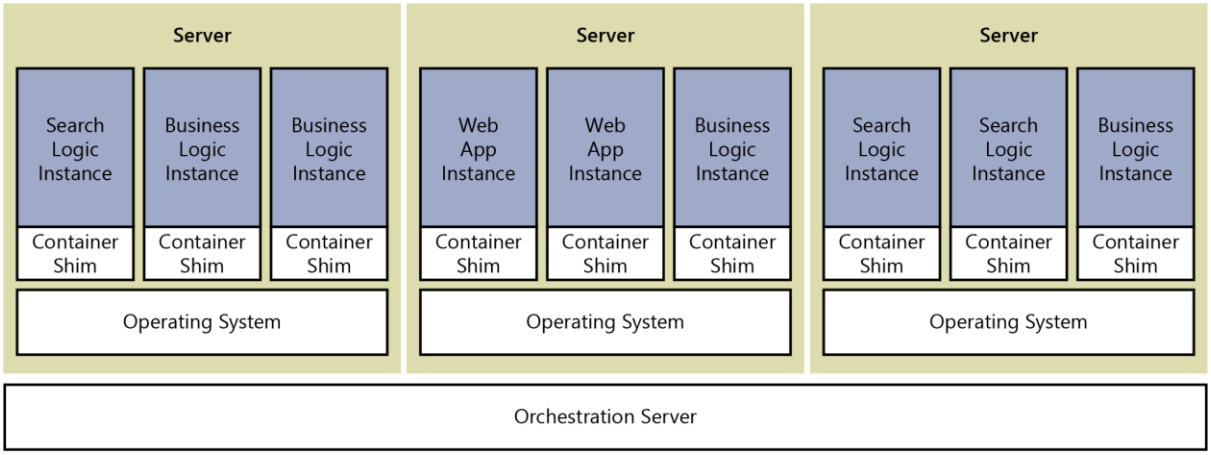


Figure 11-5: Orchestration

Containers can be a good way to move legacy applications to the cloud and gain efficiency. Microsoft's own IT organisation ported a number of applications to Docker containers running on Azure and realised four primary benefits:

- More efficient infrastructure usage
- Standardised infrastructure configuration
- Isolated application environments
- Increased application portability and reuse

At the end of this proof-of-concept of 10 applications, Microsoft IT had infrastructure that was 400 % denser than without containers and overall infrastructure was reduced by 300 %.

Microservices

The term "microservices" was coined a few years ago. It encompasses the mindset that large, monolithic applications be split up into smaller, componentised services. For example, in Figure 11-5, we have just three fairly coarse-grained pieces of our application: a web server, business logic and a database.

Consider, however, if we were building an e-commerce application. It would perhaps be a better use of resources if we had one team building the catalogue component, another building the ordering component and still another incorporating a commercial or open-source search function from a third party. Then, these components could be independently developed and updated.

Microservices are more of an architectural principle (or design pattern) than an actual technology. You can build microservices in IaaS, PaaS or using containers.

Azure Service Fabric provides a platform for building mission-critical applications that you can use for building microservices-based solutions. Service Fabric is tried and tested, in that, prior to becoming a generally available product, it was used internally by Microsoft to host core Azure infrastructure as well as other Microsoft services such as Skype for Business, Intune, Azure Event Hubs, Azure Data Factory, Azure Cosmos DB, Azure SQL Database, Dynamics 365 and Cortana.

Service Fabric hosts and orchestrates a variety of application models, including containers and actor models (described in the following section), and, having been designed for high reliability and availability, it provides automatic scaling, rolling upgrades and self-healing from faults when they occur.

Actor model

Another tool in the cloud developer's toolkit is called the *actor model* (see Figure 11-6). An "actor" is a simple, generally small object in the cloud that has a unique identity, can communicate with other actors, and maintains its state. Actors typically represent physical objects, such as people or devices; a new term that has recently been applied to this sort of technology is *digital twin*: the actor object essentially mirrors digitally what is going on in the real world. The *actor framework* abstracts out infrastructure concepts such as servers, meaning that actors can communicate with one another without having to know whether they are on different physical servers.

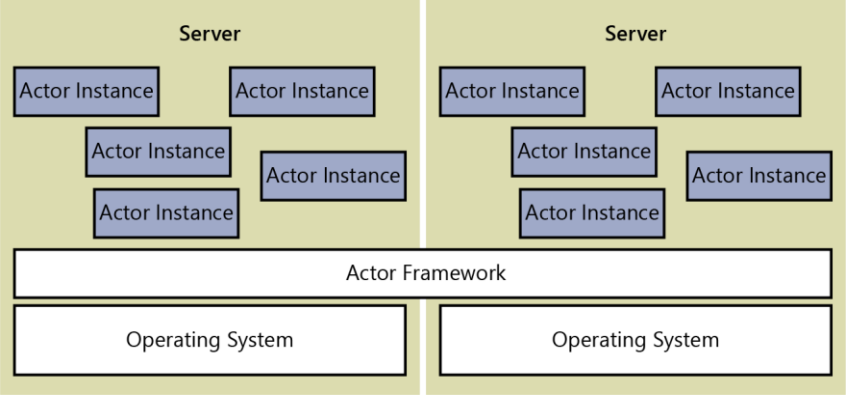


Figure 11-6: Actor model architecture

An example usage might be online games, in which each actor instance represents a gamer, and holds data such as the gamer's current score, location in the game and a list of other gamers participating.

> **Note** One of the world's largest game franchises, Microsoft Halo from 343 Studios, uses the actor model in just this way.

You can also use actors to manage large numbers of IoT devices; they are particularly efficient when there are many similar devices, for example, sensors being monitored. Each actor instance can receive status updates from a given sensor, perhaps a pressure sensor, and can automatically notify an actor controlling a valve or other device to take appropriate action.

Resilience in the cloud

It's important to note that in the cloud costs are kept low by using commodity hardware. Whereas in the past, businesses purchased higher-end machines to scale up, now, in the cloud, capacity is achieved by scaling *out*, instead. So rather than buying a large-scale supercomputer, you achieve scale by using many machines in a distributed way, and this means that you must plan for the occasional failure.

To achieve resiliency – that is, to continue to function in the face of failures – think about how to recover from individual server failures, perhaps by having multiple instances of an application or service, and how to recover from a catastrophic failure, perhaps by using Azure Site Recovery, as is discussed in Chapter 10.

Here is a model to follow when considering application resilience:[13]

1. **Define** your availability requirements, based on business needs.
2. **Design** the application for resilience. Begin with an architecture that follows proven practices, and then identify the possible failure points in that architecture.
3. **Implement** strategies to detect and recover from failures.
4. **Test** the implementation by simulating faults and triggering forced failovers.
5. **Deploy** the application into production using a reliable, repeatable process.

[13] https://docs.microsoft.com/azure/architecture/resiliency/

6. **Monitor** the application to detect failures. By monitoring the system, you can gauge the health of the application and respond to incidents if necessary.

7. **Respond** if there are incidents that require manual interventions.

The level to which you implement resilience features is partly a function of your business requirements. Consult with your business partners about the desired Recovery Time Objective (RTO) and Recovery Point Objective (RPO), which we also discuss in Chapter 10. Consider usage patterns as well, for example, if there are periods when the application or system absolutely must be available. In such cases, you might want to add additional cloud resources above and beyond less critical times.

Consider your desired Service-Level Agreement (SLA). Table 11-1 shows the potential cumulative downtime for different SLA levels.

Table 11-1: Service-Level Agreements

SLA	Downtime per week	Downtime per month	Downtime per year
99 %	1.68 hours	7.2 hours	3.65 days
99.9 %	10.1 minutes	43.2 minutes	8.76 hours
99.95 %	5 minutes	21.6 minutes	4.38 hours
99.99 %	1.01 minutes	4.32 minutes	52.56 minutes
99.999 %	6 seconds	25.9 seconds	5.26 minutes

Engineering to increase the number of "9s" might or might not be worth it, depending on the business scenario.

Another important notion is that of cumulative SLAs. Consider a system that uses a web application and a SQL database, as depicted in Figure 11-7; an Azure queue is used to hold pending updates if the database becomes unavailable.

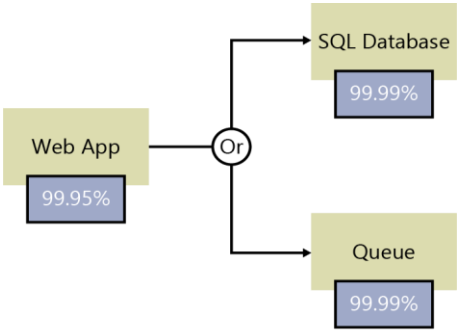


Figure 11-7: System with components having different SLAs

Each of these has a respective SLA. A simple calculation shows the cumulative SLA of all the parts taken together:

Database *or* queue = 1.0 − (0.0001 × 0.001) = 99.99999 %

Web app *and* (database *or* queue) = 99.95 % × 99.99999 % = **~99.95 %**

“Serverless” applications

Perhaps the most exciting new application model has a very misleading name, so-called "serverless" applications; of course there is a server: you just don't need to create it, manage it or pay for it. In many ways, serverless applications have the most attractive time-to-value because there’s no coding involved.

Serverless applications are applications that you can snap together from existing components with no coding required, making it possible for you to build applications quickly and inexpensively.

Here's a simple example. Suppose that, as a customer service lead, you want to know when customers report a product malfunction. Here, using Microsoft Flow, you can monitor Twitter for mentions of the hashtag #BrokenAcmeWidget; when one appears, Flow automatically copies it into your local Slack (a popular collaboration tool) channel, as shown in Figure 11-8.

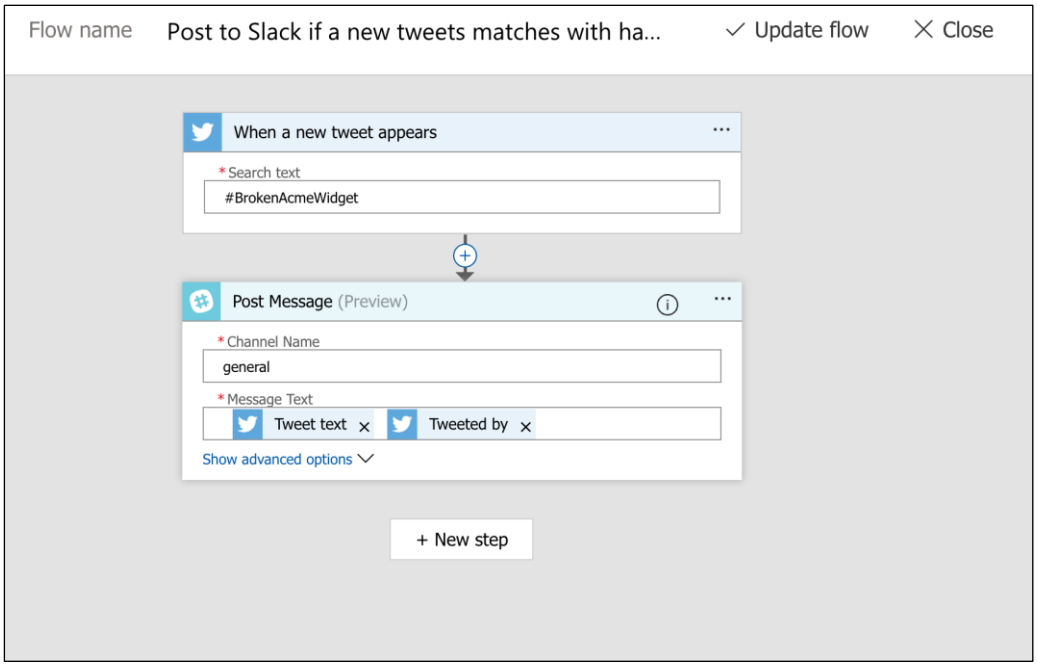
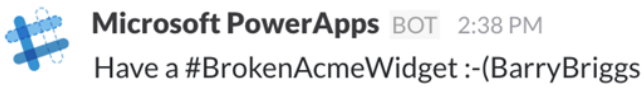


Figure 11-8: Using Microsoft Flow, a serverless cloud application

And here's what you would see in Slack:



Perhaps now you, as the customer service representative, want to open an issue, fill in a form, and kick off a workflow that ultimately provides a response to your unlucky user. Creating a form-based application used to require development resources, but with new serverless capabilities (here, showing Microsoft's PowerApps application), it doesn't. Here, the form is designed and deployed and, when used, it kicks off another Flow, as shown in Figure 11-9.

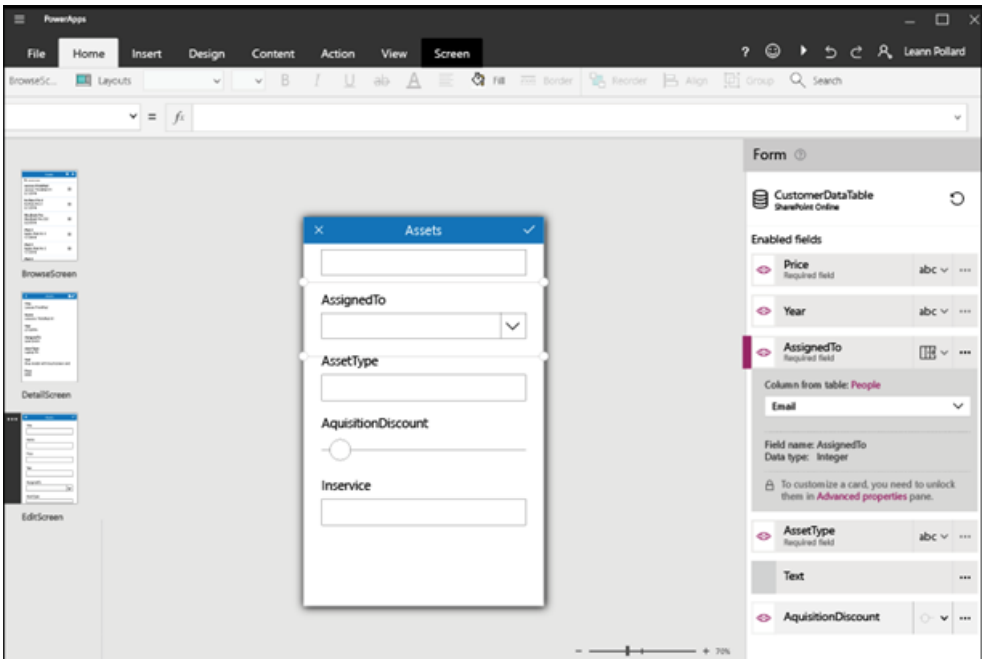


Figure 11-9: Microsoft PowerApps demonstration

Finally, there might be cases for which it makes sense to build custom processing for a particular external event, such as a Tweet. Here, your developers can build serverless "functions" that others can then use with no code to connect the Tweet to, for example, a sentiment analysis engine. In this case, development is so streamlined that no interactive development environment (IDE) such as Microsoft Visual Studio is required: the programming is done directly in the Azure portal, as demonstrated in Figure 11-10.

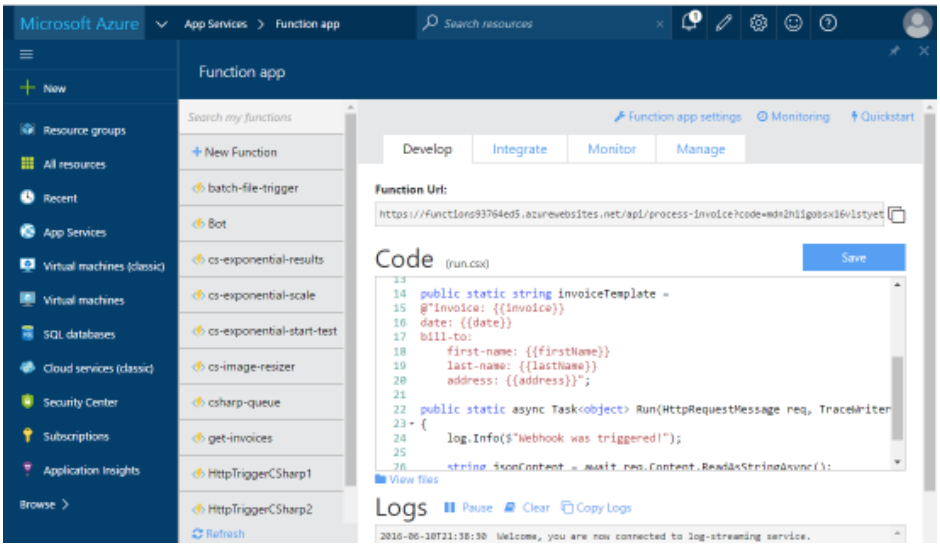


Figure 11-10: Azure Functions

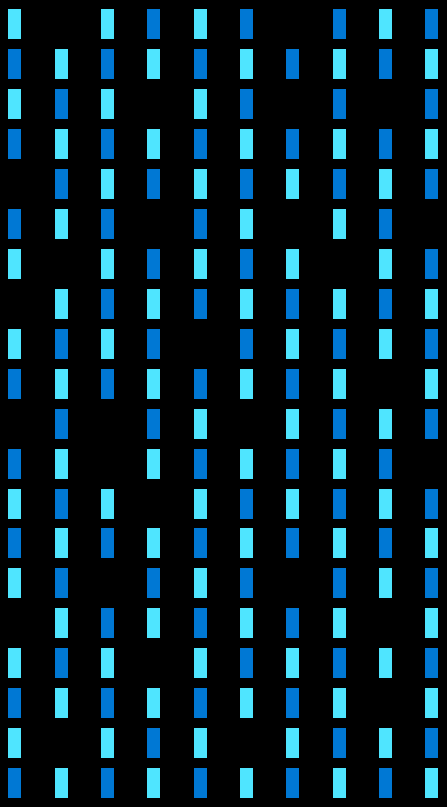
And again, after you create them, you can reuse these functions in any number of applications simply by dragging them together.

These serverless capabilities might well presage a new model of cloud applications. Increasingly, cloud apps are evolving towards a lego-block model of "serverless" computing, in which you create and pay only for your business logic, where chunks of processing logic are connected together to create an entire business application.

Infrastructure? Of course, it's there (as we've said, "serverless" might not be the most accurate term), but it's under the hood: the cloud provider manages the servers, configures them, updates them and ensures their availability. Your concern is what it should be: your business logic. And, you haven't deployed a single server. What code you've written is business logic only; not administration scripts or other code with no business value. Your developers have focused on growing your business. And, most importantly, you've created a rich, intelligent end-to-end application by simply attaching together existing blocks of logic.

Chapter 12

It's all about the data



Much of what we have discussed so far is about the "compute" side of the cloud. But, the cloud also offers extensive storage capacity at very low cost. This means that data that we once ignored, such as telemetry or user comments, we can now capture, manage and analyse. To handle all of these differing kinds of data a plethora of database technologies have emerged in the past few years. In this chapter we examine how traditional database technologies have moved to the cloud, the new “NoSQL” data management technologies, and how you can use advanced analytic and BI capabilities to derive new insights from all this data.

It is now axiomatic that information is an enterprise asset, so much so that some have proposed accounting for its value on financial reporting statements. Every decision that is made, every investment, virtually every official action taken in the enterprise is in some way traceable to information. IT managers are charged with providing decision-makers with up-to-date, accurate information and, increasingly, with in-depth analysis of the information such that consumers can derive new insights. Finally, IT managers are in most cases entrusted with safeguarding not only corporate secrets but also the personal information of employees, partners and customers. Is it any wonder, then, that data management has become a critical focus of the modern enterprise – and of the cloud?

It's common, in fact, to distinguish between raw data and information; we might say that technology provides us with the tools to process the data into information. But today, the sheer volume of data we can collect and store defies the imagination, and, increasingly, enterprises find the only way to manage this volume efficiently is to make use of cloud storage resources. Fortunately, the cloud provides many options for storing, managing, analysing and extracting value from these vast quantities of data.

Enterprise data management before the cloud

Prior to the advent of the cloud, enterprises primarily organised and managed their data in two ways. Let's take a look at each.

Structured data management

Not that long ago, the core of enterprise information management was the relational database, which held highly structured data in carefully defined tables composed of rows and columns. Relational database management systems (RDBMSs) held, and continue to hold, tremendous value for the enterprise. Perhaps their greatest value lies in their inherent *integrity,* the understanding that the information in the database could be counted upon to be consistent, no matter what happened – for example, if the server crashed after you made a withdrawal from your current account, but before you deposited the funds into your savings account, you were guaranteed to not lose money. Indeed, we often talk about RDBMS transactions as "ACID," which means atomic, consistent, isolated (from one another) and durable.

Over time, a sort of hierarchy of RDBMS applications evolved in the enterprise, as is illustrated in Figure 12-1. Master data management applications, whose data changes relatively infrequently, holds core information (i.e. customer names and addresses, or the component parts of a product), is *referenced* by other applications; the goal of master (or reference) data systems is to ensure that data that is shared across multiple applications (e.g. customer data shared by Customer Relationship Management [CRM], Enterprise Resource Planning [ERP] and support systems) is identical.[14]

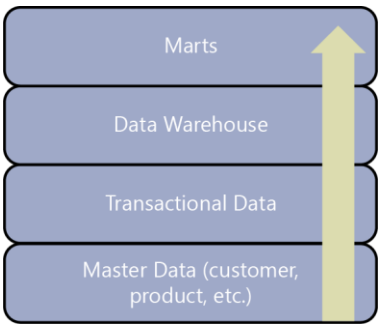


Figure 12-1: Enterprise data management in the 1990s

Transactional systems, by contrast, are frequently updated: these are the workhorses of the enterprise, handling customer purchases, financial transactions, supply-chain management and so on. In a large company, many transaction systems reflecting different parts of the business or different businesses exist. Transaction systems are coded for fast, interactive response.

For reporting purposes, transaction data would be periodically uploaded to a data warehouse, which managed large amounts of historical data. From the warehouse, data would then be sent to data marts. These are database applications that fulfil a specific function, such as predictive analytics or campaign management.

[14] We recognise that this is perhaps an oversimplification of theories advanced by data warehouse gurus Bill Inmon and Ralph Kimball. Inmon, in particular, might suggest that marts aggregate into the warehouse, as opposed to how we show it. Either is valid: our point is simply to show that for a time there was a relatively straightforward taxonomy of data management tools.

Generally speaking, data in this model always flowed upwards, as the arrow in Figure 12-1 indicates, only rarely was a change made in lower systems, such as an update to a customer address in the master data system.

All of these databases required support: a dedicated staff of database administrators (DBAs) handled backup and restore, changes to the layout of the database (new columns), ensuring error-free transfers of data from one database to another, and so on.

Unstructured data

Of course, there are many other kinds of data besides simple structured data. We humans typically consume information in *unstructured* formats – such as the book you are currently reading! Text, audio, video, images – these are all examples of unstructured data. The phenomenal growth of the web has brought with it a simultaneous explosion in the volume of these unstructured assets. Many of these come with special requirements, such as the ability to stream video, to support users moving the playhead about, to provide thumbnails and so on.

Moreover, as we have discussed elsewhere, applications in the cloud generate vast amounts of data, such as telemetry data.

Enterprise data management in the cloud era

It probably comes as no surprise that data management in the cloud brings numerous advantages and opportunities. As with computing resources, storage is a "pay-as-you-go" service, meaning that you only pay for the resources you use.

Other advantages accrue from the nature of how the cloud is designed. Because cloud servers tend to use commodity storage devices instead of expensive Storage Area Networks (SANs), resilience is handled by redundancy: cloud storage is typically backed up by two independent replicas, which might (at extra cost) be spread across geographic regions (called *geo-redundancy*).

Other benefits depend on the particular service, which we describe in the next few sections.

Core storage concepts

The most basic storage abstractions in the cloud are the *blob,* the *file,* the *table* and the *queue.* Let's take a look at each.

Blobs

The term blob is short for *binary large object*, and it is somewhat misleading because a blob need be neither large nor binary. A blob simply means an unstructured "bag of bits," which can be anything from an image to a text file to a video to, well... anything that can be stored digitally. Blobs are roughly analogous to individual files (although there is a separate "file" abstraction, which we cover in a moment).

Blobs are often used to hold web assets, such as images.

In Microsoft Azure, there are several types of blob, and the maximum size of one is approximately 4.75 TB.

Files

Many applications understand the concept of a file and in particular use the widely adopted Server Message Block (SMB) protocol to access networked files. The file abstraction is present to support this compatibility.

Tables

Tables, in Azure, are simple key-value pair lookup tables providing a basic level of structured data to applications. The maximum size of a table in Azure (at the time of writing) is 500 TB.

Queues

Queues, as the name implies, facilitate a scalable means of sending data from one application component to another. Queues are useful because they can tolerate sudden bursts of activity and prevent applications or servers from being overwhelmed.

Relational data in the cloud

Of course, the need for the powerful consistency and query capabilities in a relational database has not declined in the age of the cloud; if anything, relational data stores have grown in importance.

However, using a relational database in the cloud brings certain advantages. Many database vendors, including Microsoft, have converted their database products to platform-as-a-service (PaaS) offerings, which means that system software updates and other ordinary maintenance is handled by the platform provider. Moreover, as with all cloud-native offerings, consumers of such "databases-as-a-service," as they are sometimes called, are charged only for the capacity used.

As an enhanced feature, Microsoft's offering, Azure SQL Database, makes ongoing recommendations to users about how they can optimise the performance of their databases, including suggesting that indexes be created or deleted, fixing schema issues, and many others.

Microsoft's petabyte-scale Azure SQL Data Warehouse provides the functionality of the data warehouse described above – but again with the advantages the cloud provides. Like its cousin, Azure SQL, the Azure SQL Data Warehouse is fully managed and can scale as needed.

> **Note** With respect to "managed services," it is certainly possible to run SQL Server or many other products in infrastructure-as-a-service (IaaS) mode – that is, inside of a virtual machine (VM). In this case, you have to maintain the operating system, database software and so on, as it is in the IaaS model. With a managed service, such as Azure SQL Database, the cloud provider – Microsoft – does all the system software maintenance. Using a managed version of a software product is usually much more cost effective than running in IaaS mode.

The rise of NoSQL databases

In the past few years, a number of new architectures for storing and managing data have appeared that do not replace, but rather complement, relational databases. Such databases easily scale across many servers, which is a key requirement as the amount of data collected grows quickly.

One of these is the "document database," which holds text, usually in the form of JavaScript Object Notation (JSON). Here is a very simple "document" (record) in JSON format:

```
{
      "FirstName"  : "Barry",
      "LastName"   : "Briggs",
      "Tags"       : ["baseball-lover", "author", "coder"],
      "Profession" : "Software Person"
}
```

Document-orientated databases do not replace relational databases, but they can provide a number of advantages in certain situations. For example, document databases avoid much of the rigour and overhead of relational databases. Documents do not need to follow a rigidly typed schema; new fields (columns) can be added to records as needed, whereas such schema changes are quite burdensome in relational databases. Different implementations of document databases provide varying degrees of ACID transactions.

Although you can configure some document databases, such as Azure's CosmosDB (formerly known as DocumentDB), MongoLabs' MongoDB, or the open-source CouchDB, to have very strong consistency constraints (ACID transactions, as we discussed earlier), they can also be set to have *eventual consistency*, meaning that over time various related records will become consistent with each other (your withdrawal will *eventually* show up as a deposit).

For a banking application, a document database might not be appropriate. But for one that tracks social media posts – for example, tweets by user, time and topic, with a list of recent tweets for each – a document database is an excellent choice. A product catalogue might be another good application.

Another form of NoSQL database is the *graph* database in which data is stored not in rows and columns but as references to one another. For example, in a graph database, you might have a record describing a person. That record might then point to other friends, to records describing favourite movies or foods, or to comments on a social media site.

It's certainly possible to build a graph database in a relational model, but if you are finding that the number of foreign keys is large, a graph database such as Neo4j might be more appropriate. Different graph databases support different query languages (some custom, some based on SQL); others, such as GraphDB, base their architecture on the World Wide Web Consortium's Resource Description Framework (RDF).

Of course, the overarching point in this discussion is this: myriad high-scale data management options have emerged in the past decade or so. As you consider new sources and applications of your data, think about which works best for you.

Big data, and bigger data

But beyond custom databases is the explosion in the sheer volume of data we are now able to collect. We've all heard the statistics: that every second 1.7 megabytes of data is created for every human on the planet; that by 2020 we will have created some 44 *zettabytes* of data (that's 44 trillion gigabytes); that on a given day, a billion users visit the social media site Facebook.

It's easy to see where all this data is coming from: uploaded photos and videos; server logs; software telemetry; social media; hardware telemetry; usage tracking of cell phones, web browsing and millions of other human activities; and so on.

Indeed, many governments now place vast quantities of data in the cloud (in the United States, for example, at http://www.data.gov; in the United Kingdom, http://www.data.gov.uk; http://data.gouv.fr in France; and so on). Other companies make data available over the web for a fee, and such data can augment or even replace on-premises master data sources or can provide additional marketing insights. Use of such data can add additional insight to your models.

The challenges from big data come not just from the sheer volume, but also from the speed at which it arrives (its velocity) and the many data types it comes in (its variety). Many refer to these as the "three v's" of big data: volume, velocity and variety.[15]

[15] The concept was introduced by Gartner analyst Doug Laney in 2001.

And as we've learned, there is a lot of *information* in all that data. For example, by careful analysis of server logs, we might see that every now and then there is a failed sign-in from a distant country: perhaps a cyberattack in progress. Or, by analysing comments on Twitter, you can learn how people feel about your product, or about anything else – sentiment analysis.

So-called "big data" storage and analysis has now become commonplace in the enterprise. Big data architectures typically use Hadoop (an open-source project named for the inventor's son's stuffed elephant) or one of its successors. Hadoop's core algorithm is called MapReduce, designed specifically for querying and analysing very large datasets.

Here is a very simple example of MapReduce in action (more of a thought experiment, really: be prepared to suspend your disbelief for a moment). Imagine you want to find every occurrence of the phrase "Abraham Lincoln" on the web. There are many ways you can do this, but one fast way would be to put all the web pages that begin with "A" on one (gigantic: this is the suspending belief part) drive attached to a computer; all those that begin with "B" on a drive attached to another computer, and so on. Then each computer counts the number of "Abraham Lincolns" on its drive. Finally, each computer sends its results to a central computer which adds all the intermediate results and gives the final answer.

Here, what we've done is sent a little command – or program – to the alphabetised computers: putting the code with the data, a core concept of big data. We are not updating it, not performing transactions, just scanning it to see what insights we can gain.

Hadoop has spawned an entire family of software products and technologies, including Hive (SQL-like queries), Pig (language for writing MapReduce programs) and many others. Of particular interest is Spark, which brings faster performance using in-memory transformations. Both Hadoop and Spark are supported as a managed service called HDInsight in Azure, and also from HortonWorks and Cloudera (Hadoop) and Databricks (Spark), each with added-value features. The open-source Cassandra project (productised by DataStax) provides a big data solution with very high performance and SQL-like query capability.

The data lake

A data lake, of which Hadoop's distributed file system is one kind, is a method of storing massive amounts of data in a wide variety of formats. Data lakes are often compared to data warehouses, with a very important difference: data warehouses are an aggregation point for data, they are about *structured* (relational) data, and typically data from transactional systems must be transformed to fit the models (schemas) in the warehouse.

Data lakes, by contrast, hold raw data in whatever format and structure (or lack thereof) it originally comes in.

Analysis services and data visualisation

Business intelligence (BI) applications often require a special kind of database, one that facilitates "slicing and dicing" of the data, aggregating it in various ways, and running algorithms such as segmentation or regression against different views of the data. In the relational world, these analytic services are powered by so-called multidimensional data structures, often just called *cubes*. Cubes often bring together data from many different sources.

Azure Analytic Services, the cloud-managed service cousin of Microsoft's SQL Server Analysis Services provides this functionality in Azure, and like other cloud services, it operates on a "pay as you go" basis. Recognising that the demand for such services is often cyclic (say, at quarter end) Azure Analytic Services can be paused in between uses.

Azure Analysis Services has a close integration with the Microsoft Power BI visualisation tool (Figure 12-2), which provides simple-to-build but powerful "drill-down" visualisations of the information in not just Analytic Services, but a wealth of data providers.

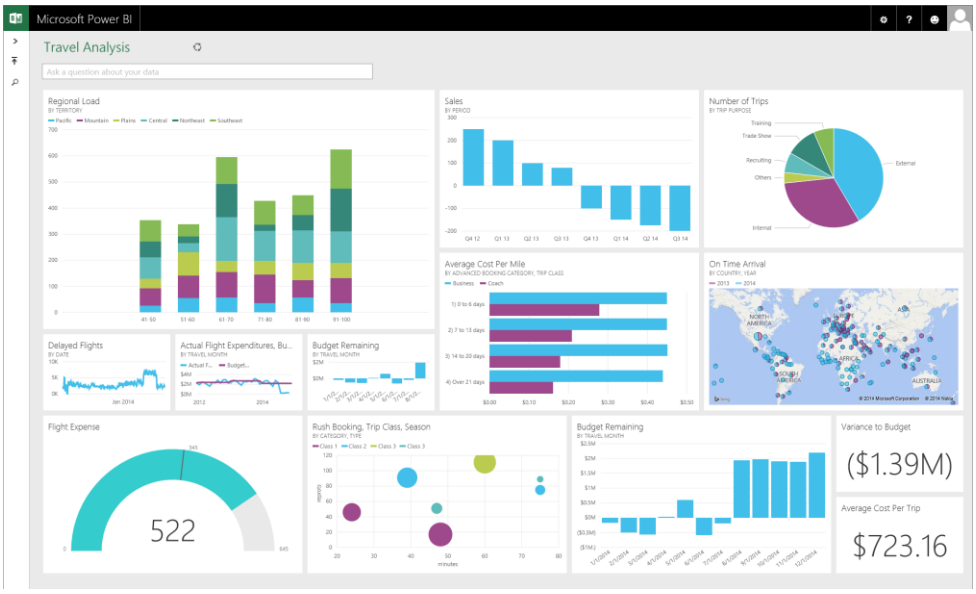
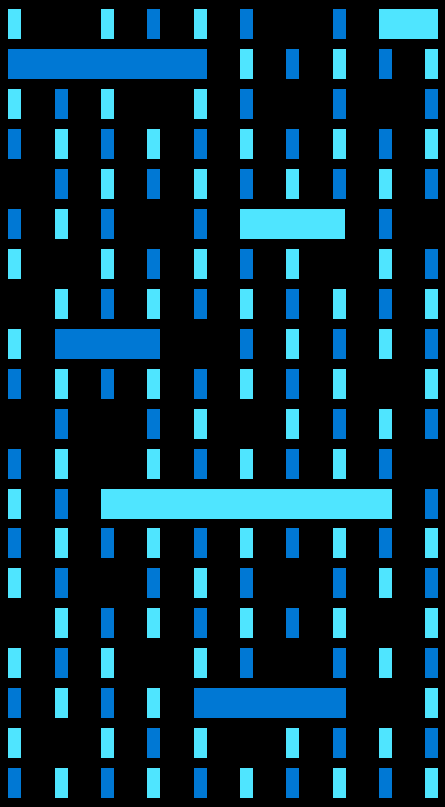


Figure 12-2: Power BI visualisation

Chapter 13

AI transforms your business



For years, the field of artificial intelligence (AI) languished in academic computing laboratories and other research facilities. But the cloud, with its massive amount of computing power and nearly infinite storage capacity, has enabled AI to enter the mainstream, and it has revolutionised computing. Talking to your computer – that is, using its speech recognition function, once fraught with errors – is now commonplace and largely error-free. Moreover, with AI technology, it is increasingly possible to predict when a part is likely to fail, or if a patient is likely to be re-admitted.

The idea of the "intelligent computer" has long been a staple of science fiction (HAL of *2001: A Space Odyssey*), and, for that matter, of research. In the 1960s, MIT professor Joseph Weizenbaum wrote a short program – barely a few hundred lines long – that emulated the methodology of a psychologist. Eliza, as the program was known, would ask users what concerned them, and then would parrot back their words in the form of a question ("My dog hates me." "Why does your dog hate you?" and so on).

Modern artificial intelligence has come a long way and in this chapter we provide a brief overview of why it has been reborn in the cloud era and the nature of the technology. Then we cover some of the problem spaces and opportunities modern AI can address.

What are AI and machine learning?

Broadly speaking, we can define AI as embedding human-like capabilities into computers: vision and speech, cognition and reasoning, and pattern recognition and prediction. Over the past few decades quite a number of approaches to AI have waxed and waned in popularity, with some enjoying mini-renaissances as the technology improved, whereas others have been discarded.

The current resurgence of interest in AI has ultimately stemmed from three important realisations:

- AI requires huge amounts of storage
- You also need enormous amounts of processing capability
- Both vast amounts of storage and processing power are available in the cloud

Many people conflate the terms "artificial intelligence" and "machine learning." To be accurate in our terminology, *machine learning* is a subset of the larger field of artificial intelligence in which we use large amounts of historical data to train models, which we then use to evaluate new data. Most of what follows focuses on the machine learning discipline within artificial intelligence.

Machine learning basics

The concepts of artificial intelligence and machine learning certainly carry with them a bit of mystique, but in fact the concepts are relatively straightforward. What follows is a brief primer into the art and science of machine learning.

Imagine that you want a program that can predict when a given part in a machine will fail. You have terabytes and terabytes of log files, and you have a list of the parts that have failed. Each record in the log notes the part, its unique identifier, the supplier, when it was manufactured, when it was purchased, when it was last serviced, by whom it was last serviced, the identifier of the machine in which the part is installed and 50 other data items.

Your task is to determine, given a part identifier, when it will fail.

A machine learning application can perhaps find the answer. The first step *splits* your historical data. The idea is that you'll *train* a model with part of the data, and then see if the model correctly predicts the results in the other part of the data.

The next step, training, is the most processor-intensive part of machine learning because it takes large quantities of data and builds mathematical correlations between all of the variables. A wide variety of such mathematical algorithms exist and are widely available; it's often useful to train with multiple algorithms and see which yields the best result.

In the final step, you take the trained model and feed it actual raw, untrained data and then see if it accurately predicts the results – that is, which parts in the untrained data (the log) actually failed – this is called "scoring." In our failure prediction example, this would take each part identifier in the untrained data and return a probability of failure and then compare that against the actual failure rate. If the results are satisfactory – for instance, higher than 95 % – you can deploy the trained model in the real world.

Perhaps in a new application you feed the logging data into your trained model and generate an alert when a part is nearing failure. Then, you can proactively send a service representative to replace the part, improving customer satisfaction and lowering your own costs.

And that is the essence, albeit just a hint, of the value of machine learning.

Figure 13-1 illustrates what developing that model looks like in the Microsoft Azure Machine Learning Studio. Note that it is a simple graphical workflow; in this case the algorithm used is a linear regression.

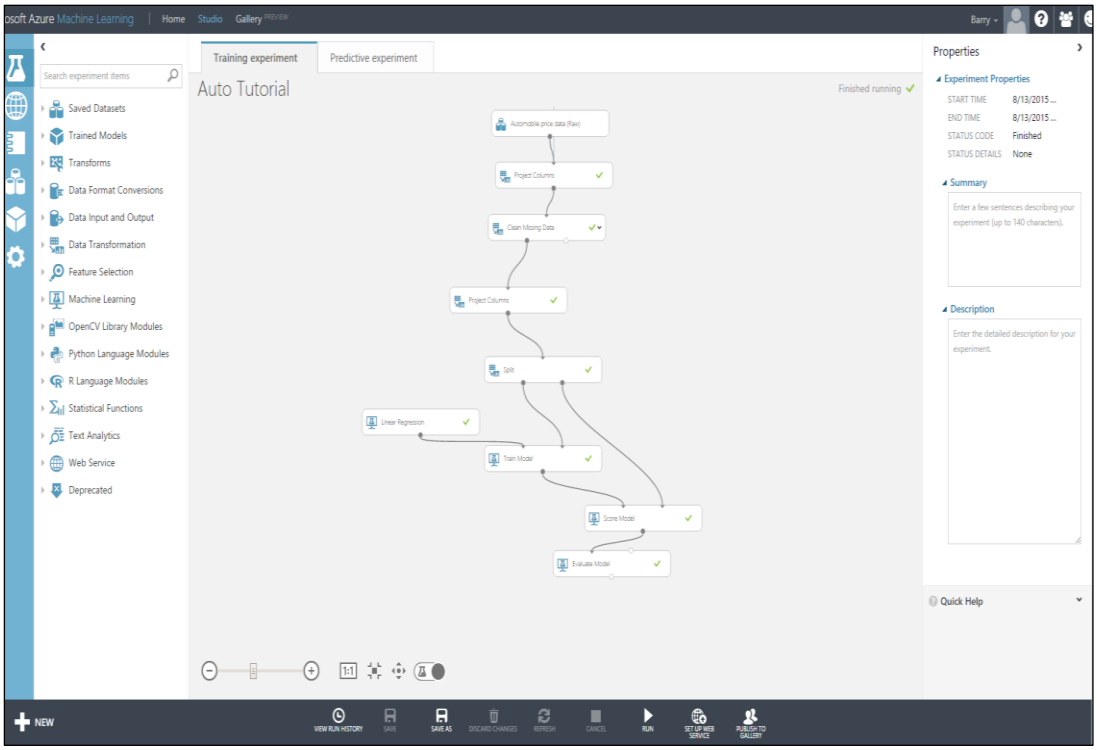


Figure 13-1: Azure Machine Learning

At this point, it's worth passing along a pearl of wisdom that every data scientist knows. Usually the most difficult part of building a machine learning algorithm such as that shown in Figure 13-1 is actually ensuring that the data is "clean", in other words, that in all the terabytes there are legal values: no out-of-bounds data and so on. (This is a problem familiar to anyone who has built a large-scale Extract, Transform and Load [ETL] process.) In the model in the figure, there is a step called "Clear missing data" – one of many built-in data cleansing capabilities.

Supervised versus unsupervised learning

Machine learning distinguishes between so-called *supervised* learning and *unsupervised* learning. In our example, we knew which parts failed, so we taught our model to find which combinations and correlations of variables would accurately predict a failure. We then scored our model based on our pre-existing knowledge and retrained it with a larger dataset, or adjusted some parameters, or picked a different algorithm if the results were not to our satisfaction. This is an example of supervised learning.

In unsupervised learning, we don't have a specific goal in mind, such as determining the probability of a part failure. Rather, we are looking for patterns in the data. There are two primary problems solved by unsupervised learning. The first is *clustering*, that is, finding groupings in large datasets that you didn't know existed previously, such as finding customers grouped by purchasing behaviour. The second is *association*, or discovering previously unknown relationships between data items in large datasets. An example of this is determining that customers who buy a given product also tend to buy another related product.

Neural networks

Neural networks are another of the mathematical algorithms that we can use in machine learning. A neural network, as the name implies, emulates the operations of cells in the brain. Essentially the idea is that a small body of code analyses a piece of the data, giving a probability that a certain condition has been met. These neurons then feed into other neurons, analysing larger and larger chunks of data, ultimately determining whether a certain condition has been recognised. Neural networks are extensively used in sensory applications such as speech recognition and computer vision.

The term *deep learning* has been applied to neural networks with many layers.

Accelerating machine learning with hardware

As we mentioned a moment ago, the most processor-intensive part of machine learning is the training step. Many of the training algorithms can benefit from *parallelism* – that is, many processors running simultaneously, operating on different elements of the data.

In fact, the more processors the better, and, as it turns out, high-end graphics processing units (GPUs) typically sport hundreds or thousands of highly interconnected processors. GPU acceleration can speed up machine learning by an order of magnitude – sometimes more. Neural networks map well to GPUs both for training and for runtime operation (inferencing) because a software "neuron" can be loaded to a single processor in the GPU.

However, GPUs have disadvantages: they tend to be expensive and they consume a lot of power. And, of course, GPUs were not originally designed specifically for machine learning applications but rather for graphics (although some vendors have removed the graphic circuitry).

Two new emerging approaches seek to provide machine-learning specific hardware. Several of the cloud providers, including Microsoft, have announced the availability of custom Field Programmable Logic Arrays (FPGAs) in their cloud datacentres. You can configure FPGAs, which are programmable integrated circuits, in a wide variety of ways to run machine learning applications, and, like GPUs, they feature parallel processing and thus much higher speeds.

An even newer technology embeds neural networks in silicon, actually creating chips composed of silicon neurons. These *neuromorphic* chips are still in their infancy (and are not broadly available yet) but show great promise.

Applications of AI and machine learning

As we pointed out early in this chapter, we believe AI and machine learning technologies are revolutionising business models. In this section, we illustrate this point with some examples.

Bots and the conversational computer

Machine learning provides applications with a means to understand natural language – that is, the way we use language to actually communicate – and that in turn gives us a way to interact with computers in a more human way. Applications that communicate with humans using language are called *bots*, and they make possible new forms of customer relationships and intimacy.

By connecting bots to corporate data sources, companies are finding many ways to use them. For example, for large sites with tens of thousands of pages, it's easy for a user to just ask a bot where to find the information needed ("How do I register a child for school?" or "Where do I pay my bill?"), thus bypassing all the sometimes-Byzantine navigation that characterises many modern websites.

Predictive analytics

In our simple machine learning example, we looked at large quantities of log data to predict part failures. The value of this sort of foreknowledge to companies is enormous: by preventing unexpected downtime, companies can operate at full capacity, increasing customer satisfaction.

In fact, some companies have used predictive analytics to change their business models altogether. Instead of selling capital equipment, they sell what might be called "equipment-as-a-service," meaning they charge a subscription fee for the device's uptime, using telemetry and predictive analytics to forestall any failures.

The British manufacturer Rolls-Royce, for example, analyses telemetry data from jet engines it sells using Microsoft's Cortana Intelligence Suite. By finding anomalies early, it can ensure that the engines are always at peak performance – and that flights are on time.

Autonomous things

Modern automobiles are being fitted with a wide variety of sensors, including LIDAR and video sensors. The goal over time is to make fully autonomous driving possible, in which the car itself has situational awareness of the road, its conditions, traffic and so on. The autonomous car – and there are projects centred around other autonomous things such as trucks, boats, drones and so on – bases its decisions on machine learning. Computer vision can recognise objects such as lane markers, speed-limit signs and other cars on the road. Other sensors inside the car can ensure that the driver remains awake and alert even if the car is in control. And the car is constantly receiving traffic reports so that it can intelligently reroute if congestion is detected ahead.

Fraud detection and other financial applications

Financial services have found many applications for artificial intelligence and machine learning, as might be expected given the volume of data generated by financial transactions. Fraud detection programs use machine learning techniques to pore over vast quantities of data looking for anomalous patterns, for example. Other machine learning–based programs can analyse details of financial reports such as 10-Ks, compare them against previous time periods or against competitors' reports to generate investment recommendations.

One of the most interesting uses of AI in financial services is the growing field of "robo-advisors", that is, AI-driven financial advisors that examine a user's portfolio and apply algorithms to make investment suggestions. Operating at a small fraction of the cost of a human advisor, robo-advisors can provide management for retirement funds such as a 401(k) and IRA as well as taxable investment accounts. They can also give automated advice on retirement savings and other investments.

Healthcare applications

In healthcare, AI finds application in a great many places. Machine learning applications can quickly scan through vast bodies of patient medical histories to provide customised treatment. By analysing a single patient's records and then comparing them to legions of other records (with the appropriate privacy protections required by law), individual treatment plans for serious conditions, including cancer, can be created. Pharmaceutical companies are using machine learning algorithms to develop new drugs quickly, sometimes avoiding laborious and time-consuming trial-and-error approaches of the past (a company called Atomwise used AI to find two drugs that have great potential for treating the Ebola virus, for example).

Summary

We hope that we have demonstrated that artificial intelligence applications running in the cloud have the potential to revolutionise all aspects of enterprise computing. It's still very early, yet already AI has transformed, or is in the process of transforming, businesses across the planet.

Summary

Microsoft

Summary

We closed the first edition of *Enterprise Cloud Strategy* with the statement, "the impact of cloud computing on the enterprise, and on business generally, cannot be overestimated." We believe that the two years between the first edition and this one have certainly proven that!

So, what have we learned?

In Chapter 1, we talked about what might be called the macroeconomics of the cloud and showed how cloud vendors can achieve economies of scale no single enterprise can. We also described some of the many opportunities for cost savings and reducing TCO in the cloud, and the opportunities for realising measurable ROI using PaaS services.

We began our cloud journey by providing definitions of key terms – IaaS, SaaS, PaaS and containers – in Chapter 2 and discussing the pros and cons of each. A key theme of the book, first mentioned in Chapter 2 and then elsewhere, is that these different approaches to cloud migration have markedly different returns: if you can replace an on-premises application with a SaaS service, you've removed tremendous expense. IaaS applications remove the burden of managing infrastructure like servers and networks, and the PaaS model goes one step further as the cloud provider is also responsible for operating system, database and other system software maintenance.

We then described the journey in three phases: experimentation, migration and transformation, being careful to point out that you might well – and should – consider doing these in parallel, rather than thinking of them as sequential steps.

In the experimentation phase, your teams move a few low-risk applications to the cloud, with the goal of learning how to engineer for the cloud, how to manage cloud operations and how to take advantage of features only available in the cloud. We recommended a best practice, which is to not limit this experiment to a simple migration, but rather to take full advantage of the opportunity and try as many cloud features as possible.

As we mentioned, the most complex phase is migration, in which the bulk of the IT application ecosystem is moved to the cloud. In Chapter 5, we provided some examples and best practices of how to organise and govern cloud migration, beginning with the creation of a cloud strategy team, responsible for setting priorities and managing the migration. You'll probably want to refer to the sections on organisational impact from time to time, because nearly every function in enterprise IT – from HR to finance to development to operations – is in some way affected by cloud migration.

In Chapter 6, we described the process of prioritising your application portfolio, and the importance of setting criteria by which prioritisation decisions are made, and we provided some examples based on our experience of how you can do this. Then, we described how to carry out your plan, described some of the tools available to help you migrate, and showed some best practices for how you can achieve financial transparency through thoughtful cloud subscription management.

As you move your applications to the cloud, and as your business begins to demand ever-faster releases, you'll increase your use of DevOps tools and methodologies, which apply lean manufacturing techniques to software. Using tools like Microsoft Azure Application Insights and Azure Monitor, you can monitor the resources your application is using moment-by-moment to ensure that you're getting the most for your money.

In Chapter 9, we discussed two of the most important areas in the cloud: governance and security, and provided a number of recommendations and best practices. With respect to regulatory compliance, as we said, it's important to understand which certifications your cloud provider has obtained, and what the division of responsibilities is between your applications and data on the one hand and the cloud datacentre, infrastructure and platform software on the other.

With respect to security, a wide array of tools and methodologies have been created to ensure your applications and data are safe in the cloud. You should absolutely be using them, and check back frequently to the Azure Trust Centre to confirm you are up-to-date, given that new threats appear frequently.

In Part III, *Transformation*, we talked about how you can use the cloud and the services available in the cloud to not only reduce costs and gain efficiency, but radically expand the capabilities of your application portfolio and drive great returns to your business.

We began by showing how you can integrate your on-premises applications with capabilities and services available in the cloud, such as backup and restore, disaster recovery and application integration.

In Chapter 11, we looked in some detail at new application models available in the cloud, including PaaS and containers. "Serverless" capabilities make it possible for your teams to create new applications without writing a line of code!

The cloud also makes a wide variety of data management tools available to you as services: not just relational databases and data warehouses, but also document databases, big data repositories such as Hadoop and Spark, and analysis services.

Finally, in Chapter 13, we showed how the emerging worlds of artificial intelligence and machine learning can, for very little cost, provide transformational benefits to your applications and businesses by – at a stroke – enabling capabilities like bots and predictive maintenance.

* * *

We believe the cloud represents an incredibly exciting set of opportunities for both IT organisations and enterprises at large. We hope that by reading this document you have learned how you can quickly realise all of the benefits of the cloud.

Enjoy your cloud journey!

Cloud architectural blueprints

We encourage you to visit the www.azure.com/solutions website to view the latest cloud solution architecture solution blueprints. We have included some in this appendix to illustrate the potential of the cloud architectures that we have discussed throughout this book.

Digital Marketing

Simple digital marketing website

Figure A-1 presents an example of a simple content management system with which you can easily maintain the messaging on your website in real time, from a browser, with no coding skills.

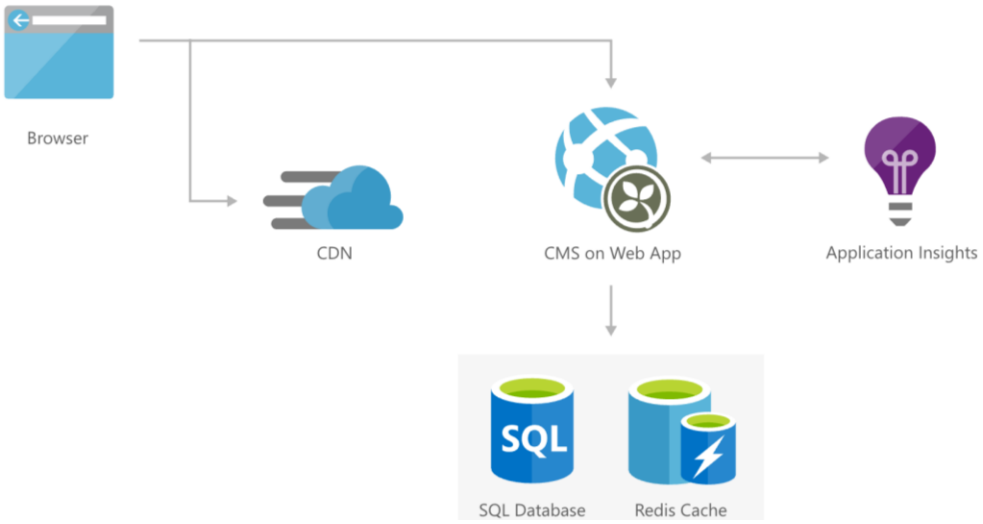


Figure A-1: Simple digital marketing website

This solution is built on the following Microsoft Azure managed services:

- SQL Database
- Application Insights
- Content Delivery Network
- Redis Cache

As with all of the services presented in this appendix, they run in a high-availability environment, patched and supported, making it possible for you to focus on your solution instead of the environment in which it runs.

Scalable Umbraco CMS web app

A larger marketing site uses the open-source Umbraco content management system web application, Figure A-2 shows that it is configured to scale and optimised for high-traffic sites. It uses two web apps, one for your front-end app and the other for your back-office app, deployed in a single region with autoscaling turned on.

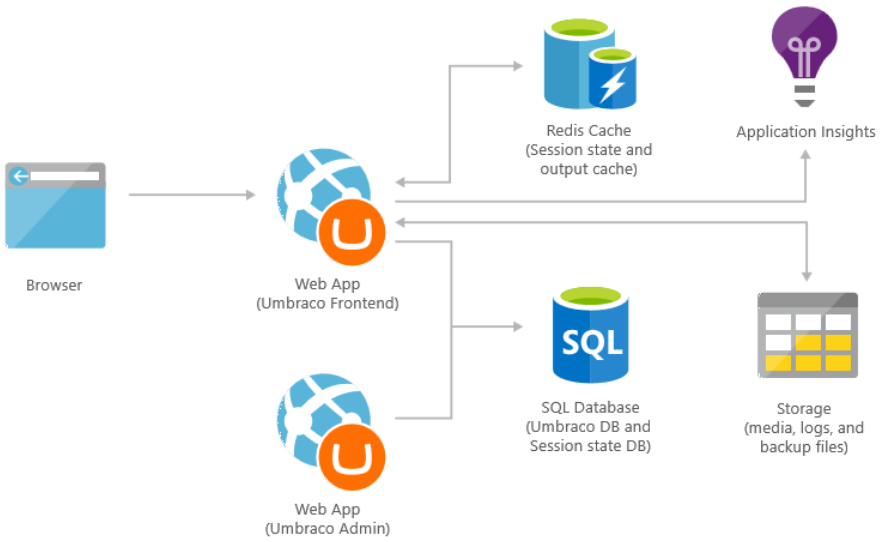


Figure A-2: Scalable CMS web application

This solution is built on the following Azure managed services:

- SQL Database
- Storage
- Application Insights
- Redis Cache

Mobile

Task-based consumer mobile app

Figure A-3 presents an example of a mobile backend that is used by iOS, Android and Windows client apps. You can use Xamarin or native client SDKs to build a mobile client app with offline sync support, including offline sync of image files. App Service Authentication is used to connect to an identity provider, and Azure Blob storage is used to store images in a cost-effective and scalable way.

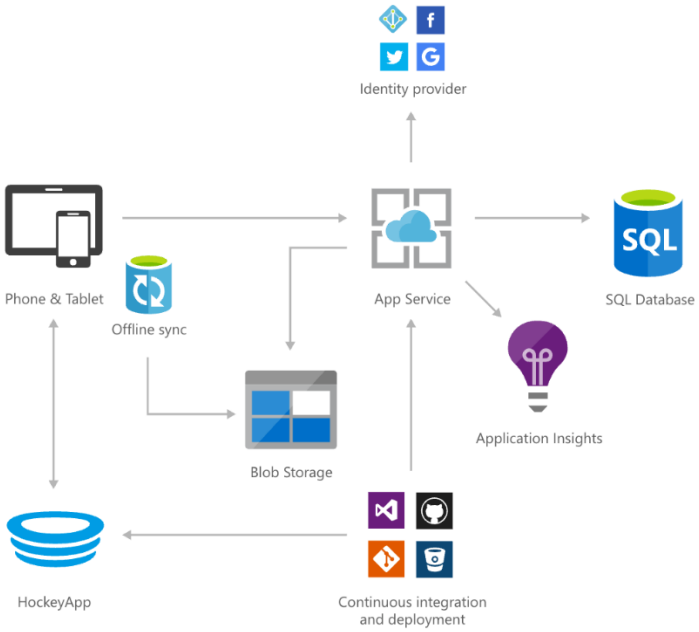


Figure A-3: Task-based consumer mobile app

This solution is built on the following Azure managed services:

- App Services
- SQL Database
- Application Insights
- HockeyApp

Custom mobile workforce app

In this example (Figure A-4), a Xamarin.Forms client app with support for iOS, Android and Windows works offline and allows field engineers to view and edit the jobs assigned to them.

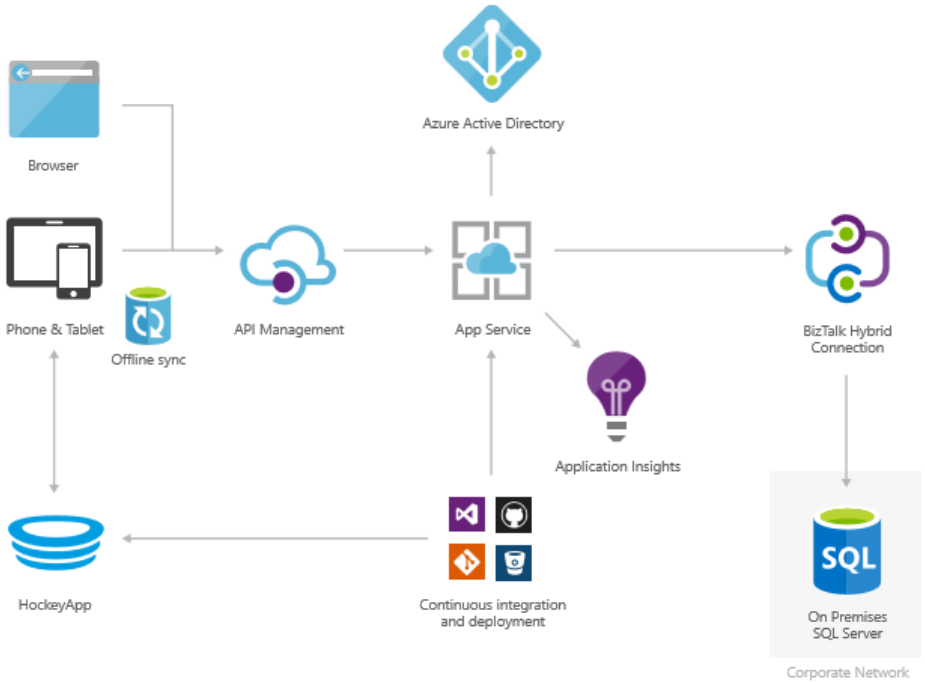


Figure A-4: Custom mobile workforce application

This solution is built on the following Azure managed services:

- App Services
- API Management
- SQL Database
- Azure Active Directory
- Application Insights
- HockeyApp

Social mobile and web app with authentication

Figure A-5 shows a mobile client app for social image sharing with a companion web app. The app backend does background image processing using an Azure Function. The mobile client app works in offline mode, giving you the ability to view and upload images even when you don't have a network connection.

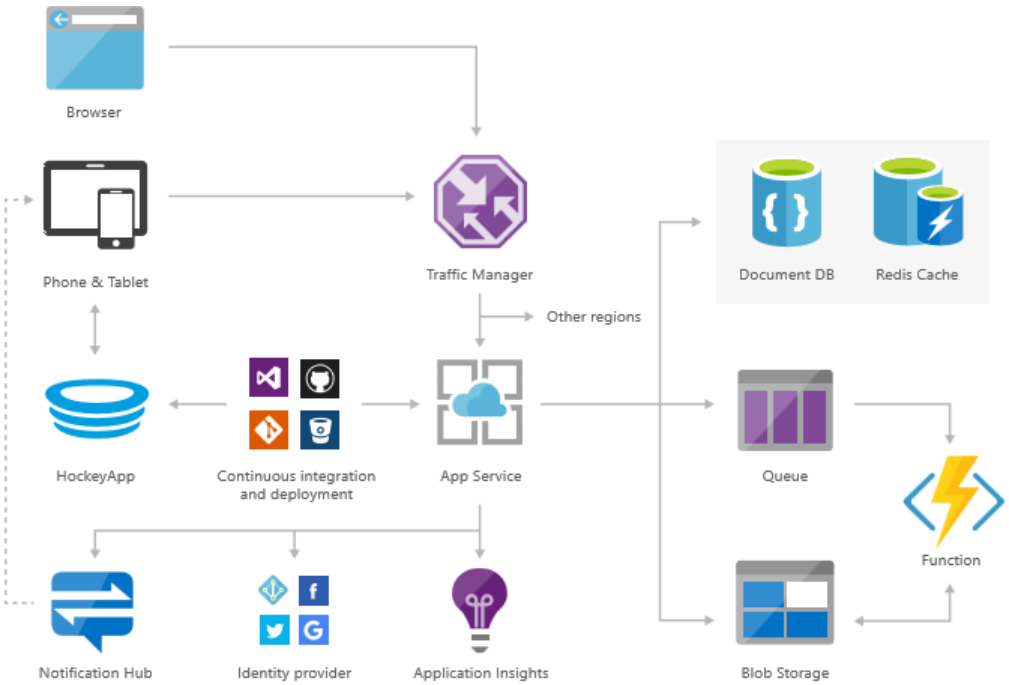


Figure A-5: Social mobile and web app with authentication

This solution is built on the following Azure managed services:

- App Services
- Traffic Manager
- Azure Cosmos DB
- Redis Cache
- Notification Hubs
- Azure Active Directory
- Functions
- Application Insights
- HockeyApp

Backup and archive

Figure A-6 illustrates how to back up data and applications from an on-premises system to Azure using Azure Backup or a partner solution. An Internet connection to Azure is used to connect to Azure Backup or Azure Blob storage. Azure Backup Server can write backups directly to Azure Backup. Alternatively, a partner solution such as Commvault Simpana or Veeam Availability Suite, hosted on-premises, can write backups to Blob storage directly or via a cloud endpoint such as Veeam Cloud Connect.

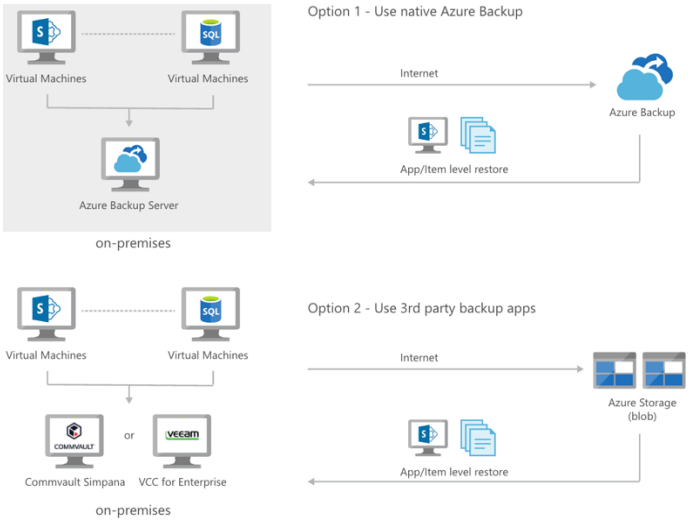


Figure A-6: Backup and archive

This solution is built on the following Azure managed services:

- Backup Server
- Backup
- Blob storage

Development and testing

Development and testing for IaaS

Figure A-7 demonstrates how to configure your infrastructure for the development and testing of a standard IaaS-based SaaS system.

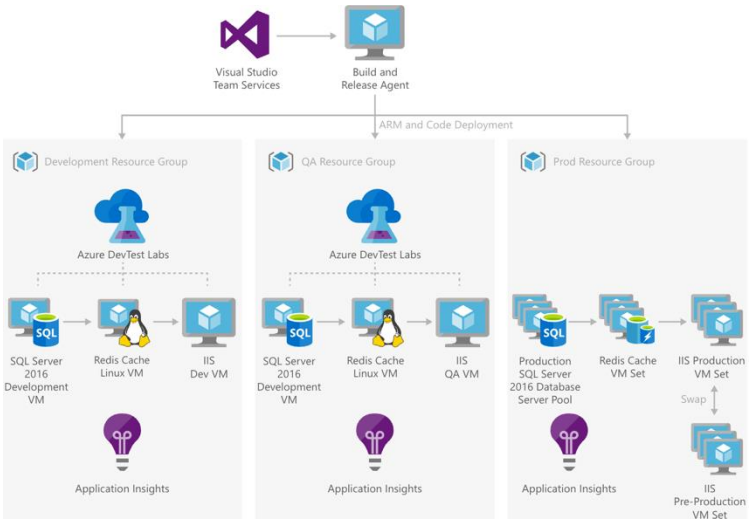


Figure A-7: Development and testing for IaaS

This solution is built on the following Azure managed services:

- Visual Studio Team Services
- Azure DevTest Labs
- Virtual Machines
- Application Insights

Development and testing for PaaS

Figure A-8 depicts how to configure your infrastructure for the development and testing of a standard PaaS-style system.

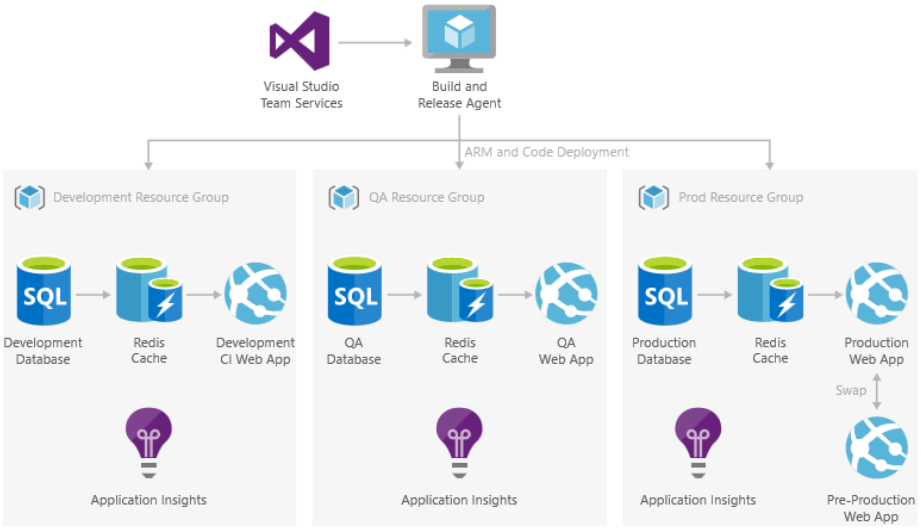


Figure A-8: Development and testing for PaaS

This solution is built on the following Azure managed services:

- Visual Studio Team Services
- SQL Database
- Redis Cache
- Application Insights

Development and testing for microservice solutions

Figure A-9 shows how to configure your infrastructure for the development and testing of a microservice-based system.

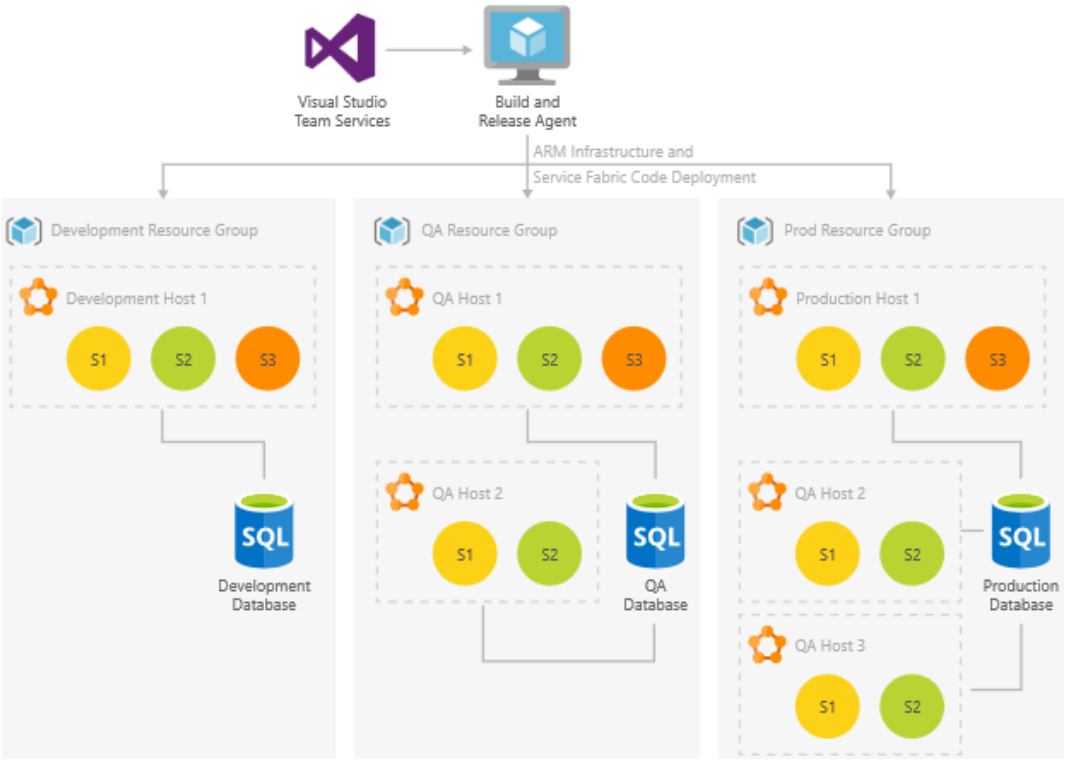


Figure A-9: Development and testing for microservice solutions

This solution is built on the following Azure managed services:

- Visual Studio Team Services
- Service Fabric
- SQL Database

Disaster recovery

Enterprise-scale disaster recovery

Figure A-10 presents a large enterprise architecture for Microsoft SharePoint, Dynamics CRM and Linux web servers hosted in an on-premises datacentre with failover to Azure infrastructure.

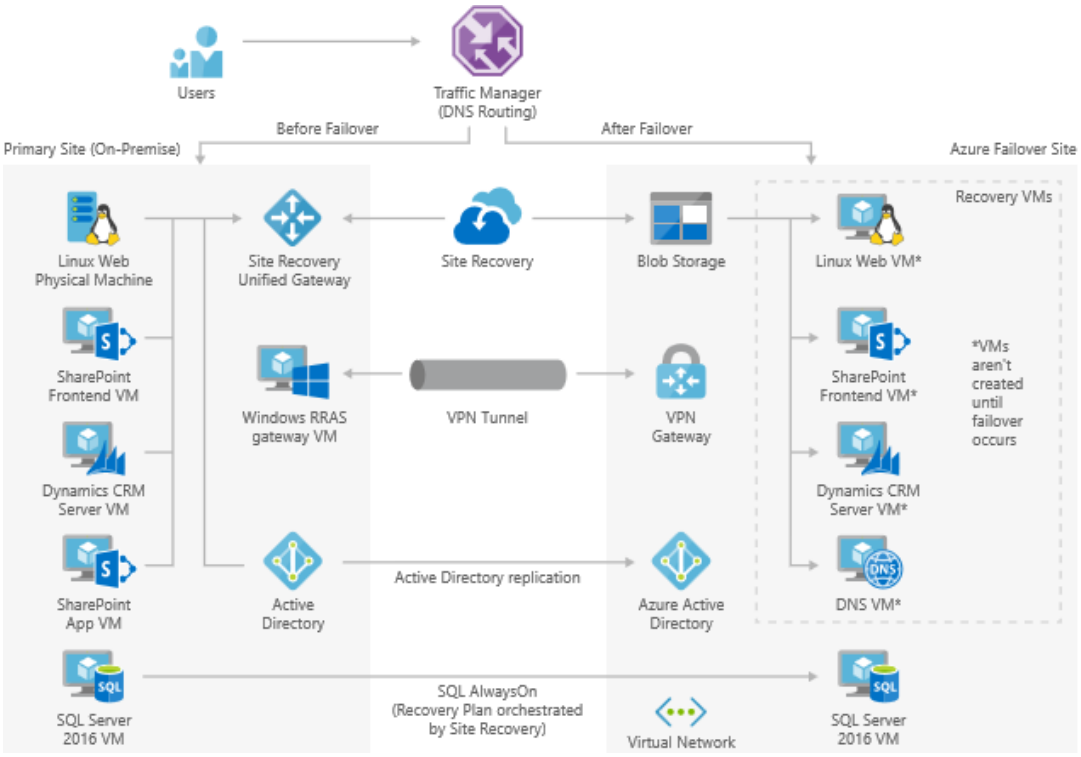


Figure A-10: Enterprise-scale disaster recovery

This solution is built on the following Azure managed services:

- Traffic Manager
- Site Recovery
- Azure Active Directory
- VPN Gateway
- Virtual Network

SMB disaster recovery with Azure Site Recovery

For small and medium businesses, you can implement disaster recovery inexpensively in the cloud by using Azure Site Recovery or a partner solution like Double-Take DR. Figure A-11 illustrates the Site Recovery solution.

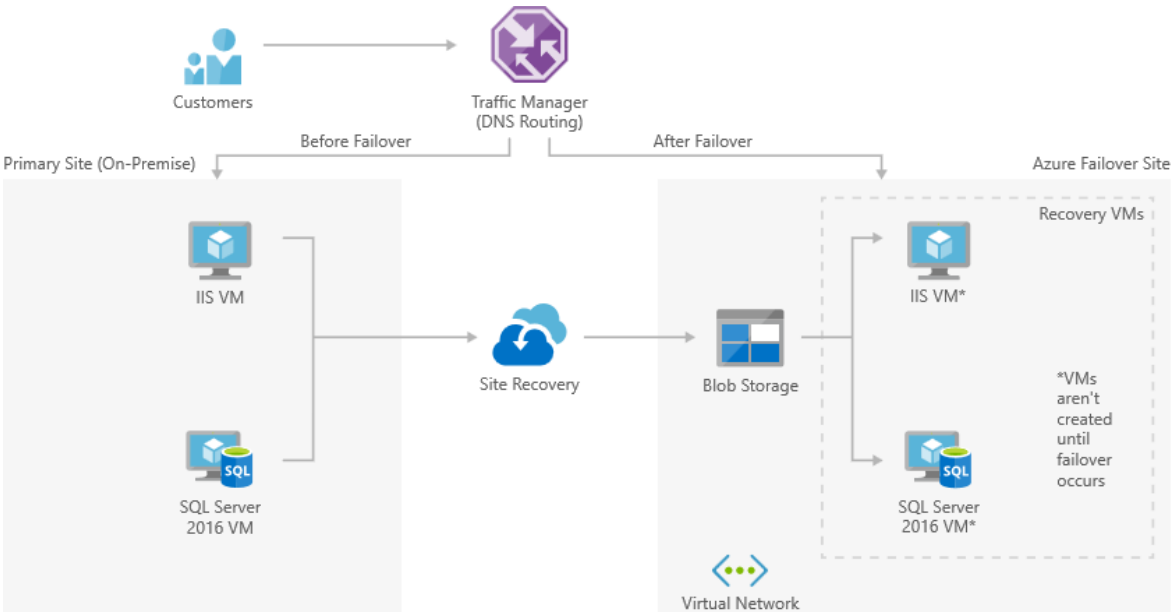


Figure A-11: SMB disaster recovery with Azure Site Recovery

This solution is built on the following Azure managed services:

- Traffic Manager
- Site Recovery
- Virtual Network

SAP on Azure

SAP HANA for Azure

The architecture depicted in Figure A-12 represents a distributed three-tier SAP system running on SQL Server in the Azure cloud platform.

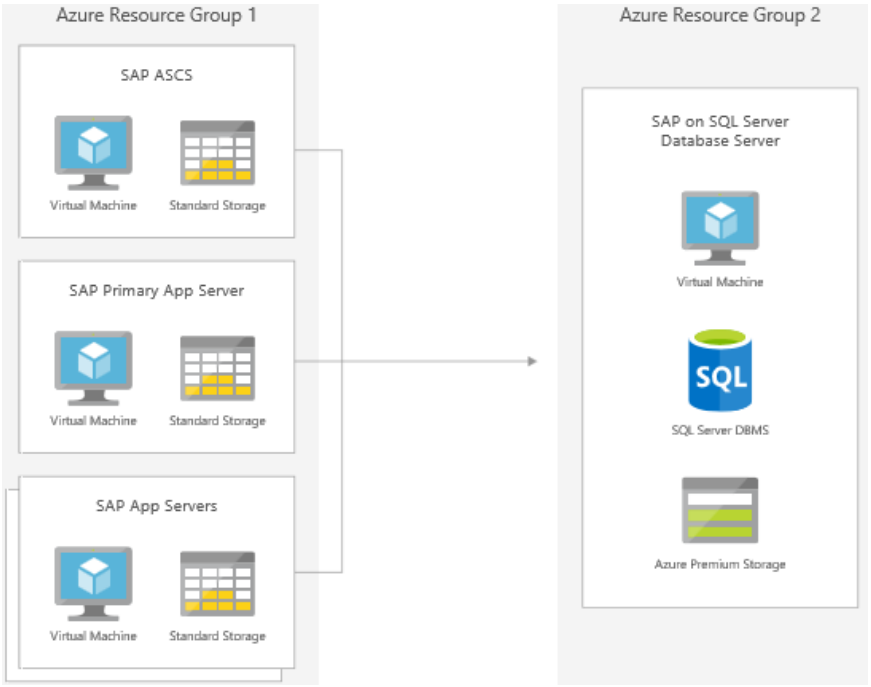


Figure A-12: SAP Hana for Azure

This solution is built on the following Azure managed services:

- Virtual Machines
- Storage

SAP Hana on Azure (Large Instance) architecture

The diagram in Figure A-13 demonstrates how to configure your infrastructure to run SAP HANA on Azure (Large Instance) that includes the Application Tier in an Azure Datacentre and HANA in the Large Instance Datacentre.

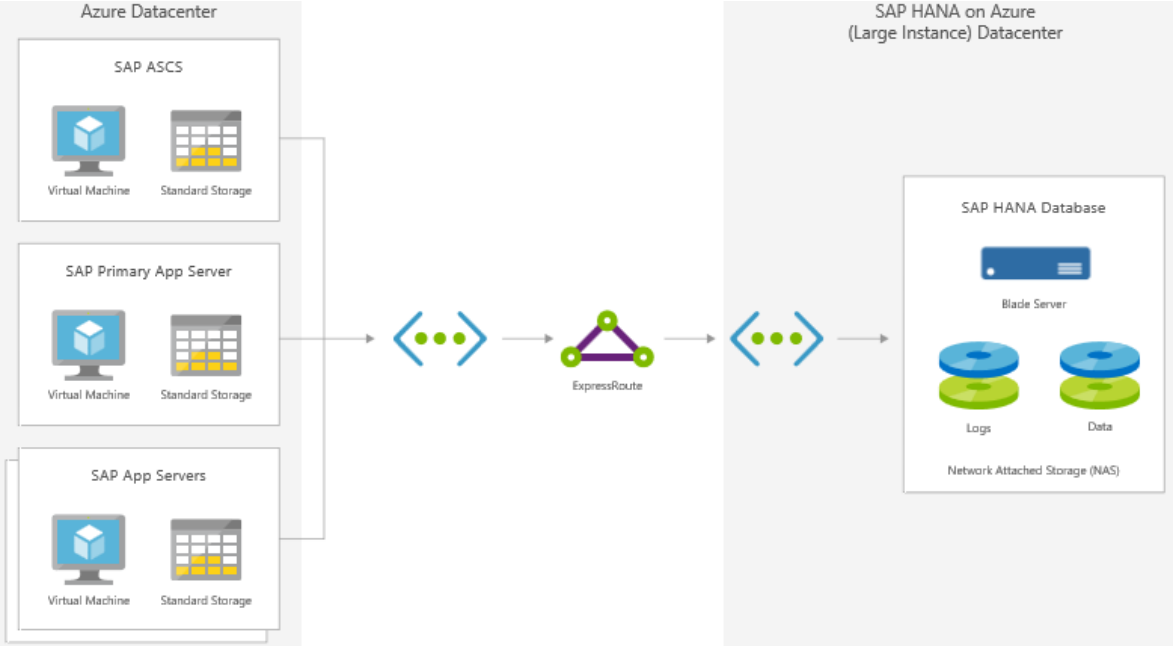


Figure A-13: SAP Hana for Azure (large instance)

This solution is built on the following Azure managed services:

- Virtual Machines
- Storage
- ExpressRoute
- Virtual Network

High performance computing

Big computing solutions as a service

High-performance computing (HPC) applications can scale to thousands of compute cores, extend on-premises big computing or run as a 100 % cloud-native solution. The HPC solution shown in Figure A-14 is implemented using Azure Batch, which provides job scheduling, autoscaling of compute resources and execution management as a platform service (PaaS) that reduces HPC infrastructure code and maintenance.

This solution is built on the following Azure managed services:

- Virtual Machines
- Storage
- Batch

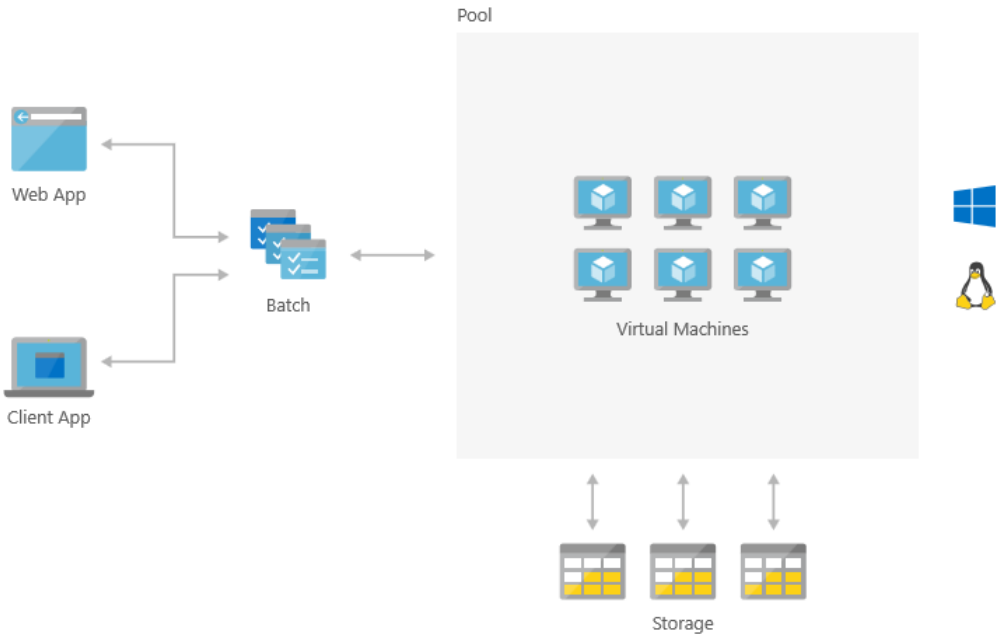


Figure A-14: Big computing solutions as a service

HPC cluster deployed in the cloud

The HPC solution shown in Figure A-15 includes the head node, computing nodes and storage nodes, and runs in Azure with no hardware infrastructure to maintain.

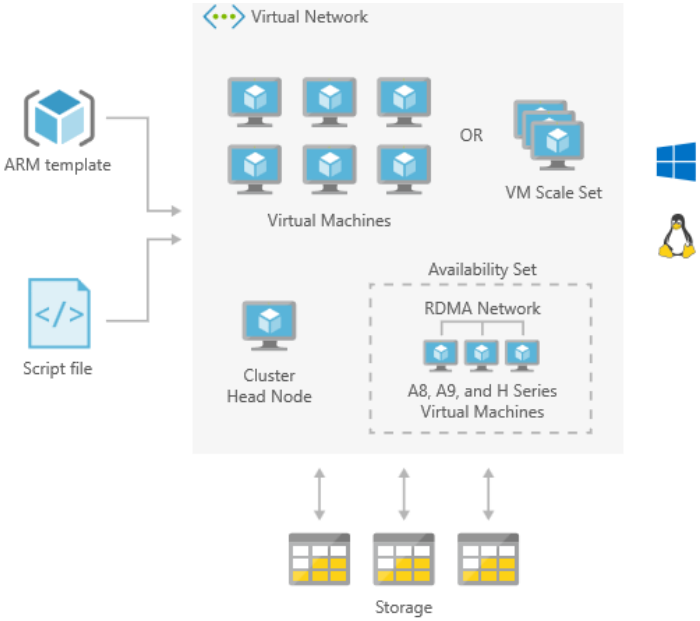


Figure A-15: HPC cluster deployed in the cloud

This solution is built on the following Azure managed services:

- Virtual Machine Scale Sets
- Virtual Network
- Storage

On-premises HPC implementation bursting to Azure

Lastly, an HPC solution can extend its computational capacity by using the computing-intensive instances of Virtual Machines running in Azure and accessed via ExpressRoute or VPN, as illustrated in Figure A-16.

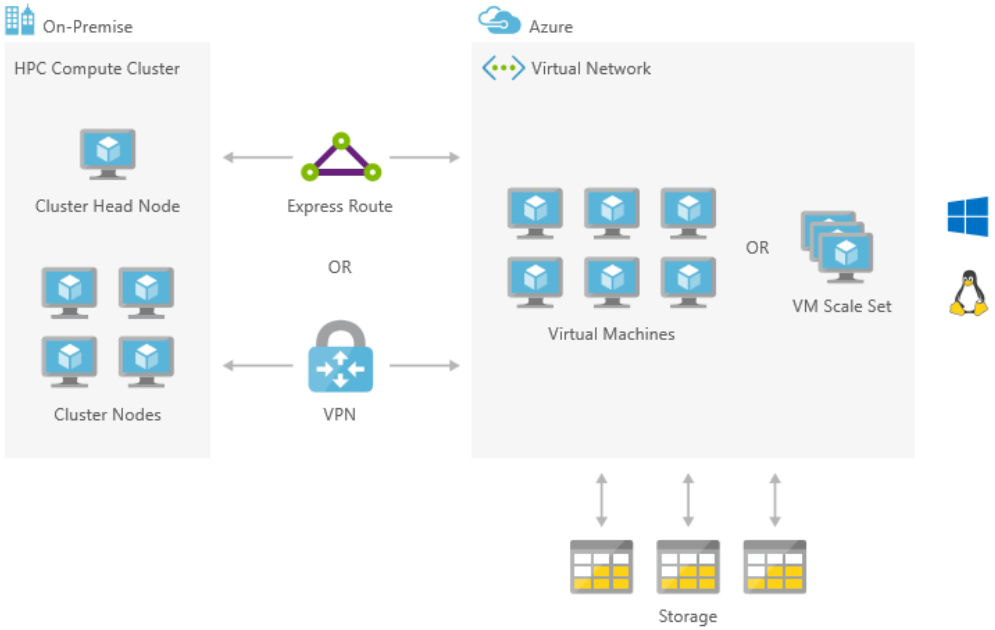


Figure A-16: On-premises HPC implementation bursting to Azure

This solution is built on the following Azure managed services:

- Virtual Machines
- Virtual Network
- VPN Gateway
- ExpressRoute
- Storage

Digital media

Video-on-demand digital media

Figure A-17 shows a basic video-on-demand solution that gives you the capability to stream recorded video content such as movies, news clips, sports segments, training videos and customer support tutorials to any video-capable endpoint device, mobile application, or desktop browser. Video files are uploaded to Azure Blob storage, encoded to a multi-bitrate standard format, and then distributed via all major adaptive bit-rate streaming protocols (HLS, MPEG-DASH, Smooth) to the Azure Media Player client.

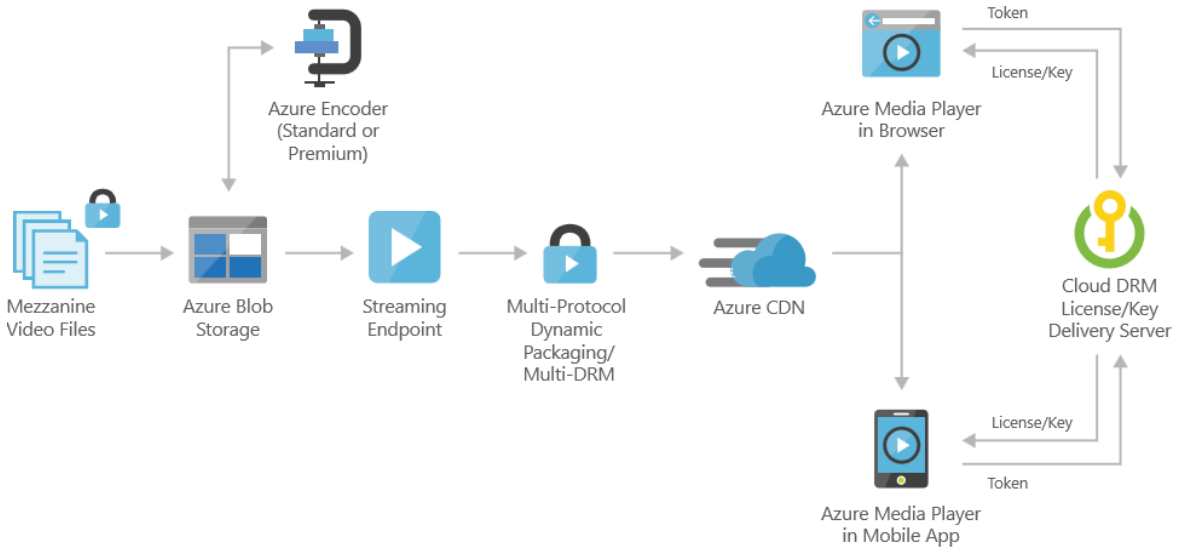


Figure A-17: Video-on-demand digital media

This solution is built on the following Azure managed services:

- Blob storage
- Content Delivery Network
- Azure Media Player

Live streaming digital media

A live streaming solution makes it possible for you to capture video and broadcast it to consumers in real time, such as streaming interviews, conferences and sporting events online. In the solution presented in Figure A-18, video is captured by a video camera and sent to a channel input endpoint. The channel receives the live input stream and makes it available for streaming through a streaming endpoint to a web browser or mobile app. The channel also provides a preview monitoring endpoint to preview and validate your stream before further processing and delivery. The channel can also record and store the ingested content in order to stream it later (video-on-demand).

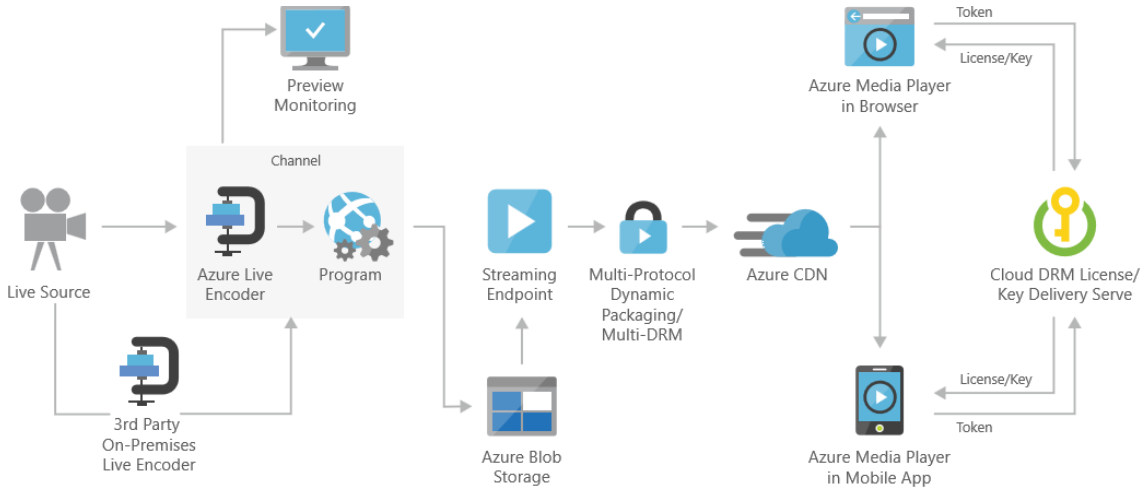


Figure A-18: Live streaming digital media

This solution is built on the following Azure managed services:

- Media Services
- Content Delivery Network

Keyword search/speech-to-text/OCR digital media

With a speech-to-text solution, you can identify speech in static video files so that you can manage it as standard content, such as allowing employees to search within training videos for spoken words or phrases and then enabling them to quickly navigate to the specific moment in the video. The solution in Figure A-19 gives you the ability to upload static videos to an Azure website. The Azure Media Indexer uses the Speech API to index the speech within the videos and stores it in SQL Azure. You can search for words or phrases and get a list of results by using Azure Web Apps. When you select a result, you can see where in the video the word or phrase is mentioned.

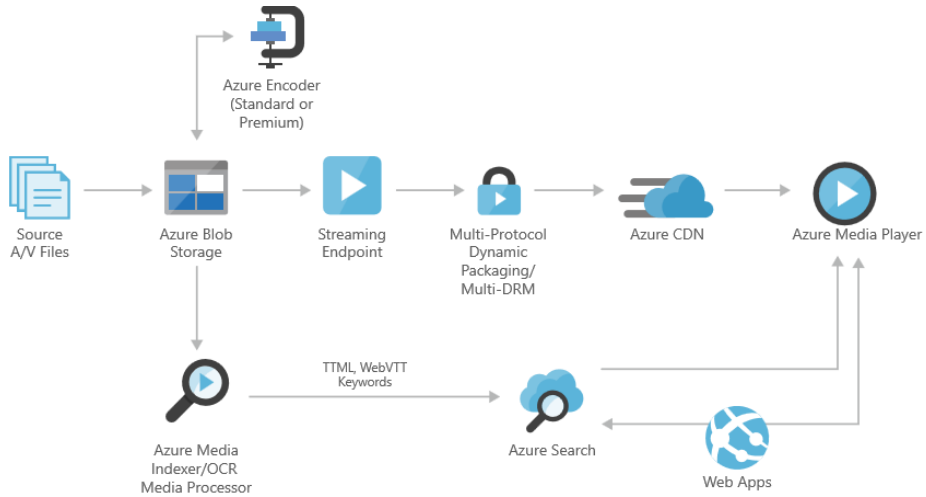


Figure A-19: Video-on-demand digital media

This solution is built on the following Azure managed services:

- Content Delivery Network
- Azure Search

E-commerce

Before you can sell it, people must want to buy it. With Microsoft's e-commerce platform, you can analyse site traffic and browse-to-buy conversion rates to define special offers and new products based on customer behaviour. Create personalised shopping experiences with targeted content and offers, and increase satisfaction through ongoing engagement – before, after and at the point of sale.

More customers means more transactions. Ensure that you're ready to handle every transaction smoothly by designing an e-commerce purchasing experience that's simple to navigate. Then deploy it to a secured and compliant e-commerce platform.

You need an e-commerce solution that adapts to the size and seasonality of your business. When demand for your products or services takes off – predictably or unpredictably – be prepared to handle more customers and more transactions automatically. Plus, take advantage of cloud economics by paying only for the capacity you use.

Your core business is selling your products, not being an IT organisation. Take advantage of prebuilt services, such as those shown in Figure A-20, in the cloud to create an e-commerce solution that enhances your sales performance and leaves the infrastructure management to your cloud provider.

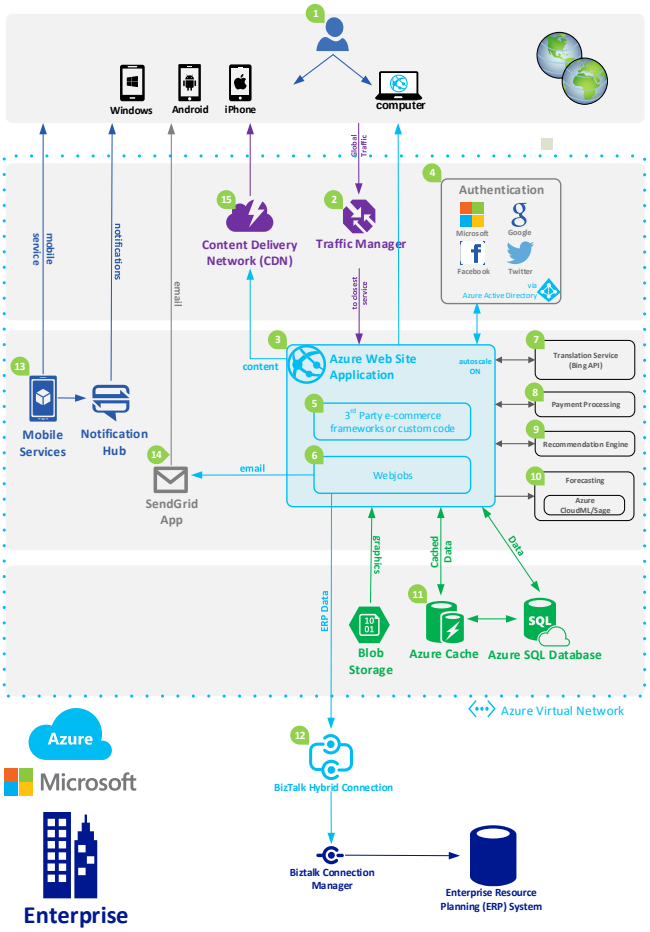


Figure A-20: E-commerce solution

As shown in the illustration:

1. Users browse and order items from phones, tablets and PCs by using HTML or native applications.
2. Deploy to multiple datacentres for global scale and use Azure Traffic Manager to route requests to the nearest one.
3. Azure Websites scale up and down automatically to manage spikes in customer shopping patterns.
4. Users log in to Azure Active Directory using credentials from Facebook, Google, Microsoft, Twitter or other identity providers.
5. Save time by using third-party commerce frameworks or your own.
6. WebJobs runs in the background, both submitting orders to the on-premises Enterprise Resource Planning (ERP) system and sending order confirmations.
7. Create a global website by using the translation service provided by Bing.
8. Azure is Payment Card Industry Data Security Standard (PCI DSS)–compliant for payment processing.

9. Targeted item recommendations are delivered from a Hadoop-based recommendation engine.
10. Forecast future demand for items by using cloud machine learning.
11. Azure Cache boosts performance of all data services.
12. Use Azure Hybrid Connections to send messages to on-premises databases.
13. Azure Mobile Services provides a unified back end for mobile ordering, including device authentication, data services and notifications.
14. Send e-commerce confirmations using a third-party app.
15. Geo-distributed CDN keeps video and graphic assets closer to users.

The Azure Marketplace offers many prebuilt e-commerce solutions, including, at the time of writing, OpenCart, Virto Commerce, AbanteCard and nopCommerce.

Internet of Things

Using the Azure IoT Hub, you can connect enormous numbers (billions) of Internet of Things (IoT) devices securely. And, like all services in the cloud, you pay only for what you use. The IoT Hub, with software development kits for Windows, Linux and Real-Time Operating Systems (RTOSs) supports a wide variety of protocols including Representational State Transfer (REST), Advanced Message Queuing Support (AMQP), and MQ Telemetry Transport (MQTT). Security is available with optional authentication-per-device and encrypted communications.

For applications needing real-time decision making and on-site artificial intelligence with IoT devices, consider the new IoT Edge service, which colocates AI, analytics and a host of other features with your IoT network.

You can collect untapped data and create predictive models when you connect your devices, assets and sensors to the cloud – from a few sensors to millions of devices. By accessing global production and supply-chain data, you'll reduce costly downtime and maintenance to increase productivity.

You can focus on what matters most to your customers: reliability. Vastly improve operations and asset availability with predictive, and even pre-emptive, maintenance by gathering and transforming data from sensors and systems.

Enhance security across physical devices, connections and data. Use per-device authentication by setting up individual identities and credentials for each of your connected devices, and retain the confidentiality of both cloud-to-device and device-to-cloud messages.

Head off potential problems while promoting equipment efficiency via predictive maintenance. Collect and analyse data from your connected assets to proactively plan maintenance, decrease downtime and improve retention of the asset value.

You can begin innovating today with IoT starter kits from Adafruit, Seeed and SparkFun, who supply development boards that are Azure Certified for IoT.

Microservice applications

In Chapter 12, we discussed applications that are based on the microservices model. With Azure Service Fabric and Azure Container Service, you can get the scale, power and global reach that your start-up demands to meet the needs of your customers as your business grows – without redesigning your applications. Run your apps at cloud scale with a rich set of services that makes it possible for you to focus all of your energy on building applications instead of managing infrastructure.

With microservice architectures, you can update in real-time as small development teams work independently, using continuous delivery pipelines and rolling upgrades to ensure customers always have access to the latest features.

Business intelligence

As we discussed in Chapter 12, the cloud gives you the ability to collect, analyse and visualise massive quantities of data. With Azure SQL Data Warehouse and Azure Analysis Services, you can offer business analysts – and everyone in your organisation – powerful, self-service analytical and business intelligence (BI) tools to drive better, faster decision making. Combine data from multiple sources to build tailored reports and create rich analytics that bring your data to life.

And, with Microsoft Power BI, you can create stunning visualisations of your data. A number of prebuilt Power BI solution templates are available from Microsoft on the Azure website.

Big data and analytics

As we've pointed out often in this book, data volumes are exploding – from traditional point-of-sale systems and e-commerce websites to new customer sentiment sources like Twitter and IoT sensors that stream data in real time using Apache Hadoop and Spark. By analysing a diverse dataset from the start, you'll make more informed decisions that are predictive and holistic rather than reactive and disconnected. You can keep your organisation's data indefinitely, no matter the size. Instead of making cost tradeoffs on what data to hold onto, you can retain your data to meet regulatory and company standards at affordable prices – now possible with Hadoop and Spark technologies and the cloud.

Azure HDInsight provides a managed Hadoop, Spark, R Server, HBase and Storm service. Data Lake Analytics provides massively scalable analytic services and, of course, you can use these tools and others from other vendors (like Cloudera, Datameer and Informatica) with Azure's machine learning capabilities.

Cloud migration

In addition to the advice provided in this book, Azure provides a number of tools to help you with your cloud migration.

As we mentioned earlier, consider using Azure Site Recovery to orchestrate the physical migration of virtual machines (VMs) from on-premises to the cloud. You also can use Azure Site Recovery to move applications from one Azure region to another.

After your applications are in the cloud, the Azure Advisor (Figure A-21) draws on Azure best practices to recommend solutions that will reduce your costs and improve the security, performance and reliability of your applications.

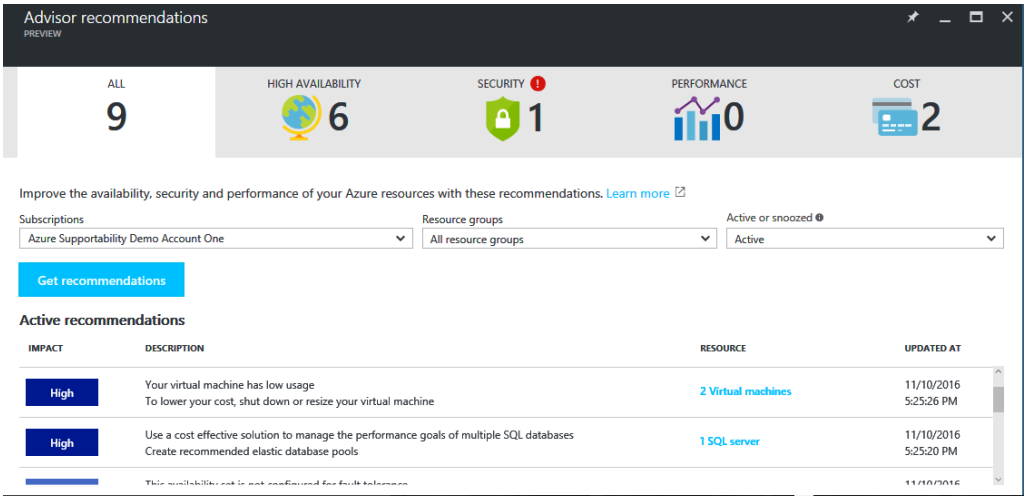


Figure A-21: Azure Advisor screenshot

Data warehouse

With Azure SQL Data Warehouse, you can transform your business through predictive analytics over all of your data with tools you already know and love – Power BI, Microsoft Excel and third-party BI tools. Plus, seamless compatibility with machine learning, ingestion, data movement and data store services ensures transformative insights over all your data.

Azure Data SQL Warehouse, a fully managed solution, can be provisioned in minutes, and as we have emphasised many times about other services, it scales elastically and you pay for what you use.

There are many data migration tools to help you move data into the warehouse, including Azure Data Factory as well as solutions from CloudBeam, BlueTalon and SnapLogic, all availabile in the Azure Marketplace.

Business SaaS apps

Want to create your own SaaS application for use inside your company or to market to your customers? Use Azure SQL Database to isolate data in separate databases to help ensure consistent performance and data security – without adding additional operational or management costs. Plus, use built-in threat detection and Active Directory authentication to help prevent unauthorised access.

You can also get predictable costs to scale your business model along with insights into usage patterns as your business grows. With monitoring tools and machine learning at your fingertips, you'll create the operational environment that your SaaS app requires.

Gaming

Some of the most popular online games today, including those from 343 Industries (the Halo franchise), Next Games, Throwback, Illyriad, Io-Interactive and Xbox Studios run on Azure.

Regardless of the platform you are developing games for – iOS, Android or Windows – use Azure to host your games' back end services, send push notifications and crunch game analytics data to drive user engagement.

With Azure, you can handle the massive scale-out requirements your successful online and social games require while maintaining a seamless experience for gamers without having to worry about downtime or interruptions.

You can build and launch your games with peace of mind. You can rely on Microsoft, which has a long history of building AAA game titles for PCs and consoles. Azure extends this deep experience, bringing enterprise-grade features to your game development efforts.

You can configure multiplayer scenarios and leaderboards with Azure Active Directory. Manage player retention and increase user engagement and monetisation across platforms using Azure Notification Hubs and Azure Media Services. With Notification Hubs, you can send personalised push notifications targeted at individual players or entire audience segments containing millions of users, across all of their devices – including iOS, Android, Windows and Kindle. With Media Services, you have the ability to manage media streaming and even insert video ads into your games.

Whether you have an existing back end infrastructure for your game or are looking to build your system from scratch, Azure provides you with choice and flexibility. If you want to lift-and-shift your infrastructure to the cloud, use IaaS offerings like Virtual Machines and Virtual Machine Scale Sets. With PaaS services like Azure Service Fabric and Azure App Service, you can focus on building your games and let Azure manage your infrastructure. And, you have a choice of storage options, from managed database services like Azure SQL Database and Azure Cosmos DB, to MongoDB, Parse Server and DataStax Cassandra on Azure Marketplace.

Blockchain

Blockchain is an emerging way for businesses, industries and public organisations to almost instantaneously make and verify transactions – streamlining business processes, saving money and reducing the potential for fraud. At its core, a blockchain is a data structure that's used to create a digital transaction ledger that, instead of resting with a single provider, is shared among a distributed network of computers.

The result is a more open, transparent and publicly verifiable system that will fundamentally change the way we think about exchanging value and assets, enforcing contracts and sharing data across industries. The applications of blockchain are almost limitless, ranging from loans, bonds and payments, to more efficient supply chains, even to identity management and verification.

Azure's blockchain as a service (BaaS), available in the Azure Marketplace, provides a rapid, low-cost, low-risk and fail-fast platform for organisations to collaborate together by experimenting with new business processes – backed by a cloud platform with the largest compliance portfolio in the industry.

Line-of-business applications

As we've discussed throughout this book, you can move line-of-business (LoB) applications to Azure easily. Figure A-22 illustrates how to best take advantage of Azure capabilities for IaaS VMs.

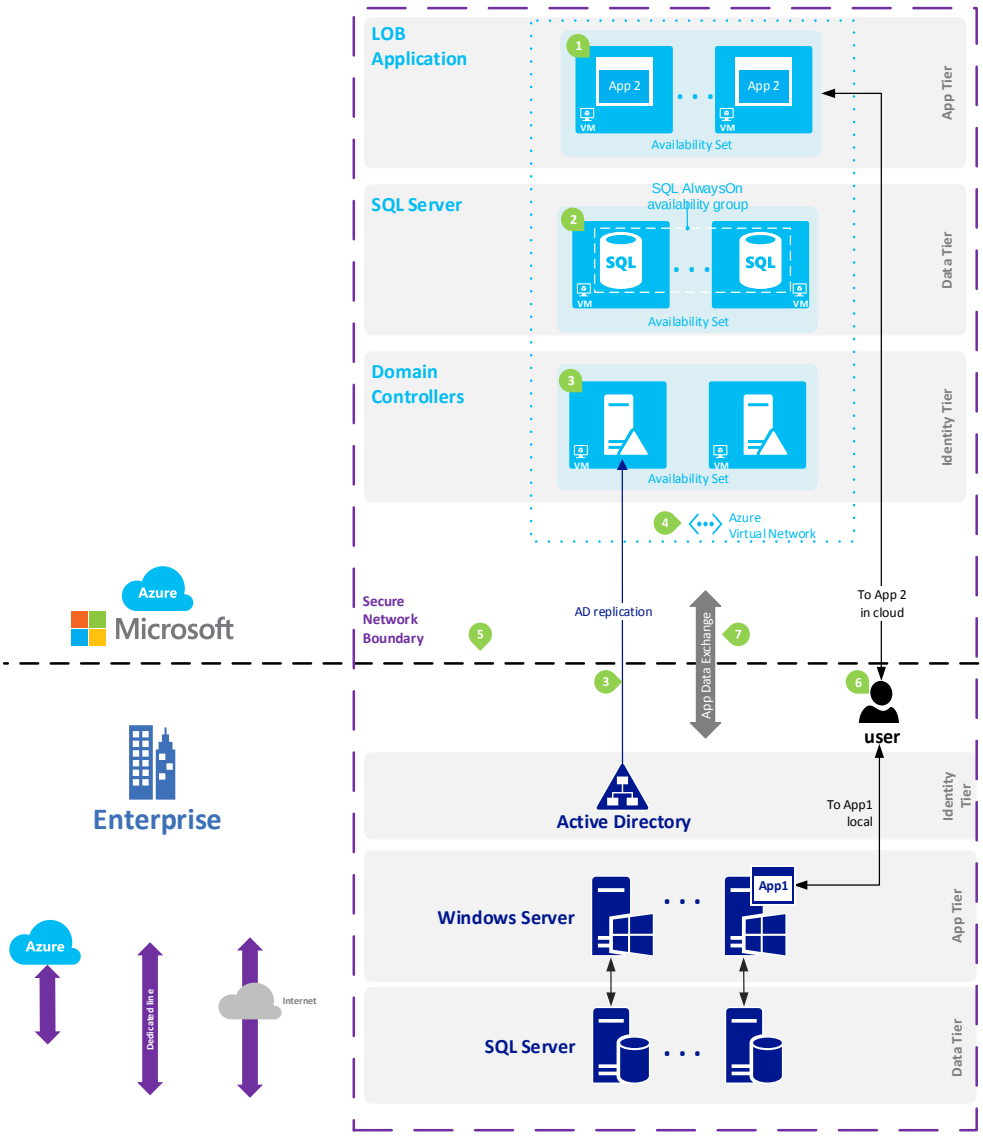


Figure A-22: Line-of-business applications

As shown in the illustration:

1. Package your application into a VM and deploy it to Azure. Run at least two copies to provide redundancy in case of failure or add more to scale out.
2. Move your data layer to the cloud for the lowest latency. Take advantage of the SQL Server 2014 AlwaysOn feature to provide redundancy and failover.
3. Run two VMs as Active Directory domain controllers and DNS servers in Azure and synchronise these services with your on-premises Active Directory domain controllers. The application can then authenticate users without the added latency of connecting to the on-premises Active Directory.
4. Connect all your VMs in the cloud to an Azure Virtual Network.
5. Connect on-premises to the cloud via a Virtual Private Network (VPN) over the Internet. For a lower-latency dedicated line, use ExpressRoute.

6. On-premises users now access their applications in the cloud with no changes to the user experience.
7. The applications in the cloud and on-premises can securely communicate and exchange data.

DevOps

As we discussed in Chapter 8, a modern DevOps process in the cloud gives you the means to release and iterate quickly, even several times a day – and perhaps more. Getting to market first can be the advantage that makes you a market leader, instead of leaving you to play catch up.

The Microsoft Visual Studio toolchain provides agile planning, source code control, package management, building, testing and release automation to continuously integrate, test, deliver and monitor your application. And with monitoring tools such as Application Insights, and deployment and configuration management tools such as Chef and PuppetLabs' Puppet, you can update as often as you need.

SharePoint on Azure

If you're not using SharePoint Online yet, you can spin up infrastructure easily for your SharePoint servers in minutes. You can set up development or test farms, or scale-out your production SharePoint deployments by instantly adding more resources. Simplify deployment and configuration with ready-to-deploy images and templates that are based on tried-and-tested configurations, and reduce the time to deploy complex SharePoint farms from days to minutes. Figure A-23 shows the configuration.

Figure A-23: SharePoint on Azure

Pay-as-you-go pricing and per-minute billing from Azure helps you to save money. For development and testing, take advantage of Azure benefits for Visual Studio subscribers to reduce software licensing costs. When testing, you can spin up additional servers as needed for scale and load testing over short periods of time and remove them when you're finished. Using resources that you need – and no more – helps you be more cost effective.

Dynamics on Azure

Dynamics 365, an online SaaS suite of business applications, gives you core business functions – ERP, CRM, supply chain, business intelligence and many others – in the cloud. Dynamics 365 gives solutions in sales, customer service, operations, financials, field service, project service automation, marketing and customer insights.

Dynamics partners have created many others. On Microsoft AppSource, you'll discover hundreds of apps developed by Microsoft and our partners to enhance Dynamics 365. Search for apps by name, industry and category – and download them to begin using them right away. Or, work with a company in the Microsoft Partner Network to get the apps up and running.

Hybrid cloud scenarios

When creating a hybrid (a mixture of on and off-premises computing) enterprise cloud application or set of applications, a number of opportunities arise to both simplify operations and cut costs. Here, we show a few ways to effectively use the cloud for common IT operational scenarios.

Hybrid cloud connectivity

In a hybrid cloud, some applications are hosted on-premises, whereas others reside in the cloud; ideally, where these applications live should be transparent to end-users. In other words, cloud-resident applications should appear to be within the on-premises network, with appropriate IP addressing and routing. Applications in the cloud are configured to be in the same IP range as those in the datacentre through the Azure portal.

There are a number of approaches to achieving this type of location transparency. The next few sections describe four separate ways to connect a datacentre to Azure.

Point-to-Site

Using the Internet, you can create such a virtual private network (VPN) in two ways. The first is called a *point-to-site* model (Figure A-24), in which the VPN is configured through software on an individual client computer in the datacentre. The least expensive of all the options, point-to-site connections are useful when only a few machines on-premises need connectivity to the cloud, or when the connection is from a remote or branch office.

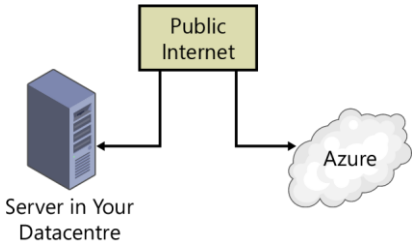


Figure A-24: Point-to-site connectivity

Site-to-Site

Another approach is called *site-to-site* connectivity (Figure A-25). In this configuration, a datacentre deploys a hardware VPN gateway to link the on-premises datacentre in its entirety with applications and data in the cloud. The hardware gateway must have a public-facing IP address and a technician must be available to perform the configuration.

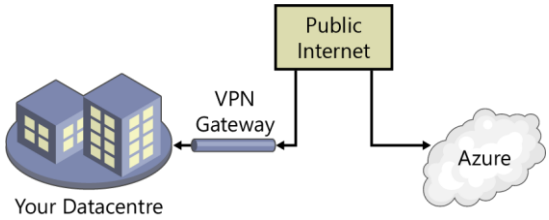


Figure A-25: Site-to-site connectivity

Microsoft Azure ExpressRoute

Many customers want configurable and deterministic network latency with their cloud applications. They might also want their network traffic isolated from the public Internet. To support these requirements, a direct connection is available from the datacentre to Azure using a partner telecommunications service provider called ExpressRoute (Figure A-26).

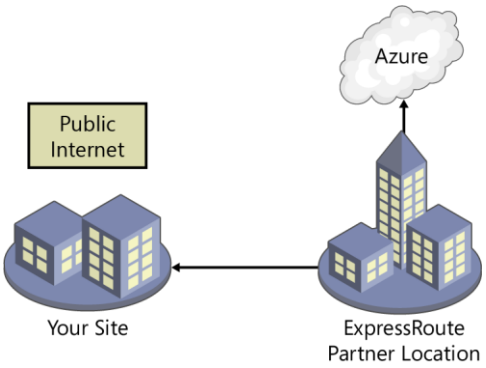


Figure A-26: Connecting to the datacentre via ExpressRoute

Although this is potentially a more expensive solution, ExpressRoute provides the fastest connectivity as well as isolation from the Internet, essentially by connecting via a "dedicated line".

A full list of supported telecom providers for ExpressRoute is available on the Microsoft website.

Wide area network connectivity

In addition, it is possible to connect through a telecom provider such that Azure simply appears as another site on the customer's wide area network (WAN), as depicted in Figure A-27.

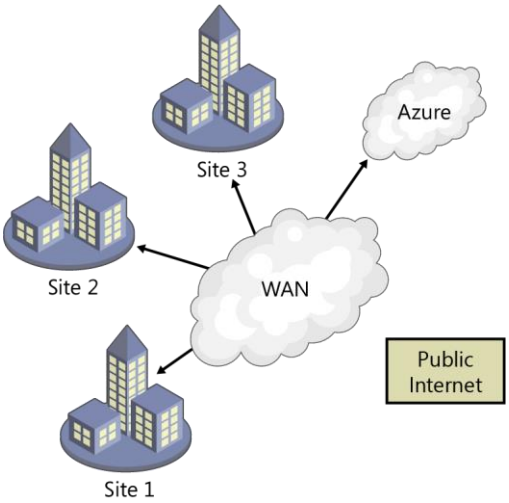


Figure A-27: WAN connectivity

As with the previous approach, by using a telecom provider as the transport, you can negotiate the bandwidth with the provider and, of course, network isolation is provided. Work with your telecom provider to find the best approach for your organisation.

In these four scenarios, we have shown a variety of approaches for connecting your enterprise datacentres to applications and data in Azure. The choice you make will depend on how you calculate the bandwidth/cost trade-off, whether it needs to be isolated from the open Internet and how geographically dispersed your sites are.

In the next few sections, we will describe a series of common application-level scenarios.

Hybrid database scenarios

Many enterprises have made significant investments in on-premises SQL Server. A number of features extend the functionality of on-premises SQL Server to the cloud, taking advantage of Azure's low cost and massive scale.

For example, an on-premises instance of SQL Server may be synchronised with either an instance of SQL Server running on a VM in Azure (i.e. in an IaaS instance) or with the cloud-native SQL Azure. This enables, for example, dispersed teams to do development on the on-premises instance, as shown in Figure A-28.

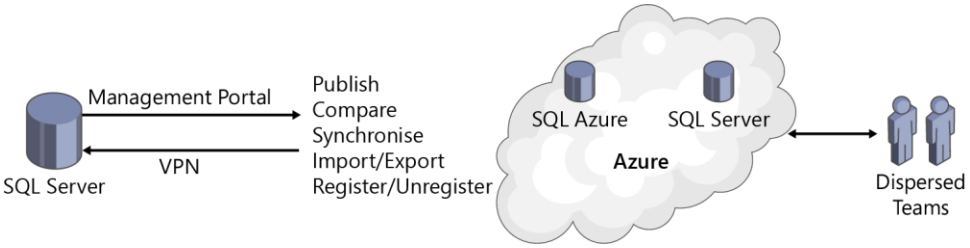


Figure A-28: On-premises SQL data publishing to cloud

In addition, as Figure A-29 illustrates, you can use an Azure instance of SQL Server as a backup destination for an on-premises instance. Alternatively, for a very cost-effective solution, SQL Server (either on-premises or cloud-based) can back up to and be restored from low-cost Azure Blob storage.

Figure A-29: On-premises SQL database backup to cloud

Finally, as another pattern, you can use the cloud to provide additional capabilities to an on-premises instance, lowering its load. In the example presented in Figure A-30, there are two cloud replicas: one being kept as a backup for disaster recovery purposes, the other being used to power business intelligence (BI) applications.

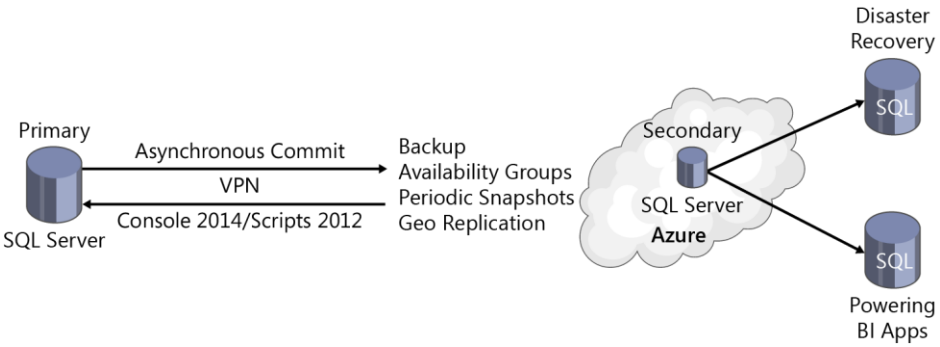


Figure A-30: Using cloud database as a replica

High availability in the cloud

Many mission-critical applications require the highest availability possible and must be resilient to hardware and network failures. Hosting applications in the cloud provides a number of capabilities, including redundancy, fault tolerance and resilient design that make high availability possible.

First, consider Azure Service Level Agreements (SLAs).[16] For example, the Azure Compute service (application services) comes with a 99.95 % SLA; SQL Database has a 99.9 % SLA; and Azure Storage has a 99.90 % SLA. Without any additional work, your application is, by default, guaranteed no more than 108 minutes of downtime in a month (out of 43,200 minutes).

However, you can do much to improve even these excellent numbers. Programming techniques such as durable queues and asynchronous communications make applications less tightly coupled to one another, improving the chances that one failure will not cause a cascade of further failures.

Using Azure availability sets ensures that different instances of VMs, and/or different workloads, are physically placed on different racks (different power supply, switch and server) in an Azure datacentre. Availability sets ensure that, should a planned or unplanned maintenance event or failure occur, at least one VM instance will be available for use.

It is also efficient to "tier" applications into availability sets. By placing all "web tier" applications into a single availability set, it becomes straightforward to reboot or upgrade the entire tier at once, with the underlying availability set logic ensuring that at least one of each application is available.

Figure A-31 shows a workload spread across three tiers; each tier is associated with a different availability set.

[16] See https://msdn.microsoft.com/en-us/library/azure/dn251004.aspx for more detail.

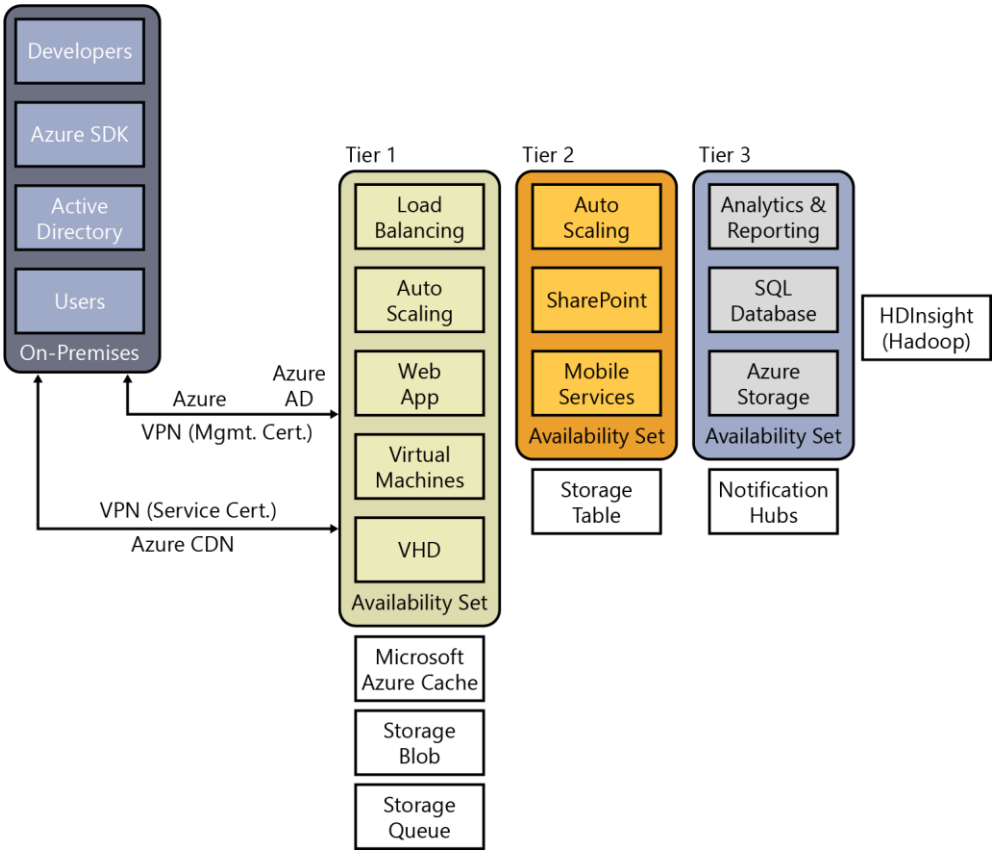


Figure A-31: High availability

In addition, you can place workloads on geographically separate datacentres, as illustrated in Figure A-32. You can use Azure Traffic Manager to switch operations from the primary datacentre to the backup in the event of a catastrophic failure in the primary.

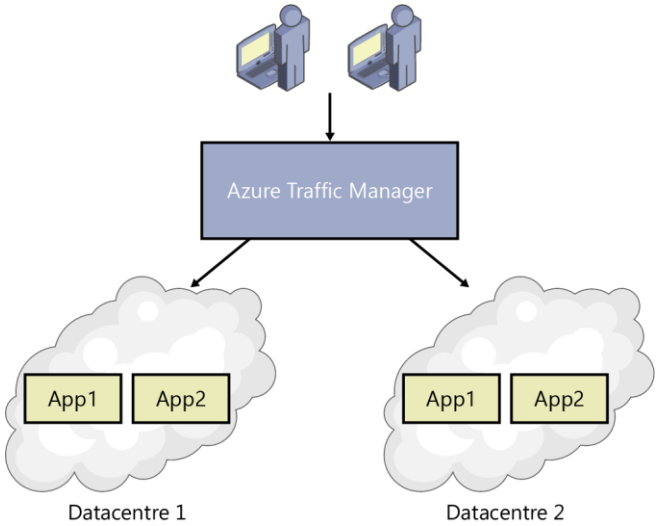


Figure A-32: Using different geographies

Design considerations

When you're thinking about the availability of your applications or workloads, consider the following:

- Do you require an SLA greater than 99.5 %, which is the default Azure SLA?
- How many instances of each application VM do you require?
- Which applications can make use of asynchronous and loosely coupled programming techniques to improve their availability?
- Would geographically redundant datacentres improve your workload availability within your cost parameters?

Connected devices

The IoT, as we discussed earlier in the book, carries great promise for – and places great demand on – the cloud. IoT devices range from medical sensors to manufacturing devices, to connected cars and aeroplanes, to building environmental sensors – and on and on. Estimates suggest that within a few years tens of billions of such "things" will be attached to the Internet in some form; in this scenario the cloud receives, analyses and takes action on data sent by IoT devices.

Azure provides a number of services that make the IoT possible. With the Azure Event IoT Hub (Figure A-33), enterprises can create a device registry listing all allowed connected devices, and can manage, configure and provision them. The Azure IoT Hub makes it possible for cloud applications to ingest very large numbers of events (billions per day, if needed) from connected devices. These events can be examined in real time by Azure Stream Analytics, which can perform filtering operations, only passing on those events of interest (such as a device failure).

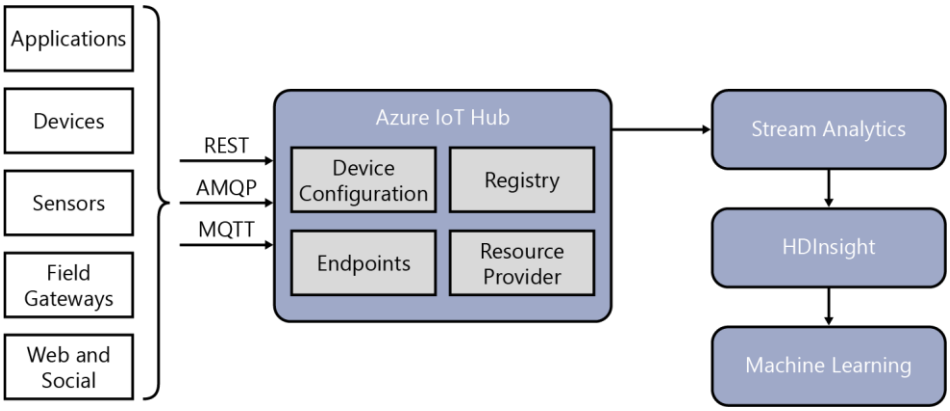


Figure A-33: Azure IoT

Other useful services include Azure HDInsight, which is capable of collecting very large amounts of data and running batch analytics programs (MapReduce) in order to find patterns; and AzureML, Azure's machine learning capability, which can be trained to detect anomalous patterns and predict future outages or downtime.

Of course, with all these devices connected to your application, security must be part of the architecture. Many IoT devices do not have the computing power to perform full public-key encryption or digital signature, so you should be familiar with and use Shared Access Signatures (also known as SAS tokens) wherever possible. A SAS signature, as the name implies, is about *access*; the token includes in its query parameters the URL being requested, an expiry time, permissions and other key data. SAS tokens provide an efficient way to guard against intruders accessing your application without authorisation.[17]

Design considerations

When designing an application that uses Internet-connected "things," consider the following:

- How many devices will be connecting? How frequently will they be sending data and how large are the messages? This will help you to determine the scale of Azure Event Hubs you need to receive and process the messages.
- What protocol (HTTP/REST, AMQP, MQTT) will they use to connect?
- What sorts of data will they send and which parts of that data are useful to applications?
- Do you need to retain the data for any reason?
- How do you want to visualise the state of your devices? Do you need a "dashboard" (such as Azure PowerBI) to aggregate and visualise the data coming in?

Identity and authentication

Identity management is the core of security in the cloud. A user's identity determines which resources they have access to, and the identity management system prevents unauthorised access where appropriate, protecting enterprise resources.

In Azure, identity management is handled by Azure Active Directory (Azure AD), based upon the industry standard Active Directory family of products. Azure AD is used to authenticate users on cloud applications, and can be synchronised with and federated to an on-premises Active Directory such that enterprise users can take advantage of single sign-on (SSO) to access both on-premises and cloud applications.

Using the OAuth/OpenID protocol, other forms of identity can optionally be configured with Azure AD (Figure A-34). Azure AD supports Facebook, Google, Yahoo and Microsoft accounts as identity providers, and each of these can be granted varying levels of access. In addition, you can integrate a wide variety of SaaS applications (such as SalesForce.com and many others) with Azure AD. Multi-factor authentication also supports compliance with NIST 800-63 Level 3, HIPPAA, PCI DSS and other regulatory requirements.

[17] For more information on Shared Access Signatures, see https://azure.microsoft.com/documentation/articles/storage-dotnet-shared-access-signature-part-1/ and https://azure.microsoft.com/documentation/articles/storage-dotnet-shared-access-signature-part-2/. It is important to recognise that SAS is not just an IoT-only technology, you can also use it, for example, with Azure Storage to provide delegated access to data.

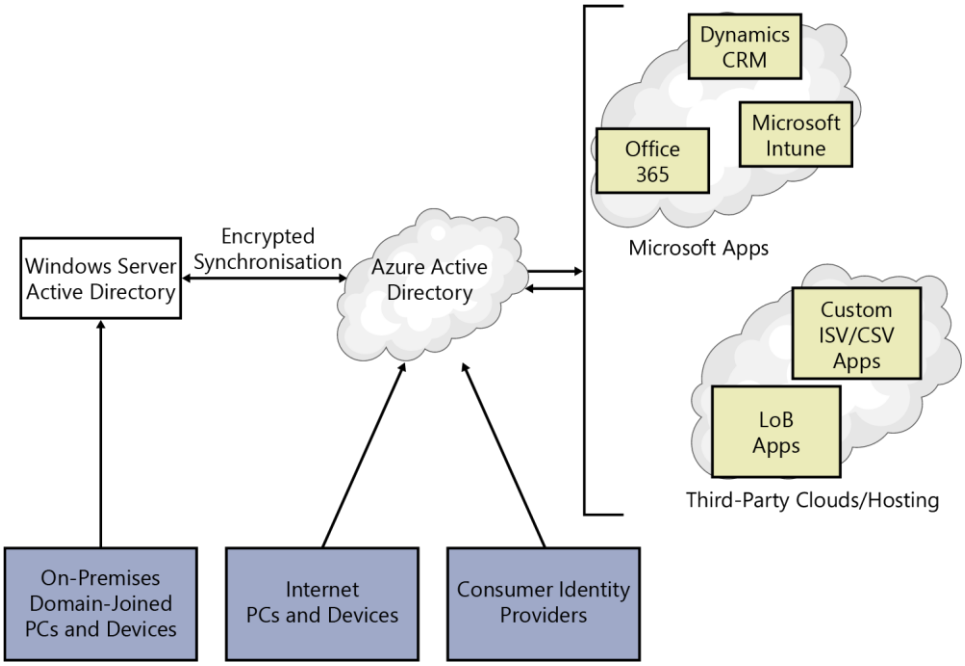


Figure A-34: Azure AD authentication

Finally, Azure Active Directory supports two-factor authentication (Figure A-35) for rigorous identity management. Typically, a user first authenticates using conventional credentials such as their username/password and then uses a physical device such as a smartphone or smartcard to complete the authentication process. Azure Active Directory can be configured to call a smartphone and request a PIN or request a badge be read, or perform a biometric authentication (for example, fingerprint).

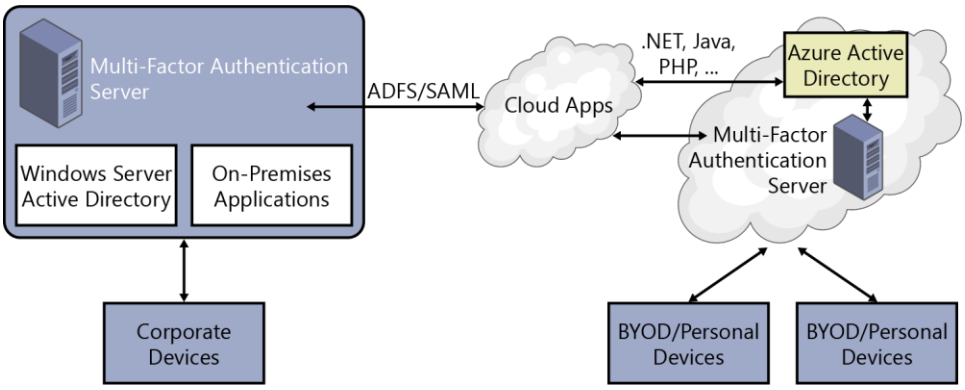


Figure A-35: Azure AD multi-factor authentication

Design considerations

It has been said that identity management is at the core of the cloud, because it controls access to computing and data resources in the cloud. With this in mind, consider the following:

- Federating your on-premises Active Directory to Azure AD to turn on SSO for cloud applications;
- Configuring consumer authentication mechanisms for certain types of access (e.g. e-commerce customers) to your cloud applications;
- Two-factor authentication for the most rigorous authentication requirements.

Enterprise Mobility Management

In 2014, a number of mobility-related services were bundled together to provide a cohesive mobility offering for enterprise IT departments. This bundle is called the Enterprise Mobility Suite (Figure A-36) and it includes Azure AD, with additional services, including the ability to perform group management and password self-service reset. It also provides preconfigured logins to a large number of SaaS applications and security reporting (e.g. for repeated failures or anomalous login patterns) and can accommodate two-factor authentication, described earlier.

The Enterprise Mobility Suite also includes a comprehensive Mobile Device Management offering using Windows Intune, which makes it possible for IT professionals to manage mobile access to enterprise resources as well as perform email profile management, selective wipe and remote lock, and password reset.

Finally, Enterprise Mobility Suite also comes with Microsoft Azure Rights Management, providing robust document protection for both Microsoft Office 365 (cloud) and on-premises information.

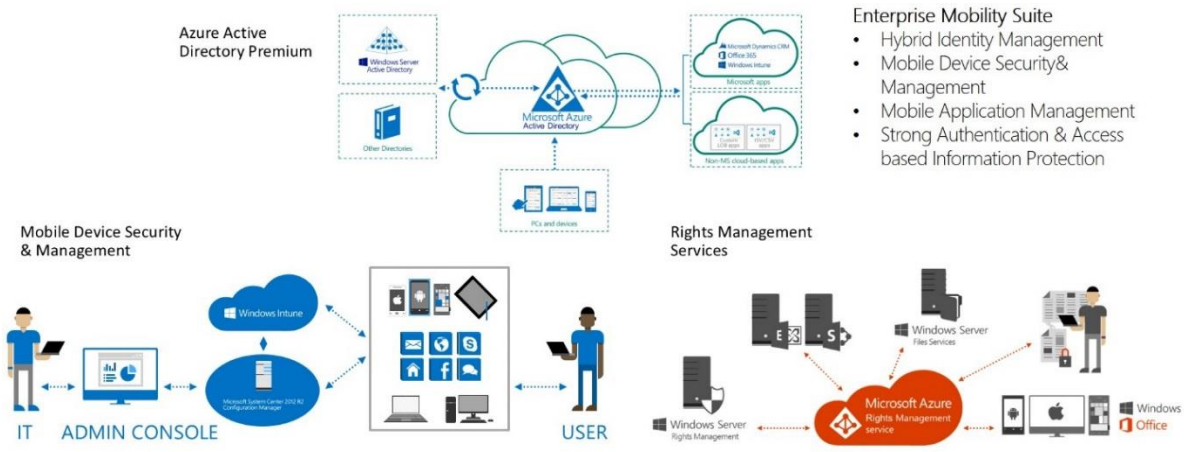


Figure A-36: Enterprise Mobility Suite

Design considerations

Consider using the Enterprise Mobility Suite if any of the following apply:

- You have a need to manage a variety of mobile devices;
- You want users to be able to set and reset their own passwords (and thus reduce the load on your help desk);
- A significant number of the mobile devices connecting to enterprise networks are actually employee-owned – for example, if your company has a Bring Your Own Device (BYOD) policy.
- You need to enforce specific data access privileges and policies for different users or classes of users.

Websites

Using Azure Web Apps and App Services, creation and maintenance of a complex enterprise website is straightforward and inexpensive. You can build advanced HTML5-based websites in any of a number of popular web application programming languages (.NET, Java, PHP, Node.js and Python). Using a wealth of tools, you can connect your site both to other web assets (such as Twitter) and on-premises data assets. Developers can create secure, authenticated web applications using Active Directory features such as Active Directory Authentication Library (ADAL) and the Active Directory Graph API, and you can secure access to documents through Azure's Rights Management Service. As mentioned earlier, you can connect and synchronise Azure AD with an on-premises deployment of Active Directory.

Figure A-37 shows website development, access and on-premises assets.

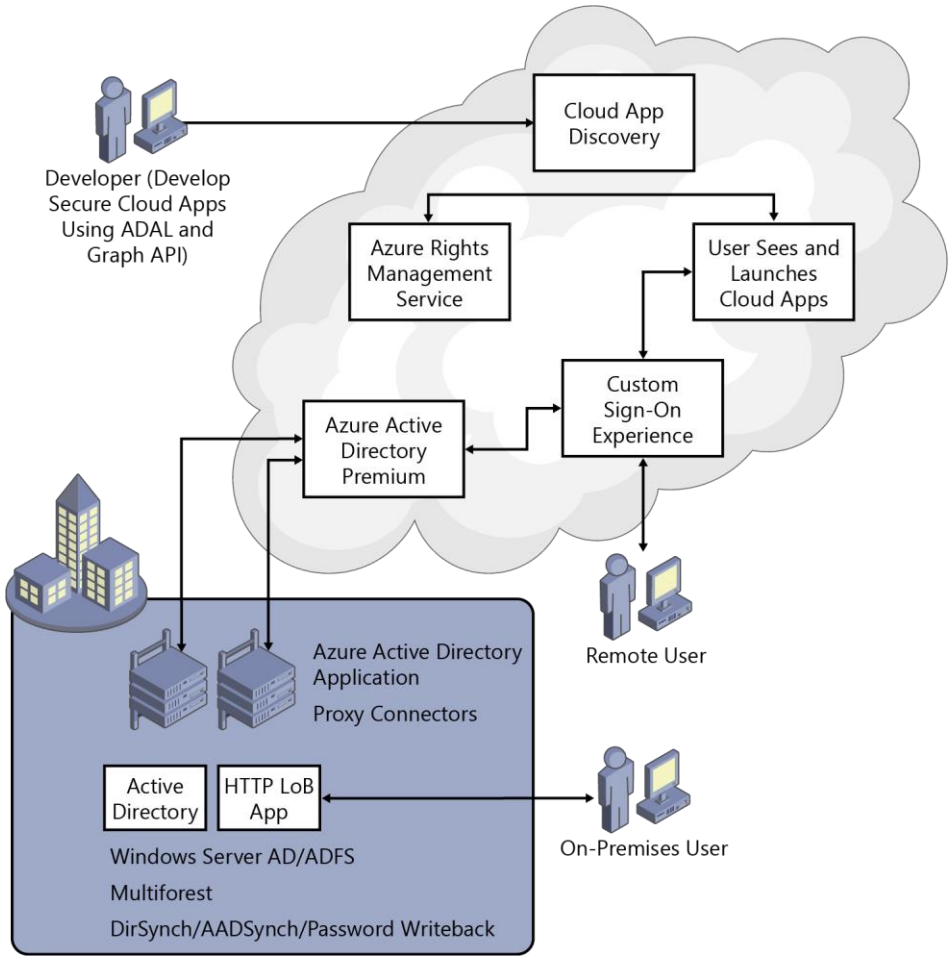


Figure A-37: Azure Websites

Design considerations

Of course, there are a plethora of design issues any time you are building and deploying a website. Here are a few that you should consider:

- Is it an intranet-only site or accessible from the broader Internet?
- How will you do content management to refresh data?
- Do you need the site to authenticate users? If so, can users authenticate with non-enterprise credentials such as Facebook, Google or Microsoft Account? And if so, do they have different access rights from enterprise users? Azure AD can provide an easy all-in-one authentication solution.
- What sorts of application integration with enterprise applications do you require? You can use BizTalk Server or Logic Apps to connect to on-premises applications such as ERP or databases.
- Do you need to perform B2B transactions on your database? Azure BizTalk Services provides the ability to connect to EDI X.12 applications elsewhere on the Internet.

Further reading

Azure Resources

Microsoft website, https://www.microsoft.com

Microsoft Azure website, https://www.azure.com

Microsoft Azure solutions site, https://azure.microsoft.com/solutions/

Azure Trust Centre, https://azure.microsoft.com/support/trust-center/

Cloud economics, http://download.microsoft.com/download/6/E/4/6E4CB3D1-5004-4024-8D90-6C66C83C17AA/The_Economics_of_the_Cloud_White_Paper.pdf

Microsoft cloud datacentres, https://www.microsoft.com/cloud-platform/global-datacenters

Forrester study, "The Total Economic Impact of Microsoft Azure PaaS," https://azure.microsoft.com/resources/total-economic-impact-of-microsoft-azure-paas/

Case studies cited in Chapter 1:

- https://customers.microsoft.com/story/rollsroycestory
- https://customers.microsoft.com/story/brainshark
- https://customers.microsoft.com/story/geico
- https://customers.microsoft.com/story/accuweather

Azure application architecture guide https://docs.microsoft.com/azure/architecture/guide/

Designing resilient applications in Azure https://docs.microsoft.com/azure/architecture/resiliency/

Resiliency checklist https://docs.microsoft.com/azure/architecture/checklist/resiliency

HIPAA: http://smb.blob.core.windows.net/smbproduction/Content/Microsoft_Cloud_Healthcare_HIPAA_Security_Privacy.pdf

Complying with GDPR https://blogs.microsoft.com/blog/2017/05/24/accelerate-gdpr-compliance-microsoft-cloud/

Cloud optimisation:

- https://www.microsoft.com/itshowcase/blog/determination-sets-8-year-old-on-path-to-save-microsoft-millions-of-dollars/
- https://www.microsoft.com/itshowcase/Article/Content/861/Optimizing-resource-efficiency-in-Microsoft-Azure
- https://www.microsoft.com/itshowcase/Article/Video/688/Managing-and-optimizing-resources-for-cloud-computing-at-Microsoft (webinar)

External sites

Martin Fowler blog post on Microservices, https://martinfowler.com/articles/microservices.html

Forbes blog post on big data, https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/

The three v's of big data: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Books

Jeanne W. Ross, Peter Weill and David C. Robertson, *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution,* Harvard Business Review Press, 2006.

Gene Kim, Jez Humble, Patrick Debois and John Willis, *The DevOps Handbook: How to Create World-Class Agility, Reliability and Security in Technology Organizations,* IT Revolution Press, 2016

About the authors

Barry Briggs, an independent consultant, has a long history in software and enterprise computing. He served in a number of roles during his 12-year career at Microsoft. Most recently he was the chief enterprise architect on the Microsoft DX (Developer Experience) team. The DX team's job is to design and build "epic" applications with Microsoft customers that exploit new capabilities of the Microsoft stack, including both Microsoft and open-source products and frameworks.

Previously, Barry served as chief architect and CTO for Microsoft's IT organisation. Principal among his responsibilities were creating and leading Microsoft IT's cloud strategy team, which put in place the strategy and processes behind the migration of Microsoft's internal IT ecosystem to the cloud. In addition, he led the Enterprise Architecture practice which aligned the business strategies to technology assets for maximum impact and agility. He drove a strategic incubations unit which builds cutting-edge software designed for IT-wide impact and technology adoption strategies, which fostered the deep relationship Microsoft IT has with its product groups. Prior to the CTO role, Barry led the team that created the world's largest Master Data Management (MDM) solution for Microsoft. He joined Microsoft in 2003 as senior architect for the Business Process and Integration Division, which built the BizTalk Server.

Prior to Microsoft, Barry served as CTO for a number of companies (Aptsoft, Wheelhouse, BroadVision and Interleaf); before that, he spent 11 years at Lotus/IBM. There, Barry was the lead architect for Lotus' famous spreadsheet product, 1-2-3, for a number of years. In addition, he also helped develop Lotus Notes and led the technology integration of Lotus with IBM following that acquisition. He also created and led the team responsible for the world's first Java-based productivity suite, Lotus eSuite. In 1995, he was named a Lotus Fellow.

You can see what Barry is up to at his website: http://www.barrybriggs.com

Eduardo Kassner is the chief technology and innovation officer at the Worldwide Channels & Programs Group at Microsoft Corporation. His team is responsible for defining the strategy and developing the programmes to drive the technical capacity, practice development, and profitability for the hundreds of thousands of Microsoft partners worldwide. He has built up 26-plus years of experience managing and designing complex IT environments and connecting IT and business objectives in real life. He has led teams that actively helped international corporations and governments with these challenges in a direct and no-nonsense approach. This involved combining structured frameworks with hard-earned experience, linking discussions from technical all the way to business value, with the ability to link the Microsoft technology stack to the way it can land and provide value in an enterprise or government environment. He recently co-wrote and published the first edition of *Enterprise Cloud Strategy* published by Microsoft Press, which has been downloaded more than 250,000 times from the Azure.com website. Eduardo has led the teams that designed, hired and managed the Microsoft Cloud Architecture role and worldwide community, Cloud Adoption Frameworks, Operations Management, Automation and Architecture Patterns. He regularly speaks at conferences worldwide on the topics of digital transformation, cloud adoption and strategy, cloud architecture best practices, public/hybrid/private cloud, operational efficiency, IT return on investment and total cost of ownership, and the overall Microsoft platform.