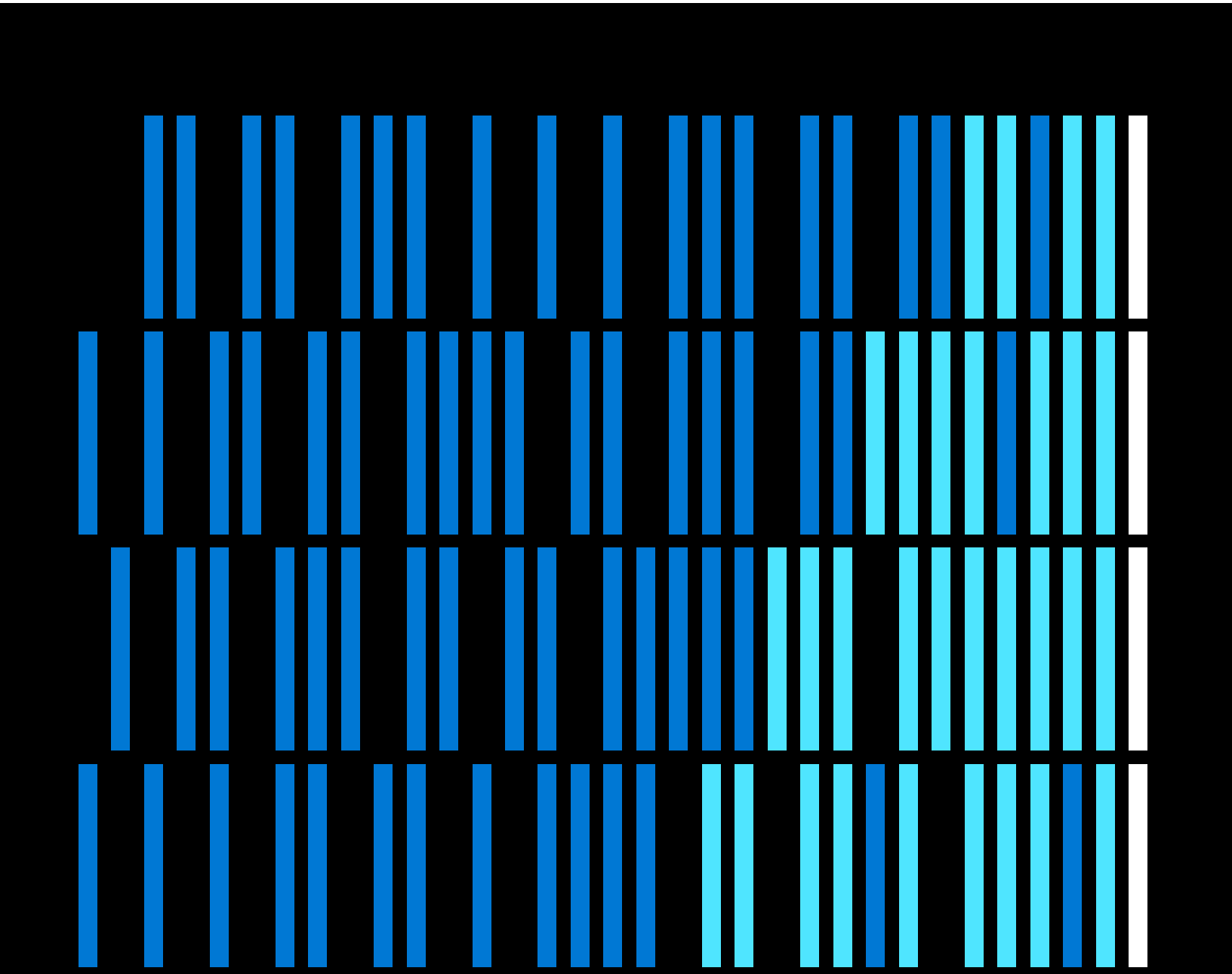
vanroey.be
TRUST DIGITAL, CREATE WONDERFUL THINGS

Microsoft Azure

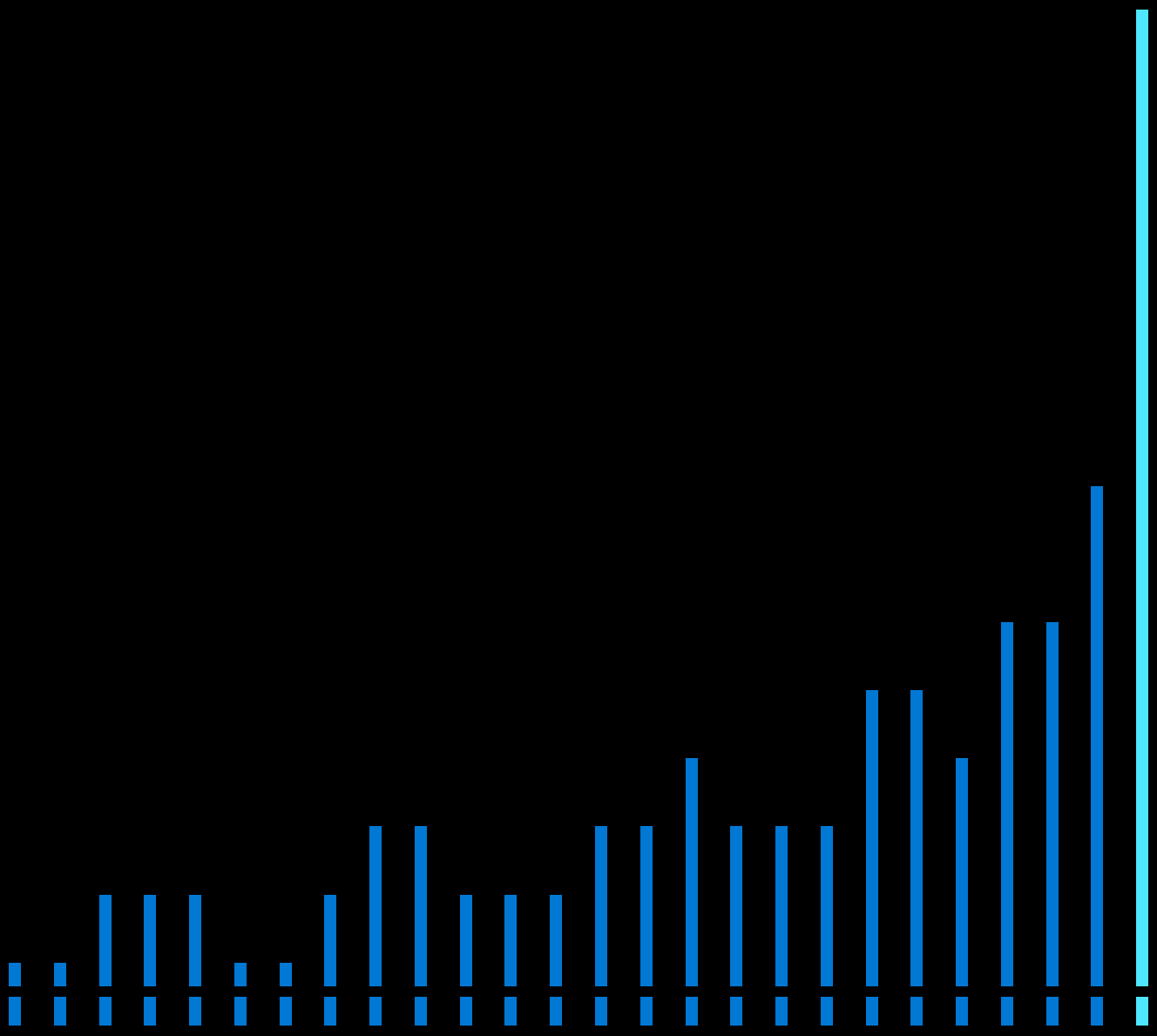
Enterprise Cloud Strategy
2nd edition



Contents

Part 1

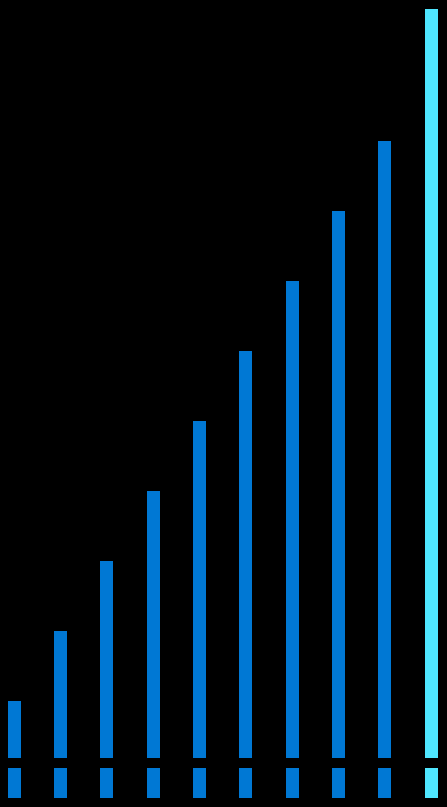
Beginnings



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

For the CIO, all of this expansion meant an IT organization 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 datacenters; procurement of new hardware; deployment of new servers, software, and retirement of depreciated hardware; network management; 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 utilization in the datacenter was, surprisingly, sometimes in single digits. Virtualization—putting multiple workloads on a single server—went some distance in improving utilization, but overall it remained low, which suggested that money was being wasted on IT assets that still were not being fully utilized.

Between operations staff, capital equipment management, and software maintenance, an IT department could easily spend 80 percent 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 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 compute, network, and storage resources that you use, the higher the bill. Of course, the reverse is also true: the less you use, the less you are charged. Indeed, most IT organizations find wide variations in system utilization: some applications (e.g., retail shopping) are seasonal; other applications (e.g., training applications) run for a short period of time before being shut down; others are simply unpredictable. The cloud addresses this variability, as illustrated in Figure 1-1, perfectly by its "pay for what you use" model.

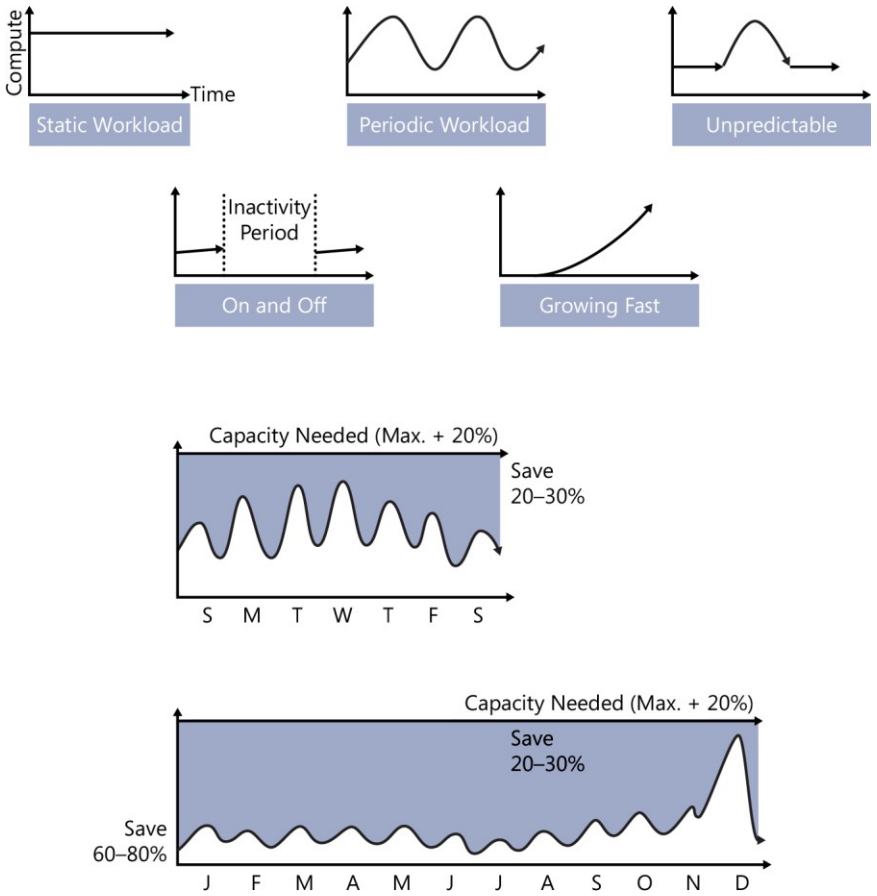


Figure 1-1: Common application utilization models

(It is worth mentioning that in the on-premises datacenter, as noted, the *maximum* utilization 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 chores in their entirety. The cloud provider in turn is heavily incented to utilize 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 datacenters hosting public clouds can also achieve economies in purchasing other resources; cloud datacenters pay only a quarter of the average cost of electricity in the United States. In many cases, cloud datacenters take advantage of local renewable energy; for example, Microsoft's datacenter in Quincy, Washington, is located near a hydroelectric facility, and other datacenters 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.

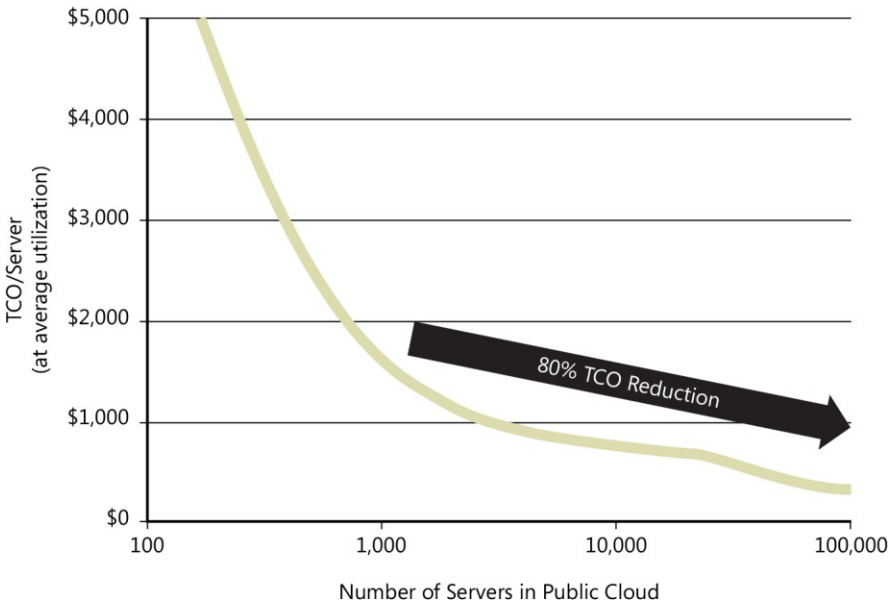


Figure 1-2: Economies of scale in the cloud

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

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

Perhaps most important, 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 datacenter 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 datacenter. 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 done. 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 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 optimize their use of the cloud on a day-to-day basis, adjusting consumption and utilization 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 makes not only IT better, but provides direct benefit to the business, making the CIO not just a cost center, 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 the Chapter 2), but the results were striking: an ROI of 466 percent, with a reduction in the amount of IT time spent on maintenance at 80 percent, among other benefits:

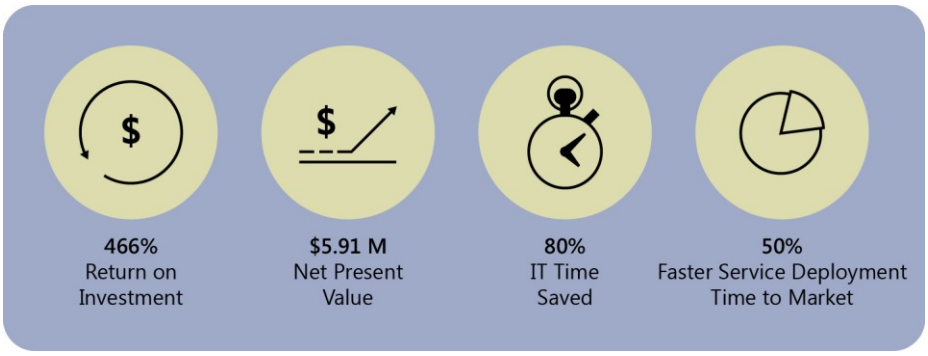


Figure 1-3: Cloud platform-as-a-service benefits

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 nonvalue-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 instantiating processes that deliver ongoing competitive advantage.

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

Transformational innovation drives a different culture and mindset than most organizations 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 examine a number of brief case studies from various global companies, all of whom 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.

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

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

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

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 24x7 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 datacenters, 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 analyze 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 modeling 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 analyze. Microsoft Cortana Intelligence capabilities are helping us filter the signal from the noise across large datasets so 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 that 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 maximizing 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 leader for its 17 years in business, 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 practice in places with which they're familiar.

By taking costs out of nonvalue-added functions associated with an on-premises datacenter, 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 March 22, 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 coordination 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 leveled 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 Program (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 visualization tool helped them to coordinate and track progress.

Figure 1-4: Nepal's debris management application

Learnings

What have we learned? 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 datacenters
- Communicate with Internet-connected devices all over the world
- Tap into big data and analytics services for personalization, better products, and more efficient processes
- Enjoy unprecedented development, test experimentation, and innovation cycles

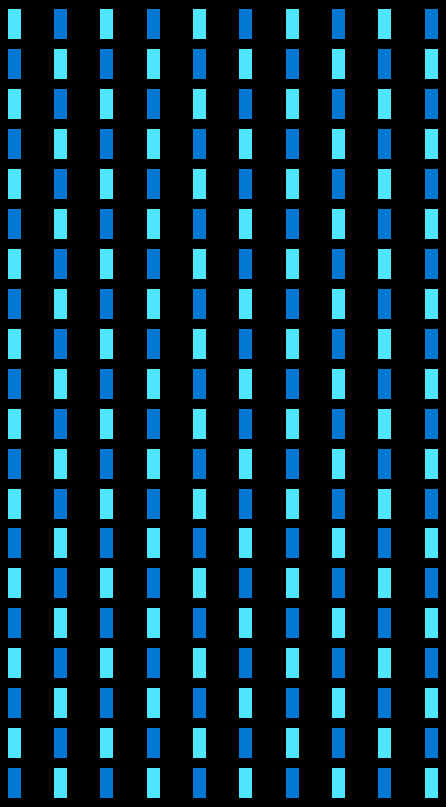
Every IT department is charged with safeguarding its company's information assets, with reducing costs, and with "keeping the lights on." These functions are, and always will be, critical components of any IT organization. Yet IT also must facilitate and foster innovation, both to make existing processes 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, which makes it possible for 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 toward 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 datacenters 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 compute capability.

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

A key measure of datacenter efficiency is called *Power Usage Effectiveness* (PUE), which measures how power coming into the datacenter 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 datacenters typically realize a PUE of 2.0, meaning that half of the incoming energy is used by noncomputing equipment such as air conditioning, lighting, and so forth. Cloud datacenters are now achieving PUE ratings of 1.1 and 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 datacenter. In fact, they are very different. In the traditional on-premises model, IT departments purchase hardware as applications need them, 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 datacenter, 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 utility. 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; thus, 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 on a rental basis. The leading public cloud vendors have datacenters 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 through their telecom provider a dedicated line linking the corporate datacenter with the cloud; bandwidth can be purchased as needed. This solution is preferable when it is desired to keep all traffic off of the public Internet or when substantially higher bandwidth is required. However, it of course entails additional cost.

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 datacenters 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 a just two.

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

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 coexist on a cloud server. A thin piece of code called a *hypervisor* ensures that each one runs in a timely and efficient fashion.

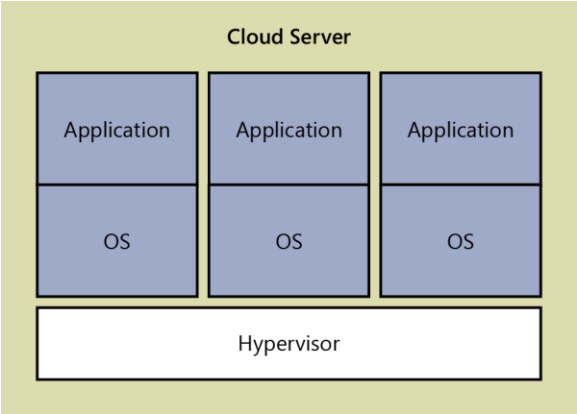


Figure 2-1: Infrastructure as a service

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

This is the easiest and fastest migration strategy; it offers many benefits, including cost savings. However, it still means that your operations staff will need to perform such tasks as patch management, updates, and upgrades. Nevertheless, IaaS is one of the most common cloud deployment patterns to date because it reduces the time between purchasing and deployment to almost nothing. Additionally, because it is the most similar to how IT operates today, it provides an easy onboarding ramp for your current IT culture and processes. As we shall see, the bulk of migration especially in the early phases of cloud adoption is to IaaS.

Platform as a service

In platform as a service (PaaS—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.

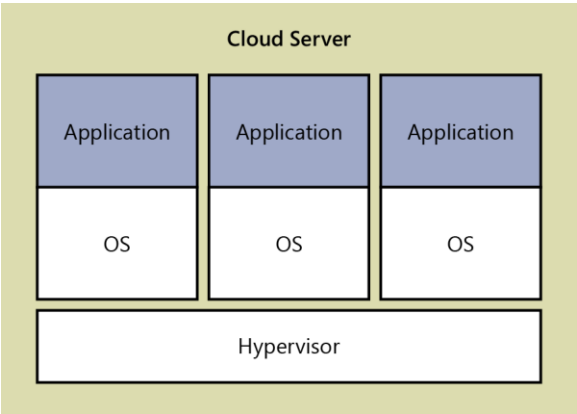


Figure 2-2: Platform as a service

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

Software as a service

In software as a service (SaaS—pronounced "sass"), you simply rent an application from a vendor, such as Microsoft Office 365 for email and productivity. This is by far the most cost-effective of all the options because typically the only work involved for the IT department is provisioning users and data and, perhaps, integrating the application with single sign-on (SSO). Typically, SaaS applications are used for functions that are not considered business-differentiating, for which custom or customized applications encode the competitively differentiating business models and rules.

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 containerized application typically does not involve loading and initializing an entire VM with an operating system, container startup can be very fast, so scale-up and scale-down can be very performant.

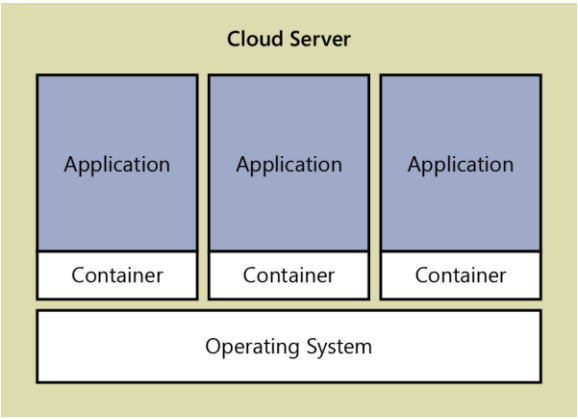


Figure 2-3: Container architecture

Containers have many advantages. Often it is possible to package an application with few or little 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 datacenter, the enterprise is fully responsible for everything, from the datacenter's operation, the facilities, electricity and air conditioning, all the way through the application. As the migration to the cloud progresses, more and more of these expenses are borne by the cloud provider.

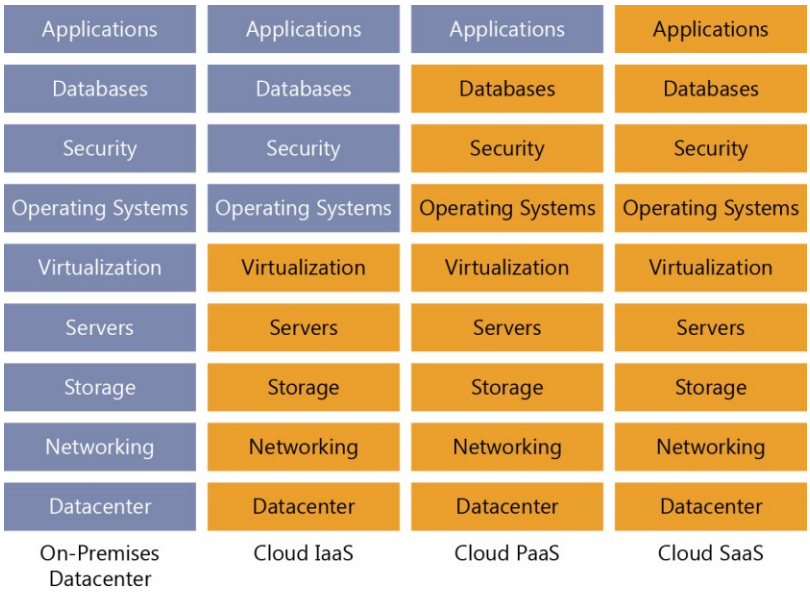
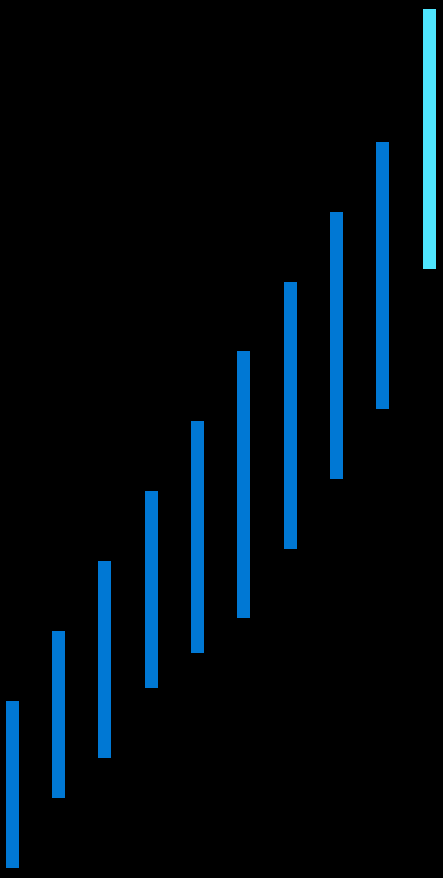


Figure 2-4: "As-a-service" compared

Chapter 3

The cloud: what is it?



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 modernize

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 *modernize* 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 their applications for far too long without assigning to 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, modernization 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 database.)

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

Evolution of the five R's of modernization

To focus our efforts on guidance for existing applications, let's proceed with the most convenient way to think about modernization, which is commonly called "the five R's":[2] retire, replace, retain and wrap, rehost, and reenvision. 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 highly on the situation, application, and types of cost involved, there is no one-size-fits-all solution.

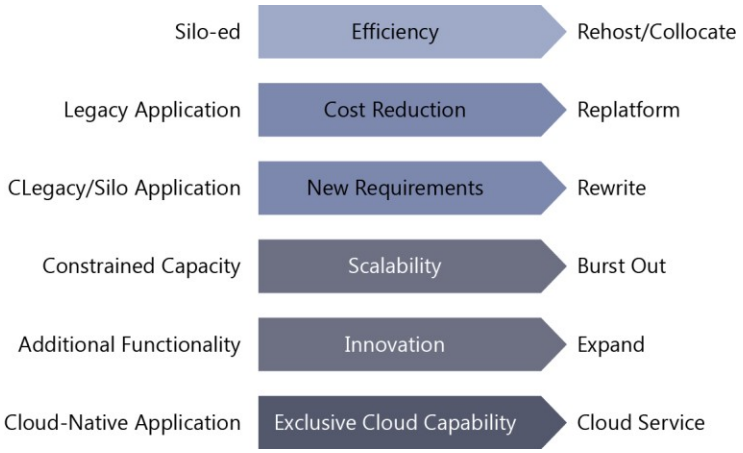


Figure 3-1: Types of modernization initiatives

- **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 modernized.
- **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—might 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, add an API so that authorized external applications can exercise the functionality of the application.
  - 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, connect them to other software components, such as analytics applications, machine learning, or mobile access.
- **Rehost** If a legacy application is providing good value but is expensive to run, it might be a candidate for rehosting. Rehosting 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 rehosting situation, the legacy application might currently be 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 easily to the cloud without some significant work. In those cases, you might want to consider reenvisioning and building the application in the cloud.
- **Reenvision** If a legacy application is providing good value but cannot be easily moved, the best solution might be to reenvision it and build it again in the cloud. Reenvisioning 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. Reenvisioning an application might require rewriting the main logic using a modern development language and tools and making it service-oriented. Reenvisioning an application can be facilitated by starting with VMs in the cloud which can be instantiated in a matter of minutes.

  There are many ways to take advantage of the cloud when reenvisioning, 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.

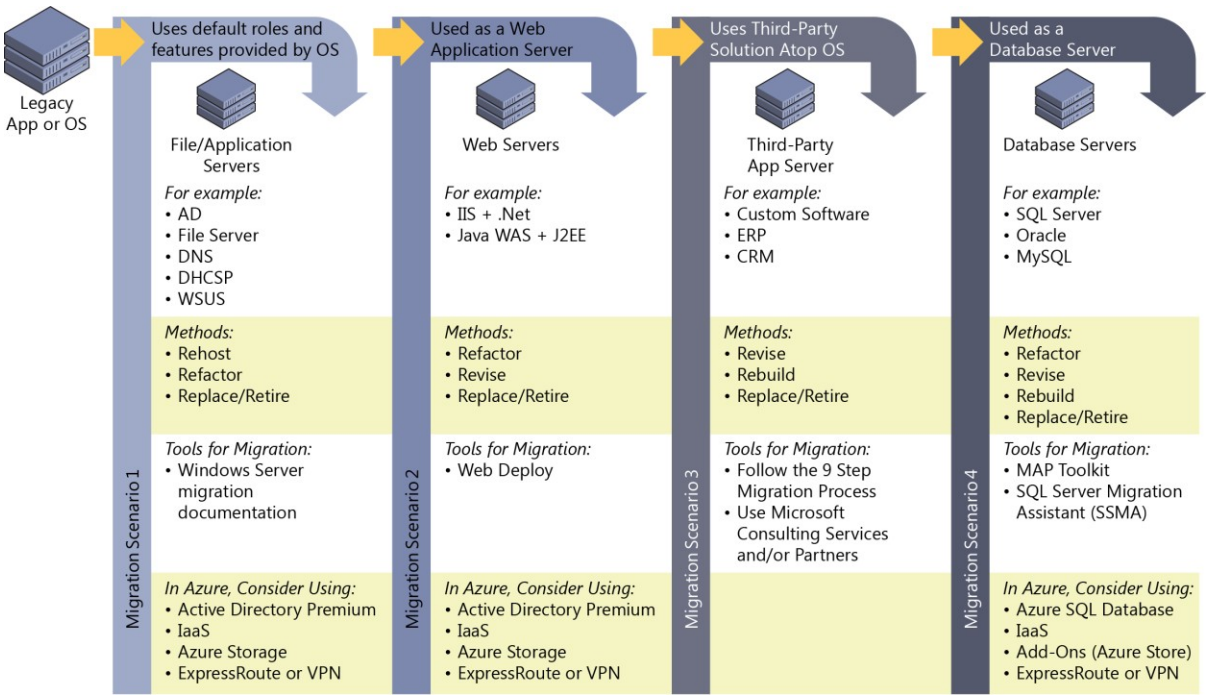


Figure 3-2: Legacy application strategies

In the figure, you can see that you can upgrade legacy software by rehosting them on more modern platforms, or you can move them to the cloud and gain more advantage. 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; rather, 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—in many ways the most demanding of the phases—the bulk of the IT portfolio is moved to the cloud in one form or another. This requires cooperation and collaboration across a number of different enterprise functions, including the technical staff, the operations staff, as well as the executive team, business sponsors, security professionals, regulatory compliance staff, legal, and HR. We spend a significant amount of time in this book covering migration in all its aspects.

Transformation

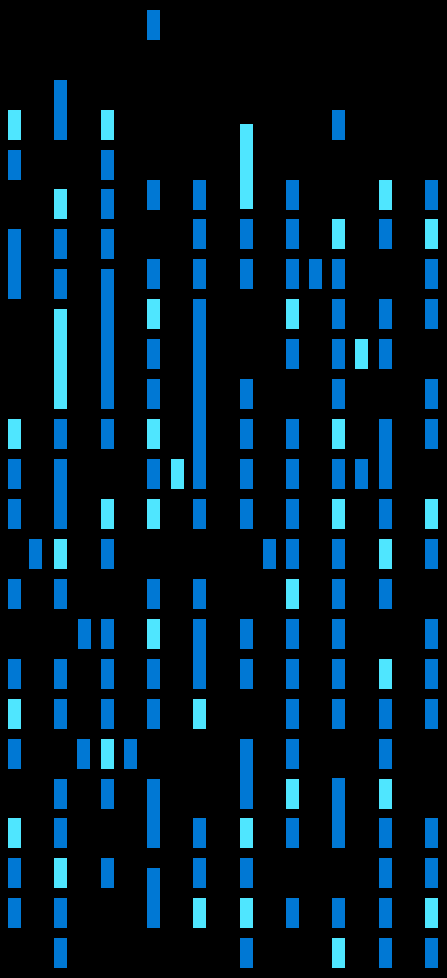
In the transformation phase (which will often coincide with the migration phase) selected applications are redesigned to take maximum advantage of the cloud—using the platform as a service model—affording greater scale, greater integration with other cloud services, and numerous other advantages.

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 function and feature than before.

The chapters that follow cover each phase in detail.

Chapter 4

Experimentation



There is always a first cloud application. In every IT organization, 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 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. First, 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.

Second, 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.

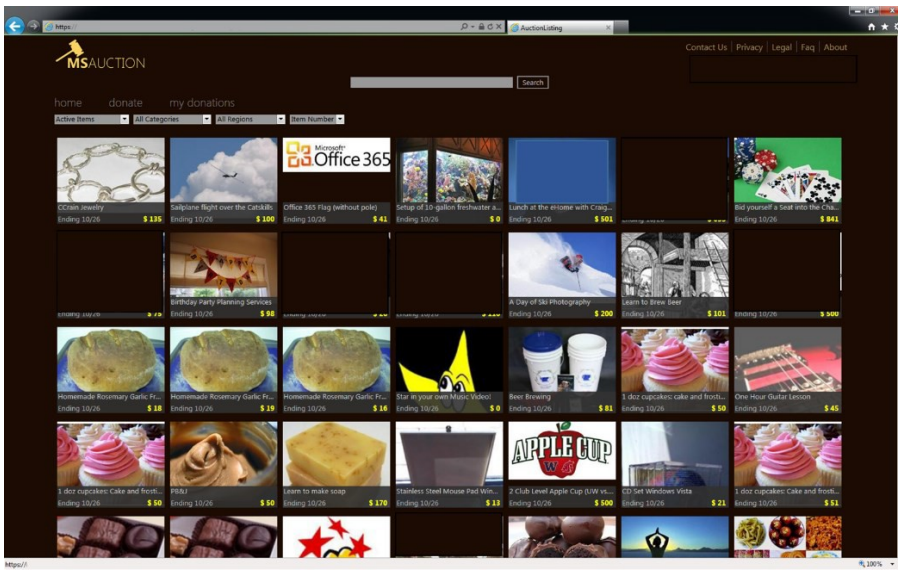


Figure 4-1: Microsoft internal auction application, circa 2010[3]

In the end, the application was very successful and the auction met its goals (over the years, incidentally, Microsoft's employee-giving has raised more than one billion dollars for charity). Microsoft IT learned 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 resources required to run it (as we would have—for servers, operations staff, and so on—had we run the application in our own datacenter). 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 analyze it using any of the big data and analytics capabilities in the cloud; build a dashboard or use machine learning to predict user behavior. These are all excellent ways to familiarize your staff with cloud services. These individuals then will become the core of your cloud center of expertise, passing along their lessons learned 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 recognize the importance—indeed the absolute necessity—of innovation: new programs 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 percent 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 publicizing a new campaign might simply create a new website on its own.

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

IT executives should realize 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 organizations to foster rapid experimentation and innovation.

Of course, it will be important to educate businesses on the importance and consequences of regulatory issues and noncompliance. 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 creates 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 both from successes and from 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 startups, 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 learnings 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, new projects can be "spun up" quickly with a few clicks rather than having to be planned, datacenter space allocated, equipment procured, 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 should be prioritized, 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 organization, with little return. Retiring them, consolidating them with applications that perform similar functions can, conversely, generate savings in a number of areas such as hardware, system software licenses, and maintenance. Consider generating metrics
- around "hot" and "cold" applications based on CPU, network, and database utilization; for example, an application that averages two percent of CPU and has few authenticated users might be 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 visualize success and impact as well as to understand the failure and the lessons learned 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 noncritical 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.

Further reading

Azure Resources

Microsoft web site, https://www.microsoft.com

Microsoft Azure web site, https://www.azure.com

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

Azure Trust Center, 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 data centers, 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 optimization:

- 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