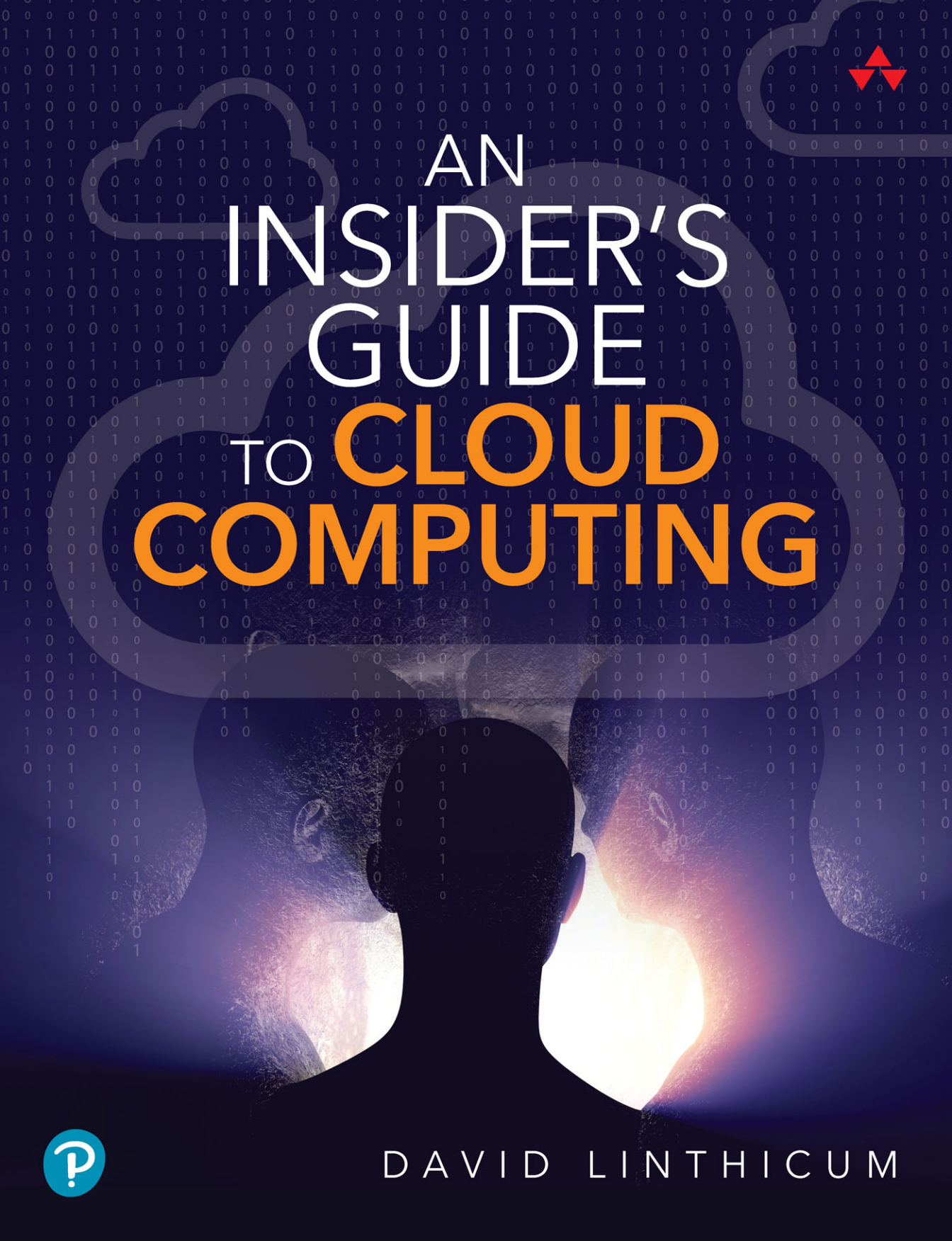
AN
INSIDER'S
GUIDE
TO CLOUD
COMPUTING
DAVID LINTHICUM

An Insider's Guide to Cloud Computing

David Linthicum

PEARSON

Finally, there are the opportunities and challenges around multicloud. Multicloud means you leverage heterogeneous public cloud services that allow you to deal with other services as the same. With multicloud, you can leverage many brands of storage and compute services using the best-of-breed services as well as least-cost services rather than deal with just a single cloud provider.

Leveraging Second-Tier Cloud Providers

Now that the concept of IaaS cloud computing is more than a decade old, we are beginning to see second-tier cloud providers enter the market. Of course, they only offer a fraction of the services you'll find within a major cloud provider, and they don't have as many points of presence (where the data centers are located), but a second-tier provider can offer a cost savings that makes them too compelling to ignore.

So, how much savings? Considering just compute, second-tier providers can offer the same compute configurations at prices 25–50 percent lower than those of the larger players, and you leverage the configurations in much the same manner. On one hand, they are not the premium brand, and thus you'll get some questions from those who consider the cloud brand inferior. But, if you save $10 million a year and your production workloads do not suffer, it may be worth the risk. Make sure to include testing, fully understand how you'll be billed, and do other due diligence, and the chances of success go way up.

Also, keep in mind that we now live in a cloud computing world where multicloud is now the norm. It's considered acceptable to leverage two or three major cloud brands. Add in the ability to place discount brands into your cloud services catalog to have them available for use, and this could be where the second-tier concept takes off. In this case, it's just as easy to attach a lower-cost resource such as storage and compute as it is to attach a major brand resource. Ease of implementation will push many enterprises to leverage the lower-cost resource, and perhaps send more savings to the bottom line. Also, by design, these lower-cost resources can work and play well with major brand cloud resources. Much of what is happening in the cloud world now and over the next several years will be a race to the bottom of the market.

Leveraging MSPs

MSPs differ a great deal in the services that each provides, but most will manage public cloud services for you, including provisioning, securing, and maintaining these cloud services in support of your workloads and data. They also support traditional systems, such as mainframes and traditional x85 such as LAMP-based platforms. In many instances, MSPs can be talked into managing more specialized systems such as edge computing systems and high-performance computing. MSPs often provide cloud-like services such as storage and compute that are less costly than the same services you might find within public cloud providers.

The advantage of using MSPs over traditional clouds is both cost and the ability to host several different platforms, including mainstream public clouds, and have the MSP manage those platforms. This means

you have a service running in front of your cloud service that removes you from much of the work and complexity of managing all those systems on your own. For many enterprises, working with an MSP is often a lower-cost choice when compared to the costs to manage all their compute and storage services. With hundreds of MSPs in the market today, this option will continue to grow in scope.

Multicloud by Necessity

We'll address the important topic of multicloud later in the book, so we won't get too deep into it here. However, when discussing the ability to leverage different cloud brands and cloud types to save costs, remember that multicloud is a core weapon to leverage to both reduce costs and focus on the use of best-of-breed cloud resources.

Multicloud by necessity means that we consider multicloud a tool to leverage different services to provide a choice that should lead to utilizing services that are more cost effective. Multicloud provides the ability to pick the exact right services your applications and systems need, or best-of-breed, as well as pick the services that are at a lower cost point. To achieve your optimization objectives, multicloud is usually a necessity.

Call to Action

Compute is a fundamental building block of cloud-based systems. I revealed some secrets here that you won't often hear from cloud providers or other sources. It's not because they don't want you to know the truth, but that the truth around using the cloud compute resource is still misunderstood on several levels. This is certainly the case when framed within the more holistic concept of building systems in clouds.

In this chapter, we focused on what these cloud computing services are, how to understand them, how to pick them, and how to obtain the best value from your cloud provider. The idea here is to learn the tricks and get the insider path on the current cloud reality. You should now understand compute and its related services in ways and with methods that can lower your costs of leveraging cloud compute instances and increase your productivity. Cost and quantity are often at odds, but they need not be.

Chapter 4

Innovative Services and Public Clouds: What Do You Really Pay For?

> *In a properly automated and educated world, then, machines may prove to be the true humanizing influence. It may be that machines will do the work that makes life possible and that human beings will do all the other things that make life pleasant and worthwhile.*
>
> — *Isaac Asimov*, Robot Visions

Now that we've covered some of the more mundane cloud services, including storage and compute, let's move on to the services that really drive cloud computing today. These more innovative and advanced services include AI, serverless, analytics, data lakes, and much more. They dominate much of what the technology press (including myself) currently writes about.

Here, we don't concentrate on the definitions of these services, which is information you can find in so many other places. Instead, we focus on how to optimize their value for the business. Most enterprises that rent advanced services from a top-tier public cloud provider only scratch the surface when it comes to their capabilities and potential benefits.

The cloud services of yesteryear were nowhere near as good as the traditional enterprise data center solutions. Those days are gone. Today's cloud-based solutions are multiples of times better than the traditional solutions, especially solutions that the top-tier public clouds can now provide. The reason is that the lion's share of all enterprise technology companies' R&D budgets goes toward cloud-based products and solutions. As a result, security is much better in the cloud, even though that seems counterintuitive to many people. There are also better databases, better development platforms, and better software packages available in public clouds.

Those who chose to stay off the cloud for any number of reasons will soon (or already) leverage inferior technology in the data center because these legacy systems receive less frequent or no updates from the provider, which means they also cost more to maintain and operate. There are reasons everyone is moving to the cloud these days. It's partly because the market evolved to give us little choice.

For our purposes, let's assume that moving to the cloud is a foregone conclusion. Your enterprise plans to leverage basic cloud infrastructure services, such as storage and compute as table stakes. The more innovative services covered in this chapter offer additional value to your enterprise, and these cloud services could even become game changers that redefine your business.

As we pointed out in Chapter 1, "How 'Real' Is the Value of Cloud Computing?" different layers define the value that cloud computing can bring to an enterprise. At the bottom layer we have basic infrastructure services (see Figure 4-1). These services support the applications and data storage systems that run the business. Basic infrastructure examples can include sales order entry, inventory management, shipping, logistics, and other systems that are common to most businesses. The target benefit of cloud computing at the bottom layer is to run these systems at a reduced cost, but these are just operational efficiencies. These are not game changer services.

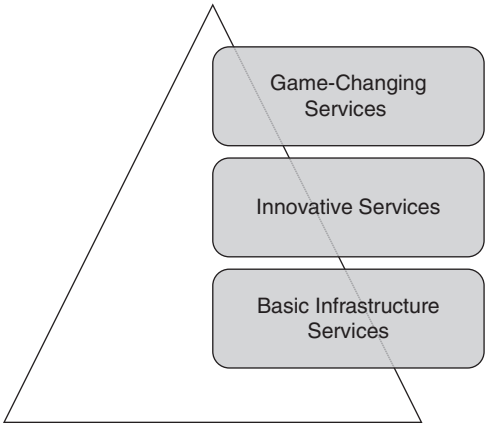

FIGURE 4-1 Cloud computing services can range from simple infrastructure services, such as those that run the business, to game-changing services, such as those that introduce innovative advantages that lead to the company becoming an industry disruptor.

Moving up a layer, we have innovative services. These services also depend on basic infrastructure services but provide true opportunities for innovation. Examples include advanced application development platforms and the ability to leverage services such as advanced data analytics, deep learning AI, massive scalability, proactive security, and anything that allows you to innovate or create unique differentiators.

Companies that leverage innovative services today include ride share companies that can leverage AI as a key differentiator to tell customers where their ride is located in near-real time, how much it will cost, and the almost exact time the ride should arrive. Ride share systems are based on simple services. By adding enhanced customer experiences using innovative cloud services, they can disrupt an existing traditional market.

For the most part, this chapter is about those innovative services. We look at what they are, what value they can bring, and how to effectively leverage them.

At the top layer, we have game-changing services. These innovative services combine to create net-new technologies and services that do not yet exist. A much earlier example of game-changing technology would be the invention of the airplane, which eventually changed how we travel long distances. These days, it could be a breakthrough in ways in which we cure diseases through genetic engineering or the ability to 3-D print construction components or entire houses. The possibilities are literally endless.

These are services and technologies that will change the way we think of things such as transportation, health care, housing, and other problems we want to solve, but we will do so in unique ways that did not exist prior to the appearance of these services and technologies.

Remember, each layer depends on the more primitive layer beneath it. Even the basic infrastructure layer where storage and compute exist needs to leverage more primitive components, such as a physical disk or a CPU, to provide the hardware services this layer needs. It's a good reference model to define different types of cloud services and solutions that exist and how they are all interdependent and work together.

AI/ML

Artificial intelligence and machine learning (AI/ML) are old concepts with new life, thanks to cloud computing.

Back in the '80s, AI was more of a high-tech experiment, with any AI advantages removed considering the high cost and lower technology capabilities at the time. You may remember computers playing chess, and IBM's Watson on game shows. It was interesting technology, but the business applications were few and far between, and it was still cost prohibitive for valid applications.

AI saw little use in the '80s, '90s, and even the early 2000s, until something came along to change all that…the cloud. Now that we can rent the infrastructure, as well as the AI engines themselves, the cost went from a minimum investment of about $1 million to a few hundred dollars per month. Today's AI technology is cheap, readily available, and with more advanced capabilities than in previous generations. It comes as no surprise that enterprises want to use AI in ways that will differentiate their business within the marketplace. Although you'll hear different definitions of AI and ML, at its essence, AI is a type of technology that leverages human-like learning capabilities. AI is the broad category of computer intelligence with implementations that can include services such as ML and deep learning (see Figure 4-2). Although there are many different definitions of ML, it's best thought of as a subcategory of AI systems that uses data to learn and find patterns. ML focuses on the pragmatic application of AI concepts that improve the value of AI for the business.

Today, most of what you see on the cloud is ML in the form of AI. ML can determine things such as a likely fraudulent transaction, in terms of patterns that may provide indicators. ML learns the patterns by learning from data specifically designed to teach the knowledge engine what certain existing patterns mean, such as data that is likely fraudulent or incorrect. It's our job, as humans, to define and set up the learning data that will "teach" the ML engine to spot specific patterns.

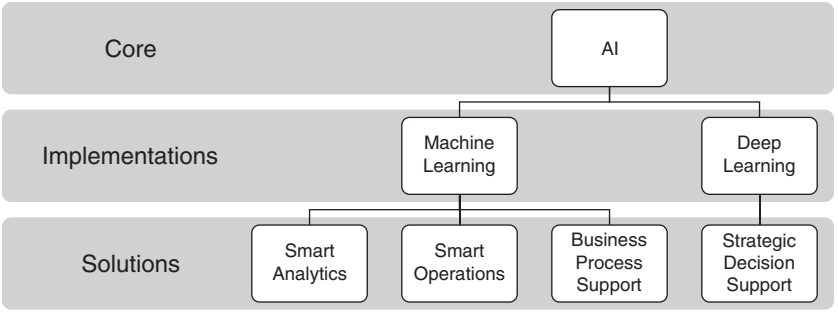

FIGURE 4-2 AI is the core concept, and machine learning and deep learning are examples of AI-related concepts or subcategories. Machine learning provides solutions related to analytics, operations, and business processing. Deep learning solutions include complex decision support and other solutions where a deeper understanding of the data is required.

Deep learning (DL) is the ability to leverage data and stimulus to find deeper meanings within data or stimulus. One can consider ML more tactically focused, meaning that we use it for more business purposes. DL can find deeper meaning within the data, sometimes finding a deeper understanding than we humans could find, such as the ability to figure out causes of some illnesses, where the problem is so complex that a single human is unable to process all of the knowledge that needs to be processed. That's why deep learning is often called "AI without limits."

It's not a good idea to get too wrapped up in the semantics of AI. You'll find that many in the world of cloud computing use AI, ML, DL, and other AI-related concepts interchangeably. While the PhD candidates may have issues with that approach, there are a lot of crossovers between the different types of technologies, which is why I just call it all "AI." Enough said about that idiom.

Overused?

Is AI overused? Perhaps initially when cloud computing providers began to offer AI as a service. The cloud made it cheap and readily available to solutions developers. As a result, AI found its way into applications that did not require AI capabilities, and the solution ended up less valuable. It's like putting high-end, high-cost racing brakes on a subcompact car. The car will stop just fine with stock brakes, while high-end brakes just waste money and resources.

These days we better understand the pragmatic use of AI: when it will prove worthwhile and when it will not. Business solutions that typically find the most value with cloud-based AI include

- **Business applications with potential patterns to find in large amounts of data.** These can be new patterns from new data, or new patterns that emerge based on what an AI engine already processes and learns from data as it processes over time. The more data that gets analyzed and the more patterns the AI system identifies, the better the AI engine gets at doing its job. We see this today in our daily lives, such as our cars learning from our driving patterns to accordingly adjust braking and acceleration. Even the smart thermostats on the wall can determine better