



## The Deloitte On Cloud Podcast

**David Linthicum, Managing Director, Chief Cloud Strategy Officer, Deloitte Consulting LLP**

**Title:** What you can do—starting right now—to optimize your cloud costs

**Description:** Who doesn't want more value from their cloud investments? In this Knowledge Short, David Linthicum discusses how organizations can increase their cloud value by optimizing their cloud costs. It starts by assessing the cloud infrastructure with FinOps practices to understand spending patterns, then developing a cloud optimization strategy. To implement the strategy successfully, however, he cautions that it's essential to build an organizational culture of optimization and accountability.

**Duration:** 00:22:07

**David Linthicum:**

Welcome to this Deloitte On Cloud podcast, Knowledge Short exploring a specific topic related to cloud computing. This is a short tutorial talking about the real-world concepts in the emerging world of cloud computing. I'm your host, David Linthicum, a cloud computing subject matter expert, author, speaker, and managing director with Deloitte Consulting and this is, "How to Optimize Your Cloud Spend."

This is becoming a very important topic now. Twenty twenty-two was a turning point in the world of cloud computing. We had lots of surveys go out, and it was consistently complaining about the cost of cloud computing, if you think about the lifecycle back in 2008/2009, we started to adopt the value of leveraging computing systems, typically infrastructure systems, storage compute, that were outside the enterprise, in other words, on a public cloud provider.

Certainly we've had lots of cloud growth in the last 10, 15 years, and lots of organizations have moved their infrastructure to the cloud. So, doing storage in the cloud, compute in the cloud, even using software-as-a-service systems and other things they were leveraging on-demand. What they discovered, and certainly this was the case in 2022, is that the cloud was a lot more expensive than they thought. So, they were looking at the bills they were getting, versus what they were paying for a traditional on-premise infrastructure, and they found that the cost was a little bit more than they thought.

So, we turn in 2023 the importance of optimizing cloud spend, and the ability to save money in leveraging cloud-based resources with this optimization process, the ability to get more governance control on how we use the cloud resources and how we account for the cloud resources, and optimizing as much as we can to make sure that we're spending the least amount of money and getting the most value out of these resources. So, if you look at the growth of things like FinOps, financial operations, it's really about that. It's the ability to put cloud technology, and accounting technology, and monitoring technology, and observability technology in place, so we can look at all aspects of how we're leveraging cloud-based systems and make sure, number one, we're accounting for the usage.

We know when they're being used, why they're being used, who is using what applications, and what resources have to be connected to that resource. In many instances, we may have a storage system that leverages an application server, and we have to account for those resources that are dependent one to another. So, it's an incredibly complex array of things that we have to track. And, certainly, we saw the growth over the last five years or so in terms of multi-cloud—and multi-cloud really makes things way more complex, because not only do you have the complexity of accounting usage within a single cloud provider, which can be difficult unto itself, but the ability to do so within multiple cloud providers who all use different billing mechanisms and rates, and they bill in different ways, and the ability to normalize that with how we're accounting for resources not only within a single cloud provider, but spanning many cloud providers.

So, that's why FinOps is so important. People are looking at this as a discipline, a set of processes and some automated things that have to be in place. So, we remove the human beings from sitting down and trying to manually monitor these systems and control these systems and govern these systems, but do so through an automatic mechanism that is able to track all these things, account for these things, even control, allocate costs, create budgets in terms of what would people leverage, what they should be leveraging, what they can spend things.

That's one aspect of it. So, that's us understanding how we're spending money on the technology. The other aspect of this that I think is more important, and that's the ability to optimize the cloud spending. That means that we're not only looking at what's being spent, but we're actively and proactively trying to put in mechanisms that are going to optimize the amount of value we're getting out of these cloud resources for the money that we're putting in.

Obviously, that's doing things such as turning storage systems off that you may not need, not over-provisioning particular compute instances or storage instances, really fundamentally getting to the fact that we're going to make these things as efficient as possible. Therefore, we're going to ensure that the value that we're getting out of this is going to be maximized, again, where the money we're putting into it is going to be minimized. We're getting the best performance out of it. We're getting the best security. We're getting everything that we need, and we're doing so by paying the least amount of money to get there. That's the core idea.

So, how do you get there? Well, a few things that we recommend would be, first, assessing your cloud infrastructure. You evaluate your cloud usage and cost. You identify your areas of potential optimization, what we can do to optimize this particular resource, even using different resources, maybe using it from one cloud provider to the other or moving something, a workload or a data store from one cloud provider to the other, the ability to right-size instances and resources, in other words, make sure that we're putting them on the instances that are most optimized for whatever that resource needs to be doing, analyzing storage usage and data lifecycles, also reviewing network and data transfer costs.

Much of the cost that we run into with cloud computing is not necessarily compute and storage, but it's moving information into the cloud and out of the cloud, ingress and egress costs. They charge you a lot for that. You have to make sure you minimize that. If you are doing a lot of that, that's typically going to be a cost overrun. You need to figure out a way to optimize that. Then identifying under-utilized or idle resources. So, we'll find out that someone is running cloud-based systems that they're allocating storage systems that they never deallocate. They may allocate a petabyte of storage, which is incredibly expensive, and they just leave it running.

Even though they may need it once or twice a day for a few hours, they're not necessarily putting that resource back. So, they're still getting billed for the resource, and then you get this big cloud bill at the end of the month. So, this is, as you can see, it's common sense. In other words, we're basically trying to optimize the infrastructure in much the same way we'd optimize our power bill or our water bill or gasoline consumption. We're just doing pragmatic things to do the right things to reduce cost.

Next would be implementing a cost optimization strategy. Once we assess the cloud infrastructure, we understand what we've got now and what we need to do to optimize it, next would be implementing a cost optimization mechanism and strategy and plan in order to get to that better optimized state. Here we need to utilize reserve instances or savings plans, if needed. And those are the ability to buy ahead of need. In other words, cloud providers will say, "If you want to have a reserved instance, you can buy this resource ahead of the need, and we're going to give you a discount for doing that." That ultimately provides savings, as long as you're able to use that reserve instance.

Sometimes people will reserve instances because they think there's going to be a need. Leverage spot instances for noncritical workloads, the ability to spin things up and spin things down as we need them, and the ability to leverage the discounts we can get from spot instances to find better cost savings there. Adopting serverless computing or containerization. This one is a bit difficult because there are some trade-offs here as well. Serverless infrastructure, by the way, really isn't serverless because what we're able to do, it's able to auto-allocate the resources you need to run a particular function.

So, if you write an application, it's going to be a series of functions that runs on a serverless infrastructure, and when we execute that function it only spins up the compute and storage systems that we need, and then puts those compute and storage systems back. So, we don't have to worry about it. In a traditional cloud environment, we have to allocate storage and allocate the compute resources that we think we're going to need for the time that we're going to need them, and hopefully we're making the right guess in terms of the number of resources that we need to allocate. Then we need to remember to put them back where we found them after we're done, and this removes the stress of doing that. In other words, we're leveraging serverless computing, and it's going to allocate the resources and deallocate the resources, so there is no thinking that we have to do.

The issue comes in sometimes that it can be more expensive than someone who does a really, really good job in sizing a resource, versus using a serverless system, where the serverless system sizes a resource for you on your behalf. In a few instances, you're going to find out that the serverless function may cost more, because they are not necessarily trying to optimize everything for cost. They're trying to optimize the serverless resource for the requirements of that particular function, and they may overestimate or underestimate the resources you need based on the profile of the function. So, you're giving the spending over to another automation that is going to do this on your behalf versus you doing it.

I'm finding that in most instances a serverless resource will be able to save you money because we run into problems with cost optimization, because we as humans aren't very good at estimating what we think a particular workload will need. We may overestimate, which means we're spending too much, or underestimate, which means that we have a performance issue or the system crashes for some reason because we don't have enough storage or compute resources there.

Then the ability to explore hybrid cloud and multi-cloud strategies. So, again, we're looking at very complex infrastructure that runs across multiple cloud brands. So, we deploy different techniques for cost optimizing on those particular clouds. Normally that's going to be leveraged through the control plane of some sort of a FinOps system. FinOps systems now don't do more than just accounting, but they're able to monitor all of these resources we're using, and then allocate and deallocate the resources in many instances on your behalf.

So, it becomes the huge control plane for dealing with cost, not only within a single cloud provider, but typically multiple cloud providers, in many instances on the existing traditional legacy systems as well, and even extending that out to edge-based systems and IoT-based systems. So, they're very functional. We've been growing this technology for a while. It's out there. You should take a look at it if you get a chance. Then also implementing automated scaling and resource management. In other words, we're putting automation in place to make sure, very much like the serverless thing we just talked about, that we have other processes that are auto-scaling, auto-allocating the resources that we need when we need them.

So, instead of having a human being look at something and estimate what we think the resources need to be, sometimes with some very sophisticated profiling tools we're going to put the automation in place. So, we scale out the resources we need when we need them, storage, compute, databases, and we scale them back when we don't. By putting this automation in place, we don't have to deal with it as human beings. We don't have to remember to deallocate those resources. It's done for you on your behalf, very much like a serverless system, but we're doing so with another mechanism within the cloud.

So, you need to monitor and analyze cloud cost to figure out where you are. So, you leverage a cloud provider's cost management tools to do that in many instances. Those have a tendency to have lots of limitations. Number one, they're set up for that native cloud provider. So, if we have multi-clouds, typically we're going to have to deal with the cost analysis and cost tools on that particular provider. So, typically it's better to have a FinOps tool that's able to look across cost across cloud providers.

Then setting up cost allocation and tagging mechanisms. So, we're tagging the resources, so we know what's being used by whom, and they have a meaning within an accounting system, so we know what they are, and then implementing budget alerts and cost controls. In other words, people are given a budget. They can only use this many dollars, or this many resources, and they're not allowed to go above it. If they get close to it, they start getting e-mails warning them of this and we don't allow them to spend more than their allocated budget.

Obviously, that can be problematic if people don't do some good cost estimates on what resources they're going to use, but you have to put some controls and some accountability there to make it work. So, also reviewing and optimizing cost reports. In other words, looking at the costs that are coming out of the system. Then as a human being, and also as an automated mechanism, look for areas where we can optimize cost. Sometimes those are going to be done through automation. The FinOps tools can do a lot of those things automatically, but again, as human beings we have to take a look at this, see what resources are being leveraged and for what purposes, and look for opportunities to shift things around and do things differently.

Also, security and compliance considerations come into the mix here. You need to ensure that we're complying with all the regulations and laws, and those are going to be depending on state, local, and federal governments that you're under control. So, we need to make sure we're doing so for the least amount of cost. We're finding many instances that the people who implement IT systems in the cloud may not understand fully the laws and regulations they're dealing with, and sometimes overestimate or don't understand something and end up spending more money than they need to on what they think is complying with the regulations, where you're just looking to just comply with the regulation without over-complying, because that's going to be additional cost. So, understanding those is going to be very important.

You need to implement security best practices to make sure that we're not being penny wise and pound foolish. So, we're going to spend money on the security mechanisms that we need and make sure we're doing so in an efficient way, but we're not underspending and then putting the infrastructure at

risk for breaches. Then evaluating cost implications of security measures. This is where trade-offs occur. There are lots of different cost points in dealing with different types and models of security. One of things that we have a tendency to do is we want to get the best security that we have available to put in place on our cloud-based systems. But in many instances, there is going to be a cost trade-off between different security models, different encryption systems, different identity access management systems and what they cost to deploy and to operate, and you need to consider those.

In other words, we always want security, but if there are different cost levels of different security approaches, we need to understand the trade-offs there and make good decisions in terms of what's going to provide the best optimization, best security of our resources for the least amount of money. Then continuous improvement and optimization. We need to review and adjust the cost optimization strategies that we're looking at continuously. This is a never-ending process of looking at the FinOps strategies, looking at the cost optimization strategies, and making sure we're adjusting to the changing needs of the business, but also its technology changes and also as we get smarter. We learn things as we go. We're going to make lots and lots of mistakes, but correcting those mistakes, learning from those mistakes, and then ensuring or basically reducing the risk that they would happen again, cost overruns.

You really need to get into the vibe of rethinking cost optimization ongoing and understand the technology is going to change. Best practices are going to change. We get better at it as we go, and that needs to be factored into how we're doing cost optimization strategies and automation. Staying up to date with cloud providers offering price changes is a big one. Now some of the FinOps tools will do that on your behalf. So, they know that the prices are changing and they know maybe the terms are changing, different ways in which we're accounting for and charging for storage, compute, but you still need to keep up with them as a human because even though the automation is doing it for you. Because there may be a time when it's cheaper to shift a resource from one cloud provider to another. If the prices have gone up on a particular resource that you need and you're able to find a good analog on another cloud provider, the transfer costs and the migration costs aren't that out of whack, and you can leverage a better, more cost optimized tool that exists on another cloud provider.

That's the nice thing about having multi-clouds. You're still locking yourself into clouds in many instances, but we're also able to compare and contrast different infrastructure prices and different support contracts, different SLAs across different cloud brands, and make a decision as to where our net new workloads, storage and compute, will run and how we're going to leverage the stuff in a more cost optimized way.

Encourage a culture of cost consciousness and optimization within the organization. This really gets to the accountability aspect of it. I think one of the reasons that we got in trouble in 2022—and many enterprises were taken aback by the amount of money that they're spending—is the fact that there was no accountability for people who were allocating these resources. So, it almost feels like it's free. If you leverage cloud computing, you're looking at a portal infrastructure. You're allocating storage. You're not having to physically bolt storage systems into a rack. It's done in a virtual way for you. Yet the costs in many instances are the same if not more. So, we had a tendency to decouple ourselves from the reality that these are very expensive resources that we're playing with.

We really need to understand people who are administering these systems, developers, designers that they're responsible for minimizing and optimizing cost as best as possible and for taking steps in making that happen. So, that comes through accountability. In other words, they're accountable for the money that they're spending and they need to justify that spending pattern.

Finally, We need to look at the key strategies for optimizing cloud spend and make sure that we're doing the optimization correctly. Think about everything I just said in the broadcast and look for ways to implement it within your own organization. Also, there are huge benefits from doing this. We can get the same value out of the services, in many instances, and spend five times less in terms of the money that's being needed to support these services. In other words, in just dealing with better optimized processes and dealing with intelligent tools, FinOps tools, cost spending observability, if we have better monitoring and better budget restrictions, and better governance in place, cost governance, we're able to reduce the costs by a great amount and not give up anything in terms of performance.

In many instances, we gain some because we're understanding what our applications and our data stores are doing, and we can better allocate the costs to make sure that we're serving those processes, but doing so in a way where it is the optimized architecture, the optimized solution, but also we're removing lots of cost from that spend because there's no inefficiencies. We're removing as many inefficiencies as we can. Enterprises are finding out right now that reducing spend by a great deal because of cost optimization processes, they're not giving up anything. They're not giving up storage. They're not giving up performance. In many instances they're increasing those things.

This is about observability monitoring and constantly improving. We need to get really good at being self-critical. In other words, what are we doing well in terms of cost optimization? What mechanisms are working well? What tools do we need to make this happen? Then also looking at the trade-offs in the different things that are occurring. In many instances, we're already good at this as humans. We optimize our own daily spend based on the budgets that we have. Therefore, we're trying to get more value out of the money that we're spending, as much as we can. This is no different when you look at really what it is, but very complex in how things are billed to us. Accountability is very complex.

How it works across cloud providers is very complex. The good news is, we have tools that make that much easier to deal with and can automate the ways in which you can optimize these systems. Get at this. I think this is an important topic for cloud professionals right now to understand how to do this well, and it can save your company a lot of money, money that can be invested in other things that are more important. So, if you enjoyed this podcast, make sure to like us, rate us, and subscribe. You can also check out our past episodes, including those hosted by my good friend, Mike Kavis. Find out more at [DeloitteCloudPodcast.com](https://DeloitteCloudPodcast.com), all one word. If you'd like to contact me directly, you can e-mail me at [dlinthicum@deloitte.com](mailto:dlinthicum@deloitte.com). So, until next time, best of luck with your cloud journey. You guys stay safe. I hope you enjoyed the podcast.

**Operator:**

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to [Deloitte.com/about](https://Deloitte.com/about).

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Visit the On Cloud library  
[www.deloitte.com/us/cloud-podcast](http://www.deloitte.com/us/cloud-podcast)

#### About Deloitte

---

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see [www.deloitte.com/us/about](http://www.deloitte.com/us/about) for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Please see [www.deloitte.com/about](http://www.deloitte.com/about) to learn more about our global network of member firms. Copyright © 2023 Deloitte Development LLC. All rights reserved.