# Deloitte.

# The Deloitte On Cloud Podcast

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

**Title:**          **This is 2023 in cloud: Current trends and a look at what's on the horizon**

**Description**:          Cloud is always changing, and in those changes lie great opportunities to return value to the business. In this episode, David Linthicum and Mike Kavis sit down to review those opportunities and debate cloud trends so far in 2023. They also try their hand at making predictions for the second half of the year and beyond. They touch on all the hot topics, including the impact of generative AI, cloud value and ROI, cost optimization via meta- and super clouds, and the ever-present cloud talent shortage.

**Duration:**          **00:23:33**

**David Linthicum:**

Welcome back to the On Cloud podcast, your one place to find out how to make cloud computing work for your enterprise. This is an objective discussion with industry through leaders who provide their own unique perspective around the pragmatic use of cloud-based technology. Today on the show, inviting him back, good friend, colleague, and host of "Architecting the Cloud," Mike Kavis. How're you doing, Mike?

**Mike Kavis:**
Doing pretty good, Dave. Long time. Long time.

**David Linthicum:**
Yeah, and catch the listeners up. People have been following you pretty closely, lots of people have read your book, love your book, "Architecting the Cloud," that you wrote a few years back. What have you been up to?

**Mike Kavis:**
Well, I had a little AWOL action. Life happens. Had to deal with some stuff, so I've disappeared a little bit from the socials and the podcasting, but I'm back ready to roll, and my focus area lately has been a lot of this gen AI stuff we might talk about today and resiliency, chaos engineering, SRE, on that side of the house, which is ironic, I grew up on the other side of the house. I was a Dev for many, many years, and then I got into cloud and I shifted to the Ops side. So, that's where I'm at these days.

**David Linthicum:**
Yeah, it seems like Ops is where the problems are to solve right now. If you look at cloud optimization, the ability to do operations better and get into FinOps and get into operation of AI systems, really coming out as things we need to do to improve the way that we're leveraging cloud computing, and other technologies for that matter, to again maximize value that's returned to the business. At least that should be the objective.

So, we're going to go ahead and take a look at mid-year outlook, and we're going to talk about where we are. Here we are in June, we'll give you a take on what's happening in the marketplace and what's likely to happen, which I think is a discussion that's even more important. Lots of things are changing right now. We see some downturns in some industries, people are contracting a bit. However, technology seems to be front and center, and certainly the revolution of generative AI systems—and, of course, cloud computing is a core catalyst to that, seems to be a big focus. So, do you think we're going to be spending more or spending less on technology now that gen AI is front and center in our investments?

**Mike Kavis:**
We're always going to spend more. That's the way this all works, and we always think that the new thing replaces the old thing, but it's always additive. So, you've still got all the old stuff, you're going to build some new stuff with new ways, going to need new skills, new tools. We're never going to spend less.

**David Linthicum:**
Yeah, I think you're right, and I think that we're gearing up for, I think, a big spend in 2024 as people understand that this gen AI stuff is finally going to be transformative to finally bring value back to the business. You've got to remember AI's been around since the 1950s conceptually. And I'm not a young person. My first job out of college was as an AI programmer, a LISP programmer, moving on to M1, taught it in college, things like that. It's just been something that's been in the background that hasn't been very pragmatic because it's been so expensive and the capabilities weren't necessarily there. So, when gen AI and some other AI revolution started people think that since Chat GPT showed up and everybody made a big deal of it, that's an evolution of things that occurred for many, many years. But the idea is we can finally make this stuff affordable and put it in the hands of small businesses, and even departments in big companies with small budgets, to allow them to finally punch above their weight.

Obviously, cloud computing had a key benefit of doing that because we were able to leverage multi-tenancy and leverage technology at scale and do all the things that cloud computing's able to do, but now we've got some real decisions in terms of we're going to bet the business on this. We're going to automate supply chains, we're going to provide better customer experiences, finally go through digital enablement of various core systems to allow us to become an innovative business where we're selling ideas and IP, and that's really the most valuable thing that we have as a business. Heavy investments need to be made in this technology, and it seems like people need to be stepping on the accelerator in 2024. What are your thoughts?

**Mike Kavis:**
I agree. This game is always about, "I need to do more for less, I need to get to market faster," and this is another enabler of that, and I'll put in quotes, "if done right" just like everything else. So, I'll give you a real simple example is when it first came out, I started messing with it. I'm a developer for many years, but I don't do it every day now, but that's just an example of coding. You go into operations and we're going to talk about FinOps and stuff, it's the same type of thing where it can accelerate decision-making, it can accelerate results, but it's a very smart engine, but you have to know how to ask it the right thing to get the right things out. That's the catch.

**David Linthicum:**
You talk about developers, we've had code generators and certainly AI-code generators for a long period of time, but your ability to become the orchestrator of developing these systems so you're asking the AI engine the correct question to get the correct code and integrating the code to make the final solution as optimized as possible.

And I think the focus is going to be on that. It's not necessarily on people who understand how to code in an optimized and productive way, and I think that's a good skill to have, but your ability to build these complete systems with all of these various generators, whether it's dealing with database design, code generation, where we're centralizing decision-making and putting in the hands of something that's going to be able to make a decision that's a lot smarter than we are because it has the knowledge of 10,000 developers and therefore can look at the best practices and what are the most optimized coding decisions, which we typically can't make as humans, but we still have to be able to ask the questions in terms of how these things are going to be assembled, and I think that's really the core focus of where we need to be right now in terms of an industry. What are your thoughts on that?

**Mike Kavis:**
I agree, and you remember when cloud computing was going to get rid of all the server admins' jobs, and now we have more cloud server administrators than we probably had physical infrastructure. But, anyway, the same thing here. It's a tool in a toolbox. It helps in some areas, but there's a skill that's almost like Sherlock Holmes type skill of figuring out the best way to extract the best value out of these things. I look at it as a copilot, and there's actually a product called Copilot built into these interfaces where you code, but it's right there like para-programming, and you have this really smart partner there, but they only know what they know. You need to fill in the gaps for them, and together you can really crank some stuff out. And people ignoring it, ignore it at your own risk. This is the future of development.

**David Linthicum:**
Yeah, and I think there's going to be some market changes that occur around this as well. I think the rise of people who can operate these AI systems, we're in a vacuum where we're building a lot of these things, but I don't think we have a good understanding of how we can operate these things at scale, the ability to provide AI-based security, and those are always challenges, and certainly we have bad actors that are weaponizing AI to come after your systems in more creative and innovative ways, and they'll learn as they go and get better and better at it. So, we have to understand how to defend against that, but that really seems to be where the focus is going to be, so I'm not sure it's going to be a huge amount of changes other than the term we're leveraging things that are more AI powered, and I think we're going through some AI washing now. I think everybody who's putting out technology—just like cloud washing. Now everything's AI powered, and yeah, why wouldn't it be. I baked that in the cake. I've done the math already. I understand that people are going to leverage this technology for whatever benefits it can find in there. But there's going to be some evolutionary changes, not necessarily revolutionary changes in terms of how we're hiring and how we're configuring AIOps and different operating models in how we're working and how we're doing development, and all these things are really going to be evolved and pushed into the existing way of doing things. I think the core thing is that we have to pick the right use cases for this. My big fear—and this is, by the way, the problem with AI when I first got into it—it was overused. In other words, they were putting it in different use cases where it wasn't necessarily needed and overcomplicated the problem. Do you think we're going to have a problem 2024-2025 where lots of things are going to get built using AI technology when it wasn't really needed and it just added cost, complexity, and risk to the projects?

**Mike Kavis:**
I mean, we've seen that. We see it with cloud, we see it with everything. It's just like, "Here's the hammer, let's go find some nails with it." I've seen it before where it's like, "Hey, let's build this proof of concept, and I've got somebody coming, "Let's implement blockchain here." I'm like, "It's a proof of concept. We've got a few weeks. We don't have time." So a lot of times it's like, "There's this new tool, let's go implement it." And there's a time and place for everything. So, we'll definitely do that. We've both been doing this for years and years and years, and it's the same thing, just new buzzword.

**David Linthicum:**
Yeah, it's about defining the business value in terms of your ability to understand what this stuff is going to bring in terms of value back to the business, and those are calculations that typically are often overlooked as we're building architectures. And you've got to remember, just like everything we've done, you talk about this in your book, and when we're building particular solutions, there's really five factorial ways in terms of the number of iterations and permutations that we can do it, but there's going to be only a few ways that are going to be completely optimized.

We're going to optimize the value that comes back to the business. I think we need to think in terms of that, that ability to get something up and running that works or something up and running that's buzzword compliant is not necessarily where we need to be pointing the objectives. It's the ability to get something that's as optimized as we possibly can, and, therefore, it runs on less cash, and, therefore, it returns the most value back to the business. I understand I'm the designated buzzkill in saying that, but I think that's what businesses are going to demand moving forward.

Speaking of business value, there's been a refocus on ROI for cloud computing, and certainly on FinOps, to get us to an operational control of various costs and usage within these systems. And what we saw in 2022 was lots of studies that came out and businesses say, "Hey, we're not getting as much back from this cloud-based spend that we thought we were. It's way more expensive than we thought, we're burning more cash than we thought, and we're not getting the business value that's coming back." In many instances, they were complaining they're getting negative business value, so actually moving to the cloud made things worse in terms of the ability for IT to bring value back to the business.

And now we're into an optimization era where people are looking at becoming more efficient with the existing systems in terms of cost burn, turning your instances off, the ability to leverage the exact amount of computing and storage that you need, versus overleveraging and overprovisioning various systems, and FinOps or financial operations, provides us tools and technology to look at those things in operational format and, so, we can understand who's consuming what for what purpose and what reason. And there are ways in which we can do to better optimize the systems, also do demand planning where things are moving forward.

So, we finally have visibility into what we're spending, I guess after 15 years of cloud. So, what's going to be the trend in time? Are we going to become more aware of what the cloud is spending and how to optimize these systems, or do you think we may fall back into the under-optimization of systems where we're doing sporadic spending and not necessarily being accountable and responsible as we should be?

**Mike Kavis:**
I think we're going to be more cost aware. I don't know if we'll ever be great at optimization, but definitely a lot of people get sticker shock. Before cloud, no matter how good or bad you were at cost optimization, you could tell what your costs were. You have infinite number of servers, you just knew, whether you were good or bad at it. Then you go to the cloud and everything is elastic, and you can get a bill that's 3X and not even know it's coming. So, I think we'll get really better at predicting and forecasting our costs. We'll probably tighten some screws on being better at optimization, but will we ever be really, really good at it?

In a lot of cases we weren't really good at it in the data center either. We just knew what it was going to be. But that's the thing is a lot of big surprises because leave a Kubernetes cluster running worldwide for a while and forget to turn it off and your bill's going to go through the roof. I've done that., That's

why we started putting controls on stuff. Those types of things we'll get good at catching. That's where things like AI could help as well, but it really comes down to architecture. You should be building systems with cost in mind, and that's not done a lot.

**David Linthicum:**
No, it's not, and I think that I'm a little bit more positive outlook. I do think behaviors are going to change moving forward. We're able to see what we're spending. Back when I first got into computing working for a timeshare company, I did the same thing. I would get into loops that would just hum—a bunch of CUs, which were how they billed in terms of money, and we're seeing the same thing in the cloud. I think as long as the thing—we're providing or delegating the capability of people to leverage or provision and manage their own components, we're going to have this issue, even if there's visibility into what those components are doing and charging, there's not necessarily been a focus on optimization. I think moving forward, the developers become better at power optimization. Certainly the focus on the sustainability stuff, ESG stuff is going to have people more focused on how much they're burning because that goes to what their carbon footprint is, so there's going to be some subtle changes moving forward.

I agree with you that we're probably not out of the woods yet in terms of our ability to understand where the money's being spent and what value we're getting back from it, but we're on our way. And I think that it's going to be a while. It's going to be some technology out there, but I think I'm a little bit more optimistic in terms of our ability to finally get this thing under control. It's like we get an electric bill every month and a water bill every month. We're very aware what we're spending and how much money it's costing us. Probably the same thing with cloud computing as well. They just weren't getting the bills. We're getting the big bill every month but not necessarily understanding details behind it in terms of what resources they're burning.

So, we're moving to public clouds and really moving to connective tissue, what's in-between the various public cloud providers. They've even given it a name called super cloud and meta-cloud, and all that is really a layer of abstraction, a logical layer that sits above the various cloud providers and even the enterprise systems and edge-based systems to provide common services like common security, common operations, common governance, common application development, and the ability to think in terms of not necessarily pushing things into the walled garden of public cloud providers, but the ability to build things that exist above them and are able to work in-between these various systems. And obviously it's able to reduce redundancy, hopefully reduce costs, become more simplistic in terms of how we're going to operate these systems. We're dealing with complexity problem now. That's where it comes. What do you think actually changes within rank-and-file enterprises that are leveraging cloud computing now and certainly in a multi-cloud configuration?

**Mike Kavis:**
When we first started this, there was these things called CMSs, these cloud management systems, and that's exactly what they were trying to do is an abstraction over all these clouds. The problem was that people were hardly in cloud, and if they were, they were either doing something in their private cloud or finally getting something to public cloud, so they were ahead of their time. A lot of those were purchased, and now they're probably called something else. And, also, when we first started, there was a lot of PaaS, but they were all public cloud like Heroku and some of these others. And then it brought that to private cloud, and pass was an abstraction over all these other clouds.

Now we're calling it meta-cloud. The concept has been there forever. Probably the tools have matured over time. A lot of tools right now just changed their buzzword to meta-cloud, but this is nothing new here. I think the only thing that is new is we're so much further down the line that people actually have a valid use case to use these tools. We're actually now in multiple clouds and all that. The name sends chills through me. This is not new. So, I get off my soapbox there, but I think this is where we're going. I always say multi-cloud's usually not a strategy; it's a reality. It's like you have multi-cloud, now, "How do I rein it in," rather than, "Let's go multi-cloud." This is one of those areas where it's, "What's the software that can help us rein it in?"

**David Linthicum:**
Yeah, one of the things I talk about when I speak on super cloud, meta-cloud, and multi-cloud management stuff is this is an old architectural trick. There's nothing new here. This is abstraction. Automation is the ability to mitigate complexity, which we've done with distributed systems for years, and we're doing the same trick in a multi-cloud environment because the same problems are emerging. We have just too many moving parts that we're having to manage and monitor, and we can't operationalize these systems with the levels of complexity that there are.

There's just too much going on, and so the ability to leverage abstraction, to combine various systems, the ability to do operations under one single layer, even though we may deploy it natively in different ways, we're dealing with one set of common interfaces and dealing with the same security, and same with AppDev tests, and all these sorts of things which will really provide us with the ability to allow these things to scale and also afford them. I think one of the things that many of these enterprises are doing now is they're rolling out multi-cloud deployments. They're hitting a complexity wall where we can't basically—and you look at the metrics—spend 2.5, and that's typically what it's going to be, times the amount of operations on maintaining these things. There's just no money in the budgets to make those happen. You can't go back to your board of directors and say, "Remember when we said cloud was going to save us money? Well, we've got to spend 2.5 times the amount of operational dollars to make this happen."

And, so, that's why people are interested in these architectural best practices that we're bringing back we've used over and over for years. I think they're proven in how it works. It's fairly simplistic in how we're solving the problems. It's going to take a while before the technology providers step up and understand what roles they play because right now I'm seeing a bunch of them saying, "Well we're a meta-cloud or super cloud." No, you're not. You need to provide some common layer that works in between these various systems that's capable of being set up quickly and operated quickly and is able to work and play well with existing technologies that are out there. So, got a ways to go, but unto itself, to the point you just made, this is fairly simple and this is where use of technology's been around for a long period of time. What are your final thoughts on this?

**Mike Kavis:**
Well, a lot of this is also because we're still in the same silos we were in in the data center, so there's all these handoffs between groups, and if we could get to, "Here's what I'm delivering as infrastructure," and provide the tools and service to provide to the next layer, which is, "Here's where we are from cloud platforms," providing cloud services up to the developer, and then you have a team that does the build and run for their app. When you get to that, it's

much simpler, but when it's the Dev team throws it over to this team, then you have the multiple complexity, and you need all these tools, and you need this single pane of glass, and everything needs to be the same. I just think a lot of this is a case of we're taking the mainframe operating model to the cloud.

**David Linthicum:**
Absolutely. And remember this is covered in the tech trends that Deloitte puts out that covered the whole super cloud, meta-cloud thing, and has an objective view as to what value this stuff is able to bring, what it can do and what it can't do. I urge you to check those out. One of the things I looked at when a bunch of layoffs were occurring and things like that, and maybe a bit of a downturn in the industry at least was described by the tech media, we're still at a two percent unemployment rate in technology, which is below natural. And, so, when I'm talking to people out there, even though you'd think that maybe demand for skills would fall, they don't seem to be falling, and the supply of skills doesn't seem to be increasing. What are your thoughts on this?

**Mike Kavis:**
Well, part of it is the speed of innovation. There's so many new things popping up, like AI all of a sudden is the buzzword. Well, how many people have ten years of AI experience? Well, the people who built the tools. So, you jump on that and then machine learning will be, and big data over there, and IOT. So, it's hard enough just to get enough people who know cloud, but all the applications there are, or all the types of use cases all have their own complexities and their own new skills, so it's just so much new stuff. It used to be a lot easier. You had mainframe and you went to client server. That was a multiyear step. Now it's like every week there's something new popping up. So, there's such a variety of skills on top of not having enough for cloud. There's just such a variety of skills we need, it's like a cat chasing its tail. You just never catch up.

**David Linthicum:**
No, you never do, and I think what we're seeing—you hit the nail on the head—is that the technology is changing extremely quickly, and I think people need to really get into the fact that if they're going to have a skill that's going to maintain their skills, the ability to learn on demand and be a continuous learner, and also don't wait for people to train you, the ability to go out and get the information that you need either on-demand training, read books, things like that to keep your skills up to date, that's probably the best way to manage your career, and I think it's going to be an 80/20 thing. I think 80 percent of the companies out there are chasing 20 percent of the talent pool because that 20 percent of the talent pool is willing to do the extra mile to keep their skills updated, and become a continuous learner, and really never be satisfied with what they know and understand. And I think that's where things are moving right now.

So, if you're in a static learning environment, you're waiting to be trained, that's probably not the best way to manage your career. It's going to be going off and reading books like Mike's book, and the ability to do some on-demand training, and the ability to even look at social media and the different news items that are coming up, what's important. You get a sense of where the market's moving to and then adjust your skills accordingly. So, if you have gen AI skills in the cloud, obviously your dance card's going to be full for the next one to five years, and same thing with operational control, and same thing with DevOps. It's just pragmatic the way you should look at how to manage your career in light of the fact that this stuff is going to be ever changing and ever dynamic. What are your final thoughts on this?

**Mike Kavis:**
I agree. The days of, "I know what my job is when I come in, and I'm going to do this checklist," are limited. It's like it's very dynamic and the technology changes, so if you like change, this is the place for you.

**David Linthicum:**
Absolutely. And we'll go ahead and leave it at that. So, if you enjoyed this podcast, make sure to like and subscribe on iTunes, or wherever you get your podcasts. Also, don't forget to rate us. Also, check out our past episodes, including the On Cloud podcast hosted by Mike Kavis on his show "Architecting the Cloud." Check out Mike. He's back. So, if you'd like to learn more about Deloitte's cloud capabilities, check out deloittecloudpodcast.com. If you'd like to contact me directly, you can reach me at [dlinthicum@deloitte.com](mailto:dlinthicum@deloitte.com). So, until next time, best of luck with your cloud projects. We'll talk again soon. You guys take care.

About Deloitte

-------------------------------