On Cloud Podcast January 2025

Deloitte.



The Deloitte On Cloud Podcast

Host, Mike Kavis, Managing Director, Chief Cloud Architect, Deloitte Consulting LLP

Title: Look back at cloud 2024: Key trends and tech that will impact 2025 and beyond

Description: In this episode, Mike Kavis spotlights key trends in cloud and software engineering for 2024. From intelligent edge computing and APIs to industry cloud platforms, hybrid cloud, and the role of AI in automation, this episode features highlights from the year's most compelling conversations with IT leaders. The overarching theme of cloud for 2024 is that technology innovations are driving significant gains in operational efficiency, competitive advantage, and business value.

Duration: 00:17:06

Mike Kavis:

Welcome back to the On-Cloud Podcast where we get real about cloud technology. I'm your host, Mike Kavis, and in this special episode, we're taking a look back at 2024, and what a great year it was. It's been an exciting year in the cloud space full of breakthroughs, challenges, and incredible insights. In today's episode, we're featuring some of the best moments from our conversations over the past year, sharing key tips, expert advice, and the most impactful trends that shape the cloud landscape. So, whether you are a cloud veteran or just getting started, you'll want to stick around as we dive into the highlights of 2024.

One of our standout episodes of 2024 explored the transformative power of edge computing with Carnegie Mellon's Satya and Deloitte's Rahul Bajpai. Together, they discussed real-world applications to interplay between edge and cloud computing, and innovations like edge-native applications in 5G. Here's a highlight from that conversation.

Satya and Rahul Bajpai

The cloud has strengths. It's very important not to lose sight of. First of all, data centers are like fortresses. It is the most secure place to put your data to perform a computation. And the cost per unit of compute is the lowest that it will be anywhere because the economies of scale drive down the per unit cost. So, in terms of longevity of data, if you need to put data in a place that you are sure will be there five years from now, 10 years from now, we'll be backed up. We'll be secure and so on, where you can execute applications which do not have any of the constraints that we just discussed that Rahul and I talked about a moment ago. The cloud is the right place to do it.

So, the goal is not to replace cloud computing by edge computing. But it is to recognize that they both have extremely important roles to play and to leverage their strengths to create end to end architectures that seamlessly combine their strengths from the viewpoint of the user at the edge or the cyber physical system at the edge.

Mike Kavis:

Another standout episode from 2024 explores the exciting frontier of AI and the edge featuring Qualcomm's Atul Suri and Deloitte's Rahul Bajpai.

Atul Suri

So, like you said, Mike, the edge is becoming more intelligent. What does that mean? In the example that you gave, the edge always had digital technology. You had sensors. Those sensors were creating an output. Intelligence at the edge really means that we're bringing real time data and analysis. And the purpose of doing all that is to make time critical decisions in real time that will bring a further improvement in productivity. It will bring a further enhancement of workflows. It will make processes that, like you said in your example, that take two days, happen instantaneously. So, this idea of immediacy, this idea of computing and real time data analytics driving additional performance and productivity gains, to me that's what this whole thing is about.

Rahul Bajpai

Yeah. And the couple of things, just from a use case perspective that come to mind, given that we have trials some of those in our Smart Factory. One is real time quality detection and sensing, which really allows right to get as narrow and as deep as we need to, in terms of monitoring an assembly line, looking at errors in machine parts in the form factor, or even looking at the thickness of a paint coat on the door of a car in a car factory. Being able to track this in near real time as the production assembly line progresses and creating that automated feedback loop or that feedback control that allows the anomaly to be detected, analyzed, and fixed without having a lot of manual or human intervention. And that is the end goal of any predictive quality sensing and detection.

Mike Kavis:

Next up is a clip from Knowledge Short episode where Jamie Sawchuk discusses groundbreaking innovations and climate mitigation from AI-enabled geospatial platforms to renewable energy transitions and how they're fostering sustainable growth.

Jamie Sawchuk:

After the disastrous 2018 California wildfires, innovators began using AI-enabled geospatial platforms to mitigate climate risks and adapt to climate change. Think about the trees in our communities that are weakened by drought conditions. When the wind comes up, there's an increased risk that tree blows into a power line and starts fire. Innovators are now using AI and drone LIDAR and satellite data, essentially geospatial data to identify the trees close to the power lines and take action before the tree falls on the line. At scale, California innovators are mitigating 85% of their wildfire risks before they occur. They're using this same technology to drive 10% to 50% operational efficiencies and increase asset resiliency.

Mike Kavis:

In this thought-provoking conversation, Abi Noda, founder of DX, challenges conventional metrics for assessing developer productivity and highlights the importance of considering developer experience and team dynamics.

Abi Noda:

If it's fundamentally flawed to be trying to measure engineering output, then what are other ways we can measure how good our engineering organizations are and how productive they are? And that kind of brought me down the path of instead of measuring the output of what is produced, let's measure the environment. Let's measure the process and if the process is optimized, then we can presume that the output is also maximized and that is what gave birth to what we refer to as developer experience, which is really the conditions for effective and efficient software delivery.

Mike Kavis:

Also, Roger Premo joined us this year to discuss reshaping the organization for cloud's third wave. He discussed the principle of hybrid by design and his philosophy of intentionality.

Roger Premo:

There are a lot of factors, but the complexity in particular got in the way of getting the business value out of those migrations to the cloud. And the real thought on this third wave of cloud is based on early adopters that have done this well and they've been very intentional about taking a hybrid by design approach. That is the notion of creating platforms that span that complexity that you abstract away some of that complexity and really can get to business value.

That intentional architecture really does have a series of different kind of technology and simplification benefits that are part of unlocking that value. And so, we see that as critical because I think in a lot of different industries, the adoption of cloud technologies, the adoption of AI now, it's not an optional exercise that to differentiate and win in your industries, you have to do that well. And if you're doing that on a kind of poorly constructed kind of base architecture, the chances of success are very low. And it's really about helping our joint clients succeed in their digital and AI initiatives.

Mike Kavis:

Deloitte's Brian Campbell joined us this year on a solo Knowledge Short episode where he did a deep dive on the "what, why, and how" of industry clouds.

Brian Campbell:

When you think about the history of technology over the last 20 years, in particular cloud platforms, they've gone through an evolution from infrastructure as a service to platforms as a service and ultimately to business outcomes as a service. And, really, the way that last piece of the puzzle is being accomplished by defining specific business problems, which when you think about the product that commercial, the operation sides of many businesses, they're so unique to the industries they operate in, whether it be healthcare, banking, software developments, oil and gas that each of those different value chains needs to be addressed in a way that the business outcomes of that industry are still being supported.

So, the portfolio of these Industry cloud solutions that are evolving and being served up across the market is expanding exponentially, particularly given today's environment. And technologies like GenAI that are increasing at the pace at which these are developed. Some of what make them unique versus sort of historical solutions, they're business oriented, but industry specific. So, again the needs of a healthcare provider are very different than the needs of a commercial bank and the different business processes across the value chain that are solved, automated, increased efficiency by applying technology to those different areas of the business are oriented towards those specific outcomes that enterprise and that industry are looking for.

Mike Kavis:

Take a listen to a clip from this compelling conversation with Salesforce's Param Kahlon and Deloitte's Kurt Anderson on how APIs drive digital transformation.

Param Kahlon and Kurt Anderson:

One thing that we're seeing our customers do is leverage those APIs to drive reusable architecture in the enterprise and then do it in a way that you're using a code-gen approach that is supported by AI. So, instead of coding everything yourself, you want to say that I've got a catalog of APIs. And I want to expose the catalog of APIs to my large language model to be able to create the outcome that I want to be able to create and what that approach gives you is that the skillset of being able to code all those APIs together can be achieved of course. But sometimes it's hard to find the skillset takes a bit longer time.

But when you apply natural language capabilities to make someone who's doing it way more productive, you're able to hit some of that gap and make people more productive. So, that's one value that we're seeing in people saying that, well, once we build the underlying investments in components like APIs, we're now exposing them to many, many more developers who were able to leverage them where they could be application developers or integration developers. They're able to leverage them in the output that they're creating more effectively by leveraging their large language model.

Mike Kavis:

This year, a good friend of mine DevOps Leader, John Willis, joined us for a three-part conversation. Here's a clip from the third episode, the series on why and how to avoid AI anarchy.

John Willis:

And, so, there's been a lot of movement in general IT of how to sort of eliminate that sort of like these ivory tower people sitting who knows where are demanding that everybody does it this way. And I've done qualitative analysis with large banks where these young kids will say, "John, just find this person that owns this control." So, I can just have a conversation with them about how insane this control is. I mean, in large organizations, they're like, "Who owns that risk?" It can't even be navigated. It's the companies are so complex and big that no one person could actually tell you own what that risk is owned by this guy. Here's his location, his number. Just call him and once you have a conversation about how insane that risk that was developed 25 years ago.

Mike Kavis:

The next clip from this episode with Deloitte's Scott Buchholz where he gave a primer on quantum computing. Let's take a listen.

Scott Buchholz:

But simply know that by and large, when you read about quantum computers today, those things are not currently capable of running production workloads. Depending on whose roadmap you believe, those may be a handful of years in the future. But the other thing to keep in mind is that quantum information science, a lot like data science, is a discipline that takes people a couple of years to master, or at least to learn, let alone start mastering. And as a result, you really need to think about getting started sooner so that you can try to shoot the curl such that when the hardware gets to the point where we can start running production workloads, that your quantum information scientists are ready to run those workloads and that you have identified the problems that are tractable to those systems.

Mike Kavis:

Google Cloud Next 2024 is always a fun conference for us to debrief about. The next clip features Bill Briggs, CTO at Deloitte.

Bill Briggs:

Yeah. So, I was with some of the Google engineers yesterday. I like the framing they have kind of the three-piece layer cake of the model garden like at the below and obviously there's a big piece of the Googles advances in Project Gemini, but also, they're open to others, so that is important. How do we think about the model building there and then Nicole, you said the agents. My new favorite word is Agentric, by the way Agentric. I'll be using that with some attribution in the weeks to come.

Mike Kavis:

Our last clip revisits the episode with Connie Fan product strategy and analytics lead at Google and Deloitte's Dr. Abdul Rahman examined the growing relationship between generative AI and cybersecurity. Let's hear from Connie.

Connie Fan:

Across the GenAI space right now, it really is just the core concept, though that LLMs are not just generating text, but they are able to call an API and so to one of Abdul's previous points of there's a lot of great gun intelligence out there, but to operationalize it as an entirely different question, to be able to securely call the API's of your Intel provider, you'll do extract from that API call what the IOCS or indicators of compromise, like Abdul was mentioning for a certain threat actor. And then have the LLM stick that into whatever syntax for whatever investigated query is relevant to the tool that you're using to be able to do your day job is like a very quick turn that was previously inaccessible. The crux of the question is that the best way to secure the data is to not touch it at all. And when we need to call it, we're calling it from the already secure API and then the ability to do that is not just purely for the standpoint of securing the data, but really to be able to access fresh and timely data and sensitive data to be able to make that very useful in office for various users.

Mike Kavis:

That wraps up our 2024 highlight reel here on On Cloud. I hope you've enjoyed these key takeaways and learned something new to help you on your cloud journey. Thanks to all of our amazing guests this year for sharing their expertise, and of course, thanks to you, our listeners, for tuning in. If you enjoyed this podcast, make sure you like us, leave a review, and subscribe. You can also check out our past episodes wherever you listen to your favorite podcast and you can always reach out to me directly, mkavis@deloitte.com. Feel free to write to me with your questions, which we can address on future episodes. Thanks for listening, and we'll talk to you in 2025.

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.

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

About Deloitte

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see 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 to learn more about our global network of member firms. Copyright © 2025 Deloitte Development LLC. All rights reserved.