



The Deloitte On Cloud Podcast

Bill Briggs, Chief Technology Officer, Deloitte Consulting LLP

Title: All about Google Cloud Next 2024 with Deloitte's Bill Briggs, Nicola Johnson, and Mamoun Hirzalla

Description: In this episode Deloitte's Bill Briggs, Nicola Johnson, and Mamoun Hirzalla discuss Google Cloud Next 2024—what's new and what's next. Hint: It's all about Generative AI. Highlights include increased capabilities for Google Gemini and Model Builder and a clear focus on reducing model hallucinations and improving results. Bill cautions, however, that data strategy and core modernization are keys to successful implementation. The team also discusses the future of cloud and Generative AI.

Duration: 00:17:44

Bill Briggs:

Welcome back to the On Cloud podcast. I'm Bill Briggs, chief technology officer at Deloitte, and I'm your guest host for today's show. Today I'm joined by good friends and colleagues, Nicola Johnson, the US chief commercial officer for Deloitte's Alphabet/Google relationship, and Mamoun Hirzalla, Deloitte's Alphabet chief technology officer. Hey, Nicola, Mamoun. Welcome to the show.

Mamoun Hirzalla:

Thank you, Bill.

Nicola Johnson:

Thanks, Bill.

Bill Briggs:

Yeah. All right. Tell our listeners, where are we today? Give them a sense of the scene.

Nicola Johnson:

We are here in the wonderful Las Vegas, and we are here at Google Cloud Next '24, which we had a wonderfully successful Google Cloud Next '23 in August in San Francisco.

Bill Briggs:

So, Mamoun, what's the energy like as you've been meeting with clients, going to sessions?

Mamoun Hirzalla:

Seventy thousand people. Electrifying energy. Everybody wants to get their hands on the latest and greatest shiny objects that they've seen on day one. And every client is asking, "How can I get that into my organization?"

Bill Briggs:

Well, so, we're going to go through some of the highlights, the announcements, what we've seen, what we've heard. Clearly, the thing for me, AI as the centerpiece, which is fantastic. As we've been pretty loudly discussing over the years, AI is an important breakthrough. And it isn't the hero of the story. It requires a lot of other pieces to come together. And so, what I love about Next is seeing the advances in infrastructure and chipset, edge, the broader

tooling that you need for engineering and experience wrapped up into of course models, Agent Builder, and the like. We'll dig into all of that. Nicola, what are some of the headlines and highlights so far for you?

Nicola Johnson:

I think one of the biggest observations overall, and as I mentioned, we were back in San Francisco in August for Next '23, the difference between Next '23 and Next '24, when we were there in August it was almost like a developer's conference that we were in the middle of. It was the cusp of what was to come. And what they've actually shown here, Google Cloud has actually shown here at Next '24, is the ability to translate from a developer form into a business-like benefit with what they can do with the Generative AI capabilities. I mean, the demos were shown along the lines of just a consumer through to an enterprise and the benefits it could bring, translated well to reality for a businessperson sitting in the audience. And I think that's really key because business owners and leaders of organizations can now go away and, as Mamoun said, get excited as to, "When can I get that in my organization? When can I bring the benefits of Generative AI to my organization?"

Bill Briggs:

Yeah. And, "I want that. And I want to apply it." And even from the folks I've talked to, it did seem like an interesting mix from last year, the architect engineer will continue and always be a part, but I've talked to more CEOs, members of boards this time, and a lot of them it's their first time to Google Cloud Next, and I think that trend's going to continue, yeah.

Nicola Johnson:

I think that somebody put the translator button on from where we were six to eight months ago.

Bill Briggs:

From text features into actual action.

Nicola Johnson:

Yeah, that's right. I mean, obviously this is predominantly Thomas Kurian, CEO of Google Cloud is an engineer.

Bill Briggs:

Yeah.

Nicola Johnson:

A very, very smart individual. Thinks in bits and bytes. And I think he's put a wonderful team around him to help bring this into the business environment and show the difference that Google Gemini can bring with all the bells and whistles that we're going to get into, to be a differentiator for our clients, their clients, and together clients.

Bill Briggs:

All right. In no particular order, headlines that jumped out to you.

Mamoun Hirzalla:

Headlines that jumped out. I would agree with what Nicola mentioned earlier. So, the fact that we were eight months ago in San Francisco, I think the writing on the wall was Generative AI. And I think there were lots of hints that were dropped in the previous conference that is one of the major trends that are coming, and I think everybody knew that. Now, fast forward eight months later, I think you see it everywhere. "Gemini everywhere" really is theme of this conference. You look at all these agents, some of them were announced earlier, some of them were massively enhanced to accommodate additional features and capabilities.

And I think a lot of people right now are still digesting, "What does that really mean to me from an organization perspective, from a software development perspective, from a productivity perspective?" I don't think honestly that all of us realize what this thing all means right now. I think everybody is intending to come and see "What is the impact? And how can I go and update my methods, procedures, to make sure that I leverage that?"

Bill Briggs:

Yeah. So, I was with some of the Google engineers yesterday. I like the framing that they have, kind of the three-piece layer cake of the Model Garden at the below, and obviously there's a big piece of Google's advances in predictive Generative AI, but also they're open to others, which is important. "How do we think about the model building layer," and then, Nicola, you said, "the agents." And my new favorite word is "agentric," by the way. Agentric. I'll be using that with some attribution in the weeks to come. But there were announcements at each layer.

Mamoun Hirzalla:

Correct.

Bill Briggs:

So, should we maybe unpack them?

Mamoun Hirzalla:

On the model side they were talking about the enhancements and access to Gemini 1.5 Pro that's being available in preview mode in 180 countries through the API. This is very fascinating because typically you wouldn't get that massive access across on day one, so this is one thing unique that they have done. The lighter models, the code, the Gemini models are being available now to you. They can run faster now for different types of purposes, so this is something also that they've added. The open-source type of models that are available to you, this is a big differentiator for Google. You are not just married

to Google-provided models. You have cloud, you have other open-source, and all of these things are available to you. If you happen to want to leverage these type of capabilities, they're right there.

Bill Briggs:

Just switch the API call and away you go.

Mamoun Hirzalla:

Exactly.

Bill Briggs:

Or reload the context and—

Mamoun Hirzalla:

Exactly.

Bill Briggs:

Yeah. I think Nicola, one of the headlines that I've seen just in the reporting, the million context window is a big deal with Gemini 1.5 pro. I'm not sure we know exactly what to do with that yet, which is exciting. What did you see—the middle piece, I think, is a sweet spot, as we're helping our clients put together the studio and their own ability to harness AI and Generative AI for sure, but seeing how that gets embedded in. So, this idea of the investments in Model Builder, what jumps out for you on that?

Nicola Johnson:

I think some of the key attributes of what they're bringing together now, the ability to combine much more the audio, video. Here we are in the public domain, the ability to leverage at the enterprise level. So, I personally think it's going to be a game-changer. Now, where I think there will be challenges, though, Bill, and I think Mamoun and I have talked about this quite a bit with our alliance team, is the one area that nobody really talks about so much is the change management aspect of the human acceptance of some of this change.

We're talking about unbelievable processing speed. We're talking about unbelievable capabilities that are completely multimodal in the way we as a human being operate. We're talking about processing power that's starting to think and can play back. That's really key. To me, one of the main things that I heard was this: I'll translate what they said: It remembers. We're moving into a completely different area of what this technology can do.

Bill Briggs:

Yeah, to me, there's a couple big things of it. One is that, the creativity of how do we just not think about what we've always done and how we've always done it to unlock new potential is a piece. So, that creative—and the force multiplier of humanity and thinking about how to apply it. But then, also, the best use cases in enterprise are taking things that historically have not been digitized or are based on years of tribal knowledge and institutional wisdom, typically in individual employees, like the artists and work process, at a factory floor for tooling, at a hospital tool for the operating theater, and we're saying we could do better. Now, we're going to amplify and empower the humans in the loop. But there's also that you have to trust the model, you have to trust the machine, and you have to actually take advantage of the advances. Yeah.

Mamoun Hirzalla:

Yeah, and I think to augment what you said, Google, everybody has some concern about the results from Generative AI. Sometimes you have to go and validate yourself and all of that. And the elimination of hallucinations is one of these key problems that makes all kinds of LLMs. So, Google basically released two key features that will significantly reduce hallucinations. The first one is you ground the results in enterprise data. And that's massive. Basically, you have your own inventory, your own data, your own customer information. You can have access of that data through the LLM, and then now you are 100 percent sure that this data is accurate based on real-time potential inventory that you're getting from it. So, one, trust level.

The second thing is you can ground data with Enterprise Search across many facets. That combined with the enterprise data and Enterprise Search, I think hallucinations will be something of the past once this method gets fully incorporated and added to the process of the results of Generative AI.

Bill Briggs:

Yeah, and in the public internet and just real-time Google search is another layer of grounding as well.

Mamoun Hirzalla:

Correct.

Bill Briggs:

Which is a great complement between—and I think, Nicola, you started with the difference eight months ago with potential energy but not a lot of real use cases. And those that were doing it was proof of concept and early pilot. This idea of how do I get from 60 percent accuracy, which makes for a great demo to get people excited, to 90-plus, 95-plus percent accuracy with grounding and reinforcement learning and tuning, that's the difference between taking something and actually vetting it in operations and transforming your business. And we've crossed that path.

Nicola Johnson:

We've crossed—I think we've crossed over that line. I think that one word that's been used in our conversation here a lot now, and I'll tell you just to take us back a step a little bit into reality type of it, is data. Every Google executive that we've had the opportunity to meet with this week as well: data, data, data. It keeps coming. So, the ability to be able to manage that data better, the ability to be able to consider holding it in one or two places rather than all over the place, and the ability to get hold of it when it is in disparate models, again, for what Google is trying to enhance here is the ability to pull and center the

right data with all these different opportunities around Generative AI, which is not what we've seen in the past. And data, just basic data, is our bread and butter at Deloitte. It's one of the key aspects of what we can bring to an organization to help them get the benefits out of this future technology, really.

Bill Briggs:

Yeah. I was on a panel in—with some of the more business execs, and one of my headlines was about that need and how none of the advances have made that fundamental—the criticality of that fundamental issue of data, data management, data strategy is more important than ever. Sorry, we don't just go and get to skip that step. And the other piece is on the core modernization, which sometimes can feel like a really loaded concept. But as we move to the agents—so, the third layer—and say we're not doing standalone in a chatbot, we're embedding it in workflow, we're embedding it in business processes, we're embedding it in customer experiences, that takes us to the sometimes third rail of you have to invest in your core systems if you expect to harness the full potential because that's where the action, that's where the data lives.

Mamoun Hirzalla:

That's right.

Bill Briggs:

And that's where work is done and customer experiences are had.

Mamoun Hirzalla:

Agreed. So, I'll give you an analogy. General practitioners, doctors, they know certain things. But when you want to go to get the root cause of certain things that you don't know, you go to a specialist. I think LLMs definitely is going into that path. There are these LLMs that are general. These are the ones that are specialized. And then, you can build these type of agents on top of these specialized ones. Now, what can make it even greater is the multidisciplinary interaction between all these specialists coming together because now we can harness not just one agent. You can harness all of them. And I can see some additional use cases later that you can look at manufacturing, you can look at supply chain, you can look at IoT, all of these things can—and logistics and scheduling of deliveries and all these things—I can see many of the manufacturing and CPG companies looking at these type of scenarios and saying, "Now I can ground it in my enterprise data. I can ground it in Enterprise Search and Google Maps and all of these type of logistics. Now, let's build something very useful together with that type of scenario."

Nicola Johnson:

And just to pile on there just a tiny bit, Google Alphabet, not just Google Cloud, the products and the solutions that the whole organization has to offer is seldom represented the way it should be. But what you could do with the combination of the media, things like the data from a media and digital advertising, marketing, as well as Workspace, ChromeOS, Pixel, YouTube. I can foresee with the movement they're making with Gemini and Generative AI in general that it won't be too much longer that we may be starting to see more of an Alphabet—

Bill Briggs:

Integrated with offerings that—

Nicola Johnson:

That's right. I mean, look what we saw this week. Google Workspace, BigQuery, Gemini all pulled together to do exactly what Mamoun has just given the example of. And if you then take it up another level, it's almost stratospheric.

Bill Briggs:

If the listener could see my technologist geek propeller spinning, because I think that's the biggest— potential. An advanced connectivity. We've talked a little bit about edge, but advanced networking connectivity. Quantum and advanced computing. So, beyond just TPUs and GPUs, into the next realm. Those are all pieces of the broader Alphabet unlock that it hasn't been announced yet, but you can see the breadcrumbs of where we're headed. Okay. So, I was going to close this with asking you each for a bit of what you're excited for, what's to come, and a bit of prediction.

Nicola Johnson:

I would say that what we want to do as the alliance is truthfully bring all this excitement in-house to Deloitte as well. I want to see us being able to leverage what those opportunities could be in our clients, educate our partners, our directors, all our practitioners and resources so we can bring this to life for our clients. That to me is that step for us at Deloitte and what we can take out of this Google Cloud Next '24.

Bill Briggs:

All right, Mamoun. What's a highlight to look out for?

Mamoun Hirzalla:

So, I'll mention two things. I'll put on my delivery hat as well and I'll say, "Okay, great. We need to make sure that we have the right assets for the right accelerators and things that will lower the barrier for entry into these type of things." That's one of the key areas that we need to bring more things to our clients. The second thing is I think the multidisciplinary use cases, how all of these things will mesh together. We've started discussions with clients on all of that. To get that to reality would require a ton of collaboration internally and making sure that we have the right minds from an industry perspective and technology perspective to bring visions that never have been thought before, to bring them to our clients.

Bill Briggs:

Okay. Well, it's been an exciting few days. All right, Mamoun. Where can we find you on the web or on social? Any place that folks can track you down?

Mamoun Hirzalla:

On LinkedIn I typically post—for day one, I did one. I will do another post later on to recap and provide some reflections on the event.

Bill Briggs:

All right. Mamoun Hirzalla—H-I-R-Z-A-L-L-A—at Deloitte. Nicola, how about you?

Nicola Johnson:

I'm on LinkedIn as well.

Bill Briggs:

If you enjoyed this podcast, make sure to like us, rate us, and subscribe. You can also check out our past episodes at DeloitteCloudPodcasts.com, all one word. You can find me, Bill Briggs, posting on LinkedIn. Feel free to give me a follow or reach out if you have any questions. Until next time, thanks for having me, all. And stay safe, everyone. Take care.

Mamoun Hirzalla:

Thank you, Bill.

Nicola Johnson:

Thank you.

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 © 2024 Deloitte Development LLC. All rights reserved.