

*This transcript is provided for the convenience of investors only, for a full recording please see the Scotiabank Global Technology Conference [webcast](#).*

**Will Grannis, CTO of Google Cloud at the Scotiabank Global Technology Conference on December 5, 2023**

**Steve Sparkes, Scotiabank:** Good afternoon. Thank you for joining us for the first Scotiabank Tech Conference. Really glad to have everybody here. We're thrilled with the participation, and I'm even more thrilled to be on the stage.

My name is Steve Sparkes, I'm the CISO and head of Enterprise Infrastructure for Scotiabank. And I'm joined today by Will Grannis. Will is just an excellent, excellent, excellent individual all around.

But before we get into the Q&A that we've prepared, I'm just going to read the mandatory Safe Harbor statement for the benefit of everyone's safety.

So some of the statements that Mr. Grannis may make today could be considered forward-looking. *[Additional comment added by moderator: In fact, I actually hope that many of them will be. But some of them could be considered forward-looking.]* These statements involve a number of risks and uncertainties. Actual results could differ materially. Please refer to Alphabet's Form 10-K, including the risk factors section in the 10-Q. Any forward-looking statements that Mr. Grannis makes are based on assumptions as of today, and Alphabet undertakes no obligation to update them. That is the end of the Safe Harbor statement.

So now I would like to tell you a little bit about Will, for those of you who don't know him. I have been lucky enough to have a few interactions over the years. And, really, again thrilled that he's here today.

He's the Chief Technology Officer at Google Cloud where he has a team of tech execs and engineers who are helping get the cloud to business. Been there for --

**Will Grannis, VP and CTO, Google Cloud:** Almost nine years now.

**Steve Sparkes, Scotiabank:** -- almost nine years, from the early days of Cloud. And as all of us have seen, Google has invested substantially in Cloud to make it a powerhouse platform, and Will is a big part of that.

Before joining Google, Will was an entrepreneur, a tech exec in a number of different sectors, so it's not just finance, although that's, obviously, our particular sweet spot today. But he's been a developer; and engineering leader, still very, very close to the tech; CEO; director; has, in the past, graduated from West Point; jumped out of a bunch of planes; landed on a bunch of people; lived to tell the tale.

And so today we're going to get to hear a little bit about what you do and how you see things developing. So we'll kick off this with a little bit of background of what does that role encompass

for you as CTO for Cloud.

**Will Grannis, VP and CTO, Google Cloud:** Great. Well, it is certainly an honor to be here with all of you.

Steve, great to see you again. Great excuse to bump into you.

So my role in Cloud, I lead a team that does two things. Number one, for our top customers and partners that are moving to the cloud, we help them navigate those first -- those first steps into the more complex patterns of adoption. And this is around the world, every geography, every part of the stack, all the way from customers that are really interested to optimizing against infrastructure all the way up to SaaS and everything in between in the platform.

An example of this -- you know, three or four years ago, we had customers asking us to create natural language interfaces to documents. Does that sound familiar to any of you, based on recent events?

And so in the period of the course of years, we worked with customers to take the best of research that we're developing at Google, and the products we're developing at Google, and help them realize the potential of that technology.

Now, you can imagine working with all of our top customers across every industry and every geography, we also learned quite a few things while we're doing that. And so the second part of what we do is we take these emerging patterns and we start testing them.

And so over the years, we created new product strategies, new entries into our roadmap based on the signals that we received from the market, our customers, and the work that we do.

An example of this is we did some work with Unilever around sustainable sourcing and palm oil and converted that into a bigger push around sustainability for all of Google Cloud that now includes climate risk analytics. And there are now APIs that Google produces in the data commons where you can get massive access to climate data through a simple API. So those are the two branches of the work that we do.

And one of the best parts of my job is getting to spend time with our customers and see the implementation in the real world because we are really an applied team. We care about making technology useful to all of you.

**Steve Sparkes, Scotiabank:** And so how -- that's a fantastic degree of customer engagement from a CTO. And so how has the scale of the team grown? You are running, what, over \$30 billion run rate? That's a heck of a growth story. How has your world grown to support that?

**Will Grannis, VP and CTO, Google Cloud:** Yeah, well, I think since Q3 of 2019, we've added roughly \$24 billion in annualized run rate. So it is a little bit different business than it was just a few years ago.

And my history dates back to roughly nine years ago, when we were in this kind of first wave of

Cloud at Google. We were just starting to create the pathways for all of this technology out into the world. And then there was a second wave where it was creating kind of the platform that would allow any industry or any customer or any geo to solve their problems.

And now we're really in the scale phase, where we've got an ecosystem that is rapidly growing, over 100,000 partners in our ecosystem today. So almost any implementation a customer asks for, we have foundational technology and we have a partner ecosystem that can help in the last-mile implementation: planning, training, and the rest.

And so for our team, what it's really looked like is making sure that our customers are getting the most leverage that they can from the technology. And I guess I could kind of break it into the stack, if you want, because this is -- in Cloud, we kind of think of this as an end-to-end platform so any company, any organization, can achieve their goals and digital transformation or critical mission.

So that starts at the bottom, which is kind of -- if you think of a typical stack for those of you that are more technology oriented, so infrastructure, we today have over 39 regions globally. For example, we have a commitment by 2030 to operate entirely carbon-free energy powering all of those data centers, and more that we will build between now and then.

So pretty much making sure access to compute is ubiquitous and also sustainable, as energy considerations weigh on people's minds and sustainability weighs more and more on company's minds.

If you move up the stack a little bit, so the second part of this -- you know, this kind of differentiated cloud that we put out is really around data and AI. Here's where you can see Google's legacy quite a bit really showing, where making data accessible and useful is super critical to us.

So to those of you that -- many customers want -- all of our customers want, they want their data -- make it so that it's easily useful to them in building experiences for their customers or in refining their processes internally. And they don't want to be constrained by the choice of a single vendor.

And so, for example, things that we've done over the years are -- BigQuery. Has anybody actually heard of BigQuery here? It's one of our hallmark technologies. Okay, great. I saw some hands.

Welcome to the technology conference, it's great to see you.

BigQuery. In 2020, we released a capability called BigQuery Omni. And BigQuery Omni allows -- basically colocates compute and binaries inside of other clouds like Amazon, Microsoft, and allows you to send a query across any cloud that you want without having to egress your data and bring that back and bring only the results back.

And you can imagine how important that is to companies who have data stores in multiple clouds. So now instead of having to deal with the cost of egress in these complex technical bits,

we've solved this through this product we call BigQuery Omni. So that's the next layer of the stack that's kind of a platform layer.

Then if you move up, we've got what we call the Security Cloud. So this is end-to-end comprehensive security, not just infrastructure security but also application security. And we brought in a ton of new capabilities, one of which is Mandiant, to round out the threat mitigation, consulting. And then we've also had VirusTotal, Chronicle, and a number of capabilities over the years to be able to characterize threats and reduce threat surfaces across an entire enterprise. Then kind of moving towards the top, collaboration cloud, if you want to think about it like that, so infrastructure, platform -- infrastructure, data, AI, and then up security, now collaboration.

This is where you get the work done. This is where you create -- this is where your workflows exist in your company. And so examples of this are like Workspace.

And what's really interesting now is how AI and what we've been bringing to market this year, how AI is now cascading. It's kind of like a thread that runs through all of those, from the infrastructure, the data, the platform, all the way up through SaaS.

**Steve Sparkes, Scotiabank:** I think -- since we're a few minutes in, we haven't specifically hit Generative AI, but we really must.

**Will Grannis, VP and CTO, Google Cloud:** There's a jar somewhere.

**Steve Sparkes, Scotiabank:** Yeah, exactly. We have to put at least another couple dollars in it. I think it's probably worth just starting with the foundational.

How do you think of Gen AI being different from 'vanilla AI', if we can call it that? And then let's talk a little bit about the foundational benefits that need to be in place for Gen AI to take off.

**Will Grannis, VP and CTO, Google Cloud:** Sure. So let's see. How do we take classical AI versus Generative AI and do this quickly so that the food that you're eating doesn't -- Okay. Let's try it this way.

All right. So in more traditional AI, what you're really doing is teaching machines to look for patterns and then you're using that to create, like, a classifier or a predictor. This includes most of the work around neural networks over the last ten years and a lot of classical applications. So you can think of machine vision in this category. So self-driving cars, you need to teach the machines to understand what traffic patterns and obstacles look like, so that they don't come into contact.

Generative AI. Generative AI is really now -- these AI systems are creating content and they're creating data that is similar to the data that's being trained on. An example of this is when you ask -- you can ask a chat application today to write you a poem. It doesn't really understand -- I guess some people would argue this. It doesn't really understand the artistic bits of poetry; but it can create content based on similarity of the words that you've used and words that are used in conjunction with tokens, words that are broken in tokens that would comprise the word "poetry."

So in this way, Generative AI is this kind of an extension and builds on a foundation of neural networks now creating content, not just thinking about predicting the next word but actually creating the next word or creating an image.

**Steve Sparkes, Scotiabank:** I think one of the challenges for all of us is at what point does it begin to hallucinate and how can you control sort of the error rate? We talked a little bit in this morning's keynote about the legal example, where lawyers fell foul of having trusted ChatGPT's output, despite saying, "Are you really telling the truth?" And, of course, it lies again.

So I think one of the things that we're thinking about is: How do business leaders embrace the power and, at the same time, have some control and have the confidence that they're not about to make a misstep?

**Will Grannis, VP and CTO, Google Cloud:** Well, in my opinion and our experience, it's really approaching it from a full stack or a comprehensive approach. And what do I mean by that?

The design principle behind Google Cloud's approach to AI is offering a complete AI stack. What does that mean? It means that the research that goes into these novel techniques, it means that the processors - both ours and others and the semiconductors and the chips, the design and the optimization, we spend a lot of time there.

But then we also spend a lot of time on the tooling. Machine learning ops is a significant burden. Gen AI ops, significant burden for any organization. So we also take it upon ourselves to create a platform that allows people to, without even knowing how to code, start their journey in Generative AI.

But then it also kind of goes up to the top of the stack, like I spoke about before. We want to give you an always-on collaborator. So even if you are operating in SaaS, like, you are in Workspace, and you are like, You know what? You are a small business. Maybe you started a dog-walking business. There's still time for you, if that's your aspiration. The good news is Generative AI is here to help.

And what it can do is you can express a desire like, 'Hey, I really want to track my clients, but I don't know how to get started.' And you can give what we call Duet AI -- that's our always-on AI collaborator -- you can give Duet AI that intent, and it will generate for you a tracking spreadsheet with the typical columns and typical fields, if you are trying to run a dog-walking business, for example, without you having to think through the structural components or get started. And so in many ways, at the top of the stack, it's about getting started with getting value without having to architect.

We also have brought the same Duet AI into our platform. So, for example, let's say you are using BigQuery, which I hope you do -- and if you're not, there's still time. But with BigQuery, what really matters is -- and what matters in all analytics systems is structuring good queries. You want to run over the right amount of data. You don't want to have a query that's not well-structured so it kind of overshoots the data, it's too costly or kind of runs on without -- without bound.

And so Duet AI and GCP, for example, Google Cloud Platform, we have the capability within BigQuery for it to give you a recommended query for a question that you have.

And so throughout the stack -- and for a company looking to get started, one, it's kind of marrying up your strategy and then being able to find your place in that stack. So if you want to build a foundational model, you care a lot about -- like, for example, Google, you care a lot about our ability to tune infrastructure and give you optionality between, like, a TPU v5e, which is what Anthropic just published this week, what they're using. Tens of thousands of chips in these clusters. You care a lot about efficiency when you're going for speed-to-market but, also, you don't want to break the bank.

But you could also be looking at GPUs, which we host from NVIDIA. And we have a range of them. We even published -- we now made available an A3 supercomputer machine shape where you get tens of thousands of NVIDIA H100s available to run the most extensive workloads you can imagine.

And so if you are a model-building company, if you're an organization that really wants to innovate there, you need that infrastructure optionality and you need a partner that really understands deep optimizations in the infrastructure layer.

Maybe if you're a company that's, like, 'You know what? We're going to take foundational models that exist and we're going to take our own data and we're going to combine them to create some competitive advantage', that's where you need that platform tier.

So what we call this is Vertex. And in Google Cloud, Vertex allows you, without even code -- You don't need to know any code at all. You can go into the console, you can find -- you know, you can find the latest model, whether it's one from Google or you can find Llama 2 there. You can find models from Cohere, Anthropic, others in our model garden and you get started combining those foundation models with your own data.

Well, then you're like, 'Well, how do I get my own data going?' Well, we have technology within Vertex that allows you to point an indexer at structured and unstructured data. Again, no code. You're just pointing it to the data sources, and it will index it for you.

Because I imagine in your work, regardless of what your background is in this room, you know that to produce content and produce insights, you also have to gather your knowledge and you have to structure it in some way. And that is actually one of the key barriers to using your own data to refine these foundational models, is structuring it, whether you are going to bring it to, like, an embedding, which is like a spatial relationship of your data, or you're going to use a Knowledge Graph. Getting you to that point quickly is really, really important. So we provide that in Vertex.

And then I mentioned earlier, if you just want to see what this stuff can do and you are not really sure where to kick the tires yet, that's where you can use Duet AI just to get immediate value from Generative AI and just speeding up your workflow, getting to an image for a presentation.

How many people in this room make presentations? And those of you that didn't raise your

hand, you're lying, but that's all right.

When you make a presentation, what is one of the key things in your workflow? Like, let's say you are trying to convey a concept with an illustration or an image. Well, you have to go find the perfect one. And if it doesn't exist, it actually slows you down considerably in the workflow of creating that content.

But also, it's the same concept for an online marketer, right, who wants to go out and put out content but needs an image that is specific to them and that they have the appropriate copyright rights and other legal rights to. So now it's really about speeding up that kind of -- that workflow.

So that's how companies can get started, is, like, picking your strategy and then picking a spot in the stack.

**Steve Sparkes, Scotiabank:** I think you mentioned the access to vast amounts of compute, which obviously is a prerequisite for the model generation. And I think it's one of those weird things about this moment where in the past, you'd see start-ups emerging and challenging the incumbents.

But in this case, the incumbents actually have access to the raw materials, whether it's data or compute. And so it's really -- the model garden route is probably the way that we will start to see start-ups emerging.

But I'm just wondering how you think about making available that capacity to emergent companies, because it feels like there's a very significant barrier to entry for them at the moment.

**Will Grannis, VP and CTO, Google Cloud:** Well, what's really interesting is that we're seeing both traditional enterprises and organizations and start-ups succeed in Generative AI. And I'll give you kind of a view of their path. So somebody like Deutsche Bank.

**Steve Sparkes, Scotiabank:** It's a small European competitor.

**Will Grannis, VP and CTO, Google Cloud:** Yeah. Just wanted to point out that --

**Steve Sparkes, Scotiabank:** For those of you who haven't heard of them.

**Will Grannis, VP and CTO, Google Cloud:** They're able to synthesize internal content -- so you mentioned what do they have. They have documents. They have content. They have analysts. They have this know-how inside their organization.

But they didn't have a really quick way to express a question or an idea or a natural-language interface to this corpus of data that they have in these documents. And to get some reasonable results back quickly enough to spend time refining it, it was just -- in the past, it probably would have been faster just to have the humans go cascade a bunch of searches, like, write all this stuff down and then put it out as content.

But now the pace and speed of a platform like Vertex allows them to index and to create these spatial relationships and this representation of their data fast enough and in a workflow that now their analysts are reviewing the content internally that's being produced, and they're using it to synthesize potentially new analysis for their customers.

Healthcare, another -- kind of a big traditional enterprise. How many of you have dealt with the health care system? Very, very kind of classic enterprise issues around innovation. But a company like HCA Healthcare is using Generative AI. They're doing live transcriptions so that doctors don't have to go back to their desks and type the notes in. They can get that live while they're with a patient, and so they can spend more time on care. So these are traditional organizations that are getting benefit from the speech-to-text and the natural-language capabilities of Generative AI today.

From the start-up perspective -- and my experience both being a start-up founder, CEO, and also kind of a big-company person as well -- start-ups, usually they're seeded by people who have really, really specific knowledge about a very specific problem. And so their advantage is speed.

And so a lot of what we're seeing right now with start-ups and their success -- I mentioned Anthropic earlier, Cohere, AI21 Labs, a lot of these organizations, they can get the scaled AI-optimized compute from us so they can focus on solving a business problem or an idea like someone wants to put out a chatbot that they view as, like, a more responsible chatbot than some of the other alternatives that are out there. They can focus on what are the guardrails they want to put in, what does the output look like that they want to generate. Not where am I going to get basically these AI supercomputers. We take care of that for them.

So in a lot of ways, we're speeding up a start-up's ability to get to their first article, their first output and then they can tune it much, much faster.

**Steve Sparkes, Scotiabank:** And on that point about exposing Generative AI content to the consumers, one of the things that we've been super thoughtful about at Scotia, one of my colleagues, Grace Lee, who leads the Data & Analytics team, has spent a ton of time on a data ethics program. And we're acutely sensitive to the appropriate treatment of all forms of data, but in particular consumer data. So we've got the human in the loop in all of our Gen AI programs and the use cases at the moment. And we've got a very bright line that we have not yet crossed for exposing Generative AI content to our consumer base.

I'm just wondering across the broader client base that you serve, where you're seeing that risk appetite and what checks and balances other clients may have been putting in place to make sure their end users don't befall some of the worst characteristics of Gen AI.

**Will Grannis, VP and CTO, Google Cloud:** Responsible AI is something that, in my opinion, is associated with Google in a way.

In 2017, we put out our first version of AI Principles, because we had been using AI in production at scale for years. And so in a way, we bumped into a lot of these issues very early on and have been developing guidance that we hope -- and is constantly evolving, but we hope



is helpful to organizations of all sizes.

But we're product people. We're technology people. We're engineers. And being a technologist myself, I'm always thinking about how our customers can get leverage from this knowledge and this experience that we have.

So, for example, I mentioned Vertex, this AI platform. What it has -- one of the capabilities that Vertex has is it includes these responsible AI filters that get added to the runs of these generative models inside of the platform. And it will actually give these customers a preview of the potential brand, toxicity, other issues that might come out with this output before they send it into the next stage of their workflow.

That may sound like a little thing to some of you, but to create these filters and the ability to create this kind of sliding scale that an organization can decide -- they can decide for themselves how much they want to push the boundaries of their own brand safety within their industry, their own competitiveness, what type of content they want to allow through and they don't. Just even getting them to a place where they can codify for themselves what the slider bar, where it should be set, is really, really important.

And so we keep embedding responsible AI into the platform so our customers can benefit from it in kind of an automation and a tooling way, because that's what scales, versus having to consistently have humans overlooking every single thing that you're doing. At some point, you need that leverage from the technology.

**Steve Sparkes, Scotiabank:** One of the things we've been finding is that it's a great augmentation for tasks. You talked earlier about the opportunities for transcription and for rapid digestion -- digesting of data. But we're not really seeing it as a displacement of actual roles in their entirety.

I think that -- do you see it as a sidekick, as an aid, as an intelligent partner in people's, in knowledge workers' activities day to day?

**Will Grannis, VP and CTO, Google Cloud:** Absolutely. We're in an era now -- I've been building data products, data center products, AI-centered products -- My undergrad was in linear algebra, which was never cool until now. And so I feel like I'm back. Like, there was a winter there for me, but now it's cool again.

And what's really interesting is the accessibility to advanced capabilities that is being created in all layers of the stack. I mean, just the computation -- one of the reasons why neural networks took off so prominently roughly, eight, nine, ten years ago is because the computational efficiency was at a place where you could reasonably now start to run neural networks at scale. So that created kind of this first wave.

Now we're in the second wave where Generative AI has kind of made AI more accessible. It doesn't matter if you're in IT. If you're in the business, if you're in finance, if you're in any part of a company now, you can get the benefit of AI because it converts your kind of natural language way of interfacing, rationalizing the world into computer language and machine language. This

is pretty profound.

I think about it as Duet AI, all you have to do is tell it what you're trying to accomplish in a surface of spreadsheets or creating a slide presentation or even within the operations of an application on a cloud; and it will help guide you to more efficient queries. Hey, maybe you're using the wrong sized disk. Based on the IO properties of this application, maybe you've overspent on this and maybe you ought to look at a different shape.

Or you are not really accessing this data all that often. Maybe you should look at archival storage versus running it on this kind of hot, high-performance storage all the time. And those little things start to really add up.

I was working with one customer in the gaming industry, and they were really interested in this concept of a digital concierge. And in this meeting with the leadership team, it was really fascinating because now -- it used to be a presentation where like, the CDO and the CTO and the CIO would go up and they would give a presentation about like, 'Here's how you do all the stuff. And everybody would be like, 'Wow, that seems, like, really weird.' There was a lot of math and a lot of computers and a lot of network and a lot of stuff. And that's really not relevant to a lot of the, kind of, day-to-day experience building that a lot of the business was trying to do.

And one of the things that Generative AI has done has kind of broken down these silos between these functions and allowed more people to participate in the creation of a digital concierge or in the creation of experiences for customers.

I guess one of the experiences for me in AI that really was an eye opener was how -- We were working with NASA. And they're like, 'Hey, we have some interns. And we'd like these interns to be more productive during the summer.' We were like, 'Cool, okay. How do we do that?'

And they -- like, literally, like, they sleep on the floor because they run these jobs and they're looking for a planet -- they're looking for exoplanets, this particular group. They are like, They have to sleep on the floor because these machine jobs, they run for a while. They may stall out. They don't work and so they have to be there all the time. It's really sad being an intern there.

And just by providing a hosted notebook with some small GPU attach and orchestrated by Kubernetes in a cloud that could kind of, like, scale up and scale down, they discovered a whole bunch of exoplanets without having to sleep on the floor in the office. Happier interns. Scientific discovery, like, ground-breaking new discoveries in science. And all just by providing kind of this simple platform for AI. I mean, that's the era that we're in right now.

**Steve Sparkes, Scotiabank:** Well, I'm going to bring you back down to something that's of immediate concern to me, which is how can we use AI to look for security holes? And do you see it as an arms race between the threat actors using AI to find those holes before the defenders can plug them?

**Will Grannis, VP and CTO, Google Cloud:** Okay, so let's see. We're in your domain area now, Steve.

**Steve Sparkes, Scotiabank:** Just a little bit.

**Will Grannis, VP and CTO, Google Cloud:** All right, big C. So I'm going to try to do my best for you.

**Steve Sparkes, Scotiabank:** We need help.

**Will Grannis, VP and CTO, Google Cloud:** A truer statement was never made.

I think -- What is it? I think \$10,5 trillion in kind of cybercrime by 2035-ish on an annualized basis. So, yeah, this is a really big deal.

And at least what we're doing to try to be helpful in this moment, and what we've been talking about is, one, making sure that there's comprehensive security for applications and programs that customers trust to Google Cloud.

So, for example, we have a Security Command Center that now integrates a lot of telemetry coming off of it. It used to be -- you kind of have -- you would know this extremely well, Steve. You have kind of applications that deal with security for different parts of the stack -- so like endpoint security, application security, network security.

And one of the things that we've done in Google Cloud is we've deployed this thing called Security Command Center. It's kind of this wrapper that gives you telemetry into every layer of the stack from an application all the way down to the infrastructure. And that's really important because these threat actors are very capable and they are constantly changing tactics. And so the telemetry and the data and the logs are actually super important, because the tools will evolve but having that foundation data is really important. So creating a lot of telemetry there.

But even more importantly, in my opinion, in gaining us some leverage is, we have created domain-specific models like Sec-PaLM 2 where we take the expertise of Google threat groups, where we take telemetry and logs and events from -- we're pretty -- it's no surprise, we're a pretty big target as well for cyber events.

And we take all this know-how, we take this data, and we train these foundational models, in this case PaLM 2, or Pathways Language Model 2, and we kind of bring the security into it and combine this data.

And now Sec-PaLM 2 is a model that's available to customers of Google Cloud, so you can bring -- it's already kind of tuned for the domain. And then you can use your own observations, telemetry and data to further refine it and create even new capabilities within your own organization, which may even be differentiated at the firm level, which is an entirely new form of leverage. So definitely focused on cybersecurity.

I mentioned that's one of the foundational layers of the stack in Google Cloud. Mandiant and the addition of really that threat action gave us insight into -- anybody deal with -- You are not allowed to raise your hand, Steve -- but anybody deal with malware detection threats in your job? So, yeah, so you may know that there's this YARA kind of framework language that is all

around malware identification and threat detection. It is like a new programming language. It's like a different language completely.

One of the things we did inside of our security console is we've used Sec-PaLM 2 to create auto-generated summaries of malware events and threat mitigation and remediation in natural language, so you don't have to be a cybersecurity professional and understand the semantics and all the nuances of YARA to actually get started on threat understanding and threat remediation.

So what does this mean? It means we're leveling up the security capabilities of an organization, both in the tooling and the people.

And in my opinion, the biggest risk is the lack of cybersecurity skills that are available -- advanced cybersecurity skills that are available. So you're going to need that leverage from your tools and your platform. And you're going to need your entry-level folks to have more capabilities without the 10 or 15 years of threat-hunting experience that you would typically need to operate at that level.

So even just these auto-generated summaries can quickly orient people on remediation. And that could save -- in real workflow terms, I mean -- this could save hours, days, weeks of threat curation and what to do next.

**Steve Sparkes, Scotiabank:** Right. And those hours, days and weeks really count, because the rate of exploits is accelerating constantly. So any time you can shave off the detection and remediation cycle is incredibly valuable. In the past, we might try and use simulations and follow some of the other more conventional learning methodologies. But being able to address the lower level of the stack is absolutely a value-add for sure.

One of the threats that we're -- one of the opportunities and the threats that we're interested in -- and I would welcome your opinion on -- is the use of Gen AI as a co-development platform, because we've seen -- I was talking to Mike earlier.

Their acceleration of commit time was impressive, and I don't want to quote the number in case it's too sensitive. But it's certainly in the double-digits percentage acceleration of the time-to-release code. And conversely, the risk of inappropriate code being injected through that process is something else that we're concerned about.

So I wonder, firstly, to what extent, within your own teams, are you using augmented development techniques? And then, secondly, how are you protecting against that risk?

**Will Grannis, VP and CTO, Google Cloud:** So code completion, code generation, understanding the provenance of certain code bits, this is all foundational to what we've been working on for years at Google. And it shows up in a couple of different ways.

One is through a partner like Replit. One of the most successful IDEs, integrated development environments for developers -- tens of millions of developers on it right now -- they're using our models underneath the hood to help their developers with code generation, code completion,

and really, really speeding things up.

But even within -- let's say you are in the GCP environment and using development environment there, this is where Duet comes back into play.

And I mentioned some earlier examples around Workspace like Slides and Sheets, but we also have -- Duet can help in code understanding, in addition to completion and generation, which if you've ever been a developer and showed up to an organization and been told, 'Go check out this codebase and tell me what this means', you have felt a significant amount of panic and pain immediately. And then trying to find out what the intent or the design principles or what this code was intended to do can actually take a really long time.

And so one of the capabilities that we're really excited about in Duet and kind of underpins this is just understanding code and summarizing what this code is trying to do.

And at Next, our annual conference this year, we showed examples of Duet being able to create summaries of this is what this code is intended to do and also be able to look for cases of risky dependencies, outdated frameworks, other things that can also kind of help you in refactoring.

**Steve Sparkes, Scotiabank:** And as a former programmer, now recovered, I can say that I amplify Will's thoughts. And I pity the poor people that had to read my code to try to figure out what it was doing. That was a long time ago.

Let's go back. A little bit earlier you talked about the capabilities of Omni -- BigQuery Omni across multicloud. I think multicloud is obviously -- it's a thing that we are all evaluating and embracing.

So in common with a lot of other organizations, we've got a hybrid environment where we've got a bunch of compute still -- and data still on-prem. We've got footprints, both data and compute, in multiple other clouds.

So maybe just spend a moment thinking about: Is that BigQuery Omni an example of how there's some -- I wouldn't say collaboration, cooperation where there's an opportunity to have some horizontal capabilities extend across the multiple clouds. And how should we be thinking about taking best advantage beyond Omni BigQuery?

**Will Grannis, VP and CTO, Google Cloud:** So I think about this, because I spend a lot of time with customers in my job -- My roles before coming to Google were all in the enterprise, large organization. Like, start-ups, I would find a problem, I would go and try to solve it, and then I would realize for scale, I would have to do it from inside a large organization.

I always start with: What's the business problem you are trying to solve? Because that really dictates how you view multicloud as an organization.

So, for example, disaster recovery, continuity of operations, that's one form of potential multicloud. You may decide that it's in your best interest that you want to run -- you want to

have data stores across multiple clouds to meet regulatory requirements for COOP [Continuity of Operations] and DR purposes.

You may also decide that you have a bunch of data in different clouds or even on your laptop or in your own data centers and you may decide that you want to -- you want to have optionality to not have to maybe converge all of them into one big data lake or data ocean or data universe. I don't know what's next.

Something like AlloyDB Omni, which we just released earlier this year, it's a PostgreSQL database and it can run on your laptop. It can run in your own data centers. It can run in our cloud. It can run across other people's clouds.

And it's the continuation of this thread of, if you want to have consistency in implementation architecture but you don't want to -- now's not the right time to make a choice to constrain the number of clouds or hybrid or even your own laptop, there are tools that we're making available and databases and analytical tools to run across all of those different surfaces without you having to commit to a convergence event.

And so the way I think about it is if you want to have -- the business problem you are trying to solve should drive how you think about multicloud and then seeking out a solution that matches that business imperative, rather than saying multicloud is attendant unto itself and then try to project it into a solution.

**Steve Sparkes, Scotiabank:** And I think it's really interesting, when you look at the -- when you shoot for the lowest common denominator, do you find -- there's some really terrific companies that are multi -- that support, in particular, the security operation.

I think of Concourse Labs, who operate across multiple clouds.

And you can see there's a clear distinction between those companies that have deliberately aimed to be multicloud from the outset, to support that capability, versus those that were sort of incubated on one and then grudgingly accepted that people were going to have multiple clouds and they were obliged to support multiple different platforms.

**Will Grannis, VP and CTO, Google Cloud:** Well, choice, again is -- as a technologist, I want to solve my problems as quickly as possible with the least friction.

This is a thread that actually runs through. I mean, if you rewind the clock, you know, eight or nine years through Google Cloud's history, you can see the earliest beginnings of this in the surfacing of TensorFlow and open-source framework for machine learning. You can see this in Kubernetes for container orchestration. You can see it through these omni versions of data analytics and data stores. You can see it across Apigee, which is a multicloud API management capability.

I mean, this is a consistent drumbeat and demonstration of a principle of being open and giving choice that now even has now extended into the Generative AI era where you pick your model, you pick the workflow, you pick what data stores you want to bring to these models, and do so

safely in your own environment without having to expose them. And so this continuous drumbeat of choice, it's core DNA for Google and Google Cloud.

**Steve Sparkes, Scotiabank:** Yeah. And that partnership engagement model is one that I think -- keen for you to double-click even slightly more on that. Do you have a group that's dedicated to figuring out how to drive those partnerships and who to bring in and what other selection criteria you use? Just a little bit more on this.

**Will Grannis, VP and CTO, Google Cloud:** This is probably the biggest -- So I mentioned there's kind of been three waves of this business, of kind of making sure the engineering primitives were there, making sure that we could serve every geo and every customer at the breadth of the platform. And now we're in the scale phase of making sure that we can scale up and our customers can scale up without any -- without any roadblocks.

And our ecosystem has grown considerably. I mean, we are at now over 100 -- and I said this earlier, I got to say it again because it actually -- it's a pretty big change. Over 100,000 partners are in our ecosystem for Google Cloud.

And this can range from the independent software vendors who use our capabilities and they kind of -- we ride the channel with them through their software deployments, like a Workday or SAP or DocuSign.

But it's also in consulting companies and integrators who help solve problems. Most companies already have relationships with other organizations. They have this trust built up over the years, and they want to continue to leverage their preferred partners.

And so we've now made it available, so with Deloitte and Accenture, Wipro, they all have thousands, if not tens of thousands, of trained individuals on Google Cloud now ready to help solve problems that are deep implementation problems.

Maybe you want -- you're looking at maybe mainframe modernization or you're looking at these problems that typically are very complex. They are multiyear. They involve integrators. They involve potentially cloud providers. They involve on-prem. Those are the types of situations that we're now capable of moving very quickly with you.

And kind of -- and to continue kind of upleveling the entire industry, we've also spent an enormous amount of time in training. Over 100,000 machine-learning courses are now in the world -- sponsored or brought to the world by Google and Google Cloud. And hundreds of thousands of practitioners are now in the world that understand both AI, machine learning, and Google Cloud. And we think that's a real business benefit to organizations at scale, right? Because these implementations are complex, and they need to happen over a period -- over a longer period of time. So that really deep understanding of your business, your industry, your geography, nuances in regulatory concerns and the rest.

We launched -- like, our sovereign clouds in Europe, we launched in partnerships with two large partners over there who help us kind of manage on behalf of the countries in which they operate.

**Steve Sparkes, Scotiabank:** One thing I've been fascinated by is to see how Phil Venables [VP, TI Security & Chief Information Security Officer at Google Cloud] has built out the CISO practice within Google Cloud. And I've known Phil for a long time, back in the Goldman days. When I was at Morgan Stanley, he was at Goldman. And we were always the healthy competitive tension. But in the cyber arena, there was always cooperation.

And as he built out the risk practice and then ultimately became the CISO for Google Cloud, I'm curious how much time do you spend hearing from him about threats that he would like you to build some capability to defend against, and vice versa? When do you go knock on his door and say, 'Here's something you ought to be taking advantage of?'

**Will Grannis, VP and CTO, Google Cloud:** Well, like, I imagine all of your organizations, the relationship between the CTO and the CISO has grown very close in Google and Google Cloud over the years.

And I go back to something like a Sec-PaLM 2. Creating a domain-specific model is the union of domain expertise and knowledge and telemetry that someone like Phil would be dealing with on a day-to-day basis and these complex policies that have to get embedded into the response and the guidance that these models might give or the content they might generate.

But then, also, the technical underpinnings of the platform and the infrastructure and the computation that makes those feasible to operate. Creating a Duet AI capability inside of a security console that can summarize threats as they come in, in near real time, is actually a very computationally complex and intensive process.

And so Phil doesn't want to sweat the technical infrastructure. And I would do a very poor job of describing in detail the nuances of the security concerns and roles.

**Steve Sparkes, Scotiabank:** Awesome.

So I'm curious, based on everything that you've seen and everything you know about what's going on, what are the things that you're most looking forward to in the next sort of -- bearing in mind the Safe Harbor statement, not necessarily Google products but the -- over the next six months to a year, what's getting you most excited?

**Will Grannis, VP and CTO, Google Cloud:** There are three things that are happening right now that I am extremely interested in and excited about and in many ways, I think, reflects how the cloud -- the cloud market itself has reset itself in the AI era.

And that is, number one, the list of companies today that we just briefly talked about, they're in almost every industry, every size, almost every use case you can imagine. And so the access that's being created to advance computation is really, really staggering.

I have two daughters, 21 and 17, and they're in school -- various schools, in college and high school right now. And they're able to utilize Generative AI sometimes with authorization, sometimes without authorization, because the next generation does find ways around.



But they're leveraging these tools and asking questions and getting answers, and they have access to, like, the world's knowledge in a way that's been distilled and synthesized so that they don't have to try to ask 5,000 questions. Because if any of you are raising kids, you know that they give up after not getting the answer to the first or the second question.

And so now there's this democratization and access to the most advanced computation that I've seen in my career, and it's just happening in a way that feels almost invisible and really easy to use. And that's happening at kind of the individual level.

Then there is this massive wave of discovery happening. For those of you that are following AlphaFold and the second generation of AlphaFold, that has already had a dramatic impact on how we think about the building blocks of life and how treatments could be envisioned and might be envisioned in the future in precision medicine, for example, which is something I'm very excited about.

I mentioned scientific discovery outside of our planet and the thing with exoplanets. Well, if any of you have heard of this organization called the B612 Foundation, they went and took these computational tools and AI, and they went and found asteroids just from old data. Not anything new, not new telemetry. But they found asteroids using some -- the first versions, they had to, like, literally look at images and use experts to try to figure out whether that constitutes an asteroid. And now they've found hundreds in a very quick period of time.

And you can think about the ramifications of being able to characterize potential existential threats in that way in the future is just so profound.

And then also kind of the third one that I think a lot about is you have medicine and you have space. But then we also have just in the last week, DeepMind at Google has released a way of characterizing crystal structures, millions of new simulated crystalline structures which might someday power the next wave of development of artificial crystals for semiconductors, solar power and the rest.

And so the pace and the cadence of these discoveries is so quick, is so rapid. And I met -- I am in such a wonderful place to be that I get to see the convergence of the research, the computational realism. And these organizations who might only be dozens of people, that now they have the output of an organization 10 to 20 times their size. It's really just an incredible time to be a technologist.

**Steve Sparkes, Scotiabank:** It really is. And I think it's also a time for us to consider what are the unique characteristics of any company, what is it that they sell, what is it that they produce, how do they do that.

And I think we're going to be facing the question of if Gen AI can deal with a lot of the basic toil in knowledge worker tasks, is it going to be like a calculator that lets people focus on value-add, or does it just make us stupid over time and do we lose the ability to innovate? And is it the friction and the toil that actually creates that struggle? Does it create new ideas and innovation? And I don't think we're going to know the answer to that one for a little while, but it's going to be

an amazing journey.

And, Will, I just want to thank you so much for the time today.

**Will Grannis, VP and CTO, Google Cloud:** Real pleasure, Steve. Thanks to all of you for being here as well.

**Steve Sparkes, Scotiabank:** Thanks, everybody.