

This preliminary transcript is provided for the convenience of investors only, for a full recording please see the [Goldman Sachs Communacopia + Technology conference webcast](#).

Thomas Kurian, CEO of Google Cloud at the Goldman Sachs Communacopia + Technology Conference on September 10, 2024

Eric Sheridan (Goldman Sachs): All right. I know people are still finding their seats, but in the interest of time, let's -- we'll get started on our next session. I'm going to start with the safe harbor. I'm going to introduce Thomas Kurian. He's going to join me up on stage, walk through a slide presentation, and then we'll do some Q&A.

Some of the statements that Mr. Kurian may make today could be considered forward looking. These statements involve a number of risks and uncertainties that could cause actual results to differ materially. Please refer to Alphabet's forms 10-K and 10-Q, including the risk factors discussed therein. Any forward-looking statements that Mr. Kurian makes are based on assumptions as of today, and Alphabet undertakes no obligation to update them.

Thomas Kurian joined Google Cloud as CEO in November of 2018. He has deep experience in enterprise software. Prior to Google Cloud, he spent 22 years at Oracle where most recently he was President of Product Development. Thomas, welcome to the stage.

Thomas Kurian, CEO, Google Cloud: Thank you, Eric, for that warm introduction. We at Google Cloud are the organization that takes the innovation that Google is making in infrastructure, our data and digital platforms, cybersecurity, and AI and brings it to enterprises, small companies, start-ups, and governments.

We do that through a global infrastructure - 40 regions, 121 zones connected around the world. On that infrastructure, we build world-leading AI training and inferencing systems, a collection of leading frontier AI models, not just from Google but also partners like Anthropic and many others, and a suite of tools. We make that available to customers in five important product offerings:

Infrastructure for them to build their own foundational models or to tune and adapt existing foundational models and modernize their IT systems; development tools to connect AI models into enterprise systems so that they can build agents and applications; data analysis to connect any kind of data to any kind of model; as well as to use AI to speed up and make it easier for people to do analysis.

Cybersecurity, integrating AI into a security operations workbench to help people prevent, defend, respond, resolve cyber threats materially faster. And then obviously we've integrated AI deeply into our applications portfolio, Google Workspace, but we're also introducing new applications powered by AI.

Starting with AI infrastructure. You know for 11 years now, Google has built world-leading AI systems. We are in our sixth generation of accelerated technology, and we offer customers a choice of our accelerators as well as NVIDIA's. They're assembled into super high-scaled systems that offer extremely good performance. 3 times competitors for training, 2.5 times competitors on a cost performance basis for inference. And as you get larger and larger clusters, reliability becomes a bigger issue, because the larger the number of machines, you end up with failures and needing to restart your model training. There are many advances we've made there.

As an example, in addition to the software on top of these systems, we've introduced water cooling. We now offer close to a gigawatt of water cooling in production. That's 70 times the number two. We've seen huge growth as a result in our AI infrastructure. 150 times more compute instances can be connected to a single storage volume, which means you get super dense training clusters. We've seen 10 times growth year over year in people using our systems for training.

Here are examples of customers. 90% of AI unicorns run on our Cloud for training and inference. 60% of all AI-funded start-ups ever use our infrastructure for training or inference. We're also seeing traditional companies now building both high-performance and generative AI models on our Cloud. An example is Ford Motor Company, they're using our deep learning technology to build simulations for wind tunnel and for virtual wind tunnel simulation, replacing a traditional approach called computational fluid dynamics.

Midjourney, a leader in AI foundation models trains on our TPU, serves on NVIDIA GPUs both in our Cloud. It's an example of having that diverse portfolio allows them to choose the best combination across the platform.

In addition to having a foundational model, you need to connect it to your enterprise systems. We offer a foundational platform for people to do that. There are a number of differentiators. I'll touch on three important ones. First, we offer an open platform. In addition to Google's frontier models, we offer leading open source as well as third-party models Anthropic, Cohere, AI21 Labs, Runway, and many others. This allows enterprises to choose a standard platform, because they get the choice of models.

We also have built AI for many years into our products, and we offer advanced capability as a result in this platform. Additional services like grounding to improve the accuracy of the answers. We've introduced something called high fidelity, adaptation, distillation for example can take a large model and shrink it down. All these additional services we offer to customers through an end-to-end platform.

You can use that to connect to any Cloud, any SaaS application, or any enterprise system. We monetize compute instances on a consumption basis. We monetize our AI developer platform by pricing on a token basis. But people also pay for the additional services: grounding, distillation, fine tuning, adaptation, and so on.

There are many developers using our platform, over 2 million of them. As people migrate from proof of concept to production, we see a ramp in usage. An example last week, we spoke with Accenture. We're working with them in many of the Fortune 500 companies. 45% of the projects have gone live and as a result we see a ramp in API requests.

Here are examples of people using our developer platform. Just a couple of examples. Bayer in health care is using our generative AI tools to power search, document creation, image analysis to help radiographers, for example, with tumor detection and comparison. Similarly, Samsung used our distillation tools to build a specific version of Gemini and Imagen to power the Galaxy S24 smartphones which are now in hundreds of millions of hands.

Models need to be connected to data. We offer two important things with our data platform. First, the ability to connect any data on any cloud to a model with super low latency. So structured data, unstructured data, semistructured data can be connected to any model with super low latency. Secondly for those people who want to do analysis, we've introduced a capability we call a data agent. It helps you do all the tasks that you need for analysis but using a conversational interface. It helps you migrate data, stage it, aggregate it, visualize it. It even builds you charts and spreadsheets.

We monetize this in two ways. First, because we've made it easier for people, it drives a lot more query volume on our analytical platform BigQuery. Secondly, because we've opened up analysis from being the domain of people who know SQL, Python, et cetera., it also drives a lot more end user subscription growth. Because we can sell more seats in an organization.

We see growth in our analytical platform. BigQuery now runs 80% more ML operations in just the last six months. It's also being used not just to process structured data, there's been a 13 times growth in multimodal data because of the capabilities of the Gemini model to process multimodal data. Many customers are using it. Two examples: UPS Capital. They run millions of calculations in real time on streaming data to analyze package movements, locations, to detect if you're going to deliver a package in an area of risk. They also adjust and drive, for example, calculations to detect if someone is doing fraudulent things like stealing packages.

Second, Hiscox, one of the largest syndicates in Lloyd's of London, introduced the first AI-enhanced lead underwriting model. So when they ensure property risk, it used to take them months to calculate the entire portfolio. A single property took three days. It takes a few seconds now. What used to take months now takes a matter of days.

What are we doing in cybersecurity? We started with a foundation that's super secure. And if you look at our reliability, we have about a quarter of the downtime of the other players in cyber as measured by CISA and others, we have a really secure Cloud. Half the number of vulnerabilities as other Clouds.

So we started with a very strong foundation. We then applied tools that we built to help organizations prevent, detect, and respond to cybersecurity threats.

How do we do that? From our Mandiant acquisition, from the millions of endpoints that run Chrome, and from our broad network, we collect, summarize threat intelligence from around the world. That is fed into an AI model to help you prioritize the threats you're likely to face. It then compares it with a configuration of your existing systems to see where are you likely to be attacked from. It then helps you generate the remediation. We call that the runbook to remediate it. It writes the audit file for your audit submission. It helps you then validate that you fixed the issue. That cycle speeds up the ability for people to identify, remediate, and respond to threats.

We're seeing growth because we have helped people speed up how quickly they can use our tools to detect and respond to threats. We've seen a 4X increase in customer adoption, three times the volume of data ingested, and eight times increase in threat hunting. We monetize this based on the volume of data we're processing and the number of threat hunts or queries that are happening on the system.

Many customers use this platform. Two examples: Fiserv is running this powered by our AI capabilities to speed up how they summarize threats, find answers, detect, validate, and respond. If you look at Apex Fintech, it's a fintech company. They wanted to speed up how fast they can run threat hunting rules. So they are using our AI system to write extremely complex threat detection processes. It took them many hours. Now it's taking a few seconds.

Finally, we have a broad applications portfolio. We've integrated AI into Google Workspace to improve productivity. I'll talk about that in the question-and-answer session. But we're also introducing new applications.

One example is applications we're introducing to help people in customer experience and customer service. Think of it as you can go on the web, on a mobile app, you can call a call center, or be at a retail point of sale and you can have a digital agent help you, assist you in searching for information, finding answers to questions using either chat or voice calls.

Our differentiation is in three important capabilities. Number one, we can handle multichannel, all the four channels I talked about -- web, mobile, point of sale, call

center -- in a single system. Secondly, multimodal. For instance, you call the telephone company to say I'd like to trade in my phone. It can send you a chat, "Please upload an image of your phone. Your screen is cracked. Here's how much I can reimburse you for." That's a multimodal conversation. Voice, text, and image. We have that capability because of the capabilities of our foundational model Gemini to process multimodal information.

Second, if you work in call centers, there are times where you need to follow a strict order of control of questions. For example, if you're in a bank, you have to verify identity, you need to be able to guarantee you're asked a set of questions to verify identity for KYC. At the same time, you may ask the bank: "Tell me what's the best mortgage offering I have? Can you compare it across these different products you offer?" The first requires deterministic control. The second requires something called generative control. We're the only ones that allow you to do that.

Lastly, imagine you call the bank and you ask a question about your balance, bank balance. You don't need it to be right 90% of the time. You need that answer to be right 100% of the time. We have a technique to answer with 100% accuracy.

All of these is driving growth in our customer experience platform. Here we monetize based on the value we save for users. Either the costs we're displacing or the reach expansion we're giving their agents. We've seen growth across all the dimensions: the adoption of agents, digital agents, the volume of traffic going to these agents, et cetera.

Examples of customers using our customer experience platform: If you call Verizon, you're talking to our chat system and our call system. 60% containment rate, high rate of call deflection. If you're driving a General Motors vehicle and you hit OnStar, you're talking to our conversational AI system.

So across this portfolio we've integrated AI, we've monetized it in different ways and we're focused on three important things: Winning new customers, winning more projects within the customer, upselling new products.

Winning more customers. For example, Radisson. We help them use our generative AI technology to speed up marketing content creation by 2X. They then bring that online.

And most importantly they saw 20% lift, because the ads were really tailored. That's step one. We then often go to other parts of the portfolio. For example, if you're working with a retailer, we can help them now that they've automated the content creation process. We can help them with retail commerce and conversational shopping. A second project. And then to tailor the ads to different customer segments, we'll sell them our analytical platform BigQuery for segmentation and personalization.

So that's the process we go through to win customers. We've trained our go-to-market organization. We continue to focus on thoughtful, geographical expansion of our field. We're teaching them to be specialists. Many of these solutions are not bought in the IT organization. They're bought by the head of customer service, the head of commerce, the head of finance and so you have to learn to sell outside of the IT organization. So we've taught them how to specialize by industry as well as specialize from a solution selling point of view. We've taught them how to do a business value methodology so they can easily measure what productivity benefits we're going to give, how much reach they're going to get, what cost savings they can generate.

And finally, we're not a big consulting shop, we're not a services organization, so we work with a broad partner ecosystem. Because we don't conflict with a partner ecosystem, we've invested in them. We've invested in technology, commercial incentives, training and certifications, as well as go-to-market incentives.

Just to give you an example, if you look at the leading ISP and SaaS vendors, 75% are building with us using our AI tools. Leading system integrators, Accenture doubled the number of people certified on our AI systems in just the last year.

All of this: Winning new customers, driving new projects, expanding and upselling product has driven growth for us. 5 times revenue growth in five years. We are now the fourth largest enterprise software company in the world on a standalone basis. While we grew topline, we also had great discipline in managing our costs. We've invested thoughtfully in engineering, we've invested thoughtfully in the go-to-market organization, and we've delivered operating income growth at the same time.

We continue to see strong business momentum. We're capturing customers faster. We're doing larger deals. Customers have increased their adoption of our products by

30% in just the last year. We have very strong support from the ecosystem. And we have been patient with AI and we think it'll monetize because of the many different parts of the portfolio in which we've integrated AI and the many different ways in which we're monetizing AI, we think it'll help us continue to accelerate our presence in the market.

So with that, Eric, happy to take questions.

Eric Sheridan (Goldman Sachs): Okay.

Thank you, Thomas. Okay. Why don't we jump right into it? Maybe start with the big-picture question. A lot has evolved over the last couple years in terms of both the industry, the public cloud space, and even within Google Cloud. To level set, can you give us your world view of where we sit in cloud adoption and usage and how you see Google Cloud's position evolve competitively?

Thomas Kurian, CEO, Google Cloud: Geographically we're seeing growth in many, many markets. For example, in Asia and Latin America many organizations are going straight to the Cloud rather than starting on premise and then lifting to the Cloud. Industry-wise, there were early movers. For example, retailers, communication service providers. Now we're seeing many other industries: utilities, mining, natural resources, a number of them are moving. So we see that happening.

We also see that historically all cloud-computing projects were controlled in the IT department, increasingly the adoption is being driven by business initiatives. For example, the head of private wealth management will say, "I want to use data and AI to streamline how my organization does research."

And so those projects increasingly are being driven not just in IT but by business buyers. We are still very early if you count all the machines and data centers today versus how many are being consumed in the Cloud. We're still early, and so that gives us -- we have a strong presence. We've obviously grown a lot. When I started, most people told me we didn't have a chance. We're now the fourth largest enterprise software company.

Eric Sheridan (Goldman Sachs): Understood. How do you frame Google Cloud's

differentiated offering to win new business and grow wallet share? Maybe you can talk a little bit about some of the products and the services that are aimed at tackling some of the themes you highlighted in your presentation.

Thomas Kurian, CEO, Google Cloud: We've always focused our product strategy customer-in. So when we started, we introduced the concept in 2019 called multi-cloud. What was that meant to be? It basically said you should be able to use our technology in concert with other cloud providers, and you should be able to standardize on our solution.

So for instance, we introduced a database that runs across on-premise environments, all of the major clouds. That's driven companies like Workday to standardize on it, because they can then use it no matter where their application is run. So that's one example. It broadens our total addressable market, because it allows us to play in all clients. So that's number one.

Second, we also said we need to offer more than just compute and infrastructure. We need to go up the stack to offer solutions and databases, analytics, cyber, et cetera. So if you look at those areas, we've also taken a different approach. If you look at our analytical system, it allows people to analyze data across multiple clouds without moving the data and copying it to Google. So you can get a single place to run your analysis no matter where your data sits.

Albertsons is an example. They've migrated to our analytical system. Many of their applications are on other Clouds. Again, it allows us to win more customers, win more projects.

Third, we wanted to upsell products and services. So when you look at our strength in data and analytical processing, now we bring generative AI along with it. It allows people to do analysis using generative AI much more quickly and efficiently. One of the largest telecommunications companies in the United States is using our analytical system along with our Vertex AI platform and Gemini to run very important but simple calculations.

Take as an example: Take all the calls coming into my call center, which are recorded

for quality purposes, summarize the calls, and compare them with the data from my billing system to tell me if the calls coming in are because of complaints about bills. So that first part, the process using Gemini, the second part is analyzed in BigQuery. And because we offer that combination, it's another one of these strategic advantages that we offer.

Eric Sheridan (Goldman Sachs): Okay. I want to turn next to go-to-market. Are there differences competitively in how you're taking your products to market and how you work with channel partners today? And how has that evolved over the last 12 to 18 months?

Thomas Kurian, CEO, Google Cloud: Great question. You know, our go-to-market, fundamentally we've been very disciplined on how we're taking to market. There are two important things. First of all, the way you sell an AI solution is not the same way as selling a compute server of feeds and speeds. So we've done three important things.

First, we've identified our go-to-market around key buyers. So for example, when we say we're selling cybersecurity, we're selling it to the CISO of a company. When we sell customer service, we're typically selling to the head of customer service or the head of e-commerce. So we specialize around buyers. We've taught our teams to sell as a solution. A solution is: What's the value proposition? How much cost and productivity advantages can it give me? And that's a different sales methodology than selling infrastructure.

In order to do that effectively, though, you can't specialize around everything. You have to be disciplined in how you specialize for specific buyers and product offerings while getting global scale through our frontlines. So we do that very well. And lastly, early on we made a decision we're not going to offer a huge number of services from our professional services. We have a very focused team. They work on our largest accounts. But it allows us to partner very well with independent software vendors, because we're not in the competing with them in core apps. It also helps us work very well with system integrators.

Eric Sheridan (Goldman Sachs): Okay. You talked a lot about AI in your presentation. Can you lay out your vision of how generative AI capabilities will be adopted and utilized

by customers across the infrastructure, model, and application layers and how you view the relative sizing of those opportunities?

Thomas Kurian, CEO, Google Cloud: So we offer, as I said, five different capabilities. Broad brush, you can think about we offer infrastructure for people to use to build their own models or to tune existing models. Second, we help people with our AI developer platform to build their own agents. We have people building insurance agents, research analysis agents, customer service agents of their own.

We've also provided packaged agents: A data agent, a cybersecurity agent, a collaboration agent for helping you write things, and increasingly we are specializing them by industry. So for example, an insurance agent is different than a nursing agent. All three monetized in different ways. Early on just like any other, if you look at the internet, first people laid the infrastructure, over time monetization moves up. And so that diversification allows us to capture value in many ways and over time I think you'll see similar things.

Eric Sheridan (Goldman Sachs): Okay. Maybe I'll turn to Vertex AI next. How are you seeing customers using Google Cloud to build scaled applications in a generative AI world? And what types of models are customers gravitating to across various use cases?

Thomas Kurian, CEO, Google Cloud: You know, we see a huge range of models. Some people are using large models because they want very complex reasoning. Others like Samsung, for example when I mentioned, they're focused on super task-specific models. We allow people to get that range of capability in the platform. And we give them tools to then tailor the model for their specific need.

In addition to that, though, you need other services to be able to use a model. And so what Vertex gives you is the breadth of services. For example, if I'm answering a question, can you validate the answer either against the internet or against my own dataset? Or for example, in financial services against Moody's, Bloomberg, Thomson Reuters, and other well, high-quality data sources. That's grounding. We make that available as a service so you don't have to do it yourself.

We offer technologies, for example, to shrink the model size so that you can say I took a general model and tuned it just for my needs. Example of a customer doing that is Samsung. And that reduces cost. It also improves latency and it shrinks down the response time for these applications.

Most importantly, we've connected all these services together. So for example, if you don't know how to ground, you can just delegate to a service we call Adaptive Grounding. And our systems will say we've got a response: This one I don't think needs to be grounded, However, the other answer needs to be grounded. And that saves people a lot of cost, because they're always worried about, "Will I have to pay each time I send a request? Will I also have to ground the answer?" So all of these are built into Vertex. As a result, we've seen huge growth with it as I pointed out earlier.

Eric Sheridan (Goldman Sachs): Okay. Maybe turning to custom silicon. Can you discuss Alphabet's strategy around silicon partnerships versus building your own custom chips for AI workloads?

Thomas Kurian, CEO, Google Cloud: We've always worked on high-scale compute systems for AI. We partnered closely with NVIDIA. We're also in discussions with many others. But just as an example, you know, we offer an integrated, high-scale system. There are a number of pieces to that system. People often get focused on the chips. It's not just the chips.

It's, for example, how much memory, high-bandwidth memory do you have? If you're doing training for a dense model, you need more. If you're training a sparse model, you don't need as much memory. What's the optical switching you have to interconnect a very large cluster with low latency? What's the cooling technology?

And then as you move up the stack, can I program in a high-level framework like PyTorch or something like that? Can I use a framework like JAX which is a compiler framework that will compile that down if you're going to a GPU to CUDA, if you're going to a TPU differently. And so that entire stack is what we offer customers. That's why many leading companies use us for high-skilled training, but also inference.

Eric Sheridan (Goldman Sachs): Okay. Can you give us an update on Workspace and

other applications in your portfolio and how you see them in the broader generative AI offerings for clients?

Thomas Kurian, CEO, Google Cloud: From a productivity point of view, we see three types of adoption patterns. With Workspace which is our communication collaboration portfolio, we do have many organizations adopting it for their professional workforce to help them write, to take notes in meetings, to create content. For example, we have many of them using it to create marketing content, to translate marketing content in many languages. We've introduced a new product called Google Vids which allows you to create short videos for training people and, you know, to do, for example, All Hands meetings. So there's a lot of that for the professional workforce. And we see people like Randstad, Woolworths, et cetera, adopting it.

Second, we have specific parts of organizations that are super high value for an organization. For example, if you're in a hospital, as a hospital company, nurses are the critical path because nurses determine how many hospital beds you can have. They control the revenue of the organization. So we work with nursing staff, for example, to do live handoffs of patients. It saves about 6 hours in a 24-hour day. And one of the leading hospitals is talking at a conference today that they estimate when rolled out it'll save them \$250 million.

Insurance. You know, handling claims. We're working with the largest health insurance company in Germany. They have a huge amount of claims coming in. On average a claim, they need to read 800 policy documents to determine if a claim is not [*sic*] is valid or not. They use our technology. It helps take 23 to 30 minutes down to 3 seconds.

So productivity in the -- in these specific places are extremely high value. And then we also see productivity in certain roles that are scarce. When I say it's scarce, for example, cyber is one example. There's not enough cyber analysts to go around. Introducing AI capability there scales productivity in a very different way.

And so we're working on all three dimensions. The professional workforce, frontline or first-line workers, as well as scarce teams. And there's a lot of thought we put into find what are the roles that are scarce, because they monetize faster.

Eric Sheridan (Goldman Sachs): Okay. Maybe I'll end on one big-picture one. When you think about the opportunities you laid out today in your presentation and in our conversation, what are the primary areas of investment that you're calling out when you're thinking about aligning investments against the goals and milestones that you and the team want to accomplish with Google Cloud?

Thomas Kurian, CEO, Google Cloud: I mean, we've shown a track record for many years now of being very thoughtful in how we're making investments. They broadly span what do we need to do in engineering to broaden our product portfolio and deepen it? What do we do with our go-to-market organization to expand it and build specialization? What do we need to do in our data centers to expand the geographical footprint as well as grow our infrastructure?

We've had very close relationships also with partners. And we incent partners and we are making investments in the channel. All of that is balanced. We do a lot of systematic planning. We look over multiple years and where technology is going. And we make thoughtful decisions based on all of that.

Eric Sheridan (Goldman Sachs): Okay. Thomas, I always appreciate the opportunity to chat. Please join me in thanking Thomas Kurian for being part of the conference this year.