

# Credit Suisse 26th Annual Technology Conference

## Company Participants

- Colette Kress, Executive Vice President and Chief Financial Officer

## Other Participants

- Chris Caso, Credit Suisse

## Presentation

### Colette Kress {BIO 18297352 <GO>}

I think it's pretty good.

### Chris Caso {BIO 4815032 <GO>}

We had some pretty good walk-in music, welcome music. So, thanks, everyone. So, I'm Chris Caso, Credit Suisse Semiconductor Analyst. So, welcome back to the conference. Our next speaker will be NVIDIA, which you all know. Here from NVIDIA is Colette Kress, Arizona-native CFO coming back home to Arizona. So, Colette, welcome.

### Colette Kress {BIO 18297352 <GO>}

Thank you. Thank you, Chris.

I just want to make an opening statement first, as a reminder that this presentation contains forward-looking statements, and investors are advised to read our report filed with the SEC for information related to risks and uncertainties facing our business.

So, I thought I'd open up in terms of some key themes and takeaways as we just finished our last quarter and gave guidance for the fourth quarter that we're in. We are working through our inventory correction that is needed for our gaming business. We are making progress. Q3 definitely showed progress in terms of reducing the inventory that we have on the channel. That was also an opportunity for us to begin the very beginnings of our Ada launch of gaming as well, and that is going quite well. We do believe that as we approach the end of Q4, we will have reached about normalized levels of inventory for our gaming business.

Secondly, theme-wise, we've got a great product line-up, new launches of architectures not only with our gaming business, but for our data center business

and our continued work in automotive with our Orin architecture. We have Ada out and running. We have Hopper out for our data center business that will begin ramping in next year, and we'll start production in this Q4 with our guidance that we have.

So, just wanted to open up with some of the key things that we're focused on right now.

## Questions And Answers

**A - Colette Kress** {BIO 18297352 <GO>}

(Question And Answer)

**Q - Chris Caso** {BIO 4815032 <GO>}

Great. Thank you. And maybe starting with some of that, and I think as I've spoken to investors in the past few weeks, there does seem to be an appetite for the semiconductor space after the pullback that we've had. And I think that a lot of folks are looking at the extent to which things are derisk in making sure that no other shoes are dropped. What you said on the call was you talked about that you believe that the end demand in gaming is on the order of \$5 billion over two quarter period, so just about \$2.5 billion a quarter. The consensus estimates of how your gaming business running at about \$1.5 billion a quarter now. So, that's a pretty sizable gap. Does that give you some confidence that again the inventory is burning through at some point, we get back to where that steady state demand is such that even if demand weakens a bit, who knows with China and that -- there are no other shoes to drop in gaming?

**A - Colette Kress** {BIO 18297352 <GO>}

Yes, that's correct. So, we had highlighted at the very onset of -- when we saw the inventory correction need to help folks understand what we saw in terms of solid sell-through in terms of gaming. So, we came up with a view of our normalized gaming of about \$5 billion stretched over two quarters.

Yes, there is always some seasonality. So, sometimes it's going to be \$2.5 billion a quarter, give or take, depending on the quarter leading in. We still see gaming is solid, and we're continuing to watch each and every day in terms of the sell-through that we're seeing particularly as we are looking at the Western holidays right now in that time.

So, we have been under-shipping. We have been under-shipping gaming at this time, so that we can correct that inventory that is out in the channel. And as we discussed, we plan hopefully by the end of Q4 -- as we approach the end of Q4 that we will be in a solid position to that. That means as we move forward, we will start to get back up, we'll get to some point of an equilibrium between sell-through and sell-in, and then we will likely get to where we'll be selling in to build the channel for our

Ada launch, and all of the additional Ada products that we will see coming to market.

**Q - Chris Caso** {BIO 4815032 <GO>}

And following from that, as we look into the April quarter and you haven't provided guidance, and I assume you're not going to provide guidance right now, but what are the puts and takes that we should think about? You've got things like seasonality in the gaming business, you have the normalization in inventory, you have some of the new H100, which starts going into cloud instances starting in the April quarter. So, maybe you could speak about some of those things and things we should consider when thinking about the April quarter.

**A - Colette Kress** {BIO 18297352 <GO>}

Correct. We have given our guidance for Q4, and we're going to start-off new year. Our new year is going to be a year focused on our new products coming to market. Ada is just in the early days right now of shipping. We've shipped one of our very high-end cards, our 4090, and our 4080 is just now starting to sell as well. So, we have plenty of opportunity with innovation that we've brought to market for the gaming.

But that's also true for our data center business as well. Our data center business is now coming forth with our H100 architecture. That's a great opportunity to really take advantage of a lot of the innovation that we put into focus on large language models. Transformers is a very big thing of it, and the overall performance improvement, the efficiencies, the total TCO that you get from H100 is there [ph]. We will start that beginning production here in Q4; and as we move into the year, we will start that ramp of H100.

As we discussed in terms of our earnings, we will see cloud instances already set-up for the H100 as we start the new year as well. So, we know it's a very important product, and we feel that we're in a great position with the new products coming to market, not just H100 in data center, but the networking, which is often coming very coupled together with what we're selling in H100. What's out there though that we can't control is the macro conditions, the macro conditions around the world, they're here in the U.S., they're definitely here in Europe and China, we've talked about in terms of the economy really, really being tough right now.

**Q - Chris Caso** {BIO 4815032 <GO>}

And China is important for you as you go into April, given the New Year holiday that occurs there.

**A - Colette Kress** {BIO 18297352 <GO>}

There is a holiday season in the Asia-Pac China area as well. And we've been still seeing solid in China, however with the lockdowns that they've been experiencing with COVID, it makes things just a little harder for the distribution, for the sale process to happen. So, we'll continue to watch that closely.

**Q - Chris Caso** {BIO 4815032 <GO>}

Maybe you could speak about the Ada launch. And I think there's been a perception that gaming had been one of the pandemic beneficiaries, we couldn't go anywhere and that there would be some impact now that we have other things to do. To what extent have you seen that? What's your perception of the success of the Ada launch as compared to the 30 Series, the 20 Series launch, how do you benchmark that?

**A - Colette Kress** {BIO 18297352 <GO>}

When we look at gaming, we've always discussed that there may have been a little bit of a pull-in as folks looked for some type of sport during the pandemic. We did see some great new gamers joining that market and taken advantage of the leadership that we have in terms of gaming. Not only is Ada a leader in the industry, so is Ampere. So, the number one and number two top architectures are NVIDIA.

We continue to watch what we see in terms of Steam. Steam is still a record level of gamers online gaming, and the top 15 cards on Steam are NVIDIA, all 15 out of 15. So, this is an area that gamers will be gamers for life. And we watch this industry both at the great times and also during some of the hard times. It is a sport that is economical, because you can choose many different form factors at many different price points for the gaming that you want to do.

Our Ada launch 4090 when it came out, yes, it was only the very first card and the high-end card, it's sold out in two weeks. So we're very pleased to see something that was a nod for us that we knew that gaming, no matter what time of year, no matter what the situation is still very important entertainment industry.

**Q - Chris Caso** {BIO 4815032 <GO>}

Okay. We'll pivot over to data center for a bit. And the H100, so maybe you can speak to what that means in terms of performance for your customers, what that means in terms of content for NVIDIA?

**A - Colette Kress** {BIO 18297352 <GO>}

Yes, the H100 in terms of its performance improvement depending on the type of workloads can be more than 6x what we see with A100. We went through the MLPerf for training the MLPerf benchmarking for inference; and in both of those cases, hands down, everything that we do here at NVIDIA shows through in terms of outperformance.

What we see in the market and what led us in the design of H100 is the continued advancements of AI. Each and every season, each and every year, there is something new. They're still in that very early days of AI, but large language models, and the size of large language models and the influence of natural language processing within that is very front and centers, not to say that recommender engines still have ways to go, and very popular with the consumer Internet companies as well to make sure their marketing is appropriately targeted. So, H100 really hones in on that; and when you think about our development of H100, it was done in coordination

with the customers, coordination with the enterprisers, coordination with the hyperscalers, so that they are seen what they would need in there.

As we move forward, it's not about H100 being a tip, it is about being a system, it is about thinking of the process of AI from the very beginning that the data enters the data center to the time that it leaves. And so, you will see a lot of work there not only with the interconnects, the links, the networking, and the memory to all assist in that full process.

**Q - Chris Caso** {BIO 4815032 <GO>}

Yes. I mean, it's a good segue into the next question, which is the sustainable advantage you have in AI. And I don't think there's much investor question about the growth rate of AI. And at this point, I don't think there's much question about NVIDIA's leadership in that space; that's why the room is so packed. Going forward, how do you sustain that? I mean what's the factor? And it has to be more than just the silicon itself. The concerns that have been voiced in the past have been things like start-up companies taking a perch, the hyperscalers themselves doing things. So, how do you fend-off those challenges and retain that advantage?

**A - Colette Kress** {BIO 18297352 <GO>}

So, when you think about our products coming to market, we may first discuss in terms of the performance availability. People ask that they talk about what will we see, but most importantly in there is the software that we've built along the way. And this is not software that in this last quarter or this last year that we were overall developing, you're talking about 10 years to 15 years-worth of developing software for this platform, and really transforming how the world thinks about computing.

The computing architecture of CPU has been here for a lot of time. But to use a GPU, you are really rethinking that application, you're rethinking the use of that hardware in the GPU. So, our full stack view of software is essential. We're able to package it up and create a full system package NVAIE that really is out there to help enterprises start their very beginning of work in terms of the AI or continue in terms of workloads that we're doing.

We have SDKs for some of the leading industries and their top applications, so that they can begin that work. It's very hard to think of a start-up or just a new chip having any work in this model, if there is no software and somebody has to work to do all that work. We have more than 3.5 million developers that develop solutions on our hardware, on our software using our full systems. Each and every time that we bring something to market, it's really about the developers and what they are looking for, and how are they going to use us.

So, I think, we've been extremely successful of continuing that model equally as much as we focus in terms of on the hardware, we are focusing on the development platform, we are focusing on the software.

**Q - Chris Caso** {BIO 4815032 <GO>}

Could you give us some help in how we should think about the growth rate of data center over time, and it's been something that's been incredibly different -- difficult to model? For the most part I think we've all underestimated the growth of it, which is a good problem to have. But going forward, you need to get despair growth rates within the cloud part of your business, which is about a half of the enterprise business, the other half. And content and ASPs are rising as well. So, I think the way a lot of folks have done is taken a look and figured some multiple on cloud growth, but it's not really a very accurate way of looking at it. And I don't know if you have another way of how NVIDIA looks at it internally.

**A - Colette Kress** {BIO 18297352 <GO>}

So, as we discussed in our earnings, we have talked about our business in data center, and we've broken it out with talking about our hyperscales, and then talking about our vertical industries and we're seeing from each end quarter, you're probably about -- have 50%-50% in there. We're going to work on providing a little bit more of a disclosure going forward. Our hyperscales are important, they're large companies, but many of them are also cloud companies in terms of serving our cloud instances for the enterprises. So, if we can take this time to look at what we're seeing also with the cloud, that gives you a better understanding of really the vast nature of enterprises that are using AI and accelerated computing as we go forward. So, stay tuned for us as we move forward on that.

What we see with the hyperscales is just that, setting-up cloud instances on a need for many different types of industries enterprise-wise. They do it also for researchers, they do it also for higher education. And what we are seeing over the period of time is that cloud is continuing to be a very important part, an important part both historically and likely going forward, as when you think about the challenges of the economy setting-up a cloud instance to get started is an easier path than working through significant amount of capital to build out your overall data center. So, it's also becoming more difficult for enterprises to have real size of IT organizations that can master some of these very complex configurations that were seen in data centers today. So, the cloud is going to be very important.

When you think about some of the announcements that we made with Q3, talking about our work with the hyperscale such as Azure, such as Oracle selling with them to the enterprises. This is a real good breakthrough of what we're seeing the important of our cloud providers, but them also realizing the importance of our relationships with the enterprises as well as our software stack. So, now, they have the ability to market the software stack, create instances that an enterprise could get, whether they built it on their own or put it in the cloud. So, we're really pleased with this in terms of the partnership and working with our cloud providers there.

**Q - Chris Caso** {BIO 4815032 <GO>}

Does that mean to say that over time now that perhaps we should see the cloud part of the business going to be a bigger part of your business, as maybe we see some migration from on-premise to cloud, as that just more efficient way of dealing with workloads?

**A - Colette Kress** {BIO 18297352 <GO>}

It's tough to say what the different growth rates will be, because you'll still have many that will desire to be on-prem. But stepping back from it, we're indifferent, right? We're indifferent of whether or not a consumer and enterprise wants to build on-prem, they want to use a cloud instance whether -- what the hyperscale wants to do. That's what we're here for.

We have many different ways in which you can use AI, use accelerated computing, and we have now set-up that throughout the world. So, we're indifferent. But yes, cloud will be a very important industry going forward.

**Q - Chris Caso** {BIO 4815032 <GO>}

Also want to speak about one of the other product cycles you're expecting this year and in Grace, the CPU product that gets added there. Maybe you could speak to in terms of timing, I think an ARM model we're expecting some moderate impact than the second half of this year. Speak to what that means for your customer in terms of performance, what it means for NVIDIA in terms of content?

**A - Colette Kress** {BIO 18297352 <GO>}

Yes. So, Grace. Grace is our first CPU, our first ARM Grace CPU coming to market. We plan to start the testing about bringing that to sampling here in the first half. Second half, we'll talk about it being into market.

What Grace is? Grace works together with our Hopper architecture. It is geared toward very large models, very high-end type of AI. It's not a general purpose CPU, it has its specifics to focus on accelerated and AI together. So, we are excited to bring it to market. We have worked to really understand the importance that that CPU will have on the workload of AI and we'll see how we do in the second half of the year.

**Q - Chris Caso** {BIO 4815032 <GO>}

How do you capture value on that? Because obviously, if a customer buys a board with Grace on it that removes the need for an accompanying Intel or AMD server, which the cost of that platform is pretty close to what would be in A100 today. Is NVIDIA in a position to capture all that value? Because you're essentially giving the customer better performance, and you actually remove some additional cost from them.

**A - Colette Kress** {BIO 18297352 <GO>}

When we are developing all of our systems, it is about a total TCO, when they have moved to accelerated computing. There isn't any one of our architectures in data center that is not improving a TCO. Enabling now a DGX incorporating Grace and Hopper together is just a better TCO for that customer that you have a very specific CPU for the workload that we're trying to do. But it's not uncommon though for us to also use our peers' CPUs as well. So, those have been within our DGX, and that is also beneficial to us just as long as we're moving to accelerated computing.

**Q - Chris Caso** {BIO 4815032 <GO>}

Right. But it's safe to say there's some content that you have to gain in that as you're adding higher performance. NVIDIA is a company generally that prices to value. So, if you drive performance, which is -- in all your market, where ASPs have been going up.

**A - Colette Kress** {BIO 18297352 <GO>}

Yes. We price for the innovation and the savings of that TCO. That's really underling whether that be gaming, whether that be data centers, that's what we're trying to do.

**Q - Chris Caso** {BIO 4815032 <GO>}

Fair enough. With some of the macro concerns, I mean several -- and some of our colleagues at Credit Suisse have some concerns about cloud CapEx in 2023, obviously there's some pressure in enterprises. Does that moderate your view of the H100 or Grace product cycle as go into 2023? Is that something we should take a little more modern expectation and expect more in 2024? How do we think about it?

**A - Colette Kress** {BIO 18297352 <GO>}

We've studied and we're continuing to watch the macroeconomic conditions. One of the things that you see during these times as enterprises will really study and work through what they're going to purchase, some of that sometimes is a lengthening of the time to overall close deals. That may happen, and there are though certain industries that AI and accelerators is so important to the work that they do, important to their monetization of their platform, important in terms of new growth for them.

So, focusing on accelerated computing and AI, whether it be sustainability, trying to think through the use of energy, it's a more efficient platform, and the leading edge of AI will be important for so many businesses. So, we're in a good position for that, with both a new architecture as well as being a leader in that industry. The macroeconomic conditions are there though, so we are watching it carefully, but we know that the cloud will be essential to keep going during this time.

**Q - Chris Caso** {BIO 4815032 <GO>}

Okay. I'll pivot to something else, which NVIDIA has spoken about pretty consistently over the last year. I don't think it's as high of an investor focus, but it's software. And I think software is interesting, because when you add software economics to a semiconductor company model, it could be pretty important. How important do you expect software to be on the financials as you go over the next 18 months, 24 months?

**A - Colette Kress** {BIO 18297352 <GO>}

Yes. So, we have started selling our software separately. For example, our data center products or gaming products, all of them have had software since the very beginning of the time. But what we've done is we've packaged up key software in probably three different areas.



One, NVIDIA AI Enterprise, which is a set of system software that helps every enterprise get started and essentially treat this what they see with many of their other enterprise software capabilities, being able to monitor jobs, being able to manage the whole data center as a whole and essentially get started on much of that workloads. That's very difficult for enterprises to start, they wouldn't necessarily have the teams on-board, the engineers on-board to do that. Essentially, this is also an area, where enterprises want that software, because with applying that to some of their mission-critical applications and work, they want to know that we are there to both support, update and keep all of that software current. So, there is a double win in terms of there. So, that's our first piece.

The second one is are Omniverse. Our Omniverse is an important area thinking about the 3D virtual world and the simulations of much of the build design -- digital-twin environment that's out there. There's many different use cases. So, this is an area, where we can sell in a pod with our enterprise customers or even in an individual basis.

Our third piece is focusing on automotive. As we've been working on automotive in terms of AV, we have agreements, where we will be selling software along with our hardware and maintaining that software over the life of the car. So, this can be a meaningful piece in these three areas for us. We're just getting started in working and selling it. You'll even see with H100 and certain forms of H100, it will come with our software coupled together with the H100. We're excited to see this. This is a driver not only of top-line, but it also an improvement can be in terms of our gross margin for the long term. So, we'll see this drop.

**Q - Chris Caso** {BIO 4815032 <GO>}

What's the chance that we get to see software kind of broken-out with a bit more granularity in the near future? Is that something you've contemplated?

**A - Colette Kress** {BIO 18297352 <GO>}

We will be giving you information of some of our key customers and software and key milestones that we reach in terms of software. We think that would be the most beneficial to help you and walk us through this journey with us.

**Q - Chris Caso** {BIO 4815032 <GO>}

Okay. Maybe I'll check to see, it's a packed room, if there's a question that somebody like to ask. Anyway, I'll continuous as you think about it. Automotive as well -- and automotive, there's been a transition from the infotainment part of the business, which has trended down, and we've been waiting for ADAS full self-driving to have an impact now. It sounds like we're kind of on the edge of that beginning to occur. So, maybe you could speak to that a bit.

**A - Colette Kress** {BIO 18297352 <GO>}

Correct. What you're seeing today, our automotive business has some very solid growth right now. Our Orin architecture came to market at a very important time,

where there's two important things in the automotive industry, EV or electrical vehicles as well as AV.

Now, when you are dealing with your EV customers, they're really thinking through their complete redesign of a car from the bottoms up, they're building a car from scratch. Important piece of that is that computing platform. Orin is the exact great product for that to do from soup to nuts. So, we have them working not only on consolidating all that different software into one computing platform, but also adding AV features to the work that we are seeing.

So, NEVs across the world, particularly in the Asia-Pac and China area have really lodged on to our Orin architecture and they're using that. We also have now agreements with very large companies, both Daimler, JLR and futures to really provide them not only the hardware, but a full software stack. Those will be coming in the next couple of years. They will start in terms of pre-production, they'll start in terms of sampling in a little bit more than a year, and then we'll see in terms of volume after that. So, the automotive has those two things occurring at the same time.

**Q - Chris Caso** {BIO 4815032 <GO>}

So, really -- and I guess the way to think about it is 2023 kind of going to '24 is more the NEV, and then '24, '25 to larger OEMs. Is that right way to think about it?

**A - Colette Kress** {BIO 18297352 <GO>}

We believe that's the general direction that it will go.

**Q - Chris Caso** {BIO 4815032 <GO>}

And will the NEV -- will it be material? And that -- and you said the auto business, there's been strong growth, but the high-class problem you've had is, you haven't had strong growth in a lot of different areas. So, it's kind of stayed as a smaller percentage of revenue. Once we get to those OEM, the larger like Daimler '24, '25, should we expect that auto grow as a more significant portion of NVIDIA revenue at that point?

**A - Colette Kress** {BIO 18297352 <GO>}

That's correct. When we look at our fiscal year '25, we do believe there will be a significant amount in there. Keep in mind, our automotive customers not only what they have within their car is important of what they're purchasing from NVIDIA, but they're also very important data center customer. In order to build what they want or AV of the future, they have our computing back in their data centers. It's a very important part of that. They've also been working of course with our workstations since the beginning of time. So, it's interesting to see, yes, we have what is within the automotive car itself. But keep in mind, we also have a great business with them through the data center.

**Q - Chris Caso** {BIO 4815032 <GO>}

Okay. We're almost out of time, and I'll end in a question on margins. And I guess in NVIDIA's case being fabulous with so many products, your margins have been fairly stable and is there anything that we should think of with regard to the growth rates of different businesses or you're spending itself that we should consider with regard to gross margins or operating margins?

**A - Colette Kress** {BIO 18297352 <GO>}

So, looking at our gross margins, we continue to see not only the manufacturing costs, but remember we have so much engineering that's working on the software. First, the value that we distribute in terms of the innovation has allowed us to have very solid gross margins. In the last year or a couple of years, you've certainly seen the cost of manufacturing, the cost of components, contract manufacturing continued to increase. We've absorbed a lot of that, and still had very, very solid gross margins. We take that opportunity each time that we look at a new architecture on the value that we're providing, but we also have to take into account the cost that it takes to do some of this very, very high-end manufacturing that we do. So, those are the important pieces.

Now, when you look at our overall operating margins, it's another case of very strong operating margins. During this time, it is always an opportunity for us to think through how we spend our money. But we're quite an efficient company, we're only 25,000 people large for a company that's operating well into double-digit billions in there. But we took this time to slow-down our hiring and focus on the engineers, focus on the teams that we had. And I think the teams rallied around that, but also continuing to look to make sure we are spending our money wisely on that. So, that will be our focus, focused not only on gross margin for the long term, but also focusing on our operating margins as well.

**Q - Chris Caso** {BIO 4815032 <GO>}

That's great. It looks like we're out of time. Colette, thanks for your time.

**A - Colette Kress** {BIO 18297352 <GO>}

Thank you, Chris.

**Q - Chris Caso** {BIO 4815032 <GO>}

Thanks, everyone.

*This transcript may not be 100 percent accurate and may contain misspellings and other inaccuracies. This transcript is provided "as is", without express or implied warranties of any kind. Bloomberg retains all rights to this transcript and provides it solely for your personal, non-commercial use. Bloomberg, its suppliers and third-party agents shall have no liability for errors in this transcript or for lost profits, losses, or direct, indirect, incidental, consequential, special or punitive damages in connection with the furnishing, performance or use of such transcript. Neither the information nor any opinion expressed in this transcript constitutes a solicitation of*

---

*the purchase or sale of securities or commodities. Any opinion expressed in the transcript does not necessarily reflect the views of Bloomberg LP. © COPYRIGHT 2024, BLOOMBERG LP. All rights reserved. Any reproduction, redistribution or retransmission is expressly prohibited.*