

## NVIDIA Corp at Nasdaq Investor Program

### Company Participants

- Colette M. Kress, Executive VP & CFO

### Other Participants

- Mark John Lipacis, Senior Equity Research Analyst, Jefferies LLC, Research Division
- Unidentified Participant, Analyst, Unknown

### Presentation

#### Mark John Lipacis {BIO 2380059 <GO>}

Okay. I think we can get started with the next fireside chat. Very excited to have NVIDIA today. And we have Colette Kress, the CFO. She has been the CFO of the company since 2013. And previously, she was CFO at Cisco's Business Technology and Operations Finance Organization and, before that, with Microsoft for 13 years as CFO of the Server and Tools division. So we're very excited to have you here. Welcome, Colette.

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

Well thanks so much, Mark.

### Questions And Answers

#### Q - Mark John Lipacis {BIO 2380059 <GO>}

I think we'll just jump right into it. Probably from an economic standpoint, the gaming business has really done phenomenally well. It grew 44% in 2016, as you guys had launched Pascal in the back half of the year. And on the -- and on your Analyst Day, you showed a slide that -- you showed that ASPs and units equally contributed to the growth over the last five years. What's going to -- is that -- what do you think drives for the next five years, the growth, units, ASPs, both?

#### A - Colette M. Kress {BIO 18297352 <GO>}

So if you think about our overall gaming business and how it's transformed probably over the last 10 years, we were well known as a very big component of the PC industry in terms of enabling overall graphics, which built that overall gaming market. That was one of the key pieces of enabling high-end graphics. But that overall business has evolved. It's a -- the market has substantially grown and the market has grown because gaming is everywhere. Gaming is no longer just a sport

with you and the overall computer. It's an entertainment arena, where you have people online for the social experience and playing with their friends. So the overall growth that we've seen over the last five years is really about that market expansion, bringing on more and more gamers, gamers to come on and play with their friends. This is both from gamers continuing to refresh to the higher-end platforms to take advantage of the higher production value of games that have been there. But also from the broad main access of broadband around the world. And the first things that those do when they come online to broadband is to game. It's a great social sport to do so. We'll probably continue to see some of the exact same drivers that we have seen in the last five years produce the industry going forward. There's new things each year that continue to dazzle the gamers in terms of both the games, the future approaching virtual reality, the mobility factor of increasing your overall ability to have high performance in a mobile platform such as a notebook. All of these will continue to grow. We get asked a lot about our ASPs and the average ASPs. I think a different way to look at this is to say the overall value that we deliver for the dollars that are spent is more and more performance on each generation in terms of our architecture. And what that means is folks go and buy even higher and higher-end platforms at the refresh or even coming in for the first time. So I think all of these drivers still are the same. The exact makeup or what that looks like over the next five years, I think that's just too hard and too specific to pinpoint.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

Okay, fair enough. I want to shift over to the data center for the next set of questions. So at Jefferies, we've hosted 4 artificial intelligence conferences around the world and -- where we typically have about a dozen artificial intelligence startups talk about their -- how they're using AI for new business models. And when we ask them what is your hardware platform almost -- to a company, they say, the reason we can do this is NVIDIA and CUDA. And so CUDA is not hardware. So what is -- what's CUDA? And can you help us understand what -- why is everybody saying this?

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

Okay. So let me start, in terms of what we're experiencing in data center. As we moved in terms of past our ability to excite in terms of the gaming and the overall graphics, the GPU is well positioned to be very influential in accelerated computing. We have been working on accelerated computing across many different platforms for more than 10 years. The first instances of things that you've seen in high-performance computing over the last 10 years was very, very essential. And 10 years ago, we took the opportunity to enable on every single one of our GPUs a programmability model, CUDA. We went out and taught that to the masses. We taught that in higher education. We taught that in research institutions. Where we are right now is probably more than 500,000 developers around the world that know CUDA. This enables additional use cases, not use cases that we have to always develop. But those that have been taught and trained on both the parallel capabilities of the GPU, the overall extensibility of a GPU in so many different cases to consider a GPU for future in terms of workloads. What's happened then over the last five years is the focus of AI. AI, for the last 20 years, software-written code, trying to get to where you'd have high-enabled AI computing structures. But new form factors, new frameworks developed using GPUs, GPUs that were very well engineered to design deep neural nets and establish additional frameworks on top

of that, allowing the ability to create AI different workloads, AI different types of applications. So our initial work just 10 years ago, that initial investment in terms of developing on CUDA and now continuing to expand CUDA with a set of frameworks, a set of building blocks that also allows AI, has really positioned us in a position to make AI available everywhere, democratize it. Allow anybody with any type of platform, any type of framework that they want to work with, the availability of CUDA, the availability of NVIDIA.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

One of the biggest questions that we get is how big is this AI market or deep learning market. And since there is -- since these tools are creating brand-new markets that don't exist, it's hard to figure that out. But you guys took a crack at it because I mentioned you were getting the question more than we were. So can you review what you did? Like how did you get that number? How did you go about assessing the -- quantifying the market?

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

Sure. So what we had focused on was an extrapolation of some of the speed of adoption that we're seeing today. We see, not in terms of looking at the overall data centers in the percentage of them that are NVIDIA GPU in them. But rather to say, how much computing? How much computing, in terms of operations per second are being computed using GPUs? How do we think about that, in terms of knowing that we're in the early stages and extrapolating that for the amount of data that we think is still out there to be trained, the amount of information in high-performance computing that can still also be highly accelerated and extrapolating that in a forward curve? We all know that we're in the early stages and I don't think anybody can say with perfect perfection, in terms of that size of the market and it's not intended to be really looked at as a forecast. But we do know that the nature of the configurations that are being so well used today in the data center are absolutely likely to change as we move forward. The use of acceleration is going to be a necessity to improve the overall throughput out of the data center going forward. There's different forms of accelerations to do that. But a GPU is so uniquely positioned to continue to capture a significant amount of compute that will go forward, whether that be for AI, whether that be for high-performance computing or whether that just means virtualizing workloads in terms of in the cloud. So those configurations will be different. It's not a replacement. It is just more condensed ability to build those infrastructures going forward.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

And when you think about that market. And I think most investors that we speak with that have done some work on it come to the conclusion that there's 2 parts to the market: there's a training market and there's an inferencing market. And I think most people are -- a lot of people are of the view that you dominate the training market and there's really no other solution. But the inferencing market, I think there's a debate. So are you guys -- how should we think about the competitive environment on the inferencing side? And maybe talk to the training side. Would you agree with that consensus view?

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

So on the training side of AI, yes, we do have a good leadership position in that training market. The GPUs from probably day 1 of the birth of overall deep learning in training were being used. And we still hold that position, not from anything other than continuing to advance the overall innovation in terms of our GPUs to take on higher-end workloads as we've moved. The overall performance increase from each of the generations since the beginning of training has vastly and continuing to improve. And we will continue to work on providing to those needs of the overall training. We're in the initial days of training. I don't even think we can fathom a day where all of the data in the world has been trained. There's the multiplication factor of you train one data but then add it to another set of data. And you've got a whole another training opportunity of adding these things together. But that moves in terms of the next large opportunity and the next large opportunity, in terms of inferencing. Now this is a vast market as well, very, very large. It has not been the case, in terms of where that has been overly GPU-led. A lot of times, this has been a CPU market. We look at this as the next opportunity, as something for us to add to the overall training environment. And we have found definitely some great cases of our very specific GPUs that have been tailored for the inferencing market. And even our new and upcoming launch in terms of V100, our Volta one, has been very well engineered for the multiple precision that may be necessary for inferencing with the tensor core inside of it to extract and do inferencing in these cases. We're working with many different types of customers on that. But yes, the inferencing, a large market and something we do plan, in terms of growing into.

**Q - Mark John Lipacis** {BIO 2380059 <GO>}

The data center market is -- has a number of buckets in it. You have not just neural network in AI. But you have high-performance computing. That business has grown 200% in each of the last 3 quarters, I believe. And so how -- what has driven the growth? And can you give us a sense of the bucket sizes and what has driven that growth?

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

Yes. So our data center business probably has multifaceted types of businesses. You touched on high-performance computing, which is our longest-standing product or part of the business in terms of working on it for 10 years. We're moving to an era where the conversions of high-performance computing in AI is front and center. The importance of accelerating high-performance computing couldn't be more predominant at this time. They really are working on the additional throughput capabilities. Acceleration is very popular in many of the key top applications in high-performance computing and most of the brand-new supercomputers that are being added all use acceleration. GPUs are one of the most important pieces on that. It did double year-over-year in this last quarter and we really wanted to demonstrate that it is not the high-performance computing of the past that was growing nicely. It is moving at a very fast pace, as well as the rest of our data center business. The other businesses, we talk about what we see in AI, AI for training, AI for specific overall inferencing. You can slice and dice that also in the manner of what is being built for internal applications, internal applications that we see Internet service providers overall building. As you know, their infrastructure is very key to their overall business

model and they have been some of the ones that have run probably the fastest, in terms of building out AI applications that you and I use every single day. Additionally, these same hyper-scale providers are also establishing cloud instances. This is really based on what they see as high demand, being asked for GPU instances in the cloud so that many people can get started in the same way that the hyper-scales are. Much different than the overall infrastructure builds that you may have seen 15, 20 years ago, they would be self-built into the IT, going and beginning and starting in the cloud, relying exclusively on the cloud or potentially bringing that on-premise allows those large enterprises to get started, research labs to get started, as well as many of the startups that you had seen on a lot of your travels. Those cloud instances are very, very key to them. So you now see every cloud provider providing GPU instances and we'll continue to see that advancement of enterprises in many of the key partnerships that we've built to build out, in terms of the enterprise as well. So there's 2 other businesses within our data center. There is GRID. Our GRID business focuses on a one-to-many, a one-to-many of one GPU to many overall users. Virtualizing the overall GPU and streaming down from the overall cloud can be used in terms of very graphic-intensive applications or just those that want to secure their overall workstations, their overall desktops and carry that in cloud and stream down the overall applications. We're seeing this to be a very important model, an important model for both collaboration, as many of the workforces are all over the world and also, an ability to just share work and share very high-intensive graphic applications. Our last business is really, again, moving back to focusing on enabling AI everywhere and what we have built is AI supercomputers, our DGX brand. What this is, is a containerized version of soup-to-nuts view of a configuration with GPUs but also a full software stack, a software stack that allows you whatever framework is your favorite and also well kept up with the latest and greatest. But it allows enterprises, research as well as startups, to essentially just plug in the supercomputer. We're starting to see mass purchases and buying in bulk of hundreds strung together for them to start their overall AI applications. This is yet another way, in terms of us fulfilling the overall demands in many different types of use cases. So early stages of the business growing but also, growing very nicely.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

Shifting over to Automotive. So historically, your Automotive business was digital dash and infotainment. And that had a very -- that had a healthy growth rate in itself. Then more recently, you had a public announcement with Tesla for -- to develop autonomous driving and the business continues to grow healthy. If you look at the more recent growth in the Automotive business, to what extent is it being driven by the infotainment versus the autonomous? Is the autonomous a big part of that business yet? Or is that still one to come?

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

Sure. So let me kind of summarize what we have in our Automotive business. Infotainment is still the lion's share of our overall Automotive revenues. We have been in that business for 8 to 10 years, focusing on high-end premium infotainment centers, whether that be that center console, the digital dashboards. You will see many of the high-end European cars that we have around here using NVIDIA. All the Tesla motor cars also use our infotainment systems and many other different types of brands. Overall, infotainment has now broad strings, I would say, that you see in

most of car manufacturers. We have focused, in terms of those higher-end cars, where they truly appreciate the graphic -- high-end graphics that we enable inside of those cars. But that has allowed us to continue to focus, in terms of what we think is a larger transformational area in the automotive industry and focus on autonomous driving. What this takes is the work that we had developed in the data center, the exact same learnings that we had learned in terms of AI, employing that to the very hard problem of autonomous driving. Now we're still in probably the early startup stages of that. Tesla Motors, as you've indicated, did choose our platform. We're on the third-generation platform of our DRIVE PX platform inside of cars. And the cars with Tesla Motors that are coming off the production line and are on the roads today are incorporating that DRIVE PX platform. Their plan is to take that to Level 3, to Level 4 using the exact same hardware today. Now that rolled out in 2016. And so there is quite a lot of interest in terms of others bringing Level 3 and beyond to the road soon. That moves us to many of the other announcements that we've made, in terms of partnerships that we have focused not only at top OEMs, including Audi, including Daimler. And most recently, at our GTC, we announced Toyota in terms of us working together. There are other key partners in Tier 1s, which are essential in terms of hooking together the overall infrastructure and they have also adopted our DRIVE PX platform in many cases. So we are working right now with probably more than 225 different OEMs, startups, research, all focused on that transformation and AI problem, as they see it, for autonomous driving. But it's a couple years out but what we are seeing in the short term is some production, definitely, regarding our Tesla Motors. And we'll also start seeing Audi later this year. Then additionally, we also have development platforms. And what I mean by that is we have development services as we are helping them. This is not just a solution that exists at the hardware, very similar to our data center. There's a significant amount of software work that we work together with the OEMs, enabling autonomous cars as they go forward.

**Q - Mark John Lipacis** {BIO 2380059 <GO>}

Okay, that's helpful. You had mentioned a Volta product before. And that's, I believe, supposed to be launched sometime in the second half of the year. What is -- what's the difference between Volta and Pascal?

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

The difference between Volta and Pascal, Volta is our most recent architecture; Pascal was our last architecture. Now keep in mind, our GPUs are all unified on the same type of architecture. What I mean by that is what we sell in terms of for gaming, pro visualization, automotive and/or data center is essentially unified on the same architecture. That allows us a tremendous ability to leverage our engineering across. So that same architecture and we continue to vertical integration for each of the different markets that we may take. So we've come out and announced our Volta just at our GTC last month and we announced the very first product, which was V100. You will see a series on focus in terms of Volta as we go forward. But V100, our first focus in terms of product, is a product focused in terms of on the data center. This product is, as we've discussed, both accessible to the overall training market but also a very key component focused on the tensor core, which is focusing on overall inferencing in the overall AI and data center. The overall performance increase is substantial versus what we've seen. And we are in production as we speak.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

Excellent. So the revenue growth has been impressive and you're throwing off cash. What are you going to do with all the cash, it's capital allocation strategy?

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

The growth continues. Our last quarter, again, a very overall healthy growth rate. We had continued to prioritize capital return. We do look at our cash. Our #1 focus is to focus on investments in the business. We do want to make sure that we are making the appropriate investments to capture these large growth markets that we have in front of us. So whatever we need for that business, that will be our first priority. Additionally, we will look at areas of potential M&A. We haven't been the overall acquisitive company. We have focused only on smart -- small, early-stage types of businesses to overall purchase but those are the things that we will do. After that, there will be free cash flow. And we take it as a priority to focus on that for shareholders. We do have a dividend program. So we will fund first, in terms of the overall dividend. And we will take an opportunistic opportunity to repurchase our shares as well. For fiscal year '18, we've established an intention or a goal of about \$1.25 billion to return to shareholders in fiscal year '18.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

Excellent. I think we have one -- time for one question from the audience, if there are any.

**Q - Unidentified Participant**

Yes. Just to return to the question about inferencing, which you brought up. We're seeing some people like Apple and Qualcomm focusing on inferencing on mobile platforms and developing hardware and also software. Now I know it's been awhile since NVIDIA's focused on mobile phones and things like that. Is that something you think you will need to return to?

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

So our focus on inferencing will probably be broader, in terms of focusing in terms of the needs in terms of the data center and also focusing in terms of on the edge. If you think about our platforms and what we refer to as a mobile inferencing, it doesn't necessarily be mobile inside of a phone. The question is, is it on the edge? If you had seen, in terms of our discussion at our Investor Day, we really focus in terms of smart cities, the overall surveillance in terms of improving that. And there's a tremendous amount of edge computing availability, in terms of there. So when we think about mobile, we'll think about it more in terms of on the edge and enabling in terms of platforms in there.

**Q - Mark John Lipacis {BIO 2380059 <GO>}**

I think that's going to have to be the last word. Colette, thank you very much for joining us.

---

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

All right. Appreciate it.

*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.*