NVIDIA Corp at Nasdaq Investor Conference

Company Participants

- Colette M. Kress, Executive VP & CFO

Other Participants

- Unidentified Participant, Analyst, Unknown

Presentation

Unidentified Participant

Okay. I think we're going to get started. I'm very delighted to have Colette Kress, who's the CFO of NVIDIA, for a fireside chat. I'm going to have some -- I have a set of questions for her. And we also have the tablet. So if anybody has any burning questions, please feel free to send them through the Meetoo app. So welcome, Colette.

Colette M. Kress {BIO 18297352 <GO>}

Thank you.

Unidentified Participant

So I think maybe we'll start off. You recently had -- Jensen was in Computex last week. And you hosted GTC in Taiwan. And perhaps, you start out summarizing the highlights of those events.

Colette M. Kress {BIO 18297352 <GO>}

Sure. So we are in the process of going through what we call our international GTCs. Normally, we hold our GTC, which stands for our GPU Technology Conference, first in the U.S. in California and then we go on a world tour. This is an opportunity for us to connect with our developers worldwide. So our first stop was in GTC Taiwan, building upon some of the great announcements that we made at GTC in the U.S., where we were focusing on our next set of products for the data center, focusing on our V100, Volta 100. And moving to 16 GB to 32 GB availability, a tremendous increase in terms of overall performance and also available for a lot different workloads now that you move to 32 GB.

Additionally, we had announced our DGX-2, which was our second version of our full containerized version system, able to essentially just be plugged in, a full stack

incorporating versus our DGX-1, which was 8 GPUs. DGX-2 incorporates 16 different GPUs.

We also, because of the additional amount of GPUs, incorporated our own NVSwitch, which serves in terms of a key part of that overall connection between a lot of the overall GPUs. So therefore, moving to our GTX Taiwan, we announced our reference platform, what we refer to as our HGX, HGX-2, which allows us to provide that same reference architecture that we incorporate in our DGX as well as many in terms of our ODMs and OEMs as they put in terms of in their overall server configurations as well as often in terms of what we see in terms of the CSPs.

So providing that overall reference architecture allows them the simplicity of learning from all of our different design and building upon that design that they can easily go to market with our new products and our introductions. Many are extremely excited about this reference architecture now being available. And we're working in terms of that availability soon.

That moves us, therefore, to Computex. Computex was a great opportunity for -- to us showcase some of the greatest things that we had in terms of Max-Q. Max-Q, if you recall, is our ability, both also with the reference architecture and also landing still on the thin and light, which is so important for our either gaming and/or in terms of our enterprise, incorporating a super high-end GPU inside the overall thin and light.

We had worked with many of the different OEM manufacturers in terms of a reference architecture that you'd be able to incorporate that. We have now showcased more than 28 different notebooks, for example, high-end gaming notebooks for the mobility of gaming and put those onto display in terms of at Computex.

Now that we've headed back, we also had a great event in Tennessee, one of the greatest states in the U.S., where we actually demonstrated -- which is now complete, the key overall supercomputer or what we refer to as the fastest supercomputer in the world, which is leveraging NVIDIA's architecture along with IBM.

Unidentified Participant

That's a lot. So thank you for that summary.

Questions And Answers

Q - Unidentified Participant

So I think also at Computex. And I'm not sure if it was -- but I read a quote that Jensen said that -- he was quoted as saying that ray tracing is the biggest thing that has happened to gaming in 15 years. And for a business that's been growing in the over 20% range, I think that's a big statement. So what is ray tracing? And why is it important?

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

So ray tracing is really some of the underlying key components which allows overall graphics. For essentially the last 20 years, we have simulated ray tracing by overall the availability of light. Light is essentially what brings together what you see as a 2-dimensional picture, what you see in terms of in film, in terms of the use of overall lighting, which brings things to light. What we have now enabled versus simulating after all of these years with the overall GPUs, with the overall performance and research over the last 10 years to enable ray tracing now in realtime, which is a huge development and what people look at sometimes as called the Holy Grail of overall graphics in that capability. So after those 10 years' worth of development in terms of research, we're bringing that to market. It is currently now in the hands of software developers worldwide focused on gaming but also focused in terms of our enterprise applications and the needs for overall graphics. So yes, we're extremely excited about this, as excited as the overall software development community is in terms of that.

Q - Unidentified Participant

And where is it in the deployment cycle? Are you deploying it in the professional market or the gaming market right now? Can you give us some idea of how to think about how this comes into the market?

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

Yes. So outside of what you see in terms of the software developers, they're working. It's a little bit of a chicken and egg that you need content as well as also naming the overall boards to support it, the overall hardware to support it. So right now, it's in the hands of many of the overall software developers. And they're busy working in terms of the content. But we've also enabled it for Quadro and our Quadro V100, which will, therefore, incorporate the use of ray tracing inside of it.

Q - Unidentified Participant

And how about on the gaming side? Is it in your gaming GPUs now? Do you -- when do you expect to have it?

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

Yes. It is not incorporated in our current gaming GPUs in terms of our Pascal. And that will be sometime in the future.

Q - Unidentified Participant

Okay. Fair enough. So on the topic of gaming, Fortnite has swept the gaming industry by storm. And one concern we've heard expressed by investors is that since the graphics on Fortnite seem to be less intense, there's a concern that perhaps there's a new era of gaming that requires less-intensive graphics. Is that a legitimate concern in your view? How do you think about that as a risk (of business)?

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

Yes. So an interesting discussion in terms of a new genre of games that have recently developed. Every couple of years there's just yet another way that people think about overall playing games. As you know, the transformation of gaming over the last 5 to 10 years has grown astronomically in terms of the influence of the social platform that gaming enables. And it really allows the bringing together of a social community, both for super, super high-end games or in sometimes sports game or in the case of others that may not be overly graphically intensive. But the important parts of it is the overall growth of gaming overall and bringing more and more now into the influence of games and socializing. In many of these games, you can have more than 100 of your friends or 100 different players at one time all working in terms of that end key type of prize. So with one game every once in a while not necessarily being highly graphic intensive, it still increases the overall focus in terms of gaming as an entertainment sport. In the same manner, there's other types in the same arena, such as PUBG, which are extremely graphic-intensive and requiring some of our mid; and/or even higher overall gaming platforms to play that at the right overall speed. So we do believe these continue to help drive the overall gaming market even though every once in a while they may not be graphic-intensive.

Q - Unidentified Participant

And Epic Games, who's the maker of Fortnite, recently gave a demo of realtime ray tracing. Did you work with them on that? And can you share with us how many games are being worked on with -- for ray tracing?

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

Yes. So essentially, when a new game comes to market and if it is geared up to be a generally popular game, we have tremendous amounts of engineers working every single day with them in terms of helping them in terms of the features that they want to put into their games and fine-tuning. So those things can really come to life as they vision them to do. So yes, those availability of ray tracing, we're aware. I think we're in the early stages of seeing how many different games are going to be coming since our most recent announcement. So stay tuned. We'll talk about more in terms of the games when they're here.

Q - Unidentified Participant

Okay, we will. Let's shift gears to the data center side. So that business has grown remarkably recent -- more recently about $100 million per quarter for the last 4 quarters. It's now running between $2.5 billion, $3 billion of annual run rate. To what extent has the growth recently been driven by the supers and the hyperscale players versus HPC versus enterprises building their own data centers?

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

So over this time that we've seen our overall business grow, I think it was -- it feels like just yesterday when we were talking about a business that was about $25 million a quarter. So it's grown astronomically. And that doesn't grow from just one thing. I think it's always key to understand the overall breadth and the depth of now what we are accomplishing in terms of the data center. You can focus in terms of the key

types of workloads that we see happening in terms of the data center. It's focused primarily on accelerated computing. That is absolutely front and center. But I think there's many different forms of accelerated computing and what is occurring in terms of the data center. Number one, the focus in terms of high-performance computing. Probably one of our first areas of focus more than 10 years ago is accelerating some of the largest supercomputers in the world. As we've discussed now having the largest supercomputer based on NVIDIA's overall architecture, there's still a phenomenal amount of growth with supercomputers that we both see today and we'll likely see going forward. Those are supercomputers that are in that top 500 but also those that are extremely important out in terms of the industry. Mathematical computations that are necessary with extremely large datasets and speed of accomplishing those jobs are extremely important. And that continues to grow widely within our overall data center business. But then we can focus in terms of the breadth and the depth that we have seen in terms of data centers that are moving from just CPU configurations to overall configurations that leverages an overall GPU. In this case, probably some of the most important workloads that we are seeing are focused in terms of artificial intelligence or focusing in terms of deep learning, deep learning training and even moving in terms of the area of inference. And I think it's important to see this vast and wide growing area, not only along the overall hyperscales or what we refer to often as CSPs but also a wide area in terms of the industry, which just focuses then on the next area of growth that we've seen, which is the overall cloud instances that have developed over the last couple years. The cloud became a very important linkage between the need in terms of AI. But the ability to get to our overall enterprises, our start-ups as well as also in terms of our research labs that we see around the world. It's an easy ability for them to get started versus thinking about building the complexity of those infrastructures inside their own overall network. So the cloud instances have popped up all over the world as well and are being very widely used in terms of the use cases in terms of AI. So we've now expanded to almost each and every hyperscale around the world, at least all of the large and major ones. We're moving to what we refer to as the consumer internet companies, which is essentially that tier down from the hyperscales, those that are broad and rich in terms of their overall applications for consumers that are necessary to incorporate AI to be competitive in their overall monetization of how they do. And now you are seeing them use the overall GPUs for their work as well. So what we've seen is just a tremendous broadening from an overall customer set but then also adeptness in terms of where they've incorporated in their infrastructures. You have it where they may be purchasing our Volta architecture in the overall Tesla boards alone. They may be purchasing overall for cloud instances. They may be repurchasing our DGXs, which is a full end-to-end system that they are allowed to just plug in and stream together a whole linkage of them together. But you can also be using our overall infrastructure for virtualized instances and using GPUs in the cloud as well.

Q - Unidentified Participant

The kind of growth you have seen in that market does not go unnoticed. And it seems you can't turn around without bumping into a new parallel processor company with a new architecture. But so far, it seems like your market share is extremely high. What are you seeing in the competitive landscape in the data center? And why -- can you help us understand why we haven't seen success from some of these new architectures?

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

Yes. So the overall growth that we've seen in the data center was fueled by a couple things that had reached the overall market. One, an understanding that the slowdown of Moore's Law is in front of us or is actually here today. The slowdown of Moore's Law puts in a question in terms of how am I going to get additional throughput and compute capacity out of the overall configurations I have today versus what I had seen in the past. The use of acceleration, therefore, is front and center. Then number two, the overall processing capability of the GPU put us in a unique position to be well-engineered for many of the AI workloads or the modern AI that we see today. Because of that overall availability, you now see many people also interested in terms of what we have built. We all know that artificial intelligence will probably be with us more than a decade, probably close to 20 years, spanning in terms of those building upon the overall capabilities of AI. Thus, others come in to say I would love to be a part of that. We're not confused that, sure, there's always going to be multiple players in terms of a lot of these different markets. But some of the work that we have been working on, not work in terms of this year. But essentially the work over the last 10 years, have enabled us to be in the position that we are, enabling more than 10 years ago, the overall software development platform on top of each and every single one of our GPUs, moved to the ability to even the birth of overall modern AI because people found a use case of the use of the overall GPUs in those instances. So when we think about the overall competition, I think there's always that thought that it's something quick and easy that is easily to solve without fully understanding that we have a company of more than 8,000 hardware and software engineers focused on one thing, overall -- GPUs for the overall different markets that we have. And that's a hard thing to kind of keep up with given the pace that we're able to bring product to market, the complexity of the overall processors themselves and our ability to not just focus on a piece of hardware but focus on the full software stack that it is. So yes, we see a lot of interest in terms of people wanting to get into the market. But at this time, we see our platform being used quite broadly because of its ability to be used everywhere, not being particularly interested in any one network, any one operating system, any one AI framework. We'll support all of those. And that's, I think, what has differentiated us a lot from where we see in terms of competition.

Q - Unidentified Participant

We've -- at Jefferies, we've hosted about half a dozen AI conferences, where we have 5 to 10 AI start-ups come in and talk about what they're doing for the new business models and disruptions. And when we ask them what hardware you're using, they say the reason we can do this is NVIDIA and CUDA, which was a lightbulb for me because it told me that this is not just hardware solution. It's a software solution, too. So can you maybe -- maybe we just go back to the beginning. What is CUDA?

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

Sure. So CUDA is our software development platform. 10 years ago, started working on it. We're on version #9. That is now available across every single one of our overall GPUs. With CUDA and all of our overall developers that we have out there, we have close to 850,000 developers focused on NVIDIA overall platforms. That's multiple years of us both teaching the overall capabilities of CUDA, how to leverage an

overall GPU as well as them coming to us in terms of sharing some of their use cases so that we can continue to build and support them in that area. CUDA doesn't stop just there. There's a tremendous amount of specific libraries that are customized for certain parts of our industries, industries focused in terms of whether that be gaming, whether that be for automotive as well as the broad area in terms of what we're seeing in terms of neural nets and deep learning and focusing in terms of AI. Our focus in terms of CUDA is optimizing many of the different AI frameworks that are out there. There are AI frameworks that are evolving every single day, evolving in the standpoint of what is popular today, changes and revises in as little as overall three months. Anything from TensorFlow, Caffe2, PaddlePaddle as well as some of the ones from Microsoft and others are extremely important frameworks. And what CUDA is capable of doing is continuing to optimize for the fluidness from the GPU to the overall application. We're essentially responsible for stitching that all together, stitching that completely, completely end to end. We have a new and important role in terms of enabling many of these overall workloads in terms of their effectiveness in terms of the applications. And CUDA is that underlying point that allows them to. Whether it be available in the framework or whether it be a new area to move to, CUDA is there for additional coding, additional changes, additional algorithms that you may need to make your overall application work.

Q - Unidentified Participant

And that starts to talk to the next question that I had. And it may be redundant. But I want to cover it again because this is the biggest question we get on you guys and your competitive position. So the other concern we hear is that since the deep learning frameworks are where most of the parallel processing or the AI, deep neural networks are coded at, CUDA and NVIDIA sits underneath those. Is it -- so the concern is that you could, therefore, be easily replaced because you're not directly seen by somebody working in Caffe or TensorFlow. I think you -- you were answering the question. But I would like to go through it just simply one more time. Why is that? What is that? How do you manage that risk of being not at the deep learning framework level but one layer underneath that?

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

Yes. I mean, there's this thought that says I'm a developer and I'm working in my framework. I may be not familiar with line 100 of CUDA. What does it do? I think that's the great aspect or the great point of it, that it is seamless in some cases to the overall developer. But seamless doesn't mean that it can, therefore, be removed. It's an important stage in terms of the compiling of the overall code together, the overall middleware and so many different other functions that it does to optimize many of these frameworks rather than it just being a additional layer. The overall ability to find what you need in terms of the AI framework may be there. But often, many of them in terms of the new capabilities are jumping into CUDA often to develop that next layer of programming and so many of it that is seen. Our teams of engineers focusing on CUDA, focusing on the frameworks, are hearing from the AI frameworks all the time that says, can you build, can you help us in terms of incorporate this within CUDA. It is surely a factor, no matter what the framework is, that they know that CUDA is behind it. And we have a role in completing that full stack. And we'll always be there.

Q - Unidentified Participant

Okay, that's helpful. So we're 25 minutes into the conversation. I'm speaking with the CFO. We haven't talked about any accounting questions or financial statement questions. So your product gross margins have expanded by 900 basis points over the last two years. What are the biggest drivers?

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

So our overall driver of our overall gross margin and growth is really about our platform and what that means. It is not about being an important chip. It's about the full platform. When we talk about the full platform, there's a tremendous amount of value that we incorporate in everything that we sell from the overall software capabilities to the overall ecosystem and working with software developers to build on top of our platform. All of that work that you see us doing is essentially not in gross margin; essentially, it's in OpEx. It's essentially with our overall engineers every single day. That allows us in terms of the gross margins that we see and to leverage that work that we see in engineering to advance our overall gross margins long term. We've seen growth in terms of our margins in gaming. We've seen growth in terms of our margins in terms of overall data center being a larger position of our overall company. And that has risen our overall gross margins as a whole. Those are some of the main overall drivers that our customers see the value of our platform in terms of both saving them time, saving them money and the overall consistency of the platform serving their overall needs.

Q - Unidentified Participant

And on that topic of adding value, when you look at the growth of the data center business or the gaming business, to what extent is that capturing more with a higher ASP or versus a unit growth over time?

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

Yes. So when you think about whether it be -- we can start with our overall gaming business, there has been the drivers of both. Just an increase in the overall market growth in terms of gaming has influenced more and more gamers to come onboard. They essentially come onboard 2 by 2 because you bring your friends because it's such a social overall sport. At the same time, we have seen an average overall price point increase. Now that's different from saying that our prices have increased because if you were familiar with our overall stack for gaming, we've got a gaming card that you can have a great experience at $100. You can even have a outstanding experience at $1,000. So we have a full stack of different products for the overall market. And what we've seen is people continue to gravitate, whether they're coming back for a refresh or coming in for the first time and buying some of our higher-end cards. That has moved the overall average ASP higher. But we tend to consistently always provide a price point so anybody coming into the market can appreciate these overall products. Additionally, when you think about our market in terms of data center, we have many different types of price points for the different types of workloads that you want. We have those that are specific for inference. We have those that are specific in terms of training. And as you know, we have different price points in terms of our DGX, which are full end-to-end systems. So we have both

a combination for almost each and every one of our businesses of a growth in terms of units as well as a growth in terms of average ASPs.

Q - Unidentified Participant

Okay, fair enough. The -- we recently published an in-depth study of free cash flow and capital return. And if I look at NVIDIA over the past four years, your sales are up 150%. Your free cash flow has grown 400%. But your dividends and stock buybacks have only grown 18%. And you have a net cash position of $5.3 billion. So what's the plan for the $5.3 billion? And under what circumstances would you consider to materially increase the capital return?

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

Yes. So when we think about our overall free cash that we have in terms of overall available, there is a part of that, that you have seen in terms of our investment back into the business as that's probably been our most focused area to be at, whether that be hiring great engineers, whether that be a focus in terms of the day-to-day things that they need to work on. But we are also not much different in terms of what we're seeing in terms of our major customers. You've seen a massive growth in terms of the overall need for capital expenditures, focused a lot of it in terms of on the data center. That's going to be a key area of our focus as well. You also (inaudible) see us focused in terms of with the overall growth that we have in terms of the engineers, focused on these large opportunities in front of us. We have to think about the overall facilities and the houses that we need to overall house our overall engineers to work. That's a good use of our overall free cash flow, may not be in our current P&L. But in terms of long terms, in terms of those types of capital expenditures, in terms of what we've done. We'll also look in terms of possibly small; and medium-sized M&A types of transactions where we find great teams around the world that can bolt on to some of the great markets that we're focusing on. The verticalization in some cases that we have, that expertise is essential to us. But it's not that we see extremely large companies out there in terms of consolidation because there's not exactly large AI companies at this point on that piece. Then comes our focus in terms of our capital return program. It has always been a long-term existence. We'll continue to use that going forward. We continue to make existing growth in terms of our dividends every single year and a focus in terms of repurchasing our stock. So nothing's overall changed in terms of there. And we'll continue to see that growth.

Q - Unidentified Participant

Fair enough. And I think we're -- we've run out of time. So that will have to be the last word. Colette, thanks for coming to London to chat with us.

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

Okay. Great. Thank you.