Piper Sandler 2021 Global Technology Virtual Conference

Company Participants

- Manuvir Das, Vice President of Enterprise Computing

Other Participants

- Harsh Kumar, Analyst, Piper Sandler

Presentation

Harsh Kumar {BIO 3235392 <GO>}

Thanks everybody for joining us for a very exciting session that's coming up now. We are very fortunate to join Manuvir Das who is the Vice President of Enterprise Computing at NVIDIA. NVIDIA is of course the largest -- single largest market cap company doing some extremely exciting things of course through all of its businesses, but I think the most exciting thing, no one will argue with this or happening with what they're doing in the data center where Manuvir is deep into it.

So with that, I'm going to turn it over to Manuvir, he has got a short slide deck that he wants to talk about, and Manuvir, the floor is yours.

Manuvir Das

Thank you so much Harsh for having me and for giving NVIDIA this opportunity to talk to the audience. It's a real privilege. I thought when I do at the outset, is just share with you the big picture view of what NVIDIA is doing and where we're headed in the data center and with artificial intelligence before we do some Q&A here. So I'll start with statement about what we are sharing in the slides as we always do. So the first picture I have here is something we've shared before when we announced a new software product from NVIDIA called NVIDIA AI Enterprise.

And I thought I would start with this to just level set. This has been news we've shared prior, and why we did this, right, so if you think about the state of the union for artificial intelligence in the enterprise or enterprise customers at large, we are at a state today where we've had a lot of success with early adopters, you know, there's are a few thousand companies across the world that have had great success improving their business, improving the experience of their customers with Artificial Intelligence. But the broad base of enterprise customers is yet to adopt AI, right. And what is the fundamental reason for this? The fundamental reason is, there are two very different sets of people within every enterprise company. On the one hand, you have the data scientists, these are the people who understand AI, who understand the tools, Jupyter Notebooks, all these kinds of things, they need the development

of new AI capabilities. And they move fast and they're pretty agile and they are the state-of-the-art cutting edge, do new things every night.

On the other hand, you've got IT administrators who are accountable and responsible for making sure that the actual applications running in the enterprise data center are safe, secure, stable because the business of the company depends on it, right. And the experience of the customer depends on it. And these two personas -- sort of these two worlds are pretty much apart because the one world of the data scientist wants to use the tools and frameworks that they're comfortable with, whereas the IT administrator is used to a different model for how to deploy applications, and there is a disconnect because IT does not know how the pick up what the data scientist would use, and the data scientist don't know how to operate in (inaudible).

And so we created NVIDIA AI Enterprise to address this gap and what we did is we took NVIDIA's AI software for training for inference for Data Science and we've made it work on top of VMware vSphere, which is sort of the de facto platform in the data science. If you look at any enterprise datacenter today, you will find virtualized servers rather than VMware vSphere.

And so that's what this picture shows, right, and it achieves two things at the same time. On the one hand for the data scientists, they see all the tools and frameworks that they are comfortable and experienced with to do their work, that's been there in green provided by NVIDIA. On the other hand, for the IT administrator, it's the same VMware vSphere environment they used with the same tooling hardware, provision hardware to access to people, but now with these new workloads for AI. And so this is really a way of bringing these two worlds together.

So this is what we've announced earlier this year in conjunction with VMware, which is NVIDIA AI Enterprise, really NVIDIA's way of becoming mainstream for enterprise customers for making AI a mainstream workload for enterprise customers. Now, this is just actually beginning. And so what I really want to share with you today was that this is something NVIDIA has been thinking about and working on for many years, right. And what we realized is this is mainstream artificial intelligence in enterprise data centers is a full stop problem. Of course, you need the right hardware, that is the layer I have shown in green, but then you also need all of these pieces of software, sort of the operating system of AI, all the essential tools so that you can run your different AI virtuals.

And then finally, if you think about it, there's just different use cases where there's vision AI detecting interesting things that are going on in video feeds or cyber security finding attacks that are happening in a data center and so you would love to have pieces of software that are customized frameworks for each of these use cases that are easy to adopt. And so I've drawn this abstract picture for you that is representative. You can think of it as a brick wall rather than if you really want to solve the artificial intelligence problem and when you need to fully construct this brick wall of all these different baskets to get a complete solution. And in NVIDIA, that is exactly what we've done.

This is the same picture that I have replaced every one of those abstract concepts with an NVIDIA product at the bottom hardware products. But in the middle in the top, all software products that NVIDIA has produced over the last few years and especially over the last year to really complete this brick wall. This is not a vision slide, this is an execution slide. All of these things I'm showing you on the slide today already exist, are already usable by customers.

The fact of the matter is that today NVIDIA is much more a software company than a hardware company. We have thousands of software engineers within NVIDIA who work on this -- on all of these things every day. And so we built this entire stack a set of frameworks for these different use cases, the essential software that allows all of this to run on mainstream servers, as I said, in conjunction with folks like VMware et cetera, all of the hardware. And then what we announced recently was, we have a partnership with Equinix to put all of this technology, the hardware and the software into Equinix data centers around the world, so that for customers as they get going, it's very easy for them to start their journey, where NVIDIA has pre-deployed all of those things for that and then as they proceed in their journey of course they can procure and deploy these things for themselves in their own data centers or in a colocation facility.

Before I come back to you, Harsh, the final point I wanted to make was that NVIDIA is pretty fast moving company, right. This is our general philosophy. And so I did this exercise for myself or if I'm going to show you this picture, the same picture from last year, but only show you the things that in the execution mode that will actually produce, what would this slide actually look like? And this is what it would look like, whereas today it looks like this, right.

And so I just wanted to end by making this point that NVIDIA is an R&D first, innovation first company. The business results we have today are based on the work we've done in the last few years and what our teams are working on every day today, all of these software stacks that we have been producing and are putting out are to unlock the opportunity in the years ahead. And that's what we're really focused on as a company.

So Harsh, that's what I had as a bit of an opening context adding statement, if you will. Artificial intelligence in the enterprise data center is a full stack problem, it's an end-to-end problem, it requires a broad ecosystem. This is where NVIDIA's focused, we put the hardware, we've put the software, we've created ecosystem. We have more than 2.5 million developers who use different parts of our stack to develop their own applications and solutions, and that's our contribution to make AI feasible for enterprise customers. And with that, we have a go-to-market motion that is in conjunction with established partners, the OEMs produce servers, folks like VMware who produce software stacks for the data center and we're really looking forward to this journey of marketizing AI for enterprise customers over the next few years.

And so with that, I will stop sharing my slide deck and hand it back to you, Harsh.

Questions And Answers

Q - Harsh Kumar {BIO 3235392 <GO>}

Manuvir, that is simply incredible to see the number of products you guys have introduced in just the last 12 months to be able to fill up the gaps of where you were and where you are trying to go. And I'll bring this to an interesting topic, there has been a lot of changes here, not just with COVID, but just generally data centers always morphing, always changing, can you talk about how it's changing, and what are the large changes that are happening in the industry that sort of you wake up and think about and say, this is the kind of direction that NVIDIA maybe needs to think about going into?

A - Manuvir Das

Yeah, that's a great question, Harsh. And it's amazing how much the landscape of data centers has changed in the last decade. You know you hear some of these buzzwords these days like cloud, Kubernetes, containers, all these things. What's the common thread to all of that, right? The common thread to all of it is that for quite some time, computing in the data center was done in a scale-up route. You take one server, you run your application on it and as the applications get more demanding, you make your server bigger and bigger and more capable, right. And then you buy a few of these servers and they're super expensive. And then what happened with the advent of the public cloud was the proliferation of a different model, which is scale-out rather than scale-up. Instead of having one giant server, let me have many small servers that cooperate to run a workload, right. This is what in computer science for decades has been referred to as distributed computing that the public cloud moved clear. And Kubernetes and containers are just a mechanism for building your application as a distributed computing original, right. And this is how the data centers have really evolved in the last decade.

So what does this mean? This means now that when you run an application, instead of running on one server, you're running on a set of servers that are working in conjunction to run your working. So when you think about computing now, you're not just thinking about building the best server. You have to think about networking. Because the data is flowing across all the servers, you have to think about security, because if a malicious thing intercepts one server, they have access to all the other servers. You have to think about how you store your data so it's accessible to all the servers, right.

So computing is really evolving to data center scale. Every workload runs within a complete data center rather than a single server and so because of that you have to solve this as a full-stack problem, you have to think about what's the right servers, what is the right networking gear, what is the right networking software so that it goes fast, what is the software stack for orchestrating the workload and running the workload? You have to put it all together, right. And this is how NVIDIA has really evolved that we have become a full-stack company for that reason.

Now, the other thing I would say, Harsh, as you know, our genesis at NVIDIA was as a hardware company, right, with the GPU. So the other insight that we had in NVIDIA was that in order to make this full stack go, you're going to need three essential components in every server. Of course, you need a CPU which is what applications have traditionally run on. You need a GPU which is the way of accelerating the workloads so you can do more in every server. And then you need this new form factor that we call the DPU, a data processing unit, which sits on the network interface and really runs, not the workload, but the infrastructure of the data center itself, okay. So every server needs a CPU, a GPU and a DPU in conjunction. This is our vision of the data center. And this is why we of course have GPUs, we have the BlueField DPU from NVIDIA, we also recently announced that we are working on the CPU optimized for artificial intelligence called the Grace CPU based on ARM technology. And we really see this as the future direction of the data center, every server will have a CPU, a GPU, and a DPU inside, right.

So just to summarize all that, I would say, because I know I said a lot there, Harsh. We really think that computing going forward in the data center becomes a data center (Technical Difficulty) rather a full stack problem. We believe every server needs to have a CPU, GPU, and DPU inside of it as essential hardware components and then you need the right layers of software that I showed on my slides to bring it all together within the data center.

Q - Harsh Kumar {BIO 3235392 <GO>}

Amazing, Manuvir, it seems like the opportunity is getting bigger and bigger as the data center compute sort of gets distributed and flattens out, if you will. So you guys, I'm sure talk to a lot of customers and I'm sure the highest end customers actually come to you with their problem and say this is kind of what we need solved. What are you seeing in terms of what's actually strategically important to the customers and what areas are these customers emphasizing versus de-emphasizing particularly as a result of, for example, COVID-19 that we're caught up in right now?

A - Manuvir Das

Right. I think you mentioned the pandemic, and that has two profound impacts, Harsh, that we've simply seen from targeting us out and the two sides of the same coin, which is namely that the amount of impulsive connection has gone down, right. A one-side of the coin means for the companies doing their own work and their own business across employees et cetera, the employees are not able to the group together anywhere, right. So the question is how can the company remain as productive as before, even though their employees are all in different places and working from home, right. That's one consideration.

The second consideration is the company's engagement with their customer base, has also not changed because of the pandemic, right. It's become much more online and digital even more so than before. And so with that change in how they're interacting with customers, what should they do, right. So let me just take a minute to break down each of these, right.

So if I take the first one, which is that employees are not all sitting in the same room together. Instead, our approach at NVIDIA with our customers is, instead of looking at this as a loss, this is actually a forcing function for a new opportunity for them, that there are actually our technology can allow companies to be far more productive by leveraging people all over the world, rather than just leveraging people in the room. And this is why we created a platform called NVIDIA Omniverse, which we now make available to enterprise customers and the way to think about NVIDIA Omniverse is, it is a digital real time remote collaboration environment for people working on the same project. It could be engineers designing a building together, it could be designers creating the facade of a display somewhere together, and with Omniverse, all of these people can essentially log into the same place, they can collaborate in real-time, one person makes a change, another person can speak with him, right, so it creates a whole new model for collaboration and working together, right. And this is why we put so much emphasis on Omniverse. So it's a big, big initiative by NVIDIA and of course, there is a bit of a bias here because for such a modern world, well, one of the core technologies you need is really good graphics and that's something NVIDIA knows (inaudible) to about, right. So it was a very natural, but it's also distributing computing problem, it's also scale, it's a data center scale problem because you're running this giant sort of thing that different people can connect to and workout, right.

So that's one change Harsh and so that is within the company's work, that's why we did Omniverse, and of course, there are the technologies for remote work like VDI that NVIDIA has been working on for quite some time with our GPU technology and we continue to do that, right.

And we see a lot of adoption for example of workstations now because if you think about, if you are an employee working from home rather, you need a proper workstation in your home, if you're going to do all your work from home, right, and you change (inaudible), right. So that's the one side. The other side is, the company's engagement with that customer base, which is now much more digital and online than it was even two years ago, right. And so when you see -- look at how we are doing this conference right now, right, we are on a video conferencing technology, Visa proliferated, right. But you see things like the need to converse with your customers, what is called conversational AI, so many more customer conversations. You don't have enough humans in your company to do all these conversations, So you need some automation, you need AI, you need a chatbot that can interact with your customers and your website, right, so you can handle more requests and more queries.

So, on that side of the coin, we've seen a lot more companies now interested in adopting AI because they see it as a way to greatly enhance their communication with their customer base in this new era, where the customer is relatively disconnected from them physically, right. So those are the two things on that.

Q - Harsh Kumar {BIO 3235392 <GO>}

What about any -- I mentioned that maybe something that's become important versus something that's become less important to the customers? Can you talk

about if you have any example I would appreciate it on first something that the customers are not as focused on today as they were maybe before?

A - Manuvir Das

Yeah, I know you'd love an answer to that one, Harsh, but I'm going to pass on that because, you know, for better or worse, I happen to be in a position where when I talk to customers, it's mostly about the things they want to do now -- and so that's where usually focused.

Q - Harsh Kumar {BIO 3235392 <GO>}

Yeah. Okay, that's fair. Let's talk about half of that first question as well, big companies, that one complex things done come to NVIDIA and I suspect that you're probably more of a partner today, and increasingly in the future than you were before because they are actually coming to you saying we need this X, Y, Z and help us with that and you're sort of involved earlier on. Is that -- am I correct in thinking about it correctly? And then, is it really happening? And are you seeing -- enhanced interaction with your customers on a daily basis with the requirements that they want fulfill?

A - Manuvir Das

I think it's a great observation, Harsh, that you've had. It is true. I will put it to you this way, right, AI actually is AI, is hard for any customer to implement. That's the truth of it, right. And we sort of went through this phase where we were just proving it out the technology was complex and we were working with a small number of customers who really need it. Yeah, like for example, of an online shopping site that they need to recommend to customers what they should buy next. And I know that AI will help me and as painful or difficult it might be, I'll just jump in and do it. And so those are the people we work with earlier, right. But that has now evolved and as we've sort of broadened our reach across the enterprise customer, they are not looking for point pieces of technology that deploy the software or put in this piece of hardware, they want a solution, right. They want to solve the business problem and so more and more we find our conversations to be of that nature that hey, this is my use case.

This is a problem I can't solve, tell me what is the recipe. What hardware do I need? What software do I need? What ISV application budget do I need to work with? What datasets do I need to acquire in order to do my training? It's a complete discussion. We believe in this so much, Harsh, that if you look inside NVIDIA, we have a very large organization, a dedicated organization of what we call solution architects and these are people, they are not sellers, they are not sales reps, they are not product engineers, they sit in the middle and what they do is, every customer conversation begins with what's the problem you're trying to solve, here is our SAs, they will sit down with you as almost consultants, right, and as partners and we design the solution with you, and as we design the solution with you, maybe you use our technology, maybe you won't use our technology, that's fine. Either way, if you adopt AI, as NVIDIA, we're super excited partner, right, and we do think we have good piece of technology even for folks like myself Harsh like when we go and have conversations at the executive level with the customer, right, I never have a

conversation as (inaudible), I never have a conversation about (inaudible) product I want to sell you, right, my conversation is what's the problem you're trying to solve, how your architect think to solve, how we think you might want to architect your infrastructure data centers to solve this problem. And if we align with that, maybe we can be of help you with some parts of that architecture, right, that's going to hold.

Q - Harsh Kumar {BIO 3235392 <GO>}

It's an amazing way to think about customer interaction because the customer in this situation will more than like you to feel you there to help them with their issues than just a vendor trying to sell a product like you put it, you put your best.

A - Manuvir Das

Yeah (inaudible) Harsh, I think if you don't mind, I might get in trouble for saying this, but my boss Jensen, who is the CEO of NVIDIA, you know, I would say the one word he uses the most in meetings with folks like myself that NVIDIA is empathy, you know, that's sort of our most important word in dictionary and that is it starts with that, have some empathy for the customer, understand the situation they're in, what problem they're trying to solve, what opportunity they are trying to take advantage of and then make it happen.

Q - Harsh Kumar {BIO 3235392 <GO>}

And you know what budget allows NVIDIA to connect at a completely different level versus the rest of the vendors. And let's move on to software. You mentioned software earlier on. So NVIDIA I've noticed more over the last three years is bringing increasingly more and more amount of software to the marketplace, specifically when it relates to AI which is I would say a core competency of NVIDIA. Can you talk about NVIDIA's AI software, what is the differentiating factor here, where are we in the adoption curve and like if I dream the dream, it's a long question, but if I dream the dream, what is the opportunity for NVIDIA here?

A - Manuvir Das

Yeah, so I'll do this in reverse order with the punch line. I think we believe that if we execute well on our plans, there is at least a multi-billion dollar incremental software opportunity here on top of what we are already doing that because today our revenue in enterprise AI is primarily based on the hardware that people buy, even the GPUs and the networking gear et cetera. But if you think about it, a simple way to think about it is if you look inside an enterprise data center, there are certain waves of software, for example, of VMware or SFE et cetera that are deployed across servers and that's a commercial model for that software and the reason is because that software solves a very important problem for the customer, which is how do I run my workloads? And the software is almost more important than the hardware because the software is what the customers' experience set and the customer has an expectation of the software is supported, it has a certain level of quality and performance. It is updated regularly, those sorts of things. And that's why it is a commercial model for the software, and we are now entering that world for the first time we can be, right. Today, we have the new software, but it's been made available into the community to other people to make their lives easier, but now for the first

time with NVIDIA AI Enterprise, we really have a similar kind of product that can be sold because that thereby our customer can allow it, right.

And there's a simple math you can do about how many servers there are in the world, how many servers we expect would be used for AI, what's our licensing you could do for the software in every server that would be fair to the customers, then we know you multiply those things out and it's at least multiple billions of dollars of incremental revenue for that layer of the software, right. So that -- so I'd stop with that. Now to your first question, as part of what are we (inaudible), right. The truth of the matter is that only need right we have this year when we have brought up the software, in fact, NVIDIA AI Enterprise went in general availability just last month, right. And it's just beginning to roll out now, there's a new version of VMware that supports that which has been rolling out. So we're in the beginning of this journey. But we certainly expect that there will be broad adoption. And you can think of that adoption on two fronts, Harsh. One is the software itself being adopted. But the other thing is what the software is really doing is it's making it possible that you can take your regular mainstream servers you have in your data center today that you would normally not think of using for AI, but now you can use them for AI.

So it also -- so it expands the balloon in two different ways. One way is there's this new thing called the software that is a commercial proposition. But the second is that the software brings a lot more servers into the picture to be used for AI and so it expands the reach, right. So that's going to what I would say, we see it as a big opportunity, we're early in the adoption. So we are in the steep part of S curve, if you will.

And then just one thing I might say about the piece parts themselves, the simplest way I would describe this is when you adopt AI, you need to do two things. Number one, with the data scientists and other people, you need to develop AI and then once you've developed it, you've got these great models, then you need to deploy AI within application, so that you can actually use the AI for example to see what's going on in your (inaudible). Okay. And so essentially we introduced two platforms. We have something called NVIDIA Base Command, which is one of the enterprise users to develop AI and we've got a platform called NVIDIA Fleet Command which you use to then deploy your AI out to all the places where you need to deploy it out, right. So that's the highest level of simplest way of thinking about our platform, there's Base Command and there's Fleet Command and we're very excited about these. But as I said, we are very much in the early stage.

Q - Harsh Kumar {BIO 3235392 <GO>}

Manuvir, just one more thing on that, do you think there's anybody in the space that's even close to the level of work you guys are doing, I know historically NVIDIA has been just a pioneer in AI on the hardware side. And now I see the focus on the software side. Is there any one even in the zip code of where NVIDIA operates in bringing the complete package together?

A - Manuvir Das

Yeah, we, of course, I'm biased here. But I would say we do not think so, right. And -- but I will elaborate on that, right. If you think about the picture shown in my slide, the point we made was this is really a full-stack problem from the piece parts of the hardware, to the systems, to the low level of software to the (inaudible) on top, we're the only company on the planet that has been working on all of these leads, and as I said, my slide was not a vision side, my slide was a reality slide of the things we have been.

Now we operate at all different levels and we are big believers in ecosystem, whether it's cloud service providers or server manufacturers and whatever, right. So our model is, we are happy to partner with anybody at anytime. For example, you must be a company that focuses on building frameworks of the top level, but then we have APIs so you can use the minimally of our software as the basis for developing effect, you might be assistant manufacturer like a Dell or an HP, you can incorporate our GPUs and our DPUs into your servers, right.

So there is certainly companies at every -- in fact, we foster that ecosystem very intentionally, but we believe that NVIDIA is the only company in the planet, Harsh, that has focused on the entire stack. And that's why we have been (inaudible) optimize DataLake services cases.

Q - Harsh Kumar {BIO 3235392 <GO>}

Absolutely, no question about it, you guys have been there at the forefront with compute and with AI for a very long time already. You brought up something Manuvir earlier on that was fascinating, Omniverse. How does Omniverse fit into your software strategy, you talked in terms of collaboration, but obviously there's got to be a longer game plan I would think if NVIDIA is putting so much upfront into it. What is the opportunity for adoption for this in the next couple of years and so then maybe I'll hit you with that first and then go from there.

A - Manuvir Das

Yeah. And I'll do this one backwards to with the punch line first (inaudible) basically when we looked at the target audience for Omniverse in the work that we've done, we think there is about 20 million designers and engineers out there. for whom Omniverse would be a great platform for them to do their day-to-day work. And if you just do some simple math of a subscription-based model that we've already put out and then fixed some norms and standards of the industry, if you will, this is again definitely a multi-billion dollar net incremental market opportunity from the use of Omniverse, right.

So that's one way to answer the question. The other way of answering is, as you pointed out, I talked about collaboration and that's certainly a use case of remote collaboration. But do we see a bigger opportunity, right? The bigger opportunity we see Harsh is that one way actually of time together, everything that NVIDIA has done from its inception as a company, whether it's graphics or AI or robotics or self-driving cars or any things, in fact, fundamentally, we had a simulation. Okay. We've built technologies in different domains that allow you to simulate something without actually having to do it. That's the core of our technology. Like for example, think

about our platform for self driving cars. Yes, you can drive cars around and you can capture what's happening on roads and make your car better, of course, we do that, but we also have a complete simulation platform that you can use to do miles and miles of driving "without actually driving" right, so you can learn a lot more.

So we really believe that going forward, no matter what industry you're in, as the world evolves, simulation will become more and more routine as the basis for how you're productive and dealing with Omniverse is it's you know and has dramatically changed the state of the art in terms of being a platform for stimulation -- for real-time simulation, so you can actually model things and see what's happening, right, and we think that is a massive opportunity that goes beyond just the ways of collaboration.

Q - Harsh Kumar {BIO 3235392 <GO>}

No. Thank you. Thank you for that. In the most recent earnings call, I think Jensen focused a lot on software for one reason or the other. And then in your presentation, you're talking a lot about software. So we see the change happening. My question is about, when do you think and if you're carrying -- how far out are we before you start generating meaningful amount of opportunity in revenues from the software stack that then it really is bringing to the table.

A - Manuvir Das

Yes, I think I'll answer that for you. I'll apologize and answer that for you in a relatively generic way instead of giving its number, right. See, this is definitely a journey that we have the beginning of, we are at the steep part of the curve, we are seeing massive interest. So we know we're heading in the right direction, but certainly right now, our revenues primarily driven by the things we have been working on over many years, right, and these things will begin to pay off as we go focused.

But as I said, you know what I quoted to you for both NVIDIA Enterprise as well as for Omniverse Enterprise of being multi-billion dollar opportunities, we see these as very real opportunities, right. I would also assume that -- I also want to paint the picture accurately for the (inaudible). There's in fact a next level of software opportunity for NVIDIA that is in some ways mobile phones and what I've discussed, right.

So what I'm talking about here is sort of the essential software you know for artificial intelligence, or for collaboration and simulation on Omniverse. But if you think of the real AI journey, what is the real AI journey about? It's about saying that in every walk of life, no matter what industry your company is in, there are certain functions that humans are performing, right. And those -- each of those functions one by one, if we can figure out a way to automate that function with AI, then you can do it much more cost effectively and you could free up your humans to focus on other things. A good example of things that you can use NVIDIA's software frameworks to look at X-rays and detect whether some -- whether there is a fracture in the (inaudible) right,

That's something that today a radiologist has to do. But you can take that function (inaudible) right. In the space of retail, you can look at the camera feeds from across

the store and determine who is shopping for what and what they're walking out of the store with, right, instead of having roomers in the backroom having to sit there and look at the videos (inaudible) right. So one by one, you can take each of these human functions and replace them with some NVIDIA software.

So now the question you asked is, what is the potential business value of the software? The business value of the software is not a function of how much did it cost NVIDIA in terms of engineers to develop the software. The business value is in terms of how valuable is it to that enterprise customers to replace that human function or augment that human function with this automated software, right, and so we see a rich landscape of business opportunity from the software there that we are yet to unlock, right. And that's a whole on the domain of opportunity.

Q - Harsh Kumar {BIO 3235392 <GO>}

So I wanted to shift gears a little bit Manuvir, we're at the last kind of seven, eight minutes. And I wanted to hit upon this, so off late since maybe the acquisition of Mellanox, we hear NVIDIA talk a lot about Smart NICs and DPUs and I guess connectivity is a core theme now, you touched upon it with the distributed compute methodology, can you update us on how you feel, A, about the importance of things like Smart NICs and DPUs, maybe what's the difference between the two, and then where you are in the roadmap as a company on these two particular connectivity products?

A - Manuvir Das

Yes. Let me do that. So firstly, starting with what's -- just that somebody gets these Smart NICs and DPUs because there is a number of people -- number of companies out there that work on Smart NICs, right. So I think the best way to think about it is, Smart NIC is sort of step one which is to say I've got a network gateway installed, the data is flowing through there, if I put a little bit of computing power maybe some ARM, ARM CPU core over there, there is some more processing I can do on the data as it's flowing through the network. Now we took this to the next level and clearly this concept of the DPU. Our DPU product family is called BlueField and the idea of the DPU it has so much horsepower in that processor that what it actually does is it takes over the functions of the data center itself. So you've heard a lot of the last decade about software-defined (inaudible). What does that really mean? What that means is that all these things you're dealing in data center firewalls and things for which you have this dedicated hardware, we're now turning to software that was running on the server itself, but as this happened, more and more this load went to the server, which means that there was less and less place with the applications themselves to actually run.

So whereas, you would have needed five servers to run an application, you now need 10 because half the servers be consumed on this job and what our DPU really does is, offload all of that work onto this other processor, move it there, you free up the CPU in the main server to run your workload and the way we fit the DPU, it actually accelerates, it's like the GPU, if you take the firewall (inaudible) and you move it from the CPU to the DPU, it's not just shifting the problem, it runs 100 times faster. And so you need much less silicon in the DPU to do the job than you would

have on the CPU, right, so that you actually save money in the data center, right. So this is why we're so higher than DPU because it can dramatically change the way data centers are architected. So our view is every server needs (inaudible).

Now two specific things we have done here Harsh that we think distinguish. The first thing is we'learnt the great lesson from when we did GPUs, we created in software SGK interface called CUDA which was a simple way for developers to interact with GPU. We said no matter what GPU you use, CUDA is CUDA, right. So it makes your work going. We've done the same thing here with DPUs, we've created an SDK called DOCA and it's a consistent SDK across our DPU family, and so again (inaudible) ecosystem is programmed to this API, this SDK and your work will translate as we make better and better DPUs and your software would just become better,

And the proof point of this, the second point I want to make is, we have a roadmap, we are already working on BlueField 3, the third generation, we've already announced the architecture pf Bluefield 4, right, and it's not just making the processor better, but we now are working on versions of that processor where it's actually the GPU capabilities inside the DPU as well, right, so you can do AI inside the network, right. So think about what that enables, right.

So that's how I summarize it Harsh. On the one hand, we have a rich hardware roadmap for how much more powerful DPUs are becoming, but we created an interface called DOCA that drives along. So for the ecosystem you just develop once and as the processor gets better, your software will just get better along the way.

Q - Harsh Kumar {BIO 3235392 <GO>}

Manuvir, it's amazing. You described it so well. I think maybe 15, 17 years ago, maybe even 20 years ago when the first NICs were coming out, I was trying to understand what they did. And the point was it takes away some of the complex functionality off of the CPU and does it for the CPU and it seems like the same thing is happening except the functions are getting more complex, they're software richer. But the basic functionality is the same but we're moving up the stack, which is great for companies like you and actually makes a data center simpler in some ways because like I said, it's more cost effective, I mean, and so anyways, it's fantastic stuff, so lot to think about there, lot to unpack.

Manuvir, as always, pleasure to have you. Thank you so much for your time. Thank you anybody that joined in and listened to this presentation. And we really appreciate your time. Thank you, Manuvir.

A - Manuvir Das

Thank you, Harsh. It was my pleasure. And on behalf of Jensen and the entire team at NVIDIA, really appreciate the opportunity to be with you today.

Q - Harsh Kumar {BIO 3235392 <GO>}

Thank you so much.Take care.