# Investor Day

## Company Participants

- Colette Kress, Chief Financial Officer
- Jay Puri, Executive Vice President, Worldwide Field Operations
- Jeff Fisher, Senior Vice President
- Jensen Huang, Founder and Chief Executive Officer
- Robert Csongor, Vice President and General Manager of Automotive
- Simona Jankowski, Investor Relations

## Other Participants

- Aaron Rakers, Analyst, Wells Fargo
- Ambrish Srivastava, Analyst, BMO Capital Markets
- Harlan Sur, Analyst, J.P. Morgan
- John Pitzer, Analyst, Credit Suisse
- Mark Lipacis, Analyst, Jefferies
- Matthew Ramsay, Analyst, Cowen and Company
- Mitchell Steves, Analyst, RBC Capital Markets
- Timothy Arcuri, Analyst, UBS
- Toshiya Hari, Analyst, Goldman Sachs

## Presentation

### Simona Jankowski {BIO 7131672 <GO>}

Okay. Well, good morning, everyone, and welcome to NVIDIA's Investor Day. I'm Simona Jankowski with Investor Relations. And it's my pleasure to welcome all of you here today, as well as all of those, who are joining us on the webcast.

Before we kick it off, I would like to read our Safe Harbor. We will make forward-looking statements in today's program regarding our expectations and other future events, which may differ materially from NVIDIA's actual results.

I'd like to refer you to our SEC filings for a description of our businesses and associated risks and other factors, which could cause the results to differ materially from these statements.

All our statements are made as of today, March 19th, 2019 based on information currently available to us. Except as required by law, we assume no obligation to

update any of these statements. Also, if we use any non-GAAP financial measures, you will find the reconciliations to GAAP on our IR website.

Okay. So with that, let me just go over very quickly the agenda for today. We are going to be starting off with a few minutes with Jensen Huang, our Founder and CEO, who I think you all know, talking about our strategy. We will then move over to our Gaming business, which will be covered by Jeff Fisher. Following that, we're going to talk about Datacenter with Jay Puri; Automotive with Rob Csongor; and finishing up with financials with Colette Kress, CFO.

We are going to have about an hour of Q&A. After all of that with Jensen and Colette, and then after that, we're going to have a lunch, which is going to be in the gold bar room if you walk out the door and down the hall to your left.

In terms of just a couple of closing items, if you need anything throughout the course of today, just reach out to myself or Shawn Simmons on the Investor Relations team. And you can find us in the back of the room or just email us. And again, I'll like to request that all of you silence your phones. And with that, it is my pleasure to now welcome to the stage, Jensen Huang.

## **Jensen Huang** {BIO 1782546 <GO>}

Thank you. This is going to be the fifth hour of my keynote. If you missed it yesterday, if you happened of missed it, you can watch it on YouTube, put it on 3x speed, because it will take about two hours if you did that Simona? Yeah, there we go.

First of all, welcome. It's great to see all of you. Here is what I'm going to do. I'm going to do a couple of things. I'm going to explain and this -- for most of you, you know this very well. There are some new faces in the room, so I thought I would do this.

I would explain what accelerated computing is. Accelerated computing is accelerating, but it's not an accelerator. And I wanted to find the difference for you. Okay. And so as soon as -- hey, guys. Fish. Hey, guys. I'm trying to give a talk -- close that door. It's an Analyst Meeting. When you guys come to NVIDIA's formal events, it always seems like home cooking, doesn't it?

Okay. So accelerated computing. Accelerated computing is particularly important today, because CPU scaling is no longer happening at the exponential rates that it used to. At a time when application workload demand on computing is growing incredibly fast, the question is how do we extend Moore's Law? How do we extend Moore's Law? Well, we came about this idea called accelerated computing, a decade and a half ago.

26 years ago when we first started the Company, we realized that accelerators could help us achieve performances otherwise impossible with a normal computer. Accelerators, and we built, we identified one particular accelerator that has what we

call at that time -- if you saw one of my presentations 26 years ago, it says -- bless [ph] you -- sustainable opportunity, sustainable opportunity meaning that this particular application called virtual reality trying to achieve virtual reality 3D graphics was going to take nearly forever. And the reason for that is, in order to create this environment, you have to simulate physics, life physics, particle physics, material physics, you have to simulate physics, and you have to do it so fast that for all practical purposes, it's going to take forever.

And the reason for that is, because at the time, simulators, supercomputers were doing it, simulating some fluid dynamic simulation or particle simulation was taking week on a supercomputer. What are the odds that we're going to be able to do it at a 120 frames per second, and to be able to simulate all the interactions with all of the agents performing artificial intelligence capabilities, all interacting together. The odds of that happening within the lifetime is approximately zero. We were not wrong.

We identified one problem statement that we said had sustainable opportunity. 10 years into it -- 10 years into it, we discovered that in fact, in order to continue to expand it, we have to expand the aperture, if you will, of the things we accelerated, no longer wasn't sufficient to just accelerate graphics. We had to first simulate the physics and then accelerate the graphics, because you have to simulate the water, you have to simulate the leaves blowing in the wind, you have to simulate things, particle physics as buildings crumbled. And so it wasn't possible to have animated all of that, we decided that you had to simulate that. So we expanded the aperture of our accelerator, and we invented this idea called CUDA. So that we could expand not just accelerating graphics, but the domain of virtual reality, the domain of virtual reality. That time when we transitioned from a graphics accelerator to a domain accelerator, we became an accelerated computing Company. An accelerator accelerates a function. An accelerated computing platform accelerates a domain of applications. Does that make sense?

An accelerator is a video accelerator H.264 accelerator. An audio codec is an accelerator. All of the stuff that runs on an audio codec with the exception of the analog can run in software. All of the things, all of the functions in a video decoder or encoder can run in software. And in fact, the first prototypes of a decoder is in software and the first prototypes of an encoder is in software.

So all of these functions -- computer functions can run in software, and it's possible to design an accelerator for that one function. You would use a video decoder to decode video, but you would not use a video decoder to compute molecular dynamics . You would use a video encoder to encode video H.264, H.265 or back in the good old days MPEG1 and MPEG2, but you would not use the video encoder to do, for example, a recurrent neural net for deep learning. And if you design a recurrent neural net deep learning accelerator, you wouldn't be able to use that for example for random forest, machine learning algorithms. If you design a functionality just for -- and accelerate it for one functionality, it would certainly be very good, but it doesn't have the necessary aperture to accelerate a large domain.

The challenge of course is if you created a product that has an aperture of infinite domains what you've done is you've created a CPU. The reason why accelerated computing is so wise, and the reason why although, many other parallel computing approaches have come before us, the reason why it has lasted the test of time is because it allowed the CPU to do what the CPU is good at and it accelerated the domain of applications that we are good at. And that disciplined of trying to figure out how to expand the aperture, while reducing the aperture at the same time, that strategic choice is ultimately the strategies you see at GTC.

There are several things you could do to test, whether something is an accelerator or an accelerated computing platform? Of course, the first thing is, it has to be a programmable architecture. On the one hand, one day you have to do molecular dynamics and the other day you do quantum chemistry and the other day you simulate a large climate science program called worth, another day you're reconstructing images out of electron microscopy called Cryo-EM, which won the Nobel Prize in physics, two years ago. It's hard to be able to do that, if it is only designed for one thing, it has to be programmable.

The second thing about all computing architecture is that it has to be an architecture, which means this, an application that you wrote for that computer runs on that computer, and on that computer. And you buy a new computer tomorrow and the application runs on it, a computing architecture has some capability of compatibility over time. And it has to have a large installed base. Otherwise applications can't find computers to run it on.

Accelerators don't have that problem, the other characteristic of a accelerated computing platform is that it has to have a rich software stack. It turns out the most important thing about our Company is our stack, that's what we talk about it all the time. If you look at this, this is our stack, our stack starts with the system architecture, I'm not showing the chip, I'm taking the chip for granted. The system architectures, the RTX is for graphics the DGX is for scale up, high performance computing, otherwise known as deep learning or supercomputing, hyperscale HGX and then AGX for autonomous computers. Little systems that are intended to live at the edge largely disconnected from the cloud, largely disconnected from the cloud.

We are currently -- because we're artificial, we're somewhat intelligent, where we can perform our jobs disconnected from the cloud. I am currently disconnected from the cloud, okay, I'm autonomous. And so that AGX is designed to be an autonomous machine. On top of that is our most important layer called CUDA. I call it CUDA here, but that layer is really complicated with a whole bunch of stuff, it's not worthwhile to go into, but it's basically if you will, our AWS, it's basically our Windows.

CUDA makes it possible for an application that runs on CUDA to run on all of these devices. Yesterday, I announced a $99 computer. A $99 full computer, it runs the same software stack as $1 million supercomputer as a coder of $1 million DGX deep learning system or a PC. There's only one computer architecture in the world aside from this that does that. And it's the x86. And so an accelerated computing architecture has a rich software stack. The other thing about accelerated computing

that's interesting is this, because of what I said earlier, there is only one computer architecture that can boil the ocean that's called the CPU. It's general purpose, that's it's nature, that's it's weakness too.

It's strength is that they can run everything, it's weakness is that it doesn't run anything super well. Now during a time when the performance is increasing by a factor of two every year and a half, it was plenty fast enough and the reason why was plenty fast enough is, because software developers take two or three or four years to complete each round of major innovation, meanwhile, the computer has already quadrupled in performance by the time that the next build comes along. It's fantastic to just ride that way to do nothing, and just let the way take you, that was the whole dynamic of Moore's law, it was fantastic well lasted.

But, now if that slows down, then all of a sudden, you can't solve new problems. If you can't solve new problems the software industry will suffocate, because they can't obviously introduce new ideas and that's why the world needs a path forward. We need a way to go forward. Now you're not going to find a way to go forward by coming up with another general-purpose computer, you have to find a way to go into it through domain acceleration, not a function accelerator, not an accelerator, but an accelerated computing architecture. So that we can take the industry forward.

Well, this computing -- this accelerated computing architecture must have vertical domains that are focused on, otherwise known as the counter of horizontal, vertical. And so we select vertical strategically, strategically and methodically, so that we can, one, make a contribution by the time that it's necessary, it's sufficiently large to be able to sustain the enormous investment that we put into it, but it's not so large, it's not so large, it's essentially a horizontal problem. For example, a web browser is so large, there is no such thing as a web browser accelerator. The only way to accelerate a web browser is to make every web browser faster. However, video games is a little bit of a unicorn. We identified the killer app 26 years ago using exactly to send methodology I just described, 26 years ago, we used the same methodology and we said, if we want to be one of the world's most important computer technology companies someday, what is the killer app that we can make a contribution to, that will take us all the way. And the killer app we found that we thought of, that we identified and focused on at the time was a $0 billion market, Electronic Arts was 14 people large. $0 billion market, that $0 billion market is called video games. It's a unicorn, because it has two characteristics simultaneously never happens, it never happens, you have a spreadsheet used by millions and millions of people, but the computation requirement is low.

You have a weather simulator, the computation requirement is enormous, but the volume requirement is very low. So in both cases, it's unable to justify an accelerated computing platform, there was this unicorn that's just stood out there. We imagined that if some day, there was a such a thing as the video game industry, it would both be large, because everybody would be gamers, who wouldn't want to play. And then two, the computation requirement of it would be gigantic.

The unicorn, 26 years ago, we found the unicorn well methods, that same method is being applied here. And you could see one segment after another, we're essentially finding verticals that are sufficiently large in domain, that could sustain more and more and more and more and more investment as we grow into one. High Performance Computing, Scientific Computing, a very important segment. Artificial Intelligence, we're going to talk plenty about it today. Drive autonomous vehicles, a very, very difficult computation problem, not just the computation problem for in the car, but computation problem before you get to the car, that gigantic computation problem I described a little bit yesterday, drive the whole platform of drive, the initiative of drive is about creating the autonomous vehicle future not about making a self-driving car. It's a little bit different. One of them is very large in scope, requires you to be a software-defined Company, it requires you to have an ecosystem, it requires you to have developers and tools. Isaac, I've described Isaac in the same way, the ultimate AI problem is both a wonderful opportunity when you solve it in the device at the edge, but getting there is a supercomputing problem. And I've shown a couple of examples yesterday, where you're essentially creating a virtual reality world, where the robot has to learn how to be a robot. That is a supercomputing problem. Clara named after Clara Barton, who started the American Red Cross is our platform for medical imaging, computational medical imaging, turning the instruments of medicine into a software-defined problem. Today, it's a bunch of instruments and widgets and things like that, today, tomorrow it's going to be largely software-defined algorithms are going to fly and they're going to be able to do things that otherwise impossible today.

And then lastly, Metropolis, the Metropolis name kind of gives it away, it's really about thinking about cities and places as one gigantic robot in the future. Our city in the future will have three characteristics, cities of the future, factories of the future, buildings of the future, we'll have three characteristics. The first characteristics is tons of sensors, the second characteristic a bunch of computation at the edge, basically, the reflexes of that robotics city, it doesn't have to go to a cognitive brain in the cloud, and then third connected to a cognitive brain in the cloud. Those three characteristics, so that they can make decisions and plan. Perception, reasoning, and planning, the three computations of an intelligent being, otherwise known as the computation loop of intelligence or robotics is going to be used for metropolis.

At this -- I think there are talks GTC between us and Microsoft, where Jetson Nano, our edge computing stack is connected to the Azure IoT stack, and some really, really exciting applications could be made possible. So this is the accelerated computing stack, very different than in accelerator. We focus on domains not functions. There's a couple of things that characterizes a company, who is a platform company. If you're a platform company, you talk about design wins less. If you're a chip company, a components company, you talk about design wins a lot.

When you are a platform company, you talk about your ecosystem a lot. And the reason for that is because you created the market or you're creating the market, and you need a lot of partners to work with you to realize the full potential of that market. You have ecosystem partners that work with you on your platform. And that platform is rich with software, and that software is domain focused not function focused. It doesn't do CNN, it does accelerated data science.

Okay. And so if you look at the comparatives, when you hear us talk, that's the reason why we talk this way, because we're a computing platform company.

I announced a couple of things yesterday. First, Fisher is going to talk more about this, but the big takeaway here is RTX is off to a great start. It is clear now that ray tracing is here, this week as Game Developers Conference and all they're talking about is ray tracing. It is clear that ray tracing is here. Remember this, ray tracing is software, and the reason why you can tell ray tracing is software is Turner Whitted, an video researcher, who invented ray tracing, iterative ray tracing recursive ray tracing did the first implementation on wax in software. We know it's software, because all the movies that are made is in software, it's called rendering software, and runs on CPU farms, otherwise known as rendering farms. What RTX does, is not do ray tracing. What RTX does is make ray tracing fast. So we love the fact that people do ray tracing. We just want to make it super-fast, and there's no question ray tracing is here and RTX is going to make it super-fast.

Number two, take -- the second thing that I showed you yesterday was the fact that graphics is going to be a new datacenter workload. This brings us so much joy as you could imagine. Graphics is going to be a new datacenter workload, you heard Matt Garman say it on stage yesterday, as well as he was talking about the need the demand on AWS, and one of the major applications is graphics, he had mentioned graphics several times. And so graphics in the cloud, we're super excited about that. It's going to be a new datacenter workload. And then yesterday we gave an update and Fish will talk more about this about our partnership with regional telcos, global telcos, and it's part of our GFN strategy, we call it the GFN alliance. He'll explain us, but very simply, they buy these servers from us. They buy these super optimized graphics servers from us called RTX. After they buy the servers, we host a service on top of it, because the GFN service belongs to us, we can host that service, and we share the revenues with them. We share the revenues with them, so they buy the servers from us, they share the revenues on the subscription fees on top. Does that makes sense? That's called the GeForce NOW Alliance. You could imagine the economics. It could be quite good.

I talked about datacenter, graphics in the datacenter. There several new workloads in the datacenter. Of course, graphics is -- we already talked about high-performance computing in the past. We already talked about deep learning in the past. We already talked about influence in the past. We are going to talk more about that today, but some of the new workloads that I talked about this week, graphics is one, and the second one, the second one is a gigantic one. This is the unicorn that we've been looking for in the datacenter.

Let me explain to you why. Remember, I explained earlier that there are two types of applications in computing, and that's why there's largely two architectures. And if you look it up, it says, there are capacity machines and capability machine. That's the way the supercomputer industry talks. The way the hyperscale datacenter people talk is they say they scale up machines and then scale out machines, scale out is hyperscale. You take a cost efficient computer, and you scale it out linearly, so that you could support a whole lot of jobs that are small at the same time, scale out. Scale

up says you build the largest computer you possibly can, whether it's the largest amount of computational capability, the amount of storage, the amount of active memory, the amount of networking, you build yourself the largest machine you can. So you could solve the largest problem as fast as possible for one person, whether simulation, climate science, these things take forever to simulate, that's called a capability machine, a capability machine, a capacity machine, a scale up machine, a scale out machine, a supercomputer, a hyperscale datacenter.

Are you guys following me? All three phrases are identical. Okay. All three are identical. Here's the Unicorn. It turns out of supercomputer, the market size for it is not very large. The computational challenge is great. And we're doing fantastic in supercomputers.

Here I showed you, I showed you a bunch of numbers -- people at NVIDIA call it CO math, and I just got to concede, it is not accurate. But it is absolutely right. Okay. It's not accurate. It's right. It's -- this is intuitive math and if you go double check it, you'll find that it's probably wrong in some area, but on the large-scale, it is perfectly right. All right. And so if you look at the numbers that each one of the numbers, each one of the ticks, if you will, is three orders of magnitude., Are you guys following me? This is three orders of magnitude here, this is extreme log, log, okay.

Now extreme log, log, says in supercomputing, I need 1 billion petaflops not in second in units. I need 1 billion petaflops in order to perform some of those simulations. In the case of concurrent users, CCUs, a hyperscale datacenter has to support hundreds of millions of people at the same time. In the case of supercomputers only, tens if not at the very, very most 100.

In the case of hyperscale, the amount of computation that it takes to perform these neural networks is small in total. It's just, there's a great deal of them, and the scale goes everywhere from hundreds of gigaflops to maybe hundreds of teraflops. In units not in time. Okay. And so what's interesting is this, data science, data science is that bubble in the middle. When we have more time, I'm happy to break it all down so that you guys get a feel for the numbers, but the important thing is this. On the upper left to the lower right, the upper left is three orders of magnitude more computation from the top to the bottom, and the lower right is three orders of magnitude in volume and the reason for that is this, data science is the only high-performance computing problem we know, where there is millions of people, millions of people, and millions of people in different fields of science, healthcare, financial services, they call them clients, insurance companies, retail, logistics, travel, you name it. Every single industry will benefit from data science, that's why there are so many people, the amount of computation you need, because the amount of data that you're working on is so gigantic. It's simultaneously a large computing problem. That's why the clients have the largest computers. Now imagine there are going to be millions of clients. And the reason for that is because there are so many industries where there's domain expertise and finally the technology is capable of being used at a large scale, the frameworks, the algorithms are sufficiently robust now and the schools are teaching it. You guys know the data science is being taught to every single field of science in a university now. From sociology to ocean

echocardiography to forestry to agriculture, it is the fourth pillar of the scientific method. This new pillar of the scientific method came about literally made possible in the last 10 years, came about in the last five years and about the same time the deep learning was happening, the same dynamics what's happening to data science. It is going to be a very large market and data science is a fourth pillar theoretical, experimental, computational and now data-driven science. Okay. This is quite a large market.

For the upper left and when doing this right upper left. On the upper left, our strategy is to create something simple for people to use. It's basically an appliance of the supercomputer, because most companies don't have the ability to build a supercomputer. It's too hard. Too much IT, too much system integration, too much software optimization, we containerized that if you will, turn it into an appliance. On the right hand side, it's a scale out problem. We have to turn basically a datacenter for an enterprise where hyperscale datacenter into a high-performance computer and there we have to break it all down in a different form factor, build different GPUs, write different software, work for different -- work with different partners and our go-to-market is different.

The go-to-market on the right side are the world's top enterprise makers -- enterprise computer makers. They're all signed up. They're also excited. On this left side, there are really deep super clients, super data scientists and there, the numbers are not in the millions, they're probably in the order of, call it, 50,000, okay, 100,000 but they need the best machine and we can reach them through experts in other specialized IT experts like storage companies, because it turns out. If you want to use one of those machines, you need a lot of storage anyhow, a lot of harmony there. And so I have something to do with our go-to-market and Jay will talk more about that. So this -- the second point is data science is a major new market.

Initial ecosystem. This is our ecosystem in one slide. Now, when I say ecosystem, I don't mean design wins. When I say ecosystems, these are all partners of ours who are taking the NVIDIA architecture to market. They're not changing it. They're not hiding underneath there's. They're taking it to market. They might integrate it with theirs.

Okay. So this is our platform and their platform coming together. Sometimes this is our platform going to market by itself, but these are ecosystem partners of ours and we are super happy that literally everybody in the larger IT industry is part of our ecosystem today. We announced two new types of computers. We announced a data science workstation and we announced a data science server, both of them software fully integrated, you buy them, you should be able to deploy them, really complicated set of software. However, it's already configured and optimized for you.

One of the things that you could see and then in the cloud, we announced a partnership with AWS and (inaudible) was very, very, very really appreciate them coming down and celebrating the moment with us and so we have workstations , servers and cloud to take this data science platform to the world. And then lastly one of the things that you might notice is it these workloads, works sets are so large that

it doesn't fit on one computer. The connectivity between the computer becomes the greatest challenge. And I showed yesterday the performance of a fast interconnect and the performance of a fast interconnect with the right type of CPU offload, the performance difference is 2x. What that says is that the architecture of the networking, not just the speed of the networking matters a great deal. The architecture of the networking, not just the speed of the networking and our vision is that someday the computing fabric will not stop at the boundaries of the server. The computing fabric would extend out into the network and the network and the compute will become one large computing fabric, especially for data centers, which is we know is the most important computer in the future.

And then lastly, we announced autonomous machines. Autonomous machines is both a edge opportunity for us, but we wanted to show you that in fact the reason why we're part of it is not just because of the edge, is because of getting to the edge is a big opportunity. That getting to the edge is a big opportunity, in order to create the ultimate AI, which is otherwise known as a robot or self-driving car or IoT device, AI IoT. When people see those things, they're saying basically of robot, an autonomous machine in order to achieve that capability and putting intelligence at the edge, the process of getting there involves deep learning systems, machine learning systems, data analytics systems, to develop the software, to simulate the robotics before you deploy it, and then of course, to deploy a very complicated set of algorithms software.

Our strategy with the autonomous vehicle is to enable the entire world of AVs to become autonomous. Whether it's robot taxes or passenger owned vehicles or trucks or cars or vans or forklifts, construction vehicles, farming equipment, they're are all going to have autonomous vehicle capability. We want to enable all of that. And so we created an open software defined accelerated platform -- accelerated computing platform.

We want to do the same for something that's even larger than that robotics. There are billion cars sold each year, but as we know based on the things that I described earlier, they'll be trillions of things around the world someday. They're all going to connect into essentially large networks that turn buildings, factories, cities, farms into essentially autonomous robots. The future factory would be a factory, would be a robot that's building other robots and so that when I say robot, I don't actually mean necessarily somebody who has limbs and walks around. The concept of robot, chat bot and AI assistant is essentially a digital robot.

Okay. So when I say robot, I just want you to hear something different than what you might be imagining an autonomous system, an AI system and so we announced a family of products and we announced yesterday that we are expanding our partnership with Toyota tremendously. We were already working with them on some early developments of cars and now from end to end from software development, AI development to simulation to computing, to algorithms, we're going to partner deeply with the world's largest car company. They have selected us to be their primary technology partner and we're very honored by that.

And so that's basically it. There are three takeaways then that one RTX has taken off. So, off to a great start retracing is here, there's a whole bunch of new workloads for the data center. Graphics is one, data sciences another, autonomous vehicles is another and of course IoT is another robotics at a very large scale and then the third is data science is that new driver for HPC and every data center in the future will be a high-performance computing datacenter.

I want to thank all of you for coming. And so with that I'm going to hand it off to Jeff Fisher.

## **Jeff Fisher** {BIO 2373419 <GO>}

Ladies and gentlemen, this is Fish. Thanks, Jensen.

## **Jensen Huang** {BIO 1782546 <GO>}

You're welcome, sir.

## **Jeff Fisher** {BIO 2373419 <GO>}

Are you sure, Chris. Yes, I love that our opinion. Chris and I have been working together for 25 years. We were children.

Welcome, everybody, to Investors Day 2019. Want to give you an update on gaming and there's a lot going on and we had an exciting year last year and for you guys all know and we look forward to even more exciting year this year, last year was a record year for gaming. We launched RTX, biggest leap in graphics in 15 years. 15 years ago, we launched programmable shaders in our firmy [ph] architecture today virtually every game is based on programmable shatters. With RTX, we launched a brand new architecture helding in real time ray tracing, and I'll talk a bit more about some of the momentum behind this next generation architecture.

Max-Q laptops driving thin and light laptops. This year we've got some of the thinnest, the lightest, the most powerful laptops driving the laptop market. I'll talk a little bit more about that. Just past last year most recently, we brought our turning architecture down into the mainstream to the 2019 price point, we now have a top to bottom stack of turning GPUs, millions more gamers coming onto the architecture. And finally, not to be missed last yea,r we mentioned Crypto came to town and this past year, it left town.

We see the Crypto hangover on track to sell through our channel inventory by the end of Q1. That is moving nicely, so this year was a record year 13% growth year-over-year, and let's dig in a bit more, first of all, the fundamentals of gaming remain very strong. Our basic core business is continues to be strong, as we've mentioned before, and I continue to say everybody born today is a gamer, every child is a born gamer. The demographics are also working in the favor of gaming, gamers continue to gain longer in life. You start out a gamer, you game longer in life, the total population of gamers continues to grow. And what's driving that? Well eSports

momentum is still huge. I wanted to -- if you don't mind for a moment, I saw in my inbox this morning, we get a weekly update from my team on what's going on in the world of eSports. I know you guys track that news very closely, but just in case you missed a few things, I'm going to read you a couple of things that came in my inbox this morning just for a note.

Call of Duty Franchise -- Franchise, eSports to sell at $225 million per team. Call of Duty franchise spots to sell for $225 million per team. Apex Legend is prime to be the next big eSports. Okay. No surprise here, Battle Royale blurs the line between Entertainment and eSports. It's not just for competitive gaming, but it's also for watching.

Snoop Dogg's you probably missed this, Snoop Dogg eSports series kicks off tonight. Walmart becomes a first major grocer chain to put eSports arenas in the stores, $5 for open play and leagues at night. ESPN announces creation of College eSports Championship, of course, Disneyland Paris to host a Dota 2 Major this year in May.

And as you know, the Dota 2 Major, the final tournament is being hosted in Shanghai this year. China's one of the biggest markets for eSports. In the Mercedes-Benz stadium that will hold 185,000 spectators and the Dota 2 to tournament is the biggest tournament in the world from a prize pool standpoint, last year, it had a $25 million prize pool. So eSports is obviously getting a ton of attention, the momentum continues to grow, and it's bringing in new gamers in the US is, but most importantly in the APAC regions in emerging markets, and in China.

The viewership for eSports is about doubling over the last four years, continues to bring in an audience, more people are watching eSports online than watching basketball, and the number of gamers continues to grow attracted by the competitiveness, the competitive nature, the social nature of eSports, and competitive gaming, about 30% over the last four years more gamers coming into PC gaming.

And of course, on the AAA gaming side, the cinematic side, game production value continues to increase. We talk about this year, but it continues to grow. Game developers are adding more realism into their games. It takes about a 5 times more powerful GPU to play today's games at 1080p 60 frames per second than games that were released in 2014. Games keep getting more realistic, you need a higher end GPU to play them, and these are the fundamentals that continue to drive gaming, and we see a strong future for gaming.

So let's take a little bit, -- take a deeper look into our business. Specifically within the gaming business, our GeForce GPU business grew 18% year-over-year last year, that is in both desktop and notebook. The contribution is both from units and ASPs are five year CAGR for units and ASPs, continues at about 14% total revenue growth over the past five years about 29%. But if you look at laptop, laptop is outpacing desktop

in terms of growth for reasons we'll talk about later, both ASP and units contributed to a laptop growth that drove about a 59% year-over-year increase.

Looking more specifically at RTX, Jensen mentioned that RTX is off to a great start. Well, I would say RTX is on to a great start, you see what I did there RTX is on. Anyway RTX on to a great start. We've now released RTX GPUs down to 349. At CES, we launched the RTX 2060 at 349, I mean time flies when you're having fun, but that was just about eight weeks ago. So I look back at our estimated sell-through of RTX from 299 up, that's a 2060 up compared to our estimated sell-through of Pascal from 299 up, that's about 1070 up starting from time zero of each of these devices.

Turning sell-through, RTX sell through has outpacing Pascal by about 46% in revenue, normalizing the time zero first eight weeks of sales. So RTX turning is definitely off to a great start, estimated sell-through.

If you look at our installed base, the installed base is ready to upgrade about half of our installed base, Pascal, the other half is older architectures. And turning is just getting its toehold in at 2%. And if I look at the performance of the installed base, 90% of our installed base is below one of our most recent GPUs, we announced the 1660 Ti is below the performance of the 1660 TI. I'll tell you in a little bit of why I pick 1660 Ti and why that's relevant. Another fact digging into our sales, the turning buyers that we're able to track that are upgrading from our installed base are buying up 90% of the GeForce RTX buyers are buying up from a lower price point. They had a lower price point GPU in their system, they bought a turning and upgraded. So 90% are buying up. So let's take a look at what's driving that, there are two types of games and not necessarily two types of gamers, but two types of games, eSports simply put, I'll say eSports, which is competitive gaming and Cinematic or AAA gaming.

Within our installed base, there is about a 50% overlap, about 50% of our gamers will play both, some will play one, some will play the other, but they all value the performance of the GPU. eSports gamers value frame rate, faster FPS means faster response time, and faster response time means more wins.

We see in our installed base, gamers that are playing eSports titles, once the play at 120 fips or higher. Interestingly enough, looking at our ecosystem, and specifically Fortnight, we can see the gamers that play at higher FPS have a higher what we call a K/D ratio. I'll call it a win/loss ratio. Gamers to play at 60 frames per second relative to gamers that are playing Fortnite at 240 frames per second win roughly 2.5 more times -- more often and their K/D ratio increases about 2.5 times more, they win about 2.5 times more often at higher FPS. It's natural, faster to point and shoot is the one who's going to win. So there is definitely a relationship of between more FPS and more wins, and the pros know this as well, there is a popular site called ProSettings.net if you've not been too it. ProSettings.net has about 900 gaming pros and streaming pros, online where they enter what all of their gaming hardware is, including their system config and GPU. 98% of those on pro settings dot net, the pros and pro settings dot net are powered by GeForce, and interestingly enough, over two-thirds of those pros are playing on systems that have the performance of

an RTX 2070 or higher, and over about a third are the performance of an RTX 2080 or 2080 Ti. So the pros know that the better the rig, the faster the system, the more often going to win and the better the game play.

Looking at our GPU stack. I mentioned the 1660 Ti, 90% of our installed base is below 1660 Ti. 1660 Ti is what's required to play Apex Legends, so I mentioned fastest growing competitive gaming title in the planet right now. At 120 FPS, 1080p high settings. Gamers -- serious gamers as well as pros value FPS, 1660 Ti is what we see as a starting point, but they don't stop there as with the pros, they will upgrade the rig to get the best possible performance.

Also mentioned AAA gamers. AAA gamers have different priorities or gamers who play AAA games have different priorities. Their priority shifts to image quality. These games are designed for cinematic's, the best possible image quality, and they'll play -- they wanted a smooth frame rate, say 45 to 90 frames per second. In order to get 45 to 90 frames per second on modern game, say, Metro, Battlefield V, you need to start with an RTX 2060, this is at 1440p. If I were to benchmark this at 4K, you would need to start at about an RTX 2080. So there is definitely upward motivation for gamers to upgrade to play the latest eSports titles at very high FPS, and to play AAA games at the highest possible image quality. But now is where the fun really begins, and that's with RTX and ray tracing. It was just past -- this past November that Microsoft launched DXR. Like I said time flies when you're having fun, but it was just about four months ago when the gates opened for ray tracing in games. Microsoft launched DXR. This week, the next huge -- big huge drop, Jensen had mentioned it as well, but epic is announcing that unreal engine for the number one game for -- the number one engine for AAA games is integrating DXR and RTX support in their game engine, and it will be shipping through the game to their thousands of game developers in the next couple of weeks. I think tomorrow, they're actually going to give a specific date and show some demos.

If you didn't notice from the keynote yesterday, real-time ray tracing is definitely the next big thing. The demos that you saw were unbelievable. If you aren't convinced then take a look at the control demo from Remedy that was posted last night on YouTube. Controls an upcoming game that looks amazing ray traced. Unity also announced that they're going to be -- Unity is the number one power's of about 50% of the world's games. Unity announced at our keynote yesterday that they're integrating RTX support and DXR into their game engine, and they're going to be handing out builds to developers starting on April 4.

In addition, most of the first party engines, including Frostbite, Remedy, CryEngine, engine from Crystal Dynamics for a games are all supporting DXR in real-time ray tracing. I'm heading up to the Game Developers Conference after this show. My team is booked solid with Dev's talking about real-time ray tracing coming into games. Real-time ray tracing is the most exciting technology we've rolled out. We've seen the best response that I have experienced at NVIDIA from developers to implement real-time ray tracing in their next generation games. We're tracking, let's say about a dozen games that are coming later this year, early next year to

implement real-time ray tracing and those are the ones that we just have visibility into with the release of Unreal and Unity I expect that to accelerate.

As Jensen mentioned, ray tracing is a software algorithm. It will run on CPUs, it's accelerated by GPUs, but we designed RTX to further accelerate to make real-time ray tracing possible in fully interactive games. If you look at Metro, we could run Metro on Pascal, I don't know if you've seen some of the coverage, but we are -- we announced that we're going to be adding DXR support to our drivers for all of our GPUs, so gamers can play with it. But the fact that it will run doesn't mean it's going to run Interactive. In fact, we're turning RTX with RT Core and RT Core plus (inaudible), RTX will accelerate over ray tracing over Pascal about 3x. In order to get fully interactive ray tracking you need an accelerator, you need a next generation architecture and that's what GeForce RTX was designed for. I should stop at the green bar, 3x. So, we're super excited about the future of RTX. We're super excited about the momentum behind real-time ray tracing. Game Developer Conference this week is I think when it all really kicks off in earnest, and we're going to see a ton of momentum coming out of the show.

We talk about notebooks now. Students want mobility. Students wants a game. Starting at CES, we launched RTX coming to notebooks, and we launched the next generation of Max-Q, thinner, lighter and more powerful notebooks than the world has ever seen. Max-Q has been driving the growth of the notebook business for the last several years. This year for the last several years, we estimate the OEM end market revenue of notebooks to be about $12 billion [ph], it's grown about 10x in five years. This is what the OEMs are seeing in terms of their total revenue and revenue growth from gaming notebooks. It's easy to think of a gaming laptop as the fastest growing game console. OEMs are so excited about gaming and notebooks in particular, they're rolling out more and more models. This year, we expect a number of notebook of Max-Q notebook models of thin and light gaming laptops to double to about 45 models. And within each model, there's going to be multiple GPU configurations, so you could easily double or triple that in terms of different notebook configurations that will be in the market this year.

Max-Q thin and light gaming laptops are taking over and driving the growth of the laptop market. Jensen also mentioned GeForce NOW at keynote, and the next billion gamers that we can address. Today, we have $200 million GeForce gamers. If you look at the entire population of gamers playing on underpowered notebooks, playing or want to play on underpowered notebooks or Mac's can reach another billion gamers. GeForce NOW has been around for about two years now in earnest. We've been perfecting the experience, quality of service, number of games onboarding, got about 500 games now available on GeForce NOW, 15 data centers, 300,000 monthly active users, about 1 million people on the waiting list, because we can't service them. The demand among gamers who are on underpowered PCs appears to be pretty huge.

Within our current monthly active users about 90% are playing on PCs that are underpowered, do not have GeForce GPUs in them. What is GeForce NOW? GeForce NOW is a GeForce gaming PC in the cloud. Give users access on lowing

clients to a high-performance gaming PC in the cloud. Fully interactive gaming, and we're rolling out VR. To simple game launch, we are not as store, it's a PC in the cloud. It's a simple game launch. You launch a game, off your desktop, just like you would any other game. And voila, it's playing in the cloud. We offer an open ecosystem. Publishers, developers, direct to gamers. We don't intermediate. We are not a store. The store is, the publishers, the developers keep 100% of their revenue. We are a service. So scaling out, we've had a ton of interest. We've seen a ton of interest from telcos, who are interested in interactive gaming and VR. It's a perfect use case for 5G, and it's a perfect value-added subscription to their broadband customers.

So we created a program called GeForce NOW Alliance. And what GeForce NOW Alliances is as Jensen had mentioned, we've developed a server that is optimized for cloud gaming. We're using that in our datacenters and we are patching it up as an end product for GeForce NOW alliance. We will sell a complete server, and on top of that, we will run our GeForce NOW service. License the telco, share revenue as it scales out. This gives us the opportunity to hit markets that we don't currently address, and it gives telcos, the opportunity to bring in more value-added customers into their ecosystem.

We announced two partners yesterday at keynote, SoftBank focused on Japan, bringing in their 6 million broadband customers, and ultimately 30 million mobile customers and LG Uplus in Korea. And as you know, Korea is a big gaming market as is Japan, bringing their 4 million broadband customers, 4 million cable customers and 13 million mobile customers ultimately into the ecosystem.

We expect to see the alliance services starting to rollout in the second half of this year. So that's gaming for me. I hope I touched on some of the things you wanted to hear about, our growth levers for this year, RTX is off to a great start. 40% -- 46% initial ramp revenue -- sell-through revenue, Pascal to Turning. GeForce laptops, fastest growing game console. That's the way I think about it. Students, gamers, kids, want mobility, they want high performance, they want thin and light Max-Q is driving this growth.

GeForce NOW, we can reach another 1 billion customers. We're super excited about the alliance partnerships, think our service is awesome. If you haven't tried it, you can log in, and I'm sure that Shawn or Simona can get you code to jump to 1 million gamer wait list, can check it out. It's really, it really is amazing. The interactivity is -- will blow your way on your Mac or enterprise notebook.

GeForce's alliance, then will let us scale out. We announced LG and SoftBank and expect to have more announcements coming over the course of the year.

So that's my story for gaming. Look forward to speaking with you all later, if you have any additional questions. Thanks so much.

I think Jay is up next for data center.

## Jensen Huang {BIO 1782546 <GO>}

Hey Jay, before you start, I got to make a quick announcement. I have quick -- make quick announcement. We are in historic grounds. It turns out, this cozy room is the location of the world's first GTC Developers Conference. This is how many developers we had. This is how it all started. This was the first one, anyways, I'm so excited to tell you that.

## Jay Puri {BIO 15036040 <GO>}

Wow. All right, great. Good morning, everyone. Welcome. It's nice to see you all. My name is Jay Puri. I'm responsible for NVIDIA's Worldwide Field Operations, and it's a real pleasure to be here. Today, I'm going to talk to you about our datacenter business. So we had another record year. We grew over 50%. The business is now $3 billion and the computing approach that we pioneered is just really taking off.

Our business is driven by applications, and you are at GTC, and you can just see the excitement that all the developers have about NVIDIA's platform. In fact, the number of developers grew more than 50% just last year. So the momentum is really terrific. Of course, we are number one in deep learning. We had a de facto platform for deep learning training and we are getting real traction inference now also. In fact our inference business last year was a few hundred million dollars. So things are actually going very well. There was a bit of a pause with some of the large hyperscalers towards the end of last year, as they digested some of their big purchases earlier in the year. But that is temporary. The amount of traction we have with them and all the announcements you heard yesterday with Matt Garman here with T4 and NVIDIA's RAPIDS platform now being incorporated into all of their machine learning platforms and so forth.

I mean the amount of stuff we're doing with these customers is actually quite mind-boggling, and so I'm sure the business is going to follow as it has to.

Okay. So let me talk a little bit about the size of the market. The overall server market today is about $100 billion, and we feel that $37 billion of that is right for high-performance computing as Jensen described it. So about a decade ago, a little more than a decade ago, we introduced CUDA to scientific computing, which was our first segment, and of course, at this point, we have a commanding position in that market. Right. All of the supercomputing centers, every major university, all the research centers, they are now deploying and really has accelerated computing model.

And then about five years ago, when deep learning came to the front and the Hyperscalers like Google and Microsoft and Facebook and all, quickly realized that artificial intelligence, deep learning was going to transform their business, and they need it a fast computing platform, and CPUs, but just not going to cut it, they all migrated towards GPUs. We quickly saw that opportunity and leaned into it big time. And so we took our CUDA architecture, widened its aperture a little more as Jensen put it, and we had libraries such as cuDNN and so on, and very soon, working with all

of the framework developers. We had the best platform for deep learning and we are doing really well with the Hyperscalers there.

And a couple of years ago after that, I think even the traditional industrial companies in automotive, healthcare, retail, financial services, the leaders began to realize, hey, AI is going to transform my business and so they all wanted to start using deep learning, and we introduced DGX, which is a supercomputing appliance that allows you to do AI really quickly, and will get off to a good start, and we're starting to make real progress in the enterprise now.

Okay. But this is just the start. As Jensen mentioned, data science is a new workload that is going to have a major impact on all of these segments. And it's going to mean that the high-performance computing part of the server market is going to more than double over the next five years, and we believe that NVIDIA's addressable opportunity there, Arkam is going to be $50 billion give or take. So we're really excited about that.

Okay. Let me talk a little bit more about the platform. Jensen did a great job of explaining to you that there's a real difference between an AI, computing platform and just an accelerator, you know, I think all of the computer science world has now understood that Moore's law is at an end and domain specific acceleration is the way forward. So obviously this is a big opportunity as I pointed out, many companies want to a part of it and so there is all types of accelerators that are being announced. And perhaps some accelerators like FPGAs and so on that would like to be platforms, but frankly they're pretty far from that if you use the broader definition as Jensen pointed out.

Now, look at our platform, right. It is we've been at it for over a decade, 12 to 15 years, and so we saw this opportunity, a lot earlier than most companies and so we've been investing in it for a long time. At this point, our platform is software compatible from the Jets and nano to the largest supercomputers in the world. We have been really disciplined about making sure that we maintain backward compatibility through all this time as we continue to innovate at a furious pace every year. And so the number of applications that are here is growing at a really rapid pace and this span multiple domains as Jensen explained earlier, right. So nobody has the maturity of this platform. Just think about the investment. We have made 10s of billions of dollars worth of investment in this platform ourselves and that's just the tip of the spear, it's really about our ecosystem. The 1.2 million developers that we have on the platform now, all of the scale-out partners. If you count the total investment in NIVIDA's platform at this time, this got to be, I don't know, hundreds of billions of dollars. So it is not easy for someone to come in at this point and try to duplicate this, right. So, we have lot of the important applications in the domains that we are addressing now whether it is scientific computing, AI going forward in data science and the performance is incredible. When you have a new domain like AI, it is important to have some industry specific benchmarks that everybody can look at to compare different options that they have. So Google led an effort to come up with a set of industry benchmarks called MLPerf recently and it's a very comprehensive set of benchmarks. They did a very good job. They are tough. In fact, we have lots of

companies that are part of the MLPerf consortium, but only about three or four companies could even submit results that met the requirements of the benchmark and I'm so pleased that NVIDIA was the leader in all six of the important benchmarks and not only that, but we beat the competition by a fairly healthy margin.

So it shows that not only do we have a very widely adopted platform, but it is the most performance platform that is out there for artificial intelligence. Okay. Our value proposition. Actually if you understand accelerated computing. I think you understand why the value proposition is so compelling. We are able to take. We are able to accelerate applications many fold. So if you can accelerate applications many fold, obviously you don't need as many servers, and so if you don't need as many servers, the acquisition cost is going to be less, and if you don't need as many servers, the energy cost is going to be less. I hope you know that in most datacenters, the energy cost actually is more over a five-year period than the acquisition cost. So as a result, you know, our value proposition is extremely compelling. And now of course, we have machine learning our data science, that's our latest workload and it's a huge opportunity and you can see our advantage in TCO is 80%.

All right, let me talk a little bit about our business model. What do we sell. We sell, we have 2 types of products we make our own systems the line of products that goes from about $40,000 to over $400,000 and then it's stacked up in racks in pods and so forth and we work with our storage partners and our networking partners to develop a complete solution for our customers. And we also take our technology to market through our OEM partners through our Tesla product line for example where Tesla cars where go from $1,000 to $10,000 and also our architecture is available through every single cloud service providers.

So let me talk a minute about why do we do this? Why do we have this product line? And what is our business model? We do our own systems for a couple of reasons. One reason is, of course, it's all about the full stack as Jensen mentioned to you. So if you're going to innovate on the complete stack, we have to have a reference architecture, and that is our reference architecture, right. It's important for us to continue to innovate and move the technology forward. It's also very important for our development partners, all the developers . They have to have gold standard, if you will, for NVIDIA's architecture, NVIDIA's accelerated computing platform.

A second reason. And just as important from my perspective, it's a great tool for business development. We have to go and create these markets, which means we have to go and engage with all of these lighthouse customers when we are first getting started and we need a way for us to be able to engage with them, and having our own product line that we can go in with and work with them on creating the first solutions and so forth is very important. So that's another reason why we have our own set of products. But I have a small sales force and really we want our platform to be ubiquitous, so the real go-to-market strategy actually is through our OEM partners and through our cloud service providers. And every single OEM in the world, every single system builder in the world is now using our platform to build their solutions and we are available in every cloud provider.

Okay. Finally, we have NGC, the NVIDIA GPU cloud, our software hub and that sort of unifies everything because it is available, our accelerated applications and know-how and so on is available there and that can be deployed whether it's on our systems or it's on systems of our OEM partners or even in the cloud. So that is sort of what we sell and how we sell it, okay .

A minute on our go-to-market strategy, right. So of course the foundation is our platform. That's where we add all the value and that's what we are really proud of, but because it is about domain specific acceleration, it's not about general purpose computing. What we do is we go and pick those domains. So we go into vertical industries whether it's transportation, or healthcare, financial services, or retail or what have you. We look at those industries. We go meet with the leaders in those industries. We try to understand what are their pain points what are the applications If you could accelerate would have a major impact on their business. And then we work with them hand in hand and see what we can do about accelerating those. And our track record of course is very good. So that's kind of how we go to market. We go and look at specific domains in specific verticals and then we go and accelerate those. Once that is done, then we have a tool such as we have a deep learning institute, whereby we can use that capability go and explain that to all of the other customers in that industry that, hey, we have a fantastic solution for you now. And of course, we spend a lot of time enabling our partners and our ecosystem -- other ecosystem players to allow us to scale out then and really go and make these solutions available widely, so that's it. Pretty simple, but actually it's a lot of hard work, but it's fun, lot of fun, because it's fun when you can offer that kind of a transformative solution to the industry that the types of discussions that you have with people, it's really, you can just see the joy that we're bringing to people and how impressed they are with what our platform can do.

Okay. Let me just go back very quickly into the three segments that I talked about Scientific Computing, Hyperscale, and enterprise, and just tell you why we think that our opportunity in each of these markets is actually growing very quickly. So Scientific Computing that's of course our beachhead that's where we got started and we are very proud of the science [ph] that is possible on our platform. It was a moment of pride for us when the Summit supercomputer is the fastest supercomputer in the world. Fastest supercomputer in the US, and there is already such fantastic science that is being done on it. I was just reading some articles about what they're doing about more -- some cancer research, some medical -- other medical research around addiction and so on, nuclear energy, fusion types of things, for renewable energy, weather prediction, just fantastic work is already being done on these supercomputers.

We also have the number one supercomputer in Europe with the Piz Daint, and last summer when Japan wanted to have a really great AI supercomputer, they wanted to have an AI supercomputer available for all of the industry to be able to use they chose NVIDIA. And so, we power that supercomputer also. The number of applications that we're accelerating is going up. We accelerate the top 15 applications that are important in high-performance computing and scientific computing, but at this stage, we actually accelerate over 600 applications. So from 450 last year to over 600 applications, and at this point almost all of the applications

that account for the vast majority of the cycles in supercomputing centers are accelerated by NVIDIA. The other reason that this market is actually going to become even larger is because it's not just about simulation anymore, people want to do AI at the same time, because as you can imagine, it's all about getting your answers fast, getting to scientific results fast, and if you can use AI to predict where your simulations are going, you can get results faster and so forth. So, in every field -- important domain of AI whether it's precision medicine, renewable energy, or all this climate weather science that's important, they are now doing both simulation and artificial intelligence, and again, because we don't have an accelerator, we have a domain specific accelerating architecture when they're using our product they can just do both types of workloads simultaneously, no problem, and it draws the overall TAM.

All right. Next is HyperScale. We are the leader in deep learning training, everybody knows that, but sometimes I get the question is that saturating, actually nothing could be further from the truth, just look at the numbers, the amount of teraFLOPS per day of training that is being done is just going straight and up to the right and not only that, but the complexity of the networks that are now being developed as people want to do more and more sophisticated AI is increasing. So today, when people at benchmarking training, it's usually MXNet 50, which they're looking at. Well, that's about 25 million parameters, okay? But the interesting AI that is going to happen is around the AI assistants and so on and for that you need networks like automatic speech recognition Jasper, that's 200 million parameters or BERT for natural language processing, that's 350 million [ph] parameters, and I'm sure we're just getting started. So, there is no question that the need for training is going to just continue to increase, in fact you can just look at the cuDNN downloads, and they just continue to go up. And as I mentioned before, MLPerf is proof if you any -- needed any that there is no better training platform then NVIDIA's. And we are available in every single Hyperscale. But I still believe that the big opportunity for us, in addition to training is inference and we're starting to get traction, but I think it's just going to accelerate. And let me tell you why, in the past when people were doing inference, a lot of that was images and it could be done in batch mode on Idle CPU cycles at night. So for example, if you have this Google cars roaming the streets, mapping an area and later on they want to label their maps with the names of businesses on the route, well, they can do that at night, there is no urgency to that, and there's plenty of idle CPU cycles they can use that. But if you want to do the types of interactions with AI assistant like I think Jensen demonstrated yesterday with an example of that with Microsoft, well, then it's a totally different story. You ask a question, the first thing you have to do is you have to go from speech to text in neural network for that, then the text you have to have some natural language understanding what is the meaning of this text, what is the context, you need some kind of a natural language processing network of -- inference on that, after you've done that, then you will do whatever is needed get a result back or search or whatever. And once you've done that, then you may display it as an image, or you may need to then go ahead and take that and put that back into speech, okay, you get it back into speech, but then that's speech sounds like a robot, you don't want that sound like a robot. So you need another network to make it sound more natural sounding. So the complexity and plus, okay, not only do you have to do all the stuff, you have to do this stuff in a few milliseconds, so that it's useful, not going to wait for idle CPU cycles to do that, right.

So you need GPU acceleration for inference going forward in a big way as AI becomes more sophisticated, and so inference is going to be a big opportunity for us.

And here is examples of many companies that are already using NVIDIA for inference, ByteDance, I don't know if you know TikTok, it's started in China, but now it's everywhere, it's short videos, they just exploring maybe one of the fastest growing company. They use us for video moderation, make sure that the content there is safe and nothing that we wouldn't want to have on their. PayPal is using us for fraud detection, billions and billions of transactions, right. But they -- by using our technology they can reduce fraud by 10%, by the way -- as I was talking to them, it's pretty interesting. The types of fraud that people think of, it's pretty amazing to all kinds of collusion between buyers and sellers then fixed stores being setup and what not, so people can be pretty creative, but they can find it out now in pretty much real-time using our technology and they can save 10%. Now they're going to save 10%, they said they can use 8x fewer servers, again, the TCO is just pretty, pretty incredible. And then, I don't know WeChat, TenCent, I mean this platform is just absolutely incredible, does everything, and a lot of the inference on that, including by the way if you end up using WeChat with somebody in China, it will do all the natural language understanding and provide subtitles in your native language, and it's -- the results are really, really great. So there's a lot of inference going on already, but as people do more sophisticated inference, it's going to be, I think it's going to be a very big opportunity for us.

And then finally in the hyperscale and in the enterprise, as Jensen said, data science is a big opportunity. It is the unicorn that only I think NVIDIA's platform is going to be able to address in the proper way. And already all of the cloud providers, whether it is AWS sage maker or Azure ML or Google ML, they have all adopted our RAPIDS acceleration into their platform, and it's being fair. So hyperscale, I feel very confident that our business in this space is going to just --

Okay. So finally, the third segment is enterprise. This was new and first, we started working with people, as I said, the leading companies is starting to work in deep learning. But then we realized, hey, these guys are actually already doing a lot of data analytics. I mean, everybody, we've been talking about the digitization of the enterprise, and they have -- they all know that to be competitive in this space, they have to collect data about their customers, about their suppliers, about their processes, and they have to get business insight from that in this extremely competitive world that we operate in. And so -- and so far, there is a lot of open source software, right, for doing that data preparation and the ETL part of it and then Pandas and kid learners who want to accelerate the models and then display them in graphs and so forth.

So that was turn off open source software already that these people are using. But frankly, they were just not able to be effective enough. And again, the reason is tons of data, but by the time, you actually to prepare it, Jensen gave the examples, yesterday, about -- I think was the Verizon network, whereby it takes eight days to

actually massage the data, so by the time to do that, it's already not current enough before they can even run the models on it and so forth.

So it really needs acceleration. It needs NVIDIA's accelerated computing platform. And that's what we've been doing. We've been working with all of these open source, the whole open source community all these algorithms and so on and making sure that they can all be sped up using RAPIDS. So that you can actually work in more of an interactive way, it's not interactive at least get results in a couple of hours rather than days and months. So you can really improve your decision-making and start making the real difference in your enterprise.

And so, data science is going to be huge. So that is the big opportunity for us in the enterprise space.

I have a few examples here of some of the work we're doing in deep learning. I mean, there is great deep learning work being done today. Continental, in the automotive industry, for example, there are big Tier 1 supplier to almost all the major car manufacturers, and they are of course embarked in trying to build self-driving cars, a great partner of ours, and they're using lots of DGXs to do everything from the data factory, deep learning training, simulation and so forth. We have a great relationship with them, whereby, not only do they buy our products, but we help them setting up the end-to-end flow for doing -- building these networks for self-driving cars, and that's just one example of I don't know how many companies in the automotive industry that we are now engaged with.

And similarly, Siemens Healthineers, they're a leader for medical diagnostics and they have lots of AI experts. They have about 40 AI applications that they are ready to deploy and they run hundreds of AI experiments today on their DGX supercomputers. And I'm pretty sure that every instrument company is going to do that and follow their example.

So we have a wonderful stuff going on in the deep learning space, machine learning and data analytics, data science, that is the big opportunity and already we do have some what we call light house accounts -- early accounts that we're working with to understand their needs and improve our platform and so forth. So Uber is using our GPUs to just to match the supply of their riders compared to the demand for -- sorry, supply of drivers compared to the demand of their riders.

I'm using these phrases even though that's kind of how they talk about it. I think what customers and drivers. But anyway they trying to match that to make sure you're going to get picked up at the right time quickly and they also use AI, and then, sorry, data analytics and machine learning for things like pricing, you're right and so forth and fraud detection and all of those things. So Uber is a great account we are working with now. Walmart is another account that is very excited about our platform. They're using it for things like forecasting, you can just imagine Walmart is the largest, largest retailer out there, the hundreds of billions of dollars of business that they do it. They can improve forecasting just by a little bit, so that we have less

spoilage or something that you want, when you go to their -- one of the stores is actually is not out of stock. That has an impact of hundreds of millions of dollars to them.

And so they need to do that in as much real time as possible. Today, they definitely use machine learning for that, but it's days behind. They don't have real-time information, and this would make so much of difference for them and they're very excited.

Okay. All right, so you can see why -- it's very evident that our opportunity in all of these segments is going to be larger and larger as we go forward. So I'm very excited about that. But the next thing that we want to work on, we want to make sure it's easy for people to adopt our technology, right. The easier, I can make it for them to buy, deploy, purchase, the faster our business is going to grow. So one of the key elements of that I think is NGC. NGC is really great. So it's the NVIDIA GPU cloud. We started by having it as a depository for our containers, but now it's more than that. Now we call it our software hub. So of course, we have now 50 plus containers that has our HPC applications that we've accelerated. It has all of the DL frameworks. It has many of the -- all of the RAPIDS algorithms, et cetera. There are so many different algorithms and we wanted to be end-to-end and so it's, this is not a simple thing to be able to pull these applications together, but we make it easy for people to use, because we just take it all, use all the best libraries to optimize the full stack and then we just containerize it and put it on NGC cloud, right?

So, and that number is just going to continue to grow. Make it very easy for people to go get that AI computing, data science computing. Okay. But we are not stopping there. We have not just a frameworks and trainers, but we have the training scripts for these frameworks. We even have pre-trained networks so that you don't have to start from scratch. You can use -- start with these and then do transfer learning on your own data and come up with networks that are optimized for your own work.

And then finally, we are even putting some of the key industry workflows up in the cloud for our customers. So two of them around medical imaging, Clara, some of those libraries train models for that and for Metropolis, some IVA applications and so forth. They are intelligent video app analytics. Those models are available in the NGC cloud now.

I think this is going to make it a lot easier for our customers to actually start doing real AI work and that will be good for our business. By the way, you can deploy these NGC cloud, I just want to reinforce is available everywhere, you can do it on-prem or you can do it in the cloud. And in fact, you can do it in any of the clouds there are on-prem on our -- any of our OEM systems that are certified for NGC. And of course, you can do it on DGX.

Okay. Some of the other things that we're doing to -- make our technology easy for people to deploy. One is I mentioned it earlier, these reference architecture partners. When we first got started, we reintroduced the DGX appliance and we said that's

great. We've got the whole stack all optimize and people can get started right away. And then we would find, they would put our DGX's in one room and they would put the storage in another room and connect it by one gigabit ethernet or something, and then they would say, hey, the performance is not very good, and so we quickly realized that we can't just solve the compute part of the problem. We've got to solve the overall data center problem, so people can deploy our technology. And we started having discussions with the leaders in storage, such as Network Appliance, Pure, EMC, IBM, and so on. And the networking companies, such as Mellanox, Arista and Cisco, and together we have developed these parts, these reference architectures. By the way, the reference, it's about domain acceleration. So, and it's about -- so it's not that you can have one reference architecture that does everything. This is a pretty -- this is an important work and take some effort. So, we have these POD's for different workloads, it maybe different POD for training versus simulation versus what have you like data science or whatever. So, we are working with them on actually putting these POD's together to accelerate not just -- to accelerate at the data center level. And it's pretty exciting to have these applications. They're all pre-configured and once it's done, we can show people what the results are going to be. So this proof-of-concepts that frankly drive you crazy, they'll buy one and then it takes six months for them to prove it out, now all of that stuff hopefully can be condensed into just a few days hopefully or at least a week or so, and then you can prove it out that, yes, you're going to get this kind of performance improvement in your workload and people -- and the customers are very happy about that.

We have some aligns now with the Colo data center providers. There is -- if you are doing scale up computing, there are certain requirements that traditional IT data centers are not used to handling in terms of the amount of power that is required, and just the density of the computing and so on. The cooling systems that maybe necessary, and so if our customers are having some difficulty working with their IT department, just go deployed in one of this Colo centers, right. They are now available and they know exactly how to build this out for you.

Finally, in terms of, again, Jensen talked about, we have two types of computing that we're focused on, early on in scientific computing (inaudible) on those capability machines, scale up computing, the supercomputers, but as we go into deployment, whether it is inference in the data center or for data science, I think it's both as you said, but today, a lot of the people are just doing inference on all of these volume servers that they have, right. So, by making T4 available in all of the high volume servers from these OEMs, we can allow them to do inference and data science right in their current data center, and millions -- I don't know 20 million or something of these servers are sold every year. And of course, you have Spark and so on to try -- as an attempt to make all of these distributed computing environment work as one, and we're accelerating Spark. So that's all great. And then over time, I think as the workloads get bigger and bigger and they want to do it faster, people are going to realize, yes, we can do it in the distributed environment with traditional servers and with T4 in them or there are many times when people are going to want, data scientists are going to want the fastest supercomputer with lots of memory and so on to go do data science and we're going to be able to address both of those capabilities. So T4 is now available from all of our -- all the major OEM suppliers, and we are no longer limited to just this -- just the capability machines, we also have the

capacity machines, the scale-out machines, which really widens the market and that's available for us, and it's again, all of the stuff is -- and you see certified, and so we know that it's going to support our platform and all the applications that have been developed on it.

Okay. So that's it. Lot is happening in our space, the data center market opportunity is a big one for NVIDIA, I'm very excited about it, at this point there's no question accelerated computing is the path forward. And if somebody ever talk to you about a new ASIC that came on, please remember, it's about the accelerated computing platform, it's not about accelerators. And I feel quite confident in our position when it comes to that. It is all about the acceleration stack and data science that is the next big opportunity. And not only is the opportunity big, but we are taking a lot of steps to make it easy for our customers to purchase and deploy our solutions, so that the business can grow faster. Okay.

Thank you very much.

## Simona Jankowski {BIO 7131672 <GO>}

Ladies and gentlemen, we will now have a brief 10-minute Coffee break.

(Operator Instructions) Good morning, ladies and gentlemen. Next we have Rob Csongor with Automotive.

## Robert Csongor {BIO 3210739 <GO>}

Hi, everyone. I'm Rob Csongor, I'm going to talk to you about Automotive. I'll give you an update on our strategy, you guys know our strategy and automotive is an end-to-end platform. It's an open platform and it's for building autonomous cars. So, what I'll do in the -- in my presentation is. I'll give you an update on how that business is growing. I'll give you an update on the market drivers, what's driving the business, what are the things that are important. I'll give you an update on our strategies and what the size of the opportunity is. And then I will talk to you about our progress. What are the things you can look at to see whether or not we're making progress towards our objectives, okay.

So first of all, in terms of growth, there's a lot of different growth factors that you can look at in our business. But there's a couple of I'll touch on. On the revenue side, I guess there was another record year for automotive, but we're on looking at a much, much bigger opportunity. So the thing that I think I would touch on and highlight as something that was very significant this past year.

So if you remember last year at Investor Day, we were just launching Xavier. We had announced that Xavier is going to be going out, we said that Xavier was the processor to power the autonomous vehicle to power the self-driving algorithms, to power the cockpit and we were just launching it. So you know that as part of our open platform. We work with literally hundreds of companies, sensors, tier ones, car makers, truck makers also to different vehicles, and during this past year, we basically

went from zero to order over 80 companies that are now building on top of the Xavier platform. So this is, I think, one of the most significant things. The Xavier platform of course is software compatible to the previous platform. Yes, people who were looking to drive an autonomous vehicle and get it out soon made the move from our previous generation to the drive Xavier platform, and that's really important. Another thing that's important if you look at this past year is that it's not just that people started developing on Xavier, but the different kinds of vehicles that are now being developed using Xavier as the base platform. Of course, Toyota this past year and you saw the announcement yesterday, but separately just individually Toyota has selected Xavier as the platform we announced that.

I highlight a few other examples just because they are interesting. Volvo selected Xavier as their platform, but they selected it for Level 2 plus. Level 2 plus translates to mass market vehicles. NVIDIA is not only being used or up until then you would seen us in mostly robotaxis are high-end Level 4 type vehicles. Now they are viewing Level 2 plus as an important first platform to engage on to get it out soon and this Level 2 plus is a very high function, fully featured autopilot solution that we call Drive AP2X and there's a number of reasons why Level 2 plus became more important this year and I'll talk about that.

On the -- at the other end of the spectrum from Level 2 plus, we announced that robotaxis are being built using Xavier as the base platform, Xavier or Pegasus as we called a development platform for Xavier, we announced this past year that Daimler -- we're working with Daimler to develop their robotaxis solutions. So not just cars ranging from Level 2 plus to robotaxis, but also different kinds of autonomous vehicles. This past year, you'll now see that there are forklifts or autonomous forklifts being developed on Xavier. There's construction equipment, earth movers now being developed on Xavier. There's last mile delivery vehicles, delivery bots, UAVs, UGV a whole world of autonomous vehicles being developed and that really brings up and illustrates the fact that the world of autonomous vehicles is much bigger to us today than it was last year. It's not just about cars and trucks.

We believe that every vehicle will be autonomous and the reasons are compelling. They're different for every type of vehicle, but in every case, they're not being developed just because this is a new feature that you'd like to add on. There is usually a critical problem or something happening where autonomous vehicles can uniquely solve a problem. So for example in cars, of course, we all aware that 3,000 people die every day in the world. We literally have a 9/11 every day in the world due to human caused accidents.

In trucking, it's a little bit of a different problem. We live in the Amazon era. Today, there is a shortage of 60,000 truckers in the United States. That's expected to triple by 2026. Furthermore, with electronic logging devices that are now required on truckers that limits the amount of time that they can drive per day and further reduces the amount of productivity that can be brought out into the road. Self-driving Level 2 plus solutions for truckers allow the truckers to extend the amount of miles that they can drive on the road, because the amount of miles where they're not driving, but resting doesn't have to be logged as driving time. This is a significant

changer -- game changer for people in that industry. In the trucking industry, of course, you're feeding the demand for delivery. There's a 120 million households in the United States, half of them 60 million or 130 million, sorry 65 million of those households are Amazon Prime subscribers. So, this is placing a demand on delivery.

Mobility services, robotaxis, buses, the cost of ownership, we have an entire generation of people, young adults today who don't want to own a car. The cost of ownership for using services is lower and also the footprint on the planet, the amount of parking lots can be reduced. And then the whole world of autonomous vehicles, it turns out that there is a thousand accidents, a thousand fatalities sorry that occur every day related to workplace accidents. 20% of those accidents are specifically related to construction. So this year, you saw Komatsu announced that they're using Xavier to develop an earth mover that can look around with cameras placed on the earth mover and make sure that you can detect workers that around it and make sure that no harm comes to them. Forklifts, delivery bots, tractors, agriculture in countries like Japan, farming has become a crisis. The average age of a farmer in Japan is 67, the amount of farmers in Japan is going to drop -- has dropped in the half in the last decade. And autonomous vehicles to help with agriculture and food production are not just a good idea. It's a strategic imperative to develop.

The result of all of these things and more is that you are seeing and this is projected autonomous vehicle shipments by 2025, 30,000 heavy trucks, 750,000 agricultural vehicles, 2.5 million commercial robots, 1 million and 1.1 million UAVs. Okay. So, the world of autonomous vehicles is much bigger than it was. To address this opportunity, there is really -- or to address this market, there's really three growth opportunities. There's three areas where Nvidia has developed a platform solution and what we call the end-to-end solution and the end-to-end solution really consists of number one, you have to build computers that go into the vehicles. So that you can autonomously drive. Number two, you have to train and develop deep neural networks to create the algorithms for those cars both in the cockpit, as well as in the car to drive. And then third, you have to test and validate those algorithms to make sure that the vehicle that you put on the road is safe.

These are not three individuals separate random pieces of equipment. The reason why Nvidia decided that we will build a car end to end is so that we could deeply understand the problem. And in the process of doing that, we of course learned that all of these things are essential to building on autonomous vehicle, you cannot build or deploy or test an autonomous vehicle without these things, and if we need them, then other people need them and that turns out to be true. The end result of this is that on the drive computer side, given all of the market dynamics I just described, we have a $25 billion TAM opportunity driven in the short term by Level 2 plus, Level 5, we have a $3 billion opportunity on the DGX side, just in terms of how many car makers are there, how many cars do you need to develop algorithms for millions of millions of images you have to collect per DNM [ph], 10 plus DNMs [ph] that have to be developed per car and then from that you can do the math.

This is only just getting started. Imagine all of these vehicles, more models, more cars, more vehicles coming out, of course this will just grow for us. And then finally, the testing and validation. The testing and validation really has to do with, you need (inaudible) to accelerate your testing and validation, because otherwise you are going to be spending hundreds of billions of driving miles for hundreds of years to test, adequately make sure the car is working. So just based on the kind of engagements we have now, the type of miles that have to be driven, we believe that this is a $2 billion opportunity for us.

At the high level, these are the opportunities. Specifically, there's a couple of things that are really driving the market for us. Over this past year, I think you're aware that Tesla Model 3 became the best selling premium car in the United States. In those cars, in model 3s, model Ss, model -- these different models, the autopilot function has an tax rate of close to 80%. They are selling that autopilot and generating an estimated roughly $1.5 billion of incremental revenue based on the fact that it is an excellent autopilot. It is operating with multiple DNNs [ph], surround cameras and it has high performance computing that's powering the whole thing. In contrast to very simple (inaudible) solutions, which can provide assistance, there's nothing wrong with them, but they just simply are not a full function driving autopilot. This is creating a market for a very full-featured autopilot Level-2 plus. So when you look for example at our announcement with Volvo of a Level-2 plus solution, you notice that this solution is being targeted at mass market. It's not just for a premium car. It's for top to bottom vehicles.

We think this is an important market driver. On the training and development side, of course you have to collect data, you have to label data, you have to train, but not only that you heard Jensen talk yesterday you saw -- you heard jay talk about the new opportunity of data science. Carmakers also collect enormous amounts of data, not just the ones that are about training and developing autonomous vehicles. So they collect data on customer behavior. They do pricing analysis. All of these things we believe are going to be opportunities for us in the data center of the automakers.

And then finally validation, simulation now is not just viewed as an option for deploying a car. You see increasingly articles coming out, now let's say that simulation is the key to accelerating the safe -- safety and arrival of autonomous driving and we believe that. We also know if you're aware that RAND Corporation issued a report where they said that they did a mathematical analysis of what it would actually take to test and validate self-driving car and they came to the conclusion that would be just about impossible.

You would have to drive billions of miles with thousands of drivers for hundreds of years. So therefore, you need an alternative solution where you can test for corner cases and a lot of the things we announced here at GTC about drive constellation are going to be the solutions for that problem. So all of these things, the opportunity, the market drivers of course, form the basis of our products and our strategies. When we say end-to-end, it means from driving, training and validation and when we say open, it means that we have a massive ecosystem of hundreds of partners that

can plug in. They can develop solutions on top of our platform. Our customers and partners are welcome to use as little or as much of our solution as they like.

For example, let me illustrate. We announced Drive AP2X yesterday. This is our full autopilot Level 2 plus solution. We have three Tier 1s that have announced -- three auto suppliers that have announced that they're building on the Level-2 plus, one is continental, one is ZF and one is Veoneer. Those three actually are the perfect example of how the ecosystem has the choice and flexibility to develop on our platform. ZF uses our software top to bottom, not just the driver west layer, not just the API layer Drive works, but also all the way through into applications.

Continental used as part of our software stack. They use our perception and then they supply a lot of their own planning and a lot of their own parking solution and then Volvo and Genuity using Veoneer are developing -- they develop on top of the driver vest software layer (inaudible) and then they develop the software stack on their own. Perfect illustration of the difference of three different partners building on top of the Nvidia platform. On the driving side, our solution starts with our platform DRIVE AGX with our software and of course all of the complexity, everything having to do with perception, localization and path planning and all of those break down into a whole bunch of different algorithms and solutions. Very, very complex, very compute-intensive and an enormous amount of software.

Yesterday, you heard us announce, I think we've shown previously that Nvidia has a world-class perception stack based on our artificial intelligence. We've also shown world-class localization to HD map working with every mapping company in every continent Zenrin in Japan, (inaudible) in China, here TomTom across North America and Europe. But yesterday we announced safety for (inaudible). We announced the mechanism for doing world-class path planning and creating computationally safe methodology for a car to navigate in a dynamic world with lots of moving objects. And then to take that methodology and transform it into driving software, that will allow an autonomous vehicle to drive safely. The end result of all of this together with our tools and then an ecosystem on top of it makes up our driving strategy and our driving platform, an enormous amount of work.

On the training and development side, in the last two years, since we first started engaging automotive companies, we went from basically a handful of customers to over 60 automotive companies today that our training and developing using DGX for automotive. This is obviously a significant increase. It's -- it makes up collectively in that number. There's 25 carmakers, 15 tier 1 truck makers, mobility service providers, mapping companies and startups. And by the way, it's not just customers of NVIDIA DRIVE. For example, here GTC, you can go listen to BMW present on training on a DGX at GDC. And BMW, of course, in their current generation are using Intel. So DGX represents opportunity and a product that the entire world can use to train and develop their-self driving cars.

And then finally on the validation and test side we announced drive constellation and it's really a three-pronged approach to how we test and validate a car. First of all, we do what we call component level so our software in the loop, imagine that you

can take the data that you have and play it back to your computer and then you can do regression testing, you can change things, you can say, hey, let me remove the radar, let me have a camera fail. Let's see how the algorithm response and you can do this in super time. Okay. For example, you can have several months of driving that occurs within a fraction of the time.

We also allow you to do DRIVE Constellation Hill or hardware in the loop. So the drive constellation box or solution is two different boxes, one box that is simulating or synthesizing the world, it is the creates the world and the other box is where you put your driving computer. It thinks that it's in a self-driving car. There are leads they come in that represents sensors. The sensor images and feedback that comes in or simulated. The drive computer there drives and sends out actuation singles back to the synthesis box -- simulation box. And as a result, we are now able to test it.

When you're driving normally, this companies are talking about driving you drive millions of miles. You know that most of the time nothing's happening. You're driving on one on one and everything is fine, it's sunny day, you're in the and you go. Obviously in simulation, you can create challenging scenarios much more quickly than waiting for them to occur in real life. So what we show here at GTC and if you have a chance to go check it out, it's amazing. At a touch of a button, we can make it rain, we can make it's snow, we can make it night time, we can make it foggy. And just confused if the (inaudible) is out of the car.

So all of this is, I think, essential to accelerating the testing and validation. Now this strategy which we came up with was born out of our needs to develop our platform. And as I said earlier, if we need it, then why wouldn't somebody else need it. Now up until yesterday, you might say how can you prove that, but how can you show a validation point that this actually is true or this hypothesis works. And today the best we have illustrated is to just highlight the announcement with Toyota.

The Toyota announcement is exactly that engagement model. It is a recognition of the fact that all of these things are essential. For the world's largest automaker to recognize that first, we need the computational power in the car to drive the algorithms. Second, we have to create the simulations to test and validate it. We have to have the computer, we have to have the development vehicle and of course, we have to have AI for the AV vehicles. This is NVIDIA's automotive business strategy applied to the world's largest automaker. It is the model for our engagement and it is the end result of what we intended with our strategy. So, we were excited to announce Toyota. We obviously believe that this is what's needed in order to scale to create lots of vehicles across all of this different world of a tons vehicles, and then of course, we look forward to making more announcements in the future.

Aside from this if you ask what are the key things that occurred this year, that you could look at that our key individual milestones or accomplishments towards our goal, I would really break it into our innovation, our product milestones, as well as partners. So, if you look at them a lot of them I mentioned Constellation, our simulation solutions, Safety Force Field, which is NVIDIA now moving to the third part of what's required for a self-driving car. We've shown world-class perception, we've

shown world-class mapping localization HD map, and now we're showing world-class solution for planning.

DRIVE AP2X, we believe Level 2 plus is now important, I think you'll see a lot of car makers make decisions on Level 2 plus this year. My Route, the reality is that HD maps don't exists everywhere. So, where they don't exists NVIDIA will create a personal map for you. It will be generated by the car based on where you drive, so that you can drive safely. All the things we show 50-mile, loop, Pegasus, we now have taken our graphics expertise and now leveraged it into creating the confidence of you, so you can trust at self-driving car. It's not enough to just have the car drive, the car has to communicate back to you what it sees. So, you can trust the car and believe that you're safe.

Hyperion, which is the extension of our development strategy from our SDK, we can now put our SDK into a car. The second, we make a change on our software stack and do an OTA you as a partner get it instantly, and that's part of our strategy. Our simulator TUV SUD, this is -- if you're -- if you know TUV SUD, they are one of the safety experts in the world. They certify various processes of developing a car, this past year they certified NVIDIA as being passing of certification for being able to develop a silicon semiconductor solution for a car, and this is an -- were the only semiconductor supplier to be able to reach this. In addition, we are now the only non-carmaker company certified to drive self-driving cars in China. China, of course very important not just to us, but to a lot of our customers and partners, global brand companies as well as the local companies in China. And then of course, global mapping. On the partner ecosystem side, you notice, I won't go through everyone, but you noticed that they're grouped into not just cars, but trucks, not just Level 2 plus, but robotaxis, autonomous vehicles, Yamaha, Komatsu, and off course Chinese companies. All of these are announcements that were made this past year, and I believe validate our approach. Okay.

So just to wrap up, our strategy is simple, we believe NVIDIA is the only company that is delivering an end-to-end open platform for building autonomous solutions as evidenced by the things I talked about. On the driving side, we believe the world of AV is bigger than ever, it's not just about cars and trucks, and I've shown you some of the design wins on these new types of autonomous vehicles. It's a big opportunity, and I believe the strategies that I talked about are game changers for a lot of the carmakers and certainly you see some of the evidence of that especially with the announcement with Toyota.

For training and development, we're just getting started. Collecting training and analyzing data are essential for autonomous vehicles. We've now grown to over 60 automotive companies on our DGX business. And like I said, it's just getting started. And then finally, on the validation side DRIVE Constellation simulation systems are now available and the DRIVE simulation system like every other part of our platform is open. We have multiple partners from IPG, developing physics models and sensor models to Cognata, who's developing traffic scenarios existing simulation solutions that already exist in the market that can now tie-in to our platform, because of our open platform strategy.

Okay. Thank you very much. And at this point, I'm going to introduce Colette Kress, our CFO.

## Colette Kress {BIO 18297352 <GO>}

Okay. Still morning. We're a little bit behind, but we can catch up. I'm going to try and just summarize in total, what you've heard throughout the teams and then we'll take that time afterwards open up for Q&A. But let's just talk through a couple of numbers. How about that? All right. So another record year, this is actually our fifth consecutive record year in terms of revenue. As we finished fiscal year '19, of $11.7 billion and growing more than $2 billion year-over-year. Growth rate of about 20% fueled by all of our different platforms, which we'll talk about. Our gross margin also a record in terms of its overall growth and reaching 61.7%. Keep in mind, there is still in there, we would have been higher except having to write-down some of the overall inventory later in the year. But since the absence of our overall IP licensing, our value-added platforms continue to drive our overall gross margin up.

Our operating income, also a record year and reaching $4.4 billion and growing faster than our overall revenue at 22%, overall profit whether you look at overall net income or EPS growing significantly faster at approximately 35% as well. Now, when we think about the market platforms that we just addressed throughout the room. You heard from three of them, four of them in terms of here all reaching overall record level. And this is in a view to look at our overall growth rate over the last three years and the compounded growth rate that we have seen. First starting with gaming. Gaming, in terms of its long-term growth rate has been growing 30% over this period of time. Even this last year growing 13%. But as you think about this going forward, you should think about the overall gaming as being an overall entertainment industry. (inaudible) was up here talking about what you should see in terms of the growth drivers as we move forward. RTX is here, a new overall architecture to take us forward for the next couple of years, and we now have a full portfolio of RTX available, talked quite a bit in terms of the overall ASPs [ph] and how they have overall helped our portfolio in the past, but as you can see there's even more opportunity as we move forward.

The overall unit growth in terms of gaming is definitely there as well, as we're thinking about the refresh opportunity of our existing gamers, and as we know there are more gamers coming onboard every single day. Those in terms of starting at a younger age and also staying in terms of longer in terms of, well, in terms of their 40s. So this in terms of will continue as we hope moving forward. We also talked about new opportunities and things that we have seen most recently, the growth of overall notebooks and the use of notebooks in the mobility to continue their overall gaming experience. Additionally, we talked about streaming gaming, and now we have an opportunity to again address this very wide and growing market in a new form factor and for gamers that have not actually been in touch with us.

Now, pro-visualization, pro-visualization also extension in terms of the graphics that we see on the gaming side, but taking that to the overall enterprise. We've seen an expansion of this market as well, largely focused in terms of the mobility of their overall workstations, the thin and light, the overall performance improvement has

expanded and you see in terms of the growth rate that we see in ProVis 15% over the last three years growing quite nicely. But you also have RTX coming to overall pro-visualization. You also have heard in terms of yesterday, our focus in terms of the creatives out there and how they can improve the overall rendering process with ProVis.

Data center, a business over the last three years has pretty much almost 10x increase. Just three years ago this is a $300 million business and were now approaching $3 billion. I think the whole day today as well as yesterday was really focused about the breadth and depth in terms of the overall solutions that we have for overall data center that means in terms of focusing not only on supercomputing, focusing on high performance computing, something that we've been working on for 10 years. But the addition of hyperscales over the last couple of years, but now the growth that we can see in terms of the enterprise, that focuses on many different types of workloads, focus in terms of deep learning, which you know us very well by in terms of overall training, but also what we have been able to do in terms of expand into overall inferencing. Our growth in terms of high performance computing and adding overall AI, an acceleration in there as well. But then lastly, we're focusing on many of the different workloads that the overall enterprises uses and the expansion of the market from data scientists to the overall focus in terms of rendering as well.

Automotive, on the surface in terms of, we're just getting started, we're still looking at a three-year CAGR of 26%. That 26% is largely due to our base of overall infotainment systems. But over the last couple of years, you've seen us also grow in terms of incorporating AI with in terms of the cockpit and our initial overall work in terms of what we can do for autonomous driving.

This is going to be broad and far in terms of where we can actually address the market using our solutions in terms of automotive. Not just thinking about what will be inside of the car, but what will be in their datacenters and what we will do to help them as they continue to have these cars on the road in terms of the testing, the validation and other pieces. So again, our overall portfolio, all in terms of in growth opportunities as we move forward.

Our gross margins. Our gross margins continuing to grow over this three-year period of time, and our value-added platforms continuing to be the most important part of our overall gross margin, and what has driven that. We'll talk about this further in terms of the need of overall software in terms of our platforms to bring them to market to allow people to overall use that. But as you know, the software is not necessarily included in terms of our gross margins that will be incorporated in terms of our OpEx.

So overall growth in terms of our gross margins and definitely an opportunity to continue overall growing. So we broke out here. Our gross margins in a slightly different view, in terms of our overall gross profit, where do we get the majority of our overall gross profit? More than 70% of our overall gross profit stems from gaming and overall datacenter which obviously takes up a good portion of our overall business. But keep in mind, one of the highlights that we talked about on our

last earnings calls was the impact of inter and intra overall segments in terms of both.

Mix is the largest driver in the near term of our overall gross margins. Mix both in terms of between our overall segments as well as in our overall segment, the black lines here, indicate in terms of the ranges that we can see based on the portfolio that we could sell in, in those two major overall segments. So these overall drive our gross margins as we continue to build a larger and larger proliferation of products in the terms of the datacenter as well as the different overall gross margins in ASPs that we have in terms of our growth gaming business.

Operating expenses. Our operating expenses business, excuse me -- our operating expenses here grew about 27% this last year. Trying to keep up with the growth that we have in terms of our product portfolio. Very well structured overall OpEx, because we can have an overall architecture consistent across net unified architectures allows us to be quite efficient in terms of the amount of spending that we need to do. Our outlook for fiscal year '20 as we move forward, this is slightly lower rate in terms of what we had seen in this last couple of years. We're expecting about a high single-digit growth rate or a little bit over $3 billion, $3.1 billion overall growth.

Our operating leverage, so we talked about this a bit n terms of what we have seen in terms of the leverage that we get from having a single overall architecture. Just five years ago, our engineers that we had, were mostly focused in terms of our hardware, meaning we had a larger organization in hardware than we did in terms of software. As you've seen us talk about the overall software over the last couple of days, you see now in fiscal year '19, we have a larger percentage of software engineers, a significantly larger amount of overall software engineers than we do overall hardware.

When we think about our R&D therefore by those platforms, and starting at the bottom in terms of the underlying architecture of the GPU architecture, that makes a 40% of our overall hardware, excuse me, our overall R&D costs. Our software layer is there for about 30% of the overall costs as we string that across all of the different GPUs and all of the different systems that we have.

On top of that, we just have a small percentage of about 25%, that allows us to go industry-specific market-specific in terms of building out our individual solutions, whether that'd be for automotive, whether that'd be focused on AI or whether that'd be focused on in terms of what we need for graphics as well.

Our operating margin expansion has been focused on this unified model that allows us to overall expand our margins quite nicely over the last three years and continue to effectively invest in our businesses without having to worry about the overall margin increase, and we'll probably see this continue as we go forward, as well as we look at this as a very key area for us to focus in terms of growth.

Our cash flow and overall cash balances. Our cash flow has grown quite about 3x over the last three years and we're reaching about $3 billion or $3.1 billion of this last year. That's allowed us to produce an overall cash balance, a 7.4 billion by continuing though with our overall capital return program.

Our capital return program is an integral part of our overall shareholder value and delivery, and since 2013, we've delivered more than $7 billion to shareholders or approximately 70% of our free cash flow. What this has allowed us to do in this last year is we started out the year with a little bit smaller in terms of capital return. We initiated our intent for capital return for the new year and started out at the end of fiscal year '19. So what we have remaining in terms of our intent for fiscal year '20 is about 2.3 billion to return to shareholders over this period.

Where and use of our overall cash. As we look backwards in terms of '19 vary in line with where we had talked about the last time we met, we'd focus primarily in terms of investing back into the business. You can see us was $2.8 billion invested back. We focus in terms of also CapEx. A lot of that CapEx is focused on our engineers and allowing them to give the tools, the supercomputers that they need to build in order for them to eventually sell them, but also our focus in terms of the capital return is a key area that we did -- that we focused on.

As we move into fiscal year '20. Fiscal year '20, you'll see about the same side of overall OpEx, a little bit higher, maybe about $100 million to 4200 million more. You'll see about the same amount of CapEx of a raw approximately about $600 million, focused not only on our internal engineers, but also in terms of the facilities that we need, but you'll see a large amount that we'll be able to take the cash that we have on the balance sheet to execute our overall transaction for $6.9 billion. We'll continue with our capital return and finish that out as well of used of our overall cash.

Highlighting here the title says, our outlook remains unchanged. We're in the middle of Q1. Just to remind you that our Q1 was not necessarily about normalized and in terms of overall returning back to where we believe we have in terms of the growth opportunities in front of us. $2.2 billion in overall revenue. We are still working through the excess channel inventory that we have in gaming. We indicated back in November that we thought that would take about one to two quarters to work through. We're on track and we feel confident by the end of Q2 that we will be completed with our overall excess inventory that we have in the channel. You've seen the initial signs of that as we've continued to start selling in our newer platforms into the market from the 2060 to 1660 and 1660 Ti. Our overall gross margin for the current quarter is at 59%, which is up 300 basis points from where we just finished this last quarter as well.

Our operating expenses will remain flat with last quarter. We'll see that slightly uptick in the next couple of quarters as we go, but that's what you'll see to get and reach that overall growth rate for the full year. We get questions quite a bit, such as you're often giving us overall full-year guidance on overall OpEx to help steer us on something that you can definitely control. We provided our full year in terms of operating expenses, in terms of looking at high-single digit overall growth over the

prior year, but we also took this opportunity to provide full year revenue range of overall guidance. We look at that to be flat to slightly down. The flat to slightly down, was to help the teams understand what we saw in fiscal year '19. We took this opportunity after overall cryptocurrency to find a quarter that was not tainted the cryptocurrency to come up with what we believe is a normalized run rate for overall gaming. That means we took Q2, Q3, Q4 as well as our Q1 guidance and looked at that in terms of the overall desktop business and concluded on average we'd look at about a $900 million quarter. On top of that, we have our overall console and notebook business, which equates to approximately $500 million.

That's a $1.4 billion normalized gaming baseline for us to start. And again remember Q1 doesn't necessarily reflect our overall normalized, as we're still working through that excess inventory. But that allows us as we move forward to grow from this point forward. It allows us to look at the back half of the year as reaching some of the growth potential of the great opportunities that we have produced today.

So that's what we have in terms of our full year overall guidance for revenue. We're excited to announce that we have signed an agreement to acquire overall (inaudible). This is part of the overall transaction summary and the key points about we will purchase it for $6.9 billion in overall enterprise value. We expect this deal to close at the end of our overall calendar 2019. And right now, we will work through the overall regulatory approvals that we need in terms of -- in the US and overall China. We're excited to bring the Company on-board and we'll be working now to get to better understand how will overall integrate them forward. But again we'll have to wait in terms of the overall regulatory approval. At the time that we close, we'll have a discussion in terms of what we expect in terms of guidance afterwards, how we will incorporate in overall -- incorporate overall (inaudible) in terms of our reporting structure.

That was our short summary and we are here for Q&A. I'm going to invite Johnson up here and we will open up hopefully turn on the lights out here right now little dark. For us to take questions from.

### Jensen Huang {BIO 1782546 <GO>}

Hey, good job Collette.

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

I enjoyed listening to my team talk. I'll first turn on the lights. We can see there -- see there much, -- much better.

## Questions And Answers

### Q - Toshiya Hari {BIO 6770302 <GO>}

Thank you for the presentation. Toshiya Hari from Goldman Sachs.

## A - Jensen Huang {BIO 1782546 <GO>}

Hi Toshiya.

## Q - Toshiya Hari {BIO 6770302 <GO>}

Hi Jensen. And one of just slides, I think you showed the trailing five-year CAGR for the gaming business both in terms of units, as well as ASPs. It was encouraging to see the ASP number. I think it was 14% accelerate from what you had showed last year. More importantly, considering all the things you guys talked about in terms of the eSports momentum, the (inaudible) initiative of the traction you've seen so far entering. How do you think about the next five years for that business both in terms of units and ASPs? And related to that does Intel's intention to re-enter the market over the next couple of years impact, how you think about or how does that impact your thought process if at all?

## A - Jensen Huang {BIO 1782546 <GO>}

Thank you. I'll answer the second one first. The -- we have to pay respect to all of our competition. I mean, we stay alert and we compete with -- we've competed with 120 graphics companies in our company's history. At one point in time, we competed against 35 at the same time, and they were large companies, there were small companies. And so we're quite a debt that competition. And this is -- you're looking at a company that's incredibly focused and incredibly intense, and from the leadership all the way down, there's just so much technical debt and so much passion for this business. I think you were going to remain quite competitive. But nonetheless we always should stay alert. In terms of growth rate, here's the way I think about it.

There is a couple of -- there is some -- there are some numbers that should inform us. On the one hand, it is recognized that the PC as a gaming platform host for gaming platform and a GeForce is essentially a game console. A game console has a reasonable price point in people's head of somewhere at the end of life, at the end about $300 and at the beginning around $400 to $500. That's kind of the ASP in the head of a gamer, does that make sense? If you're a gamer, you're going out to buy a game platform to play games. In the case of PC because it's a good host for the game console, they can upgrade that host several times with the new game console. So every couple of years they can buy into a new GeForce and they can imagine paying some $300 to $500 somewhere in that range for something that delivers performances much better than our game console would.

It's a very logical sensible thing for them that informs that. There's a couple of other ways to inform it Unlike a game console, that's largely for playing games, PCs could be used for eSports and there's two types of people did not do that. There are many ways that you can play sports. You can play sports because you enjoy it. And you can play sports because you want to win.

And I think that the another way to think about ASPs is for the people who are athletes or aspirational athletes or they just really love to win. They need to have better gear and that's one of the reasons why you see in eSports, the high-end GPUs

are like 20 ADTI's. And they want 20 ADTI's because they want to run never missing a heartbeat at 120-150 frames per second. Many gamers can click 300 clicks a minute. And so when they click, they want to make sure that they get a shot off before the next person.

And so that kind of frame rate, you're not going to miss a click. And so this is buying the best gear is another reason for that. The rest of it is production. Calue is increasing all the time. Max-Q increases ASP's. Max-Q increases ASP, because you're using higher end GPUs running at a much lower voltage. Much higher end GPUs running at a much lower voltage to deliver a great performance.

Max-Q's great innovation is really about running -- using silicon in a way that is about running at lower at the most energy efficient point. Max-Q increases ASPs as well. So production increased -- production value increases ASP, Max-Q increases ASP, competitive gear increases ASP and those kind of factors don't play into game consoles and the game consoles that sensibility provides for me, I think, the long-term floor. So I don't know if these numbers help you, but it's kind of in that space for us and that's those dynamics is what's causing ASPs to grow over time.

## Q - Toshiya Hari {BIO 6770302 <GO>}

Yes. Thanks.

## Q - Aaron Rakers {BIO 6649630 <GO>}

Aaron Rakers with Wells Fargo. Great presentation. I think one of the most interesting things that we heard is this idea of revenue sharing this GeForce NOW alliance. I'm curious kind of first question, how do we think about the proliferation of your partnership ecosystem are? How are you thinking about it in terms of the service providers. And can you help us understand the attributes of the revenue sharing model? How we should think about that from a financial perspective? And then one real quick follow-up question. Any updated color on kind of your visibility on the datacenter side would be helpful? Thank you.

## A - Jensen Huang {BIO 1782546 <GO>}

Sure. Every country has a different telco. From many countries in Eastern Europe to Western Europe to Asia, Southeast Asia, Latin America, India. This is the first time we've been able to create a game platform that can scale out to those regions. The other one billion gamers. Most of the gamers we've been able to reach are in the western and China, but there are so many emerging countries that would love to have access to PC gaming. PC gaming is particularly great, because it's free to play, it's social, it's easy to access, it's open. They want a PC anyways. They need a PC anyways. So, there are a lot of characteristics about PC gaming that makes it vibrant and unique. Using the GeForce NOW

Alliance, we can reach them. You buy a server from us and then we operate the network on top of it for you. You buy the server from us, we operate the network on top of it. We take in terms of relative to the -- when we go into a subscription model, when we go and right now it's a beta, when we go into a subscription model. So out

of a few dollars, call it $10 a month of subscription fee, maybe they'll keep more than half, and we'll keep less than half. And the reason for that is because they bought the server, they're operating it, they're running it on their network, does that makes sense? All of the capital investment is theirs. On top of that, we're bringing the network, we're bringing the service, we're developing all the software, we're operating it for them, we're enhancing the QoS or onboarding all the games, we're doing all the marketing, because NVIDIA is the gaming platform. And so, they get to benefit from it as well. One of the things that's really great for them is in order to capture, gosh, that's a terrible way of describing it. In order to win a new customer, the economic benefit as many of you know is quite significant like time. And so for them, this is a pretty fantastic way to differentiate their service over somebody else's service. How many -- the way that we see it is we'll probably enter into at least one of these relationships per country. And for the larger ones, maybe two or three. So this is quite a scalable approach. That's one of the reasons why we've built the whole stack. We could do this, nobody else in the planet can. We've built the whole system, architectured to whole server, develop all the software, everything is in one and everything is a one shop. And then of course, we've been operating the service now for a couple of years and we're getting quite good at it.

You had a second question.

## Q - Aaron Rakers {BIO 6649630 <GO>}

Data center visibility?

## A - Jensen Huang {BIO 1782546 <GO>}

Data center visibility. Our data center business is in a grid in my mind. There's high performance computing. There's high performance computing. There's CSP for training, CSP for inference, CSP for cloud and now enterprise high-performance computing. Enterprise high-performance computing, for example, data sciences, cloud computing, all the things that we do quoted today. Deep learning you know very well. Inference, Jay talked about, last year we kicked it off, we're doing fantastic. And then supercomputing, you know very well. So that's one way to think through it. And then you have all of the go-to market by industries, and you overlay that across. We monitor the intersections of this for everyone of those grids, because how they use it and how they go-to market is different. Jay told you, our way of going to market, basically several ways. One is of course direct sales to the cloud service providers. Second, basically a high performance convergence -- hyper-converged, high-performance computing solution, we call it reference architectures, DGX POD reference architecture. We also go through the market through enterprise partners. And so, we have all these different ways of going to market and we just track the pipeline for each one of those. Some of those, we get better visibility and some of them we get lesser visibility. For example, last year we had a little bit less visibility in the hyperscale data center, because they -- in retrospect we all realize now, they bought too much capital earlier in the year and they had to really, really slow down. We didn't know about that at the time, and by the time we found out, it was well into the quarter. And so, some areas we have less visibility, but we try to have as much of a pipeline as we can and monitor the pipeline on a weekly basis, so we feel pretty good about the year.

## Q - John Pitzer {BIO 1541792 <GO>}

Yes. John Pitzer with Credit Suisse. Thank you for the presentation. A couple of questions, one kind of near-term, one longer-term. On the near-term front you spent a lot of time yesterday and today, really focusing on your investments in software platform and ecosystem. There is one of your competitors that places a bunch of emphasis on process technology and mine [ph] with nodes. Would love to hear you kind of talk about where that sits in kind of your quiver of IP and maybe talk about the path to 7-nanometer for you, that's the near-term question. And I guess longer-term, last week, you clearly demonstrated that you think interconnect is going to be very important going forward in data center architecture. Wondering if you could make the same sort of comments around memory, because clearly there's not one of your competitors is looking at memory and persistent memory as perhaps a way to really lower the TCO. How do you view that as a competitive threat and what could you do on the memory side of things to help out?

## A - Jensen Huang {BIO 1782546 <GO>}

Yes. You don't hear us talk about process technology, packaging technology, memory technology. And the reason for that is, even though we are world-class at using it and oftentimes to earliest. For example, 3D packaging, the world's first is SXM, the largest chip that the world makes. HPM, we used it before anybody else. The reason why we don't talk about it that much, is because we are just as good at buying all that stuff as anybody else. I don't find it particularly differentiating -- by 7-nanometer. It's available for anybody who wants to buy it. They want to sell it to you. And so, that is not a point of differentiation to me. What is the point of differentiation is architecture efficiency. For example, the fact that 2080 Ti or 2080 or Turing is so much more energy efficient compared to somebody's 7-nanometer GPU. It's shocking to everybody, but not to me, that's the whole point. To be able to use something cost effective, so that we could -- and cost effective, cost efficient and get the most architectural innovation out of it, that's what we hire our engineers for.

TSMC hire their engineers for building 7-nanometer, our job is to get the most efficiency out of any silicon that we purchase. And our goal is to be able to deliver the best energy efficiency, the best performance, the best functionality at any given point in time. Turing is just crushingly good. Just got to measure it, it is that good. And that's one of the reasons why is off to a great start. In terms of the data center, where you see us really differentiate is of course we buy all the best. We're one of the world's largest consumers of HPM-2. In fact, we are the world's largest consumer of HPM-2. We're the world's largest consumer of 3D packaging of TSMC CoWoS. We shipped more 3D packages than anybody, we just don't talk about it, because our customers don't care, what they care about is the functionality they get, the efficiency they get, the performance they get, the TCO they ultimately get, that's what they care about, and that's what we focus on.

In order to overcome the slowing Moore's law, in order to overcome it in a dramatic way, and I don't mean improve it by 10%. If you want to overcome it by X-factors, which is what we're about. If you overcome CPUs by 10%, you might as well what just wait for the next CPU, because accelerated computing requires software optimization, you would only do so if there's an X factor in there. And I mean 10x

factor, because it's a fair amount of work, that's 600 plus applications all of those frameworks, all of those deep learning neural network models we now accelerate. Engineers worked on it really, really hard, ours, there's, the ecosystems, everybody working super hard. It wasn't because of the pervasiveness of CUDA, nobody would lift their finger to do it. And so, now that this is no different MapReduce is for Hadoop, what essentially what we just announced called RAPIDS for accelerated Hadoop, think of it that way. Okay.

Hadoop comes into memory is called Apache [ph] memory, on top of it is called RAPIDS. RAPIDS, the way to think about RAPIDS is essentially MapReduce except accelerated by GPUs. Well, you don't build that unless you have a great deal of computer science expertise, and that's what NVIDIA is, that's our differentiation, that's why we're not addressing a percentage share of a market someone else created, that's why our company is always talking about new markets that we're creating and those new markets tends to be -- tens if not hundreds of billions dollars large industries. You can't do that unless you go and reshape it, re-factor it, come up with new algorithms, you can't build faster ships to do that alone.

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

Hi. Mark Lipacis from Jefferies. Thanks a lot for the presentation.

## A - Jensen Huang {BIO 1782546 <GO>}

Hi, Mark.

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

I found the accelerated computing platform framework and vision particularly compelling. But it seems like some of your customers, your biggest customers also use that same Lexicon platform, and they also have lots of resources and I was wondering if you could help us maybe share with us a framework for thinking about the frame -- the platform that some of your customers are developing, is that -- is the NVIDIA platform, is it -- is that -- is your customer platform sitting on top of the NVIDIA platform or is it sitting next to the NVIDIA platform, let's just say 5 or 10 years down the line? Thank you.

## A - Jensen Huang {BIO 1782546 <GO>}

Yes. Excellent. And the reason for that Mark is if you look at -- you look across CUDA-X, two of the squares are horizontal platforms and in that case a customer, a partner -- a partner of ours ecosystem partner would tend to jigsaw puzzle and interweave with it. Parts of our platform will stick out, parts of our platform will not stick out, but accelerate parts of their platform. Okay. So, let me give you an example. In the case of cloud machine learning platforms like Google Machine Learning Cloud or AWS SageMaker or Microsoft Azure ML. Okay. In those cases, our XGBoost library sticks all the way up to the top. Our RAPIDS, which is essentially the modern version accelerated version of MapReduce goes all the way to the top. Our cuDF is basically like pandas for one user or Spark for data centers, cuDF is basically like Spark, but accelerated in the pipeline [ph] ecosystem. cuML is basically psychic learn. Okay. So these are platforms go all the way to the top in some cases.

In many cases like TensorFlow, our Tensor Core architecture, Tensor Core AMP basically and cuDNN, CUDA cuDNN Tensor AMP, it sticks into and is deeply integrated with TensorFlow, but what you see is Tensor Flow, and so it just depends. The way we come out of this, we try to create a platform where if the ecosystem prefers another platform suppliers approach, we would integrate into theirs. If one doesn't exist and one never will exist, for example, if we didn't write RAPIDS, the MapReduce of GPU accelerated data centers would never exists. Nobody is -- nobody knows how to do it. Nobody has -- nobody of engineers to do it and nobody has the will to go do it, it's too much work. MapReduce sitting on top of yarn sitting on top of Hadoop is very complicated stuff.

GPU accelerate that is beyond comprehension, nobody is going to go do it, that's why we had to go do it, and it took about four years to go do that. And so we -- the first part is when there is a platform like data science, we integrate into it depending on how they like. Okay. And so, Google has some of our stuff sticking out, notice RAPIDS is now in virtual machines on the Google cloud platform for their machine learning. It sits next to TensorFlow, in the case of SageMaker, some of it more of RAPIDS integrate into SageMaker and some of its (inaudible) out. In the case of Azure, the vast majority of it's takes out. Okay. So that's one answer. The second answer is in some vertical markets, like for example, large scale medical imaging, computational software defined medical instruments. The future of medical imaging, multi modality, image reconstruction, AI, visualization, segmentation in 2D and 3D multiple disease, multiple sensor modalities. We've created a platform for that because it one doesn't exist on the planet. We call that Clara. We would now integrate that platform into our partners. For example, GE, has their medical imaging platform it's very, very good.

Siemens has an excellent one. Sony has parts of it. Canon has some of it. Toshiba has some of it. So Philips has a lot of it, and so we will integrate Clara in pieces into those. We'll integrate all of it into Nuance, which is the text annotation standard practically of radiologists. Okay. So that's a Clara example. We also gave you a drive example. We developed a whole stack from top to bottom and end-to-end, and then everything is open. So that if somebody would like to use our simulation platform, but not our physics platform, they rather have their own car physics simulator for example IPG, we're delighted to plug that in. If somebody would like to have our visualization in our physics simulation, but they would like to have somebody else's traffic AI simulator Cognata, we're delighted to plug that in and so we create APIs all over our platform so that the ecosystem could adapt to it.

Now, the positive way of thinking about it which is the way we think about it is of course, we would like to enable the ecosystem to shape our platform in the way they'd like to use it. Okay. The benefit to us of course is -- we're more central part of the ecosystem. If you look at the transportation ecosystem every day that goes by more and more and more and more people has some of our stuff all over the company, whether they're buying our chips for the car or not buying our chips for the car, they have our development system, sometimes they built their own development systems, but they have our chips in the car. Sometimes they have our software in car as well. And so, all kinds of ways to working with people.

So Mark the answers is this, there's nothing more powerful than a platform of platforms. That's how we can reach, that's why the NVIDIA ecosystem is sticky, that's why the platform is sticky, because we have other people's platforms integrated with our platforms. Our platforms is also I don't -- and together we're helping the ecosystem, helping that industry move forward. Okay. Simplistically that's how.

## Q - Timothy Arcuri {BIO 3824613 <GO>}

Hi. It's Tim Arcuri at UBS. Thanks. I had two questions. First in gaming, Jensen, if you read a lot of the websites, they sort of talk about the fact that most of the gamers that are playing AAA games they have pretty old monitors three to five-year old monitors. So how do you think about maybe whether the display technology becomes a ramp or a gate on how fast turning might ramp, number one. And number two, in terms of manufacturability, you're already radical [ph] limited on a lot of your designs. So how do you think about how to combat that, do you move to a Chiplet design and Intel's already sort of moving in that direction, so can you talk about that too? Thank you.

## A - Jensen Huang {BIO 1782546 <GO>}

Sure. The vast majority of the world's gamers are currently at 1080p. And the first thing that they want to do is in the world, once that market is at any given resolution in the case of 1080p, the first thing that they wanted to get to 1080p, but then they want to increase their frame rate within that 1080p and they want to increase their frame rate and then they want to increase the beauty of the images of 1080p. Okay. Now increasing their frame rate is not just about seeing its smoothly, it's about reducing latency. So 100 frames per second is much, much lower latency than 30. Right? 30 frames per second, 33 milliseconds, which is quite a large number of millisecond's in the world of competitive sports. And so, that's within 1080p. Once they achieve over a 100 frames per second and the visual fidelity, all the options are turned on, then the next thing is they would like to go to the next resolution, which is 1440p. When you go to 1440p, everything gets cut in half. And then now you've got to double -- you've got to increase your graphics processor, so that you could start getting your frame rate back. Meanwhile, we just added ray tracing. And so, we're going to keep on making their gaming experience better, every two or three years the resolution of monitors kind of clicks up another 2x. The next one after that is 4K. But right now, people are at 1440p. So, I don't find that monitors are an obstacle at all, because they're too -- as you know I just mentioned, there are four factors, there's monitor resolution, there's latency and frame rate, there's visual fidelity and then there's new features, and we've got some really great new features coming.

I'm sorry, your second question? I just turned 56, and it's like point.

## Q - Timothy Arcuri {BIO 3824613 <GO>}

(inaudible)

## A - Jensen Huang {BIO 1782546 <GO>}

Oh, yes. Right. Chiclets, I think Chiclets are good. They are yummy, I like the orange version. We are at -- your question was actually reticle limits. We are at reticle limits,

Pascal, the P100 was near reticle limits, VOLTA reticle limits. VOLTA is reticle limits, it is the reason why we invented NVLink. So that we could take 16 GPUs that are reticle limits and connect them altogether. Okay. And then that's number one. And there are limits to 3D (Technical Difficulty) fitted into one node, no matter how big that node is. And so, we have to find a way to connect that through smart interconnect, and that's the reason why we decide (Technical Difficulty)

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

(Technical Difficulty) Unit growth in total, and we'll probably announce that as we work through the rest of the year.

### Q - Timothy Arcuri {BIO 3824613 <GO>}

And -- I'm sorry, just the inventory flush Q1 to Q2?

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

Yes. So, we indicated in one to two quarters starting back in November, that we would work through our overall inventory. So that means at the end of this quarter, that's the second quarter, one to two quarters that we get through -- Q1, Q2, but starting back with where we started in November.

### Q - Timothy Arcuri {BIO 3824613 <GO>}

Got it. So by the end of Q1, then it should be done.

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

That is correct.

### Q - Timothy Arcuri {BIO 3824613 <GO>}

Thank you.

### A - Jensen Huang {BIO 1782546 <GO>}

Stacy [ph], there's -- you might have -- you might be -- you may have an assumption that isn't quite right. Where it's or we've not been very clear. We have a GPU at every price point. One of the confusions that we created for ourselves is the price point of 1070 to 2070. There's an impression that we have -- those two numbers should always be the same price. That somehow a BMW 5 Series would be exactly the same price over the history of time and unfortunately that's not possible. A 1070 was higher price than 970, 970 was higher price than 570. And so, it's just that because people are paying so much attention now these days. They just nobody paid any attention to it in the past, these numbers were only numbers to us in the past, it has become numbers to society, it's a little bit like the three series, the five series, the sevens, it's a bit, it has gotten that kind of notoriety. And so, I think that, that was partly surprised me too. But we have a GPU at every price point. There's a GPU at 299, 399, 499, 599, just like the four. And at every single price points way better. And so if people are buying the 45%, that the simple answer is, it could have been just for

all from ASPs, it has to be from a lot from units, because our ASPs can go up that far. The simple answer is, yes, I mean, the simple answer is, yes.

## Q - Ambrish Srivastava {BIO 4109276 <GO>}

Hi. I'm Ambrish from BMO. Colette, I had a question on the op-model. And as we think through gross margin for this year and I wanted to go back to especially in context of qualitatively, you've talked about the positive impact from the crypto business, so, and you gave us the delta between the inventory and then the 300 bps improvement, but as we think through the year. So near-term, what's the right way to think about gross margin trajectory and a little bit longer-term when you talk about the op margins focus to get those margins up. OpEx is kind of growing in line with what you've said consistently that you would be investing in the business. So, is it going to be more on the margin front mix, what's the right way to think about it? Thank you.

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

So, when you think about our gross margins, as we move forward. Moving from the conversation we'd just had about gaming. As we look at the ranges of overall gross margin in that business, as we see people upgrade and upgrade higher into a higher overall GPU, that helps us in terms of overall op margins. Additionally, when we think about our overall Datacenter business, we know that there is a significant amount of software that is incorporated in the platforms that we sell. Now, you have an opportunity again to improve the overall gross margins as our Datacenter business becomes more a larger percentage of our business as a whole.

Let's not forget our overall Automotive business, we're continuing that transition from just our overall Infotainment business to move to AI within the cockpit, move to the overall development services that we are working with them and then long-term when we think about the overall production piece of it as well. All of these things continue to change both the mix of what we're selling, and overall -- improve our overall gross margins.

Okay? When we focus -- your focus in terms of on the OpEx and where we are focusing on the OpEx, is that the nature of the question. Meaning, how do they -- in terms of the total [ph] pieces? In this most current year, we take a look at this on a yearly basis to say what is the appropriate amount of spend. We have great opportunities in front of us that we need to make sure that we have properly invested in, but we have the uniqueness of that unified architecture to probably get the most out of the spend that we do in terms of OpEx. Going forward, we're not here at a model that says we look at OpEx as a percentage of revenue. It's a little bit to massive Company and numbers focus. What we actually do is look at the workloads, what is it going to take us to get that work done. We focus in terms of redeploying even our internal headcount towards these projects, so that we can better utilize our workforce for more greater things.

So right now, I would look at, we will always keep OpEx front and center as a key area of investment, but keep that in mind in terms of focusing on operating profit in terms of how we can produce the best overall profit and leverage that we can as

well. Those are the two things that we keep in mind, rather than just an absolute overall OpEx or an OpEx as a percentage of revenue.

## Q - Matthew Ramsay {BIO 17978411 <GO>}

Hey, it's Matt Ramsay from Cowen. A couple of questions. I guess, first, Jensen, you guys made a bid for Mellanox last week and it's no big secret that there were a couple of other folks that were also involved in that bid, and I think you guys came out a bit late. So, I wonder if you could give us a little bit of an update as to the industry and partner reaction to you intending to acquire that business and what steps you're making to keep the InfiniBand standard open?

And then, Colette, maybe you could talk to us a little bit about the infrastructure your Group may be putting in place to monitor inventory levels across the business and across the channel? You may be getting a little bit less granular information now from the GeForce software stack as to when GPUs are actually activated for gaming. So whatever infrastructure you put in place there to monitor inventory and update would be helpful? Thank you.

## A - Jensen Huang {BIO 1782546 <GO>}

We are super excited that Mellanox decided to accept our offer. Wow, was it competitive? And the reason for that is because there is such a unique company, 20 years in the making, 100% focus on high-performance computing networks, a software stack that's been -- that's integrated into high-performance computing software stacks all over the world. You know that this is -- when you're building a high-performance computing system, this is a great company to work with. They have a lot of expertise.

It is the only -- you should also highlight that when you look at all these press releases of systems being built, they seem to be the only other company aside from us mentioned, it's actually kind of interesting. And the reason for that is because their engineers work hand-in-hand at the datacenters on all the software engineering that's necessary to get the performance, lowest latency processing, the best performance, uploading necessary and the data structures be moved around in these -- in the distributed computing these days is really, really complicated stuff. And so, they are really a super special company.

The customers in the industry is just delighted. They're delighted because they really feel that this important company is going to be well cared of in our hands, because we understand computer architecture. This is a computer architecture question. This is a system architecture question. This is not a chip question. It's not a components question. It's an architecture question. And they understand that we care about this area very, very much. From the highest points of leadership all the way through this company, Mellanox knows, the industry knows this is something that we're very good at, something we care very much about. And we're going to continue to invest in this and we're going to invest in it, leveraging many of the things that our Company has, they can take advantage of all that to accelerate their development. And so, this is an area that the industry is just delighted by.

And when, of course, keep it open and our whole platform, as you know, is an open platform. What NVIDIA is about is creating open platforms that everybody else can build their companies, their market, their applications, their datacenters around it. This is an open platform company.

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

So the comments on the overall inventory in our process that we have done both in terms of our inventory that we have on hand, with the sudden drop off in terms of overall cryptocurrency many of the work that we had started for the demand that we -- followed that had started as early as six months prior in terms of the work with our overall fabs, our works in overall purchasing the components and the pieces that we need to put that together.

So, at the time that Q3 and Q4 came around and we had seen the drop off of crypto, it became that opportunity to look through primarily just the components and the overall amount of components that we had associated with that. We feel that's a thorough process that we do from time to time, and this was even more thorough process to make sure we fully understood. But again, looking in hindsight, probably nothing we could do given a lot of those purchases were done more than six months ago.

The other focus is focus in terms of on our channel and our focus in terms of where they are in this process in terms of channel. Now, what we've done is looked at not only just the weeks of what can we get in terms of reporting in terms of the weeks, but where they are in the life of the overall product, where are they before it launches, where are they at the time that launches, where they six, eight weeks into it, to ensure they have the appropriate amount to both feed the market and that we have enough inventory and that we haven't gapped out, but also on the side that says, is there the right amount levels if we are a year or two years down. The overall cadence is continuing but the rigor in terms of at the life cycle at any stage is probably where we put more of the focus. Jen-Hsun probably have more to add here.

## A - Jensen Huang {BIO 1782546 <GO>}

No, I know every chip by name now and I have a relationship with every one of them. And so, I monitor all of them from birth to the next life.

## Q - Harlan Sur {BIO 6539622 <GO>}

Good morning. Thanks for of things for hosting this presentation. Harlan Sur with J.P. Morgan. We had the head of your healthcare team, Kimberly Powell, present in our Healthcare Conference recently. And the team is doing a lot here, right, medical imaging, patient diagnosis, drug discovery, genomics and they're leveraging all of the systems platforms within your pro viz and datacenter portfolio and driving healthcare-specific platforms like Clara, that you mentioned Jensen, and I know that you're targeting the platform approach across other verticals, industrial, retail, agriculture. So, wondering if you can just size these vertical targeted businesses with your TAM outlook of $50 billion. Could the vertical focus represent 20% to 30% of the overall $50 billion TAM? That's my first question.

And then second question, if you could just give us an update on China, have you seen demand fundamentals starting to improve with the more relaxed government stance on gaming bans?

## A - Jensen Huang {BIO 1782546 <GO>}

I know the second question better, yes. The first question, the way we do it, the way we do it is this, we never talk to you about TAMs until we have clear side of it. So notice we've not one time talk to you about Clara TAM. We just assume it's zero. Until we really, really understand deeply like DRIVE and we are engaged deeply with the ecosystem, that's when we start sizing it, otherwise we go to zero. Industrial, we assume zero. But there's no question it's not zero. There's no question it's not zero. But we largely assume it's zero. Let's see what others, robotics we assume zero. There is no question it's not going to be zero. There is no question it can't be zero. It's -- it will very likely be the largest AI market. Everything -- sensors are literally everywhere, temperature sensors, vibration sensors, camera sensors, microphone sensors, unfortunately, sometimes. And so, it's going to be everywhere. And so, I think we assume it's largely zero until we have a really clear side of it. And then we can talk to you about it with some amount of expertise, until then we just assume it's very large based on intuition. We got -- most of the markets we go into in the beginning it's all based on intuition.

Let me give you an example of the intuition that led us to Clara that is very clearly divine intuition. The intuition was that in the future, healthcare, its most important instrument, imaging, medical imaging of any modality will be software-defined. That was the intuition that it will be software-defined in the future, and that was spot on. When we start -- we had that intuition about five years ago when we started working on it and we tried to not over invest in it in the beginning, so that we could do a lot of discovery work and do some prototyping work. And if you take a look at some of the early versions of Clara that I showed you, I mean, it was rickety, but it, at least gave us the opportunity to engage with doctors and research universities all over the world and get a lot of feedback and now we're in deployment. And so, that's how we -- that's kind of how we do it.

10 years ago when I started working on DRIVE, the early version of it was kind of rickety, but I knew that there is no question in my mind that a self-driving car was going to be a software-defined problem. You're not going to connect 17 chips, separate chips from 14 different vendors together into a what is apparently, a self-driving car, that is not how it works. And so, there is no question in my mind it was going to be software-defined. And so, we just kind of ticket methodically and the timing has to be right, there are some other things that we're working on, that I don't think the timing is quite right. And so, we underinvested slightly and -- but I keep an eye on it and dabble on it so that this company has a future beyond what we currently described to you and we have a future 10 to 15 years out, that we're working on at all times.

Okay. So that's the thing I really love about our company is the ability to, on the one hand, execute incredibly well on today's work, realize the dream for tomorrow and start to explore the day after that and to find the right balance of all of that, that's -- I just love working with the management team on this.

## A - Simona Jankowski {BIO 7131672 <GO>}

I think we have time for one last question.

## A - Jensen Huang {BIO 1782546 <GO>}

Sure. Oh boy, the pressure is high, sir.

## Q - Mitchell Steves

Yeah, thanks. I'm Mitch Steves from RBC. So I just want to turn back to the gaming, I really had two questions. So, first, it's good to hear that the Turing launch has gone well for the beginning. But how do we get comfortable, I guess, around the content increases going forward without any visibility, the kind of the games we made?

And then secondly, if I recall, about a couple of years ago, you really just emphasized VR the unit opportunity there, what the ASPs will be. But I notice now it's kind of like not as topical, so I'm guessing -- I'm wondering why that is and kind of what the unit opportunities going forward, since that was -- ASP somewhere around a 2020 opportunity.

## A - Jensen Huang {BIO 1782546 <GO>}

Yeah. Great. Let's see, why am I so absolutely certain? I'm as certain about ray tracing as I am that this is the last question because I'm in total control of it, because you said so, because Simona said so. So number one, the reason for that is this, there is no question that ray tracing is the right answer because it was always the right answer. Using -- mimicking the physical behavior of light, the physical modeling of light is what computer graphics is all about. And the issue with ray tracing was never -- was it the right answer, was it the more elegant answer, is it the simpler answer. It's all true. It's just that -- just was too computationally intensive. And so, we found a way to use this hybrid rendering approach of half -- some rasterization, some ray tracing and that's what RTX means, mix mode rasterization and ray tracing.

We invented this technology, invented this approach and we evangelized it to the ecosystem, to the world. And you saw some of the things that happened: Microsoft with DirectX R; VulkanRT, engines built on top of it, Epic's engine is now 4.22, it is now DXR ready, RTX ready and Unity's next build coming out on April 4 is also RTX and DXR ready. These are the engines of the game industry. This is the operating system of the game industry. And if the engines has it and it works fast, it's -- you just use it, that's how it works, you just use it. You don't have to invent it, you just use it. It's in the toolkit. And so, there's no question in my mind that is going to happen. I'm absolutely certain of it. Okay.

And so, games keep coming out, games keep coming out as they come out on almost monthly cadence, several hundred games a year, as you know, and not to include, not to mention China, there's a whole bunch of games being -- and Korea bunch of games being made. So there's lots of games being made. There's no question in the years' time ray tracing will be literally everywhere. This conversation is worthwhile to capture in a years' time, we'll come back as a gosh you're right.

It was the last question. That is world-class humor, sir. And so, what was the second question? Oh, VR. We don't talk about VR very much, but VR is really, really still very important. It's particularly we work with Microsoft on HoloLens, we -- because industrial design in the professional market between the two of us where we do really great work there. VR is used in industrial design all over the place, styling, architectural engineering. It is a very important part of our Quadro business. It's probably one of the reasons -- one of the drivers that's causing ASPs to go up.

In the consumer world, I think what we really would love to have is a VR headset that is less cumbersome with less cables, and Turing has a special connector that comes out of it that -- it's called VirtualLink that connect into a head-mounted display that reduces the amount of cables in the way that the cable tremendously. And so, a whole bunch of new head-mounted displays are coming out, it is starting to show up now. And I think you'll be surprised. I think there's no question that the experience is fantastic.

And then the next step beyond that will likely be some form of head-mounted display that is VR, AR-ish and stream from the cloud. If somebody can figure out how to stream VR from the cloud and you might have seen some of our work in this area. Some collaboration we've done with AT&T and Verizon to test NVIDIA's wireless VR from GFN from GeForce NOW, we could stream VR directly out of the cloud. This technology is still in development. We're still very early -- I would say, beta quality, but the experience is really quite phenomenal. When you take that and you connect it up to a head-mounted display now that's wireless, then you have no cables at all. And if it's semi-translucent, then where AR starts and VR starts and then it's going to be quite interesting. Okay. So, don't take your eyes -- keep asking me this question, we're continuing to work on it, the ability to mix reality and virtual reality is going to come. It's absolutely going to come and I'm excited about it.

I want to thank all of you guys for joining us today. And GTC, we're at -- this is all where it started and it's kind of fun to sit up here and chat with you guys where it started, and now you guys know what GTC turned into. Last year, we had over 30,000 GTC attendees, and I'm looking at 200, that's fairly fast growth in a matter of 10 years. So I want to thank all of you for your support. Have a great GTC.

## A - Simona Jankowski {BIO 7131672 <GO>}

And lunch, if you head out the doors turn to the left in the Gold Room and we'll be here, as well as with the executives from the Company and we'll join you for lunch. Thank you.