Investor Day

Company Participants

- Ali Kani, VP Automotive
- Colette Kress, Executive Vice President and Chief Financial Officer
- Ian Buck, VP, Hyperscale Computing
- Jeff Fisher, SVP, Gaming
- Jensen Huang, Founder, President and Chief Executive Officer
- Manuvir Das, VP, Enterprise Computing
- Rev Lebaredian, VP, Omniverse & Simulation
- Richard Kerris, VP, Omniverse Developer Platform

Other Participants

- Aaron Rakers, Wells Fargo
- Ambrish Srivastava, BMO Capital Markets
- Atif Malik, Citi
- Christopher Muse, Evercore ISI
- Harlan Sur, J. P. Morgan
- John Pitzer, Credit Suisse
- Matt Ramsay, Cowen and Company
- Stacy Rasgon, Bernstein Research
- Tim Arcuri, UBS
- Vivek Arya, Bank of America Merrill Lynch

Presentation

Operator

Welcome to NVIDIA's Investor Day, at GTC Spring 2022. I hope you had a chance to listen to Jensen's keynote kicking off GTC this morning. It was packed with new products and amazing innovations. We issued forward team press releases this morning, which you can find on our website. We have an exciting Investor Day planned for you over the next two and a half hours.

Before I go over the agenda, let me quickly remind you of our Safe Harbor statement. During today's presentations, we may make forward-looking statements based on current expectations. These are subject to a number of significant risks and uncertainties and our actual results may differ materially.

For a discussion of factors that could affect our future financial results and business, please refer to our most recent forums 10-K and 10-Q and the reports that we may file on form 8-K with the Securities and Exchange Commission. All our statements are made as of today, March 22nd 2022, based on information currently available to us. Except as required by law, we assume no obligation to update any such statements.

We have a packed agenda for today. Jensen will start with an overview of the highlights from our announcements this morning, as well as our strategy. You will then hear from Manuvir Das on enterprise computing, Ian Buck on hyperscale computing, Ali Kani on automotive, Rev and Richard Kerris on Omniverse, Jeff Fisher on gaming and finally our CFO Colette Kress on financials. We'll leave plenty of time for Q&A with Jensen and Colette at the end. You'll be able to find our presentation on the investor relations website later today.

And now, I'd like to turn it over to Jensen.

Jensen Huang {BIO 1782546 <GO>}

Thank you, Simona. She's amazing. Welcome to GTC. We have a packed GTC, 1,600 speakers, representing technology, retail, consumer Internet, pharma, finance, the auto industries and researchers from over 100 universities. GTC Talks cover AI, digital twins, climate science, quantum computing, protein engineering, 6G research and more.

NVIDIA is accelerating computing, across full stack and a data center scale. The compound effect has sped up computing by a million x over the past decade, a million x has democratized AI and open the opportunities to tackle grand challenges, like drug discovery and climate science. NVIDIA's full stack computing platform is open and in four layers chips and hardware, system software and acceleration libraries, the NVIDIA platforms RTX, HPC, AI and Omniverse, and AI and robotics applications and frameworks.

Each layer is open to scientists and researchers, computer makers, software developers, service providers and then customers to integrate into their offerings however best for them. NVIDIA is built like no computing company. Our open, full stack four layer data center scale platform lets us partner with companies across health-care, energy, transportation, retail, finance, media and entertainment to apply accelerated computing and AI to revolutionize $100 trillion of industries.

We announced a giant wave of products this GTC. New GPU, CPU and networking chips, new systems and new software products. NVIDIA SDK's are the heart of accelerated computing. These SDK's tackle the immense complexity at the intersection of computing, algorithms and science. With each new SDK new science, new applications and new industries can tap into the power of NVIDIA computing. NVIDIA SDK's connect us to new opportunities, and new growth.

We launched 60 plus new and updated libraries of nearly 500 at GTC, for millions of developers, scientists, AI researchers and tens of thousands of startups and

enterprises. The NVIDIA systems they run just got faster. NVIDIA now offers licensable software products and NVIDIA AI Enterprise and video NVIDIA Omniverse Enterprise with enterprise service levels, access to experts and multi-generational stability.

AI is racing in every direction, new architectures, larger more robust models, new science, new applications, new industries, all simultaneously. And transformers, and AI model architecture that opens self supervised learning has unblocked the need for human labeled data and boosted AI into warp speed.

NVIDIA AI is the engine of the AI industry and is used by 25,000 companies and startups. NVIDIA Omniverse is integral to robotic systems. The next wave of AI Omniverse is a simulation engine for a physically accurate virtual worlds, and digital twins. And just as (inaudible) are essential frameworks for perception oriented AI Omniverse will be integral for robotics AI.

The Omniverse ecosystem is growing fast. In just one year Omniverse has over 80 third party tool connectors. And downloaded nearly 150,000 times and integrated into Bentley systems Luminar T, our first third-party integration. We announced new GPU, CPU and networking chips and systems. AI applications like speech, conversation, consumer service, recommenders, computer vision, robotics and self-driving cars are driving fundamental changes in data center design.

AI companies process mounds of data to train and refine AI models. Their data centers are essentially AI factories. A whole new type of data center has emerged because of AI.

Today, we announced Hopper architecture H100 the new engine of the world's AI infrastructure. The performance of Hopper H100 is a giant leap over ampere and order of magnitude. H100 has a new tensor core with for petaflops, 4000 teraflops of AI processing, transformer engine multi-instance GPU with complete isolation, confidential computing, DPX dynamic programming instructions, and the fourth generation NVLink with sharp in network computing.

A DGX connects 8 H100s and a new NVLinks switch system connects up to 32 DGXs into a massive exaflops DGX super pod. Hopper H100 power systems at every scale from the H100 CNX for mainstream servers to DGX and DGX SuperPOD. H100 is in production with availability starting in Q3. When we announced Grace last GTC, we only told half the story. The full Grace is truly amazing. The grace CPU is a superchip connected by 900 gigabytes per second NVLink.

Grace CPU superchip has 144 course, and an insane 1 terabytes per second of memory bandwidth. Grace is on track for production next year. Grace moves and processes mountains of data and is ideal for AI infrastructures, scientific computing and Omniverse digital twins. One of Grace's best features is the rich ecosystem of servers (inaudible) libraries NVIDIA software platforms, RTX, HPC, AI and Omniverse and a world of partners that we will bring to Grace.

NVLink will be coming to all to all future NVIDIA chips, CPUs, GPUs, DPUs, and SOCs. We announced NVLink is open for customers and partners to build custom chips. NVLink opens a new world of opportunities to build semi-custom chips and systems that leverage NVIDIA's platforms in ecosystems. We announced the Spectrum 4 400 gig Ethernet switch and end-to-end platform. Spectrum 4 is a major new product, the world's first 400 gigabytes per second switch, a massive jump in performance translates to higher data center throughput and lower cost and power.

Spectrum 4 with CX-7 and Bluefield-3 Smart (inaudible) end points and the DOCA infrastructure software will be the highest performance in Ethernet platform which Spectrum for Ethernet, Quantum for InfiniBand, NVLink for multi-node DGX, and our DOCA networking, storage, security, infrastructure software stack, NVIDIA is ready to help build out the world's AI infrastructure end-to-end. Spectrum 4 samples in Q4.

The next wave of AI is robotic systems that perceive, plan and act. NVIDIA Avatar, Drive, Metropolis, Isaac, and (inaudible) are robotics platforms built end-to-end and full-stack around four pillars; ground truth data generation, AI model training, robotic stack, and Omniverse digital twin. We engage partners and customers in any or all four pillars. Our ability to add value at every stage of the AI and robotics workflow gives us many ways to partner with the AV and robotics industry.

The demand for robotics and industrial automation is increasing exponentially, NVIDIA works with 1000s of customers and developers building robots for manufacturing and retail, healthcare and agriculture, construction, airports and entire cities. One of the fastest-growing robotics segments is AMR, Autonomous Mobile Robots, essentially driverless cars for indoors. There are tens of millions of factories, stores and restaurants and hundreds of millions of square feet of warehouse and fulfillment centers. We announced a major release of Isaac, Isaac for AMR's.

Like the NVIDIA Drive, Isaac for AMR's has four pillars; deep map, NVIDIA AI on DGX, Isaac reference AMR robot powered by Orin, and Omniverse for digital twins. Orin, our robotics computer chip is a great success. Drive Orin started shipping production this month. Isaac Orin developer kits are available now. And (inaudible) scan developer kits are available in May.

Omniverse is central to our robotics platform and the next wave of AI and like NASA and Amazon, our customers in robotics and industrial automation realize the importance of digital twins and Omniverse. Last time, we showcased BMW, Siemens and Ericsson, this time Pepsi -- the Pepsi Company and Amazon Fulfillment Center digital twins. Modern fulfillment centers are evolving into technological marvels, facilities operated by humans and robots working together.

The warehouse is also a robot, orchestrating the flow of materials and the route plan of the AMRs inside. This is the busiest GTC in our history, the largest wave of new CPU, GPU, networking chips, new systems, new software products, new AI and robotics models. Today's presentations will cover our growth drivers, strategies and

opportunities. NVIDIA management will discuss five areas, Enterprise, Hyperscale, Omniverse, Auto and Gaming. Every group builds its products and strategies on one NVIDIA architecture leveraging the full platform and all our technologies to serve our markets.

This intense focus on platform leverage, let's just direct the full might of NVIDIA to serve every industry. In computing, we will distill our opportunities in serving $100 trillion of industries, cloud computing, consumer Internet, healthcare, financial services, energy, retail and logistics, manufacturing, industrial automation, higher education, scientific computing, digital content creation, and more into chips and systems, and are two major software platforms and NVIDIA AI and NVIDIA Omniverse.

We estimate our own available market opportunity at about 1% of the industries we serve. Over the years and decades ahead, our TAM will grow into this opportunity as you will hear today. We will start with Manuvir, who will talk to you about our opportunities in enterprise computing. And I'll be back in a bit for Q&A. Manuvir?

Manuvir Das

Thank you, Jensen. In this section, I'll talk about our opportunity with enterprise companies at large with a focus on our AI software. We've seen over the last few years, that AI has really does occur everywhere. Internet scale companies doing AI in the cloud, large companies doing AI in their data centers, and more use cases every day at the edge. This is why our data center business which includes all of these has grown in the way it has.

This view here on this slide shows the growth over the previous seven quarters. Of course, if we have chosen to project further back, the growth would look even more dramatic. We expect this trend to continue and we are prepared for this by nurturing a sustainable ecosystem.

Developers and startups everywhere have integrated with our AI software. It's way over 25,000 companies use our technology. Here's one example, Snap is using Reva our software for speech AI in their lens studio product. They use are pre-trained models and our inference software (inaudible) using our software. Now the way we usually talk about our opportunity is by looking at data center infrastructure, and how much of it will be accelerated by NVIDIA over time.

But the reality is that AI is a full-stack problem. There is tremendous value to customers from the software of AI. We have created more AI software than anyone. We see this as our business opportunity, both hardware and software. But also AI is about use cases that change industries either saving money or enabling building new business. For example, in retail AI is being used for automated checkout, a new experience that simplifies the shopping experience, driving more customers to stores. And it is also being used for loss prevention, saving money, in financial services for fraud detection, in logistics for optimizing delivery.

So this is not just about the envelope of traditional IT spend rather there is an opportunity for an AI provider to participate in the revenue of the industry itself. To that end at NVIDIA we have developed a full stack of AI. The best hardware of course, that's what makes AL (inaudible) practical in the first place. Then, the essential tools and libraries that underlie any AI use case.

Think of this layer as the operating system of AI. Every server used for AI will run this software, this engine of AI regardless of use case. And then finally skills created for specific use cases. Let me take a minute to unpack this stack. The lowest layer is the infrastructure underlying NVIDIA AI.

We have a wide ecosystem of OEM server builders, the public clouds, and our own systems, all growing rapidly of course. Like other enterprise software platforms, we have a certified hardware program. So customers can choose and deploy hardware with confidence. Notice that I included the DPU in the infrastructure layer. We are seeing early success with our BlueField to DPU.

On this slide I've shown three examples. And of course, we are working closely with VMware on Project Monterey moving software-defined data center services from the CPU host to the DPU. We see dress as the go forward security architecture for data center servers a DPU in every server. The operating system of AI is what makes AI go. The tools for data processing, training inference.

Based on our experience to date, we know that every server used for AI will benefit from having this software installed. And finally, the skills frameworks that implement particular use cases that are applied broadly. For example, Riva is a framework for speech AI. As a regulated bank, you can use it to translate audio recordings of customer conversations to text. As a retailer, you can use it to convert product documentation into human voice.

So, this is the full stack then, NVIDIA AI software on industry standard hardware. At this GTC, we announced version 2 of NVIDIA AI Enterprise the operating system of AI representing the next big step in enterprise adoption of NVIDIA AI. Whereas version 1 focused on virtualized servers running VMware, version 2 runs on both virtualized and bare metal servers running VMware, Red Hat or the platforms, as well as on all of the major public clouds.

And where is version 1 focused in servers with GPUs, version 2 runs on either CPU or GPU. Together, these enhancements bring NVIDIA AI to every server. Every server will run this operating system. We already see this in data centers where AI is developed. Going forward, we expect to see wide deployment of this AI at the edge for a variety of use cases shown on this slide. Cameras on the roads, AI detecting traffic violations, (inaudible) drive-throughs AI taking orders and recommending menu options.

We've been preparing for this growth for some time now by fostering an ecosystem for edge AI, just as we did with AI in the data center. The chart here shows the

growth in our ecosystem and also the components we have added to NVIDIA AI over time to enable this ecosystem, NVIDIA AI enterprise on every server, data center or edge. Along with us the flywheel of the ecosystem has been gearing up to sell NVIDIA AI enterprise.

I've highlighted some companies whose sales teams are working together with our sales team to sale NVIDIA AI. Earlier, I mentioned NVIDIA certified systems, hardware underlying our NVIDIA AI software now, we add to that NVIDIA AI Accelerated, a similar program for AI applications built on top of NVIDIA AI. Over 100 software providers are already in the program. The flywheel is connecting.

It comes back to a simple view of our business opportunity. One that we know already exists from our own experience with AI to date, the engine of AI on every enterprise server in the data center or at the edge. $150 billion of software opportunity to go with the hardware opportunity.

With that, I'll hand over to Ian Huang to talk about hyperscale.

Ian Buck {BIO 18454865 <GO>}

AI is transforming large markets. And everyday we work closely with our cloud partners to help bring new AIs to life. We collaborate on the system's, the physical and software infrastructure, the AI frameworks and AI applications both for their internal cloud services and with their cloud customers. And it's a platform that's continuously growing.

It is estimated that the cloud server market install base is 20 million servers and analysts project this number will grow to 35 million by 2025. Driving that growth is the consumer Internet, the apps, the websites, the services that each of us use everyday and is built on the cloud. Not surprisingly, 100% of the consumer Internet applications will be adopting the AI, From Meta to PayPal, Pinterest, Snap and Twitter. AI is being developed everywhere to process every engagement, every product, every recommendation to deliver great customer experiences.

As a result, AI recommenders are becoming the engine of e-commerce with over $7 trillion worth of sales projected by 2025. And these are just some of the customers using NVIDIA AI today. NVIDIA's growth in hyperscale computing is continuing as more companies and developers find new ways of adopting AI for their applications and the introduction of new GPU architectures turbocharges that adoption.

New GPUs do this in three ways. First, by reducing the time to train, we speed up the productivity of AI developers helping them deploy more AI in the cloud and drive faster growth for AI infrastructure. Second, by improving the scalability of our architecture, we expand the scale and size of AI supercomputers to help our largest customers, as well as NVIDIA ourselves to build the next generation of AI infrastructure and push the limits of what AI can achieve.

And third, by improving AI inference, the production use case of AI, we widen the aperture to allow, even larger and more powerful AIs to be deployed into production. Just as we saw a 3x revenue growth from the launch of the NVIDIA V100 to the A100 GPU so, to will the Hopper H100 enabled a new wave of AI models and applications. Hopper is the new engine for AI infrastructure and will be the platform for innovation for large language models, recommender systems and the complex digital twins in the cloud.

To advanced AI is important to understand the trends in AI. Over the past few years, a new type of neural network has emerged. Invented by Google, the transformer has become the dominant building block for neural networks. Built on the idea of attention, transformers help AI understand which parts of a sentence, an image or disparate data points are relevant to each other. And unlike CNN's, which typically only look at immediate neighboring relationships, transformers are designed to train on the more distant relationships, which is important for applications, like natural language processing.

And transformers are transforming AI 70% of the AI papers published in the last two years incorporate transformers into their work. Transformers are also the building blocks of the world's largest neural networks for large language models, like opening AIs, GPT3 and NVIDIA's own megatron turn NLG 530B. This neural network has 530 billion parameters trained on the corpus of the internet, to build intelligent chat bots and other intelligent language applications.

Hopper with its new transformer engine is explicitly designed to accelerate these transformers. It can train GPT3 6x faster than A100, reducing the time to train from five days down to just 11 hours and it gives the latest mixture of expert model -- transformer models from Google a 9x boost, reducing time to train from a week to less than a day.

Hopper's innovations, don't just benefit training, when deploying these models for inference Hopper delivers a 30x higher throughput compared to the A100. And Hopper's ability to accelerate transformers, will not only help bring new AIs to market, but will turbocharged AI productivity and as a result, the demand for AI infrastructure in the cloud.

There is a second equally important AI use case that is taking shape in the cloud, AI based recommender systems. Recommender systems are the commercial engine of the Internet. Hyperscalars and the cloud service providers use recommender systems to connect literally trillions of items with the billions of consumers. Even the simplest search query today, involves a complex recommender system that attempts on the first try and only a few milliseconds to connect you with the right product, article, tweet or advertisement.

NVIDIA Merlin is an open source framework for building large scale deep learning recommender systems. NVIDIA's Merlin is envy tabular library can accelerate feature engineering and pre-processing to manipulate the many terabytes of unstructured

data sets into AI tensors that can be operated on by an AI. In addition, Merlin supports distributed training with model parallel and bedding tables and data parallel neural networks, when across multiple GPUs for these giant models.

Snap used in NVIDIA GPUs and Merlin software to achieve a 50% increase in the cost efficiency and an improvement in serving latency by 2x for the content delivery. Training and operating recommenders with NVIDIA GPUs saves money, enables smarter, more intelligent consumer interactions and activates the $7 trillion worth of e-commerce coming to the cloud.

Inference, once you've trained an AI model, you need to deploy it. 5 years ago, AI could still be run on legacy CPUs within the hyperscale datacenter. The overall amount of AI workload was small enough and these models simple enough that one can use the millions of existing CPU servers to deploy AI. That's not true today, as AI has gotten smarter, AI models have gotten larger and more complicated. CPU simply cannot meet the real time inference requirements of modern AI.

Furthermore, as AI has become an increasingly larger part of the cloud workload optimizing infrastructure for the AI throughput of the data center matters. We have seen a rapid shift to GPUs as a result. Starting with the NVIDIA P4 GPU in 2016, to T4 GPU in 2018, and now with ampere based A2, A10 and A30 GPUs, we've experienced a 9x growth in our inference revenue.

We invested heavily in software for inference. A software platform for inference needs to handle all the different types of models used across the company to deliver inference in real-time, while maximizing the throughput of an infrastructure, as well as handling the increasing complexity of AI models.

To tackle these challenges, we built the open AI inferencing software solution called Triton. Unlike training -- unlike the training frameworks, Triton is designed exclusively for AI inference. It is an open-sourced framework supporting every AI model running on CPUs and GPUs and has become the de-facto framework for AI inference across the cloud and on on-prem deployments.

Last year, we announced Grace Hopper, the ideal processor for giant scale AI and HPC. This year, we've announced the new Grace CPU superchip, the world's fastest most efficient CPU for the data center. For markets where CPU performance is paramount Grace shines. And as AI models continue to get bigger and our GPUs get even faster, CPU performance plays an important role in managing the execution as well as the pre and post processing of data for AI operations.

The Grace CPU superchip is designed to be the CPU for AI infrastructure. Its performance and efficiency will allow GPUs to train faster and larger AI models without ever letting the CPU get in the way. Furthermore, Grace's configurability for new CPU, GPU system configurations optimized -- allows us to optimize AI infrastructure for different workloads, leveraging both existing (inaudible) CPUs and GPUs attached with the new envy link chip-to-chip interconnect.

The Grace CPU superchip is of course, an amazing CPU all by itself And we're seeing strong interest in scientific computing, data analytics, hyper scale computing applications, where absolute performance, energy efficiency, data center density matter for the CPU applications.

For NVIDIA, we see the data center as the new canvas of innovation, with every generation of the AI we innovate unconstrained, studying all aspects of how AI operates inside the data center, knocking down barriers and inventing new technologies and products to optimize that infrastructure. With new kind of computer servers based on our Hopper H100 GPU's and our HGX, GPU-based (inaudible) products, optimizing communication with BlueField 3, Spectrum 4, Quantum 2, and even integrating networking and GPU's into a single products, like our conversion accelerators we have added compute to the network itself, offloading teraflops of computation from the compute nodes and saving gigabytes of network traffic.

This year, we're even taking it a step further by taking envy link, previously, used to connect GPU's inside the server and now unleashing it into the data center itself, scaling interconnect beyond the server with NVLinks switch systems. With Grace, we can broaden that opportunity to build novel CPU, GPU system designs for the variety of AI workloads. And of course, working closely with our hyperscale partners NVIDIA is inventing the data center of the future together.

Of course, now that we've opened NVLink C2C our chip-to-chip interconnect, we are open for business for custom IP integration, which brings new custom silicon opportunities within video technology to the cloud. The available market opportunity has expanded and we've gone beyond the GPU and are now three chip company. The new data center reader is redefining the nine million hyperscale servers deployed each year. And this opportunity opens up $150 billion market for NVIDIA and hyperscale and with just the infrastructure opportunity alone.

NVIDIA is the only AI company that works with every other AI company. And we are at the center of the AI ecosystem, working with AI companies to bring AI to the cloud enabling new AI applications and accelerate innovation across industries powered by NVIDIA.

Thank you, and I'll now turn it over to Ali Kani on Automotive.

Ali Kani

I'm here to talk about our automotive opportunity NVIDIA. The automotive industry is industries large with 100 million cars sold a year and an installed base of over 101 billion vehicles on the road. Auto is at the beginning of a few inflection points that together create compelling opportunities for NVIDIA. First, advancements and electrification have pushed OEM's to re-architect their cars from the ground up into software-defined vehicles that use centralized high performance computers, that can provide a large and growing list of new features and services over the life of the vehicle.

Seconds, these services give OEM's an exciting opportunity to transform their business model like we have seen Tesla do with their autopilot software. That is grown in price from less than $5000 to $12,000 a car today. It's part of this AV software disruption, we're seeing an order of magnitude 10x in compute increase in vehicles as partners used twice the number of sensors in their cars with each sensor supporting five to ten extra resolution of current sensors in production.

These cars are also being developed with more advanced machine learning algorithms that enable the development of vehicles that support L2 plus all the way up to advanced full self driving capability. Now, we're just at the beginning of these inflection points. Today, electric vehicles and vehicles with L2 plus or higher software represent less than 10% of cars sold a year. But in the next decade, a majority of the vehicles being each year should be electric, software-defined vehicles that support L2 plus or higher capability.

Now, we've invested heavily to develop a full stack solution for the automotive market. We offer our Hyperion platform in vehicles, that includes our Orin SOC and reference compute and sensor architecture. We also offer DGX and OVX servers, and data centers that partners can use for AI training, map generation and system validation.

We try to make our car to cloud experience seamless by supporting common SDKs, APIs and libraries end-to-end. We have three layers to our automotive software stack above our hardware layer. All -partners can take our core operating system that runs on our hardware. Many take our drive works acceleration middleware that makes it easy to efficiently run advanced AI applications on our platform.

And some partner's like Mercedes-Benz and Jaguar Land Rover use our full stack application software across their car and cloud infrastructure. In such cases, the entire end cab, parking, mapping, autopilot, and even some IX software is developed by NVIDIA in partnership with our OEM partners.

Autopilot and AI cockpit application development is a grand challenge. It requires high performance computing, platform programmability and scalability, advanced machine learning and robotics know-how, as well as expertise and functional safety and cybersecurity. NVIDIA is unique in our ability to help our partners across this entire stack, from chips and software in the vehicle to data collection services in cars, to AI training, map creation services in the cloud to application software from AV to cockpit in cars, and finally, onto simulation for vehicle validation.

We invest to improve our solution across the entire stack, because we believe what will most differentiate automotive companies is the speed of their end-to-end development flow, from finding an issue in a car to root causing it, providing a fixed if it's quickly validated, and then securely OTAing better software into every vehicle.

We estimate Auto to be a large $300 billion market opportunity. And NVIDIA's opportunity spans both hardware, software and services in the car and in the cloud.

Inside the vehicle, there are nearly 100 million cars sold a year that will each need a high performance computer. We offer these OEMs, our Tegra SOCs and discrete GPUs along with our operating system and drive work acceleration SDK and libraries.

We -- when these partners go to production, we have the ability to support them with long-term software services that ensure they have a safe and secure experience for their customers over the life of their vehicle. We also offer application software to partners which gives them the ability to increase the revenue opportunity in their cars. Our business model here is to share in the revenues generated by the software we provide to our partners.

We're especially excited about this software opportunity as they can even be larger than the hardware opportunity in each vehicle. Now, on the infrastructure side, there are close to over 100 OEMs that need DGX systems for training OVX service for validation. And we also can offer them a range of software services. For example, Drive Replicator can be licensed to build up virtual vehicles and create synthetic data for our partners AV software development.

And with our recent acquisition of DeepMap, we have the ability to build map at scale for partners own AV stocks worldwide. Both our car and cloud to market is well, positioned to grow as the investment needed for L2 Plus is many times larger than the (inaudible) only cars. And truthful self-driving will also require an order of magnitude larger investment than L2 Plus. We have a large pipeline of wins in automotive that we have announced will be going to production in the next few years.

Our traction is strong across all the segments of the automotive market, we have one designs in 20 of the top 30 EV car OEMs.We're working with seven of the top trucking companies, eight of the top robotaxi companies and we help all of the leading OEMs with their infrastructure in the cloud. Beyond working with OEMs, we have an open platform strategy that we believe is a big differentiator for our partner ecosystem.

We have an inception team that targets all automotive startups worldwide. We work with many of the major universities who used to drive platform for their automotive and robotics education programs. We also partner with Tier-1 software providers, sensor and simulation partners to help them better develop their application on our platform.

We learn a lot from each of these engagements and use the learnings to make our drive platform roadmap even better for our future automotive partners. We announced last year that our automotive pipeline was $8 billion over the six-year period from fiscal 2022 to 2028. Orin has been a huge success with all the wins we have announced over the last year we're providing an update that our six-year pipeline from fiscal '23 to 2029 is now estimated at $11 billion.

You'll start seeing this ramp scale up this year with production of Orin and EVs from partners like Neo, Lee Aut,o X bang and SAIC's R brand. We'll see a bigger ramp in the following years as OEMs like BYD, Hhyundai, Volvo ramp up, and we'll even be able to scale even further when Mercedes And JLR, as well as our partners in the L4 commercial trucking, and robo taxi market scale beyond fiscal 2025.

Now, I would like to introduce you to Rev who's going to tell you about Omniverse.

Rev Lebaredian {BIO 21360517 <GO>}

It's well understood that modern AI is built with enormous amounts of data. Up until recently, we've relied on data captured from the real world and painstakingly labeled it by hand. The next layer of AI requires a scale, diversity and accuracy of data, that's impractical and in many cases impossible to capture through traditional means. The only way to produce the data we need is by synthesizing it.

Like humans and all creatures, AIs learn continuously from their environment. Babies learn how to perceive depth and identify objects by experiencing their environment. They learn the rules of the world, physics through experimentation. AIs learn in precisely the same way. They experience the world through the images, sound and information we feed them. They learn physics through continuous experimentation, trial and error.

But unlike AIs are born and raised inside a computer. The most natural place for them to learn and experiment is not in the real world, it's in a virtual world. In virtual worlds AIs can learn in super real time where one second of our time can be days worth of life experience. In the virtual world, they are free to experiment learning how to operate heavy machinery and drive multi ton vehicles without risk of physical harm.

Once an AI has learned a skill well in the virtual world, we can move its brain to a robot where it can operate in the real world. For this brain transfer to work, the virtual world must be indistinguishable from the real world. It must look, sound and feel the same. The rules of physics must match the real world closely. Otherwise, the AI will have learned poorly.

We have built Omniverse for this very purpose. Omniverse is our platform for building and simulating virtual worlds that are indistinguishable from the real world, leveraging the full might of NVIDIA's accelerated computing. Four key technological advancements have recently converged making the ideal conditions for the creation of Omniverse. First, with the introduction of NVIDIA RTX in our touring generation of GPUs, we transformed real-time 3D rendering from a system that produces images that merely look good into a physically accurate simulation of how light interacts with matter.

Previous techniques based on rasterization had hit a wall in terms of physical accuracy. But, with Ray Tracing, we know we can simulate all aspects of the behavior of light. Virtual world simulation is always been limited to relatively small computers

at the edge mobile devices, gaming consoles or gaming PCs. With recent advancements in data center GPU computing, we have the opportunity to leverage graphics supercomputers in the cloud running simulations that are too large and compute intensive for traditional computers.

Omniverse is designed as a cloud native and scalable engine that can utilize the full capabilities of the data center. Pixar invented universal scene description or USD an open source that in 2015. USD provides a common standard that allows us to describe virtual worlds with physically accurate pieces, that can be composed into a large virtual worlds. Omniverse is built with USD at its core enabling easy and lossless interchange of 3D data between the rapidly increasing set of tools and simulators that support it. USD is to Omniverse what HTML is to the to the 2D web.

In addition to creating a large market for world simulation, AI is the key to building a virtual worlds. The construction of high fidelity, physically accurate and large virtual worlds is currently limited to a small group of artists, who have spent decades mastering the craft of 3D design and visual effects and video games. Every nook and cranny of the virtual worlds, we enjoy in films and video games has been touched by an expert artist.

For 3D worlds to be as ubiquitous as 2D web pages, we need everyone to participate in the creation of such worlds. Fortunately AI has advanced to the point where we can train them to help us build virtual worlds, augmenting average people with skills that would otherwise take decades to master. AI will make creating virtual worlds, as easy as creating a web page is today.

All things that are designed and built by humans are typically first built in a virtual world. Bicycles, cars, bridges, factories are all designed with various CAD tools well before they are built in the real world. Physically accurate and extremely fast simulation is key to designing the best and most efficient products. We can quickly test many iterations of a design in the virtual world at a fraction of the cost of what it would take to build them in the real world.

Once the digital version of the product is complete its transformed into its real world duel, one built from atoms instead of electrons. In most cases today, that's the end of the road for the digital version. But, if we link the two manifestations digital and real, they can evolve with each other. We can capture data from the real world through IoT sensors and devices, and feed it into the digital model keeping the twins in sync.

Applying accurate physical simulation to the digital twin gives us incredible superpowers. We can teleport to any part of the digital twin just like we can in a video game and inspect any aspect of it reflected from the real world. We can also run simulations to predict the near future or test many possible futures for us to pick the most optimal one. We've built the Omniverse platform to unlock the full potential of digital twins, from design and creation, to physically accurate simulation, on our super computers.

And now, I'd like to welcome my good friend and colleague, Richard Kerris to tell us about the growth of the Omniverse ecosystem.

Richard Kerris {BIO 16447189 <GO>}

Thank you, Rev. We are super charging a huge ecosystem. Starting with developers, worldwide there are over 25 million developers and we see great opportunities to expand our Omniverse developer platform to encompass developers across all verticals that NVIDIA serves and beyond.

NVIDIA currently has over 3 million developers in our program, growing from just -- from 2.5 million from just over a little year ago. Omniverse is a robust and modern SDK. And with the recent release of Omniverse Code, there are now opportunities for hobbyists to professionals to create extensions, connections and even full-blown applications on the platform. And with artists and designers, that are close to 45 million creators in the world and that number is growing faster than ever, with the onset of growing demand for the content and world creations as we move into the next generation of the World Wide Web, commonly referred to as the meta verse.

Now a large percentage of these artists and designers are already familiar with NVIDIA, and are using partner applications that have been NVIDIA GPU-accelerated. Omniverse is a platform that extends and enhances existing work flows. Meaning, we don't replace partner applications, we bring new features and capabilities to them, like true to reality simulation, and real-time photo realistic rendering, essential features in the growing need for -- content for virtual worlds, and digital trends.

And with enterprise there are over 150,000 warehouses and over 10 million factories in the world today many of whom are moving to automation and digital twins. Matter of fact that global digital twin market is forecast to grow over 40% in the next five years. Omniverse is designed from the ground up as a platform for the enterprises who serve these industries, as you saw some amazing examples of in the keynote earlier today. From developers to artists and designers, we are super charging the ecosystem of Omniverse.

We have some great examples of early indicators of success. For individuals downloading and making Omniverse part of their workflow, we have had a growth of over 10x where they were just a year ago. Many of the leading 3D software applications across media and entertainment, architecture, engineering, construction and operations, manufacturing and industrial designs are connected to Omniverse with more coming every month. Plus, we're seeing a growing number of startup companies interested in using Omniverse as part of the platform for their work.

There are many other ways to connect to Omniverse as well, not just software applications, things like sensors, cameras, light out scanners, many of which are essential for digital twins, and we're seeing great growth with connections here as well. Over 10x where we were a year ago, with hundreds more on the horizon for the year ahead. And lastly, Omniverse is a compute engine is something that can be

licensed to power the next generation of software products for our NVIDIA partners. You saw the first one earlier today with the launch of LumenRT by Bentley Systems.

Bentley is leading provider of software and services to design, build and operate the world's infrastructure and they license Omniverse to power their next generation of I-20 applications. We are an active negotiation with other leading software companies looking to use the power of Omniverse for their product lines as well. These are some great examples of how it's going with the Omniverse ecosystem.

Now, with on Omniverse Enterprise, we are empowering a global network of Omniverse enterprise as a platform for them to sell and expand their product lines. We currently have over 65 partners worldwide and 30 of them have Omniverse demo labs being set up all over the globe for our enterprise customers. This is building on an existing strong (inaudible) foundation, which has been serving the artist and designer markets for many years with NVIDIA.

Omniverse is supercharging, thire already active networks. And we're seeing momentum with the leading companies making Omniverse as part of their work flow, such as BMW, Siemens, Erickson, and those we featured today in the keynote PepsiCo and Amazon. And we have over 700 more companies in our pipeline.

Omniverse can serve all these opportunities from individuals to the world's largest enterprises, because it's run on RTX systems from laptops to servers and even directly from the cloud as you saw announced today. So, Omniverse is for consumers, professionals, developers and researchers across all industries.

Omniverse enterprise software is $150 billion market opportunity. And we estimate the available market available market opportunity for Omniverse software at this $150 billion because it's based on two main use cases immediately in front of us. First, we estimate that there are 45 million designers, those designers that are creating an industry such as meeting entertainment, architecture, engineering and construction, manufacturing, and industrial design. Many of these are end-user customers for our -- products already. And Omniverse and enterprise can help them modernize their existing workflows.

And second, we have Omniverse opportunities for our digital twins with Omniverse. This will serve millions of factories, warehouses and fulfillment centers across the globe, and we already have active engagements with hundreds of them in these early days of Omniverse enterprise. We're investing in our Omniverse ecosystem, we're empowering our (inaudible) partner network, and we're building a long-term subscription-based model that will provide even more opportunities in the future. Omniverse software and chips and systems equals a $300 billion market opportunity and we're going to go get it.

Thank you very much. Next up is Jeff Fisher to talk about games.

Jeff Fisher {BIO 2373419 <GO>}

Thanks, Richard. Hi, everyone, I'm excited to have this opportunity to talk to you about our gaming business. Gaming is huge. The industry is on fire, the number of gamers, eSports athletes, creators and broadcasters engaged in new shared experiences is exploding. With 3 billion gamers, no one is asking if gaming is growing. The question is, how big will it get? And you can start by looking at generation Z. By most by most measures it's the largest generation ever.

When surveyed 80% of gen Z are gamers. And gaming is their favorite activity, twice as high as music, watching TV, movies or social media, Jen Alpha is up next. The 2022 game developer conference annual survey once again, ranked PC as the most important platform for developers. Looking at the openness and technology leadership of the PC ecosystem and the fact that we are adding 50 million to our ranks every year, we couldn't agree more.

And GeForce is a lot more than playing games. We estimate 80 million creators and broadcasters, who are designing, building and sharing their work. There are now 24 million twitch channels doubling in the past two years. Last year, there was 29 billion in YouTube ad revenue twice that of 2019. Minecraft the ultimate game fusing playing and creating, reached 140 million monthly active users in 2021, growing one and a half times over two years. And Minecraft content has been viewed over 1 trillion times.

3D creators are the construction workers of virtual worlds. Blender is the tool of choice among this growing class of 3D casual and pro creators with 14 million downloads 1.5x times more than 2019. Over the past 25 years, we have dedicated ourselves to building the best platform for gamers and creators. It is enjoyed by hundreds of millions of gamers on desktops, laptops, consoles and streaming from the cloud.

At the heart of our platform is our GPU and a history of revolutionary architectures, each delivering new innovations for developers to create amazing games and creators to do their best work. Our new GPUs are programmable. Hobbyists and professionals regularly discover new applications for them and crypto mining is one example.

On top of our hardware comes a massive investment in software. Game ready drivers is our commitment to the best possible gaming experience. Whether on PC or in the cloud, we will release an optimized driver with every major game to provide gamers maximum performance and stability. And we have now extended this commitment to creators, with our studio driver.

This investment also delivers new technologies, like DLSS, Reflex, Max-Q and NVIDIA Broadcast along with SDKs to enable a broad ecosystem. Our newest architecture, Nvidia RTX reinvented graphics, featuring real-time ray-tracing and DLSS AI image upscaling. Game developers, creator ISVs, even other GPU suppliers have gotten on board.

For eSports athletes, we introduced Reflex, removing latency between the game and the gamers. Over 20 million gamers are competing on ReflexAlt with Reflex on each month. There are now over 250 RTX-accelerated games and applications. And this comes at a time of strong gaming growth.

The pandemic introduced millions more to PC gaming. And we expect they're here to stay. Valve continues to highlight gaming momentum. Last year there were 30 million more gamers buying games on Steam. And the number of engaged, concurrent gamers on Steam has more than doubled in five years. This past weekend set yet another record.

The Epic Game store has shown similar strength, adding almost 100 million users in just two years. All of this and more has led to record results in our gaming GPU business. And with about 30% of our installed base on RTX, there are a lot more gamers yet to upgrade.

The opportunity grows when considering all the gamers on Steam and elsewhere who are not yet on GeForce GPUs. And one more thing I'd like to share, looking into the millions of desktop GeForce gamers, who we know have upgraded their GPU to a 30-Series, they are buying up. The GPU is offering more value than ever. Based on our data, they are spending $300 more than they paid for the graphics card they replaced.

Now let's look at gaming laptops, the fastest growing PC category fueled by gamers, creators and students looking to work and play from anywhere. This year we introduced our Fourth-generation Max-Q. Working with OEMs and CPU manufacturers, we use AI to instantly optimize the CPU and GPU for every workload. These are our thinnest and lightest laptops ever.

This year we have announced 170 new RTX 30 series laptops starting at just $799. Our gaming laptop business continues to deliver record growth, revenue units, and ASP and GeForce gaming GPUs are driving the overall consumer laptop market approaching 25% attached with plenty of room to grow. We estimate there are over 80 million creators and broadcasters who are fueling a crater economy in excess of $100 billion.

As more careers are built around content creation their tools become more important. We built NVIDIA Studio for these creators. NVIDIA Studio starts with RTX GPUs powering a range of laptops tailored to the needs of creators. On top of that is our Studio software stack that includes specialized drivers and dozens of SDKs, which accelerate over 200 of the industry's top creative applications. This includes Adobe Premiere. Adobe Photoshop, OBS, the number one broadcast app and blender the number one 3D design app.

And there is a strong connection between gaming and creating. Today, a quarter of GeForce gamers are also creating or broadcasting. And they value performance investing more in graphics than other gamers. There are also creators and

broadcasters expanding the reach of our platform, which likely well extends beyond those who share their profile.

RTX AI is making creation more approachable for everyone. It is easier than ever to create like a pro. Like nache AI pose estimation, you can animate a 3D character or avatar using your body, your body motions in a webcam or adobe substance that converts a photo into a 3D texture you can add to any design. NVIDIA Canvas is our app that lets you draw a simplistic image and use AI to create a photo realistic picture.

And of course NVIDIA broadcast powered by (inaudible) our popular application that uses AI to enhance your video streaming with features like background noise removal. With billions of devices unable to play the latest games and apps, it's no surprise that cloud gaming is projected to grow to over 100 million users in 2024. Our cloud gaming strategy is to offer an RTX gaming PC in the cloud through our own GeForce now service and expand the reach globally through alliance partners and third parties.

GeForce now offers user access to our most advanced RTX gaming GPUs starting at $999 a month. And like on PC, we want to offer a full stack of our latest gaming GPUs. So we recently announced an RTX 3080 tier for $1999 a month. GeForce now opens the PC gaming ecosystem to any client, including Android and iOS phones and RTX gaming rig in your pocket. There are over 1,200 games on boarded from Steam, Epic game store, Ubisoft connect and others.

And we are extending our footprint through alliance partners like SoftBank, LGU Plus and Taiwan Mobile, who operate the service regionally and help offer GFN to over 80 countries worldwide. Several marketing partners also feature GFN in their products and services, including LG, Samsung and AT&T.

Finally, we are offering RTX graphics, our game-ready software stack and cloud gaming expertise to third party gaming services like Tencent game metrics. And cloud -- and gaming in the cloud is not just for playing games. As announced today with GFN we will have an expanded opportunity to offer Omniverse running on RTX to every creator on every client.

Looking forward, the opportunity for gaming and graphics is almost endless. There are a ton of headlines about a multi verse, billions of dollars, a real estate, NFT's and crypto economies, but one thing is very obvious, gaming is leading us there and creators will build it. And we are just at the beginning. The graphics required to deliver a cinematic VR experience in a massive multi player physically accurate world will likely require three to four orders of magnitude more than the performance of our highest end GPU's, plus continuing advancements in algorithms for rendering , physics AI and animation.

There are three billion gamers and creators and it's growing. We believe over time a quarter of them will spend over $100 a year for high performance GPU's in desktop,

laptop, cloud or console. In total, this translates to a $100 billion opportunity. The fundamental strength of gaming has never been stronger.

I will now turn it over to our CFO, Colette Kress.

Colette Kress {BIO 18297352 <GO>}

Fiscal year 2022 was a record-breaking year for NVIDIA with revenue, gross margins, operating income and earnings per share all achieving records. Revenue increased 61% year-on-year to $26.9 billion, driven by our incredible ramp of our a architecture across our graphics and data center platforms. We achieved record revenue in gaming, data center and professional visualization.

Gross margins increased 120 basis points year-on-year to 66.8%, as we benefited from gamers, and craters buying up our stock. Gross margins expanded against the backdrop of industry-wide supply chain disruptions and rising costs. This speaks to the strength of our business model and execution. We drove strong operating leverage, as operating income increased 87% year-on-year and $12.7 billion and earnings per share increased 78% year-on-year to $4.44.

I'd like to talk about our market platforms and the opportunities that we see ahead of us. Starting with gaming. Fiscal year '22 was a phenomenal year with revenue increasing 61% year-on-year to 12.5 billion driven by broad-based strength across desktop, notebooks and consoles. Strong demand for RTX in our a GPU's, help drive tremendous unit and blended ASP growth. This is consistent with what we have observed over time.

Gaming revenue has grown at a four year 23% compound annual growth rate, with both units and ASPs contributing. We see these trends continuing in the future. The universe of gamers continues to expand and the creator, economy will be further turbo charged with Omniverse now available for individuals. With RTX enabled content, nearly ubiquitous and over 70% of our installed base yet to upgrade to an RTX GPU, we see a tremendous revenue opportunity ahead of us.

Our Max-Q technology transformed the notebook PC into a new gaming device and unlocked for us one of the fastest growing and largest PC gaming markets. Longer-term, GeForce now expands the reach of our GeForce platform to billions of gamers. NVIDIA is the only gaming platform to address every way a gamer plays desktop, notebook, console and the cloud. Over time we see a 100 billion available market opportunity and a long runway for growth.

Turning to data center. Fiscal year '22 built upon the great momentum we saw in fiscal year '21 with revenue increasing 58% to $10.6 billion and exiting at a $13 billion run rate. Strong and broad-based help from the A100 helped fuel strong revenue growth in hyperscale and vertical industries, a natural language understanding and deep recommendator models are uniquely enabled by our full stack approach. The diversity compute intensity and latency requirements of these models also helped to

drive accelerating growth in inference revenue and widespread adoption of our Triton Inference Server download it by over a million times by 25,000 customers.

Our data center business has grown at a four year compounded annual growth rate of 53% and we entered fiscal year '23 with great visibility into demand and supply. Announced this morning the H100 GPU and the Hopper architecture is set to build on an incredible success of the A100, and the ampere architecture. The H100 has an engine to speed transformer networks, the most important deep learning models invented. These models are helping to give rise to an emerging type of data center, the AI factor (inaudible) and where data center is the raw materials and intelligence is the end product.

These factories require large amounts of compute and networking, and a full stack approach for both training and inference. Our Grace CPU is perfect for these environments. Finally, our BlueField DPU rounds out the three chip strategy and is set to ramp this year with interest high, among our CSPs and major OEMs.

We also see strong interest in support for graphic, intensive workloads from car gaming to virtual worlds and industrial digital twins. Just as DGX runs on NVIDIA AI software for processing, machine learning and deep learning workloads our just announced OVX server will run on Omniverse software for processing industrial digital twins. This is a perfect example of our ability to extend our platform and add new growth factors of growth. Our large ecosystem continues to grow helping to unlock new markets and faster adoption of our platform.

Turning to professional visualization. Fiscal year '22 revenue increased 100% year-on-year to $2.1 billion. We saw strong demand from hybrid work-related deployments and the ramp of our ampere GPUs into workstations. RTX and AI have completely revolutionized the computer graphics and can drive continued growth as adoption expands within the estimated enterprise end user base of about 45 million creators and designers.

Omniverse adds a new tremendous new growth opportunity and interest is high with some of the world's leading companies such as BMW, Siemens Energy, Ericsson developing in Omniverse. We have more than 700 companies in the pipeline. Not only does Omniverse present a large software opportunity, but it also will help drive a large hardware opportunity as Omniverse runs on NVIDIA RTX powered desktops, laptops, workstations and servers.

Finally, turning to automotive. We believe this will be our next multi-billion business and we're on the cusp of an inflection. Autonomous driving is a significant technological challenge and NVIDIA uniquely enables the entire workflow. This comprehensive yet flexible approach is helping to drive rapid adoption of our drive platform and Orin SOC across the transportation industry unlocking new business models for us and our customers.

Our design win pipeline measured over the next six years is now 11 billion, reflecting our great momentum with new NAVs, traditional OEMs, truck makers and robotaxis. This is up from $8 billion a year ago. Our opportunity is large and we expect our revenue momentum to build in the coming quarters and hit an inflection point in the second half of this year.

Let me talk about our software strategy. As you know, software has been integral into our platform for over 15 years since the introduction of CUDA. It helps us enable and create new markets and is a key competitor differentiator. So far software has been largely included as a part of our broad platform offering and not sold standalone. And we've been offering some standalone software and services including VGPU subscription for professional graphics, GeForce now subscriptions for cloud gaming, as well as software support.

All in these types of reoccurring software and services revenue are currently at an annual run rate in the low hundreds of millions. Building on this foundation, we see a much larger software revenue opportunity going forward across three key opportunities. First, the NVIDIA AI enterprise software suite brings NVIDIAs AI tools and SDKs to enterprise IT. It is offered as an upfront license plus maintenance or as a subscription and it is available through our broad channel partners.

Second, our Omniverse enterprise software platform enables collaborative product development and operation of digital twins. Omniverse is offered to enterprise customers as an annual subscription with additional licensing opportunities for the operation of digital twins. Third, drive software revenue will be enabled by our end to end and full-stack autonomous driving platform as described by Ali.

This new business model can drive one of our largest revenue opportunities as millions of vehicles become software-defined and capable of delivering software and services over their lifetime. Each one of these software offerings has a multi billion revenue potential and to contribute positively to our gross margins over time. Our software go-to-market leverages many of the relationships in existing channels, we have built over decades. This will help us scale these businesses with an efficiency unique to NVIDIA.

I'd like to spend some time discussing our long-term available market opportunity. Keep in mind, this is not a reflection of our TAM in any given year, as adoption and penetration curves can vary. For example, cars have a much longer refresh cycle than PCs and servers, so the auto opportunity will take longer to realize. For brand new businesses like Omniverse, it's hard to predict the pace of adoption, but we can help sighs the overall available opportunity based on the number of gamers, designers, engineers, servers, cars and other devices that we can power with our technology.

For gaming, we see a total available market opportunity of $100 billion, as we can reach gamers, anyway they play. There are 3 billion gamers globally growing every year and we believe a quarter of them can be addressable over time. Whether we

serve them with GeForce GPUs in their systems or GeForce now in the cloud, the per the per-user pricing is similar and analyzes at over $100 per year.

For chips and systems, we estimate a total available market opportunity of 300 billion. This spans all 3 processor types, GPUs, DPUs and CPUs, as well as networking infrastructure to power AI graphics and high-performance computing, from the cloud to the edge including public and private clouds, enterprise and edge locations, and workstations.

Our estimates assume that all servers overtime will address the GPUs and the DPUs at a portion in the high-end will be addressable with our CPUs. As Manuvir described in a new era of AI and virtual worlds, we believe companies will direct a greater portion of their capital and operating expense budgets to their technology infrastructure spent. Running on top of our hardware stack are our enterprise software offerings. Enterprise, AI and Omniverse enterprise.

For NVIDIA AI Enterprise, we estimate the total available opportunity at 150 billion based on the installed base of enterprise servers and our per server software pricing. For Omniverse enterprise, we also estimated 150 billion software opportunity based on two opportunities. First, a per seat software subscription for professional designers and craters which we estimate at 45 million. And second, upper robot software subscription for digital twins based on more than 10 million factories and warehouses.

Finally for automotive. Our opportunity reflects three components. The drive software for autonomous driving, the in vehicle hardware, and the data center infrastructure for training and simulation. The vast majority of the estimated 300 billion opportunities comes from software for two reasons. First, our software content per vehicle can be in the 1,000s of dollars over the lifetime of the vehicle, compared to the 100s of dollars for the hardware.

And second, software scales with the installed base of vehicles, not annual production. So in total, we see a 1 trillion available market opportunity in front of us. We believe our opportunity will increase in time as we roll out new products and offerings, unlocking new markets that previously were not available or did not exist.

We have been investing significantly to address this opportunity. Fiscal year 2000, we have invested cumulatively $29 billion in research and development, growing at a 24% compounded annual growth rate. We have also invested significantly in capital expenditures 1 billion cumulatively since fiscal year 2010, growing at a 24% compounded annual growth rate over this same period of time.

We have also extended our platform and added talent through a handful of acquisitions. Given the wave of new products and offerings discussed today and the opportunity ahead of us, we will continue to scale investments to support continued strong revenue growth.

Our full stack approach not only is the key competitive differentiator, but enables us to innovate quickly and profitably. We have architected our software to work across all products, systems, platforms and applications. This single platform allows us to develop and launch new products at a rapid pace and efficiently enter new markets as each innovation takes advantage of NVIDIA's enter entire body of work and go to market capabilities. This approach enables a business life like no other, allowing us to invest at scale with confidence while also driving strong operating leverage.

Compared to fiscal year '18, fiscal year '22 operating income increased 3.5x to $12.7 billion. Over the same period of time, operating margin has expanded 1,000 basis points to over 47%. We will continue to balance investing for growth with driving operating leverage over time. With our software rich business model and inherent operating leverage, we generate a lot of cash.

Free cash flow has grown significantly at a four year compounded annual growth rate of 29%. Free cash flow growth accelerated in fiscal '22, almost doubling to $8 billion. We anticipate significant growth in our free cash flow over time.

Let me update you on our capital allocation priorities. First, after pausing for over a year we resumed our stock repurchases this quarter. We've repurchased $2 billion. We have $5 billion remaining under our authorization through calendar year end. Second, we plan to maintain our dividend, which is currently a use of cash of around $400 million per year. And third, we will continue to make strategic investments, where it makes sense to grow our talent, platform reach or our ecosystem. Note however, that our number one focus will continue to be investing organically for growth.

I'd like to close with our commitment to ESG, NVIDIA's building one of the world's greatest companies, by focusing not only what is good for business, but is what is good for our employees, our partners, the environment, and society at large. We have strived to create a company and a culture, where employees will want to come and stay and do their life's work. We were ranked number one on Glassdoor's best place to work list for 2022.

We are building Earth-2, the world's most powerful AI supercomputer dedicated to predicting climate change. And we are committed to strong corporate governance with NVIDIA to receiving a number of recognitions for the strength of its management team and diversifying our board.

That wraps up our presentations for today. Please enjoy this short video that we have for you and then we'll move to the Q&A portion of our event.

(Audio-Video Presentation)

Questions And Answers

Operator

(Question And Answer)

We will now start the Q&A session. (Operator Instructions) Our first question comes from Aaron Rakers from Wells Fargo. Aaron, please click okay on the pop-up and you will be admitted to the room. Please unmute yourself and start your video using the buttons in the bottom there, online.

Q - Aaron Rakers {BIO 6649630 <GO>}

Can you hear me?

A - Jensen Huang {BIO 1782546 <GO>}

Yes.

Operator

Okay.

Q - Aaron Rakers {BIO 6649630 <GO>}

Sorry about that. Thanks for doing the detailed presentation. I guess correct the thing that most notably stands out that you're really kind of leaning in on sizing the software opportunities that the company is developing and discussing as far as the PAM opportunity. So I guess my question to you is that how do we, as investors, think about the progression of -- I think you said $100 million or so ARR in your comment, how do you define success of that? And where do you think that the earliest success would show up? And does it become a material driver even looking through this next fiscal year? Thank you.

A - Colette Kress {BIO 18297352 <GO>}

Great. So let me start off, Aaron. Great question to start on the software. We articulated today so many other software opportunities available. We're talking definitely about three key areas for the enterprise, NVIDIA enterprise AI software, Omniverse as well as DRIVE. But correct, today, already, we have been selling software to our enterprises, and this is a couple $100 million today. And we believe this is a growth opportunity for us, but a growth opportunity in many ways, not just on the software line, but the infrastructure that will be important in terms of building this out as well. So they'll both go hand in hand, but we do believe it's an important growth factors as we go forward.

I'll move Jensen if he wants to add more.

A - Jensen Huang {BIO 1782546 <GO>}

Well, Aaron, the thing -- the important thing about our software is that it's built on top of our platform, meaning that it activates all of NVIDIA's hardware chips and

system platforms. And secondarily, the software that we do are industry-defining software. If you understand -- you understand well that NVIDIA AI is a collection of libraries that make it possible for you to do data processing, to machine learning, to deep learning, to inferencing at hyper scales. In the cloud, there are large engineering organizations that help the clouds do that themselves. But for the world's enterprise, you have to do it with them and for them and help them maintain this really complicated AI engine across multiple platforms and multiple generations of platforms. And so the amount of value that we've encoded into NVIDIA AI over the years is really quite tremendous. And that's just one.

Second, as I mentioned, when you start to move into the edge or the industrial edge or what people call robotics, those systems required the simulation, a digital twin, if you will, that models your products out in the field, because if you can't do that then you can't develop new software, optimize the software and very importantly, do what is called continuous integration and continuous development and integration so that you could deploy the software into your fleet. You have to be able to simulate the results of that deployment before you deploy it. And so the people are really coming to grips with the idea that if you want to deploy AI out into the edge, if you want to put robotics out into the world, you really need this concept of a digital twin.

We're years ahead of the industry in this, and because it leverages NVIDIA's entire body of work, Omniverse is really an industry-defining piece of software. These two products, as you know, are one of a kind, and it runs on top of our platform, enables AI to go to the world's enterprises into all of these industries, and so that's really the reason why we've productized them, built internal organizations to be able to productize them, support them and deploy them over time, and we're really quite excited about it.

Q - Aaron Rakers {BIO 6649630 <GO>}

Thank you very much.

A - Jensen Huang {BIO 1782546 <GO>}

Thanks, Aaron.

Operator

Thank you, Aaron. Our next question comes from C.J. Muse at Evercore. C. J., please okay pop-up and you will be admitted to the room.

Q - Christopher Muse {BIO 18608702 <GO>}

Hey, how are you? Thank you for today, really appreciated. I guess my question is on the hardware side. I think you've tripled the size of the TAM there from $100 billion to $300 billion. And I'm curious if you can kind of walk us through what you're seeing from a core GPU perspective in terms of increasing the size of that TAM as well as what kind of assumptions you're making around penetration for both Grace on the CPU side as well as on the DPU side? And I guess lastly, the synergies that you

see from offering all three pieces of silicon and how that can drive overall revenue growth as well? Thanks so much.

A - Jensen Huang {BIO 1782546 <GO>}

Yeah. Thank you, C.J. First of all, remember that or note that we have GPU, CPU, and the DPU, what the Mellanox architecture, the Mellanox platform if you will, all three platforms are unique in their richness of ecosystem. These are not just three chips. There are three platforms. And the body of work and software that's on top of each one of them and the ecosystem of systems and servers and computers and the partners and the go-to-market partners and all the third-party developers, everybody that's working on these three platforms is really, really unique. So I'm delighted to have three of the most important data center chip technologies in one company. However, these three platforms are wonderful all by themselves, all individually.

There are several different growth drivers for today's GPUs. The number one, of course, by far, is AI being put into operations all over the world. Inferences for recommender systems, conversational AI, speech AI, the number -- a natural language understanding the number of new models that are based on deep learning is just growing exponentially. And this is really the modern way of doing software development and there's no question at this point that what has now taken over the vast majority of the cloud will go forward into all of the world's enterprise. So that's number one driver, AI with all of the different models for training and for inference.

The new thing that we -- and I think in our last conference call, we said several times that our visibility of data centers through the year is really, really excellent and that is just driven by today's continued expansion, continued use of deep learning. The new growth drivers that we talked about today, there are couple. We spoke about of course the Grace CPU Super Chip. No CPU has ever been designed this way, and two very, very powerful CPU dies that are then connected using NVLink, 900 gigabytes per second NVLink, that's memory coherent, makes for one super CPU chip.

And so this particular CPU is going to be really great for moving data processing data, which is really consistent with all of the core business of our company, AI, scientific computing, and in the future, Omniverse digital twins. And so all of these applications are going to benefit from a CPU that's not just incredibly good at single-threaded processing but very importantly moving data. It's not 50% better or something like that over the best today, but it's many times more memory bandwidth than what's available today. And so that's a new business driver for us.

I also spoke about Spectrum 4. You know that our business in NICs and endpoints, whether it's SmartNIC or the BlueField-4 -- BlueField-2 DPU is doing fantastic. I'm going to add to that with Spectrum 4, which is a 400 gigabits per second ethernet switch that in combination with CX-7 and BlueField 3 turns it into an end-to-end 400 gigabits per second Ethernet platform. That's going to be a major new driver for us. We're super successful already with InfiniBand. We're super successful with end-to-end InfiniBand. This is going to be a new journey for us, and I'm super excited about

it. The performance is unrivaled and the software stack on top of it is incredible. So we have a new data center driver with CPUs, new data center driver with ethernet switch end-to-end platform, and so we should have some pretty exciting times ahead for data center hardware.

Operator

Thank you. Our next question comes from Vivek Arya from Bank of America. Vivek please consent to join the virtual room. It looks like Vivek has -- having some issues, so we will return to him in a moment.

Our next question comes from Matt Ramsay at Cowen.

Sorry, there is Vivek.

Q - Vivek Arya {BIO 6781604 <GO>}

Hi. Yes. Sorry. Thank you, Jensen, thanks for listing, I understand. Really appreciated. Actually, Jensen, I wanted to go back to the Grace CPU -- server CPU. So from what you're suggesting, you're only targeting the high end of the market, and I'm curious why only limit yourself to the high end of the market? Why not go after the cloud and the broader enterprise market as well? What's stopping you from doing that, because do you not leave x86 competitors who can kind of come up the stack and continue to challenge you at the high end of the market? So that's kind of part A of the question. And I thought I heard you say that you're using the off-the-shelf kind of the new worst scores that ARM has developed. Do you have any plans to do your own custom implementation of those cores over time that can give you a bigger competitive advantage in that market? Thank you.

A - Jensen Huang {BIO 1782546 <GO>}

The answer to the second question first. There are more surprises for Grace that will be coming out. And we'll have plenty of time to describe all the characteristics of Grace over time. And so today, I thought we would focus on the super chip architecture, and it is such a sensitive fundamentally different way of designing chips and systems and it provides incredible capabilities for us to modularize and combine and create different types of systems to diversify the platform in a lot of different ways. So the number of different types of configurations that you're going to see from Grace Super Chip, Grace Hopper and Hopper and CX-7 and BlueField 3, the combination with -- of those chips with the switches that are behind them, the combinations and the configurations of systems are going to be pretty staggering. And so I'm super excited about that, and I'll talk to you -- we'll describe more about that over time.

With respect to the target market of Grace, the area we're most focused on and we're -- first of all, the CPU core is incredible that as you saw that our estimated spec-in performance is off the charts compared to what's available today. And so the CPU performance is fantastic. However, what really distinguishes Grace is a couple of things. Its memory bandwidth is unrivaled. The memory capacity and the memory

bandwidth that available on that capacity is like nothing the world's ever seen. And second, the energy efficiency of the entire CPU subsystem, which includes the CPUs and all of the memories associated with it and all the securities. The energy efficiency is probably about two times may be more than what's available -- what will be available in the market at that time. And so that's a giant leap in those couple of factors.

The areas where we're going to focus Grace initially, and as you know, we'll have plenty of time in the beginning of our journey into providing discrete CPUs. And we'll have plenty of time, and the market for discrete CPUs is quite segmented and quite fragmented. And so we have to respect that. The areas where we're going to focus are also happens to be the fastest-growing segment of CPUs, the fastest-growing segments in today, which is AI infrastructure. We -- as you know, we're one of the fastest-growing data center companies in history, and yet all of that data center growth is rather new. This idea of an AI factory is a new thing that came about because of AI. This is a data center that most companies historically didn't have and many companies in enterprise still don't have.

And so as we grow into this new class of data centers, call AI factories or AI infrastructure, this is an area that would really want to focus Grace on. You could use it for training very large models. You saw earlier that in training large language models, Hopper is going to be some order of magnitude higher than Ampere. The way to think about that is no one really builds data centers, AI factories with more than a couple of 1,000 or 4,000 GPUs today. Well, you can now extend that to 10,000, 20,000, and the reason for that is because the efficiency, the utilization of the processors now made possible by the new architecture of Hopper and all the interconnects makes it possible for you to scale up your infrastructure so that you could do the training of these really valuable models from weeks to days to hours. That's just game-changing. And so this is one way that we're going to scale up.

The other way that we're going to scale out is that the ability for us now to build very, very dense, Grace Hopper, as you see, is incredibly dense. It's the most dense AI inference computer the world has ever seen. One, an incredibly dense server in just one super chip. That one dense server replaces about 14 servers, each instance replaces two T4s, and so that seven instance times two, 14 servers can be replaced by this 1 single super chip. So whether it's AI infrastructure for training large models or AI infrastructure for large-scale deployment of AI, we're going to have plenty of market to go after, just and just really, really giant markets to go after. And so that's where our focus is.

Thank you, Vivek.

Q - Vivek Arya {BIO 6781604 <GO>}

Thanks.

Operator

Our next question comes from Matt Ramsay from Cowen. Matt, please join us in the room. Matt, please unmute yourself.

Q - Matt Ramsay {BIO 17978411 <GO>}

Yes -- Yes. Thank you very much, Jensen and the whole team for a very helpful day.

A - Jensen Huang {BIO 1782546 <GO>}

Thank you.

Q - Matt Ramsay {BIO 17978411 <GO>}

I wanted to ask a couple of quick follow-up questions on the software business because that was something that was of emphasis and new today. The first one is quick. Colette, you mentioned a couple of $100 million today. You shed any light on the growth rate of that number in the recent periods.

And Jensen, the longer-term question, irony, the Omniverse is, I guess, come in to the investor Lexicon about your company over the last six, nine months, but some of the work that we've done, I think I'm a bit clearer on the TAM and the ability to potentially monetize the Omniverse than I am, maybe on the enterprise AI opportunity in software for your company over time. Maybe you could talk a little bit about how you guys put that TAM together? I think some of that is priced on a per CPU basis today because you kind of meet in with the VMware pricing model, just the inputs of how you're thinking about the sizing that enterprise AI TAM for the company, revenue per seat, revenue per CPU, just examples of how you're going to penetrate that market over time? Thank you.

A - Colette Kress {BIO 18297352 <GO>}

All right. Well, let me first start, Matt, on the question on what do we think it will grow? What do we think software will grow moving forward? It's an important part of what we're planning in both this next year and a decade going forward. Probably the best way to think about the growth is the growth that we'll see in enterprise. Enterprise overall hardware, overall systems, and seeing software being an important part of that complete stack that we're going to be needing for them. So it's tough to say how much will the enterprise growth be versus some of our other components. But I think it will track quite well to what we'll see in enterprise.

A - Jensen Huang {BIO 1782546 <GO>}

NVIDIA AI, remember what's inside it. There are several stages in building an AI model and deploying an AI model. The first stage is processing the data. It's amounts of data, terabytes of data, petabytes of data, just incredible amounts of data. You have to find a way to refine that data, process that data, refine the data, get clean that data, augment that data. There's just a lot of it's related to sequel when it comes to structured data. When it comes to unstructured data, a lot of it is image processing, signal processing. You're doing a lot of processing of data. Number one.

Number two, you have to do feature engineering try to figure out what our predictive features. Number three, whether you're using -- if you're using classical machine learning, which is the vast majority of the industry today graph analytics, all of that you would like to do 1,000 times faster, 1 million times faster because the amount of data that you have is just torrentious. And so number three is machine learning.

Number 4 is deep learning, and deep learning is where TensorFlow comes in, where Pietorch comes in. And then when you're done with that, you have to deploy it, inference. And so that entire workflow is unlike any software development that's done today. The vast majority of the world software development until now has been human writing coding, testing it against some data set or some test suite, and then deploying it. That's the vast majority of development today. It's done by humans developing software and laptops. And yet in the future, the way software is going to be developed is engineers developing software and laptops, but connected to supercomputers in the back. If you look at the number of -- the amount of infrastructure per software engineer at the largest Internet companies or NVIDIA, you will see that the amount of computational infrastructure beyond the laptop is enormous, that's how machine learning is done. That's how AI is done.

And when you're doing this in the cloud, when you're doing this in the hyper-scale companies, they have a lot of engineers who could do that. For the rest of the world's Fortune 100, Fortune 5000, the other 100,000 companies around the world who needs to do this and would like to do this either on-prem or at the edge, somebody has to go develop that software suite. Somebody has to bring the NVIDIA AI software engines that are running in the cloud today and putting it into on-premise, and it's really a body of software that's really quite complicated and end-to-end. And so that's number one.

Now how many companies in the world will be doing data processing, feature engineering, classical machine learning graph analytics to deep learning? Well, I happen to believe that every company -- every company's fundamental production, fundamental output is intelligence, a recommendation for our financial strategy, a recommendation for some health regimens, some recommendation for therapy, some -- it's a recommendation. And so in the future, almost every company will be a tech company and every company will be an AI company producing intelligence.

If so, then every company servers will have some part of this pipeline running on it. And you run that pipeline, if you want to run that pipeline, you want to run it well with -- on an enterprise, we have a library and engine. The engine has a suite of libraries but an engine that allows you to run it on every server. There are about 50 million installed servers in the world's enterprise today. It's going to be a lot more in the future, especially as we move on to the edge. But 50 million installed today. If every single server and the way you count a node is the CPU inside and that's why we use CPU, but basically a node. For every single node, if you want to run NVIDIA AI, we have an engine for you and that engine is proper CPU per node, it's $2,000. So 50 million, $2,000.

On top of that is all of the NVIDIA SDKs all of the other AIs and AI frameworks, maybe it's an recom -- it's an AI framework just for recommenders and AI framework just for speech or AI framework just for large language models or AI framework just for computer vision or robotics or whatever it is, we're going to have a whole bunch of software on top of that, but they all on top of that one engine NVIDIA AI. So I just gave you the NVIDIA AI story.

The NVIDIA Omniverse story is really about connecting designers and artificial intelligence. It's about connecting designers and artificial intelligence. The artificial intelligence could be a self-driving car. It could be a robot that's roaming around inside a logistics warehouse, 1 of the 100 million square miles of fulfillment warehouses around the world, they're going to be -- they're just too big for humans to walk it. So you're going to have a whole bunch of AMRs move stuff around. And so those -- all of those AMRs are going to be sitting in digital twins. You have to have a digital twin because you're going to reprogram the AMRs.

And when you want to reprogram the self-driving car fleet or the AMRs or the pick-and-place robots or the last-mile delivery -- pizza delivery robots, grocery delivery robots, when you want to reprogram them, optimize the software before you do it, you want to see how that software build is going to do in the real world. You don't want to just develop software and roll it out and hope for the best. You want to simulate it somehow in virtual reality, virtual worlds. We call that Omniverse, Omniverse is an engine for you to simulate all these different types of robots. And so designers, robotics, AI developers are going to all be connected into this virtual world, and they're going to develop software, optimize the fleet, optimize the factory, and when they're ready, they deploy it.

The way that we benefit from Omniverse is the connections of the robots and the connections of the designers, and hopefully, I would expect that, in fact, more things will be designed in Omniverse long term than the physical world because you have many versions of cars and houses and cities and buildings and factories and so on and so forth, the number of designers that are connected to it, hopefully, starts from the $50 million today, and hopefully, it's a lot more. And the number of robots there, I think, it's fairly clear now that the world will have billions and billions of robots, not humanoid versions like us, but they are autonomous robotic systems that are moving around, they could even be medical imaging systems, surgical systems, AMRs, and so on. And so we have two different industry-defining software platforms, NVIDIA AI and omniverse. They have different business models because they're used in different ways. And they -- both of them leverage all of our platforms, which means the entire network go-to-market that we've developed over the years are super excited to take these two platforms out to market with us.

And so we have large channels already built up. We have a large network partners are built up. We had a large number of third-party software developers that are hooked into it. And so -- so these two platforms, I think, are going to be -- that's one of the reason why we're focused so intensely on these two. And then one more thing. With respect to NVIDIA AI, remember, I think, Matt, it was you that in the beginning, why now that we're -- and just now we're going into it. Remember,

NVIDIA AI runs on NVIDIA Gear. Even though a lot of software, even a lot of software runs on CPUs, a lot of its most important features can only run on NVIDIA's hardware. This is the groundbreaking work that we did with and our GPUs and so on and so forth. And so now is really quite ideal because we've had several years about six or seven of building an installed base of NVIDIA hardware in the world's enterprise. And remember, software wants installed base. And so we have the benefit now of going to market with a known large enterprise installed base, and that installed base hopefully doubles every year. And so that's the plan.

Q - Matt Ramsay {BIO 17978411 <GO>}

Thanks for all the doubts, Jensen.

A - Jensen Huang {BIO 1782546 <GO>}

Thanks a lot, Matt.

Operator

Our next question, apologies, comes from John Pitzer from Credit Suisse. John, please join us in the room. John, once you've joined us, please turn on your audio and video.

Q - John Pitzer {BIO 1541792 <GO>}

Perfect. You guys, can you hear me, okay?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Nice to see you, John.

Q - John Pitzer {BIO 1541792 <GO>}

Perfect. That's a test. Thanks, guys. Along with everyone else, thanks Jensen for all the information today. I'm kind of curious if you could talk a little bit about the transition from Ampere to Hopper? How quickly do you think that's going to happen in the last couple of GTCs as you brought out new products, especially in the data center, incremental performance gains are not measured in percentages as much as multiples. And so is there a risk that A1 demand falls out more quickly than A100 are then you can ramp Hopper and then collect maybe as a back half of that question, you reiterated, I think, last quarter, gross margin progression every quarter this year despite all these new product introductions. I'm just wondering if that's still the case?

A - Jensen Huang {BIO 1782546 <GO>}

We have excellent visibility into our data center business because of the breadth of AI products that we offer and the number of services, AI services and applications that are built on top of it that the world provisions for their own business. It is the case that when we launch a new consumer product, the transition is rather crisp. However, because the world's enterprises and the cloud service providers, they are running their business on top of Ampere today, they've got their businesses forecasted out for some time and their expansion forecasted out for some time.

They're going to keep on building it because they -- every single system they put in place, provisions more services and more growth and more customers. And so they're anxious to get that in place, and they want stability and security on the forecast they give. That's one of the reasons why we have so much visibility today.

Now when we first started in the data center going from Kepler to PASCAL, it was super spotty. It was because we -- the number of applications on top of our GPUs wasn't that many. And when we went to Volta, that was really still kind of the beginning. But Volta built a great base, Ampere built a phenomenal base and now the number of deep learning services that are sitting on top from imaging to video to language to speech to recommendation systems, just the recommender systems that drives the world's commerce on the Internet, the number of recommenders in the world. I mean it's not one recommender per company, it's hundreds of recommenders per company that recommend products and ads and things like that, right? They're recommending all kinds of things. That is so vital to their business. They forecast it out, they planned that out and that gives us the visibility we need. Okay.

A - Colette Kress {BIO 18297352 <GO>}

So when talking about gross margin, moving forward, we've done a tremendous job with gross margin up to this point. We're probably looking even this quarter at 67%, and we know that the future in front of us is going to incorporate software, software stand alone which will assist our gross margins. Our products and systems that in the data center can also help influence and the right mix of growth can also influence our gross margins as well. So we'll stay focused on margin going forward and looking from the growth from software to probably be one of the largest drivers that will increase our gross margin.

Q - John Pitzer {BIO 1541792 <GO>}

Thank you.

A - Jensen Huang {BIO 1782546 <GO>}

Thank you.

Operator

Our next question comes from Stacy Rasgon at Bernstein. Stacey, you should see a pop up please and join us in the room.

A - Jensen Huang {BIO 1782546 <GO>}

Stacy, please join us and unmute.

Hey. Stacy.

Q - Stacy Rasgon {BIO 16423886 <GO>}

How are you?

A - Jensen Huang {BIO 1782546 <GO>}

Terrific.

Q - Stacy Rasgon {BIO 16423886 <GO>}

Great. I have two questions, one on longer-term and one on the shorter term. For the longer term, the $300 billion in chips and systems opportunity in data center, can you give us some feeling for how that -- how you see that breaking out between the enterprise side and the hyper-scale side? And I guess more generally like where is it -- where does all that come from to the entire server arm today is, what, $100 billion, maybe networking is about the same. I guess what's just underlying that $300 billion? And how does that split out between enterprise value?

A - Jensen Huang {BIO 1782546 <GO>}

Long term, I expect enterprise and edge to be bigger than hyper-scale. I believe that there will be not just hundreds of data centers in the world, but there will be millions of data centers in the world. And I believe millions of data centers will be out of the edge, and they have to be built, designed, orchestrated like it's a cloud computer, but it's all over the place. So that you can ensure guarantee nanoseconds, surely less than a millisecond of latency, and guarantee that service every single time, not best effort, no excuses doing high traffic times because there's an industrial application connected to it.

They're robotic applications, working hand-in-hand or machine-to-machine and they're communicating with each other, and they just can't afford to get behind a test drop of some new show on Netflix. I mean I just -- that can't happen. And so they're working together, they're doing important things, humans are working among them, and that world, you need data centers right at the edge. And so long term, I believe that hyper-scale will continue to be very, very large, of course, and it's going to continue to grow from here and the industrial edge will be quite large.

I also remember that NVIDIA is not a chip-only company. We're a chip and systems company. We build some of the world's largest systems. And those largest systems are not one-off supercomputers for a particular nation, but they are supercomputers that are built as AI factories. You saw recently a very large company announced a very large installation of an AI factory. And it's really about processing data and try to refine the data and trying to produce the most valuable commodity that we have, which is intelligence.

And we now have the ability as a computer -- as a form of information science to be able to harvest data, to process data, and turn it into intelligence invaluable intelligence. And so I believe that kind of data center the DGX SuperPOD type of AI factories are going to continue to grow. It's already been spectacularly successful. And I think we're the only company in the world that builds that. We offer the blueprint to all of our partners so that everybody could build it, but of course, we build it ourselves as well.

And so the second thing is that, remember, inside it, our systems. Now our systems are, of course, has CPUs inside, which is a discrete CPU, which is a brand-new growth opportunity for us. We have next inside the hyper-scale cloud called CX-7. We have smart at the edge of the called BlueField, okay. And we have the switches that connect everything together, and we have three types of switches. Those three types of switches connect basically the end to end of an entire modern data center. At the core, where the nodes want to be connected, we have this brand-new MV switch, a new class of switch that doesn't exist anywhere on the planet. Second, we have InfiniBand switch to Quantum platforms. And third, we now have, for the very first time, a world-class Ethernet switch platform, absolutely world-class.

And so these three platforms allows us to connect every company, whether it's hyper-scale or enterprise from the core all the way out to the edge. We have the end-to-end solution, we have the compute solution, and very importantly and probably the most importantly, we have the software capability to go all together. Otherwise, how do you even assemble all this stuff. It's just way too much gear. Nobody without software has the encourage to invest $200 million on a bunch of hardware to connect it together if you don't have software, and when it comes to this type of software, AI factory software, NVIDIA is singular. We are -- this is our focus. And so that -- all of that plays. And I think the $300 billion basically represents all of that, okay?

Q - Stacy Rasgon {BIO 16423886 <GO>}

And on the short term, Colette, I hate to ask this question in this forum, but I've got 10 emails in my inbox from investors asking. So I'm just going to ask it. Yes, we're about two-thirds of the way through the quarter. Do you have any updates on the quarter itself, any changes at all that you're seeing? And again, I apologize for asking it in we have (inaudible)

A - Colette Kress {BIO 18297352 <GO>}

Well, for your 10 questions that are out there, we don't have any update for you. We provided guidance at the beginning of the quarter. I feel that our guidance was quite solid, even in what we'd say has been a lot of world dynamics over this period of time. But at this time, no change from the guidance, nothing to add. And I guess, in that perspective, we decided to concentrate here on GTC and the great announcements and you should say status quo always looking fine.

Q - Stacy Rasgon {BIO 16423886 <GO>}

Got it. That's helpful. Thank you so much, guys.

A - Jensen Huang {BIO 1782546 <GO>}

Thank you. Thanks, Stacy.

Operator

Our next question comes from Tim Arcuri from UBS. Tim, please go ahead.

Q - Tim Arcuri {BIO 3824613 <GO>}

Hello, can you hear me?

A - Jensen Huang {BIO 1782546 <GO>}

Yes, sure, we can. Hi, Tim.

Q - Tim Arcuri {BIO 3824613 <GO>}

Well, thanks. Hi, Jensen, how are you? I had a couple of questions on autos. And I know you said it's the next multibillion-dollar business with cost some inflection. So my two questions are, first, can you sort of help us shape the curve for that $11 billion that's in the pipeline over the next six years? I guess maybe one way I was thinking about it was, if you split it sort of into two different three-year parts, it's reasonable that maybe 25% of that pipelines in the first three years and 75% is in the back three years, that's the first question. And then the second question is, it sounds like most of that is software versus hardware. So I'm wondering if you can break that down for us. Thanks.

A - Jensen Huang {BIO 1782546 <GO>}

I'll do the second one and Colette will do the first part, okay? So on the second, if you -- the autonomous vehicle, the software-defined car movement took one generation longer than I expected, but it is all here now. And part of it has been aggravated by vision of what a software-defined car could do and the business models that it could enable. Every single car company in the world wants to be a high-tech company and a tech company doesn't chip a product and never, say, never connect to it. Again, a technology company today is a connected device company and the car is one of the greatest opportunities for a connected device because it stays in your connection for 20 years.

Once it's on the road, you're connected to it for 20 years, the installed base that you could build over 20 years is incredible. And I think that car companies, especially the state-of-the-art car companies what is called the new electric vehicle companies, the NEVs, they all see this. And they're piling on as much computation as they can into the car because they're going to provide new services for two decades after that. And so it took a while but the time is now here. They see the vision, they see the excitement, they see the opportunity for transformation, they see the business model opportunities that the economics after they sell the car is going to be way, way better than the economics at the point of sales. And so that's the big realization.

So although it took us a little longer to get here, we are all here now. And so Oren, that started production this month, it's just a home run. It's potentially one of the most successful products in our company's history. It is singular in the marketplace, it has the benefit of all of NVIDIA software stack that sits on top of it so that you could program all of this complicated robotic software. It has the benefit of having other four other -- three other pillars aside from the robotics computer inside the car, you have NVIDIA's architecture to help you train the model, you have NVIDIA's architecture to help you develop synthetic data to train that model. You have the

opportunity to use NVIDIA architecture to do the digital twin simulation so that you could orchestrate and manage your fleet. And so we have four pillars of opportunities besides what goes into the computer. That $11 billion doesn't include the other three pillars. The $11 billion is just what goes into the car. And so now you can imagine how big this business opportunity is for us and the largest robotics opportunity near term. And so that's -- I think I answered this question.

A - Colette Kress {BIO 18297352 <GO>}

I think you did. So your second part of the question was regarding the $11 billion and how to think about it in terms of the years. And I would look at it in multiple inflection points, okay? From inflection point now as we begin the ramp on or in ramp with the NAVs, the EVs is using this as a computing platform and this is what you will see even today, even this year. Now the second part of it comes into calendar 2024, calendar 2025, software begins. So yes, you're correct. It is over time and very much influenced by the software when that ramps, but that will be a very important part of this growth in terms of 11 billion.

Q - Tim Arcuri {BIO 3824613 <GO>}

So is it -- so the reasonable collect to say that like 75% of its part into the out period? Is that a reasonable estimate?

A - Colette Kress {BIO 18297352 <GO>}

It's reasonable It's reasonable. But again, we're not done. I'm sure we'll continue to update that pipeline over time as more and more partners become very locked in, in terms of this platform. But for right now, yes, it's a reasonable assumption.

Q - Tim Arcuri {BIO 3824613 <GO>}

Thanks a lot.

A - Jensen Huang {BIO 1782546 <GO>}

Thank you.

Operator

Our next question comes from Ambrish Srivastava from BMO. Ambrish, please go ahead. Ambrish, please unmute yourself and turn on your video.

Q - Ambrish Srivastava {BIO 4109276 <GO>}

Hi, can you hear me?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Perfectly.

Q - Ambrish Srivastava {BIO 4109276 <GO>}

Thank you, folks. Glad to Jensen, Thank you. That was very informative. A lot of information to digest. I have a question on the software side. I'm -- just pardon me for failing to understand the opportunity. And really, it's pretty big, $150 billion in both sides of the enterprise as well as the Omniverse. So really, the longer-term question is, a, how big is this opportunity today out there? There are other surveys obviously, the market is nascent. But how much are you competing to get that? Are there other players participating in the market? So that's really longer-term trend to understand these are really big numbers and big numbers at track competitors. So that's what I'm trying to understand how should we see NVIDIA's position. And more on a tactical term, Colette, you for sharing the -- if I got the number right, $200 million today. At what point would you consider giving us metrics on backlog or any other metrics you had in mind? So, thank you.

A - Jensen Huang {BIO 1782546 <GO>}

There's two types of software if I could simplify it. There's the application software and then if you will, the operating system software. In the case of a data center, the operating systems of a data center, the operating environment of a data center is VMware and Red Hat. For example, the operating environment of a computer, a client computer, Windows, Mac, for example, Android, for example, that's the operating environment. On top of it, there's an engine. That engine is, if you will, the operating environment of a domain of applications.

In the case of AI applications, the engines are built on top of CUDA, as you know quite well, we pioneered this whole space. And so CUDA has an engine on top of it called CUDA NAND[ph] has a library called TensorRT, has a library called Triton and the list goes on, okay? There's Dali. There's a bunch of stuff inside for doing all the things that I mentioned earlier, which is the ingestion of data, the preprocessing, the processing of data, to the feature engineering to the machine learning to deep learning to inference. And every one of those stages of that workflow has engines associated with it, libraries associated with it. That library engine today runs in everybody's cloud, it runs in hyper-scale companies all over the world. Pieces of that engine runs all over the place. For -- up until now and for hyper-scale, we'll continue to as part of our product, if you will.

For the world's enterprise, they will need a different level of support. They need a different level of support because the world's enterprise doesn't have the type of DevOps and MLOps that's needed to maintain this engine, and so we will do that continued innovation bringing new features and capabilities to it, updated for new GPUs like hoppers coming out. So we'll create a new version for Hopper. We'll connect their existing services to our last gen -- new services to last generation, old services to new generation, that entire body of work that is fairly intensive work for operating an AI factory, that work, that technology, all of those services, if you will, we embody into this thing called NVDAI AI.

Does anybody else do it today, that engine? I think it's reflected in our success with NVIDIA GPUs in the world's enterprise for data ops, data science, machine learning, deep learning, we're quite successful, as you know, and quite singular, as you know, engine sits on top of our GPUs. This engine sits on top of our DGXs and servers and

all of our in-network computing, distributed computing. It sits on top of all of that, okay?

And so we've now finally produced a product that an enterprise can license. They've been asking for it, and the reason for that is because they can't just go to open source and download all the stuff and make it work for their enterprise. Not -- no more than they could go to Linux download open source software and run a multibillion-dollar company with it. That's why Red Hat exists, that's why VMware exists, that's why so on and so forth, okay?

And so even though we have a lot of our software in open source, the enterprises really need us to turn this into a product, support it like a product, enter into service level agreements that gives them 24/7 access, teach them how to use it and help them operate it, deploy it into their own data centers, turning every enterprise's data center into a state-of-the-art cloud, and so that's what they would like and that's what NVIDIA AI is about. We have an installed base of GPUs in the world today. We support them with NVIDIA AI as we already are. But going forward, we've turned it into a product and a licensable product called NVIDIA AI enterprise, okay? As far as alternative, we are -- NVIDIA AI is really quite industry-defining and it is the case with NVIDIA Omniverse as well quite industry-defining.

A - Colette Kress {BIO 18297352 <GO>}

Ambrish, your question that you indicated on software in the future, do we see metrics being eligible for discussion? Absolutely. If this is a growth driver for us as we see going forward, providing you insight in terms of what drove that software growth and how to think about both what the licensing is, what the maintenance is of it going forward, we're happy to share.

Q - Ambrish Srivastava {BIO 4109276 <GO>}

Okay. Thank you, folks.

A - Jensen Huang {BIO 1782546 <GO>}

Thank you. Good questions.

Operator

Our next question comes from Harlan Sur from JP Morgan. Harlan, please unmute yourself and start your video using the buttons at the bottom-left corner. Please go ahead.

Q - Harlan Sur {BIO 6539622 <GO>}

Can you guys hear me?

A - Jensen Huang {BIO 1782546 <GO>}

Yes, perfectly.

Operator

Let me check that for you.

Q - Harlan Sur {BIO 6539622 <GO>}

Can you hear me?

A - Colette Kress {BIO 18297352 <GO>}

Yes.

A - Jensen Huang {BIO 1782546 <GO>}

Yes, still can.

Operator

Bear with me Harlan. I am just looking for the video feed.

A - Jensen Huang {BIO 1782546 <GO>}

Your video feed got lost on the internet.

Operator

Apologies, Harlan, we have lost the connection to the video feed. Please bear with me while I find out what's going on.

A - Jensen Huang {BIO 1782546 <GO>}

Harlan, we can hear you. We could check.

Well, while we're waiting, I like to say, I thought the NVIDIA management team presentations were pretty fabulous.

Operator

Okay, Harlan. Please --

A - Jensen Huang {BIO 1782546 <GO>}

What an amazing management team. I love my team.

Q - Harlan Sur {BIO 6539622 <GO>}

Do you guys hear me?

A - Jensen Huang {BIO 1782546 <GO>}

Yes, hey Harlan.

Q - Harlan Sur {BIO 6539622 <GO>}

Thanks for hosting this a very informative event. My question is, is the version of your Hopper GPU, the H100 that you announced today, is that optimized for your great CPU and other ARM-based CPUs that are currently in the market today? Or do we have to wait for a follow-on version of the Hopper because Jensen I know that you had talked previously about having an x86 optimized GPU version and an ARM-based optimized GPU version. And then outside of the early wins that you have with Grace on computing platforms like ALPS, you mentioned broader expansion into AI infrastructure. Would this include your successful DGX platform, maybe DGX SuperPOD powered by your ARM-based race architecture and GPUs in the future?

A - Jensen Huang {BIO 1782546 <GO>}

We -- our company's business is about accelerating computing, which means we like computers of all kinds, x86 kinds, arm kinds, any kind, okay? And so wherever there's a CPU, there's an opportunity for us to accelerate that CPU. That is really the core of our business, and we'll continue to support whatever CPU, the market best desires. And there's all kinds of different CPUs out there for different types of configurations and different use cases, and we'll support all of them. We'll support all of them. That's kind of the nature of our company, and we'll continue to be open and support whatever the market needs.

Grace has just off the charts phenomenal capability. And its performance is unlike any others for the type of AI applications that we're targeting. So for large data movement workloads, Grace is really quite ideal. And so for AI infrastructure, whether it's in our DGX, OEM servers, computer makers in the cloud, wherever there's AI infrastructure, we're going to offer Grace as well as support for x86. And so let the market decide, and we're delighted by adoption of accelerated computing wherever it is and whatever microprocessor comes along.

Q - Harlan Sur {BIO 6539622 <GO>}

Thanks, Jensen.

A - Jensen Huang {BIO 1782546 <GO>}

Yup. Thanks.

Operator

We have time for one last question. And our last question comes from Atif Malik from Citi. Atif, please unmute yourself and start your video.

Q - Atif Malik {BIO 7312618 <GO>}

Hi, can you hear me?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Nice to see you, Atif.

Q - Atif Malik {BIO 7312618 <GO>}

Hi. Thanks for taking my question. I have a question on gaming. Last year, you were supply-constrained. I wanted to get your thoughts on supply and demand for this year. There has been disruptions in both the supply side with the Shenzhen lockout as well as the demand side in Russia-Ukraine contract impacting European gaming demand. So how should we think about your supply and demand dynamics for this year? And then the second part, if you can talk about RTX, the installed base doubled from last year, 15% and 30%. And how should we think about your refresh on the gaming products, given rising competition from AMD's RDNA 3 and Intel Arc? Thank you.

A - Jensen Huang {BIO 1782546 <GO>}

I'll go backwards. It's hard to comment about things that don't exist. And so I'll look forward to them when the time comes. The -- with respect to the two dynamics that you mentioned, they're disproportional by an enormous amount. We -- our supply constraint is by far the greatest impact of this last year and it continues to be.

There are several -- just really terrific dynamics that are happening in gaming. Number one, there are more gamers than ever as Fisher was saying earlier. The way that people game is changing, not only are they playing games for the game itself, but gaming is also a way to hang out with friends and spending time with friends. Gaming is a form of art now and gaming, of course, as you know, is a very important form of sports. And so gaming now cuts across leisure to social to our art to sports. And no -- very few. I mean, I can't think of one right now, very few other entertainment genres is as broad as broadly impactful. More gamers, more way to game, and of course, very importantly, gamers don't just game and gaming is not just about games anymore and the creative part of gaming has really done so well.

We are quite unique in our ability to serve every segment of gaming, whether it's PC desktop, PC Laptop, the most successful game console in the history of game consoles to cloud gaming, first-party cloud gaming with GeForce NOW to third-party cloud gaming partnerships. So we have the ability to reach televisions and tablets and phones and PCs and wherever, whatever operating system it happens to be. And so our gaming strategy is incredibly broad. And because gamers are such a creative bunch is so much more than gaming. So our dynamics is really great, which explains the reason why channel inventory remains low, and we expect it for some time.

With respect to RTX. RTX is a big deal on multiple dimensions. Number one, if not for RTX, Omniverse wouldn't exist. If not for RTX, it would not be possible to make something like Omniverse exist, which is a simulation. It's not prebaked art. Everything that you see is not prebaked like most video games have a lot of prebaking. It's called prebaking. And movies, as you know, is largely prebraking. Omniverse is real-time. It's synthesized in real-time. It's simulated in real-time. The

materials, the life, the shadows, all of the really impressive effects that you see that makes things beautiful come about because it's just beautiful. The physics is beautiful. And so RTX made that possible.

RTX reset computer graphics altogether. And if you look at NVIDIA's installed base today and you look at the world's installed base of gaming platforms, the number that our RTX level of ray tracing is really, really small. And so in a lot of ways, we have completely reset because of discontinuous invention in computer graphics, we have reset the world's installed base of hardware. And the combination of the rich dynamics of gaming and the fundamental invention of RTX has really caused the demand to be just through the roof during this time. And so I think the gaming dynamics overall are just really terrific, and I really appreciate that question.

Operator

These are all the questions we have time for today. And I would like to hand back to Jensen Huang for any closing remarks.

A - Jensen Huang {BIO 1782546 <GO>}

Thank you for joining NVIDIA GTC and our Analyst Day. I would like to say one more time how incredible the NVIDIA management team did. It is so fantastic to be on stage with Colette and to share the stage with the NVIDIA management team. As you could see, I'm super proud of them. You could also see why I should be, they're incredible. And it is the reason why NVIDIA is such a great investment.

From NVIDIA's management presentations, you could see the exciting growth drivers. Gaming dynamics are excellent, as I mentioned, more gamers, more ways to game. RTX has reset the gaming installed base and games are so much more than games now. Demand continues to exceed supply, keeping the channel inventory low. We have strong demand for our data center platforms driven by AI training and inference across all of those different models that I mentioned earlier and across just about every cloud computing company and now going into the world's enterprise, and we have excellent visibility into our data center business.

NVIDIA is the engine of the world's AI infrastructure and our software business now augments our platform. The platform that we've been developing over all these years, we've been developing software on top of it, and now we turn them into software products that customers can license for the enterprise level that supports that the desire. We're offering two industry-defining platforms, NVIDIA AI and NVIDIA Omniverse, and they both come with world-class software support and licensable support.

Auto is on its way to be our next multibillion-dollar business, and I'm super excited about the work that we've done. It took us nearly a decade to reach this point where the entire automotive industry is now ready to be a software-defined industry and become a tech industry.

And today, we announced and launched a giant wave of new products, the Hopper H100, the DGX SuperPod with our brand-new NVLink switch system, our Grace CPU super chip and the enabling technology that made it possible, the NVLink chip-to-chip incredibly energy-efficient, high-speed world-class link is now open for our partners. Spectrum 4, 400 gigabits per second Ethernet switch.

And of course, the whole bunch of software that connect scientific challenges, markets, and new growth to our company. It's software that activates all of this hardware. It's software that connects all of this software to interesting challenges, groundbreaking work by developers and scientists, and, of course, very importantly, new growth of our company. With each of our four-layer stack on top of our one NVIDIA architecture, we engage more opportunities. And now where our company and the accelerated computing platform has grown so much to 500 libraries, we're able to serve the world's $100 trillion dollars of industry.

Thank you all for joining us today.