FINAL

Bloomberg Transcript

Q4 2022 Earnings Call

Company Participants

- Colette Kress, Executive Vice President and Chief Financial Officer
- Jensen Huang, Founder, President and Chief Executive Officer
- Simona Jankowski, Investor Relations

Other Participants

- C.J. Muse
- Harlan Sur
- Joe Moore
- John Pitzer
- Matt Ramsay
- Raji Gill
- Stacy Rasgon
- Tim Arcuri
- Toshiya Hari
- Vivek Arya

Presentation

Operator

Good afternoon. My name is David, and I will be your conference operator today. At this time, I'd like to welcome everyone to NVIDIA's Fourth Quarter Earnings Call. Today's conference is being recorded. All lines have been placed on mute to prevent any background noise. After the speakers' remarks, there will be a question-and-answer session. (Operator Instructions). Thank you.

Simona Jankowski, you may begin your conference.

Simona Jankowski {BIO 7131672 <GO>}

Thank you. Good afternoon, everyone, and welcome to NVIDIA's conference call for the fourth quarter of fiscal 2022. With me today from NVIDIA are Jensen Huang, President and Chief Executive Officer; and Colette Kress, Executive Vice President and Chief Financial Officer.

I'd like to remind you that our call is being webcast live on NVIDIA's Investor Relations website. The webcast will be available for replay until the conference call to discuss our

financial results for the first quarter of fiscal 2023. The content of today's call is NVIDIA's property. It can't be reproduced or transcribed without our prior written consent.

During this call, 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 the disclosure in today's earnings release, our most recent forms 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, February 16, 2022, based on information currently available to us. Except as required by law, we assume no obligation to update any such statements.

During this call, we will discuss non-GAAP financial measures. You can find a reconciliation of these non-GAAP financial measures to GAAP financial measures in our CFO commentary, which is posted on our website.

With that, let me turn the call over to Colette.

Colette Kress {BIO 18297352 <GO>}

Thanks, Simona. We had an excellent quarter with revenue up 53% year-on-year to $7.6 billion. We set records for total revenue as well as for Gaming, Data Center and Professional Visualization. Full year revenue was a record $26.9 billion, up 61%, compounding the prior year's growth of 53%.

Starting with Gaming. Revenue of $3.4 billion rose 6% sequentially and was up 37% from a year earlier. Fiscal year revenue of $12.5 billion was up 61%. Gaming has become the top entertainment medium that continues to show strong momentum. Just last month, Steam hit a record 28 million concurrent users, up 50% in two years.

Record desktop revenue in the quarter was led by our growth in our GeForce RTX 30 Series products with continued strength in the high end. At CES, we announced the RTX 3050 GPU, which hit retail in late January, bringing NVIDIA, RTX, and AI technologies, to more mainstream audiences.

Laptop gaming revenue also set a record, driven by the ramp of the new GeForce RTX 3070 Ti and 3080 Ti GPUs, which were also announced at CES. These leverage our fourth-generation Max-Q technology to enable quiet, thin and light gaming laptops.

All in, we announced over 160 new laptop designs for NVIDIA Ampere architecture RTX 30 Series GPUs. These include a number of Studio systems targeting the tens of millions of creators driving the future of design, innovation, and virtual world.

In addition to supporting the new RTX 30 Series GPUs, Studio laptops feature support for NVIDIA software like Omniverse, Fundus [ph], and Broadcast.

Availability of our gaming products in the channel remains low. NVIDIA RTX ecosystem continues to expand with over 30 new RTX games and applications added this quarter, including blockbuster hits like Battlefield 2042, Grand Theft Auto: The Trilogy, Call of Duty: Vanguard, and, God of War. In addition, several new titles support NVIDIA Reflex for low-latency gameplay.

Our GPUs are capable of cryptocurrency mining, though we have limited visibility into how much of this impacts our overall GPU demand. Nearly all desktop NVIDIA Ampere architecture GeForce GPU shipments are Lite Hash Rate to help direct GeForce supply to gamers. Cryptomining processor revenue was $24 million, which is included in OEM and other.

We continue to expand the NVIDIA GeForce NOW cloud gaming ecosystem with new hit titles including EA's Battlefield 4 and Battlefield 5. At CES, we announced a partnership with Samsung to integrate GeForce NOW in its smart TVs starting in Q2 of this year. This follows last month's beta release of the GeForce NOW app for LG smart TVs. In addition, we teamed up with AT&T to bring GeForce NOW to 5G mobile devices in the US. We also added our first GFN data center in Canada.

Moving to Pro Visualization. Q4 revenue of $643 million was up 11% sequentially, and up 109% from a year ago. Fiscal year revenue of $2.1 billion was up 100%. Sequential growth in the quarter was driven by a shift to higher-value workstation and the continued ramp of our NVIDIA Ampere architecture. We believe strong demand is fueled by continued buildouts for hybrid work environment as well as growth in key workloads, including 3D design, AI, and rendering.

For example, Sony Pictures Imageworks is using NVIDIA RTX to accelerate ray tracing for rendering related applications. Volkswagen is using NVIDIA RTX for AI to assist in predictive maintenance of their vehicles, and Duke Energy is using NVIDIA RTX for AI and VR to map, view and maintain energy facilities.

NVIDIA Omniverse Enterprise software entered general availability and while it's still in early days, customer feedback so far has been very positive with multiple significant Enterprise licensees already signed. In addition to software licenses, Omniverse also drives computing opportunity for NVIDIA RTX in laptops, workstations, and on-prem servers, and the cloud. Omniverse can be used by individuals for free and by enterprise teams via software subscriptions.

At CES, we made the free version of Omniverse for individuals general availability. Omniverse allows creators with RTX GPUs to connect leading 3D design applications to a single scheme and supercharge their work with AI and physics. We also announced early access to Omniverse Cloud, which adds one-click capability to collaborate with other artists, whether across the room or across the globe.

For digital twin applications, we announced the Isaac Autonomous Mobile Robot platform using Omniverse and securely orchestrated and cloud delivered with Fleet Command. The platform optimizes operational efficiency and accelerated deployment from logistics

robots. It consists of several NVIDIA AI technologies and SDKs, including DeepMap for high-precision mapping, Metropolis for situational awareness, and ReOpt for real-time route optimization.

Moving to automotive. Q4 revenue was $125 million, declined 7% sequentially and 14% from the year ago quarter. Fiscal year revenue of $566 million was up 6%. We have just started shipments of our Orin-based product platform and expect to return to sequential revenue growth in Q1 with more meaningful inflection in the second half of the fiscal year and momentum building into calendar 2023 and beyond.

I will now hand it over to Jensen to provide more color on this morning's automotive news.

Jensen Huang {BIO 1782546 <GO>}

Thanks, Colette.

Earlier today, we announced the partnership with Jaguar Land Rover to jointly develop and deliver fleets of software defined cars. Starting in 2025, all new Jaguar and Land Rover vehicles will have next-generation automated driving systems plus AI enabled software and services built on the NVIDIA DRIVE platform. DRIVE Orin will be the AI computer brain running our DRIVE AV and DRIVE IX software. And the DRIVE Hyperion sensor network will be the central nervous system.

This new vehicle architecture will enable a wide spectrum of active safety, automated driving and parking systems. Inside the vehicle, the system will deliver AI features including driver and occupant monitoring, and advanced visualization of the vehicles surroundings. We are very much looking forward to partner with Thierry Bollore, JLR's CEO, and his team to reinvent the future of luxury cars. Our full stack end-to-end approach is a new business model that offers downloadable AV and AI services to the fleet of JLR vehicles with a shared software revenue stream for both companies over the life of the fleet.

This partnership follows the template of our announcement with Mercedes-Benz. Our shared software revenue opportunity with both OEMs will scale with the size of their NVIDIA powered fleet, which combined, can exceed 10 million cars over a decade.

Colette, back to you.

Colette Kress {BIO 18297352 <GO>}

Thanks Jensen. Moving to Data Center. Record revenue of $3.3 billion grew 11% sequentially and 71% from a year earlier. Fiscal year revenue of $10.6 billion was up 58%. Data Center growth in the quarter was once again led by our compute products on strong demand for NVIDIA AI. Hyperscale and cloud demand was outstanding with revenue more than doubling year-on-year. Vertical industries also posted strong, double-digit year-on-year growth, led by consumer internet companies.

FINAL

Bloomberg Transcript

The flagship NVIDIA A100 GPU continued to drive strong growth. Inference focused revenue more than tripled year-on-year. Accelerating inference growth has been enabled by widespread adoption of our Triton Inference Server software, which helps customers deliver fast and scalable AI in production.

Data center compute demand was driven by continued deployment of our Ampere architecture based product for fast growing AI workloads such as natural language processing and deep learning recommendation systems as well as cloud (inaudible).

For example, Block, Inc, a global leader in payment, uses conversational AI in its Square Assistant to schedule appointments with customers. These AI models are trained on video GPUs in AWS and perform inference 10x faster on the AWS GPU service and on our CPUs.

Social media company Snap used NVIDIA GPUs and Merlin deep recommendator software to improve inference cost efficiency by 50% and decrease latency to 2x.

For the third year in a row, industry benchmarks show that NVIDIA AI continues to lead the industry in performance. Along with partners like Microsoft Azure, NVIDIA set records in the latest MLPerf benchmarks for AI training across 8 popular AI workloads including computer vision, natural language processing, recommendation systems, reinforcement learning, and object detection. NVIDIA AI was the only platform to make submissions across all benchmarks and use cases, demonstrating versatility as well as our performance.

The numbers show performance gains on our A100 GPUs of over 5x in just 18 months thanks to continuous innovations across the full stack in AI algorithms, optimization tools, and system software. Over the past three years, we saw performance gains of over 20x powered by advances we have made across our full stack offering GPUs, networks, systems and software.

The leading performance of NVIDIA AI is sought after by some of the world's most technically advanced companies. Meta Platforms unveiled its new AI supercomputer Research SuperCluster, with over 6,000 A100 GPUs into an NVIDIA Quantum InfiniBand Network. Meta's early benchmarks showed its system can train large natural language processing models 3x faster and run computer vision jobs 20x faster than the prior system. In a second phase later this year, the system will expand to 16,000 GPUs that Meta believes will deliver 5 exaflops of mixed precision AI performance. In addition to performance at scale, Meta cited extreme reliability, security, privacy, and flexibility, to handle a wide range of AI models as its key criteria for the system.

We continued to broaden the reach and ease the adoption of NVIDIA AI into vertical industries. Our ecosystem of NVIDIA-certified systems expanded with Cisco and Hitachi Vantara which joined Dell, HewlettPackard Enterprise, Inspur, Lenovo, and Supermicro, among other server manufacturers. We released version 1.1 of our NVIDIA AI Enterprise software, allowing enterprises to run [ph] accelerated AI workloads on VMware vSphere, and mainstream IT infrastructure as well and we expanded the number of system integrators qualified for NVIDIA AI Enterprise.

Forrester Research in its evaluation of Enterprise AI infrastructure providers recognized NVIDIA in the top category of leaders. An example of a partner that's helping to expand our reach into enterprise IT is Deloitte, a leading global consulting firm, which has built its center for AI computing on NVIDIA DGX SuperPODs. At CES, we extended our collaboration to AV development, leveraging our own robust AI infrastructure and Deloitte's team of 5,500 system integration developers and 2,000 data scientists to architect solutions for truly intelligent transportation.

Our networking products posted strong sequential and year-over-year growth, driven by exceptional demand across use cases ranging from cloud computing, supercomputing and enterprise to storage. Ethernet adopters led growth driven by adoption of our next-generation products and higher speed deployments. While revenue was gated by supply, we anticipate improving capacity in coming quarters, which should allow us to serve the significant customer demand we're seeing. Across the board, we are excited about the traction we are seeing with our new software business models, including NVIDIA AI, NVIDIA Omniverse, and NVIDIA DRIVE.

We are still early in the software revenue ramp. Our pipelines are building as customers across the industries seek to accelerate their pace of adoption and innovation with NVIDIA.

Now let me turn it back over to Jensen for some comments on Arm.

Jensen Huang {BIO 1782546 <GO>}

Thanks, Colette.

Last week, we terminated our efforts to purchase Arm. When we entered into the transaction at September 2020, we believed that it would accelerate Arm's focus on high performance CPUs and help Arm expand into new markets, benefiting all Arm customers in the entire ecosystem. Like any combination of pioneers of important technologies, our proposed acquisition spurred questions from regulators worldwide. We appreciated the regulatory concerns. For over a year, we'd work closely with SoftBank and Arm to explain our vision for Arm and reassure regulators that NVIDIA would be a worthy steward of the Arm ecosystem.

We gave it our best shot, but the headwinds were too strong and we could not give regulators the comfort they needed to approve our deal. NVIDIA is working to accelerate computing and our overall strategy will continue as before. Our focus is accelerated computing. We are on track to launch our Arm-based Grace CPU targeting Giant AI and HPC workloads in the first half of next year. Our 20-year architectural license to Arm's IP allows us the full breadth and flexibility of options across technologies and markets.

We will deliver on our three-chip strategy across CPUs, GPUs and DPUs. Whether X86 or arm, we will use the best CPU for the job, and together with partners in the computer industry, offer the world's best computing platform to tackle the impactful challenges of our time.

Back to you Colette.

Colette Kress {BIO 18297352 <GO>}

Thanks, Jensen.

We're going to turn to our P&L and our outlook. For the discussion of the rest of the P&L, please refer to the CFO commentary published earlier today on our Investor Relations website.

Let me turn to the outlook for the first quarter of fiscal 2023. We expect sequential growth to be driven primarily by Data Center. Gaming will also contribute to growth. Revenue is expected to be $8.1 billion, plus or minus 2%. GAAP and non-GAAP gross margins are expected to be 65.2% and 67%, respectively, plus or minus 50 basis points.

GAAP operating expenses are expected to be $3.55 billion, including the Arm write-off of $1.36 billion. Non-GAAP operating expenses are expected to be $1.6 billion. For the fiscal year, we expect to grow non-GAAP operating expenses at a similar percent as in fiscal 2022.

GAAP and non-GAAP other operating, other income and expenses, are both expected to be an expense of approximately $55 million, excluding gains and losses or non-affiliated investments. GAAP and non-GAAP tax rates are expected to be 11% and 13%, plus or minus 1%, excluding discrete items.

Capital expenditures are expected to be approximately $350 million to $400 million.

Further financial details are included in the CFO commentary and other information available on our IR website.

In closing, let me highlight upcoming events for the financial community. We'll be attending the Morgan Stanley Technology Media and Tech Telecom conference in-person on March 7. We will also be hosting a virtual Investor Day on March 22, alongside the GPU Technology Conference. This will follow Jensen's opening key note, which we invite you to tune in for.

Our earnings call to discuss the results for our first quarter of fiscal 2023 is scheduled for Wednesday, May 25th. We will now open the call for questions.

Operator, will you please poll for questions.

Questions And Answers

Operator

(Question And Answer)

Thank you. (Operator Instructions). We'll take our first question from Toshiya Hari with Goldman Sachs & Company. Your line is open.

Q - Toshiya Hari {BIO 6770302 <GO>}

Great. Thank you so much for taking the question. Jensen and Colette, I wanted to ask about Data Center. Colette, based on your guidance, you're probably guiding Data Center growth on a year-over-year basis to accelerate into the April quarter. You talked about Hyperscale, Cloud growing more than 2x and Enterprise verticals growing strong double-digits in the January quarter. Can you kind of speak to the drivers for April, and perhaps speak to visibility into the second half of the fiscal year as well in Data Center? Thank you.

A - Jensen Huang {BIO 1782546 <GO>}

Sure.

A - Colette Kress {BIO 18297352 <GO>}

I'll start and I'll turn it over to Jensen. For Q1, our guidance can include an acceleration of Data Center from where we left in terms of Q4. We will have growth across several of our market platforms within Q1, both Data Center, Gaming and probably a couple of others. But yes, there is expected to be accelerated growth as we move into Q1.

I'll turn it over to Jensen to talk about the drivers that we see for the quarter and also for the full year.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. We have several -- first of all, Toshiya, great to hear from you. We have several growth drivers in data centers. There's Hyperscale, Public Cloud, Enterprise Core and Enterprise Edge. We're seeing growth across the entire spectrum. There are several different use cases that are particularly exciting, large language models, language understanding models triggered by the invention of transformers, which is probably one of the most important AI models that's been invented in some time. And conversational AI used for customer service, chat bots, a whole bunch of customer service applications.

It could be web-based. It could be point-of-sale based, it could be cloud-based. Recommender systems, deep learning-based recommender systems are making groundbreaking improvements. And cloud graphics, all of the work that we're doing and putting rendering or putting simulations up in the cloud, cloud gaming, Android cloud gaming, are really driving adoption in the cloud. And so, so many different use cases but across all of the different platforms in data centers.

Operator

Next, we'll go to C.J. Muse with Evercore ISI.

Q - C.J. Muse

Yes. Good afternoon. Thank you for taking the question. I guess another question on the data center side. Curious if you can speak to supply constraints on the wafer side and

whether that played a role in terms of capping revenues in the January quarter and how you see that becoming less of a headwind for you as you proceed through the year?

A - Colette Kress {BIO 18297352 <GO>}

Thanks, C.J., for the question. I'll start on the data center supply. We discussed last quarter and discussed today, we still have some supply constraints across some of our businesses. Networking in the Data Center business has been supply constrained. We're improving every single day and we do expect to improve supply each quarter as we enter into fiscal year '23.

So that is probably the key area within our Data Center. But from time-to-time, there can be other types of focus from data center. So I'll turn the rest of the question to Jensen in terms of how (inaudible) about the rest of the year as well.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Colette captured it well. We are supply constrained. Our demand is greater than our supply. But as you know, our data center product line consists of GPUs, mix, Bluefield DPUs, Quantum and Spectrum switches, HGX, if you will, system component, meaning that the entire motherboard or the entire GP board is delivered in combination because it's so complicated. And so we have products that span a broad reach of use cases for data centers from training of AI models to inferencing at very large scale, to universal GPUs for public cloud, industry standard servers, commodity servers for enterprise use, and supercomputing systems that use InfiniBand and Quantum switches.

And so the applications space is quite broad. We saw demand constraint pretty much across the entire range. Our operations team did a fantastic job this year, both in executing in all of these complicated products, but also in expanding our supply base. We expect supply to improve each and every quarter going forward. And this quarter, this coming quarter, the April quarter is based on guidance that Colette just made, is consistent with an increasing supply base. We expect to still be demand constrained, but our supply base is going to increase this quarter, this next quarter, and pretty substantially in the second half.

Operator

Next, we'll go to Joe Moore with Morgan Stanley.

Q - Joe Moore {BIO 17644779 <GO>}

Great. Thank you. I wonder if you could talk a little bit more about Grace now that the strategy is kind of separated from the acquisition of Arm. The -- what are your aspirations there? Is it going to be primarily oriented to the DGX and HX Systems business versus merchant chips? Just how are you thinking about that opportunity long-term?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Thanks, Joe. We have a multiple Arm projects going in the company, from connected from devices to robotics processors such as the new Orin that's going into autonomous

vehicles and robotic systems, and industrial automation, robotics and such. Orin is doing incredibly well. It's started production. And as we mentioned earlier, it's going to drive an inflection point starting in Q2, but accelerating through Q3 and the several years after as we ramp into all of the electric cars and all of the robotic applications and robotaxis and such.

We also have Arm projects with the CPU that you mentioned, Grace. We have Grace, and we have surely have the follow-ons to Grace, and you could expect us to do a lot of CPU developments around the Arm architecture. One of the things that's really devolved nicely over the last couple of years is the success that Arm has seen in hyperscalers and data centers. And it's really accelerated and motivated them to accelerate the development of higher-end CPUs.

And so you're going to see a lot of exciting CPUs coming from us and Grace is just the first example that you're going to see a whole bunch of them beyond that. But our strategy is accelerated computing. That's ultimately what we do for a living. We, as you know, love it where there's any CPUs. If it's an x86 from any vendor. So long as we have a CPU, we could connect NVIDIA's platform to it and accelerate it for artificial intelligence or computer graphics or robotics and such. And so we love to see the expansion of CPU footprints, and we're just thrilled that Arm is now growing into robotics and autonomous vehicles and cloud computing and supercomputing and in all these different applications, and we intend to bring the full spectrum of NVIDIA's accelerated computing platform to NVIDIA Arm CPUs.

Operator

Next we'll go to John Pitzer with Credit Suisse.

Q - John Pitzer {BIO 1541792 <GO>}

Yes. Good afternoon, guys. Thanks for letting me ask the question. Colette and Jensen, just on the inventory purchase obligations. I think this was the fourth quarter in a row where you've seen greater than 30% sequential growth and it is the first quarter where that number is now eclipsing kind of your quarterly revenue guidance. And so, I guess I'm trying to figure out to what extent is this just a reflection of how tight things are across the semi industry? To what extent it is the poker tale of kind of how bullish you are on future demand? And relative to your commentary, that supply starts to get better throughout the year, should we expect that number to start to level off? Or as the mix moves more to data center and more to longer cycle times, more complicated devices, should that number continue to grow?

A - Jensen Huang {BIO 1782546 <GO>}

The factors, the drivers that you mentioned in the supply chain, we expanded our supply chain footprint significantly this year to prepare us for both increased supply base and supply availability in each one of the quarters going forward, but also in preparation for some really exciting product launches.

As mentioned, Orin ramping into autonomous vehicles is brand new. This is the inflection point of us growing into autonomous vehicles. This is going to be a very large business for us going forward. It was already mentioned, Grace is a brand new product that has never been on NVIDIA's road map. And we already see great success with customers who love the architecture of it and desperately in need of the type of capability that Grace brings. And this should be a pretty exciting year for new product launches. And so we're preparing for all of that laying the foundation for us to bring all those exciting products to the marketplace.

Operator

Next, we'll go to Tim Arcuri with UBS.

Q - Tim Arcuri {BIO 3824613 <GO>}

Thank you very much. Obviously, there's a lot more talk from you about software and I think it's still kind of a little bit of a black box for a lot of investors. And I know, Jensen, that you've talked about software as a medium to basically open up new markets. But I'm wondering maybe if you can sort of quantify how big the software licensing revenue is today? And maybe when you might start to break it out like you did data center, which really got the stock moving in a huge, huge way. Thanks.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. NVIDIA is a software-driven business. Accelerated computing is a software-driven business. It starts from recognizing what domain of applications we want to accelerate and can accelerate, and then building an entire stack from the processor to the system, to the system software, the acceleration engines, and potentially even the applications itself, like the software that we were mentioning earlier, NVIDIA DRIVE, NVIDIA AI and NVIDIA Omniverse.

These are applications that sit on top of system software and are really valuable to the marketplace. The way to think about our software licensing -- so we've always been a software-driven business. But for the very first time, we have packaged licensable software on -- available to customers. The way that we license software for NVIDIA AI Enterprise is per node of server.

There's some 20 million, 25 million servers that are installed in the world today in enterprises, not including clouds. We believe that every single server in the future will be running AI software. And we would like to offer an engine that enables enterprises to be able to use the most advanced, the most trusted, the most utilized AI engine in the world. And so that is essentially the target market, if you will, for NVIDIA AI.

The NVIDIA Omniverse is targeting, is designed for creators contributing content to a virtual world and connect it to robots that are contributing to content in a virtual world. And so it's based on connections. There are 40 million designers and creators around the world. There are going to be hundreds of millions of robots. Every single car will essentially be a robot someday. And those are connections that will be connected into a

digital twin system like Omniverse. And those are -- so the Omniverse business model is per connection per year.

And in the case of NVIDIA DRIVE, we share the economics of the software that we deliver, and if it's AV software or parking software or cabin-based AI software, whatever the licensing is or whatever the service, if it's a upfront license, we share the economics of that. If it's a monthly service subscription, we share the economics of that. But basically, for the cars that we are a part of, that we're developing the end-to-end service, we will get the benefits of the economics of that for the entire life of the fleet of the car. And so you could imagine, what 10 million cars with a modern car lifetimes of 10 to 20 years, the economics and the market, the installed opportunity is quite high. And so our business opportunity is based on those factors.

Our software business really, really started several years ago with virtual GPUs, but this year was when we really stepped it up and offered for the very first time NVIDIA AI Enterprise, Omniverse, and DRIVE. And so I -- watch the spot, I think this is going to be a very significant business opportunity for us, and we look forward to reporting on it.

Operator

Next we'll go to Vivek Arya with Bank of America.

Q - Vivek Arya {BIO 6781604 <GO>}

Thanks for taking my question. Jensen, in the past, you mentioned about a 10% or so adoption rate for AI among your customer base. I was hoping you would quantify where we are in that adoption curve that you tend to differentiate between the adoption differences between your hyperscale and enterprise customers? And then kind of related to that, is there an inorganic element to your growth now that you have over $20 billion of cash on the balance sheet? How are you planning to deploy that to kind of accelerate your growth also?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. The applications for AI is unquestionably growing, and it's growing incredibly fast. Whether in enterprises and financial services, it could be fraud detection, in cases of consumer pointing businesses, customer service, conversational AI, where people are calling chat bots. But in the future, every website will have a chat bot, every phone number will have a chat bot that -- whether it's a human in the loop or not human in the loop, will have a chat bot. And so customer service will be heavily, heavily supported by artificial intelligence in the future.

Almost every point of sales, I think, whether it's a fast foods or a quick service, businesses are going to have chat bots and AI-based customer service. Retail checkouts will be supported by AI agents. And so all of this is made possible by a couple of breakthroughs -- computer vision, of course, because the agents, the AIs have to make eye contact and recognize your posture and such, recognize speech, understand the context, and what is being spoken about, and have a reasonable conversation with people so that you could provide good customer service.

The ability to have human in the loop is one of the great things about an AI, much, much more so than a recording, which obviously is not intelligent and therefore it's difficult to, if you will, call your manager or call somebody to provide services that the AI can't. And so the number of different applications that have been enabled by natural language understanding in customer service in just the last couple of years has grown tremendously.

I think we're -- we remain early days in our AI adoption. It's incredible how fast this has grown and how many different applications are now possible with AI. It pretty much says that almost all future software will be written with AI or by AI. And when it's done, it will be an AI. And we see it in all these different industries. And so I'm pretty certain we're in the early innings yet of AI, and this is going to be one of the largest industries of software that we have ever known.

With respect to capital, we -- as you know, we had just terminated our Arm agreement. We have a regular capital strategy process, and we'll go through that, and we'll make the best judgment about how to use our capital in helping our growth and sustaining our growth and accelerating our growth, and we'll have all of those sensible conversations during those capital allocation meetings. We're just delighted to have so much capital. And so just to put it out there.

Operator

And next we'll go to Aaron Rakers with Wells Fargo.

Thanks, guys. This is Michael on behalf Aaron. Can you guys talk about how the launch of the RTX 3050 is going so far? And maybe more broadly, your view of where we are in the Ampere product cycle on gaming? Thanks.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Thanks, Michael. Let's see. We -- RTX is an unqualified home run. RTX completely reinvented modern computer graphics. It made -- it brought forward ray tracing about a decade earlier than anybody thought possible. The combination of RTX with artificial intelligence, which enabled this technology we call DLSS, is able to not only do a ton more computation using our Shader processors, but also engage the powerful Tensor Core processors that we have in our GPUs to generate images, beautiful images.

RTX is being adopted by just by every game developer on the planet now. It's being adopted by just about every design tool on the planet now. And if not for RTX, Omniverse wouldn't be possible. We wouldn't be able to do physically based path tracing and simulate sensors like radars and LiDARs and ultrasonics and of course, cameras, and simulate these cameras physically and still be able to deliver the type of performance that we deliver. And so RTX was a game changer for the industry. It reset modern computer graphics and it was an enabler for us to build an entire new platform from called Omniverse.

We're about, I think about a third of the way through upgrading an installed base that is growing. You know that video games is now the world's largest gaming genre. And Steam, over the last two years, has grown by 50%. The number of concurrent players on Steam has grown tremendously. And in just the last couple of years, a brand-new game store from Epic came on, and it's already a multi-hundred million dollar business. I think it's close to $1 billion that they're doing incredibly well. I'm so happy to see it. And so the overall gaming market is growing and it's growing quite nicely.

But the thing that, in addition to resetting computer graphics for our entire installed base, the growing of our installed base because gaming is growing. There are a couple of other growth dynamics that are associated with GeForce and RTX that's really quite brand new. One of them is hybrid work. This is a permanent condition. And we now are seeing across the board people who are designers and creators, now have to set up essentially a new workstation or new home workstation design studio so that they could do their work at home. In addition, the creative economy, the digital economy, the creative economy is really, really doing fantastically because everything has to be done in 3D now. Print ads are done in 3D. So 2D print is done in 3D, video is done in 3D, and live video, broadcast video, the millions of influencers now augment their broadcast with rich augmented reality and 3D graphics.

And so 3D graphics is now not just for video games and 3D content, it's actually used now for all forms of digital content creation. And so RTX has all of these different drivers working behind it, and we're definitely in the early innings of RTX.

Operator

Next, we'll go to Stacy Rasgon with Bernstein Research.

Q - Stacy Rasgon {BIO 16423886 <GO>}

Hi guys. Thanks for taking my question. So you said that the growth in the next quarter is about $450 million, give or take, driven by data center. Can you give us some feeling for how that growth is being driven by units versus pricing versus mix? And how those drivers might differ between gaming and data center, if at all, for Colette?

A - Colette Kress {BIO 18297352 <GO>}

It's really early in the quarter to determine, Stacy, our exact mix that we will have based on the unit and an ASP across the board [ph]. Our overall growth quarter-over-quarter going into Q1 will be driven by data center primarily. We will see a little bit of growth there in gaming. I think that's important to understand that even after Q4 holiday moving into Q1, we'll still probably see growth there in gaming, which is different in terms of what we've seen seasonally.

And we'll probably have growth in automotive as well sequentially between Q4 and Q1. There are still some areas that are supply constrained. We are working again to try and improve that for every quarter going forward, but that's how you should look at our earnings for Q1 primarily from data center.

FINAL

Bloomberg Transcript

Operator

Next we'll go to Harlan Sur with JPMorgan.

Q - Harlan Sur {BIO 6539622 <GO>}

Good afternoon. Congratulations on the solid results and execution. The networking connectivity portfolio addition has been pretty solid for the NVIDIA team, especially, in enabling scaling of your GPU systems, improving connectivity bottlenecks in yours and your customers' accelerated compute platforms. So in a year where soft spending is growing 30%, you've got a strong networking upgrade cycle, which is good for your NIC products and just continued overall good attach rates. If the team can unlock more supply, will the networking connectivity business grow in line or faster than the overall data center business this year? And then for Jensen, have you driven synergies between Mellanox's leadership in networking connectivity, you know, for example, leveraging their capabilities for your internally developed NVLink connectivity and switching architectures?

A - Jensen Huang {BIO 1782546 <GO>}

Yes, absolutely. If not for the work that we did so closely with Mellanox, the scalability of DGX and DGX SuperPOD and the Research supercomputer that was just installed at Meta would just not be possible. The concepts of overlapping networking and compute, moving some -- moving computing into the fabric, into the network, the work that we're doing with Synchronoss and Precision Timing, so that we could create Omniverse computers that obey the laws of physics and space time, these things are just simply not possible.

The work that we're doing to bring cloud-native, secure multi-tenancy to supercomputing wouldn't have been possible. The number of innovations that -- are countless. And so I am so thrilled with the combination and so thrilled with the work the Mellanox team are doing. We've accelerated road maps as a result of the combination that we could leverage a much larger base of chip design. BlueField's road map has been accelerated probably by about a year. The switch in -- the Quantum switch and the Spectrum switch, the SerDes are absolutely world-class, shared between Ethernet and InfiniBand and NVLink, absolutely the best SerDes in the world.

And so the list of opportunities or the list of combination benefits is really quite countless. And so I'm super thrilled with that. With respect to networking growth, we should be growing. If we weren't supply constrained, we should be growing faster than overall CSP growth. And the reason for that, the reason for that is two-folds.

The first is, because the networking leadership position of Mellanox, Mellanox is highly heavyweight in the upper end of networking, where the adoption of higher-speed networks tends to move. And so it's sensible that as new data centers are built, the first preference is to install it with higher-speed networking than last-generation networking. And Mellanox's networking technology is unambiguously world class.

The second reason is because the areas where the overall NVIDIA is strong, has to do with the areas that are growing quite fast, which related to artificial intelligence or cloud AI and

such. And so those different applications are growing faster than the core. And so it would be sensible that we have the opportunity as we expand our supply base to continue to grow faster than CSPs overall.

Operator

Our next question will come from Matt Ramsay with Cowen.

Q - Matt Ramsay {BIO 17978411 <GO>}

Yes. Thank you very much. Good afternoon. Jensen, I maybe wanted to expand on some of the things that you were just speaking about in your last answer with respect to the data center business. It's not often maybe ever that you have both x86 server vendors having new big platform upgrades in the same year, which will probably happen later this year. There's a lot going on there, DDR5, PCIe5, some CXL stuff. I wonder if you could talk a bit about your data center business broadly? And what you feel might be memory and I/O constrained currently that these systems might unlock for you, both in the cloud and enterprise side, but also in the DGX business? Thanks.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Thanks, Matt. The -- there are several bottlenecks, and let me just highlight some of them. One of the largest bottlenecks is memory speed. And memory speed, that's the reason why we use the fastest memories in the world, HBM and GDDR, et cetera, et cetera. We are the largest consumers of the fastest memories in the world and not even by -- there's not -- with no close second that I know. And so our consumption of fast memories is important to the work that we do.

The second is networking performance. It is the reason why we have the fastest networks. It is also the reason why we have the most fastest networks in any system. We will have, for example, 8 InfiniBand at the highest speeds connected right into a HGX or DGX server. And so the work that we do in GPU direct memory, RDMA, the work that we do with GPU direct storage, the work that we do with in-network computing and all reductions and moving data around inside the network, is absolutely world-class. This is an area that we are just -- I am just incredibly proud.

All of that is so that we could be less bottlenecked by the CPU. Remember, inside our DGX system is one CPU and eight GPUs. And the fundamental goal is to offload as much as we can and utilize the resources that we have as much as we can.

This year, we expect a transition in PCIe Gen 4 to Gen 5. We are constrained on Gen 4. We'll be constrained on Gen 5, but we're used to that. And that's something that we're very good at. And we'll continue to support Gen 4 well through next year, maybe well through the next couple of years. And all of the installed base of Gen 4 systems that are going to be all over the world, and we'll take advantage of Gen 5 as much as we can. But we have all kinds of new technologies and strategies to improve the throughput of systems and avert the bottlenecks that are there.

Operator

Our final question comes from the line of Raji Gill with Needham and Company.

Q - Raji Gill

Yes. Thanks. And congrats on the good quarter and guide. Colette, a question on the gross margin and to Jensen's point about really creating a software business driven by Omniverse, DRIVE and Enterprise. When you're kind of contemplating your margin profile over the next couple of years, how do we think about that? Is it really going to be driven by an increasing mix of software as a percentage of your revenue over time? Is there more margin upside on the hardware side in terms of some of your segments? The software opportunity is very exciting, but I'm just curious how that would translate to your kind of more of a longer-term margin profile? Thank you.

A - Colette Kress {BIO 18297352 <GO>}

Yes. Thanks for the question on gross margin and the long-term. When we think about the long-term gross margin, we have incorporated software in many of our platforms even today, and in our high-valued platforms in data center, ProBiz with the parts of [ph] our business, have really helped us with our gross margin to this point, and we've done a really solid job of managing that and the growth over the years.

I believe these businesses will continue to be a growing opportunity for us, but now also with the ability to package up software and (inaudible) as that scales with our Enterprise customers, in the data center and with our already procured deals, a lot of (Technical Difficulty) we've got a great opportunity (Technical Difficulty) in the future and (Technical Difficulty) margin (Technical Difficulty). So we're going to work on that. We've set the stage for having been able to package it up to be able to sell it separately to create the business model, to create the partners that are helping us sell it. But yes, we do believe this will be a driver in the long-term.

Operator

Thank you. I'll now turn it back over to Jensen Huang for closing remarks.

A - Jensen Huang {BIO 1782546 <GO>}

Thanks, everyone.

The tremendous demand for our computing platforms, NVIDIA RTX, NVIDIA HPC, and NVIDIA AI, drove a great quarter, capping a record year. Our work propels advances in AI, digital biology, climate sciences, gaming, creative design to autonomous vehicles and robotics, at some of today's most impactful fields. Our open computing platform, optimized across the full stack, architected for data center scale, is adopted by customers globally, from cloud to core to edge and robotics. I am proud of the NVIDIA operations team as we make substantial strides in broadening our supply base to scale our company and better serve customer demand.

And this year, we introduced new software business models with NVIDIA AI Enterprise, NVIDIA Omniverse, and NVIDIA DRIVE.

NVIDIA DRIVE is a full stack end-to-end platform that serves the industry with AV chips, data center infrastructure for AI and simulation, mapping, and the autonomous driving application service. Our data center infrastructure is used by just about anybody building AVs, robotics, robotaxis, shuttles, and trucks. EV companies have selected our Orin chip across the world. And our partnership with Mercedes-Benz and Jaguar Land Rover has opened up a new software and services business model for millions of cars for the life of the fleet.

NVIDIA Omniverse is a world simulation engine that connects simulated digital worlds to the physical world. Omniverse is a digital twin, a simulation of the physical world. The system can be a building, a factory, a warehouse, a car, a fleet of cars, a robotic factory orchestrating a fleet of robots, building cars that are themselves robotic. Today's internet is 2D and AI's in the cloud. The next phase of internet will be 3D and AI will be connected to the physical world. We created Omniverse to enable the next wave of AI, where AI and robotics touches our world.

Omniverse can sound like science fiction, but there are real-world use cases today. Hundreds of companies are evaluating Omniverse. We can't wait to share more of our progress at next month's GTC, learn about new chips, new computing platforms, new AI and robotic breakthroughs, and the new frontiers of Omniverse. Hear from the technologists of Deloitte, Epic Games, Mercedes-Benz, Microsoft, Pfizer, Sony, Visa, Walt Disney, Zoom, and more. This GTC promises to be our most exciting developers conference ever.

We had quite a year, yet nothing makes me more proud than the incredible people who have made NVIDIA one of the best companies to work for and the company where they do their life's work. We look forward to updating you on our progress next quarter. Thank you.

Operator

This concludes today's conference call. You may now disconnect.