Q1 2023 Earnings Call

Company Participants

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

Other Participants

- Aaron Rakers
- Ambrish Srivastava
- C.J. Muse
- Chris Caso
- Harlan Sur
- Mark Lipacis
- Matt Ramsay
- Stacy Rasgon
- Tim Arcuri
- 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 First 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 first quarter of fiscal 2023. 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 Investor Relations website. The webcast will be available for replay until the conference call to discuss for

financial results for the second 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, May 25, 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 delivered a strong quarter driven by record revenue in both Data Center and gaming with strong fundamentals and execution against a challenging macro backdrop. Total revenue of $8.3 billion was a record up 8% sequentially and up, 46% year-on-year. Data Center has become our largest market platform and we see continued strong momentum going forward.

Starting with gaming, revenue was $3.6 billion rose 6% sequentially and 31% year-on-year, powered by the GeForce RTX 30 Series product cycle. Since launching in the fall of 2020, the RTX 30 series has been our best gaming product cycle ever.

The gaming industry has grown tremendously with 100 million new PC gamers added in the past two years according to new Zoom. And NVIDIA RTX has set new standard for the industry with demand from both first time GPU buyers, as well as those upgrading their PCs to experience the 250 plus RTX optimized games and apps, doubled from last year.

We estimate that almost a third of the GeForce Gaming GPU installed base is now on RTX. RTX has brought tremendous energy into the gaming world and has helped drive a sustained expansion in our higher-end platforms and installed base with significant runway still ahead.

Overall end demand remained solid, through though mixed by region and demand in America has remain strong. However, we started seeing softness in parts of Europe related to the war in the Ukraine, and parts of China due to the COVID lockdowns. As we expect some ongoing impact, as we prepare for a new architectural transition later in the year, we are projecting gaming revenue to decline, sequentially in Q2.

Channel inventory has nearly normalized, and we expected to remain around these levels in Q2. The extent in which cryptocurrency mining contributed to gaming, demand is difficult for us to quantify with any reasonable degree of precision.

The reduced pace of increase in ethereum network cash rate, likely reflects lower mining activity on GPUs. We expect a diminishing contribution going forward. Laptop gaming revenue posted strong sequential and year-on-year growth driven by the ramp of the NVIDIA RTX 30 Series line up with this year's Spring Refresh and ahead of the upcoming back-to-school season. There are now over 180 laptop models featuring RTX 30 Series GPUs and our energy efficient thin and light Max-Q technologies, up from 140 at this time last year.

Driving this growth are not just gamers, but also the fastest growing category of content creators from whom we offer dedicated NVIDIA studio drivers. We've also developed applications and tools to empower artists from -- numerous for advanced 3D in collaboration to broadcast for live streaming to canvas for painting landscapes with AI. The creator economy is estimated at $100 billion and powered by 80 million individual creators and broadcasters.

We continued to build out our GeForce NOW cloud gaming service. Gamers can now access RTX 3080-class streaming, our new top tier offering with subscription plans of $1,999 a month.

We added over 100 games to the GeForce NOW library, bringing the total to over 1,300 games. And last week, we launched Fortnight on GeForce NOW, with touch controls for mobile devices, streaming through the Safari, web browser on iOS and the GeForce NOW android app.

Moving to pro visualization. Q1 revenue was $622 million, were down sequentially 3%, and up 67% from a year ago. Demand remains strong as enterprise has continued to build out their employee's remote, office infrastructure to support hybrid work.

Sequential growth in the mobile workstation GPUs was offset by lower desktop revenue. Strong year-on-year growth was supported by the Nvidia RTX Ampere architecture product cycle. Top use, cases include, digital content creation, at customers, such as Sony pictures animation, and Medical Imaging at customers, such as Medtronic.

In just its second quarter of general availability, our Omniverse enterprise software is being adopted by some of the world's largest companies. Amazon is using Omniverse to create digital twins to better optimize warehouse design and flow, and to train more intelligent robots.

Kroger is using Omniverse to optimize store efficiency with digital twin store simulation. And PepsiCo is using Omniverse digital twins to improve the efficiency and environmental sustainability of its supply chain. Omniverse is also expanding our GPU sales pipeline, driving a higher-end and multiple GPU configurations. The Omniverse ecosystem continues to rapidly expand with third-party developers in the robotics, industrial

FINAL

Bloomberg Transcript

automation, 3D design, and rendering ecosystems, developing connections to Omniverse.

Moving to Automotive. Q1 revenue of $138 million increased 10% sequentially and declined 10% from the year-ago quarter. Our DRIVE Orin SOC is now in production and kicks off a major product cycle with auto customers ramping in Q2 and beyond. Orin has great traction in the marketplace with over 35 customer wins from automakers, truckmakers, and robotaxi companies.

In Q1 BYD, China's largest EV maker and Lucid, an award-winning, EV pioneer were the latest to announce that they are building their next generation fleet on DRIVE Orin. Our automotive design win pipeline, now exceeds $11 billion over the next six-years, up from $8 billion just a year ago.

Moving to Data Center. Record revenue of $3.8 billion grew 15% sequentially and accelerated to 83% growth year-on-year. Revenue from hyperscale and cloud computing customers more than doubled year-on-year, driven by strong demand for both external and internal workloads.

Customer's remains supply constraint in their infrastructure needs and continue to add capacity as they tried to keep pace with demand. Revenue from vertical industries, grew a strong, double digit percentage from last year. Top verticals driving growth this quarter include consumer internet companies, financial services, and telecom.

Overall Data Center growth was driven, primarily by strong adoption of our A100 GPU for both training and inference with large volume deployments by hyperscale customers and broadening adoption across the vertical Industries. Top workloads includes recommender systems, conversational AI, large language models, and cloud graphics.

Networking revenue working revenue accelerated on strong, broad-based demand for our next generation 25, 50 and 100 gig ethernet adapters. Customers are choosing in videos, networking products for their leading performance and robust software functionality.

In addition, networking revenue is benefiting from growing demand for DGX, super pods, and cross-selling opportunities. Customers are increasingly, combining our compute and networking products to build what are essentially modern AI factories with data as the raw material input and intelligence as the output.

Our networking, products are still supply constraint, though we expect continued improvement throughout the rest of the year. One of the biggest workloads, driving adoption of Nvidia AI is natural language processing which has been revolutionized by transformer based model, recent industry breaker troughs, traced to transformers include large language models like GPT-3, NVIDIA MegaMolBART for direct discovery and deep mine alpha fold or protein structure prediction.

FINAL

Bloomberg Transcript

Transformers allow self-supervised learning without the need for human labeled data. They enable unprecedented levels of accuracy for tasks such as text generation, translation, summarization and answering questions. To do that, transformers use enormous training data sets and very large neural networks well into the hundreds of billions of parameters. To run these giant models without sacrificing low inference times, customers like Microsoft are increasingly deploying NVIDIA AI, including our NVIDIA Ampere Architecture based GPUs and full software stack.

In addition, we are seeing a rising wave a Customer Innovation using large language models that is driven by increased demand for NVIDIA AI and GPU instances in the cloud. At GTC, we announced our next generation Data Center GPU, the H100 based on the new, or upper architecture packed with 80 billion transistors H100, is the world's largest most powerful accelerator offering an order of magnitude leap in performance over the A100.

We believe, H100 is hitting the market at the perfect time. H100 is ideal for advancing large language models and deep recommender systems. The two largest scale, AI workloads today. We are working with leading server makers and hyperscale customers to qualify and ramp H100. It as well as the new DGX H100 AI -- computing system will ramp in volume late in the calendar year.

Building on the H100 products sale, we are on track to launch our first-ever data center CPU Grace in the first half of 2023. Grace is the ideal CPU for AI factories. This week at COMPUTEX, we announced that dozens of server models based on Grace will be brought to market by the first wave of system builders, including Asus, Foxconn, Gigabyte, 2CTA [ph], Supermicro and Wyvern.

These servers will be powered by the NVIDIA Grace CPU Superchip, which features two CPUs and the Grace Hopper Superchip, which pairs an NVIDIA Hopper GPU with an NVIDIA Grace CPU in an integrated model. We've introduced new reference designs based on Grace for the massive new workloads of Next Generation Data Centers.

CGX for cloud graphics and gaming, OVX for digital twins or Omniverse, and HGX for HPC and AI, these server designs are all optimized for NVIDIA's rich accelerated computing software stacks and can be qualified as part of our NVIDIA certified systems line up.

The enabler for the Grace Hopper and Graces Superchips is our ultra-energy efficient, low latency, high speed memory coherent interconnect called NVLink which scales from die-to-die, chip-to-chip and system-to-system. With NVLink we can configure Grace and Hopper to address a broad range of workloads.

Future Nvidia chips, the CPUs, GPUs, EPUs necks and SOCs will integrate NVLink just like Grace and Hopper based on our world-class 3D technology. We are making NVLink open to customers and partners to implement custom chips that connect to NVIDIA's platforms.

In networking, we're kicking off a major product cycle with the introduction of Spectrum-4, the world first, 400Gbps end-to-end ethernet networking platform, including the

FINAL

Bloomberg Transcript

Spectrum-4 switch connect X7 smart neck, who filled three DPU on the DOCA software.

Build for AI Nvidia Spectrum-4 video spectrum arrives as Data Centers are growing exponentially and demanding extreme performance, advanced security, and powerful features to enable high performance, advanced virtualization, and simulation at scale. Across our businesses, we are launching multiple new GPU, CPU, DPU, and SOC products, over the coming quarters with a ramp in supply to support the customer demand.

Moving to the rest of the P&L. GAAP gross margin for the first quarter was 65.5% and non-GAAP gross margin was up, 67.1%, up 90 basis points from a year ago and up 10 basis points sequentially. We have been able to offset rising costs and supply chain pressures, we expect to maintain gross margins at current levels in Q2.

Going forward as new products ramp and software becomes a larger percent of revenue, we have opportunities to increase gross margins longer-term. GAAP operating margin was 22.5% impacted by a $1.35 billion acquisition termination charge related to the arm transaction.

Non-GAAP operating margin was 47.7%. We are closely managing our operating expenses to balance the current macro environment with our growth opportunities, and we've been very successful in hiring so far this year and are now slowing to integrate these new employees. This also enables us to focus our budget and taking care of our existing employees as inflation persist.

We are still on track to grow our non-GAAP operating expenses in the high 20s range this year. We expect sequential increases to level off after Q2 as the first half of the year includes a significant amount of expenses related to the bring up of multiple new products, which should not reoccur in the second half.

During Q1, we repurchase $2 billion of our stock. Our Board of Directors increased and extended our share repurchase program to repurchase an additional common stock up to a total of $15 billion through December 2023.

Let me now turn to the outlook for the second quarter of fiscal 2023. Our outlook assumes an estimated impact of approximately $500 million relating to Russia and China COVID lockdowns. We estimate the impact of lower sell-through in Russia and China to affect our Q2 gaming selling by $400 million. Furthermore, we estimate the absence of sales to Russia to have a $100 million impact on Q2 in Data Center.

We expect strong sequential growth in Data Center and automotive to be more than offset by these sequential decline in gaming. Revenue is expected to be $8.1 billion plus or minus 2%. GAAP and non-GAAP gross margins are expected to be 65.1% and 67.1% respectively plus or minus 50 basis points.

GAAP operating expenses are expected to be $2.46 billion, non-GAAP operating expenses are expected to be $1.75 billion. GAAP and non-GAAP other income and

expenses are expected to be an expense of approximately $40 million, excluding gains and losses on non-affiliated investments.

GAAP and non-GAAP tax rates are expected to be 12.5% plus or minus 1%, excluding discrete items, and capital expenditures are expected to be approximately $400 million to $450 million. Further financial details are included in the CFL commentary and other information available on our IR website.

In closing, let me highlight the upcoming events for the financial community. We'll be attending at the BofA Securities Technology Conference in person on June 7, where Jensen will participate in a keynote fireside chat. Our earnings call to discuss the results of our second quarter of fiscal 2023 is scheduled for Wednesday, August 24.

We will now open the call for questions. Operator, could you please pull for questions? Thank you.

Questions And Answers

Operator

(Question And Answer)

Thank you. (Operators instructions) We'll take our first question from C.J. Muse with Evercore ISI. Your line is open.

Q - C.J. Muse

Yes, good afternoon. Thank you for taking the question. I guess would love to get an update on how you're thinking about the Gaming cycle from here. The business has essentially doubled over the last two years. And now we've got some crosswinds with crypto falling off, channel potentially clearing ahead of a new product cycle.

You talked about macro challenges. But at the same time, only one-third of the installed base has RTX and we're moving out from under supply. So we'd love to hear your thoughts from here once we get beyond kind of the challenges around COVID lockdown in the July quarter. How are you thinking about Gaming trends?

A - Jensen Huang {BIO 1782546 <GO>}

Yes. C.J., thanks for question. The -- you captured a lot of the dynamics well in your question. The underlying dynamics of the Gaming industry is really solid, net of the situation with COVID lockdown in China and Russia. The rest of the market is fairly robust. And we expect the Gaming dynamics to be intact. The several things that are driving the Gaming industry.

In the last two years alone, 100 million new gamers came into the PC industry. The format has expanded tremendously. And the ways that people are using their PCs to connect

with friends, to be an influencer as a platform for themselves, use it for broadcast. So many people are now using their home PCs as their second workstation, if you will, second studio.

Because they're also working from home. It is our primary way of communicating these days. The need for G4 PCs have never been greater. And so I think the fundamental dynamics are really good. And so as we look into the second-half of the year, we look -- it's hard to predict exactly what -- when COVID and the war in Russia is going to be behind us. But nonetheless, the governing dynamics of the Gaming industry is great.

Operator

Next, we'll go to Matt Ramsay with Cowen. Your line is open.

Q - Matt Ramsay {BIO 17978411 <GO>}

Thank you very much. Good afternoon. Jensen, I wanted to ask a bit of a question on the Data Center business. In this upcoming cycle with H100, there's some I/O upgrades that are happening in servers that I think are going to be a fairly strong driver for you in addition to what's going on with Hopper and the huge performance leaps that are there.

I wanted to ask a longer-term question, though, around your move to NVLink with Grace and Hopper and what's going on with your whole portfolio. Do you envision the business continuing to be sort of card-driven attached to third-party servers, or do you think revenue shifts dramatically or in a small way, over time, to be more sort of vertically integrated all of the chips together on NVLink. And how is the industry sort of responding to that potential move? Thanks.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. I appreciate the question. The -- let's see, the first point that you made is a very big point. The next generation of servers that are being teed up right now are all gen 5. The I/O performance is substantially higher than what was available before.

And so you're going to see a pretty large refresh as a result of that. Brand-new networking cards from our company and others. Gen 5, of course, drives new platform refresh. And so we're perfectly timed to ramp into the Gen 5 generation with Hopper.

There are a lot of different system configurations you want to make. If you take a step back and look at the type of systems that are necessary for data processing, scientific computing, machine learning and training, inference done in the cloud for hyperscale nature, done on-prem for enterprise computing done at the edge.

Each one of these workloads and deployment locations, the way that you manage would dictate a different system architecture. So there isn't one size that fits all, which is one of the reasons why it's so terrific that we support PCI Express that we innovated chip-to-chip interconnect for that before anybody else did, this is now some seven years ago, we're in our fourth generation of NVLink that allows us to connect two chips next to each other,

two dies, two chips, two modules, two SXM modules to two systems to multiple systems. And so our coherent chip-to-chip link, NVLink has made it possible for us to mix and match chips, dies, packages, systems and all of these different types of configurations.

And I think that, over time, you're going to see even more types of configurations. And the reason for that has to do with a couple of very important new type of data centers that are emerging. And you're starting to see that now with fairly large installations, infrastructures with NVIDIA, HPC and NVIDIA AI. These are really AI factories where you're processing the data, refining the data and turning that data into intelligence. These AI factories are essentially running one major workload and they're running at 24/7. And deep recommender systems is a good example of that.

In the future, you're going to see large language models essentially becoming a platform themselves. That would be running 24/7, hosting a whole bunch of applications. And then on the other end, you're seeing data centers at the edge that are going to be robotics or autonomous data centers that are running 24/7. They are going to be running in factories and retail stores and warehouses, logistics warehouses all over the world. So these two new type of data centers are just emerging, and they also have different architectures.

So I think the net of it all is that our ability to support every single workload because we have a universal accelerator running every single workload from data processing to data analytics to high-performance computing to training to inference that we can support ARM and x86 that we support PCI to multisystem NVlink to multi-chip NVLink to multi-die NVLink, that capability for us is -- makes it possible for us to really be able to serve all of these different segments.

With respect to vertical integration, I think that system integration, the better way to make me saying that is that system integration is going to come in all kinds of different ways. We're going to do semi-custom chips as we've done with many companies in the past, including Nintendo. We'll do semi-custom chiplets as we do with NVLink. NVLink is open to our partners. -- and they could bring it to any fab and connect it coherently into our chip. We could do multi module packages. We could do multi-package systems. So there's a lot of different ways to do system integration.

Operator

Next, we'll go to Stacy Rasgon with Bernstein Research. Your line is open.

Q - Stacy Rasgon {BIO 16423886 <GO>}

Hi guys. Thanks for taking my question. I wanted to follow up on the sequential. So Colette, I know you said the $500 million was a $400 million hit to Gaming and a $100 million hit to data. So I'm assuming that -- that doesn't mean the Gaming is down $400 million. I mean it's Gaming -- do you see Gaming actually down more than the actual Russia and lockdown hit. And I guess just how do I think about the relative sequentials of the businesses in light of those constraints that you guys are facing?

A - Colette Kress {BIO 18297352 <GO>}

FINAL

Sure. Let me start first with what does that mean to gaming. What does that mean to Gaming for -- we do expect Gaming to decline into Q2. We still believe our end demand remains very strong. Ampere has just been a great architecture, and there's many areas where we continue to see strength and growth in both our sell-through and probably what we will see added into that channel as well.

But in total, Q2 Gaming will decline from last quarter from Q1 that it will probably decline in the teens. As we try and work through some of these lockdowns in China, which are holding us up. So overall, the demand for Gaming is still strong. We still expect end demand to grow year-over-year in Q2.

Operator

Next we'll go to Mark Lipacis with Jefferies. Your line is open.

Q - Mark Lipacis {BIO 2380059 <GO>}

Hi. Thanks for taking my question. If you listen to the networking OEMs, this earnings season, it seems that there was a lot of talk about increased spending by enterprises on their Data Centers and sometimes you hear them talking about how this is being driven by AI. You talked about your year-over-year growth in your cloud versus enterprise spending. I wonder if you could talk about what you were seeing sequentially. Are you seeing a sequential inflection in the enterprise. And can you talk about the attach rate of software for enterprise versus data centers. And which software is -- are you seeing the most interest. I know you talked about, is it Omniverse. Is it natural language processing. Or is there one big driver. Or is it a bunch of drivers for the various different software packages you have. Thank you.

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Thanks, Mark. We had a record Data Center business this last quarter. We expect to have a record -- another record quarter this quarter, and we're fairly enthusiastic about the second-half. AI and data driven machine learning techniques for writing software and extracting insight from the vast amount of data that companies have is incredibly strategic to all the companies that we know. Because in the final analysis, AI is about automation of intelligence and most companies are about domain-specific intelligence.

We want to produce intelligence. And there are several techniques now that have been created to make it possible for most companies to apply their data to extract insight and to automate a lot of the predictive things that they have to do and do it quickly. And so I think the trend that you hear other people are experiencing about machine learning, data analytics, data-driven insights artificial intelligence.

However, it's described, it's all exactly the same thing. And it's sweeping just about every industry and every company. Our networking business is also highly supply constrained. Our demand is really, really high. And it requires a lot of components aside from just our chips. Components and transceivers and connectors and cables.

And just it's a complicated system, the network, and there are many physical components. And so -- the supply chain has been problematic. We're doing our best and our supply has been increasing from Q4 to Q1. We're expecting it to increase in Q2 and increase in Q3 and Q4. And so we're really, really grateful for the support from the component industry around us, and we'll be able to increase that.

With respect to software, there are two, Well -- first of all, there are all kinds of machine learning models, computer vision, speech AI, natural language understanding, all kinds of robotics applications, the most -- probably the largest -- the most visible one is self-driving cars, which is essentially a robotic AI.

And then recently, this incredible breakthrough from an AI model called Transformers that has led to really, really significant advances in natural language understanding. And so they're all these different types of models. There are thousands and thousands of species of AI models and in all these different industries.

One of my favorite, I'll just say it very quickly and I'll answer that question about the software. One of my favorites is using Transformers to understand the language of chemistry or using transformers and using AI models to understand the language of proteins, amino acids, which is genomics. To apply AI to understand -- to recognize the patterns, to understand the sequence and essentially understand the language of chemistry and biology is a really, really important breakthrough. And all of this excitement around synthetic biology, much of it stands back to the -- some of these inventions.

But anyhow, all of these different models need an engine to run on. And that engine is called NVIDIA AI. In the case of hyperscalers, they can cobble together a lot of open source, and we provide a lot of our source to them and a lot of our engines to them for them to operate their AI. But for enterprises, they need someone to package it together and be able to support it and refresh it, update it for new architecture, support old architectures in their installed base, et cetera, and all the different use cases that they have.

And so that engine is called NVIDIA AI. It's almost like a sequel engine, if you will. And except this is an engine for artificial intelligence. There's another engine that we provide and that engine is called Omniverse and it's designed for the next wave of AI, where artificial intelligence has to not just manipulate information like recommender systems and conversational systems and such. But it has to interact with physical systems. Whether it's interacting with physics directly, meaning robotics or being able to automate physical systems like heat recovery steam generators, which is really important today. And so Omniverse is designed to be able to sit at that interface, that intersection between simulation and artificial intelligence, and that's what Omniverse is about.

Omniverse has now -- let's see some -- we're still early in the deployment of Omniverse for commercial license. It's been a couple of quarters now since we've released Omniverse enterprise. And I think, at this point, we have 10% of the world's top 100 companies that are already customers, licensing customers, substantially more who we're evaluating. I think it's been downloaded nearly 200,000 times.

It has been tried in some 700 companies. And Colette highlighted some of the companies, you might see some of the companies that are using it in all kinds of interesting applications at GTC. And so I fully expect that the NVIDIA AI engine, the Omniverse engine, are going to be very successful for us in the future and contribute greatly to our earnings.

Operator

Next, we'll go to Vivek Arya with BofA Securities. Your line is open.

Q - Vivek Arya {BIO 6781604 <GO>}

Thanks. Just wanted to clarify, Colette, if your Q2 outlook includes any destocking benefits from the new products that you're planning to launch this year. And then Jensen, my question is gentleman for you. You're still guiding Data Center to a very strong, I think, close to 70% or so year-on-year growth despite all the headwinds.

Are you worried at all about all the headlines about the slowdown in the macro economy. I like is there any cyclical impact on Data Center growth that we should keep in mind as we think about the second-half of the year?

A - Colette Kress {BIO 18297352 <GO>}

Yes, Vivek. Let me first answer the question that you asked regarding any new products as we look at Q2. As we discussed about it, most of the ramp that we have of our new architectures, we're going to see in the back half of the year. We're going to start to see, for example, Hopper will probably be here in Q3, but starting to ramp closer to the end of the calendar year. So you should think about most of our product launches to be ramping in the second-half of the year on that part. I'll turn over for Jensen for the rest.

A - Jensen Huang {BIO 1782546 <GO>}

Thanks. Our Data Center demand is strong and remains strong. Hyperscale and cloud computing revenues, as you mentioned, has grown significantly. It's doubled year-over-year. and we're seeing really strong adoption of A100. A100 is really quite special and unique in the world of accelerators. And this is one of the really, really great innovations as we extended our GPU from graphics to CUDA to Tensor Core GPUs. It's now a universal accelerator.

And so you could use it for data processing for ETL, for example, extract, transform and load. You could use it for database acceleration. Many CPL functions are accelerated on NVIDIA GPUs. We accelerate Rapids, we accelerate which is the Python version a Data Center scale version of Pandas, -- we accelerate Spark 3.0. And so from database queries to data processing to extraction and transform and loading of data before you do training and inference and whatever image processing or other algorithmic processing you need to do can be fully accelerated on A100.

And so we're seeing great success there. on at the core and closer to what is happening today, you're seeing several different very important new AI models that are being

invested in at very, very large scale and with great urgency. You probably have heard about Deep Recommender Systems. This is the economic engine, the information filtering engine of the Internet, if not for the Recommender system, it would be practically impossible for us to enjoy our Internet experience shopping experience with trillions of things that are changing in the world every day constantly and be able to use your 3-inch phone to even engage the Internet.

And so all of that magic is made possible by this incredible thing called recommender system. The second thing is conversational AI. You're seeing chat bots and website customer service, even live customer service being now supported by AI, conversational AI has an opportunity to enhance the customer service on the one hand. On the other hand, supplement for a lot of labor shortage. And then the third is this groundbreaking piece of work as related to Transformers that led to natural language understanding breakthrough. But within it, is this incredible thing called large language models, which embeds human knowledge because it's been trained and so much data. And we recently announced Megatron 530B. And it was a collaboration we did with Microsoft, the foundation of, I think they call it Turing. And this language model and others like it, like open AI, GPD three are really transformative and they take an enormous amount of computation.

However, the net result is a pre-trade model that is really quite remarkable. Now we're working with thousands of start-ups, large companies that are building who are using the public cloud. And so it's driving a lot of demand for us in the public cloud. I think we have now 10,000 AI inception startups that are working with us and using NVIDIA AI, whether it's on-prem or in the cloud, it saves money because the computation time is significantly reduced.

The quality of service is a lot better and they could do greater things. And so that's driving AI in the cloud. And so all of these different factors, whether it's just the industrial recognition of the importance of AI, the transformative nature of these new AI models recommender systems, large language models, conversational AI. The thousands of companies around the world that are using NVIDIA AI in the cloud -- driving public cloud demand, all of these things are driving our Data Center growth. And so we expect to see Data Center demand remains strong.

Operator

Next, we'll go to at Tim Arcuri with UBS. Your line is open.

Q - Tim Arcuri {BIO 3824613 <GO>}

Thank you very much. Had a question about this $500 million impact for July and whether it's more supply related or demand related. And that's because most others in semis are sort of setting this China stuff, in particular, is more of a logistics issue, so more of a supply issue. But the language Colette you were using in your commentary cited lower sell-through and gaming and sort of absence of sales and Russia.

To me, that sounds a little more demand, which would make sense in the context of this new freeze on hiring that you have. So I ask, because if it's supply related, then you could

argue, it's not perishable and really just timing. But if demand related, that might never come back, and it's to be the beginning of falling night So I wonder if you can sort of walk through that for me. Thanks.

A - Colette Kress {BIO 18297352 <GO>}

So thanks, Tim for the question. Let me try bet here on the China and Russia. Two very different things. The current China lockdowns that we are seeing interestingly has implications to both supply and demand. We have seen challenges in terms of the logistics throughout the country, things going in out of the country, it puts a lot of pressure, on just logistics that were already under pressure.

From a demand perspective, it is also been head from the gaming side. You have a very large cities that are in full lockdown, focusing really on other important things for the citizens there. So it's impacting our demand. We do believe that they will come out of COVID and the demand for our products will come back. We do believe that will occur. The supply will sort of doubt it's very difficult to determine how.

Now in the case of Russia, we're not selling to Russia. That's something that we had announced earlier last quarter. But there were plans and Russia has been a part of our overall company revenue of probably about 2% of our company revenue historically and a little larger percentage when you look at our Gaming business. I hope that helped.

Operator

Next, we'll go to Ambrish Srivastava with BMO. Your line is now open.

Q - Ambrish Srivastava {BIO 4109276 <GO>}

Hi, thank you very much for letting Jensen, and I actually really appreciate it that you call our demand for most two companies. It seems like it's Harry Percy[ph] to say, the demand is a problem. So refreshing to hear that. I had a question on the second-half and it relates to both Data Center as well as Gaming.

So last couple of times you have talked publicly, you have made comments that your visibility into the Data Center has never been better. So I was wondering if you just take out the Russia impact, is that still true, all the orders that you have been getting there intact and you did say that business will see a strong momentum. I just want to make sure that statement of confidence you have made stays. And then on Gaming, Colette, do we expect second-half to be up year-over-year just based on the guide for second quarter. It seems like it could be up sequentially, but may not return to year-over-year growth in Q3. Thank you

A - Jensen Huang {BIO 1782546 <GO>}

Yes. Ambrish, thanks for the question. On first principles, it should be the case that our visibility of Data Centers is vastly better than a couple of years ago. And the reason for that is several. One, if you recall a couple two, three years ago, deep learning and AI was starting to accelerate in the most computer science deep companies in the world with CSPs and hyperscalers.

But just about everywhere else, it was still quite nascent. And there was a couple of reasons for that. Obviously, the understanding of the technology is not as pervasive at the time. The type of industrial use cases for artificial intelligence requires labeling of data that's really quite difficult. And then now with Transformers, you have unsupervised learning and other techniques, zero-shot learning that allows us to do all kinds of interesting things without having to have human labeled data.

We even have synthetic generated data with Omniverse that helps customers do data generation without having to label data, which is either too costly or quite frankly, oftentimes impossible.

And so now, the knowledge and the technology has evolved to a place that most of the industries could use artificial intelligence at a fairly effective way and in many industries rather transformative.

And so I think, number one, we went from clouds and hyperscalers to all of industries. Second, we went from training focused to inference. Most people thought that inference was going to be easy. It turns out the inference is by far the harder. And the reason for that is because there are so many different models and there are so many different use cases and so many quality of service requirements, and you want to run these inference models in a small of a footprint as you can. And so when you scale out the number of users that use the service is really quite high. So using acceleration using NVIDIA's platform, we could inference any model from computer vision to speech to chemistry to biology, you name it. And we do it so quickly and so fast that the cost is very low. And so the more acceleration you do, the more money you will save. And that, I think, that wisdom is absolutely true.

And so the second dimension is training to inference. The third dimension is that we now have so many different types of configurations of systems that we can go from high-performance computing systems all the way to cloud to on-prem to edge.

And then the final concept is really this industrial deployment now of AI that's causing us to be able to in just about every industry find growth. And so as you know our cloud are growing very, very quickly. However, the vertical part, vertical industries, which is the financial services and retail and telco and all of those vertical industries have also grown very nicely. And so in all of those different dimensions, our visibility should be a lot better.

And then starting a couple of years ago, adding the Mellanox portfolio to our company, we're able to provide a lot more solution oriented end-to-end platform solutions for companies that don't have the scales and don't have the technical depth to be able to stand up these sophisticated systems. And so our networking business is growing very, very nicely as well.

Operator

Next, we'll go to Harlan Sur with JPMorgan. Your line is open.

Q - Harlan Sur {BIO 6539622 <GO>}

Hi, good afternoon. Thanks for let me ask a question. I just want to maybe just ask this question a little bit more directly. So it's good to see the team being able to drive navigate the dynamic supply chain environment, right? You look strong sequential growth in Data Center in April, here in the July quarter, even with some demand impact from Russia, right?

And so as we think about the second-half of the year, cloud spending is strong, and it's actually, I think, accelerating. You're getting ready to ramp H100 later in the year. Mellanox, I think is getting more supply as you move through the year. And in general, I think previously, you guys were anticipating sequential supply and revenue growth for the business through this entire year. I understand the uncertainty around Gaming, but does the team expect continued sequential growth in Data Center through the remainder of the year?

A - Jensen Huang {BIO 1782546 <GO>}

Either one of those answers. The answer is yes. The answer is yes. We see a strong demand in Data Center, hyperscale to cloud computing to vertical industries. Ampere is going to continue to scale out. It's been qualified in every single company in the world. And so after two years, it remains the best universal accelerator on the planet, and it's going to continue to scale out in all these different domains and different markets. We're going to layer on top of that, a brand new architecture Hopper. I'm going to layer on top of that brand new networking architectures. Quantum 3, CX-7, BlueField 3, and we have increasing supply. And so we're looking forward to an excellent quarter next quarter again for Data Centers and going into the second-half.

Operator

Next, we'll go to Chris Caso with Raymond James. Your line is open.

Q - Chris Caso {BIO 4815032 <GO>}

Yes. Thank you. Wonder if you could speak a little bit about the purchase obligations, which seemed like they were up again in the quarter. And how that -- was that a function of longer-dated obligations, or a higher magnitude of obligations. And maybe you could just speak to supply constraints in general. You've mentioned a couple of times in the call about continued constraints in the networking business. What about the other parts of the business? Where are you still constrained?

A - Colette Kress {BIO 18297352 <GO>}

Yes. So let me start here, and I'll see if Jensen wants to add more of that. Our purchase obligations as well as our prepaid have two major things to keep in mind. One, for the first time ever, we are preparing to make sure that we have that supply and those commitments long-term. And additionally, on our purchase obligations, many of them are for long lead time items that are a must for us to procure to make sure that we have the products coming to market.

A good percentage of our purchase commitments is for our Data Center business, which you can imagine are much larger systems, much more complex systems and those things that we are procuring to make sure we can feed the demand both in the upcoming quarters and further. Areas in terms of where we are still a little bit supply constrained are networking. Our demand is quite strong. We've been improving it each time. But yes, we still have demand -- supply concerns with networking still. Is there others that you want to add on Jensen?

A - Jensen Huang {BIO 1782546 <GO>}

No, I thought you were perfect. That's perfect.

Operator

Our final question comes from Aaron Rakers with Wells Fargo. Your line is open.

Q - Aaron Rakers {BIO 6649630 <GO>}

Yes. Thanks for fitting me in. And most of my questions around Gaming and Data Center have been answered. But I guess I'll ask about the Auto segment. While it's still small, clearly, you guys sound confident in that business starting to see significant sequential growth into this next quarter. I'm wondering if you could help us kind of think about the trajectory of that business over the next couple of quarters.

And I think, in the past, you've said that, that should start to really inflect higher as we move into the second-half of the year. Just curious if you could help us think about that piece of the business?

A - Jensen Huang {BIO 1782546 <GO>}

Several data points. We are just starting. We have just started shipping ORAN in the first quarter of shipping production ORAN. ORAN is a robotics processor. It's designed for a software defined robotic card or robotic pick and placer or a robotic mover, logistics mover.

We've been designed into 35-car and trucks and robo taxi companies and more others. If you include logistics movers and last mile delivery systems and farming equipment and the number of design wins for ORAN is really quite fantastic. ORAN is a revolutionary processor. And it's designed as a -- if you will a Data Center on a chip. And it is the first Data Center on a chip that is robotic, processes sensor information, it's safe, it has the ability to be rather resilient as confidential computing. It is designed to be secure, designed to be all those things, because these data centers are going to be everywhere. And so ORAN is really a technological marvels in production.

We experienced very likely the lowest auto quarter for some time. And the reason for that is because over the next six years or so, we have $11 billion and counting of business that we've secured estimated. And so I think it's a fairly safe thing to say now that ORAN and our autonomous vehicle and robotics business is going to be our next multibillion dollar business. It's on its way surely there. The robotics and autonomous systems and autonomous machines, whether they move or not move, but AI systems that are at the

physical edge is surely going to be the next major computing segment. It is surely going to be the next major Data Center segment.

We've been working in this area, as you know, for a decade. We have a fair amount of expertise in this area. And ORAN is just one example of our work here. We have four pillars to our strategy for autonomous systems.

Starting from the data processing and the AI training part of it to train robotics AIs. Second, to simulate robotics AIs, which is Omniverse. Third, to the memory of the robotics AI otherwise known as mapping. And then finally, the actual robotics application and the robotics processor in the system, and that's where ORAN goes. But ORAN is just one of our four pillars of our robotics strategy and the next wave of AI. And so I am really optimistic and really enthusiastic about the next phase of the computer industry's growth. And I think a lot of it is going to be at the edge. A lot of it's going to be about robotics.

Operator

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

A - Jensen Huang {BIO 1782546 <GO>}

Thanks, everyone. The full impact and duration of the war in Ukraine and COVID lockdowns in China is difficult to predict. However, the impact of our technology and our market opportunities remain unchanged.

The effectiveness of deep learning AI continues to stand. The transformer model, which led to the natural language understanding breakthroughs is being advanced to learn patterns with great spatial, sequential and temporal complexity. Researchers are creating transformer models that are revolutionizing applications from robotics to drug discovery. The effectiveness of deep learning AI is driving companies across industries to adopt NVIDIA for AI computing.

We're focused on four major initiatives. First, ramping our next generation of AI infrastructure chips and platforms, Hopper GPU, BlueField DPU, NVLink, InfiniBand, Quantum InfiniBand, Spectrum Ethernet networking. And all this to help customers build their AI factories and take advantage of new AI breakthroughs like transformers.

Second, ramping our system and software industry partners to launch Grace, our first CPU. Third, ramping ORAN, our new robotics processor, and nearly 40 customers building cars, robo taxis, trucks, delivery robots, logistics robots, farming robots to medical instruments.

And fourth, with our software platforms, adding new value to our ecosystem with NVIDIA AI and NVIDIA Omniverse, and expanding into new markets with new CUDA acceleration libraries. These initiatives will greatly advance AI. And while continuing to extend this most impactful technology of our time to scientists in every field and companies in every industry. We look forward to updating you on our progress next quarter. Thank you.

Operator

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

FINAL

Bloomberg Transcript