

Seeking Alpha^α

Transcripts | Technology

NVIDIA Corp.'s (NVDA) CEO Jensen Huang on Q3 2020 Results - Earnings Call Transcript

Nov. 14, 2019 9:12 PM ET7 comments | 10 Likes

by: SA Transcripts

FQ3: 11-14-19 Earnings Summary

 *Press Release*

EPS of \$1.78 beats by \$0.21 | Revenue of \$3.01B (-5.25% Y/Y) beats by \$97.14M

Earning Call Audio

 Subscribers Only

0:00:00

-1:00:26

NVIDIA Corp. (NASDAQ:NVDA) Q3 2020 Earnings Conference Call November 14, 2019
5:30 PM ET

Company Participants

Simona Jankowski - VP, IR

Colette Kress - EVP & CFO

Jensen Huang - Co-Founder, CEO & President

Conference Call Participants

Vivek Arya - Bank of America Merrill Lynch

Aaron Rakers - Wells Fargo Securities

Christopher Muse - Evercore ISI

Harlan Sur - JPMorgan Chase & Co.

Toshiya Hari - Goldman Sachs Group

Stacy Rasgon - Sanford C. Bernstein & Co.

Mitchell Steves - RBC Capital Markets

Joseph Moore - Morgan Stanley

Harsh Kumar - Piper Jaffray Companies

Operator

Good afternoon. My name is Christina, and I'm your conference operator for today. Welcome to NVIDIA's Financial Results Conference Call. [Operator Instructions].

I'll now turn the call over to Simona Jankowski, Vice President of Investor Relations, to begin your conference.

Simona Jankowski

Thank you. Good afternoon, everyone, and welcome to NVIDIA's conference call for the third quarter of fiscal 2020. With me on the call 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 fourth quarter of fiscal 2020. 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 Form 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, November 14, 2019, 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

Thanks, Simona. Q3 revenue was \$3.01 billion, down 5% year-on-year and up 17% sequentially. Starting with our gaming business. Revenue of \$1.66 billion was down 6% year-on-year and up 26% sequentially. Results exceeded our expectations driven by strength in both desktop and notebook gaming. Our GeForce RTX lineup features the most advanced GPU for every price point and uniquely offers hardware-based ray tracing for cinematic graphics. While ray tracing launched a little more than a year ago, two dozen top titles have shipped with it or are on the way. Ray tracing is supported by all the major publishers, including all-star titles and franchise such as Minecraft, Call of Duty, Battlefield, Watch Dogs, Tomb Raider, Doom, Wolfenstein and Cyberpunk. Of note, Call of Duty: Modern Warfare, had a record-breaking launch in late October that came on the heels of CONTROL, an action adventure game with multiple ray trace features. Reviews have praised both for their ray tracing implementation and game-play performance. With last week's PC release of Red Dead Redemption 2 as a strong gaming lineup for the holiday season, our business reflects this growing excitement. RTX GPUs now drive more than 2/3 of our desktop gaming GPU revenue.

Gaming laptops were a standout, driving strong sequential and year-on-year growth. This holiday season, our partners are addressing the growing demand for high-performance laptops for gamers, students and prosumers by bringing more than 130 NVIDIA-powered gaming and studio laptop models to market. This includes many thin and light form factors enabled by our Max-Q technology, triple the number of Max-Q laptops last year.

In late October, we announced the GeForce GTX 1660 Super and the 1650 Super, which refresh our mainstream desktop GPUs with more performance, faster memory and new features. The 1660 Super delivers 50% more performance than our prior-generation Pascal-based 1060, the best-selling gaming GPU of all time. It began shipping on October 29, priced at just \$229. PC World called it the best GPU you can buy for 1080p gaming.

We also announced the next generation of our streaming media player with 2 new models, Shield TV and Shield TV Pro, which launched on October 28. These bring AI to the streaming market for the first time with the ability to upscale video real time from high definition to 4K using NVIDIA-trained deep neural networks. Shield TV has been widely recognized as the best streamer on the market.

Finally, we made progress in building out our cloud gaming business. Two global service providers, Taiwan Mobile and Russia's Rostelecom with GFN.ru joined SoftBank and Korea's LG as partners for our GeForce NOW game-streaming service. Additionally, Telefónica will kick off a cloud gaming proof-of-concept in Spain.

Moving to data center. Revenue was \$726 million, down 8% year-on-year and up 11% sequentially. Our hyperscale revenue grew both sequentially and year-on-year, and we believe our visibility is improving. Hyperscale activity is being driven by conversational AI, the ability for computers to engage in human-like dialogue, capturing context and providing intelligent responses. Google's breakthrough, introduction of the BERT model, with its superhuman levels of natural language understanding, is driving a way of neural networks for the language understanding. That, in turn, is driving demand for our GPUs on 2 fronts. First, these models are massive and highly complex. They have 10 to 20x, in some cases 100x, more parameters than image-based models. As a result, training these models requires V100-based compute infrastructure that, in orders of magnitude, beyond what is needed in the past. Model complexity is expected to grow significantly from here.

Second, real-time conversational AI requires very low latency and multiple neural networks running in quick succession from de-noising to speech recognition, language understanding, text-to-speech and voice encoding. While conventional approaches fail at these tasks, NVIDIA's GPUs can handle the entire inference chain, in less than 30 milliseconds. This is the first AI application where inference requires acceleration. Conversational AI is a major driver for GPU-accelerated inference.

In addition to this type of internal hyperscale activity, our T4 GPU continue to gain adoption in public clouds. In September, Amazon AWS announced general availability of the T4 globally, following the T4 rollout on Google Cloud platform earlier in the year. We

shipped a higher volume of T4 inference GPU this quarter with V100 training GPUs, and both were records. Inference revenue more than doubled from last year and continued a solid double-digit percentage of total data center revenue.

Last week, the results of the first industry benchmark for AI inference, MLPerf inference, were announced. We won. In addition to demonstrating the best performance among commercially available solutions for both data center and edge applications, NVIDIA accelerators were the only ones that completed in all 5 MLPerf benchmarks. This demonstrates the programmability and performance of our computing platform across diverse AI workloads, which is critical for wide-scale data center deployment and is a key differentiator for us.

Several product announcements this quarter helped extend our AI computing platform into new markets, the enterprise edge. At Mobile World Congress, Los Angeles, we announced a software-defined 5G wireless RAN solution accelerated by GPUs in collaboration with Ericsson. This opens up the wireless RAN market to NVIDIA GPUs. It enables new AI applications as well as AR, VR and gaming to be more accessible to the telco edge.

We announced the NVIDIA EGX Intelligent Edge Computing Platform. With an ecosystem of more than 100 technology companies worldwide, early adopters include Walmart, BMW, Procter & Gamble, Samsung Electronics, NTT East and the cities of San Francisco and Las Vegas. Additionally, we announced a collaboration with Microsoft on intelligent edge computing. This will help industries better manage and gain insights from the growing flood of data created by retail stores, warehouses, manufacturing facilities and urban infrastructure.

Finally, last week, we held our GPU Technology Conference in Washington, D.C., which was sold out with more than 3,500 registered developers, CIOs and federal employees. At the event, we announced that the U.S. Postal Service, the world's largest delivery service with almost 150 billion pieces of mail delivered annually, is adopting AI technology from NVIDIA, enabling 10x faster processing of package data and with higher accuracy.

Moving to ProVis. Revenue reached a record \$324 million, up 6% from the prior year and up 11% sequentially driven primarily by mobile workstations. NVIDIA RTX graphic and Max-Q technology have enabled a new wave of mobile workstations that are powerful enough for design applications yet thin and light enough to carry. We expect this to become a major new category with exciting growth opportunities.

Over 40 top creative design applications are being accelerated with RTX GPUs. Just last week, at the Adobe Max Conference, RTX accelerated capabilities were added to 3 Adobe Creative apps. RTX-accelerated apps are now available to tens of millions of artists and designers, driving demand for our RTX GPUs. We also continue to see growing customer deployment of data science, AI and VR applications. Strong demand this quarter came from manufacturing, public sector, higher education and health care customers.

Finally, turning to automotive. Revenue was \$162 million, down 6% from a year ago and down 22% sequentially. The sequential decline was driven by a onetime nonrecurring development services contract recognized in Q2. Additionally, we saw a roll-off of legacy infotainment revenue and general industry weakness. Our AI cockpit business grew driven by the continued ramp of the Daimler as they deploy their AI-based infotainment systems across their fleet of Mercedes-Benz vehicles.

In August, Optimus Ride launched New York City's first autonomous driving pilot program powered by NVIDIA DRIVE. Urban settings pose unique challenges for autonomous vehicles given the number of density of objects that need to be perceived and comprehended in real time. Our DRIVE computer and software stack allows these shuttles to safely and effectively provide first- and last-mile transit services. We remain excited about the long-term opportunity in auto. Our offering is -- consists of in-car AV computing platforms as well as GPU servers for all AI development and simulation. We believe we are well positioned in the industry with leading end-to-end platform that enables customers to develop, test and safely operate autonomous vehicles, ranging from cars and trucks to shuttles and robo-taxis.

Moving to the rest of the P&L. Q3 GAAP gross margins was 63.6%, and non-GAAP was 64.1%, up sequentially, reflecting a benefit from sales of previously written-off inventory, higher GeForce GPUs average selling prices and lower component costs. GAAP

operating expenses were \$989 million, and non-GAAP operating expenses were \$774 million, up 15% and 6% year-on-year, respectively. GAAP EPS was \$1.45, down 26% from a year earlier. Non-GAAP EPS was \$1.78, down 3% from a year ago. Cash flow from operations was a record \$1.6 billion.

With that, let me turn to the outlook for the fourth quarter of fiscal 2020, which does not include any contribution from the pending acquisition of Mellanox.

We expect revenue to be \$2.95 billion, plus or minus 2%. This reflects expectations for strong sequential growth in data center, offset by a seasonal decline in notebook GPUs for gaming and Switch-related revenue. GAAP and non-GAAP gross margins are expected to be 64.1% and 64.5%, respectively, plus or minus 50 basis points. GAAP and non-GAAP operating expenses are expected to be approximately \$1.02 billion and \$805 million, respectively. GAAP and non-GAAP OI&E are both expected to be income of approximately \$25 million. GAAP and non-GAAP tax rates are both expected to be 9%, plus or minus 1%, excluding discrete items. Capital expenditures are expected to be approximately \$130 million to \$150 million. Further financial details are included in the CFO commentary and other information available on our IR website.

In closing, let me highlight the upcoming events for the financial community. We will be at the Crédit Suisse Annual Technology Conference on December 3, Deutsche Bank's Auto Tech Conference on December 10 and Barclays Global Technology, Media and Telecommunications Conference on December 11.

We will now open the call for questions. Operator, would you please poll for questions?

Question-and-Answer Session

Operator

[Operator Instructions]. And your first question comes from the line of Vivek Arya with Bank of America Merrill Lynch.

Vivek Arya

For my first one, you mentioned that you were seeing strong sequential growth in the data center going into Q4. Jensen, I was wondering if you could give us some color on what's driving that and just how you think about the sustainability of data center growth going into next year and what markets do you think will drive that. Is it more enterprise, more hyperscale, more HPC? Just some color on near and longer term on data center. And then I have a follow-up for Colette.

Jensen Huang

Yes. Thanks a lot, Vivek. We had a strong Q3 in hyperscale data centers. As Colette mentioned earlier, we shipped a record number of V100s and T4s. And for the very first time, we shipped more T4s than V100. And most of the T4s are driven by inference. In fact, our inference business is now a solid double-digit, and it doubled year-over-year. And all -- most -- that is really driven by several factors. The -- as you know, we've been working on deep learning for some time, and people have been developing deep learning models. It started with computer vision. But image recognition doesn't really take that much of the data center capacity.

Over the last couple of years, a couple of very important developments have happened. One development is a breakthrough in using deep learning for recommendation systems. As you know, recommendation systems is the backbone of the Internet. Whenever you do shopping, whenever you're watching movies, looking at news, doing search, all of the personalized web pages, all of just about your entire experience on the Internet is made possible by recommendation systems because there is just so much data out there putting the right data in front of you based on your social profile or your personal use patterns or your interest or your connections, all of that is vitally important. For the very first time, we're seeing recommendation system based on deep learning throughout the world. And so increasingly, you're going to see people roll this out. And the backbone of the Internet is now going to be based on deep learning.

The second part is conversational AI. Conversational AI has been coming together in pieces; at first, speech recognition, which requires some amount of noise processing or beam forming. Then you go into speech recognition. Then it goes to natural language understanding, which then gets connected to a recommendation system, which then gets

connected to text-to-speech and a speech encoder. And then that has to be done very, very quickly. Whereas images could be done off-line, conversation has to be done in real time. And without acceleration and without NVIDIA's accelerators, it's really not possible to do it in real time. It takes seconds to process all of the handful of deep learning models, and now we're able to do that all on an accelerator and do it in real time.

And so the combination of these various breakthroughs from deep learning-based recommenders, the speech stack as well as natural language understanding breakthrough in what is called a bidirectional encoded transformer, that breakthrough is really quite significant. And since then, derivative works have come from that approach. And natural language understanding is really, really working incredibly well. And so what we're seeing people do is -- the hyperscalers across the world, we work with just about everybody, this area of work is really complicated. The models are very, very large. There's a whole bunch of models that has to work together, and they're getting larger. And so that's one large category, which is the hyperscalers.

The second, which we introduced this quarter, is really about taking AI out to the edge. And the reason for that is because there are many applications, whether it's based on video or other types of sensors of all kinds where there's a vibration sensor, temperature sensors, barometric sensor, there's all kinds of sensors that are used in industries to monitor the health of equipment, monitor the conditions of various situations. And you want to do the processing at the point of action. This way, you don't have to screen the data, which is continuous back into the cloud, which costs a lot of money. You want to take the action at the point of action because latency matters. Maybe you're controlling gates or vehicles or robots or drones or whatnot.

And then lastly, one major issue is data sovereignty. Maybe your company doesn't own all of the data that you are processing and, therefore, you have to do that processing at the edge, and you can't afford to put that into the cloud. And so these various industries: retail, warehouse, logistics, smart cities, we're just seeing so much enthusiasm there around that. And this -- so we built a platform called the EGX, which basically is a cloud native, completely secure, takes advantage of NVIDIA's full stack of every single model. And it's managed with Kubernetes remotely, and you could deploy these services at the edge in faraway places because IT departments can't afford to go out there to manage them. And

we've seen some really great adoption. We announced this last quarter. Walmart is using our platform. BMW is using it for logistics, Procter & Gamble for manufacturing, Samsung Electronics for manufacturing, visual inspection. And then last week, we announced probably the largest logistics operation in the world, the United States Postal Service.

And so those are -- I would say that intelligent edge will likely be the largest AI industry in the world for rather clear reasons. If you just kind of estimated the size of retail, it's nearly \$30 trillion. And if retail stores could be made a little bit more convenient, it could save the industry a lot of money: warehouses, logistics, transportation, farming. I think there's like 0.5 million farms in the world, covers 1/3 of the world's land mass. And so there's a lot of places where AI could be put at the edge and could make a big difference. And I think this is going to be the grand adventure that we started this last quarter with the announcement of NVIDIA EGX.

Vivek Arya

Right. And Jensen, as quick follow-up, on PC gaming, how are you looking at growth going forward in that you had a very good quarter in October? I think in January, you're probably guiding to some seasonal declines, but I imagine a lot more of that is due to console decline. Just how are you looking at PC gaming growth going into October -- into January and then next year as you get competition from 2 new consoles that are also supposed to come out?

Jensen Huang

Yes. During Q4 and Q1, we see normal seasonal declines of console builds, and we also see a normal seasonal decline of notebook builds. And the reason for that is because the notebook vendors have to line up all their manufacturing in Q3 so that they could meet the hot selling season in Q4. And so we're seeing -- what we see in the Q4 and Q1 time frame are just normal seasonal declines of these systems. Overall, for PC gaming -- and RTX is doing fantastic. Let me tell you why it's so important. I would say that at this point, I think it's fairly clear that ray tracing is the future and that RTX is a home run. Just about every major game developer has signed on to ray tracing. Even the next-generation consoles had to stutter step and include ray tracing in their next-generation consoles. The effects -- the photorealistic look is just so compelling, it's not possible to really go back anymore.

And so I think that it's fairly clear now that RTX ray tracing is the future. And there are several hundred million PC gamers in the world that don't have the benefits of it, and I'm looking forward to upgrading them.

Second, and this is a combination of RTX and Max-Q, we really created a brand-new game platform, notebook PC gaming. Notebook PC gaming really didn't exist until Max-Q came along. And our second-generation Max-Q, this last season, really turbocharged this segment. Over 100 laptops now are available for PC gaming. And my sense is that this is likely going to be the largest gaming platform, new gaming platform that emerges. And we're just in the beginning innings of that. And so the combination of upgrading the entire installed base of PC gamers to RTX and ray tracing and this new gaming segment called notebook PC gaming is really quite exciting, and it's going to drive our continued growth for some time. And so I'm excited about that.

Operator

Your next question comes from the line of Aaron Rakers with Wells Fargo.

Aaron Rakers

I have a follow-up if I can as well. Just thinking about the trajectory of gross margin here, solid gross margin upside in the quarter, you also noted that you had the benefit of selling through some written-off components. So I guess first question is what was that impact in this most recent reported quarter. And how do we think about the trajectory of gross margin here even beyond the January quarter? What should we be thinking about in terms of that gross margin trend? And again, I have a quick follow-up.

Colette Kress

Sure. Thanks for the question. In the current quarter, the net benefit, as we refer to as the net release of our inventory provisions primarily associated with our components, was about 1 percentage point to our overall gross margin. As you know, going forward, mix is still the largest driver of our gross margin over time. Over the long term, we do expect gross margins to improve, and we'll continue to see, outside of the benefit that we received, gross margin improvement for the long term.

Jensen Huang

Yes. As you know, just to add to that, as you know, NVIDIA's really become a software company. If you take a look at almost all of our products, the GPU -- having the world's best GPU, of course, is the starting point. But almost everything that we do, whether it's in artificial intelligence or data analytics or health care or robotics or self-driving cars, almost all of these platforms: gaming, rendering, cloud graphics, all of these platforms start from a really rich stack of software. And you can't just put a chip in these scenarios and they work. And so most of our businesses are now highly software-rich, and they address verticals that we focus on. And then secondarily, we're a platform company. And so our platform is available from all the OEMs and cloud providers. And as a platform company that has a great deal of software intensity, it's natural that the margins would be higher over time.

Aaron Rakers

Yes. Very helpful. And then you mentioned in your prepared remarks that you've seen hyperscale -- your hyperscale business within data center grow both on a quarter-over-quarter as well as year-over-year basis in this last print. You also mentioned that your visibility is improving. Can you just help us understand what exactly you're seeing in the hyperscale guys because it feels like there's some mixed data points out there? What underpins your improved visibility? Or what are you seeing in that piece of your business?

Jensen Huang

Yes. We had a strong Q3. We're going to see a much stronger Q4. And the foundation of that is AI, it's deep learning inference. That is -- this deep learning inference is understandably going to be one of the largest computer industry opportunities. And the reason for that is because the computation intensity is so high. And for the very first time, aside from computer graphics, this mode of software is not really practical without accelerators. And so I mentioned earlier about the large-scale movement to deep learning recommendation systems. Those models are really, really hard to train.

I mentioned earlier about conversational AI. Because conversation requires real-time processing, several seconds is really not practical. And so you have to do it in milliseconds, tens of milliseconds. And our accelerator makes that possible. What makes it really complicated and the reason why -- although so many people talk about it, only we demonstrated -- we submitted all 5 results -- all 5 tests for the MLPerf inference benchmark, and we won them. And the reason for that is because it's far more than just a chip. The software stack that sits on top of the chip and the compilers that sits on top of the chip are so complicated. And it's understandably complicated because a supercomputer wrote the software, and this body of software is really, really large. And if you have to make it both accurate as well as performant, it's really quite a great challenge. And it's one of the great computer science challenges. This is one of those problems that hasn't been solved, and we've been working hard at it for the last 6, 7 years now.

And so this is really the great opportunity. We've been talking about inference for some time now. Finally, the workloads and a very large diverse set of workloads are now moving into production. And so I'm hoping -- I'm enthusiastic about the progress and seeing the trends and the visibility that inference should be a large market opportunity for us.

Operator

Your next question comes from the line of C.J. Muse with Evercore ISI.

Christopher Muse

I guess I'd love to follow on, on that last question. So clearly, your commentary, Jensen, here is much more bullish than I've heard you, I think, before on inference, particularly as it relates to this first benchmark. And so I guess can you talk a bit about how you see mix within data center looking out over the next 12, 24 months as you see kind of training versus inference as well as cloud versus enterprise, considering, I would think, inference over time could be -- could grow into a large opportunity there as well?

Jensen Huang

Yes. C.J., that's really good. Let me break it down. So when we think about hyperscale, there are three parts: training, inference and public cloud.

Training, you might have seen the work that was done at open AI recently where they've been measuring and monitoring the amount of computation necessary to train these large models. These large models are now only getting larger. The amount of data necessary, therefore, has to scale as well. The computation is now growing and doubling every 3 months. And the reason for that is because of recent breakthroughs in natural language understanding. And all of a sudden, a whole wave of problems are now able to be solved. And just as AlexNet 7 years ago kind of was the watershed event for a lot of computer vision-oriented AI work, now the transformer-based natural language understanding model and the work that Google did with BERT really is a watershed event also for natural language understanding. This is, of course, a much, much harder problem. And so the scale of the training has grown tremendously. I think what we're going to see this year is a fair number of very sizable installations of GPU systems to do this very thing, training.

The second part is an untapped market for us, and this untapped market is really inference. The reason why I haven't really spoken about it until now is because we've never really been able to validate our intuition that inference is going to be a large market opportunity for us, that it's going to be very complicated. The models are very large. They're very diverse. They require large amount of computation, large amount of memory bandwidth and large amounts of memory and large and significant capabilities of programmability. And so I've talked about this before, but I've never been able to validate it. And of course, with MLPerf and sweeping the benchmarks and, frankly, only -- the only one although so many have attempted, they submitted results and some of them resented it, that this benchmark is just really, really hard. Inference is hard. And then finally, our business results also validated the -- our intuition. And so our engagement now with CSPs are now global. We're working across natural language understanding, recommendation systems, conversational AI, just a whole bunch of really, really interesting problems.

Now the cloud is the third piece. And the reason why cloud is growing so well and represents about half almost of many of our CSPs, particularly the ones with the public cloud, the reason for that is because the number of AI start-ups in the world is still growing so incredibly. I think we're tracking something close to 10,000 and more AI start-ups around the world. In health care, in transportation, in retail, in consumer Internet, in Fintech, the number of AI companies out there is just extraordinary. I think over the last 3

or 4, 5 years, some \$20 billion, \$30 billion have been invested into start-ups. And these start-ups, of course, use cloud service providers so that they don't have to invest in their own infrastructure because it's fairly complicated. And so we're seeing a lot of growth there.

And so that's just the hyperscalers. The hyperscalers give us 3 points of growth -- 3 areas of growth: training, inference and public cloud. And the public cloud is primarily AI start-ups. Then there's the intelligent edge, which we recently ventured into, and we've been building this platform called EGX for some time. And it's cloud native. It's incredibly secure. You can manage it from afar. It's -- the stack is complicated. It's performant. And we saw some -- we've been working with some early adopters. And this last quarter, we announced some of them: Walmart and BMW and Procter & Gamble and the largest logistics company in the world, USPS. And so this new platform, I think, long term, will likely be the largest opportunity. And the reason for that is because of the industries that it serves.

Operator

And your next question comes from the line of Harlan Sur with JPMorgan.

Harlan Sur

There are a lot of concerns around China trade tensions, economic slowdown. But history has shown that gamers tend to be less sensitive to these macro trends and, in fact, also somewhat insensitive to price changes, at least at the enthusiast level. So given that China is such a big part of the gaming segment, can you just discuss the gaming demand trends out of this geography?

Jensen Huang

Gaming is solid in China, and it is also the fastest adopter of our gaming notebooks. This gaming RTX notebooks or GeForce notebooks is really a brand new category. This category never existed before because we couldn't get the technology in there so that it's both delightful to own as well as powerful to enjoy. And so we saw really great success with RTX notebooks and GeForce notebooks in China, and RTX adoption has been fast.

Your comments make sense because most of the games are free-to-play these days. The primary games that people play are esports, which you want the best gear, but you could - after you buy the gear, you pretty much enjoy it forever; and mobile, which is largely free-to-play. You invest in some of your own personal outfits. And after that, I think you can enjoy it for quite a long time. And so the gear is really important. One of the areas where we've done really great work, particularly in China, has to do with social. We have this platform called GeForce Experience. And as an extension of that, there's a new feature called RTX Broadcast Engine. And it basically applies AI to broadcasting your content to share it. You could make movies. You could capture your favorite scenes and turn it into art, applying AI. And one of the coolest features is that you could overlay yourself on top of the game and share it with all the social networks without a green screen behind you. We use AI to stitch you out, basically, to cut you out of the background and irrespective of what noisy background you've got.

And so as you know, China has really a super hyper social community -- communities back there and they have all kinds of really cool social platforms to share games and user-generated content and short videos and all kinds of things like that. And so GeForce has that one additional feature that really makes it successful.

Operator

And your next question comes from the line of Toshiya Hari with Goldman Sachs.

Toshiya Hari

I wanted to ask on automotive. Colette, in your prepared remarks, you talked about your legacy infotainment business being down in the quarter. Just curious, what percentage of automotive revenue at this point is legacy infotainment versus the newer AI/ADAS solutions? And more importantly, Jensen, if you can speak to the growth trajectory in automotive over the next 1.5 years, maybe 2, that would be appreciated. And I do ask the question because it feels like we've heard many, many announcements, customer announcements, collaborative work that you're doing with your customers, yet we haven't quite seen sort of a hockey-stick inflection that some of us were expecting a couple of years ago. So just kind of curious when we should -- how we should set our expectations going forward.

Colette Kress

Yes. Toshiya, let me address the first question regarding our legacy infotainment systems for our automotive business. It is still representing maybe about half or more of our overall revenue in the automotive business. We have our AI cockpit continuing to grow and grow quite well, both sequentially as well as year-over-year, as well as our autonomous vehicle solutions that we may be doing, including development services.

Jensen Huang

Let's see. The -- we're the first -- probably the first AV car that's going to be passenger-owned on the road, and I think we've talked about it before, is Volvo. And we're expecting them to be in the late 2020, early 2021 time frame. And I'm still expecting so. And then there's the 2022, 2023 generations. Most -- I would say most of the passenger-owned vehicle developments are going quite well. The industry, as you know, is under some amount of pressure, and so a lot of them have slipped it out a couple of years or so. And this is something that I think we've already spoken about in the past.

Our focus, our strategy consists of several areas. One area, of course, is passenger-owned vehicles. The second part is robot taxis. We have developments going with just about every major robot taxi company that we know of. And they're here in the states. They're in Europe. They're in China. And when you hear news of them, we're delighted to see their progress. And then the third part has to do with trucks, shuttles and increasingly a large number of vehicles that don't carry people, they carry goods. And so we have a major development with Volvo. That was Volvo Trucks. Volvo Cars and Volvo Trucks, as you know, are 2 different companies. One of them belongs to Geely, Volvo Cars. Volvo Trucks is the heritage Volvo. And we have a major program going with them to automate the delivery of goods.

You also see us during various GTCs, I'll mention companies that we're working with on grocery delivery or goods delivery or within a warehouse product delivery. You're going to see a whole bunch of things like that because the technology is very similar, and it's starting to -- the development -- the technology we develop for passenger-owned vehicles

has started to propagate down into logistics vehicles. I continue to believe that everything that moves eventually will have autonomous capability or be fully autonomous. And that, I think, is, at this point, fairly certain.

Now our strategy is both in developing the in-car AV computing system, and it's software-defined, it's scalable, as well as the AI development and simulation systems. And so when somebody's working on AV and they're using AI, and most of them are, there's a great opportunity for us. And when they start ramping up and they're collecting miles of data, it becomes a very large market opportunity for us. And so I'm anxious to see every single car company be as progressive and aggressive in developing AV. And they will be. They will be. This is a foregone conclusion.

Operator

Your next question comes from the line of Stacy Rasgon with Bernstein.

Stacy Rasgon

I have two data center questions for Colette. The first question, I want to return to your kind of outlook for strong sequential data center growth in Q4. Now this business grew 11% sequentially in Q3. And you didn't actually call out strong growth as we were going into the quarter. You are calling it out for Q4. Does that suggest to me that you expect sequential growth in Q4 to be stronger than Q3 given you're calling it out in Q4 and you didn't call it out in Q3? Or would you define like what you saw in Q3 as well as already being strong sequential growth? Like how do we think about the wording of that in relation to what we've seen in Q3 and what you expect for Q4?

Colette Kress

Sure, Stacy. When we had provided guidance in Q3 and how we finished the quarter in Q3, we had indicated that our growth would stem from both gaming and data center. We completed that. And we also had stronger than expected from guidance from both gaming and data center in our Q3 results. Moving to Q4, Q4 is a sequential decrease in totality versus Q3. We have reminded the teams about our overall seasonality that we sometimes have in gaming associated with our consoles as well as also with our notebooks that seem

to be primarily in Q2 and Q3 being our strongest quarters and likely, therefore, a seasonal downtick as it move to Q4. What we wanted to do was, if we have in totality overall decline associated with that, we did want to emphasize what we are expecting in terms of data center with the overall strong growth sequentially.

Stacy Rasgon

So I guess to ask the question again, would you define what you saw in Q3 as being strong growth as well?

Colette Kress

I would believe our growth of 17% was higher than we expected to Q3. Again, when we get into Q4, we'll see how the quarter ends in terms of data center, but we are expecting strong growth. Thanks, Stacy.

Stacy Rasgon

Okay. And for my second question, hyperscale you said was up year-over-year. Now -- and that's after, off of last year, where it was the peak. Inference doubled year-over-year. And this suggests to me -- I know you said enterprise was down year-over-year. But this suggests to me that it wasn't just down year-over-year, it was down a lot year-over-year. How do we think about that in the context of like the growth that we've seen very strongly over the last few quarters in enterprise. And going back to your commentary at the Analyst Day, which was almost entirely about the opportunity coming from enterprise growth, what's going on there? What drove that? And what should we expect going forward?

Colette Kress

Sure. Our enterprise business has been beginning to ramp from over a year ago at a very, very, very small base. We've continued to see great traction in there with a lot of the things that we've announced throughout. But keep in mind in our year ago quarter, we also had very strong systems and a very large deal associated with our DGX. So when we look from a quarter-over-quarter period or just looking at 1 quarter, we can have a little bit of lumpiness. So that year-over-year impact is really just due to an extremely large deal in the prior year Q3.

Operator

Your next question comes from the line of Mitch Steves with RBC.

Mitchell Steves

I apologize for any background noise, but I just have one question, just for Jensen. So in 2018, can you give us a rough update on what the GPU utilization was for deep learning application? What it is today? I'm just wondering how the -- how that's advanced over the last couple of year or two.

Jensen Huang

Let's see. I would say 2018, it was nearly all related to training. And this year, we started to see the growth of inference to the point where we now -- we have now sold more -- this last quarter, we sold more T4 GPUs for inference than we sold V100s that's used for training, and both of them were record highs. And so the comment that Colette just made, comparing to year-over-year, we had a large DGX system sale a year ago that we didn't have this year. But if you excluded that, the V100 and the T4 is doing great. They're at record levels. And T4 didn't hardly existed a year ago, now it's selling more than V100s, and both of them are record highs. And so that kind of gives you a feeling for it. I think that's really the major difference that inference is really kicking into gear, and my sense is that it's going to continue to grow quite nicely.

Operator

And your next question comes from the line of Joe Moore with Morgan Stanley.

Joseph Moore

I wonder if you could talk a little bit more about the 5G opportunity that you announced at Mobile World. And I guess you talked a lot about AI and IoT services in a C-RAN environment. But is there -- how big is that opportunity? And can you address kind of the core compute aspect to C-RAN with the GPU?

Jensen Huang

Yes. If you look at the world of mobile today, there are players that are building DRAMs and their radio heads in the BBU, basically the baseband units. In the data center where people would like to move the software for radio networks, it's really an untapped market. And the reason for that is because the CPU is just not able to support the level of performance that's necessary for 5G. And ASICs are too rigid to be able to put into a data center. And so the data center needs a programmable solution that is data center-ready that can support all of the software richness that goes along with the data center, whether it's a VM environment like VMware. And we -- recently, during the quarter, we announced another partnership with VMware. They recognize that increasingly, our GPUs are becoming a core part of data centers and cloud. We had a partner -- we announced a partnership with Red Hat. They realize the momentum that they're seeing us in, in telcos, and they would like to adapt their entire stack from open stack to OpenShift on top of our GPUs. And so now with VMware, with Red Hat, we're going to have a world-class telco enterprise stack that ranges all the way from hypervisors and virtual machines all the way to Kubernetes.

And so our strategy is to -- our goal is to really create this new world of C-RAN, vRAN centralized data centers and software-defined networking. And the software-defined networking will, of course, include things like in the data center networking as well as firewalls. But the computationally-intensive stuff is really the 5G radio. And so we're going to create a software stack for 5G and basically exactly the same way that we've done for creating a -- excuse me, a software stack for deep learning. And we call it Aerial. Aerial is to 5G essentially what Cuda-NN is for deep learning and essentially what optics is for ray tracing. And this software stack is going to allow us to run the whole software -- run the whole 5G stack in software and deliver the highest performance, the incredible flexibility and scale to as many layers of MIMO as customers need and to be able to put all of it in the data center.

The power of putting it into data center, as you know, is flexibility and fungibility. With the low latency capability of 5G, you could put a data center somewhere in the regional hub. And depending on where the traffic is going, you could shift the traffic computation from 1 data center to another data center, something that you can't do in basebands, in baseband units in the cell towers, but you can do that in the data center. And that helps them reduce

the cost. The second benefit is that the telcos would love to be a service provider for a data center's computation at the edge. And the edge applications are things like smart cities and whether it's warehouses or retail stores or whatever it is because they're geographically located and is distributed all over the world. And so to be able to use their data center to also be able to use AI in combination with IoT is really exciting to them. And so I think that that's really -- this is really the future that we're going to see a lot more service providers at the edge. And these edge data centers will have to run the data center, the networking, including the mobile network and software as well as run 5G and IoT -- AI and IoT applications

Operator

And your last question comes from the line of Harsh Kumar with Piper Jaffray.

Harsh Kumar

I apologize for the background noise. But Colette, maybe you could give us an idea of gaming. In the guidance, it's down. And I was wondering, could you maybe give us the impact of the console business versus the laptop and give us an idea of what might be the bigger driver there?

Colette Kress

I would say, for our Q4, both of them are expected to be seasonally down. In the case of the consoles, we do wait for Nintendo to assist in terms of what they need. So we will have to see how the quarter ends on that. But in both cases, in totality, these businesses have ranged, maybe in totality of the 2, of about \$500 million a quarter. And we'll see both of them sequentially decline. Thank you.

Operator

I'll now turn the call back over to Jensen for any closing remarks.

Jensen Huang

Thanks, everyone. We had a good quarter driven by strong gaming growth and hyperscale demand. We're making great strides in 3 big impact initiatives. The world of computer graphics is moving to ray tracing, and our business reflects that. Some of the biggest blockbuster games this holiday season and beyond are RTX-enabled, including Call of Duty: Modern Warfare; and the best-selling game of all-time, Minecraft. Design applications used by millions of artists and creators are rapidly adopting RTX ray tracing. We're reinventing computer graphics and look forward to upgrading the hundreds of millions of PC gamers to RTX.

Hyperscale demand was strong this quarter, and our visibility continues to improve. The race is on for conversational AI, which will be a powerful catalyst for us in both training and inference. And lastly, we have extended our computing platform beyond the cloud to the edge, where GPU-accelerated 5G, AI and IoT, will revolutionize the world's largest industries. We look forward to updating you on our progress in February.

Operator

Ladies and gentlemen, this concludes today's conference call. Thank you for participating. You may now disconnect.