

Q3 2018 Earnings Call

Company Participants

- Colette M. Kress, Chief Financial Officer & Executive Vice President
- Jen-Hsun Huang, President, Chief Executive Officer & Director
- Simona Jankowski, Vice President, Investor Relations

Other Participants

- Atif Malik, Analyst
- C. J. Muse, Analyst
- Chris Caso, Analyst
- Craig A. Ellis, Analyst
- Hans Mosesmann, Analyst
- Joseph L. Moore, Analyst
- Matthew D. Ramsay, Analyst
- Stacy Aaron Rasgon, Analyst
- Toshiya Hari, Analyst
- Vivek Arya, Analyst

MANAGEMENT DISCUSSION SECTION

Operator

Good afternoon. My name is Victoria, and I'm your conference operator for today. Welcome to NVIDIA's financial results conference call. The phone lines have been placed on mute to prevent any background noise. After the speakers' remarks, there will be a question-and-answer period.

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

Simona Jankowski {BIO 7131672 <GO>}

Thank you. Good afternoon, everyone, and welcome to NVIDIA's conference call for the third quarter of fiscal 2018. 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. It is also being recorded. You can hear a replay by telephone until November 16,

2017. The webcast will be available for replay up until next quarter's conference call to discuss Q4 and full-year fiscal 2018 financial results. 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, November 9, 2017, 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 M. Kress {BIO 18297352 <GO>}

Thanks, Simona.

We had an excellent quarter with record revenue in each of our four market platforms. And every measure of profit hit record levels, reflecting the leverage in our model. Data center revenue of \$501 million more than doubled from a year ago amid strong adoption of our Volta platform and early traction with our inferencing portfolio.

Q3 revenue reached \$2.64 billion, up 32% from a year earlier, up 18% sequentially, and well above our outlook of \$2.35 billion. From a reporting segment perspective, GP revenue grew 31% from last year to \$2.22 billion. Tegra processor revenue rose 74% to \$419 million.

Let's start with our gaming business. Gaming revenue was \$1.56 billion, up 25% year on year and up 32% sequentially. We saw robust demand across all regions and form factors. Our Pascal-based GPUs remain the platform of choice for gamers, as evidenced by our strong demand for GeForce GTX 10 Series products. We introduced the GeForce GTX 1070 Ti, which became available last week. It complements our strong holiday lineup, ranging from the entry-level GTX 1050 to our flagship GTX 1080 Ti.

A wave of great titles is arriving for the holidays, driving enthusiasm in the market. We collaborated with Activision to bring Destiny 2 to the PC earlier in the month. PlayerUnknown's Battlegrounds, popularly known as PUBG, continues to be one of the year's most successful titles. We are closely aligned with PUBG to ensure that GeForce is

FINAL

Bloomberg Transcript

the best way to play the game, including bringing ShadowPlay highlights to its 20 million players. Last weekend, Call of Duty: World War II had a strong debut, and Star Wars: Battlefront 2 will be out shortly (4:43).

E-sports remains one of the most important secular growth drivers in the gaming market, with a fan base that now exceeds 350 million. Last weekend, the League of Legends World Championship was held in Beijing's National Stadium, the Bird's Nest, where the 2008 Olympics Games were held. More than 40,000 fans attended live, and online viewers were set to break last year's record of 43 million following in 18 languages.

GPU sales also benefited from continued cryptocurrency mining. We met some of this demand with a dedicated board in our OEM business and a portion with GeForce GTX boards, though it's difficult to quantify. We remain nimble in our approach to the cryptocurrency market. It is volatile, does not and will not distract us from focusing on our core gaming markets.

Lastly, Nintendo's Switch console continues to gain momentum since launching in March and also contributed to growth.

Moving to data center, our data center business had an outstanding quarter. Revenue of \$501 million more than doubled from last year and rose 20% on the quarter amid strong traction of the new Volta architecture. Shipments of the Tesla V100 GPU began in Q2 and ramped significantly in Q3, driven primarily by demand from cloud service providers and high-performance computing. As we have noted before, Volta delivers 10x the deep learning performance of our Pascal architecture, which has been introduced just a year earlier, far outpacing Moore's Law.

The V100 is being broadly adopted with every major server OEM and cloud provider. In China, Alibaba, Baidu, and Tencent announced that they are incorporating V100s in their data centers and cloud service infrastructures. In the U.S., Amazon Web Services announced that V100 instances are now available in four of its regions. Oracle Cloud has just added Tesla P100 GPUs to its infrastructure offerings and plans to expand to the V100 GPUs. We expect support from V100 from other major cloud providers as well.

In addition, all major server OEMs announced support for the V100. Dell EMC, Hewlett Packard Enterprise, IBM, and Supermicro are incorporating it in servers. And China's top server OEMs, Huawei, Inspur, and Lenovo, have adopted our HGX server architecture to build a new generation of accelerated data centers with V100 GPUs.

Our new offerings for the AI inference market are also gaining momentum. The recently launched TensorRT 3 (sic) [TensorRT 3] programmable inference acceleration platform opens a new market opportunity for us, improving the performance and reducing the cost of AI inferencing by orders of magnitude compared with CPUs. It supports every major deep learning framework, every network architecture, and any level of network complexity. More than 1,200 companies are already using our inference platform, including Amazon, Microsoft, Facebook, Google, Alibaba, Baidu, JD.com, iFLYTEK, Hikvision, and Tencent.

During the quarter, we announced that the NVIDIA GPU Cloud container registry, or NGC, is now available through Amazon's cloud and will be supported soon by other cloud platforms. NGC helps developers get started with deep learning development through no-cost access to a comprehensive, easy to use, fully optimized deep learning software stack. It enables instant access to the most widely used GPU accelerated frameworks.

We also continue to see robust growth in our HPC business. Next-generation supercomputers such as the U.S. Department of Energy's CRN Summit systems, expected to come online next year, leverage Volta's industry-leading performance, and our pipeline is strong.

The past weeks have been exceptionally busy for us. We have hosted five major GPU Technology Conferences [GTC] in Beijing, Munich, Taipei, Tel Aviv, and Washington, with another next month in Tokyo. In a strong indication of the growing importance of GPU accelerated computing, more than 22,000 developers, data scientists, and others will come this year to our GTCs, including the main event in Silicon Valley. That's up 10x in just five years. Other key metrics show similar gains. Over the same period, the number of NVIDIA GPU developers has grown 15x to 645,000, and the number of CUDA downloads this year are up 5x to 1.8 million.

Moving to professional visualization, third quarter revenue grew to \$239 million, up 15% from a year ago and up 2% sequentially, driven by demand for high-end real-time rendering, simulation, and more powerful mobile workstations. The defense and automotive industries grew strongly, as did demand for professional VR solutions, driven by Quadro P5000 and P6000 GPUs. Among key customers, Audi and BMW are deploying VR in auto showrooms. And the U.S. Army, Navy, and Homeland Security are using VR for mission training.

Last month, we announced early access to NVIDIA Holodeck, the intelligent VR collaboration platform. Holodeck enables designers, developers, and their customers to come together virtually from anywhere in the world in a highly realistic, collaborative, and physically simulated environment. Future updates will address the growing demand for the development of deep learning techniques in virtual environments.

In automotive, revenue grew to \$144 million, up 13% year over year and up slightly from last quarter. Among key developments this quarter, we announced DRIVE PX Pegasus, the world's first AI computer for enabling Level 5 driverless vehicles. Pegasus will deliver over 320 trillion operations per second, more than 10x its predecessor. It's powered by four high-performance AI processors in a supercomputer that is the size of a license plate. NVIDIA DRIVE is being used by over 25 companies to develop fully autonomous robo-taxis, and DRIVE PX Pegasus will become the path to production. It is designed for ASIL D certification, the industry's highest safety level, and will be available in the second half of 2018.

We also introduced the DRIVE IX SDK for delivering intelligent experiences inside the vehicle. DRIVE IX provides a platform for car companies to create an always-engaged AI copilot. It uses deep learning networks to track head moment and gaze. And it will have a

conversation with the driver using advanced speech recognition, lip reading, and natural language understanding. We believe this will set the standard for the next generation of infotainment systems, a market that is just beginning to develop.

Finally, we announced that DHL, the world's largest mail and package delivery service, and ZF, one of the world's leading automotive suppliers, will deploy a test fleet of autonomous delivery trucks next year using the NVIDIA DRIVE PX platform. DHL will outfit electric light trucks with the ZF ProAI self-driving system based on our technology.

Now turning to the rest of the income statement, Q3 GAAP gross margin was 59.5% and non-GAAP was 59.7%, both up sequentially and year over year, reflecting continued growth in value-added platforms. GAAP operating expenses were \$674 million and non-GAAP operating expenses were \$570 million, consistent with our outlook and up 19% year on year. Investing in our key market opportunities is essential to our future, including gaming, AI, and self-driving cars.

GAAP operating income was a record \$895 million, up 40% from a year ago. Non-GAAP operating income was \$1.01 billion, up 42% from a year ago. GAAP net income was a record \$838 million and EPS was \$1.33, up 55% and 60% respectively from a year earlier. Non-GAAP net income was \$833 million and EPS was \$1.33, up 46% and 41% respectively from a year earlier, reflecting revenue strength as well as gross margin and operating margin expansion.

We have returned \$1.16 billion to shareholders so far this fiscal year through a combination of quarterly dividends and share repurchases. We have announced an increase to our quarterly dividend of \$0.01 to an annualized \$0.60, effective with our Q4 fiscal year 2018 dividend. We are also pleased to announce that we intend to return another \$1.25 billion to shareholders for fiscal 2019 through quarterly dividends and share repurchases.

Our quarterly cash flow from operations reached record levels, surpassing \$1 billion for the first time to \$1.16 billion.

Now turning to the outlook for the fourth quarter of fiscal 2018, we expect revenue to be \$2.65 billion plus or minus 2%. GAAP and non-GAAP gross margins are expected to be 59.7% and 60% respectively, plus or minus 50 basis points. GAAP and non-GAAP operating expenses are expected to be approximately \$722 million and \$600 million respectively. GAAP and non-GAAP OI&E are both expected to be nominal. GAAP and non-GAAP tax rates are both expected to be 17.5% plus or minus 1%, excluding discrete items. Further financial details are included in the CFO Commentary and other information available on our website.

We will now open the call for questions. Please limit your question to one. Operator, would you please poll for questions? Thank you.

Q&A

Operator

Your first question comes from the line of Toshiya Hari with Goldman Sachs.

Q - Toshiya Hari {BIO 6770302 <GO>}

Great, thank you very much for taking the question and congrats on another very strong quarter. Jensen, three months ago you described the July quarter as a transition quarter for your data center business, and clearly you guys have ramped very well into October. But if you can, talk a little bit about the outlook for the next couple of quarters in data center, and particularly on the inferencing side. I know you guys are really excited about that opportunity. So if you can share customer feedback and what your expectations are into the next year in inferencing, that would be great. Thank you so much.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Excuse me, I cleared my throat. Toshi, thanks for that. As you know, we started ramping very strongly Volta this last quarter, and we started the ramp the quarter before. And since then, every major cloud provider, from Amazon, Microsoft, Google, to Baidu, Alibaba, and Tencent, and even recently Oracle, has announced support for Volta and will be providing Volta for their internal use of deep learning as well as external public cloud services.

We also announced that every major server computer maker in the world has now supported Volta and in the process of taking Volta out to market. HP (sic) [HPE] and Dell and IBM and Cisco and Huawei in China, Inspur in China, Lenovo have all announced that they will be building servers, families of servers around the Volta GPU. And so I think this ramp is just the first part of supporting the build-out of GPU-accelerated servers from our company for data centers all over the world as well as cloud service providers all over the world.

The applications for these GPU servers have now grown to many markets. I've spoken about the primary segments of our Tesla GPUs. There are five of them that I talk about regularly. The first one is high-performance computing, where the market is \$11 billion or so. It is one of the faster growing parts of the IT industry because more and more people are using high-performance computing for doing their product development or looking for insights or predicting the market or whatever it is. And today we represent about 15% of the world's top 500 supercomputers. And I've repeatedly said and I believe this completely, and I think it's becoming increasingly true that every single supercomputer in the future will be accelerated somehow. So this is a fairly significant growth opportunity for us.

The second is deep learning training, which is very much like high-performance computing. And you need to do computing at a very large scale. You're performing trillions and trillions of iterations. The models are getting larger and larger. Every single year, the amount of data that we're training with is increasing. And the difference between a computing platform that's fast versus not could mean the difference between building a \$20 million data center or high-performance computing servers for training to \$200 million. And so the money that we save and the capability we provide is really - the value is incredible.

FINAL

The third segment, and this is the segment that you just mentioned, has to do with inference, which is when you're done with developing this network, you have to put it out into the hyperscale data centers to support the billions and billions of queries that consumers make to the Internet every day, and this is a brand-new market for us. 100% of the world's inference is done on CPUs today. We announced very recently, this last quarter in fact, the TensorRT 3 inference acceleration platform. And in combination with our Tensor core GPU instruction set architecture, we're able to speed up networks by a factor of 100. Now the way to think about that is imagine whatever amount of workload that you've got. If you could speed up using our platform by a factor of 100, how much you could save.

The other way to think about that is because the amount of - the networks are getting larger and larger and they're so complex now. And we know that every network on the planet will run on our architecture because they were trained on our architecture today. And so whether it's CNNs or RNNs or GANs or auto-encoders or all the variations of those, irrespective of the precision that you need to support, the size of the network, we have the ability to support them. And so you could either scale out your hyperscale data center to support more traffic or you could reduce your cost tremendously or simultaneously both.

The fourth segment of our data center is providing all of that capability what I just mentioned, whether it's HPC, training, or inference, and turning it inside-out and making it available in the public cloud. There are thousands of startups now that are started because of AI. Everybody recognizes the importance of this new computing model. And as a result of this new tool, this new capability, all these unsolvable problems in the past are now interestingly solvable. And so you could see startups cropping up all over the West, all over the East, and there are thousands of them. And these companies don't either - would rather not use their scarce financial resources to build high-performance computing centers or they don't have the skill to be able to build out a high-performance platform the way these Internet companies can.

And so these cloud providers, cloud platforms are just a fantastic resource for them because they can rent them by the hour. We created in conjunction with that, and I mention all the cloud service providers have taken it to market. In conjunction with that, we created a registry in the cloud that containerizes these really complicated software stacks. Every one of these soft frameworks with the different versions of our GPUs and different acceleration layers and different optimization techniques, we've containerized all of that for every single version and every single type of framework in the marketplace. And we put that up in the cloud registry called the NVIDIA GPU cloud. And so all you have to do is download that into the cloud service provider that we've got certified and tested for, and with just one click you're doing deep learning.

And then the last - and so that's the cloud service providers. The way to guess that - estimate that is there are obviously tens of billions of dollars being invested in these AI startups, and some large proportion of their investment fundraise will ultimately have to go towards high-performance computing, whether they build it themselves or they rent it in a cloud. So I think that's a multibillion dollar opportunity for us.

And then lastly, this is probably the largest of all the opportunities, which is the vertical industries, whether it's automotive companies that are developing their supercomputers to get ready for self-driving cars or with healthcare companies that are now taking advantage of artificial intelligence to do better diagnostics of - diagnosis of disease to manufacturing companies for inline inspection, to robotics, large logistics companies. Colette mentioned earlier DHL. But the way to think about that is all of these companies doing planning to deliver products to you through this large network of delivery systems. It is the world's largest playing Crumble (24:30). And whether it's Uber or DD [BeMyDD] or Lyft or Amazon or DHL or UPS or FedEx, they all have high-performance computing problems that are now moving to deep learning. And so those are really exciting opportunities for us.

And so the last one is just vertical industries. All of these segments we're now in the position to start addressing because we've put our GPUs in the cloud. All of our OEMs are in the process of taking these platforms out to market, and we have the ability now to address high-performance computing and deep learning training as well as inference using one common platform. And so I think we've been steadfast with the excitement of accelerated computing for data centers, and I think this is just the beginning of it all.

Operator

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

Q - Stacy Aaron Rasgon {BIO 16423886 <GO>}

Hi, guys. Thanks for taking my question. I had a question on your gaming seasonality into Q4. It's usually up a bit. I was wondering. Do you see any I guess drivers that would drive a lack of normal seasonal trends given how strong it's been sequentially and year over year? And I guess as a related question, do you see your Volta volumes in Q4 exceeding Q3?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Let's see. I'll answer the last one first and then work towards the first one. I think the guidance that we provided we feel comfortable with. But if you think about Volta, it is just in the beginning of the ramp and it's going to ramp into the market opportunities I talked about. And so my hope is that we continue to grow. And there's every evidence that the markets that we serve that we're addressing with Volta are very large markets, and so there's a lot of reasons to be hopeful about the future growth opportunities for Volta.

We've primed the pump. So cloud service providers either announced the availability of Volta or they announced the soon availability of Volta. They're all racing to get Volta to the cloud because customers are clamoring for it. The OEMs are - we primed the pump with the OEMs, and some of them are sampling now, and some of them are racing to get Volta to production in the marketplace. And so I think the foundation, the demand is there. The urgent need for accelerated computing is there because Moore's Law is not scaling anymore. And then we primed the pump. So the demand is there, there's a need, the need is there, and the foundations for getting Volta to market is primed.

FINAL

With respect to gaming, what drives our gaming business - remember, our gaming business is sold one at a time to millions and millions of people. And what drives our gaming business is several things. As you know, e-sports is incredibly, incredibly vibrant. And the reason why e-sports is so unique is because people want to win, and having better gear helps. The latency that they expect is incredibly low, and performance drives down latency, and they want to be able to react as fast as they can. People want to win, and they want to make sure that the gear that they use is not the reason why they didn't win.

The second growth driver for us is content, the quality of content. And boy, if you look at Call of Duty or Destiny 2 or PUBG, the content just looks amazing. The AAA content looks just amazing. And one of the things that's really unique about video games is that in order to enjoy the content and the fidelity of the content, the quality of the production value at its fullest, you need the best gear. It's very different than streaming video. It's very different than watching movies, where streaming videos, it is what it is, but for videogames of course it's not. And so when AAA titles come out in the later part of the year, it helps to drive platform adoption.

And then lastly, increasingly social is becoming a huge part of the growth dynamics of gaming. People, they recognize how beautiful these videogames are, so they want to share their brightest moments with people. They want to share the levels they've discovered. They want to take pictures of the amazing graphics that's inside. And it is one of the primary drivers, the leading driver in fact of YouTube, people watching other people play video games, these broadcasters. And now with our Ansel, the world's first in-game virtual reality and surround and digital camera, we have the ability to take pictures and share that with people.

And so I think all these different drivers are helping our gaming business, and I'm optimistic about Q4. It looks like it's going to be a great quarter.

Operator

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

Q - C. J. Muse {BIO 6507553 <GO>}

Good afternoon, thank you for taking my question. I was hoping to sneak in a near-term and a longer-term question. On the near term, you talked about the health on the demand side for Volta. I'm curious if you're seeing any sort of restrictions on the supply side, whether it's wafers or access to high-bandwidth memory, et cetera.

And then the longer-term question really revolves around CUDA. And you've talked about that as being a sustainable competitive advantage for you guys entering the year. And now that we've moved beyond HPC and hyperscale training to more into inference and GPU-as-a-service and you've hosted GTC around the world, I'm curious if you could extrapolate on how you're seeing that advantage and how you've seen it evolve over the year and how you're thinking about CUDA as the AI standard. Thank you.

Bloomberg Transcript

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Thanks a lot, C.J. Well, everything that we build is complicated. Volta is the single largest processor that humanity has ever made at 21 billion transistors, 3D packaging, the fastest memories on the planet, and all of that in a couple of hundred watts, which basically says it's the most energy-efficient form of computing that the world has ever known. And one single Volta replaces hundreds of CPUs. So it's energy-efficient. It saves an enormous amount of money, and it gets its job done really, really fast, which is one of the reasons why GPU-accelerated computing is so popular now.

With respect to the outlook for our architecture, as you know, we are a one-architecture company, and it's so vitally important. And the reason for that is because there are so much software and so much tools created on top of this one architecture. On the training side, we have a whole stack of software and optimizing compilers and numerics libraries that are completely optimized for one architecture called CUDA.

On the inference side, the optimizing compilers that takes these large, huge computational graphs that come out of all these frameworks, and these computational graphs are getting larger and larger, and their numerical precision differs from one type of network to another and from one type of application to another. Your numerical precision requirements for a self-driving car where lives are at stake to counting the number of people crossing a street. Counting something versus trying to track - detect and track something very subtle in all kinds of weather conditions is a very, very different problem.

And so the type of networks are changing all the time. They're getting larger all the time. Their numerical precision is different for different applications, and we have different computing performance levels as well as energy availability levels that these inference compilers are likely to be some of the most complex software in the world.

And so the fact that we have one singular architecture to optimize for, whether it's HPC for numeric - molecular dynamics and computational chemistry and biology and astrophysics all the way to training to inference gives us just enormous leverage. And that's the reason why NVIDIA could be an 11,000-people company and arguably performing at a level that is 10 times that. And the reason for that is because we have one singular architecture that is accruing benefits over time instead of three, four, or five different architectures where your software organization is broken up into all these different, small, sub-critical mass pieces.

And so it's a huge advantage for us and it's a huge advantage for the industry. So people who support CUDA knows that the next-generation architecture will just get a benefit and go for the ride that technology advancement provides them and affords them. So I think it's an advantage that is growing exponentially, frankly, and I'm excited about it.

Operator

Your next question comes from the line of Vivek Arya with Bank of America.

Q - Vivek Arya {BIO 6781604 <GO>}

Thanks for taking my question and congratulations on the strong results and the consistent execution. Jensen, in the last few months, we have seen a lot of announcements from Intel, from Xilinx and others describing other approaches to the AI market. My question is how does the customer make that decision whether to use a GPU or an FPGA or an ASIC? What can remain your competitive differentiator over the longer term? And does your position in the training market also then maybe give you a leg up when they consider a solution for the inference part of the problem?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Yes, thank you, Vivek. So first of all, we have one architecture. And people know that our commitment to our GPUs, our commitment to CUDA, our commitment to all the software stacks that run on top of our GPUs, every single one of the 500 applications, every numerical solver, every CUDA compiler, every tool chain across every single operating system in every single computing platform, we are completely dedicated to it. We support the software for as long as we shall live. And as a result of that, the benefits to their investment in CUDA just continues to accrue.

You have no idea how many people send me notes about how they literally take out their old GPU, put in a new GPU and without lifting a finger, things got two times, three times, four times faster than what they were doing before, incredible value to customers.

The fact that we are singularly focused and completely dedicated to this one architecture and in an unwavering way allows everybody to trust us and know that we will support it for as long as we shall live. And that is the benefit of an architectural strategy.

When you have four or five different architectures to support that you offer to your customers and you ask them to pick the one that they like the best, you're essentially saying that you're not sure which one is the best. And we all know that nobody's going to be able to support five architectures forever. And as a result, something has to give. And it would be really unfortunate for a customer to have chosen the wrong one. And if there are five architectures, surely over time, 80% of them will be wrong. And so I think that our advantage is that we're singularly focused.

With respect to FPGAs, I think FPGAs have their place, and we are use FPGAs here in NVIDIA to prototype things, but FPGA is a chip design. It's able to be a chip for - it's incredibly good at being a flexible substrate to be any chip, and so that's its advantage. Our advantage is that we have a programming environment, and writing software is a lot easier than designing chips. If it's within the domain that we focus on, like for example, we're not focused on network packet processing, but we are very focused on deep learning. We're very focused on high performance and parallel numerics analysis. If we're focused on those domains, our platform is really quite unbeatable. And so that's how you think through that. I hope that was helpful.

Operator

Your next question comes from Atif Malik with Citi.

Q - Atif Malik {BIO 7312618 <GO>}

Hi, thanks for taking my questions and congratulations on the good results. Colette, on the last call you mentioned crypto was \$150 million in the OEM line in the July quarter. Can you quantify how much crypto was in the October quarter and expectations in the January quarter directionally?

And just longer term, why should we think that crypto won't impact the gaming demand in the future? If you can, just talk about the steps NVIDIA has taken with respect to having a different load and all that. Thank you.

A - Colette M. Kress {BIO 18297352 <GO>}

So in our results, in the OEM results, our specific crypto boards equated to about \$70 million of revenue, which is the comparable to the \$150 million that we saw last quarter.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Longer term, Atif - first of all, thank you for that. The longer-term way to think about that is this. Crypto is small for us, but not zero. And I believe that crypto will be around for some time, kind of like today. There will be new currencies emerging. Existing currencies would grow and value. The interest in mining these new emerging currency crypto algorithms that emerge are going to continue to happen. And so I think for some time, we're going to see that crypto will be a small but not zero part of our business.

When you think about crypto in the context of our company overall, the thing to remember is that we're the largest GPU computing company in the world. And our overall GPU business is really sizable. We have multiple segments. And there's data center and I've already talked about the five different segments within data center. There's ProVis, and even that has multiple segments within it. Whether it's rendering or computer-aided design or broadcast, in a workstation, in a laptop, or in a data center, the architectures are rather different.

And of course, you know that we have high-performance computing, you know that we have an autonomous machine business, self-driving cars, and robotics. And you know of course that we have gaming. And so these different segments are all quite large and growing. And so my sense is that although crypto will be here to stay, it will remain small but not zero.

Operator

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

Q - Joseph L. Moore {BIO 17644779 <GO>}

Great, thank you. Just following up on that last question, you mentioned that some of the crypto market had moved to traditional gaming. What drives that? Is there a lack of availability of the specialized crypto product, or is it just that there's a preference being driven for the gaming-oriented crypto solutions?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Joe, I appreciate you asking that. Here's the reason why. So what happens is when a crypto - when a currency, a digital currency market becomes very large, it entices somebody to build a custom ASIC for it. And of course, Bitcoin is the perfect example of that. Bitcoin is incredibly easy to design a specialized chip for. But then what happens is a couple of different players starts to monopolize the marketplace. And as a result, it chases everybody out of the mining market and it encourages a new currency to evolve, to emerge. And the new currency, the only way to get people to mine it is if it's hard to mine, you've got to put some effort into it. However, you want a lot of people to try to mine it.

So therefore, the platform that is perfect for it, the ideal platform for new emerging digital currencies turns out to be a CUDA GPU. And the reason for that is because there are several hundred million NVIDIA GPUs in the marketplace. If you want to create a new cryptocurrency algorithm, optimizing for our GPUs is really quite ideal. It's hard to do. Therefore, you need a lot of computation to do it, and yet there's enough GPUs in marketplace, it's such an open platform, that the ability for somebody to get in and start mining is very low barriers to entry.

And so it's the cycles of these digital currencies. And that's the reason why I say that digital currency crypto usage of GPUs, crypto usage of GPUs will be small but not zero for some time. And it's small because when it gets big, somebody will build a custom ASIC. But if somebody builds a custom ASIC, there will be a new emerging cryptocurrency. So it ebbs and flows.

Operator

Your next question comes from the line of Craig Ellis with B. Riley.

Q - Craig A. Ellis {BIO 1870408 <GO>}

Thanks for taking the question and, Jensen, congratulations on data center annualizing at \$2 billion. It's a huge milestone. I wanted to follow up with a question on some of your comments regarding data center partners because as I look back over the last five years, I just don't see any precedent for the momentum that you have in the marketplace right now between your server partners, white box partners, hyperscale partners that are deploying at hosted, et cetera. And so my question is relative to the doubling that we've seen year on year in each of the last two years, what does that partner expansion mean for data center's growth?

And then if I could sneak one more in, two new products just announced in the gaming platform, 1070 Ti and a Collector's Edition on Titan Xp. What do those mean for the gaming platform? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Yes, Craig, thanks a lot. Let's see, we have never created a product that is as broadly supported by the industries and has grown nine consecutive quarters. It has doubled year

over year, and with partnerships of the scale that we're looking at. We have just never created a product like that before. And I think the reason for that is several-fold.

The first is that it is true that CPU scaling has come to an end. That's just the laws of physics, the end of Moore's Law. It's just the laws of physics. And yet the world for software development and the problems that computing can help solve is growing faster than any time before. Nobody's ever seen a large-scale planning problem like Amazon before. Nobody's ever seen a large-scale planning problem like DD before. The number of millions of taxi rides per week is just staggering. And so nobody's ever seen large problems like these before, large-scale problems like these before. And so high-performance computing and accelerated computing using GPUs has become recognized as the path forward. And so I think that that's at the highest level the most important parameter.

Second is artificial intelligence and its emergence in applications to solving problems that we historically thought were unsolvable. Solving the unsolvable problems is a real realization. This is happening across just about every industry we know, whether it's Internet service providers to healthcare to manufacturing to transportation and logistics. You just name it, financial services. And so I think artificial intelligence is a real tool. Deep learning is a real tool that can help solve some the world's unsolvable problems.

And I think that our dedication to high-performance computing and this one singular architecture, our seven-year head start, if you will, in deep learning, and our early recognition of the importance of this new computing approach, both the timing of it, the fact that it was naturally a perfect fit for the skills that we have, and then the incredible effectiveness of this approach, I think has really created the perfect conditions for our architecture. And so I think - I really appreciate you noticing that, but this is definitely the most successful product line in the history of our company.

Operator

Your next question comes from line of Chris Caso with Raymond James.

Q - Chris Caso {BIO 4815032 <GO>}

Yes, thank you. Thanks for letting me ask a question. I have a question on the automotive market and the outlook there. Interestingly, with the other segments growing as quickly as they are, auto is becoming a smaller percentage of revenue now, and certainly the design traction seems very positive. Can you talk about the ramp in terms of when the auto revenue, when we could see that as getting back to a similar percentage of revenue? Is that growing more quickly? Do you think that is likely to happen over the next year with some of these design wins coming out, or is that something we should be waiting for over several years?

A - Jen-Hsun Huang {BIO 1782546 <GO>}

I appreciate that, Chris. So the way to think about that is, as you know, we've really reduced our emphasis on infotainment, even though that's the primary part of our revenues, so that we could take literally hundreds of engineers, including the processors

FINAL

that we're building now, a couple - 2,000, 3,000 engineers working on our autonomous machine and artificial intelligence platform for this marketplace to take advantage of the position we have and to go after this amazing revolution that's about to happen.

I happen to believe that everything that moves will be autonomous someday. And it could be a bus, a truck, a shuttle, a car, everything that moves will be autonomous someday. It could be a delivery vehicle. It could be little robots that are moving around warehouses. It could be delivering a pizza to you. And we felt that this was such an incredibly, incredibly great challenge and such a great computing problem that we decided to dedicate ourselves to it.

Over the next several years, and if you look at our DRIVE PX platform today, there are over 200 companies that are working on it, 125 startups are working on it. And these companies are mapping companies, they're Tier 1s, they're OEMs. They're shuttle companies, car companies, trucking companies, taxi companies. And this last quarter we announced an extension of our DRIVE PX platform to include DRIVE PX Pegasus, which is now the world's first auto-grade, full ASIL D platform for robot taxis. And so I think our position is really excellent, and the investment has proven to be one of the best ever.

And so I think in terms of revenues, my expectation is that this coming year we'll enjoy revenues as a result of the supercomputers that customers will have to buy for training their networks, for simulating all these autonomous vehicles driving, and developing their self-driving cars. And we'll see fairly large quantities of development systems being sold this coming year.

The year after that I think is the year when you're going to see the robot taxis ramping, and our economics in every robot taxi is several thousand dollars. And then starting, I would say, late 2020 to 2021, you're going to start to see the first fully automatic autonomous cars, what people call Level 4 cars, starting to hit the road. And so that's how I see it. Just next year is simulation environments, development systems, supercomputers. And then the year after that is robot taxis, and then a year or two after that will be all the self-driving cars.

Operator

Your next question comes from the line of Matt Ramsay with Cannacord Genuity.

Q - Matthew D. Ramsay {BIO 17978411 <GO>}

Thank you very much, good afternoon. I have I guess a two-part question on gross margin. Colette, I remember, I don't know, it was maybe three years ago, 3.5 years ago at an Analyst Day you guys were talking about gross margins in the mid-50s, and that was inclusive of the Intel payment, and now you're hitting numbers at 60% excluding that. If you could, talk a little bit about how mix of the data center business and some others drives gross margin going forward. And maybe, Jensen, you could talk a little bit about - you mentioned Volta being such a huge chip in terms of transistor count, how you're thinking about taking cost out of that product as you ramp it into gaming next year and the effects on gross margin. Thank you.

A - Colette M. Kress {BIO 18297352 <GO>}

Okay. Thanks, Matt, for the question. Yes, we've been on a steady stream of increasing the gross margins over the years. But this is the evolution of the entire model, the model of the value-added platforms that we sell and inclusive of the entire ecosystem of work that we do, the software that we enable in so many of these platforms that we bring to market. Data center is one of them, our ProVis another one, and if you think about all of our work that we have in terms of gaming and that overall expansion of the ecosystem. So this has been continuing to increase our gross margin.

Mix is more of a statement in terms of each quarter, we have a different mix in terms of our products, as some of them have a little bit of seasonality. And depending on when some of those platforms come to market, we can have a mix change within some of those subsets.

It's still going to be our focus as we go forward in terms of growing gross margins as best as we can. You can see in terms of our guidance into Q4, which we feel comfortable with that guidance, that we will increase it as well.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

With respect to yield enhancement, the way to think about that is we do it in several ways. The first thing is I'm just incredibly proud of the technology group that we have in VLSI. And they get us ready for these brand-new nodes, whether it's in the process readiness through all the circuit readiness, the packaging, the memory readiness. The readiness is so incredibly important for us because these processors that we're creating are really, really hard. They're the largest things in the world. And so we get one shot at it.

And so the team does everything they can to essentially prepare us. And by the time that we tape out a product for real, we know for certain that we can build it. And so the technology team in our company is just world-class, absolutely world-class. There's nothing like it. Then, once we go into production, we have the benefit of ramping up the products. And as yields improve, we'll surely benefit from the cost.

But that's not really where the focus is. In the final analysis, the real focus for us is continue to improve the software stack on top of our processors. And the reason for that is each one of our processors carry with it an enormous amount of memory and systems and networking and the whole data center.

Most of our data center products, if we can improve the throughput of a data center by another 50%, or in our case oftentimes we'll improve something from 2x to 4x. The way to think about that is that \$1 billion data center just improved its productivity by a factor of two. And all of the software work that we do on top of CUDA and the incredible work that we do with optimizing compilers and graph analytics and all of that stuff then all of a sudden translates to value to our customers, not measured by dollars but measured by hundreds of millions of dollars. And that's really the leverage of accelerated computing.

Operator

Your next question comes from the line of Hans Mosesmann with Rosenblatt.

Q - Hans Mosesmann {BIO 1522582 <GO>}

Thank you. Hey, Jensen, can you comment on some of the issues this week regarding Intel and their renewed interest in getting into the graphics space and their relationship at the chip level with AMD? Thank you.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

Hey, Hans. There's a lot of news out there. I guess some of the things I take away, first of all, Raja [Koduri] leaving AMD is a great loss for AMD. And it's a recognition by Intel probably that the GPU is just incredibly important now. And the modern GPU is not a graphics accelerator. The modern GPU, we just left the word G in there, the letter G in there. But these processors are domain-specific parallel accelerators, and they're enormously complex. They're the most complex processors built by anybody on the planet today. And that's the reason why IBM uses our processors for the world's largest supercomputers. That's the reason why every single cloud, every major cloud, every major server maker in the world has adopted NVIDIA GPUs. It's just incredibly hard to do.

The amount of software engineering that goes on top of it is significant as well. And so if you look at the way we do things, we plan a roadmap about five years out. It takes about three years to build a new generation, and we build multiple GPUs at the same time. And on top of that, there are some 5,000 engineers working on system software and numerics libraries and solvers and compilers and graph analytics and cloud platforms and virtualization stacks in order to make this computing architecture useful to all of the people that we serve.

And so when you think about it from that perspective, it's just an enormous undertaking, arguably the most significant undertaking of any processor in the world today. And that's the reason why we're able to speed up applications by a factor of 100. You don't walk in and have a new widget and a few transistors and all of a sudden speed up applications by a factor of 100 or 50 or 20. That's just something that's inconceivable unless you do the type of innovation that we do.

And then lastly, with respect to the chip that they built together, I think it goes without saying now that the energy efficiency of Pascal GeForce and the Max-Q design technology and all of the software that we created has really set a new design point for the industry. It is now possible to build a state-of-the-art gaming notebook with the most leading-edge GeForce processors and be able to deliver gaming experiences that are many times greater than a console in 4K and have that be in a laptop that's 18 millimeters thin. The combination of Pascal and Max-Q has really raised the bar, and I think that that's really the essence of it.

Operator

Unfortunately, we have run out of time. Presenters, I'll now turn the call over to you for closing remarks.

A - Jen-Hsun Huang {BIO 1782546 <GO>}

We had another great quarter. Gaming is one of the fastest growing entertainment industries, and we are well-positioned for the holidays. AI is becoming increasingly widespread in many industries throughout the world, and we are hoping to lead the way with all major cloud providers and computer makers moving to deploy Volta. And we're building the future of autonomous driving. We expect robot taxis using our technology to hit the road in just a couple of years. We look forward to seeing many of you at SC17 next week. And thank you for joining us.

Operator

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

This transcript may not be 100 percent accurate and may contain misspellings and other inaccuracies. This transcript is provided "as is", without express or implied warranties of any kind. Bloomberg retains all rights to this transcript and provides it solely for your personal, non-commercial use. Bloomberg, its suppliers and third-party agents shall have no liability for errors in this transcript or for lost profits, losses, or direct, indirect, incidental, consequential, special or punitive damages in connection with the furnishing, performance or use of such transcript. Neither the information nor any opinion expressed in this transcript constitutes a solicitation of the purchase or sale of securities or commodities. Any opinion expressed in the transcript does not necessarily reflect the views of Bloomberg LP. © COPYRIGHT 2021, BLOOMBERG LP. All rights reserved. Any reproduction, redistribution or retransmission is expressly prohibited.

FINAL

Bloomberg Transcript