JP Morgan Healthcare Conference

Company Participants

- Kimberly Powell, Vice President of Healthcare

Other Participants

- Harlan Sur, Analyst, J. P. Morgan

Presentation

Harlan Sur {BIO 6539622 <GO>}

All right, good morning and welcome to JP Morgan's 38th Annual Healthcare Conference here in San Francisco. My name is Harlan Sur, I'm the semiconductor analyst here for the firm. And for the second consecutive year we have the team from NVIDIA presenting. And for those of you that don't know NVIDIA, they're are leader in visual and accelerated computing silicon and platforms in areas like artificial intelligence and deep learning, powering some of the world's most powerful supercomputers and driving compute innovations for the cloud and hyperscalers as well as driving large vertical markets like healthcare.

Here with us today from NVIDIA is Kimberly Powell. She is the Vice President and General Manager of Healthcare at NVIDIA. She is responsible for the Company's worldwide healthcare business including domain targeted hardware and software platforms for accelerated computing artificial intelligence and visualisation that power ecosystems of imaging, genomics, drug discovery, and precision medicine. Ms. Powell has been with the NVIDIA team since 2008.

So Kimberly, thank you very much for joining us today. Let me turn it over to you.

Kimberly Powell {BIO 22145194 <GO>}

Thank you so much. Good morning, everyone. Harlan, thank you. It's an honour and a pleasure to be here at J.P. Morgan again as a presenter. It's always been a fantastic conference. So today, I would like to give you a little bit more insight into what NVIDIA is working on in the area of healthcare. I think you're all familiar with our Safe-Harbor statements. So I'll leave this up here for a quick glance. And then I'll hop right into it so that we can get through some great insights on this fourth day of J.P. Morgan. So NVIDIA is at the intersection of three mission critical Computer Sciences. If you think about it, the intersection of computer graphics, high performance computing or accelerated computing and artificial intelligence, and these three critical computer science domains are all capable of operating on NVIDIA's one architecture. It's is a very unique where developers of all kinds can invest in a singular architecture and also innovate at the intersection of all three of these computer science domains.

And if you think about that every large, major and important industry really relies on it. You or your children know that NVIDIA is known for our incredible gaming industry. Gaming wants to be immersive. It wants to look like you are inside of a virtual world. You can only do that through the power of that inner section of computer graphics, accelerated computing and artificial intelligence. Cloud computing, today's cloud computing industry is innovating at a pace in artificial intelligence that I'll brief you on quickly, where they need to train up the artificial intelligence at massive scale with massive data and then they need to serve millions if not billions of users simultaneously.

Cloud computing is a high-performance computing challenge. Think about transportation, transportation autonomous vehicles, impossible without all three of those computer science domains. You have sensors bringing in data that needs to do real time artificial intelligence. You need to stimulate the world in order to know if that car can safely drive itself. You need high performance computers to build up the infrastructure in order to build the safest artificial intelligence that are going to drive the transportation industry moving forward.

Manufacturing is an incredibly exciting space as well. This is where you have large warehouses and robots and robots with humans that want to interact. If you want robots and humans to interact, there is a massive amount of artificial intelligence that needs to go into these systems. And it will revolutionize manufacturing. And I'm here to talk to you of course today about health care. If you think about healthcare, the thing that we like to do first and early intervention is to see inside the body.

This is a graphics and visualization problem. Then we want to stimulate how drugs are responding inside the body. This is a accelerated computing high performance computing challenge and then we want to incorporate all of patient data, whether it be genomic data, imaging data, health record data in an artificial intelligent manner. So that we can drive towards that dream of precision medicine. And as you can see that one architecture can serve multiple industries and it's really, really critical for the innovation in those industries as well.

So if I could take a moment to indulge you on this fourth day, and wake you up a little, let that coffee start running through your veins. NVIDIA, this past year invented a new technology called RTX. It's real-time ray tracing, and what that means in simple terms as we can simulate light in the most photo realistic real way that you can imagine. How does light bounce off of a material? How do shadows appear? Such that what we create in the graphics industry is completely real. So let me talk you through these different areas where it's happening. Project Sol is a project and video built to show cinematic as if you're doing motion pictures and it's completely done through computer graphics, real-time ray tracing with artificial intelligence is what made that come to a reality

It's been our dream. Researchers of ours dreamed for 35 years. We thought that it was going to become available to us in 10-years and we made it a reality last year in 2019. This same technology is revolutionizing the most popular game on the planet Minecraft. If you imagine what Minecraft is, its users inside this environment who are building the environment in real time. There is no way to pre-bake the beautiful illumination that happens inside these games. You need to do it in real time and we've enabled that as well.

This same technology allows architects and builders to simulate light of the city, 365 days of the year. So they know exactly what it's going to look like before they build it. And even into an incredibly exciting area where we've cooperated with NASA, using the Summit computer, which is the world's fastest supercomputer powered by NVIDIA GPUs, creating 150 terabytes of simulation data, we're able to stimulate the Mars Lander, landing on Mars. And this is just breathtaking. A 150 terabytes of simulation data visualized. This gave the researchers and scientists working on this just an incredible amount of insight to see all that computation truly visualized.

So we got to think about NVIDIA as there is one architecture that has a job. Its job is different than the job of a CPU. It's accelerated computing. There is a program that wants to do some level of sequential computing and another level of parallel or accelerated computing. We are a complement to the CPU, but essentially a time machine. There are X-factors involved, 10 times, 50 times, 100 times, 500 times speed up in the application performance.

So you want to give the application and developers the right tool for the right job, but you have to go beyond the processor itself. You have to think about it from a complete full stack. Because the idea of accelerated computing is now a known understanding with the ending of Moore's Law. You have to think about going into a full stack to realize the Super Moore's Law experience. It's a new programming approach that we have been pioneering for the last 20 years called accelerated computing and we take that full stack approach to take the acceleration beyond just what the processor is capable of. And you've probably seen in video over the past five or more years entering into the systems architecture business.

We created the world's first AI supercomputer called the DGX. This is a full system approach. Our DGX 2 has 16 GPUs in it, interconnected by high speed interconnections . So that it's an AI Supercomputing appliance. It's allowed us to do amazing things. We have edged supercomputing systems. We have embedded device supercomputing systems. And because we have one architecture and because the system resides everywhere in every cloud service platform in every server OEM, it's available to every developer that needs it and wants it.

And we now have over 1.5 million developers who are operating on this platform. So NVIDIA is the engine of artificial intelligence. AI is in need of accelerated computing every single day that is becoming more and more true and more and more apparent. The compute is doubling every three and a half months in the realm of artificial intelligence. And if you think about what NVIDIA has done in the processor technology, the full stack approach into the system architecture, we're able to do

Super Moore's Law. If you take a benchmark by the whole industry, it's called MLPerf, one of the benchmarks there is ResNet-50 to do that training.

It would take you 25 days, just five years ago. We can now execute that same training with the full stack approach in two hours. And this is why NVIDIA is leading in both the training and inferencing of AI platform and we are number one in the MLPerf benchmark. This has kicked off what we call the AI Renaissance. There are two significant ways that I would like to mention here, because it's so applicable to what needs to happen in healthcare with artificial intelligence. With the advent of AlexNet, we gave the capability to see things in superhuman ways and create all sorts of amazing applications.

And just about 18 months to 24 months ago in the 2018 time-frame, we also had a critical breakthrough in natural language understanding, which gives us this new level of understanding, being able to do question and answering, being able to summarize, being able to actually write things, have a dialog. This natural language understanding is now a superhuman capability. So what does this all mean in healthcare? Hopefully, we're asking that. So our question to that -- that question of ourselves. So NVIDIA has been in healthcare for over a decade. Thank you, Harlan for mentioning that. We've been in this space for a long time. Diagnostic Imaging was one of the early adopters of our new computing approach called accelerated computing. They were creating sensors, CT devices, MR ultrasound that needed a new level of computing to do real time imagery construction, to do 3D and 4D visualization of ultrasound.

Even into the molecular simulations and drug discovery space, we can simulate how diseases manifest themselves inside the body and how molecules and drugs can attack them. We had a early indicator that AI was going to be incredibly useful and powerful in health care, when there was a breakthrough and a challenge in the year 2013, where deep learning was applied to pathology to detect mitosis.And that really kicked off our understanding in investment of how we could make a difference in healthcare.

We worked with the community, the ecosystem to create open data-sets. The data science (inaudible) put on by Booz Allen Hamilton kicked off, how can you automatically detect ejection fraction. We established a partnership with Massachusetts General Hospital and Brigham Women's Center for Clinical Data Sciences and we we invented a way to use against Generative Adversarial Networks to create synthetic data for MRI to do training. It kicked off our thinking that we need to create a domain specific platform that we now call Clara. We've been applying that now with an applied research team that is established at NVIDIA to win challenges and put all that understanding of how to build as well as deploy artificial intelligence into the compute platforms that we hope are going to fuel the future of AI in healthcare.

So let's take a minute to understand that artificial intelligence is domain specific. If you think about any one of our intelligence, it is domain specific. And so we need to do is, we need to understand what world are we living in. There are sensors in that world that allow us to generate data and then we need to create AI programs with computers that can automate a portion of that industry. Here we're talking about self-driving cars. We sensor a car and we build applications that need a computer inside the car to automate the driving process.

The same holds true for manufacturing. We have warehouses, we want robots to be able to navigate those warehouses and automate the work inside them. And the same is for healthcare. We have radiologists and doctors that need to see, they need to understand and we need to take the data that we generate, build AI programs and deliver levels of automation, so that the massive demand that we have on our physicians today can be met. And that's why NVIDIA created the NVIDIA Clara platform. It is a domain targeted AI application framework. What does that mean? What that means is, we have built a whole software application suite that allows you to bring domain expertise into a software platform already built for you and create an AI application. It's almost like zero coding. We've written all the software for you to be able to label the data, transfer learn on that data and then deploy that data to a computing platform. We call this the NVIDIA AI application framework.

So what you can do is bring your data and create AI applications. This platform and the artificial intelligence capabilities is kicking off the smart medical device revolution. There are going to be medical devices of all kinds. They could be smart sensors something like be AliveCor or the care.ai, where we're sensing patients through very accessible sensors like RGB camera technology or radar technology and being able to make real time understanding of what is happening in the world around them.

We also have smart imaging devices, because of artificial intelligence we're going to see sensor technology completely redefined. Because our artificial intelligence and accelerated computing can do things with the sensor data that we couldn't do before, you can imagine, creating smaller, cheaper, sensor technology. Mobile MRI is now a real thing. HYPERFINE is a company who has invented that. It's a small as 3 feet high, 3 feet wide, it's on wheels. It's a mobile MRI device, it no longer takes complete infrastructure, a dedicated room in a hospital.

It can image, a three-month-old baby without having to give it any anaesthesia. These are breakthroughs you're going to see come day after day in the months and years ahead. It also applies into treatment, achieving such a critical area where if you think about a Varian using their complete intelligent platform to make a personalized approach to radiation therapy. Of course robotic surgery, another incredibly important domain. There is one challenge or many challenges, actually I should say that is in healthcare that may not be in other industries that we have to think through.

Data is a very sensitive topic in healthcare. Data is very different in health care. If you think about having to label some of the data in the healthcare system, you need trained experts. It's a critical challenge. To label data in radiology, generally you have to use radiologist to do it. And they're already oversubscribed. So we've got to think about tools to facilitate that. There is data diversity issues. We have medical devices that live 15 years in the installed base and they create data that looks like it was generated 15 years ago and then you have very cutting edge data. That diversity can introduce some instability into AI platforms. And I would say the very third one, and probably the most important is the data privacy issues, data sovereignty issues.

So we created a way, a new AI application development framework and capability, we call Federated Learning. Federated Learning is going to become the new standard in AI development in healthcare. The reason for that is this. The data remains in its residency. It never has to leave where it is, whether that be a hospital, a pharmaceutical company, an insurance company, but you want to be able to collaborate. You need to see the diversity of a hospital patient network in California versus in Texas versus Massachusetts and never mind expanding that to the demographics of the rest of the world.

And so now with this AI application framework, we've created this capability of Federated Learning. You introduce your data to the system, you locally train on your data, and all you share with the Federated Management System is the model weights. You could reserve all the privacy of the data and you allow that data to remain resident. This is going to be truly transformational. And we're already seeing massive adoption. We have the medical imaging industry, who is taking this incredibly quickly into building AI applications.

The American College of Radiology, the Clinical Center for Data Sciences, the UC Systems. We have the NHS that we're already piloting at four NHS hospitals, even in a public healthcare system like the NHS, they have to keep the data where it resides. Another incredible project that just announced last year is the MELLODDY Project. Pharmaceutical industries, their data is their IP. Yet, they all know they need to innovate. So 10 of the largest pharmaceutical companies have created a consortium called MELLLODDY, and they're are -- we're going to create essentially a machine learning ledger orchestration of a drug discovery, leveraging this same application framework for Federated Learning.

So NVIDIA Clara is our AI application framework for imaging. We help the industry create pre-trained models, really great starting points, all the tools to label the data, develop the applications and deploy those applications wherever they want to be. We can make a contribution in another very important and I think a very, very inflection point that we're all excited to see come to fruition and that is in the area of Computational Genomics. There is a phenomenon as we all know that the cost has been decreasing and that's fantastic. But there is another super interesting data point, which is the instruments that are being created. The throughput is beyond what we've ever seen before. MGI has an instrument that can do 60 genomes per day, 20,000 genomes per year by a single instrument, that means we're going to be generating so much data. And because the cost is decreasing, we're going to move to where genotyping and whole-genome sequencing are approximately the same cost, so we're going to hold genome sequence.

It has so much more data than 0.5% of the whole-genome when you're doing just genotyping. So whole-genome sequencing and population genomics is on the rise. And in fact the data shows it. We're doubling the data every seven months and what that means is by 2025 we should have well over 200 million whole-genome sequences. If we wanted -- excuse me, we wanted to do Variant calling on that 200 million, it's 2 trillion CPU [ph] hours. It would take the entirety of every CPU in the cloud globally, 200 days to do that computation.

Obviously this is a place where we need accelerated computing. We need accelerated computing and we need artificial intelligence. So just a few weeks back we acquired a fantastic genomic software company called Parabricks. They have GPU accelerated. The industry standard GATK Varian calling pipeline and we have lots of roadmap innovations along the area of artificial intelligence for Varian calling as well as somatic to bring us into the very exciting and very powerful and very important area of cancer genomics.

There are breakthroughs happening in computational genomics and last year it was just a fantastic year for it. We saw MGI and NRT7 [ph], as I said 60 genomes per day. NVIDIA GPUs are inside these instruments because there's no other way that you could create the base calling at the throughput these devices are happening without them. Similarly, Oxford Nanopore, the next-gen -- next generation of sequencing, where AI is at the core of being able to get accurate based calls out of the signal of nanopore sequencing.

Parabricks and Personalis for cancer genomics because they need to do it faster, and they need to do it more accurate to deliver their service. At the Sanger Institute, the Mutographs project, where we're trying to understand what are the DNA signatures that cause cancer in a population genomics program, and we're applying artificial intelligence and machine learning, getting 30 times speed-up things again that would have taken upwards of 20 days into the hour time frames. This is going to be revolutionary for the research in cancer genomics. De Novo Assembly, because of nanopore sequencing, long read sequencing, De Novo Assembly is now going to become the norm, but through accelerated computing. We want to sequence every species on the planet, but we also want to be able to use long read technology, so that we can understand that the reference genome is -- maybe not enough to be able to have reference genomes for all of our populations. So just a lot of exciting times ahead. We're even working very closely with the Single Ccell ATAC-Seq community. If you think about the future of genomics, whole-genome sequencing, population sequencing incredibly important and we're going to see that rapidly evolve here, but there is a huge road-map ahead in sequencing.

We want to do sequencing especially in cancer genomics, you don't do it once, you do it many times. As we enter into the liquid biopsy phase, you're going to potentially want to do it every single day. You're looking for that floating (inaudible) that certain circulating tumor. We need to build. Now is the time to invest and build accelerated computing in AI applications. So NVIDIA and healthcare and accelerated computing is fast forward. Powering the next generation of medical devices, whether they be smart sensors, mobile imaging devices, high throughput devices of all kind that let us see into the what makes the bodywork is critical.

We are creating domain targeted applications and hardware applications that are either in a data-center at the edge, in the cloud, in the device, wherever the developer, the business model, the application wants to reside because we have one architecture that lives everywhere. And the breakthroughs in the industry are proof that this is happening and it's super exciting.

So with that, I would like to conclude, and thank you for your attention.