Patrick Moorhead: Dario welcome to the Six Five Summit, 2021. And I just want to thank you for speaking here on day four.

Dario Gil: Thank you so much, Patrick, for inviting me.

Patrick Moorhead: Yeah. We have some really great IBM speakers and you are one of them and it's great. We had Head of Strategy and now we have Head of Research. But before I dive in, maybe I'll let you introduce yourself and what you do at IBM?

Dario Gil: Yeah. As you said, I'm responsible for research for the entire company, globally and I would say overall, for the technology roadmap and the technical community of all of IBM.

Patrick Moorhead: Well, that's great and that spans so many different areas, hardware, software, cloud. I'm really impressed at what you guys do here. Can you talk about what IBM research does and why it exists? One thing I love to point out to everybody is, we always say R&D. Right? And they're so different, I don't know why we call it R&D because research is farther out, it's risky, it's not definitively tied to an exact product. Sometimes it works, sometimes it doesn't. And development leverages all the research and hard work that is done out there, but I guess we put it on P&L's, as R&Ds, so we call it R&D?

Dario Gil: I mean, you are so right about that and the horizons are different and the risk is different. And the first observation I would make is, the research division of IBM has existed for 76 years. So the first reflection I would put forth is, that given that we've been in the information technology business as a company for 110 years, and is the only company that has been able to do information technology for over a century. It begs the question of, why others haven't? Right? Or, what is the difficulty? And that is, that integrated over time, in an industry that changes so dramatically, is the normal course of events is to disappear. And what I would say is, the reason why IBM has been able to survive and to be able to do these advances throughout all these periods, is really because we've had a research division that is able to look across different horizons.

So I agree that one characteristic of it is, you can take on a lens that is more risky, some things don't work out, that's the nature of science and the nature of research activities. And another one has to do with the composition of the teams that conduct the work. Right? So, it's composed of lots of scientists that are really focused on applying the scientific method in our context, for the world of new ways to process, store, and use information in creative, new ways. So in the end, I mean, there's a linkage, let's not over push it, right? You've got to then derive the value from research to development and make products, but they are distinct fields.

Patrick Moorhead: Yeah. And I think it would be valuable to maybe talk about how ultimately, the research that you do, does end up positively impacting your customers or what IBM lovingly refers to, as clients?

Dario Gil: Yeah. No, that's right. I mean, the ultimate value is to make sure of that impact. So let me give some examples. So let's say, since I know we're going to be talking about semiconductors in the world of hardware, if we look at the world of systems, and if you look at our mainframe franchise. So you will go and you say, "Wow, how can a product, a franchise, last that long, that many decades of generation after generation of delivering value?" And underneath it, is because we reinvented every generation. And there is a heavy, heavy element of research that goes into it, from the design of architectures for the microprocessors, to the core semiconductor technology, that we're able to use.

So if you look at many of the seminole innovations in the semiconductor industry that have taken place over the years, from copper interconnects to High-k metal gate and many technologies that were created by research in IBM as an example, in the context of impacting that product. If you look at the world of software, we see similar things happening, right? With our Watson portfolio in AI. But I'll give you another extreme, which is a CodeNet new market, a whole new industry, like quantum computing, that originated here in research, and after many decades of work, we're now seeing its commercial realization.

Patrick Moorhead: Yeah. I have to point out that, so Intel CEO, Pat Gelsinger, he led the keynote a couple of days ago. I know he's a customer, and that was super exciting to see them sign up with IBM research and also on the slides is Samsung. And so you have the number one and two largest semiconductor manufacturers publicly working with you and I think we both know, that what's not public, is a lot more people, than the number one and number two semiconductor folks.

Dario Gil: You know what that testament is, is on their respect and the value that our very broad ecosystem of industry leaders pay to the research powers of IBM. Across the research division, we employ 3,000 scientists. Right?

Patrick Moorhead: Yeah.

Dario Gil: We're well over 2,000 PhDs and if you look at our almunia ecosystem that we have created now over 20 years, to create the world's most advanced semiconductor collaborative environment for R&D, in these cases, attracted leading equipment makers, the leading EDA companies. But you're talking about nowadays, semiconductor manufacturers, right? Companies like Samsung and now Intel.

Patrick Moorhead: Right.

Dario Gil: And what's beautiful about this, is that we've got to leverage the strengths, the complementary strengths of all the different players, all the different companies. But it's a reflection of that respect and that value, for investing fundamentally, in research, that is the reason why we keep collaborating and expanding the ecosystem. And indeed, the Intel announcement-

Patrick Moorhead: Right.

Dario Gil: ... if you asked me and I'm sure like you, a few years ago, I'd say, "Wow, would you imagine Intel and IBM working together on semiconductors?" We'd say, "Oh, that's not possible." But it's a testament of, I think our leadership and the approach to ecosystems, of PATs, the transformation agent-

Patrick Moorhead: Right.

Dario Gil: ... that's coming to Intel and look, what would seem impossible, now it's his reality. Right?

Patrick Moorhead: Yeah. I had to go back and search for IBM and Intel alliances. And I think I had to go back to before when I started working in the '80s, it had been a while. So it was good to see. And you had a really big announcement with 2-nanometers that I know, sent shock waves for the industry. I know that it's funny, you always read, "Oh, we're not going to get to this lithography or we're hitting a brick wall." And the industry always finds a way and IBM found a way to, and you actually had a working product, the 2-nanometer. Can you talk about that a little bit?

Dario Gil: Yeah. I mean, we are incredibly proud of the work of the team, it is the fruit of four years of work. So this is based on a transistor technology called nanosheets, it's a form of gate-all-around transistor. We had announced actually, about four years ago, the introduction of the nanosheet technology for the 5-nanometer node, as well. And now we've been able to demonstrate after four years of work, that the whole technology, could be designed for the 2-nanometer node. And what that is going to mean, is that we are going to be able to design chips that are going to either improve the performance by about 45% compared to 7-nanometer technology or reduce the power consumption of the same performance by 75%. So it's a real breakthrough, we're incredibly excited and we really believe that it is going to be the core technology that is going to be used by the whole industry.

Patrick Moorhead: Yeah. And I may have used the wrong word device, 2-nanometer device, which that technology will make its way into both power and Z and the rest of the industry. But we can't talk about that, of course. So other announcements you made, I think were pretty exciting and sometimes people gloss over these. But you actually have developed a system that uses artificial intelligence to inspect code and to be able to help developers move to the cloud. So let's say, you had an application that was written 20 years ago, you put it into CodeNet and essentially, it tells you what's used and what's not, and I'm kind of stealing your thunder here. But that's so exciting, I can't help myself. So why don't you talk about it?

Dario Gil: I mean, it's actually had a tremendous impact and there's just so much excitement, externally in the community and the open-source. So here's an

analogy I would provide. We're all familiar with the impact that ImageNet had, I would say in the late 2000s, early 2010s, in the world of AI. So what was it? It was the curation of a massive data set of images that have labels and combined with large-scale neural networks, deep learning, and compute power, it transformed how computer vision was being done and image processing and we've witnessed the last eight years, of the deep learning explosion. So this thing by analogy, we actually purposely named the data set CodeNet, like ImageNet. And what we've done, is a massive data curation engineering effort, where we have taken tens of millions of lines of code, of all sorts of languages I mean, we're not talking only about a contemporary Python and C++ and Java, but we're talking about Cobol. Right?

Patrick Moorhead: Yeah.

Dario Gil: Now that we're there, we've learned that people use it a bit, still in payroll benefits in New Jersey. Remember, in the middle of the pandemic, when they were expanding unemployment benefits and people were saying, "There's a shortage of Cobol programmers." Right? So, actually, if you look at the economics of this, it takes about, something like 50 cents a line of code, when you have to do this manually to convert it. So what have we done? We've curated a massive amount of sample programs that are written in different languages and just like that curation in the past with languages enabled simultaneous translation, using statistical machine learning, the idea is to use AI and train large neural networks to be able to do these code translation. And it begins by curating these data sets and using computing power. So I think that if you just look at the productivity potential of software engineering assisted by AI, software is like the Lingua Franca business. So enhancing the productivity is massive.

Patrick Moorhead: Yeah, it is, and getting people to a cloud model is important. And it's so funny, I remember Y2K, we had a shortage of Cobol programmers. I'm thinking, "Okay, we're in 2021, there's plenty, an extra, extra, extra, shortage out there." Actually, IBM and the Z Team has done a great job on keeping COBOL moving forward. Let's hit one more. It's funny, real-time or near-time technology here are called Qiskit. So we sometimes think of quantum computing as 10 or 15 years from now, and we talk about, "Hey, isn't this hardware beautiful." And by the way, you have some of the most beautiful hardware that exists out there. But there's real software and runtimes that go along with that. Can you talk a little bit about the Qiskit runtime?

Dario Gil: Yeah. So Qiskit stands for quantum, information, science, software kit, Qiskit and it's the world's most popular open-source environment to do development in quantum computers. And what we announced, is the creation of a runtime, where now you have the combination of classical and quantum resources, tightly interlinked, so that when the user creates a program, all of that happens behind the scenes, to be able to execute the program very, very efficiently. And what we demonstrated in doing this, it's 120 X speed up, compared to what you could do when you're using your laptop as the environment, to be able to

orchestrate the productivity of the quantum program. And what we have to remember, is behind the scenes, you're going to have these quantum circuits that are going to be running in the quantum computer, and you're going to need to execute hundreds of millions or a billion at these quantum circuits, to be able to create value from an application.

So the productivity of the software environment, it's a massively important effort of it. And we're just incredibly excited about the fast-growing ecosystem and really the advancement of the runtime as we mentioned, to be able to create a whole new generation of applications. I'll just give you one fact. Today, at any given day, we run now, over 2 billion quantum circuits a day. So it's fantastic, we've seen this exponential growth of the community, of learning how to program and creating these new generations of applications. So we need to envision these frictionless software development environments.

Patrick Moorhead: Yeah. So literally real-time, we're talking about shrinking lithography down to 2-nanometer with nanosheets, we have machine learning and AI writing its own programs. And we have real quantum runtimes out there to make real applications in the future and actually, develop algorithms for today. But this is where you get paid the big bucks here, Dario, which is about the future. I want to talk about the future and get out your crystal ball here. Actually, I'm sure you have a plan, it's more than a crystal ball, but let's say beyond two years. I know you go out a lot farther, but every year you got out, it gets a little foggier. What are we looking at two years from now?

Dario Gil: So these three trends that we have talked about, which I summarize about, the future of computing is basically, bits + neurons + qubits, right? So high precision transactional processing, the world of neural networks on artificial intelligence, and the world of qubits. I think that the part that is not sufficiently understood by the industry, is that each one of those, we understand is a massive market and a massive roadmap and we work and invest very aggressively in advancing each. But I think the part that people don't sufficiently understand, is what's going to happen with their convergence? What happens when you actually architect systems in a hybrid cloud environment, to be able to bring together the best of those three worlds? And I just want to tee off an opportunity that we're very, very focused on, which is just to elevate it.

What do I really believe? What I believe is, that science and the scientific method needs to be elevated and diffused more broadly across society, to be able to solve the most existential problems that we face. I mean, and the public is experiencing the power of science in a way that I don't think they have experienced it in their lifetime, which is with the development of the vaccine in record time. So imagine a world where we can compress that time to discovery. So what we're really focused, is about creating these next generation of architectures and systems to be able to bring the best of bits, neurons, and qubits, to accelerate scientific discovery. And if we did it, it will be like the mRNA platform for vaccines, but for the world of scientific advancement at large. So

that's the mission that we have, the component technologies, their convergence, and a purpose, which is the scientific method accelerated.

Patrick Moorhead: Now, let me make sure I get this right for the audience. Bits, neurons, and qubits?

Dario Gil: That's right. Bits + neurons + qubits is the future of computing.

Patrick Moorhead: Wow. You need to get that on a license plate or something, BNC before somebody jumps on it. No, that is super exciting. And you can't get everything from the past, but it can give you an indication of how some things can come together and then using the art of the possible to pull that together. And this combination of creativity, right brain and left brain coming together to create the future, I think is fascinating and it makes pedestrian things like quantum, look easy.

Dario Gil: I wouldn't say that, no.

Patrick Moorhead: I know, I'm joking of course, but the way that I've seen GPUs leverage traditional computing in a similar way, that this is my expectation. I think it's years of how we get our first practical toe in the water on quantum, beyond very important algorithms that we're doing. It'll be a combination of a traditional program accelerated by quantum. Okay? And whether you want to call those, the bits that are being used, or you're referring to things like 2-nanometer, but it's this combination of these three things, that creates some pretty exciting things.

Dario Gil: You're exactly right. And I think, the industry talks so much about what happens at the limits of Moore's law and what's after? And what I think is exciting is, we actually have an answer to what's happening. Is that A, we keep pushing the limits of it, you used the 2-nanometer example and beyond. But beyond that, it's really this, it's really architectural diversification and the combination, and is the combination of these three architectures that are going to represent the next many, many decades of advancements. And we're seeing that innovation is happening through the entire stack on each one of those elements.

I mean, the nano technology is a good example of that but if you look at AI, beyond reduced precision architectures, we're investing very aggressively in almunia, as well, on analog architectures for AI. And then you look at quantum, is a whole other set of devices based on the transmon qubit, with superconducting technology, and so on. So what you're seeing is across the stack, these beautiful new integrations are possible. And I guess the point I'm making, is that in addition to that, we're going to have the implications of the architectures of combining them. How much classical versus how much quantum?

Patrick Moorhead: Right.

Dario Gil: How much AI versus how much of the high precision? That in its own form is going to be a form of engineering differentiation?

Patrick Moorhead: I want to dial us back a little bit. Let's pretend we're two to five years out. How do bits, neurons, and qubits make an impact on the cloud? And listen, there's new things in the cloud all the time. Just when we're all settled on containers, we see serverless. Right? And new things will come up but how do bits, neurons, and qubits, how might that affect our cloud environments, moving forward?

Dario Gil: Yeah, it's a great question. So what we need to do with a hybrid cloud platform, is two fundamental trends, are important to wrangle the complexity of this heterogeneity that is underneath and mask it for the user. Right? We don't want the user to have to think about quantum physics. Right? That we want to enhance their productivity as developers and as users of the overall system. So what do we need to do, to do that? My view, I like on this and the philosophy that we take at IBM, is that you need a thin layer of interoperability and it's based on open technologies and the right abstractions. So open, there's an answer to what the operating system is for this, which is Linux. And then on top of that, in terms of wrangling that complexity, on top of that, you need the right abstraction.

Containerization is a wonderful example of the encapsulation of software and abstraction. Right? And we've seen in this evolutionary battle of this abstraction, that Kubernetes has emerged as the dominant form of these kinds of orchestration. But you use the right trend of what is happening, which is a world of serverless, which is, you need to raise even further, that level of extraction. So that division, in the end, needs to be that massive amount of compute power, heterogeneous that is across multiple clouds, private and public, multiple public ones, for a user should feel, this is the ultimate vision, as if it was a single computer. And we're accustomed to sitting in front of our laptops and doing file transfers, doing things like that, programming, and so on.

Patrick Moorhead: Right.

Dario Gil: Imagine that, but behind the scenes, is this hybrid cloud infrastructure. That is where we need to be able to get to through open technologies and abstractions.

Patrick Moorhead: So Dario, every time I talk with you, I pull something back and I love this whole trifecta of bits, neurons, and qubits. You may have said it before, but it didn't internalize with me and I'm sure it's going to be new to our audience. But doing this in an open and at the right level of abstraction to deliver this one, looking computer, that can be many computers, but it looks like one. Wow, I'm going to have to think about that for a while. I think our audience is going to have to think about that, as well. So I just want to thank you for coming on. I mean, I could do this with you for the next hour, but people might not want it to be that long, but if they do, they can come find all of your videos.

You did a pretty cool one that I thought was very cool, that gives the full double-click on what you're doing. But listen, for the sake of the industry and IBM clients, I thank you for what you and your team are doing. There's not a whole lot of research being done anymore. I mean, I could count on both hands when I first started my career over 30 years ago, on who was doing research and I can name which ones, one of them just got out of it for various reasons. But thankfully, IBM has hung in there and delivered results because listen, if there aren't results, then there's no funding and that's life.

Dario Gil: Right.

Patrick Moorhead: But I'm really glad to see that you have because you just keep cranking them out there. But anyways, Dario, thank you so much for coming on the show.

Dario Gil: Thank you for inviting me to this event and Patrick, the feeling is mutual. I always learn a great deal from you as well, right in the conversation. So I look forward to future ones. Thank you.

Patrick Moorhead: I appreciate that. So this is Patrick Moorhead with Moor Insights and Strategy. Hope you're enjoying your day four at the Six Five Summit. Take care.