Research @ Citi Podcast, Episode 34: Follow the AI Chips to the Data Centers

Recorded: May 14, 2025

Published: May 21, 2025

Host: Anne Malone, Head of North America Equity Research, Citi

Guest: Michael Rollins, Communication Services and Infrastructure Analyst, Citi

Transcript:

Michael Rollins (00:00)

So, we developed a methodology, and I would phrase it as the chips have to go somewhere. The risk here is we overbuild. And at some point, there's going to be a digestion phase if you overbuild. And that happened during the cloud phase.

Anne Malone (00:16)

Hello, and welcome to Research @ Citi Podcast. Today, I'll be speaking to Mike Rollins, our Communication Services and Infrastructure analyst. I am Anne Malone, Head of North America Equity Research here at Citi. Today, the topic is data centers. Mike, thanks so much for joining me. I appreciate the time.

With everything else going on in the markets, it seems like there's doubt creeping into the data centers. So, what's all of that about?

Michael Rollins (00:41)

Yeah, so I think it started beginning of the year. So, we come into this year, and all of a sudden, there's the emergence of these what we'll call lower-cost Large Language Models. There was some pullback in near-term demand from one of the hyperscalers. And then there's also been some supply limitations in terms of the quantum of capacity that could come online and the speed over which it can come on.

And so, all of those things, I think, have raised questions about what demand is going to look like. And, you put on top of that then, you know, the macro uncertainty, the tariff uncertainty, and all of that together has really created some questions about how fast data-center demand is going to expand.

And when we looked at this, and we just recently published a report with our global Super Sectors, we remain positive on the pace of demand through 2030 and certainly can get more into that. But I think those were the factors that got us to this question or bigger question of what things are going to look like.

Anne Malone (01:52)

So certainly, we're living in different times than usual, but isn't there an old adage, right, trees don't grow to the sky? Is anyone really surprised that something new and innovative didn't go in a straight line? Is there something different we should look at or is this just normal fits and starts?

Michael Rollins (02:09)

You know, the hard part in this situation is trying to quantify demand. So we hear all of these ideas about how companies want to use AI, whether it's to drive revenue or to drive efficiency. And the challenge, though, is trying to figure out what all of that means in terms of the chips and the data storage and the networking, which all goes into a data center and drives the amount of power or what we would call IT load needed to fuel all of those services. And so that's what's made this situation really complicated.

Anne Malone (02:47)

Okay, so more variables, perhaps than usual and tougher to predict. But you say till 2030, which shockingly, is sooner probably than we all think, you feel good until 2030. What makes you constructive, and then if we could talk a little about what is consensus in the market thinking?

Michael Rollins (03:05)

Yeah, so it's a great question, and it's been really hard because the data-center industry is so fragmented. All of these use cases and customers are so fragmented. So, we developed a methodology, and I would phrase it as “the chips have to go somewhere.” So what we've been able to learn from our semis and global super-sector colleagues in technology and communications is how they're viewing the demand for these chips, for AI and high-performance compute.

And what we've done is we've taken those forecasts and we've thought about market shares for one specific large player in the market – and from there, we've estimated the amount of total chips and the associated power needed for those chips to drive AI and high-performance compute workloads.

And from that, we've been able to think about, OK, well, how much data-center capacity are we going to need on a global basis? We've done our forecast in global ex-China, and that's what gives us a lot of conviction in the multi-year opportunity to grow. And on average, through – we'll call it 2030 – it's over 10 gigawatts. We do have a little bit lower expectation for 2025 at about 9.5 gigawatts because we had a really strong 2024, so there could be some digestion. And we also wanted to contemplate some macro and tariff uncertainty within our forecasting.

But if you look at all of that, that's what really gives us the conviction, and the really interesting part of this is we've segmented the workloads into two pieces – what's been core and cloud, what's been growing for years, and what specifically is needed for the AI and high-performance compute. And through 2030, to get to our forecasting of demand, you need very little core and cloud growth. Most of it's going to come from AI, and that's how we've triangulated around this.

Anne Malone (05:06)

That phrase, though, “chips have to go somewhere.” We have to think about putting that on a t-shirt or something, but the chips have to go somewhere, but that implies you've done all of it based on the supply, and is your thought on demand is that this AI is so life-changing and you and I are going to come up with new ways to use it, and so is everyone else, that demand will find that home.

Is that the right way of looking at it or am I wrong?

Michael Rollins (05:32)

Yeah, it's one way of looking at it. And what we've triangulated that against, because you're exactly right, it's supply side. And these supply side numbers can change, and they have been evolving, and we can talk more about what we're seeing out of that in a moment.

But what we have also triangulated this against is the hyperscalers, their commentary on capex, their commentary on demand. And each of these hyperscalers have a bit of a different story in how they're trying to use AI and what they're doing over the next couple of years to invest, create the capacity and the opportunity for companies to then tap into these resources and go after whether it's the incremental revenue or the cost efficiencies that we were talking about a little bit earlier. And so that's one way that we've tried to look at the demand side of the equation.

You know, we're also tracking this against some of our coverage and also some third-party sourcing for data on demand as well.

Anne Malone (06:33)

And is something like this, since it's so different, is it a straight line between ’25 and ‘30? Do you see any if you're going to talk about digestion, I'm going to talk about indigestion and, you know, along the way. Anything you could predict or things like that always have to be surprises?

Michael Rollins (06:49)

Yeah, that's the hardest part because there can always be surprises with this. However, if you look at it from a growth-rate perspective, the CAGR from ’25 to ‘30 that we're looking for is about 14%. And it's not that different than years ago when we tried to contemplate what the cloud growth was going to do for data-center demand in terms of that double-digit type of growth opportunity.

And, you know, in terms of a digestion phase, so that's the risk. Like the risk here is we overbuild. And at some point, there's going to be a digestion phase if you overbuild. And that happened during the cloud phase, where things got ahead of itself.

Based on the way that the conversations we've had, the visibility that companies have had, we think that the risk of a digestion phase would be in the 2028 to 2029 time frame, at least from a data-center IT load perspective.

One of those things that gives us comfort in calling that near- to medium term as to why that risk of digestion is still a few years out is because most of the development that's happening right now, most of the construction is highly pre-leased. So, there's a lot of leases in place for what's built. A lot of new capacity – like, if we just wanted to lease a data center for ourselves, you know, do our own AI thing.

Anne Malone (08:07)

Let's talk after this about that, OK?

Michael Rollins (08:11)

Yeah, it could be interesting. That's right. You know, it would take at least till the end of ‘26 or ‘27 to start getting that new capacity. It's taking time to construct. The power availability is limited in a number of markets, and it's more from what we hear transmission versus supply.

It's, you know, plugging the data center into the grid with the needed amount of power. And it's also in some markets municipal or government related, where just some governments are questioning the pace at which this capacity is growing in their markets, and some of them are putting controls on it. So, I think it's all of these things that have slowed what could have been maybe even a faster pace of supply expansion in the category. And so that's helped keep this demand supply, I would call an equilibrium in the sense that it feels like demand is still greater than supply, but in terms of development, versus what's filling it, it's been moving close to in step for at least the last year from what we've seen from the data.

Anne Malone (09:15)

Not to get too far off-track, but you made me think of a question – geography then. You just talked about some municipalities, etc. So what part does that play when we're thinking about data centers, where it should be, where it's being built, what those think about it coming into their backyard?

Michael Rollins (09:31)

So, from a location perspective, we've seen this kind of like a pendulum just swing back and forth. We saw it during the cloud phase. We're seeing it now. So, you typically see data-center deployments where there's a lot of fiber, where there's people and GDP. And when you get into these big hyperscale builds that are like this ginormous warehouses of data-center capacity – when these things first start popping up, they tend to go into, whether it's the middle of the country or certain markets – that may be not too far from these population GDP centers, but where there's more land, there's more power, there's maybe even more green and renewable power, things tend to be cheaper because if you're going to do a lot of processing there, but you're not doing a lot of networking, back and forth of the data, that's a great place to have that kind of data warehouse to efficiently manage and process large datasets.

And that's been embraced during this early AI phase, where there's been big builds in markets that just have those characteristics. But then we also see over time as consumption for those services grow, then you start bringing the workloads closer again to those GDP and population centers.

So, the biggest market in the U.S., and I think it may still be the biggest – one of the biggest, we'll call it, in the world, is northern Virginia. And we still see healthy demand in northern Virginia. There's each continent, we'll call it has several key

markets where all of these things kind of come together and have been pretty robust from data-center growth. But the number of markets as things get regionalized seems to be expanding in the U.S., and it's something that can also happen in EMEA and Asia over time as well, where you just see a little bit more regionalization of those deployments.

Anne Malone (11:27)

A few times now you've referenced back to the cloud era, which is hard to believe that's historical now, but that there was a time of indigestion. Any other lessons learned from that?

Michael Rollins (11:39)

You know, I think that's where the hardest part is predicting how the use cases will then translate into the amount of power and IT load that we're going to need. And that's the one thing, you've asked about this in terms of the risk of a digestion phase. I mean, that is the one thing that's really hard.

I feel like in this phase, though, AI is something that's potentially measurable, where you can look at what you're going after, whether it's in terms of opportunity and cost save and measure that against what you're investing in AI as a platform. And I feel like that's something from a business standpoint, when they're looking at budgeting and they're looking at how to move forward with projects, that ability to measure investment versus return feels like an important aspect of this, emerging technology.

Anne Malone (12:32)

No conversation can ever be had on Wall Street right now without directly talking about the impact of tariffs. So how about tariffs on data centers? Is there a first order impact? Is there a secondary? Is it all combined? You know, what do you think there?

Michael Rollins (12:48)

Yeah. So for the data-center companies that we cover themselves, it's lower-risk because a lot of what they're building is like ginormous warehouse space, refrigerating it with commercial-grade air conditioning and chillers, backup generators, and the associated gear to keep those up and running 24/7. And so a lot of this is construction, materials, people. So yes, there is some risk in tariffs in the sense that if some of those inputs have inflationary factors, that is something that we need to keep watching over, and we've looked at some sizing of that.

But when you look at it in an aggregate, it feels low-risk in general, whereas there's that second-order risk from a data-center perspective of the customers. So, the chips and the semis and the investments that the customers are making, that's where there could be, based on what our global colleagues have published, greater sensitivity on the tariff side. And so that is something that we need to keep a watch over because that could affect demand. And that's why when we came out with our

recent update to our global forecast with our global team, we put some macro and tariff uncertainty into the numbers.

And one of the things that's interesting is just today actually, we had a chance to review the first-quarter industry data that we track. And what's interesting is we have a lower forecast for ‘25 of 9.5 gigawatts of global IT demand. The first quarter we're seeing an estimate of 1.55 gigawatts of absorption that took place in the first quarter. And if you take it comparable to what our numbers are because there's some differences in markets, it's about 1.43 gigawatts. That was ahead of our quarterly forecast of 1.05.

So, we're seeing at least from the first-quarter data, it's running ahead of our first-quarter expectation. And you might then think, well, wait a minute, 1.05 you had for the quarter, 9.5 for the year. What's going on there?

Anne Malone (14:55)

The math doesn't work, Mike.

Michael Rollins (14:57)

Yeah, what's going on with that? That's more than four times. And we did anticipate a back-half-loaded year, you know, after the party that we had, we'll call it in 2024, with a much stronger leasing absorption outcome. And then we felt like it would be back-half-loaded partly as just we had a stronger exit in ‘24, but also the factors that we were talking about earlier, just including the supply questions and time to get stuff up and running.

We felt we just needed to get to the back half of the year for things really to pick up, but at least the first-quarter number was better than what we had expected. So that was a help.

But back to your larger question on the tariffs and the macro is something that we've thought about. We have to continue to track it. It's clearly dynamic. So we're watching the news every day to see what comes next. But that's where the sensitivity could be. It could be first felt on the customer side, but then we have to watch that downstream implication for the data centers.

Anne Malone (15:55)

This is obviously going to continue to be a hot topic. We always have earnings season, we always have what the hyperscalers have to say. We, of course, have the Citi TMT conference at the Hilton September 3rd to the 5th in Manhattan, so we'll get a lot more then.

Thanks so much for your time, Mike. Very interesting, very insightful. Appreciate it.

Michael Rollins (16:15)

Thanks, Anne.

Anne Malone (16:16)

Thank you for joining today. This podcast was recorded on May 14th, 2025. Our next Research @ Citi podcast will feature Beata Manthey, our European Head of Equity Strategy for a comprehensive overview on global markets. Thanks so much.

[Disclaimer] (16:31)