



The Deloitte On Cloud Podcast

David Linthicum, Managing Director, Chief Cloud Strategy Officer, Deloitte Consulting LLP

Title: AI at the Edge holds great promise, but it'll take work to get there

Description: Edge computing isn't a new concept, but pairing it with artificial intelligence holds new promise. However, there are significant challenges that companies must meet to realize the promise of Edge AI. In this episode, David Linthicum talks with ClearBlade's Aaron Allsbrook about how Edge computing, paired with AI, works, and how it can give companies that leverage it a competitive advantage. He also lays out some challenges they'll face to reach their Edge goals.

Duration: 00:25:20

David Linthicum:

Welcome back to the On Cloud Podcast. Today on the show I am joined by Aaron Allsbrook. Aaron is founder and CTO of ClearBlade, is the architect of the industry's first and leading IoT Edge AI platform. Aaron, we're excited to have you on the show today. How are you doing?

Aaron Allsbrook:

Very well. It's a pleasure to be here, Dave. Thank you very much for having me.

David Linthicum:

So, give me the Aaron Allsbrook story. So, how did you get to working at this company and founding this company and what's your background and what kind of gets you up in the morning?

Aaron Allsbrook:

Yeah, so I guess I started as an EE, electrical engineer, and then basically left college and immediately started building enterprise applications so at IBM and building enterprise applications and the tools around them. And that I guess took me down an interesting path of understanding how things get architected, and the whole world's changed—obviously you know well—with the cloud coming around versus those old big server stacks that we had. And I left that and joined with my cofounder, Eric Simone, and we founded ClearBlade.

ClearBlade has been focused really on IoT at the beginning, but as we learned with many conversations, many conversations with folks there was this missing spot between the cool stuff they were doing that I could hold in my hand and measure and this cool stuff that we saw in clouds and scaling, and that was the Edge. And, so, we really launched headlong into Edge in the 2016 timeframe, one of the first products, at least one of the first ones I can find, in the market to be building these Edge applications. And that's just been an exploding area for us and an area where I spend most of my time today. How can we get this data and process it and do as much as we can before we roll it all back to our big clouds? So, that's a nutshell of how I got started in this Edge AI and Edge space.

David Linthicum:

So, tell us a bit about your technology. How does it assist developers in building these solutions?

Aaron Allsbrook:

Yeah. I mean, really, I come from this Dev tools background, and how do you build applications faster? And really when we started this core stuff for IoT and Edge, we asked, "What are the elements that you need? And a lot of folks are doing IoT platform. I actually just spent this week with a couple folks that were with us way back when with interoperability with MQTT, and we were going over the history of how our companies have all kind of grown over time. But we have these core development tools that basically abstract away the challenge of how to get data in the right spot, how to get processing done and deploy to the right locations as quickly as I can, how to put a pattern in place that's easy to use and start with, but yet streams to hundreds and hundreds of thousands of devices and I don't have to worry about it. That's what we're providing.

So, we're providing that abstraction to this stuff. We're not forcing people to go through with heavy, heavy DevOps to figure it out. And we'll run it anywhere. So, our customers are all really focused on enterprise, so when I say that server side, that means any cloud, or it means running on my customer's servers or their OpenShift or Kubernetes technologies. And when I say Edge, I pretty much mean we'll put our stack down and run it on anywhere where there's an operating system. So, we don't quite go all the way to the embedded side when I said Edge AI—it's certainly a big part of that—but we'll run on very, very small, single-core, old ARM processors up to obviously some of the very kind of large box stuff we see getting rolled out at the bases of cellphone towers and stuff.

So, really about providing that layer, and then really trying to serve the users of that infrastructure without having to make them understand all that technology. A lot of what we're focused on now is operationalizing it and serving that end user, not necessarily the developers who have a lot of knowledge and can extend it and take it where they want it to go. But really the end user, the mechanic, the maintainer—those folks are who we're after helping now.

David Linthicum:

So, what is the typical Edge platform you're dealing with? What's the physical hardware, software that runs on it and what purpose does it have? I know it's all over the place, but what are you seeing most often?

Aaron Allsbrook:

Yeah. I mean, so most often we're looking at industrial gateways and I guess we have partners—Phoenix Contact is a great partner, MultiTech, Eurotech. They make these boxes that have different levels of processing power and different price points, designed to—some live on the North Pole and some are designed to be nice and carpeted retail stores, but they all have different values. And, so, that's really what Edge means to me, because that's most typically where suddenly you're pulling off a SCADA protocol or you're pulling in a wireless BOE or LoRaWAN device. Everything after that looks much more like our IT in the cloud.

You've got IP addresses and stuff. But when you're working with how you ingress this mess of IoT and nimDim and wireless protocols, that's the Edge. And that's where we really focus our solution of how we bring all that in and then use what we can down there. before we pay the cost of shipping it back, what can we do and what can we process before we leave that box?

David Linthicum:

So, moving forward ultimately the enterprise needs some sort of a business case for doing this. And so, how do most enterprises weaponize this technology? I'm talking about specifically not just Edge, but Edge-based AI technology. And what are some of the major use cases you see out there?

Aaron Allsbrook:

Yeah. So, one of the things that I've learned—and again, I came from IT and now I kind of live in this operational technology world. So, when you say weaponize, I think operationalize. How do you put this stuff to work? If people are going to pay for this, right—and you look at it. There are both CapEx and OpEx costs. Somebody's got to buy some hardware to get this stuff started, and someone's got to keep it running and pay for it over time. There's really got to be a reason. And we look at—we call them capabilities; I'm sure bigger analysts have better names. But the capabilities—are we going to predict the machine's going to fail? Are we going to understand its utilization so that we can own less or own the right amount of these machines? Are we going to automate when we need to predict maintenance or refilling. There's just a lot of tasks that go into the fieldwork.

So, when you say weaponize—how do we do a better job at that and bring those margins down? That's always one. Another thing that we always see folks are interested in is how you can improve safety. So, first we improve the operational processes and make that work better. The other thing is we can make

a safer environment. That's always got a lot of ROI associated with it. And, so, in that case, you're looking at do we have people in the right spots? Can we be sure that people are using the equipment correctly, the right number of hours? Those sorts of use cases are really valuable. Is the equipment activating like it's supposed to for its own safety devices? So, I see a lot of value there. So, when you weaponize or operationalize, can you get after safety?

And then the big one is—and it goes broad—is, “Can we cause there to be more revenue generated from what we've got?” And then we see that everywhere. maybe we make a better product. We make a product that people want to buy over other products. That's an easy one by some special features. You see people drive new revenue because now they realize that the product they have consumes energy, but if they consumed less energy then they'll be able to reuse it for other cases. They'll be able to save that consumption cost. So, we see that a lot, and we resell that consumption cost.

Then we see things like, “Can you change how the functions are going to drive either more customers in your stores and engagement?” And we definitely see things like, if we can know where our customers are and engage them the same way Amazon or Shopify engages us as we're browsing the website and watching our mouse, if we can watch their interest in products and stuff and engage them at the right times and with the right ancillary also-likes, there's new revenue to be seen.

So, I really come to those three when I think of how do you weaponize this. You're taking the new data you've got, you're understanding it in basic ways, the business domain. This is stuff that I spend most of my time learning, is how do these businesses work. And then you're saying, “All right, what is it? Is it operational efficiencies? Is it safety, the environmental that we can do? Or is it driving new revenue?” And, so, that's what operational and weaponizing means to me.

David Linthicum:

Yeah, and maybe I should watch the military analogs. I tend to do that more than I should. It used to be sports analogs and now it's military analogs. So, this is kind of exciting to me, because number one, you kind of hit the nail on the head. If we're going to serve the customer better, we're going to become an innovative differentiator and we're going to leverage AI—we'll call it a force multiplier instead of a weapon this time. But the ability to kind of push that out closer to where the information is going to be processed seems to be key to me.

And, so, it never made sense—and I'm a cloud guy, a cloud geek, been in the industry for the last 20 years. We had to transmit everything centrally and we store it and we process it and we also make decisions on it, and we may bind an AI engine to it to train the knowledge base and then figure out how to make certain decisions and then learn as we go and things like that. But that round trip—it becomes kind of an issue, because number one, you have to assume someone's going to have good network connectivity, and you can't always assume that –

Aaron Allsbrook:

You got it.

David Linthicum:

But, also, you just don't have this instantaneous response time that I think customers are looking for. You've got to remember you're trying to delight the customer, and it would delight me—not necessarily doing away with a centralized gathering of information because I think we still need that, and even some centralized knowledge process, but the ability to make more tactical knowledge base decisions down in the Edge-based system, whether it's a mole computer, whether it's in my car, whether it's in my tractor, or whether it's a business application or personal application. We just kind of get out of this stuff we have to centralize everything. What are your thoughts on that?

Aaron Allsbrook:

Yeah. I mean, you make a great point. So, we've lived in a world where everyone's centralized. I was actually just at a conference this week that's an oil and gas kind of networking conference, and I'm meeting all these people that have been doing—moving data to their back offices forever in all kinds of ways and satellite. And if anyone knows how expensive it is, it's someone who's got an oil and gas platform trying to move that data off of it. So, I see it's incredibly important to do this at the Edge, especially now that we understand it. I mean, ten years ago we were still watching AI on *Jeopardy!* trying to figure out if that's what AI was. Now we're understanding it. Now we can know that if we have these five features, we can predict a leak. We can predict a need. We can predict a broken part because these are the parts we've seen break. So, we understand that AI is more mathematical.

I think the real challenge that I've seen with Edge AI is—and again, I call myself an applications guy. And, so, everything to me is solved by an application. I've met data guys; everything's solve by a better view or query against the database. Data scientists are always showing me their tools. And data scientists do great stuff, but they tend to write them in languages that work well for cloud and infinite compute and infinite time to wait. And Edge AI is, like, the opposite of that. Now we have extremely limited compute and, really, we're trying to solve a problem we don't want to wait for.

And, so, I think that's the big leap that's happening right now as we've gotten all these cool tools, we've gotten data scientists using all kinds of Python libraries and TensorFlow libraries and they've got gigantic TensorFlow containers that have every tool possible in them. And then we're saying, “Okay, take what you did but put it on a tractor,” right? And now suddenly it's like is it all for naught? How do we do that? And I think that's where Edge AI—and that's where we're playing, is, how can we take that knowledge and move it and run it? We like ONNX and—but there's other runtimes for these things. Can we move it and run it on something quite constrained and make an inference within the two seconds necessary to say and alert people to a safety issue?

That's the big part of this. I mean, AI—lots of stuff being solved, very cool. The only way to use it in that cloud world is to shift every bit of data up there, and you may need that data for a while to train. But at some point, you should have an understanding of that problem that you can move it back down. And that's Edge AI to me, efficiently running all that stuff we've been uncovering and digging out of the ground for the last ten years with regard to AI space.

David Linthicum:

So, why is this not more pervasive? We've kind of had the technology around for a long time and had broadband networks out for a long period of time. We seem to be spending all our time centralizing these resources when there hasn't been a lot of balancing to the layers, in other words the ability to look for other ways to run the code and to store the data and to run the knowledge processing that's more efficient than they did, and also cheaper than we've done it before. One of the things that I see is people are getting sticker shock when they get their cloud bills.

And suddenly these very nice utilities are able to leverage where we can store data and build applications and things like that. These things cost a lot to run. And your ability to repatriate—and that's probably a bad word, but your ability to push things out to the layers at the Edge, as long as you're able to manage those systems efficiently—and that's kind of a core if—are going to be much cheaper and much more efficient. So, do you think this is going to pick up speed in the next few years?

Aaron Allsbrook:

Definitely. And again, you never know where networks are going to go and—Low Earth Orbit networks and such, but even if you have the ability to send more data with the next G, it still costs you money. Even if you get it all into the cloud the clouds are doing a good job of giving people a little easier services, and every time there's a new service it's, like, 7X the cost if you use the base metal kind of thing underneath. So, they're making their margins on these cool services. And you're right. The sticker shock on these things where you thought you were paying a micro-fraction of a cent run for a whole year wide open, suddenly you're paying \$50 for a product that costs \$1,000 for a one-time purchase. So, a ten-year, you've lost all possible margin to try to run this thing for folks. So, the cloud bills are killing you. Whether you're trying to get that money back as a product or wherever that ROI is, it makes it much harder.

The other thing that's happening is as our eyes are opening up if you think about these systems before, they were complex process control systems managing things like submarines and power plants. That's what IoT was before we called it IoT. It was complex M2M sorts of things. Those systems were incredibly expensive. So, if you were going to try to compute at the Edge, you kind of knew you were walking into hundreds of millions of dollars. Now everybody's got a Raspberry Pi laying around and you're seeing the vision of if I toss this here in our business, we can begin to open up stuff.

So, the use cases are proliferating; we weren't expecting that. We're finding that we can get a lot more data off of these things. No one really expected our pumps to generate terabytes of data that we were going to pay for in the cloud. So, while a gig seemed cheap, one terabyte per pump is crazy. We're not running jet engines. So, I think you're right. We weren't really expecting—we wanted all the data because our data scientists were excited, but we weren't really expecting that operationalizing, that value and what it was going to cost us to get.

But now with these cheap processors, we do. We have a latent target to move that back to. And, so, I think that's important. This journey is, yes, you may bring it to the cloud, and you may pull lots of data up, but we do say folks saying, "Okay, enough. Let's turn off the spigot. We've learned what these things do in the field. We still want to know major events when things go long, but let's just push down this logic and let it all run down in the field." And I think that's a huge part. I'll say security's always important, you have to update for. But I think updating these applications to change that profile for how you turn down the cloud knob, how you turn up the AI knob, how you turn up the local processing knob—that's a big part of where these architectures and applications are going, and Edge AI in general.

David Linthicum:

So, are we giving anything up by doing things at the Edge? In other words, we have Edge-based AI, but is that almost as good or as good or 20 percent good than the stuff that exists in the cloud right now? Or is this pretty much going to move to parity pretty quickly?

Aaron Allsbrook:

I would say we'll struggle to get parity. Whenever you've got bigger computers with more data, I truly believe they're going to be able to come to stronger, higher accuracy conclusions. So, I do believe that there will always at least be somewhere in the lifecycle of getting your AI going, you're going to want to be centralizing and pulling something up to the cloud to work with and tools that your data folks love.

There's a lot that goes on with that data that's never been seen before. even businesses like railroads or connected products, there are people in there that have never actually seen what happens to their product after it's been in the field, after it's been through a Canadian winter. How is it behaving after that moment? And, so, that's the value prop. Bringing that in and then looking at those failures because anytime you're trying to predict failure it's use and it's events and it's the environmental that have had experience. So, if you're going to do that, you've got to bring it in at least once for those folks.

I believe there's a good enough bringing that back to the edge and letting it kind of retrain itself with just its data it knows about itself. I think that will happen—that'll be the lifecycle we see. Bring it up, learn something, push something down, and let it keep itself in line and calibrated with the environmental that it focuses on locally. But it's not going away. The cloud is not going away.

David Linthicum:

All right, pull out your crystal ball. Let's figure out where this stuff is going. So, Edge computing—obviously a lot of momentum right now, and certainly IoT is going to be a part of that. And, certainly, if you look at the growth of 5G, it's going to enable more Edge-based processing, more Edge-based technology like yours, and there's more adoption of Edge-based systems.

So, what are the challenges then in the future? And I'll put one out there first and let me get your opinion on this. I think the management and operations of these things is going to be a challenge. I wrote a blog in InfoWorld three or four weeks back on the fact that a lot of people are kind of running into walls where if they're getting to 1,000 Edge devices which are doing things like monitoring oil pumps in some remote area of South America, they're finding it troubling to keep up with the updates and make sure the data is resolved and solve hardware issues, things like that, and doing so at scale. Is that going to continue to be a problem? And, also, what other trends do you see moving on in this world?

Aaron Allsbrook:

[Laughter] I'll start with the first one. The one thing I think a lot of us, especially folks like me that came from enterprise apps—we thought the enterprise was slow in how they did things, because they've been managing thousands of computers and stuff. And, so, there was some slow inertia in that. When you go to the Edge, you now not only have just that activity whether you use enterprise tools for managing that stuff or not. That task is still going on, has to be done, has to be secure.

But you've got somebody that has to go out there and bolt this stuff on. I mean, like, that's a whole other world that—most of us just got our computer from IT when we started. In this world, somebody has to think about the right way to attach it, the right frame, the right screws, the right bolts, the wiring

diagram. So, even if we have the exact right software and hardware technology, there's someone going out there with a wiring diagram putting it all in place. It has to have the exact right—my green dot said everything worked; my red light said it didn't.

And if the red light comes up, I try one more time and I bring it back. There is a ton of process that I think a ton of folks in the Edge space never even expected. A lot of the Edge talk is about how do you manage the containers and applications that might be running out there.

So, I think you're right. I see one of the biggest challenges—supply chain aside—is just getting the right rollout processes. And I love folks like the industrial automation companies and stuff because that's the challenge they know. They've been connecting up SCADA systems and doing it well for many, many years. And that technology, that breadth now needs to be brought to Slushy machines and freezers and all kinds of stuff. So, I think that's a big one.

I think a lot of folks are maybe taking too many lessons from our cloud world. Quite frankly, I think DevOps has exploded because of how complex we let our cloud implementations get. I'm not sure Edge needs to be as complex as that. I see a lot of folks that are trying to build solutions that are going to run on a Raspberry Pi. I don't quite understand why we're taking so many lessons learned from high scale, high availability, and trying to put them in places where we know that if lightning strikes, it's going to be gone anyway.

So, I think that's a challenge, is finding the right amount of architecture that fits these use cases and not just bringing—because it was a good practice here, taking it everywhere else. I see that as a running challenge. That's something that our communities are going to have to address together, is to decide why are we doing these things? Why are we pushing down large-scale high-availability tools to incredibly constrained little devices that we need to spend much more time doing AI versus running middleware kind of behaviors.

Those are just two, David. What else?

David Linthicum:

No, you see anything else coming in the future that you're excited about?

Aaron Allsbrook:

I'm really excited—AI has been—we haven't talked a ton about AI, but AI is a huge part of where we're going. We've seen some amazing successes. I mean, I've been a big doubter about what can AI really do in some of these places, and we've seen accuracies that predict 72 leaks on—or flying 72 leaks on 70 alerts. I mean, that's just crazy to me, that it can be so successful. But getting to the Edge—for us, we've been playing with the ONNX AI runtime. It's letting us move so much stuff down that I thought we would never be able to get unlocked from the data scientists' brains and put them down and predict things right in the field.

So, if you had to ask me, I'm really excited about what we're able to do in that space, in that open community right now. We're able to put that on our partners like Phoenix Contact and MultiTech and others and really turn over some interesting things, and even to our partners that are in—heavy in verticals already. So, that's probably a big part of what I'm most excited about in the Edge AI space, is that community that's going to let us reshare a single run time across all of this AI.

David Linthicum:

Okay. Where can our listeners find you and your company on the web?

Aaron Allsbrook:

Sure. So, you can find—I'm at ClearBlade. You can find us at ClearBlade.com. There's many ways that we'll try to get to engage you once you're on our website. Definitely you can find me on LinkedIn, Aaron Allsbrook, or you can find me on Twitter. @AAllsbrook is my Twitter handle as well and I'd love to hear from your listeners.

David Linthicum:

Yeah, and I'd love to learn more about this technology moving forward. This has been an area of research of mine for the last four years and I think it's really exciting as to where are the technologies going and also what key enabling good things we can do with this technology, whether driving safer cars or living healthier lives or having remote diagnostics of various health issues, things like that. This is really exciting, where it's going.

So, if you enjoyed this podcast, make sure to like us, rate us, and subscribe. You can also check out our past episodes including those hosted by my good friend, Mike Kavis. Find out more at DeloitteCloudPodcast.com, all one word. If you'd like to contact me directly, you can e-mail me at DLinthicum@Deloitte.com. That's L-I-N-T-H-I-C-U-M. So, until next time, best of luck with your cloud journey. You guys stay safe. Cheers, 'bye.

Operator:

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to Deloitte.com/about.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Visit the On Cloud library
www.deloitte.com/us/cloud-podcast

About Deloitte

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.
Please see www.deloitte.com/about to learn more about our global network of member firms. Copyright © 2022 Deloitte Development LLC. All rights reserved.