Deloitte.

The Deloitte On Cloud Podcast

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

Title: **Cloud-native principles power the edge's need for speed**

Description: Cloud-native design principles make cloud apps scalable, flexible, and more quickly deployable. Those principles can be applied to situations other than cloud, as well. Take edge computing, for instance. Edge can be thought of as a point-of-presence cloud, so applying cloud-native principles to the edge can make app performance, speed, and scalability cloud-like. And it can help solve the "edge war" of the future: the battle to reduce latency to its vanishing point.

Duration: **00:26:03**

David Linthicum:
Welcome back to the On Cloud Podcast. Today on the show I am joined by Jason Bloomberg, president of Intellyx. Jason, welcome to the show, and man, we've known each other for a long time, right?

Jason Bloomberg:
Yeah, thanks. Yeah, ever since you invented platform-as-a-service before anybody have heard of it.

David Linthicum:
Yeah, I did. I did. I also invented SOA and cloud computing, and I think I invented the internet, so.

Jason Bloomberg:
You know I remember Grand Central. That was PaaS before it was PaaS and you were key to that, so.

David Linthicum:
Yeah, yeah, that was a lot of whitepapers that just went nowhere. And also I learned the fact that you can be early in the process and still lose out. So, you can have the best idea before everybody else but not necessarily win the game, because the idea has to come out at the time that there's a need.

Jason Bloomberg:
Exactly.

David Linthicum:
I've discovered that a few times in my career. Sometimes I hit it, most of the time not so much. So, catch us up. You've been on the podcast before. What have you been doing since then?

Jason Bloomberg:
Well, Intellyx—we're an industry analyst firm, a boutique firm focused on broad topics of enterprise digital transformation, which really means that we talk about many different disruptive trends in enterprise IT and really try to help tie them together and help our enterprise audience understand how they all connect together and support business goals. And hopefully we'll be doing that on this podcast.

David Linthicum:
Yeah. So, what's typically—is this you predicting things? Is it you providing feedback for people who use your analyst service, writing reports, all of the above?

Jason Bloomberg:
Well, we work with the vendors and our audience are primarily enterprise practitioners. So, we help vendors tell their stories in a way that the audience can understand what the value proposition is and whether or not they should actually buy something from a particular vendor. So, vendors often have a hard time explaining not only what they do, but why anybody should care, and that's where we come in.

David Linthicum:
So, what attracted you to this field? I think you have a background in mathematics, if I remember correctly, and now you got into the—actually you're definitely one of the better analysts out there. You're always on the top ten of who's who and being listened to and influencers and having the best content out there, and you've kept that up for many years. I mean, I've known you for at least 20 years, and I think you've always been on those lists and always been in the top ten of people who are listened to out there. So, what was the Jason Bloomberg journey that got you from where you were to where you are now, where you're just like a listened-to expert in terms of technology and where technology's going?

Jason Bloomberg:
Yeah, well, I appreciate that. Thanks for the kind words. Well, in the '90s I was a web developer for a while and got into Java in the very early days of Java and tried my hand at coding. And I just found that writing was more enjoyable to me than coding. Coding is great fun. You have a challenge. There's a problem you have to solve. If you don't get it right, the code won't run. So, it's a real sort of puzzle-solving kind of activity, but just writing in English as opposed to in Java was just something that I found more enjoyable.

So, I made the decision toward the end of the millennium to shift from writing software to writing about software, and that really led to a brief stint at IDC, and then ZapThink. ZapThink, quite well, an analyst firm in the 2000s, focused on service-oriented architecture. And then later on at Intellyx—I founded that as well following the ZapThink model, right, a boutique analyst firm. But I didn't want to be typecast, right? The problem with ZapThink—we were doing great when SOA was the hot topic. But once the audience sort of moved on to other topics, then we were typecast as the SOA guys, and it sort of—we sort of struggled there for a while.

So, in Intellyx we'll talk about any emerging technology trend as long as there's confusion about it. We love confusion; it gives us something to write about. And if a topic sort of falls by the wayside and people aren't talking about it, we just move on to another one. So, and for a while there SD-WAN was the hot topic, and everybody was talking about it, and now nobody talks about it, so neither do we. We'll move on to something else.

David Linthicum:
So, speaking of hot topics, some of the ones I'm commenting on, and we talk a lot about on the podcast, would be cloud native, what it is and how to use it and how to leverage it effectively on the enterprise, and also edge computing. And recently you worked on some great content where you talked about the cloud-native edge, and ultimately that is kind of a new market trend moving forward and a new technology pattern. Really tell us what it is, and why is it interesting?

Jason Bloomberg:
Yeah, cloud-native computing—there are sort of different ways of looking at the topic. The starting point is Kubernetes, right, the most popular container orchestration platform out there, opensource. And many large enterprises, really most of them at this point, are revamping their IT infrastructure leveraging Kubernetes and following cloud-native best practices.

But from our perspective, cloud native is more than just Kubernetes. It's really a new paradigm for enterprise IT where we're saying we want to take the best practices of the cloud like scalability, elasticity, and dynamic software development, and really extend those to all of IT. So, it's no longer just about the cloud itself, but it's saying, well, we can do all this good stuff across all of enterprise IT, in public cloud, private cloud, on premises, or whatever

conversation, leveraging these best practices. Kubernetes is a part of the story, but it isn't necessarily about containers. It could be traditional virtual machines.

It could be serverless functions. So, whatever is the right tool for the job, and we're going to build these comprehensive abstractions, control planes that enable us to run software and deploy software rapidly and taking advantage of these cloud-native best practices. So, it's much broader and just the cloud and it's much broader than just Kubernetes.

David Linthicum:
So, what are the examples of the use cases out there that you see for this technology?

Jason Bloomberg:
Well, for cloud-native computing, the key benefits are rapid scalability and elasticity and ability to deploy software very quickly. So, not every situation qualifies, right? There are plenty of instances of enterprise software that you don't deploy very quickly, that you don't require massive elasticity, right, scale up, scale down, right? The core transaction processing in banks is stable; you don't change the software very often. There's no need to change things there.

But increasingly what enterprises are finding as they revamp their organizations to be more digital, so to leverage technology to provide better value for their customers, they're finding that the rapid scalability, elasticity, and rapid deployment of software are becoming increasingly important, and as such there's this massive set of business drivers for cloud-native computing. And, so, we're finding that across traditional enterprises as well as web-scale companies, born-in-the-cloud companies who are all over this stuff.

David Linthicum:
So, moving forward—and I like the way you discuss this because you talk about this isn't applicable everybody. You need to have a requirement, you need to understand the value that it is, and there's going to be a cost and risk in moving this technology, like any technology. So, how should we look at the cost and risk as relative to what the value's going to be? Because I think businesses don't understand, ultimately, what value this stuff can bring. They can't do the value assessment, those sorts of things. What are some of the macro-understandings that you have around value of cloud-native technology and what are some of the key things that enterprises should think about?

Jason Bloomberg:
Well, perhaps the Achilles's heel of Kubernetes and the Kubernetes ecosystem in particular in cloud-native computing broadly is that it can be enormously complex, right? Kubernetes is a very complex set of technologies and it's an ecosystem of opensource projects, and as such the various projects are at different states of maturity. They all have to work together and sometimes they do; sometimes that working together bit is a work in progress.

And then you have this broad community of commercial vendors who are leveraging opensource and building solutions on top of this cloud-native opensource ecosystem. But as a result, there's a lot of moving parts and a lot of complexity. So, that's probably the greatest risk, is that you really need to have sufficient motivation in terms of the benefits I discussed in order to have the wherewithal to deal with the complexity.

This also means that a lot of vendors are in this space trying to build solutions that reduce the complexity, that abstract out the complexity. And cloud-native computing is all about building these comprehensive control abstractions that, if you get them right, reduce the complexity for the users that are interacting with them. But that's a big if, right? Getting them right is a challenge.

David Linthicum:
You said something really profound that I see is kind of the key inhibitor for folks leveraging this technology, and that is dealing with the complexity and certainly the operational complexity that this brings. And we do have the ability and suddenly we're going from say 1,000 cloud services under management, some of them may be microservices on Kubernetes and whatever, to 3,000 services, and we have a particular application we're deploying on Kubernetes, and it's complex unto itself with lots of different moving parts and lots of things that we can operate, and we have to operate it effectively for it to finally bring value back to the business.

And, so, you're right. The complexity challenge seems to be where we're breaking our pick, and the vendors are stepping up and trying to remove the complexity through abstraction, automation, things like that. So, where do we stand with that and what are they working on right now macro-wise, not naming a particular vendor?

Jason Bloomberg:
Well, there's an old joke that there are only three numbers that computer programmers deal with. There's zero, one, and N, where N is any number larger than one, right? And that's it, right? There's no other numbers because you would never code 47 into your program, right? You'd never see numbers like that in a computer program, right? There's no point in doing that because maybe it's going to be a different number; maybe it's a variable.

Well, the same is true for Kubernetes clusters, right? And you either have no clusters at all where you're planning your deployment, and so you're still at zero, or you have one cluster so you're getting up to speed. But then after that, nobody ever goes to, like, a certain number of clusters. It's N clusters. It's having many clusters at that point. And, so, this is obviously a key part of the scalability benefit as well as the challenge of Kubernetes and cloud-native computing generally, is that it really is intended to scale up massively. Why would any company need dozens or hundreds or thousands of clusters? Well, the question is really not just why would they need it, but what could they do with it once it becomes cost-effective to deploy? And that opens up areas of innovation, and that ties to the edge computing story that I'm sure we'll get to in a little bit here.

David Linthicum:
Yeah, so let's talk about edge computing. We're seeing a rise of edge computing as a concept and different flavors of edge computing. We certainly have the micro-clouds, what they call edge-based clouds, with Azure Stack, Amazon Outpost and other equivalents that are in essence kind of working to the systems. We're also building edge-based systems that are remote-connected via 5G. I worked on designing lots of these things over the last five years. They seem to be increasing in momentum right now. So, where is that trend going and what should we be watching?

Jason Bloomberg:
Yeah, well, it's good that you brought up offerings like Outpost, right, where a big cloud provider is basically extending the cloud to the edge. And that really gets you scratching your head as to what we really mean by the word cloud, right? Traditionally we think of cloud as a big datacenter, right, with sufficient software to abstract the hardware so that the cloud provider can offer abstracted services to their customers. So, an instance type would be an abstraction of a virtual machine on a server. Virtual machines are themselves abstractions. So, that's where you—but it still all has to run on physical hardware in a datacenter.

Well, now we're saying on the edge, well, what if we rethink what we mean by a datacenter? And now it isn't necessarily just some massive building with huge power and cooling requirements. It could be a smaller building, right? If you're a telco, it could be that little building that you see at the base of a cellphone tower. Well, what's in there? Well, it's telephony equipment, but it could easily be a server rack and it could easily be this mini-cloud, or a point of presence.

And people aren't necessarily familiar with telco points of presence, but if you go through any town in this country or really anywhere in the world, you'll see a nondescript little building in your downtown area that probably has a small sign with your local telephony provider. Nobody goes in. Nobody goes out.

But what's in there? Well, it used to be switching equipment, and now it's all racks of servers and other parts of the telephony infrastructure, but it's also cloud infrastructure, right? So, what is the difference? Well, it's still a datacenter. It still has the same kind of power and cooling that it needs in order to run servers in racks. But it's at this smaller footprint, and it is also closer to its customers, right?

So, back when it was switch telephony, you needed to put these things close to customers because you'd actually have physical wires running from businesses and homes to this building. And then you had, in the good old days, the speed of your internet had to do with how close you were to this point of presence, so that was important. And now it's about lowering the latency, where we still have limitations of the speed of light. So, if we want very low latency, very high performance and responsiveness at our user interfaces say for real-time behaviors, we want to have the physical infrastructure close to us, which is one of the primary motivations for this part of the edge story.

But there's more to the edge than just these little mini-datacenters. Those mini-datacenters, they could be at retail establishments. They could be in phone closets, what we used to call a phone closet because we used to—every retailer or every office building had to have all this phone equipment in the closet somewhere. Well, now we don't need all that because we do our phones over the internet, so what do you need? We put in a little datacenter in there instead, a rack of services. And it's remotely managed and built to be part of this broader fabric of abstracted server capabilities that we think of as cloud computing. So, that's part of the—oh, sorry, go ahead.

David Linthicum:
Yeah, I love—yeah, I love this way of thinking because if you think about it, back in the day when I built systems for retail establishments when the internet wasn't around, everything was kind of built on the edge, but it was a disconnected edge. In other words, we had a PC that gathered data during the day, and at night it talked to some other centralized PC and uploaded the information. And hopefully it was uploading the right information. Hopefully it didn't fail. I remember backing a few of those out.

And the reason we were doing that was because—not only latency—we didn't have connectivity moving forward. And I don't think it makes sense for us to send every piece of data and every request that goes to some sort of a centralized server to be returned back to the system, traveling in some cases a round trip of 5,000 miles. It does make sense to have these different points of presence where we not only can reduce the latency but increase the reliability, and also manage them as if they're centrally located. What are your thoughts on that?

Jason Bloomberg:
Mm-hmm. Yeah, well, that's very important and also highlights the needs for different types of edges, right? The edge is not just a single thing. So, we like to talk about three different edges. You could—how you define them, you might have more than that. But we talk about the cloud edge, which is actually in the cloud. That's where CDNs, content delivery networks, live, right? So, a CDN will put servers close to the customer in order to reduce latency and to help distribute the load on B2C websites, right? And that's part of the edge. They think of themselves as being on the edge, but it is still in the cloud.

And then we have the near edge. That's what we were talking about before, the telco points of presence and equipment in phone closets and also internet of things gateways which would support internet of things, IoT capabilities in facilities that require them. So, it could be factories or smart cities or other things. So, you'd have this gateway which is essentially a piece of middleware running on appropriate hardware that you would also locate in this near edge.

Now the far edge then are the end-user endpoints. So, it could be smartphones. So, all of our phones now are in the far edge, but it could also be IoT centers and actuators, so all of those IoT devices that really make the things part of the internet of things, right, that make that a reality. So, all of that is at the far edge.

So, what we have now is we have this traditional cloud, centralized datacenter-based cloud. We have the cloud edge, we have the near edge, we have the far edge, and we are doing massive compute on all of them for different purposes depending upon what we're trying to accomplish. So, for example, if we have a situation where we're doing video surveillance, so you have some facility. It could be—I don't know—let's say an airport, and you have video surveillance.

You have all these cameras all over the place. Well, in the old days back in 2010 in ancient times, we would have to have human beings running—sitting in these rooms, watching dozens of screens, looking for suspicious behavior. Well, we've gone way beyond that and now we have AI doing it, right? We have AI that is sophisticated enough that it can detect suspicious behavior or whatever kind of behavior you're looking for in video feeds.

Well, that's massive quantities of data coming from the cameras to that AI inferencing engine. We don't want to put that in the traditional cloud; we want to move that close to the processing. So, we want to put that on an IoT gateway in the airport, and maybe have many of them in the airport because we have many cameras. And we just don't want to send all of the raw data any further than we need to, right, because it's expensive to move, expensive to

process. So, we process it locally and close to the cameras themselves. We use AI to drive our inferences, right, so the insights, what we need the suspicious behavior we're looking for.

That's the information now we send upstream to whoever needs it. That's far easier to do than sending all of the raw data. So, that's one example of where processing at the near edge in order to support high-volume data collection at the far edge is especially useful, and that's so that's sort of driving a lot of this innovation in this space.

David Linthicum:
So, kind of the elephant in the room around edge computing would be how we're going to manage and operate these things. Ultimately, we talked about our complexity. This thing is going to lead to complexity. We could have—I have a client that has a thousand devices out there monitoring oil pumps. And ultimately, you have to update the software, and you have to replace some hardware stuff, replace a network interface. All these things are 5G connected, cellular connected, whatever. And managing these things seems like it's going to be a complex undertaking.

What are we doing as an industry to make sure that we have the capabilities to scale operations up to deal with the number of edge devices that are going to be—near edge, far edge, IoT? You named them all. I think we're going to see a world like that if it's not there already. We're seeing this stuff happening today. But as we inflect, how do we operate it?

Jason Bloomberg:
Well, that is the perfect question, because the answer is fundamentally to move toward a cloud-native paradigm, right? So, the cloud-native story is now becoming the answer to that complex IoT management or edge computing management challenge, because what we really want to do with cloud-native computing is we want to abstract all of our processing capabilities, wherever it is and whatever it actually looks like, whether it's small datacenters, whether it's IoT gateways, whether it's virtual machines, whether it's containers running in Kubernetes.

Whatever it happens to be, we want to be able to abstract all that in a consistent way in order to build a control plane, which is a declarative representation of the policies that drive the behavior of the underlying software. We want to do that in a consistent, comprehensive way so that we can now handle that management in a centralized way, if that's what we want to do, or we want to distribute the management we can do that as well, and we're doing it in a consistent way that is policy-based, that makes it more straightforward to implement the appropriate security and compliance as well, which can be a very important part of the story depending upon the situation.

If there's personal information or data sovereignty issues, the compliance story is important, as well. And cloud-native best practices can help you with those challenges, the guardrails you need for security and compliance, as well as supporting the sophisticated control plane that enables you to manage a diverse, geographically-distributed set of technology resources.

David Linthicum:
Yeah, I like the idea of a control plane from an architectural perspective because we're not necessarily replicating things across various processers. We're providing control and federation of various processes, data, security, governance, operation system. We just hit it, and I think we're going to be able to address all problems. And I think even using this model in kind of the multicloud use cases, because we kind of haven't solved how we're going to operate things across those systems as well, and ultimately if we get to a point where we're able to have common control planes where we're able to manage this edge computing, able to do operations, able to do security, able to do governance, we're kind of getting more toward this stuff being more pragmatic and cost-effective. Am I wrong?

Jason Bloomberg:
Well, I mean, this is the pattern of technology innovation, is things get better, faster, and cheaper. So, yeah, we're going to get the cheaper part as well as the better and faster part. It's just the nature of the beast, right? We're going to be working out some issues. You were talking about the 5G rollout. That's in progress, right? Is that complete? So, then we start looking toward even further, generation 6G, et cetera.

We're going to be in a situation where we have greater consistency at that protocol level. That's one of the limitations to innovation in the IoT, is that there are so many different protocols, and now the technology has to struggle to deal with protocol resolution. And those sorts of problems—we saw that back in the network wars Netware versus TCP/IP, and that got resolved, and it's the same idea. Eventually it sort of gets cleaned up, and then there's a certain set of protocols that make sense to everybody, and that simplifies now the next layer up, which is the control of how all these things work.

So, as time goes on, we solve some of these problems that are providing roadblocks in front of us. We move on to the next set of challenges. But whatever it is we were trying to do, yeah, we have better, faster, and cheaper approaches to dealing with those problems.

David Linthicum:
So, what do you think are going to be the real-world challenges in rolling out this technology? You and I have been in this business a long time. We write papers and do a lot of talks and talk about kind of an innovative way to improve upon the way in which we're doing architecture and implementing technology, which makes a lot of sense. But rolling it out for example, a business actually implementing this technology is going to run into some obstacles. What do you think those obstacles are going to be and how are they going to be overcome?

Jason Bloomberg:
Well, we've talked about security and compliance, so—but those are boring topics, so here's one that's a little more interesting, what I like to call the coming latency war, right? We talked about how moving technology closer to the user lowers latency because you have less physical distance, and really the whole reason that that is a challenge is because of the pesky speed of light, right? There's no way to go faster than the speed of light, so if you want to get a bit from one place to another, it can only go as fast as light can go from one place to another, and generally a bit slower because it's not just light over a piece of fiber, but there's other pieces of equipment along the way that slow things down a little bit.

So, if you want to get as close to instantaneous as possible, you're going to have to do all sorts of tricks about moving the technology around and dealing with special ways of handling caching and other specific sort of tricks that you can provide in order to get better behavior at those user interfaces. But

there's no perfect solution because the speed of light is inviolable, right? So, there's—so there's always going to be tradeoffs in terms of how low you can drive your latency.

So, the question is, well, just how important is low latency? It's important for certain things. If you're doing real-time gaming with your buddies, you obviously don't want it to pause because your buddy's going to come around the corner and shoot you while your screen in paused, right? So, you don't want that, right? That's a latency problem.

But there could be—and there are any number of different business situations where low latency is becoming increasingly important and it becomes an area of innovation, right? If we have ultra-low latency as well as high performance and all the other benefits of next-generation technologies, then it becomes a battle over how to build the best solutions in particular categories. And that's always what innovation is about, right? It's a circle of life thing, right, predators chasing after prey, companies that are innovating better than the other ones.

David Linthicum:
Yeah, this is fascinating stuff, and certainly I think it's the direction where I think the industry is going. And I'm really glad that folks like you are leading thought in that space. So, where can we find out more about Intellyx and yourself and find your content on the web and how can our listeners understand more about Jason Bloomberg?

Jason Bloomberg:
Yeah, well, Intellyx is spelled I-N-T-E-L-L-Y-X, so Intellyx.com is our website. You can follow me on Twitter at @TheEBizWizard, and I'm on LinkedIn as well. We do quite a bit of social media on LinkedIn as well. So, that's probably the best place to find us, is LinkedIn, Twitter, and our website.

David Linthicum:
Yeah, follow Jason. He's one of the thought leaders in the business. It's going to be like that for a long time. And he kind of knows where things are going and really looking at the big pictures and the big patterns and what you can use to be successful and pragmatic use of this technology, which is even more important.

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. And 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. Everybody stay safe. Cheers.