
# Architecting the Cloud, part of the On Cloud Podcast

Mike Kavis, Managing Director, Deloitte Consulting LLP

**Episode: Title**: Going to the Edge: Edge computing what it is and how is it helping shape the future of IT
Duration: 0:27:13

**Operator:**
The views, thoughts, and opinions expressed by speakers or guests on this podcast belong solely to them and do not necessarily reflect those of the hosts, the moderators, or Deloitte.

Welcome to Architecting the Cloud, part of the On Cloud Podcast, where we get real about Cloud Technology, what works, what doesn't and why. Now here is your host Mike Kavis.

**Mike Kavis**
Welcome to the Deloitte Architecting the Cloud podcast. I'm your host Mike Kavis, and I am here with Rob Hirschfeld Founder and CEO of RackN, and I think we've probably done this a couple times in the past, so great to have you back.

**Rob Hirschfeld:**
Thanks Mike, I love it, you and I have a very colorful discussions.

**Mike Kavis**
We do, and we agree mostly on most things it's good.

**Rob Hirschfeld**:
Most things, I love that, makes sense, that's what makes it good.

**Mike Kavis**:
What's that?

**Rob Hirschfeld**:
That's what makes the conversation good.

**Mike Kavis**:
Yes, I'm a little too pro-cloud I think sometimes, so, tell us a little bit about your background, about RackN, and then we'll dive right into it?

**Rob Hirschfeld**:
I'd be happy to do it, so RackN is a company I cofounded four years ago now, to really tackle what we see as a core problem in data centers, which is the physical layer, so we focus on creating cloud like experiences, but for physical servers. Not virtualize, we don't, you know, we can manage things like that but really making regular servers, commodity servers do the things that you expect, infrastructure to do in the Cloud, Image-Based Deployment, BIOS configuration, Pooling, Real API-driven automatic workflows, that type of thing, so that's what RackN does. It really comes out of a vision for us, because back 10 years ago, when we were early in the Open Stack and Hadoop days and SolveIt. If we can improve data centers, how they operate than everything else we get easier and we've been really working on that vision, so we're excited about that. It's a nice tier up for us about some new opportunities that we see coming around Edge Infrastructure and things like that. And as a side project, I'm doing what you do in that we have a podcast we started actually going to a year now called the "L8ist Sh9y" and where we talk about Edge and Infrastructure and things like that to so well bring flavors that into that into this conversation if that's alright.

**Mike Kavis**:
Yes, actually, it's pretty much what we are going talk about today is Edge, so the question: you were at Edge Conference—Edge Congress Conference—and I wanted to get some of your key takeaways on that, but before we do that, there was a lot of talk about Edge and Fog. Are they the same, and are they different, so maybe we just clear that up for everyone?

**Rob Hirschfeld**:
Oh boy, so Edge and Fog are just - people are having trouble defining Edge let alone what Fog is? "Fog" was a marketing term that Cisco introduced to describe what they thought would be Cloud in remote locations or "Cloud Supporting IoT Devices," sort of more distributed closer to the ground. What's the idea with Fog? And then it ran into trouble because it was a marketing term and then it's slowly being getting some revival because it feels like people who talk about Edge can't resist the comparing it to Cloud and making it feel like it's part of the Cloud, and it is, and it's not it's sort of funny like that. So do you want me to define Edge a little bit for people that are helpful?

**Mike Kavis**:
Yes, that'd be helpful.

**Rob Hirschfeld**:
Sure so "Edge," there are lot of definitions for Edge and it's typical for an Edge Conference to spend about half the available time defining Edge, which I'm not very patient about that, so Edge to be very simple about it, any device in the environment is usually considered an Edge Device: your phone, your thermostat, a car, a camera, anything. What we typically talk about and mean with Edge is not the IoT or the device Edge, but the Infrastructure Edge. So what gets IT people excited is this idea that we're going to be putting Data Center Infrastructure near these IoT devices distributed in the environment, and so that's a lot of times when people think about Edge they think it's like data centers and stores, or data centers and cellphone towers. I've sort of evolved a more generic definition, which is, it's IT infrastructure with constraint. So, if its bandwidth constrained, or power constrained, or processor constrained, anytime you're dealing with IT environment that's constrained it's really Edge. Cloud is unconstrained IT, where you can consume as much as you want, or you

assume that you can consume as much as you want. That to me is the big difference between Edge and Cloud.

**Mike Kavis**:
Yes so to put that in a context, one of my favorite use-cases for Cloud and Edge, I wrote about long time ago, I keep bringing up, this is the windfarms right, so the windfarms have all kinds of sensors, on the blades and the long stem and it looks at weather patterns and tries to adjust the blades to maximize output when it detects wind and vibration. So you have these Edge devices making real quick decisions, but like there's too much latency going back to a Cloud, so they have some level of infrastructure in the field that is a communication mechanism and then, at then at the same time its trickling all that data back to the Cloud where they do some heavy analytics, so that's a use-case I think that speaks to this. Anything, any comments on that?

**Rob Hirschfeld**:
I think that's a really good one, right, you've got a lot of devices collecting a lot of data, don't send all the data back to the Cloud. You need to do an analysis right there no latency; you're adjusting blades by…algorithmically, usually, with some AI would be my expectation. And then there are safety concerns and all sorts of things that require you to have real IT onsite. It might not be in the windmill or in the turbines; it's going to be in a little cabinet that's deployed nearby, and that's usually people think of that as "Captive Edge Infrastructure." What gets people very excited right now is "Shared Tenant Infrastructure" where you could imagine that windfarm being powered from a Shared Data Center Infrastructure that's in close proximity to it, and then you could start basically selling new applications. The challenge with a captive data center for the windfarm is it's really just the captive data center there's no -- you can't introduce new applications without talking to the people who designed the windfarm, and so what we want to have is a way to have multi-tenancy, or a platform where we can deliver something new into that environment, and that's where it's windfarms people talk about tractors, autonomous cars of course drones, but there is huge applications for the sort of shared infrastructure that anybody could write an application for and run globally distributed with low latency.

**Mike Kavis**:
So one thing I find interesting and kind of…I'll say myself, but ourselves, is when I first started in this industry there was mainframe and I was working at a steel plant and they had all these devices on hot strip mill which was pouring the steel, and it collected. All this information…it was IoT before IoT. The problem was to get thrown on the VSAM file, and we read I think the next day in 24 hours we turn around report. So the use-case has been there with technology change but what was different was you didn't have the connectivity, you didn't have the RAM, you didn't have the CPU, and I'm like this is that all over again. Right now, you have really small devices out there with limited capacity. If you look at a self-driving car, it has a data center on wheels but sometimes you're not connected, so it's kind of pushing us back to how we had engineer stuff back in the day, but with better technology. What's your thoughts on that?

**Rob Hirschfeld**:
There's definitely a technology piece to it, but when I was in my early career, I was doing Programmable Logic Controllers, or PLC's, with factory floor oil rigs – I've actually done oil rig systems, and while it's great to be able to think of more Compute and AI and things like that, the thing that I feel like is revolutionary is the breaking out of the "Captive Sensor Model," and so let me -- I want to try and explain that a little bit, because it's really where I think people need to think about Edge. What we're doing is we're trying to make it where you can buy sensors and devices from a wide range of vendors and plug them together super inexpensively, so what we don't want to do is say all right, I have to have a physical wire attaching to every device, and I have a certain number of ports in my compute infrastructure certain amount of capacity. If we can open that up and drive the cost of adding a sensor or a new platform or a new capability down, then we really open up an explosion in immersiveness, in the environment, and that's to me what people need to think about with Edge It's fun to talk about the tech, but the use- case here is that we should get a real sensor explosion where we can use all sorts of data from all sorts of sources: cameras, wind, weather, passing cars, shared information…that's the explosion to me that is the next phase and then what really drives Edge to be exciting, because otherwise it's just Captive IT Systems, which are great, but not a commercial opportunity for everybody.

**Mike Kavis**:
And what's interesting now is typical IT we own all these devices. In this world we don't, right? If we're communicating with your and my cellphone or whatever. So how does operations look in this world? Is it totally different than what we are used to?

**Rob Hirschfeld**:
It's very different and I don't think people know. We'd like to hear a lot of people talking about it being mini-cloud, which I think is completely wrong. The sense that I have for Edge Infrastructure is that it's not a human scale problem, and it has to be managed differently; we have to treat it differently, and let me explain that in a couple different context, because just from an IT infrastructure perspective, if you have hundreds of thousands of mini-datacenters running workloads, you can't afford to have a crash cart go to fix things. So they have to be completely lights out, they have to be zero touch, they have to be managed in a distributed way that we don't

manage datacenters. That way, right? Actually, we've been doing the opposite we're giving everything over to Amazon and saying you take care of it in a handful datacenters. What Edge looks like from an infrastructure perspective, it's going to look like something that requires a truck roll to go fix a problem, if there's a problem, and they can't, we can't do that it doesn't work. But the other thing that's really different between Cloud and Edge is that you're not going to log into an Edge site and do maintenance on your application. So today we sort of have this, you go get a cloud server but you own it, you log and you manage it, you control what zones and regions you're in, you control how systems are distributed by and large. And with Edge you're not going to have that luxury, you can't, so what we have to be able to do from a platform perspective is for an application author to upload an app, and then tell the platform how to run that app in a distributed way. And so the platform has to be able to say, oh you have a user in Detroit, I'm going to run workload near that user when they're not using the application anymore. Destroy that, so you're not paying for resources. That type of spinning up and spinning down applications, it's got to be completely automated, because it's going to be constrained by the resources that are available, but also by the network latency. And so those are both operational, critical operational components that need to be figured out and don't really exist in current paradigms.

**Mike Kavis**:
And those platforms you're talking about are they living on the Edge datacenters or back in the datacenters and Clouds, or both?

**Rob Hirschfeld**:
It's got to be sequenced in both, so when people develop these applications they're going to develop it in Cloud-like infrastructures, but then, and then interact with the management platforms. It's not that different, if you want to think about it, than serverless what' we're doing with Lambda, and things like that, or Containers and Kubernetes, but because really what we say here's a workload go run it, we talk to our platform and it does it, the difference is, is that we know where that work is going to be done. In the Edge Infrastructures, we're going to have to give that up. So it's going to have to make the system—the platform is going to have to make the decisions about globally where that data is run, and then that creates networking data access latency. There's a whole host of problems that have to be also addressed, right? How do I have secure data? I can't trombone my data from an Amazon, or I keep saying Amazon, but it could be any Cloud Platform, back to the Edge site, and if my data is in the wrong location, that could break the whole latency, break the whole application, so we have layers of problems to solve to build effective apps here.

**Mike Kavis**:
So what do you, what do you think are the Killer Apps for the Edge?

**Rob Hirschfeld**:
So, I love that question in that it's not entirely clear yet what one is specifically, but the one that gets everybody excited is autonomous cars, and I think that that is a potential a huge market driver; there's going to be a lot of money and funding, and I've seen drones discussed too, but I actually like to answer the killer app question a little bit more meta, if you will. What I think is a constraint for the killer app is that battery, and compute, and price of device are really problems. The amount of compute that we're putting in an autonomous car, and the cost of that car, and the battery drain of running all that compute in the car are problematic. The same would be true for Augmented Reality Goggles, right? If I want Augmented Reality Goggles, and I have to put a powerful processor and a big battery in those goggles, I can't make it affordably and I can't make it wearable. So the killer app really comes into a play, where you can take things that we're doing today on a device and move it to an Edge Infrastructure, and augmented reality could be that. So the way to think about that is, I can sell you a pair of goggles for $50, because the processing for that is actually done as a service, and the AI and the video analytics are all done as a service, and that drives the cost of the goggles way down, and then you're paying for the compute as you use it, which transfers the cost to the consumer, rather than the device producer. It makes the device more affordable. Same is going to be true with drones. Drones have very fixed payloads and so if I could make a smart drone, or a prolonged battery life, by moving processing away from the drone and on to near compute infrastructure, then I could radically transform the cost or the power of drones, right, which could turn them into better delivery vehicles. And so those, to me, the killer app here is ones where we have a physical constraint: battery processing, weight, that are going to be addressed, and I just named a couple of them, but I think the killer app will fall into sort of that meta envelope.

**Mike Kavis**:
What's interesting, remember your Comdex Conferences back in a day?

**Rob Hirschfeld**:
Oh yes.

**Mike Kavis**:
I remember going to one and it had to be right around 2000, and they had cars, talking cars talking refrigerators, all this stuff, and they

were talking about smartphones, probably in a different context. It was all there; just as you say, everything was too expensive, right? It was way too expensive and but over time the cost of storage and compute and bandwidth, and even bandwidth got more available, and that's where we are today, where some of this stuff is feasible. What's it going to take, now, when we're talking about Edge, to do that same turnaround, right? You're talking about, you know, all these cool things, but they're still a little too expensive. What is that, Moore's Law? What is it that's going to turn this around, so five years from now, all this stuff just normal and cheap.

**Rob Hirschfeld**:
So, the safe answer is 5G, which is the new spectrum that's getting open for telecom that allows higher bandwidth, with lower latency. It's more of a data focused cellular spectrum, and I think that that's, having the medium, is going to be a huge piece for this. To me the challenges are operational, because that's where the costs really come in. If I'm going to put datacenters in a whole bunch of locations, the hardware is not that expensive right now, but actually having a way to manage that hardware that lets me keep up with the pace of IT, and IT innovation is the missing element with that, because right now telcos, telephone systems, even Cloud vendors and enterprise vendors, are really pretty slow, and they move faster by going to one location and limiting their scope. This problem isn't going to necessarily be solved, as you know, in one location although like I said should amend that. It's, the way people are approaching this today, is they're taking metro areas and they're building prototypes within a metro area and creating really significant multi-tenant Edge infrastructures in a major city, and so you might see cities...I think this is going to happen even faster in Asia, where there's fewer regulations to control, like Drone Propagation and Autonomous Cars and safety. So in those environments, you could build a metro area, low latency network with a fair bit of capacity and then create a decent innovation lab. And I think you'll see those come out as reasonable markets. It would only take 20 reasonable metro markets to cover a significant and profitable part of the country in the US. And I think with some of the mega-cities that we see outside the US, maybe even less, and so I don't think there's a lot of barriers to making this stuff happen. And then that will, in turn, make it possible to have the devices. I have a small aside. I don't want to run us too long, but the killer feature for the iPhone was actually Internet access, right? The magic of smartphones today is that what you've really done is you've put all of the Cloud Infrastructure in somebody's pocket, and so from that perspective, the power on the phone is not the capabilities of the phone, although that's been progressing. The capability on the phone is its ability to access these close-by IT Infrastructures, and we should expect that trend to continue.

**Mike Kavis**:
Absolutely, so last question here this is a thing I struggle with, because I do a lot of consulting in the space of Cloud, and I see the struggles companies are having transitioning from datacenter computing to cloud computing. Now, they have to go do Edge stuff, and the things are changing; the innovation is changing in such a pace, if I'm a CIO how do I even get my arms around it, and how do I move the needle on us, for so much change coming at me so fast? How do we keep up with this stuff?

**Rob Hirschfeld**:
Yes, this is a real challenge and it needs -- I want to give a thoughtful answer for this, because telling people to run faster isn't working. And frankly what I see, and what we deal with when we talk to customers, right, our production, it was plugged, right, will improve people's operational efficiency 10 times. It's a significant improvement, right? It's a system-changing type of improvement, so we're like earning SSDs to build systems or CI/CD Platforms, right, where you literally change the way people interact with these systems. And the way that people need to keep up with this is to look for these types of opportunities where they can change the fundamental costs of doing business. And when you can do that, when you take out a major unit of your production costs or a major block, you actually change the way people interact with things. I want try and define that. It it goes back to Theory of Constraints and Goldratt, but what happens is, if you can find an opportunity within your infrastructure where you can change people's behavior regarding how they look at something...Cloud really did this where it used to take weeks to get a server and build something, and now with Cloud, it took minutes, and then people said well, I could I just automate that. Now we have a CI/CD pipeline so you don't even get servers; it just happens. Those are places where we've really transformed our thinking about things, because we've changed the fundamental cost, human cost, more than the physical, the actual dollars cost, but human cost of doing that type of work. People need to think that through. The thing that keeps people from thinking it through is that they get overworked, they get too busy, and so most of the people that we deal with are so far underwater working150% right, not being able to take risks because if they took a risk and it went bad it would -- they don't have any slack in their schedules to deal with it. You really can't innovate when you're running a 100-plus percent. Innovation means taking some risks, innovation means having time to get away from the tedium of what you do, to innovate and I'm sure there are people listening, but my job is innovating I am constantly innovating! You need to be careful. It can be very easy for you to run 100% doing the work that you think you have to be doing, and not leave time to transform that work, and so Mike, to me when people present, how do things are changing so fast I need to keep running...you're not going to win the race if you're running at a certain speed and somebody says you got to run faster to win the race. Sometimes you got to stop, sharpen the saw, get better shoes, train, spend money to hire a trainer, fix yourself and take and sort of take this attitude of, to keep up with technology I actually need to get my team to go a little bit slower so that they can find the real improvements, not the incremental ones.

**Mike Kavis**:
Yes, I spend money…I think you need to invest, and doesn't always have to be spending money, maybe not doing something else, but you know it seems to be, and this you know, there's a new thing, put that on your plate.

**Rob Hirschfeld**:
And that's it's a real hard problem. We see that. I mean it's a funny, it's an oxymoron here in that I'm saying go faster, change things to go faster, and to do that you have to make time to add the new thing in. It's not a simple problem, or people would be doing it. You need to figure out how to give teams, your team, a chance to find improvements within their infrastructure and then make them happen. We see that all the time. It's sort of classic like for us, the Spectre BIOS firmware should have caused a lot of people to go fix their firmware deployment techniques, right. Just as a history, right, a couple months ago we had a major firmware bug. People had to patch every system they had, and nobody went back and fixed firmware patching. They just ran through and did it is fast as they could by hand, on this sort of expectation that it wouldn't…there wouldn't be a second event, but there will be, and so this is a classic case of, its urgent, I put out the fire, move to the next fire. If we keep doing that, we're going to be, you know, that's where security happens, that's where these problems occur. For Edge, specifically, to bring it back to that, it's really about getting good operational practice, and this is what scares me when I look at some of the industry work going on around edge, where we're just trying to take existing virtualization platforms and stuff them into smaller Edge infrastructures without thinking through the real problem is, that's a slower road to make this work. Just because you have some existing tech doesn't mean it's well suited for this environment, and I really hope that companies looking at Edge work are thinking through, how do I manage that, before I get hooked into a deployment fix. Figure out your management story, figure out how you're going to scale it. It's very expensive at scale. It's incredibly even more expensive to add zero touch automation back into a system after you've done it wrong, and so I hope people take that, so I'm sort of like, you can hear my voice -- I hope people take it seriously, because they never do. But they do it, we need to we need to address this and fix it. We're doing great work in Cloud improving efficiency and automation. We just need to take those lessons and bring them back into these new opportunities.

**Mike Kavis**:
Yes, well, great conversation Rob I appreciate your time today. Where can we find you on Twitter and any of your blogging and your podcast. Where we can get a hold on that good stuff?

**Rob Hirschfeld**:
So on Twitter, I am @zehicle Z-E-H-I-C-L-E, that's like vehicle but with a "Z." Goes back a long time for my electric car days, and so I'm very active on Twitter and easy to get a hold of their "L8ist Sh9y" podcast we're doing weeklies. It's a lot of fun, and that's a good place to catch up with what these conversations were having about Edge, and Open Source, and some other tech items and that is L-8-I-S-T-S-H-9-Y. Our company is RackN, R-A-C-K-N. Try and keep things short and simple with that, and the Open Source Project that we maintain around that that does IT Physical Infrastructure in some really novel ways, that is "Digital Rebar" Rebar. Digital is the domain. So whole bunch of ways that people can go find out more and interact and pester me with I didn't agree with that which I love to hear and discuss.

**Mike Kavis**:
Following about they are in Twitter is always good information. So that's our show for today you can find this podcast and others by my colleague, David Linthicum, by searching for Deloitte On Cloud Podcast, so we'll see you next time on Architecting the Cloud.

**Operator:**
Thank you for listening to Architecting the Cloud, part of the On Cloud Podcast with Mike Kavis. Connect with Mike on Twitter, LinkedIn and visit the Deloitte On Cloud blog at www.deloitte.com/us/deloitte-on-cloud-blog. Be sure to rate and review the show on your favorite podcast app.

Visit the On Cloud library

www.deloitte.com/us/cloud-podcast