

DR: [00:00] I mean, when a customer calls you at midnight on Friday night, they're pretty serious. They're like, "Yeah, we've lost everything. We need a cloud solution. We need it right now. Can you do it? And can you get this running in two weeks?" Now this kind of system takes at least six months to install. Usually, it's about a year. And the world's biggest telco companies are usually the best at it—Ericsson, Huawei, Nokia, Amdocs. But they didn't have cloud solutions. And we're like, "I don't know if we can do it in two weeks. I like the way you're thinking. Let's go for it." And we did it in 18 days.

Announcer: [00:39] This is "Telco in 20", a podcast that helps telco execs achieve a competitive advantage with AI and the public cloud. It is hosted by Danielle Rios, also known as DR. Today we're talking with Jillian Ford, Senior Solutions Architect at AWS.

DR: [00:58] Hi guys, I'm DR. Are you headed for Las Vegas for AWS re:Invent? It's that time of the year when thousands of cloud nerds head West to learn about all the latest hacks with AWS.

[01:11] I recently had the opportunity to sit down with Jillian Ford on "The Official AWS Podcast," where I shared the incredible story of how Totogi used its amazing Charging-as-a-Service product to restore connectivity to millions of people in Sudan. It was the kind of challenge that would take any other vendor in telco six to twelve months to do. But not us. To solve this problem, Zain Sudan needed a 100% public cloud-native solution. And since we built Totogi as an AWS-first product, we were able to do what literally no other vendor could. Migrate over 20 million subscribers to our system and get them up and charging in record time.

[01:50] So as we roll up to the start of re:Invent, I'm sharing my conversation with Jillian on this podcast. We talk about how Totogi did it, why being AWS-first made all the difference, and how other telcos can absolutely use the public cloud for mission-critical workloads like charging. The AWS version of our conversation is a little longer than this one. We've created a condensed version for our "Telco in 20" listeners. You can find a link to "The Official AWS Podcast" episode in the show notes. So let's take 20.

Jillian: [02:26] Welcome to “The AWS Podcast.” I’m your host for today, Jillian Ford. And we’ve got a remarkable story about a company that was able to onboard 23 million new customers in 18 days. That is absolutely remarkable. We’re going to talk about how they were able to do that. And of course, using AWS. I’ve got a phenomenal guest with me today, Danielle, who goes by DR, CEO of Totogi. So DR, tell us more about you and Totogi.

DR: [02:55] Yeah, I am telco famous, really well known in the telco industry. But yeah, everyone calls me DR, my initials, I’m not a doctor. I am CEO of a vertical SaaS company named Totogi that focuses on rewriting all of the software of the telecom industry. And this is one of the last industries to really move in a big way to the public cloud. And so I’m so lucky to be seen as a visionary in the industry. It’s been super fun and a great ride and I’m super excited to talk about it today.

Jillian: [03:24] Well, I can definitely tell you that after this interview, people who are not in the telco industry are going to know you, so you’ll be cloud famous.

DR: [03:32] Awesome.

Jillian: [03:34] Let’s kind of start with really this story of a SaaS company being able to onboard 23 million new customers in 18 days. And why it was that Totogi had to be able to onboard so many customers so quickly?

DR: [03:46] Yeah, I mean, when you’re a startup, you’re really trying to find your big breakthrough. And our very first product is a product that the telco industry refers to as charging. Everyone is familiar with charging. I’m in my fifties now and I remember the AT&T commercials about calling grandma on nights and weekends, and that’s because the rate or the charge for the call would be lower. The business day was the expensive time to call long distance. And so the product area is old, it’s been around since the eighties. And so all the telcos out there have a charger. And the charger is determining how much I’m going to charge you for the minutes of this phone call or that text message or the data that you’re pulling down.

[04:28] Now in the United States, we’re post-paid. Meaning we pay a flat fee, we pay \$60 a month or whatever it might be, and

we get unlimited on all these elements and we don't really get charged by the minute. But the majority of the world is prepaid, meaning it's almost like a debit card. Where you have to put money into the telco first and then it deducts. And so the charger is really kind of like an ATM in the sky. This is their main monetization engine of their biggest asset, the network, super mission-critical. And so we had this situation happen with a customer of ours, and really the only vendor in the world that could have solved this problem was literally Totogi because of our AWS-first cloud-based SaaS architecture.

Jillian: [05:12] Yeah, I think, tell us more about this company that you end up working with. And what came about to decide like, hey, now you need to be able to quickly onboard so many customers?

DR: [05:23] Unfortunately, there's humanitarian crises in the world. In Sudan, there's a civil war going on in the country, and every country has about three telcos. Most of these telcos still run their operations and the majority of their business software in on-premise data centers. But they're on the ground, they're old school, they're like the data centers of the nineties, before the public cloud. And so they were worried that there would be some sort of political or war action aimed at their data centers. Either electricity would be cut or they would be bombed, and what would be their business continuity plans? And so we had been talking to them, we have a cloud thing, maybe we could solve it that way. And internally at Totogi we're like, if data center one goes down and data center two goes down, then you'd use Totogi. We were taking it seriously, but we're like, "Are they really going to buy this?"

[06:14] Well then the unthinkable happened where both data centers lost electricity simultaneously. Not just at this one telco, but also all the telcos in the country. Literally, you can see on the graph, mobile network goes dead February 4th of 2024. I mean, when a customer calls you at midnight on Friday night, they're pretty serious. And they're like, "Yeah, we've lost everything. We need a cloud solution. We need it right now. Can you do it? And can you get this running in two weeks?"

[06:43] Now this kind of system takes at least six months to install. Usually, it's about a year. And the world's biggest telco

companies are usually the best at it—Ericsson, Huawei, Nokia, Amdocs. But they didn't have cloud solutions. And we're like, "I don't know if we can do it in two weeks. I like the way you're thinking. Let's go for it." And we did it in 18 days. And it was onboarding all of their customers into our system. And we had to work 24/7, we were using the global clock. I was using Asia workers, North American workers, EMEA workers, people sleeping whenever they could. But we got them up and running and really saving lives. It's insane. People come up to me and they're like, "I want to take my picture with the woman that saved millions of people's lives because you restored mobile network connectivity to the country." It's insane.

Jillian: [07:33] It's absolutely remarkable. And when you decided, we're going to have to onboard all these customers in such a short period of time, from the actual cloud perspective, was there anything that you had to do? Whether it's change in the architecture, maybe some of the technical processes in order to be able to get all those customers onboarded.

DR: [07:53] Yeah, I mean we had some of the processes because we had been onboarding customers, but nothing at this scale. When you're talking, "Hey, here's a file with 23 million customers," it's really going to put your organization and your processes to the test. And so the first thing we did was like, "Yo, AWS watch out because DynamoDB is about to go crazy and it's intentional."

[08:13] And so it was really important for us to partner with AWS and keep them abreast of what we were doing. It wasn't like a switch and we turned on the network all at once. We really kind of rolled out city by city very quickly. But those workloads were scaling. And so even though they take minutes in telco time, that's too long. We need milliseconds. And so we were kind of front running and a little bit over-provision just to make sure that the workload wouldn't be throttled or have any issues.

[08:41] And then over the course of the last couple of months, we've been right-sizing it to make sure that it can scale fast enough and we have enough room. But yeah, I mean we didn't really have to change our product at all. It was ready to go. I'll be honest, I was a little bit like, "Is this going to work guys?" And

they were like, "Theoretically it should." And we had worked with amazing solution architects at AWS who had reviewed our design. And we kind of take this approach of WWAWSD, like what would AWS do?

Jillian: [09:12] I love that.

DR: [09:13] We're not designing to be cloud-agnostic. We're not trying to be multi-cloud fancy. This is a service, kind of like Salesforce, but for a mission-critical workload like charging. But design this the way AWS would do. They would not use Kubernetes, you guys would use Lambda. You wouldn't use an agnostic database, you use DynamoDB. And so our engineers are like, "We can write that." I'm like, "Shut up. We're going to use AWS services."

[09:36] And it's made not only our product better and more resilient because we're writing the innovation wave of AWS. I mean, we're just betting so big on you guys, that you continue to expand on regions, you continue to invest in your chips, you continue to build amazing services that are going to keep coming out. And we're riding that wave. And then we just build our telco industry-specific solution on top of it. And so our product keeps getting cheaper and it keeps getting better. And our customers benefit from it. And so yeah, I'm not just cloud-first, I'm AWS-first.

Jillian: [10:08] I love definitely being able to hear that. And I think that also resonates with a lot of the listeners here who are really thinking about the right approach to be able to design their business. And every company, I guess, really hopes that they're able to onboard new customers or just planning for the unexpected. Like with COVID where maybe some people had a lot more customers to their business. Could you tell us a little bit more about when you started working with AWS and what was the architecture like? What was the actual planning process? Just so people here can be able to learn from your experience.

DR: [10:41] It was a conscious choice to pick AWS as our hyperscaler, and a previous job I had been working with Google Cloud. The key, especially in my industry because it's so global in nature, it's highly regulated. Where data sovereignty and data regulation is actually a pretty important topic. And right or

wrong, we had this thesis that Amazon retail were continuing to build out fulfillment centers coupled with data centers. I mean, this is a trillion-dollar CapEx investment. We needed a hyperscaler that was going to build out a lot of regions and a lot of countries. So that our telco customers could legit use it.

[11:19] That was one of our first guiding principles around engineering. It's like, "It's going to be AWS." And we're like, "If it's going to be AWS, then we're all in." We're not going to hedge your bets and change cloud providers later. So yeah, use Graviton. Don't go Intel chips because we might leave later and we're not sure. These are bold bets, we're not on the fence. And so from there, there's a lot of decisions that come in terms of our architecture. But it became very easy. You just keep picking AWS. There's Kinesis, let's use that. And I always tell people they're like LEGO pieces, they're building blocks. It really accelerates your application development time. They're battle-tested, they're robust. So we love all of that stuff and we try to infuse that culture and that thinking into our company. We're so invested with AWS, for sure.

Jillian:

[12:05] I'd love to hear more about the architecture and engineering team, as well, that was able to handle all of these new customers that were on the platform. On the architecture side, you said you were using Graviton, Kinesis, DynamoDB.

DR:

[12:18] We'll just start with the most important component in the application, which is the database. This is a transactional database. Like I described earlier, you're checking if the consumer has balance. And if they are, you've got to record it in a call record. Back in the nineties, we would get packages from your cell phone provider of all your call records to prove the billing, right? I charged you a dollar for that text message, I charged you 36 cents for the call. And so they would send you the call records, CDRs, we now call them EDRs, Electronic Data Records. This is the key to this product. Any charger, it starts with the database. And a lot of the old-school legacy databases use Oracle. Because it was like the world's best enterprise database. But DynamoDB handles Prime Day and that's way more transactions than what we're doing. And so we're like, "If it's good enough for AWS, it's going to be good enough for us." And so we use a single table database, DynamoDB, and that's

really a big part of the product and everything else is kind of driven off of that.

Jillian:

[13:16] It sounds like the architecture and your perspective on how you were going to best utilize AWS was able to allow you to be able to onboard so many customers within such a short period of time. So that's truly remarkable. But I'm curious if there was anything else that maybe you had wished you had done before that.

DR:

[13:36] Any advanced warning whatsoever would've been nice. Like I said, we had designed how we would do this and how this product would work in a business continuity setting. And we had vetted it with some of the primary architects of DynamoDB and they had given us ideas on how to better design it. We had vetted it with AWS architects and engineers like WVAWSD, like what could we do better? Okay, let's go do that.

[14:01] We had done some scaling tests with synthetic data. We had scaled to a billion transactions per second. Which is less than what DynamoDB does on Prime Day. And that would approximately handle all the charging volume of a continent. You might need a second instance for India because they're like at 1.3 billion subscribers. You might need a second instance for Africa. The telcos have one charger or multiple chargers per telco. And when you think about this on-premise workload, you have your primary key mission-critical product. You have bought and provisioned capacity to handle fluctuations and workload. You have these unexpected events like holidays, but it could be a natural disaster, could be a flood. And so the IT team doesn't want to be the guys that under provision to save money. So they ask the vendors, "Hey, provision this for our capacity." And then they bump it up by another 20, 30%.

[14:54] So usually you have over-provisioned capacity. They're running hot, but they're not being used. Now you think about business continuity. Okay, now you're just going to double it. It's a whole copy of another system. Software, it's hardware, it's maintenance, it's support, it's security testing, patching. All of that work is going on. This is just such an interesting use case for the public cloud.

[15:17] For all the enterprise executives listening to this, where they're like, "Oh, my mission-critical workloads are insane, DR." I'm like, "Try it with your business continuity stuff that's not even being used right now." Half the time when there is a disaster, you don't flip over to your DR system, not me, but disaster recovery. You don't flip over and use it because it doesn't work. You haven't tested it in months and you're afraid to flip over. But imagine if the system was essentially free. Because it ran in the cloud and you only paid when you used it.

[15:49] So when we first priced this for Zain Sudan, we're like, "You're going to need this one or two days a year. And so we'll charge you a little fee to make sure everything's plugged in. But when you flip it over, it's just going to work." And so it's such a great use case to start to dip your toe into using AWS. But when you're like, "I flipped over to public cloud, I kept our network up and our revenue going and we didn't miss a beat. And then we could fail back to our old safe systems and it didn't really change anything." You'd be insane not to have it. We think this is a really cool idea for telco, but any highly regulated, slow-moving industry that's afraid of the public cloud, why don't you start with disaster recovery and business continuity? But again, business value workloads that really matter. The easy stuff you've already moved.

Jillian:

[16:33] Totally. Yeah, I couldn't agree more with that piece of advice.

[16:37] So DR, I've got one last question. And this one is really for the listeners who either work at a company in a regulated industry or maybe they work at a company that still has a lot of applications that are on-prem and they're starting to kind of dabble with the cloud, and now looking at how they can adopt generative AI. So I'd love to understand how you're thinking about this. And any advice for those types of companies to look at generative AI to be able to innovate?

DR:

[17:07] In a nutshell, I would say if you weren't already in the cloud, then you really are going to want to be in the public cloud because of generative AI. Getting access to multiple foundational models, making sure that your data is safe and not getting into the training models, being able to experiment



rapidly, and use customized chips like Inferentia and Trainium. These are just gifts from Amazon and AWS.

[17:31] I mean, this is a marathon, but everyone's running it like a sprint. Public cloud was a big change, but this is going to change work as we know it in the biggest way and certainly my lifetime. And so I talked to telcos and it's a highly regulated industry, and they're like still fighting about, "Oh, the hyperscalers trying to steal my data. Is my data safe?" And I'm like, the hyperscalers don't want to move into telco. They are growing like crazy making sure that they're supporting the enterprises of the world. The hyperscalers have the cash and the capability and the technical chops to move into any industry in the world that they want to. So they can, but you have to think, will they?

[18:13] And they won't because if they started to do this, then everyone would completely contract from the public cloud and it would be over. So it'd be dumb for them to do this. So while it's possible, it's not likely. So now that you've gotten over the trust factor with the hyperscalers, they're bringing you gifts to accelerate your move to the biggest change that we've seen in technology. If I'm in a regulated industry, first mover advantage to use this in a major big way, I'm doing it. I mean, absolutely. Everyone moves all slow and they're like looking at what are my competitors doing? But if one person moved in a big way and did it right, the other guys would have to follow. And so I'd be moving aggressively forward and I would be all in on public cloud and just going crazy on generative AI. For sure.

Jillian:

[18:57] I love it. With that mic drop moment, DR, I am so happy and grateful that you came to "The AWS Podcast" today. I appreciate your time and thank you so much for sharing your advice to the listeners here on "The AWS Podcast."

DR:

[19:10] Go cloud, go generative AI, and thank you so much, Jillian, this was super great. Thank you.