Ishwar Parulkar: [00:00] What we learned with Telefónica is how do you use the cloud infrastructure for running networks in a brownfield type of a scenario. And we created this portfolio that whatever your deployment requirements are, you have a choice of AWS infrastructure to host those functions on. So it could be the big regions. It could be our local zones. Or it could be these Outpost servers that we announced, rack of servers on your on-prem facilities, or these can be the far edge for running the RAN DU workloads.

DR: [00:35] So this sounds like AWS now has the layers it needs to support all the network workloads to run a network on the public cloud.

Ishwar Parulkar: [00:42] Absolutely. That's exactly the way to look at it.

Announcer: [00:50] 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 to Ishwar Parulkar, Chief Technologist for Telecom at AWS.

DR: [01:09] Hi, guys. I'm DR. There's an ongoing debate in our industry whether network workloads can be run on the public cloud. As telco's public cloud evangelist, I've been pushing operators to go all in on the public cloud for years, but old-school network diehards insist it can't be done, citing either latency requirements, data sovereignty issues or their need for "five-nines of reliability" that they claim only purpose-built telco infrastructure can deliver. But then Dish Wireless happened, a greenfield 5G network in the US totally built on the public cloud. And while it hasn't realized commercial success, the network does work. Still everyone dismisses it.

[01:51] Now there's Telefónica Germany, a brownfield network who last year announced they were going to start testing a cloud native core on AWS with 1 million subscribers. Fast-forward to MWC this year where the story evolved even further. AWS announced a new Outpost 2.0 specifically designed to run network workloads. And while it's not a pure public cloud design, this hybrid approach might finally be the missing link we need to run all kinds of network topologies across multiple generations to truly achieve a cloud native network.

[02:24] Today I'm talking with Dr. Ishwar Parulkar, the Chief Technologist for Telecom at AWS. We're going to dig into why

AWS is creating a continuum of infrastructure options for network workloads, how Telefónica Germany is using this mix from AWS to develop its cloud-native network, and whether telcos should leverage their edge data centers and additional RAN capacity for enterprise workloads. So let's take 20. Dr. Ishwar Parulkar is the Chief Technologist for Telecom at AWS. Hi, Ishwar. Welcome to "Telco in 20."

Ishwar Parulkar: [02:58] Hi, DR. It's great to talk to you today.

DR: [03:00] Yeah. I'm so excited to have you on the podcast. So many exciting things going on with AWS. And so MWC was a few months ago. Still feels like yesterday. But coming out of that show, there's always a lot of conversation about new technologies and tons of announcements. And so what are the major trends you're seeing from AWS's perspective?

Ishwar Parulkar: [03:21] So if you look at telcos, their main challenges are threefold. One is really how do you engage with customers, your consumers better? The second one is cost, just cutting down cost of operations. The industry is one of the largest spenders of money in operations. Lastly, looking at really new ways of building products and generating new sources of revenue. So I see three fundamental trends with telcos right now. One is generative AI, how you would use it, how you would apply it to gain productivity in internal systems and in networks, and how you can offer new products. It's a transformational technology that's going to change all industries across the board, all enterprises. With telcos, it's going to be a big needle mover if the telcos really take advantage of it in the right way.

[04:07] The second trend I see is modernizing the network. And this is a journey we've been on with the telcos for a long time, is how do you build networks in a way that are easy to operate and can be dynamic, more flexible, can give you the possibility of building new consumption models? Something similar to what we did with the cloud with compute and storage. So this is all about how you run network functions on the cloud. And the third big trend I'm seeing is how can network be designed for AI applications. If you see the patterns over the network, they've changed over the years from simple connectivity, voice. Couple of decades were about video. Now it's about really running large scale AI applications across the fabric. And I think that's a big opportunity for telcos to really start looking at how to build networks to handle AI applications. So there's a lot of interesting challenges there.

DR: [05:00] Yeah. No. Super exciting. Like you said earlier, the opportunity for telcos to really reduce their costs, not only operationally for the cost of business but also in the network. And so at MWC, AWS announced new Outpost rack and server offerings specifically designed for telcos. So tell me about this big announcement that you guys made.

Ishwar Parulkar: [05:21] So at MWC, we announced two products. One was a preview of Outpost servers that could be run on-prem and can run network workloads. If you look at how the network is built, it's a series of different points connecting the end users or devices to data centers where your OSS, BSS and the core sits. So you need different locations where you can run these network functions. That's the reason we really looked at building an edge continuum. This is how can you offer exactly what we have in the regions, the same infrastructure, the same services, the same APIs in locations that are outside of the regions, but really tuned for network workloads. What that means is they have a super high performance, high bandwidth. We have crafted instances and network interfaces that allow you to run very high throughput network type of workloads, and they can be installed on-prem where you might want to run functions like the UPF, which either need to be close to the endpoints for latency reasons or at a certain location because of your internet peering points.

[06:27] The second one is a cloud RAN server. So it's a small footprint optimized servers for edge locations, like very far edge locations. And this is where you could run the DU stack. This particular server is again an AWS infrastructure, so same services, same APIs that we run in the regions, but it also has an L1 accelerator integrated into it, the likes of ISVs like Nokia and others use.

DR: [06:53] The big network equipment vendors.

Ishwar Parulkar: [06:55] Exactly. So this is all integrated into that server and the AWS services that are on the top. The silicon we're using for this is Graviton3, which is an ARM-based processor. We believe that power efficiency is super critical in this part of the network. If you look at networks, 70% or more of the power is consumed in the radio access networks. We believe this would be a big differentiator in terms of really changing the silicon architecture that the RAN stack runs on to allow for those power efficiency benefits as well.

DR: [07:27] Yeah. And I know one of the customers you guys have been highlighting and has been big news has been Telefónica Germany. They completed their trial of loading a million subscribers into the public cloud using Nokia and AWS. And I think they're also using these new Outpost offerings. So how's that going? Was that a discovery that you guys figured out was necessary because of Telefónica Germany and so that's what really led to this announcement?

Ishwar Parulkar: [07:51] Yeah. So it's been a two-way street. And it's not just telcos. It's also the ISVs like Nokia and Ericsson because they build the software stack and we had to understand what is required in the cloud to make those network stacks run at scale reliability with performance. And then from the operators like Telefónica, we had to learn how the network topology works, how their current footprint is of data centers and transport fabric connecting the data centers. So it's been learning from both these aspects and then from the cloud side really seeing what is needed to build cloud services or enhance the existing services to meet these requirements. So it's a two-way thing. What we learned with Telefónica is how do you use the cloud infrastructure for running networks in a brownfield type of a scenario.

DR: [08:41] Situation.

Ishwar Parulkar: [08:42] Yeah. So we had done Dish Wireless before, which is a greenfield network. We had a fair bit of flexibility there in how we wanted to deploy network functions. What happens in a situation like a brownfield network is they already have data centers. They already have fiber laid in the ground. They have agreements for internet peering points. So your peering of the telco network to the internet needs to happen at those points. So that ends up determining where do you want to put which functions. And we created this portfolio with that intent that whatever your deployment requirements or considerations are, you have a choice of AWS infrastructure to host those functions on. So it could be the big regions. It could be our local zones, which are metro level sites. Or it could be these Outpost servers that we announced, rack of servers on your on-prem facilities. Or these can be the far edge for running the RAN DU workloads. So with Telefónica, we're using a combination of these depending on their data center footprint and their internet peering points and where it's most efficient to place these workloads.

DR: [09:45] And so you brought up Dish Wireless. So in a greenfield situation, because nothing's built, they don't have these agreements, there's no fiber in the ground already, are you using more public cloud than say you would at Telefónica because they have the edge data centers already built?

Ishwar Parulkar: [10:02] Yes. We don't call it public cloud versus private cloud. To us, it's all cloud and there's an edge continuum. So when I say Outpost, it's the public cloud on-prem.

DR: [10:12] Yeah, yeah, yeah. It's a nuance.

Ishwar Parulkar: [10:14] It's a nuance. Exactly. But you're right in the sense that when you don't have data center assets, there's more opportunity to look at local zones and regions to host the network, whereas if you actually have a data center, you can use Outposts. And that's one of the reasons we created Outposts was if you already have a data center that you want to utilize, how do you bring in the public cloud into that data center? So one way of looking at these Outposts racks is you're bringing the public cloud onto your premises.

DR: [10:42] Help me understand. Is it a technical reason to make the network run that it needs to be on-premise? Or is it more like, "We have these assets. Let's leverage them because it's more of a commercial reason."?

Ishwar Parulkar: [10:54] No. It's definitely technical. So in addition to whether you have assets or not, the geography and the topology also makes a difference. So if you look at the US, you cannot reach all your users with a certain latency unless you have 20, 30 of these sites where you host the UPF or the data plane of the 5G core. It's just not possible, technically. You need that latency. If you look at a country like, say, Switzerland or Austria, which are smaller, you can use the region to run everything.

DR: [11:25] So it depends?

Ishwar Parulkar: [11:26] Yeah.

DR: [11:27] So this sounds like AWS now has the layers it needs, whether it's brownfield or greenfield, you can support all the network workloads to run a network on the public cloud?

Ishwar Parulkar: [11:36] Absolutely. That's exactly the way to look at it.

DR: [11:38] That's amazing. Okay. Cool. Well, you mentioned obviously one of the big trends you're seeing is generative AI and AI in general. And so what are some of the success stories that your telco customers at AWS are seeing? How are they using AI in innovative ways? I mean, we could talk... That's a whole podcast right there.

Ishwar Parulkar: [11:55] Yeah. Exactly. Or three maybe depending on the categories. One big area is around where genAI can transform the overall customer experience, not only from understanding intent of the customer, but having very interactive conversations with customers to really address their problems, as well as understand them to a level where you can offer personalized services. One of the key successes we've had in the recent past have been SK Telecom, which has built a model trained specifically for Korea and for telco use cases of customer experience that was trained with Anthropic cloud model on AWS. So, that's one example.

[12:33] The second, which I talked about earlier, is using cost of operations, business, as well as the network operations, which is a big piece. For example, with British Telecom, we deployed Amazon Q Developer, which is a genAI application that can generate code with natural language prompts or you have existing code which you need to modify. We're seeing that type of interest with ISVs as well, who actually write software stacks for the different network functions. And lastly, we're looking at how telcos can use genAI to offer new products. And this is an early space. The big thrust right now is looking at how we can personalize some of these offerings tied to understanding the customer better. As telcos get comfortable with the technology, we see some solid proof points and business value emerge, you know, you'll see adoption grow.

DR: [13:23] Are you seeing a correlation between the telcos that are more cloud forward and embracing the cloud earlier with the people who are able to implement AI more quickly? Is there a correlation between that? Or really no? I mean, even if you really haven't started on your journey on cloud, still they're able to use some of your AI tools?

Ishwar Parulkar: [13:41] You can definitely start using AI tools without having hyperscaler partner, but the real benefit, acceleration and scale, comes with partnering with the cloud. If you look at AI, I mean, there's obviously the AI/ML layer on top, but at the foundation

is data. Telcos are very old siloed systems. And that is one of the challenges that they're facing in the adoption of AI, is really getting a better handle on data.

DR: [14:08] For sure. And I think it's even lower level than that, just the access to the right chips. I think the big NVIDIA H200 servers, the line is out the door and wrapped around the block to get access to those machines. But you guys have Inferentia and your special AI chips, and so you're giving the ability for businesses of any size access to the kind of chips and machines they need to use to run these workloads and you pay for it as you use it.

Ishwar Parulkar: [14:34] Absolutely. So the pay as you go model of the cloud works very well for some of these very cost-intensive AI pieces. If you're looking as an example, training a model, this is not something you do all the time and it just doesn't make commercial sense to invest in GPUs that you're not going to use later.

DR: [14:52] Exactly.

Ishwar Parulkar: [14:52] Once the model is trained, you want to move to inference, which requires a very different pattern of using compute. So the cloud model is fundamental to being very efficient with using AI.

DR: [15:03] Yeah. Well, one interesting thing that's come from a chip company, we talked with Chris Penrose from NVIDA about AI RAN and using this extra capacity to run AI workloads at the edge. As we've been discussing, they've invested in these edge data centers and they always have the RAN capacity over-provisioned. And there's some telcos out there, like Verizon, that are starting to talk about monetizing enterprise AI workloads at the edge. And so from where you sit, what's your opinion of this? Do you think this idea will work?

Ishwar Parulkar: [15:34] So to me, not yet completely clear whether the RAN edge makes sense for running any of the AI applications. We've looked at the edge for a long time, and when we talk about the edge, I think it's important to understand the different points in the edge. So there's what I like to call the middle edge or the metro level edge. These are data centers, but smaller that are closer to the users. And there's the far edge, which is the RAN. We don't see at this point any applications that would need to run at that point because it's very close to the devices and a lot of the stuff today devices can do. So it's a question of what is

the right balance between the device where you can run some of the inference today, if you look at your mobile devices or VR glasses, a lot of the actual inference happens on the device itself.

[16:23] So there's the device. Then there's the cloud at the other end where it's relatively much cheaper to run these things because of the scale and utilization and multi-tenancy. And then there's an edge in the middle where it makes sense because either you have latency considerations or the fact that you want to keep data, that you don't want to move the data. We see these three as the key points, the device, the middle edge, which is the on-prem data centers.

DR: [16:48] And the cloud.

Ishwar Parulkar: [16:49] And the cloud. Yeah. So to me, the applications fall in some combination of these three. Not yet completely clear whether the RAN edge makes sense for running any of the AI applications.

DR: [17:02] Yeah. I agree with that. And so this has been an insane time with AI, kind of feels like you're living in the future and living in a movie. And I hear you're a big movie buff. I love movies myself. I used to go to the movies all the time. I watched 86 movies in the theater the year before my son was born. And after he was born, I think I watched like four. And so I really love the theater experience, popcorn, everything about that. And so my question for you, Ishwar, is where's your favorite place to watch a movie? A plane? Telco people are always on planes. A theater? Or are you more of a Netflix and chill kind of guy?

Ishwar Parulkar: [17:38] So I watch movies wherever and whenever I can. So a really huge movie buff myself. But I have a combination of the two, the theater and watching them at home because I have a pretty big home theater with a large screen. I have antique posters. I do the whole thing when we watch movies, popcorn. And it's not just Netflix. I have like 10 streaming platforms.

DR: [17:59] Right. We unbundled cable to all buy it separately and spend way more money.

Ishwar Parulkar: [18:03] Right. But I do love going to the theater as well. I love that experience, just going out, getting popcorn, sitting in a big room with a bunch of people.

DR: [18:10] It's the best.

Ishwar Parulkar: [18:11] Absolutely.

DR: [18:12] Well, Ishwar, this was an awesome conversation, learning about what AWS is doing to support the network workloads and obviously with AI. And so I want to say thank you so much for coming on to the podcast.

Ishwar Parulkar: [18:22] Thanks for having me. It's been a pleasure talking to you.

DR: [18:24] Awesome. Stick around. We end each podcast with a "Telco in 20" takeaway. I've got two minutes to tell you something you need to know. Ishwar and I just explored the AI landscape for telcos and I want to be crystal clear about something. Whether your network runs on the public cloud or not, your AI strategy absolutely should. Here's why. The public cloud gives you access to powerful AI platforms and tools that you can apply to your specific telecom challenges from network optimization to customer experience. What public cloud providers have solved for you are the complex infrastructure challenges that would take your team months or years to figure out. GPU orchestration, model training at scale and prompt optimization frameworks are all ready to go. And while you're waiting six to 12 months for hardware procurement approvals and GPU deliveries, the cloud users are fine-tuning LLMs and deploying solutions today.

[19:25] Remember, AWS, Google and Azure collectively operate the largest AI infrastructure on the planet. They've invested billions and still growing in foundation models, fine-tuning capabilities and inference optimization that you can leverage without building from scratch. The most innovative telcos aren't the ones building their own LLMs. They're the ones who recognize competitive advantage comes from implementing AI solutions faster than everyone else. The public cloud makes that possible. Full stop. You don't need to move your entire network to the cloud to benefit from cloud-based AI. Jumpstart your AI strategy with Totogi's BSS Magic and implement high impact use cases to transform your business.

[20:08] I'll be talking about all this and more at Telecom TV's DSP Leaders World Forum on June 3rd and 4th in Windsor. I'm speaking on Tuesday on the panel "Leveraging AI Throughout The Network," and I'm hosting an amazing garden party Tuesday night. I promise not to cause as much trouble as I did last time.

Until then, tune into more "Telco in 20" episodes, like and follow, and leave us a five star review. Connect with me on LinkedIn and X @TelcoDR. And don't forget to sign up for my awesome email newsletter on TelcoDR.com and be sure to check out our killer YouTube channel. Later, nerds.