Intel

Panel- Agile Network Slicing with AWS, Capgemini Engineering, Cloudify and Intel

CORPORATE PARTICIPANTS

Petar Torre
Intel – Principal Engineer

Sameer Vuyyuru
AWS – Head of WW Business Development for Communication Service Providers

Shamik Mishra
Capgemini Engineering – CTO Connectivity

Nati Shalom
Cloudify – CTO and Founder

Wie Yeang Toh
Intel – Sen. Director Network Platforms Group

PRESENTATION

Petar Torre

Welcome, everyone, to the Intel Network Builders Webinar Program. I am Petar Torre, principal engineer at Intel's communication service provider vertical for telco cloud transformation, and your host and moderator for today's webinar panel. Thank you for taking the time to join us today for our panel titled "Agile first" approach to end-to-end networking slicing with AWS, Capgemini Engineering, Cloudify and Intel.

Before we get started, I want to point out some of the features of the BrightTALK tool that may improve your experience. There is a Questions tab below your viewer. I encourage our live audience to please ask questions at any time and we will try to answer them during the panel. At the end of the presentation, please take the time to provide feedback using the rating tool tab. We value your thoughts and we will use the information to improve our future webinars.

Intel Network Builders Webinar Series takes place twice a month, so check the channel to see what is upcoming and access our growing library of recorded content. In addition to the resources you see here, we also offer a comprehensive NFV and 5G training program through Intel Network Builders University. You can find the link to this program in the Attachments tab, and as well as the link to the Intel Network Builders newsletter.

For the panel today, we are pleased to welcome-- switch to the next slide so we'll see. It's still stuck. So, for the panel today, we are pleased to welcome Sameer Vuyyuru from AWS, Shamik Mishra from Capgemini Engineering, Nati Shalom from Cloudify, and Wei Yeang Toh from Intel.

Sameer Vuyyuru is AWS head of worldwide business development for communication service providers. Sameer is a customer obsessed veteran of the telecommunications industry and prides himself on delivering value through innovation to the telecommunications ecosystem worldwide.

Shamik Mishra is Capgemini Engineering VP and CTO Connectivity. As the global lead for 5G and edge at Capgemini, Shamik drives connectivity industry offers in areas like virtualization, cloud native, edge computing, artificial intelligence, and carrier cloud platforms.

Nati Shalom is the founder and CTO of Cloudify, a serial entrepreneur and thought leader in open source, multi-cloud orchestration, network virtualization, DevOps, and more.

Wie Yeang Toh is senior director for Network Platforms Group at Intel, leading the business strategy and market development organization to support the industry's supply chain and 5G network transformation journey.

Thank you for joining us today. Please do not forget to rate this session at the end. And with that, we will turn it into our panel discussion. And maybe for beginning, as introduction, Nati, if you could please give us a summary of NFV automation, the current status in deployments that you're -seeing, and how should all that improve going forward.

Nati Shalom

Thanks very much, Petar, and glad to be joining this webinar with all these distinguished guests here. With regards to NFV, I think people are familiar with the term. NFV is network function virtualization. It's all started by the motivation to move from hardware appliances into virtual appliances, and today, we're really talking about moving into a completely different game, which is cloud.

5G is becoming a major category around this. The reason why is because 5G, unlike, I would say, its previous generation, it's a much more sophisticated network and much more, therefore, complex. Network slicing is one use case, I think, that represents the level of sophistication, but also complexity that is needed for that platform, and that drives, really, for a lot of automation. And when we look at that from an automation perspective, we really see a move of the entire NFV towards something that is more cloud native as a result. So, the next-generation NFV would be very cloud native-centric.

The other thing that I think is interesting from, I would say, a more high-level perspective is the thing that we're seeing is a strategic partnership between telcos and public cloud being also driven to that, and reason why is because I think, also, telcos realize that to really deploy effectively a 5G network, you have to really collaborate. You cannot just build on your own local network and you have to think more globally, and this is where the partnership with public cloud providers is becoming more essential. You also need to be able to integrate 5G with the rest of the ecosystem. It's not a siloed network anymore and there's a lot of talk about connecting it to IoT and other use cases, including private 5G, which also brings-- and that's another segment. It also brings new players into the market, start-ups, and so forth.

The last thing that I wanted to mention is with regards to the question of what needs to be improved. So, those who've been active in the telco space and seen, you know, other projects like ONAP and other Linux Foundation projects that try to adopt a lot of those new concepts. A lot of those discussions were very much API and standard driven, and as a result, a lot of the operational aspects were kind of left as an implementation detail, kind of left behind to the vendors or to the operators to actually deal with. And that led to a lot of frustration because at the end of the day, what you're doing on a day-to-day type of basis, you're not really changing APIs. You're really operating, managing change, and managing the environment itself, and that's why there's been a lot of lag in tooling and a lot of lag in standards, and that's kind of where we came in into this "agile first" approach that I believed aimed to change some of that dynamic.

That's pretty much it. Petar, back to you.

Petar Torre

Or to comments from other panelists, please, on it.

Nati Shalom

Oh, sure.

Petar Torre

Or we can also ask something related that is relevant here. So, how frequently do we see changes into the environment that is already deployed, like configurations or new network slices, and how often do we see new network functions getting deployed?

Sameer Vuyyuru

Let me take that, Petar. So, over a decade ago, when we started on the journey with AWS, cloud essentially took an expensive, inflexible, complex IT industry and turned it on its head. It put the power of the world's most advanced technology into the hands of developers

and leveled the playing field for all organizations to have access to the same level of technology, right. Now, there's transformational business innovation possible by the development of the industry's first set of cloud native 5G networks, some of which have been announced already.

With cloud-based 5G networks, we're not only help telcos deploy networks faster than ever, but we're also ensuring that these networks can offer richer connectivity services on-demand to the end user. Building the cloud native 5G network with an agile outcome in mind is the first step to offer an intent-based end customer service experience, on-demand and in real-time. For instance, highly latency-sensitive use cases such as remote surgery would benefit by being able to pair AI/ML, IoT, robotics, with a low latency, high bandwidth 5G network slice, provisioned as and when required for the actual surgery.

With an entire 5G network in the cloud, those capabilities are integrated into the 5G network, and customers can quickly provision those capabilities through a dedicated slice, and then turn them off after the application is done.

So, to summarize, the application is going to drive how often the RAN configurations and the core configurations are going to change and, what slices get deployed. They're going to be provisioned within seconds and pulled down within seconds, and you will redefine how customers order and consume these 5G services. It'll allow telcos to scale network resources up or down on-demand to meet customer requirements, and simplify the process for developers to create new 5G applications.

I hope that answers your question. Essentially, it's on-demand, immediate, as many times as it's needed.

Petar Torre

Any comments, please, from other panelists.

Shamik Mishra

Well, just to add to that, perhaps, I think, in general, the overall agility of the network needs to be-- that needs to be better with different kinds of use cases coming in. If you look at the 5G today, it has to be driven through use cases. I mean, you cannot possibly make money out of 5G without use cases, and these use cases are basically cloud native applications.

The cloud native applications today are developed by the developer community who are used to using hyperscaler platforms, and for them, it's quite natural to look for a platform or an alternative within the telcos space, which gives the same kind of tooling and experience which they are normally used to. And in order to do that, I think it's a natural choice for telcos to go towards a common platform that can host a RAN, a core, and an edge, and applications for 5G.

So, the baseline for a platform is kind of changing today. The baseline is becoming more cloud native, more developer-centric, and just because it might be too expensive to have different kinds of platforms for running different kinds of use cases. Now, it would be wonderful, at some point, if the same platform can run latency sensitive, highly performant radio network, and at the same platform can also run an application like a cloud gaming. Now, if that kind of a situation happens, then I think, from an economics point of view, it would make total sense for a telco, for a carrier, or even private networks, because you have a common baseline for everything. The tooling, the automation, the operations, the whole-- the whole upgrades of software, the DevOps, and the pipelines, all can be common across all kinds of applications and workflows.

So, that's-- from a Capgemini Engineering perspective, we're a strong believer that there is a need to have, or at least move towards, common platforms in the short-term.

Nati Shalom

And I wanted to maybe add what is called today private 5G, it's something that I think is really shifting a lot of things. For those who are less familiar with the term, it's really-- think about it as an extended Wi-Fi, or a better Wi-Fi with a more guaranteed SLA, with a wider range of coverage that organizations, especially if you look at that in manufacturing facilities and other things, especially today, when

you're using robots, and when you're using a lot of IoT connected devices, if you plug them into Wi-Fi, you're not necessarily guaranteed to have network access across the entire area, and therefore, having a private 5G becomes very sensible and meaningful.

Now, the reason why I mentioned that, it's also changing the entire way we think about network and carriers network, because we don't necessarily need to have this service being delivered by carriers themselves. And what we're seeing is, I think that's where I believe a lot of the public cloud providers will play a very important role, and I believe start-ups will enter into this game, so we're going to see a lot of that interesting dynamic.

So, 5G is not just about adding applications into the network, and changing the underlying architecture into cloud native. It's really changing the game in many ways. It's changing the landscape and I strongly believe that 6G will look very different in terms of the ecosystem and players than 5G, and if we kind of move a little bit faster in that regard.

Petar Torre

How does all that relate to DevOps? Is this then—

Shamik Mishra

Yeah, I mean, it is actually. If you look at the overall picture, as I mentioned, having common platforms is critical to the overall ecosystem, and that would mean that we need more agility in the network in the way workloads are deployed. If you can treat a radio network workload in the same way as an application, like a cloud native application, that would be wonderful, but we can't do that today, primarily because there are too many different regulatory, as well as processes involved.

So, DevOps primarily, it was about integrating or converging operations and developers, in some ways. That was the original idea behind DevOps. If you can actually get the network operations of, say, a telco or a private network of an enterprise integrated with the developer ecosystem, I think that would be the real solution or the real innovation going forward.

Now, in order to do that, there would be a requirement to adopt new kinds of toolchains, new kinds of operational models, automation models, how to upgrade software. I mean, can we upgrade the software in the same way as we upgrade an application software in the cloud? Can we upgrade a RAN software, for example? As of today, no. I mean, there's too many different complexities involved in it.

And also if you look from an operational standpoint, more and more in disaggregated RAN, particularly, there would be a huge amount of edge computes globally. At the same time, there would be-- even for edge computes, where applications are running, there would be many such edge computes. The far edge hosting RANs would be huge, or many in number. So, how do we manage all of this as a managed services provider or a system integrator without having a DevOps? I mean, it’s going to be a humongous task, next to impossible, if there is no way of automating all of the operations across all these different locations.

And remember, when we go to network slicing and private networks, the question of tenancy comes in, multiple tenants working across multiple ledgers. That itself is a hard problem, and without an automation, it's going to be even harder. So, to put all of that in perspective, DevOps is required. Maybe it will evolve into a new kind of DevOps going forward, a new kind of innovation will come there, but definitely some kind of a model is required to handle all of these challenges.

Wie Yeang Toh

Yeah. just to add on top of that, right, the DevOps we can look at-- almost most of the company, they are implementing DevOps right now, right, and really depending, are you looking at your own in-house software developments to comply with DevOps cycles, or from a telco perspective, they look at different layers, which layer they decided to separate it out to implement DevOps. So, almost most of the companies in the industry is self-implementing DevOps, and heading towards the right direction. For mission critical, say a network function, like RAN would still take some time, right, but yes, it's part of the process, and telco itself will decide which layer really makes sense to suit them the best in order to move faster. So, that helped them not to slow them down, yes. So, today, DevOps is critical.

Petar Torre

Could you maybe also comment on, Wie Yeang, how did your teams work with major communication service providers in rolling out commercial vRAN? I mean, what were the learnings there?

Wie Yeang Toh

Oh, yes. So, the... we have made tremendous progress with the major telco and ecosystem, right, over the past few years in rolling up the vRAN, right, for example, say-- I'll just they just throw a couple of examples out there how we partnered with the fellow telcos. Rakuten Mobile is one of the most noticeable ones in the market. They've been using a proven cloud technology and underpinning their network with Intel's solution technology. Things range from FlexRANs, Xeon series processors, their FPGA software toolkits to architect their network rollout. Rakuten has shared publicly about their solution that's building based on a virtualized, fully virtualized, cloud native solution that leads them towards a pretty significant economies of scale to keep their capitals, CapEx and OpEx, right, pretty effective.

Looking at back in the home countries of Verizon, say, we've be working with Verizon, Samsung, Wind River to successfully complete the world-first virtualized end-to-end 5G data session to get established. This is last year to enable a faster response to customer latency and computing. So, I can see around the advertisement right now, Verizon has claimed the fastest 5G network supported by third-parties' research.

We also partner with Deutsche Telekom in Germany, which we started off the xRAN with them together, and later on, on O-RAN, by working with, of course, the industrial partner, together, to deploy a complex RAN solution based on the common off-the-shelf, say, a platform virtualization technology, and recently Deutsche announced this O-RAN Town, which is very interesting. They selected a town in Germany. With live networks, they are deploying using Open RAN and virtualized RAN, and underpinned technologies out there coming from Intel, and with our partner, of course, right?

Lastly, Reliance Jio a pretty interesting one in India, so our partnership focusing on supporting the network transformation from call to access, right, as well as the co-innovation not only at the radio, as well as the edge computing, AI, right. This will help the whole vision of the Digital India, right? So, the momentum is very strong, and we are humbled as a trusted partner with the telco, fellow telco, and ecosystem partner to move vRAN forward with more agile abilities as the Intel technology that underpins the solution that we're offering.

Petar Torre

Nice, and Nati, your team's also worked over years on that in this type of layered stack, multi-vendor solution. So, what are those, some of those learnings?

Nati Shalom

Yeah, I think multi-vendors, there are a couple of areas where you could talk about multi-vendors. We can talk about multi-vendors in the context of the CNF, like the container network functions. Also, the 5G core multi-vendor in the context of the infrastructure itself, et cetera, and the reason why it is important is because, especially when you're developing a software solution like ourselves, you want to be able to make sure that the platform itself, or in our case an orchestration, can play with a multi-stack type of solution.

The work that we've done with the AWS and Capgemini, and Intel yourself involved, kind of took on interesting phases, which proved what that actually means in reality. In this case, we had to be able to switch between one commercial vendor that was using one carrier, and then for demonstration, we needed to use a more open source alternative, in this case Open5GS, and to do that, we have to switch the core, 5G core. And what we found is that because we had an abstraction layer on how we interact with the 5G core using the CI/CD pipeline, the Amazon CI/CD pipeline, changing the 5G core didn't really mean changing the entire orchestration scheme, and we were able to actually do that within a couple of weeks.

The other thing that we added is when we introduced FlexRAN, the Intel FlexRAN, and when we introduced, obviously, the Capgemini implementation of that, plugging a vendor like yourself and Capgemini into this was also very easy. Why is that? For two things. One of them is because we're using a more standard stack, like cloud native. So, for you to plug in your software, you didn't have to know much about our platform, or much about the Kubernetes cluster, because you already built your software around it. So, that in itself simplified the ability of vendors to play into this ecosystem.

The second part was the use of Amazon as a shared environment. We're actually a global development team. We're based in Israel. We have a team also in the US. There was a team involved in India, and obviously, Europe, all of them working collaboratively around the project. And because everyone had the same access to the AWS environment, it was much easier to do the integration each on its pace, each on its own time. You're working on the FlexRAN integration, and god knows, you had quite an amount of work to actually integrate that. Within a couple of weeks, we were able to integrate all of those pieces without too much engineering or work, to be honest. And a lot of that had to do because there is much more consistency across the layers, possibly because of cloud native or because we could use both in Outpost and the core network, the SIEM, in this case, infrastructure which was, in this specific case, AWS. And both because we are using standard tooling, not something that is very telco specific.

And that's another very important thing. We talked about DevOps, but that's another important thing is the skillset. So, getting a developer to work against a telco standard, it's much harder than finding someone in the street that already knows how to work with DevOps tools and the AWS stack and the Kubernetes stack.

And that, in itself, enabled us to bring people that were not really that savvy about 5G, and they were able to be up and running very shortly and delivering a very complex use case in a very short amount of time.

I hope that answers your question, Petar.

Petar Torre

It does very nicely. Any comments from the team?

Sameer Vuyyuru

I'd be happy to go, Petar, right. I mean, at AWS, you know, 90% of what we build is what customers have explicitly asked us to build. And so, what are they asking for? What are my CSP customers asking us for?

The single biggest ask is agility, right. If you think about traditional telco networks at a very simple level, there's a lot of infrastructure and hardware to deploy and manage. Now, how do you abstract away all that complexity and put it in the hands of any developer anywhere with the right permissions to manage the network?

The second one is cost. How do you reduce the cost of deploying a network and how do you maximize a network once it's deployed?

The third is the ability to innovate, which is what are the over-the-top applications, the line of business applications? What is the actual network level innovation that you can do?

The fourth one is elasticity. I mean, we were all part of network builds in our past life, and you always provision for peak, right, like for Mother's Day when everyone was calling their mother. But on-demand elasticity, you don't really have to worry about provisioning for peak, you provision for a capacity that you need.

And finally, global reach. I think our CSPs now want to expand just beyond their home country and really use the same infrastructure, the same manageability, whether it's across SD-WAN or across 5G or other services that they offer wherever their customers go. And their customers are global.

So, those are the five things that our telco customers have asked us, and that's what we're laser focused on delivering. And, yes, just wanted to complement Nati's take on this.

Petar Torre

Nice. Could you also share – try to quantify this agility, maybe some numbers, examples?

Sameer Vuyyuru

No, sure. Agility is not just in the network is something we've got to understand, right. It's across the whole value chain of a telco. Whether it's customer service where we work, for example, with T-Mobile in terms of reducing their call resolution time by 30%, just by deploying customer care on AWS. Or to provision the network, which is what we're doing with a cloud native provider that we announced a few months ago. I will not go into too much detail on that one, I guess because it's – it's a little confidential. But suffice to say that we are able to bring up and instantiate a network about 10 times faster than it did in the past.

For example, you heard Nati talk about AWS Outpost, right. Outpost extends AWS infrastructure, services, APIs, tools to virtually any location and it's built and is well-suited for deploying and managing tens of thousands of cell sites across a RAN. And to Wie Yeang's point, you can run an application on an Outpost in parallel with a RAN today, right. So, there's the agility right there in terms of being able to say, "Look, this slice of the network is serving a manufacturing customer so I'm going to do a manufacturing IoT application, and this slice of the network is doing a mobility tracking application, so I'm going to use an IoT service", and so on.

And then we've also got the AWS Local Zones, which are well-suited for deploying and managing the network core, which is, you know, as you know, responsible for actual user traffic routing and the virtual control unit of the RAN. And of course, you always have the regions that provide the necessary computes to a range of supporting services.

So, in terms of quantification, we have literally deployed more than 100 private networks today, to Nati's point earlier, with the line of business applications running on these networks. And customers can scale this up and down on-demand. These are not – these are real breathable networks. These are not fixed capacity networks. And new services can be provisioned within minutes, literally within minutes, and customers can self-service, which means there is no call into a network operations center. They go on the AWS console and they provision both the network and the line of business applications that they do today.

Nati Shalom

Thanks, Shamik, and Sameer. I think it's also worth mentioning the fact that you actually – with all that – opened the door for developers to actually deploy this new application. Because without it... like if you look at any operator, they all work very hard to say, "How do we open our network to developers?" and fail, basically, because it's very hard to do it when it's a private network and the access to that is very specific, and each carrier would have its own tricks and whatever. So, why would a developer worry about, you know, putting the effort of getting there. But once you plug it into more, if you like, commonly used environment and ecosystem that already have the developers plugged into it, the ability to do, I think, what Shamik you mentioned, which is really getting applications deployed and developers accessing the network, which is almost unheard of, like even a year or two years ago. That's becoming critical. And I don't think that you could quantify it yet, Petar, but in my view, I think it's very easy to see the dollar value that you could get out of that, even if we can't really put a number yet.

And maybe, Shamik, you could say a few words about it.

Shamik Mishra

Yes, absolutely. I mean, it's hard to give a dollar value to it right now, but definitely what it does mean is that the whole application ecosystem, as we know, has evolved quite much in the last year or so. And... I mean, if you go back two to three years ago, application developers, and if you would have asked them that you would have to deploy this on a carrier network and it would be an absolute no-no for them. There are many reasons for that, and one of the primary reasons for that is that the telco network is extremely complex. The orchestration is complex. It takes days, if not months, and sometimes quarters to get anything to run on a telco network, primarily

because there are processes involved in it. And there are certain lack of automation was there earlier. And now, these things are changing today.

If we expect a developer in an AWS environment to just go in there, create a virtual machine, and install their application and run everything by themselves in seconds or maybe in minutes, reaching that stage for a telco network is still going to take time. But I think unless we move towards that direction, attracting developers to build applications for 5G private networks, or edge computing will not be easy.

So, the focus has to be entirely on how do we simplify the developer experience, and that's absolutely critical going forward.

Petar Torre

And maybe if you could comment also on this journey. How many standards are needed? Which organizations? How is O-RAN going to continue?

Shamik Mishra

That's a trick question. I mean, there are a few standards... I mean, I believe that whether we call it standards or ecosystems, there would be certain things that definitely will be required, and I'll give some examples starting with applications.

So, if an application developer has to build an application for Operator A who uses a Platform X, and has a completely... a kind of API ecosystem around it. The same application developer goes to another operator, and who has a completely different stack and a different set of APIs to manage the applications or launch the applications. Then you don't expect the application developer to maintain different versions of the same software just to manage multiple operators. I mean, it's not possible. It's technically, or theoretically, it's a nightmare, right.

If I'm building a cloud gaming application and there are 25 operators, do I maintain 25 different versions of the same software just because the operators have different platforms and different APIs? I might as well run this on AWS, right. I mean, that's a common set of APIs and platforms.

So, definitely the operators will have to look to build a common set of APIs at least for the developer-facing services. And in the GSMA Operator Platform Group, for example, where I happen to co-chair, we are trying to build a common set of minimum APIs, which an application developer can leverage, so that he doesn't have to continuously build applications or different versions of the same applications to get it built across different operators.

There is a role of standards, but whether we call it standards or just API ecosystem, it doesn't matter. But there must be some basic guidelines and common agreements across operators to build APIs as far as applications are concerned.

But when we move deeper into the network and go towards the radio network side, or to the core network side, standards are unavoidable. I mean, you cannot possibly have a disaggregated radio network where there is a different container-as-a-service platform, there is a different orchestrator, there is a different radio unit, the CU/DU is from a different vendor, there are six different, you know – or a plethora of vendors available for a certain network. And a system integrator like us, we need to integrate all of that. Without standards, it's going to be a nightmare, because there will not be interoperability, it will be impossible to pinpoint where the problem is. It is impossible to build some kind of an automation.

So, definitely for network functions we need standards. And I think O-RAN Alliance is doing a good job in getting that started. Now, of course, it's still work in progress, there are a lot of stuff that needs to be done still. But if we have to tackle the elephant in the room, that is really your network virtualization, we need to have standards for sure.

Now, how that would evolve over a period of time is still open, in my opinion, but it's unavoidable.

Petar Torre

So, maybe, Shamik, a question on – with regards to standards. If we look at the cloud scale kind of world, let's look at CDN like Content Delivery Network as an example. There's not a specific standard there and still the industry we're able to deliver a global service without necessarily agreeing on a common API. So, why do you think that we have to have standards to manage a radio network? Why is radio network different than CDN?

Shamik Mishra

Well, I would say it is in the sense that when you are talking about integrating multi-vendor and disaggregated networks, I think it's required to have standards, because you cannot possibly have protocol terminations, which are split across two different boxes to talk to each other without a common language. I mean, if... we both know English, so we talk. At the same time, I think, in the network – in the network itself, protocols play a big role, and that's the only reason I would say.

Had it been the same vendor providing everything right from the antenna right up to the orchestration, there would have been absolutely no need for a standard. But if we want to go and build a disaggregated network where many different vendors will come in, software companies can enter the market, which so far has been dominated by large network equipment providers, we'll have to have some kind of standards in order to make these companies to enter the market.

And these standards don't need to be hard standards like 3GPP, these can be APIs, simple DevCloud native, RESTful APIs, which can be easily defined and then people can start to use that to integrate.

On the other hand, the protocol part, I think, is very important to have standards. I mean, you cannot possibly split a protocol into two parts and expect them to talk to each other without a common language.

So, some part of the standards are required, but if you talk specifically about orchestration, I think it's going to be cloud native, and that's the de facto standard, so Kubernetes or whatever you call it. But within the protocol, if you go deep inside the network functions, some standards will be required.

Wie Yeang Toh

And even for CDN, right, you have Open Caching API as well, which is a standard working group, right, in between, right, but they still need a form of API agreement, Open Caching API, just the classic one that industries believe in right now moving forward. Come to RAN, it's even more complex, right. It's low latency, it's a high reliability, so some form of standard will require to guardrail the interfaces, right, to make sure the latency was able to well cap-- to be able to deliver as it promised, right. So, a mix of standard working groups have to happen. But, yes, I still believe some standards will need to keep things intact, right, and not to slow down the industry, but bring more innovation. Yes, yes, so good question, right, but I just want to point out even for CDN, you have Open Caching API that the industry is coming on strong, saying, hey, let's move that forward, yeah. It's about the API.

Sameer Vuyyuru

So, you're basically saying it's a unified – a uniform programming environment as opposed to a codified standard.

Wie Yeang Toh

Yes.

Sameer Vuyyuru

I think there's a very big difference between the two, right.

Wie Yeang Toh

Yes, you need to have a mix there. That's my point. The standards, you still need it for interfaces, latency, protocol, that stuff, right, and the upper layers, right. You go for – often go for agreeable API and a couple of the key vendors or key drivers in the market will make it happen, the rest will come and follow, right, yeah. It will be a mix, of course, depending on which layer they start drafting, from here above, let's go and work on it, and then from here downwards, let's keep it as a standard, because it is very critical for multi-vendors interoperability.

Petar Torre

Could you, Wie Yeang, comment on the Intel options and support offer for vRAN implementation?

Wie Yeang Toh

For vRAN? Yes, so if I summarize it, we bring three key things for telco ecosystem, right, from Intel perspective, right, over the actual experience here.

First, we offer a full suite of feature rich silicon, optimized software, and tools, right, those are like the bread and butter that Intel brings, right, often as it underpins their technology.

Second, right, we have fostered a merchant developer ecosystem that's constantly optimizing and innovating for our fellow telcos, yes, customers based on the common architecture that we deliver, the first thing that I talk about, right, and the common software tool suite in an optimized way.

Third, we bring the experience in helping customers to transform their network, right, over the past decade as well, right. It started off from NFV, right, as part of the co-founders there, right. It is important to have that consistency where this thing gets started, right, and where are we heading next, right.

So, to accomplish those three things, recently we did announce an expanded line up of the platform and software offering across the network infra to focus at the agility, right. For example, the latest, Xeon Scalable to SOC, right, to allow our customers to have a common, say, architecture across various workloads to give them, ultimately, scalability. We also have a frequent update of the FlexRAN software stack, right, that with more than 100 licenses out there to enhance the optimization on massive MIMO wireless pipeline UI and so on. This goes beyond CPU.

And then we also put the flexibility back to the customer, so that they are able to choose if they need to have the most compute-intensive processors pipelines run on an accelerator, right, we do have a dedicated accelerator option, ranging from the FPGA to custom ASIC, right, to offload the specific function that they wish to take it out from the general processing, say, platform, right, in order to increase their agility for the edge services.

So, all in all, it is about the platform, it's about the optimized software, it's about when to accelerate, right, the interconnect technology, the tools, as well as the merchant ecosystem knowhow. All those is the strong value of the consistency that we're always committed, right, to bring different options for our customers, right. And the same thing applies for vRAN, it's even more complex than the classic NFV.

Petar Torre

Great. And Nati, you were mentioning this practical integration work. Could you describe what were parts of it and what was the demo stack?

Nati Shalom

Oh yes, of course. So, we actually started the journey back in December last year. That was based on a specific carrier use case. And we were called to be able to orchestrate this architecture, I think, originally put by the Amazon Professional Services team for that specific carrier, which was based on EKS on both the core and the edge, and obviously, Outpost as the edge platform and pipeline is the – I

would say the… taking part of the orchestration and Lambda services and abstraction – API abstraction layer that was called to interface and mediate the API into that.

So, our first role was really to deliver an end-to-end automation that will provide a much simpler way to integrate with those interfaces. And the challenge was how do we plug in into something that was already developed, and not coming and say, "OK, let's rip and replace everything and just plug in to that". So, I think the use of those Lambda services and pipeline was an easy plug in for us, starting point. And I think later, as I mentioned earlier, we added the ability to switch, that was for an MWC demo and for other demos that we took from that experience into other carriers. And that became the multi-vendor approach that we added into that.

And I think towards MWC, we joined forces with Intel, you specifically, and Capgemini and that context, we wanted to demonstrate not just how do we deliver core network and automate all of that, but also do an end-to-end network slicing between core and edge. And for that, we really needed everything to be cloud native. Obviously, FlexRAN fit well into that mix, and with the help of Capgemini, we could cover not just the DU side, but also the CU side, which is the central part. And I think it took some time to get it to the point where we can also run it in an environment that is not necessarily more developer-oriented, I would say, and that enabled us to run FlexRAN also on a standard, if you like, EKS deployments and do the integration very quickly into that.

So, I think the nice thing about all this architecture is that it really provided very fully featured network slicing, which is considerably a very complex use case in terms of orchestration, for those who are less familiar with that. And doing it in a multi-vendor type of landscape, doing it globally, doing it across different providers and different partners, in this case. Being able to join forces in such a short amount of time, for me, was even more exciting than the use case itself of network slicing. Because I've been in different POCs and experiences with other carriers where we've been using this type of architecture and demonstrating network slicing in, I would say, traditional carrier kind of network and that was months and months of experience. And we didn't get a fraction of what we were able to achieve here.

And that, for me, was kind of the click of the moment where I said, "OK, there is something here that is much bigger than network slicing". And that's where the term "Agile First Approach" came into play, and we said, "OK, that's it". And it's, interestingly enough, not necessarily specific to network slicing on 5G. It can be applied to any edge use case. And that's another, I think, interesting part of that. You can easily extend that same architecture and apply it to a very rich set of use cases, almost the same use cases or the same building blocks that we're using here.

Petar Torre

Where is this heading directionally in next steps?

Nati Shalom

So, I mentioned a couple of things here in terms of the technical features that we added. I think that we've proven that this architecture can actually deliver its promise in terms of work. Right now, we're working on extending the ecosystem behind it. Obviously, Intel and Capgemini has been the first to actually join it and join forces around that, and being here on a panel together is, in itself, a milestone that I think is very important. But we're now also extending it to other vendors. I can't really mention their names, because I didn't ask for permission, so I'm not sure, but I can say that we're already demonstrating other vendors. I think some of them have been mentioned running in large production 5G networks already, supporting this type of architecture, and some of them, you know, covering the OSS side of the spectrum, some of them covering the edge side of the spectrum. Both of them, I think, are partnering with you guys in different shape and forms already.

So, for me, it's really to expand this reference architecture and expanding the ecosystem behind it, and the integration that we can deliver through that, both with edge, I would say, CNF edge type of use case, but also the OSS of the use case that would be the main thing that we want to deliver here. And obviously, get more carries around it, so that would be the ultimate goal.

Petar Torre

Nice. Let's touch on one of the topics, which is hardware acceleration needed in some of these use cases. And like keep it agile and then have hardware accelerated container-as-a-service.

Shamik Mishra

Yes, maybe I'll take that. So, from an overall application standpoint, particularly if you're looking at CU/DU as an application in itself, then it has to be cloud native in the first place. So, we are actually looking at a microservice-based approach there. But at the same time, there would be some level of hardware acceleration required in order to achieve performance.

For example, the Layer 1 of the GU would have to perhaps run on some kind of a hardware accelerator, in this case it's based on the Intel FPGAs on top – where the Intel FlexRAN software runs on top of it.

So, hardware accelerator is essential to have the right kind of performance required for a radio network, for example. At times, even in application space on the edge side. For example, if you're running an augmented reality or a virtual reality application or even edge analytics, we also need hardware acceleration, for example, GPUs.

So, hardware acceleration is required. Different kinds of use cases are required. And it's not anything special. I mean it is what is used to be called earlier, or what we used to call as Commodity Hardware. FPGAs and GPUs are now part of that commodity hardware ecosystem, in my opinion, at least in terms of software development. Most of these hardware accelerators today have Kubernetes operators, so you can actually integrate this quite easily with Kubernetes operators. There are projects in open source which actually enable such build or development of Kubernetes operators to ensure that the hardware accelerators are accessible by cloud native workloads. Including a project that we are running as Capgemini Engineering called Adrenaline which is actually enabling hardware acceleration for different kinds of Kubernetes workloads, not just FPGAs and GPUs, but when we want to go beyond that, at some stage.

So, the point of having a hardware accelerator is important for performance, but at the same time, it's ability to integrate with the Kubernetes ecosystem is very important, and I think we are already there. So, now, we can benefit the – or benefit from the agility provided by microservices architectures and cloud native architectures and orchestration that Kubernetes kind of ecosystem brings onto the table. And also, benefit from the hardware acceleration that FPGAs bring for, say, a DU workload.

So, to me, it's quite good. It's a well-functioning ecosystem. And as and when more and more hardware accelerators gets built for such development, I think it will be easier to on-board them on a cloud native world.

Petar Torre

And then Wie Yeang, from Intel there are hardware accelerators, if you could please comment on some of those and where we are heading with all that.

Wie Yeang Toh

So, just a quick one, right. It was well covered just now, but just real quick, right, it is all about flexibility and consistency, right, which matters for the customer. As I mentioned just now, right, the high performance Xeon platform itself today, it offers the – it brings the consistency and the flexibility to the network infra, right, along with the FlexRAN as a software development environment. It allows customers to have the development environment to go and work on the agility that they need. When it comes to certain processing, say, capabilities that they wish to offload, right, they would like to have the control back in their hands, where we give them, ultimately, the scalability option and choice, right. So, you have things that range from, again, FPGA-based or the ASIC-based of the accelerator they can choose from, primarily is to focusing at a very specific function that the end user or ecosystem would love to have acceleration to come in, in order to save the precious computing capability especially running at the edge, right, for other more mission critical, say, new services provisioning.

So, we do offer the flexibility, right. I think it's important about flexibility and consistency just a closer look.

Nati Shalom

Maybe just on that, let's talk about it in terms of developer aspect, because some developers would hear "hardware acceleration" and would be scared of it.

I think the important thing is that I think this is where Kubernetes plugs into is that when you're talking about hardware acceleration in the context of Kubernetes, it means that from an application perspective, you could still run without it, you can still run natively on Kubernetes and if you don't have the hardware acceleration, it would just function less well, or less performance. And that's a big difference, because I think when we're running an open stack and we're running other types of platforms, we have to write the software differently to actually use those accelerators. And that's very different when it comes to Kubernetes, because it has this layer. And I think Shamik you mentioned it, has this layer of abstraction that is much more granular, that it enables the hardware vendor to provide those accelerations, but in a way that is transparent to the actual workload itself. Even in this case, we actually experienced that we could use FlexRAN in a commodity type of environment and still function and do the API integration. But obviously if we need to run it in production, we would need to run that same thing on environments that do provide the hardware acceleration.

But the difference between the environment wasn't that we couldn't really run it, it was mostly the behavior would be different in terms of SLA, that's really the difference.

Petar Torre

Nice. And then some of the next topics, maybe Sameer if you'd like to share where AWS is going with some of these machine learning toolkits and some of these related topics.

Sameer Vuyyuru

Yes, absolutely, Petar. So, at AWS we believe that in the fullness of time, virtually every application will be infused with ML and AI, right. Today, tens of thousands of customers are running machine learning on AWS, you know, really spurred by the broad adoption of Amazon SageMaker.

But in the context of a telco network, and the whole telco ecosystem, what I'd really like us to view the network and the ecosystem is as a set of applications, where the RAN is essentially an application, the core is essentially an application, the BSS is an application, the OSS, the customer care, everything, right.

So, if you take that lens and we say that every application will be infused with ML and AI, that's really our vision.

And so, if you look at what we offer, we offer really three sets of tooling and capabilities for the network to take advantage of the ML/AI capabilities, but with the extra processing power that our partners at Intel offer, for example.

So, the bottom layer is for expert machine learning practitioners, advanced developers, data scientists who are comfortable building, training, managing models and working at the framework level. And these are all for the major frameworks, right, TensorFlow, MXNet, PyTorch, Caffe and all of that. 90% of the data scientists use multiple frameworks because different frameworks are better for different types of workloads.

Today, TensorFlow has the most resonance, and about 80, 85% of the TensorFlow workloads running in the cloud today run on AWS. But Inference is also a very big part of it, right, which makes up about 90% of ML costs. And people use Elastic Inference to save up to 75% on those costs by adding a little acceleration to any EC2 instance, right.

But if you want the machine learning to be as expansive as we really want it to be, the middle layer of the stack is where we make it accessible to people who aren't machine learning practitioners, and most of our CSPs today use SageMaker. It's a fully managed service that removes the heavy lifting, the complexity, and the guesswork from each step of the machine learning process.

So, if you look at how our CSPs are using it today, most of them have, honestly, started on the customer experience side, the customer care side of things where they have learnt how to use machine learning to anticipate fraud, to anticipate churn, to reduce customer call times and so on. But the next step is to take all of the learnings from that and basically flow it back into the network so that the network can essentially self-adapt, learn, and pre-provision.

For example, you could go – pre-provision for an autonomous driving route that is highly used by doing a predictive quality of service model with SageMaker, which is something some of our customers have done already, right.

And then, finally, at the top layer of the stack we have AI services that are fully managed and I think this is where, again, in the fullness of time, our telco customers will end up where there is a fully managed AI/ML service provider that is most likely integrated with the orchestration environment for people like Cloudify, and that combination of agility driven by these models that get tuned over time is where we're headed with our customers.

I'd love for others to chime in.

Petar Torre

That would be fantastic, but we are running out of time. So, thanks for this last very detailed comments.

So, with this, we came to the end of our nice discussion, and here on this last slide, you can see the list of contacts per company. If you would like to have any types of follow-up, please use that. So, thanks to our panelists. Thank you for – thank you all for joining today and please do not forget to give our team a rating for this session, so that we can continuously improve the quality of our webinars.

And again, thank you for joining and this concludes our webinar today.

Sameer Vuyyuru

Thank you for having us.

Wie Yeang Toh

Thank you, Petar.

Shamik Mishra

Thank you for having us.

Wie Yeang Toh

Thank you all, yes.