accenture

The Best Run Cloud: SAP's Approach on Cloud Infrastructure

Ozren Kopajtic | SAP

Ozren Kopajtic

Just very briefly, my name is Ozren. I'm working with SAP. I'm running a unit called the Global Cloud Services. You're rolling into our CTO, Jürgen Müller. You probably never heard of it. You heard about our business application and stuff like that.

We are the largest unit and largest division of the company of SAP in terms of our PNL. More than 2 billion is the size of the budget or PNL we are managing. The reason for that is because we host, all of the SAP cloud infrastructure like flows through this unit. We are custodian of the infrastructure for the SAP.

We also run a number of other central capabilities the company needs to deliver cloud services such as observability, such as aspects of security, such as additional central services, such as service management, so topics around incident management and stuff like that. They centralise with us. We're in a bit of unusual position because we have a unique view on how SAP is delivering our cloud services.

I'll talk about it a bit today. It's not only about infrastructure. It can be a bit unusual when somebody from SAP goes and talks about infrastructure. But I'll come to explain why we are talking about it. Why is it so consequential?

One thing to understand about SAP's cloud delivery strategy is the following. Of course, we talk about infrastructure in a minute, but fundamentally, the baseline of the company is what we call BTP, Business Decimation Platform. It's like SAP's platform as a service capability, which is supposed to augment our cloud services and cloud capabilities in general.

It is a large revenue generating unit of the company itself. But fundamentally, the reason why SAP invested into BTP is twofold. One is that when you're delivering cloud services, you're almost in a space of manufacturing. When you're doing manufacturing, you're deciding how much you're going to vertically integrate, how much of the cloud delivery chain you're going to capture.

It's like a technical level. You are deciding, am I going to go to maybe a hyperscaler and use a high-order API from there, or a managed service from there, or am I going to build it in branded service myself, because I have the scale, I have the skill, and I'm going to then capture the margin? Because when you're running on hyperscalers, they like making money as well, and they like making profit as well, and we like making profit, and sometimes it's a bit of a problem.

Another thing for why BTP is so critical for the company is that we want to move on from SAP history. You have this massive on-premises software, you deploy it, you customise the hell out of it, you develop on top of that. Of course, customers love it. It's not suitable for software-as-a-service delivery model at all, because software-as-a-service delivery model is standardisation, standardisation, standardisation, and scale.

Where do the customisations sit? It is very, very difficult to build extensive customisation capability on top of a software as a service stack. What you do then? You build a platform, you build the BTP. So the idea is that, in the future when you're moving from S/4-on-premises-system into a software-as-a-service world, you have S/4 public cloud plus BTP and customisations are built in BTP.

Of course, we can go and set, okay, then all these customisations, how about we use hyperscaler platforms? How about we use Azure or Google and so on? But then you lose the revenue, you lose the profit, you lose the control, you lose the data.

While BTP is the second largest line of business in terms of revenue for the company and one that is growing

almost the fastest. It's also really critical to cloud delivery. It's really critical for generally a future of SAP as such, because if we turn out to be just a application provider, it reduces the barriers to entry, and I'll talk about later on in the context of GenAI and ability to have smaller developer teams producing much faster, our unique value becomes a problem.

We are looking at the whole thing holistically from the perspective what our customers need, but also as the business perspective of how are we delivering our cloud services, and what do we control, and how we're building our manufacturing chain.

In terms of how we run infrastructure… First of all, why is infrastructure so important? You would think we are a "build applications" company. The thing is we are quite big and quite very big. The amount of infrastructure we consume deliver our services is staggering. I can't talk about the numbers. They're strict confidential, but they're massive.

For each one of these three, Google, Microsoft, and Amazon, we are among the top 10 customers globally. Each one, top 10. Then we are still top 10 for Dell, and Cisco, and VMware, and others. But we are quite huge.

Infrastructure or how we consume infrastructure is really critical for our profitability because fundamentally, we have three kings: revenue, margin, and service quality. Two of those are heavily determined by infrastructure: service quality and the margin. They are connected.

I'll just give you one example. If you want to deploy Microsoft Azure and achieve a availability required from a mission critical service… Because remember, what we do, we do software which underpins manufacturing, which underpins supply chain, underpins mission critical capabilities of our customers. It's not like if our software doesn't work, nobody cares. Our software doesn't work, Goodyear doesn't get trucks out of the warehouse. It's a different priority.

You look at availability and uptime in different manner. If you want to deploy in that way on a hyperscale like Microsoft, you will have to deploy on three availability zones, and then you're going to go cross-region for DR. That's very bloody expensive at our scale. It costs a lot of money, a lot, and it blows your margins apart.

We are looking into being very intentional. We approach it in an engineering manner. That's why infrastructure is so critical. We are doing a lot of work around it to be able to deliver the level of uptime inequality at margin that a company needs. If you mess it up, you hear it on the news, because we need to restate our financial statements like promises to the market. It's quite critical.

We run something called four plus one, where we run at scale at four of these biggest hyperscalers. The fifth one is Oracle. Of course, we don't run on them. We're like weird. But Alibaba, we run only in China. Globally,

we run on Microsoft, Amazon, and Google. We do it in intentional manner. We're doing a lot of central engineering and so on, so that we can run across them in a standardised manner, while not abstracting fully—I'll talk about it in a second—but making sure we don't create the levels of lock-ins, which would remove our strategic ability to decide and change and reposition ourselves.

We also still fully design and build around our own infrastructure platform. No, you cannot buy it as some people asking why. No, we are not selling our infrastructure. We are not going to enter the business, but we are running our own infrastructure platform in data centres we fully own, as well as data centres we rent, and then we basically run around stack. It's an open source based stack. I'm not talking about it too much, but we are doing it also in intentional manner.

Because while we can get a globally scalable infrastructure with the hyperscalers, which is really good, and we can also get a lot of innovation velocity from the hyperscalers because they are way more basically platforms, we can also get much higher profitability in our own infrastructure. In some cases, we can get higher uptime infrastructure.

We have decided to continue running it to retain that competitive advantage, to retain that capability, also taking into account that we are effectively the only European big software company. Everybody else is in the US. While right now we are friends, that's awesome, we have won elections away from not being friends. You got to look at it from a perspective of business sustainability and not just leaving the world of illusion.

Aside from that, I talked a bit about infrastructure and so on, but fundamentally for all of this to come together, there is also a couple of other topics which are really, really critical for us. One is of course, security. They refer to us as ERP hyperscaler, meaning we have a bull's eye on our head quite literally with state-sponsored actors as well as organised crime. Security also to significant extent is tied to infrastructure, is tied to observability, is tied to data collection analysis. We are doing a lot of work around that one. It's a very big part of what we do.

Then the observability is truly critical. Generally, just to give a broader view, when you talk about cloud infrastructure, we make a lot of decisions like what we buy, what we build ourselves. Typically, what happens is we start by buying a capability which we need from market. In many cases, though, that proves to be financially unattainable once we start scaling out. We just cannot afford it. We just cannot afford it at the level of verbose data collection we need to do to light up our observability and the way that observability is the backbone of the full cloud delivery, it's a mission critical system for us.

If observability is down, I cannot patch teams when there is an outage. I cannot diagnose an outage. We are

heavily invested into that stack. We are building their own just for the sake of the size, because with verbose data collection start getting problems, we are seeing that in some other parts or other parts of the company, so that's a big one for us.

Another really big one is this I talked about cost efficiency in effect. Again, I cannot talk about numbers like how much is infrastructure part of our overall cost structure, but it's massive. We are, again, biggest customer for all three hyperscalers as well as for a lot of suppliers in the data centre space. We need to optimise that ship for the company in governance, infrastructure usage for the company so that we actually hit the profitability targets, which are very, very tight in the coming years.

Another big one, of course, is a topic of scalability. This is where the hyperscale is a really, really beneficial for us because they allow us to enter markets very rapidly, to scale up very rapidly. But you also have to do it in a smart space because the topic of runaway costs is true. It's there.

This is another area where we are heavily investing our own engineering capabilities, which is FinOps, financial operations. We do have a couple of products we have bought out in the market, but fundamentally, we engineer our own system up. It's quite unique. We actually benchmarked it against the market. We had some smart people coming in and telling us, and basically, they told us the stuff you're doing is fairly unique, even in the global level, because there is very few, very few companies around the globe which are running at three hyperscalers of this scale, trying to solve the problem. Not many folks are really doing it.

This is where we are really investing a lot of engineering resources. We are also investing a lot in terms of a generative AI, our engineers and stuff like that, looking what we can do more because the upside for us is attaining the margin, attaining the profitability. The failure means basically, the company cannot achieve the financial promise that we made to the market.

I'm almost done. A few more minutes and we go for questions. In terms of AI, you probably all read about Jewel and what SAP is doing toward the customers. There is a huge effort across all units in the company, where we are trying to light up for the capabilities in our products. There are some concerns there particularly in the context of hallucinations, because again, most of the cloud services we do as a company are mission critical and business critical, so high quality is expected.

You can't have the thing hallucinating and lying to people. When it's lying, it's lying in a very confident manner, so how do you get to hallucination zero without turning this whole thing into an expert system, which is like the '90s? You don't want to do that. That's the R&D problem that we are having on our product side. On the

back end where we are, again, I told you, we have a huge scale.

GenAI for us is a productivity and acceleration tool primarily. I'll give you a few examples. Let's say I have a big, big outage, which is crossing multiple line of businesses, a combination of infrastructure and application-level outage maybe related to BTP because it underpins everything. You go in. We have colleagues from Service-Now here. We are heavily using ServiceNow as our backbone. We page 300 people, 100 SREs. You go into the bridge goal is 100 people there. You try to orchestrate them.

You detect an outage, but to interpret it takes a while. What we have, for instance, learned and we are working on it right now is that if we are able to fine tune or even retrain a model on our own data, and we have a verbose data collection, again, observability, we are able to identify and get to the root cause faster than any human SRE we can do. Then the SRE is actually taking action and restoring it.

We are in the process of lighting it up and talking about something which is currently research. We actually have some patents pending as well, so I'm not going to talk more about details, but it's quite impressive what can be done and how fast, how much you can accelerate your reliability engineering through use of that. We actually have SREs building the thing. It's not built by somebody outside.

Another example is we, of course, are heavily audited. We are submitting to number of certifications, like PCI but tons of other ones. Doing that is very labour intensive. I don't like spending engineering effort, engineering muscle on doing checkbox. We have actually an engineering team working on pushing really hardcore automation in any security control space, so to say.

Because we don't have that many people. If I can get an engineer to do smart stuff and not doing dumb stuff, it's a huge benefit for us. This is what we're investing a lot into. But it's always around topics around uptime and optimisation, those kind of things, because we're in the back end channel.

In this space, we always start in a way that we go in to consume somebody else's API, like an OpenAI's API or on Azure or something. But very rapidly, you see that, for you to be able to really play the game, the difference is in the data sets. We actually saw the marketable change. When you are going in asking ChatGPT, you feed it with your data coming out of the observability system in a prompt engineering form, and then it tells you what happens versus it has been trained on your data, and then you just trigger it to get you the analysis of an outage or something. It's night and day. It's not even comparable.

But then that means you have to train it on your data. You have to retrain it. You have to fine tune it. You got to run it somewhere where you control it. Running it is also very expensive as well as training is very expensive.

I think that's going to be a very interesting thing around GenAI in general because yes, it's a very cool capability. It's a very expensive capability.

If you want to run it at scale, are you rich enough? None of us is. I would love to be like Aramco and have a black cheque like 2 billion budget and do whatever you want. It's not. For us, it's a big question because also, if you are succeeding with your GenAI product in the market, your costs are going like this. Your revenue better goes like this. If it doesn't, you're going to run out of money really quickly. That's also infrastructure.

I'm joking. I'll finish with this one. My background is engineering, but I think by the time I'm done with this, I'll personally be able to open an accounting company. I think I'm going to change career because it's like counting beans, to be honest with you. I'll finish with that.

Please, I would like to open for questions. Anything you would like to ask, I'm more than happy. If I cannot say, I cannot say, but I'll try. Please, I want to.

Audience/Host

Thank you for the presentation. I'm Philip, I'm working with Amazon, which is why the question is maybe a little bit biased, but what is the upside for SAP running multicloud, and from your perspective as SAP, what is the upside for other enterprises running multicloud?

Ozren Kopajtic

This is the thing, very open. If I was starting my own startup, I would run on Amazon probably, or I would run on Google. Why? Because Amazon works really well. You guys just work. Google has a better stack, in my opinion. By the way, just for the record, I was 10 years at Microsoft on Azure. Just to be very clear. Just to be clear, I never worked for Amazon, not for Google.

If you are starting a startup, that's the way you roll. Why multicloud? Very open, it's not something that happened at SAP intentional. We are a massive company. We had a lot of acquisitions. Many of them came with their baggage, with their own plans and stuff like that.

Also, we have go-to market and commercial relationships with all three hyperscalers. We co-sell with Microsoft like crazy. We are doing a bit of that with Amazon, and we do it with SAP, so it's a combination of. If I would not care about that side, I would probably run on one or two, probably on two, and I would have on infrastructure because of the scale, if you guys go and try to bust me.

But four plus one, very openly, I do have a bit of a problem with it, especially because of the Alibaba side, because they're running only in China. We are seeing more and more political challenges in the context. Alibaba in China is not considered to be a cloud supplier that is trusted by the government. They're not sovereign in China.

That aspect of having them on board is like, we have a reason why we run it. I cannot talk about it. It's a business reason, but that stretches you a bit. The thing is, the following is, that if you want to run on true multicloud and you're running it at scale, you're going to have to invest engineering resources to be able to run. You will not be running just like running. Otherwise, you're going to lose control.

Then you have to make an assessment. Does investment and this engineering resources make sense for my skill, for my go-to market, for my business? There is no one answer. But again, if I'm running a startup, I'm running only one, for sure.

Audience/Host

Thanks.

Ozren Kopajtic

Next one, please.

Audience/Host

There it is. Perfect. Thank you very much. Once again, Steffen from Accenture. You mentioned also in the keynote lots of different tools, new technologies, new capabilities, be it in the GenAi space, be it in the observability space. SAP is one of the largest software companies, and for sure in many areas, very much advanced.

How do you decide or how do you dig through all of that slew of different technologies? Where do you pick them? Where do you decide for building something?

Ozren Kopajtic

Money. I'll give you a very, very, very blunt example. Anybody from Splunk here? It's going good because I pissed him off now. Basically, very openly, I'm going to take a specific case. Take Splunk, for example. Their tool is amazing. It's kick ass.

If you are able to do verbose data collection, and then you just built up the stuff they already have, just light them up. You don't even have to do smart stuff. It's wild. Their analytics stack is awesome.

I can't afford to run them at a scale. I am not rich enough. I'm not rich enough. I don't have enough money. Castro is not giving us money below the table. This is the thing. Our problem is scale.

Everything which scales and bounds for us has the potential of scaling unbound. We're going to run it in open source ideally. We're going to engineer ourselves, and we're going to try to run it in the data centre where it is cheaper. Because if I run it in one of our SAP own data centre, where I own the land, it's half the price of what Amazon can give me the best day, and you guys are the cheapest, I'm telling you. Others are more expensive.

Again, one thing, I don't care. There's nothing they can do to me. That's a fact of life, half the price. Not

because hyperscalers are bad, but because they are a cloud provider who has to make profit, and they have to make a cloud profit, which is 79-82% margin is what the financial markets expect on the cloud business, otherwise you're not cloud. It ramps up in our scale. It's like petabytes of collection. That's how we decide.

If I look at something, and I'm like, I'm not going to go and build my own chips, I'm not crazy. Why? Because I have Dell, and I have Lenovo, and I have HP, and the hyperscalers are kicking their butts anyway to make a high-quality product. I just ride a wave. I'm not going to do that, and I'm not going to build my GPU chips as well.

But in some other areas, yes, we will. This is how we decide. Money and scale is the biggest driver for us. You can argue that correct or not, but that's the way we do it now. We have a bias for open source.

Audience/Host

One more question.

Ozren Kopajtic

Just one more.

Audience/Host

Microphone's coming.

Audience/Host

Question for Merck. We are currently running also the go-to cloud strategy together with SAP. But the question and also one comment in the form of question ring a bell with me regarding China. We heard in the keynote the tensions rising between US and China and where to store the data, how to build up the instances, or what is the better strategy. Should we go for one instance on one cloud? Or you mentioned the multiple cloud setup. Is it in multiple cloud, one instance, several instance, or what are the ways also to consider the situation with more diversification, getting that goal?

Ozren Kopajtic

[crosstalk 00:23:59] the following. If you run in China, you run in China. It doesn't matter. You can run on Azure Amazon and Alibaba in China. There's no problem. You design a strategy. The point is that you run in China for China. You don't run outside of China for China. You don't run Chinese stuff anywhere.

I have a situation, where I cannot have a company, any entity which is exposed to the US market because US is extraterritorial in principle. I don't have to be a US company. You can have a traffic traverse if you have a device. You're breaking actually a particular rule of the US government.

The way we think about it is everything for China, we run in China. Nothing for China I'm running outside of China, and outside of China I have nothing Chinese. I

know it sounds a bit nuts, but that's the way it is. The way we look at it, it can only get worse. Again, you're one election away from even bigger problem. The businesses have to prepare.

I just read an article today, the last month was the first time in a history of recording since '92, where the amount of investments exiting China was higher than the amount of estimates coming into China. I'm not making any conclusions either way. I'm just saying we got to pay attention. If you're exposed to the market, we all are, you got to pay attention.

I don't know what the future is going to be, but what we're going to do? We are doing stuff to protect our business so that we can react and we don't… Russia exit was a massive exercise for us, and we did outstandingly well. But it was a heroism, and I don't want to do it again, if I have to. I don't want to do it again.

I'm so proud of our teams, all of us, but it was no fun. I hope that answers a bit.

Audience/Host

I think we need to wrap up. Thanks a lot. Very inspiring and authentic presentation. Good to have you here, and thanks a lot.