



The Deloitte On Cloud Podcast

David Linthicum, Managing Director, Chief Cloud Strategy Officer, Deloitte Consulting LLP

Title: Going global: Building an extended enterprise multi-cloud

Description: For companies that operate across geographies, IT architectures and regulatory requirements can make it difficult to access widely distributed data for analysis. Extended enterprise clouds (EEC) can help provide access to disparate data via augmenting the enterprise multi-cloud with third-party services such as vertical clouds, AI clouds, or data clouds—all with a common control plane. Building an EEC takes planning and a cultural change around data ownership, but it's well worth the effort.

Duration: 00:24:11

David Linthicum:

Welcome back to the On Cloud podcast. Today on the show I'm joined by David Bernstein, Chief Global Cloud Architect at Roche. Hey, David, how're you doing?

David Bernstein:

Hi, David. Fine, thank you. How are you?

David Linthicum:

Tell me about your role at Roche. I know you've been on the podcast before, so catch us up with what you've been up to and what your passions have been lately, including some of the technology stuff but also things you do as a hobby.

David Bernstein:

Yeah, well thank you. A couple years ago, I joined Roche. As you know, Roche is a large pharma company with a lot of wholly owned subsidiaries you've heard of like Genentech, or Foundation Medicine, or Flatiron. We're pretty big in the space and very big in competition biology and using IT for pharma. So, that was an exciting opportunity, and I am glad to say that I'm doing a lot in the cloud architecture area here, so I've got a lot to talk about.

David Linthicum:

So, what are the three things you like about your job? I always love asking that question.

David Bernstein:

Well, number one, I get to work on cloud and we have lots of resources and lots of projects, so lots of variety from scientific computing to enterprise or IOT applications. Also, the mission is good. I mean, the idea of actually improving life for people through life sciences is great. I'm glad to be able to do that in this part of my career. And of course, we're a global company, so it's really fun to work with people from all around the world on these projects.

David Linthicum:

Yeah, I love that too. I love working in the healthcare vertical just because of that and it has such a direct impact on the quality of people's lives. And the better we get at that, I think the better we're going to get at solving disease problems, have people living healthy lifestyles, actually finding issues and just kind of hacking the way in which we have to deal with our own biology. I always think it's an exciting aspect of that. So, we're going to talk today about enterprise extended multi-cloud. So, give us a brief explanation for maybe people haven't heard you before or haven't read what you've been writing and things like that in terms of what is an extended multi-cloud.

David Bernstein:

Well, in a smaller company, a company generally will say, okay, let's concentrate on using one cloud provider or another, one hyperscaler or another, and kind of all rally around maybe a single application, a classical SaaS company. But in a company like—a large multinational company like ours, we have just such diversity of different teams and different projects, so not only do we use Amazon, but we also use Google cloud, we use Azure cloud, we use Alibaba cloud, and we use them in different regions for different purposes all around the company.

So, this is in addition to a pretty large on-premise infrastructure, but just public cloud alone has lots of diversity in it, so that's something that is just a fact of this type of size of company. And, so, we cope with that. That's the first half. The second half is we use these other—the reason we were kind of talking about extended multi-cloud is because we also use other clouds, the data clouds for example: Snowflake, Mongo DB, and so on, which are clouds in their own right. So, we have potentially data all over the place. That's sort of the statement of the problem right there.

David Linthicum:

So, in other words, what we're doing is we're building out kind of a traditional enterprise multi-cloud, typically two or three public cloud brands, but also extending that out so we're dealing with third-party cloud providers which have a tremendous amount of value, so either they're a vertical cloud provider, in other words, they're working with vertical processes and vertical data models, or as you mentioned, they're a data service provider unto themselves like Snowflake and many others. So, it's the ability to kind of leverage a cloud or a multi-cloud provider that's much more heterogeneous than a standard multi-cloud. Is that a good way to describe it?

David Bernstein:

Yeah, that's right. And what's of great interest, of course, is that having an internal data mesh, or an internal data marketplace, where what's really happening, at least in places like Roche where we do computational biology, combining data sets to fuel new discoveries, clinical and genetic data together to try to understand what new medicines might be able to do for certain groups of people. This is really powerful stuff, and so being able to know all the data sets we have, combine them, and do really large data science on them is really the place where we get an enormous amount of value acceleration. But it takes a lot of data and it takes a lot of computation.

David Linthicum:

So, what I like about this is we're not force fitting things into a particular bucket, so what we're saying is if we can't find the best ways to manage data, even extend it out to AI, or processing, and even transaction processing within our multi-cloud environments, there are three main clouds—Amazon, Microsoft, and Google—but of course other clouds are involved as well. We're able to extend it out, so we're not necessarily—we can add a data cloud, we can add an AI cloud, we can add a monitoring and management cloud, perhaps a specialized verticalized service around healthcare information, data providers, things like that.

And if you think about it, when we first envisioned using cloud, we had API-based economies and service-based economies. Back in 2001-2002, we were talking about everything being loosely coupled services that we bring over the internet to assemble and reassemble applications. And we even had the programmable web. I think sure it's still out there, which is a list of these APIs or a list of these services we could leverage to do anything discrete, and then build these composite applications. Little crazy because you're thinking about you're building an application with hundreds of services that are hosted all over the world.

But, in essence, we're kind of getting back to that same value here with not going too crazy about it. In other words, we're extending our abilities and our capabilities by leveraging different heterogenous—then to be more openminded and acceptable to leverage many different services that are typically cloud services. So, instead of going through three to five, we may go to 20, and I think that's perfectly fine. What are your thoughts on it?

David Bernstein:

Absolutely. The challenge is, of course, that the data is not as fluid as the composite application components. The data is big, expensive to move, or hard to move, and in some cases, we've got geopolitical or data sovereignty issues where we really can't move the data. So, we've got to address this in a way where we bring the computing to the data for all those reasons. And, so, that really makes you think about the way you organize global data in the context of this multi-cloud. You have to really think about this in advance a little bit so that you know that you're doing a good job being a curator of data for all the owners of it. In this case, this may be, of course, personal or medical data, so this is a very interesting challenge. And, actually, it's pretty new across a lot of these vendors that are able to do processing on where the data is, and that's I think something that might be interesting to talk a little bit more about.

David Linthicum:

Yeah, tell me about how we manage the information, how we deal with data management, data security, data governance, metadata management, things like that.

David Bernstein:

So, when you've got a lot of different teams generating a lot of different data sets and they need to stay in a particular country, or they need to be traced very carefully as to who uses them for what, of course, what you want to be able to do is allow scientists and enterprise researchers to use whatever data capabilities they need—BigQuery, Snowflake, Mongo, or Athena, all of these different great engines, machine learning. But you have to operate where the data is, so nowadays what you can do is you can actually have your Snowflake run next to where your data already is on Amazon, for example, and operate directly on it in your own S3s, not move it at all.

And I don't know if a lot of people know, but BigQuery can operate actually the same way, on Azure actually next to the data that you already have sitting on Azure. So, you can take a philosophy where, yes, we can use a lot of these clouds, but because, yes, we've thought about this in advance and we've kind of created this federated query-in-place architecture, and that's a lot of what we see as sort of the silver bullet in being able to do this without having to move petabytes around across countries, which just doesn't really work out.

David Linthicum:

Yeah, I never liked redundant storage of data even though in some instances it couldn't be avoided, and certainly when you deal with compliance issues and regulatory issues around how you're going to deal with data, HIPAA for example, European security standards, it becomes kind of a common fight. So, this is about not placing the same data in two different places and therefore having a single source of truth and having one way in which we can process the data, which is a much more healthier way to use data. Is that right?

David Bernstein:

Yeah, but letting different teams or different scientists use the right tool for the job. The right tool for the job is not necessarily connected with where the data exists and you can't move it. So, the idea is, should a new, interesting analytics school come along, point it at the data where it's at and go for it. That has really become a lifesaver for us to allow all this diversity in cloud without having to lose control of where the data actually is. So, that's a big difference for companies.

David Linthicum:

Yeah, and this has always been the vision but difficult to achieve. I remember writing about this in my EAI book back in the '90s, the ability to get at any piece of information at any time, for any reason, for any purpose, and use it into different analytical domains. So, it could be used for decision support, high-speed transactions such as sales transactions, and the ability to aggregate the information with other information that may be external to the enterprise, external to multi-cloud, to get more value out of the data. And you think about it, we're not getting as much value out of the data, and moving to this approach, we now have that opportunity. What are your thoughts on that?

David Bernstein:

I think that is totally a change, and one of the interesting changes in this is that it's turned out to be pretty difficult networking wise to actually move the scale of data around that we have. So, the idea that I have some petabytes sitting in a geopolitical area, or a country and I want to access those, I can go ahead and utilize the diversity of different clouds to access that data in place, and I can do that simultaneously with my colleagues accessing that same data. So, what I've created is added value to my data set but still being a good curator of that for the owners of the data, the patients in our case. So, this is really different. I mean, in the case where now we have a lot of ability to stay on the service provider networks through this sort of storage federation, it's a huge cost advantage as well, so it's been pretty much a breakthrough in the way we've been able to deal with making the most out of these large data sets.

David Linthicum:

So, besides doing looking at data regulation issues and things like that, the ability to deal with that, and certainly the need to keep data in certain geolocations and the ability to look at federated data that's actually dealing with this consolidated, what are some of the other use cases that you're finding where this is more applicable?

David Bernstein:

I think specifically what we're seeing is that folks are able to access and combine data sets that they might not have been able to before. For example, as we begin to combine some of the diagnostics data that we have with some of the clinical trial data that we have, these are data sets that, because they're in this new world sitting near each other, or they can be accessed by a compute engine that we hadn't anticipated in advance, that's what gives us the flexibility then to find new relationships or trends.

And, also, to work with other companies as well. So, we can actually now do research with other universities or other life sciences companies and share our data in place by providing specific access to a data set that we don't need to move, we can provide access to a number of co-researchers to operate on that data with us simultaneously. So, this opens up a huge new opportunity for collaborative work across researchers, across companies, or universities that I think was really not possible when you thought about actually having to move the data around and own it and associate the data with each different compute project. Now we have the data all on its own and we bring the compute projects to it. That's a very big paradigm shift.

David Linthicum:

So, what about other sectors like finance and manufacturing and retail and those areas? Is this technology applicable in their scenarios as well?

David Bernstein:

Absolutely. I hear a lot about ideas of collaborative data sets and supply chain. I hear a lot about that in finance. I can't really talk as much to those industries, but I do know that combining data sets across different stakeholders and being able to create this enhanced visibility is something that I think is very topical. I think we're going to see a lot more of this as time goes by.

David Linthicum:

Yeah, I do too. I think that now that we have connectivity issues solved, the rise of 5G, and the ability to have access and connectivity to many different locations, the ability to leverage federated data, and by the way, not just distributed databases—those have been around for a long time—but the ability to have a common control plane that goes across different data storage nodes that exist all over the world, and either leverage that as the node unto itself or leverage that as a unified data set and combine and contrast and abstract the information so it's more useful to you is something that's been coming for a long time.

I never understood why we weren't able to hit the accelerate on that problem. We're still combining data, we're still doing batch jobs where we're taking one big mega data set and we're doing a batch job to extract, translate, and load it into another data warehouse set, things like that. Now we're moving into more advanced analytics and more real-time analytics, but we still have trouble in dealing with data that's very complex and distributed where reality is just takes some thinking on how you do that and to kind of take things to the next level.

So, how would enterprises approach this now and what are the steps that you think they should consider to kind of, number one, see if it's right for them in terms of the requirements, and how should they approach the technology and how should they approach changing their existing data storage technique and data retrieval techniques to something like this where it's more federated?

David Bernstein:

Yeah, great question. I mean, it's a question of planning ahead. It's kind of thinking inside-out as opposed to outside-in because oftentimes the data scientists or application sets are thinking about the data they're starting with and what tool they want to use to solve their problem. And then you've bound together your analytics or your machine learning environment with the data set that you have. And you probably pick the vendor and a country to do that in. I would start thinking from the middle out.

Start thinking, "Okay, we've got the following areas around the world, we're going to collect data of these kinds, we're going to put them in a place, we're going to make that the place that we control and share out of, and then everybody can bring their engines, point them to that place, and then we'll start building up a much more synergistic as opposed to siloed data regimen." That's really it. Plan ahead and—because we all know that we're going to have 100 times more data than we do now pretty soon, and so don't try to look backwards. Try to plan ahead as to how to create a global data architecture and build the applications around that.

David Linthicum:

So, even though we're generating more data all the time, we see lots of stats with that, shouldn't this make things better in the fact that we're not storing redundant data, insisting on storing redundant data, so have copies of the same customer information, transactional information, inventory information which exists in 30 different places within the enterprise, and therefore there's no single source of truth. So, if we eliminate this redundancy, won't that kind of normalize our data congestion problem where we're just storing this stuff everywhere like we're data hoarders?

David Bernstein:

Well, that's the holy grail, David. I think you've nailed it on the head. So, that's what we should strive for, creating these data lakes, creating your immutable data lakes where you sort of keep all the original data, and then immutable data lake where you can even do transactions on now with things like Apache Iceberg. You could have all these teams operate on the same in situ, query in place, or transaction place or data set, combine multiple ones together. And this will save cost. That's sort of an interesting other piece to this.

You're right, having the data set replicated is not only expensive in terms of operation syncing things and so on or finding all the copies, but you're paying for them, and you're paying for the network to move them around. So, this is a win-win in a number of ways, but it requires planning ahead and requires folks to set—as they do their projects, it's a little more difficult always to set them up to use the federated data, but that's an investment well worth doing up front.

David Linthicum:

One of the issues I think we're going to run into as we start normalizing data, the data ownership within enterprises typically is not a single entity. It doesn't roll up to a single person. It's owned by different divisions that are owned by different personalities and personas, and it's going to be hard to get all of these organizations on the same page in how databases—how data should be managed and data should be stored and using a more open concept. I look at this as kind of a pragmatic data architecture. In other words, we're not reaching here in terms of what technology's able to do.

The technology's able to do this for a while. This is about accepting the fact that these capabilities exist and we're going ahead and move to this particular architecture. How do you think we maneuver the politics, the cultures, and the organization that really want to hold onto their data and want their data to be in a certain place and don't mind it being replicated all over the enterprise and don't mind the way applications are addressing their data access in kind of a dysfunctional way? How do we change their minds?

David Bernstein:

Well, this is a cultural change, you're right. A number of organizations you'll see a Chief Data Officer, and you'll see a team that's sort of responsible for global data strategy, and it's sort of a first-class concern of the organization to really do its scientific or its enterprise data footprint in a particular way. Make this—don't just allow entropy to take over. Create a strategy and create a strategy that's going to enable the most synergy of your data going forward as possible, and the least risk too.

Because this is a topical issue to make sure that the data is—follows the law and also that you can maneuver with that, know where your data is when, and so if you need to do something about it, move it or shut off access or—I mean, the world is a changing place. This should be one of the things you think about. So, I think making—just as a Chief Digital Officer or Chief Data Officers, this should be something that is within their charter and they can help create the culture change to make this something that people really sign up to, just like other big concerns in corporations like sustainability really takes a chief officer somewhere. So, data is really valuable for corporations. Let's make a proactive, thoughtful strategy as to where to put it as opposed to leave it to a bunch of cloud data scientists. We know how that turns out.

David Linthicum:

Yeah. No one can agree on anything, and I think in some instances this should be a board level, CEO level decision, and because it has ramifications for the business, your ability to have the business strategically focused. If you think about it, we're moving into an economy where the businesses are going to be valued by not the products they produce, but the customer experience around the shipment of the products and the services they deliver and things like that. And their ability to do that in a much more innovative, creative, and different ways, Netflix versus Blockbuster, and all the other disruptors that are out there.

And if you want to become a disruptor, your data becomes the disruptive technology you're able to leverage and able to leverage as a force multiplier, to be an innovator in your industry, you can go into a way that's more creative than your competitors and become a disruptor before you get disrupted, and that should be something as strategic to the business. Am I reaching here?

David Bernstein:

No, you're not at all, and the key thing is being a disruptor sometimes involves making a connection between things that others weren't able to see or weren't able to make, and so if you have data from all these different parts of your business and you can kind of put them together and get a better sense of what's going on either in a supply chain or in a research area where you're bringing different insights together to get—find out something new, or as a customer experience, which is a classic example. If you don't plan ahead to try to get insights out of the synergy of your data, then it's much, much harder to do.

David Linthicum:

Yeah, I certainly get that. So, where can we go and learn more about this concept on the web?

David Bernstein:

Well, there's a lot going on around data mesh. I would definitely check out the—all the different data cloud vendors in particular, their moves towards what they call storage integration. It's kind of a new part of how they've launched their global data clouds on all the three different major hyperscalers as well as in all kinds of geographic locations. So, that I would go—I would definitely go there. And of course, there's a number of different data marketplaces which are very intriguing inside of places like Snowflake and Amazon that are showing when different corporations or universities or research institutes get together and share their data what happens there. So, there's a number of very exciting open source and industry alliances going on around data sharing using this kind of capability. So, it's a pretty interesting phenomenon. For the gear heads involved, take a look at things like Apache Iceberg, which is sort of a big innovation in this space.

David Linthicum:

So, where can our listeners find out more about you on the web?

David Bernstein:

Best place to find me is on LinkedIn, so just check it out and I'll be happy to talk.

David Linthicum:

Yeah, this is an important topic. We're considering new distributed federated architectures now that we have the capability to do it. Now's the time to start making your move as we start moving to multi-cloud and other more complex distributed architectures and figuring out how we're going to deal with federated data and how we're going to deal with data differently. And, certainly, leveraged data is the first-class citizen that it needs to be leveraged for within your organization within your enterprise.

So, if you enjoyed this podcast, make sure to like us, rate us, and subscribe. You can also check out our past episodes and include those hosted by my good friend, Mike Kavis. If you enjoyed this podcast, make sure to like us, rate us, and subscribe. You can also check out our past episodes including those hosted by my good friend, Mike Kavis. Find out more at deloittecloudpodcast.com. And if you'd like to contact me directly, you can e-mail me at dlinthicum@deloitte.com. So, until next time, best of luck in your cloud journey. Everybody stay safe. Goodbye.

Operator:

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to Deloitte.com/about.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Visit the On Cloud library
www.deloitte.com/us/cloud-podcast

About Deloitte

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms. Copyright © 2022 Deloitte Development LLC. All rights reserved.