



Transcript provided as a free resource by



Data Mesh Radio Episode #189: Our Data is In the Cloud... Now What?

Interview with Vikas Kumar

Listen ([link](#))

Transcript provided as a free community resource by Starburst.

To check out more Starburst-compiled resources about Data Mesh, please check here: <https://www.starburst.io/info/data-mesh-resource-center>

To get their Data Mesh for Dummies book (info gated), please see here: <https://starburst.io/info/data-mesh-for-dummies/>

List of all previous episodes (including links to available transcripts) [is available here](#).
The folder of all previous transcripts [is available here](#).

0:00:00 Scott Hirleman

The following is a message from George Trujillo, a data strategist at DataStax. As a reminder, DataStax is the only financial sponsor of Data Mesh Radio, in the Data Mesh Learning Community at this time. I work with George and I would highly recommend speaking with him, it's always a fun conversation.

0:00:19 George Trujillo

One of the key value propositions of a Data Mesh is empowering lines of business to innovate with data. So it's been really exciting for me personally, to see Data Mesh in practice and how it's maturing. This is a significant organizational transformation, so it must be well understood. Empowering developers, analysts, and data scientists with downstream data has been part of my personal data journey that reemphasized the importance of reducing complexity in real-time data ecosystems, and the criticality of picking the right real time data technology stack. I'm always open and welcome the opportunity to share experiences and ideas around executing a Data Mesh strategy. Feel free to email or connect with me on LinkedIn if you'd like to talk about real time data ecosystems, data management strategies, or Data Mesh. My contact information can be found in the notes below. Thank you.

LinkedIn: <https://www.linkedin.com/in/georgetrujillo/>

Email: george.trujillo@datastax.com.

0:01:11 Scott Hirleman

A written transcript of this episode is provided by Starburst. For more information, you can see the show notes.

0:01:18 Adrian Estala

Welcome to Data Mesh Radio with your host, Scott Hirleman, sponsored by Starburst. This is Adrian Estala, VP and Field CDO at Starburst and host of Data Mesh TV. Starburst is the leading contributor to Trino, the open source project and the [Data](#)



Transcript provided as a free resource by



[Mesh For Dummies](#) book that I cowrote with Colleen Tartow and Andy Mott. To claim your free book, head over to starburst.io.

0:01:48 Scott Hirleman

Data Mesh Radio, a part of the Data as a Product Podcast Network, is a free community resource provided by DataStax. Data Mesh Radio is produced and hosted by Scott Hirleman, the cofounder of the Data Mesh Learning Community. This podcast is designed to help you get up to speed on a number of Data Mesh related topics. Hopefully, you find it useful.

Our data is in the cloud. Now what? Bottom Line upfront, what are you gonna hear about and learn about in this episode? I interviewed Vikas Kumar, AVP and Head of Data, AI and ML At CNA Insurance. To be clear though, he was only representing his own views in this episode. So some key takeaways or thoughts from Vikas' point of view, number one, in Data Mesh, make sure to keep focused on bringing the business domains along. You aren't building for the sake of building. If users can't derive value from the data work being done, why is it being done? Number two, the 2010s through the early 2020s have been about moving data to the cloud, but we are starting to see people really leverage the data that's now in the cloud to generate value. The cloud unlocks many new possibilities around data due to flexibility, scalability and unit economics. Number three, with moving to cloud, there is much less focus on specifically managing the data itself and more focus on getting value from the data. SaaS data product offerings really unlock people's time to focus on driving value. Number four, cloud gives us the scale and data availability, but there is a long way between having the data available and leveraging the data for driving significant value.

Number five, cloud can be a double edged sword. It gives you the flexibility and scalability, but without good controls, you are likely to do a lot of duplicate work. Lots of cost embedded in that duplicate work. Be careful that ease of data product creation or at least proof of concept creation doesn't create chaos and data product overlap. Make sure to have good governance here including strong communication. Omar Khawaja talked about that in his episode a lot. Number six, potentially controversial, we aren't very good yet at making it easy for business domain users to leverage data in many of their decisions. Where do we fall on the spectrum between we need to teach them how to do everything data and we need to curate everything for them? Are we teaching them to fish? Are we fishing for them? How far between those do we go? Number seven, another potentially controversial one, it's easy to focus too much on the short term quick wins in data. You need to think about your overall data landscape and build a foundational approach so you can go after big picture, big impact bets with your data work. You should think about building every data product from a foundational approach too to make them more extensible.



Transcript provided as a free resource by



Number eight, we need to get people out of their functional silos with business people only speaking business and data people only speaking data. Number nine, to do Data Mesh well, we have to focus on the operating model of the organization around creating and maintaining data products. There is too much focus on the technical aspects instead of, how does this actually get done in a way that fits with the organization's ways of working? Number 10, data producers must assess data consumer's data fluency levels. If they aren't very strong with data, should you really be delivering them raw data instead of maybe curated insights? I think this is a big kind of ongoing question as to how far does that ownership extend and are we just dropping off raw data and the people are going, "What do I actually do with this?"

Number 11, for any data product, you should start by mapping it to a target outcome, but it shouldn't stop there because with reuse, new outcomes may emerge that drive additional value, but you have to have something that you're trying to build this for, right? You don't just build it for the sake of building. Number 12, data product owners are crucial to building good data products. It's their job to identify, and then satisfy that objective of building the data product. What are you trying to achieve? Number 13, we shouldn't focus only on the data product. The work to create that data product is what makes it valuable. The data product is merely the vehicle for delivering the value, the output of real product work around data.

Number 14, potentially controversial, many companies doing Data Mesh appear to be trying to leave data governance until later and that is likely to bite them. The governance, meaning the security access control, but also the interoperability, you might not need to implement all of your data governance upfront, but you should plan out your general governance strategy very early in a Data Mesh journey. This one I kind of fall on both sides of this. I think a lot of people are using governance as an excuse to not get going, but at the same point, if you're just thinking about each individual data product living on its own and driving value on its own, you're kind of missing the point of the mesh. The mesh is that it's not... Service mesh isn't about a bunch of services. It's about how these services interoperate. That's why Zhamak titled it Data Mesh, right? And finally, number 15, access control is a really hard problem. Many organizations don't have good communication or visibility into who is using what data, and especially the kind of how and why. We need to be asking these questions, and then setting access policies that expire too. We should check in to see if people still need to access. That's just good governance. It's not set it and forget it. With that bottom line upfront done, let's jump into the interview.

Okay, very, very excited for today's episode. I've got Vikas Kumar here, who's the AVP and Head of Data, AI and ML at CNA Insurance. But to be clear, he is only representing himself. We're gonna be talking about a lot of different things. We're



Transcript provided as a free resource by



gonna be talking about something that I don't think we've had anybody really state specifically, but it is kind of a trend that I think a lot of us have noticed of the three phases of data, it's kind of, "Data was important," to now moving data to the cloud and we're really starting to focusing on actually unifying the data and sharing information instead of just having data somewhat available, but not really usable and not really leverageable and we're not getting the full value about it.

And then we're gonna talk a lot about data products and Vikas' specific view on where we have to go with that, how we can do Data Mesh and Data Mesh, so many people think of it as your strategy versus your North Star and that's... I've said that's silly and Vikas, you said you agree that it's not your data strategy. And so we're gonna just be talking about a whole bunch of different things. But before we jump into that, if you don't mind, if you could give people a bit of an introduction to yourself and then we can jump into the conversation at hand.

0:10:16 Vikas Kumar

Oh yeah, thank you, Scott, and thank you for having me actually. Really, really excited about this opportunity. So, as you rightly said, actually my background, I have around 25 years of IT experience and not throughout this one... It wasn't the data experience, I started like everybody back in late '90s and Unix Admin, Ingress, Sybase. So I worked on many, many different technologies in my career, even COBOL, if some of you might even remember these technologies. And from there on started Java and somewhere around a decade ago, I moved to data. And as Scott was saying briefly is that back in 2000, data was deemed as, we require data, let's use data to determine what happened in our business and make use of those things, but after 2010, people really realized that this is really something, is so exciting and we can utilize this to accelerate our business in meaningful way.

And then from there on, people started to realize data is important. Let's move to that, and that's where the AWS of the world and the cloud of the world came into picture. And we moved off from DB2, IBM Mainframe, and Oracle and Sybase into let's put that in cloud. We can extend that and scale it there. So that's where the data cloud journey actually begins in my point of view. And cloud providers rightly so and essentially so, start providing a scale and high availability of these things and here comes the data ecosystem all together coming together where you have AI, ML coming up into picture and everything all together coming into picture. Now from 2010 to 2022, most of the focus was on moving data to cloud, and when you move it to cloud, that's where you start to generate value from the cloud.

So you start to give you a good value proposition from this. And from there on what I see is now happening in 2020, is people try to see, okay, how can we scale this further? And Zhamak rightly said in 2018 in one of the videos I've heard about it is it's



Transcript provided as a free resource by



all about scale. It's all about how you make the data available to the business, so they can make data-driven decisions on top of that. So it's throughout a journey, if you look at it from 20 decade from... Data was important, important, but we can't really scale it, so that's where the cloud comes into the picture, so let us scale the data to the cloud, and now from cloud, a lot of technologies are emerging to scale to fuel that further and promote it further to a place where we can start to involve business in this, and that's where I think that the Data Mesh is really scaled.

Data Mesh is really talk about how you scale, how you make business and domain participate in this. It's not about just data people, it's all about business and bring them together, and it's all about scaling it further. So I'm thrilled and excited about this, I've been part of this and we went through this journey, but I think also it's all about how you share this data, it's not about scaling, how do you scale the data, how do you make sure how you share data among not just enterprise without sort of enterprise, so really you can bring the value from this data. So if you look at the journey started in 2000. Data is important, yes. But people really didn't know what they can do with the data, and in 2010, they talk about cloud and in 2020 is more of AI and ML and scale and altogether. So I've been part of it, and I think it's fascinating how this has evolved so far.

0:13:55 Scott Hirleman

And I think one question I would have in there is... Was there a specific thing that you think is the tipping point? Is it simply that... 'Cause you said once we've gotten the data to the cloud, that means that we can do more with it. So is there something where you think that we're still missing? Like there's that, have we already hit the acceleration point of figuring out, we're focusing on that unifying and sharing the information, or are we still kind of figuring that out and is that people process or is that tooling basis or something that you think is still missing for us to really accelerate that? Or do you think that the tooling is in place and the people processes, we've kind of started to figure those out, and it's just, we've gotta have people that are kind of focusing on it and actually doing the thing, or are we still kind of figuring that out, is that gonna take another couple of years?

0:14:52 Vikas Kumar

Yeah, I think what you highlighted is very important and I want to, actually like to expand on that one a little bit more. So in 2010, until 2010, when the cloud wasn't the thing, wasn't the accelerator thing, data people tend to focus so much on technology aspect as well. How do you scale my own platform? Like I've dealt with DB2, DB2 drivers, DB2 versions. People just focus on data to... We just manage the data, they didn't know what they can get out of the data, if you know what I mean, right? So now with cloud, their focus is shifting from not managing the data is to get the value out of the data. So a paradigm shift happens all of a sudden and gives you the



Transcript provided as a free resource by



capability just so that people... Like early days, there were DBAs. How many of you know we have DBAs now, it's more about data managed by somebody, you just get the value out of the data. The manageability aspect has just gone.

You still do certain things, but you don't spend 60%/70% time managing the data. I would argue now you spend 60%/70%/80% time to get the value out from the data. That's 10%-20%, it just goes into the block on the cloud, manage it for you. So that's where the tipping point I saw that is basically, it's just the data is available for you to now it's available in front of your eyes and then you start to manipulate that, you start to see meaningful things out of that. So big data started from that. I worked with Cloudera at that time. Now, big data has started, now the cloud has started. NoSQL. There was nothing NoSQL at that time, right? So it all started and people started to see value from this and see okay, how are you gonna structure this data? There was no such thing as NoSQL in 2000. It's all about how it's propelled together. So instead of going from managing the data is how you are getting the meaningful insights out from the data, so that's a paradigm shift for me.

0:16:42 Scott Hirleman

And do you think that... So, somebody I work with, Sam Ramji, DataStax, has talked about... I think the 2000s to 2010s, he calls the time of scale out, I wanna say scale out compute, and then the next one is scale out network. I can't remember which order he goes in, but the 2020s is about scale out data and what you were talking about there is something where I think there's a lot of promise, but there's also a lot of risk in that because we are able to... Just because we can doesn't mean we should.

0:17:20 Vikas Kumar

Right.

0:17:21 Scott Hirleman

Like when we think about the cost benefit of how to do data, I've started talking to some people about this online even about measuring the value of your data work and things like that, and people just keep pointing back to vacuum statistics about like, well, what is your usage, is the usage of this above something else? And it's like, well, does that matter? If this thing is used once a quarter, but it's used to make an incredibly strategic decision and it's incredibly accurate at making that incredibly important strategic decision, it doesn't matter that it's literally accessed once a quarter, you probably don't need to update the data in real time, but it's a value. Do you think... We saw this a lot when people move to the cloud of wasted resource as a former cloud cost manager, oh man, there was a lot of wasted resources. Do you think that that's going to be the thing that we've got too much at our fingertips, or do you think that you're seeing people are doing this in a systematic way where they're not going, well, I can, so I'm just gonna go kind of crazy or... What are your



Transcript provided as a free resource by



thoughts there?

0:18:35 Vikas Kumar

I think what you are touching on is a two fold, I think one is you still have to as you rightly said, the scale is there, but you still have to manage that. Right? And then second is I think I can guarantee and tell you that even the cloud gives you the scale and data availability. There are many things which you are still not doing with the data. I think it all comes down to do you really understand your data? I think that's very important, right? And that's where the business domains come into the picture, right? We as a technologist can give you the data and then all of the grains about the data. By the end of the day, how do you slice and dice that data so that business people speak a language so they can really, the language they are speaking and you are able to suffice the data for them.

So I think I see this two fold. One is the scalability is there, at the same time you have to be really careful about it, that you don't just go and open the... Generate so many EMRs and scar cluster and things like that because you think everybody's doing it, let me to also do it. It doesn't have to be that. And as you rightly said, right, as long as you... And the way I see it, you work backward. You start to see what business is looking for, and then you backward towards and then generate the data rather than I'll give... I do all of the things like this company is doing AI/ML and GPU, CPU, do you really need that? We don't need all of that. Unless there is a need for that. So I think it all comes down to, and we are gonna talk about data product in a few minutes. But I think it all comes down to do you really need that? If you really need that, then work about it.

0:20:21 Scott Hirleman

Yeah. And I think one thing that's coming out of that that has been a little bit of a through line, but I'd like to get your thoughts on as well, is the cost of trying something and the cost of iteration. When you're on prem, you had to order all the servers, you had to kind of tee everything up. You had all this CapEx and all this planning versus cloud, you can spin something up and spin it back down. But in data, the concept of spinning something down means it was a failure. Versus where we are from a microservices standpoint, you can spin something up, you can try a new feature. If it doesn't get any traction, you close that feature down. That's product thinking. That's trying, iterating, taking in the feedback what's actually happening. Do you think when you're talking to people and when you're thinking about this yourself, do you think consumers are ready for that evolution to happen? Because anytime there's been a change, it's been a breaking change.

So change is always bad when you're a data consumer, 'cause it means something broke upstream or you've got something new. But there we haven't had that kind of



Transcript provided as a free resource by



product-y type of thought process. And are data producers ready for that? Are the data engineering teams, the people kind of in the middle, in Data Mesh, we're obviously trying to move away from that super centralized team, but even in early days of all these implementations, there's still some centralized teams. So like where do you think... When the rubber meets the road, do you still see that as a major challenge? Do you think that's a mental challenge, a people process challenge? Do you think? Like how do we get through that? Because that seems to be a through line in all these conversations when I talk to people, especially people who aren't in Data Mesh, the concept of change is frightening and terrible and bad versus we have to iterate and understand and experiment.

0:22:23 Vikas Kumar

Yeah, I think it comes down to a concept, which I like to say is a data with context and data literacy. I think your journey has to incorporate these things, without that businesses speaks business language, data speaks data language. So there has to be a combination of these things. You have to blend this together, then only get out, you can drink that smoothie out. So I would say that it has to be a... When you move the data to cloud and whenever you are working with business, you have to mix the context of the data. So content has to have context around it. And you have platform which can associate, doesn't give the data literacy to the business consumers.

So they can give you a feedback loop, which is, it's very important to get the feedback loop for business what you're doing. It has to be a two way street. Doesn't have to be one way street, otherwise you won't, you're not gonna be successful in this. It's gonna be a constant churn around this if you don't do that. So I think not just you have to involve business from what they're looking for, but constant feedback is very important, so that you get the feedback loop and Data Mesh actually allows you to do that. That because you have to work with business and make sure that they are seeing what they want to see from the data.

0:23:39 Scott Hirleman

So on that side, who has to move, right? Is it that the data people have to move a hundred percent of the way? Or is it that the business has to move 20% of the way and the data people have to move 80% of the way? Or is it getting the data people out of the conversation? And it's an entirely different paradigm or... Because so many of these things, it seems like the advice is... I had an episode with Scott Taylor, the data whisperer and the title of it was pop your data bubble and stop the data babble. And it's exactly what you're talking about. But at the same point, it feels like we're saying this is the reason people haven't been able to use data is data team, it's been your fault versus we haven't developed these processes, we haven't had the tools and the capabilities to do this. So like...

I don't wanna want to throw the past under the bus. But at the same point, are we



Transcript provided as a free resource by



saying data people, you have to do a hundred percent of the work? Because that's where I feel like people are really bucking and saying, no, I don't wanna do this.

0:24:43 Vikas Kumar

Yeah, I think you bring up a good point. The way I would like to see this is as an operating model too if you will, right? There has to be operating model and a data product mindset. And data product has to be from business side who really understand and gives you the input for the teams, the decentralized team. Hey, I want to... I want to see metrics X, Y, Z, but I want insert in this way and I want this frequency, right? So I think it has to be in an operating model, when I talk about feedback loop, that's where the operating model will give you that.

So it has to be... I don't know whether there is, you can say 50/50. I think it's whatever works for your organization, I would say, but identify who can give... Who can be part of this operating model so they can communicate to the business and give you the inputs because what you are building makes sure it's consumable, as you said, right? Producer, data producer has to make sure at the end of the day, right hand side the way I see data gets generated from left hand side, it gets generated or consumed on right hand side. So you have to really make sure the right hand side people are in the mix all the time. Otherwise it's gonna be very difficult to be successful in this journey.

0:25:56 Scott Hirleman

And I think this brings up something that's... Marisa Fish on her episode mentioned this and I kind of extended her framework, but why are we exchanging information? What is the point of this? Exactly what you said in there of like, we need to get very crisp on being clear, what is the point of this information exchange? Am I providing you the data, the raw data in some form or somewhat cleaned but rawish data so you can create your own insights so you can inform your own opinion based on this, you can combine it with other things? Am I giving you the insight itself and then you have to do the so what or are you expecting me to do the insight and the so what, and I'm just presenting that at you, that you're not doing anything with it yourself other than maybe going and executing on it?

And that we have to get super crisp about as the consumer, the consumer, like the... I feel like so much of this needs to be on the consumer to make it extremely clear, what are your fricking expectations? What do you want from this? Do you want me to literally do your job of telling you the so what? Okay, I'll try and do that. But like, you need to tell me instead of, you need to say, this is what we're trying to do, this is the data we need. And then it just kind of sits out there because the data's been created, but nobody is in charge of analyzing it or nobody's capable of actually doing the analysis to get to the so what on the execution part.



Transcript provided as a free resource by



0:27:29 Vikas Kumar

Right. So I think it all comes down to the North star as well, right? I think it all depends what... In the data journey where you are as an enterprise, right? I think some enterprise, I would say that they can give what you just said. I can give you curated data. I can give you unified data, I can give you raw data, and I have all of the platform and capabilities I can give that. Some may have to just start, let's start minimum, let's start a... Let's talk to business and find out, "Hey, I can give you curated data and if you need more than that, we have the flexibility and ability to give you that". So I would say it all depends upon the scale of the organization and where you are in this journey. It doesn't have to be start from the North star.

You start it as much as simple as, okay, what are your needs? And then identify their needs in terms of data availability, and then in terms of frequency and things like... And then also what type of use cases they're trying to solve. It's very important to understand use cases as well. Are these analytic use case? Are these AI/ML use case? Because I can argue that AI/ML use cases might, you have to go all the way to raw data to flesh that information out. They just can't get insight from the curated data because the nature of the business they are in. So I would say it's not a single answer for that one. It all depends what business is looking for. And also and the scale of your organization. I think that's where the Data Mesh, I would say that people in the Data Mesh journey wanna say, "Hey, let's do everything what Data Mesh tells you". No, you don't have to do that. You have to start small and iterate over this and see how it's working for you. So that's where I would think that it should evolve from.

0:29:08 Scott Hirlleman

Yeah, and I think the data people wanna play with their toys, right? I don't want to throw people under the bus, but it's kind of the case. And so a lot of this is focusing on the what matters, like what are we trying to do? But I think that making that implicit explicit, right, of having that conversation, I think this does transition well into like, what are some examples? We were gonna talk about like what are the signals that you should extend a data product versus create a new one? And I think this kind of comes into that of like, what were you trying to solve with this? Do we wanna create this frankensteinian monster of a data product that's trying to solve 70 different use cases? Like how do you have that conversation? How do you think about that versus, the consumer doesn't necessarily care as long as they get what they want.

And so they don't care about the mechanism. So like, how are you thinking about that and how are you having that conversation with them to not just say... You know, they say, well, isn't it easy we just... But you've got the data. Just give me the data. Versus like how are you thinking about that process of measuring where are boundaries? Where should we do this? Like how do we go forward?



Transcript provided as a free resource by



0:30:28 Vikas Kumar

And actually the one you just said, I would like to say it's pretty clear in some people head, it's pretty gray in some people head. So I think people struggle with the data product concept. And I thought maybe a simple analogy might be helpful to provide an insight about this. And since I worked on Unix I came up with this, some analogy is in Unix's word the way I see all of the Unix command, you can treat them as a product. For example SED is a said command of many of users is Unix is a text stream editor, which allows you to identify quickly and efficiently, but you can pipe with something and get something, some other capabilities out of that. So what I'm trying to say here, and same for AWK command in Unix, AWK commands in Unix allow user to process and manipulate data and produce in different reports format. But if you pipe with something else, you get different results out of it.

So what I'm trying to say is you always try to map a data product with some sort of outcome, but it doesn't have to stop there. You can always extend that outcome to something else. So start with something which gives you the meaningful and measurable output and meaningful output to the business. But you end there? Do you stop there? No, it doesn't have to be, you can extend it further with some other capabilities and expand it. So that's how I see it. And I'm sure there are different camps of thoughts about this one. People see it differently. People say, okay, what, where should I start? How should I start? And people talk about platform as a product. People talk about data set as a product, right? So they have different definitions around it. The way I like to see it is as long as a product, data product gives a insight to the data, to the business, what they're looking for, you can call as a data product.

0:32:26 Scott Hirleman

Yeah. I think that within Data Mesh, I've used it for a while and I've kind of moved away from it, but like, stop using the phrase exact data product because data product means so many different things to so many different people.

0:32:41 Vikas Kumar

Exactly.

0:32:42 Scott Hirleman

But what you're talking about is true of like when you think about a product, it has a point, it has a reason for existing and so much in data... Like, do you create a data product that is a test of a... Is this valuable and then you shut down that product quickly? People in data don't have that concept. You know, Monzo Bank, I reference this a lot because it just kind of made... It sounded so crazy to me. They had at the time about I think 2500 people and they said, we have 4500 data assets in



Transcript provided as a free resource by



production in a bank of 2500 people. And it's like, okay, how many people are engineering or data people, let's say even it's over a thousand, right?

Still way too many per each person, right? And are you tracking the usage or... They were so super proud of their lineage and it was like, that's great because it is a mess for anybody to track what's going on. So like what you're talking about there of that you've gotta have a point for the data product, that you need to think about that. Have you found that that communication goes well to the business people? Or have you found that they're kind of not in a lot of cases in that product thinking as well, and you have to explain kind of, and it's both sides learning together, or is it that they... A lot of times when I talk to business people, they get business concepts and product as a business concept, so, yeah.

0:34:18 Vikas Kumar

No, I think again, and it comes down to you should have a product owner for this data product. It's very essential to have a product owner for this data product. And their job essentially is to identify what's the objective of this? What's the end result of these? What are we trying to achieve, and what are the characteristic of this? What are the fundamental of this? What are the governance of these, right? How do you wanna scale this? Or who's gonna use this? What outcome you're gonna get it? I think once you identify, and I haven't listed out all of it, but once you identify some of these things, I think you start to see the picture and you start to see the value. I think the value proposition has to be there to begin with. Otherwise you're gonna waste time and you're gonna throw in the trash. I think the value proposition is important, right?

And it also comes down to, I can give you from my experience they are consume... Plus business users go to 10 different things to perform one exercise. Can you combine into one? Can you give them one interface to do that? That's itself a value proposition. Instead of going into 10 different things, I'm just doing it from one place. So that's a value proposition for the business. They don't have to toggle to different places to get just one meaningful information out. It can be as simple as that. And I think a product mindset has to be there. You start simple, you iterate over this, you prove the point and you say, "Yes, this is something we can expand upon". And then you iterate over, it's a journey, product is a journey. It's not a project. So you have to iterate over this and keep adding things on top of that. So business has to be part of that, what I'm trying to say, and the product definitions and product ownership has to be part of it as well.

0:36:05 Scott Hirleman

Yeah. And one thing in there is, so we talked about it a little bit earlier, but data, historically we haven't had that product life cycle and things. And so we're seeing it's



Transcript provided as a free resource by



crucial to see that this has a value, but that value can be that we tried it and found out it doesn't have a value. Is that a mindset that you're having people like, that blows their mind of like, the value was in figuring out, does this have a value to continue with it forward? And the second you find the answer is no, or that the cost exceeds the benefit, that you shut it down and that that's okay. Is that something that you're finding people are actually willing to engage with that thought versus you're telling me, I'm creating a product and I may have to kill it?

You know, I mean, even in data, you're a AI/ML side, right? So you're familiar with the feature factory side of AI and how bad that that can be. Like, we kind of have that on the data side too of this is an interesting "insight." Is this what we actually call an insight? Is this ongoing something that we should continue to do? Like, are you finding that people are okay with the idea of testing to see if something has value and if that testing has value, but that it doesn't have to mean that we actually end up with a net value versus we try this... Like, should we enter this new market? We went and we did all the information work and we found out we shouldn't. That has a value from preventing a bad business decision, but will people assign a value to that versus, well, it didn't mean that we netted a bunch of revenue or things like that?

0:37:49 Vikas Kumar

Yeah, I think it's interesting question. The way I would like to see this, I think when you start a journey of a product, I think you, of course, start with set of goals in mind. Well, I want to achieve this X, Y, Z goals with this one. What I see ends up happening sometimes is this X, Y, Z goals might be present in some other data product, so that's where I see the mindset shift. Okay a data product, which we thought will do X, Y, Z, but there is another data product is already doing X, Y, Z plus, and then should we keep them together? Should we combine it? What do we do around those scenarios? So I think the way I see this is evolving, in this context for me at least.

So one of the things we have to make sure we're not duplicating the product, and that's where I think what I'm saying is, well you might have an opportunity to combine these or get rid of one if you have to say. So I think what you have to be very very careful about is make sure that only a single responsibility principle is there so you're not duplicating things into whether it's a data set, whether it's capability, whether it's a product feature, whatever it is, you're not duplicating this. That has to be part of that conversation as well. So I haven't seen anything like that you truly trash the whole thing, you may learn something from what you build, and then if you think there are other things you're building in this and the other thing can incorporate what you just built previously, so be it, let's do that.

0:39:17 Scott Hirleman

Yeah, that's interesting, I think I am hoping that we see more people that throw



Transcript provided as a free resource by



things in the bin, because I do think that we have to get realistic about ongoing costs and that tech data debt, data debt can be that you did something wrong, but it can also just be, you've got something that's running where the usefulness is not there anymore, and it's okay to pivot away from that, and I hope we can get to that, but it's not a conversation that people are willing to engage right now.

0:39:50 Vikas Kumar

Yeah. And then you very easily can get into SOL with the cloud, because it gives you the capability to just quickly turn on the things and do it. So you have to be very careful about, as we talked about in the beginning, cloud gives you the capability and flexibility and scalability, but it also can easily give you duplication things. There are duplicate things out there, and people are not governing those things out there as well. So you have to be careful about cataloging these things, providing meaningful definitions around this, make sure you provide the context around this, make sure who your API consumers are. It all comes down to governance process as well. And the operating model which we talked about.

0:40:32 Scott Hirleman

Yeah, I think the question around white space of is it worse to have two things that have a bit of overlap? You don't want them to have a ton of overlap, but if there's white space where you're not covering something that could be important, is the balance between white space and overlap and how we think about that, but I think that's a very broad philosophical question. Happy to give you space to react to that, but I also think this transitions well a lot of what we're talking about of focusing on the long term instead of the short term, a lot of people when they're thinking about data work, they keep saying, go for the low hanging fruit, go for the winds, go for quick time to value. And if you're only focused on these quick times to value, you're getting just the small point total.

Whatever sport you wanna use as an analogy, that if you're really only focused on these kind of small wins, in the long run, you're losing out to the people that are focusing somewhat on the long term and still getting that short term to get the momentum, but how do we actually give ourselves the space and the room to focus on that longterm when we could just keep racking up these small wins that are pretty easy, but that don't play into the greater whole? And that you're not betting all your thing on one kind of long term bet versus multiple long term bets, but how do you think about balancing that? How do you think about having that conversation, especially the world is now quarterly driven, right? It is so much quarterly financially driven, so how do you think about having that conversation to make it so people are okay that you are focusing more on long term things, but that you're not putting all your eggs in one basket. Sorry. Complicated question.



Transcript provided as a free resource by



0:42:25 Vikas Kumar

No, I think it's a fascinating question and the way that I would like to see is... Let me try to answer it. The way I would like to see is you create the base, data basis, not data basis, data base. I would like to call it more of abstraction layer, for lack of a better word. You create this abstraction layer, virtualization layer, whatever you want to call it. And then that has to be built with the product long term mindset. On top of that, once you have identified that base, what I mean is you curate the data in such a way that you can quickly, hey, let me utilize this 20% to get quick win out of that, let me utilize a 20%. What I don't want people to just say, hey, let me just curate this data for just 20%. I think your mindset should be when you get the data into the cloud, you look at it holistically, you look at all of the lines of the business data.

You look at it universally, all of these things, because you can quickly get into the silos and then you cannot be able to interoperability, you will miss the interoperability if you don't look at the holistic picture. So I think interoperability and the holistic picture with abstraction data is very, very important, and that's where the long term comes to the picture. So everybody, whoever is doing that, I would say based on my experience is, look for opportunities to holistically design the data in such a way that you can quickly turn that into value streams with small data products. So your data product base essentially is very important from your building piece, on top of that. I would like to see more of a pyramid. If you see the pyramid, base of the pyramid is the base where you're building these up. It doesn't have to take the pyramid shape, you can quickly iterate it over on top of that, but it's very important and also it's...

Since you're asking me that, in my experience, somebody asked me, can you give me intelligence data? Let me ask, okay, how many accounts are there? So accounts. In any company accounts can be many things and in many places account exist. So what do you do about this use case, right? So what I'm saying is you build an abstraction layer, you bring the data sources from multiple cases into this abstraction layer, for a given LOB, and you make intelligence, that layer is intelligent enough. And then that's where the context and APIs come into the picture. Hey, I'm gonna give you this query, can you give me the data out? I don't know where you're gonna give from this data source, but give me that information out. So that's where the long term come into the picture. You create this abstraction layer and then you derive the data from there. Does it make sense what I'm saying?

0:45:14 Scott Hirleman

Yeah, and it really does. And I think my question from that becomes, how do you build a base of these data products? I've kinda talked about in your phase one of Data Mesh, you're putting out initial data products around your use cases, and then once you're heading towards phase two, you start to think of your suite of data products around certain capabilities and things like that. I like the concept of a base.



Transcript provided as a free resource by



I was gonna say, what about a building's foundation? But you're building multiple buildings on top of what this is in this analogy, the building, it's like you have a foundation for a block and they are all built on top. But one thing that I would ask you is so scalable and flexible and agile, not capital A "Agile," but ability to kind of change with the times quickly and not lock yourself into anything. How do you build that base so that... Or how do you think about building it so that you aren't tied into anything? And as new things emerge, you don't have to create an entirely new base every single time.

This is what I think around taxonomies and ontologies and things like that, where Zhamak is so... I don't think she's against them as much as she's against the way that they have been used because they've been used as these very, very rigid structures. And I'm trying to think of who... I think it was Alla Hale who was talking about, it might have been Elena Samuylova who was talking about bridges and engineering and that rigidity versus flexibility, and that you have to have that balance, otherwise, if something's rigid, then it's very brittle, but if something is too flexible, it can't support anything. So how do you find that? How do you create that base to build on top of, but that isn't like these are the only way that this can be built on top of versus give the capability to shift things.

0:47:21 Vikas Kumar

Yeah, I think it's a good question, and I think if you remember what I talked about, about the data product, extendability of data product, you can extend that, you can extend from that base. It doesn't have to use, everything is attached to that base. So what I'm trying to say is you have a foundation, but that foundation is extendable to a data product. So it's become detectable from there, or what that base will allow, your foundation allows you to do, to quickly develop these things, quickly iterate these things. So you're not struggling to find out, okay, from where I should build this from? That's what I was saying. So it's not a big gigantic shape, but you're still building a smaller shape of this product for the analogy purpose, but you knew a product type, you knew a blueprint and from there on, you're building it up.

0:48:11 Scott Hirlleman

Yeah, I'm thinking about a lot of the things people are talking about with templates and things like that, that even what you're talking about is that kind of templating thing of like... We have something that makes it so you can get 80% of the way there quickly. If you don't, if 40% of it doesn't actually even work, we're gonna make it so that you can throw that out, that it's not that if you throw out the 40%, it blows out the other 40%, but that you're focused on making it so that there are... The whole building block analogy and things like that, that there are many different of these components that are loosely coupled to each other, but that gives you that capability to quickly get to a value stream or get to some value from the value stream as you



Transcript provided as a free resource by



iterate. Right?

How do you get something out there and capture 40% or 50% of the value in two weeks as you build towards it actually being a product over the next three to four weeks, and you capture the rest of that value, that you get that out there quickly. And you get the feedback 'cause it might be while you were thinking that we were going northeast and instead we're gonna go northwest, and it's like, oh okay, we started to go northeast and we went three miles along the path, and we talked to them and said, is this where you wanted to head? And they're like, no. And so then that divergence, that move is much less than if you went 150 miles and then you've gotta go all the way, you've gotta go 100 miles back across to the northwest. Yeah, I really like that concept. Do you have any advice on how to actually figure out how to do that? Because it's great to say do this, but somebody trying to go, okay, Vikas I want you to tell me how to actually do this, do you have advice on building that extensibility.

0:50:04 Vikas Kumar

I think it also comes down to value stream in terms of platform capabilities, your governance models, and how the actually data is flowing in. So I think once you establish that ecosystem, I think things will start to come up. And then also I would like to also see that this product has to be driven from a stream where you're quickly doing the prototyping and POCs and proving out that, okay, yes, this is really what you want to see, and more of an assembly line, a factory approach where you're building these small small parts and you're building this all together.

You quickly realize that oh this is something I want to build, or this is something I don't want to build. So I think I would say this is more of Agile methodologies and quickly driven from each part and process, but that multiple teams are contributing to it, so they're quickly realizing and giving you the feedback, hey this is something I wanna build it. Right? So think of assembly line concept, but the many part come together from many places, what you're doing is create a small POC, I wanna build this. Yes, this makes sense, let's continue to build that and then prove it to the assembly line. That's how I see this.

0:51:20 Scott Hirleman

Yeah, you're just unlocking a lot of through lines, that then connect a lot of these conversations for me, 'cause I think... We talk about be extensible, do this, do this, do this. But I think exactly what you're talking about of like, you can't lock yourself into too much, but you also can't just do everything custom each time, so you wanna find more and more of these places where, what are the things that we found that worked? Let's automate those, let's consider that toil for 80% of the people, so let's figure that out, but if this was toil for 10%, we're not gonna automate it. We're gonna



Transcript provided as a free resource by



say that if you run across this, we've got something that might help you, but we're not gonna automate it.

We're gonna do that and just... It seems like... Yeah, and that value stream approach. How have you found thinking about searching out these value streams? The divining rod, the little stick that you hold that is supposed to help you find water but doesn't actually. But how are you looking for those? How do you evaluate, is this value stream worth it? Is it something where we just wanna go with a big old bucket and just try and grab all of the water once and move away from it 'cause that's all the value of the water in that one thing, versus we wanna tap the stream and put a waterwheel on it and get some power from it? How do you think about that?

0:52:51 Vikas Kumar

Yeah. I think the way I would like to see is value stream is one component of your data product. You have many streams and value stream is just one component of your end result of a data product. So once you establish value stream, and then you establish this value stream is doing what it's supposed to do, as you said rightly about you wanted to fill the water in this, but make sure that water is not leaking and it's carrying forward all the way to the end. And all of the value from whatever you're building it is capturing and kept there. So the way I would like to see is you have a certain... In the export analogy on this, let's say you got a union station, and there are many, many tracks come into the union station. The union station is a value stream at the end of the day. It's a product you wanna build it. So I think the way I'd like to see analogy perspective, there are many things coming together.

But the end of the day, they had to come to become a tool or a product or a data set, whatever you wanna call it, which you can consume it, which you can utilize it at the end of the day. So this value stream, essentially is a component or a vehicle of that journey, which gives you that capability to, when you reach to the destination, then you can start to see the value from these things. But the value stream could be a stream of streaming work, a stream of governance work, a stream of data work. And then you quickly do the POCs along the way and quickly realize that, A, by doing the POCs in Agile way, you quickly realize that whether you are producing the result which you want to produce at the end of the day, and then that's where I was talking with a feedback loop to the business. It essentially is a need as well. So you need to make sure that whatever you're building it, and the business is giving you feedback along the way, so you make sure that whatever you're building is consumable, and then give you the value out of that.

0:54:49 Scott Hirleman

Right. And so I think I've got one aspect that I think is correct, and I think I've got one where my brain kind of went that is completely incorrect. So I would like to do the...



Transcript provided as a free resource by



So what you're saying is each product should have multiple value streams, and that there isn't a single... If you've got a single small scale point, that's probably too small for a singular product. Maybe not, but probably. And so you wanna think of there being multiple value streams 'cause there are multiple capabilities of a product. 'Cause otherwise, what you've got is a feature. And then the other aspect of it was when you first started to talk about that, I was almost thinking you were saying that the value of the product isn't the point that you were trying to create from the product. And I think I've got that wrong because I think you moved into the multiple value streams. But I wanna make sure that I don't have that right, that there's a much bigger point to the product than what you were building it to achieve. I don't think you were saying that, but I wanna make sure that I...

0:56:00 Vikas Kumar

No, I'm not saying... Yeah, no. I'm not saying that. I think that the stream is essentially is a component, that's all. I may have maybe not articulated correctly. Yeah.

0:56:08 Scott Hirleman

Okay. No, you got there and that's where my brain went and I was like, "I don't think this is what he's saying, but this would be a really interesting point that I really wanna dig into" 'cause my brain cannot comprehend how that would work but I wanna know if he's saying that. 'Cause I think you're dropping some pretty good information on us and some pretty good thought processes, so I was like, "If he's saying that, I need to know more 'cause my brain can't..." "

0:56:36 Vikas Kumar

Let's dig into that more.

0:56:37 Scott Hirleman

Yeah. So, we've covered a whole bunch of different things. Is there any... We've still got 10 minutes or whatever before we generally look to wrap up but is there anything that you wanna head into as we're generally wrapping up the conversation? We were looking at kind of maybe empowering the domain with resources, or how do we actually get people to iterate towards good and not be perfect and things like that, or the self serve platform? We were looking at doing a bunch of those different topics but is there any one that you think is really important that you want, that you think people have... It's been undercover or you think people really need to know, or?

0:57:24 Vikas Kumar

Yeah. I think self serve platform is interesting one. I think a lot of people undermine the importance of that, I think. And not many people focus on that. I think it's something and because... The reason I'm saying it, it incorporate governance, security, accessibility to this data product. I think it's very important aspect that every



Transcript provided as a free resource by



organization should pay a heavy amount of time there, in terms of, how do you want to achieve that? Because there are many ways to do that. Right? You can do as simple as you start YAML or JSON or whatever you wanna do it, and then you tell the form and do that but you can go as fancy as you create own product to achieve that.

So, I would say it goes back to our previous... Beginning of the call. How do you achieve self service platform to the business? I think it's very important that you create something which is just clicks. You just do clicks and then I have the self service platform ready for the consumers and for the business but also have the security, it also have the governance, it also have the context. And there are many ways to build that. I think people try to establish... There are some tools out there which does something, but there are many... And cloud gives you different capabilities to build this. You have to be savvy about what is out there and how do you build such a platform which gives you the flexibility because I can tell you many organization have small, small capabilities, but they don't know how to bundle this together to create a self service platform.

0:59:10 Scott Hirleman

How do you think about risk when you think about self serve? And not just that somebody gets access to the wrong... To data they shouldn't have access to, versus, Ust Oldfield that made this point and it's really started to sink in more and more as I've talked to more people, that if somebody doesn't know what they're looking at but can get access to many different things that they don't know what they're looking at, they can make a lot of real bad decisions based on data they don't understand.

So, how do you think about building in? Is it building in some breaks to make sure that people are not just going and poking at things and going, "I now know what this means," or how do you think about balancing? Because we want everybody to be able to embed information, analytics, and data and things into their day today to improve what they're doing but in a lot of cases, they can be information overload or they can completely misinterpret what that means, what the data that they're seeing means and make bad decisions based on it. So, how do you think about cost benefit there and how do you think we can maximize the benefit while containing the risk, or at least making people aware that there is a risk if they're not paying attention to what they're really learning?

1:00:37 Vikas Kumar

So, actually, you ask a couple of questions there, so let me just try to answer...

1:00:42 Scott Hirleman

Sorry.



Transcript provided as a free resource by



1:00:43 Vikas Kumar

Each of them. No, it's totally fine. So, one is a platform. Okay, so let's say I'm a data scientist and I want to access certain dataset, so I think you have a... When I talk about the platform, you should have the capability to define the functions in organizations. Okay, who is who? Right? That's very important and then when you build this platform that actually already capturing that information, who is Scott? It belong to this LOB and he operate in the dysfunction so we already know that, who you are. And there's the second thing which you asked, and actually, I don't know if you ask or not the other thing, which I would also say, "How do I know Scott needs his data access for one month, two month, six month, one week?"

So, I think it comes down to governance. So I think I am truly... I'm building a data, self service platform, I would incorporate that, as well. Do you need access... And I think one of the things people have struggled with, and I've seen it in past, is Infosec and other organizations, just... They don't know who's accessing what and what's the timeframe. So, you build a platform which gives you capabilities to decommission that access once that timeframe has elapsed. And it can be built. There are tools out there, there are the capabilities out there. It can be built and that's where it is and also the second portion of you ask... So, you just asked a people person, and I'm gonna talk about even the platform and services concept, as well. How do I know this service that's accessing this data, so, you give them service account, IAM and then all of that but how do I know that they are not just attached to all the time? How do you make sure you decouple those? So, there is a two component of that. One is a human component, other is just service component so you have to establish a platform which can decommission these once the time has elapsed. And then this also brings up an important question which I see most of the time is a production data access. Okay, I want to access production data and then many organization doesn't want to give the production access to anybody. So I think this platform also gives you the capability to satisfy those use cases, as well.

1:03:03 Scott Hirleman

Yeah, and I think one thing that's come up a lot is... So Sarita Bakst was the first one that really brought this up to mind which I kind of... The more that I dig in, the more that I don't love automated access, unless the data product itself is pretty harmless, if anybody in the organization gets access to it. And our back... Things change and evolve in a large organization. If you're in a small enough organization where a data engineer is a data engineer is a data engineer, or if you're in a giant organization and you treat all of your people as if they're exactly cogs and exactly the same, you're not doing Data Mesh right anyway. But what Sarita mentioned was there's sensitive information in data products, if anybody were to just go ahead and get access.



Transcript provided as a free resource by



So what they have is is somebody who registers their use case, and they tell the producer, "This is exactly how I'm going to use it. Give me the access for this, in this way, and it will get connected downstream and you have visibility into how I'm connecting it with other information, how I'm dealing with it," and that means that that producer is far more comfortable. And then that consumers actually signs a DocuSign based contract or whatever. If not, just, "Okay, I'll give you access to it and I trust that you're gonna do this." They're the ones that then take on regulatory risk, where that consumer is now the one that if they misuse the data, they're the ones that are in trouble with the regulator.

They're the ones who hopefully not go to actual jail, but who their organization within the greater company is gonna get fined and get in trouble and all of that. But that conversation, going between somebody saying, "I would like access," and making that happen with that conversation is incredibly quick and that self service is also about granting access. It's not about just everybody having access to everything 'cause that just creates ridiculous unnecessary risk and there's not a lot of reason to do it versus the kinda concept of making this just extremely low friction for people to exchange context and to figure out. And say the producer doesn't know, is that allowable by law or does that fit anything?

I'm gonna go to my central governance team and we're gonna have a conversation and I'm gonna share the context with them but I'm not asking the governance team to understand my domain. Where if somebody says, "Hey, can we get access to X, Y, Z column or whatever," and it's like, "Well, that's sensitive information." "Well, we're gonna use it in this way. We're trying to get access to everybody's phone numbers so that we can say, 'How many people have an area code that doesn't match their zip code?'" That's totally different from... And that's all we're trying to do. It's like, "Oh, yeah. We can do that and we can truncate it so you only get the area codes," or we can do stuff like that where it's just not an issue, and boom. It just makes it easy and golden and you move forward. I'm not even sure what my question is here, but like...

1:06:26 Vikas Kumar

No, I think what are you trying to... I mean, if I got this right. How do you make sure that right people get access to the right thing? I think. So, I would say that a couple of functions, if you are a data analyst, if you are data product, you are consuming some of it and you have it all but you don't want to... You don't have it all, but you can see social security number, date of birth, and things like that, for example. So, I think that's where it comes down to, your data products should have capabilities to filter those things out based on their functions or who's accessing what and you build intelligence around it because I can tell you that data scientists can really come and stay on your head.



Transcript provided as a free resource by



"I want access to everything because I'm running certain algorithm and I want to actually see. And I have been in situations where certain state, such as California, doesn't allow you to even see the whole zip code. So, you can only have... So, I've been in a situations where I can give... The easy way to deal with it you get first. So, you will be surprised to know that first three will give you a whole county so you're not pinpointing any individual person and their zip code. So, there are ways to handle such information, you have to be aware of it and you have to make sure you incorporate that functionality in your data product when you're building the self service platforms.

1:07:57 Scott Hirleman

Yeah, yeah. I think that makes a lot... I think the answer that comes from that, or the question of like, "Well, then how do we do it?" is we're figuring it out. It's tough, it's not that there's exact answers but there's stuff where you just look for the toil and you automate that out. And the toil, especially in Data Mesh, is gonna be stuff you don't expect 'cause you think that it is the tooling aspect or it is this or is that, and a lot of it's like just so many people are talking about, "Okay, the toil is that somebody feels like they have to design their schema, their data model entirely from scratch," and it's like we've just got a standard schema, a standard data model. Does your data fit this? Hey, yeah, we can bring somebody in to even poke at it and say, "Does this fit this," but boom. Here's the thing.

That we just find the ways that we look for the friction points because even finding who's been using your data and everything like that, people don't have it, and so when we actually start to look for the friction, we can't tell where the friction actually is. And so, I think having these conversations be a lot more like front and center that we say, "There will be friction, there will be pain. Let's find it, let's get rid of it but we want you to tell us," instead of, "We want to ignore you when you try and tell us that there's a problem."

1:09:24 Vikas Kumar

That would be... You just talk about a whole separate paradigm, schema, basically. Every industry have their one schema so do you try to fit everything? Do you live with 40%/50%? That's a whole set of conversation that I can talk about hours and hours on that but because I've dealt with such things.

1:09:43 Scott Hirleman

Yeah. There's a panel that we're recording relatively soon with some folks on data modeling and Data Mesh and my point around that is that I think we need to reinvent certain aspects of it, or we need to test certain aspects. But yes, there's also an easy path 90% of the time and it's fine to take that easy path.

1:10:09 Vikas Kumar



Transcript provided as a free resource by



I can tell you... Let me know if you want to be part of that because I can tell you that the biggest complex schema is the travel industry and I have dealt with it. It's huge and it's like you can't even wrap your head around with this one.

1:10:24 Scott Hirleman

Yeah, Carlos Saona is at eDreams ODIGEO and he was on and I kinda liked their approach, especially in the travel industry 'cause they have just a single domain event. That's the maximum size of a data product is a single domain event, so that way it is if people want to cross and create all of this stuff, it's all on the consumers because otherwise it becomes so difficult to try and put everything together all at once 'cause... Yeah, a trip means something completely different in nine different domains and so yeah, yeah. Totally.

1:10:57 Vikas Kumar

I dealt with three different product on the one travel industry. We can talk about it any time if you would like to. It's really complex, if you just combine it one. We have to separate. I think it all comes down to microservices versus monolithic. You don't want monolithic in travel. It will be nightmare for you to understand because trip could be something else in airlines, trip could be something else in hotels. It's very, very complex out there. So I would like to see as the consumer, if you can't filter it where it's truncated and have a schema for that, then it'll be good for you to consume it later on because you don't want when it comes and land it there, then you try and do schema, schema-fy things, and it will be very complex, so I would say that use case, you try to fix at producer side, not consumer. Actually at producer side and then have some sort of schema enforcement there.

1:11:53 Scott Hirleman

Yeah. And I think that's what they're making it as easy as possible to do like foreign key combinations but they're not trying to make it overly complex because that's exactly what you said, so. But, yeah. So, well, Vikas, this has been a really, really great conversation. Is there anything we didn't cover that you wanted to or any way you'd kind of wanna wrap up the episode in general?

1:12:14 Vikas Kumar

No, I think I would just wrap it with the Data Mesh. I think Data Mesh is trying to solve the scale problem I would say. And the scale problem and then not just the scale and how do you make sure that business is part of your journey, I think, domains are part of your journey? Because I truly believe they are the one who knows about the data better than you, the data people know. And they're part of the journey. If you empower them to make and self sufficient and they can be part of that ecosystem and that'll help establish the decentralized phenomena. I think that's where the Data Mesh is really scaled and then give you really true value out of that. But it takes a lot



Transcript provided as a free resource by



of effort.

It takes a lot of operation stuff. It takes a different mindset shift but I think Zhamak has done a fantastic job and I think we are all trying to put that out there that yeah, there is a shift, mind shift you need to have if you really want to do the Data Mesh. And also you don't want to be, as I talk about, not the start of that. You start small, you iterate over that. And then also, you have to really see whether you are a true fit for that, I think with any other data, in any product, you see what's the fit for this? If you are fit for this, I think you should truly adopt it, embrace it. But it's fascinating how people are starting to see this and I'm glad people are starting to adopt that more and more.

1:13:53 Scott Hirableman

Yeah, I get really concerned when people just talk only about a Data Mesh architecture instead of Data Mesh involving the organization and the domains and things like that, that it's like, it's not just a technology problem. Well, so, I'm sure there's gonna be a lot of people that would love to followup with you after this. Where's the best place to do that? Anything specific you'd like them following up about?

1:14:17 Vikas Kumar

I am on Twitter, I am on LinkedIn, so Twitter I'm @VikasKumar9 is my Twitter handle. Vikas Kumar Sharma is my LinkedIn profile. My email is with vikaskumar9@gmail.com. Happy to talk and then interact with people based on that. So all three platform I'm actively available.

1:14:42 Scott Hirableman

Yep. And we'll drop links to those in the show notes to make it easy for people to reach out, as well. But again, Vikas, thank you so much for your time here today and, as well, thank you everyone out there for listening.

1:14:53 Vikas Kumar

Thank you, Scott. Really appreciate it.

1:14:56 Scott Hirableman

I'd, again, like to thank my guest today, Vikas Kumar, AVP and Head of Data AI and ML at CNA Insurance. You can find a link to his LinkedIn and Twitter, as well as his email in the show notes, as per usual. Thank you.

Thanks everyone for listening to another great guest on the Data Mesh Learning Podcast. Thanks again to our sponsors, especially DataStax, who actually pays for me fulltime to help out the Data Mesh Community. If you're looking for a scalable,



Transcript provided as a free resource by



extremely cost efficient, multi data center, multi cloud database offering and/or an easy to scale data streaming offering, check DataStax out. There's a link in the show notes. If you wanna get in touch with me, there's links in the show notes to go ahead and reach out. I would love to hear more about what you're doing with Data Mesh and how I can be helpful. So please do reach out and let me know, as well as if you'd like to be a guest. Check out the show notes for more information. Thanks so much.