Transcript provided as a free resource by 

Data Mesh Radio Episode #32: Applying a Historical Lens to Data Mesh
Interview with Azmath Pasha
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?utm_source=DataMeshRadio

Scott Hirleman
Welcome to Data Mesh Radio, produced and hosted by Scott Hirleman, the founder of the Data Mesh Learning community. Data Mesh Radio is a vendor independent resource for learning more about Data Mesh. Let's jump in.

Scott Hirleman
In this episode, I interviewed Azmath Pasha, a member of the Forbes Technology Council, who has 25 plus years of experience in the software and data space, with a special focus on implementing large scale data projects, consulting companies like Capgemini and Paradigm Technology. We covered a lot, but one interesting insight was Azmath's belief that we should have some people in the loop aspect to data discovery. I'm seeing everyone trying to solve the data discovery challenge exclusively with tooling and it's causing a lot of strife. Maybe a data concierge type of role might be helpful to supplement the tooling. It's an interesting avenue to explore. Just beginning to think about how that might actually work, and how it might also not be very scalable depending on how many people you've got in your org trying to use the data.

Azmath talked about a few things he sees as crucial to getting a large scale implementation in the data space right. First one is the data product experience, what Zhamak calls the experience plane for Data Mesh. Azmath sees this as a crucial aspect for driving organizational buy-in. If your product, really any product, has a bad experience, it's gonna limit adoption. Number two is data discovery, as we mentioned a bit ago. Number three, staying tool agnostic. There's a lot of reasons for this, some of it is to "Future-proof yourself." Some of it is just to not get too tied into any one vendor from a lockin perspective, but also from a capabilities perspective. You should always be out there looking at what are we trying to solve, not what tool are we trying to use. Number four will be looking at your investments over a five year time horizon, not just right now. This will help you prevent from kinda constantly iterating simply because you weren't focused on the mid-to-longer term. Number five, reducing time to delivery for your data product producers, especially for their initial delivery. And number six, data democratization, broadly speaking. We've gotta lower that bar to using and accessing the data, and then we've gotta raise the quality

of the data that you're putting out there. Azmath had an interesting insight on the modern data stack. We're driving towards commoditization of the data ecosystem, and that's great and important for getting to a much larger data usage scale. If we continue to have too much friction in adopting, tooling or scaling our usage of tooling, we're gonna constantly be held back. So the more that we can commoditize and really move forward with the capabilities of scaling our usage of any one tool or adopting new tools, that's very crucial.

Lastly, we covered Azmath's three value measurements in data. Cost savings, business value, kind of driving new business and things like that, and data reuse. But Azmath believes we're entering a new era of using data and that we will need to find new metrics and measurements to assist the value of our data. I think you'll get a lot out of this from somebody who's got a more broad perspective rather than somebody who's super specifically focused in the Data Mesh implementation or on the Data Mesh space. With that bottom line up front done, let's go ahead and jump into this interview.

Scott Hirleman
Hi! Really excited about this episode today. I've got Azmath Pasha who's from the Forbes Technology Council, and we're gonna be talking about... He's got a long history of doing IT projects, and we're gonna be talking about what he's learned over 25 years and how that can apply to Data Mesh so that we can prevent ourselves from going down bad paths, and a lot of advice from a career of learning of how to do these types of large scale transitions and projects. So, if you don't mind Azmath, if you could give yourself, or if you could give the listeners a bit of an introduction to yourself and we can jump from there.

Azmath Pasha
Thank you, Scott, I appreciate it, and thanks for having me on the show here. My name is Azmath Pasha, and I've been in the industry for about, going on 26th year now, and fortunately been in the data and analytics field for a long time. So pretty much, I'm representing Forbes Technology Council here and recently, I've been with Paradigm as the Chief Digital Officer and with Capgemini running the Snowflake Global Competency. So Data Mesh is, although it's been the buzzword that has been making it in the last two, three years, it's been around with us. The technology industry as such as you see it regularly cycles through trends on its endless path toward innovation. And one of the latest trends is the concept of Data Mesh. So basically, what is a Data Mesh? It's an architectural design to decentralize and make it available closer to the products and make more data available in a more productised format to the customers.

Scott Hirleman

And Zhamak has talked a lot about the organizational side, and it's not as much in the written content as it is in her presentations, but what paradigms have you seen in the past that have that organizational component, and is that something where that kind of forever pendulum between centralization and decentralization that seems to swing back and forth constantly. Have you seen many past paradigms that really focus much on the organization as Data Mesh has?

Azmath Pasha
Yeah that's a great question, Scott. So, pretty much, you should look at it from more of a centralized single source of truth. That's what we're trying to drive. We as practitioners, talk to our customers, talk to many of our stakeholders in trying to bring that single view of the data, right? So looking at those levers, they are like... There's data quality, there's metadata, there's data architecture, data modeling, data interoperability which is very critical, and of course, the data warehouse and the business intelligence which basically fosters all that, right? So these are always going to be there. And this is where the lever is pretty much leading towards that single source of truth and how the Data Mesh nets out to making data more accessible. Increasing availability, improving discoverability, and ensuring governance quality and uplifting security. So this is... It's been around with us, and the concept is more towards centralizing the source of the data, but at the same token, decentralizing and making it closer to the customer in a much more domain driven distributed architecture.

Scott Hirleman
Right. Yeah. Well, and closer as well to the producers where, by putting it on the producers, and then they have to communicate with the consumers. I think that's something that... It's funny how many times the conversation of, when I'm talking to practitioners, it goes towards, we got the people in the room and they really started to... The producers and consumers, and their eyes lit up because they could actually communicate well, and they just had never spoken. So, it's kinda crazy that change aspect of just trying to change the attitude.

So, we wanted to talk about Data Mesh from a practitioner view. A lot of what's been out there has been somewhat theoretical, right? We're still early days in even the most advanced journeys towards Data Mesh. So I would love to get your view on how you think an organization might set themselves up to be successful, rather than, "I'm trying to rush towards my PoC," or "I'm trying to rush towards getting as many data products as possible." That kinda slow down to speed up aspect. What have you seen historically from projects, that have made projects that are this large and this kind of longtailed, long winded be successful so that it's not kind of a flash in the pan and that it's not, a year in, everybody goes, "Okay, we're gonna move to the next thing." What have you seen that organizations have done to make themselves successful?

Azmath Pasha
This is a great question. So you have to look at the way things have been operating in the old way of doing things, right? Where the data warehouses were created, and they were talking about a more centralized data team that supports multiple domains. Code management was a challenge, data and policy. Then the concept of more optimization for control and security, that became very pivotal and more of the domain awareness started getting forced. And these were all basically in a traditional monolithic data model, right? So that's the old way of doing things. Now, what we're seeing customers talking about and where we have implemented projects in large scale implementations have been as more autonomous, more serving the domains, being able to bring in that pipeline concept, right? Like how Snowflake has no pipe, which basically trying to orchestrate the layer where the data is coming ingested, whether it's real-time or batch processing, whether it sits on an AWS architecture or Azure, it doesn't matter. We're looking at much more decentralization in optimization, and making it more domain agnostic. And these are the things that I'm seeing that the market is shifting towards, and more and more as we are talking about it, people are doing it without realizing that what in fact they're doing is a Data Mesh, although it's part of their roadmap.

Scott Hirleman
They're definitely headed that way. And that governance aspect, I think, is an interesting place to doubleclick into, of how have you seen it be successful when you're federating that governance? And it's not fully federated, it's not just that the domains now have to own all of governance. How have you seen... Especially large organizations, transition from the super centralized governance towards having governance in the hands of the domains?

Azmath Pasha
Yeah, so basically I would say you have to look at the compliance aspect of it from a governance standpoint. You're looking at data quality and the security of those data products, right? So calling off those few things, I'll just stress upon the fact that when you're doing a self service enabled, and when you are enabling your business users to access catalog, transform, prepare, you'll have to look at it much more of an agile manner of looking at your requirements, which are much more now. If you look at the use cases, and MVPs that our customers are working towards before they stress test and productionalize, they're looking at much more process driven where they can share the results and look at the data which can be harnessed across different ecosystems, right? And then you're looking at the regulatory aspects to it. Because you said about data governance, regulations have always prohibited the way that data transfers between different architectures, between different departments, companies and geographies. You have to look at that, and then sensitivity of which

data can go in if you are in a virtual private connectivity mode, whether you can share the concept of data exchange. Lot of folks are doing a data exchange right now between the consumers and providers. Those are the reasons why I feel the governance is going to be very critical, and the whole concept of the governed Mesh topology, where how you will have the data products and how you'll have a data integration hub, share the data between those from a data source and the data teams is going to be even more critical when you start looking at it from an architecture standpoint.

Scott Hirleman
Yeah. And I think there's this emerging concept of access by default, but one thing that Zhamak has talked about a lot is, when somebody runs a query, that the governance comes with the data. So if you have access to this data and you run a query and then you decide to share that data with somebody else, they might not have the right to access that data, so how do you manage that? And I think nobody's figured that out at all. That right now is kind of a pipe dream type of thing. So, happy to talk further about governance if you'd like. But I guess you've got such a rich history of working with so many things. What pitfalls would you tell people to avoid here? Is it just like, "Don't get ahead of yourself," or things like that? Or what do you see as the most likely places that people will get into a bad situation when trying to implement Data Mesh?

Azmath Pasha
Okay, so this is an interesting question. So this is going to be more of those battle scars that you'll have.

Scott Hirelman
Yes. [chuckle]

Azmath Pasha
Implementing large scale decentralized, centralized platforms. So, pretty much you're looking at your data product experience. How has that been implementing across the enterprise? So, large financial services companies, they look at it from, "Okay, if I'm bringing in that new nomenclature of a Data Mesh, and then I'm also productizing my data domains, what are the things that I have to avoid?" So, the key thing is, you have to look at it from how you get the organizational buy-in into it. Is Data Mesh only appropriate for large complex organizations and large complex projects? That's one thing that you have to always look into, because clearly, it is meant for the future, you have to future-proof your architecture. And then you'll have to look at eliminating your enterprise data warehouse in totality over the course of time. And if you're going to be bringing in larger cloud-based ecosystems like RedShift, like Snowflake or Synapse, you'll have to look at it from future-proofing your

architecture, what kind of standards are there that need to pretty much accelerate your Data Mesh adoption?

And one key thing that I also told you is, you'll have to make sure that data is the new oil in your company and make sure that it is a true asset, which will show up in your balance sheet somehow. So that's a kind of stakeholder buy-in that you need to have, and be very tool agnostic. Let's be honest, there's plenty of tools out there that can literally help you do your Data Mesh and make it much more domain driven, but you have to stay tool agnostic and look at it as what best fits the architecture and future-proof it.

Scott Hirleman
Yeah. I think you've talked a lot about... Or you hit on a lot of really good points. And I think the first one you talked about was the data product experience of... You can't have every single data product be its own experience, because if you don't have that kind of common experience as to how you're going to get to data, how you're going to access data, then each... Every time somebody goes to consume, they have to learn how can I consume data from this data product, versus how can I consume data from the Mesh? Consumers shouldn't have to care about that specific, "What exact data product am I consuming from?" versus, "Is this something I trust and is this something I wanna use?" So how have you seen that kind of UI/UX kind of thing happen around data? Have we really nailed that in the past? The enterprise data warehouse has many challenges, but it at least had the common models that people had that kind of positive user experience aspect to it. It might not have had all the data the way they wanted it with all the contacts, but how have you seen user experience play out, good and bad?
[chuckle]

Azmath Pasha
So this is a loaded question. The end user at the end of the day, when... If I come in the morning and I see my dashboard functioning, I'm happy. But you don't know what goes behind the scenes to have a close to near real-time data that is made available depending upon what type of user you are. You could be an executive, you could be a day-to-day person running reports.

So I will pivot this question in a little different way. Let's look at the data lake. Everybody knows about data lakes. I've been fortunate to be affiliated with two large scale implementations for Fortune 100s where they did data lake implementation and moved towards Data Mesh architecture without even knowing that that's what they were doing. So the key things where you have to democratize the data, you have to take advantage of all that big data that you have in your cloud data, your Hadoop systems and all that stuff, maximize availability and accessibility across your

analytics investment, achieve a balanced business technology outcome that displays measurable benefits.

And the three core principles that I always go talk to my customers about is cost savings, business value, and data reuse. From a cost savings, we have to identify those use cases where possible to reduce your mainly infrastructure cost, while still providing business value and data reuse. Now, from a business value standpoint, you'll have to look at it from a, what are those use cases that can unlock business value through removing pain points, and creating those net new opportunities for you? The net new opportunities can be things that you have not been thinking about. From a KPI standpoint, you never wanted to look at it, but now, there is an option when you are looking at it from an infrastructure standpoint.

And then the key important thing is data reuse, so you have to align those use cases with associated data domains to identify, maximize, and get that ownership and then centralize the governance and bring your stakeholders along. So this is the three principles I look at, cost savings, business value, and data reuse. From a UI/UX standpoint, once you take care of that when you already have the marching orders to put the right ingestion services and the orchestration, your metadata managed, your cataloging, your telemetry, whatever data that you have... All that you've done is much more easier for going to the customer and showing that value that, hey, from a usability standpoint, and this is all part of your digital ecosystem that I'm able to bring in that confidence that my data is close to near real-time and once that confidence comes in, even you can have the basic rudimentary report in a self service mode, the customers would be happy. Even if it is Excel, they'll be happy. It doesn't have to be a jazzy tool like Power BI or Qlik or Tableau and all that stuff.

Scott Hirleman
Yeah, and I really like that framework. And we've been talking in the community about, how do you measure the ROI of Data Mesh? And I think these are good points to point to, what have you found if somebody is trying to drive buy-in internally, what have you found has the higher pull? Is it to kinda hit on that cost savings quickly, if you can get there upfront and so then you can prove that value more quickly versus the business value is the long run, and that's the dreamer thing, and that's where you're pitching the CEO type or the data reuse. I think a lot of what you talk about is also just being able to trust that data where somebody can... They're not discovering data in a data lake and going, "What is this?" It's telling them what it is. It's gotta be documented well. So what do you see is somebody trying to do a POC, what have you found has been successful in driving beyond that POC that if you were able to show that value upfront?

Azmath Pasha

So that's basically trying to bring that product thinking in the mindset, right? Data as a product, that's going to be very, very critical. We went about, I still remember about 13 years ago, EDW was not an Agile project, when we had Agile concepts introduced, a lot of enterprise data warehouses were all more SDLC driven, I will basically get the data moved to build that ODS and make it available through data marts and go in. Just about 13 years ago, we started talking about bringing in Agile, keep the lights on on certain use cases, certain systems, start to build new products, bring in a product owner from a domain and making that person responsible for making sure those use cases have been successful.

So now we are looking at more modernization aspects, right, how do you modernize a data platform with the cloud being a very strong player in the last three, four years with COVID, we have seen how much cloud has become so important for everything that we're doing. The things that I always go and tell my customers is find a way to democratize that data, liberate your data, make it much more a distributed architecture, have the self service concept and the empowerment given to the people that are going to be working on it as well as the people who are gonna be consuming from your DBAs, to your data analyst, data engineers, to the people like the data scientist and to the end users who are gonna be using.

So the consumption is going to be very, very strong, but doing all the stuff, you'll have to look at it, making it as a product, bring in that end to end ownership mindset, meaning you don't have the product owner just focusing on the data moving from the source, all the way into the ingestion layer. Make the product owner own the entire thing, all the way from the source to the consumption, and that's gonna be that mindset and that's going to be that game changer which I'm not saying that they're not doing it, but if you make that productised mindset in, I think you will start seeing that value coming in and the consumption and usability side.

Scott Hirleman
So, but if I were to just get you to pin just one, two, or three, if you are advising somebody on a PoC and they're trying to pitch it to the rest of their organization, what do you think has the most cache to kinda move forward if somebody is trying to drive that buy-in internally, it's not that it's coming top down, that somebody's trying to pitch it... Do you think it's the PoCs, the cost savings, the business value or the data reuse? Would you go for cost savings just 'cause it's got the easiest to measure?

Azmath Pasha
Cost savings is an immediate sell because most of the ID budget is owned by the growth officer and the CMOs, and so that will sell fast. The business value across the board to different domain owners that would be the second one. And the third one is

the technology one, if you make it more of a cloud first, meaning really you have to be bought in to implementing a cloud based system, data system, then you have bought into it and you're able to justify the cost, because the cost will take care of it. Your consumption grows over time and you have less ingestion challenges, you don't need that many data engineers as you build your CICD orchestrated layer using your XOps and there you go, your puddles have become ponds, your ponds have become lakes, and your lakes have become oceans.

Scott Hirleman
I also like that you said XOps, just 'cause it is that X faster...

Azmath Pasha
It's getting... It's the buzzword now... Everybody is talking about it. DataOps and DevOps and XOps.

Scott Hirleman
And DevSecOps and DevSec DataOps. Yes. So yeah, I really like that and I think that's kind of what I'm seeing is a lot of the use cases, some of them, the initial use cases have been business value, but that the cost savings ones get a lot more play early and then you go. This is just an easy win for us, and it also builds that muscle early. So, one question that I've been seeing a lot that I don't know that we've historically solved very well in data is data discovery and data discovery means a whole lot of different things to a whole lot of different people, but have you seen anything historically of trends of where we've done that kind of... You know it is that data democratization, it's not just can you find the data, but can you use the data? Can you understand the data? Can you interoperate the data? All of that.

Have you seen anything historically from an organizational or a technology or an architecture basis that people could look to get some inspiration? Because this seems to be the challenge that comes up pretty often in a lot of conversations.

Azmath Pasha
I see that because I tend to, you know, get into these workshops and trying to understand the customer, for example, there was a recent customer, we were doing a revenue recognition, we were doing a workshop for seeing if they are the right suitable customer for moving towards a data cloud environment on Snowflake, and the importance was all like how tightly the architecture was coupled in ways they're doing things and they want to be nearest to the dollar and cents. So, the aim should always be to take a pause, look at data discovery and spend a lot more effort upfront before taking the plunge. So, data discovery has to be part of your engagement model, whether it's a two week, three week or a four week engagement, depending

upon your use case, you have to do that, and then you have to basically treat it as like, okay, I have to look at the domain specific things. What kind of anomalies that I'm gonna see? What kind of volumetrics that can I derive from if I were to look at moving from a legacy platform, whether I'm using my lift and shift strategy versus my greenfield migration, look at all of those things.

I think a two week, three week data led, discovery led approach is mandated, regardless. Even if you have jumped in and you've already gone into an architecture where the company is bought in and your teams are ready to go, take a pause a few days, look at it, have a meaningful discussion, and look at if, it is that domains are operating, providing the value, make sure the customer is, the guys who are on the product ownership are participating in those discussions and get most of the value from those discovery sessions. So Scott basically, discovery is important, and I think every project that we have on the practitioner's side, we always recommend our customers to do discovery led engagements.

Scott Hirleman
I think that's an interesting point that I haven't heard made as much because it is the domain driven design and it is kind of doing the internal data product marketing, but I think... If people are looking to solve discovery via the technology, right of, okay, this data scientist needs to be able to find this data versus if you have that domain to domain communication of hey data scientist, here's the access to the people and let them actually even just talk to you rather than it's just the data products are there and you can find it in that way. I think that people to people communication hasn't been really specified for discovery in a lot of the conversations, and I really like that because it does mean that the people know, not just that there are data products or what might kind of be in there, but they can have that person-to-person conversation and you might have the domains go, oh, you really want this extra information that I wouldn't.

So, it's almost the pre-discovery as to what people are gonna wanna consume and that you have that person to person conversation, that's a really, really interesting approach that I hadn't heard or thought of before, so I really like that.

Azmath Pasha
And you also know right, Zhamak in one of her articles, she said that the distributed data architecture requires an enterprise view, right. So, how do you get to that enterprise view is bringing in the stakeholders close to the decision making while you start your engagement, and that's why the discovery is going to be very critical because an architecture just alone won't fulfill its full potential, right? The teams that are gonna deploy it, the data services themselves, that they're gonna be working with, the siloed aspects, those all need to come together and consistent

communication, commitment and strong governance are required to scale this up.

Scott Hirleman
Yeah, I fully agree on that, and I think you had mentioned earlier some stuff about standards as well, when you were talking about moving away from the enterprise data warehouse, we do have to have standards about how we work together and how data should be interoperable and how it should be stored. We wanna give domains the autonomy. But, any time I talk to people in Data Mesh about standards, everyone goes, 'Yeah, yeah, we need to have them.' But then it's you're trying to get to specifics, there's nothing there. What have you seen around, historically around standards for data, not just the standard model for a company in the enterprise data warehouse, but how we collaborate together and share each other's context with or share a similar model so that we can interoperate but that we don't drop context on the floor, which has been a challenge with the enterprise data warehouse.

Azmath Pasha
I think you gotta go back to what the four core principles are that we all know around the Data Mesh. One is pushing the accountability to the domains, that is something that we're seeing... We're struggling day in, day out with the data governance councils, the organizations who are trying to get the data governance going, and that's gonna be very, very critical because you are pushing that accountability, and that means you have to transition from a mindset of being very siloed and more of the enterprise view that I talked about. Then the second thing is you have to look at it from a self service standpoint, where you are having a self service, not just to your data generalist, but to your catered audience because catered audience, if you were able to have your data scientists and your report generalist work together and provide that value back to the business saying that this is what we found in 2022 versus 2021, in our year over year data sets. Those kinds of things are gonna be very, very critical.

And then the data platform itself attempts to mobilize the generalist, so you have to be catching up for that. And then the other one is automating the decentralized government. So now you talk to the council, you had something where you are trying to emphasize that we want accountability, then you talk to your end users saying that, "Okay, selfservice is a push" and they got into it. The third one is, how do you automate it? And that's going to be the most critical thing. And automation means literally the whole engine has to work seamlessly, and then harmonizing the data from the source system all the way through the transition layer and making it available for the end users, that's gonna be super critical. So just to summarize this, I would say domain oriented ownership, federated computational governance, which is gonna be very important, then data as a product and self service data platform, those are critical ones.

Scott Hirleman
So that harmonization aspect, when we've tried to do that historically, when I've seen things on harmonization, do you have anything where you've seen harmonization go well, 'cause any time I hear people talking about it, it's that it's not an easy thing. And it's not gone well. So how do we harmonize?

Azmath Pasha
So five years ago, I got used to this concept called Data Fabric, you're familiar with the Data Fabric right?

Scott Hirleman
Yeah.

Azmath Pasha
One of Forrester's analysts, Noel Yuhanna, she was one of the first individuals who started talking about Data Fabric, bringing that Data Fabric layer, how do you build that concept between managing and making it much more available through the transactional systems and all that. So now, let's be honest, we had several vendors, we had Talend, we had Informatica, we had capabilities such as Atacama, Denodo, and now with Google Cloud, Dataplex and all that stuff, has that nirvana state with Data Fabric been achieved? And has that governance ever gotten the benefits? Absolutely, yes, but I won't give it a rating of five out of five, I would say three because it's just been around for almost seven/eight years, and we're doing that, but now we have Data Mesh now, Fabric is all uniting everything together. Okay?

So you're going from the cataloging layer to the orchestration layer, going from all the way to the reporting, analytics, and data science groups, and bringing all that together. Now, when you look at the Data Mesh, Data Mesh is just very different. It is basically saying, okay, we tried uniting you all together, we had capability three out of five, now with Data Mesh what we are saying is, okay, all the pitfalls that we had doing the Data Fabric layer, we are going to put it back in and making sure that this is going to be robust. The challenges that we had with the domain owners, with challenges with productization, we're gonna address it by decentralizing, giving you guys the control, and this way you have the power to make the decision and help us push the value of data across the enterprise. And you have to think it through that way, because Data Mesh is now two/three years old, Data Fabric has been there, what are we doing, how do we connect the dots and how do we bring both together?

Scott Hirleman
Yeah, exactly. I think people are expecting Data Mesh to be the unifying theory of

everything, and it's like, it's not... It was never meant to be. If you look at it, it's for data analytics. A lot of the first things that Zhamak was talking about was much more even ML, of we need to have this data available in a clean trustable, harmonisable, interoperable way that is shared, so that way we're agile when it comes to ML and as people have picked up on it more and more, it's that more people are excited about it for the analytical side as well, and I think that makes sense, but yes, it's not trying to extend into the operational plain and it's not trying to be... The DevSecOps and all that stuff of how we're trying to do absolutely all of the data across the operational and the analytical plane.

And I think it's interesting as well. I don't know like how clear that split between operational and analytical is because there's so much blurry lines there, where you think about an ML model that's serving to realtime and it's doing analytics in real time, but it's an operational system and everything like that, and you think about some real time analytics use cases that are recommendation engines and all sorts of different things, the rise of Apache Pinot and things like that. Where do you see those splits happen, or do you think that that completely goes away or is this kind of, again, that pendulum where... It swings where it's like they're very, very separate and then they're kind of mixed and then they're gonna go back to very, very separate or?

Azmath Pasha
So a very interesting question. I think I want to tie it up to my previous commentary on Mesh and Fabric. So Data Mesh is all about bringing people and processes rather than architecture itself, unlike data Fabric, which is more bringing the architecture together, but the key thing is the operational versus analytical thing, so let's take a step back, right? We are in the age of AI, we're doing a lot of machine learning algorithms. Five years ago, it was a black box, nobody knew, everything was like, wow, we're gonna have data scientists, wow we're gonna have machine learning engineers, statistics and PhDs and all that, now ah it's okay, I can get the data scientists because we now know where to plug them in the organization.

We now know where their value comes from, they are sitting in between the data that is getting generated out and the data, the consumer, and they play a very critical role, and the MLOps and everything else that happens is on that layer. So I'm gonna draw upon a very quick architecture that we presented to one of our customers, which is basically Snowflake Driven on AWS. The customer asked us, how do we enable our data scientists, how do we make it much more operational, where are our algorithms going to sit? If that was your question, that's going to be outside the entire data processing itself. So in the olden days, we used to have, depending upon whether you are have not modernized your data platform, we used to have a data mart, the data mart was going to be the one which is going to make you more analytical, more slice and dicing everything in a much more granular, the roll up and

moll up if you remember those concepts that came out.

Now with this, what happens is, after the data comes in through the Snowflake and you have the virtual warehouses, the data is now going out and sitting in in a specific layer which is just meant for your data scientist to come and play their algorithms on. So what will happen is that as the divide has now clearly been put in place, the operational view of the architecture, and now the ML view, which is the analytic view, is sitting in the consumption side. The knowledge about this wasn't there five years ago because we were all carried away with, wow, this is a hot technology, we can cut to the chase, we don't need to go to the complete EDW. We can go straight to the source system, plug on it, do some R and Python-based algorithms and boom, get the data. No, it's not gonna happen. You have to follow it, but we have the segregation now.

Scott Hirleman
Yeah, and I think the rise of MLOps, so it's kind of there's another XOps for you, but I like Demetrios Brinkmann and the MLOps community a lot, because a lot of it is the practicality of this isn't just hot stuff anymore, it's how do we actually leverage this like business value. Why? Why are we doing this? And you kinda touched on a couple of things that I think would dovetail nicely as well with the concept of the modern data stack. Are you a fan or not a fan of the modern data stack, or do you think it's just yet another kind of buzzword that doesn't really have a very clear definition, or do you think that that's something that's helping people to be productive and that's useful in a lot of cases?

Azmath Pasha
I am a fan of whatever my customer likes. But as a practitioner, I have to advise and do the right thing, so doing the right thing is very simple, chase the business value, chase the cost savings. The modern data stack is all about commoditizing your data ecosystem, okay. The modern data stack today tells you that I don't need to have X number of ETL engineers, X number of folks doing reporting, X number of DBAs running my Teradata platform, my Informatica and all the stuff, admins. So I am a big fan of that because truly, you're bringing value to the business with limited investments, plus the technology is there.

So Snowflake talks about zero administration, time travel, where you have archived data available for 90 days. A lot of those things were not there before, when we started doing an implementation, I remember 14, 15 years ago for an airline industry, this was all we were doing Teradata Based SQL, standing up in the Teradata environment and running a private cloud, all that, and it was very costly.

Now, the cost savings are there. Now, that's one aspect of it. The other aspect of it is

how seamless things can be and how easy it is to go get that data and get that value, that's the most important thing for me, cost is for let the number of people work on it, for me, I'm more about efficiency and the efficiency from being a practitioner from managing large teams is how fast my team members can get things done with all the open source technologies like Dataform, using crowdsourcing platforms for coming and doing the technology work. We have TopCoder and those kinds of platforms available where we can put our use case out, we have the data scientists coming in showing that value. So that's the modernization stack, right? And you will have to think modern, because you can be paying $400,000, $500,000 for these monolithic models and the source systems that you have operationally, you have to save the cost, you have to have a compute and storage where you can wax and wane based on usage and commoditize the ecosystem and your architecture.

Scott Hirleman
How do you drive resiliency and reliability within that as well? If you're focused on the agility, the integrations can be a little fragile, a little spaghettiish, so how have you worked towards... You get kind of that MVP out and then you're building the resiliency as you go, but you could also just get pulled in a million different directions to just keep trying to add value in new ways instead of just creating some tech debt. So have you seen a good way to balance that, or is it just totally organization dependent?

Azmath Pasha
It's organization dependent. It is also about how much do you have the appetite. Suppose if I am a financial services company, I have to do loss prevention, I have to go to my customers and offer my customers better UI, UX, more data in their hands, customize products for them, whether it is a mortgage customer or a private customer. So those are my appetite. It's all about how much I want. So, for me what is modern is, I can put an AWS stack with Snowflake, with an API layer, I've done with MuleSoft, get a lot of this data modeling exercise done using data well. Fine, I can do all that stuff, but is that going to suit the need of the hour or the need for my customers that I have? That's the answer that you have to bridge. So the other thing is, resiliency is all about the future-proofing your architecture.

The last three years, which we've been in this COVID phase in the pandemic side, we have seen everybody trying to jump on to systems, getting into the cloud was a lot... We saw cloud investments go up, a lot of companies have come in and making that value added pitch. We've seen people doing more orchestration using the CI/CD platform, we've seen the API driven architectures coming through, SnapLogic and MuleSoft. So those are all things that you would see consistent investment coming from depending upon which industry you are in and what your use cases look like. Now security is playing a very vital role now. In the last three years, I'm seeing

Immuta and SecuPi and all these companies are coming and adding more and more of that value along with your cloud and modernization pitch. So that's also gonna be very important, and you have to take that into account because there are breaches every day, and it is all about what you want, at the end of the day.

Scott Hirleman
Yeah, but the governance side and the security side, a lot of people think about the governance and internal, and it's not a...

Azmath Pasha
We're not in a perfect world, you have to start somewhere. And there will be some failures, right? But the key is to remember that showing everything into a monolithic data architecture won't support the masses. It has to be very surgical. Yeah.

Scott Hirleman
Well, this has been, I think, really great, really helpful for a lot of folks, I think this will give them a lot more perspective rather than just... A lot of people coming towards Data Mesh are coming from the microservices and the software engineering side rather than the data side, so somebody who has been focused on a lot of this. This is a lot of very useful information to not just get attached to the new stuff, but we covered a whole lot of different things here. Is there anything that we didn't cover that you wanted to make sure that people looking at Data Mesh are aware of that you think would be good advice or good things to watch out for?

Azmath Pasha
Yeah. So from an information lifecycle management standpoint, you should look at your investments five years from now. Look at it not just about enabling better insights to the data that you're harnessing, it's more than that. And Data Mesh points us to a new era of data value and which requires new metrics around how you will monetize those data products and make it available to your consumers, that would be my key takeaway.

Scott Hirleman
Yeah, I think that's very helpful. Well, this has been really great Azmath. I really appreciate this context and everything that we've kind of talked about here today. If people wanna follow up with you, what's the best way and what would you like them to follow up with you about?

Azmath Pasha
Absolutely. If they have questions about their journey, how they have invested in decentralized architecture or they want to do discovery sessions or they want to have a particular use case to see if they can stress test on their current architecture that

they have, they can always reach out to me, I'm available. And the best way is at linkedin.com/azmathpasha.

Scott Hirleman
I'll drop that in the show notes, so that way people can have an easy link to that. You're a consultant as well, so didn't know if you wanted to let people know about what you're doing in that realm.

Azmath Pasha
Absolutely. So, I'll be more than happy to render services for you folks, been having plenty of work I've done in the consulting field. I was with Capgemini, I was with Paradigm Technology. And as a CDO recently, I've offered services across the whole gamut, from data analytics to cloud to AI and ML, so I'll be more than happy to give you focused sessions and advisory.

Scott Hirelman
Okay. Well, this has been very great. So thank you so much Azmath and thanks everybody for listening.

Azmath Pasha
Thank you. Thank you, Scott. Thanks everyone.

Scott Hirleman
I'd again like to thank my guest today, Azmath Pasha from the Forbes Technology Council. If you'd like to get in touch, his LinkedIn is in the show notes.

Starburst Sponsor Spot:
With a Massive Move to distributed data architecture, it's essential to have access to all of your data wherever it is. A data mesh emphasizes domain driven data ownership, data as a product, self-service infrastructure, and federated computational governance, giving you faster time to value without needing to transport your data. Starburst allows you to achieve this distributed architecture by allowing you to run SQL queries across distributed data that connects sources, regions, and clouds. For more information on how your team can benefit from a data mesh strategy, check out our data mesh resource center on our website.

Scott Hirleman
Thank you so much for listening to this episode of Data Mesh Radio. Hopefully, it was useful to you. If you'd like to connect with the show, you can find us on LinkedIn or Twitter. If you'd like to connect with me, you can do the same. If you have feedback or especially if you'd like to be a guest, we've got some links in the show notes to tell you how to do that, would love to hear what questions people have and how I can be

useful.