Sara Armstrong (00:04):

Hello and welcome to this episode of the AWS podcast series, Innovation Ambassadors. I'm Sara Armstrong, your host and head of Innovation and Transformation programs at AWS. Along with my co-hosts from around the world, we'll act as your ambassadors to some of the most interesting prototyping and innovation engagements with our AWS customers. We provide you with a roadmap to innovation and cloud technology solutions. We're so glad you joined us on this journey.

(00:33):

On this episode of Innovation Ambassadors, we feature Parameta Solutions, a global provider of over-the-counter market data within the TP ICAP Group. Their mission is to improve global financial and commodities market access through innovative solutions. We'll dive into their collaboration with the prototyping team, where they harnessed the power of generative AI and a serverless approach to their architecture to create governance at scale.

(01:04):

Well, I'm excited to welcome from Parameta Solutions, Rushmi Katyal, Chief Governance Risk and Controls Officer. Rushmi, thanks so much for being with us today.

Rushmi Katyal (01:15):

Happy to be here. Thank you.

Sara Armstrong (01:17):

And also from Parameta Solutions, we have Jacob Marlow, Senior Data Engineer. Hey Jacob.

Jacob Marlow (01:23):

Hey Sara.

Sara Armstrong (01:25):

And joining us from AWS Prototyping, we have Patrick O'Connor, Prototyping Engineer. Always great to catch up with you, Pat.

Patrick O'Connor (01:32):

Likewise. Thanks.

Sara Armstrong (01:33):

So Rushmi, for our international audience, can you tell us a little bit about the mission of Parameta Solutions?

Rushmi Katyal (01:42):

Absolutely. So Parameta solutions are the data and analytics arm of TP ICAP. Essentially we leverage the breadth of our group brands such as ICAP, PVM, Tullett Prebon, which have a RAF of OTC unbiased over-the-counter data. And we create possibilities with that using a variety of different data and analytics products and our new regulated business benchmark and indices.

Sara Armstrong (02:06):

And this new regulated aspect to your business brought additional challenges or opportunities, didn't it?

Rushmi Katyal (02:13):

Absolutely. So we actually became regulators the first inter-dealer broker to become a benchmark administrator in May '22. So it's been quite a journey upscale and an upscale for us in terms of what we've been doing. And since then, one of the key things that my team has been focused on is ensuring that we can provide the first best-in-class first class governance for our customers when creating benchmarking indices. And the challenge that we had with that is making sure that we were providing the most detailed documentation and the best methodologies in the industry whilst also being able to scale and deliver in a fashion that allows us to.

Sara Armstrong (02:52):

And tell us a little bit about how you came to work with AWS in this context.

Rushmi Katyal (02:58):

So I was extremely fortunate and I don't think many business people actually get this opportunity. And I was certainly in the front row, but I was invited to an Immersion experience whereby technology colleagues of mine invited me and I was able to see at firsthand some of the case studies that AWS had created it with respect to AI. And that really spurred some ideas within me. And in fact, I came back to our Account Manager about 7:00 AM the next day with seven ideas on how we could leverage that within Parameta Solutions and one of those opportunities we're going to be discussing today.

Sara Armstrong (03:31):

Oh, that's wonderful to hear that plug for Immersion Days. So Jacob, maybe you can set also a little bit of the context of you were already using AWS and some of the other areas of the business. Maybe set that context for us of what the systems that you're already working on.

Jacob Marlow (03:48):

Yeah, sure. So as Rushmi already alluded to, we're a data solutions provider for an inter-dealer broker, which is basically a financial exchange. So we have huge data sets. We're receiving millions of financial markets data points every day, primarily in over-the-counter indicative pricing, which is basically a price of an asset as well as trade and order data, which are trades going through from the brokerage. So working on diverse data sets across the energy and commodities markets as well as capital markets. So one day we can be working with oil data, we could be working with interest rate data, pricing stocks, working with foreign exchange data. So it's a diverse data set we're working with and we have a fairly mature footprint in the cloud. So we have a number of streaming and batch analytics workloads already in the cloud. And we host a number of our applications with Amazon Web Services.

(04:38):

So a lot of day-to-day is delivering data to clients over a number of different mechanisms, but it's also about providing enhanced insights the top of this unique data, so to provide risk and regulation analytics that allows clients to manage their risk. And I think from this project we had quite well-defined best practices in Amazon Web Services already, but it was really important for me just to be able to experiment and push the boundaries with the prototype. So kind of using the excuse to step outside of these boundaries and actually experiment the resources that Amazon has to offer. As I'm sure everyone knows there's a vast majority and you don't look at half of them. So it's great to have some guidance

from a Protype Engineer and actually point you in the right direction. And all the learning from this we can kind of feed back into the business. I'm sure we'll hear about that later.

Sara Armstrong (05:24):

Wonderful. We're Rushmi, this particular project involved compliance of regulation, is that right? Can you maybe tell us a little bit about the use case?

Rushmi Katyal (05:34):

Absolutely, I'd love to. So my team, what we do is focus very much on the first line business controls and ensuring and working with our product teams to make sure that when they create benchmarks and indices, they're doing so in a compliant fashion for benchmarks and indices, you have this document called a methodology, which is a public facing document. It's actually for your end clients. It's got regulatory requirements around it as to what you have to disclose. And essentially what you need to make sure is that document is providing enough transparency to the market so that they know how your index is being calculated so they know that when they're using it in another product, they can kind of work out how it's going to be tracked. And so there's certain requirements is quite... Regulation's always in the gray and there's quite a lot of detail in that.

(06:18):

And one of the things that can be challenging for a team like mine is ensuring that we are making sure that these documents are compliant. As I said, we're not making mistakes and we're doing so at an efficient speed and an efficiency when the team is quite lean in itself. What we've done historically, or what we did before we came up with a solution was actually create an Excel spreadsheet where we'd copied regulation and it wasn't the most efficient process. So when I saw at the Immersion Day some of the systems and applications that AWS had to offer, I realized very quickly that we could automate this process and make it much more efficient and scalable going forward.

Sara Armstrong (06:57):

So very manual, potentially error-prone and not very scalable for your business. And maybe that's a great segue, Pat, to talk a little bit about the approach in prototyping. How did we decide what we are going to do and what areas we're going to explore in the prototype?

Patrick O'Connor (07:16):

Yeah, yeah. No, I think Rushmi started off very well in saying that the grounds of this prototype really started with an Immersion Day, an idea, an innovative idea. And my approach to this prototype as we go into prototypes is to maintain that energy, that creative energy that started with an idea and keep that iteration going throughout the project. So some of the approaches that we implemented going onto this engagement or this prototype was to look into mechanisms that would include the very amazonian way of working backwards. So my priority is before I even step on the ground, what is it that I'm solving for? What is that core challenge that we're working towards? And in order to forecast where we're going, we need to work backwards to find that underlying challenge. Rushmi made mention to it saying that there's a lot of process in particularly the compliance area, a lot of hours going into creating Excel spreadsheets.

(08:17):

So understanding that that is part of the solution is creating an automated system that reduces the hours that optimizes and creates new opportunities, perfect. We now have that iteration one, or at least

the goal that I'm working towards. From there I perform or part of the prototype involves discovering. So running discovery workshops. Week zero ,as I like to formally announce it, is that period of time where it's all about investigation. No sort of commitments, it's about discover, create, invent. It doesn't necessarily have to be the solution we're taking, but it's a pathway forward. At the end of that week, we then take on a very rapid period of building and this isn't necessarily building towards the solution, it's building towards a solution. So a perfect illustration, which we'll talk about later is the fact that we went through many iterations or failures as you might say. And that's another methodology of failing fast. So I'm sure we'll make mention to it later on, but the whole goal of this prototyping is to keep the momentum, is to keep ideating and then once you run into those learnings, you then pivot fast and keep moving.

Sara Armstrong (09:35):

So starting with the end customer what they need and working backwards from there to determine what technology approach will work best. So Jacob, I understand you were involved in that workshop, in that discovery workshop. Can you talk through some of the different directions that we looked at?

Jacob Marlow (09:58):

Yeah, sure. So I think initially we set out to do the natural language processing, so use traditional machine learning models and I think we spent a lot of time discussing the classic NLP issues, so how much data do we have, how are we going to label it, how are we going to train the models, kind of pipelines associated with that. And we actually did spend a lot of time working on that at the start, and I personally thought that was going to be our biggest risk factor for not completing the prototype. So how are we going to get time from the business sponsors to label the data? Are we going to have enough time to create these models and that kind of thing really. And we did spend a lot of time looking at what machine learning models we're going to use. We read a lot of academic papers and we did try some open source models, but the basic Q and A didn't quite go as planned and we didn't have the kind of quality of answer that we're looking for.

(10:48):

So at the start we did kind of initially see that we could do a stretch goal of the power, sorry, of the GenAI. So it did evolve as Patrick was saying, quickly moving through the prototype and we moved at such a velocity, we got to the point where we could actually start to investigate the GenAI part. And I think it was really powerful for us to see the comparison between the traditional machine learning methods from the power of GenAI because we spent a lot of time on the machine learning side and if I'm honest, I don't think Rushmi was particularly happy with what we came up with. And I think we then quickly moved on to the generative AI and we actually started to use Bedrock, so leveraging the pre-trained models there and we could start to generate solutions immediately without having to train the models.

(11:32):

And I think as soon as Rushmi saw that she was starting to get a bit happier there, I think. And the only kind of thing we had to worry about there was a model selection, tweaking a couple of parameters and then looking at how we could generate the prompt engineering. So how we constrain the model to get accurate results, which for us is really important because we're talking about regulations here. So we spent a bit of time on that and we started to get impressive results straight away there. And actually we were able to start to implement the peer review feedback cycle within the prototype so a user could

start to upload a document and get feedback on whether the document was going to meet the regulation they desired.

Sara Armstrong (12:08):

So Rushmi, that traditional ML approach required some work from you and your team.

Rushmi Katyal (12:19):

I have to say before we got to the machine learning piece, and there's a funny anecdote about that, we had a keyword position whereby I think the first process was there was going to be a key word and it would give you a percentage of how compliant it was as no full document. And I remember they kind of showed it to me on screen and I said, ah, that's it? And I was a bit underwhelmed to say the least, and it wasn't really going to help the efficiency and scale of my team because we'd still have to do a line by line and extract the right words and that sentiment wasn't going to be enough. And what I love about this working backwards model is it was really focused to what my needs were and what the end goal were for the organization. Jacob is absolutely right when he said about the traditional machine learning, when the idea of me having to create hundreds or review hundreds of methodologies and tell you what's good and what's bad along with my day job did not inspire me.

(13:13):

I did try for a few and we realized very quickly this was going to be a very difficult process if we were going down this route, particularly as, I don't think we've mentioned it before, but it was only a three-week prototype. So from what we've achieved, I think it's really important to put into context the short time in which we've all worked together to achieve this and the team did a phenomenal job. But yes, I was very excited when we got to the GenAI position mainly because firstly it meant that I could really see how it could scale from my business, but also I realized that this tool is no longer just for me. The use of GenAI has expanded its users to Sales because they could use things and part of it, to the product team who could use it unilaterally and then improve their documentation before they even come to my team. So I saw the value add across the organization and it was no longer just focused on the Governance Risk and Control team, which is exciting.

Sara Armstrong (14:05):

That is exciting. And Patrick, maybe you can take us through a little bit, dive in a little bit to the architecture. What were some of the key elements that you brought in?

Patrick O'Connor (14:17):

Yeah, from an architectural perspective, I think it's one of the highlights that I'd like to point out is really the serverless mechanisms that we've leaned on to really scale with this business. The prototype itself might be in of itself a small building, but when we leverage onto these serverless technologies, and I can make mention to some of those technologies, is we're able to make a scalable system where the organization is not maintaining those services, but those services are auto provisioned in a way to grow with the business. Let's start off with the end user. So from an end user perspective, we used Amplify to create a React website. We then use GraphQL, AppSync from an end user's perspective to build this user interface which the business uses to interact. Behind the scenes, some of the key services that we've leaned on, Rushmi made mention to the keyword, this was our first iteration.

(15:21):

So we use Comprehend to extract keywords. We then use a series of step functions, so an orchestrating tool to then do that matching and then saving it into a Dynamo DB. Now the more interesting parts of this solution is really the GenAI piece. For GenAI we use the quite in preview mode right now, Bedrock. Bedrock is a managed service where you can come and use prebuilt models. In our particular case, we used the Claude V2 model, it was giving us really good results and all we have to do from a usage perspective is just point to the UI and simply ask the question. Now a bit before that was the Textract piece. So we are using Amazon Textract to pull information, save it, parse it into a Dynamo DB, and then once it's in that ready state, we lean on some technologies that, I'll just put out there, LangChain that then allows us to communicate efficiently to Bedrock and then connects that to the user to give this feeling of a full integrated system, even though it's a combination of jigsaws that come together.

Sara Armstrong (16:35):

Nice. And one of the things, Jacob is the prompt engineering with GenAI right now. I understand you and the team did quite a bit of work in that.

Jacob Marlow (16:46):

Yeah, definitely. So I think because of the ease of interacting with Bedrock, I think all of the effort tends to go into the prompt engineering and actually trying to constrain what results you're getting from the models itself. So I think kind of off the back of that, we ended up creating a generative AI chatbot because one day we were preparing for a demo, I kind of started coding up in a notebook and I thought this is extremely close to the chatbot, and I did notice that the AWS Bedrock had some starter code available on GitHub. I ended up pulling in that and actually I was really quickly be able to go from this notebook into a working chatbot within a day and now we're able to integrate it into the product for the demo. And I think that's probably the joy of prototyping is the fact that you are kind of moving fast and you can pull things in and start to work with it and if you see it's working then you can go forward with it and if not, you just kind of discard to decide and carry on moving.

(17:41):

So I think it was kind of a critical element for us. It was really tangible for the senior management to actually look at this and see what we've done and I think get the approvers off the back of that because they can actually see us interacting with the chatbot and we can start to ask questions, provided context, ask questions about the regulations, for example, what currencies are covered this regulation. And it was really, really great for them to see that. And the little Easter egg there is that we did actually name the chatbot after Pat, so our internal PatBot as it's called, is going to live on.

Sara Armstrong (18:12):

Nice. Wonderful. And Rushmi for your team and maybe even the broader user base, this really opened up the ability for, well the prompt engineering as well as the PatBot really opened up the services for other users, is that right?

Rushmi Katyal (18:34):

Absolutely. And so Jacob gave great context there and actually what happened was we were creating our demo to present to our CEO of Parameta Solutions and we were talking about a strapline is creating possibilities and how we could go beyond and stretch what we'd created in the prototype and go beyond. And one of the ideas was a chatbot, we leave Jacob for 24 hours, he calls me over and then we had this chatbot and what's so great about it or the PatBot, apologies Pat, the PatBot as we absolutely

did call it logo and all. What's so great about it is other people can now use the tool and ask questions whether it be on regulation or it be on a particular methodology or benchmark, and that will help our Sales division and other people in the organization that no longer have to read the whole document, which can be quite lengthy. They can ask specific questions and go back and answer queries in a very efficient fashion.

Sara Armstrong (19:31):

We often talk in this podcast about prototyping being about democratizing experimentation. It's not an experiment if you know what the answer is, right. So that idea, and Patrick, you kind of hit on this in the beginning of iterating and failing fast. I love that you had these three different approaches that you tried, these three different technology approaches to see what was possible and what worked and then refining that GenAI solution through. And as I understand it, Rushmi, in addition to all of the prompt engineering that was done to really hone and ensure that we're getting proper responses, you also had a human in the loop. You always have someone on your team that's doing that final validation. Is that right?

Rushmi Katyal (20:18):

That's absolutely right. So we were very aware of the risk of hallucinations and so we very much at every stage had that human in the loop, that interaction and I think that's really helped us move from prototype to approval. This idea that we've started small, we've got a human in the loop and that we're very aware in doing it in a very structured fashion has really helped us, I think go on our journey to not only go from prototype, which was in that very short three week window, but now to approval and we're hopefully going to be in production soon.

Sara Armstrong (20:52):

Nice. I was just about to say, where are we now? So going through that approval process, is that right?

Rushmi Katyal (20:57):

Absolutely. So we've been through the approval process, we've got internal governance committee to go through in terms of technology and that will help us move into the production phase. What's been really helpful for us is Bedrock's obviously very private and that has really helped us get the approvals that we need. And I know Jacob will be able to explain that in a much better technical fashion, but the private and enclosed space has really helped us with overcoming the risk that I think other areas and other people have found when using GenAI.

Jacob Marlow (21:29):

Yeah, I think from a technical side, it is great to be able to use and leverage the bedrock service. So it's almost like Amazon are holding your hand through the GenAI experience. So I know there's been a lot of talk in the past about using generative AI, but there's a lot of concerns about sending your data across to third parties. So I think what's great about Bedrock is that it's a service offered by Amazon, all the compute and all the data storages in Amazon, and we're already doing that. As I've explained earlier, the vast majority of our data, the vast majority of our workloads are in Amazon. So it makes it much easier from my perspective to take it to an architectural design board and say, we would like to use this service that's delivered by Amazon, but also we have this exciting new technology, which is generative AI that's hosted in Amazon and it makes my life much easier, which I'm very happy about.

Sara Armstrong (22:17):

That's wonderful to hear Jacob and Rushmi, and one of the key elements of prototyping that we hope is that it not only solves that direct workload that you're, that project that you're working on in terms of compliance, but also opens up the possibilities elsewhere for the use of this technology. So that's wonderful to hear. I'm wondering if as you reflect on the process and this prototype in particular, if you have anything that you'd like to share with listeners? Pat, I'll start with you.

Patrick O'Connor (22:55):

Yeah, I think the key takeaway from my perspective being working on the Amazon side is at the end of the day we want to build cool technologies. And with new technologies coming out such as GenAI, what I've taken away from this experience, and hopefully you've heard it in this podcast, is these are a cool group of people. They're building inventive, innovative practices. It makes my job a lot more interesting when 24 hours later, I can still remember the days of coming on board and we were talking about storage and tiers of S3 and how it functions. We fast-forward three weeks later and there is this self-guided curiosity in building chatbots. Jacob didn't even open the can of worms. I think he was looking into creating avatars at one point.

(23:49):

And so as a Prototyping Engineer, it's very rewarding to be working with the customer, but having the customer come back at you with innovative projects. And so my key takeaway is just the passion and energy of this customer is super exciting. I'm very keen to see six months down the track where PatBot might evolve to or what avatars might be created. So that's my key takeaway.

Sara Armstrong (24:15):

Nice. And Jacob, you, what were some of the reflections that you had?

Jacob Marlow (24:20):

So I think just to follow on from Pat, really, so I've done a lot of talking about the actual products and the tangible output of the prototyping project, but I think from my perspective there was more to it than that. So obviously having Pat embedded in our team, I think we learned a lot from him and I think it was a great experience. It was great to have him in our team and we connected with him and I think that's what helped us deliver this project so quickly. And I think the other thing is it started conversations within the business. So we can now see we can use this tech and we've proved out the concepts that people are now starting to ask. We have other data around the business, so can we start to set this technology on top of it and start to garner value from that. And I think that's what the prototype's there for is we had help to deliver this product, but now we're going to take it away, implement what we've learned and actually build future products on top of that.

Sara Armstrong (25:12):

Wonderful Rushmi for you. The final words.

Rushmi Katyal (25:16):

For me, I would say what's really important when engaging in these sorts, especially these exciting new technologies, is really to do the working backwards method. If you have a business sponsor and a business case that will truly work for your organization, you're very much more likely to get buy-in. And it will really help you deliver something that is tangible. And as Jacob said, when we show the

organization, everyone's going to want to use it and it will in true Parameta Solutions fashions create possibilities.

Sara Armstrong (25:45):

Well, wonderful. Rushmi, Jacob and Patrick, thank you so much for being here today and sharing your journeys and sounds like you really are creating possibilities for your users.

Jacob Marlow (25:58):

Thanks for inviting us.

Rushmi Katyal (25:59):

Thank you.

Patrick O'Connor (25:59):

Thank you.

Sara Armstrong (26:05):

I'd like to thank our listeners for coming on today's journey with us. Look for future episodes of our podcast on your favorite podcast platform. Share your ideas for future episodes or comments on this one via the email and the description. Thank you.