David Starr: Welcome to the Microsoft Industry Experiences team podcast. I'm your host David Starr, and in this series you will hear from leaders across various industries discussing the impact of digital disruption and innovation, sharing how they've used Azure to transform their business. You can find our team online at Aka.ms/indxp or on Twitter at industryxp.

So listeners, welcome to the show. Today, we're going to be discussing the internet of things because this is industrial and we're talking about real uses for it in industry, including what types of IOT data exist and how they differ, aggregating IOT data streams and connecting data for faster insights.

Joining us today on the show is Sameer Kalwani, Sameer is the founder and head of products at Element. Before joining Element Sameer ran product for a company that delivered a full industrial IOT solution loop for over 250 water treatment plants in India. Those solutions provided sensing and gateways, email processing and analytics in the cloud and even cloud to edge based control. And this was 10 years ago, back in 2008. So welcome to the show, Sameer.

Sameer: Thanks. Thanks for having me.

David Starr: You Bet. And we're also joined by Diego Tambourini, he the principal lead for Azure Manufacturing and the Microsoft Industry experiences team where he focuses on developing technical content to help manufacturing companies and software developers deliver their solutions on Azure at scale.

He also champions partners who deliver manufacturing solutions using Azure. So welcome Diego.

Diego: Thank you very much, David for having me.

David Starr: You Bet. I wonder if we could just start off Sameer with asking you a little bit about your company and can you tell us about it and its mission?

Sameer: Yeah, so Elements focuses on making industrial data incredibly easy to use in order to unlock insights never before possible. So I started Element back in 2014, 2015, basically focusing on how do we bring data science and data science capabilities to this rich industrial data.

And what we found was that there's a high amount of friction between data scientists, analytics, engineers who want to have wide scale data access to be able to do the analytics that they need to do any given day at any given time. There's a need for customers and owner operators to want to get that data to those folks and those folks want to be able to access that data, but there's a high level of friction in getting the data in the right context and seeing the data in the right standards and being able to actually manipulate the data.

And so we have used our software and our services to be able to bring that together into a solution to make that data a lot easier to use for engineers, analysts, anyone within the organization or external to the organization to deliver new insights that they never had before.

Speaker 2: What about that friction? What is it that you're describing? When you use the word friction?

Sameer: One of the key things that's different about IOT data and obviously there's many different types of IOT data, but one of the key things that that's there with IOT data is that it's a time series data.

You're actually measuring things that are happening in the real world. It's not software generated data and so what happens is you don't know whether ... if I'm looking at a pump that's there at Brownfield facility, that's been around since World War II, how do you actually know what sensor that sensor's connected to? What piece of equipment is that connected to? There's a lot of things that have happened across that entire time with that one sensor with that one pumps, for example, across those 50, 60 years of time.

And how has that pump evolved? How people change the naming conventions? How has that pump been changed in the various turnarounds that have happened on that pump? And then you think about that across an entire facility. So oil refinery or mining site and all these things have changed and evolved and the data is just not easy to use.

And so if you bring in a engineer who's fresh out of their Ph.D. and knows are in Python and wants to be able to work with that data, they have no idea about the context. They can't look at a pump versus another pump. And so you want to be able to make sure the data is easy to use. It's easy for them to have that rich set of contexts of history of what it is today, what's gonna be happening to it and what's happened around it.

Not just a time series of that sensor, but what's the functional location of it? Where is it in the process hazards? Who's worked on it? All these various sets of context are surrounding our physical world and it's really important to create a digital representation of that physical world in order to be able to let someone get the insights that they need to get at any point in time.

Diego: Element analytics brings to the table is that it connects the dots between all these roles sensor and device data to actionable analytics. So there is that realization now that a lot of people have have jumped into the IOT bandwagon and are collecting data from their devices and that they are realizing that it's not a matter of just blogging all that data somehow into a black box that is going to start speeding insight or or predictions.

There is a lot of sausage making. IOT even though is essentially a big data analytics problem is very unique in that like Sameer mentioned is time series data and it comes in different formats at different frequencies through different protocols and you don't always get what you need. If you're trying to say train a machine learning model or calculate a KPI such as OE, you may not have the data that you need to do that so there it's a lot of sausage making again in the process and the solutions like Element analytics help and customers do just that.

Sameer: Yeah, and Diego I think you're highlighting a really key piece here. So if I'm Fitbit and I manufacture Fitbit watches or Fitbit devices, I know my sensors. I have exactly the same piece of equipment that I'm selling to every single person as a consumer device and if it fails it's okay. I can always replace it. Right?

When we were talking about in industrials it's a very, very different animal. So you might be putting in a raspberry pie or in video jetson or something like that. That's using a lot of edge based data collection and that's getting sent to the gateways and that's getting sent to the cloud or you might be using your existing control systems and your policies in your control network and sending that data to the cloud. How do you build all that context about everything? How do you get that data into a central spot?

How do you augment that? Not just sensor data with ... but you also want to augment that data with all the other sorts of things are going on about maintenance, supply chain integrity hazards, you name it. Piping and instrumentation diagrams, all that into one spot so you have that full context.

Then you want to be able to give the folks who are analysts that same sort of schema on read capabilities that they have with web traffic data or ask their queries immediately and get the data relevant to that query to be able to get the insights using tools like Azure data bricks, but as your SQL data warehouse, we will to get the insights and then displaying those insights in Power BI or Azure ML.

David Starr: So what I'm hearing is there needs to be a lot of thought upfront about the time series data structure and what it's going to report, but people have been recording data for a long time and we've got a pile of Brownfield data there. How useful can that be?

Sameer: Element's actually connecting directly to that Brownfield data. So for keeping your data in a data historian, you have your EAM system that has all the edits. So enterprise asset management systems that has all the maintenance work orders and the functional locations of your equipment. All of that is rich, rich context about how your physical operations or are taking place.

It's important actually miscited to not take that information in because you don't want to just look at sensor data. You want to look at all the rich context

that's there, augment that with new sensing around IOT. Put that together into a form that's actually easy to use and to be able to perform analytics upon and so that is absolutely critical to be bringing in all the Brownfield data and even if you're at a Greenfield site and you have other sets of data that add context to it, context is key when you are talking about the physical world and wanting to do analysis on this sorts of data.

David Starr: And the Brownfield data provides the context, the new more structured data provides exact sensor readings that you're able to then extrapolate back in time.

Sameer: Yeah, exactly. So I would say the analogy I always like to make is like think of the IOT devices as stethoscope. You're getting a lot more fine grain information that you were not able to get before.

So I'm talking about a facility that's a water treatment plant or a an oil refinery or a mining site or power generation plant. Each of these systems, each of these facilities already have a great deal of control network. So they have control systems and DCS control system that's there and that's connecting to the actual equipment.

And it's using the sensing that Sarah's making determinations of what needs to happen, but if you want to start to add non-controlled data, so that's where a lot of where we're seeing a lot of customers trying to add IOT data on say their tanks or things that a distributed site where you might have a gas pipeline and you want to start to have more IOT sensors there.

You want to be able to put all that information in, in addition to your control data and get that to the cloud. So that way you're able to do fleet wide reporting or do what if analysis or do reactions on failures or whatever or optimizations.

All of those things are now possible where you're not making decisions just with data, but you're using data to enable decision making for the user to be able to make the right things happen at the right time to be able to reduce operational hazards significantly at that water company that you mentioned.

In my introduction, we reduced maintenance costs by 60% because of the ability to do machine learning and analytics and in helping our field service workers actually be able to get insights from data to be able to make the right action at the right time, knowing what equipment to bring with them when they're acting on these new insights.

Diego: So send me one question specific to customers that want to use machine learning on IOT data and I've heard that that a lot of the roadblock, if you will, is that there is not enough data science expertise to go around and and figuring out, selecting the right machine learning algorithm, and tweaking it with the

right parameters. It's not trivial at all. So how do your tools help address that problem for, for those who don't necessarily have a data scientist in in-house?

Sameer: My recommendation for a lot of folks is to think about analytics as not a maturity model or think of it as a capability model. You want to increase your capabilities and being able to deliver analytics to your organization.

So just like with Industry 3.0 when we had automation and control come in to help add a new capability to industrials, analytics and data and the digital capabilities are a new capability and you think about that starting to increase its slope and in your ability to adopt it.

The machine learning is sort of further along in that capability model. My initial recommendation is for folks to stop looking at data in a sensor by sensor basis and compare with temperature sensor at one place versus a temperature sensor another place. We're looking at one piece of equipment or one process. Look at the fleet and so what you need to do there is basically standardize all your pumps based off of functional location or based off of make or model, whatever it might be, but at least get all your pumps or whatever.

[inaudible 00:13:05] target equipment type and start to standardize that and so that way you can start to benchmark things and say, "Okay, what is my pump efficiency across the entire fleet? Why is this group worse than the others?"

And you can start to ask questions about that. So get your data into a fully contextualized, standardized model. That's what we call descriptive analytics. The second stage in the analytics journey is around what we call diagnostic analytics. So start to label your data with key pieces of information about what's happened when.

So if you start to say, "Well, I'm starting to see on these pumps there's a frequently occurring failure, maybe once a week or every month or so, there's a failure on this one pump. Can we start to diagnose what's going on right before that?" Because once you have that labeled information and your time series, you can now start to say, "Okay, well, let's look at what's going on in the maintenance activities? What's going on the the operator round activities? What happened right before that issue, so that we can diagnose these issues using data, not just a fishbone diagram, but using data to help understand why this is happening.

Then the third step is around what I call predictive analytics, and so that's where machine learning starts to come into play. You have a high sense of data quality. You have standardized, contextualized data. You have labeled data. Now what you can start to do is say, "Okay, well if I know what's going on in this dataset and I have enough statistics, either history or a lot of equipment, now I can apply machine learning to be able to tackle that."

And yes, a lot of folks might not have data scientists, but folks like Element and many other service providers are able to provide data scientists as a service to be able to have that. But data scientists can't tackle this unless the data's in the right form and there's a frequently mentioned statistic that Harvard Business Review said that's saying 80% of the time in getting to machine learning or AI is around getting the data ready.

That's only one part of the equation. There's also parts of the equation which says that you need to have the right infrastructure in place. There's other parts of the equation that say that you need to make sure that the data is actually in a form that people can even understand.

And so you have to be able to solve those issues before you can even think about or bringing on data scientists to solve your problems.

David Starr: And we haven't even mentioned yet that you have a product that lives in this spacecraft studio. Can you tell us about that?

Sameer: Yeah, so the graph studio is all about connecting to all these different Brownfield data sources and new IOT sources to be able to bring all that data together and to help folks to be able to create that digital model of the physical environment or what many people call the digital twin.

So you include all the physics into the model. You encode all the people who are working on that, on the equipment. You start to create a fully standardized, fully contextualized model of the data. And on top of that you start to ensure the data integrity. So making sure that the data model is accurate, making sure the unit of measures are highly consistent and converting those if they need to be converted, identifying where there's bad signals in the data so as things are known, miscalibrated, noisy, identifying all that and we're constantly checking that so we can make sure that we're able to maintain the data model.

So if things get moved around or things get replaced, the data's up to date. We keep that history as well and then expose that out into a data lake. So we put that onto Azure Data Lake.

And then from there customers are able to bring in data bricks, SQL data warehouse, Power BI, Azure ML, and because the data's in Azure Data Lake, you can now bring in the fancy data science tools. You can also bring in the basic BI tools, any sort of ... or if you just want to do SQL queries, you're able to pull that information off of that data because you include the schemas of whatever insight you want to go after.

What we're seeing with customers that everyone has. When they're thinking about analytics. It's not just a one use case. Today it's a one use case, but tomorrow's going to be thousands of use cases. And how do you make sure that your organization and your IT capabilities are able to feed all of those insights

that need to be driven, be able to let your newly minted engineers ask Python queries, ask R queries, be able to do all the various analyses that they need to do.

You need to make sure the data is ready for that. And so this graph studio really helps organizations get that data into that standardized, contextualized model for getting those levels of insights to the data store on Azure Data Lake. And in addition to that, the graph studio produces a graph database that lets you easily slice and dice the data on Azure Data Lake so you have the context around where to pick and pull the data that's stored in the raw format from all those source systems that we're pulling from.

So the data stored on Azure and folks are able to ask the queries that they need to in order to be able to get the insight that they need.

David Starr: So once the data's there, do you then leave it up to the customer to come to the data itself and put whatever reporting mechanisms on top of it they want.

Sameer: So yes and no. Some customers are more mature in their digital capability and their digital transformation. And so they might want to be more self serve. Other folks, this is their first set of steps in this analytics journey.

And so we have a set of folks who are focused on making sure that, that deployment goes well, that they're able to utilize the technologies on Azure accessibly. The other sets of areas that we have is that we have folks who are focused on more of the analytics piece.

So if you're looking for data science help or machine learning help or whatever it might be, just one of these into Power BI. We want that to be self served. So we are here to be teaching folks how to fish. So we have a model where we basically let folks go through. We will be a player for them. There will be a player coach and then we'll coach them to make sure that they can be self running in this whole new analytics environment.

Then we also have a set of folks that are much more strategic and focused and say, "Okay, maybe we're helping build this discipline of analytics in one part of the organization. Let's start to push out other parts of the organization and make sure that people are much more self serve and can get this spread across the organizations. [inaudible 00:19:29] working in one geographic unit or the downstream part of your oil and gas organization? Let's move that to other BUs and become much more of a analytically driven organization going forward."

David Starr: And so it truly changes the culture of the company it sounds like?

Sameer: Oh yeah, for sure. I mean, when you talk about digital transformation, you need to be thinking about like any other transformation.

When you think about change management, when you think about change management, there is technology changes, there's organizational changes and there's process changes. Yes, our software can help you with the technology pieces, but you need to be thinking about this more holistically.

And so we have a set of delivery services that we provide to make sure that we can help folks go down the analytics journey going forward. Because the most of our folks on our team have done this in the ad tech industry. As we all know that's been massively transformed. The commerce industry, so e-commerce now being a big thing, they know how to deal with the digital transformation there.

And so they're helping guide now the industrial folks and we also have a set of industrial focus services folks who know that area so they know the business questions that are there and are taking that design thinking approach and and the machine learning approaches that are being used in other industries and applying it.

David Starr: So not to go too far in [Tangent, 00:20:50] but you guys are actually using design thinking in your solutions?

Sameer: Yes, yes. I actually ... we do quite a few of those. So we will usually help folks go through a design thinking workshop. So that can last between one day to three days depending on the scope of what the organization and what they're looking at.

What we typically do is we go through a process of identifying what the use cases are. So there's things of ... there's a lot of post-it notes getting things and ideas on the whiteboard, making sure that we're bringing people from the business side, from the IT side, from the OT side, and making sure that there's a full alignment that's there and then making sure there's some sort of executive who's sponsoring this and saying this is something that we need to go down and pursue, and then what we do from there is we usually from everyone just throwing up post-it notes, we can come up with 50+ use cases in about a half hour period.

Then we'll go through a process of whittling that down and saying, "Well, what are the low hanging fruit that we can go after?" And so we usually get down to about 20 or so and then we'd go through a set of criteria of saying, "All right, is this the right thing that we should be going after right now? Do we have the data availability today? Do we have the right people on board with this?"

All those sorts of checkpoints that need to happen, make sure that there's proper alignment. We go through a value modeling exercise and then we also go through what we call crazy eights, which is where everyone starts to sketch up what they want that solution to look like from a visualization standpoint or a process standpoint, whatever it might be. Then we start to say, "Okay, well let's

hone in on these two or three ideas. Let's start to get the organizational buy in and then narrow that down into one idea that we're going to go off and tackle."

When we go through that process with new customers and even with production customers, we do that sort of checkpoint every six months and say, "Okay, well what are the objectives that we have for the next six months? Where the key results and what are the actual tactical items that we need to go off and get done in order to achieve that?"

And so it's actually really helpful for folks to keep pushing along the analytics journey using design thinking as an approach to bring the best ideas up to the top and percolate those. So we can go off and tackle them and make sure that we're pushing forward on this capability.

David Starr: That's fantastic. I'll put links to the Stanford program for design thinking in the show notes.

Sameer: That's where a lot of that has come from our side. We have quite a few folks from the D school on the team here.

Diego: Could you tell us a little bit more about on the data ingestion side? So I understand that graph studio helps customers. I do data joinings and digital twin mappings. On that data ingest side, how close to the device do you get and what assumptions do you make us too in the protocols they speak and the way they tag data? Can you elaborate on on that side closer to the where the rubber meets the road, if you will?

Sameer: Yeah, so we actually only connect to systems. Systems being like an enterprise asset management system or a process historian. We're not connecting to the actual edge devices or gateways. The rep or someone has a new set of raspberry pis and we want to send that to their gateway and then put that into IOT hub, go for it.

Then we'll connect to IOT hub and pull that information in. Yeah, if you already have a data historian or a processes throwing like the OSI top Pi System or Honeywell Ph.D. or Aspen Tech IP 21 or Schneider Wonder .... I can go through the whole long list of them. We connect to those systems.

Diego: Well, that actually answers my question clearly that you assume that when it comes to IOT telemetry or data records you start from the actual IOT hub. That's where you pick up the data and you focus on what happens to that data from that point onwards.

Sameer: Exactly, exactly. Our take on it is that there's a lot of folks building out sensing and devices and hardware or gateways to be able to push that data to the cloud. And that's great. And there's enough sort of use cases around that.

Then there's lots of folks who are doing bespoke analytics for specific use cases. Around six slips for oil and gas drilling or whatever it might be, and that's great as well, and they can be applied models, but connecting the data and the context of everything that's going on at the edge, not just one sensor, but everything and not just one piece of equipment, but the whole process, the whole facility. It's all in context.

Things are upstream of each other and downstream of each other within a process and we need to make sure whether it's manufacturing, power regeneration, oil and gas, whatever it might be, all that is all these things working together and you can't just look at one manufacturer's analytical application or one sensor packages analytical application. You got to look at the whole thing in context.

And so we're bringing all that data together to make sure that you can go off and tackle this one or two use cases, but also look at the thing in a broader scale as well.

David Starr: I'd love to keep going with this, but I'm afraid we've reached the end of our time, so I want to think Sameer and Diego very much for joining the show today. I appreciate it you guys. It's been very informative for me.

Sameer: Thanks for the time guys. I really appreciate it. These were great questions and I'm happy to maybe come back some other time or something, whatever it might be to help answer anything else that comes up.

Diego: Yeah, and thank you Sameer for your insight.

Sameer: Yeah, it was great chatting with both of you guys.

David Starr: Thank you for joining us for this episode of the Microsoft Industry Experiences team podcast, the show that explores how industry experts are transforming businesses with Azure. Visit our team at Aka.ms/indxp, and don't forget to join us for our next episode.