
# RR010

## The big data unlock for AI with Edward Calvesbert, IBM

Capgemini

# RR010

# The big data unlock for AI with Edward Calvesbert, IBM

Disclaimer: Please be aware that this transcript from the Realities Remixed podcast has been automatically generated, so errors may occur.

(00:01.422) Well, there's nothing wrong with me, so there must be something wrong with the universe. (00:18.862) I'm Dave Chapman. I'm Emee van de Giessen, I'm Rob Kernahan. And this is Realities Remixed, an original podcast from Capgemini.

This week, we go back to the big AI unlock, which is data. Has anybody fixed this yet? Spoiler alert, no. Joining us later in the show to talk about this with us is Edward Calvesbert . He's a VP Product Management, IBM watsonx AI & Data Platform. If you want to jump right to our conversation with Edward. You can find the point of the show in the show notes, but before that, we'll just try and unpack a little bit where we are with the state of data. So Robert, you say no. What's your perspective that had you jump in on that? He wants to go for a very short episode, I think. Yeah, no, that's it. Just turn it off. Data has been the bane of... most organizations lives. It's the great enterprise architect, I told you so moment, you should have always taken care of your data. Now some organizations, they have good control of their data, but they're generally well managed, disciplined, and they're in that game. So their game is data. Enterprises that maybe build things or they've had it as a bit of a side thing, haven't really taken care of it. So it's siloed, it's potted, it's all over the place, it's expensive to manage, it's duplicated, they don't really govern it or... tag it so they don't understand it correctly. So bringing it all together has always been this great big monstrous task. And so we've had these systems that we've built up over time that try and help us deal with it. But again, the strategy is starting to evolve because of the learnings we're seeing from the difficulty of just managing this enterprise. And if you haven't got your data in order, don't feel bad about it because you're like 99 % of the rest of the organizations that struggle with this. Chief Data Officer's career has gone down on the ship called Master Data Management, hasn't it? Oh, my word. Yeah, the original way of let's all try and tightly control data. That didn't work. It's too much work for the organisation. It's too intensive and it's difficult to understand the return on it. And then the other side of just throw it all in a bucket. Well, that hasn't quite worked as well. So we see some new strategies rising out which can make a bit of a difference and they kind of had a green shoot. you know. (02:32.918) maybe there's some hope for the future. let's come to a couple of those in a second. And I think we're to talk about things like the lake house concept, the fabric concept and maybe an emergence of new ways to think about data and data governance. But before we get to all that stuff, just as a reminder, or maybe just for anybody who hasn't tuned into one of our shows before, data and AI clearly have a close relationship. They have a close relationship because Your AI responses are only as good as your data. There are obviously well documented at this point issues with things like copyright. There are issues with things like hallucination. So effectively, you don't get your data right. Your AI is shaky at best. And data, of course, is not just the big open internet troll. Data will come from varying trusted sources. And I think one of the things that we perceive likely to happen is a true market for data that is being resold by other organisations in a trusted way. monetisation point is a big one. Some people are sat on gold mines of data and they don't realise it. If you've been collecting it historically for a long time, aged data can have huge value in it, especially when we think about climate change, changing environment, all this sort of stuff. So be able to analyze it see what's actually happened. So a lot of organizations have a load of stuff and they could monetize it and they don't because they're not quite understanding what it actually means to them or what it means maybe to some other organization that might seek value in something that you don't really see. Yeah, exactly. People have talked about data as being the new oil for years, I think. in some cases it is. You know, there are obviously organizations at the moment that that sell data, like aggregate and then transform data and then sell data. But I would imagine that organizations that aren't about selling data, that might be about selling other services, may well start to sell their knowledge in a different way, do you think? Yeah, potentially. So if you take those organizations that really understand, understand it well, like

  

the Met Office, right? So you've got... (04:52.686) High velocity, high frequency, high volume data, incredibly difficult to manage. They've put mastery into managing it to try and predict the weather. They then sell that on, you get it through your app sort of stuff, know, that type of thing. So there's an example of an organisation that has built its life out of better data and data processing. But they're at the fringe of what it is. when you take the sort of I build something or I run retail or I do warehouse and supply chain, your data's kind of more scattered and maybe you don't appreciate what might be there. And I think for that organisation, which is the fuzzy middle, so you've probably got like 90 % plus, they struggle to understand it, they struggle to have the skills to be able to manage it. And it does sometimes take huge investment to sort it out because you've let it rot maybe. And I suppose you get into a situation where you go, you know... I should fix it, but it's all a bit too hard. I absolutely agree with what you guys are saying because it's also with a mindset that you want to design the systems to remove ambiguity, right? That's what we're all aimed for. So that the data is clear, that there's no doubt of what it is actually saying. So if the data is right, the output is predictable. But if we now look, especially in the world with AI, it's not structured. You know, you have deterministic outputs, probabilistic models, incomplete data. patterns instead of rules. And we still have leadership and organisational structures that are completely built around control. And I think there's huge tension in there. Yeah, it's a very good point. And you've got this, you've got one side, which is the data warehouse crew, the one complete accuracy and create this perfectly refined data. Hugely expensive. You've got the other side of the divide, which is let's throw all the data in a big bucket and shove algorithms at it and hope something comes out what we see now in the middle, it's the lake house term, is the merger of those thinkings coming together. So sort of mixed mode data. Well, let's use that as a bridge onto where we are now. So as I touched on it earlier, I think we see two paths that emerging, the lake house and the fabric. Go on Rob, this is absolutely sweet spot for you. It's architectural stuff. architectural stuff. Level four. Yeah. I mean, looks like this is what you're here for. (07:11.094) I just get so excited. So the lake house is structured unstructured and you kind of work on a mechanism of trying to bring those two worlds together. the fuzziness in the centre, I'll go back and use that phrase again. And then the fabric is actually how do I connect all this together and make it interoperate and it has to flow effectively. So there's this concept of I'm going to work on the data and create value, but there's also this concept that says, I have to get hold of the data. I have to either ingest it or get access to it, et cetera. Because there are certain things like when you're dealing with aged data, it can be a massive issue. So in logistics or in other sectors, if your data is out of date, it's a real problem. So you need to go to the source system straight away and get it second by second. Whereas other data, it can be a day old and it's all right. So you might handle it differently. So there's this... This is mixing of methodologies that are coming out and that's what the lake house is. This merging of thinking to try and get a better way or I should use the word pragmatic way forward because the totally purist view is too expensive and is hard to do and the totally unstructured world doesn't always get the results that you might think you want. Okay, so that's lake house. the other, more of an emergent, visible in the market approach to this is fabric. the data fabric, which is an approach that allows you to unify disparate data sources across an organization into a cohesive layer. Is that sympathetic to Lakehouse, different to Lakehouse? there was always the strategy of, I'm going to put all me data in a big lake and I'm going to process it from there, but that creates duplication. It can create data aging like we've just discussed. What data fabric does is it kind of reaches into your organization. pulls the data out from its source and then can present it in a way that is easier to manage and use. So you don't have to understand the complexity of where it came from, but you are able to get hold of the data items that you're actually interested in. But there's still a place where you can dump large unstructured data as well and you can process it there because you don't know where the source came... Sorry, you

might not have control of where the source came from, might be ad hoc, etc. So it needs a good place to rest. So it's a bit of a mix of both. (09:25.858) The Lake House then takes the two ideas of unstructured and structured data and sort of merges it together to go over the top. fabric is essentially a layer of abstraction that allows us to deal with the concept of distributed data, but we don't have to think about that when we're doing the sort of data products that we would then use as the output that would create some value. good. Now within that data architecture, of course, you would have systems of record as well as potentially vastly more distributed data sources of varying different sorts and in a kind of emerging, you know, kind of new version of the tech stack, you're going to have your data sources, if you imagine it as a layer cake in the bottom, and then ultimately you're going to have some form of data layer in the way that we've been describing it that can pull and munch data together above that. Above that, you may well get things like agentic layers. Now, One thing that's been very interesting this calendar year, and we'll be talking in, where are we, April 2026, is this notion of SaaSpocalypse. And this is a term that was put out by the media in response to a recent release of some technology by Anthropic that seemed to have a very big swing effect on the market with more traditional application organizations, even modern ones. in the SaaS world being impacted because people began to question that once you've got these layers that we're talking about here, that you no longer need or may not no longer need traditional apps. What do you think of that? Is that a real thing? We've touched on it on the show before. It is absolutely a real thing. And it's really important to understand that basically if you've got the data or the access to the data, the fabric, and you've got something that... understands what that data is, so a semantic approach to it. And then you have the ability to plug an agent in and then importantly orchestrate between agents, which is MCP and A2A and all that stuff coming together. Then actually your application thinking can radically change. Essentially, you don't need the concept of traditional transactional applications. And then the, I'll use the word again, fuzzy, the agentic layer can deal with fuzziness. (11:51.202) well-known architectural term, isn't it? It's not a one, it's not a zero, it might be somewhere in between. It's moving out of its binary, which is a strict transactional architecture, to something that can vary. So the agentic layer can vary the way it operates based on the circumstances, the context and the situation it's facing. So you then get into this ability to say what a human would do at I use the word judgment, right? But you'll look at something, a situation, and you'll decide the best way to deal with it based on your scenario, circumstances, understanding and past history. Strict transactional architectures can't do that very easily, agentic and the way they work can. However, the counterbalance to that is there are some issues with modern AI around hallucinations and things like that, so we need careful guardrails in place and a control plane that makes sure we can continue to trust this system. But it's rebooting the way we think about how work gets done. And that's quite important because it's a very different way of thinking about how a business might operate in the future. Does it look like a teddy bear? Does it look like a teddy Well, we can make it look like a teddy bear if that's what you Because it's fuzzy, right? Yeah, it's fuzzy. it is a... That's a phrase from the 80s. Remember that fuzzy logic. I am not old to remember that. But it is a word that we should use more because it's not black and white. It's shades of grey. A lot of the things you come up with are from the 80s, aren't they? 80s were the greatest decade. Yeah. So like, let's bring up Jump by Van Halen. We haven't got enough time to dig into that, right? Actually, I'll tell you what we should do. Do you know how listens you've done to Jump by Van Halen this year? No, I need to look. It's down It's down. Was it in your top 10 in your Spotify and rap last year? It was, but it was lower this year. What was my top song again? I can't remember now, but I'll have to... go back and have a look. But it wasn't Jump Violin, it was still in there. But it's such a happy song. Such a happy song. would call that progress. You've shamed me into not listening to it. Is that what you're saying? Something like that. That's what it took. If that's what it took. Are you just worried about my

  

street cred? Is that what it is? You are desperately trying to drag me up by the hills to make me be more presentable to the wider community. Yes. Very good Dave. Something like that. (14:07.618) What you need to do is have a vote on whether it's working or not. That's the main thing. All right. Let's jump to our conversation with Edward Calvert-Burt, VP Product Manager at IBM Watson, and go back to that tricky conversation about getting your data right. Something we are still as an industry struggling with 20 years in, maybe even 30 years in to the subject. And let's see where we're up to. (14:42.25) Edward, good to see you. How are you today? I'm doing well. It's great to be here, Dave. Well, thank you for spending a bit of time with us. And I understand you're in Texas, not country. I am in Texas. It's a great time to be in Texas right now. Yeah, not quite as warm as it can get during the summer months, I think. That's right. Very good. So we are going to talk. Actually, let's talk a little bit initially. about Watson. Just a little, give us a little bit of a pen picture of when we say IBM Watson today. What does that really mean? Yes. Yes. So Watson, course, one of IBM's most popular brands, really a granddaddy of AI brands. Watson for us means AI for business, right? That's what IBM's mission is and that's what the Watson brand represents. How do we apply AI to business problems? And in 2023, we launched Watson X, which is really kind of the evolution of Watson to incorporate the new GNI capabilities, right? As they were coming online in 2022 and 2023. so Watson today, so when I always think of Watson sort of historically, I almost sort of, you know, I almost picture a mainframe that was being kind of developed very specifically to run things like machine learning and, and, know, play chess and things like that. I presume today we're talking about a sweetened portfolio of products, are we? It's a software product, yes. It's a software platform. Of course, have IBM is the hybrid cloud and AI company. So on the hybrid cloud side, we deliver and manage our software products across a very broad range of hardware platforms, including on-premises infrastructure, including mainframes. x86 servers, Nvidia GPUs, and of course all of the cloud infrastructure across the various hyperscalers. So really those are the two ingredients, Hybrid cloud and AI. I think we're going to talk a little bit about scaling AI and what I'm interested in when you look at consumption today, there's obviously AI is more than with us. It frames every conversation it seems at the moment. It's all over conferences, but one of the things that we observe (16:57.878) on the show is a thing that we refer to as the adoption lag, which is the technology is innovating at such a tremendous pace that it can almost be difficult for organizations to keep up with. And then of course there's a profusion of things to try and get your head around when it comes to then scaling AI, which is increasingly possible today. So what's your, being right at the heart of some of this stuff, what's your observation in terms of adoption maturity at the moment? Yeah, well, I think the adoption has been very fast. I think enterprises saw the opportunity clearly and early and got their hands on the technology and started experimentation as early as 22, 23, right? And then of course it's evolved very, very rapidly. And we can definitely talk about some of the inhibitors and opportunities as that evolution. is taking place and as it scales. But I think companies leaned into it quite aggressively because there's a sense that the competitive advantage and the business value is so high. And if you think about it from a competitive advantage perspective, how quickly and how well a company implements the technology and really differentiates their implementation from the competition. could mean the difference between winners and losers in certain industries. The productivity gains are there, the client lifecycle value and unique user experiences and interaction patterns are there. So I think that's why everybody is so excited. That's why it's so pervasive in the business conversation. And then of course there's the whole impact that it'll have on all of our lives as consumers, right? Interacting with companies that are implementing the technology and how it changes many aspects of our lives. But I'm very impressed with how quickly enterprises have adopted and adapted to the technology. think it's that thing about cycle time in organizations continues to decrease for the pace of adoption of technology and the rate of increase in its capability. feels

  

like AI might be the thing that is then the precursor to some companies rising fast and others maybe not surviving so long or getting consumed by others. It's like the event that occurred that then said you can track it back to the (19:19.214) the whole system changing or realigning to those who embrace the capability better. And it goes back to the adoption lag. Those who are more successful probably will get a massive boost. I wonder, Ed, if you've seen the beginning of best practice yet. to my mind, when things work really, really well, it's because somebody has spent enough time to make them simple and straightforward to adopt. therefore that cycle time to both implement, adopt, and then get to business value is, you know, kind of... intracable in maybe months, not years. Are you seeing that level of maturity of adoption happening yet? Absolutely. And I think a lot of it has to do with where you were before AI. So did you have your house in order? When it comes to data quality, data governance, data access, structured and unstructured data, right? So we can talk about that. Did you have your house in order? when it comes to developer productivity, CI, CD, separation of duties, just good, basic software management and implementation practices. Did you have your house in order when it comes to infrastructure? Do you have flexibility? Are you carrying a lot of technical debt? I think COVID probably did a pretty good job of advancing the state of hybrid cloud infrastructures. So I think a lot of how quickly they are able to implement it and adopt the technology. So they has to do with where they were before it started. But having said that, there absolutely are AI specific best practices that have to do, for example, with making data a part of the design process, right? Not an afterthought. Same thing applies to governance, right? How do you have governance built in so that it doesn't become an inhibitor to putting a system into production down the road? of course, your people, right? And how you provide access to the technology to your employees. A lot of companies, I think, are kind of caught in a little bit of a decisional challenge, right? Wanting to maximize access, but having some concerns about how to do that safely, how to do that cost-effectively, how to do that to enable collaboration. (21:43.206) as opposed to everybody working in their own little windows. So I think there's definitely some best practices evolving there as well. think it would be an industry first, wouldn't it, to adopt new technologies in a really elegant, quick way and everything be really nicely done and you're not doing repeated investment cycles and things like that. And it seems to me that as part of the adoption curve this time around for the pioneers, you know, they're going to find, you know, potential issues out, aren't there down the road? Like what are you seeing there? Not necessarily that you can avoid them. I think it's a sort of a natural part of maturing, isn't it? To sort of, to go through a few iterations to start with. Yeah. So I think about one particular design principle on this point and is it's what are you assuming is going to happen with the models? are you assuming that they are going to get better quickly? Or are you assuming that we're gonna achieve some kind of leveling off at call it where we are today? If you assume that, then you probably have to build a certain amount of scaffolding around the models, supporting systems. And we'll talk a lot about what those systems can be, Semantic layers and guardrails and this and that. Because... because the data state and your data state and the models are at some level of equilibrium. Or do you think that the models are going to get much better, much quicker, which is the track record over the last few years? And if you assume that, then at some point, you're going to a platform, build a frameworks and put systems into production, you're going to do it based on what's available today, right? But if what's available tomorrow is significantly better and makes what's available today obsolete, then how do you manage that, right? So I think we're kind of trying to shoot a bit of a moving target. That's why I think the first focus needs to be on the business case, right? So what is it that you're doing? (24:00.91) from an automation perspective and how that impacts your cost or your top line. And do you see on that business case, Edward, do you see that as being like a large scale enterprise-wide group level business case where it's making very macro pronouncements about the sort of transformation that may come about? Or do you

  

see it at the moment as being more an evolution of the POC situation we've been where things are still very sort of tangible, small use cases, or maybe not small use cases, but clear use cases. Yeah, I think it's the latter. So I think it's a lot more organic, a lot more bottoms up, right? If, again, if you're able to provide the technology to your frontline innovators, right? We think of this new user emerging who's technical enough, not a developer, right? Really just a huge... increase in capability enabled by AI. Of course, developers also and analysts. So if you can get the technology into the hands of enough skilled innovators inside your company with the right guardrails, with the right kind of direction, but really unleash that creativity and that productivity, that's where they're solving their own problems. And that's where the value becomes quite immediate. There's a thing there in the business case, which is like, organisations that try it and fail often learn as long as it doesn't sink you. And that's a very valid thing to build capability in an organisation. And often it's very hard thing to put into a business case to say, it's great that we're going to have the technology, but can our organisation cope with what that technology can do? So you have to push them through that learning cycle. think that, what's your view on how, how best an organisation to do that? Because if you don't get a go when suddenly this big enterprise thing lands potentially, everybody goes, I don't know. how to deal with this. It's like preparing the organisation, giving them the skills, making sure they're aware of what's going on. What's your view about how organisations can probably approach that a bit better rather than get the employee shock where something arrives and it's all changed? well I think whether organisations like it or not, like to admit it or not, their employees are using GEN.AI. That's happening. They may even be using it in ways that are not sanctioned. (26:23.578) So they may be copy pasting company information into the chat GPT and using it. I think that is actually a significantly underestimated practice. considering that, how do you provide an enterprise alternative? How do you provide a solution to that? And at IBM, for example, we have a technology that we call Bob, Bob the Builder. And it is kind of our IDE, right, so our internal development environment started with developers but has now moved into product managers, marketers, and really a broad range of users. It uses our IBM account, right, so the tokens are paid for by the corporation, right. Of course everything has guardrails, is monitored, multiple models are interacting at the same time. We have modes. right, for different types of users. Those modes have skills, which are recipes of, you know, what the model should be doing, so on and so forth, right? Our products are all getting MCP-ified so they can be easily consumed through models. But it really started with saying, okay, we need to provide a tool to our employees. And that tool needs to accommodate an increasing number of use cases, personas, and skills. And now we've deployed it to over... 40,000 of our, you know, close to 300,000 employees and that continues to scale, you know, in multiple areas. I feel, I feel I need to say that you just did a really eloquent description of scaling agentic infrastructures there, but in the back of my head, I'm just hearing the Bob the Builder theme. (28:06.222) I assume you're aware of the actual TV show Bob the Oh, of course. And I didn't name it, but I assume that whoever did was... It wasn't a coincidence. Was there an American variant of Bob the Builder? Very good question. Do we know this? I don't know. Like the American version of The Office. We a producer here. Maybe they could try and find out for us. Yeah, the chaos gremlin. We've got someone on that. I think it's the same in all over the world. It's Bob the Builder, I think. Yeah, no, but what did they check? Like the office in America was different to the office in the UK. Is it the same? Did they make an American variant? But I mean, this is the question of the age, isn't it? It's like now you brought it up Dave. I mean, this is what we're here for. It's really get to the bottom of these sort of situations, isn't it? It's Well, Edward's talking about making a connection to your end user or the ones who's viewing your stuff. You're also talking about different user personas. but also from the client side, so the ones you're providing the platform to. That's different, right? So you have software engineering, but you also have people that are less technical, but

  

also should be able to work with the platform. Can you give some insight on how do you do that? Is there a feedback loop? Do you have a user group that you just sit by and watch how they're using it? I think the classic old change management and new technology introduction techniques are... as relevant as ever. I think there's an interesting concept in in GNI tooling now called skills, which is basically how do you provide the model, right? A set of best practices, could be knowledge, instructions, tools, right? How do you provide them the model skills in a particular mode, right? So you end up having the same model being able to behave differently for different tasks, right, for different use cases, and also start to incorporate some of the user's context, the user-specific preferences, the user-specific experiences. So I think that's a critical thing to enable on, right, so that users can start to customize the model, make it feel really like an extension of their own personal productivity and style. I think that gets people, you know, really excited about how it helps. (30:29.614) their day-to-day life be easier and more productive. And then it's finding a balance between what's prescriptive and top-down and what can be customized more on an individual basis. I think that's where you kind of find the sweet spot to really bring it into the day-to-day activities and have people using it on a daily basis and become more and more part of their working tool set. I mean, one of the things I observe with Gen.ai in corporate world is they often unleash it and they've put all the guardrails in place and the safety and the data leaking and trying to make sure it's used responsibly and safely. But often they don't tell the stories about the end users where they've been successful on how it massively changed their day to day job or whatever else. And I often think that's undersold as the benefit of... humans seeing other humans using it successfully then go, I now understand because here's the tool, off you go, and use it, the whole training point and the capability. think we should do a lot more about that storytelling around AI and what it can do and how it should be used. I'd be interested in your views on that as well about have you seen anybody do it well and it worked from an adoption perspective? On that point, I see how we're doing it internally. I haven't spent too much time on change management and individual onboarding. I'm the data for AI guy, right? So what I'm doing is a little more plumbing oriented in the builder analogy, but there's no wrong way to do it, right? It's lean into it, right? And celebrate success stories, learn from failures, right? And just commit, right? Because anything short of that, I think is a risk. from an organisational perspective. Do you hear that David commitment? Commitment, that's what we need. We do need a bit of that. Can you try that? I'm willing to give it a shot Rob. A couple of hours on Tuesday morning maybe. Right, let's then move on to your area of specialism Edward. I really want to come to data. So think we've talked a little bit about (32:39.288) you know, kind of where we are with business casing, this sort of stuff and thinking about adoption beyond POC, but perhaps before giant, ultra mature enterprise level cases can be written at this stage. How sort of traditional change management techniques still work in terms of driving user adoption and helping users adopt these technologies. But data, data is really at the heart of this thing. I think getting it right means you have very accurate very useful tooling, getting it wrong, get one end hallucination, at the other end, of dangerous misinformation. So where would you say the current state of the art on data is? It's sort of data, I feel, has plagued the enterprise, the traditional enterprise over 20, 30 years where it's just been very difficult to get right. Many careers have been sunk on master data management. you know, attempts that kind of thing. So where are we on state of the art data leverage? Yeah, state of the art. is a journey, right? I think enterprises have been on a data unification, on a data quality, data governance journey for many years, right? Decades even. So I think that the biggest challenge is the same challenge has been around forever, which is data silos and data fragmentation. So much data is in a database and the database is serving an application. Or in spreadsheets. Or in spreadsheets. I'm starting from the core and then getting out to the edge. absolutely, there's fragmentation in the core. There's copies of that data. If you think about data, it originates usually in an application. That's where data is being...

  

created records are being updated, transactional systems, operational systems. Then it starts to propagate out into analytical systems, the first copy of that data that usually gets unified with other first copies of that data. So now you have the problem of keeping those two in sync. Then you start to add value to that data from an analytics perspective, which means joining it, enriching it, starting to build. (35:01.452) Data products are a popular word, really kind of higher level, higher value representations of that data. Now you're talking about your third copy, right? And I'm being very generous here by keeping the numbers small. then you've got the user that's taking that and making a local copy, sharing it. So you're easily talking four or five copies of the data kind of from its source originally. And again, a source that's dynamic, right? So then you've got that downstream propagation. And then you probably want to take those analytical insights and feed them back into the application, right? So there's actually a cycle where analytics and insights are now driving application interactions and automations as well. that's a real problem. And Gen.ai is not changing that. And enterprises need to get that. data house in order, there have been some important innovations, particularly when it comes to that first step, right? The transactional systems, the operational systems, in my view, they are what they are, right? That's really what's driving data creation, data update, and that's gonna continue. There's maybe some reduction of all the different variables, and there's probably some abstractions that have happened, particularly when you think about SaaS applications. where you've got application integration and data integration kind of happening at two different levels. But let's leave that one there. Is there though, I'm tempted to use the term silver bullet, but I don't mean it as being that simplistic, a more straightforward way to access data these days rather than, you know, kind of the heavy lifting that's needed to be required in data migration and centralization and warehouses and lakes and varying different iterations of trying to get all your data in one place is fabric. approaches both the solution to this and get you going faster, but actually are they real? Fabric and I'll say Lake House are, I think, the two competing, and they don't have to be mutually exclusive, but they're the two competing analytical approaches to data management or approaches to analytical data management. They're that first step, post the transactional system. (37:24.718) Now we want to analyze, we want to process, we want to enrich that data, we want to use it to train machine learning model, we want to use it to feed, create an MCP tool and feed an agent. So that first step, which is where a lot of data unification happens also, which is important. I think there's some best practices on how to get data in there. For example, Kafka, I think is a very popular Pipelining technology, you may know that IBM announced the intention to acquire Confluent, which is the main provider of Kafka. It's also very popular in the open source, right? So essentially available to everyone. But that's one way of getting data close to real time. That's the other thing, right? If you think about traditional ETL and batch systems, traditionally they run overnight. So your data is always one day old. And that's okay for a lot of cases, right? But it's not okay, right, for real-time insights and agentic automation. So, you know, we're closing the gap in that first hop from the transactional system to the analytical system. And in the analytical system, you've got the lake house and you've got the fabric. Like I said, they're not mutually exclusive. I think the lake house is really built on some very solid foundations, right? Commodity, cloud object storage as your storage tier. It's the cheapest, most resilient, like physical storage of data. And you're putting data in open file formats, Parquet file formats, Avro, CSV, right? So these are things that the enterprise can really control, right? And you have very, very high confidence that it's cost effective. Okay, so you're not getting locked in, you're not making a bad decision. You can go big on that, because really what you wanna do is standardize at that tier. Then you've got the Open Table Formats on top of that. Iceberg really has emerged as primary standard there. There's of course a couple others, Delta Lake, Hooty. But that's really what brings order to the old data lake or data swamp, right? Idea where it was just a bunch of files on object stores, right? And sometimes not even object stores, but the

  

Open Table Format starts to bring some of those relational structures to that data lake. And then on top of that you have (39:50.47) all the different ways of accessing and querying that data, different tools, different languages, different optimizations for volumes of data, right? So I'm thinking SQL query engines and Spark and Python notebooks, all sorts of different ways of accessing that data. But if you've built it on that solid foundation, now all that data is available for you down the road for whatever requirements, whatever use cases. whatever company you acquire that is using something else, it really starts to simplify the architecture. The fabric comes in as an overlay on top of that in my view. So the fabric then now gives you integration capabilities when you don't wanna unify the data and you want to access it in place with Data Federation, you can do that. It brings you overlay data policy and data governance management. So really to me, it's value added on top of that Lake House foundation. in a world where historically humans have been consuming that data and therefore are making sort of value and accuracy judgments about the data as they're consuming it and as they're looking at it. And I think most people who've been in business for, I don't know, a decent part of their career will have had that. moment where they trusted data and then in a setting where they were trying to use it, it wasn't right and you only ever do that once, you know what I mean? In organizations though that are going to be human-agent hybrid organizations, what do we need to do there to address that issue? Does the human have to be in every loop? Is it some kind of adapted governance that needs to work to ensure that agents as they're taking data and using them, there is something there to assure the accuracy? It is quite a stressful moment when you rely on... When have any suspicion that that may be happening, I start asking the same question over and over again. Are we sure? that the right data? like asking different people. When you get that, when you're presenting and you've just kind of, you know, like confidently gone through something as if it's absolute truth and someone goes, Stu, hang on a second. Just go back one slide. (42:11.278) That number doesn't look right. Where did you get that number from? you're like, no. It started. The crumble begins. Exactly. Yeah. And I think, I think we've all experienced that. Hopefully not every day. I think, agentic consumption at that level and human consumption are not that different. I think the big difference is agentic consumption happens 24-7, 365 potentially, right? Can happen in highly parallelized capacities, right? So I think there's a massive potential increase in the volume, right? The number of interactions. And I do think agents, and there's been some good papers written about this, probe the data, right? They don't have necessarily... the one SQL query that they're looking to run, right? Agents explore the data, right? They do reasoning, know, reasoning loops on the data. They may be leveraging data from multiple sources. And then they kind of come at it with a, I want to maybe a hypothesis, maybe that's too much of a weighted term. But then you get to the actual query that's being run. And then that's where the human in the loop may come in and say, well, let's understand if that's the right query to answer this question, right? So there's a, you know, in the agentic consumption, right, you have the problem of natural language to your actual programmatic access, right? We APIs, SQL or other, right? So there's that translation that has to get right. And then if the question is maybe a little open-ended, right? Or subject to interpretation, there may be multiple translations and iterations that happen to get to the right. answer, really to the right question and then the right answer. Whereas with humans, right, we are usually working, you know, with a, maybe a BI tool, right, which already has kind of very prescriptive dimensions, prescriptive ways in which a prescriptive object model that's been curated, right? So your boundaries are much smaller, right? You're kind of operating inside a box if you're using (44:34.222) tools like that. If you're using, you know, if you're a SQL developer, then you are familiar with the structure of that, you know, of that language and may take a little bit of time, but you are careful, right? You are very deliberate in structuring those queries and then you go through whatever iteration process. And then once those things are set, they're usually set, right? Unless you have things like schema evolution, right? Data

drift. you're starting to bring in new sources or some of those sources change upstream. And there's tools, Like data observability to be able to manage that evolution. There's flexibility built into say the Lake House architecture. So you don't break the queries. You may just not be able to take advantage of the new data. So we've come a long way, I think from the days of, know, structured ETL into relational data warehouse. where if anything changes, you're done, right? You have to redo the pipeline, have to redo the warehouse, you have to redo the dashboards. So I think we have a much more fluid environment, but there, there's also potential for data quality issues. So you still have to have that observability component, you still have to have that data governance component. And by the way, I think that's becoming even more important with agents now because data governance used to be a bit of a nice to have. Sometimes it would come from a data quality initiative or a compliance initiative. It was pretty top down. were data stewards as a persona. Whereas I think now, new data comes in, you have to attach it to the business terms glossary, to the data dictionary, to the emerging context graph or semantic model. And if that process isn't happening pretty dynamically, then the agents will really struggle being able to use that data. And we haven't even talked about unstructured data yet, which we'll get to. What's that merger of the warehouse and the data lake, isn't it? You sort of bring the both of collapsed it together. Yeah, that's the lake house, right? Yeah, that's the lake house. That's the very definition of the two. So highly structured and the sort of shush of it all in an object store. Great for AI and ML, but you don't get the preciseness of the reporting and all that stuff. There's that how it's that governance shift around it is you still got to allow people to (46:57.198) embrace the openness of a lake, but have the rigidity somewhere of the warehouse. And it's that careful balance of governance that you want it to be open, but you need to control it because of the whole trust point. You can shove any old data in and you don't trust it, then it can do a load of downstream corruption. And then Dave stands up in front of the board and they go, but Dave, that figure's wrong. And it's that which becomes embarrassing, doesn't it, about the... Where's your view on the balance between data warehouses very rigid, data lakes a bit more open, and the balance between the lake house view? As you say, it feels like a very careful path we have to tread. It's just enough. It's like the Goldilocks governance problem. Well, funny you said Goldilocks because the answer to that is called the medallion architecture. Oh, love a good architecture. This is new. It's bronze, silver and gold. Okay. So, so your, your, your bronze tier is that Ryan Jess, right? Where you just putting the data into the, into the data lake, into the files and getting the initial schema, right. Which, which is not prescriptive, right. Which is, which, which can be, which is, which can be adapted right from, from the data itself. That's your bronze tier. The advantage there is just go, right. It's simplicity. It's speed. Then you have your silver tier, which is where you're starting to apply more quality, right? You're starting to bring more unification. You're starting to normalize things a little more, right? So you're starting to want to call the same thing the same way, even if it's represented differently in the source systems. I think there's a huge opportunity for agents to intermediate in that process, by the way. I think that's kind of the cutting edge of AI on data management, having agents perform some of those quality assessments, perform some of those transformations, right? And simplify that process, almost fully automate that process. And then you've got your gold tier, right? Your gold tier is where, you if you've heard the word data products, right? It is something that, you know, has been established to be high quality. However you go through that process of establishing it, right? It becomes authoritative. That's its nature, right? So, and then it has... (49:14.422) It may have an SLA, a data contract. may have like a commitment to update. It may have, it almost becomes an API, right? So it's managed and maintained so that you can consume it with confidence downstream, right? And again, this is like the prototypical three tiered approach. It's how you use a Lakehouse, how you implement a Lakehouse. It's also what defines the difference between ETL, which is doing a lot of that work. upfront before you put

  

it into the warehouse and ELT, right? So extract, transform, load versus, now I'm getting my letters confused. Do the transformations in place, right? So you're reading and then writing and then that's how your tiers are progressing. So I think that's a very, very solid. starting point to manage those trade-offs. can relate that probably to a persona of user who's interested in the different tiering based on what they're trying to do in the activity and objective. It gives a nice sort of bronze will be good enough thanks versus no, this has to be gold because we're making a critical decision and it has to be repeatable and blah, blah, blah. It's a fundamentally different user. You don't have consumers on bronze. Bronze and silver are the domain of the data producer. They're the domain of the data engineer. Gold, right, is, and maybe silver at some level, is the domain of the data analyst, of the data consumer, right? I think you'll have agents operating across all of that, right? And really that's where the definition of the MCP tools for data skills and data producers versus data consumers really make a difference, right? Because you want to have an agent that's looking to answer a question interacting with the gold tier as much as possible. That's also where you're to have the best descriptions, the best semantics, right? That are going to assist the agent in doing that. want to move on to talk about the changing nature of apps in a second. But before I do that, I have an update on the big question of the day, was, was Bob the Builder adapted for the US market? Yes, it was. Oh, of course it was. So the British show underwent several changes to make it more accessible to a North American audience, including... (51:36.726) Redubbing the voices, changing terminology and changing the gender of a character called Muck. So there we go. Muck and Dizzy. See the problem is when my kids were young, they used to obsess over that and post from Pat. So I know a bit too well. was Paw Patrol in my day. Paw Patrol. I missed Paw Patrol. Yeah. Yes. Your kids are a bit older, aren't they? Yeah, they are a bit older. I've got a five year old and it's super kitties. what's that? Is this the new It's like the cats from the Paw Patrol. Ah, alright, this is the thing, innit? It's funny how things like that, in time, permeate into your life and then you never forget them. I went through this phase of we had the post from physical village and there were certain characters you couldn't get and we were sort of like completionists, we want to build the whole village and at one point I found myself on eBay potentially going to bid... like 75 quid on a two pound plastic character just because I wanted to complete the set. And it was Dylis from the local store and I sat there with a glass of wine. Well known as being the rare collectible one. So hard to get. mean literally it was an I'm pleased to say I didn't buy it, but there was a moment in my life where I went, I about to spend an awful lot of money on a very small crap plastic character? We'll dig into that some other time. Psychological profiling. Just before we end our conversation today, I do want to just touch on applications. So applications, be it mainframe or client server, or even more cloud native and Sassified in more recent years. I've had certain things in common, I think, over the years, which is that I had a UI with some form of experience, sometimes great, sometimes, let's face it, absolutely shocking. And what they did was encode, you know, processes, you know, things like standardization of ways of working, and then access a data set to allow you to do that. So it did a couple of different things. allowed like unification of things like... (53:50.27) I don't know, job codes and stuff like that across large scale organisations. And also it meant that lots of jobs could be quite well ordered because they've got software effectively running it and you might have humans processing things, but it was very clear what needed to get done. It seems to me as we're going to an AI world that the way that we think about applications and application architecture is very much going to change. Edward, I wonder if you've seen the beginnings of that anywhere yet. Absolutely. I think we go even further back, The original applications were command line interfaces, terminals, right? So you had to know what the system could do and then instruct it to do that. Then came, of course, the point and click user interfaces that we've all grown up with. And now you didn't have to know what the system could do. The UI would tell you what was available. You just have to navigate it and

  

select it, right? And interact with it. I think the next phase is conversational, right? So you don't even have to know by experiencing the user interface what's available. You say, this is what I'm trying to do. This is my intent. Intent is actually, I think, a really powerful concept interacting with models, right? Because it... gives the model the goal, right? And it allows the model to then have a little bit of creativity instead of being so prescriptive, which is effectively how we operate with software systems today. They are very prescriptive or limited. And that was seen as value, I think. And I think it still is value at some level because it encodes the business process well. It encodes the essential. data elements and decisions and actions well, but its strength is a weakness, That's what it does. So you need 1,700 of those to do what an enterprise needs to do. now, personas, skills, job roles, all these things that go into software design are different at every company, right? A company with... (56:13.678) 50 employees is different than a company with 200, is different than a company with 2,000, and so on, right? Different people wear different hats. And if the software systems are designed for certain assumptions, right, then they may not be the best fit for a company that's a little bit different. So really like abstracting that and putting in a new layer, right, which is the conversational layer, and letting the conversational layer or the agentic layer. leverage multiple applications at the same time. Also, I think really is kind of the bridge, right? It's, it paves over those silos, those application silos, those overly prescriptive, you know, processes, capabilities that is pretty, you know, standard today. So, I think it's really exciting. And of course, a lot of software companies are disrupting themselves in that way, So looking to evolve their own experience and capabilities to be conversational first. And then of course, there's the new overlay, right? Which is the agentic experience, the personal co-pilot or co-work, and then how that starts to subordinate the applications to really be more. a series of APIs and databases, not so much kind of that prescriptive workflow management. (57:52.878) Now we end of our episode of this podcast by asking our guests what they're excited about doing next. And that could be something in their personal life, like they've got a good restaurant booked at the weekend, which I Robert has, or it could be something in their professional life. So Edward, what are you excited about doing next? Well, I've scheduled to hike Mount Rainier in July of this year. I heard a podcast, some years ago that it was a good practice to do something really hard, maybe once a year and really hard to find that's not sure you can actually get it done. So it has to be a real challenge. What kind of altitude are we talking about? Is it a two day walk? Is it a three hour walk? It's a two day walk with an overnight, it's a 14,400 feet summit. Permanently snowcapped. there's a good amount of ice trekking, which makes the trip a little slower. at 14.4, you've got to contend with the oxygen issue as well. I've started training this year and I'm up to about 150 stories at a time carrying 25 pounds. gotta get up to 400. and then it'll be 400 each day on the two days when we get up there. So I've got a few months left. I mean, I get the whole like pick a stretchy thing to do, but there's maybe a middle ground to you somewhere. Yeah. have a friend who's done it before and he assures me that it can be done and proper training. is essential. that's what I do with my Sunday mornings, climb stairs. Slight word of warning, I've been in several situations with Mr Chapman here where he's assured me I'll be able to, it'll be fine and it will be good and it'll have a brilliant outcome and I'll really enjoy it. And let's just say history keeps repeating itself in a particular way. You know you love it, Robert. You wouldn't have had that burrito in San Fran. (01:00:09.91) That is, you know what, and the mariachi band that turned up. That sounds like a cracking thing to do. I afterwards when you've completed the climb and everything, you'll be really sort of pleased as punch with yourself actually. You'd be quite proud of being able to achieve it. Yeah, I'll probably be laid out on a bed for a bit. Even the training leaves me walking a little funny. So yeah, it's definitely a physical mental challenge. And then I'll have to think about what's the next big thing. difficult thing to do. Marcel, our producer who sat there. Goblin, Dave. his new title changed. Send us Photoshop pictures of him up the top of mountains. he does, he? In dangerous positions. Yeah,

  

I definitely did it. I'm like, did you do that? You know, that looks pretty dangerous. He's like, oh, did it. What's the thing that gives away the AI? it, there's a giraffe climbing next to him or whatever. He's got like six fingers. (01:01:10.062) All right, Edward, thank you so much for spending a little bit of time with us this Friday. It's been a real pleasure talking to you and just sort of really digging in into the state of the outscaling AI at the moment. So thanks very much. Absolutely. It's been a pleasure. Thank you. If you would like to discuss any of the issues on this week's show and how they might impact you and your business, please get in touch with us at realitiesremix.capgemini.com. all on LinkedIn. We'd love to hear from you. Feel free to connect on DM if you have any questions for the show to tackle. And of course, please rate and subscribe to our podcast. It really helps us improve the show. A huge thanks to Edward, our sound and editing wizards Ben and Louis, our producer Marcel, what are we calling him now? Chaos Goblin. Chaos Goblin. And of course to all our listeners, see you in another reality next week.

## About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, generative AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2024 global revenues of €22.1 billion.

Make it real.

www.capgemini.com



Capgemini