Capgemini

CLOUD REALITIES

CR026

Becoming a data driven
organisation pt.2 with Mark
Jones, Reach PLC

CLOUD REALITIES

LISTEN NOW

Capgemini's Cloud Realities podcast explores the exciting realities of today and tomorrow that can be unleashed by cloud.

CR026

Becoming a data driven organisation pt.2 with Mark Jones, Reach PLC

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

[00:00:00] To our sound and editing wizard, Ben, and of course to all of our listeners. We're on LinkedIn. Our Did you just, did you just mix up LinkedIn and Twitter? Yeah. I'm gonna say that's quite exciting that maybe we've just hit on a business model.

Welcome to Cloud Realities, a conversation show exploring the practical and exciting alternate realities that can be unleashed through cloud driven transformation. I'm David Chapman. I'm Sjoukje Zaal, and I'm Rob Kernahan.

This week we'll be taking a view on how to become a data-driven organization. What are the steps that you need to take and what are some of the challenges that you need to overcome, and actually what's the scale of organization you need to be in place to become a data product organization?[00:01:00]

Joining us this week, I'm delighted to say, is Mark Jones. He's the head of engineering for Data and Analytics at Reach, UK's largest commercial, national, and regional news publisher. Mark, welcome. Great to see you. You just wanna introduce yourself and say a little bit about Reach. Hello. Thank you very much for having me.

Reach, as we've mentioned, is a large news publisher. We have a variety of print titles and obviously a lot of websites and digital properties. We're doing a lot of interesting things with technology and data in particular. Let's get a start mark by just digging right into reach and how they're using data.

So why don't you just before we sort of drill into the complexity of actually, you know, making your data make sense and making it work for you, tell us a little bit about what reach is actually trying to do with data and how you're trying to leverage it. So Reach is the largest commercial national media news publisher.

We have lots of print media, but a a, a very large digital presence. It's news. So we've got everything from online websites like the Mirror and the [00:02:00] Express through to regional things like Cheshire, live North Wales, live My London, that kind of thing. So there's a huge digital footprint and obviously we, we've got a lot of content that we wanna get in front of the right people.

So, I mean, going to our purpose reaches purposes to enlighten. Empower, entertained through journalism, and that's all underpinned by data. So it's about making sure that we provide relevant content and more engaging experiences to our customers so that they remain engaged with us. We build a deeper relationship with them.

They show, we show them news that's important to them, and that's basically at a national and a local level. So, you know, obviously we. If we want to be able to show you things like local news such as our, in your area brand, we need to know where you live. So a lot of that is about value exchange between us and our customers.

Would that be, I go to the website of say, one of your publications and I would get content that's aggregated from all of your publications into that one brand front end? No. No. So the brand front end are, are separate. So we do obviously see what you're doing with each of the [00:03:00] brands. Right. But there's also the opportunity for, for us have, we have like an aggregated.

Put in where you live and, and, and brings together content from across all of that space. I see. But what's really powerful, all for us with the data is if you tell us more about you, then we can serve you a more, you know, content that we think you'll be interested in. So we can say, you know, here's a story that would be interested in, here's something that impacts your particular area.

And in return, That's kind of the, the value exchange really, which is if you tell us a bit more about you, we will give you, you know, more relevant content and we can recommend things for you. So we do a lot of stuff around recommendation and other data-driven tools. So you know, what you interacting with, do you want to see more?

Things like that are the related topics that will be interesting to you. And obviously on the other end of that, you know, We are funded by advertising. So we, we, you know, in, in terms of, it's not just, you don't just pay for physical papers. Obviously when you visit our websites, we show you adverts. We want to make sure that they are tailored to you and that you know they're relevant to you.

And so that we can both make it so that it's not something that's gonna straight our [00:04:00] customers, but also has a better yield for us. Right. And just again, before we go and dig into some of the architecture behind that and such, then the challenges. Just briefly, your team in the middle of that, how are you structured around providing data services to that ecosystem?

So I sit within the wider data and analytics team, which you know, has the kind of standard team you expect to see. There's a BI team, there's data scientists, there's a data management team, et cetera. I look after the data engineering side of things. I've got multiple. The sort of sub teams in there. So I've got a platform team that is responsible for running all of our data platforms and for ingesting data into our lake.

So data platforms are things like our lake, our enterprise warehouse reporting tools, that kind of thing. So obviously that's the core of it. I've then got a data services team, which provides services to the business based around data. So that's things like, you know, APIs to consume and also. Applications to help with the maintenance of that data as well as, um, you know, tools to help drive some value for the business around that.

In terms of [00:05:00] things like newsletter subscriptions and what have you, I've also got another team which is responsible for actually, you know, loading data into the E D W and doing the transformations in there. And I've got a team that's responsible for all machine learning engineering. And they are very, very clever individuals.

Not that everybody else isn't, but those guys in particular are, you know, really good at what they do. Um, and they drive all the things like the recommenders on the site. So obviously data scientists come up with the different sort of mechanisms that they want to use and the algorithms. But in terms of how does that get productionized and how do we run experimentation and ab tests on, on different recommenders all sits with them and.

Very cool. So if you step back then and look at how you're trying to leverage data across the digital front end and a quite complex digital front end with a number of different brands and requirements going on within it, where would you say the organization is on its data journey and maybe like in maturity terms, so you like level four, five data maturity, or are you actually on a journey towards that?[00:06:00]

I think, I think we're on the journey towards that. I think different areas are at different maturity levels, and I, I would say that's the same for every organization. I don't think I've ever been in any organization that can claim to be at level Maturity five or anything, and I almost, almost everybody feels like one.

It feels like we've made hardly any progress. It's really, really hard. Yeah, exactly. So I think in terms of where we are, I think we're in a good place. We've got a good foundation in terms of our infrastructure, so we've got a good set of data platforms in there. We're happy with the

technology. You know, we've got patterns and standards for ingesting data and transforming it.

And obviously we we're doing some very advanced things with that data in terms of machine learning algorithms and other bits and pieces. So I'm quite happy with that. I think data governance is a good area as well. We take it very seriously. So, People inside reach understand the value of our data and also the need to make sure it's used appropriately and kept safe and secure, which I think is very good.

'cause not all organizations take data quite so seriously. In terms of that. I think where we could be better is, I, I'm, I'm [00:07:00] very keen on self-service, but to do self-service, you've gotta have a huge amount of collateral in terms of. Data catalogs, data lineage, documentation, that kind of thing. So I think we could improve in those areas, but I'm, I'm pretty happy with where we are in terms of the capabilities we're providing and that we've got the right foundation to go to our next stage of growth.

It sounds pretty mature to me, mark. Thank you. It does. And I was gonna say, was there a point when the business really cottoned onto the power data? It sounds like happened a while ago because you, you're well on your data journey. Was there a threshold moment or was it something that there was a flywheel effect that went on slowly?

They understood that if you apply this stuff more and more suddenly there is a really good reason to to use it as part of your business model. So I think that happened. Um, unfortunately that happened before I joined Reach, I think. So it's quite hard for me to talk to about that, but I think there's definite awareness.

And one of the reasons I actually joined Reach was that the, the second point on their sort of strategy was data layer to opposition. And so to see something like data at the second, as the second point on. On a corporate strategy is actually very [00:08:00] rare in my experience. Usually you'll see there's four or five key things, and then there'll be something in a blurb at the bottom of your average company thing saying, oh yes, and we'll use customers.

We have some data. Oh yeah, we mustn't forget about data. Yeah, that's it. Where's what I really like about reach is a very upfront, and so this is really important. Obviously the journalism's the most important thing. But data's number two, and I think that's fantastic. Yeah. It's the same as, uh, security in many organizations where they are.

Yeah. We best put security on that statement as well, because that's important. Exactly. Yeah. Yeah, yeah. So, so I'm interested in, uh, in reflecting on the journey to this point then. So what have been the sort of core challenges and hurdles that you've had to overcome? Presumably one of them at least must have been aggregating data in the first place, given that you've got multiple brands, which I assume.

At one point will have been multiple different publications. Just tell us a little bit about the journey to here. Sure. So it started off again before my time with the data lake and then there was the, the initial parts of, of an enterprise data warehouse. And so it was very focused on sort of capturing data.

What it's about now is I think [00:09:00] trying to, and where we've tried to get to is how do we make it so that we can combine that data? 'cause there's lots of different sources and so the challenges are. How do you get all of that data? How do you know what you need and what has value, and how do you get it in quickly?

And obviously that's where data lakes help because ingestion is a bit easier. But the problem

is then when you do find that data is valuable, you need to transform it and make sure it's available so that people can actually use that. In terms of one of the big challenges, it's about sort of how do you do identity resolution across websites and different systems.

So we obviously have a central, you know, we have some tools that centralize things around user accounts, but you appear in so many other systems. So you appear in web tracking systems, you appear in an email system, you, you know, there's lots of different. Things. Plus you've got things like ad tech and MarTech systems.

And so how do you, how do you identify a user across all of that estate? It has been a big challenge that we've got a reasonable solution to. Um, and so it's, it's not just that, but also how do you transform and clean data? 'cause I'm sure, as you've probably all encountered, you go off, you scrape a bunch of data [00:10:00] out the system, that's fantastic.

You can report on it. Then all the anomalies are discovered, you know, One of these fields isn't very well populated. Somebody's put Jeff in there as their, as their, um, date of birth, you know, if you haven't got the right validation in place. So it's how do you get to this whole thing of not just how do you cleanse data after the fact, but how do you make sure that you, you can feed that stuff in and work with your, you know, your partners across the business to put things like validation in the source.

And maybe actually we could just dwell on on that a little bit because the clearing up problem and inconsistency problem, I think. Lies at the, at the root of why most organizations tend to find a data journey extremely difficult. There's another governance layer to that in a second, which I think we'll come to, but I'm just interested in how you guys got after that.

Was it literally like a hard yards sleeves rolled up, almost physical exercise? Or did you manage to create some automation around it? Maybe just go a level deeper on how you actually rooted that out. Sure. So our data management [00:11:00] team we're responsible, uh, responsible for looking for a lot of data quality stuff.

So they do a lot of good stuff around sort of producing quality reports where they report on, you know, some of the metrics in terms of key fields, you know, what they populated, what's their correctness and that kind of thing. And that's. Has helped to drive us to sort of prioritize sort of engineering tasks.

So you started with sort of case study like this. These are the sort of views that we need, like business driven perspective on it rather than data model up, if you see what I mean? Yeah, and I can't take credit for that. In fairness, uh, the head of data management's done a really good job in that area.

The engineers have then responded that and said, okay, well we can obviously do some clearing up. We can try and identify where these things have come from because,

Over here. Where did it come from? Where was the problem introduced? Often it's at source and then of course it's, it's how do you work with your colleagues in product across, you know, know, because you've got product managers across the whole piece to get them to prioritize and, and put things like validation steps in and that kind of thing.

And do you get the, do you feel [00:12:00] like you've got over that data trust issue? Often when you start this type of journey, there's some reticent to trust the data because of all the issues that, that you talk about where if I use this data in the wild and. Push it out to a consumer if I get it wrong. That's a bad thing, but if I get it right, it can be an extremely compelling experience.

I think so. I think a lot of it comes down to how reliable is the source system. Yeah. Um, so part of the thing is that you can't, the key thing is to make sure people don't, don't perceive data as a whole. You know, you have to be able to educate the, you know, the consumers and say, well look, you know, there are some, there's different tiers of systems and there's also, you know, different levels of reliability because you get into things like when you, everyone's on things like single version of.

Every organization has multiple versions of truth. Because often what you'll see is most organizations have multiple ways of tracking user behavior and visits, in my experience. So you'll often see there's like a core central one, and then there's companies often have GA because you end up integrate with [00:13:00] AdWords.

So it's about being able to articulate and say these data sets are clean and they are reliable. These may be less so, but there's still value. But you know, it's, it's how do you tier those systems and how, you know, 'cause you have finite resource as well, which is how do you make sure that you. Yourself sitting across all systems trying to get to a certain level of quality.

You need to make sure that, obviously things that are gonna be used for things like financial planning and business strategy are, you know, absolutely rock solid and other things that provide useful insights about certain things are perhaps, you know, less focused on. Do you find that puts a lot of pressure on the quality of the algorithm that you build so that that understanding of which data set is absolutely the truth and others, which may not be quite correct because you've got this massively distributed view really on your responsibility in multiple organizations.

You've got to consolidate that together and then server an experience out there. There must be quite a lot of pressure on the quality of the algorithms that you use and how you constantly tweak them as the data sets move around you that may not be completely under your control. To a certain extent. So obviously in [00:14:00] terms of gathering data and process input into general reporting, that's, that's fairly straightforward.

And in terms of we can apply, you know, standardized set of rules across things, when you get into the usage of that data for thing like recommender, that becomes much more nuanced. And what's the appropriate. Things to use is something that the data scientists and machine learning engineers are experts in and they know how to, how to go through that and identify, and that's where things like testing becomes really important because, you know, how, how do you experiment with that and see whether or not, you know, an an approach or a new feature is going to yield a, a better result in terms of, you know, people clicking through and seeing more stuff as opposed to, you know, internet, an engineering perspective and reporting perspective.

It's fairly straightforward to turn around and say, you know, does this data match this rule? Does this, does this data provide as additional value? Is a more difficult question to answer. Just returning to your journey then and your journey to here, we dwelt there a little bit on that data cleansing issue.

What are the other elements that have taken you to the point of maturity that you're at? So I think it's about making sure that you've [00:15:00] got clear roles or responsibilities. I think. I think team structure's really important. Um, not necessarily in terms of you, there is only one structure to do it, but when you do settle on it, you've got very clears and roles and responsibilities for teams.

So, you know, there's, there's multiple ways of doing that. So for us, we are tended to focus

more on particular areas. So it's, it's kind of a bit more of a vertical split to a certain extent for a lot of. The team. So there's a team that looks after the platforms and datalink. There's another ones looking after thew, there's another ones looking after some data services, stuff around that.

So things like, you know, apps for our business users and um, APIs for services. It might seem like an obvious question, but why is that structure working for you? What is it about that sort of shape that means that it's pushing forward the agenda around data? I think it's working for us because it enables us to have clear roles and responsibilities and we're able to align that around, um, the people that we have and the skills we want to build.

'cause you get into this whole thing of, well, you can go the other way and say, well, you can go with completely [00:16:00] cross-functional teams as an example. Yeah. Right. And try to look at things like data mesh and data products. And I think that's, that's, I think data products is, is absolutely the best way to do that.

But you need a very large, very mature organization for that work. You can't. Build something like a data mesh without every team involved in that understanding a lot about data, um, in terms of a platform level and transformation. And just for listeners who maybe don't have a mental model of what a data product structure would look like, maybe just give us a sense of what that would look like so we can compare the two.

So a data product is when you decide that you want to try and introduce product thinking. So a lot this got. Publicized quite widely, um, with the, um, excellent paper. I can't remember the last name about, uh, the author's name about, uh, data meshes. So the idea was that you would treat data as a product in the same way that you would treat things like website areas as a product.

So for example, if you were running like a retail website, You might have things like, you know, product pages or the customer, the unbook customer [00:17:00] onboarding as like a product that somebody looks after. You might have something like logistics as another products that somebody looks after or an area, and so you'd have those products.

As we look after logistics chain, we provide logistics services to the business. We provide customer onboarding, we provide product information, that kind of thing, so you can have those different product teams. Probably not the best explanation I've given. Hopefully it's suffices. It's also more sort of blended skill within a, within a team.

Yeah. Yes, absolutely. Yeah. So the whole idea of of multi, of cross-disciplinary teams, whereby each of those teams has got backend engineers, frontend engineers, testers, bas, delivery managers, et cetera. The idea with data products is you, you also go there with data and potentially you can either have separate things or more likely you try and say that that.

Data is part of that product. So what happens there is I could say, well, if I was responsible for, say, the logistics data or the logistics function, I will provide logistics, data products, which tell you things like stock levels as a data product and um, you know, uh, [00:18:00] warehouses and delivery lead times and things like that.

And so I could create that data set as a sort of mini. Data lake, if you like, and say, I will publish all these things here, and these are high quality pieces of information as opposed to the traditional model, which is that it's just exported or left in a database somewhere, and a central data team collects that.

Now, to get to that level, you've gotta have a lot of organizational maturity in terms of product management and a lot of data skills in each of those teams. Unfortunately, at the

moment, data skills are still relatively rare compared to general software engineering. It's still quite a very specific specialism, so you end up with quite large things, and it's where, sorry, quite large teams.

And where do you get to in terms of things like economies of scale, because that's always the trade off, which is. Do you have Crossdisciplinary teams that could do everything or do you have certain centers of excellence? And I, it's a bit like, um, DevOps as an example, which is a principle in terms of, you know, you want developers to operate their code and be able to write infrastructure as code and that sort of stuff.

What often [00:19:00] happens is you have a small central team of DevOps as well that responsible for looking at overall management of your cloud estate. It's the thing that I've seen. Now, obviously you guys know much more about this than I do, so I dunno if you've seen that kind of model where you, you have a mix of embedded resource and central.

Teams center of Excellences. Yeah, and I think there's, there's definitely the, the adage products and platforms. So platforms give you capability, products serve the business. The beautiful thing you talk about there with the data is the product, is the business. Understand what you're serving them and it, they can value it and they can govern it because you are serving it to them in a way that they understand it.

So it's not just random technology in the corner that does magic. It is, oh, I understand that function that that gives me because it helps me achieve my role and my job and my outcome. I. The power of the product is permeating throughout the IT organization because it, it, it's the, it helps bring that fusion of business and technology, um, together.

'cause everybody understands its same lexicon, same language, and you can talk about the same topic. So it's, it's, it's a powerful thing and I think the maturity is actually, you say you've got [00:20:00] you, you're well on your maturity journey and it's the shows, the sophistication that you built into your operating model to show that, that you can value and govern and exploit the data that you've got.

I think also one of the reasons for centralizing it is because ultimately data becomes really valuable when you combine it together. And that's why we've seen lots of interest in things like data fabrics. And so kind of having teams with specialists that can bring all of that data together and in a way and expose it, it kind of makes more natural sense for me, particularly around things like enterprise data warehousing.

Um, so I think unless you've got a very large organization with a lot of engineering cloud across each of the different divisions, centralization is a good starting point. So I wanna talk briefly then about the governance structure you've got in around the data and how you're linking, say like data infrastructure concerns and data clarity to a business objective.

I. So one of the key things is obviously we want to make sure that, you know, we are processing data in a way that is in line with, you know, various regulations, but also in ways [00:21:00] that that are sensible. And so what I mean by that is, you know, there are certain things you're allowed, so you're allowed to process data in terms of legitimate interest and in terms of consent, but also you want to make sure that you're doing things that are sensible with that data.

So not just in terms of making sure that you're treating customers fairly and that kind of thing, but. If you are trying to, to determine something, make sure you're using the right data to do that, because one of the problems is that you get this sort of misunderstanding of people talking about things like correlation implies causation, where they go, well, I've

plotted these two disparate points together and therefore this is happening and you've used the wrong data and you've got completely the wrong conclusions.

So it's about making sure that our data governance covers off. Is there a data protection impact assessment in place? Is this project using data in a way that it's allowed to, can we source the data we need to actually, you know, answer this question or, or, or provide this particular value? Because obviously one of the things we're trying to do is say, can we use data to tell us X?

Or we want to know if y have we got the data to do that? [00:22:00] Mm-hmm. Can we use that data, you know, in terms of we think it's the right, it's the right thing to do, and there's legal basis for the processing and that kind of thing, and is that data reliable? And it's kind of hard to talk about it in sort of general terms because obviously it really does depend on.

On exactly what you're trying to answer and what data you know you have. And so it does vary per project, but you have to sort of look at it and turn around and say, well, what's, it's almost like a little mini feasibility study, really, which is, is it feasible to use this data for this purpose? And. Have we got the right checks in there and how do we get an answer on that, like a, a finger in the air idea with that before we start an initiative.

And do you find the business, are the ones now requesting, is it feasible too or is it the tech and the scientists going, did you know it might be feasible too? Which, which ways the pool coming from in your experience? So we get a lot of it from the business. Now. I, I can't talk to a lot of that because there is a separate sort of bi function that looks after that mainly.

But obviously we know we get requests in from them and we do [00:23:00] get some directly from. From stakeholders and systems. So the business is very interested as they are in a thing. It just, well, this thing has happened, what can you tell us about it? Yeah. Um, or we want to understand this, or can we use data to, to determine if we did this?

This particular action we would see an uptick in, in, in positive behaviors or something like that. But that's quite powerful when the business have the data curiosity and they're coming to you 'cause they understand the potential that you can serve to them. So that's, again, that's that psychological, uh, or I should say maturity step that happens in the business that says, hang on a minute, I can value this data and then I can exploit it and get more out of it.

And that's a nice thing to see that they get excited about it as well. Definitely cast in the net forward. Then what do you think the next couple of years look like and have you set out a vision for what data usage in the organization looks like? So, yep. In particular, from the engineering point of view, what I'm interested in is making sure that we ingest as much data as possible and we try to move more stuff to real time [00:24:00] ingestion, because I think that's where some of the real power is coming from.

So obviously if you go back. 10, 15 years, a lot of data ingestion was just done as batch. But now that we see, there's so many things now that exist that enable us to do real time stuff, it's about trying to move as much of that into real time ingestion because not only can we then ingest that and do, you know, our standard daily reporting and that kind of thing, but you have an opportunity to react to signals and things as they happen.

And I think like things like streaming analytics are gonna really interesting and really powerful. You know, if you look at things like, you know, we event driven models and you know, so Kafka for example, has been a classic thing for this, and that's been around a while and I remember using it very early on.

The ability to basically pump all of your data into a central topic and then start to look at

signals is really useful. Particularly if you look at like streaming analytics and windows where you turn around and say, well, And, and so for example, if I use like a retail example, if you can see somebody putting things into a basket and then you notice that they haven't, they've dropped off the [00:25:00] site, the ability to do things, like, to try and reengage them and abandoned basket type stuff is really powerful for us.

Things like people looking at content give us real time signals on, on and allow us to change our recommendations on what you should look at next. So yeah, you've read this story about politics or, you know, um, a scandal, love island or whatever. Maybe you'd be interested in this thing. There's all that sort of stuff that can happen at an operational level, but also if you think about things like anomaly detection, that becomes really powerful if you can stream that data.

The big challenge there is of course, having sources that can actually stream that data to you. So internally that's not a problem because you can do a lot of stuff with events, but when you're dealing with third party systems, not all of them are rigged up for that, and that is changing. So we're seeing things like there's much more stuff in terms of being set webhooks and a calls third party platforms.

That's where I see a lot of investment going in terms of trying to get that data in as fast as possible, as well as more around governance. Good sketch there. I think then of an organization that's actually [00:26:00] probably well into three or four level maturity, I would say in the way you describe it in terms of.

Organized around it more than the first few steps in terms of clearing up the data and getting in the right place and actually clarity on, on what the path forward looks like. So, yeah, I think I, I really do think you've, I think to Shay's point earlier, it feels like a pretty mature setup with a clear direction forward and probably a number of other organizations would look at that and probably go, God, I wish I was there.

M VSIs, yes. Yeah. Yeah, exactly. Like, so what advice would you give people who are, uh, beginning this journey and actually maybe a bit bogged down and, you know, clearing up the data or something along those lines? I. Um, I think I would suggest by starting small and come up with a clear goal. So just don't try and boil the ocean is something I think is always really important.

I like the, the creation of things like exemplars, as you know. So this idea of saying, well, let's, let's find a use case and let's try and deliver that to the best of our ability rather than try and play whack-a-mole with data quality issues across the [00:27:00] entire estate. So pick a, a question you want to answer.

And that could be something as, as nebulous as like, what's customer lifetime value? Um, up to, you know, or things like, um, you know, if you're dealing with say like, you know, retailers, it might be something like, you know, logistic turnaround times or goes lost in transit or that kind of thing. And to take, take us a use case like that where you think you can probably get the data you need and build out something that enables you to do that.

So. You can build all the right things and build that up because think it's start off with one that is quality. Try and on. It would be things like build out a really good platform to enable that, build out all the, the most important thing with that is all the documentation, the cataloging, the lineage, the descriptions of that data.

Because if you want to get to the point of self-service, people have to understand data and know what they want to do. And if all you've got is a set [00:28:00] of tables or you know, like

a Tableau dashboard and a data set behind it with no description of what those things are, you are never gonna drive that change.

Um, so it's about giving people the skills and the information they need rather than having the best possible, most efficient E T L pipeline.

Sjoukje, what you've been looking at this week. So each week I will do some research on what's trending in tech, and this week I want to focus on why becoming a data-driven organization is so extremely hard. Being a data-driven organization has been a priority for many organizations for decades now, but we have seen mixed results in that.

And why is that? So according to a survey of executives, company culture is a harder hurdle to clear. Than any technical problem that they have. And on top [00:29:00] of that, the continuing explosion of the amount of data and also the growing concerns of privacy and data ownership makes this more difficult as well.

But there are a couple of principles that you can keep in mind when you do want to become a successful data-driven organization. And these are, firstly, you really need to think different. Because it requires a different mindset to become a data-driven organization. And you also need to fail fast and learn faster as well.

Learn through experience, which often entails trial and error. And lastly, you need to focus on the long term because it's a transformation effort that really unfolds over time. So I have a question to you, mark. Did you also apply these principles to your organization? Um, yeah, I think, I think a lot of them, I think one of the problems is going back to saying about thinking differently with, with engineering, for example, software engineering, you're very focused on, on things that you can [00:30:00] easily unit test, um, and integration test, and you're not worried so much about state is the main thing.

Whereas obviously with data it's all about state. So again, you know, you can unit test functions and transformations work very well as function processes and. But it's when you get into, well, you know, if you make a mistake or you need to go and reprocess or reacquire some data, you could be going off to get multiple terabytes and that isn't something that you're gonna basically load in in an afternoon.

So I think those kind of things are, are difficult. I think the challenges you mentioned around sort of governance and ownership are key as well. I think it's very hard to. To govern effectively if you unclear on who your data owners are and what you can and can't do with that data and what the nature of the consent is.

And, and people, particularly product managers, have to understand what data has been consented to be done with. Because just because you've got a piece of data doesn't mean you can do whatever you want with it. Um, you know, you've got very strict rules on what you have, what people have consented for you to do with that data.

And I think that's, that's another big challenge [00:31:00] as well as getting people who are owners, as I said earlier, so, How did you bed that in Mark? Was that something that. Was part of a human to human governance conversation, or are you managing it to build that into the logic of the platform somewhere? So obviously we have, we have things where we capture consent.

A lot of it is, is about understanding that projects are doing the right thing with our data. And so we've got, you know, obviously we've got a data protection office, we've got a data management function. The engineers are well versed, you know, everybody's well versed and understands what our responsibilities are.

Data training. The experience is built up over time around this and focusing on the long term. How long are you already into that transformation from a data perspective? Oh, we're, we're, we're many years in. I think it'll never end, but I think we are many years in, so that's why we're reasonably mature. I think it's difficult because, Obviously data, the number of data sources will only grow.

The volume of data will only grow. [00:32:00] And like any other engineering discipline, particularly in software engineering, I mean, technology's constantly changing. I think part of the problem is going to be, 'cause we are several years in it's worth at the time where we're looking at refreshing our technology estate and reevaluating where we are.

And that gives you a bit of a headwind because you are now to have we still got the right platforms we need, do we need to be considering what future platforms can give us? Because if you look at data warehousing now, Compared to 10, 15 years ago. It's predominantly in the cloud. I think the point you make about it never ends is absolutely spot on.

As organizations go on their journeys, whether it's a data transformation or whether it's a cloud transformation or whether it's a, you know, the other D word, the digital transformation. I think what, what a lot of organizations miss is the switch into a new way of working that more sort of exploratory.

Constantly innovating, constantly engaged way of working is not the transformation journey. That's the end state. Mm-hmm. And the exploration continues. Right. [00:33:00] Absolutely. Also at the, at the end of your point, you were saying now about is predominantly in the cloud. Where are you on the cloud journey? Is it something that you guys are now looking at or is it already done?

Uh, we're done. We're, we're, we're all in the cloud. The challenge with it is, of course, obviously as ch, as cloud services grow and more things become available, what do you, what do you look to adopt? So for example, you know there's a lot a w s, which is. Our main cloud is, is always introducing new services.

How do you find the time to evaluate those and look at what those changes are and if you should adopt them, as opposed to, you know, just ignoring them and staying focused on, on doing something else rather than choosing to adopt a new service. Should you invest back in something you've already adopted?

Not been like bright light, bright light, bright light, bright light, bright light. That's like, that's. The innovation cloud, the mag of cloud. I was gonna say click on. Yes. I love a bit of that. Thank you very much. I'm was prone to Mag Magpie development as the next person. It's a challenge. Yeah. I think it's that thing though, that is that, um, everybody went to cloud and went great.

This is new fresh evergreen. But then realized [00:34:00] that actually cloud now has an issue with technical debt. Or uh, backlogs that say, we did all that stuff but we haven't kept it up to date, et cetera. And there's this new cycle of, hang on a minute. Some of the things that we used to associate with legacy technology are now ideas that we still have to keep very conscious in our mind with cloud as well to stop that magpie effect causing proliferation of technology and then just a chaos basically.

It sounds like you've got good control over that, but is there anything in particular you've got around the best way to do that so that the engineers get the technology they need, but you keep. So you prevent proliferation of things that you don't need. We, we have clear rules

on what engineers, um, are allowed, what their remit is and what they can do in terms of technology.

So, um, what requires architecture, governance, and what doesn't? So, you know, we don't want to be really prescriptive in terms of blueprinting, everything. So, I, I. You know, you don't want to get into an argument over which Jason passing library is the best one to use and that kind of thing. Engineers know that and they can solve [00:35:00] that on their own.

Rob loves that sort of conversation. Mark, you're on. You're on really thin ice here. I'm worried if we get into it, you won't have any listeners left, but don't get stuck next to me at dinner party people. Me too. Matt, me too. Other listeners. Um, so. Those kind of things. You want to leave at a level and, and certain services you can whitelist if you like, in terms of, I should say allow list.

Um, so there are a bunch of things you can sort of pre-approve if you like, for people to be able to use and not be too restrictive. I tend to show the lines of programming languages because I think there needs to be conversations about that because, What you don't want is 20 engineers and 20 programming languages, and then somebody leaves and we can't maintain something else.

You know, and that kind of thing. You don't wanna, you don't want a very low bus factor for your team. What we do want is to enable the right amount of self-direction and autonomy in a role. So that's why I say we try to pre-approve things and say, well, these things, these technologies, yep, you can use those as, as you [00:36:00] see fit.

Um, you can't change cloud on a whim. So I don't, you know, we don't wanna come in one day and find out everybody's migrated. Azure or G C P or something like that, without any, without telling anybody. Although that would be a, a massive governance value surprise in the morning with migrating plan, A finops team would be like, what's happen?

Done? What? What about the reserved instances? I love, I love the idea of an engineer just going, yeah, all right, let's do it Friday night. Yeah, let's go. I, I do know of an organization that happened to actually, really? Yes, I did. Oh, my words moved. Cloud. Oh shit. You'd be like, well, what's the meeting on the Monday morning when you find out?

You go, you've done what now? Hey, my understanding was that there was a robust conversation. Luckily, I wasn't anywhere near that and I wasn't in that organization at the time. But yeah. Well, mark, look, thank you for that. I think on that note, The note of cloud chaos will, uh, will draw our conversation to a bit of a closer Mark, thanks so much [00:37:00] for spending some time with us and sharing your insights and journey.

Been a great chat. Thank you very much for having me. Pleasure. Now we end every episode of the show by asking our guests what they're excited about doing next. Now that could be watching. The new Ed Sheeran documentary on Disney Plus, for example, all the way through to something you're excited about in your professional life.

So, mark, whatcha excited about doing next? I suppose this is kind of sad, but I, I'm actually quite excited. I wanna play around with some LLMs, so I'm very interested in seeing what GitHub Co. I thought I'd give you enough warning about Rob's interest in this. Earlier, mark, God, you, you again. I'm, I'm feeling, I'm feeling at risk.

I'm sorry. I'm just really interested in, in seeing whether or not it lives up to the hype. Um, um, because I've seen some very interesting things, so I know that it can obviously stitch things, words together and I know there's been some interesting research done into LLMs are great because they can generate a lot of text, but they don't necessarily know what's

going on.

I kind of wanna try that for myself really. It just seems like really interesting when you read like the articles about how it can generate some quite. [00:38:00] But then it falls down when you ask, like chat, g p d questions about, I think there was one, some researchers asked it, what's better as a hair net if you're working a fast food restaurant, a, a used wrapper or a bun, and it suggested the bun because you didn't understand the under think lines.

So I just kind of wanna kick the tires on it and play with it, if that makes sense. It fully does. Yeah. And, and as sophisticated as it is, It is not actual intelligence, at least not yet. So I think you're right. It's, it's ability to distinguish about around these things when it's asked to distinguish, I think is still lacking at the moment, thankfully.

But yes, a few years ahead of us. I think with technologies like that, that are actually very unpredictable, both in terms of the power it will ultimately get to, but then something that concerns us on the show quite a lot at the moment is like the the duty of care that also needs to come into. Using it appropriately, not necessarily because it will, you know, lead to a Terminator two style singularity moment.

Though it might, [00:39:00] um, it, it's more about the actual societal impacts that will have on the way to that, if you see what I mean. Yep. Yeah, definitely. I think it's gonna be very, very interesting times. Fascinating stuff. Fascinating stuff. So, mark, thanks again for joining us. Uh, a brilliant conversation. So a huge thanks to our guest this week.

Mark. Thank you so much for being on the show. Thanks to our producer Marcel, our sound and editing wizards, Ben and Louis, and of course, to all of our listeners.

We're on LinkedIn and X, Dave Chapman, Rob Kernahan, and Sjoukje Zaal. Feel free to follow or connect with us and please get in touch if you have any comments or ideas for the show. And of course, if you haven't already done that, rate and subscribe to our podcast.

See you in another reality next week

[00:40:00]

About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of over 360,000 team members in more than 50 countries. With its strong 55-year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fueled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2022 global revenues of €22 billion.

Get The Future You Want | www.capgemini.com