FEDERAL TECH PODCAST

Ep. 25 Controlling the Hybrid Infrastructure

SUMMARY KEYWORDS
data, cloudera, cloud, rob, dod, people, agency, understand, infrastructure, lake, navy, decisions, run, hybrid environment, systems, security, analytics, managing, data centric security, civilian agencies

Welcome to the federal tech podcast. My name is John Gilroy and I'll be your moderator. Our guest today needs no introduction. I think everyone in the federal technology community here he has knows who rob Carey is he's currently the president of Cloudera Government Solutions. So because I'm a dutiful host, I went to Rob's LinkedIn profile, we all know he was the Navy, but pretty long time in the Navy and other companies as well. And so I have to come up with some common Navy reference. And we're gonna talk about data today. So I got to have a navy and a Data Reference. What can I come up with? It's creative, I know what it is, we'll go all the way back to Samuel Taylor Coleridge. And he wrote the Ancient Mariner, he said, water water everywhere and Not a Drop to Drink. So my lament is data data everywhere, and not a data point to consume. I mean, this is what we have, like these people behind me all kinds of data everywhere. And how do you handle it? How do you ingest it? What do you do with is it secure or not secure? So we're going to start off by going back to an old poem. Rob, how do you think of that? Well, let's do it. Yeah, so here's the situation, lots of new data points, a lot of information coming in every federal agencies, they're pulling their hair out, if I had hair, I'd pull my hair out to you know, there's a federal data strategy 2021, the State Department has a data strategy 2021, even the guys, the DOD, they have a data strategy. So so how can Cloudera help my listeners in handling all this data has been deluged?

03:11
In many ways, John, in many ways. So let's go back to the fact that there, there is now a call for and there is activity in this space. Years ago, when I was a CIO, I like to consider myself a recovering CIO, because you never quite get it get it off yet. But you know, the data person worked for me the data architect, right, so there are people worried about data structures, data formats, and how do I get my hands around? What is considered off authoritative data, right? And so let's fast forward five years to the early you know, 2010 and 2011 2012. We are now experiencing the benefits of Hadoop which the forefathers of Cloudera brought to, to bear in the market, a way of identifying and managing data, right? Although that is eons old, that is the start of this data revolution. Now we come into 2015 2016 2017, we're starting to see, hey, I got a data centric security, I have to take care of this, I have to be able to trust it. I have said in many speeches, that when we sit down in front of our computer, we look at it we believe everything it tells us correct. How would you know that it wasn't real? Well, you know, once you see really good guys manipulate things, you would you would understand that that's not necessarily a true statement that everything is real. But nevertheless, the genesis of decision making is data. Right? So now let's fast forward a couple more years. We're starting to see strategies being written

we're starting to see positions the chief data officer position is a new position. It did not exist five years ago. It exists at It aplenty today that

05:02
I gotta jump in 90 members, Chief Data ops Council and going

05:07
now. But if you went back three years, no one will own that title. So so the the acceleration from 2008 2009 through time now is Lightspeed compared to what some things do even though we're in the information technology domain. So now you have Chief Data and Analytics Officer, Chief Digital and analytic officers, as in DOD, for example. And so you have people sitting over top this structure of a strategy and an expected outcome, right, which is powerful, right, somebody actually now is caring about what you open up with. And that is their job. Five years ago, it was no one's job. There was a lot of people trying to work the problem. Now you had to look back five years, what tools existed five years ago, some, you know, Hadoop and sun of Hadoop. But now there are many companies in this space, there are companies like us who now can manage data wherever it happens to be. So you started out with, you know, hybrid data. And so data has moved from Legacy datacenters, old fashioned mainframes to now it's a hybrid multi cloud environment. What do we mean by that? We mean, it's in the cloud, it's in another cloud, it's on legacy Prem. And it may, in fact, still be in some mainframes that organizations cannot afford to get rid of. So now, if you're the CTO, or you're the CIO, you now have a really complex data architecture to manage. And now expect to bring your agency or department into the 21st century. With this, I don't say hodgepodge, but I'll say a hybrid architecture. So now you need tools that connect those dots. And that's one of the things that we do very well.

06:55
Now some people say data first, other people say, apps first, and then we said another company, I'll say, no, no, it's the infrastructure for so. So how do you rank them? Rob?

07:07
You know, I would say, Oh, they're all important. Don't get me wrong. But it's true. But but at the end of the day, you do not render decisions, let's just let's just, you know, play three moves here. First of all, we don't make decisions on infrastructure, right? Infrastructure is merely the necessary compute store environment to hold our data. And our next thing, our applications and our systems, right, okay, they're now actually performing functions on data that allow you to render decisions or conduct transactions. But the data is this Genesis of what is critical, right? Because we do not protect it. Yes, we used to protect the infrastructure in cyber, but now people understand data centric security is really important. Back to my comment, I want to be able to trust what shows up on my screen. How do I do that? Well, I know that that data, in addition to the systems and infrastructure are secure, are held to a reasonably secure standard, ergo, we have the FedRAMP ecosystem blossoming now, right. It's a little challenging at times to get through it. But it is a necessary evil to provide in the public cloud equivalent security, as you might find inside your firewall. So back to data as being that thing of utmost importance.

08:32
And the data can reside in a public cloud, private cloud, meta cloud, mega clouds, some big cloud, Rob's cloud. I mean, there's so many names for it. But I think what everyone understands is survey after study shows, most people understand that most of the data is migrated to some type of a hybrid cloud environment and didn't matter what name it is. And so then you have to have some kind of a tool to to manage that spread all over the place, don't you?

08:56
Well, you do. And so again, I hate to walk the dog back here. But remember when the vet Qudra announced cloud first strategy, he was in my show, I remember distinctly so and that and that started the genesis of hey, you guys gotta go there right now. We struggled with security, we created FedRAMP, we still struggled with the implementation of FedRAMP in that and there was like, are you doing security? Or am I doing security? Right, so now that we have that under control, we now have the ability to we fund information technology by program, right? We don't fund it in the federal government. Rarely is it funded at an enterprise level like anyone in navy doing everything with one tool. There are a few but but for the most part programs get funded, certain programs that got funded, move themselves into the cloud as they could because it fit what they were trying to get done for their program. And then others a were not funded or cloud didn't work for them. So So you know, you started out right away with In a migration to a hybrid environment, we didn't use that term. But we had cloud based programs or we had non cloud based programs. Now there's a push to get into the cloud. And again, there are certain conditions that that works for there, there are certain conditions that that worked for doesn't work for everything. And you wouldn't put you wouldn't spend money to continue to spend money, you would just let your infrastructure run its course and then decide. So today we've gotten from, we've gotten more into complex hybrid environments. So we have multiple clouds that government agencies use. And we have multiple legacy environments we have on prem private clouds that we use. And so you think about this, again, back to the statement, this is what the CIOs are managing today. Right? The agencies are running in a hybrid environment, whether they like it or not. So now they just have to, you know, sort of get their arms around it, along with the Cisco to protect it.

11:02
Rob, you and I both live in Virginia, if you get your car and drive out, you'll find lakes Smith Lake is one. So I have a friend who has guessed what a little cottage or a house on the lake or lake house. And that's what I bring, I got to bring this up because there are terms associated with Cloudera. And one of the terms is Lake House. So So how's this lake house term? fit this discussion?

11:26
So, you know, Lake House is is something that, you know, there there used to be data lakes, and there used to be data warehouses,

11:33
right. I remember four years ago, yeah. And the function

11:36
out. Again, this is very quick. So the functionality, basically, of a data lake house is the best of both worlds, right? So you can do store and compute, you can do transactions in there, you can do data warehouses are best suited for analytics, data lake store, right. So now, I want to be able to move data in run analytics on it provide access, and then I don't care where that data is, right? That's what CDP one does. The data could be wherever it resides, we connect to it, we bring it into a central location. And not only can I store it, I can run analytics on it, which is the genesis of providing insight to decision making, right? I mean, information technology exists to support decision making by whomever, right? And so you start to understand and that that's what it's there for, then you have a whole different perspective of why you have all these tools. Why do you care about data, because without the data, you can't get the insight to then make that accurate or insightful decision to help you move your agency forward.

12:47
If if you go to Google like I did this morning, and type in cloud aerospace lake house, inevitably, the next three words or four words is going to be cloud era data platform platform one. And so. So we're talking about here is Cloudera has an offering that can assist federal agencies and achieve their goals by managing the data effectively. And and that's really what clutter is all about is that a good summary

13:11
is and the CDP one enables a lake house as a SAS, right. So software as a service, right? So you're not, you're not buying the traditional way you're buying an offering that really does give you a fast and easy analytics on the fly. And you're you're just consuming cloud services, you are not investing in this infrastructure permanently. So as your needs scale up, you're consuming more if they scale down and you consume less, but you know, there's a, there's a rate at which you're now going to be able to gauge how much is data services going to cost you which is makes it a much better investment over time.

13:54
That's a phrase that's not used on interviews like this. But if you're sitting and having a coffee with someone, and you talk about the cloud and the hybrid cloud, and this kind of a casual phrase I'll use anyway, some people say is, Hey, Rob, you know, you're not I may get pulled for this. But hey, you know, we're not on the hook for security. You just dump your stuff in our cloud, you still gotta worry about this phrase is on the hook. So even if you move to the cloud, you're still on the hook for that security, aren't you? I mean, that doesn't just transfer somewhere in some cloud mysteriously someplace else. I mean, you're on the hook for that you have to understand exactly how to lock it down.

14:29
That's correct. And we use a product part of our product called STX, which is a shared user experience to provide a fundamental cybersecurity, I'll call a foundation. But you're right, the cloud providers provide a certain level of security they do, right. So if I went into FedRAMP, for example, and I went into moderate and there's 325 or 320, some controls, there is a specific set that each of the cloud providers would we got that covered but To your point, John, there are quite a few that are not. So they must be embedded in, in the cloud

infrastructure by the by us or by a an application, right. So you must cover down on these controls one way or the other. But you also to your point must understand who's covering it, either the cloud service provider as part of infrastructure, or you. And once you understand that, now you're ready, you're in a good space to take advantage of

15:27
cloud. But if you're listening to this podcast, and there's snow on the ground, and you type in Rob's name, you're listening to this, you might want to get information. And I read an article this morning by one of your colleagues, Gunny Rick Taylor. I mean, he's seems like a bright guy. He wrote a blog over at Cloudera. And it talks specifically about universal cyber keys cases for the cloud. It talks about specifically cloud era data platform. One is one subsection, universal data distribution solves DoD data transport challenges. And so we're just hitting the mountaintops during this discussion. But if you want to pursue this further, I would type in brick Taylor over to cloud era. And he explains it in very, very detailed manner. I think he's speaking at some events to you speak at events, Robbie, any events coming up for you? Yeah, I've

16:12
got a Fed scoop event I think I'm coming up on and there's a couple of things in the future. Rick just spoke at the FCA cyber conference down in Augusta. That's where this

16:25
makes sense. I mean, FCA and this whole system explained correctly,

16:29
correct? Yeah, so so it is important to understand, again, universal data, that shooping is a broad term to describe any data and its source to any destination in its source. That's easy to say. fairly difficult to do. But But that's something that we have tackled the ability to then gather up data, it could be endpoint data in terms of log files from all the endpoints, gather them up from wherever they may be, and put them in a place a central location, a data lake house, that now the organization can decide to just do what they want with it, right, they can push it into the SIM, they could push it into another place, they could run analytics on it streaming analytics along the way. There's lots of things that can be done that provide you with insights you didn't have, before you connected these dots, what's going on on 100,000 computers, or a million computers, right? The important thing is in the DoD space back to them is scale, right? So so to be able to handle edge data at scale is very crucial to to them protecting themselves from bad guys, and they are under attack every day.

17:45
I was in LA in August, I did a little swim race out there. And if you use the word agency to someone out there, they think of advertising agency. And there's one and two man shops and huge agencies in Washington, DC if use word agency, it can be big, small cabinet level all kinds of agencies, I would guess that there's different maturity levels for handling data within the civilian agencies. And so where does Cloudera fit in with these varying maturities of data sophistication with the different civilian agencies?

18:14

So good question, I would say all of them have a level of expertise, at least maybe centrally located, they all now understand that have to get on top of this data horse and ride it. It may be a small agency that runs a policy organization, it may be Department of Transportation, it may be Department of Homeland Security, it may be the Department of the Army, right? That agency is a generic term, but they do have the understanding of I have to protect this data, I have to then use it to conduct mission. Right? That's it, they're doing business and us as we say, you know, they're actually performing business operations. And we are conducting mission, right. So in supporting the mission of that agency, it depends on what they might want to do with it. The FAA, for example, might be tracking everything airborne, and then being able to analyze what's happening in the air, excuse me and provide insights on that to then make better decisions on what radar should look like, for example, the DoD could be doing it on lots of things that are related to national security warfighting, the IRS may be doing it based upon Okay, let's look at all these tax returns coming in. And let's see which ones are real and which ones appear to not be real. Let's see if we can use data to help us discern that without somebody looking at the return with their eyeballs. Right. So you know, automation starts to take root here. And the basis of automation is artificial intelligence and machine learning, which is all about how do I look at this data and draw conclusions from it? And then validate those conclusions in a model and now run it on live data to see what what pops out.

19:57

Now, if we're talking about the IRS, they're gonna hire 1000s of people and they maybe look Hiya, John Gilroy, he's tax returns. And they'll come in on Tuesday morning and have a cup of coffee and go through them and maybe in a couple of weeks, we'll send me a note and, and have to explain something in the DOD. They're not gonna be sitting in come back in two weeks. If the sensor brings in something, they have to make sure that data is accurate and make it actionable. And so it's not a matter of you know, put your feet up and waiting two weeks it's it's right now so you have a sensor coming in from a satellite and you get some kind of an image you get something from something in the water somewhere. So managing data seems all nice and almost academic except in the DoD will know it's life or death, isn't it?

20:36

Well, it is. So today you know, general crawl who is the J six little while back started the journey of jazzy to is the acronym but joint all domain command and control what is that? Sounds like alphabet soup, but it is very crucial to be able to take the now five services right Army, Navy, Air Force, Marine Space Force and try to integrate the way they will talk to each other and how they move data back and forth in support of warfighting missions. We we we do joint warfighting today. And we have to continue to do it in the future. So So now there's some level of standardization that is a must, right for the Department of Defense to be able to conduct its national security mission. So So general Kroll has embarked on this overarching framework to help the Army Navy Air Force marine Space Force come together the Navy and the Marine Corps are using Project overmatch. The Air Force has ABMS systems, so, so they are taking the guidance and implementing it in their space. Now, the question is okay, we will need to reconcile how we use data to like, as you said, John, you inferred you know, we have a airplane or a missile, we're tracking it, how do we hand that off from one platform to another, and then conduct our mission, whatever that mission may be for that particular target. So they're working through those standards and working through the systems that they currently have, because they can't

just flip those systems over into a new paradigm, they actually have to have some rules setup, which is what jazzy too, is doing. So I applaud these efforts, I think they're on the right path to delivering the ability to make decisions with that data in near real time.

22:28

And I think their topic comes up about volume and complexity of the data that's coming in. And the military doesn't deal with one blip on the screen, it could be 1000s of screens simultaneously, on Christmas morning at 2am. I and decisions have to make quickly. And if they're not responsible to handle your state, the data could be corrupt, something could happen. It may be heaven forbid, there's there's unexpected delay, or there's some kind of a problem in delivery. And so what you have to understand is that, I guess I'm an old guy, and I've seen a lot of things that maybe the system isn't online, maybe there's a person in a country and and they have to communicate something and the communication sounds. So how do you make it resilient? And resilient, I think is a really important word for this joint domain. It's resilient, resilient, resilient, because you're not may not always have connections, do you? Where you don't

23:17

you don't and, you know, the let me go back to basics here a little bit. The first thing when they wrote their data strategy, and there was of course, a federal data strategy, I think 2020 Yeah, they recognize it's a strategic asset. That's not a loosey goosey Uber, right, hey, this is how we make decisions, like you said, when something's coming at you, you must decide, is that a flock of seagulls? Or is that a flock of drones that have weapons on them, you need to be able to decide pretty quickly so. So now the technologies associated with discerning those things like AI, right, we stood up the jig, right, the joint Artificial Intelligence Center, and then and now there is a CDO, right, who is responsible for taking data and making sure that these decisions can be rendered as rapidly as possible using state of the art technologies like AI and ML? So you know, the question is, okay, I still have to do some work. AI is not instantaneous AI is fairly complicated to build a model, you know, run your data through the model, get the expected results, then you can take live data, and that takes little bit of time, but it is becoming more and more, I don't want to say rote, but they're able to quickly learn from what they've done in the past or certain types of data to then apply it, tweak it and get it done. And so, you know, DOD is is a heavy player in this space, as all and all the services are making heavy investments and how to render data useful in at the speed of war, which is, as you said, near instantaneously.

24:50

Yeah, I had to bring that question up because I went to your company Twitter feed, and one of the hashtags was hashtag data in motion. And the data could be sales of shoes. By From your perspective, all important than shoe size or types of shoes coming in the port in Los Angeles is something so hashtag data in motion, hashtag machine learning and hashtag, data lake house. So ending this interview here. You certainly have a varied career and seen all kinds of different things. What do you see on the horizon next five years easy transition, a lot of people kicking and screaming, maybe some people dissenting so what do you see the transition taking place in the next five years from so

25:27

in this space? John, you know, what I'm hopeful of as we commoditize data, and we commoditize AI? What do I mean by that? Today, data is a field that you must be expert in. So you're a data scientist, a data engineer. There are not 1000s of those guys on the street today. A few years ago, if you remember, there was a shortage. And there still is a shortage of cyber experts, right? People understood network security. So now data is in that same spot where we need more of those folks. But now I also need to get, I need to get data out and about to the average analyst, whether it's an intel analyst, a financial analyst, what have you, I want to be able to allow them to engage the platform's right with writing simple, simple applications that would engage in engine like CDP one, for example, and then draw conclusions as a matter of their job today, that's a little bit of a hop, skip and a jump to get to the data scientists who can then turn the engine on run what you want and get back to you. It's not as simple as everybody can do it, right? Because these engines are fairly complex, but they need front ends, we need to be able to commoditize data commoditize AI to an to a degree, and then more people can use it. And now we start to really accelerate the mission of government, whether it's national security, or it's the civilian governments or state governments or or local governments

26:58

who want an interview. We started off talking about the Ancient Mariner, we ended up talking about data scientists. So we've gone from the 19th century to Star Trek and sounds like it's pretty good. Unfortunately. Here, Robert running out of time. I'd like to thank my guest Rob Carey, President Cloudera government solutions on the federal tech podcast with John Gilroy. Thanks, Rob. You bet. Thanks, John.