FEDERAL TECH PODCAST

Ep 50 Data Storage Strategies for Complex Federal Systems

00:14
Welcome to the federal tech podcast where industry leaders share insights on innovation for the focus on reducing cost and improving security for federal technology. If you liked the federal tech podcast, please support us by giving us a rating and review on Apple podcast.

00:38
Welcome to the federal tech podcast. My name is John Gilroy and I will be your moderator. Our guest today is Zach Duncan, Director at SFB at a company called Pure Storage. Today we're gonna talk about data strategy for large federal systems. But before we begin, I would be remiss if I did not say we're recording this from the floor of elastic con in lovely downtown Washington, DC, all kinds of people here you can hear but the background, it's about 500 people here eating lunch, we're getting like stuff thrown at us and muffins and pieces of bread and everything else. And so my question to you, what have you learned the conference, please act?

01:14
Well, John, I've learned that logging analytics. And cybersecurity is a huge push in the federal government, especially last year after the Biden Administration released the executive order on improving the nation's cybersecurity operations. And then the mandate came out around that around logging analytics and storing your your analytics and machine generated data to better ensure that you're protecting your agency and mission critical data. I've had a lot of conversations with customers about that elastic fits heavily into that log analytics, cyberthreats tracing IoT data, stuff of that nature,

01:52
I don't remember the exact executive order, I think it was in July of 2021. When they talked about protecting the data. It's all about protecting the data because that's where the secrets are. That's where the information, that's where the personally identifiable information is. So that's what the focus is, isn't it?

02:05
Absolutely. And with ransomware being a huge topic, not just in the federal space, but across commercial enterprise healthcare, any industry, you know, CIS a DHS, any of those entities that deal with cyber and data protection have mandates out there on ransomware how to protect your data, how we're gonna deal with insider and outsider threats. So that's been a huge topic as well today.

02:29

Well, I never was a spelling champion in grade school. But I sure can spell Pure Storage pu r e s, t o r H, E. And because I'm very deft with a computer, I went to your website today. And what that means is big headlines. 96%, less Rackspace. 85%, less energy. So tell me exactly what Pure Storage because how's the Met manage all this magic?

02:50

Yeah, so pure storage. If you're not familiar with us, I'll give you the high level overview we were founded in 2009, out of Mountain View, and what we really wanted to do was uncomplicate, your data storage. So we built from the ground up our hardware and software to run on all flash, right. So our claim to fame is being the best and first market all flash data center storage. So now we've infiltrated over half the Fortune 500 We work with in all industries and verticals. And then the federal government, we work with DOD, civilian and intelligence agencies across all security and classification levels, to help them improve their data operations make more efficient use of their data, and really move at the speed of mission with the data that they're capturing and harnessing.

03:36

Now there are certain agencies in town here, we're in Washington, DC, there's the historical Smithsonian, and if they have data there, maybe they can record things about Thomas Jefferson, you can just sit there and that's fine, you don't have to worry about retrieving it. However, if you go across the river to the Pentagon, there are certain organizations there that they have to retrieve that data quickly. If they don't retrieve it quickly can come out dated. So this is very, very important for many of the people within Washington, DC area, isn't it?

04:01

It is data. If you take a look recently, go on LinkedIn, read any reports from DOD or any agencies out there in the civilian or intelligence agency space, they either have or are implementing Chief Data Officers, right, they really want to make better use of that data. And typically, traditionally, I believe the federal government has been seen of a hoarder of data. And I don't mean that in a negative connotation, but what we really want to do is not only store that data for the long amounts of time that we have to whether it be for mission or citizen data, but we really want to take that data and we want to gain insights from it moving forward. It's no longer just about storing your data. But how can we better use the historical data that we have to make better real time decisions, better improve citizen engagement, and better improve mission and agency outcomes?

04:48

I'm thinking about tears here and different levels of importance for retrieving data. If you go to my house and you want to retrieve a knife and a fork, you can get that in kitchen instantly. Okay? Now if you want to retrieve maybe a few I owe from my office, it takes a bit longer than if you want to retrieve the pictures of my kids and 20 years ago, you got to know the basement and dig all this stuff out. And I think the federal government maybe has to structure with tears as well, because why should you afford to have very expensive fast storage for something regarding Thomas Jefferson? Were something maybe from a satellite imagery was more important. So I think you're offering and elastic really fits in with different tier concept of of importance of information, doesn't it? It does,

05:26

I would say, there's a paradigm shift, right? We talked to federal customers all the time who say maybe flashing it is too much for our data needs, or it's to perform it. But if we've looked at how CPU in the speeds of CPU, and now GPUs infiltrating the data center, and the rise of AI and ML, and the this pace, it would net networking speeds have increased, you know, it used to be one gig. Now we're 1025 40 100 gig 400 Gig networking speeds on the market, right? We believe that flash is going to overtake spinning disk and hybrid arrays. And in the very near future, Flash will be the predominant storage in your data center to keep up and not be the bottleneck with the increase in speed for CPU, and networking.

06:16

There are about 500 people in the room, I would guess about 10 people in the room know what this term is. And that term is NV me perhaps you could explain that for the the smarty pants in the audience. Folks,

06:27

well, NVMe is non volatile memory Express, it's just a faster way for data in SSDs. Right, think back this data. One thing we'd like to talk about a pure is serial versus parallel. And when I'm talking serial, John, I'm not talking about Captain Crunch. Okay. So when I say serial, think of legacy technologies that had minimal connections going in and out. What we've done is we've parallelize the way that we talk to the flash with our direct flash modules. And we have more streams going in and out for your applications to talk to your data storage. What does that mean, in layman's terms? Think of when you go to the DMV, if you had 100 people, but you only had one teller waiting on those people, how long would it take for you to get that up there to renew your ID? Now what if you spread out 20 tellers to handle those same 100 people? Think of the efficiencies you get from that. So when we say serial versus parallelism, nbme versus Stata, those are the things that we really want to hit on.

07:27

I remember years ago on those things called scuzzy drives. If you remember that far back, that's a whole different technology. But again, it's involved with mSATA as well. I got up early this morning, and I did some research on Pure Storage. So I went to your website, but I also went to Google Trends. Okay, so I typed in Google, I went to Google Trends, and I typed in Pure Storage. And I see 2015 really started. I mean, there's up until the right, I mean, it's a great shift. Also, I typed in NVMe, and almost parallel with that exact same thing. So it looks like you're well known for this area of NVMe, inspect, you must have some expertise in that right.

08:02

I wouldn't say I'm an NVMe. Expert, I would say I'm a data expert. Ah, so let me ask you this. Do you like waiting on things? Think of all the applications you use in your day to day life? Now, John, that he didn't use in the past? Do you like waiting on the results? Yeah, no, I remember what happens when you start waiting, people get impatient, especially in today's society. So what we've really done a pure is we have built our solutions to address the need of modern applications and modern workloads, across file block and object storage. Okay, structured versus unstructured data. And we want to be the bridge or the conduit from on premise to hybrid cloud solutions, and be able to uncomplicate your data storage, and handle all the types of

data that your agency or enterprise is interfacing with, so that you can make better insight and utilize the modern applications that your customers and constituents are using today.

09:01

Well, we got to talk about Tom Cruise every what's your favorite Tom Cruise movie? Well, here's we're gonna talk about, we're gonna talk about the need for speed. Remember that we talked about the need for speed. And that's what we're talking about here, the need for speed, Google did a study, and they looked at people who looked at websites on a phone. And if it takes more than three seconds for that phone to load, 53% of the people are going to bail now 10 years ago, that wasn't the case. 20 years ago, it would take two or three minutes to load a machine. And so I think this is exactly what's happening in the commercial marketing world. People are looking at low times and think it's very, very important. But look at the federal government. There's also a e o initiative, talking about user experience and citizen experience. And so we can get something from Pure Storage on here and talk about meaning data, talk about these complicated storage memory terms. But really, the end of the game is user experience citizen experience. It's what you're telling me. If we use peer storage, then our website will load faster and the better responsive for our citizens. It really kind of It's a long connection, but I can see it

10:01

absolutely, I actually can talk about that customer constituent experience as well. So not only is the data important, right, because you want to be able to make real time decisions. But I used to interface with a company called ServiceNow, who does a lot with customer and citizen engagement. And what you really want to do is you have to have the data stored on infrastructure, or appliances that can move at the speed of real time decisions. But what you also want to do when you get to that end user experience is think about the apps you use on your phone or day to day life, right? If you have an issue with your cable, or you have an issue with your cell phone bill, you can interact with a chatbot, or digital agent or someone will text you how are we getting those digital experiences we have in our daily consumer life with the apps we love to interact with? And how are we emulating that in the federal space for our constituents and our consumers. So I think there's a mesh amongst the applications and the data that are needed to get to those experiences.

11:02

This is Malcolm Davis, from deputy deputy listen to episode 50 of the federal tech podcast to hear how IT Service Management can reduce costs for federal projects.

11:14

Sec Duncan here with Pure Storage, I went to your website and I took some notes. And there was a phrase on there may was a blog post or category something's called uncomplicate data storage. Maybe that's the title of this uncomplicated data storage because it's one count, it's going to be faster, it's uncomplicated. It's not going to be lost, or very difficult to find at all. At your website. There is a big fancy phrase again, we're going to throw out some big terminology. Here were 500 people in the room, maybe there's half of them. It was it's called Flash blade. So what is flash blade?

11:46

Well, Flash blade is one of piers, flagship products. And 2016. We realize the need for modern applications and modern workloads, and the exponential growth of unstructured data. So when I say unstructured data versus structured data, just a real general overview. Think of structured data as your back office it data your VDI, your VMware deployments, right. Unstructured data, think of stuff like Video, Audio sensor data, geospatial data, GIS data, machine generated and IoT data like we're talking about here and elastic on today, AI and machine learning, unstructured data, the way it differentiates itself from structured data is it each piece of unstructured data is its own unique snowflake, so to speak. And so when you store all this unstructured data across all the types and workloads that I just discussed, it becomes more difficult to analyze and gain insights from so what we saw at pure in 2016. And why we develop flash blade from the ground up for unstructured data was the rise and all those modern applications and modern workloads. And another thing that we saw was that across all those unstructured data workloads, video, AI, ml, analytics, high performance computing, we were creating data silos not we as pure but the data community. And by creating data silos, whether it be built on direct attached storage, or network attached storage, or in the cloud, whatever you were storing in any recreating the silos and it made sharing data across your organization or agency inefficient. So what we did with Flash blade is we built it from the ground up with our own hardware and software to address unstructured data and break down those data silos. So now instead of having multiple environments, you have to maintain, we can run analytics, AI, ml HPC, healthcare impacts and imaging data, data protection, the ability to rapidly recover it all on the same infrastructure, what we call that is unified fast file an object or UFO. And that's the benefit. Our customers are really seeing out a flashlight,

14:01
we're at elastic con and I listened to the keynote this morning guy named ash and bow. I ran out of ink taking notes, there's so much going on. He concluded with his excitement about what they're looking at as far as vector. And when he said vector he said video and audio. And so what they're looking at search now is not just searching for some simple file, like a database or something. But this is searching for an audio or video again, it plays exactly right into your concept of flash by doesn't it?

14:32
Absolutely. And you start thinking about why people want to rapidly search audio and video, I'm pretty sure we could all come up with a ton of hypothesis of what that would be used in in the federal government. We can look to some of our countries or continents outside of the United States and see what they're doing with that data. Flash blade absolutely helps address the need to store that as well as rapidly search it think about sports car In a semi truck, if we put them together, that's what flash blade is right? You have the ability to move very quickly and carry a large capacity of data at the same time.

15:09
So let's continue this analogy. So where does elastic fit into this got sports car got a truck. So what's elastic is elastic, the boss with the motor, what's the last elastic

15:17
is one of the engines that drives that data creation? I would say. So think about the unstructured data use cases, I talked about analytics, elastic, specifically log analytics, or IoT and machine generated data. Think of

the massive amounts of unstructured data that our sensors, anything that we're running telemetry on whether it be networking devices or other devices on our agency's network. Elastic captures all that data and how we really help the traditional elastic user or someone who cares about cyber threats or log analytics, traditionally, that would be built on what we called direct attached storage nodes. Think of a server with Compute and Storage in it right. And so your elastic environment would be spread out across that what we did with Flash blade and pure to really help is we disaggregated that compute and storage node. So we took your storage, we gave you a performance storage subsystem, that you could grow as your data grew, and disaggregated the compute and storage so as your elastic environment grew, and you needed to increase your number of indexers or nodes, you could just add more compute as needed. Or if you wanted to add more storage, you could just add more storage as needed. And as your environment grew, we gave you that ability to do it non disruptively. So we're performing it that we're the performance storage subsystem, we're easy to use, we literally have video somewhere floating around of a seven year old installing a peer system somewhere. And the thing that really helps all of our customers, whether it be federal enterprise, or customer is what we call our evergreen business model. I don't know if you've ever bought enterprise grade data storage in your life, John, but for the people that are out there that have you buy, let's say 100 terabytes or a petabyte, and then three, four or five years down the road, whether it be spinning disk or hybrid arrays, it's starting to get a little long in the tooth, right? So they come back to you and they say, Hey, your storage systems getting hard to maintain, we're gonna have to raise your maintenance costs. Or we have this newer system over here, which maybe is equal to or about the same cost if you just renewed your maintenance. Well, we did it pure with our evergreen business model. Think of products like the iPhone or Tesla. You buy the hardware upfront, and it gets better over lifetime as you upgrade the software and features get enhanced. So with pure you never rebuy the same terabyte, we give you all the performance and future proofing you need. And as our product gets better as we upgrade the software, we do our own software and hardware, your system gets better over the life that you own it. And we've had federal customers and enterprise customers that have taken us we're first generation FlashArray to where we're at now, or flash blade, and they have migrated non disruptively, and really uncomplicated their data storage like we've talked about, and the products gotten better with age, we began

18:18
this discussion with some statistics from your website. 96%, less Rackspace. 85% less energy. I think these two concepts would tie into sustainability because you're producing the same results at less cost, but also using less energy. So it fits into sustainability and green, doesn't it?

18:37
It does. I don't know if you've ever heard of the federal tech podcast.

18:43
I hear it's fantastic. It is. It's

18:44
great. I was listening to it the other day and there was an episode about 70% of internet traffic flows through Ashburn and in northern Virginia area and the data centers. When we think about the public cloud providers, there's less than 1% vacancy in them. Is that correct, John? Yes, it is. So we think about cloud smart cloud first

data center optimization and consolidation efforts of the federal government, like maybe we don't have to go straight to the public hyperscalers are cloud providers. What if we were able to reduce our footprint on prem with modern technologies that consume less water power, cooling, all that Swapsies stuff that we traditionally talked about? And that's what Pierre is able to achieve? We have a ton of customers that use us because of our environmental issues as well. And if you actually go to our website, and Google Pure Storage, ESG or pure storage, environmentalists, they have a calculator where whatever your current data platform is, you don't have to put in specifics you just put in generics about your amount of data, Rackspace, all that and it will put out how much savings we produce. over a six year TCO, I believe, it's very beneficial information.

19:55
I'm trying to see where I'd see a pure storage logo. Imagine if I went to one of the those data centers out in Ashburn, I would see some logos out there. But so are your customers only in data centers? Do you have other customers? I would imagine, not just hypothyroid. Here, there may be three letter agencies in town here that maybe can't use cloud solutions and may have some of your products in there. Is that a logical conclusion?

20:16
Absolutely. We're across every branch of the federal government. We're, obviously data storage is going to reside in the data center a lot. But when we go out and talk, we're talking from sea levels down to the system architects, right? Everyone cares about data. Everyone cares about environmental ills. Everyone cares about using simpler solutions that are efficient and more performant. So we talk to everyone across industries, from the people administering the systems all the way up to the board level decision makers that care about these other things we've been discussing.

20:51
So if, if you're listening to this, and there's, you know, baseball's being played, maybe the Atlanta Braves are built playing, where can they get more information? Or they're like white papers at Pure Storage, they can read? What what's the best source of information about peer storage for my listeners,

21:05
pick your poison, I would say peer storage.com and go to YouTube, you can YouTube, peer storage videos on a variety of topics. We have a podcast is not as as famous as the Fed tech podcast, but it's called the pure report. I think we have an Instagram page, you can follow if you're into that. I don't know if we have a tick tock yet, but I'm not going to administer it if we do. So there's a ton of ways to interface with this. Or you could just reach out to your local Pierrot.

21:35
Before we leave, I want to give a prediction. Here we are at elastic con. And we have a lot of people Ed here kind of walking out of the room. We're recording this, what do you see this whole business transitioning here in the next four or five years, as far as elastic Connor as far as how we're handling our day together, and the whole idea of data storage and, and vector and all kinds of trends we see. I think

21:55
elastic con, especially with the amount of data that we're generating from more than just human sources, right? With the proliferation of IoT and machine generated data. Tools like elastic con and log analytics, and tracking your events and Incident Management are only going to get better. And the amount of data is only going to continue to grow. I don't know what the stat is, I'm paraphrasing. But in the last two years of human society, we produce more data than the combination of all the years we've existed before So data is not going anywhere. If we want to improve customer or citizen experiences, we're gonna have to use data, not just for store it and keeping it forever, but to gain insights and move at the speed of mission that we've talked about. And I think federal leaders are going to see that modern applications, modern apps, that customer experience that we want to provide for the people that work in our agencies as well as citizens are going to require more and more data that we're going to have to secure and take care of.

22:54
And Pure Storage can help all of our listeners achieve that time to insight that they're really seeking, aren't they time to insights important? You've been listening to federal tech podcast with John Gilroy. I'd like to thank my guest Zack Duncan, Director ETS FB at Pure Storage.

23:11
Thanks for listening to the federal tech podcast. If you liked the federal tech podcast, please support us by giving us a rating and review on Apple podcast.