Harold Byun:

Hi. Good morning. Good afternoon. Good evening. Welcome to today's webinar on Cloud Data Risk. My name's Harold Byun. We're going to hold for just another moment or another minute or so before we kick off as people are still getting in, so appreciate your patience. We'll be back with you shortly.

Harold Byun:

All right. Well, welcome again to Cloud Data Risk and How to Better Protect Your Data in the Cloud. My name's Harold Byun. I head up product management for a company called Baffle. Today, what we're going to do is provide you with an overview of some frameworks that people have been using for cloud security and data security in general. The agenda for today is to cover that at a high level, look at some of the trends and security concerns for data in the cloud, which would seem to be fairly obviously for many people. Then we'll look at what we consider to be some common misunderstandings of security controls and the overall threat model for security for cloud data. Lastly, then, we'll look at current options that are available for you.

Harold Byun:

The demo today, I've been having some connectivity challenges, so I will be walking through some of the capabilities and referencing them. We have a live demo that we do every week, actually, that's about 10 minutes long. If you're interested in the live walk-thorough where we set up multiple datastores and actually protect data on the fly in about 10 minutes, you're welcome to review that. I apologize for that in advance, but hopefully, the material will be interesting to you. I encourage you to ask questions throughout. You can use the QA chat as well as you can always email us, as well, if you have private questions that you would just like to discuss more on a one-on-one basis.

Harold Byun:

Without further ado, let me go forward here. Just a little background on me, I've been working in security for about 30 years at this point, both on the practitioner side from a security architecture standpoint as well as on the vendor side. From a focus area, I would say that I've been concentrated more categorically in data containment, in general, how to better prevent data leakage. I had an extended stint in data loss prevention, worked in mobile containment for a while as well as a stint with CASB or cloud access security brokers, so that's been the bulk of my career. I've been working with Baffle for a number of years now really around core-level data-centric security, so that's just my background in this area.

Harold Byun:

Let's start off by looking at some frameworks for cloud security. When we look at one of the more familiar frameworks for how to manage data in cloud and really looking at what we would... This is a picture, obviously, of the Shared Responsibility Model from AWS. It really is a comparison of infrastructure management versus data management.

Harold Byun:

Really, the bottom half of this is what AWS really focuses on and what they claim is their portion of the responsibility, which is to provide you with the infrastructure, and the hardware, and the networking, and the ability to manage and secure that environment, whereas you, the customer, are responsible for the security of the cloud and the infrastructure itself, so configuration controls, managing vulnerabilities,

managing open services and, obviously, the data that you put in the cloud. This is one of the core frameworks that we see people relying on and have been relying on for the better part of a decade in terms of how they approach the infrastructure provisioning of services versus the actual security and configuration of those services.

Harold Byun:

Within that context, you can look at Gartner's landscape of cloud security. Really, the way that they've outlined this is, on the one hand in the top, you have a category known as CASB or cloud access security brokers. These vendors...

Harold Byun:

Well, and hopefully, you can hear me now that we're back. I apologize. It looks like I just got dropped off there. Hopefully, you can hear me now. I appreciate you letting me know that.

Harold Byun:

CASB, obviously, cloud access security brokers, and they are largely focused on discovery and monitoring of cloud access services.

Harold Byun:

CSPM are cloud security posture management vendors which are really looking at configuration controls and the ability to ensure that appropriate controls are applied for different components or services running in the cloud infrastructure.

Harold Byun:

Hopefully... It looks like I'm still having a problem with the audio here. Bear with me. All right. Okay, it looks like I am back now.

Harold Byun:

The cloud security posture management vendors, as well, looking to lock down different types of services based on policies and automatic configuration control checks.

Harold Byun:

CWPP is really referencing cloud workload protection platforms, are really looking at things like containers and the overall posture of runtime environments operating in the cloud, so vendors like StackRox or Twistlock or Aqua in the CWPP space, CSPM really referring to folks like Evident.io and RedLock. CASB, some of the more familiar vendors might be Netskope or Skyhigh Networks, folks like that.

Harold Byun:

The interesting part about this framework and this view of data is that it is, obviously, missing actual data protection. These are all focused on, again, securing the infrastructure, securing methods for people to access different services, really talking about the overall posture and availability of those services and how people access it, but it doesn't really necessarily speak to how people are securing the data. We always found that a little bit interesting.

Harold Byun:

Then when we look at another framework known as a cloud data protection platform, we really look at different ways where we are focused more data-centrically on what is being put in the cloud. There's a lot of mechanisms available for this. There's traditional encryption at rest. There's access monitoring. There are de-identification and data privacy solutions. There are methods to secure the migration or lift and shift to cloud.

Harold Byun:

What we're seeing emerge, as well, is a lot more adaptive data access controls. These are things where we're really starting to look at how people are accessing which data values and should be getting title to those values. Then there's a fair amount of emerging technology in what we would call the advanced encryption or secure computation space. These are things like confidential computing or enclave computing technologies. Secure multi-party compute is another methodology, and homomorphic encryption, which is still this far-off holy grail, still available as one of these advanced modes.

Harold Byun:

Within these frameworks, we look at how people can potentially better secure the actual data that they're putting in the cloud. Then what we're going to walk into is what are some common gaps beyond what's available? This is another framework that people have been talking about a lot more in recent years, so hold your own key or HYOK, BYOK, bring your own key, using this for SaaS and different cloud infrastructure models and how you can actually support this in your environment or have your SaaS provider support you, as a customer, in terms of utilizing your own key material to encrypt your data as it is living in somebody else's cloud or in somebody else's VPC.

Harold Byun:

As data becomes more distributed, we're starting to see a larger trend around how can the infrastructure and the services offering really enable an overlay for these customers in multi-party scenarios or multi-organizational or multi-tenant scenarios where data is spanning multiple organizations across your information supply chain, and so how can you better support that while still holding on to your own key?

Harold Byun:

Again, encourage you to ask questions as we continue to go along.

Harold Byun:

This is a view of a data security governance framework from Gartner. I'm not going to walk through any of this in depth, but I think the main takeaway here is, obviously, having some kind of methodology to look at what your desired posture end state is, looking at potential data assets that you're looking to protect. Probably, I think, the piece that we're going to focus on in this next section is really the bottom, which is really what are the threats and how is the data exposed? How is the data secured? What are gaps in the methodology and the way people are thinking about these types of problems?

Harold Byun:

When we think about these types of problems, there is a fundamental question that we're going to get to as we cover some of these key trends in security gaps. I'm sure many of you are familiar with what's

on this slide, which is that, obviously, data breaches continue unhindered. There's over a billion records that have been leaked from cloud storage in recent years, and there's a large trend as data, again, continues to be distributed across multiple organizations, vendors and the information supply chain where third parties have introduced an excessive amount of risk in terms of data sharing and data access.

Harold Byun:

Within this, this is just a survey. This was put together by the 451 Research group. I think they're actually associated with Standard & Poor's now, but really asking the question in the survey of what the biggest data management and analytics challenges were for organizations in data security and data privacy, really covering close to the top of concerns of executives looking at data management and analytics challenges.

Harold Byun:

Then, obviously, within the backdrop of this, there's a whole more stringent regulatory environment in terms of how people are actually regulating data privacy, and the rights for consumers, and the right to be forgotten, the right for people to control their records and have a right to data revocation. That has massive financial implications.

Harold Byun:

We have a number of data privacy resources that are available to you, as well, on our website. There's also, obviously, attachments in this webinar. You're free to download any of that material. If you go to baffle.io/privacy or /dfp, there's a number of different materials there for data privacy research as well.

Harold Byun:

Within that regard, you really look at what is happening within the backdrop of some of these privacy regulations that have been passed more recently. Walmart was very recently sued just days after the California Consumer Privacy Act took effect. It actually was passed or took effect on January 1st of this year. Then there was a six-month, I guess, kind of free-rein for vendors to get their house in order before they were actually going to start enforcing it and then... That was July 1st. Then, once July 1st hit, within days, Walmart was [inaudible 00:14:08] for CCPA, so fairly significant there.

Harold Byun:

More recently, Capital One was also fined roughly $80 million as well as getting additional compliance oversight for their AWS breach that they had. I think it was earlier this year or late last year. There's definitely things from a financial implication that have impact, potentially, on your business and, ultimately, the revenues that your companies are trying to drive.

Harold Byun:

When we look at other key trends in terms of what is really driving this cloud adoption, it comes as no surprise it's continuing to be on the rise. Just some basic statistics, so by the end of 2024, 75% of organizations are going to be looking at leveraging AI...

Harold Byun:

Okay, I should be back again. Apologies again.

Harold Byun:

Then, obviously, other additional stats. Hopefully, you're able to read this. 35% of organizations are going to be using data as a service at some level. We're also finding a number of customers that are just running out of space given the data growth that they're seeing, and so lots of challenges on that front. We're just seeing a continual move to migrating more and more data. I think the more recent stat that I saw is a 94% growth in data warehousing in cloud environments. Obviously, a lot of that may be driven by the Snowflake factor, but a significant move to cloud-based analytics.

Harold Byun:

This is just a diagram to show what we've been seeing, as well, for a lot of organizations that are moving from on-premise models into cloud. They are leveraging a cloud data lake and some kind of storage mechanism. This is obviously moving to S3 to look at using that as a more economical data lake and then using a bunch of ML and AI services and forecasting services as well as data warehousing services that may be available.

Harold Byun:

Similarly, Microsoft Azure has a similar scenario where people are moving to Azure Data Lake Services and Azure Blob Storage as a method to also drive that same type of analytics model going forward. Google Compute is no different in that regard.

Harold Byun:

When we look at that overall backdrop and how people are moving more and more data to cloud, it becomes even more important, within the context of regulation and overall data security, to find better methods to actually, A, understand what the gaps are and also find ways to improve the security model around data that you're placing in that cloud. That is ultimately your responsibility. In many cases, it may be your customers' data as well.

Harold Byun:

Really, what we're looking at is this fundamental question of how data is accessed and secured. If you look at the traditional model and gaps of data being accessed, one of the bigger challenges that we saw and one thing that obviously keeps a lot of people up at night is who is the user who's accessing this information? Are they good? Are they malicious? Are they a compromised user where the credential's been stolen and they have a large blast radius in terms of the ability of resources that they can access or the number of resources they can access, rather?

Harold Byun:

What is the profile of the application that's accessing data? Is it a known application? Is it a good application? Is it a known process? Is it a morphed malicious process? Is it a web shell console like the Equifax scenario where somebody was able to leverage Apache Struts vulnerabilities and install a web shell console? What is actually the footprint of the code that is accessing the data? Then, on the back end, what is the data request, and what data is being exposed and sent back within this model?

Harold Byun:

This access model has actually continued to evolve. This is more of a traditional application structure. What we see here is that, when we introduce this notion of third parties, the notion of microservices or

serverless code functions or API gateways and we put the database in the cloud, the footprint of good users, and known users, and compromised users and the footprint of good code, and known code, and what code is actually known and profiled appropriately and secured, as well as where the database is living and operating and who has access to that database, has just blown up from an access-point methodology standpoint.

Harold Byun:

It really becomes an access channel, a modern-day access channel that needs to be better secured. This is where we think... gap in understanding what the existing security controls do for you versus what is actually locking down the data and what you can do to leverage things like adaptive data access controls to better secure that information.

Harold Byun:

When we look at one common misunderstanding that we hear, time and time again, is, "Well, as long as I have encryption at rest." The auditors will tell this to you too. Your auditors are probably all telling you, "You got to have encryption at rest. Everybody's got to have encryption at rest enabled for the sensitive data." The fundamental challenge here is that encryption at rest doesn't do anything to protect your data in the cloud. It does nothing to protect you against a modern-day attack. When you encrypt data at rest, all the logs are in the clear, any of the data in the instance in memory is in the clear. The actual data is in the clear, and when it's accessed by any privileged user or any attacker moving laterally in your environment, they get the data in the clear.

Harold Byun:

Then the fundamental question then becomes what has encryption at rest even bought you? The answer is, in the modern-day world, it has bought you nothing. It is protecting physical disks and data centers. That is what we jokingly call the Tom Cruise threat model where Tom Cruise is in Mission Impossible and breaks into the data center, crawls through the HVAC tunnels, and drops in from the ceiling and steals the hard drives from the data center or gets everything onto a USB stick and walks out. That is not how people are stealing data today. They're coming in over the wire.

Harold Byun:

That's why you see this prevalence of this over-hyped zero-trust network model. It's why you see a lot of people talking about what are people doing to contain attacks laterally or constrain attackers who are known in the environment and quarantine them so that they can't move laterally or isolating and remediating them from the environment? It's fundamentally because, once they have access to the network or assuming that they do have access to the network, then you have to understand that these data-at-rest solutions are not really protecting you at all. Encryption at rest, checking that box, checking it on the S3 bucket, doing things at the tablespace or transparent data encryption layer do nothing to protect you against a modern-day attack.

Harold Byun:

What is fundamentally different with data-centric protection is that, as you can see on this slide versus the prior slide where the data was in the clear, anybody getting access to an environment laterally is going to get de-identified data. That means the logs are protected or tokenized or de-identified. It means that you get safe harbor if the data is somehow leaked in this form. You get a much better or a tighter data-centric control model versus checking a box, which is relatively easy to do but really doesn't

buy you anything. It may buy you a compliance checkbox for the time being, but I think that those days are short-lived as well.

Harold Byun:

This is a notion of a different... It's just a comparison, the bottom being cleartext data, which is what you would get with at-rest or object-level encryption, again, just to show you some of the differences versus what you might get using a data-centric encryption or tokenization or de-identification method. This is where you can actually protect data values in a datastore, in an S3 object, in a data pipeline storage model where you can actually de-identify selected data to ensure whether that be PCI compliance or HIPAA compliance and still drive a lot of the data warehousing and AI analytics that organizations want to enable.

Harold Byun:

The classic example of this, I was on a security thread the other day where somebody was saying, "You should have no reason to have data that is dormant in the cloud. You should just get rid of all of that data." While that's great for the security person, that's a nightmare for the data scientist in today's world. The data scientists and the data science groups want as much data as they possibly can get.

Harold Byun:

I think that what we're seeing the overall market move to and what we're seeing a lot of businesses move to is much more of a de-identified data set that still allows people to identify patterns, still allows people to perform aggregate analysis on the data, still allows organizations to share data across multiple parties in a manner where data privacy is not violated, where if data was leaked, you wouldn't have the type of exposure that we're seeing today, and yet it still doesn't put the brakes on the business.

Harold Byun:

I believe that there's a happy middle ground for a lot of organizations to embrace a model where security can live with the data privacy model...

Harold Byun:

Sorry about that again. I believe that we are finding a place where organizations and security can find a happy middle ground and really work to enable the business and also accelerate adoption of cloud given the flexibility that it's going to drive for the business. Let's cover some of those methods now as well.

Harold Byun:

I do think that there is a fundamental question here where you're really asking this. We've heard this in many scenarios now where it's really one of who can see what data and under what conditions, so effectively, who are you, what data class or data type are you asking for, and should you be entitled to that data?

Harold Byun:

There's a number of examples across a lot of different security verticals, per se, or security technologies where this is applicable. Effectively, who is really an identity question or relates to IAM controls. What conditions is really talking about context and attribution that you have around the user if you're doing

any type of user risk scoring or user behavioral analysis. What is the overall posture of that user? What is their location, if you have that awareness? Are they MFA-enabled, things like that?

Harold Byun:

Other questions that come into play here are what data, so what is the data type, what is the classification, frequency of access, the amount of data being accessed? In an HR application for, let's say, a governmental personnel management group, should somebody be able to enumerate 22 million, 21, 22 million records? Is that a valid use case? If that's not a valid use case, then what can you do, from an adaptive security model, to prevent that type of enumeration from occurring?

Harold Byun:

Because if you really look at the threat model beyond what we were covering earlier in terms of how applications and good or bad users are accessing data, the other thing that is known, as it relates to any of the zero-trust network methodology, and low-lying attacks, and the overall phases for a kill chain and a breach, you're really looking at an attacker whose performing a lot of discovery before they make the go call and want to actually take data.

Harold Byun:

The go call to exfiltrate data is actually incredibly rapid, and so you really want to look at, well, what types of adaptive controls can also prevent somebody from when they make that decision to exfiltrate as much data as they possibly can as fast as they can? Is that a valid and legitimate request within the environment, and particularly as it relates to things like HR or massive financial record repositories? There's a pretty big question as to whether or not that should be a legitimate access request.

Harold Byun:

What are the current options for cloud data security, and what are we seeing? There's, obviously, the Shared Responsibility Model. I'm not going to cover it again, but just from a reference perspective, again, the infrastructure provider's going to give you the tools. You're responsible for locking it down.

Harold Byun:

When we look at existing infrastructure control methods, there's a number of different solutions and tools that the infrastructure providers have made available to you. This is just this... a lot of lists, but there's a lot that you can do to, obviously, restrict public access. This has been the Achilles heel for S3, in many cases, and many publicized data leakages that have come across. There's the overall ACLs and, within Active Directory, there's similar secure access signatures and things that you can do in terms of roles to lockdown Azure Blob Storage.

Harold Byun:

There's this notion of IAM roles, which is a best practice in terms of how you actually manage instance access or service access across different services in cloud. The Secure Access Signatures Model is kind of an equivalent in the Azure model. Obviously, a ton of investment in terms of overall monitoring and logging and visibility that have been put in place across multiple infrastructure providers.

Harold Byun:

Then there's, obviously, the more traditional encryption at rest. I think I've talked ad nauseam about our thoughts around that. There is a lot that you can do. Yes, it's a best practice, but again, in the modern-day world, it's really not buying you that much from a threat model standpoint.

Harold Byun:

There is an exception around client-side encryption and customer-provided keys. We see a lot more interest, these days, around how to leverage that. The challenge with those methods, and client-side encryption, and customer-provided keys is it often requires significant development and development overhead to enable. Then HTTPS and TLS 1.2 and higher should be a standard for, pretty much, any organization operating today. Then there's a number of things that you can do from a VPC and a private endpoint standpoint.

Harold Byun:

There's a wide, wide range of controls that are available to you. You should be using all of these in combination from an approach, but this still leaves a lot of data exposed even when in this model. The fundamental challenge is really that, even with all of the security solution that's going on and the whole premise of zero-trust even though it is over-hyped, is that, inevitably, people are going to get into the network. They're going to find gaps. There's going to be something misconfigured. Then that model...

Harold Byun:

If you're willing to grant... again, apologies... that these attackers will get into your network, then inevitably, the logical step that you need to take is to find a way to mitigate the risk to your data in a data-centric manner.

Harold Byun:

This is an on-demand access model that is presented, in this case, by Amazon. Microsoft Azure has a similar model. You can use HashiCorp Vault and Consul. It uses a secrets manager or a vault to deliver credentials and connection strings dynamically to microservices and more dynamic application tiers. In this case, we're showing a lambda function. What basically is happening is the credential is requested on demand as well as the connection string to a repository via a secrets manager, and so nobody really even knows how to connect to the backend resource, and so that's something that can be available to you as well as a model.

Harold Byun:

Then, ultimately, we believe that there needs to be more of a data-centric data protection service layer that needs to be put in place. This is really something that is evolving from a market space, but we're seeing a lot of customers moving towards a common security architectural layer known as this data protection service that is a consistent layer of protection that looks at access control, a role-based access control, how to better control the presentation layer, how to easily and consistently de-identify data and manage data across multiple repositories and data stores. That's something that we're seeing that's only become more prevalent across a number of organizations operating in the cloud, because similar to your entire CICD model, if you're in a DevOps mode, and how you're rolling out and deploying environments, this becomes a consistent layer of data protection.

Harold Byun:

These are other models that we see with data protection services, so being able to drive a consistent layer that autoscales with the application environment wherever that's being accessed, so something that can either sit behind a virtual IP in a load-balancing environment.

Harold Byun:

Another architecture model with this is, obviously, Kubernetes Pods running in a sidecar mode. People are using containers like Fargate in cloud. These are all models where the DPS components basically can just run as stateless components and spin up and down on demand as well as driving a high availability or HA model.

Harold Byun:

Then when we look at this as it's applied to principles like data pipelining, and this is just a sample data pipeline for how people might be processing data using an S3 bucket, millions and millions or billions of records sitting in an S3 bucket that may be sitting there in the clear and then being processed using an AI or a modeling solution like SageMaker and then exposed for consumption and analytics, it can leave a lot of data exposed.

Harold Byun:

An alternate approach is to use this type of data protection service to de-identify the data and then re-identify the data. In this case, moving left to right, we are de-identifying the data as it comes from an on-premise solution. It hits an S3 bucket in a de-identified state and then, as analytics or warehousing solutions are looking to access that, we can selectively re-identify the data, again, using a more of an adaptive access model.

Harold Byun:

This is definitely something, if you are concerned about data living in the cloud and you want to mitigate some risk, that is a potential tool in your arsenal for better managing some of that data as it goes to cloud. It can potentially unlock a lot of concerns around what the business is putting in the cloud and how you can better ensure that it's actually protected.

Harold Byun:

This is another model, an advanced security model around a secure data sharing. There are other methods [inaudible 00:35:17]. I talked about secure computation, and so there's a lot of emerging work around how to leverage this across multiple parties and what can be done to share data across multiple parties while using the cloud as that vehicle and not revealing the underlying private values to parties that shouldn't be authorized to it.

Harold Byun:

There's a bunch of information on our website around this. It is an emerging space, something particularly of interest for healthcare providers, for example, who might want to share data across multiple hospitals because maybe there's, oh, I don't know, a virus or a disease that may be rapidly being transmitted, and it might be good to accelerate calculations on how many people may be afflicted with such a thing in different regions. Those are scenarios where this type of solution could be potentially very highly valuable. Fraud and threat intelligence sharing are other scenarios where this type of access model can apply.

Harold Byun:

You will get these slides in post. These are just an overview of different methodologies for de-identification and data-centric protection mechanisms that may be available to you with definitions. I'm not going to read them ad nauseam.

Harold Byun:

Lastly, I'll give you a quick walk-through. If you are interested in seeing the live demo, I apologize. As you can probably gather, my connection pretty much sucks today. If you search Baffle Tokenize on BrightTalk, skip to the 20-minute mark. You'll see a live demo of multiple RDS instances with [AWS KMS 00:36:54] in about 10 minutes or less. It's just to show you how you could set something like this up.

Harold Byun:

I'm going to give you a quick walk-through. What we do is we provide just the vendor pitch here, apologize for that, but we provide a data protection service layer that's invisible to the application and to APIs and microservices and serverless code functions. They simply do not know that we're there, but we perform this de-identification, re-identification function on the fly in conjunction with [CloudDB PaaS 00:37:24] and clouds.

Harold Byun:

The way we do that is we basically are just an inline reverse proxy layer that is, effectively, invisible to the data storer, and we can encrypt and decrypt on the fly. This is an example of us interacting with Database Migration Services from AWS. We do the same thing with Microsoft Azure where data can be pipelined on the the fly to the cloud for a lift and shift, and we can de-identify it on the fly.

Harold Byun:

This is a quick walk-through, more or less what you would see in the demo. This is an example of AWS KMS that was set up, and we use an envelope encryption model with a master key or CMK and encrypted DEKs or data encryption keys.

Harold Byun:

This is just some of the configuration stuff. We support multiple key storers, cloud HSMs, other HSMs, [inaudible 00:38:14], on-premise stuff, HashiCorp. We've got an integration with HashiCorp. We also support Secrets Manager, Azure Key Vault, a number of different solutions.

Harold Byun:

This is how you would actually put in a data protection policy where you can select things at a data-centric manner for what you want to actually protect.

Harold Byun:

This is an example of a migration where we've actually ended up encrypting that solution. I think I might have skipped a slide here.

Harold Byun:

That is the quick walk-through. I don't want to spend a lot of time of the vendor pitch, but more importantly, really focusing on what we think are some key takeaways here. That is, ultimately, you can look to leverage the flexibility of the cloud and not compromise on data privacy. There are a number of risks, obviously, putting your data in the cloud, but there's a broad range of solutions available to help you lock it down, but you do need to understand the [inaudible 00:39:16] for that model and some of the key gaps and [inaudible 00:39:18].

Harold Byun:

That's one of the key takeaways that we would like you to understand is that encryption at rest, you may have had that drilled into your head. I mean there's more on the CISSP exam around different encryption algorithms than there is on the threat model. I find that personally disappointing, but the reality is that there's a big push for everybody to do encryption at rest and not a lot of people talking about how, quite frankly, encryption at rest isn't doing anything for you. Hopefully, you can come to that realization and understand that it isn't really going to buy you a lot in terms of hack or breach protection or data leak protection in today's world.

Harold Byun:

Then, lastly, you can really look at different operational models that are going to make you a lot more efficient from a DevOps perspective. Quite frankly, your developers may actually like you more for it. That's some of the takeaway from today.

Harold Byun:

I'll open it up. More resources, obviously, available to you as well as all the attachments that we have. We're happy to talk with you one on one, as well, if you want to reach out. Promise we won't badger you to death if you have any questions or want a technical consultation one on one.

Harold Byun:

With that, let me open it up for Q&A. Hopefully, I will be able to stay alive and not get disconnected yet again. I'll open it up for questions, at this point, so feel free to chat away on that.

Harold Byun:

I'll take the first one here, "If you're using these types of de-identification methods, what happens for disaster recovery and backup and restoration?" It largely should be seamless in these types of models. Ultimately, whenever you're de-identifying or tokenizing or encrypting data, it's no different than when we used to... not to show my age here, but when you used to backup to tape. If you backed up to tape and you encrypted it and you sent it off to Iron Mountain, if you didn't have the encryption keys when you got the tapes back in your DR center, you weren't going to be able to restore. That's the same principle that applies to things like disaster recovery or HA failover or replication in these types of models. If you don't have the re-identification capabilities enabled for you, then you're not going to be able to, obviously, re-identify the data. It defeats the purpose there.

Harold Byun:

"What types of performance overhead are you seeing in use with these solutions?" Yeah, that's always a tricky question, I think. Well, not a tricky question, but I think that the way we approach this is, look, mileage is going to vary. I think that, depending on the solution, you're going to see differences in

behavior. We've had a lot of success based on the performance that we've been able to enable. I can say that for us, but with very, very minimal overhead, we've been measured at scale like one to two milliseconds overhead and the rest at, pretty much, close to wire speed. But again, your mileage is always going to vary. It depends on the application load. It depends on the result set. It depends on the number of concurrent connections, but we've been highly optimized in that space.

Harold Byun:

Another question here, "What key stores do you integrate with?" Pretty much anything that supports KMIP, which is an industry standard protocol for HSMs or hardware security modules. They pretty much all standardize on a protocol known as PKCS 11, which we also support. Then, for the cloud key managers and secret managers, it's largely REST API calls.

Harold Byun:

"Why should a customer choose Baffle instead of a cloud service provider data protection tool?" It's a good question. I think a lot of it depends on the data classification. A lot of it depends on the overall security posture for an organization. There are, for example, a lot of customers that we work with that don't use the cloud provider database migration services. I'm not making a judgment here. Right or wrong, they don't trust the fact that the cloud vendor has given you this database migration service by virtue of the fact that this third party, which is also going to store your data, gave you the service to help you move the data into their service is already a level of distrust for some of these organizations.

Harold Byun:

We can talk about paranoia all day long and whether or not that's appropriate or non-appropriate. It really depends on who you talk to. There have been instances where some cloud vendors have actually seen people's data, and so you could say how legitimate is that? I can't get into specifics in terms of how paranoid people are and justified they are, but those are scenarios.

Harold Byun:

Key ownership is a big one going back to hold your own key and bring your own key. We can support any key manager, and we basically consume that in using that, and so when a customer disables our master key, they effectively have disabled their data in the cloud, and so that gives you that right of revocation.

Harold Byun:

Other mechanisms that we can enable... and there's some scenarios like... so we do record-level encryption where we can segment data at the record level, which gives us a data entitlements mechanism. We can support encrypting S3 objects with different keys for different parties. Personally having configured S3 bucket [inaudible 00:44:53], it is not the easiest thing. That's, quite frankly, why I think that a lot of S3 buckets are public, and so that's another reason where you can use, again, a third-party encryption model to better protect data access. Hopefully, that answers some of your questions, but again, it depends on your overall security posture and model.

Harold Byun:

"What kind of deployments do you have for encrypting in the cloud?" We work with GCP, AWS, and Microsoft. We're partners with all of them. We probably have the deepest integrations with AWS. That has been the bulk of the customer base, but we have a good footprint with Microsoft Azure as well. It

basically mimics market share, and so what we do is we support AWS RDS, and S3, and Azure Blob Storage from a de-identification encryption standpoint.

Harold Byun:

Classic one, you want to save on Oracle licenses and you want to go to AWS [Postgres 00:45:51] RDS but you have concerns over security, we can lift and shift the data and de-identify it on the fly, and your application and your developers won't know the difference. That's one classic one.

Harold Byun:

Then the other one is securing the data pipeline for S3 cloud data lakes where people want to store and warehouse a ton of data in S3, but obviously, if that data got exposed, you basically would have hundreds of millions of records exposed. We can de-identify CSV and parquet file format structures at the field level so that those can be consumed in something like an [Athena 00:46:25] or Redshift environment for analytics purposes or Snowflake and then be selectively re-identified as well. Those are the types of deployments that we can support.

Harold Byun:

I think that that might be it for the questions. If there is other questions, please don't hesitate to reach out to us. My email's here. You can always reach us at info@baffle.io. I hope some of this information was useful for you today and really appreciate your time. Apologize for the connectivity issues. Hopefully, you were able to stick with it and we didn't lose too much of the flow. Thank you very much. Have a great day.