Harold Byun:

Hi. Good morning, good afternoon, good evening and welcome to today's weekly webinar actually on tokenizing your data and AWS RDS with AWS KMS. My name's Harold Byun. I head up product management for a company called Baffle. And today we're going to walk you through some different scenarios around data privacy, some key gaps in what we believe our current data protection methods and how people are protecting data in AWS. The bulk of the time is going to be on a live demo where I'm going to walk you through three setups with AWS RDS - Microsoft SQL server, Aurora MySQL and RDS PostgreSQL. We support all flavors of AWS RDS and Aurora. So you can use us to basically enable a simplified de-identification method for AWS. We see a predominant number of deployments with AWS KMS but there's a lot of variety, but anyway, the point being, I'm going to try not to slide where you to death and focus mostly on the configuration and setup of this.

Harold Byun:

If the demo gods are with me, everything should go well, this is going to be a live setup. We'll see how that goes. A quick agenda overview of kind of data privacy and gaps in the data protection model. We're going to cover how we do what we do from a data protection and de-identification standpoint. And time permitting, so usually a little bit pushing the limit depending on how fast things are responding here. We could cover some of the privacy preserving and advanced analytics capabilities that we have. I encourage you to ask questions throughout the presentation so you can use the chat for that. You can always email @info@baffle.io. My personal email's, Harold@baffle.io happy to answer any questions that you all may have as well.

Harold Byun:

So why don't we jump right into it. I'm going to try and get you out of here in about 45 minutes or less, including all the live demo. A little background on me, I've been doing security for close to 30 years. Both kind of from a security architecture and hands-on management perspective early days, incident response, and even some honeypot and e-commerce architecture. I've been on the vendor side in what I would consider data containment technology. So data loss prevention, mobile data containment. I was working at Skyhigh Networks for a while where I was head of product for a number of users as well. So really looking at how people are accessing data and ultimately how it may be leaked. And that's kind of led to a number of years here with Baffle working on developing this data protection services capability.

Harold Byun:

An overview of key data privacy challenges. Again, I'm going to try to gloss through the slides. We'll send these out to you in post so you'll have all the slides. Obviously, there's a number of data breaches that continue. It's estimated that there's over a one billion records we took from cloud data storage at this point in time. There's this notion to adopt cloud to move faster from a DevOps and CICB perspective that is kind of pushing the boundaries and limits on security teams and the ability to that and perform security architecture reviews. And then there's the backdrop of privacy compliance like GDPR or CCPA and a number of other pending data privacy regulations that require [de-identification of key data](https://baffle.io/advanced-data-protection/) as well as a right of replication or right to be forgotten on behalf of the user. And so we look at that, obviously the privacy concerns continue to grow globally.

Harold Byun:

There's probably an estimate of, I don't know, 30 states roughly that are going to be passing their own regulations within the United States. And from our perspective, really, as you see that in conjunction with this broader adoption of [cloud infrastructure](https://baffle.io/solutions/secure-cloud-migrations/), it places a lot more focus and onus of responsibility on you, the master of the data or the owner of the data that you're putting in the cloud. And that's effectively what the shared responsibility model is showing you here. For those of you, I'm sure many of you are already familiar with this, but effectively it's stating that the infrastructure provider is actually responsible for providing you with the infrastructure and the capabilities and controls to secure that infrastructure. They're obviously responsible for securing the data centers and additional capabilities to secure the configuration of the infrastructure that's running, but you are responsible for configuring the security controls and you are responsible for actually securing the data that you're storing within that infrastructure.

Harold Byun:

And so that's kind of the general delineation of the line between customer responsibility versus provider responsibility in that shared responsibility model. Tons have been written on that. This is just a quick overview of Gartner's view landscape of cloud security. CASB is cloud access security brokers. Folks like Skyhigh Networks or Netskope, CSPM clouds, cloud security, posture management. I think RedLock and Evident.io were vendors that were operating in that space to provide automated security control validation of cloud, make sure that buckets aren't wide open that certain security controls are in place from an infrastructure management and then cloud workload protection platform or CWPP is really focusing on kind of microservices and workload protection. So folks like Aqua, Twistlock, StackRox who are focused on this kind of container based security model and vulnerability posture assessment of those workloads.

Harold Byun:

The interesting thing that we always felt that was ironically missing from this landscape was data and data centric security controls and what are you actually doing. I mean, great that you're blocking people from going to cloud or monitoring their cloud usage. Great that you're trying to do you're best to automate lockdown of buckets and things are secured, but inevitably, if you believe in this over-hyped zero trust model but there is evidence around this notion of an assumed breach posture, then you have to grant the fact that attackers inevitably will get to the data and in that type of threat model, the logical step is to protect it in a data centric manner. And so this is something that we feel is fundamentally missing in terms of how people look at the security model, and that's probably part and parcel why breaches continue to occur ad nauseum.

Harold Byun:

Again, you're going to get these slides. I'm not going to go through every single bullet point. Privacy regulations are out there. Encryption, deidentification, tokenization, lot of words thrown around what you can do with the data. The net net of it is if you use one of those methods and you do have a leak or a breach, it does give you some safe harbor as an organization. The fines are continuing to grow. I don't know, for recent news Capital One who was breached in AWS they, I think, took an $80 million fine most recently plus additional regulatory oversight. I'm sure that's something that everybody wants is more regulatory oversight in terms of how to actually run their operations and do their jobs. And then Walmart was recently fined under the CCPA regulation. I forget the dollar amount, but it was pretty extraordinary.

Harold Byun:

So the revenue and financial implications are definitely getting higher. Very quickly I'll cover this throughout the demo. We provide a simplified no-code method that allows you to de-identify encrypt tokenized data in AWS RDS. So if you're looking to move to AWS RDS, if you're looking to migrate data, you're looking to operate there, we've a number of different protection models that basically can do that without modifying application code. And we simplify that and we also integrate seamlessly with AWS KMS and other key management vendors. Quickly going to walk through the threat model and again, I will get to the live demo. Why do breaches occur? There's a multitude of reasons and we just kind of click through all of this. The net net of it is that breaches continue to occur for a ton of reasons.

Harold Byun:

It's not that attackers are necessarily getting better, but inevitably they're going to get into your network via some method. And if they do, then we need to look at how you can best mitigate against that type of threat model. And the reason that we think this occurs is because people traditionally have looked at this as a siloed type of data access pattern. So if you have your data structure and you have an application and you have a user, what are the current gaps? Well, you have good users, bad users, compromise users or unknowns. You've got your application tier, which could have the standard pre-approved embedded application code that you want to run. Somebody could have landed on it and installed a Web Shell console and that would be a malicious application. And so what is the application that's actually making a request on behalf of that user and then how much data are they actually asking for?

Harold Byun:

And if you look at this as kind of an end to end chain of a channel where information and data is provided, it really represents an end to end access model that needs to be secured. And I think part of the challenge is, is that people are looking at ways to secure different components within each of these peers. And that's all very logical and it makes sense, but it often leaves some significant gaps in terms of how people can actually get at the data in the clear at the end of the day. And so if we evolve this to the cloud model, the challenge with this is that you don't have just your own users, you've got third parties that may be accessing data. You've got microservices, pods, and containers that are accessing along with serverless code and API gateways. And ultimately that database is living in the cloud.

Harold Byun:

So it really creates a pretty distributed data environment that becomes even more of a challenge to secure from an agenda and perspective. And so now we have good users, bad users, code running all over the place and your databases in the cloud. It's a more untenable problem and again, represents an opportunity to better secure that channel. So when we look at kind of common methods at the data layer that people are using, the most common methods that people are using are two techniques. One is data at rest, and then the other is a database encryption or referred to as database encryption or a tablespace encryption or a database container encryption, or transparent data encryption or TBE.

Harold Byun:

The challenge with each of these methods, which is what this slide is showing you, is that when you get access to the data tier, you see the data in the clear. And so for those of you who understand what encryption and tokenization does, it's supposed to make the data unreadable, but in this case, you've selected an encryption option for at rest or TDE. And in both of those scenarios, the data is in the clear for anybody who makes it to the database tier. And that's a problem. And so it's not just clear when you access the database, it's in the clear in logging it's in the clear in memory. Is there anybody who gets access to that system basically gets unfettered access to your data in the clear. And so what we're basically suggesting is that these methods of encryption do absolutely nothing to protect you against the modern day hack. We jokingly refer to them as the the Tom Cruise jet model. I was actually just watching Mission Impossible the other day and inevitably he breaks into a data center and drops down from the ceiling and steals the hard drives.

Harold Byun:

And that's great for those types of super advanced attacks, where people are dropping in a strike team to break into data centers. But most commonly that's not what's happening in modern day hacks. People are coming in over the wire and actually stealing data over the wire. And in that model, that is where a data centric method such as one shown on the screen now is a field or record level encryption model that protects the data at the data layer. It encrypts the data in memory, and the data is encrypted in the logs, but seemingly any authorized user from the application tier can still access that data. So those are some of the fundamental differences. These are different data protection modes that we support. I'm not going to cover them ad nauseum just in the interest of time. You can read them.

Harold Byun:

Shortly [field level, encryption, record level encryption](https://baffle.io/database-encryption/) for multi-tenant SaaS providers, multi-tenant environments. We have file level encryption. We can de-identify semi-structured data going into S3. There's an API based model and then we have dynamic data masking as well, and role-based data masking for the access layer. A number of different options for you, again, you'll get the slides. You can read through that at nauseum, it's also on our website. So how do we do what we do? Well, this is kind of the series of data protection services that I was just covering. We basically offer you a portfolio of data protection services and capabilities to better protect and secure your data. Whether that be 3D identification and tokenization or format preserving encryption, whether that be through dynamic data masking, to prevent who can actually read the data. And then we have a core field level and record level encryption method, as well as the data pipeline encryption mechanism that we offer.

Harold Byun:

We have a BYOK model for SaaS providers that are running us in multitenant environments to secure specific tenants data within a multi-tenant model. And then the last two boxes in the lower right are advanced encryption modes where we can enable homomorphic encryption operations at speed. We actually don't use homomorphic encryption. We think that that's kind of a bit of a myth from a performance and delivery perspective, we use a different technique, time permitting, we can cover that. There's also more webinars on that. And then we do this by providing you with an industry standard key virtualization layer that can talk to virtually any key store or manager or HSM to retrieve key material. So you, the customer, own the keys while you encrypt the data as it lives in cloud infrastructure. So that's kind of the net of our offering.

Harold Byun:

How do we do it? Well, let me kind of skip over some of these. We talked about this. This is an example of some of the masking and redaction. I think in the left-hand column, you can see we have a fixed string dynamic data masking where we are replacing a fixed string value. In the right hand column, we are exposing the last four digits and selectively masking the the rest of the prefix.

Harold Byun:

So different options different flexibility in terms of data format. And the way we do this is through the data protection service architecture. And so the main component here is the Baffle shield, which is the orange circle in the middle of the screen. And that is a TCP wire protocol reverse proxy that can run in between the app tier and the database tier. So we're below the driver, we're below the interface layer. So we're transparent to any application call or Lambda function call or API service call. It's basically you just modify the connection string to point to us. So it's an IP address change, which a DNS host name change and all the traffic pipes through us. We've had customers take ECS far gate application tier container layers, and through a single value change, run the entire infrastructure through a Baffle shield or a set of Baffle shields.

Harold Byun:

The application doesn't know the difference. It thinks where the database, but we're performing that encrypt decrypt function seamlessly for AWS RDS Baffle managers, our UI, I'll give you a walk through of that as we go through the live demo very shortly. And the secure multi-party compute is the advanced decryption mode. We're not going to cover that in today's call. Well, we might cover it go into that. But it basically allows for operations on encrypted data without ever decrypting the underlying values. So things like wild card search, which in theory were impossible on encrypted data, things that we can support. I know that sounds crazy, but we can cover it another time. This is a view of a data transformation de-identification pipeline. So where on the left, we have an on-premise database. We can integrate seamlessly with AWS database migration services or a third party data migration tool.

Harold Byun:

We basically pass that data through that same Baffle shield that I was talking about, and it would hit an S3 bucket and a de-identified state. And from there it can be consumed by other third party data analytics solutions like a Redshift or [inaudible 00:17:19] or snowflake. So that's a common data pipeline and migration strategy where people are trying to use cloud data lakes as a storage point, and then use these analytics solutions to analyze the data. We can support that in a de-identified state pretty easily, and again, with no code change. And this is just encryption as a service, another variation where you can call into us via an API, a similar structure.

Harold Byun:

I encourage you to ask any questions again, almost at the demo. I know it's a lot of content to take in. I'm hoping that it's somewhat interesting or relevant kind of the common architecture. So we have the Baffle manager UI, we never store your keys. We just talk to the key manager and establish a key mapping. That key mapping has been applied to the Baffle shield and the Baffle shield uses that to perform encrypt decrypt functions or tokenization functions using the source key from your key store of choice. Encryption modes, I already covered some of this. I'm going to kind of gloss over it. Again, you'll get the slides. This is an example of format preserving encryption, which we can also support. So the bottom is the clear text. The top is encrypted. It's probably a bit of an eye chart, but you can see that credit cards we can randomize credit cards, maintain the length, credit cards can still pass.

Harold Byun:

What's known as a lunch check, which is a standard check or algorithmic check to see it's a valid credit card type. We did social security numbers and agree to dates based on available date range, email there's different options. And we can do this for virtually any string, but if you make out, I think the top on the top version and the second to left-hand column, it's all randomized tax except for the app sign and the period sign. So everything else is randomized, but it still passes any email validation field check. In the right hand, most column, we have encrypted the names and preserve the domains. There's different options in terms of how you would choose to apply something like this. But point being that there is flexibility in the tokenization modes and options and de-identification methods. All we are at the end of the day is a data transformation tool that allows you to do this in VM no code model.

Harold Byun:

We use mist standards for all of our encryption, AES 256 and then for format preserving encryption, there's different mist approved algorithms that are used. Here's a quick glimpse of record level encryption. I think for the bulk of you, it may not apply, but this is what some of our SaaS providers use. So what we provide is an off the shelf, bring your own key service that allows a SaaS provider to segment a multi-tenant environment and encrypt data in that environment with different keys and again, without architecture and code changes. And so the way that we do this is we source multiple keys from different customer key stores. We use those keys to reference different rows or records in a multi-tenant environment and are able to encrypt those and this item on the left, which you can probably barely make out how some asterisks and that's where we've obsessed skated the data or mask the data because the key is not present.

Harold Byun:

And so if the customer kills a key in this model, the data is in the cloud encrypted and can never be reidentified. And so that gives you the right to be forgotten or this right of data revocation. Alternate deployment model, again, just showing you some of the breadth of the solution. One of our key customers where Kiva is a SaaS provider, they do all the financial reporting and SEC filings for publicly traded companies. So they cover roughly 80% of the fortune 500 were deployed globally through North America and European operational spheres in AWS and a multitenant environment. They have over 5 billion data records for these SEC filings that they handle on behalf of these companies. And we are in production with two of the top five largest US banks in that environment. So from a performance and scaling perspective, the core of the technology is a very, very efficient and can scale very easily in global deployments.

Harold Byun:

All right. So that's enough of my slides, more or less. I got like two more slides and then I promise we'll get into the live demo. So hopefully, you get a sense of kind of the different capabilities that we have. I'm going to walk you through a live demo now in ideally less than 20 minutes. Again, encourage you to ask questions as we go through the chat, if you have an interest in that. And then the way that this is going to run is I've got a Baffle manager set up, it's going to talk to AWS KMS. I'm going to set that up on the fly, and we're going to use DBeaver as our application client. It's a SQL client that typically database users or admins use, and we're going to access a set of three RDS databases, Microsoft SQL Postgres, and Aurora MySQL. Let me switch gears here and share my screen.

Harold Byun:

It should come up in just a moment here. I think I'm sharing now. So you should be able to share through this. This is the slide that I was just on. It was a little different in that it has these IPS in case I actually need them. And so this is the demo that we're going to run through. Not this, that's after the demo. Let me go into my environment here. So this is the Baffle manager that I was talking about, just going to log in.

Harold Byun:

I've obviously set up the initial config on this, just in the interest of time. There's documentation on our website and this getting started, where I'll send that to you along with a copy of the docs. You can probably set this up while the security groups are set up within ... You gotta read the docs, but within an hour, roughly, I mean, if you're super fast, you can do it. You can do it in 15 minutes. So what I've got set up, I've got a couple of databases that I've already connected to, and I've got a couple of Baffle shields that I've already set up, but I'm going to walk you through all of that. And then I've got a local key store, but I'm going to add the AWS KMS key store into the way that this is set up is we basically have all these different key stores that we can support.

Harold Byun:

And I'm basically going to point to AWS KMS and it choose a region. And I select this S3 bucket where we store encrypted DEKs. It's just going to specify to us. I think that the bucket is this. I hope it is. We'll see. And I'm going to basically create this key store. So let's see if it succeeds. I may actually have to go over this three and see what I actually created, it did. Okay. So the way that we actually did that is we are communicating to AWS KMS via the rest API APIs and using an IAM role that has refined permissions. There's some leased, privileged documentation that we can make available for you, but it's using standard IBM bowls, which is best practice for AWS KMS. I have that set up and what I'm going to do is on the fly, I am going to also set up this Postgres database. And so I'm going to establish a connection to this Postgres database. I do need to pluck the URL for that. Fortunately, don't remember that.

Harold Byun:

And then I'm going to enable an SSL connection, and this is Amazon or AWS root CA for RDS. I downloaded that from my website and I'm going to set this up and of course my credentials are incorrect. Okay, there we go. So we've established this connection to the database and I've set up a key store with AWS KMS. I'm going to add this third shield, just so you guys get a flavor of it. In this option, we have a EC2 instance. It could be a Kubernetes pod or a Docker image, and it doesn't really matter to us what the deployment or the infrastructure vehicle is. I'm going to use SSL. So we run SSL end to end. I have an SSH key that I was using for these other shields, and then I'm going to connect into the shield and we should spin up and ideally be connected.

Harold Byun:

All right. So now in my third shield, I typed it a little differently than the others, and you'll see it's idle, the others are running because I've already enrolled this application. And so these applications, I've set up enrollment. So I'm going to kick those off right now, just so you all get a sense of what we're doing and then I'll come back and do ... Actually, you know what I'm going to do, Postgres first anyway, because Postgres has, I think the most number of records or data. So you'll see here, I'm setting up this mapping. Now I'm choosing this data store Postgres, I'm picking AWS KMS, and we have a column method and a record level method of encryption. I'm going to enroll this application. And so what's going to happen is when I go in here, we have the option to walk through encryption options for this data structure now. So I've got this environment set up or I've set up the scaffold demo one table, and it's not the most exciting data, but just to kind of give you a flavor of what it looks like in this Baffled demo data, sorry, it's this environment.

Harold Byun:

I have this bunch of tables, but I've just Baffle demo one data, and it's not huge, but it's a little over a million records. And if I do a query on this you'll see that it's not the most exciting data, it represents an IoT environment. This is an IoT reporting environment and somebody storing that data in RDS. And so if I do a direct connection right now, this data is in the clear. And so what I'm going to do is I'm going to kick this off where I'm going to pick a couple of these columns that I think are sensitive, potentially. And so these columns are for, again, not IoT reporting. I'm going to add a key here to select a different key from KMS and I'm going to hit next. And I get a confirmation screen. Is this the stuff that you actually want to encrypt? I'm going to say, yes. I'm going to enable parallel processing to try and make it go a little faster, but in order to save money, I'm running off of a really dinky Baffle shield.

Harold Byun:

But I'm going to kick that job off just because it'll give you a flavor of us doing roughly a million records. And I think that's going to take probably three minutes or something like that to run. But while we're doing that, I'll show you the other environments. And so the other environments that we're working with, so this is a Microsoft SQL environment. You can see that I've got a database called Baffle demo with a table called superstore. And again, if I look at that data, it's potentially sensitive data. If I go to AWS Aurora, let me see this environment, superstore.

Harold Byun:

I have a superstore table and very similar dataset, but this is obviously AWS Aurora. And what I'm going to do in these environments is select some columns for encryption as well. So let's go through. You'll see that Postgres is kind of in motion. I'm going to do Microsoft SQL now. So I select this Baffle demo environment and select superstore. And then what I'm going to do is pick something like a customer ID and customer name in this environment maybe a order ID. And so what I'll do here is those three columns and hit next. This is for Microsoft SQL and hit encrypt. We'll kick off that job. And then I'll go to Aurora MySQL, which again, we're already set up, we're picking superstore not Haroldstore and I'll pick different ones. I'll do city and product name and product ID. And I'll kick that off. I have a database site item here.

Harold Byun:

I guess the demo gods might be on my side. Let's see. Okay. I do have to cramp now, so this should actually be fine. And when I kick this off we'll be able to run encryption on these columns. So that's actually how. And you see the Microsoft SQL one already completed, the Postgres one's still in motion. For those of you that will have questions about performance, we roughly run at the speed of DMS. I'm using really undersized instances just for this demo. And then on kind of an in-place migration, we're running on a well sized, properly sized instance. We'll see roughly throughput over around 60 to 70 million rows per hour. So it depends on your mileage is always going to vary, but that's kind of these estimates. So if I go back to my Microsoft SQL server, I can basically rerun the same select and we should see that customer ID, customer name.

Harold Byun:

And I think I picked order ID are encrypted. And so this is where, again, we've seamlessly encrypted that data in an RDS environment with AWS KMS. And then so what I'll do here is I will go through this Baffle shield. And this Baffle shield is basically just going to the Baffle shield instance on a different port. That's the connection string change. And when I run this we decrypt it. So it's the same dataset but it's being decrypted by us using that Baffle shield transparent data protection layer. So let's go back here, you'll see that AWS, MySQL is also completed now. So if I go to Aurora and I do the same select, you'll see that Citi. And what else did I select? Product ID and product name are encrypted. And if I go to this Aurora shield ... And if I set this one up, but we'll find out

Harold Byun:

I'm not connected here, bare with me. [inaudible 00:33:15]. And I think that this is shield two. Okay so now I'm connected. And if I look at this, we should be in a flat door address and so now we run this. And you'll see I'm connected now. So that is the connection string change. Just for those of you basically just connecting to whatever instance is running up Baffle shield. And when I rerun this query, we decrypt that same dataset. That's kind of that item. And then lastly, if I go to Postgres, which is now completed on 1,000,000 records, and I go direct on Postgres, I can do this count where we have the same record count, a little over a million run the same query. And you'll see that these values are then encrypted in the Postgres environment. If I go direct. And then if I go to baffle shield and connect. And this environment, open this.

Harold Byun:

And we're performing to decrypt. That was running encryption or tokenization jobs again AWS said yes. Using integration with AWS KMS on three database platforms with no code changes and about 13 minutes in terms of kind of setting that up. Aurora gives you some flavor of it. If I go to the key store and AWS KMS you'll see that we're referencing this master key item, which is the CMK that we use for encrypting the DEKs, So that is actually generated using KMS service and then it's used to encrypt the DEKs that we're using. That's kind of the demo, quick down and dirty. I mean, just going kind of go back to just some architecture slides and then again, encourage you to ask any questions off of here. So in terms of high availability, we often get asked about that. We can be deployed and you see two and stints behind the load balancer basically running a bit which can be utilized with auto scaling groups, if you want to kind of minimize your costs.

Harold Byun:

And if you've got bursty traffic, can be handled that way. This is an example of a reader writer or a master replica set up behind an EOB that can also be utilized with EC2. This is an example with Kubernetes pods, where again, you can deploy a [inaudible 00:37:17] so we've got customers that will run us in a pool of three to five pods and then auto scale to 25 or 30 based on kind of load requirements. Different models there, and there's other customers that are using us the DCS far gate. So a lot of different options from a deployment perspective and architecture on establishing this data protection service layer and then using that to de-identify the data and selectively reidentify for RDS. Let me pause there and go back to kind of the environment and see ... Sorry if the slide change here, let me catch up and see if other people have any questions before we go into the privacy preserving analytics.

Harold Byun:

What key managers do you integrate with? There's a number of AWS KMS is the one we've used in today's demo. Gemalto KeySecure, which is now owned by Thales is an option. There's SafeNet [inaudible 00:38:22] HSMs that are available. We split cloud HSM as your key vault, if you're an Azure customer. Hashicorp AWS secrets manager. So there's a lot of different integration points that we have there and we use industry standard protocols. So PKCS 11 is the protocol for HSMs and keymap is the protocol for traditional key management stores. So we support both of those as well as any rest API integration to a third party key manager or key store, which is what we use for things like Hashicorp. Another question, what kind of performance overhead is there?

Harold Byun:

Sorry to fumble here. We basically, I'd love to say that there's no performance impact. We have a very minimal performance impact. Again, your mileage is always going to vary based on network latency and the application results that that is returned. Those are the kind of the two key variables that we've been highly optimized. And so in transactional environments, we are seeing a one millisecond to two millisecond overhead. That's what we've been measured at. There are customers that have run API level access tests from an application to you're in environments that exceed a billion records and they claim no overhead. I find that hard to believe. There's gotta be some overhead, but it's not noticeable. So it's incredibly minimized by and large. The reality there is no free lunch and so there is always encrypt decrypt overhead, but we've heavily optimized kind of how we handle traffic and the bulk of the traffic passes through us at wire speed.

Harold Byun:

So hopefully that answers your question on that. Lastly, just kind of a quick glimpse into the privacy preserving analytics, what is it? It's an advanced mode of analytics that allows you to operate on a good two data, perform things like aggregate or shared data analysis on encrypted data to leverage intelligence from the data, whether that be for BI reporting or AI or ML running. And so what we can do is de-identify the data and maintain it in a de-identified state and still allow these operations to occur. These are some architectural models where we see people trying to leverage us. There's a hospital vendor rather that has a presence in over 2200 hospitals that is using us as a patient database service protector for these multiple parties.

Harold Byun:

And so this is a classic example where you've got in this case, it's not vendor one or vendor two, it's hospital one, hospital to and hospital and accessing a data structure or queering against it with the goal of better sharing patient data and information more securely. That's kind of one model that we see from an architectural perspective. There's other webinars. There's a lot of material on our website around how we do this. Threat intelligence and fraud detection or types of cases is more along the lines of the healthcare data sharing use case that I was just explaining. And this just walks you through it. We encrypt the data from one party. The keys are never on the database tier, the second party encrypts similar data or the same data using different and then encrypts differently but we're still allowing this aggregate analytics to occur.

Harold Byun:

There's white papers on this, happy to make those available to you. In closing, what can we do? We can easily enable de-identification and field level encryption and tokenization for your data in the cloud. We integrate with a number of key management solutions to make it easy for you and we basically have offloaded a lot of the complexity that's required in order to implement this type of data centric protection. So we can help organizations move more workload to cloud faster and you don't need to burn application development cycles or a ton of security review time when you deploy this type of model. Last question, is there anything else that comes across what is the method for the advanced encryption mode? This is the last, I guess question has come across. So we use a technique that is known as secure multi-party compete or NPC or SMTC.

Harold Byun:

It basically is a method that distributes the operation across a distributed set of computational nodes that never allows the nodes to see the original values of the data or even the encrypted values of the data. And so it sounds like it's impossible, but effectively what happens is the encrypted values are inferred by these nodes using this protocol. And without ever seeing that data, they're able to run operations and calculations and ordering and mathematical operations on that dataset return the results in encrypted form and the data tier in the application never know the difference. I know it sounds a little bit like mad science, but we've proven that out at scale and there's a number of other materials available on that particular capability. In closing, I hope that this was helpful for you.

Harold Byun:

If you have an interest, we'll send out the doc to all of you. We'll send out the documentation, the links for the marketplace. You can go to AWS marketplace and search for Baffle. There's only one entry there for Baffle and you can download our marketplace image, the documentation will give you a quick walk through on how to basically do what I just did in the 10 or 15 minutes that we had on the demo. And hopefully that'll be helpful for you. And we're available to answer any questions or if you need assistance in the setup, we're happy to walk you through it really quickly. Again, thank you for your time. Hope this was somewhat valuable for you. And if you, again, need any help or have other questions, feel free to reach out to us info with Baffle.io and Harold@Baffle.io. And thank you very much for your time. Take care. Bye.