Harold Byun:

Hi, and welcome to today's webinar from Baffle. We still have a number of registrant's getting signed in, so we're going to get everybody settled and then we'll kick off the webinar, cover the agenda, and also cover a brief demo today, as well. So thank you for your patience. We'll be back with you in just a moment.

Harold Byun:

Okay. Well, good morning. Good afternoon, and thank you for joining today's broadcast: Preventing Data Leakage. Let's get started. Today's broadcast is being recorded and all attendees are in listen-only mode. During this broadcast, we will be taking questions, which you can post in the question and answer module in the communication panel. And we'll also follow up with any emails if we don't get to your question in today's broadcast. You'll also get a link to the recording, as well as the ability to download our advanced data protection data sheets, and additional materials on our website.

Harold Byun:

So with that, we'd like to kick off today's webinar. Thanks for taking the time to join us today. My name's Harold Byun, and Baffle has been at the forefront of simplifying data protection and data encryption, particularly for Cloud environments. And one of the areas where adopting the solution is really to secure a lot of their Test/Dev environments, or staging environments in Cloud, so that they can obviously assume a faster release cycle and more of a DevOps posture and embracing the nimbleness and flexibility of the Cloud, but still with an eye to securing their data. I'm sure, as many of you are familiar and as we'll cover in today's webinar, there are a number of significant gaps related to data as it resides in the Cloud, as evidenced by the latest data leakage scenario that you may read about. And, there's been a multitude of them with some pretty marquee names of folks who have had several million records exposed.

Harold Byun:

So, today's agenda, we're going to cover some common gaps in the Cloud data protection model that we're seeing in the market, as well as some key trends that are ultimately affecting, not only the migration or the move to Cloud, but what is leading to a greater consternation and focus on how to best secure that data in the Cloud, and still be responsive to the business in an agile and accelerated fashion. We'll talk about how you can support Test/Dev and staging spin ups in the Cloud, the process around that, that we're seeing. People adopt from a Test/Dev environment spin up and the downstream effects of data pipelining, and what that does to ultimately lead to potential data exposures, or, again, data leaks in the Cloud.

Harold Byun:

We'll review our single-click secure Cloud migration capabilities. The way we facilitate a faster and simplified spin up and integrating data protection into the migration process. We'll talk a little bit about Cloud Key Management options and then open up for Q and A. This should be a relatively brief webinar, but it should give you a sense of some of the capability sets and how some people are adopting best practices around the overall Cloud data protection.

Harold Byun:

When we look at some key trends impacting Cloud data security in general, there, obviously, are many more, these are some of the major trends that we're seeing, the first being fairly obvious around data breaches and data leakage. These events continue unabated, whether it's a half a billion records from Marriott or 800 million from a mortgage company, I think First American Financial Corp. was the latest one, potentially, and maybe there's one more recent than that. And then as it relates to overall security priorities, data loss and leakage is the number one Cloud security concern for folks, according to the 2019 Cloud Security Report.

Harold Byun:

The other key trend that's really impacting the overall Cloud data security attention level is how people are embracing DevOps, overall. This notion of shifting-left in the release cycle, being able to move faster from a continuous integration, continuous release model, is something that is very, very appealing for a lot of organizations given that the pace of the world is moving faster. But, it also means that, in many ways, as we've heard from certain customers, it's leaving security behind in many aspects and that people are just releasing at a certain pace and security becomes very much an afterthought. Hence, given the rise also to terms like DevSecOps, where security is more interwoven into the overall DevOps release process, but, ultimately, the business wants to move faster, developers want faster spin ups, and access ultimately so that they can respond to the business requirements and competitive requirements in an accelerated fashion.

Harold Byun:

But, the real question is, how do you actually secure that data while you're still meeting that accelerated timeline and deployment model? And then, the third major trend is privacy, privacy, privacy. We hear about it all the time. GDPR is really the poster child. There's a lot of attention with the California Consumer Privacy Act, or CCPA. And, I think there are around 10 to 12 additional states that have regulation in progress around data privacy, as well as additional international nation states that are passing their own privacy regulations, as well. Obviously, the financial penalties have stiffened and the brand impact is much more severe. You can just point to the latest Facebook or Apple privacy ads in terms of the importance level that people are placing on that in today's world.

Harold Byun:

When we look at overall privacy, it's more of a global view of privacy regulations and where we're starting to see emerging privacy regulations. Just a more of a global visual view. But, more or less, saying the same thing that we were saying in point number three of the slide before. And then, within the overall context of the shared responsibility model for Cloud, I think many people are familiar with this. Many people, we find, still are not. This is a view of this from AWS where the security of the infrastructure, or provisioning of the infrastructure, and securing of the infrastructure is the responsibility of the Cloud provider. And this example could be AWS, or Azure, or GCP, or other Cloud provider. The responsibility of the customer is the security in the Cloud and, specifically, security of the data in the Cloud. And that's a fundamental difference in that.

Harold Byun:

You can even look at the overall security landscape of vendors working on Cloud-related technologies. There's a ton of capabilities around securing, and configuring, and notifying around Cloud infrastructure,

or misconfigurations. There is very little around actually securing your data in the Cloud. And, obviously, once you put that data in the Cloud, it does become your responsibility to protect it from a regulatory standpoint and just from a best practices and security standpoint.

Harold Byun:

When we look at Gartner's view of the landscape of security, they have broken it out into some major categories. So CASBY, or Cloud Access Security Brokers, organizations like Sky High or Netskope, have traditionally been focused on monitoring Cloud access. There is the CSPM space, which is really around overall security of the infrastructure, so people who focus on configuration and compliance management. AWS, themselves, at Reinforce yesterday, just made an announcement around tower control, or Control Tower, rather, which is basically an automated compliance solution, as well as their announcement around Security Hub, which is a Cloud-based SIM model for security events and incidents. CWPP, or Cloud Workload Protection Platform, is really a focus around the processing workload or containers and microservices, and how those are secured.

Harold Byun:

Again, a portion of the infrastructure, a method for how you're deploying processing and compute, but not a lot of focus on the data itself. When we look at how a Cloud data protection platform could evolve is really, what are some of the capability sets that we find people asking for, as it relates to a set of functionality to protect their data as they move to Cloud. At the foundational level, and I think everybody has a common understanding of this, is there are encryption at rest capabilities built into S3, and RDS, and different TableSpace or database-level encryption options that people have. But, these are all container-based approaches to encryption, which do very little to nothing, to protect against the modern-day attack or a modern-day data leak. And so, that's really kind of the continual gap that we have in the market and, quite frankly, why we believe that the breaches continue to happen. The security model is, quite frankly, protecting against the wrong risk, or trying to protect against the threat in an incorrect manner. And that's really what you get at this foundational level.

Harold Byun:

There's a lot that's being done in terms of monitoring and visibility. There's this focus that folks are starting to apply around how to really protect data once it's in Cloud or as it's passed across parties. Then there's data-centric protection mechanisms, which is where Baffle, obviously, is focused very much on simplifying data-centric encryption with no code changes. Then we also have an advanced encryption capability, which we won't be talking about today, but it provides basically homomorphic-like functionality on encrypted data, never decrypting the data in memory, in the search index, in use, or in process, while still allowing mathematical computation and analytics to occur on the data. That's out of scope within the test of realm that we're targeting today, but happy to engage with anybody on discussions on that at another time.

Harold Byun:

So, let's talk about Test/Dev spin ups, and data leaks, and where we see the overall migration of data, and how people are handling that today, and, inevitably, where things may go awry. We look at a basic Cloud migration flow. Typically, what happens is you have a source and target endpoint. So, you have a source endpoint and target endpoint. Sometimes, in the case of AWS, there's an appliance to help move

large data sets. But, inevitably, what happens is that the source database has some type of migration process, or data dumping, or data extract function that runs against it. And the extract of that data actually lands in a clear text CSV in the Cloud. So at 0.1, for a lot of folks, that's already a compliance violation, and you've already just taken a dump of your sensitive data and put it in a CSV in the Cloud, before it gets ingested into a data structure.

Harold Byun:

And then, ultimately, what ends up happening is people tend to spin up production data clones for development in that DevOps or shift-left release cycle to quickly spin up a Dev environment on demand, and the production data clones sit out there in the Cloud, in the clear. And this is a problem that we find pretty much endemic across multiple organizations today. As you're probably well familiar from just the general news around data leaks, there's a tendency for that data to go unmonitored or be handled in a less secure manner. Let alone the fact that in many cases, developers will have access to sensitive personal information in the clear while they're building their applications. So, this is fundamentally problematic on a number of levels from a data security and security access control standpoint. In addition to that, there are a number of data pipelining capabilities that exist in the Cloud.

Harold Byun:

The ability to not only spin your database up in a Cloud environment via database platform, as a service, but the downstream effect of that is actually pipelining, or running extracts on that data, from those databases in the clear to alternate sources, and ultimately, in many cases, the data pipelining efforts stages that data in an S3 bucket, and that S3 bucket, oftentimes, it can be used in many ways. A lot of people are using S3 for ingestion or storage into multiple other analytic sources. But traditionally, we find it going to either MapReduce, or an EMR set up, or some type of elastic, or some type of unstructured store with elastic layered on top. So, people leveraging ELK Stack and things like that. Inevitably, though, this is effectively a chain of events that is leading from your on-premise environment into a data landing in the clear, potentially in another database environment in the Cloud, and then being pipelined again in the clear to additional tertiary data stores.

Harold Byun:

As you can see, inevitably, this can prevent a significant data exposure across your organization and your Cloud footprint. When we look at some of the key data risks, there's some basics here that developers, obviously, have direct access to the production data fields. This is fundamentally problematic for a lot of organizations that we're working with. The sensitive data sitting in the Cloud environments may obviously violate regulations. You still, obviously, have a core data breach risk than the fact that you have a data structure out there with production data. And arguably, probably not necessarily, getting the same level of security attention.

Harold Byun:

The data pipeline into S3 is ultimately leading to another data exposure point and, rephrasing a common phrase, when bad things happen, misconfigurations happen, right? And so, a misconfiguration happens and that typically seems to be what is happening. And so, how can you better put some wraps or controls around this type of data flow? When we look at what's available in Cloud, aside from the rest that we've already mentioned, most of the data protection solutions today are really focused at

encryption-at-rest or container-based encryption models. And again, I think that there's a fair amount of market education that needs to happen here on the security practitioner side, and so, we'll be very blunt about it. These solutions do absolutely nothing to protect your data in the Cloud. They do absolutely nothing to protect your data from a modern-day attack or a hack.

Harold Byun:

That is fundamentally the problem, where people are checking a box because it's easy and it meets a security check box, but these solutions, again, the threat model that they're protecting against is somebody in a data center, or somebody breaking into AWS's data centers, or Microsoft data centers, or Google, and ripping out hard drives. They're really designed to protect against a lost device in a taxi, versus a modern-day attack. We jokingly call this "protecting against Tom Cruise" or "the Tom Cruise threat model" because Tom Cruise and his team are obviously going to break into the data centers and steal all your data. And that's how most of the breaches are happening today, right? So, we don't think that that's the way breaches are happening today. And this is a view of container-based encryption, for those of you who may not have as much familiarity at the database level, but anybody who access to that system, including your developers and DBAs or CIS admins, basically, are going to see data in the clear, in an environment that's using an at-rest encryption solution or a TableSpace or transparent data encryption to protect that data.

Harold Byun:

This is an example of application-level encryption. You can see here that the fields are actually encrypted at the field level, and so anybody working with that data, as it goes to Cloud, will not get access to sensitive data fields in the clear. A couple other data points in these lines... DOT was probably the most recent poster child, but they had TDE in place, and again, it doesn't do anything to protect against a modern-day data theft.

Harold Byun:

How can we simplify data security for Test/Dev? The way we actually can do this is if we look at the way that we support Cloud migrations. We basically have what we call a "single click secure migration". And what that does, is it specifies any data source or database source on-prem, it could be in another Cloud, it could be in a production environment, and we basically allow you to specify a target in cloud. And we encrypt the data on the fly at the field level, at the field or record level, actually. And so, what that means is that you can basically take a production spin up and point to a stage spin up in Cloud. Basically, it's a point-and-shoot operation. And on-the-fly as that production environment is, as the production data set is migrated to a staging or Test/Dev environment, the fields are protected on-the-fly in Cloud.

Harold Byun:

It never hits the Cloud in the clear, and we can also integrate with virtually any key manager. So, AWS, KMS, Azure, Key Vault, any other third-party HSM or key manager. We also support AWS Cloud, HSM, and Secrets manager. And so, this is a relatively simplistic solution. I will walk you through it on a couple of screenshots here really quickly, and then I'll also just do it for you live, so you can actually see what's going on here. So you'll see here that we have the ability for us to specify that source database, pick a key manager, pick a migration plan. We can do it in place or a source to target, which is the single-click

method. And then we specify the target, pick a field, and basically click "Go". And you're off to the races, in terms of securing your data.

Harold Byun:

So we've tried to make it, in the words of some other folks, stupid-simple is the way we've approached this problem. And so, hopefully you'll see that this is also, potentially, a stupid-simple way to actually do this. Let me see if I can actually get over and screen-share here, it might just take one moment to actually get this going through our little application here. I will share my entire screen, and that will put us back into here. So what I'm going to do here is simulate, or walk through, not simulate. I'm actually going to walk through the data migration process so you can see it live. And so, what I have stood up here is a sample table, which represents our source database. And you can see that there's potentially sensitive data in this table, called HB megastore, which is my megastore application.

Harold Byun:

And then I have an RDS spin up, here. This happens to be running SQL server. It could be running whatever flavor of RDS you want to run. And you'll see that within this database, there are no tables down here. I just refresh this in the lower left-hand corner. There is no table, and there's no data in the Cloud in this particular instance. And so, if I go to our interface from a management perspective, you can see that this is the Baffle admin interface. I've already entered some credentials, just because I don't think that that's a valuable use of time, and we've specified what we're going to use for our Baffle Shield, which is the main component that we use for encrypting and decrypting, and also select a target. It looks like that time down here, so I guess you'll get to see the whole process here.

Harold Byun:

When I go into setting this up, what I'll do is we'll do a test of migration. And, what we'll do is walk you through how we actually do this. So, I'm going to select my source endpoint, or source database. I select the key store in this case, I'll use AWS KMS. Again, we can support a third-party HSM or cloud HSM. And then we also have an in-place migration plan. We also have a source-to-target migration plan. And so, for those of you that are looking to support Test/ Dev spin ups, this is the source-to-target plan and I'll specify a target server. And we basically pick whatever columns we want to protect. In this case, I'll protect city, country, and customer name. And then what I'll do is basically click "Go".

Harold Byun:

When I click "Go", what's actually happening underneath the covers is we are talking to this AWS KMS set up. And what we do is we utilize the customer managed key, which you, as a customer, would create. And we use that to encrypt data encryption keys. Your customer-managed key is always owned by you, and you use that for, ultimately, securing a lot of what's going on with KMS. And what we do is we produce a data encryption key that is encrypted with your CMK, and you can always rotate one or the other, if you choose to, from a compliance perspective. And we would, obviously, rewrap keys at that point. So, we never persist the keys. We're just effectively creating a mapping. And what's going on underneath the covers here is effectively this process.

Harold Byun:

So I just showed you the AWS KMS integration, but what is going on under the covers is I was in this admin console talking to AWS KMS in this instance, and we're using Baffle Shield to support that point and shoot migration to Cloud. And so, that's effectively what's been going on here. If I go back to the application set up, you can see that we selected this source-to-target migration plan, and that we've specified that certain keys are going to be encrypted in this model. If I go to the instance and we can see that if I refresh in the RDS instance now, we have this table that's been migrated. It was not there before. And when I select the data, we've actually encrypted that in the Cloud. This is where, again, point-and-shoot, facilitating quick spin ups, enabling a production to stage, to performance, to Test/Dev, or whatever order or cycle that you guys want to support, is available to you.

Harold Byun:

And in the event that you ever do need to actually see the data in the clear, we basically would just execute a connection string change. You go through Baffle shield and you obviously would limit this based on whoever needs to have access. And when we do that, we can access the database and Baffle's function for applications or other types of access to the data is to obviously decrypt that seamlessly for the application. So in this manner, we're invisible to applications. We're invisible to work bench in management studio applications, but in the need to actually protect that data from a compliance perspective, again, it's pretty much a simplified operation.

Harold Byun:

Some other aspects of the solution that may be of interest to folks. So, in addition to just the simplified migration, we can also support data masking. So, in addition to the single click migration, we encrypt the data underneath, but then it's also masked in the Cloud from developers optionally using that Baffle shield. So, instead of rendering that data in the clear, we can also choose to mask the data on a field by field basis. And the benefit of this is, again, going back to data pipelining and the downstream data stores in that event, when you're running these types of extracts and pipelining activities, the data is going to retain its masking and not be exposed in the Cloud environment. So, it's one of the other benefits of this type of approach.

Harold Byun:

We can apply similar principles, so we don't just do this at the field level or the column level. We can also do this at the record level. So if any of you are SAS providers that are hosting in Azure, or AWS, or GCP, there's a fair amount of interest from multi-tenant SAS providers who are using us to basically encrypt data at the record level, tied to a data owner or entity ID. And the benefit of this is that, in a co-mingled data store, you can basically segment your data based on the data owner or the subscriber. And then in the event that the subscriber chooses to no longer do business with you when you delete their key effectively, it facilitates data shredding, which is evidenced here, in the right-hand side, the data is encrypted. When you delete the key that that data is encrypted with, we mask it within the environment or suppress the output of it, but effectively, it can no longer be rendered in the clear, which fulfills the right to be forgotten, which is a key requirement for GDPR and CCPA.

Harold Byun:

So, there's a fair amount of interest in this capability, as well as you choose to operate within a cloud environment. The overall structure for how Baffle actually deploys and does what it does is, again, we

have three major components. One, is the admin console that I was showing you earlier. There's the Baffle shield, which is basically a TCP SQL wire protocol, reverse proxy. So, it supports SQL and no SQL. It sits below the driver layer and is completely invisible to any application. So, it also interfaces well with APIs and service mesh for containers and microservices. By doing this, we become a consistent data abstraction or control plane, as it relates to data in Cloud. Obviously, we can do it on premise as well, but in this case, we're talking about Cloud. So, we become a consistent data control plane in there.

Harold Byun:

It can also, in the third major component of our solution, which we're not going to talk about today, is our secure multi-party compute capability, which, again, provides for mathematical computations, sorting wildcard searching on encrypted data without ever decrypting the underlying values. So, that's a stronger security method that protects from advanced attacks, but also doesn't have any impact on the functionality of an application. So we can fully support application functionality, like search or even aggregate analytics and those types of scenarios.

Harold Byun:

So, in summary, how can we help? Well, one, we really help customers fulfill their portion of the shared responsibility model for protecting data in Cloud. The single-click secure migration encrypts and masks your data in the Cloud. It's a great way to support the developers in your organization with quick Test/Dev spin ups, and facilitate faster response to the business.

Harold Byun:

We fully support a range of Cloud-native services, all the Cloud key managers, different templates for Cloud spin ups, high availability modes. We have an out-of-the-box, for example, our Baffle componentry and high availability model spread across two availability zones, registered with an elastic load balancer in about three to four minutes. So, we've vastly simplified the deployment model and integrated it thoroughly with Cloud-native services. And, ultimately, we do this because securing data, as many of you are well aware, is not the easiest thing, and things get broken, and things get misconfigured, and steps get missed. And so, part of the objective for us, as an organization, is to help make it quick-easy, I guess, if that's even a term, and make it easier for you to secure that data. So with that, what I'll do is, I will move into Q and A, here. So I'm going to go back into our Q and A portal, here, and see what may have cropped here.

Harold Byun:

A couple of questions that we do have are, do you support Cloud HSM and other key managers? The answer is yes, we, absolutely, do support cloud HSM, Azure key vault, other key managers. We have a good relationship with Talas, who owns Gemalto KeySecure, as well as running their own HSM. It is really your option in terms of how you want to store your keys. Hopefully, that answers that question. There's more information on our website, but, again, we use industry standard came up client four key managers, PKCS 11 libraries for HSMs and for the Cloud key management solutions, like KMS or Key Vault. We Use the rest APIs because those are how those are accessed.

Harold Byun:

Homomorphic encryption is a question that we're getting from the audience. So, we support... We say, cause I think it makes it easier for people to understand, is homomorphic-like encryption. And, what we mean by that is, we can support all of the functionality, all the functional benefits that you would get from homomorphic encryption, but we use AES encryption. And that sounds impossible for many people to swallow on the surface, but effectively we're using AES encryption, but we're still able to perform operations on the data without ever decrypting the encrypted values. It's a method known as "secure multi-party compute". We have a patent on this capability. We've proven this out in multiple scenarios and I'll be happy to walk you through the solution.

Harold Byun:

Just as another data point, Google Cloud actually implemented multi-party compute last week in their environment, to achieve some similar principles in terms of functionality. We're able to support ad hoc queries, ad hoc mathematical computation, and aggregate analysis via our methodology, including wildcard search on AES encrypted data. I know it sounds crazy, but we're happy to engage with you on additional discussions, if you have an interest there.

Harold Byun:

Additional questions... I think a couple more here. How are keys protected and do you support key rotation? We mentioned this a little bit earlier, we do support, we effectively take a master key, and we use that to encrypt the data encryption keys, and we can support key rotation for both DEX master keys, or CEX and DEX, but we facilitate a full hierarchy of key rotation, in that regard.

Harold Byun:

Do we support GCP and Azure? Yes. We support GCP and Azure. We're independent in terms of Cloud, we're agnostic in terms of the Cloud provider, we're agnostic in terms of the key manager, we're agnostic in terms of the database.

Harold Byun:

Another question here, if you encrypt the data by segmentation-like key, and the key is removed, how does the base business get access to the data if there's a legal requirement? Well, the answer is that if the key is removed with no recovery process, then people are not going to get access to the data. Even if there is a legal requirement, unfortunately, I mean, effectively you could try and break the encryption. We use quantum safe mechanism methods, right now. So, that's not going to be an easy way, that's not going to be an easy path forward. Ultimately, when you do make the choice to encrypt data, it really does become a key availability question more than anything else.

Harold Byun:

And I think that that's important to bear in mind, but that is also no different than if you go back, I don't know, maybe I'm showing my age a little bit, but when you go back to backing up to tape and encrypting that and sending it off to Iron Mountain, those tapes were encrypted. And if you didn't have the key in your DR center, then you were pretty much hosed. And so, the same principle applies. It's a key availability problem.

Harold Byun:

If there is a scenario where you expect that a business is going to need to get access to the data, then you basically need to find a way to have that key available to them. I think one of the benefits that we have, again, is we can segment the data at the record level. So, if the business unit is an owner or a data owner of a set of records, then at least you can use that business unit's key, or a group level key, to restrict access to that data. Whereas other business units or other data owners in a co-mingled environment would not be able to see that data. And so that is one of the other benefits. So, you may not necessarily need to remove the key, you just need to restrict access to the key. That could be an alternative approach, as well.

Harold Byun:

Hopefully, this was helpful. There's just other events, so we'll be at the Microsoft Inspire conference in a couple of weeks. We'll be at Reinvent towards the end of the year. We were just at Reinforced earlier this week, and FSI SAC. Related to the homomorphic encryption question, there's a homomorphic webinar, as well, that we have on BrightTALK. We also have another webinar on securing data in the Cloud, and you can always reach us at info@baffle.io, if you want more information. We're happy to engage and promise not to waste your time. This is my information, I'm harold@baffled.io. I'm on Twitter with my alter ego, Howard, and feel free to reach out to us if you have additional questions or want additional information. I really want to thank you for taking the time out to join us today. And hopefully this was some decent information for you. So have a good day. Thank you.