Harold Byun:

Hi and welcome to today's webinar on CCPA Data Privacy Compliance Simplified. My name is Harold Byun. I'm the VP of Products at Baffle, and we're going to be walking through some of the CCPA requirements that are soon to go into effect at the beginning of next year. We'll be walking through some methods that we believe will help simplify compliance for your organization, assuming that you are affected by the regulation. We're just going to hold on for just another minute here, give some folks who are looking to get into the session, and then we'll jump into the presentation. Thanks for your patience here.

Harold Byun:

And as we get rolling here, just some housekeeping items. We'd like to make this as interactive as possible, so if you have any questions as we go forward, please use the chat function to go ahead and ask those questions. We can always be reached at info@baffle.io. My email personally is harold@baffle.io if there's other questions as well that you, for whatever reason, don't want to ask in this particular forum. So why don't we jump right into it? Many of you are probably familiar CCPA compliance, or the California Consumer Privacy Act, is going into effect in January of this coming year. A lot of organizations are obviously going to be impacted by this, and there are certain specific regulatory requirements around that that may require you to obviously take action as an organization and maybe change some of your privacy policies as well.

Harold Byun:

Some of the things that we're going to be covering today is obviously an overview of CCPA. We're going to go into the data breaches and the penalties and fines that are associated with that, as well as any type of... Some methods to simplify compliance in the [inaudible 00:02:27], one run through a quick demo, as well as go through some Q&A as well. These slides will be made available for you as well. Hopefully that will be something that'll be a useful resource going forward, too. So in terms of key timelines and dates, many of you, as we've already talked about, this is going into effect in January 1st, 2020, but one of the things obviously to be aware of is that the consumers actually have a right to request data back 12 months. This actually goes back to the beginning of January 1st in 2020. I mean, January 1st, 2019, rather. So something to be aware of that they do have the ability to make a request to see how data is being collected and stored and used or shared or sold back until the first of 2019.

Harold Byun:

In terms of enforcement, the attorney general is kind of, I don't know if I would call it a grace period, granting a type of grace period for six months, so enforcement action won't begin until July 1st, 2020. It is worth noting that the consumers still have a right to file complaints once this goes into effect. So just something to be aware of, but in terms of perhaps buying yourself a little bit of time from an enforcement perspective, July 1st is when the AG will start to take action there.

Harold Byun:

Is your organization actually affected? There's some key things to consider here in terms of whether or not you're affected. There's basically three major criteria, and if your organization meets at least one of these criteria, then yes, you would be mandated to comply with CCPA. That is generating more than $25 million in revenue, collecting personal information, and there's definitions of personal information that are going to come here on 50,000 or more California residents, households, or devices annually. You

make more than 50% of your annual revenue selling California residents' personal information, so this doesn't even really apply to whether or not you're located or based in California. There's a fair amount of write-ups around whether or not you need to be a California corporation or a California organization. It really is geared towards who is the consumer base and who's interacting with your organization from a consumer residency perspective.

Harold Byun:

In terms of CCPA personal information, this is not an exhaustive list, but there's a pretty broad defined view under CCPA as personal information that identifies or describes an individual, or could be directly or indirectly even inferred with a particular consumer or household. The basic PII fields are here pretty much at the top, the standard ones that you would normally expect to see. Some other ones that are potentially of interest are obviously the online tracking technologies or cookies or IP addresses.

Harold Byun:

That's something that's become much more evident with GDPR, biometric information, geolocation data. Those are things that are maybe not immediately obvious in terms of what people are necessarily collecting, obviously professional or non-public educational information. The bottom bullet, I also particularly find interesting. It's any inferences or profiling that can be drawn from behavior or market personas or other characteristics or attributions that could be associated with an individual consumer. Anything that anybody is doing around big data and AI, or some other type of advanced targeting, would particularly come into play under that kind of all-consuming bullet.

Harold Byun:

There are methods that we believe will help still allow for what we would call privacy preserving analytics, or big data analytics that still allow for privacy to be insured so that companies don't have to completely abandon their investments in these areas, but it remains to be seen how this is actually policed. The fines associated with the compliance factor pretty significant when we get into the meat of it. Some exclusions here.

Harold Byun:

Things that are excluded, obviously information that's publicly available, things that are registered with the government, property tax information, any public educational information, the CCPA is also excluding aggregate data. It is something that is geared more towards the individual consumer. Also, medical or health information that's governed by HIPAA is also excluded, so if you're a healthcare organization, you already have regulatory controls under the guidelines of HIPAA, but that type of data set is excluded from the CCPA Act.

Harold Byun:

Third-party risks. One of the other aspects of this, there's been a lot of focus on third-party supplier risk over the better part of a decade at this point, maybe even more, and really trying to assess where data leakage is occurring, or where data exposures are occurring. The challenge with CCPA and third-party risks is that ultimately your company is responsible for any third party that houses your data. In the event of a data breach, which from a lot of the data points that we're seeing in industry it's upwards of 60% of companies have had a third-party data leak within the last two years. This obviously represents a significant risk factor, and ultimately it is your responsibility. It's probably going to require even more

third-party review and risk assessment, and probably modification of contracts in terms of data handling, so something to consider in terms of how you're going to manage your overall supply chain as you move forward.

Harold Byun:

Consumer rights under CCPA, they could pretty much be bucketed into these three major areas. At least, I've bucketed it this way. It's a right to know, a right to be forgotten, and a right to control. Within the context of a right to know, you really want to be asking the questions around what information are you or your third parties collecting about consumers? What categories of information need to be made available to the consumer if they request that? How is that information being used? And ultimately, are you going to be sharing it or reselling it? And then within the realm of the right to be forgotten, similar to GDPR, the consumer has the ability to request that their data be deleted. There are some exceptions here. There's limited analytical use cases. We talked about the healthcare scenario earlier, aggregate data, and certain research scenarios. A lot of that is still going to have to be weighted through in terms of some of the legal documents, but it'll be so those are called out as some limited use cases.

Harold Byun:

The right to control. Consumers can opt out of the sale of their information, but requires a modification to your privacy policy. If you don't already have that in place, you must provide an option or present an option for the consumer to opt out on that. Also, the consumer cannot be discriminated against for asking to opt out. This probably also applies to any type of tiering of service offering that you may offer, where those consumers obviously should not receive a lesser grade of service. There is a question here. Is the third-party risk for the CCPA the same as GDPR? It depends on your view of that. The risk is that ultimately, it's actually greater under CCPA because there's a lot more stringent guidelines around tertiary sharing or resharing of data.

Harold Byun:

So it is any service provider, and there's a specific section on service providers, which is basically the Act's definition of a third party. Any service provider that is responsible for handling or processing data on your behalf, consultants, contractors, billing services, collection services, any other type of customer support or service augmentation or marketing, those all fall under the realm of your organization. I actually think it's more extensive, whereas GDPR is in many ways, I feel like a lot of that was much more focused on cross-border transference of data within the general data protection capability.

Harold Byun:

Keep going here. Data breaches and penalties. In terms of data breaches and penalties, one could argue that it's less egregious than GDPR. I think that it's not. I think from a regulatory standpoint, it may be less egregious depending on your interpretation. From a consumer right of protection, I think it's definitively more than GDPR in terms of penalties and fines, and actually a lot of the more significant fines that we've seen over the last year even has ventured into the 10 to 20% of revenue number versus the 4% of GDPR revenue that everybody seems to be really familiar with. What CCPA provides is a right of action for the actual consumer, the individual, when a data breach occurs. The consumer doesn't even necessarily have to formally sue an organization, but they have a right to receive $100 to $750 for a breach without providing any proof that they were harmed in the data breach. So the net of that is that

if you actually have a data breach, the consumer would then actually be in a position where they could make a claim against your company for $750.

Harold Byun:

Within the guidelines of CCPA, they specifically call out this paragraph. Any consumer whose non-encrypted or nonredacted personal information, as defined in whatever the subparagraph in the section we're talking about. This section that's referenced in the CCPA Act, it's making a specific reference to a data-centric capability in terms of encryption or nonredaction, so if that breach occurs, then you would be on the hook for that. When we do the simple math, $750 times 10,000 consumers comes out to $7.5 million. That's obviously at the top end of the fine, but even if you had $100, it'd be $1 million for 10,000 consumers. For a lot of businesses, that's pretty egregious. The Attorney General can fine companies $2,500 per violation of noncompliance, as well as $7,500 per intentional violation.

Harold Byun:

It's still not really clear on what is going to be the counting mechanism as a quote unquote violation, but one could expect that there could be potentially multiple violations incurred based on how the non-compliance is viewed. We do have another question here. When requesting your data to be deleted, does the vendor need to send proof to the client? Yes, there needs to be an affirmation or confirmation that the data has actually been deleted. This is as part of the process that you're going to need to build out in terms of the handling of this. There's other people that have talked a lot about call center support and how people would actually go about starting to build out process for handling these types of inbound requests from consumers.

Harold Byun:

Methods to simplify compliance. The first kind of approach that we're looking at is just process and methodology in a framework around this. This is a version of a checklist around how to go about assessing your overall risk and scope. There's a variety of different methods that one could follow, but obviously identifying data that you're collecting and storing, evaluating that with your third parties, looking at the security risk or threat to the consumer data, and looking at this from a data breach impact and probability and risk perspective, really putting that within the context of a risk-based approach to managing your data, you're going to have to review privacy policies and procedures again, and probably adapt them for CCPA, assessing whether or not their consumer data is actually being collected in a compliant manner.

Harold Byun:

Implementing additional data-centric protection methods is definitely something that we would advocate very strongly, and then creating processes, as we just chatted about, to address when consumers exercise the right to be forgotten, or to learn more about what data is actually being collected. There's obviously additional training of staff and employees to handle this type of request going forward. We fully expect another 15 to 20 states to pass additional regulation in this area, so it is going to become the norm going forward. While people, obviously, are saying this is the strongest data privacy law in the nation, there's a greater push for a national data privacy law. We have been tracking roughly 15 to 20 states that have a regulation in process as well, so this will be the norm more or less going forward.

Harold Byun:

When we look at data-centric protection and how to comply with CCPA, there are two major points that we wanted to also emphasize. The first is that you need to make sure that you're using the correct technical controls to secure the data, and they call it out in that article that we referenced earlier. That is that non-encrypted, nonredacted data is subject to the data breach and the associated consumer-driven fines that you would be exposed to. The second point around that is, all encryption does not actually protect at the data level. That's a common misnomer that we've kind of learned over the past few years in talking with a number of customers and potential clients, and there's a lot of confusion in the market over what encryption is actually protecting. If I go back two or three years, I would venture to say that 80% of the people that we talked to didn't know the difference between a data-centric method and a standard encryption at rest method.

Harold Byun:

There's a lot of confusion in the audit community over that and in terms of compliance, so it's something that I would encourage people to dig in a little bit. I know encryption can be hard and confusing, but it's important to know what you're actually protecting against. Some of the key benefits of a data-centric approach, it protects the actual consumer data value. The data value in question, when we talk about personal information, what that means from a violation perspective, that becomes pretty important. It addresses encryption and reduction of the data, it can easily enable the right to be forgotten, which obviously is something that is paramount in terms of deletion of that data, and it's a truly data-centric, simplistic model that requires no application code changes or architectural modifications, which allows for faster compliance.

Harold Byun:

This is a view of a non-data-centric protection model called TDE, or transparent data encryption, and I just kind of show this because this is something that we think that people should get a better handle on. As you can see, when somebody gets access to the system, whether this be an attacker and a breach or an accidental type of access environment, all of the data is exposed in the clear. It does nothing to protect against a modern day hack. If you believe there's a lot of noise around zero trust and the modified perimeter or the new perimeter list world, you believe that people are moving laterally in your network and they get access to the database environment or the data store. Using these non-data-centric methods, they get access to all the data in the clear. All the logs are in the clear, and if they do a memory dump, the information again is in the clear.

Harold Byun:

Some of the more high profile breaches were using this type of method, and our view is that it does nothing to actually protect the data. As it relates to CCPA, it will do pretty much nothing to actually ensure that you're not going to be paying breach penalties to the consumer because their data would be actually in a non-encrypted state. This is a data-centric method. A data-centric encryption method is really looking at things at the field or the record level. It is encrypting that data and it will protect it in that data-centric fashion, so obviously if it's stolen or exfiltrated, then in this type of case, you can make the case that this is going to give you safe harbor under CCPA, as well as other data privacy or data protection regulations.

Harold Byun:

This is a data-centric masking and redaction capability. In terms of how this actually appears to the application, it could be encrypted underneath, but you, as an organization, could easily choose to redact it or mask it in other forms or fashion, again, providing you additional end-to-end control by encrypting at the field level in the actual value, and then representing that data in a redacted form at the presentation layer. It really gives you more of an end-to-end approach to better controlling your data.

Harold Byun:

Enabling the consumer right of revocation. When we look at this, supporting the way to do this at scale where we can actually encrypt data on a per record basis tied to what we would call a data owner, or in this case, the consumer. If the consumer decides that they want their data to be deleted by killing their respective key, you effectively have erased the data.

Harold Byun:

You're effectively performing a form of data shredding because the data can no longer be accessed in your environment, which is executing this right to be forgotten. In many ways, we could even extend that control to the consumer going forward. It gives you selective data masking for different data owners. It was originally designed for multi-tenant SaaS environments, but it's something that obviously applies the same principle of a record level segmentation of data on a per owner basis. That owner could be a subscribing organization, or it could be an individual consumer. This is a case study from one of our customers who's implemented our data-centric record level encryption. It's a company called Workiva. They have an environment with over five billion records, and they manage financial reporting for the Fortune 500. There's a lot of sensitive data, and they basically use us as an off-the-shelf Bring Your Own Key service, using record level encryption tied to a data owner at scale with no application modification and no architectural overalls. Again, accelerating the ability to comply or deliver on the security requirements that are being pressed on you.

Harold Byun:

This is more of a technical architecture of how we actually do this. It's a model where we basically take keys from any type of disparate key source. It doesn't really matter to us where the key source is. Even though this says AWS KMS, it could be HashiCorp or Gemalto or any key provider. It could be file-based keys. It doesn't really matter to us. We're able to apply those in any type of shared data store at the record level, again, and that would facilitate that record level shredding. The last kind of use case here around this is, and I'm going to show you all of this really quickly, too, and then we'll wrap up with any additional questions, is privacy preserving analytics.

Harold Byun:

Privacy preserving analytics is something where it is a method to perform analysis in aggregate on subsets or aggregates of the population without actually exposing the individual consumer record, and being able to derive intelligence or analytics out of that information. Particularly, a lot of organizations looking at data as a service, or data modeling as a service, or what I would call multi-party scenarios, are places where we believe that this type of technology is still an emerging space, but there's a ton of interest in it in terms of being able to facilitate different informational access models, again, without violating the confidentiality contract. One of them is third-party data access control. We talked about third-party risk and your vendors and how people are actually accessing information, and how they're

using this consumer-based information. We believe that the privacy regulations are going to make companies move towards more of a consolidated information supply chain.

Harold Byun:

In this model, there is a method to basically allow third parties to get constrained access to a shared data store, predominantly your data store with your consumer's data, and you can control or gate what access they have and what is actually presented to that third-party risk player, which is ultimately going to give you more control and vastly reduce your risk and exposure under compliance regulations such as CCPA. Another variant of this is what we call cross-party data sharing. These are scenarios where we see disparate organizations that want to share information because everybody recognizes that the value of the combined data set is much greater. It's a one plus one equals three scenario, but under existing company policies, as well as privacy regulations or data protection regulations, there's no way that these two disparate organizations could actually ever share information.

Harold Byun:

There are methods using these types of privacy preserving analytics capabilities that facilitate advanced data sharing, or what we kind of call the art of sharing data without really sharing it. It's a dynamic pseudonymization process that allows for aggregate analysis on shared data without ever actually sharing the data values that are in question. Another variant of that is anonymized threat intel sharing. We're in progress with two of the largest banks in the U.S. who are looking to share threat intelligence using this anonymized model of known threat IOCs that could be validated without either party fully revealing what it is that they're tracking. It's just kind of another variant of what I was talking about. These are some of the models there.

Harold Byun:

Before we get into Q&A, let me kind of give you a quick glimpse of how we believe we make this easy. Again, feel free to fire away on questions as we go forward here. qI think I have to share the entire screen.

Harold Byun:

Let's see if this actually comes up. All right. Looks like this is hopefully sharing. I think it's sharing. Yes, it is. Okay. So what I'm going to do here is, I'm just going to walk through a couple of quick demo scenarios to kind of give you a flavor of how we can simplify this for your organization, potentially using these data-centric methods. I'll also show you some of the privacy preserving use cases as well. This is obviously a sample dataset of potentially sensitive data. It could be consumer data, it could be whatever data set you wanted it to be. It's more or less a dummy data set here, and you can see that it's non-encrypted and it's in the clear. What I'm going to do is, I'm going to walk you through our solution and how we simplify that. What we do is, we provide the ability to map simplified encryption and do it without any code changes, so what I'm going to do here is I'm going to set this up.

Harold Byun:

I'm basically going to do a bare-bones set up installing our Baffle shield component, which is an encryption decryption mechanism, and we're going to encrypt the data on the fly while we're talking through this. I was on a Microsoft SQL server. I'm going to use, in this case, AWS KMS. It doesn't really matter. What you'll see is that we provide you with a field picker, and the field picker basically says, "I

want to utilize encryption on certain sensitive data fields that are subject to this regulation." When I hit next and go, we're basically going to talk to a container or a server or wherever we want to deploy this, basically spin it up, lay down the Baffle shield component, and then what we're doing is we are talking to the encryption key store, which in this case is AWS Key Management Services, or KMS.

Harold Byun:

I'm using key ID six, and we're going to use that key ID six to actually encrypt the data in the respective app or database without making any modifications to the application. The benefit of this approach is, again, accelerated compliance, no changes to the application. There's a cost savings and a time savings because you're not engaging with developers, and basically the process that we're going through here is effectively this, where we are talking to this key management store and we are installing this Baffle shield and saying, "Use these keys to encrypt the data." It's a pretty small data set, so we'll see if that is actually completed at this point in time.

Harold Byun:

You'll see that what we've done here is we now have these two fields that are encrypted. If I go back to the database and refresh, we've actually encrypted that data again, without any changes to the application. If I go through our Baffle shield and run that same task on the data, we perform the decrypt function, which is, again, this simplified data-centric encryption and capability without modifying any applications. The other variant of this is a database where we have the data in the clear. This is not what you would want to do for CCPA, but some organizations want to obviously step through things. There's one column that's encrypted. The rest is in the clear. If I connect through a Baffle shield here, what you'll see is that we're able to mask in various formats.

Harold Byun:

When I connect in here, we have that same database available to us, and table. You'll see, in this particular example, we masked the encrypted column with X's. We used the confidential string in this field. We randomly generated numeric data in the customer ID field. We could do dates and timestamps, and we have a partial redaction with the last four showing on this other field, so highly configurable, highly flexible. Again, no code changes, but combined with encryption, it really gives you an end-to-end access control model. The last thing that I will show you very briefly is some of the privacy preserving use cases. In this case, what I'll just show you very quickly is we have a database structure here where there is threat data, and threat data is considered sensitive. We've encrypted it, and the reason we show this is because we are front-ending this dataset with Tableau.

Harold Byun:

Tableau is a third-party application, and what we're able to do is render the data and present it visually, show you trending of the threat data decrypted with frequency analytics, as well as across tabulation count with the target IP and threat data, which were both encrypted fields and driving those frequency counts. This is a mathematical operational capability on encrypted data, which is what facilitates some of these aggregate operations or analytics. As you may recall, within the CCPA discussion, aggregate data is considered out of scope for compliance, so this is another way that you can still drive analytics or intelligence models in aggregate without necessarily violating the CCPA compliance regulation.

Harold Byun:

The last use case that I want to show you is, again, our advanced encryption capability. This is another variant of what I just showed you with Tableau, and it has the ability for us to do things like sub-string searches or wild card, where I could say, "Show me emails that contain 'the,'" and you'll see that we get a sub-string of our sub-search of 27 records. I can also say, "Well, show me the last name where it contains a sub-string of 'th,'" and run a search, and you'll see that we're down to seven records. The interesting thing about this is if you look at this IP address, and then I go in through a direct access, so it's The same IP, and I'm going to go in and go direct here. If I actually access this database, you'll see that all the data is actually encrypted underneath. If I do this and run this query, you can see that all of that data is actually encrypted, but again, using this advanced capability, we're still able to perform different types of analysis and comparison.

Harold Byun:

That's kind of what we wanted to cover in terms of some methods that we think will help you simplify some of the compliance requirements here. What I'm going to do here is take a look at any last questions and see if people have other kinds of things that they wanted to cover here. Just some other resources for you as we go through this. We will be in AWS re:Invent in Las Vegas for the first week of December. If you're interested in booking a meeting, you can reach out at info@baffle.io. We are also slated to present on a panel as part of the IT GRC Forum, December 17th. It's going to be multiple panelists covering critical steps to manage CCPA compliance and risk, so those are a couple other events that may be of interest for you.

Harold Byun:

There's a White Paper on simplified encryption approaches that we have available on our website, which is at this link below. It's baffle.io/simplified-encryption. We also have a Gartner Cool Vendor Report on privacy preserving analytics and different capabilities in that space at baffle.io/gartner-cool-vendor. These will be available to you in posts. We'll make sure that we send out the deck. Let's see if there's any additional questions here.

Harold Byun:

What encryption methods are prescribed for CCPA compliance within the guidelines of the Act? None of these regulations actually specify a method of encryption. In this case, they simply specify one. What was talked about on that slide, which is non-encrypted or nonredacted data, is going to be considered non-compliant within the realm of a data breach, and it's going to subject you to those fines. If you're using these alternate data encryption or physical encryption methods, the data is in the clear, it is not encrypted. I don't think that you're going to have any safe harbor protection not using the correct technical controls, so that's what those points are about.

Harold Byun:

Are there specific key managers that you require and support? Well, we've talked about some of them today, so the AWS KMS for encryption, we support CloudHSM, we're supporting AWS Secret Store Manager, HashiCorp, virtually any HSM that's available on the market. There's a multitude of different key management solutions out there, but we are basically using industry standard protocols to integrate with all of those and use that as a source of key material. And either you or the consumer always owns the key in our model. We don't want your keys. We don't want your data. We just simply are providing a

data-centric method to simplify protection that we provide to you. It's software that you guys deploy. Any last questions here? Going once, going twice.

Harold Byun:

Well, I really appreciate your time. I hope you found some of this information useful and, again, feel free to reach out to us if you have additional questions. Again, we have the other events and webinars ongoing. There's other webinars that are available on our channel as well talking about a number of different approaches to data privacy and security, if you have the interest. So again, enjoy the rest of your time today, and we look forward to hearing from you. Thank you. Bye-bye.