EPISODE 968

[INTRODUCTION]

[00:00:00] JM: Logging provides raw data that can be abstracted into higher level information. Logs are generated at every layer of infrastructure. Physical host, virtual machine, container, pod and Kubernetes cluster. Logs are generated by network proxies, edge servers and API requests, and there's far too much logging information to be read by humans.

Log messages need to be refined into statistical metrics that can be put into charts. A high-volume of log messages can be used to detect anomalies across a system. If unusual behavior is present in a system, the relevant log messages can be identified and sent to a human operator for that operator to triage and respond to.

Kalyan Ramanathan works at Sumo Logic, a platform for log management and continuous intelligence. Sumo Logic recently published the Continuous Intelligence report, which is based on a study of over 2,000 technology companies. It's a useful dataset for anyone who is looking to understand adaption of cloud products and Kubernetes, and it can be found at softwareengineeringdaily.com/sumologic.

Kalyan joins the show to discuss log management, continuous intelligence and the data that Sumo Logic has gathered in the Continuous Intelligence report.

Full disclosure; Sumo Logic is a sponsor of Software Engineering Daily.

[SPONSOR MESSAGE]

[00:01:31] JM: Kalyan Ramanathan, welcome to Software Engineering Daily.

[00:01:34] KR: Oh, thank you very much.

[00:01:35] JM: Today we're talking about logging, and log management, and some higher level discussions beyond that. To start with the idea of logging, logging is raw data. That raw data

gets abstracted into higher level information. Logging is often high-dimensional. It's high-volume. Give some examples of common logging information and how that high-volume, high-dimensional data gets aggregated into something more abstract and usable.

[00:02:09] KR: Yeah, it's a great question, Jeffry. Let me start out by explaining what a log is for many of your listeners. So a log is essentially a record of noteworthy events that are generated from software applications, operating systems, infrastructure, you name it. Log essentially keeps tracks of what is happening behind-the-scenes of the application and the system so that you have a detailed list of events that are happening. So when you have a malfunction, you can go back, you can copy your logs, you can see what is working, not working, etc.

Now every application stack component right from the application, to the infrastructure, to the platforms, maybe on-prem or cloud for that matter that the application is deployed on are essentially emitting logs. What these logs do is, as I mentioned before, provides the ability to understand how these components of the application are working.

So applications or logs are typically written out by app developers and they are tracking just about anything that is happening within an application. It could be a login. It could be a transaction. It could be errors or warnings that are happening in the application. Infrastructure logs generally come pre-packed from infrastructure vendors, and they generally give you visibility into the health and the performance and any other interesting events that are associated with the infrastructure itself.

[00:03:44] JM: My sense is that while logs have been used for as long as we've been doing software engineering, the idea of aggregating all of these logging data and storing it and making sense of it is a newer phenomenon. It has something to do with the cloud. But I'd like to hear your perspective. You've definitely been in the industry for longer than I have. You've seen how the world looked pre and post-cloud. Was log management a thing before the cloud?

[00:04:16] KR: Log management has been around for a longtime, but it has existed in a very different form than perhaps what exists today with the advent of cloud and high-scale infrastructure. This notion of infrastructure, particularly network devices and operating systems, writing logs and perhaps using a syslog type protocol, dropping logs into a centralized store that

you could perhaps query and get some visibility and perhaps do some troubleshooting has been around for a longtime. This is not a 2010 or a 2020 phenomenon. This has existed from the days of when cisco routers and switches and whatnot have been a common place infrastructure within a data center.

What has changed I would say in the last 10 to 15 years is the ability now to actually collect all these datas at very large scale, and then more importantly, the ability to analyze this data at an even larger scale so that now you can, one, identify needles in a haystack when you're running into an application malfunction or a security issue within your application, or two, you can do very long-term trending of this data so that you can understand how your application is perhaps performing over, let's say, the last one year, or how many users are you getting to your application in the last six months.

Where are these users coming from? What are they doing with their application? Or perhaps look at security incidents that you may have had in your application going back months and months of time, right?

What the cloud provides you and what these high-scale storage and analytics engines are providing you is the ability to indeed do this at scale, which was not quite available once upon a time. So once upon a time, log management used to be the purview of maybe a team of 5 or 6 people. Now, we have customers of products like Sumo Logic where we have hundreds or even thousands of people who are actively and concurrently using these solutions.

[00:06:34] JM: So the trends of the cloud have made modern log management solutions possible. The cloud has also changed how infrastructure itself is laid out and how we're consuming software. You have more layers of infrastructure. You have physical hosts. You have VMs. You have Kubernetes. You have pods and containers. You have black box proprietary tools like DynamoDB that you don't really have a tone of introspection into.

With all this heterogeneity, can you describe some of the best practices for how to monitor all of these different layers and all of these different systems?

[00:07:20] KR: I mean, it's an interesting question. The basic hypothesis of monitoring starts out with visibility, right? There multiple terms that people have applied to visibility. Some people call it observability. Some people call it transparency. You name it. But at the heart of it, you cannot manage what you don't see and what you don't understand.

So the essence of all of this starts out with, "I need to bring back all the signals," maybe logs, and metrics, and traces, and you name it, into a central repository so that I can understand the various components that makes up my application all the way from my application to the platform of other service component, to the infrastructure as a service component, [inaudible 00:08:14] the hosts, and the VMs, and the containers that you just mentioned. So the essence of monitoring and the essence of understanding an application is that visibility.

The next thing that you do once you get control of visibility, once you have a good understanding of the application, is to start monitoring the application. What that means is to understand what is normal and what is abnormal in the behavior of the application itself. Now, there are multiple techniques by which you can do this. You can do it via static methods. You can do it via dynamic methods, and there are even machine learning or AI-based approaches to understanding what is normal and what's not normal about an application.

Then, finally, when you have identified an abnormal condition, obviously that's where you need to take some action. You need the right tools and the right analytics, the right troubleshooting systems in place so that you can very quickly understand the root cause of the problems and then you can fix those problems itself.

Now, this problem has always existed and the basic process has always been the same whether it is on-prem or the cloud. What changes in the cloud is exactly what you say, Jeffrey. It's that a lot of things change in the cloud, right? Starting from the cloud platform, to the infrastructure that you deploy, to the architectures of the applications that you're driving, to even the teams that are responsible for keeping the application healthy, secure, you name it, and the complexity of the cloud environment, the newness of the cloud environment, the number of components in this cloud environment. That's what makes this the act and the task of keeping cloud applications so much harder.

[00:10:01] JM: With like the addition of a system that allows me to have some observability over my infrastructure, how am I using that? Who are the people that are using a log management system? Who are the constituencies? Is it just engineers or is it also product managers? Does everybody on the team want to be using this kind of observability software?

[00:10:29] KR: Yeah. I mean, I think it's interesting. Here at Sumo Logic, we see many personas who use the data that comes from applications. Signals from applications such as log, right? We obviously see DevOps teams and SRE teams that are very interested in monitoring and ensuring the performance and availability of applications.

Logs are also at the heart of ensuring the security and compliance posture of an application. So if you were to walk into a typical enterprise today and look up the security operations team, chances are that they are looking at logs and events coming from these applications to identify security threats, respond to these threats to ensure the compliance state, maybe PCI, HIPAA, you name it, for the application itself.

But we're also starting to see logs being used in other interesting ways. A lot of the development tools, anything from GitHub, to CICD tools, to even deployment tools write out logs. So if I am a VP of engineering and I want to get complete visibility into my DevOps process, the number of tools that I am making from GitHub, the number of times I'm running my CICD workflows, the number of deployments that I'm doing to my production environment. Chances are that a log management system that integrates too many of these systems can indeed provide you that complete end-to-end visibility.

Finally, we are seeing logs being used in many newer applications. For example, IoT applications. We have customers who write out IoT logs, and guess what? If you can bring these IoT logs into a centralized system, you can learn a lot about your IoT systems. You can see where these systems are being deployed. You can see how they are working or not working for that matter, and it provides you a lot of visibility into not just your application, but also the business that is being driven by your application.

[00:12:38] JM: You mentioned anomalies a little bit earlier. You want your log management system to detect anomalies, because you cannot manually look at everything that is coming in to your log management system. I mean, that's the whole point.

But the problem with out-of-the-box anomaly detection is that anomalies vary from infrastructure to infrastructure. It might make sense for some companies to have a system with a QPS of 10 billion per minute, and some other system, if it hits 10 billion QPS per minute, then that's dramatically – Or I guess QPS is queries per second. So per minute doesn't even make any sense.

But in any case, anomaly is kind of a subjective phenomenon, and an anomaly depends on what the infrastructure is. So how do you want to define anomalies for a log management or an observability platform?

[00:13:43] KR: Yeah. I think you're hitting on a problem that's been around for a longtime, and many observability systems, and it doesn't have to be logs in particular. Even systems that collect signals and metrics from infrastructure and applications have long been stymied by these problems.

There are some standard ways by which you can try and address this. Let me sort of walk you through the different approaches here. There is the age-old way of setting up what I call static thresholds, and static threshold is where you may say that, "Look, if my infrastructure produces more than X number of warnings or errors per minute, fire off an alarm; or if my application has a latency of let's say more than 20 seconds, fire off an alarm," right?

There are many systems that sort of work with these static anomaly-based thresholds, and that's been around for a long time. The problem is exactly what you mentioned, right? Which is how do you know what is that right threshold set? How do you know if 10 seconds is the right threshold, or should it be 20? Should it be 30? That's pretty much dependent on the infrastructure, the application, the particular use can. You name it, right?

We've seen system sort of obviously improve from the static anomaly-based detection mechanism, and systems generally do that through one of two ways, right? One of two ways.

One is the applied dynamic thresholds, and dynamic thresholds is where the system itself understands as to what is the right threshold that needs to be applied.

Now, this can come through multiple mechanisms, or multiple methods I should say. You can do this based on historical basis. So for example, say you're an e-commerce site and you want to set threshold for let's say 10AM in the morning, and let's say 7PM in the evening, right? Typical e-commerce sites generally will see more traffic a little bit later in the day given that people are going home. Now they're going to be transacting on this e-commerce system. You name it.

The time or temporal system can look at patterns of usage of the system overtime and can apply thresholds based on patterns of usage. So every day at 6PM, I expect a certain threshold or a certain performance benchmark for my application, and I can go back over the last one week. I can go back over the last one month and then I can arrive at what that threshold ought to be and I can set that as a benchmark. So here's an example of an automatic threshold or benchmark that is set by the system itself.

Now, another way of setting up these dynamic anomalies is arriving at it through a means. So what you can do is you can look at a typical dataset and then you can arrive at what is the normal for that dataset and then set your threshold as a certain standard deviation from that dataset.

For example, if you expected a certain performance at 6PM every day, look at your last 10 hours. Arrive at what is a normal for the last 10 hours and then say, "Anytime you have more than 1, 2, 3 configurable standard deviation from that point in time, you set up a threshold and you set up an alert as a result of this.

So there are multiple ways as you can see by which you can set up a threshold and thus identify an anomaly within your platform. Now, the most recent way, and I'd say the way that most people are starting to experiment with anomaly management right now is by using advanced technologies like ML and AI, where you can apply machine learning technologies to now start to identify what is normal and what's abnormal. This is obviously a brand-new ground. Still in experimentation phase. Some of these techniques work. Some of them don't work that well. But

we're starting to see more and more companies now start to work with ML and AI technologies to identify anomalies in their system.

[00:18:02] JM: Can you give an example of how you would want your log management system or your observability system to expose machine learning functionality?

[00:18:16] KR: Yeah. Let me give you a good example of how Sumo Logic does this, right? Sumo Logic collects a lot of logs, tens of terabytes of logs from unique customers, right? So it's impossible to sift through tens of terabytes of logs, this is on a daily basis, to identify patterns of things that you may see in a log, right?

So what Sumo Logic does is apply machine learning technologies, like clustering for that matter, to then very quickly summarize these logs into patterns. So imagine a log dataset that may have let's say 30,000, to 40,000, to 50,000 logs that may be arriving into your system by applying a technology, propriety technology, what we call log reduce in our system. You can quickly now summarize these 30,000 to 40,000 logs in about 10 to 20 patterns.

So rather than sifting through each one of these logs or trying to search through these logs and identifying what might be happening in this log set, you are better off by looking at these ten log patterns identifying that – Summarizing that either of these 10 log patterns a perfectly acceptable and normal for your system.

But, aha! These two things that were recollected by Sumo Logic and that were reduced by Sumo Logic into patterns are indeed unique log patterns, and perhaps those are the things that you should be paying attention to as you are trying to troubleshoot perhaps a problem in your system. So that a perfect example of how machine learning can be applied to large-scale problems, in this case, log analytics, to help you quickly identify the needle in a haystack and then focus on just the right things other than get drowned by the complexity and the scale of your log analytics problems.

[00:20:12] JM: The vision that your company, Sumo Logic, has for how log management and observability tools and metrics tools of today evolve is something you call continuous intelligence. Can describe what continuous intelligence means?

[00:20:34] KR: Yeah. I mean, it sort of goes back to the discussion that we've been having, Jeffrey. What we are seeing is that many of our customers are making the move to modern applications and to cloud-based systems, and when you run your applications in cloud systems, what we're seeing is that just about everything changes with your application, right?

Your core platform is different. Your application infrastructure is different. Your architecture is different. Teams that are involved in this management are different in some sense. You're going from dev, sec, ops team that were unique teams, different teams, siloed teams, to perhaps even dev sec ops teams, where there is one team that is responsible for managing this application.

The one other thing that we see in this world is that your applications are also starting to be developed and released at a much, much faster pace than you ever did before. So what we're seeing is that – And this is our hypothesis and this is the company thesis, I should say, is that you need continuous intelligence to build and run and secure these applications in a different way and in a better way.

So what Sumo Logic's continuous intelligence platform enables you to do is to essentially provide you that continuous intelligence so that you can manage these applications better, that you can build, run and secure these applications better and meet the expectations of the business.

[00:22:05] JM: Whenever there is kind of a new term like this, like continuous intelligence, it's often summarizing some lower-level technologies that have come before it. Can you identify what are the specific technologies that are composed into continuous intelligence? What makes that possible?

[00:22:28] KR: Yeah. I mean, at the heart of continuous intelligence for Sumo Logic, is our service itself. Our service, the Sumo Logic service, runs in the cloud and it's a cloud native service, and it's really the underpinning of what delivers continuous intelligence to our customers. Our cloud native service leverages every aspect of the cloud that we run on. We are cloud native on AWS, Amazon Web Services.

What it does is to provide our customers incredible scalability and elasticity. So our service collects data from your applications, your applications and your infrastructures, which can run anywhere. These applications may sometimes run in the cloud, may sometimes run on-prem, may sometimes be in a hybrid environment. Because we run in the cloud, very much like a Salesforce, all you have to do is send us your data. Our service collects this data and you can start to analyze this data in real-time within our service.

I mentioned that our service runs in the cloud, and therefore provides incredible scalability and elasticity. Let me sort of emphasize that a little bit. The problems that we are trying to address, our big data problems, the problems that we are trying to address are problems where scale and elasticity is very important. We have customers who have dynamic apps that some days may see, let's say, 10,000 users, and other days may see a million users.

When you see that level of dynamism in application and application usage, the machine data that comes from these applications, the signal that come from these applications also varies dramatically, and therefore if you need a management system that can collect all these signals and that can analyze all these signals, the management system should also scale with the applications and with the users of the applications. So that's what the Sumo Logic service provides. Because we run in the cloud, we can leverage the power of AWS, the cloud that we run on, and can support our customers, whatever dynamic load that they may want to use within the Sumo Logic service.

The one last thing I do want to highlight here is that the Sumo Logic service, while it runs in the cloud, is built with security-first principles in mind, right? So we started 10 years ago building our service in AWS. We realized that our customers and their data is extremely sensitive and extremely critical for our customers.

So we built a service with security-first principal. Our services, SOC 2, type 2 compliant. It's got all the right attestations, including PCI, HIPAA, GDPR. We are FedRAMP ready. So we make sure that we take utmost care of our customers and their data.

[00:25:30] JM: So there was a report that Sumo Logic put out fairly recently about continuous intelligence, and it was just basically aggregating some survey data, some conversations that

you've had had with customers. What's your sense of how infrastructure usage is changing in 2019? What are the biggest changes that you're seeing?

[00:25:55] KR: It's a great question, and let me sort of just give you a bit of a background in this report, and then we can talk about some of the changes that we did see and we're highlighted in this report, right?

Look. I mean, the fundamental reason we put out this report was to really provide our customers and practitioners at large a roadmap of how to build, run and secure mission-critical applications in cloud environments. The way I always talk about this report is that had we had this report 10 years ago when we were starting to build our applications in the cloud, we would have loved to learn as to how do companies build and run applications in the cloud. We didn't have that back then.

What we have learned over the last 10 years working with over 2,000 customers, I would say almost 75% of them running mission-critical applications in the cloud, is that the people who build applications in the cloud think about their infrastructure differently. They think about architecting their applications differently. As a management system that manages these applications and these infrastructure, we have a unique vantage point, a unique visibility into how these customers are doing what they do. So what we have done is collected this data, anonymize this data, and then presented this in a form that every practitioner can perhaps use, learn and implement in their own environment.

Now, in terms of the report itself, right? I think what we saw were really five key observations. First and foremost, what we see is that the whole notion of multi-cloud environments, it's starting to happen. It's becoming more and more real now. When we did this report the first time around in 2016, those were the only dates of multi-cloud deployments. There was one cloud to speak of at that point, that was AWS. Azure was just getting out of the gate. GCP was still very small. What they can tell you is that things have changed quite dramatically right now.

I mean, we have almost 13% of our customers who are running multi-cloud deployment, applications running on AWS, Azure or GCP for that matter. So multi-cloud is not just a buzzword. I think there is definitely some reality in multi-cloud.

The other thing that we saw in this report is that many of our customers are starting to adopt Kubernetes. Jeffrey, I mean, I'm sure somewhat surprising, right? I should say surprising and not surprising at the same time. Kubernetes is still a fairly new environment, and yet what we see is that one in five customers of Sumo Logic and AWS are already using Kubernetes.

What is also interesting is that when customers are thinking about multi-cloud deployments, you see a sharp adoption in Kubernetes. So if you think about a customer running on just AWS, as I mentioned before, one in five is using Kubernetes. On the other hand, if you think of a customer who's perhaps deploying on all three environments, AWS, Azure, and GCP, we almost see 8 in 10. That's 4 and 5, however you want to slice it, that use Kubernetes.

The take away from this is that Kubernetes is becoming this sort of grand equalizer and a key enabler of multi-cloud deployments. So if you are an enterprise architect today and if you want to build an application that is no longer beholden to one cloud AWS, Azure, or GCP, it will behoove you to think about using Kubernetes as your underlying platform, because that gives you the ability to port your application rather seamlessly from one cloud to the other.

Now, I got a few more observations, but in the interest of time, Jeffrey. Let me know how you want to proceed next.

[00:29:49] JM: No. I'd love to hear more. Tell me more. I mean, what I thought was interesting about this report, was there is a lot of information that was – Aggregates of I think – What? Like 100 or 150 companies? Something like that?

[00:30:03] KR: Oh! No. Let me definitely correct you on that. I mean, this is 2,000 companies, right? 2,000 accounts. It's rather a large and statistically, what's the word, accurate sample size. We've gone to through great lengths and pains to ensure that we only put out data for which we have a statistical significance in our data. So the data that we have presented in this report are quite – I would say they're quite the norm for most enterprises and something that every practitioner out there should definitely pay attention to.

But let me continue since you wanted to hear more. Point number three, and this is something that anybody who goes to AWS Reinvent, AWS's big show, will definitely relate to. So we all go to AWS Reinvent, and here is AWS announcing another 15 services, or 20 services for that matter, pushing the envelope of all these stuff that you can do in AWS.

So there's always this question. AWS is this relentless engine that is pushing innovation every year. How many people do adopt these new services that are coming at such a fast clip from AWS, right? So looking at our data, what we realized is that while AWS is indeed pushing out a lot of innovative capabilities, the core services that are adopted by many of these AWS customers are still very limited, right?

What we found out is that on average, and this is looking at over 1,500 customers so far who are using AWS. On average, we see that only 15 of these –150 AWS services are really adopted by a large plethora of these AWS customers. So customers go to AWS. This is sort of the hypothesis coming out of this data, this data point, is that customers like the innovation pace at AWS. Customers like the fact that AWS is pushing the envelope when it comes to releasing new capabilities, providing new services to enterprises. But then a vast majority of them are still in the early phases, still in the innings 2 to 3, so to speak, of adopting these services.

We saw a few customers, few enterprising I should say, who adopted many, many, many AWS services, but those are far and few in between. A large majority of these customers end up going to AWS for their infrastructure as a service platform, the EC2 service, the S3, the cloud formation, the RDS and IM, the identity management services, are still cutting their teeth at these basic services. The fact that AWS offers all these cutting-edge services obviously is a good carrot and a good icing in the cake, but the majority of these customers are still using just the core of the cake at this point.

Point number four, let's talk about serverless, and everybody is talking about serverless. Serverless has being proclaimed as the next programming paradigm that's going to change everything about the cloud and systems and how you use systems. What we are starting to see is that at least in the AWS environment, serverless has reached a tipping point.

What AWS's serverless implementation is called Lambda, and what we see is that the Lambda adoption has indeed gone up dramatically over the last few years. In 2017, which was the second year when we were doing this report, we saw a 12-person adoption of Lambda. Right now in 2019, we see almost 36- person adoption of Lambda. That means one in three enterprises that are running applications in AWS today are using Lambda or serverless in some way shape or form. So that is an interesting data point, and I think that's sort of tells you that people are really experimenting and deploying and implementing serverless technologies at a much faster flip now.

I have one last point, and then this is sort of switching gears from more of the architecture, let's say talking about security technologies within the cloud. Jeffrey, I'm sure you must heard that security and cloud security is perhaps the number one obstacle, or blocker, or challenge as enterprises think of adopting cloud.

There's some method to it. I mean, enterprise soc teams are used to seeing their infrastructure. They're used to counting their infrastructure. They're used to seeing it in their data center. That gives them the comfort of their data being secure within the confines of that data center. Now moving everything to the cloud does bring up these concerns that maybe my data is in a place that I don't control. What happens to security then?

So what we're starting to see is that, again, this is within the purview of AWS, a lot of the enterprises that are deploying their applications in the cloud and in AWS are starting to use many of these services that AWS offers in order to better secure their applications in the cloud. So AWS has these interesting technologies. I'll sort of rattle out a few names here. It's called Cloud Trail. Cloud Trail, think of it as providing the who, what, when, where, how of access of APIs and services within AWS.

AWS offers VPC flow logs, which gives you visibility into who is accessing your systems. AWS has a new technology called guard duty, which provides you visibility into the security state off your AWS applications. What we're starting to see is that many Sumo Logic customers who are deployed in AWS are using these services and are using these services at a fairly significant rate. So that gives us confidence that enterprises are indeed paying attention to security as they are moving to the cloud. They are building in the right in capabilities. They are using the right

services that are offered by the cloud platform vendors themselves, and all of these obviously [inaudible 00:36:27] well for applications that are running in the cloud.

So soc teams, if there's anything that you should take away from this podcast is that security in the cloud is a little different, but it's definitely doable, and there are all the right capabilities available from vendors like AWS and also from third-party vendors, like Sumo Logic, that can ensure the security of your apps and infrastructures in the cloud.

[00:36:56] JM: I assume you've been to Reinvent.

[00:36:58] KR: Absolutely. We do go to Reinvent. There's a big contingent of Sumo Logic folks who go to Reinvent every year.

[00:37:03] JM: So I'm going for the first time this year, and just after seeing pictures and hearing stories, it sounds completely overwhelming. Do you have any advice for surviving Reinvent?

[00:37:18] KR: Oh man! Hydrate, hydrate, hydrate, and get out every now and then from the booths and from the sessions. So what can I tell you? That applies to Las Vegas as a whole. Reinvent has become the technology Mecca to be and to go to. I'm dating myself here, but I –

[00:37:38] JM: Is it more so than CES at this point?

[00:37:41] KR: CES is a different kind of show, right, Jeffrey? So CES is a bit more consumer-oriented. I'm sure if you are selling the latest IRT gadget or a phone, a CES or an MWC, the show in Barcelona is perhaps the show to be at.

If you're doing B2B applications, if you are targeting anybody who is in the dev world, the DevOps world, a site reliability engineer or perhaps even a security person. Reinvent is definitely the place to be. The vibe in that place, the energy in that place is off the charts. Hence where technology is today and where technology is headed for the next few years.

Rightly or wrongly, AWS is the 8,000-pound gorilla in the cloud space. They're building cool stuff. They have some amazing customers. It may sound like an advertisement for AWS, but the company does really innovate well. Sumo Logic was deployed in AWS. We like the innovation that they bring to the IT world and we like what our customers are able to do on AWS too.

[00:38:51] JM: It sure is amazing. I mean, it's a sign of what Amazon does right, that it's like as a byproduct of the accidental business that they created with AWS, they created like a billion-dollar conference industry.

[00:39:09] KR: Yeah. I mean, what is it? I hear they're on a run rate of 25 billion or whatever right now.

[00:39:16] JM: For Reinvent?

[00:39:17] KR: No. I meant AWS as a business itself. Not Reinvent. I don't know what's the actual revenue of –

[00:39:24] JM: I mean, you got to imagine, it's pretty profitable.

[00:39:27] KR: Oh! [inaudible 00:39:28] it's profitable given the number of vendors like ourselves who are there and the amount of money that we spend in terms of demonstrating, getting booths, sending people to that event. I'm sure it's a profitable venture for AWS.

But, look. I mean, at the end of the day, I mean, this is a place where you meet the people who are building your next gen apps, right? Whether it's Fender, the guitar player, who is building their apps on the next gen mobile app on AWS, whether it's Airbb, whether it's Netflix, or whether it's a tiny little company with like three guys in a garage. Today, nobody would question you if you were to put your credit card and build your first app on AWS. So AWS has, and Reinvent has become the show to go to to see what's going on in the IT world for sure.

[00:40:21] JM: All right, as we begin to wrap up, I have a few other high-level subjects I want to discuss. So the changing usage prototype of a log management system or an observability system, we touched on this a little bit earlier, but the fact that today not only do you have

engineers and operations people working with your log management tool, your metrics tool. You have data scientists, you have customer service people, you have product managers. You might even have marketing people.

As these different constituencies want access to this logging data or maybe they want access to the post-processed log metrics, how does that change the interface and how does it change the product design?

[00:41:12] KR: That's a good question. We're an early, I would say, phases of understanding it ourselves. But I think what Sumo Logic is starting to see is that many organizations, our customers are not just the typical people. I mean, the SREs and the DevOps people or the security people who use our systems, but they're also these product people, like exactly like you said, or customer support people who may not have the sophisticated understanding or our query language. May not have the ability to create dashboards and alerts and other things that you would take for granted if you're someone who lives and breathes within the Sumo Logic system.

So what we're starting to see more and more of is two modes of operation. One is this notion of what I call as templatized offer, right? Where I, as an administrator of the product, or as an expert user of the product, I may templatize what I may want to do in the product and then expose a few very simple to configure things within my product and then offer it to a larger swath of users.

So for example, imagine if i have let's say 300 customer support people, and I need to support these 300 customer support people and they need to look into Sumo Logic system every time somebody calls in with a problem and it will look up an ID and that they need to see what's wrong with the product of the of the system at that point.

Rather than expose the Sumo Logic query language to all of these 300 people who I'm sure they come with different knowledge and different skillsets, what we can then do is offer templatized interface or dashboard where all you have to do is drop the product ID. When you drop the product ID in, or the customer ID in, you get all the details about the customer, the

product, the usage of the product, the logs from the product, the errors, and the debugs, and the warning messages in the product itself.

So this templatized offering, essentially, in some sense, simplifies. I don't want to use the word dumps, because that has a pejorative meaning to it. But at least simplifies the access and the analytic that these non-experts can also get from our product.

The other approaches that I've also seen from our product is the ability for us to export data from our product into other systems. So we have customers who export our data into Tableau, into Click, into Looker, so that they can start to slice and dice this data in somewhat different ways, which maybe more in line with what these personas maybe used to end and may also want to see it. At the end of the day, we are in the business of getting insights from data and to the extent that there are other systems that can perhaps do some of these stuff. We're are more than happy to explore this data, to integrate this data into other systems.

We also see some of our data signs customers use notebooks, and Sumo Logic has a simple interface to export our data into the data science notebooks where you can do other forms of analysis on this data, which may not be core to Sumo Logic, which may not be the focus area of Sumo Logic itself.

Look, at the end of the day, I mean, we're in the business of getting best insights from our data for our customers and we'll do our utmost to make that happen.

[00:44:38] JM: Kalyan, I want to thank you for coming on Software Engineering Daily. It's been great talking to you.

[00:44:41] KR: Thank you very much, Jeffrey.

[END]