Speaker 1 (00:00):

This is episode 610 of the AWS podcast released on August 7th 2023.

Simon Elisha (00:09):

Hello everyone. Welcome back to the AWS podcast, Simon here with you. Great to have you back and I'm joined by not one by two amazing co-hosts who are both, funnily enough, co-located in the Big Apple, but we'll get to that for a moment. Firstly, get a Hawn Nguyen-Loughren. How are you doing?

Hawn Nguyen-Loughren (00:23):

It's fantastic and beautiful in the NYC.

Simon Elisha (00:26):

And Jillian Forde. You're also in the biggest of apples.

Jillian Forde (00:30):

That's right. And always excited for 51 updates that we're going to talk about today.

Simon Elisha (00:35):

It's a beautiful thing. Now Hawn, you're not normally in the Big Apple, you're normally in Atlanta, so what brings you to the city that never sleeps?

Hawn Nguyen-Loughren (00:45):

One of the most exciting event ever besides bring an event is the AWS Summit in New York City.

Simon Elisha (00:51):

New York, New York. It's a hell of a town and I'll stop singing from now on, I promise.

Hawn Nguyen-Loughren (01:00):

We're going to spread in the news.

Simon Elisha (01:00):

Oh God, this is going to go badly. Alrighty, we have lots of things to get through and of course we like to start with analytics and Amazon Redshift ML, which is a service that allows you to create, train, and deploy machine learning models using familiar SQL commands. And you all know how I love SQL. With Amazon Redshift ML, you can now leverage Amazon Forecast an ML based time series forecasting service without having to learn any new tools. You can now create and train forecasting models known as predictors from your time series data in Amazon Redshift, and use those models to generate forecasts about revenue, inventory, or demand in your queries and reports. And you also have the ability to generate forecast based on the probability metrics that you define. Amazon Redshift has also announced support of the qualify clause in the select SQL statement. And the Amazon Database Migration Service now supports Amazon Redshift Serverless as a target endpoint.

Simon Elisha (01:53):

With this new support, you can securely migrate your data to Redshift Serverless, where it can be stored, queried, and analyzed without the need to provision or manage infrastructure. AWS Lake Formation now supports delegation of LF Tag Management and AWS Glue crawlers now support Apache Hudi tables. Amazon on EKS now supports Apache Spark with Java 17, so you can now leverage Java 17 as a supported runtime to run Spark workloads on Amazon EKS and get all the benefits of the Java 17 language enhancements and performance improvements. And Amazon EMR on EKS has also added programmatic execution of Jupyter notebooks when running interactive workloads via managed endpoints. This enables you to run open source big data frameworks like Apache Spark on Amazon EKS, and Amazon EMR on EKS can set up the use of a managed endpoint. This available in preview to run interactive workloads and use your integrated development environments like EMR Studio. And Amazon QuickSight has added access customization options for small multiples and also a radar chart. And finally on this one, Amazon QuickSight has launched a unified color experience for analysis and dashboards.

Jillian Forde (03:06):

All right, two updates on application integration. Customers who use Amazon's symbol notification service for sending mobile push notifications can now host their applications in 12 additional AWS regions. Amazon's simple notification service can send mobile push notifications on your behalf to mobile devices and desktops using supported push notification services like Amazon device messaging, Amazon push notification service, and Firebase Cloud Messaging to Android devices among others. Using Amazon SNS, customers can send a message directly to one endpoint, send platform specific messages by destination type or send a message to multiple endpoints at once by subscribing those to a topic and sending messages to the topic. Amazon Location Service now supports a publishing track device position updates on Amazon EventBridge, allowing customers to leverage position updates to deliver features tailored to the physical location of tracked devices. Developers can create applications that show devices as they move on a map or store movement data in long-term storage to be used for purposes like asset movement insights, predictive analytics or compliance.

Hawn Nguyen-Loughren (04:30):

And on some exciting topics on compute. AWS Lambda can now detect and stop recursive loops in Lambda functions customer built event driven application using Lambda functions to process events from source like Amazon SQS and Amazon SNS. However, in certain scenarios due to resource misconfiguration or code defect, a process event may be sent back to the same service or resource that invoked the Lambda function. This can cause unintended recursive loops and result in unintended usage and costs for customers. With this launch, Lambda will stop recursive invocation between Amazon SQS and Amazon Lambda and Amazon SNS four 16 recursive calls. When a function sends the event to Amazon SQS or Amazon SNS using supported AWS SDK version or higher Lambda tracks the number of times a function has been invoked based on that event. If an event is invoked by the same triggering event more than 16 times, lambda will the next invocation and send the event to a dead letter queue or on failure destination.

Hawn Nguyen-Loughren (05:32):

If configured customers will also receive an AWS Health Dashboard notification with troubleshooting steps. AWS simplifies writing .NET 6 Lambda functions. We are happy to announce the general availability of a Lambda Annotations framework for .NET. This new programming model makes the experience of running Lambda function in C# feel more natural for .NET developers using C# source generator technology. The AWS Lambda Annotation framework provides annotation that enables developers to define a Lambda specific configuration and behaviors directly in their application code.

Developers no longer need to manually synchronize their Lambda functions defined in code with projects CloudFormation template. The framework does it automatically, which is pretty sweet. AWS Lambda and Amazon EventBridge pipes now supports enhanced filtering. Amazon Route 53 Resolver now available in AWS Outposts rack, which is super handy. AWS AppConfig agent simplifies Feature Flag and configuration used for Amazon EC2.

Hawn Nguyen-Loughren (06:32):

AWS AppConfig, a capability of AWS system manager announces support for AWS AppConfig agent on Amazon Elastic Compute Cloud, amazon EC2. AWS AppConfig allows customers to deploy faster and safer using Feature Flags and dynamic configuration, decoupling software releases from co-deployment. The AWS AppConfig agent is a client software that manages configuration data on behalf of customers. Previously, customers needed to manage their own polling and caching of data when running applications on Amazon EC2. Now the agent takes care of these tasks. With this update, the AWS AppConfig agent now has native support for following compute services, AWS Lambda, Amazon Elastic Container Service, Amazon ECS, Amazon Elastic Kubernetes, EKS, and Amazon EC2. Amazon ECS adds domainless gMSA authentication for Linux and Windows containers. Amazon Elastic Container Service, ECS, announces domainless Group Managed Service Account, gMSA, support for Linux and Windows containers running on Amazon EC2.

Hawn Nguyen-Loughren (07:36):

This integration allows application hosted on Amazon ECS or on EC2 to easily authenticate with Microsoft Active Directory, AD, to access network shared resources. With this launch, customers can run containers that require AD authentication without joining the ECS nodes to the domain, even during auto-scaling events. Group Managed Service Accounts, gMSA, is a managed account that provides automatic password management service principle name, the SPN management and the ability to delegate management to administrator over multiple servers or instances. This allows multiple containers or resources that share AD account without having to authenticate each container or resources individually or without having to access network shared resources such as SQL server hosts or file shares. Previously customer have been able to run ECS containers with gMSA by joining underlying nodes to an targeted AD domain. Now, customers can also use a built-in plugin on the latest ECS-optimized Windows AMIs that enable underlying nodes to retrieve gMSA credentials, portable user identity and plugin mechanisms, instead of hosting computer accounts.

Hawn Nguyen-Loughren (08:47):

Read these guides for a step-by-step walkthrough on how to use a feature on the Linux containers and Windows containers on ECS. AWS Fargate enables faster container startup using Seekable OCI. AWS Batch now supports Linux ARM64 and Windows x86 containers in AWS Fargate. This feature helps simplify the adoption of modern containers for AWS Batch customers by expanding their architecture option for scheduling Linux ARM64 and Windows x86 containers in Fargate compute environments. Support for ARM64 architecture also gives customers sustainable benefits of Graviton Instances in Fargate, which can help improve performance and price over comparable x86-based instances for a variety of workloads including high performing computing. Karpenter with a K, now supports Windows containers. AWS announced Windows support for Karpenter, starting with version 0.29.0, the open source node provisioning and management project.

Hawn Nguyen-Loughren (09:45):

Using Karpenter customers can now quickly scale their Amazon EKS clusters, running windows with right size Amazon EC2 Instances in response to changing application load. Karpenter helps customer optimize their operational overhead and costs during this process without manually adjusting the compute capacity of their clusters, over-provisioning compute to meet rapidly changing demands or creating multiple EC2 Auto Scaling groups, EC ASG. Karpenter helped customers optimize their operational overhead and costs during the process without manually adjusting the compute capacity of their clusters, over-provisioning compute to meet rapidly changing demand or creating multiple EC2 Auto Scaling group.

Jillian Forde (10:29):

Now onto customer engagement and we've got five updates from Amazon Connect. Amazon Connect launches CloudFormation support for routing profiles and queues. Amazon Connect Cases now provides case assignment. Amazon Connect launches, pre-defined Contact Lens conversational analytics metrics. Amazon Connect Wisdom is now available for chat agents. And finally, Amazon Connect launches APIs to programmatically delete routing profiles and queues.

Simon Elisha (11:03):

And that teams have been definitely busy working for their customers. Let's talk about where the data lives, databases. Amazon Aurora PostgreSQL compatible edition now supports the pgvector extension to store embeddings from machine learning models in your database and to perform efficient similarity searches. Now embeddings are numerical representations vectors created from generative AI that capture the semantic meaning of text input into a large language model. Pgvector can store and search embeddings from Amazon Bedrock, Amazon SageMaker and more. By using pgvector on Aurora PostgreSQL, you can simply set up operate and scale databases for your ML enabled applications. The pgvector extension allows you to build ML capabilities into your e-commerce media health applications and more to find similar items within a catalog. For example, a streaming service can use pgvector to provide a list of film recommendations similar to the one you just watched.

Simon Elisha (11:58):

Aurora machine learning enables you to add ML based predictions to applications via the familiar SQL programming language, so you don't need to learn separate tools or have prior machine learning experience. The pgvector extension is available on Aurora PostgreSQL 15.3, 14.8, 13.11, 12.15 and higher in AWS regions including the AWS GovCloud regions. And Amazon Aurora has also added some additional PostgreSQL support 15.3, 14.8, 13.11, 12.15 and 11.20 versions. My regular reminder to all of you patch your stuff, it is the best way to keep up to date from a security standpoint and you also typically get performance and function improvements. And speaking of improvements, the Amazon DocumentDB has launched index improvements to enable faster index builds on collections and the ability to view index bill statuses. Amazon DocumentDB index builds can now be sped up by up to 14 times, that's a lot of speeding up, when using parallel workers compared to using a single worker. The index creation process now uses two workers by default and you can configure the number of workers on Amazon DocumentDB version 4.0 and 5.0 Instance-based clusters.

Hawn Nguyen-Loughren (13:08):

Now we have developer tools. A couple of updates here. On July 18th 2023, Amazon announced quarterly security and critical updates for Amazon Corretto long-term support, LTS, and feature, FR, version of OpenJDK Corretto 20.0.2, 17.0.8, et cetera are now available for download. Amazon Corretto

is a no-cost multi-platform production ready distribution of OpenJDK. Amazon CodeCatalyst now supports workflows triggered by GitHub pull requests. AWS announces support for starting Amazon CodeCatalyst workflows based on pull request events in linked GitHub repositories. When a workflow is triggered by GitHub based pull requests, users will be able to see the name of the PR that triggered it in the CodeCatalyst workflows UIs and click a link that takes them directly to the pull request in GitHub. I've actually used Amazon CodeCatalyst and I love how smoothly it creates my environment project and CI/CD pipelines and deployment, which is super slick. And finally, AWS Proton introduces deployment history.

Jillian Forde (14:15):

Up next front and web and mobile. We're excited to announce significant improvements to the AWS Amplify JavaScript library by reducing the bundle sizes for our auth, storage, notifications and analytics categories. We've improved our code base so that you get leaner bundles resulting in faster load times and improved performance for your applications. You can benefit from these in performance by installing and updating to version 5.3.4 or beyond of the amplified JavaScript library. This is the first in a series of enhancements aimed at reducing the bundle sizes for apps built with a AWS Amplified JavaScript. A key focus for upcoming releases of the amplified JavaScript library is further bundle size reductions and improved tree shaking capabilities. I had no idea what that was actually until I looked it up related to JavaScript.

Simon Elisha (15:16):

Shake that tree.

Jillian Forde (15:18):

That's right.

Hawn Nguyen-Loughren (15:19):

It's amazing what it brings down when you shake that tree actually when you run the node for... But anyways, on the things of internet of things, a quick update here. AWS IoT Device Defender now supports monitoring of device disc duration metrics to monitor the duration of disconnected of each device. With this additional metrics, you can now track how long a device has been disconnected to learn whether is operating as expected. You can also configure alarms as predefined threshold levels and be alerted in the case of persistent device connectivity issues.

Simon Elisha (15:52):

Now let's talk all things machine learning. Lots of updates here. Llama 2 foundation models from Meta are now available in Amazon SageMaker JumpStart. This is a machine learning hub that offers pre-trained models built in algorithms and prebuilt solutions to help you get up and running fast. And you can now use Llama 2 foundation models with just a few clicks in SageMaker Studio or programmatically through the SageMaker Python SDK. Now, Llama 2 is an auto aggressive language model that uses an optimized transformer architecture, comes in a range of parameter sizes, seven billion, 13 billion and 70 billion, as well as pre-trained and fine-tuned variants. The tuned versions used supervised fine-tuning, SFT and reinforcement learning with human feedback, RLHF, to generate more relevant responses. Developers can leverage Meta's responsible use guide outlining their best practices for responsibly building each layer of the stack of a gen AI product and understand the importance of addressing risks associated with commercial use of LLMs.

Simon Elisha (16:51):

You can now derive the combined advantages of LlAMA 2 performance and ML Ops controls with SageMaker features like SageMaker Pipelines, Debugger, or container logs. The model will be deployed in an AWS secure environment under your VPC controls, helping ensure data security. Llama 2 is intended for commercial and research use in English. Tuned models are intended for assistant like chat, whereas pre-trained models can be adapted for a variety of natural language generation tasks. Amazon Translate is a neural machine translation service that delivers fast, high quality, affordable, and customizable language translation. Today we're enabling real-time translation of DOCX files. With this launch real-time document translation can synchronously HTML, TXT and DOC files while preserving the source document format. Customers can use the API via the SDK, CLI or the console to submit a translation request and get it straight back with a source formatting intact.

Simon Elisha (17:50):

We're also happy to introduce analytics on Amazon Lex. Now, if you've not come across Amazon Lex, it is a service for building conversational interfaces into any application using voice and text. With Amazon Lex, you can quickly and easily build sophisticated natural language conversations, virtual agents or IVR systems. And now we have the general availability of analytics on Amazon Lex and analytics gives access to rich insights and prebuilt dashboards related to Lex conversations, intense slots and utterances. This empowers developers with data-driven insights to improve the performance of your bot. Amazon Personalize now makes it easier to add columns to existing data sets so you can kind of even personalize your personalize.

Jillian Forde (18:31):

Management and governance. First up, AWS Config Advanced Query supports 65 new resource types. AWS PrivateLink announces integration with Amazon CloudWatch Contributor Insights. Amazon CloudWatch is excited to announce pattern. This is a new logs insight command, which enables customers to more easily identify trends and patterns in their logs. Customers often need to analyze thousands of log lines to identify relevant information. The pattern command uses AWS machine learning algorithms to automatically recognize patterns in log data, aggregate related logs, and then summarize thousands of log lines into a few easy to visualize groupings. Pattern helps customers quickly surface emerging trends. Monitor known errors, increase cost visibility by identifying frequently occurring log lines and more. And last Amazon Managed Grafana now supports in-place update for version 9.4.

Hawn Nguyen-Loughren (19:43):

Up next, we got media services, a couple of updates here. Amazon IVS now supports rendition filtering and higher frequency thumbnails for the record of S3 feature. With Amazon Interactive Video Service, Amazon IVS, you can now have the option to filter the specific HLS rendition stored in your S3 buckets as well as control the resolution and pixel count of thumbnail JPEG files while using the auto-record-to-S3 feature. These control allows customers to control the amount of data stored while recording to optimize storage costs. Additionally, customers can reduce the thumbnail generation interval down to one second, previously five second, limit to have a lower latency thumbnails for discovery or post-processing use cases. These enhancements apply to Amazon IVS standard and advanced channel types without additional costs. And finally, AWS Elemental MediaTailor now supports queues and ad tags in channel assembly.

Simon Elisha (20:41):

Now moving on to the concept of doing migrations and transferring data around. A quick update here, but an important one, AWS Mainframe Modernization is now available with a range of new capabilities for easier management and operation of the services fully managed runtime environments. These capabilities include role-based access, application level monitoring and application deployment updates that enable seamless user authorization control, greater application performance visibility and faster application deployment respectively for modernizing mainframe applications.

Jillian Forde (21:15):

Now networking and content delivery, Amazon CloudFront announces support for 3072-bit RSA certificates. Customers can now associate their 3072-bit RSA certificates with CloudFront distributions to enhance communication security between clients and CloudFront edge locations. RSA is an encryption algorithm widely used in digital certificates to secure internet communications through digital signatures and data encryption. Prior to this update, CloudFront customers could only use RSA certificates with 1024-bit or 2048-bit strength or an ECDSA P-256 certificate. ECDSA P-256 certificate provide greater security than 1024-bit or 2048-bit RSA certificates. Yet they might not be supported by legacy clients and devices. With the introduction of 3072-bit RSA certificates, customers can now achieve the same security level in CloudFront previously exclusive to ECDSA P-256 certificates.

Simon Elisha (22:32):

Am I dating myself here if I say that I remember when certificates were 128-bits?

Hawn Nguyen-Loughren (22:39):

A little bit.

Jillian Forde (22:39):

A little, yeah.

Simon Elisha (22:40):

Okay, then I won't talk to you about triple dares then. Okay. As you were.

Hawn Nguyen-Loughren (22:44):

And now for a quick but critical update for security identity and compliance. AWS WAF now supports URI path as an aggregation key for rate-based rules, providing customers with enhanced control and simplicity in managing request rates for URI paths. With this feature, customers can track request rates for any URI path and apply rules action based on the number of requests received.

Simon Elisha (23:08):

And wrapping up today with a bit of storage, Amazon S3 Inventory can include ACLs as object metadata in inventory reports and Amazon FSx for NetApp ONTAP, which is a service that provides fully managed shared storage built on NetApp's popular ONTAP file system is announcing two additional monitoring capabilities that let you monitor file system events and diagnose network connectivity. You can now access ONTAP event management system logs and collect packet captures on your file systems. And also the FSx for NetApp ONTAP now supports SnapLock, which is an ONTAP feature that prevents data from being modified or deleted for a specified retention by transitioning files to a right once read mini worm

state. And this allows you to help you meet regulatory compliance and protect business critical data from ransomware or just not accidentally delete stuff.

Simon Elisha (24:02):

And finally, Amazon FSx for NetApp ONTAP now supports IPsec encryption of data in transit. With this additional option to encrypt your data end to end FSx for ONTAP offers even more flexibility for you to protect your data. Plenty of cool and interesting updates there today. I think there's something for all of us to have a bit of a muck around with at some stage. Hawn, what about yourself? While you're in New York City, if people want to reach out to you about the podcast or about cloud stuff, how do they do it?

Hawn Nguyen-Loughren (24:31):

You can reach out to me, at my twitter handler @hawnsolo_1, or come out to the Java Center to come and check out what's going on.

Simon Elisha (24:38):

Fantastic. And Jillian, if folks need to understand where to get the best slice in NYC, what do they do?

Jillian Forde (24:45):

Oh, definitely Miss Jill Ford on Twitter. I've got plenty of food recommendations

Hawn Nguyen-Loughren (24:49):

I might have to ask you.

Simon Elisha (24:51):

Good idea. And if you want, send us feedback the old-fashioned way, awspodcast@amazon.com is the way to do it. Of course, on the website, you can also record vocal feedback, voice-based if you want to as well. And until next time, keep on building.