AWS Architecture Monthly

Advertising Technology
September 2021

aws

Editor's Note

This issue of the magazine includes solutions, case studies, videos, and blogs that illustrate how companies are using AWS to store, automate, and scale huge volumes of data, identify consumer trends, improve workloads, and most importantly, serve their customers well.

We'd like to thank our experts, Gerry Louw, Worldwide Tech Leader for the Advertising & Marketing Technology Industry, and Clark Fredricksen, Head of Worldwide Marketing for the Advertising & Marketing Technology Industry, as well as our AWS Media Series Technical Editor, Bonnie McClure.

Please give us your feedback! Include your comments on the Amazon Kindle page. You can view past issues and reach out to aws-architecture-monthly@amazon.com anytime with your questions and comments.
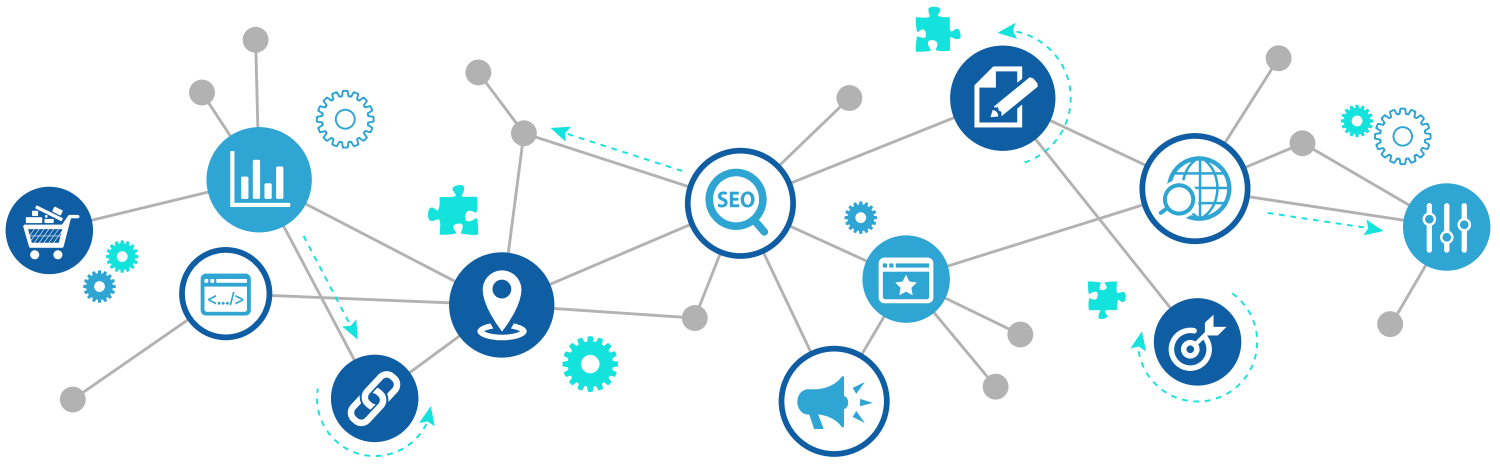
Jane Scolieri, Managing Editor

Table of Contents:

Notices

ASK AN EXPERT

Gerry Louw, Worldwide Tech Leader, Advertising & Marketing Technology Industry
Clark Fredricksen, Head of Worldwide Marketing for the Advertising & Marketing Technology Industry

What are the general architecture pattern trends for advertising and marketing in the cloud?

The advertising and marketing industry is in a major transition. The deprecation of consumer identifiers such as third-party cookies and the expansion of data privacy laws like General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) have disrupted decades-old technology. As a result, industry leaders like The Trade Desk, Nielsen, and Publicis Media accelerated their use of the cloud to develop new, interoperable standards for consumer identifiers and analytics. These standards support personalized advertising, measurement, and attribution. This shift drove customers like TripleLift and FreeWheel to invent new technologies to create audience segments,

forecast inventory, predict attribution, and identify contextual signals that improve ad relevance. It has also driven expanded use of first-party data collection and customer data platforms (CDPs) to support personalized advertising without reliance on third-party identifiers. We see several trending architectural patterns as a result:

- **Scaled data pipelines** processing first-party data for advertising intelligence use cases, such as ad measurement and event logging. Often customers are adopting serverless containerized compute to reduce the time and increments used for highly variable (spikey) ingest patterns. We also see customers adopt a stateless, streaming log analytics pattern to enable the use of Amazon Elastic Compute Cloud (Amazon EC2) Spot Instances to reduce costs and automatically scale.

- **Customer data platforms** that unify first-party data into "golden" user records with identity resolution for anonymized and known data. They also allow for audience segmentation, analytics, and the activation of defined audiences. Industry customers with scaled, real-time datasets typically adopt NoSQL databases like [Amazon DynamoDB](#) or [Aerospike](#) in these cases.

- **Shared data lakes and data clean rooms** with permissions that enable privacy-safe data sharing and collaboration between entities, without revealing underlying raw data, and in compliance with data sovereignty laws.

- **Machine learning pipelines** that take advantage of [Amazon SageMaker](#) to reduce heavy lifting through automation of feature engineering, training, and deployment.

When putting together an AWS architecture to solve business problems, specifically for advertising and marketing customers, what are some of your considerations?

The first priority for most customers is to grow revenue. The most important consideration in achieving this is, "How do we improve time to market?"

We address this question three ways:

1. **Reduce development time spent on undifferentiated work**, like managing commodity services like databases, Kubernetes clusters, or streaming services. This frees up your engineers to develop new features and products.

2. **Automate your application** deployment and testing.

3. **Migrate your applications to the cloud.** The ability to deploy a new service, instance type, database, or toolset in minutes allows you to experiment faster and with lower risk.

Additionally, we must also consider that advertising and marketing is unique in that applications must achieve these three goals:

1. **Scale to millions of transactions.** Our applications must be able to scale up and down automatically, without limitation or decrease in performance, based on demand.

2. **Deliver performance as cost effectively as possible.** Optimal cost efficiency in the cloud is obtained with cloud-native architecture. This starts with being stateless, using Spot Instances, and extends to the use of serverless compute to further reduce cost.

3. **Normalize data so that it can be used for advanced machine learning and analytics.** This includes use cases like inventory prediction, media analysis, audience modeling, data clean rooms, or federated measurement.

Do you see different trends in advertising and marketing in the cloud versus on-premises?

Customers usually face the same challenges, but customers using the cloud can resolve challenges quicker with lower costs. We see customers in the cloud investing in these workloads:

- Implementing data lakes with [AWS Lake Formation](#) to allow central governance and data sharing.

- Investing in clean room services with solutions provided by the Amazon Partner Network.

- Investing in machine learning platforms to deliver optimization, segmentation, attribution, targeting, and personalization with integrated services and tools.

- Using machine learning services such as SageMaker and Amazon SageMaker Studio allows customers to establish a machine learning foundation in their organizations. These services automate their machine learning pipelines, which reduces the time from development to production.

- To reduce costs, cloud customers port their applications to Amazon EC2 instances with AWS Graviton2 processors. These are the fastest processors in the cloud and enable up to a 40% improvement in price performance compared to similar generation x86 instances.

On-premises infrastructure provisioning and application development are generally much slower, which impacts time to market significantly and reduces engineering capacity for product development and delivery.

A common practice for on-premises customers is to expand into new geographic regions using the cloud. This is especially true when industry customers onboard large advertising partners. They risk losing customers to competitors if they are unable to provide the required services in the required time frame.

We also see on-premises customers investing in refactoring their workloads to be stateless and to containerize their workloads. This optimizes their compute resources and prepares them for a future cloud migration.

What's your outlook for advertising and marketing, and what role will the cloud play in future development efforts?

We have two predictions about the role of the cloud for advertising and marketing:

1. Uncertainty in how advertisers will be able to market to consumers on a personalized, one-on-one basis will continue. The resulting loss of revenue will force publishers to adopt a walled garden approach, which will require users to log in to access their favorite websites and applications.

2. Increased legislation to protect consumers is likely. Using the current availability of third-party identifiers to build consumer profiles while results can still be attributed to individuals is critical. This also applies to building the capability to define, target, and activate audience segments. Transforming organizations that are dependent on third-party cookies to be effective requires immediate and urgent action.

We have four recommendations about the role of the cloud for advertising and marketing:

1. **Stop moving data.** Create a data lake that can be governed and shared in a single location with query federation and analytics initiated at the point of consumption. You'll be more compliant with data privacy regulations and reduce heavy lifting and transfer fees.

2. **Evaluate capabilities like Amazon SageMaker** to accelerate innovation and reduce heavy lifting for machine learning use cases like audience modeling and inventory forecasting.

With the pace of innovation in the industry today, customers are moving faster using these capabilities. This allows data science teams to spend more time on actual data science.

3. **Prioritize cloud-native architectures and instance types** if you are considering migrating to the cloud. Examples like Nielsen Marketing Cloud's serverless data pipeline processing highly variable (spikey) traffic at volumes of 250 billion events per day is a good indicator of what's possible.

4. **Experiment with Graviton2**, which enables up to 40% price performance on comparable x86 instance types as a primary cost-optimization and performance driver.

Anything else to add?

Visit our industry website at https://aws.amazon.com/advertising-marketing/ to learn more, and see additional customer examples. Reach out to me — louwgerr@amazon.com — if you want to connect and ask questions.

ABOUT THE EXPERTS

Gerry Louw is head of worldwide solutions architecture for the advertising and marketing technology industry at AWS. Gerry has spent his career in advertising technology and media, with previous roles as Chief Technology Officer at Smaato, SVP of Engineering at RhythmOne, and CIO at VMS.

Clark Fredricksen is head of worldwide marketing for the advertising and marketing technology industry at AWS. Before joining AWS, Clark spent a decade at research firm eMarketer, where he sat on the company's executive management team and held leadership roles in marketing, product management, and communications.

This is My Architecture
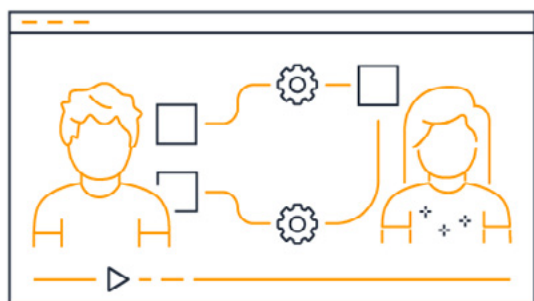
A technical video series that showcases unique or innovative cloud architectures
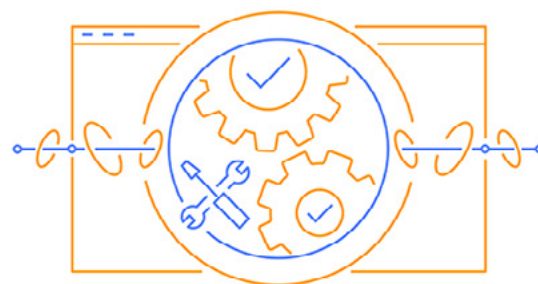
How to Build This

A video series designed for builders of all skill levels to start building with AWS

CASE STUDY

Nielsen Processes 250 Billion Ad Events per Day Using Serverless Computing on AWS

2021

[Nielsen Marketing Cloud](#), part of [The Nielsen Company](#), a global measurement and data analytics company, leverages Amazon Web Services (AWS) to process hundreds of billions of advertising measurement events per day. According to Matthew Krepsik, global head of analytics at Nielsen, the company has been able to scale up and down its platform to right size the compute needed to support its advertising and publishing customers. "It helps us democratize the work we do in the measurement space and provide access to all advertisers and platforms, no matter how big or small they are," says Krepsik. Nielsen is shifting its attribution business to be 100 percent cloud native and leveraging [Amazon EMR](#) to achieve up to 20 percent daily efficiency for its overall compute utilization. This helps Nielsen "drive down runtime and drive faster insights and data flow back to our clients," says Krepsik.

Nielsen led two AWS re:Invent 2020 sessions on using AWS to handle workload demands. Learn how Nielsen built a serverless, containerized data pipeline using [AWS Lambda](#) and [Amazon Elastic Kubernetes Service (Amazon EKS)](#) to process 250 billion events per day in [this 2020 re:Invent presentation](#). Then, discover how Nielsen uses Amazon EMR as a data mart in [this follow-up re:Invent presentation](#).

[*Read case study online*](#)

BLOG

Watch the re:Invent 2020 Sessions for the Advertising and Marketing Technology Industry

by Clark Fredricksen

We were lucky to share some amazing customer stories at re:Invent 2020 focused on the advertising and marketing technology industry. With all of the [announcements](#) and [keynotes](#), you'd be forgiven for not having time to watch each industry-specific session. That's why we've compiled all of the sessions specific to advertising and marketing technology for you to watch in this post—as well as an [industry playlist](#) for you to see everything in one place. Enjoy!

Breakout sessions

How Salesforce CDP unifies consumer data at exabyte scale

Discover how Salesforce CDP enables marketers to know everything about their consumers at massive scale by using AWS. Get an inside look at how Salesforce builds on AWS to process exabytes of data, resolve known and anonymized IDs into highly accurate user profiles, and deliver audience segments across activation channels. Learn how one of the largest software companies in the world thinks about systems architecture, network topology, security, and elastic computing in order to achieve performance, cost efficiency, and reliable performance for its data platforms.
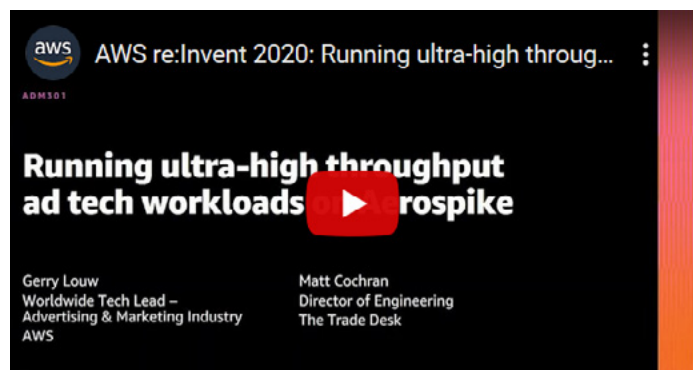
The Trade Desk — Running ultra-high throughput ad tech workloads on Aerospike

Dive deep into the AWS high-performance playbook for running ad tech workloads in the cloud using Aerospike, a low-latency NoSQL database platform. See how The Trade Desk implements Aerospike on AWS to support millions of queries per second at the edge for real-time bidding and peak loads of 30 million writes per second in its cold storage of user profiles. Learn best practices for configuring Aerospike on AWS for ad tech workloads, including recommendations for data engineering, architecture, and Amazon EC2 instance types for performance and cost efficiency.

FreeWheel, a Comcast Company – Distributed machine learning for digital video and TV ad serving

Discover how FreeWheel, a Comcast company, uses Amazon SageMaker to predict advertising inventory for digital video and linear TV months in advance for billions of ad serving records per day. Learn how FreeWheel built an end-to-end distributed ML pipeline for long-range, time-series inventory prediction across audience segments, geographies, and media types at massive scale. Additionally, take away best practices and lessons that FreeWheel learned to improve accuracy, reduce training time, optimize costs, and avoid pitfalls for ad inventory prediction with cloud-based solutions on AWS.

Cox Automotive — Building the post-cookie identity graph for marketing

See a production example of how Cox Automotive combined data from millions of car shoppers across 40,000 auto dealers and brands to create an identity graph with Amazon Neptune. Learn how to bring together customer datasets, anonymized web events, ad data, purchase logs, and other data. Take away best practices on modeling graph data, querying with Gremlin, avoiding graph pitfalls, right-sizing compute instances, and achieving cost-efficiency at scale.

Nielsen Marketing Cloud — Processing 250 billion ad events a day with Amazon EKS and serverless

Learn how Nielsen Marketing Cloud built a stateless, serverless data pipeline processing 250 billion events per day for their data management platform. See how AWS Lambda and Amazon EKS with Spark can reduce costs on logging and event processing at incredible scale, and come away with pitfalls to avoid and powers to exploit when using serverless and containers. Take away tips and tricks on how to tame the serverless beast, handle containerized processing, untangle problems delivering to real-time consumers, and avoid bugs when building your own system.

Nielsen Marketing Cloud — How Nielsen built a multi-petabyte data platform using Amazon EMR

In this session, learn how Nielsen used Amazon EMR to build and operate its multi-petabyte data lake and date warehouse. Nielsen discusses the growing pains of building a data lake, explains how to avoid them, and shares Amazon EMR best practices to improve performance in order to gain insights, reduce the cost of operating analytics workloads, and improve operational efficiency. Nielsen also walks through how it performs data exploration, sets up and shuts down Spark clusters using Jenkins, manages batch workloads through Airflow DAGs, and writes queries using notebooks.

Integral Ad Science — Contextual targeting and ad tech migration best practices

Hear how Integral Ad Science (IAS) migrated to AWS to process 100 billion events per day for its ad verification workloads—including brand safety, fraud detection, viewability measurements, and contextual advertising—supporting multiple regions and languages. IAS shares how it migrated to the cloud, overcame common industry challenges, and provided best practices on running ad tech workloads in the cloud, including architecture, data engineering, machine learning, cost analysis, migration strategy, and recommended services.

Live sessions — AWS on Air

Publicis Media — Automating Audience Segmentation with Machine Learning

Patrick Houlihan, PhD, SVP Decisioning at Publicis Media discusses how Publicis Media used Amazon EMR and SparkML to build a machine learning pipeline that trains models with petabytes of audience datasets. He shares how Publicis provides media buyers with highly accurate recommendations on audience segments. Learn how Publicis leverages machine learning to augment job roles and pairs them with natural language processing for rapid consultancy as a service.

TripleLift — Inventing Programmatic Product Placement Ads with Amazon SageMaker

TripleLift presented in "AWS on Air 2020: Industry Live Advertising and Marketing" about inventing contextually aware native ad formats for OTT and Connected TV using Amazon SageMaker, Amazon Rekognition, and AWS Elemental MediaTailor. Check out the video below to learn their architectures and see how they used machine learning with AWS. Read more about how TripleLift uses machine learning for programmatic product placement in TV advertising with AWS.

We hope you enjoyed this year's sessions focused on the advertising and marketing industry. If you're hungry for more re:Invent content, check out the full list of re:Invent 2020 news here or rewatch Andy Jassy's keynote to see our biggest announcements. See you next year!

Read blog post online
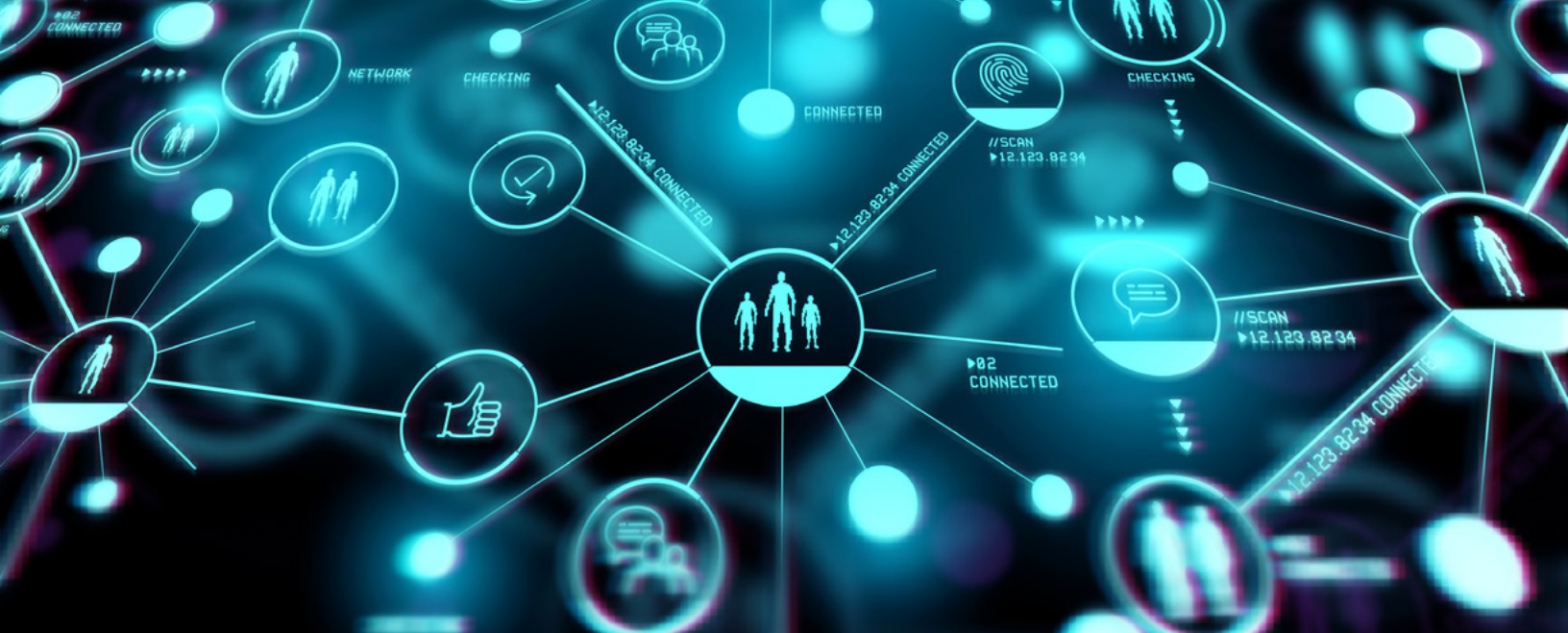
SOLUTION

AWS Streaming Data Solution for Amazon MSK

What does this AWS Solutions Implementation do?

The AWS Streaming Data Solution for Amazon MSK provides AWS CloudFormation templates where data flows through producers, streaming storage, consumers, and destinations. To support multiple use cases and business needs, this solution offers four AWS CloudFormation templates. Similar to the AWS Streaming Data Solution for Amazon Kinesis, the templates are configured to apply best practices to monitor functionality using dashboards and alarms, and to secure data.

Benefits

- **Automated configuration**
  Automatically configure the AWS services necessary to easily capture, store, process, and deliver streaming data.

- **Four template options**
  Choose from four different AWS CloudFormation template options. Test new service combinations for your production environment and improve existing applications.

- **Real-time use cases**
  Capture high-volume application logs, analyze clickstream data, continuously deliver to a data lake, and more.

- **Customizable source code**
  Customize the solution's boilerplate code, and then use the solution's monitoring capabilities to quickly transition from testing to production.

AWS Solutions Implementation overview

The diagrams below present the four AWS CloudFormation templates that you can automatically deploy using the solution's implementation guide.

Option 1 – AWS CloudFormation template using Amazon MSK
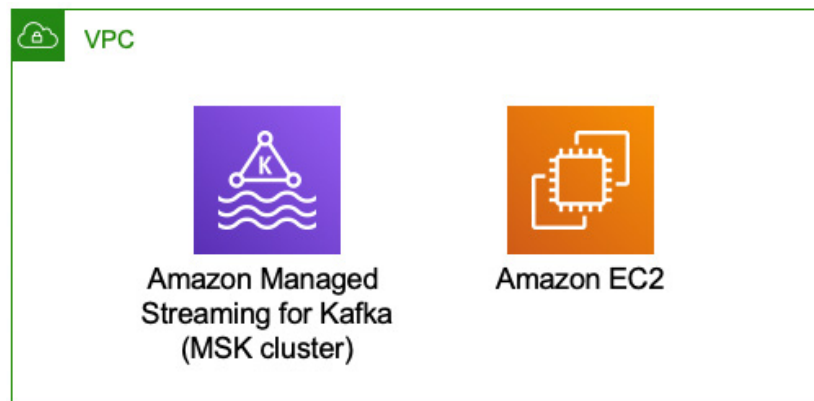
This AWS CloudFormation template deploys a reference architecture that includes the following:

1. An Amazon MSK cluster.

2. An Amazon EC2 instance that contains the Apache Kafka client libraries required to communicate with the MSK cluster. This client machine is located on the same VPC as the cluster, and it can be accessed via AWS Systems Manager Session Manager.
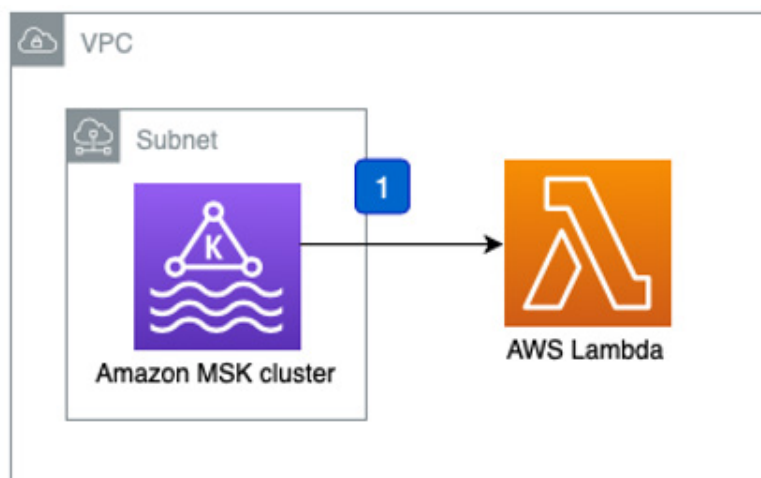
Option 2 – AWS CloudFormation template using Amazon MSK and AWS Lambda
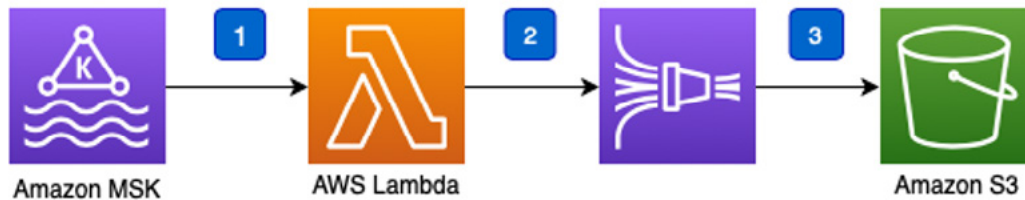

This AWS CloudFormation template deploys a reference architecture that includes:

1. A Lambda function that processes records in an Apache Kafka topic. The default function is a Node.js application that logs the received messages, but it can be customized to your business needs.

Option 3 – AWS CloudFormation template using Amazon MSK, AWS Lambda, and Amazon Kinesis Data Firehose

This AWS CloudFormation template deploys a reference architecture that includes the following:

1. An AWS Lambda function that processes records in an Apache Kafka topic.

2. An Amazon Kinesis Data Firehose delivery stream that buffers data before delivering it to the destination.

3. An Amazon Simple Storage Service (Amazon S3) bucket that stores all original events from the Amazon MSK cluster.

Option 4 – AWS CloudFormation template using Amazon MSK, Amazon Kinesis Data Analytics, and Amazon S3
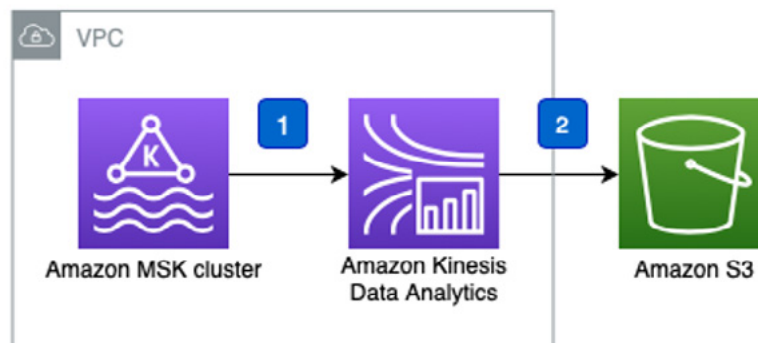
This AWS CloudFormation template deploys a reference architecture that includes the following:

1. An Amazon Kinesis Data Analytics application that reads events from an existing topic in an Amazon MSK cluster.

2. An Amazon Simple Storage Service (Amazon S3) bucket that stores the output of the demo application.

View solution online

Integral Ad Science Goes All In on AWS to Scale Ad Verification Globally

2021

Integral Ad Science (IAS) is a global leader in digital ad verification, ensuring that ads are viewable by real people in safe and suitable environments. The company's services help advertisers improve the quality and effectiveness of their campaigns, such as through tools that help place ads next to unobjectionable content for brand safety; verify that ads are viewed by humans, not bots, to avoid fraud; and measure that ads are in clear view of consumers.

Until 2019, IAS operated data centers in the United States and Europe and hybrid on-premises and cloud models in Australia, Singapore, and Japan. The company had used a hub-and-spoke model, in which all data was pulled from those dispersed centers into a New Jersey data center to be analyzed. To support peak traffic, IAS had to buy more hardware, which required weeks to provision and remained unused during slower seasons.

Due to its continued global growth, the company wanted faster, more efficient architecture, so it chose to migrate to a cloud-based infrastructure on Amazon Web Services (AWS). IAS began its migration in early 2020, moving its global edge services first, followed by reporting platforms; extract, transform, load pipelines; continuous integration / continuous delivery environments; event collection; and data science tooling.

Migrating to AWS to Scale and Accelerate Innovation

Prior to the migration, AWS Premier Consulting Partner Onica helped IAS perform a readiness assessment to understand which applications required architectural changes and which applications could migrate in a lift-and-shift format. After this assessment, IAS scheduled immersion days to train staff on using AWS.

IAS runs its applications using a suite of AWS services, including Amazon Elastic Compute Cloud (Amazon EC2), a web service that provides secure, resizable compute capacity in the cloud, and Amazon EMR, an industry-leading cloud big data solution for processing vast amounts of data using open-source tools. "On AWS, we have access to several services and Amazon EC2 instances, and we can test new technologies quickly," says Danny Rathjens, senior director of technical operations at IAS. "This helps our business remain agile, and it's a key factor in working with infrastructure as code in an optimal, modernized way."
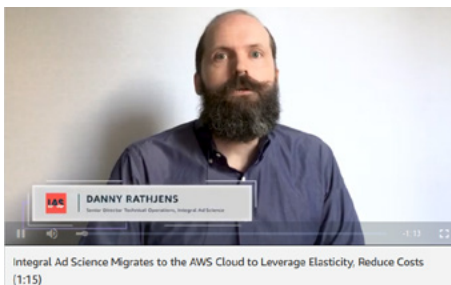
Integral Ad Science Migrates to the AWS Cloud to Leverage Elasticity, Reduce Costs (1:15)

"Having access to AWS resources, subject-matter experts, and solutions architects helped us a lot in our migration path."

Danny Rathjens
Senior Director of Technical Operations, Integral Ad Science

Since its migration, IAS has improved agility and cost efficiency, enabling it to launch in new regions and deploy products in hours instead of weeks.

Scaling to Handle Billions of Ad Events

IAS processes over 100 billion web transactions per day on average for ad verification—trillions of data events per month. It analyzes millions of pages per day for brand safety and provides real-time recommendations for bidding, responding to API calls in 10–50 ms. It also provides known contextual analysis scores in under 10 ms for 99 percent of requests.

"When an advertiser is going to serve an ad, it wraps the ad with a JavaScript snippet that makes a call to our system," explains Rathjens. When the ad is served, JavaScript also runs in the browser, sending out 16 PB daily on Amazon EC2 C5 Instances, which deliver cost-effective high performance at a low price per compute ratio for running advanced compute-intensive workloads. JavaScript captures and sends information about the ad to IAS's data lake on Amazon Simple Storage Service (Amazon S3), an object storage service that offers industry-leading scalability, data availability, security, and performance. "With all our data in Amazon S3—which has a very high level of failover and more ephemeral processing—we always have those resources available," says Rathjens. "We can run it in different locations for disaster recovery."

Once the data is pooled into the data lake, ephemeral Amazon EMR clusters run jobs to combine, enrich, and aggregate the data into a data warehouse, a workload powered by Amazon EC2 I3 Instances, the next generation of storage-optimized instances for high-transaction, low-latency workloads. "We're using a variety of instances for different workloads, and it's great to have that flexibility," Rathjens says. The transactions are analyzed for fraud and viewability using machine learning models built on H2O, an open-source machine learning and artificial intelligence solution that enables IAS to use a single tool to analyze its network and detect fraud. From the data warehouse, IAS provides ad campaign performance reports to its customers.

IAS reduced costs by 12 percent since migrating to AWS. "We consulted the AWS team about determining the cost-optimal setup that made the migration worthwhile," Rathjens says. Partly responsible for cost savings is Amazon EMR Managed Scaling, which automatically increases or decreases the number of instances or units in a cluster based on workload. IAS uses it for a mix of Amazon EC2 On-Demand Instances, Amazon EC2 Spot Instances, and Amazon EC2 Reserved Instances, as well as Savings Plans, a flexible pricing model that offers lower prices on Amazon EC2 instance usage.

Scaling on AWS helped reduce the need to spend up to 6 weeks provisioning hardware to meet peak traffic levels; now, IAS can provision new capacity in a few hours, increasing speed to market. And by reducing costs, IAS uses its additional resources to pursue innovative projects. IAS can spin up instances for rapid experimentation and testing without purchasing hardware. In-house staff no longer needs to update hardware, and the time staff previously dedicated to maintenance is now focused on core business.

IAS uses Elastic Load Balancing, which automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances and Aerospike—a flash-optimized in-memory open-source NoSQL database—as a distributed cache in each region. This architecture enables low-latency databases in each edge location, which improves the experience for advertisers and their customers. Further reducing latency is Amazon CloudFront, a fast content delivery network service that securely delivers data,

videos, applications, and APIs to customers globally. With low latency and high transfer speeds, IAS can use Amazon CloudFront to improve speed to integration. "We can onboard new clients even faster," says Rathjens. "The global scale and reach help us put our systems closer to our clients so that they get better performance."

Exploring Future Applications of Cloud-Based Architecture

IAS's migration to AWS has saved costs, increased agility, and strengthened fraud protection and disaster recovery. In the future, IAS plans to incorporate even more machine learning into its analytics processes and launch a new brand safety project on AWS.

The company is continuing to explore using AWS Graviton Processor –powered Amazon EC2 instances to reduce its costs. IAS is also using AWS to advance further developments, engaging the AWS digital innovation team to envision new products that aim to drive revenue and optimize the customer experience.

"Having access to AWS resources, subject-matter experts, and solutions architects helped us a lot in our migration path," Rathjens says.

About Integral Ad Science

Integral Ad Science (IAS) is a global leader in digital ad verification, ensuring that ads are viewable by real people in safe and suitable environments. IAS's mission is to be the global benchmark for trust and transparency in digital media quality for the world's leading brands, publishers, and platforms.

Benefits of AWS
- Reduced costs by 12%
- Processes 100 billion web transactions per day on average
- Provides known contextual scores in under 10 ms for 99% of requests
- Responds to API calls in 10–50 ms
- Removes responsibility of hardware maintenance and upgrades from staff
- Scales to enable simple experimentation
- Improves disaster recovery
- Onboards new customers faster

Read case study online
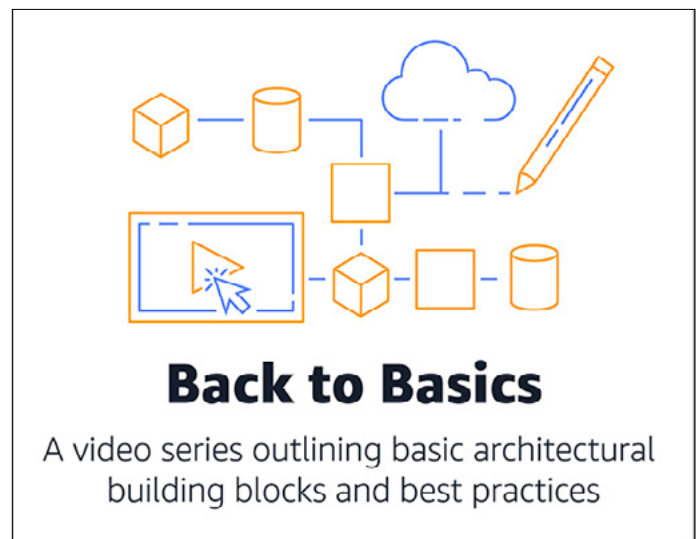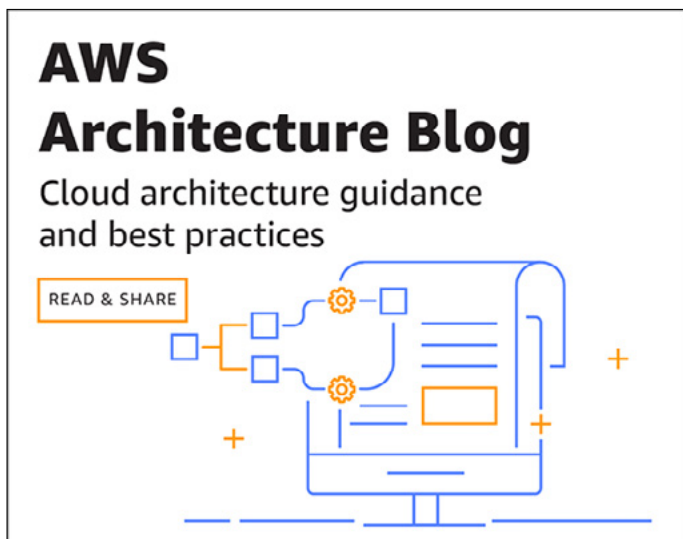
Using Machine Learning for Programmatic Product Placement in TV Advertising

by Christopher Kuthan, Akhil Aendapally, and Anita Snyder

TripleLift is a technology company rooted at the intersection of creative and media. The company's real-time advertising platform inserts products and brands natively into media across desktop, mobile, TV, and streaming video.

Research shows that native advertising that matches the look and feel of media experiences drives improved consumer recollection and impact. For example, product integrations showed a 50% increase in brand recall when paired with 30-second ad spots, according to a 2020 study on native advertising by MediaScience Audience Research Labs and TripleLift.

This post covers how TripleLift used AWS to create four new types of native ad products for streaming TV and OTT video. Each product fits seamlessly into the video experience, can be bought in real time, and is delivered with the same tracking capabilities as standard digital ads. We'll show how TripleLift uses Amazon Rekognition and Amazon SageMaker to perform video analysis and find available surface areas for ad placements. We'll also walk you through how TripleLift used Amazon SageMaker to train and deploy custom models for finding placements.

Video Analysis Workflow

TripleLift's video analysis workflow is triggered when we onboard publisher video content by uploading a video file into an S3 bucket (shown in the left side of the figure below). This kicks off the following steps:

1. A Lambda function triggers when a file is uploaded into the S3 bucket. The Lambda function starts our analysis pipeline by triggering an EC2 instance running the TripleLift OTT application from a fleet of instances. This Lambda function also serves to load balance and keep a record of the jobs assigned to each server machine.

2. After the EC2 instance receives the video details from Lambda function, it retrieves the video from the S3 bucket and calls the Amazon Rekognition Shot Detection API to break the video up into individual shots. Shots are the non-fungible currency of our Brand and Product Insertion formats, in which brands and products appear in any number of shots.

3. Once the Shot Detection results are available, the server will create new video files for each shot within the TV episode.

4. Using Amazon Rekognition video analysis (media asset search and indexing features), we analyze shots for various contextual elements that could be used to exclude candidate shots. For example, we avoid shots with images that brands may want to avoid—such as guns or violence.

5. We use Amazon SageMaker to run a custom OTT DeepStream model that analyzes the video for flat surfaces and selects surfaces that are appropriate for product and brand

insertions (referred to as "placements"). SageMaker provides an annotated video file and unique tracking ID for each detection, which are then used to minimize the search space for viable placement candidates, as well as compare similarities with previously successful placements. The files are then pushed to an S3 output bucket.

6. In parallel with the SageMaker models, the server also pushes some video frames to an SQS queue that triggers a Lambda function for Amazon Rekognition Image inference to analyze the image for additional points of classification, detection, face recognition and contextual information.

7. The output of Amazon Rekognition is pushed to S3 output bucket by Lambda function. All of the output data is written to a database for later access via an interface we provide our creative and business development teams.
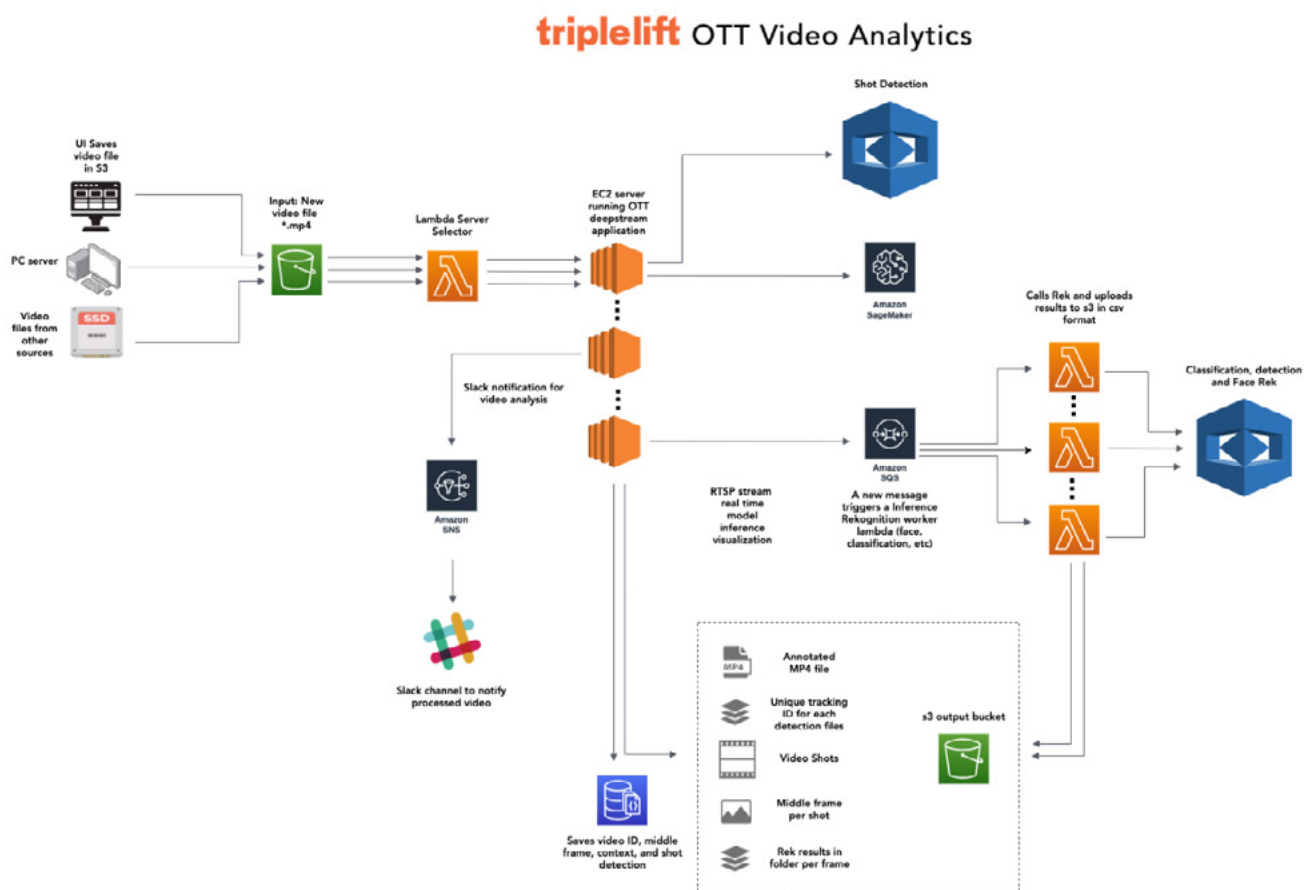
Inserting ads into shots and video streams in real time

After our machine learning pipeline identifies appropriate locations to place native ads, TripleLift's creative teams then go through a compositing process to insert brand images and ad units into frames. The team uses EC2 instances running VFX software such as Nuke to perform the compositing. Once brand assets have been composited into a shot, they're able to be swapped for other brand assets, truly showing the programmatic nature of this advertising.

The final step to deliver the ad units to a viewer requires server-side ad insertion (SSAI) technology. When an advertiser buys an ad unit through TripleLift's real-time advertising platform, TripleLift's servers call AWS Elemental MediaTailor or another SSAI service to request insertion. The ad-enhanced shot—which will include the ad creative natively composited into it— will be injected into the video stream and seen by the end user.

Results

We've seen several noteworthy results from developing this pipeline and these new products with AWS. At a business level, we've been able to expand into a new market and invent a never-before-seen type of ad format that's seamless inside video and has all the measurement capabilities of a display ad. With tools like SageMaker and Amazon Rekognition, our team can also process videos faster with fewer people—and we've been able to expand our capacity to provide meaningful ad placements more quickly than before. Overall, AWS allows us flexibility, control and convenience in building, deploying and maintaining our Content Analysis Workflow. Over the next year, we look forward to continuing to add to this pipeline with new analysis modules that build on our existing ones.

To learn more about this workload, check out TripleLift's AWS on Air presentation from re:Invent 2020.

Read blog post online

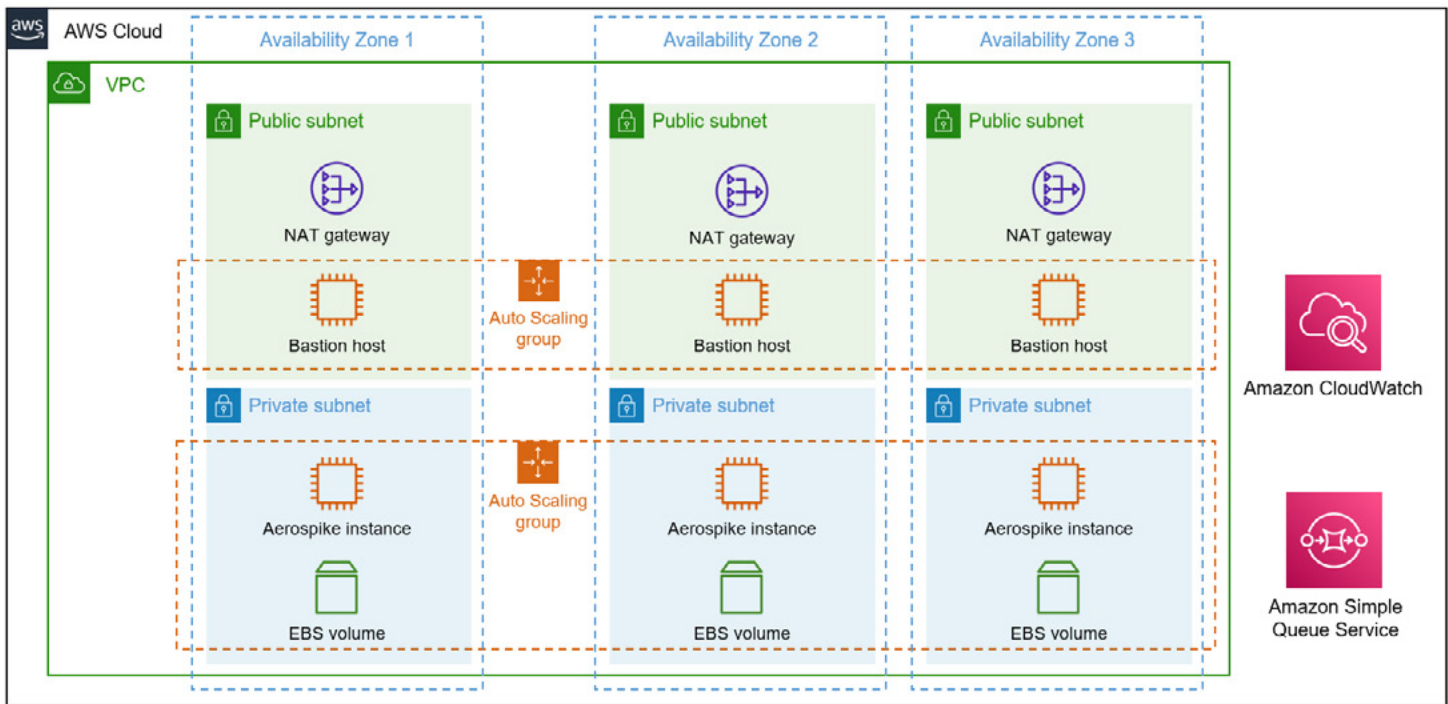## QUICK START

Aerospike Database Enterprise Edition on AWS

Predictable performance for globally distributed applications at petabyte scale

This Quick Start is for people who want to get started with Aerospike Database Enterprise Edition on the Amazon Web Services (AWS) Cloud using current best practices, high availability, and security-related features. Aerospike is a real-time NoSQL data-storage platform. It delivers predictable performance at petabyte scale with microsecond latency across billions of transactions.

When you use Aerospike with data-intensive, geodistributed applications, it performs like a cache with the persistence of a NoSQL data platform. It provides multi-site clustering, cross-datacenter replication, fast restarts, and rapid rebalancing.

Aerospike features a hybrid memory architecture where indexes, including the primary index and optional secondary indexes, are stored by default in dynamic random-access memory (DRAM) for fast access. You can also store data on solid state drives (SSDs) by configuring each namespace separately. Small namespaces can take advantage of DRAM, and larger ones gain the cost benefits of SSDs.

This Quick Start sets up the following:

- A highly available architecture that spans two or more Availability Zones.*

- A virtual private cloud (VPC) configured with public and private subnets, according to AWS best practices, to provide you with your own virtual network on AWS.*

- In the public subnets:
  - Managed network address translation (NAT) gateways to allow outbound internet access for resources in the private subnets.*

  - A Linux bastion host in an Auto Scaling group to allow inbound Secure Shell (SSH) access to Amazon Elastic Compute Cloud (Amazon EC2) instances in public and private subnets.*

- In the private subnets:
  - An Auto Scaling group of EC2 instances configured with Aerospike Database Enterprise Edition and Aerospike tools.

  - An Amazon Elastic Block Store (Amazon EBS) volume in each subnet.

- Amazon CloudWatch for logging and monitoring the Aerospike instances.

- Amazon Simple Queue Service (Amazon SQS) for Aerospike migration.

View quick start online
View deployment guide for details

The template that deploys the Quick Start into an existing VPC skips the components marked by asterisks and prompts you for your existing VPC configuration.

CASE STUDY

AppsFlyer Builds a Predictive Analytics Solution for iOS 14+ Using Amazon SageMaker

2021

The advertising industry has been upended by new standards for data privacy, cookies, and usage of ad identifiers. To improve the measurement of marketing campaigns in this privacy-centric landscape, marketing measurement company AppsFlyer used Amazon Web Services (AWS) to deliver PredictSK, a predictive analytics solution that uses machine learning (ML) to accurately predict mobile user lifetime value (LTV) of iOS SKAdNetwork campaigns based on anonymous data to prevent specific user identification.

PredictSK uses ML and serverless services from AWS, including Amazon SageMaker, which helps data scientists and developers prepare, build, train, and deploy high-quality ML models quickly. The product uses predictive modeling to avoid tracking users while providing higher accuracy on campaign performance, producing insights based on the first 24–48 hours of user interaction. The solution also protects user privacy in compliance with Apple's iOS 14 privacy changes for advertising.

Solving an Industry Challenge Rapidly on AWS

AppsFlyer has over 12,000 customers globally across ecommerce, financial services, gaming, and more, and its services are used in over 89,000 mobile apps. PredictSK uses ML to predict the LTV of mobile app users on IOS 14+ in a marketing campaign based on SKAdNetwork signals. The company began building its initial solution in the fourth quarter of 2019. "PredictSK significantly optimizes savings and campaigns for marketers," says Michel Hayet, senior product marketing manager for predictive analytics at AppsFlyer. "It gives our customers a lot of knowledge about their campaign performance much sooner than the traditional LTV evaluation cycle, which can take weeks or months to know if a campaign is working well."

Starting in 2018, data privacy regulations in the European Union and the United States have set new guidelines that restrict the types

of personal information that can be collected, shared, and used for marketing. Major browsers have deprecated third-party cookies used for tracking, targeting, and measuring data. In June 2021, Apple rolled out iOS 14, which features SKAdNetwork, a framework for privacy-preserving mobile install attribution, and App Tracking Transparency, a framework that prevents consumers' device identifiers from being used in ad tracking and targeting without their consent. "With iOS 14, we only have about 24 hours to look at each user's behavior and decide if the user will be valuable and whether to invest in the media source that provided the user," says Elena Levi, product team lead for PredictSK. "But with predictive analytics, all we need is 24 hours to get the long-term insights required."

AppsFlyer used AWS to adapt its predictive analytics solution to meet new industry requirements. The solution went from an initial idea to a working product in 6–8 months. On AWS, AppsFlyer cut time to production by 66 percent with the same number of staff, and in June 2021, the company prepared to onboard the first customer to PredictSK—1 month after App Tracking Transparency went into effect and 1 week after SKAdNetwork was implemented. AppsFlyer is the first mobile attribution provider to deliver predictive analytics abilities.

Accelerating Development and Deployment on AWS

To build its novel solution, AppsFlyer relied on Amazon SageMaker and various AWS serverless services that shortened the path from research to production, such as AWS Lambda, a serverless compute service that lets companies run code without provisioning or managing servers. "The serverless architecture on AWS reduces development time and maintenance," says Benjamin Winestein, senior software developer at AppsFlyer. "We can start with a small scale but then increase the scale with the confidence that everything will still work."

PredictSK gives mobile app users a prediction score on a scale of one to nine. The solution uses Amazon SageMaker automatic model tuning—which finds the best version of a model by running many training jobs on the dataset—to automatically adjust thousands of combinations of algorithm parameters to rapidly improve the accuracy of predictive ML models. Each model then deploys; or if it needs additional training, AppsFlyer repeats the training flow using AWS Step Functions —a serverless function orchestrator that makes it simple to sequence AWS Lambda functions and multiple AWS services into business-critical applications. To maintain accuracy, each model is retrained monthly using tens of gigabytes of data.

AppsFlyer creates a custom ML model for every app that enters PredictSK. It does this for both security and user privacy reasons; no information is shared between apps. Additionally, AppsFlyer must train different ML models for different use cases because user behavior varies depending on the type of app. For example, users won't behave on a healthcare app as they would on a gaming app. To meet that demand, AppsFlyer heavily relies on Amazon SageMaker multimodel

endpoints, which provide a scalable, cost-effective way to deploy large numbers of custom ML models. AppsFlyer runs Amazon SageMaker on Amazon Elastic Compute Cloud (Amazon EC2) P3 Instances —which deliver high performance compute in the cloud, powered by NVIDIA V100 Tensor Core GPUs— and uses p3.2xlarge instances.

In the scheduled prediction flow, AppsFlyer uses Amazon SageMaker batch transform to run inferences every 1–24 hours on large batch datasets using a simple API. The prediction results are then stored in a database and made available to customers. "For a normal advertiser that uses no predictive or advanced insights, it would take around 30 days to receive any kind of LTV insights for a user," says Hayet. "PredictSK cuts this time frame to as little as several hours."

PredictSK's near-real-time prediction flow operates on a serverless architecture and provides AppsFlyer's customers with a predicted user benefit score nearly instantly—in 10–30 ms per inference from request to return. It processes several hundred gigabytes of user data each day, at tens of thousands of events per second, and will scale to hundreds of thousands of events per second. The solution consumes the day's relevant events and saves that data to Amazon DynamoDB, a key-value and document database. It then uses AWS Lambda to prepare the data for inference and send it to an Amazon SageMaker

multimodel endpoint. The inference results then are written to another table on Amazon DynamoDB to be sent to customers.

Optimizing a Novel Solution on AWS

AppsFlyer next plans to streamline—and potentially fully automate—the process of how user acquisition managers use PredictSK insights to optimize their campaigns. It is also exploring distributed training using Amazon SageMaker, which could help shorten training time and scale to support larger datasets.

On AWS, AppsFlyer quickly reacted to a significant industry change with a solution that improves campaign performance accuracy and protects consumer privacy while providing actionable insights earlier than other existing solutions. "We are happy to provide a kind of service that none of our competitors are offering our customers now," says Hayet. "It's an amazing position to be in, but we're not resting. We keep progressing, improving, and working on this solution." Levi adds, "Predictive analytics changes the mobile marketing game in a number of important ways, and we're looking forward to offering the ecosystem more innovations moving forward."

About AppsFlyer
Founded in 2011, AppsFlyer provides software as a service for mobile-marketing analytics and attribution. Operating out of 20 global offices, AppsFlyer helps over 12,000 customers measure how users interact with brands through various apps, channels, and devices.

Benefits of AWS
- Implemented idea for ML algorithms in 6–8 months
- Produces insights based on the first 24–48 hours of user interaction
- Scales to hundreds of thousands of events per second
- Processes several hundred gigabytes of user data per day
- Retrains models monthly using tens of gigabytes of data
- Sees 10–30 ms per inference from request to return
- Predicts LTV in 1–24 hours compared to at least 30 days
- Cut time to production by 66% with the same number of staff

Read case study online
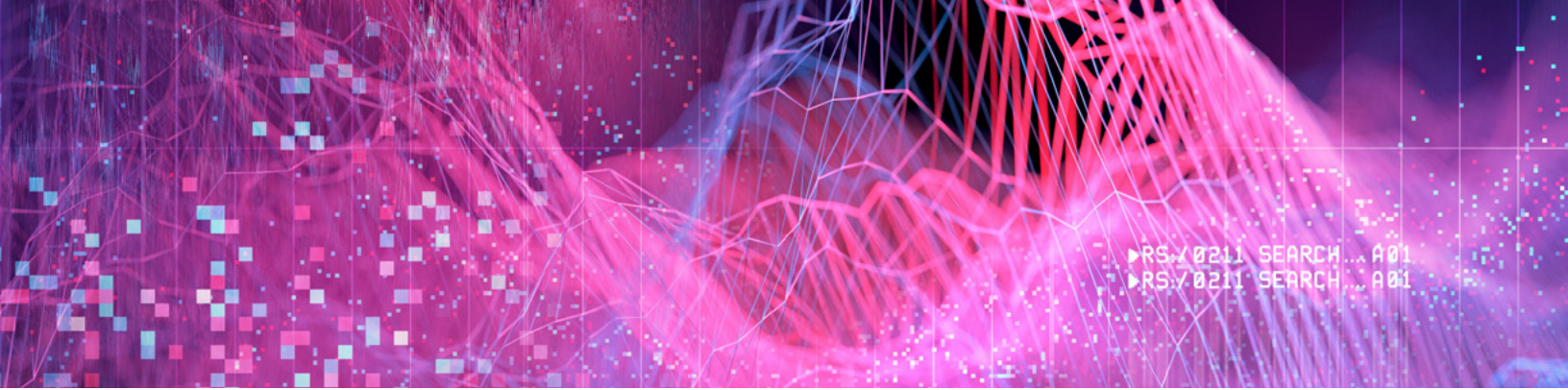
BLOG

Cox Automotive scales digital personalization using an identity graph powered by Amazon Neptune

by Carlos Rendon and Niraj Jetly

Cox Automotive Inc. makes buying, selling, owning and using cars easier for everyone. The global company's 34,000-plus team members and family of brands, including Autotrader®, Clutch Technologies, Dealer.com®, Dealertrack®, Kelley Blue Book®, Manheim®, NextGear Capital®, VinSolutions®, vAuto® and Xtime®, are passionate about helping millions of car shoppers, 40,000 auto dealer clients across five continents and many others throughout the automotive industry thrive for generations to come.

Auto dealers hosting their e-commerce websites on platforms like Dealer.com need innovative ways for targeting website visitors with relevant and personalized content. To deliver personalized content, dealers need tools to segment shoppers, display relevant advertisements, and trigger personalized email drip campaigns for different shopper segments.

Historically, website visitors were tracked across multiple domains using third-party cookies. Several browsers have already phased out third-party cookies with the remaining ones phasing out by 2022. This change will heavily impact the way Cox Automotive delivers personalized content to its online shoppers.

Cox Automotive's disparate business units were brought together via acquisitions and have evolved in silos. They have a growing need to combine cross-business unit data to create a holistic 360-degree view of the consumer household.

The Consumer Insights team is focused on providing shopper personalization services across brands. Their software serves consumers, car dealerships, and automotive OEMs.

In October 2019, the Consumer Insights team decided to use an identity graph approach to be less reliant on third-party cookies while also addressing the growing need for building a 360-degree view of households that can be utilized across business units. The team decided to use Amazon Neptune to address their identity graph needs.

Neptune is a fully managed graph database service that makes it easy to build and run applications using highly connected datasets. Neptune is a purpose-built, high-performance graph database engine optimized for storing billions of relationships and querying the graph with milliseconds latency. Neptune supports both the Property Graph and the Resource Description Framework (RDF) standard.
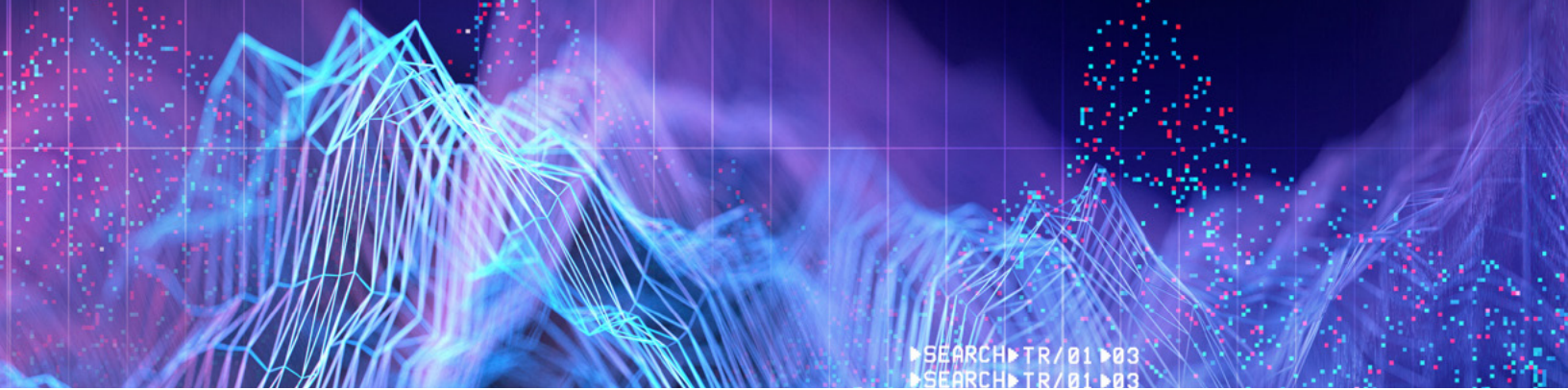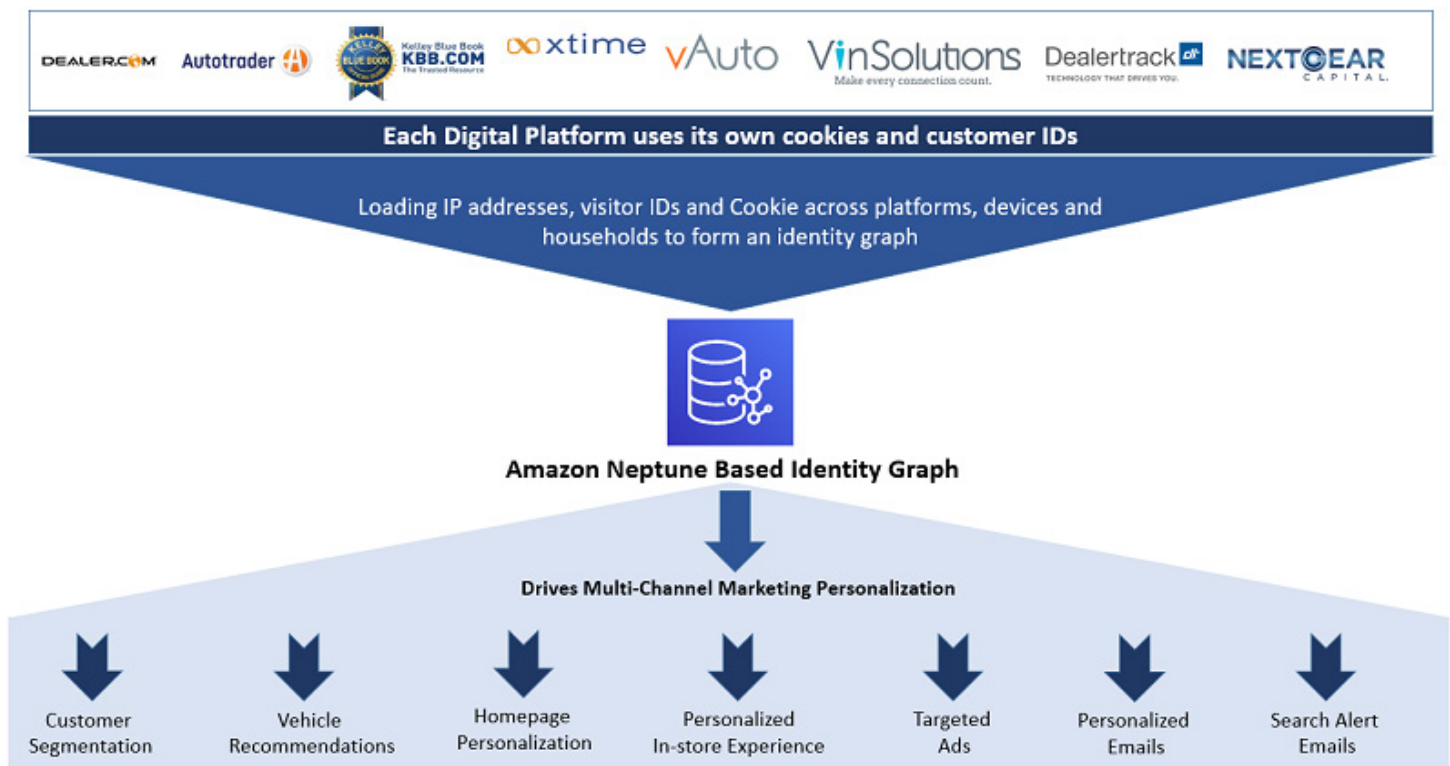
The Consumer Insights team uses several data sources in the identity graph for building personalization capabilities. These include:

- Cox Automotive's proprietary data, referred to as Pixall Data
- Vehicle inventory data
- Consumer browsing history
- Vehicle transactions
- Vehicle leads

Each Digital Platform uses its own cookies and customer IDs

Loading IP addresses, visitor IDs and Cookie across platforms, devices and households to form an identity graph

Amazon Neptune Based Identity Graph

Drives Multi-Channel Marketing Personalization

| Customer Segmentation | Vehicle Recommendations | Homepage Personalization | Personalized In-store Experience | Targeted Ads | Personalized Emails | Search Alert Emails |

As phase one, the Consumer Insights team built an identity graph that ties together consumer browsing history data with CRM data (leads and transactions). The following visualization outlines the team's approach to addressing current challenges while crafting a vision for an identity graph driving all aspects of multi-channel marketing personalization.

Reasons for choosing Neptune

The Consumer Insights team ran experiments to store the connected datasets in their identity graph with managed relational databases, key value stores, in-memory databases, and graph databases. The experimentation focused on performance and TCO. Two key technical aspects stood out in favor of Neptune:

1. Simplified data modelling – Traditional relational databases with their rigid schemas and relationships posed a challenge in building a data model for a graph problem. The Consumer Insights team took several passes at building the right data model. The data model was hard to scale as new vertices and edges were identified both in terms of query performance and difficulty of writing the desired query in SQL. Modeling the data as a graph mimicked the business use case. This was the first indication that Neptune was a natural fit for the use case.

2. Query performance – Neptune offered out-of-the-box query performance that met the team's needs and saved them time for doing any optimizations. Query performance scaled well as additional edges and vertices were added.
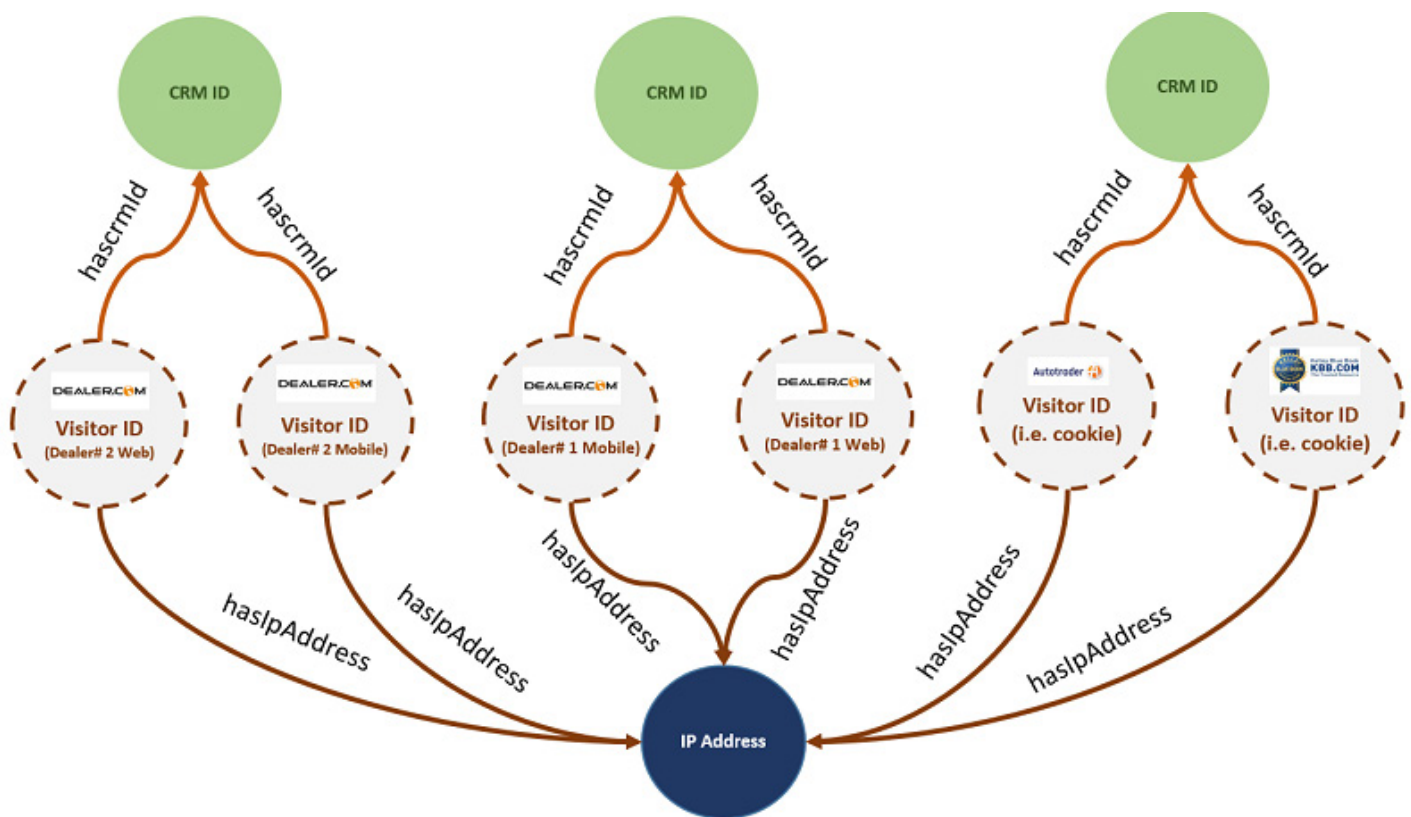
The identity graph

To demonstrate the data model simplification offered by Neptune, we provide the following visualization of the Consumer Insights team's identity graph, which shows the relationship between three entities (vertices) – cookie, IP Address and CRM ID.

This diagram shows the actual Gremlin query used in the identity graph. The following query shown facilitates the identity resolution of a visitor in simple steps. This query traverses the graph for a given visitor (request.uid), finds the CRM ID vertices with edges to the visitor, and finds all visitor IDs with edges to that CRM ID.

At this point, two visitor IDs have been identified in the preceding. From those two visitors, all IP addresses with edges to them are identified and from those IP addresses, all

visitor IDs with connected edges are found. In the preceding diagram, one visitor ID becomes six visitor IDs after walking the graph.

Business processes powering the identity graph

After choosing Neptune as the preferred database, the Consumer Insights team embarked on the next step of actually building the identity graph.

The identity graph is the heart of personalized marketing offering identity resolution. It's built on top of two distinct steps to create a 360-degree view of the consumer and the household: householding and lead mapping.
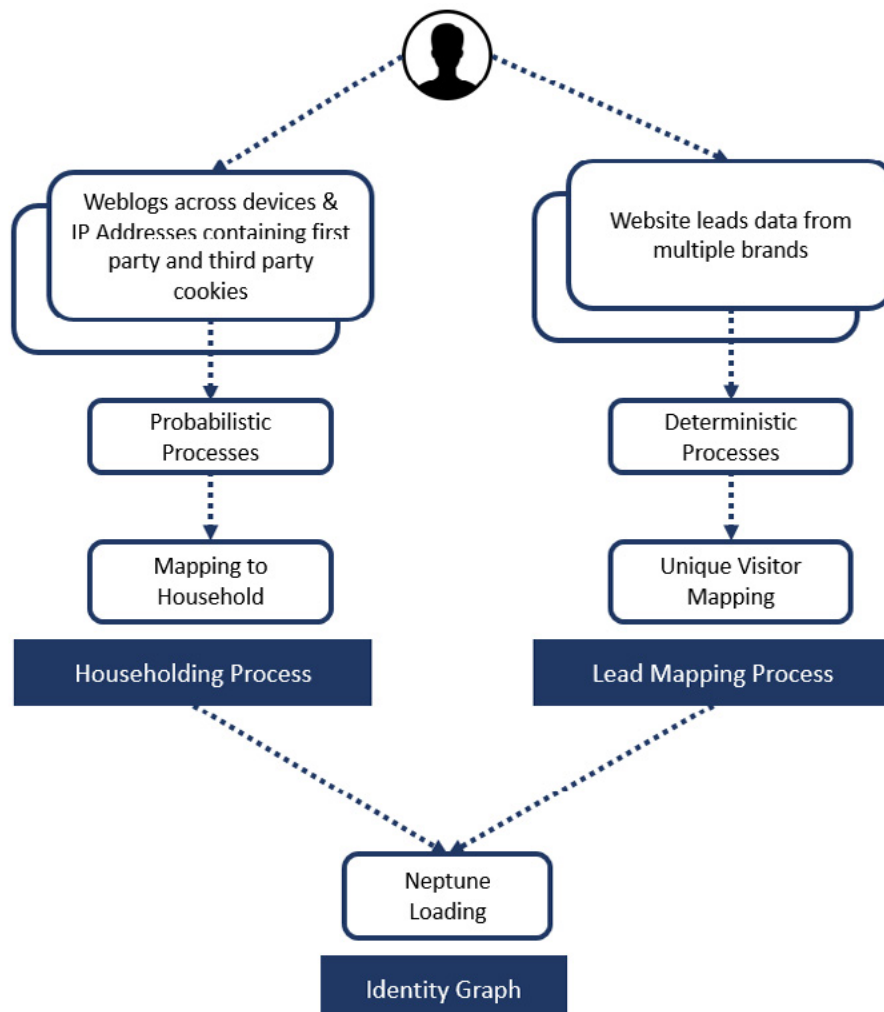
The following diagram illustrates the solution architecture of the identity graph and shows the steps for incrementally refreshing the graph (using Amazon S3, Amazon EMR and bulk loader), Multi-AZ Neptune read replicas and the built-in load balancer.
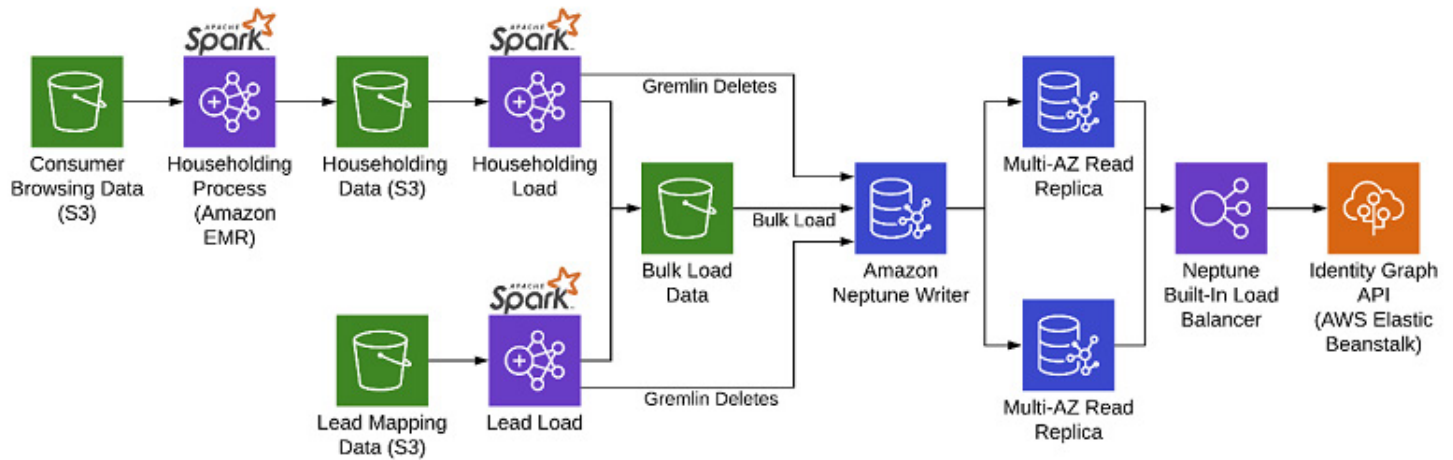
[Read full blog online](#)

CASE STUDY

Merkle Builds Customer 360s and Data Clean Rooms on AWS, Enables Personalized, Privacy-Safe Marketing

2021

Data privacy regulations have forced the marketing and advertising industry worldwide to find new systems that enable compliant collection and use of customer data. Merkle, which uses technology to help Fortune 1000 companies transform their customers' experiences, uses Amazon Web Services (AWS) to help marketers build solutions that protect consumer privacy and maintain compliance with regulations while enabling personalized advertising.

Using AWS, Merkle developed Merkury, an identity resolution solution that enables marketers to unify customer data and run analytics for audience segmentation, activation, and measurement in a data clean room—a privacy-safe data sharing workspace.

Helping Customers Rapidly Adapt to Data Privacy Regulations

Through data and digital transformations, Merkle helps companies across healthcare, financial services, manufacturing, and other industries to create personalized omnichannel marketing that transforms their customers' experiences. It offers services for creative, analytics, technology, consulting, and data management. Merkle manages more than 150 marketing databases and 3.7 billion customer records and has delivered 150,000 campaigns.

The onset of far-reaching data privacy regulations, such as the General Data Protection Regulation in the European Union and the California Consumer Privacy Act, led to a seismic shift in digital marketing. Major internet browsers have removed third-party cookies, which were once the standard source of audience data and identifiers that companies would use to inform their advertising and marketing campaigns. Now, marketers can only use consented and anonymized customer data, not personally identifiable information. That data must be collected and used in compliance with regulations. "Marketing teams have to migrate to solutions that enable them to do that successfully," says Sunil Rao, senior vice president of analytics at Merkle. "They need innovation in consented ways of reaching customers and of measuring the impact of marketing in a secure and privacy-compliant manner."

An AWS customer since 2016, Merkle decided to build Merkury on AWS. In 2020, Merkle became an AWS Select Consulting Partner, signed an AWS strategic collaborative agreement, and accelerated the development of a simple-to-use solution for its customers. "AWS brings the technology, and Merkle brings

"We use Amazon Redshift on a frequent basis to bring the data to a usable state. We run structured query language queries and point business analytics against it."

Ankur Jain
Senior Vice President and Global Cloud Practice Lead, Merkle

our marketing expertise so that we can create something of value to our end customers," says Ankur Jain, senior vice president and global cloud practice lead at Merkle.

Personalizing Marketing While Protecting Customers' Data on AWS

Companies can use Merkury to own, build, and activate marketing campaigns based on customer 360s—detailed profiles of their customers—without cookies. The solution instead uses an organization's first-party customer relationship management data and interactions such as website visits, logins, outbound email campaigns, and addressable media reach. The customer 360 practice informs cross-channel targeting, personalization, measurement, and more.

When a global entertainment company reopened its theme parks worldwide, for example, it wanted to elevate guests' experiences through targeted marketing and personalized offerings. The company used Merkury to build a 360-degree view of high-value guests, who are four to five times more valuable than the average guest, by capturing data such as where they stayed, which rides they went on, and whether they had seasonal or day passes. "We used AWS to bring those interactions together to build a customer 360, then used advanced analytics to personalize the experience on the website, in the park—anywhere that particular guest is interacting with the brand," explains Jain.

Another feature of Merkury is data clean rooms, which enable marketers to analyze and join shared first-party data with partners in a privacy-safe environment and to control how much of the underlying raw data is exposed to other parties. "Clean rooms remove a layer of overhead and restrictions when we put these analytical environments together," says Jon Regan, vice president of technology and data

management at Merkle. "They simplify the approach to compliance and security because there's no personally identifiable information in there."

On AWS, Merkle can avoid infrastructure maintenance and yearlong development times and can focus on delivering quality to customers. "Everything is prepackaged within the fully managed services of AWS," says Jain. "The pain of setting up the infrastructure, installing software, and managing the environment on a daily basis is taken away. Our innovation cycle is shortened." Customers of Merkle and of AWS can seamlessly purchase Merkury from AWS Marketplace, a digital catalog that makes it simple to find, test, buy, and deploy software that runs on AWS. "Developing this offering for every customer in a unique way would be very costly," says Jain. "For our customers on AWS, we can have them up and running with our solution within weeks, if not days, whereas it would take months if we did it from scratch."

The data for customer 360s resides in Amazon Redshift, a fast cloud data warehouse that makes it simple to gain new insights from data. "We use Amazon Redshift on a frequent basis to bring the data to a usable state," says Jain. "We run structured query language queries and point business analytics against it." Housing the raw data from customers' marketing systems is Amazon Simple Storage Service (Amazon S3), an object storage service that offers industry-leading scalability, data availability, security, and performance. To transform that data for processing, Merkle uses AWS Glue, a serverless data integration service that makes it simple to discover, prepare, and combine data for analytics, machine learning, and application development.

As real-time engagement becomes more important, Merkle uses AWS Lambda—a serverless compute service that lets companies run code without provisioning or

managing servers—for near-real-time data transformations. And Merkle enables near-real-time streaming using Amazon Kinesis, which makes it simple to collect, process, and analyze near-real-time streaming data so that companies can get timely insights and react quickly to new information. Merkle also uses artificial intelligence and machine learning services from AWS. For example, it identifies objects within creative advertisements using Amazon Rekognition, which makes it simple to add image and video analysis to applications using highly scalable deep-learning technology.

Using AWS to Build Next-Generation Marketing Solutions

By using AWS, Merkle can offer its customers an innovative solution that enables them to seamlessly adapt to industry changes, protect consumer privacy, and continue to deliver personalized, relevant marketing messages. In the future, Merkle plans to use the sharing capability of Amazon Redshift to enable its customers to exchange data nearly instantly. Merkle is also working to build intelligent marketing solutions on AWS. Jain says, "We're planning several strategic initiatives to help brands take advantage of the power of AWS to build next-generation marketing solutions."

MERKLE

About Merkle
Merkle is a technology-enabled, data-driven customer experience management company that helps Fortune 1000 companies build and execute customer-centric business strategies. It operates in 25 countries and more than 50 offices globally.

Benefits of AWS
- Implements a solution for customers in days or weeks compared to months
- Cuts costs in development
- Shortens innovation cycle
- Simplifies security and compliance
- Removes burden of infrastructure maintenance from teams

Read case study online

VIDEOS

Eyeota: Scalable ML Model Lifecycle Management for Advertising Audience Discovery

Learn how Eyeota architected a scaleable ML pipeline on AWS, to drive improvements in ad-tech data quality. The solution utilizes Sagemaker, Glue, Airflow and Cloudwatch to orchestrate an ML model lifecycle that iteratively improves over time.
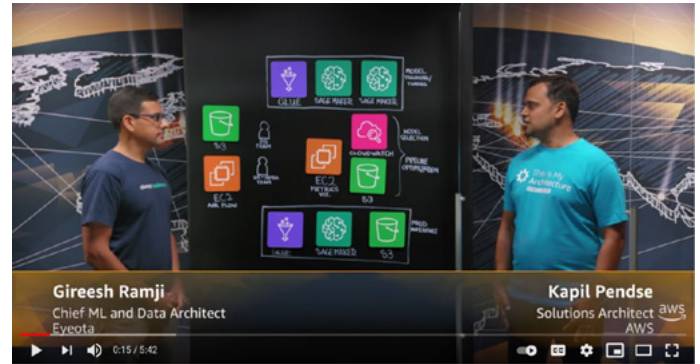
Nielsen: Processing 55TB of Data Per Day with AWS Lambda

Learn from Nielsen Marketing Cloud how to process 55TB of data per day while maintaining quality, performance, and cost using a fully automated serverless pipeline.

## How Salesforce CDP unifies consumer data at exabyte scale

In this session for executives and engineering leaders, discover how Salesforce CDP enables marketers to know everything about their consumers at massive scale by using AWS. Get an inside look at how Salesforce builds on AWS to process exabytes of data, resolve known and anonymized IDs into highly accurate user profiles, and deliver audience segments across activation channels. Learn how one of the largest software companies in the world thinks about systems architecture, network topology, security, and elastic computing in order to achieve performance, cost efficiency, and reliable performance for its data platforms.

FreeWheel, A Comcast Company Uses Amazon SageMaker for Digital Video and TV Ad Inventory Prediction

In this session for technology leaders, data scientists, and machine learning (ML) engineers, discover how FreeWheel (a Comcast company) uses Amazon SageMaker to predict advertising inventory for digital video and linear TV months in advance for billions of ad serving records per day.
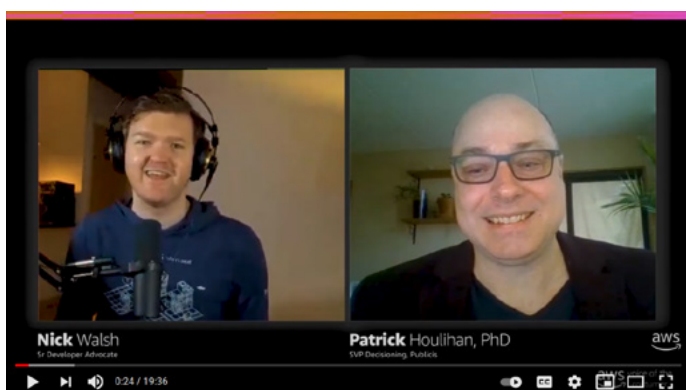
Publicis Media Automates Audience Segmentation Using Machine Learning on AWS

AWS On Air host Nick Walsh, Sr Developer Advocate and Patrick Houlihan, PhD, SVP Decisioning, Publicis discuss Automating Media-Buying with Machine Learning. When time-consuming, manual and intensive processes for ad planning and buying are automated using machine learning, planners can focus more on elements that benefit from human touch.
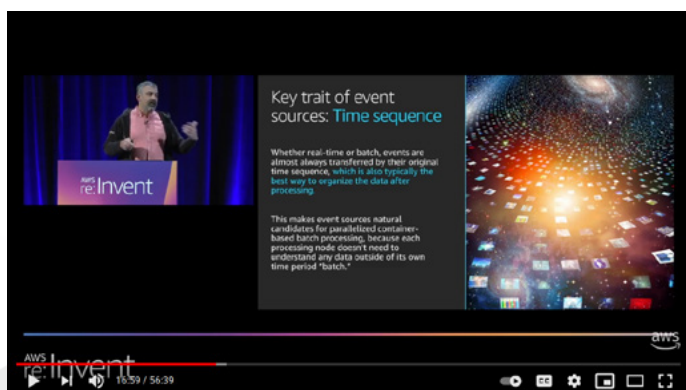
Annalect Uses Containers and Amazon Redshift Spectrum to Cost-Effectively Process Trillions of Events

Most companies in Advertising & Marketing run some kind of big data workload. But are they taking advantage of the latest cloud technology? In this session, learn how AWS customers can optimize data collection, analytics, and identity resolution using containers, serverless computing, and graph databases. Customers share detailed technical best practices for big data and advertising analytics at massive scale and low cost.