aws

ADVERTISING & MARKETING TECHNOLOGY

How four leading businesses optimized real-time advertising performance

Achieving optimal price performance and accelerating innovation with real-time advertising on AWS

Table of Contents

Introduction 3

Scale in seconds for real-time advertising 4

Success stories 8

- The Trade Desk transforms its real-time bidding strategy with AWS 9
- GumGum saves 62% on real-time bidding 10
- Nielsen processes 250 billion ad events per day 11
- The Trade Desk processes 30 million items per second for profile storage 12
- AppsFlyer prevents mobile ad fraud by processing 100 billion events per day 13

Next Steps 14

Appendix 16

The right ad, the right place, in real time

With unmatched opportunities to optimize compute performance and costs, AWS is the cloud standard for advertising platforms that buy, serve, and measure hundreds of billions of ads per day. Customers can move fast and spend less with the broadest set of solutions to migrate real-time advertising workloads, expand into new markets, and improve development velocity with cloud-native best practices and architectures.

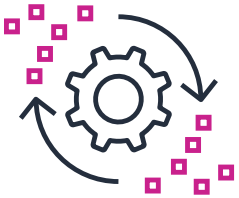
50%

Reduce your real-time advertising unit costs per billion requests by up to **50%**

70%

Decrease your advertising analytics workloads by up to **70%**

Scale in seconds for real-time advertising



Unmatched price-performance

In advertising analytics, companies need to ingest terabytes or petabytes per month of fine-grained data about digital events such as ad impressions or ad clicks—all while controlling costs. Leading organizations leverage the unmatched compute capabilities of the AWS Cloud to achieve the performance needed for advertising and audience data pipelines. They use services like **Amazon Kinesis** or **Amazon Managed Streaming for Apache Kafka** (**Amazon MSK**) for streaming data into a data lake on **Amazon Simple Storage Service** (**Amazon S3**) and **Amazon Elastic Compute Cloud** (**Amazon EC2**) instances with **Graviton2 processors**, which enable up to a 40 percent better price over comparable current-generation x86-based instances for their underlying compute.

Businesses traditionally run batch processes at periodic intervals for other types of data, such as advertising or marketing partners pushing web events, purchase events, or demographic data. Serverless technologies, which are seeing rapid adoption for data processing, can simplify the logic for these tasks and enable greater flexibility for handling large spikes in event data driven by partner ingress or advertising demand. Leading organizations use tools like **AWS Lambda**, **AWS Glue**, and **Apache Spark on Amazon EMR** to coordinate data flows and transformations—as well as containerized capabilities with **Amazon Elastic Kubernetes Service** (**Amazon EKS**) to scale quickly.

For real-time bidding (RTB) and ad serving, the performance and cost requirements are even more extreme. Ad-tech companies must respond to partners within tens of milliseconds or lose an opportunity to buy an ad. Ad-tech firms use Amazon EC2 with Graviton2 processors to spin up hundreds of instances in a single region in seconds to execute millions of queries per second cost-efficiently.

Ad techs also run NoSQL data stores such as **Aerospike**, **Amazon DynamoDB**, and **Amazon ElastiCache for Redis** to read and write data about user profiles, audience data, frequency capping, and other information at ultra-low latency and high throughput. Low-latency read-and-write performance is critical for these data stores since ad techs receive millions of bid requests per second and write new consumer data continuously into event logs and profile databases. RTB workloads also include bid logging and user information, which are streamed toward a centralized analytics pipeline.

Agility to experiment and expand

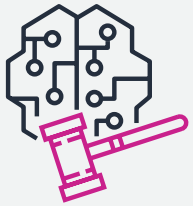
As connected TV and streaming media usage expand globally, industry leaders continue to accelerate dealmaking globally for new advertising inventory. This drives a greater need for agility to provision new real-time advertising infrastructures near consumers and partners in local markets.

On-premises infrastructure provisioning and application development can often be too slow to meet these requirements, impacting time to market significantly and reducing engineering capacity for product development and delivery. Leading organizations increasingly turn to cloud-based solutions to move quickly in new regions and set DevOps and developers free from costly build-outs and lengthy advance planning.

Cloud-native optimization

Companies moving to the cloud sometimes run into challenges meeting price-performance goals with lift-and-shift migrations that maintain colocation architectures designed for peak capacity instead of cloud elasticity.

Ad techs who invest in refactoring workloads to use cloud-native architectures and capabilities achieve up to 50 percent greater cost performance compared to colocation for real-time ad workloads. For example, companies use containerized ad serving and bidding instances with **Amazon EKS** and **AWS Auto Scaling** groups to scale compute up and down on an hourly basis throughout the day to reduce unit costs per billion ad requests. Stateless architectures for event logging enable customers to use **Amazon EC2 Spot Instances**, which enable up to a 90 percent discount on on-demand instances, to further reduce costs without worrying about interruptions.



Solution:

AWS offers multiple services, solutions, and partners that address many of these challenges for building advertising platforms running at petabyte scale and millisecond latency.

The following success stories demonstrate some of the ways companies have significantly improved performance, reduced latency, avoided bottlenecks, and optimized costs using AWS solutions for Advertising Platforms.

Success stories

The Trade Desk transforms its real-time bidding strategy with AWS

The Trade Desk is an advertising technology company that provides a self-service platform through which media buyers can purchase digital advertising. The company recently began shifting its real-time bidding workload—which receives 10 million queries per second, 800 billion queries per day, and requires an average response time of less than 15 milliseconds—to the AWS Cloud, building four new sites in 2019.

It used to take The Trade Desk up to six months to build out a bidding site for a new market. But, after migrating its bidding workload to AWS, the company has cut build time to less than a week and production deployment to about three weeks.

"Even though AWS is obviously a highly virtualized environment, we have much finer-grain control over that environment than we do in a managed physical site," said Zak Stengel, SVP of engineering at The Trade Desk.

There's much more agility (with AWS). If we need to tune a site, we can make changes much more quickly. We'll do much more iteration in a single week than what we would do in a full year with a managed physical site."

- Zak Stengel, SVP of technology, The Trade Desk

AWS services used:

Amazon EC2 Spot Instances: Run fault-tolerant workloads for up to 90 percent cost savings

Amazon EC2 Auto-Scaling: Add or remove compute capacity to meet changes in demand

AWS Real-Time Bidding in the Cloud Solution: Deploy in regions around the world with just a few clicks, minimizing RTB latency and cost

theTradeDesk®

GumGum saves 62% on real-time bidding

GumGum, Inc. (GumGum) is an advertising technology company that uses computer vision and natural language processing (NLP) to deliver contextually relevant advertising campaigns for brands and agencies around the world. To do this, the company runs an advertising exchange and supply-side advertising platform that conducts 30 billion transactions and processes 100 terabytes of data per day.

GumGum recently embarked on an organization-wide cost-optimization initiative, focusing on reducing costs for its compute-intensive advertising technology workloads such as real-time bidding, advertising analytics, and contextual advertising analysis.

The company swapped its CPU instances for GPU clusters on Amazon EC2, which are now processing thousands of events simultaneously, saving GumGum $12,000 per month. The company also prioritized using stateless architecture, enabling it to leverage Amazon EC2 Spot Instances, which let deployments take advantage of unused Amazon EC2 capacity at heavily discounted rates.

Amazon EC2 Spot Instances are an integral part of our architecture. Anything we launch these days, we design to run on Amazon EC2 Spot Instances."

- Vaibhav Puranik, SVP of engineering, GumGum

AWS services used:

Amazon SageMaker: Get the most comprehensive machine learning (ML) service, with purpose-built tools for every step of development

Amazon EC2 Spot Instances: Run fault-tolerant workloads for up to 90 percent cost savings

Amazon EC2 G4 Instances: Enjoy excellent price performance with the industry's most cost-effective GPU instances for ML inference and graphics-intensive applications

gumgum

Nielsen processes 250 billion ad events per day

Nielsen Marketing Cloud, part of The Nielsen Company, a global measurement and data analytics company, leverages 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 rightsize 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," Krepsik said.

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," said Krepsik.

What we get from AWS is a partner that…(is) willing to lean in and invest—not just to deliver a product, but to help us innovate collectively and move the ecosystem forward."

- Matthew Krepsik, global head of analytics, Nielsen

AWS services used:

Amazon EMR: Easily run and scale Apache Spark, Hive, Presto, and other big data frameworks

Amazon Elastic Kubernetes Service (Amazon EKS): Gain the flexibility to start, run, and scale Kubernetes applications in the AWS Cloud or on-premises

AWS Lambda: Run code without thinking about servers or clusters. Only pay for what you use.

Amazon S3: Store and retrieve any amount of data from anywhere

nielsen

The Trade Desk processes 30 million items per second for profile storage

The Trade Desk is a technology company that empowers buyers of advertising through its self-service, cloud-based platform.

"We have a phenomenal platform, in that we actually make the decisioning that decides which ads show up where on the internet," said Matt Cochran, director of engineering at The Trade Desk.

The company uses Aerospike, a low-latency NoSQL database platform running 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.

The company leverages AWS services, such as Amazon EC2 and Amazon EMR, and Aerospike to support both data storage use cases.

The elasticity that we get from using Amazon EC2 or Amazon EMR has been really helpful for allowing us to scale up and down as we need to."

- Matt Cochran, director of engineering, The Trade Desk

AWS services used:

Amazon EC2 Spot Instances: Run fault-tolerant workloads for up to 90% cost savings

Amazon EC2 Auto-Scaling: Add or remove compute capacity to meet changes in demand

Aerospike Database Enterprise Edition on AWS: Get predictable performance for globally distributed applications at petabyte scale

theTradeDesk®

AppsFlyer prevents mobile ad fraud by processing 100 billion events per day

In an effort to combat mobile app fraud, mobile attribution company AppsFlyer manages, measures, identifies, and blocks fraudulent installs with its Protect360 anti-fraud solution powered by AWS.

Primarily using the cost-effective compute power of Amazon EC2 Spot Instances—combined with savings plans or Reserved Instances and on-demand instances—AppsFlyer provisions its machine learning algorithm, which processes 100 billion events per day, and saves its customers $8.1 million per day by preventing fraudulent activity.

AppsFlyer is always looking to add new services, features, and functionality using other AWS services. “We are in charge of another product in AppsFlyer called Validation Rules, which enables advertisers to define rules about how to handle their own traffic,” said Ido Berkovitch, research and development director at AppsFlyer.

As it continues building Validation Rules, AppsFlyer expects to use Amazon DynamoDB, a key-value and document database that delivers single-digit millisecond performance at any scale.

If we reach a certain threshold, we can provision Amazon EC2 Spot Instances immediately, and we’re able to handle the load.”

- Ido Berkovitch, research and development director, AppsFlyer

AWS services used:

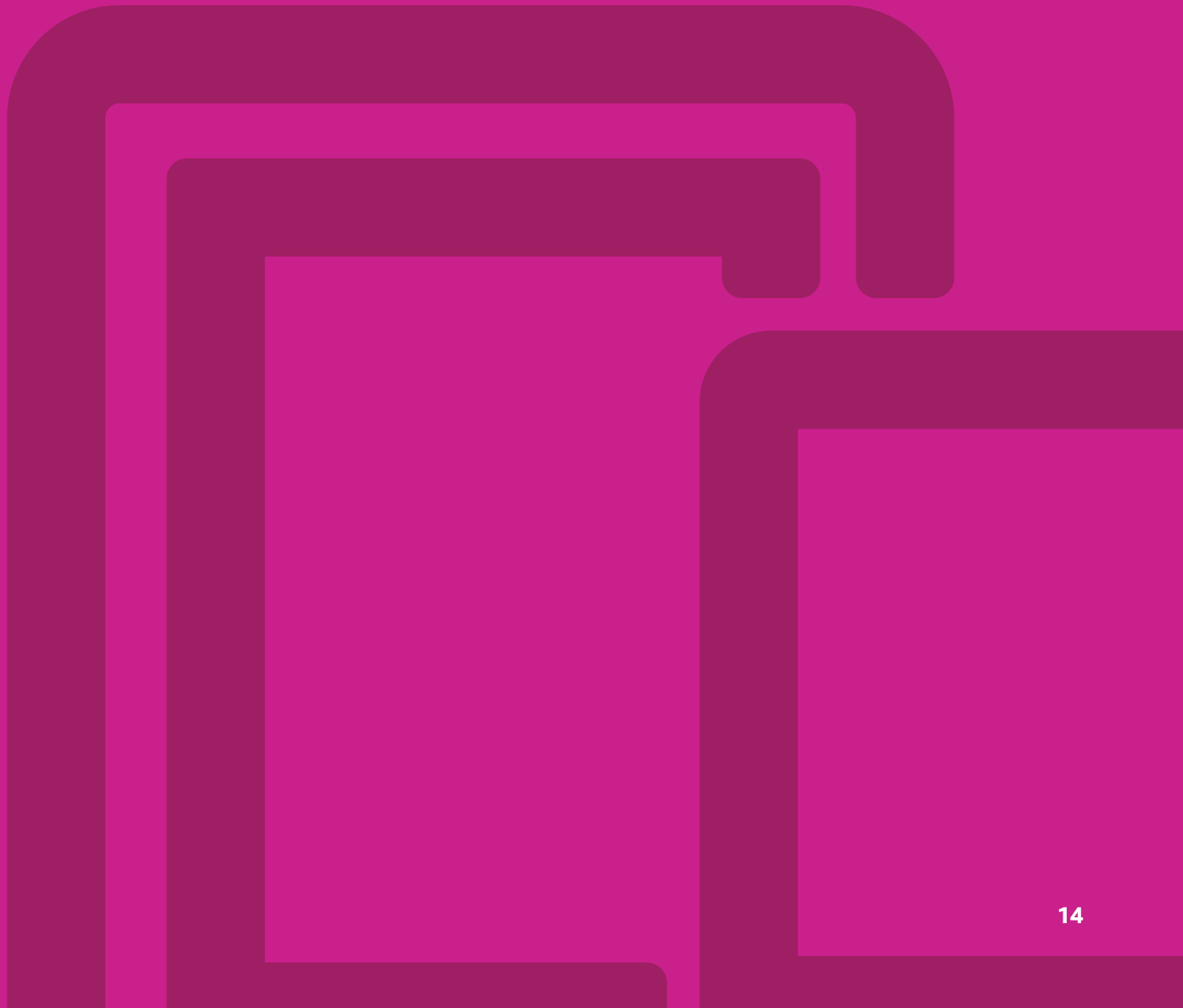
Amazon EC2 Spot Instances: Run fault-tolerant workloads for up to 90 percent cost savings

Amazon DynamoDB: Get fast and flexible NoSQL database service for any scale

Amazon S3: Store and retrieve any amount of data from anywhere

Amazon EC2 G4 Instances: Enjoy excellent price performance with the industry’s most cost-effective GPU instances for ML inference and graphics-intensive applications

Next steps



Achieve next-level advertising performance

AWS provides solutions to help you design cloud-native applications for real-time advertising—so you can take advantage of architectures that perform at scale with the most cost-efficiency.

With unmatched compute capabilities, more sizes and types of instances, and the fastest processors in the cloud with **AWS Graviton2**, AWS enables more opportunities to optimize costs for high-performance, high-scale, ultra-low latency advertising platforms.

Learn more about running Advertising Platforms on AWS ›

Appendix

Aerospike Database Enterprise Edition on AWS: Get predictable performance for globally distributed applications at petabyte scale

Amazon DynamoDB: Use a fast and flexible NoSQL database service for any scale

Amazon EC2 Auto-Scaling: Manage compute capacity to meet changes in demand

Amazon EC2 G4 Instances: Get the industry's most cost-effective GPU instances for ML inference and graphics-intensive applications

Amazon EC2 I3en instances: Tackle data-intensive workloads with dense SSD storage instances

Amazon EC2 Spot Instances: Run fault-tolerant workloads for up to 90 percent cost savings

Amazon Elastic Kubernetes Service (Amazon EKS): Choose Amazon EKS—the most trusted way to run Kubernetes

Amazon EMR: Easily run and scale Apache Spark, Hive, Presto, and other big data frameworks

Amazon Managed Streaming for Apache Kafka (Amazon MSK): Enjoy a fully managed, highly available, and secure Apache Kafka service

Amazon Rekognition: Automate your image and video analysis with machine learning

Amazon S3: Store and retrieve any amount of data from anywhere

Amazon SageMaker: Put ML in the hands of every developer

AWS Graviton2: Get faster processors in the cloud, with up to 40 percent better price performance over comparable current-generation x86-based instances

AWS Lake Formation: Build a secure data lake in days

AWS Lambda: Run code without thinking about servers or clusters

AWS Real-Time Bidding in the Cloud Solution: Deploy in regions around the world with just a few clicks, minimizing latency and cost in the real-time bidding process

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.