ANT208

What's new in Amazon Athena

Scott Rigney

Sr. Product Manager
Amazon Web Services

Ofer Eliassaf

Director, REM Cloud Infrastructure
Mobileye Vision Technologies

ANT208 ▶️

What's new in Amazon Athena

Speakers

- Ofer Eliassaf, Director, REM Cloud Infrastructure, Mobileye

- Scott Rigney, Sr. Product Manager, Amazon Web Services

Announcements

- ❖ Lots of incremental announcement. See notes and slide 12 for a summary.

Takeaways

- ❖ Athena get's faster and new capabilities with SQL v3 engine and Iceberg.

Azure

- ❖ I don't believe Azure has Athena like capability. Interestingly Azure Data Lake Storage Gen 2 is a new Athena Connector.

New Features and Capabilities

- New – Athena for Apache Spark.

- New – Athena SQL Engine v3

  - 50+ New SQL Functions, 20% Faster Queries, 100% backwards compatible
  - Faster Queries, on more information, for lower cost

- New – Faster queries with result reuse

How MobileEye Use Athena

- MobileEye is an Autonomous Driving software company. The give an overview of how they chose Athena, and how they use it.

New Data Lake, Connector, and Integration Features

- Presented AWS's view of Modern Data Architecture

  - Data Lakes – Store all your data in open formats. Cost effective storage, Decouple Storage from Compute, Choice of Analytical and ML engines, process data in place.

- New – Capabilities for Apache Iceberg (underlying engine)

  - Create Table as Select (allows you to create new tables as a result of a Select query), Views to hide complex logic, Performance 10x faster queries
  - Merge to sync tables and simplify data prep, Vacuum to manage storage, New formats Avro & ORC.

- New – Expanded fine-grained access controls

- New – Data source connectors

Slide 1

Amazon Athena for Apache Spark

Get started with interactive analytics on Apache Spark in under a second

GENERALLY AVAILABLE TODAY

- Harness Apache Spark for complex, powerful analytics
- Spend more time on insights instead of waiting for results
- Build applications without managin resources or configuring software

Slide 2

Amazon Athena SQL engine v3

Upgrade today to experience multiple improvements and new features

- Faster queries at lower cost
- More frequent updates and features
- Based on open-source Trino, Presto, and AWS-developed innovations for both

Slide 3

SQL engine v3 new features and enhancements

FASTER QUERIES AND MORE FREQUENT UPDATES

NEW	NEW	NEW	UPDATED	NEW	UPDATED
Query result caching	Aggregation functions	Date & time functions	Prepared statements	Number formatting	Memory usage

NEW	UPDATED	UPDATED	UPDATED	NEW	UPDATED
More frequent updates	Geospatial functions	Syntax enhancements	AWS Glue metadata	Array functions	Right-join dynamic filter

Slide 4

Faster Queries with Result Reuse

New caching feature to accelerate repeat queries and reduce time to insight

5x

Up to 5x faster queries when result reuse is active

$0

No data scanned, so queries are free to you

The process that led us to Amazon Athena

Design phase

Initial idea of developing two systems for short/long term storage (cost vs. efficiency)

Started with the long-term storage

Research conclusion: Amazon Athena + S3 parquet files for long term

Solution verification with AWS solution architects

Data-lab in Seattle

Implementation phase

Conclusion verification and POC implementation in the data-lab

Finalizing implementation

Production phase

Usage grew exponentially (usage pattern, amount of data, cost)

Iterative optimizations

Up until now, no need for the short-term storage

Optimizations toward our current state

Iterative approach

- Query-level optimizations
- Partnered with AWS to reach massive concurrency
- Refactored data model to eliminate joins and align with Athena best practices
- Injected partition projection (geospatial + time)

Our latest optimizations led us to get:

90% Query time decrease

90% Cost decrease

0 bottlenecks across different map activities
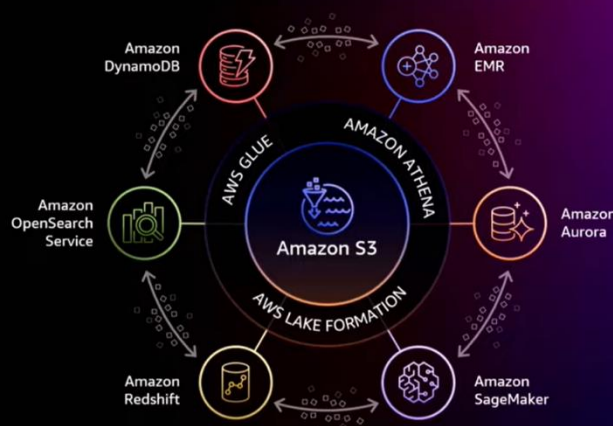
10s → 100s concurrent map cell runs per minute

0 concurrency issues

Modern data architecture on AWS

Amazon DynamoDB — Amazon EMR — Amazon Aurora — Amazon SageMaker — Amazon Redshift — Amazon OpenSearch Service — AWS GLUE — AMAZON ATHENA — Amazon S3 — AWS LAKE FORMATION

New features and capabilities with Iceberg

NEW

CTAS

Easy and fast way to create new tables

NEW

Views

Create views to hide complex logic

NEW

Performance

10x faster queries when using SQL engine v3

New features and capabilities with Iceberg

NEW

Merge
Sync tables to simplify data prep and updates

NEW

Vacuum
Manage storage and delete records to meet regulatory requirements

NEW

Avro & ORC
New support for Avro and ORC formats

NOW AVAILABLE

Expanded fine-grained access controls

Apply AWS Lake Formation access control policies to more of your data

Centrally defined table-, row-, and column-level policies in Lake Formation

Now supports all table formats and file types available with Athena

Filtering operations performed natively by Athena for increased performance

---

New data source connectors

DATA SOURCE CONNECTORS FOR IN-PLACE ANALYTICS

Easy to set up and query
Use the Athena console to create and configure your connectors, then run SQL queries

Secure cross-account access
Deploy and configure one connector, then use flexible IAM policies to enable cross-account access to other analysts

Open-source and fully supported
Athena connectors are open-source, so you can use our code as a starting point to building your own

NEW

Azure Data Lake Storage Gen2
Azure Synapse
Cloudera
Google BigQuery
Hortonworks
IBM Db2
Microsoft SQL Server
Oracle
SAP HANA
Snowflake
Teradata

What's new with Amazon Athena

KEY TAKEAWAYS TO APPLY TO YOUR WORKLOADS TODAY

- **Amazon Athena for Apache Spark** – Available today! Sub-second startup for interactive Spark analytics. Check out **ANT209** to learn more

- Upgrade to **SQL engine v3** for faster queries and more analytics features

- Turn on **Query Result Reuse** for cached results

- Expanded support of **Apache Iceberg** and **AWS Lake Formation**

- **New data source connectors** and **Amazon AppFlow** help you expand analysis and break down data silos

- And many more …