
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT208

What's new in Amazon Athena

Scott Rigney

Sr. Product Manager
Amazon Web Services

Ofer Eliassaf

Director, REM Cloud Infrastructure
Mobileye Vision Technologies

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

New features and capabilities for analytics

How Mobileye uses Athena to enable autonomous driving

New data lake, connector, and data integration features

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Athena

Interactive analytics for all your data



Simple, instant start



Interactive analytics



Open and flexible



Cost effective

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Simple, instant start

Serverless, no setup

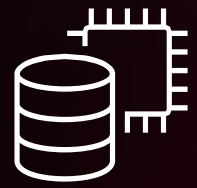
Start quickly, optimized for fast results



Interactive analytics

Integrated tooling works where users are

Run queries on **25+** data stores



Open and flexible

Based on open source, optimized for AWS

Supports multiple formats, compression types, data types, and more



Cost effective

Pay only for what you use

Save **30%–90%** on per-query costs through compression

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use Cases for Amazon Athena



Interactive Analysis

Fast, scalable interactive queries on petabyte-scale data lakes



Business Intelligence

Visualizations and scheduled reporting on your data lake



Data Workflows

Lightweight data workflows to prepare data for downstream use cases



Analytics Apps

Build applications that use Athena to surface insights



Machine Learning

Combine raw data from multiple sources to train more diverse models

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer challenges with SQL

"SQL alone is not expressive enough. Users need additional languages and ecosystem of add-on libraries."

"Open source frameworks like Apache Spark have many settings to tune to get optimal performance."

"Setting up infrastructure for interactive workloads is hard. Job profiles are hard to predict and warm pools are needed for quick-start."

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NOW AVAILABLE

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 managing resources or configuring software

aws

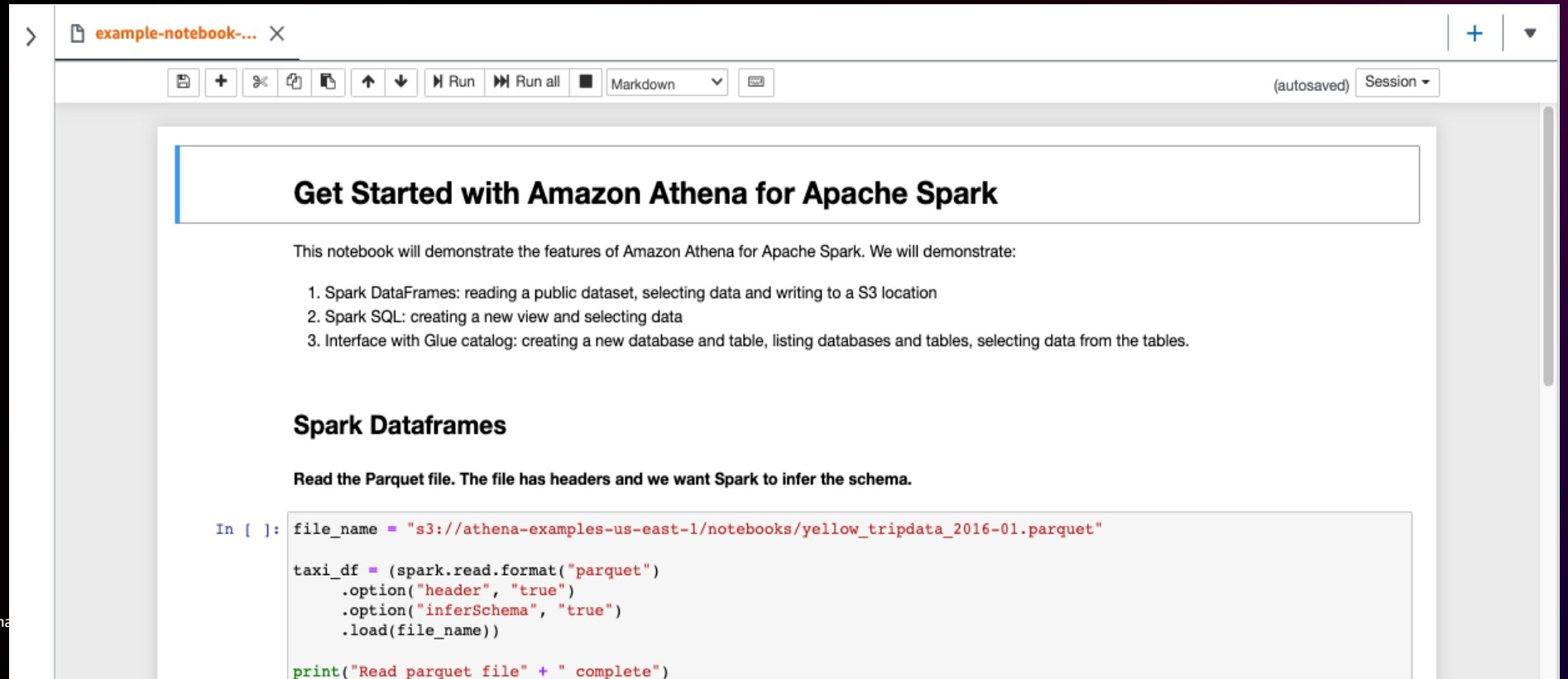
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Attend ANT209 to learn more

Build interactive analytics applications on Apache Spark

- Thursday, December 1
- 3:30 PM PST

New notebook experience in Athena



© 2022, Ama

NOW AVAILABLE

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

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SQL engine v3 benefits

FASTER QUERIES AND MORE FREQUENT UPDATES

50

New SQL functions to expand analytics capability

90+

Enhancements to engine behaviors producing faster performance

20%

Faster queries at the same cost with up to 10x faster query planning

100%

Functional parity with engine v2 and easy to upgrade

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“We can now query our massive petabyte-scale data lake more efficiently and at lower cost. We’re excited about using the latest Trino features in our ML-driven anomaly detection solution”

Arie Teter

VP R&D at Orca Security

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SQL engine v3 new features and enhancements

FASTER QUERIES AND MORE FREQUENT UPDATES

NEW



Query result caching

NEW



Aggregation functions

NEW



Date & time functions

UPDATED



Prepared statements

NEW



Number formatting

UPDATED



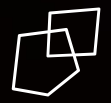
Memory usage

NEW



More frequent updates

UPDATED



Geospatial functions

UPDATED



Syntax enhancements

UPDATED



AWS Glue metadata

NEW



Array functions

UPDATED

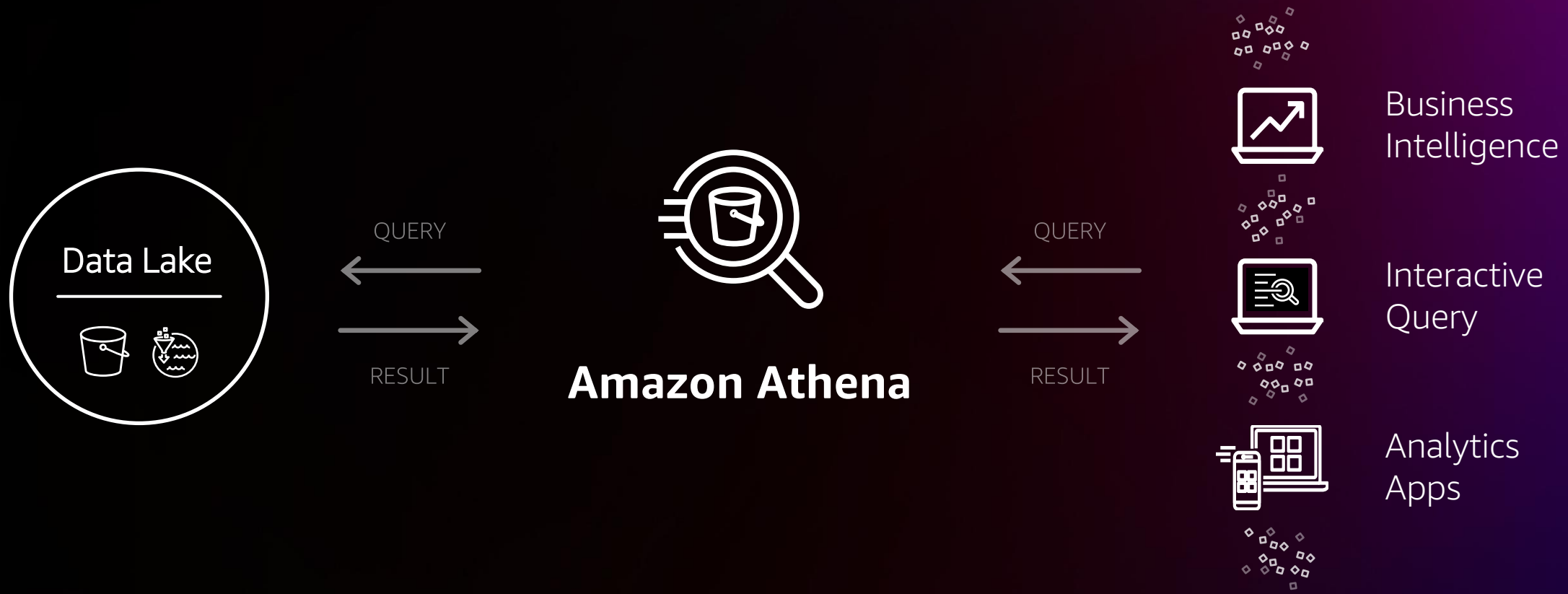


Right-join dynamic filter

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Athena in multi-user applications



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NOW AVAILABLE

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

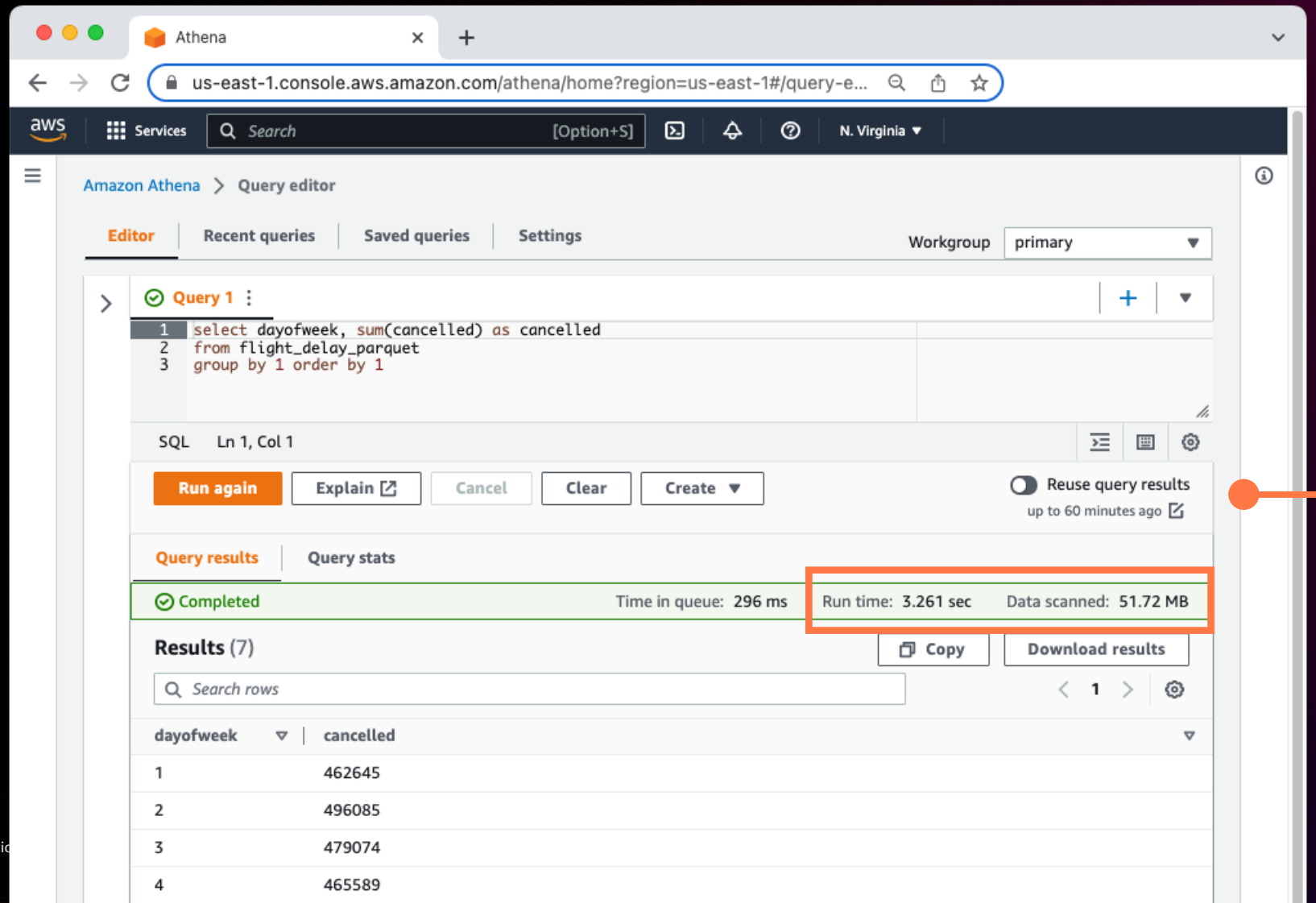
No data scanned, so queries are free to you

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Accelerate queries with result reuse

UP TO 5X FASTER QUERIES WITH NO DATA SCANNED

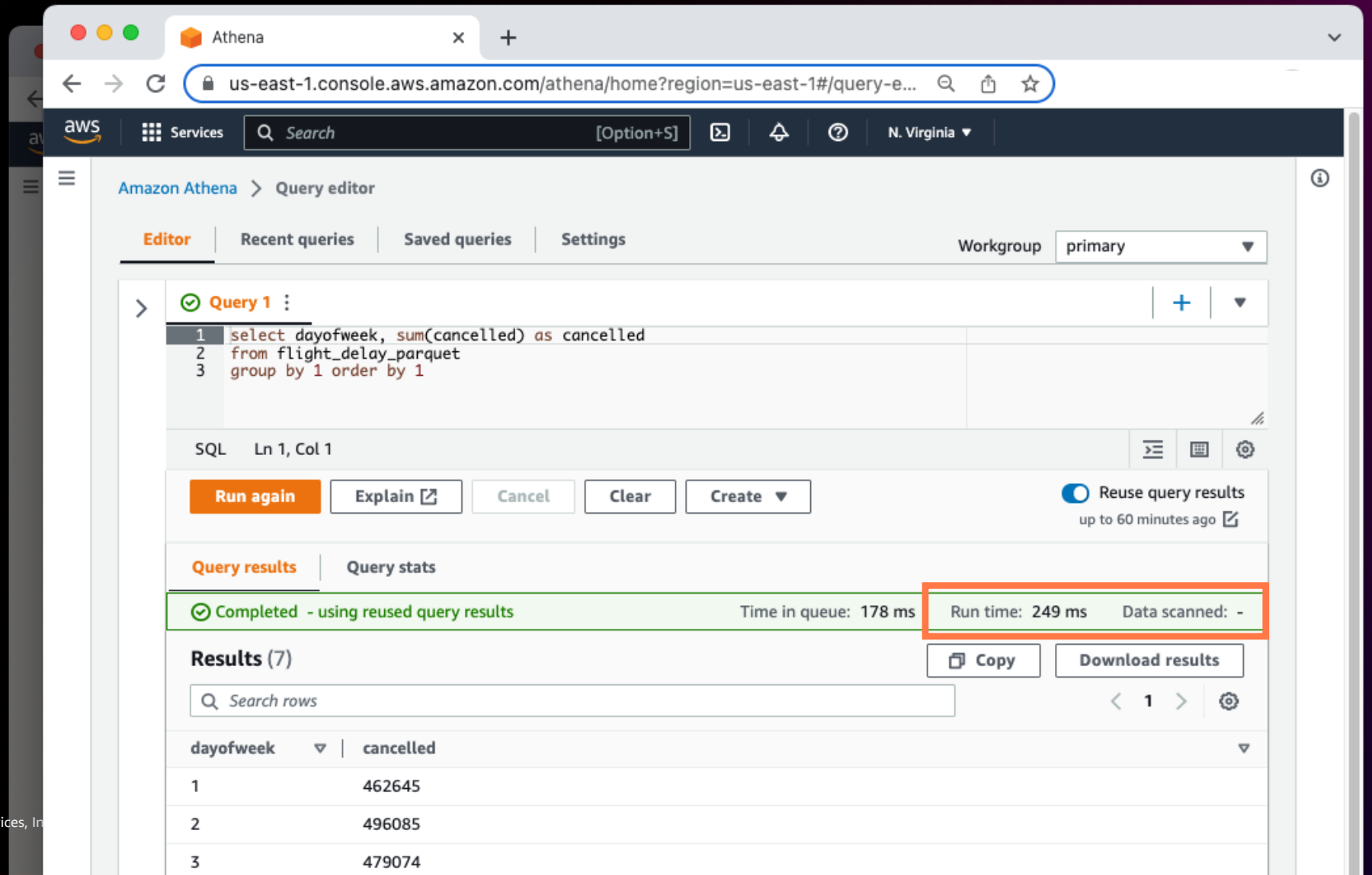


Turn on result reuse

© 2022, Amazon Web Servic

Accelerate queries with result reuse

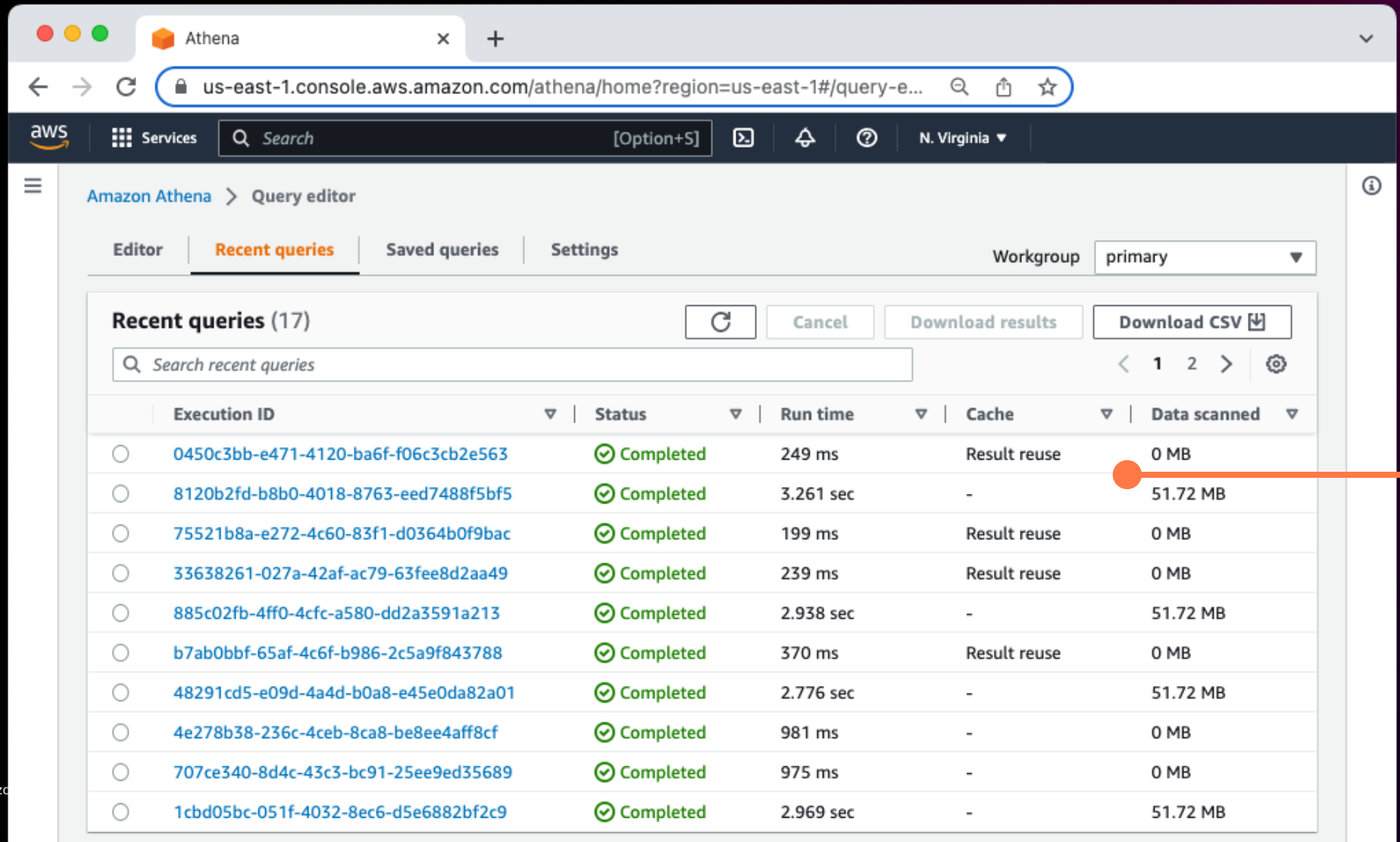
UP TO 5X FASTER QUERIES WITH NO DATA SCANNED



© 2022, Amazon Web Services, In

Accelerate queries with result reuse

UP TO 5X FASTER QUERIES WITH NO DATA SCANNED



© 2022, Amazo

Query Result Reuse

UP TO 5X FASTER QUERIES WITH NO DATA SCANNED

Easy to use

No changes in your queries required. Simply turn on query result caching in the console, driver, or API

Automated

Athena automatically identifies queries that can be sped-up through result reuse

No data scanned

Queries executed with result re-use don’t scan any data

Multi-user benefit

Multiple users get the performance benefit from wherever they access Athena

aws

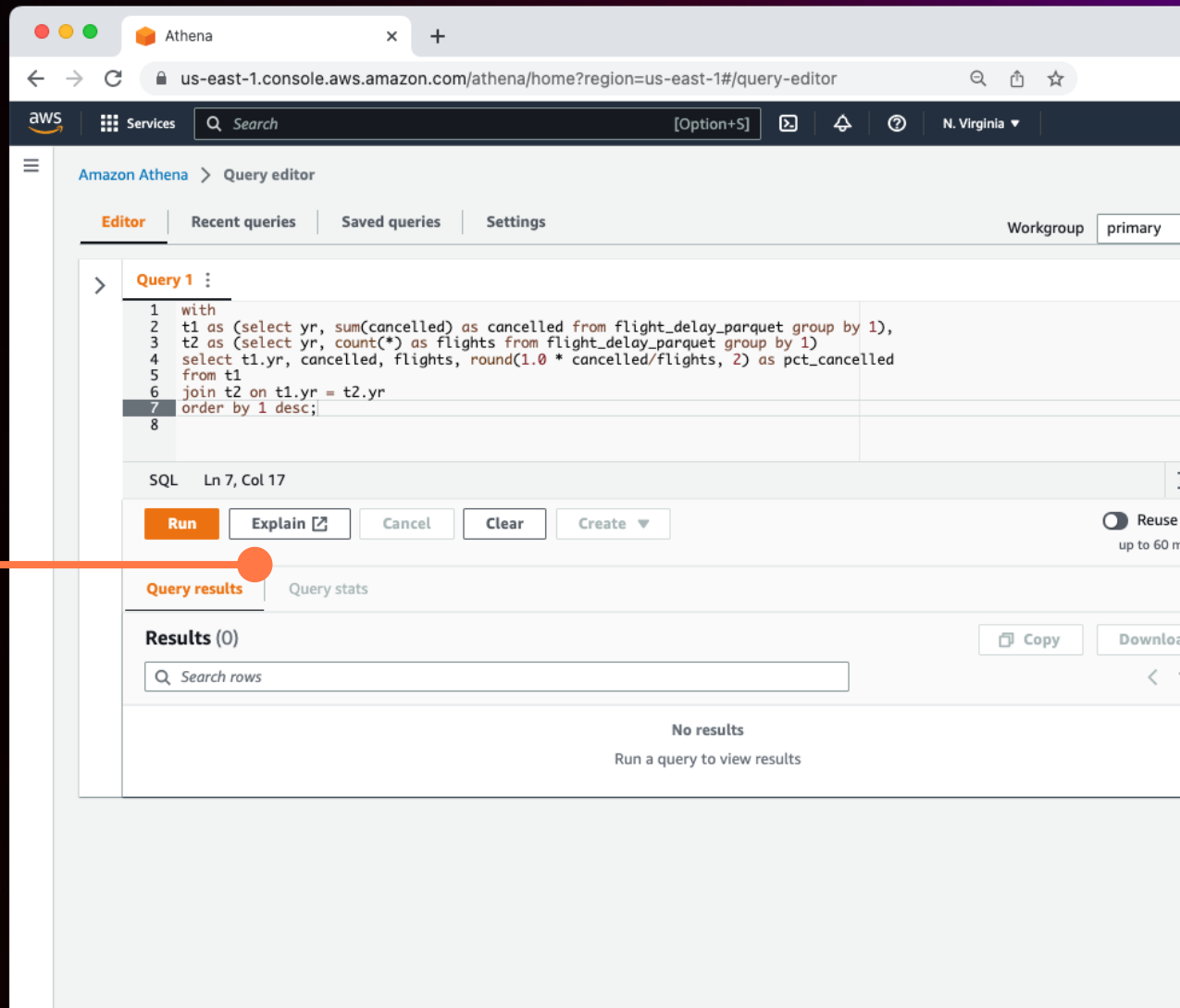
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Query analysis and tuning tools

INTERACTIVE QUERY PLAN AND RUNTIME DETAILS TO HELP YOU OPTIMIZE QUERIES

Visual Query Plans

Interactive query plan visualizer allows you to easily inspect queries before execution



Single-click query plan

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

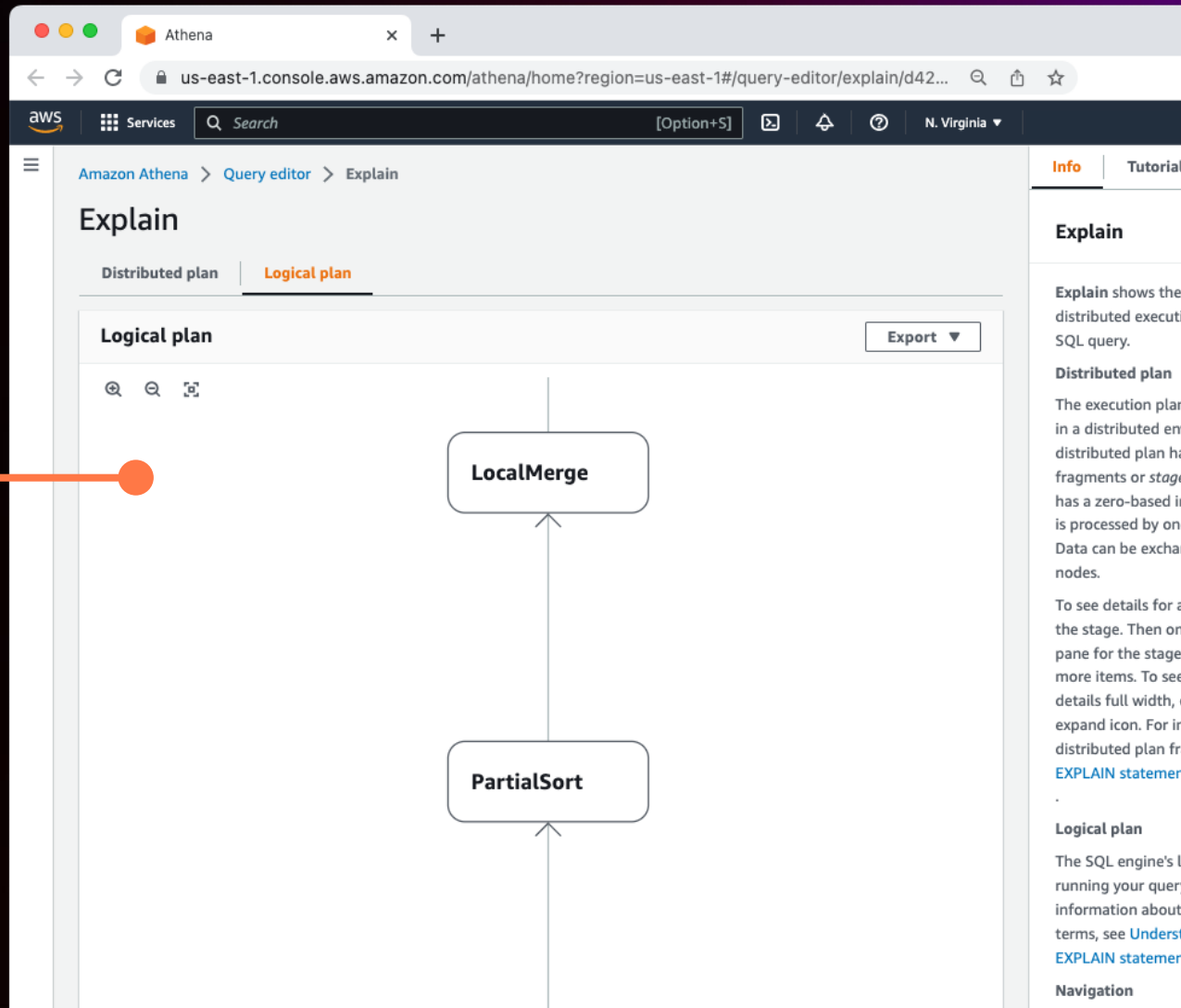
Query analysis and tuning tools

INTERACTIVE QUERY PLAN AND RUNTIME DETAILS TO HELP YOU OPTIMIZE QUERIES

Visual Query Plans

Interactive query plan visualizer allows you to easily inspect queries before execution

Visual query plan



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Query analysis and tuning tools

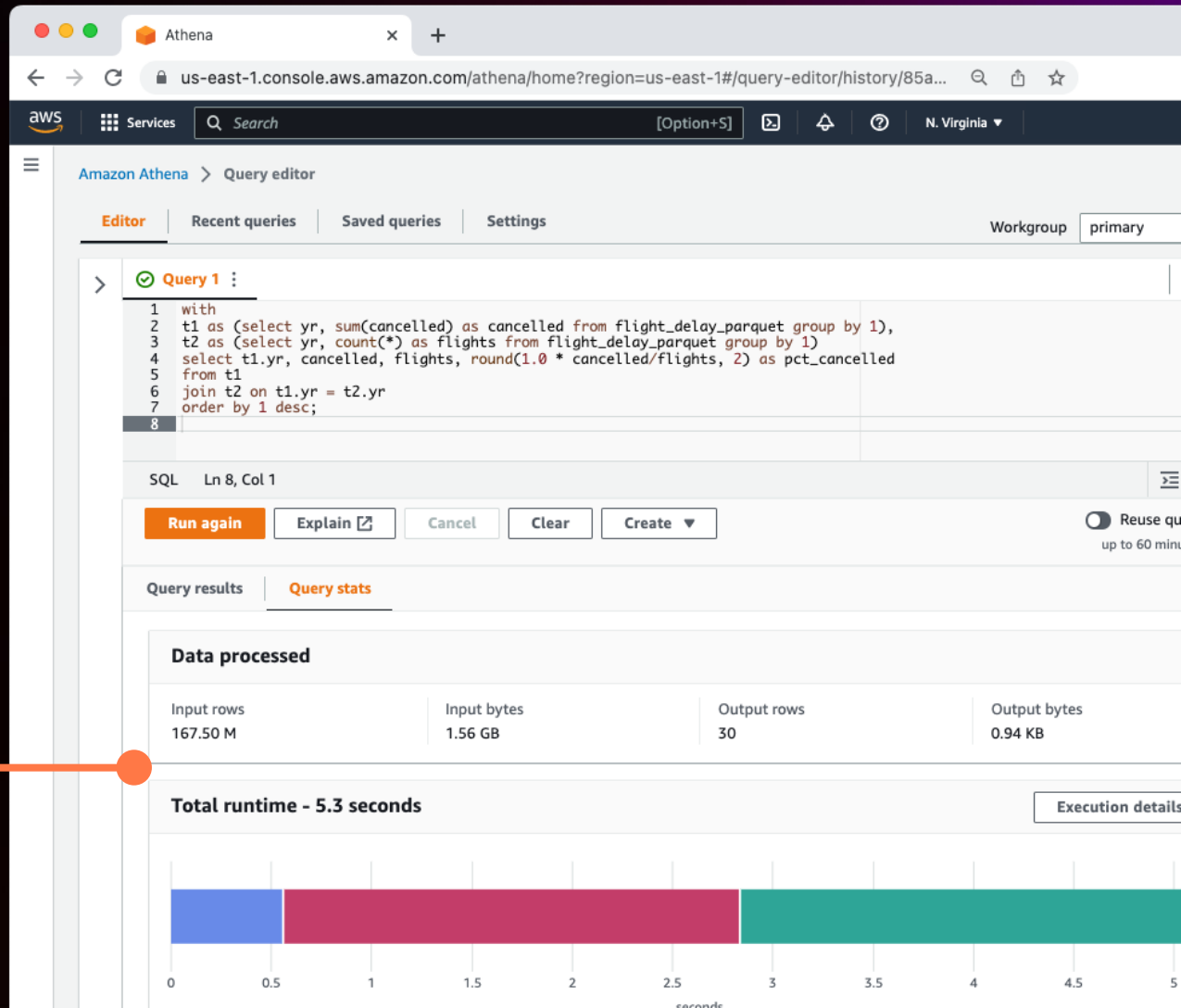
INTERACTIVE QUERY PLAN AND RUNTIME DETAILS TO HELP YOU OPTIMIZE QUERIES

Visual Query Plans

Interactive query plan visualizer allows you to easily inspect queries before execution

Runtime Statistics

Easy to digest performance statistics help you validate queries and understand the impact of changes in SQL



Interactive execution statistics

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Query analysis and tuning tools

INTERACTIVE QUERY PLAN AND RUNTIME DETAILS TO HELP YOU OPTIMIZE QUERIES

Visual Query Plans

Interactive query plan visualizer allows you to easily inspect queries before execution

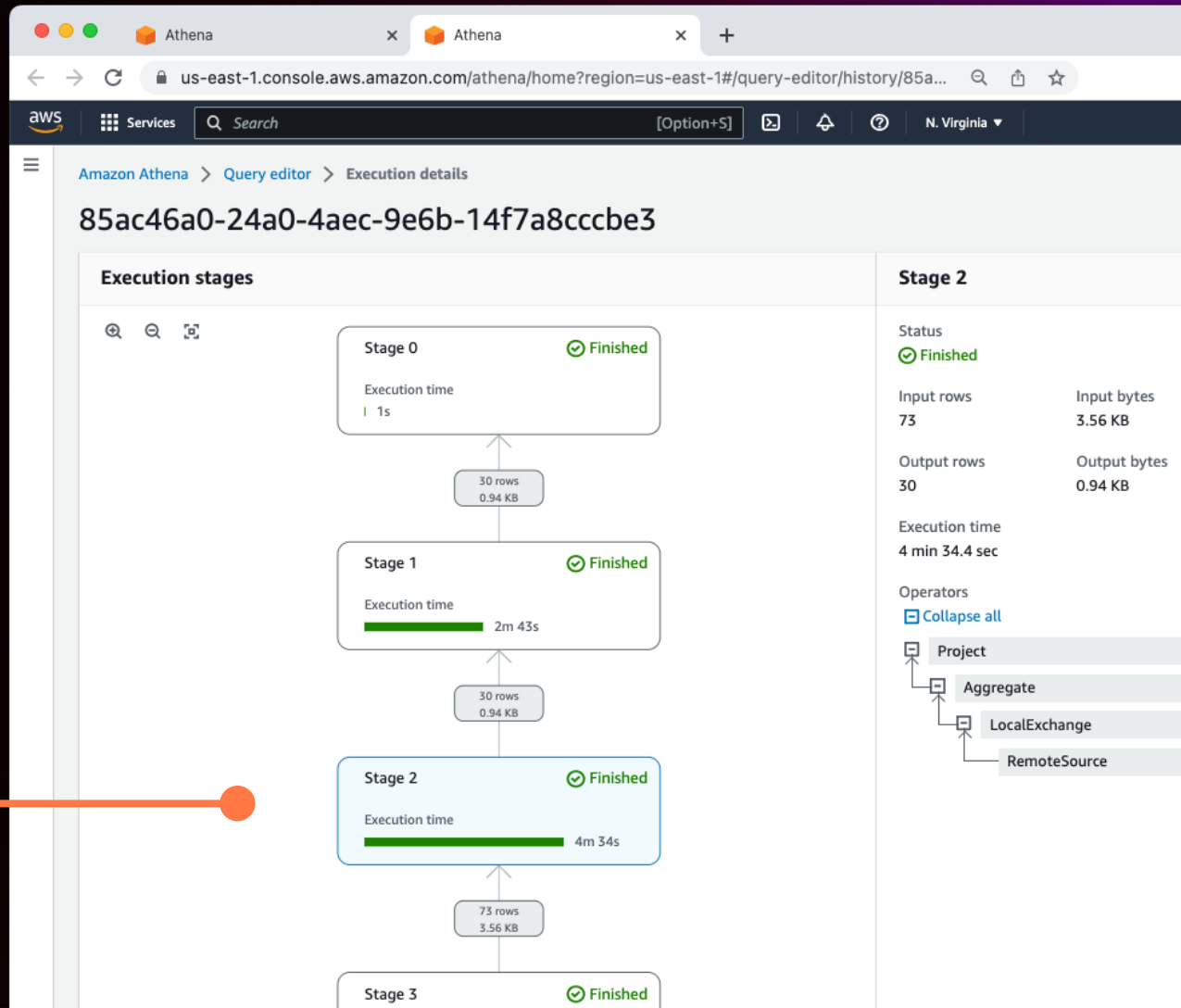
Runtime Statistics

Easy to digest performance statistics help you validate queries and understand the impact of changes in SQL

Execution Details

Examine your query with interactive, stage-level details to uncover areas to tune your queries for greater performance

Detailed, stage-level operator performance data



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Query analysis and tuning tools

INTERACTIVE QUERY PLAN AND RUNTIME DETAILS TO HELP YOU OPTIMIZE QUERIES

Visual Query Plans

Interactive query plan visualizer allows you to easily inspect queries before execution

Runtime Statistics

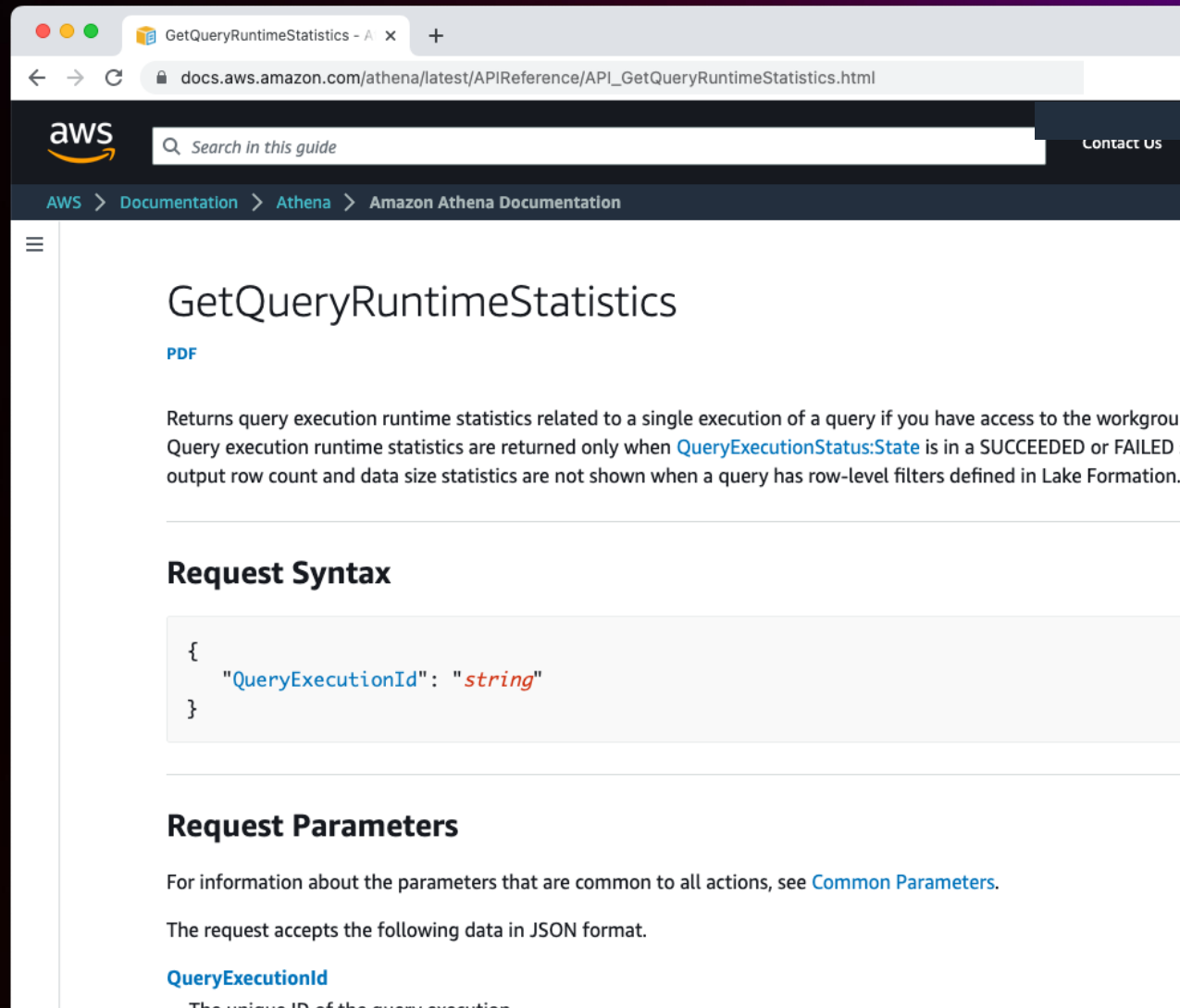
Easy to digest performance statistics help you validate queries and understand the impact of changes in SQL

Execution Details

Examine your query with interactive, stage-level details to uncover areas to tune your queries for greater performance

Query Statistics API

Access execution details through Athena's API for more advanced query analysis use cases



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Optimize data structure for performance

FASTER QUERIES AT LOWER COST WHEN YOU USE COLUMNAR FORMATS, PARTITIONS, AND COMPRESSION

Use Columnar Data

Query performance improves if you convert your data into columnar formats. Use Athena's CTAS feature or use AWS Glue jobs to convert your existing data to a columnar format

Supported Types

- Parquet
- Orc

Partition Data

By partitioning your data, you can restrict the amount of data scanned by each query, thus improving performance and reducing cost

Available Methods

- Partition Projection
- Glue Partition Indexing for multi-engine support
- Bucketing

Compression

Athena bills on compressed bytes, so you can save up to 90% on queries that scan data through compression

Supported Types

- BZIP2
- DEFLATE
- GZIP
- LZ4
- LZO
- SNAPPY
- ZLIB
- ZSTD

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How Mobileye uses Amazon Athena to enable autonomous driving

Ofer Eliassaf

Director, REM Cloud Infrastructure
Mobileye Vision Technologies

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

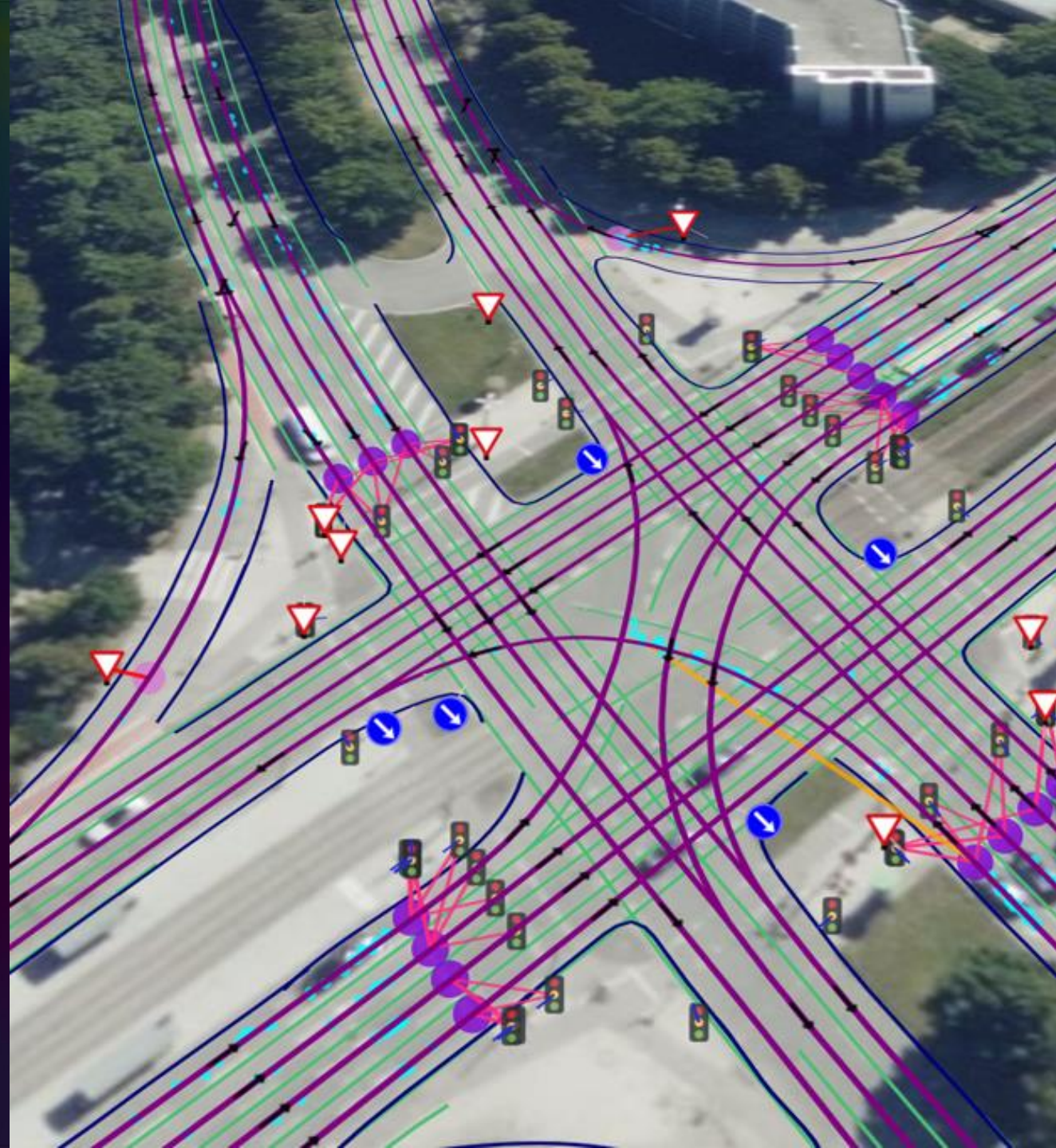
Agenda

- Overview of Mobileye’s REM AV mapping technology
- The data ingestion challenge and need for data lakes
- Our journey with Amazon Athena

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why do autonomous vehicles need maps?

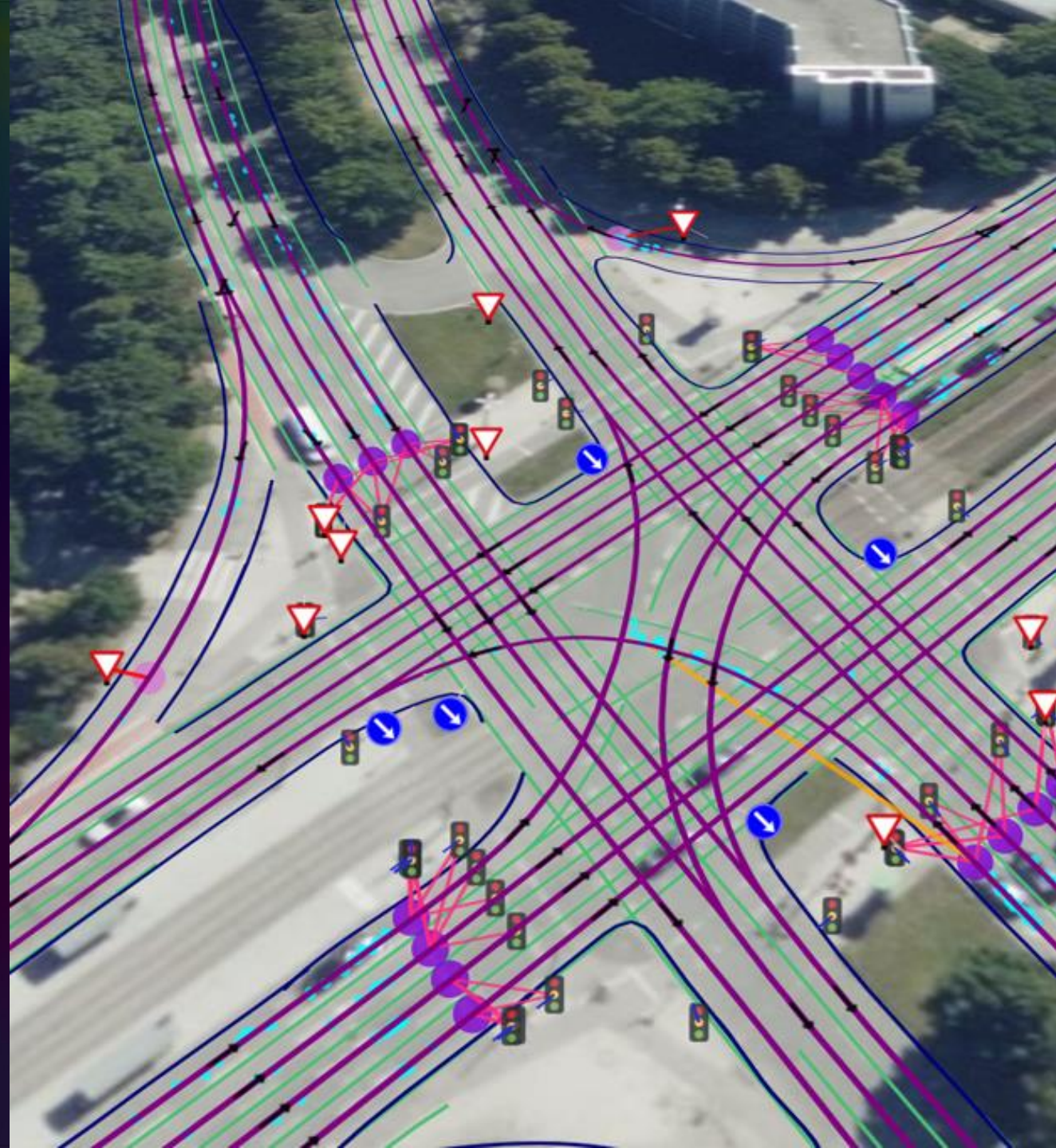
- REM™ is Mobileye's technology for AV mapping
- Roadbook™ is Mobileye's AV map product



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why do autonomous vehicles need maps?

- REM™ is Mobileye's technology for AV mapping
- Roadbook™ is Mobileye's AV map product



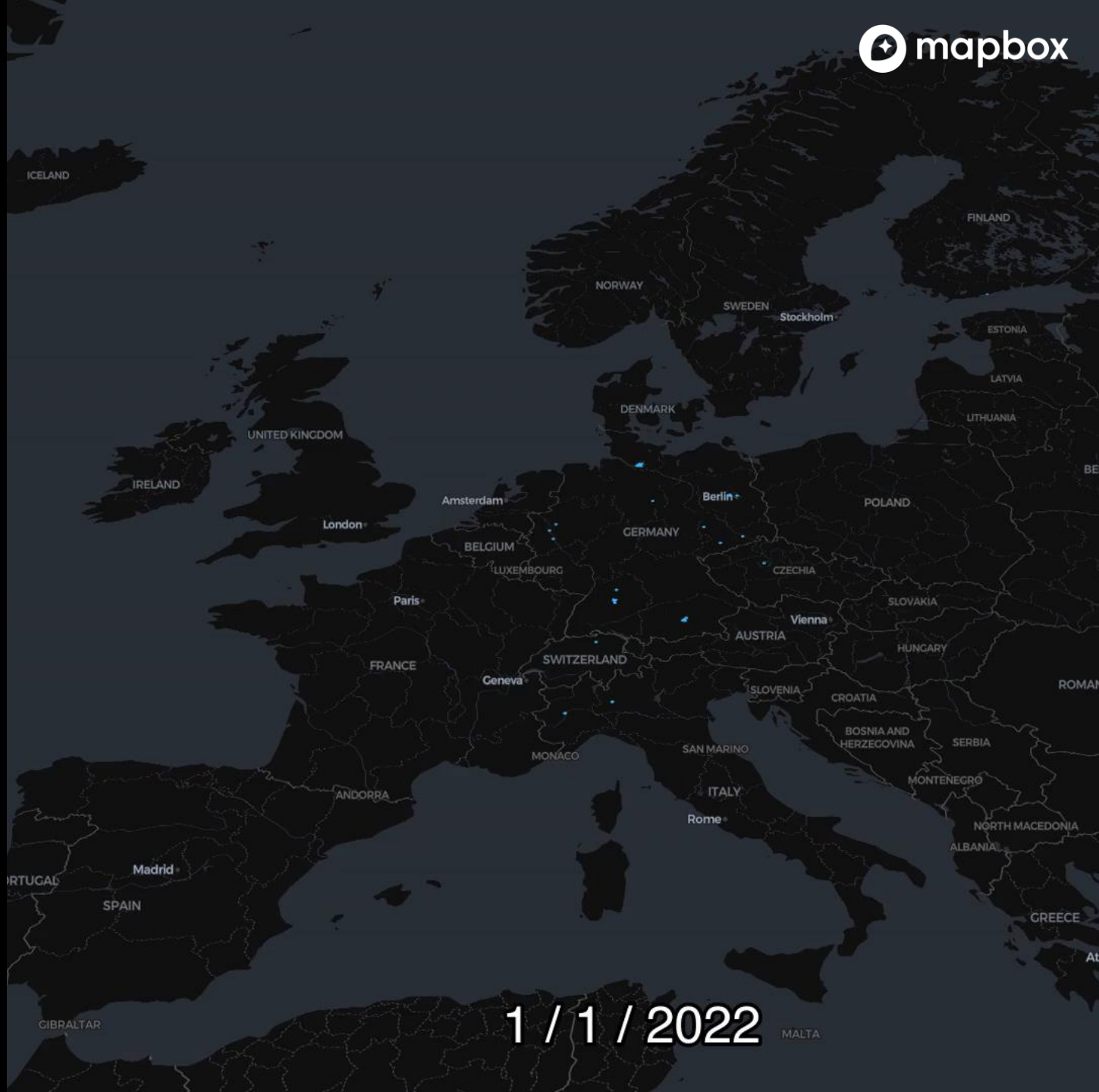
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Roadbook example



REM’s data scale

- Many car companies (OEMs) and a few car models per OEM
- More car companies and projects are expected in the coming years
- Our current ingestion rate is ~69M KM/day with global coverage
- ~14B KMs of road data collected as of July 2022
- ~1.5M (est.) contributing vehicles
- Our ingestion covers vast majority of the United States and Europe
- Growing exponentially year over year



REM crowdsourced mapping

Scalability

Unlocking millions of “mapping agents” in every relevant region

Accuracy

Use novel state-of-the-art algorithms to achieve high accuracy levels where it matters

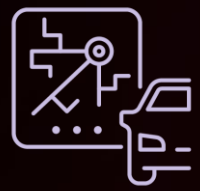
Detailed semantic features

Use explicit attributes and crowd-sourced data to generalize traffic rules and driving culture

REM Process:



1 RSD Harvesting



2 Automatic Roadbook Creation



3 Localization & AD

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

REM crowdsourced mapping

Scalability

Unlocking millions of “mapping agents” in every relevant region

Accuracy

Use novel state-of-the-art algorithms to achieve high accuracy levels where it matters

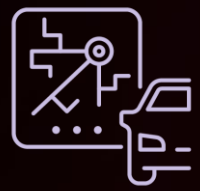
Detailed semantic features

Use explicit attributes and crowd-sourced data to generalize traffic rules and driving culture

REM Process:



1 RSD Harvesting



2 Automatic Roadbook Creation



3 Localization & AD

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The need for an ingestion data lake

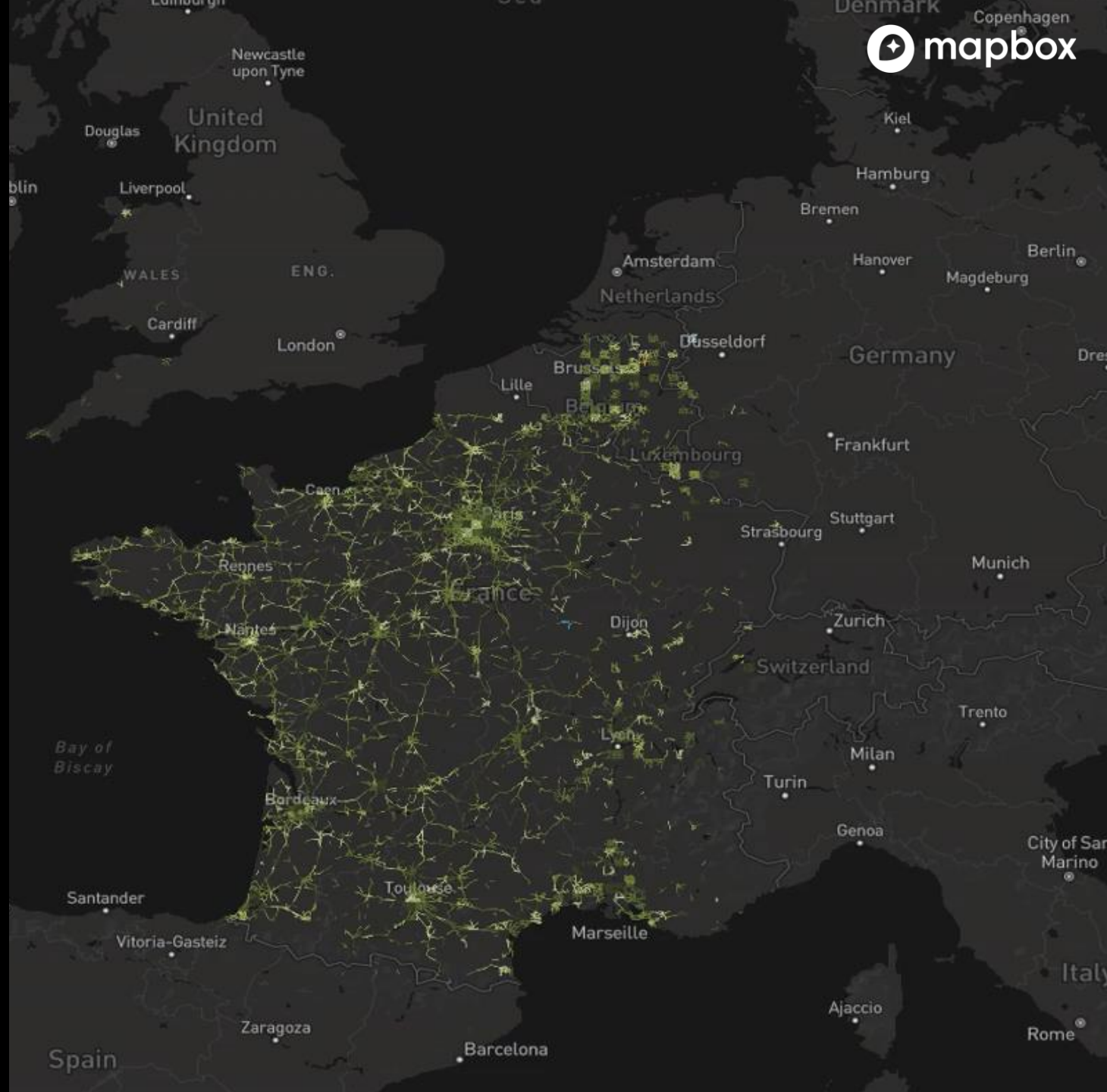
A need to store metadata of ~150 attributes per ingestion payload (RSD)

- Roadbook creation
- Various analytics queries
- UI visualizations
- Data services usages such as smart cities
- More and more usage as time goes by

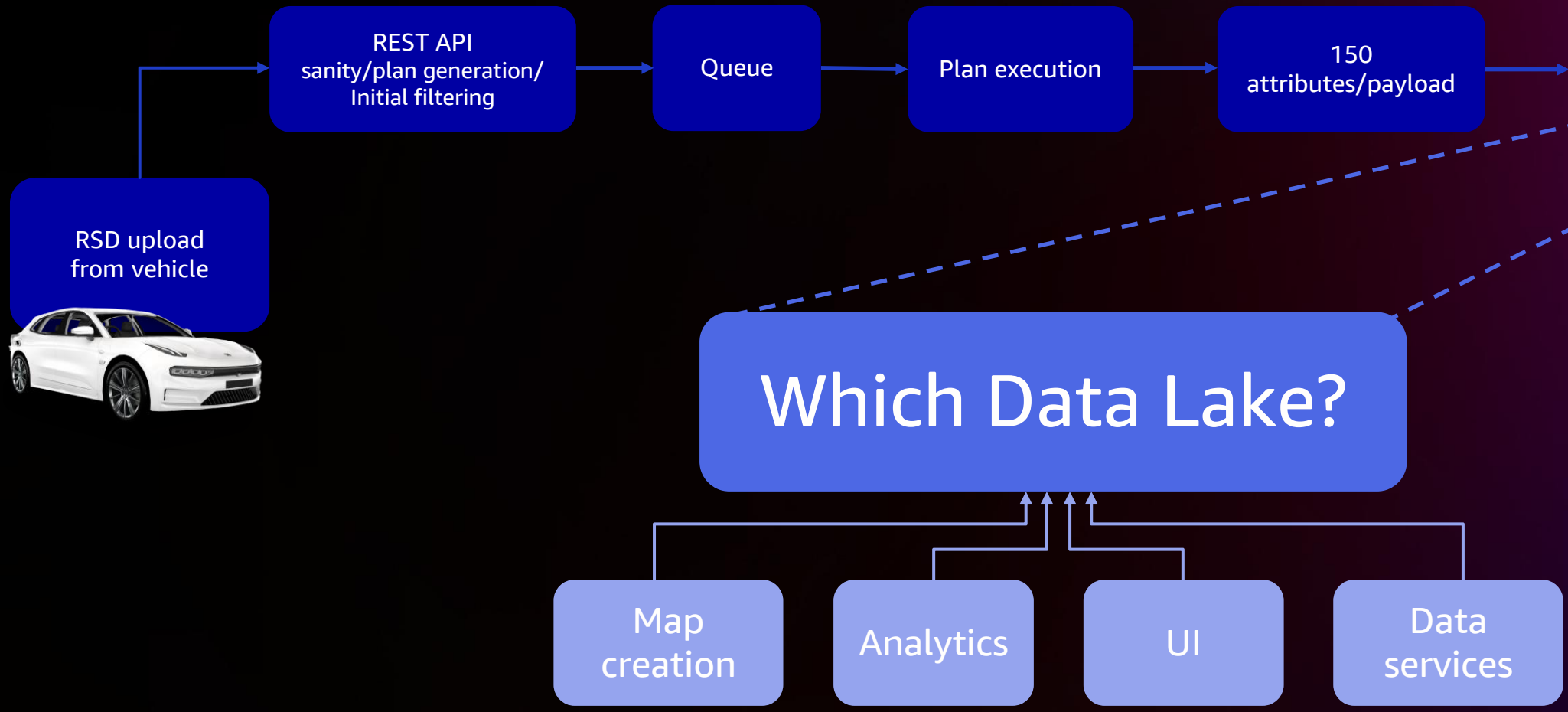
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Roadbook usage of the data lake

- Map creation is built from multiple geographical cells of ~10x10 KM
- Mapping each cell starts with a query returning millions of records (RSDs)
- Most are scanning 3-5 months of data in specific regions
- Multiple mapping jobs in parallel
- Up to tens of thousands of queries per day



We started with ingestion pipeline



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

REM requirements from the data lake

- Ability to do join/visualizations/BI → relational data by nature
- Data freshness
- Geographical queries
- Reliable (mission critical) and easy to maintain
- Reasonable query speed
- High query concurrency
- Long term storage (cost effectiveness)
- Future usage pattern is unknown

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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)

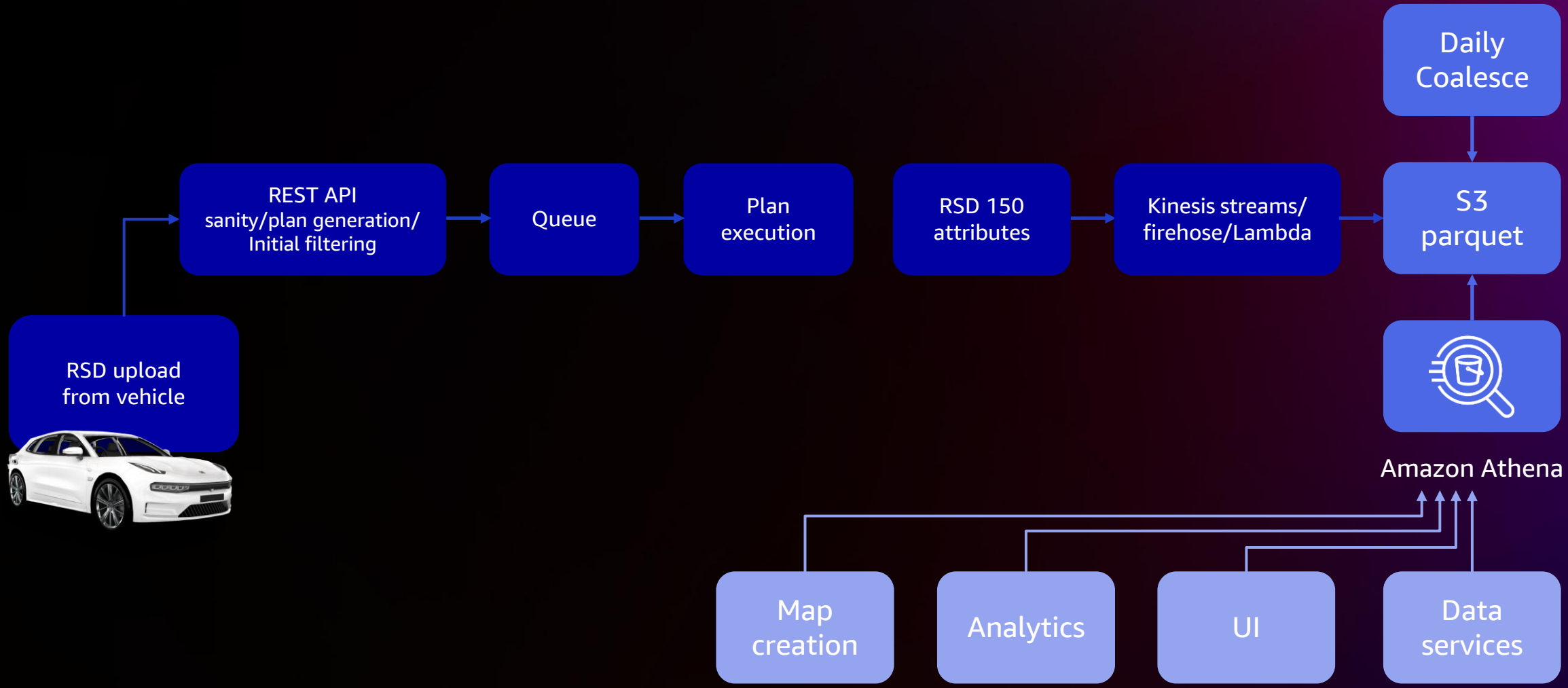
Iterative optimizations

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

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our current ingestion pipeline



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How Mobileye uses Amazon Athena to enable autonomous driving

Ofer Eliassaf

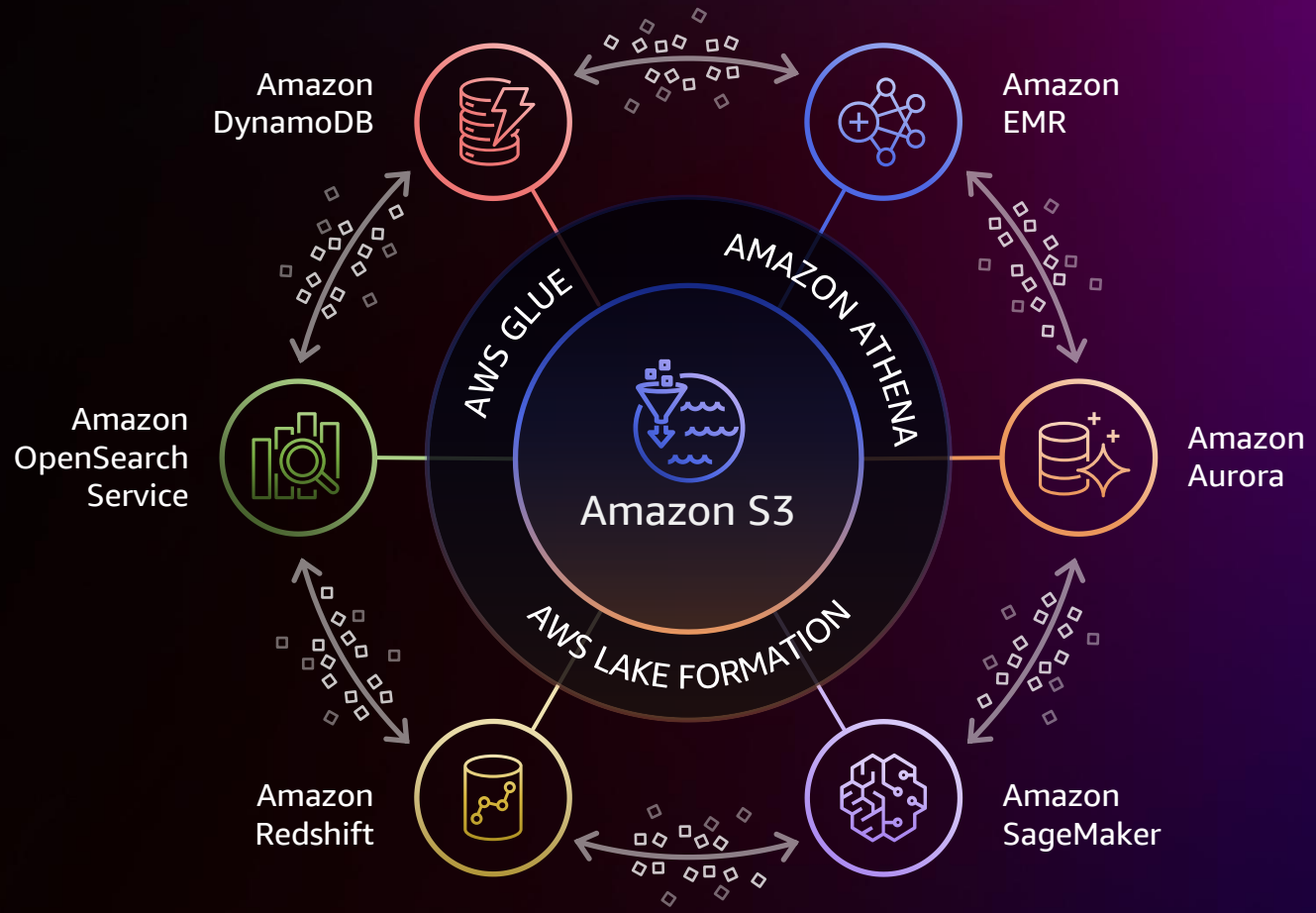
Director, REM Cloud Infrastructure
Mobileye Vision Technologies

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern data architecture on AWS

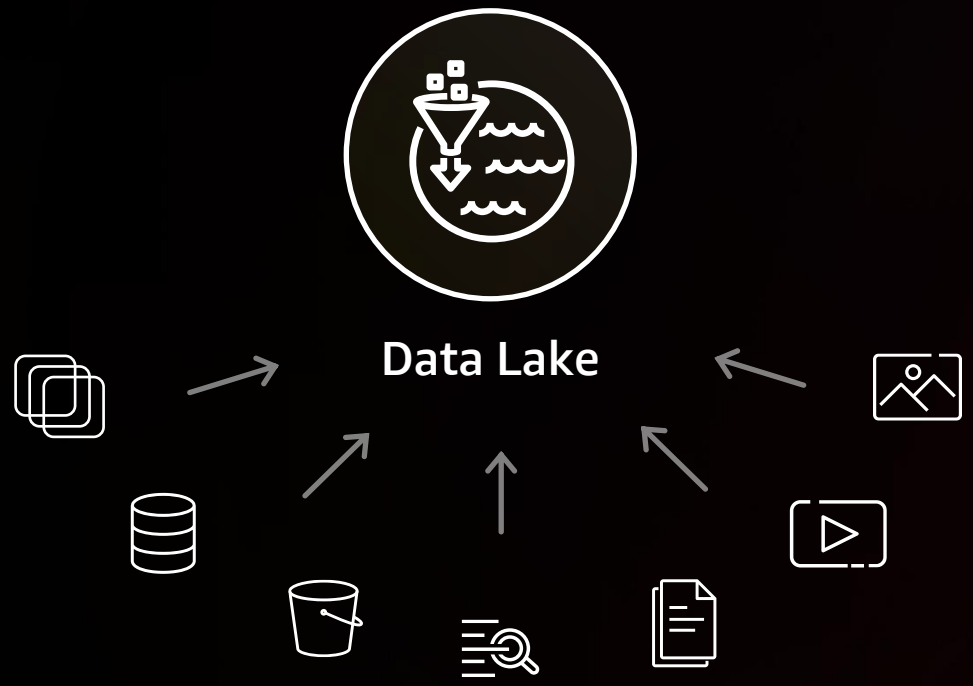
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern data architecture on AWS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Athena ACID Transactions Powered by Iceberg

- Record level insert
- Update
- Delete
- Compaction
- Time-travel queries
- Schema evolution

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

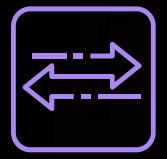
10x faster queries when using SQL engine v3

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

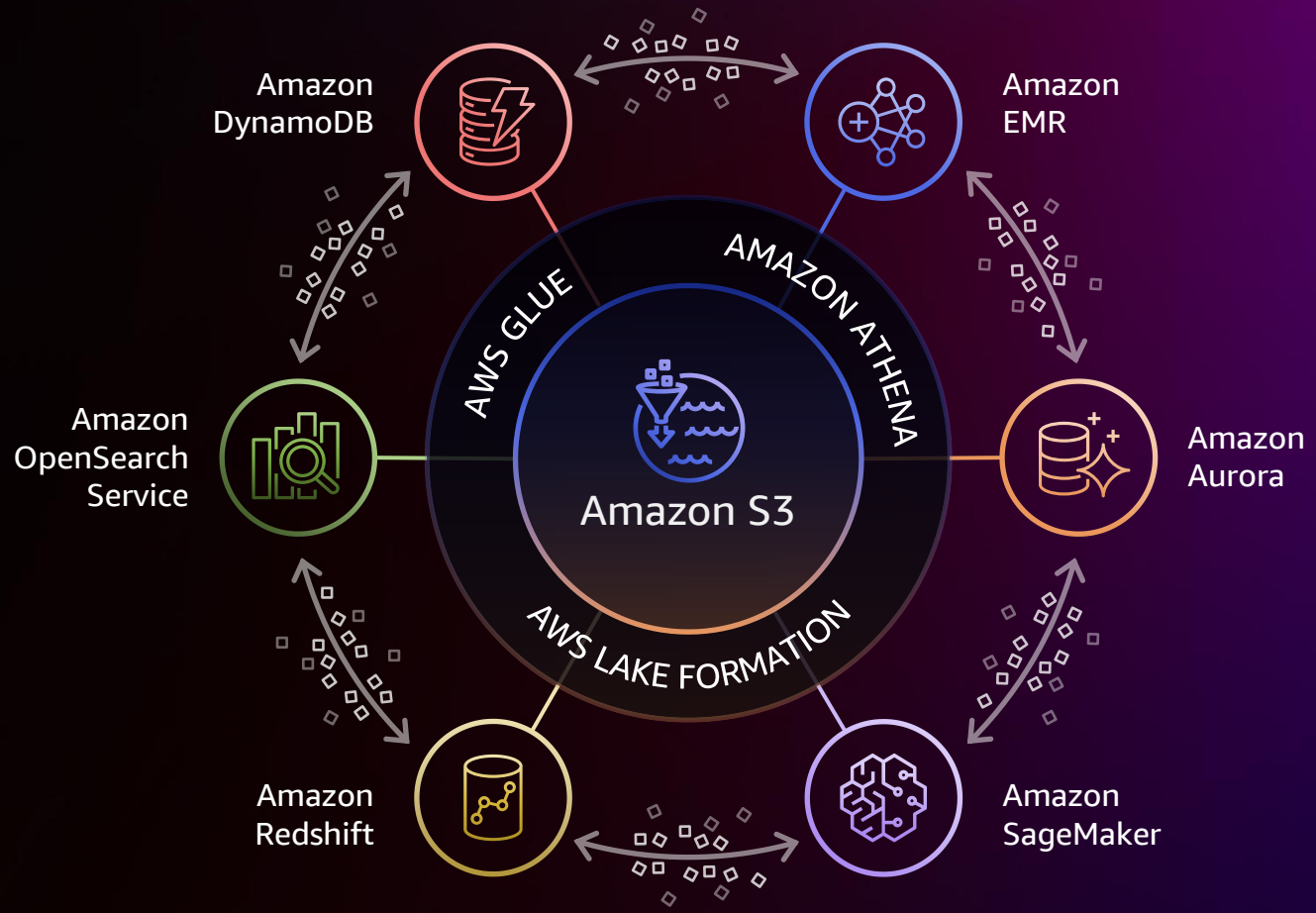
Now supports all table formats and file types available with Athena

Filtering operations performed natively by Athena for increased performance

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modern data architecture on AWS



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Federated Query

DATA SOURCE CONNECTORS FOR IN-PLACE ANALYTICS



Integrate external sources with your lake

Flexible and easy way to ingest new data

Process data in place

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Supported data sources

AWS

- Amazon Redshift
- Amazon DynamoDB
- Amazon RDS
- Amazon DocumentDB
- Amazon OpenSearch Service
- Amazon Timestream
- Amazon ElastiCache
- Amazon Neptune
- Amazon CloudWatch
- Amazon CloudWatch Metrics

THIRD PARTY

NEW

- Azure Data Lake Storage Gen2
- Azure Synapse
- Microsoft SQL Server
- Cloudera
- Google BigQuery
- Teradata
- SAP HANA
- Hortonworks
- Snowflake
- Oracle
- IBM Db2
- Redis
- HBase
- MySQL
- PostgreSQL
- Vertica
- TPC-DS
- Custom

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon AppFlow Connectors

Amplitude
Datadog
Dynatrace
Facebook Ads
Google Ads
Infor Nexus
Jira Cloud
Marketo
Mixpanel
Salesforce
Salesforce Marketing Cloud
Salesforce Pardot
SAP OData / ODP
ServiceNow
Singular
Slack
Snowflake
Trend Micro
Upsolver
Veeva Vault PromoMats
Zendesk (Ticketing)
Zendesk Chat
Zendesk Sell

AWS


Amazon S3


Amazon Redshift


Amazon EventBridge


Amazon Connect

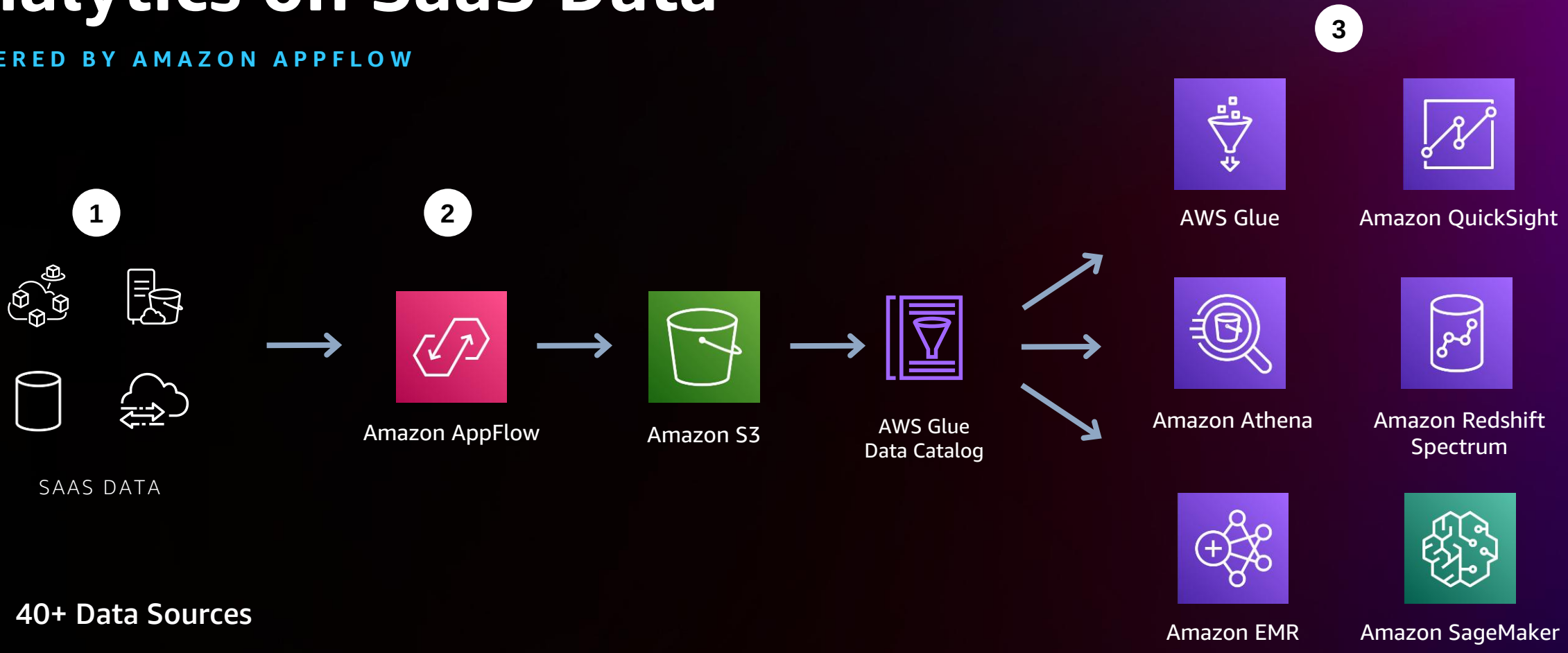

Amazon Lookout for Metrics


Amazon Honeycode

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Analytics on SaaS Data

POWERED BY AMAZON APPFLOW



1. 40+ Data Sources
2. Automated Data Cataloging and Partitioning
3. Integration with AWS Analytics and ML Services

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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 …

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Scott Rigney
athena-feedback@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.