AWS re:Invent

ANT391

Building Serverless Analytics Solutions with Amazon QuickSight

Quinton Alsbury
PM Amazon QuickSight

AWS re:Invent

aws

AWS database and analytics stack

One of the broadest and deepest portfolios purpose-built for builders

Business intelligence & machine learning

Amazon QuickSight (BI)

Amazon SageMaker (ML)

Relational databases

Aurora

Amazon RDS

Non-relational databases

DynamoDB
(Key value/Document)

ElastiCache
(Redis, Memcached)

Analytics

DW | Big data processing | Ad hoc

Amazon Redshift

Amazon EMR

Athena

Real-time

Amazon ES

Kinesis Data Analytics

Data lake

Amazon S3/ Amazon Glacier
(Storage)

AWS Glue
(ETL & Data Catalog)

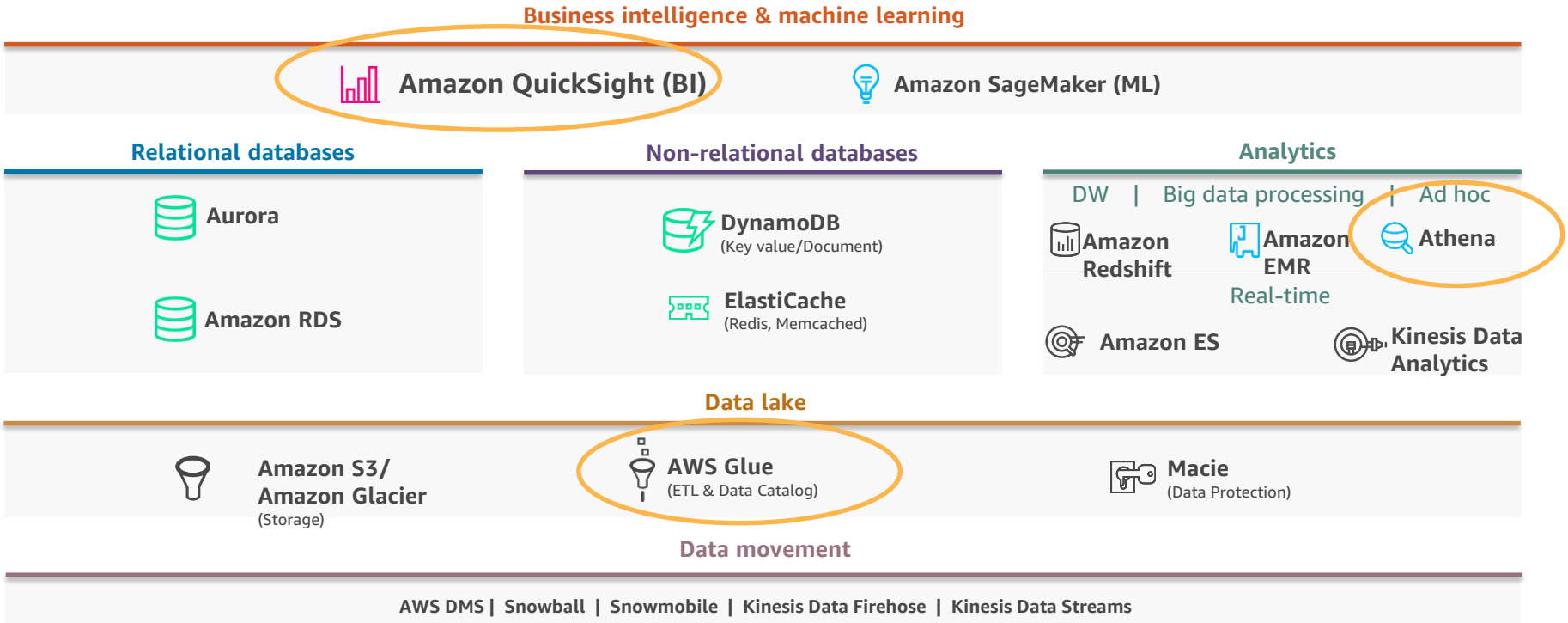
Macie
(Data Protection)

Data movement

AWS DMS | Snowball | Snowmobile | Kinesis Data Firehose | Kinesis Data Streams

AWS database and analytics stack

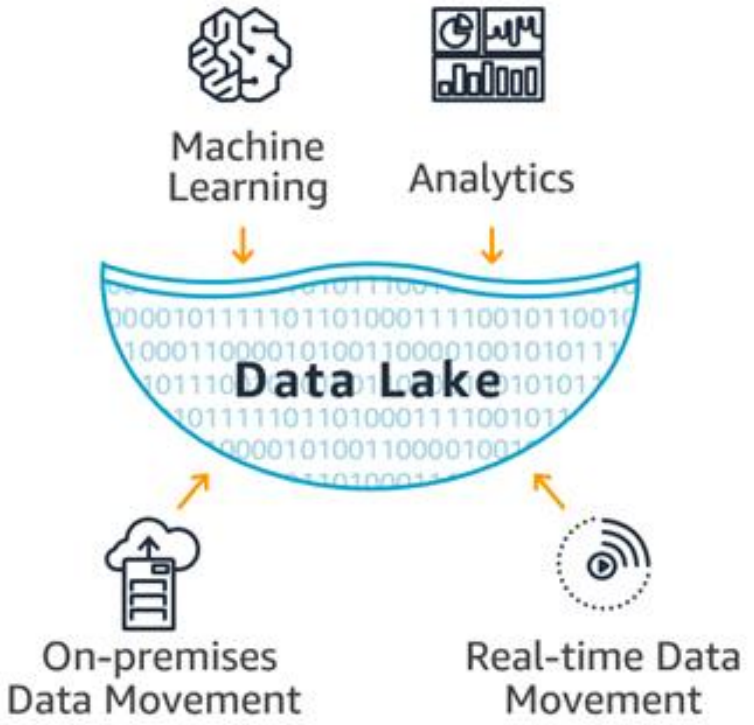
One of the broadest and deepest portfolios purpose-built for builders



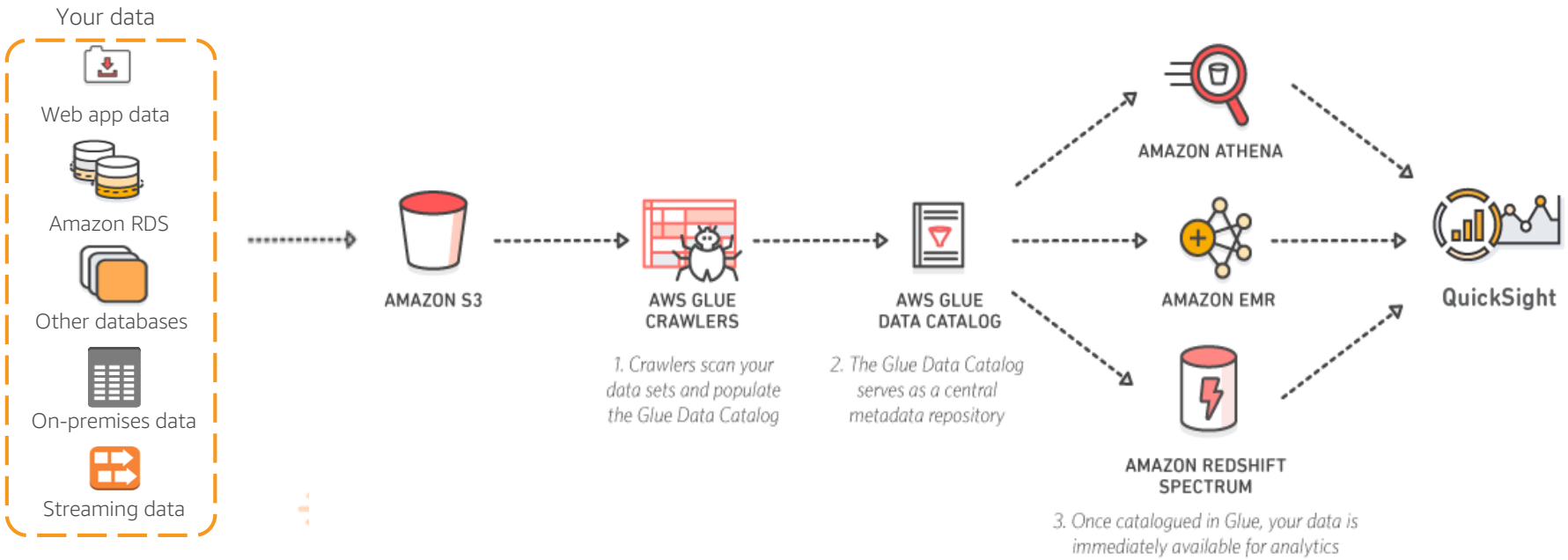
What is a data lake

Architectural pattern enabling

- Ubiquitous storage at any scale
- Consolidated data processing
- Collaborate and analyze data in different ways, leading to better, faster decision making



Serverless data lake & analytics with AWS



AWS serverless data lake & analytics overview

- AWS Glue – Serverless ETL and Data Catalog
- Amazon Athena – Serverless SQL query engine
- Amazon QuickSight – Serverless business intelligence / visualization platform

AWS Glue

Simple, flexible, cost-effective ETL

- AWS Glue is a fully managed ETL (extract, transform, and load) service
- Categorize your data, clean it, enrich it, and move it reliably between various data stores
- Once catalogued, your data is immediately searchable and queryable across your data silos
- Simple and cost-effective
- Serverless; runs on a fully managed, auto-scaling Spark environment

AWS Glue enables you to

Discover

• Automatically discover and categorize your data, making it immediately searchable and queryable across data sources

Develop

• Generate code to clean, enrich, and reliably move data between various data sources. Easily customize this code or bring your own

Deploy

• Run your jobs on a serverless, fully managed, scale-out environment. No compute resources to provision or manage

What is AWS Glue Data Catalog?

Unified metadata repository across relational databases, Amazon Relational Database Service (Amazon RDS), Amazon Redshift, and Amazon Simple Storage Service (Amazon S3) . . . with support for more coming soon!

- Get a single view into your data, no matter where it is stored
- Automatically classify your data in one central list that is searchable
- Track data evolution using schema versioning
- Query your data using Amazon Athena or Amazon Redshift Spectrum
- Hive metastore compatible; can be used as an external Hive metastore for applications running on Amazon EMR

What are crawlers?

- Crawlers automatically build your Data Catalog and keep it in sync
- Scan your data stored in various data stores, extract metadata and data statistics, and add table definitions to your Data Catalog
- Classify data using built-in and custom classifiers
- You can write your own using grok expressions
- Discover new data, extracts schema definitions
- Detect schema changes and version tables
- Detect Hive style partitions on Amazon S3
- Run as needed or on a schedule; serverless – Only pay when crawler runs

Amazon Athena

Amazon Athena is an interactive query service that makes it easy to analyze data directly from Amazon S3 using standard SQL

Athena value proposition

- Decouple storage from compute
- Serverless – No infrastructure or resources to manage
- Pay only for data scanned
- Schema on read – Same data, many views
- Secure – AWS Identity and Access Management (IAM) for authentication; Encryption at rest
- Standard compliant and open storage formats
- Built on powerful, community-supported OSS solutions

Familiar technologies under the covers

Used for DDL functionality
Complex data types
Multitude of formats
Supports data partitioning

Used for SQL queries
In-memory distributed query engine
ANSI-SQL compatible with extensions


QuickSight
Amazon QuickSight is a fully managed, serverless, cloud business intelligence system

Amazon QuickSight overview

Scalable

Amazon QuickSight automatically scales with your usage and activity, with no need for additional infrastructure. From 10 users to 10,000, Amazon QuickSight seamlessly grows with you

Pay for what you use

Pay monthly or annually. Now with pay-per-session pricing, data consumers only pay when they access their reports and dashboards with no up-front costs!

Serverless and fully managed

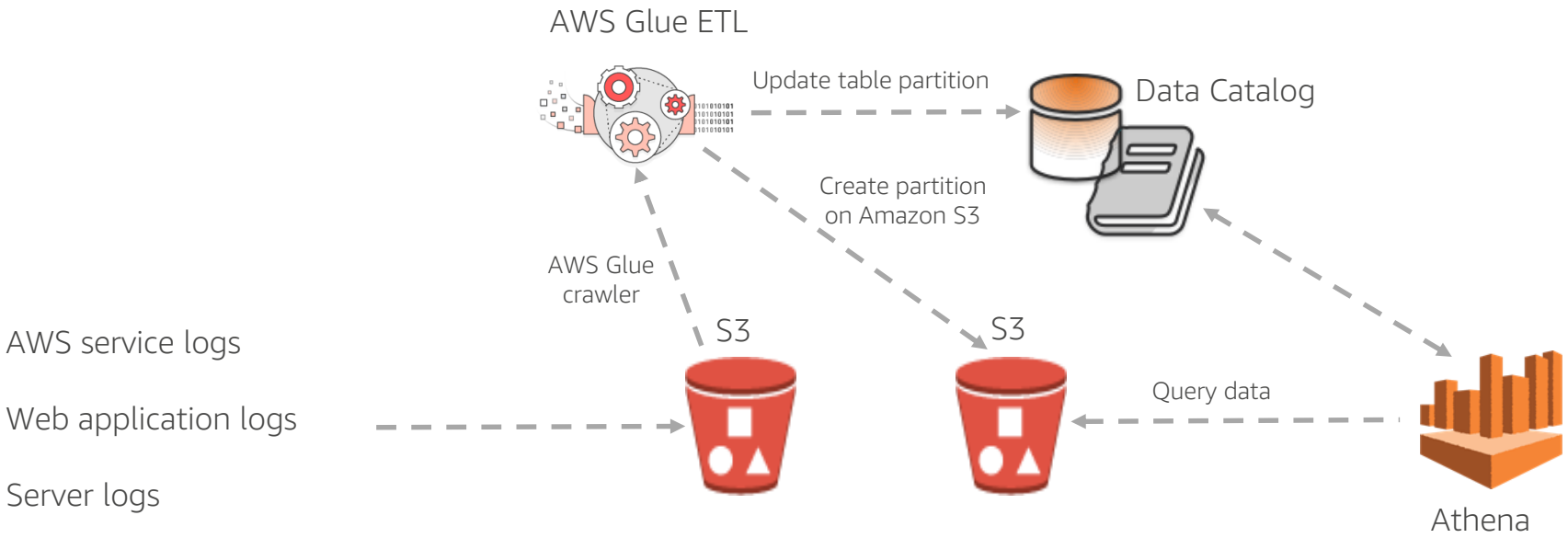
Amazon QuickSight is a fully managed cloud application, meaning there's no upfront cost, software to deploy, capacity planning, maintenance, upgrades, or migrations

Fully integrated

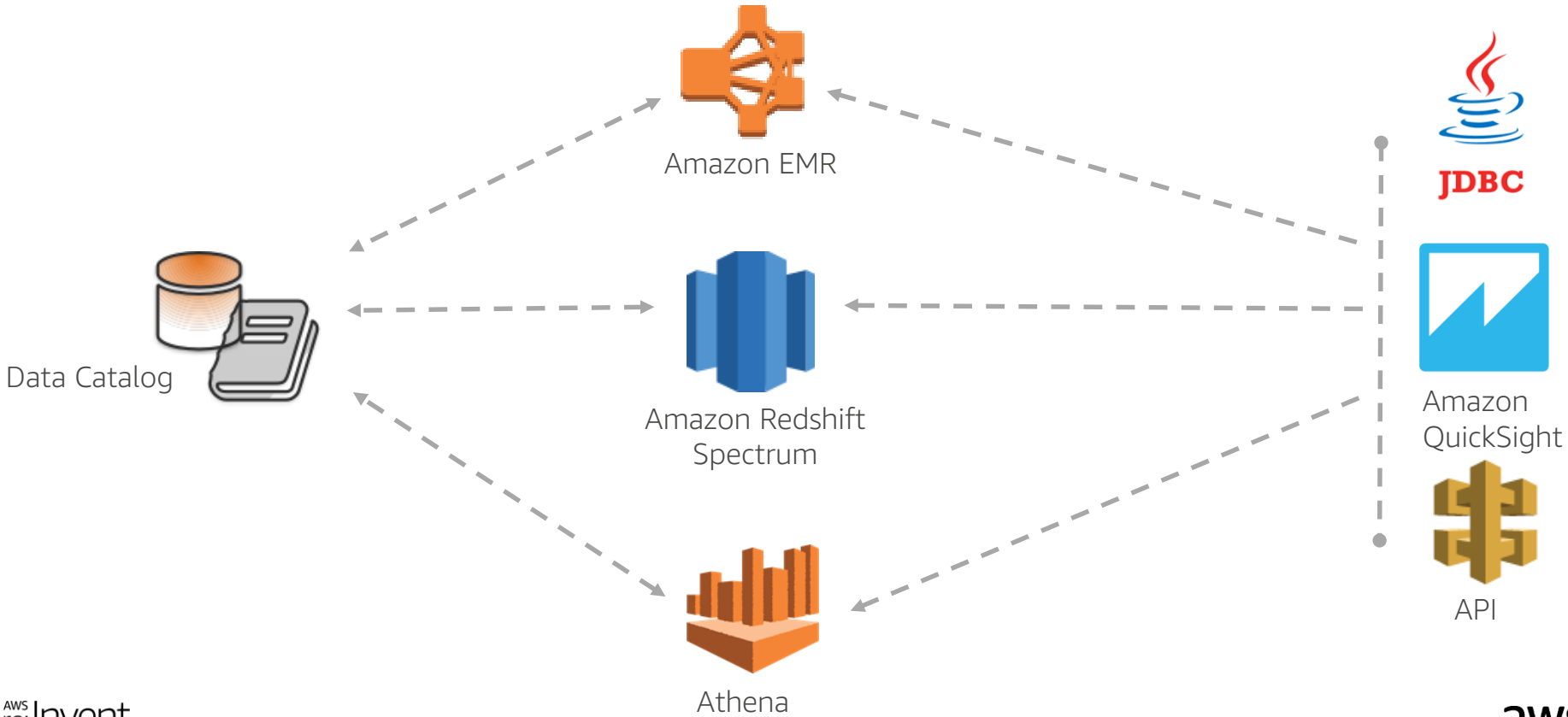
Amazon QuickSight is deeply integrated with your data sources and other AWS services like Amazon Redshift, Amazon S3, Athena, Amazon Aurora, Amazon RDS, IAM, AWS CloudTrail, Amazon Cloud Directory, and more – providing you with everything you need for an end-to-end cloud BI solution

Demo

Log aggregation with ETL



Analytics reporting



Thank you!

Quinton Alsbury
qalsbury@amazon.com

AWS re:Invent

aws

Please complete the session survey in the mobile app.