AWS State, Local, and Education Miami Learning Days

Building a Modern Data Strategy

Gabriel Brackman

Sr. Solutions Architect
Amazon Web Services
brackmg@amazon.com

Hina Siddiqui

Sr. Manager – Customer Solutions
Amazon Web Services
hinasid@amazon.com

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

The data challenge

The opportunity

The modern data strategy journey

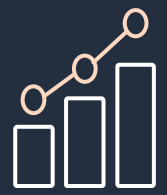
Modern data architectures on AWS

Sample reference architectures

How to get there

© 2023, Amazon Web Services, Inc. or its affiliates.

The data challenge



Availability of electronic data is growing exponentially

Data coming from new, disconnected sources

Increasingly diverse in file type and volume

Used by many people (e.g. policy makers, researchers, etc.)

Analyzed by many applications

© 2023, Amazon Web Services, Inc. or its affiliates.

A data-driven organization means...

Data is an organizational asset

No longer kept in silos or as the property of individual departments

Data is accessible

Available easily and securely to anyone who needs access to it

Data is put to work

Used in analytics and ML to make better decisions, create efficiencies, and drive new innovations

© 2023, Amazon Web Services, Inc. or its affiliates.

Create better business outcomes with data

Make better, faster decisions

Improve customer experience

Prepare for the future

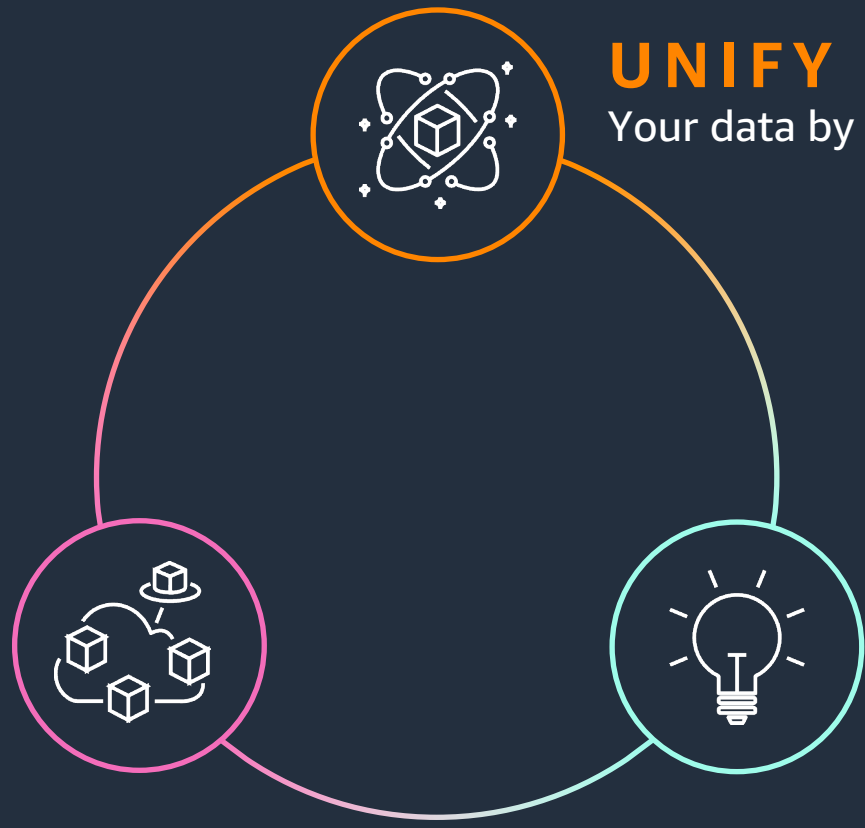
Reduce costs and improve productivity

© 2023, Amazon Web Services, Inc. or its affiliates.

© 2023, Amazon Web Services, Inc. or its affiliates.

Modern data strategy for better business outcomes

Start anywhere



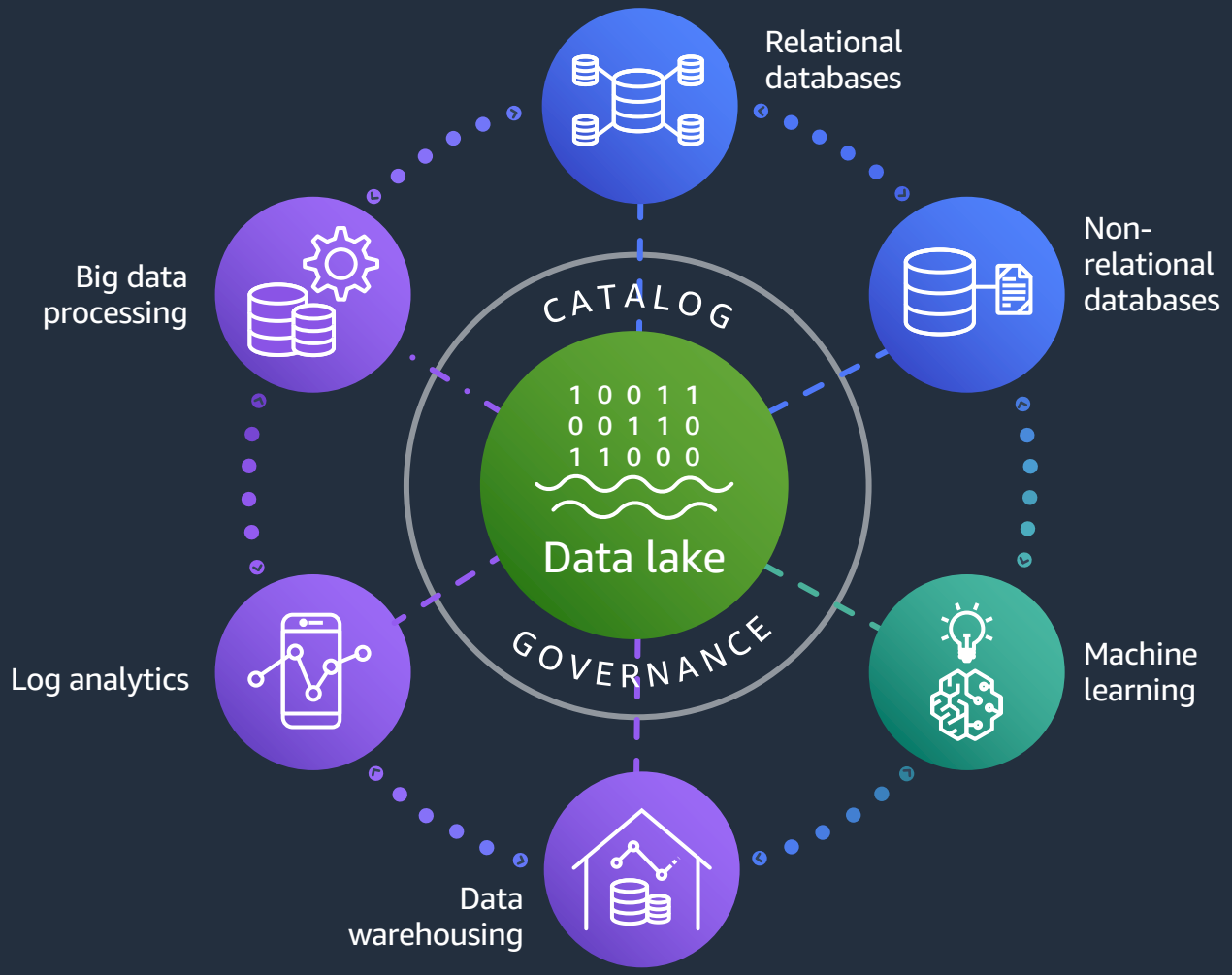
UNIFY
Your data by breaking down silos

INNOVATE
By inventing new experiences and reimagining existing processes

MODERNIZE
Your data infrastructure to a scalable, trusted, and secure cloud provider

© 2023, Amazon Web Services, Inc. or its affiliates.

A modern data strategy



Data at any scale

The best price-performance

Seamless data access

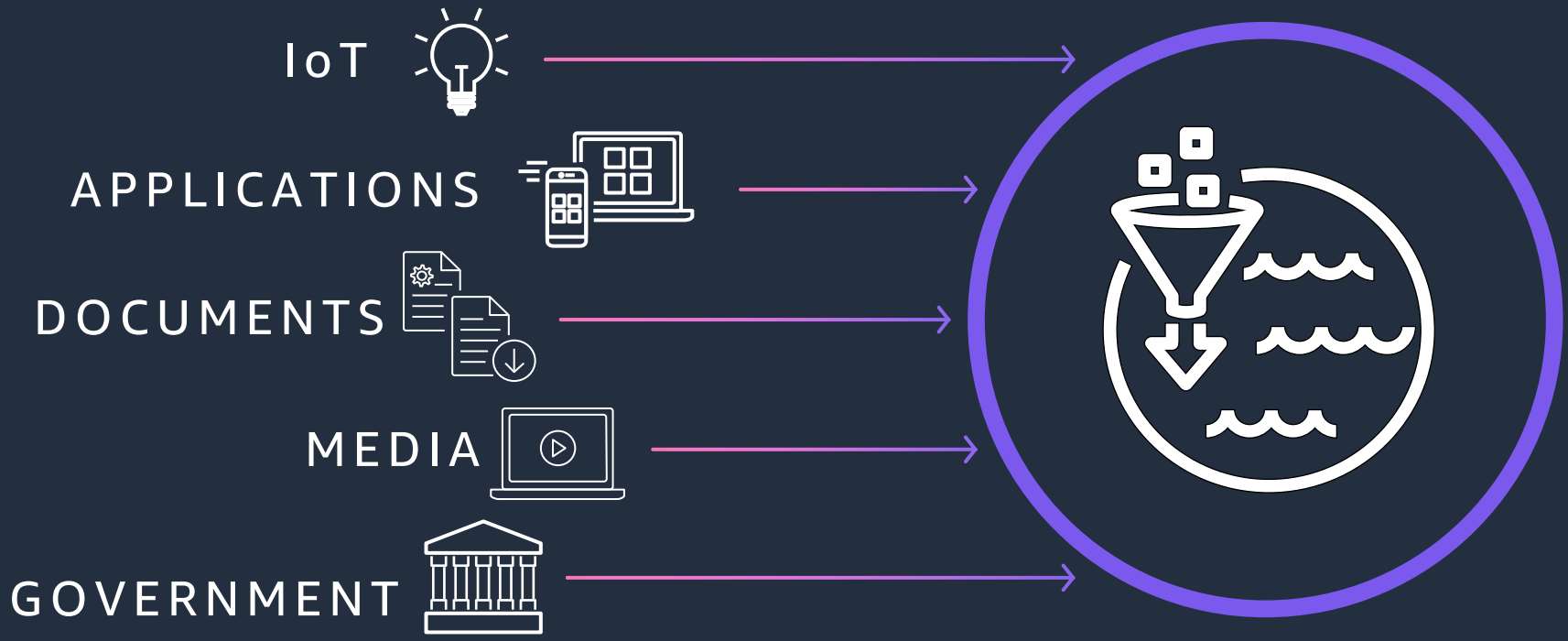
Unified governance

Artificial Intelligence (AI) & Machine Learning(ML) to solve business challenges

aws

© 2023, Amazon Web Services, Inc. or its affiliates.

Data lakes are foundational for data unification



© 2023, Amazon Web Services, Inc. or its affiliates.

To get more value from their data, customers are..

Breaking free from legacy databases

Moving to fully managed database and analytics services

Modernizing your data warehouse

Building modern applications with purpose-built databases

© 2023, Amazon Web Services, Inc. or its affiliates.

Purpose-built data services

Optimize performance, cost, and scale for your use cases

Amazon Athena

Interactive query

Amazon EMR

Big data processing

Amazon OpenSearch Service

Log and search analytics

Amazon Kinesis and Amazon MSK

Real-time analytics

Amazon Redshift

Data warehousing

© 2023, Amazon Web Services, Inc. or its affiliates.

Build new experiences and reimagine old processes with AI/ML

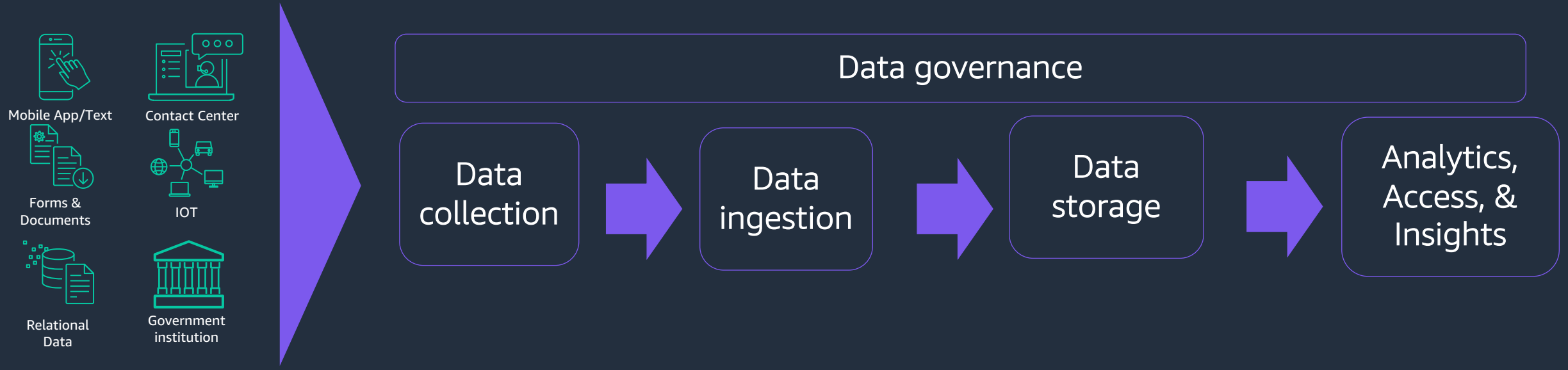
- Make accurate predictions, get deeper insights from your data, and improve customer experience
- Create ML predictions without any ML experience or writing any code
- Build applications with our pre-trained models
- Train and apply your own models
- Use your own algorithms by working directly with ML-optimized AWS infrastructure
- 100,000+ customers use AWS AI and ML services to make predictions from their data

© 2023, Amazon Web Services, Inc. or its affiliates.

Putting it all together

© 2023, Amazon Web Services, Inc. or its affiliates.

Key components of modern data architecture



Key considerations:

1. Ability to handle the increasing volume, velocity, and variety of data
2. Each component should be independently scalable
3. Make data easily accessible and sharable

© 2023, Amazon Web Services, Inc. or its affiliates.

AWS data lakes provide a flexible foundation for analytics and innovation

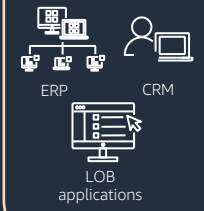
Data catalog
Crawls and catalogs your data; discover, prepare, and combine data for analytics and ML

AWS Glue

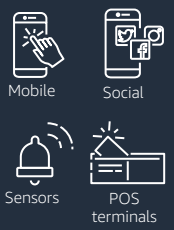
Data governance
Centralized authorization layer to define data sources and data access & security policies

AWS Lake Formation

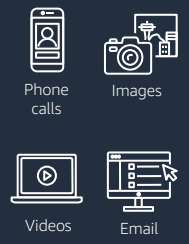
Structured data
Data that are highly normalized with common schema and stored in relational databases, powering transactional line-of-business applications



Semistructured data
Data that contain identifiers without conforming to a predefined schema



Unstructured data
Data that do not conform to a data model and are typically stored as individual files



Data collection
Purpose built databases for different workloads

- RDS
- Aurora
- DynamoDB
- DocumentDB
- ElastiCache
- Neptune
- Timestream
- QLDB

Batch load
Extracts data from various data sources at periodic intervals and moves them to the data lake

- Amazon EMR
- AWS Glue
- AWS DMS
- AWS Transfer Family

Streaming
Ingests data that are generated from multiple sources such as log files, telemetry, mobile applications, and social networks

- Amazon Kinesis
- Amazon MSK

Amazon S3 data lake
Cloud-scale centralized and scalable architecture that enables enterprise data science

Amazon S3

Analytics
Leverage data warehouses, Spark, and graph databases to gain insights from your data

- Amazon Redshift
- Amazon EMR
- Amazon Neptune

And data stored in the data lake can also be made directly searchable and queryable

- Amazon Athena
- Amazon QuickSight

Machine Learning
Storing data in an Amazon S3 data lake enables customers to leverage predictive or prescriptive analytics; perform ad-hoc analyses; and use AI/ML for automation and efficiency

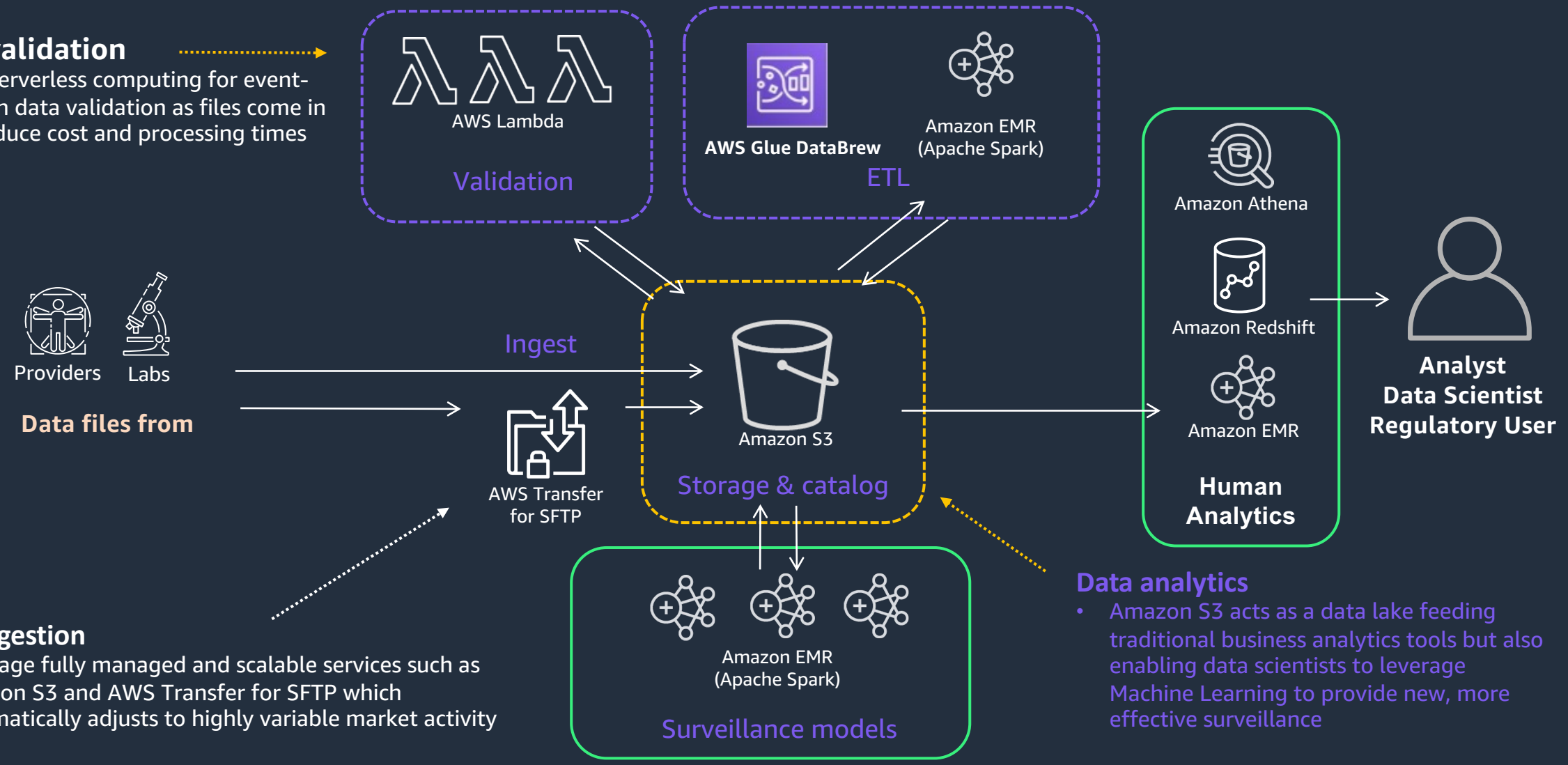
- Amazon SageMaker
- Amazon Comprehend
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- Amazon Personalize

Security – Reliability –Operational Excellence – Performance Efficiency – Cost Optimization

© 2023, Amazon Web Services, Inc. or its Affiliates.

aws

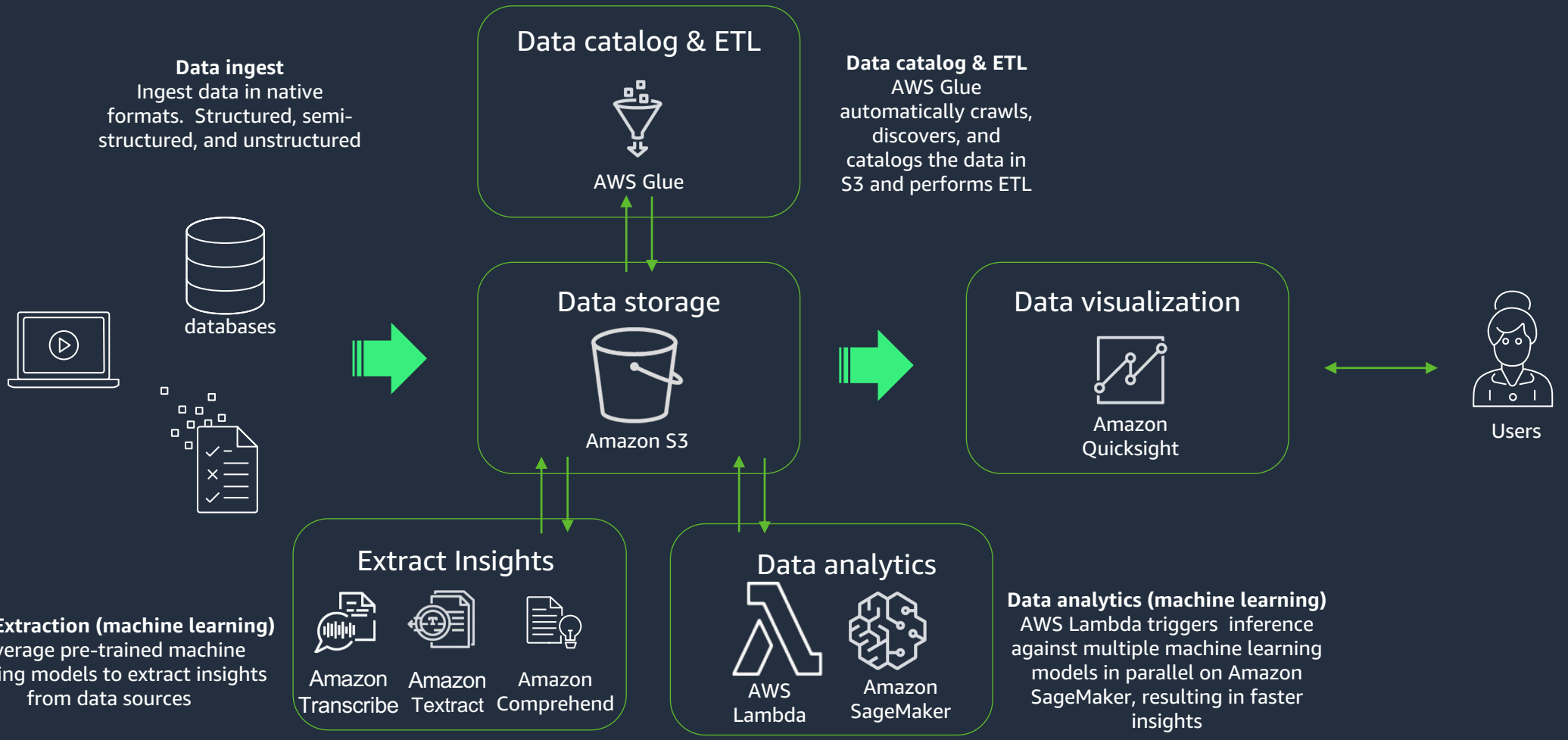
Sample reference architecture for disease surveillance



© 2023, Amazon Web Services, Inc. or its Affiliates.

Sample architecture for AI/ML machine learning

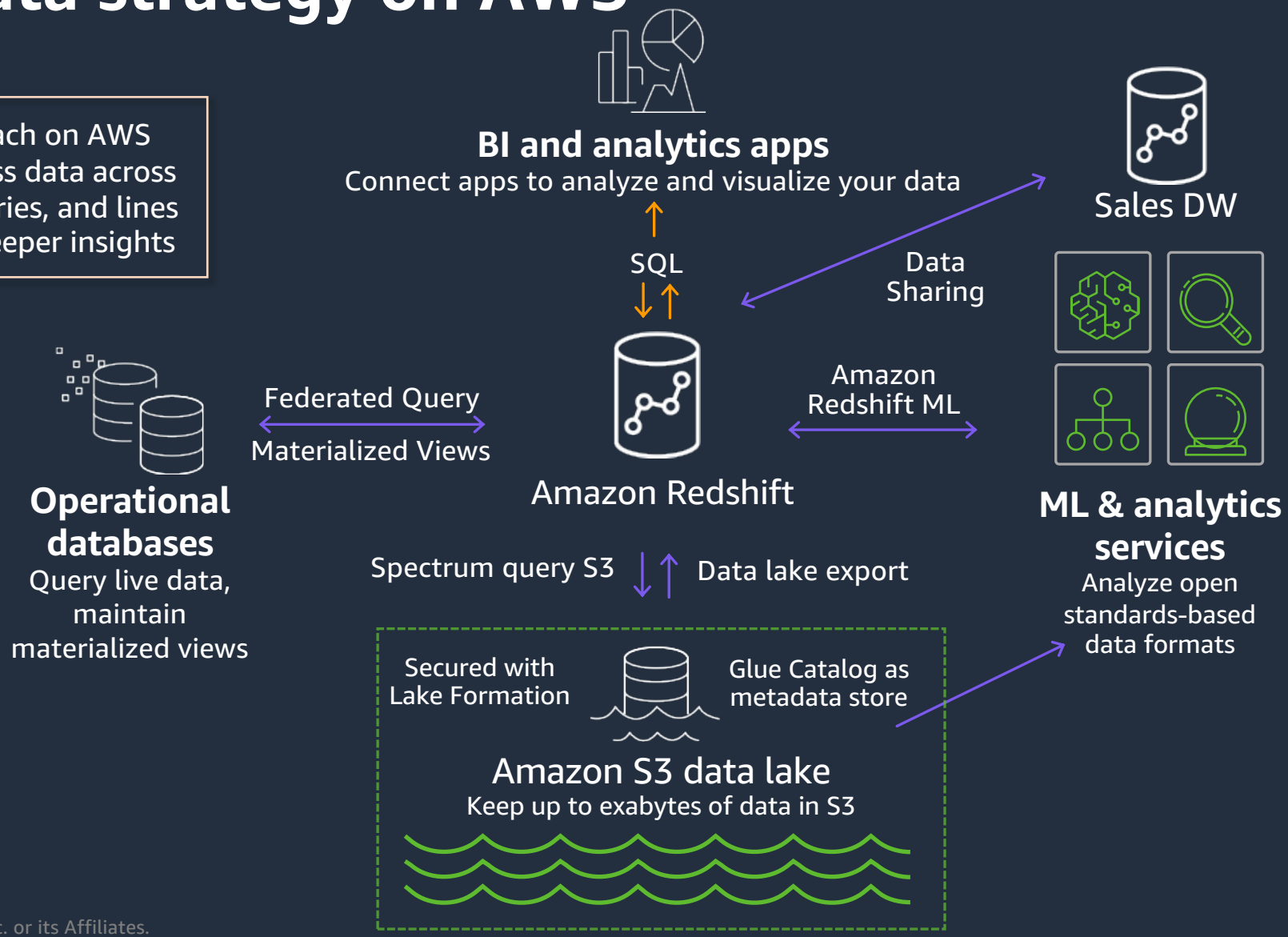
Use serverless and event-driven architectures for automatic scale to enable faster insights without the need to manage infrastructure



© 2023, Amazon Web Services, Inc. or its affiliates.

Extend the capabilities of your data warehouse with a modern data strategy on AWS

A lake house approach on AWS makes it easy to access data across applications, repositories, and lines of business to gain deeper insights



© 2023, Amazon Web Services, Inc. or its Affiliates.

Get started

BUILD WITH US

Data Lab

ML Solutions Lab

AWS Professional Services

AWS Immersion Day

Data-Driven Everything

Migration Assistance Program

BUILD WITH PARTNERS

AWS Partner Network—
100,000+ partners

AWS Marketplace (ISVs)

UPSKILL YOUR TEAMS

AWS Training and Certification

ML Embark Program

aws

© 2023, Amazon Web Services, Inc. or its affiliates.

Thank you!

Gabriel Brackman
brackmg@amazon.com

Please take the survey:

Data & analytics track
Building a modern data strategy

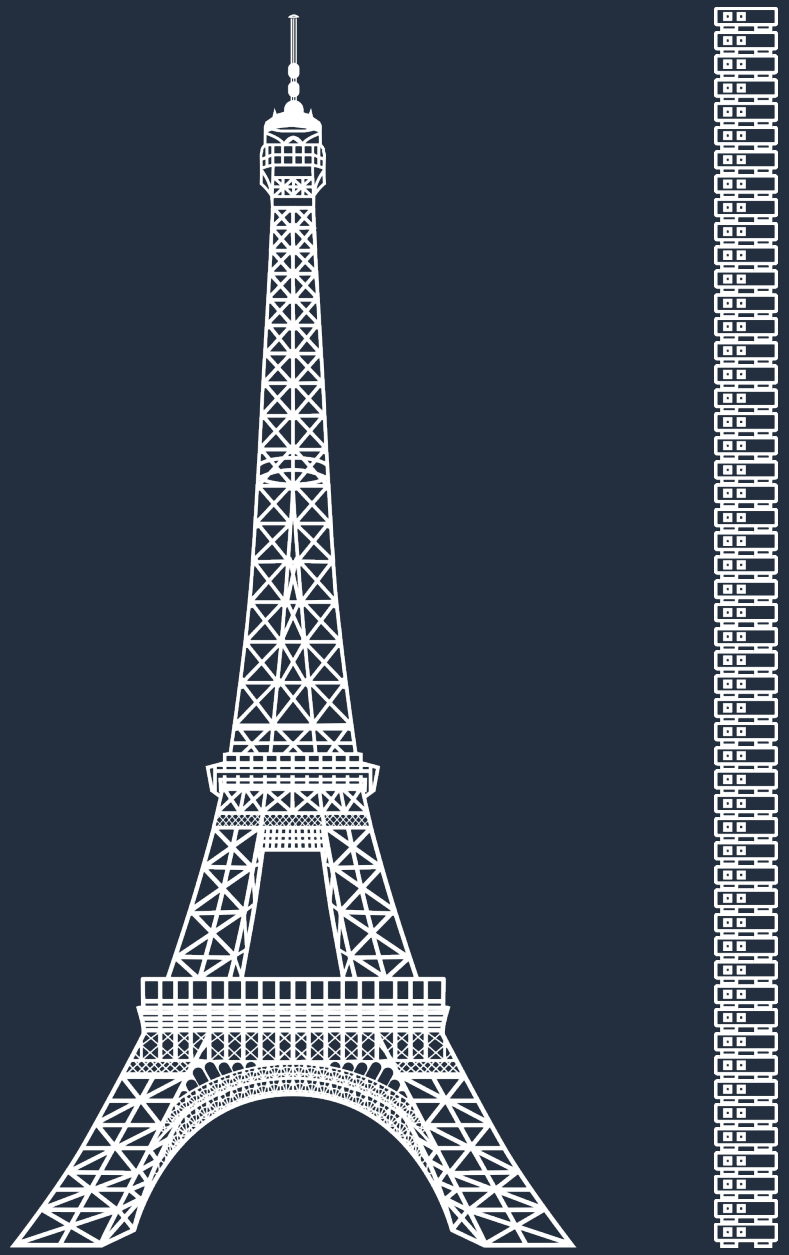
© 2024, Amazon Web Services, Inc. or its affiliates.

448 ZB

of data will be created, consumed, and stored in next three years

Statistic provided by Statista.com

© 2023, Amazon Web Services, Inc. or its affiliates.



over

20 million

times

© 2023, Amazon Web Services, Inc. or its affiliates.

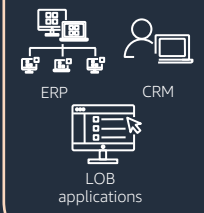
AWS data lakes provide a flexible foundation for analytics and innovation

AWS Glue

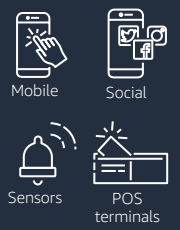
Data catalog
Crawls and catalogs your data; discover, prepare, and combine data for analytics and ML

AWS Lake Formation

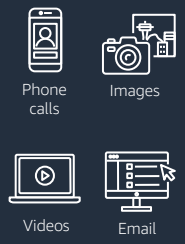
Data governance
Centralized authorization layer to define data sources and data access & security policies



Structured data
Data that are highly normalized with common schema and stored in relational databases, powering transactional line-of-business applications



Semistructured data
Data that contain identifiers without conforming to a predefined schema



Unstructured data
Data that do not conform to a data model and are typically stored as individual files

Data collection
Purpose built databases for different workloads

- RDS
- Aurora
- DynamoDB
- DocumentDB
- ElastiCache
- Neptune
- Timestream
- QLDB

Batch load
Extracts data from various data sources at periodic intervals and moves them to the data lake

- Amazon EMR
- AWS Glue
- AWS DMS
- AWS Transfer Family

Streaming
Ingests data that are generated from multiple sources such as log files, telemetry, mobile applications, and social networks

- Amazon Kinesis
- Amazon MSK

Amazon S3 data lake
Cloud-scale centralized and scalable architecture that enables enterprise data science

Amazon S3

Analytics
Leverage data warehouses, Spark, and graph databases to gain insights from your data

- Amazon Redshift
- Amazon EMR
- Amazon Neptune

And data stored in the data lake can also be made directly searchable and queryable

- Amazon Athena
- Amazon QuickSight

Machine Learning
Storing data in an Amazon S3 data lake enables customers to leverage predictive or prescriptive analytics; perform ad-hoc analyses; and use AI/ML for automation and efficiency

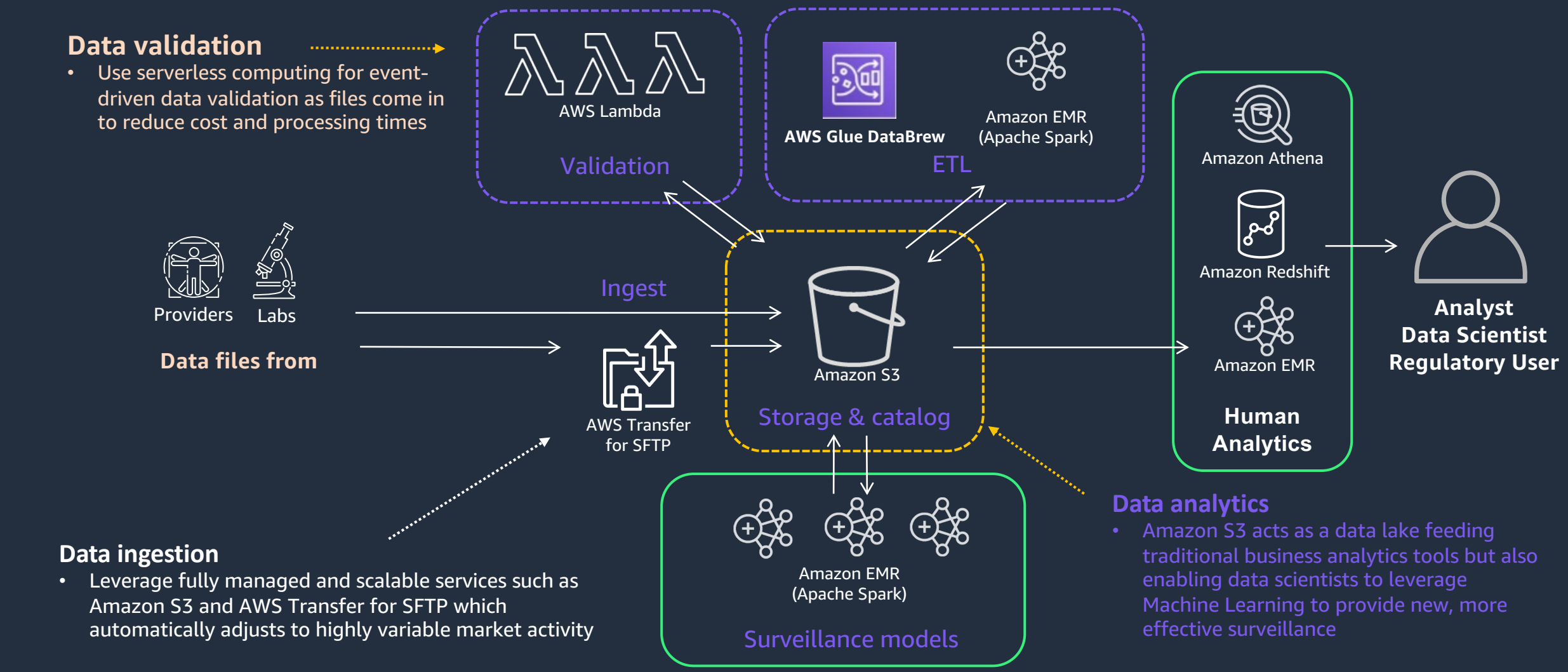
- Amazon SageMaker
- Amazon Comprehend
- Amazon Textract
- Amazon Transcribe
- Amazon Translate
- Amazon Personalize

Security – Reliability –Operational Excellence – Performance Efficiency – Cost Optimization

© 2023, Amazon Web Services, Inc. or its Affiliates.

aws

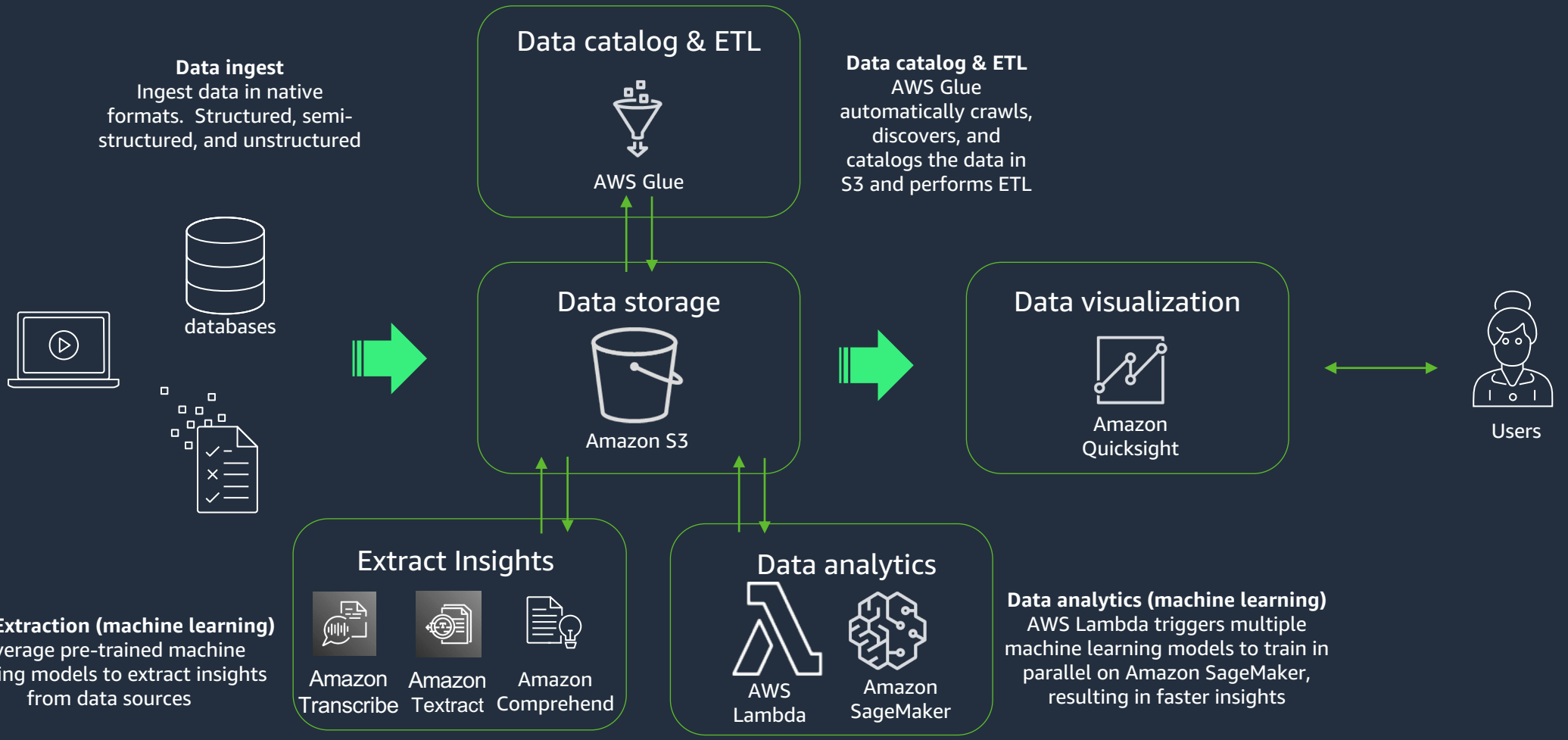
Sample reference architecture for Syndromic Surveillance



© 2023, Amazon Web Services, Inc. or its Affiliates.

Sample architecture for AI/ML Machine Learning

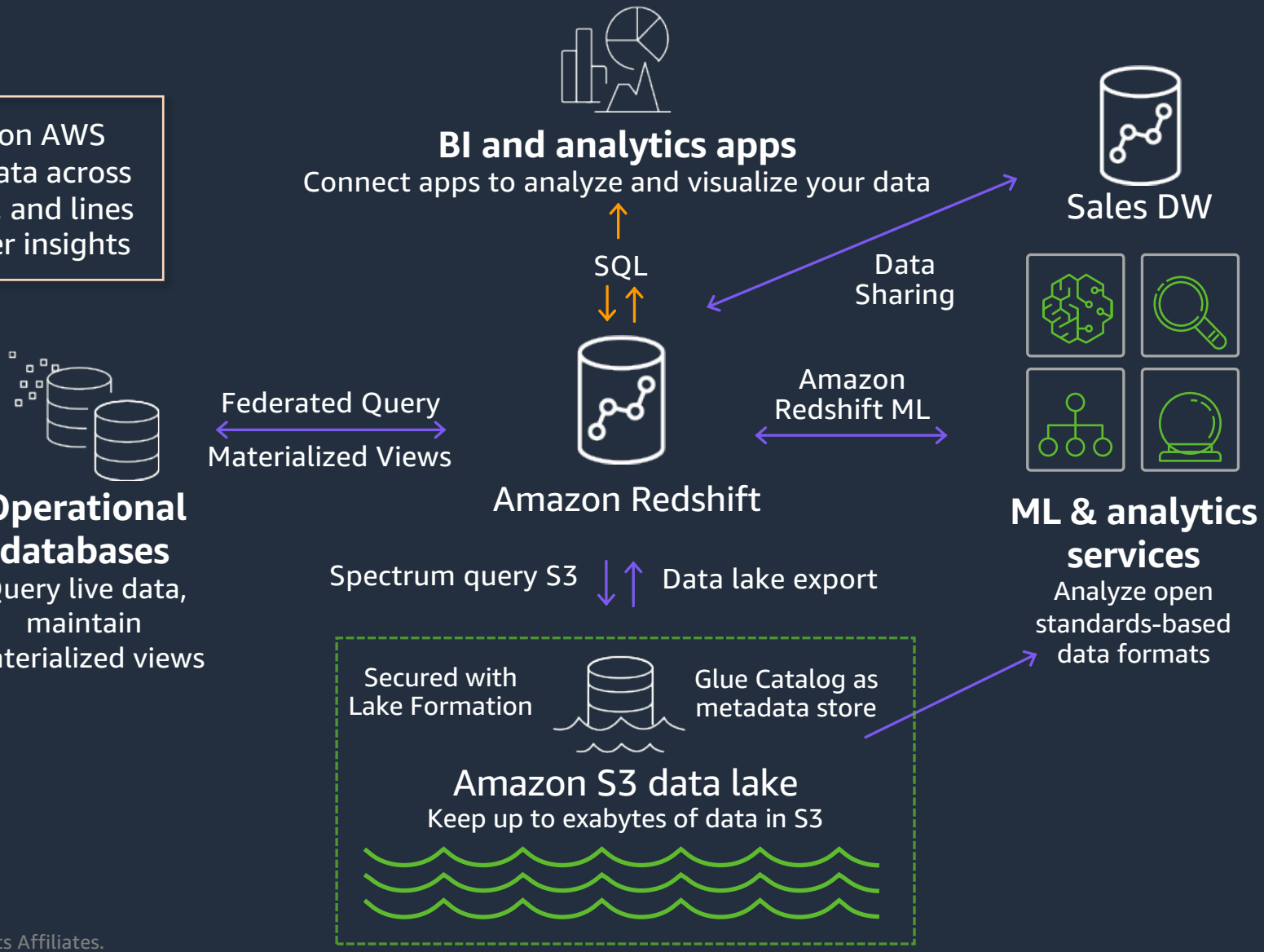
Use serverless and event-driven architectures for automatic scale to enable faster insights without the need to manage infrastructure



© 2023, Amazon Web Services, Inc. or its Affiliates.

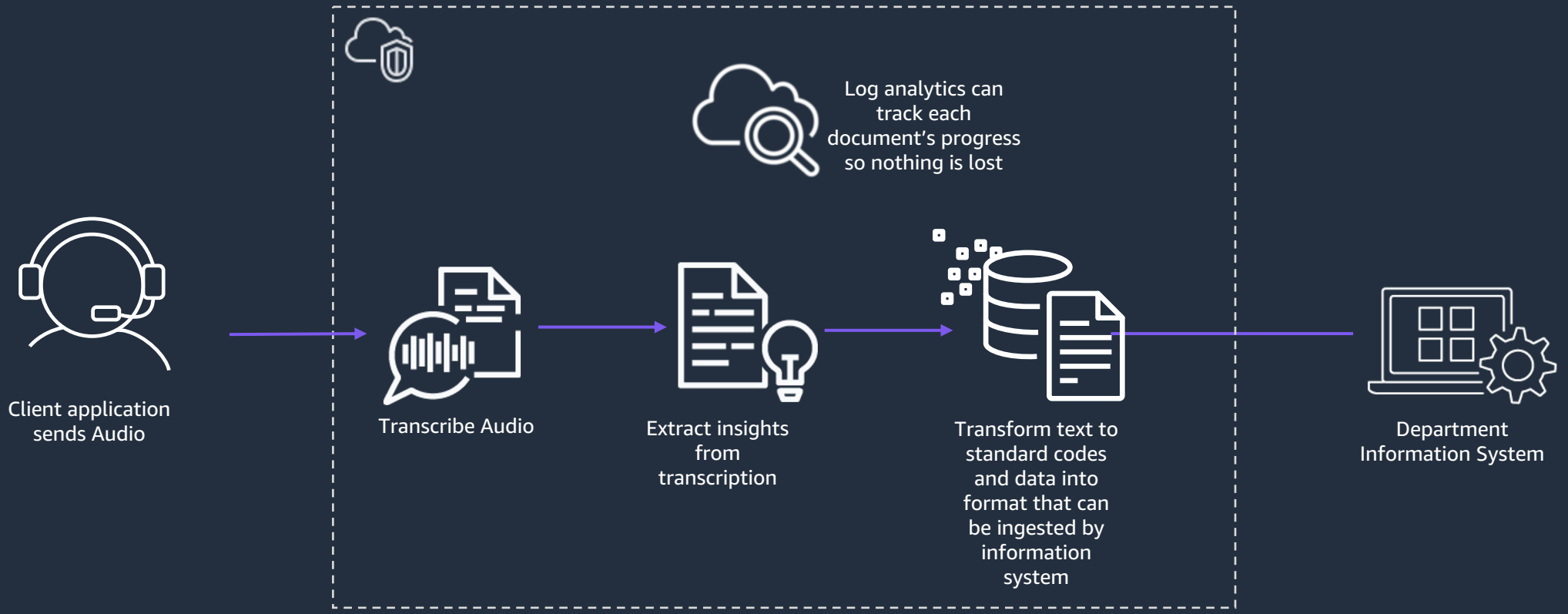
Extend the capabilities of your data warehouse with a Modern Data Strategy on AWS

A lake house approach on AWS makes it easy to access data across applications, repositories, and lines of business to gain deeper insights



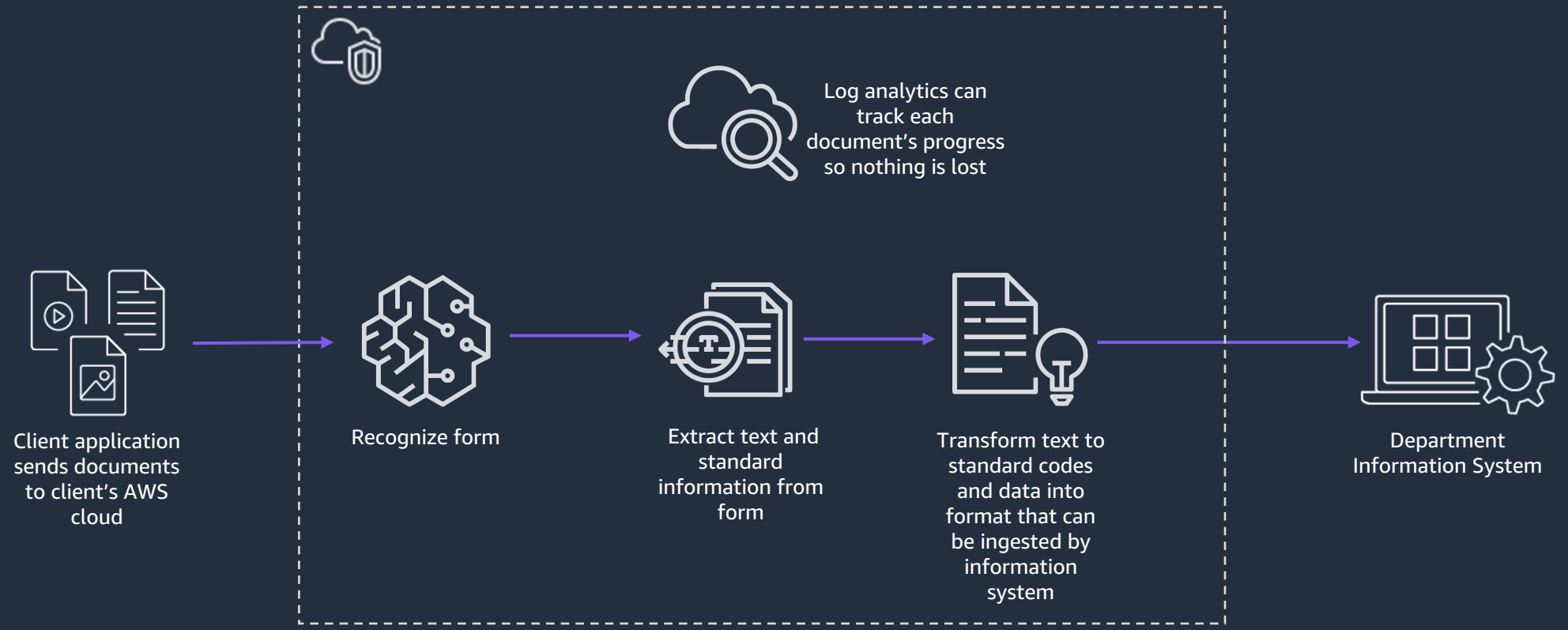
© 2023, Amazon Web Services, Inc. or its Affiliates.

Insights from Audio - High Level Process



© 2023, Amazon Web Services, Inc. or its affiliates.

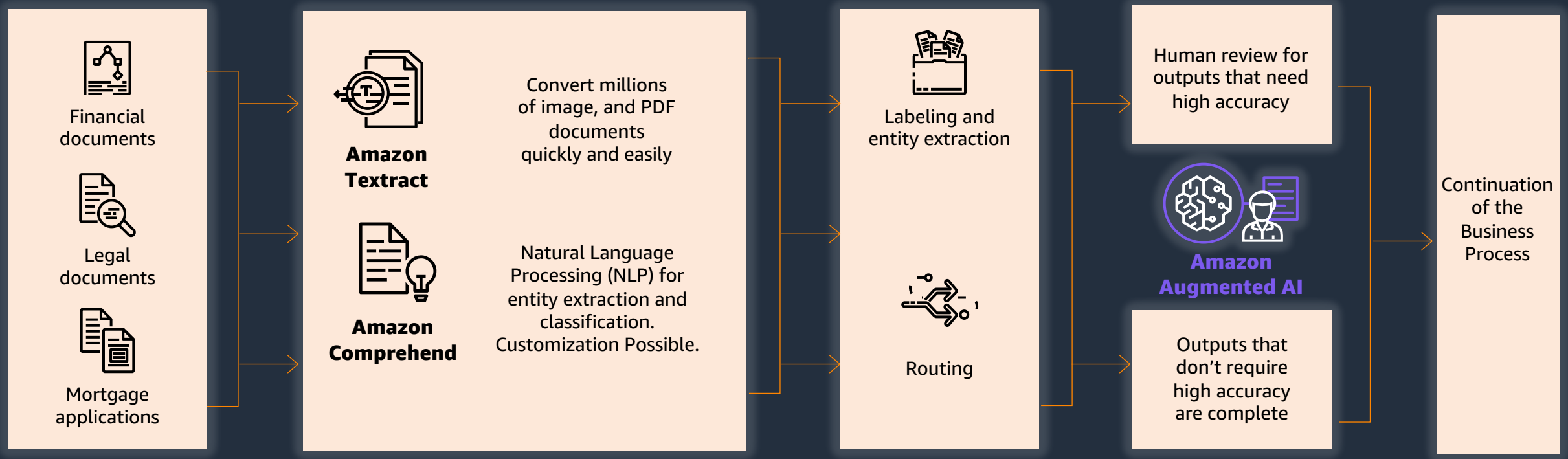
Text extraction - High Level Process



© 2023, Amazon Web Services, Inc. or its affiliates.

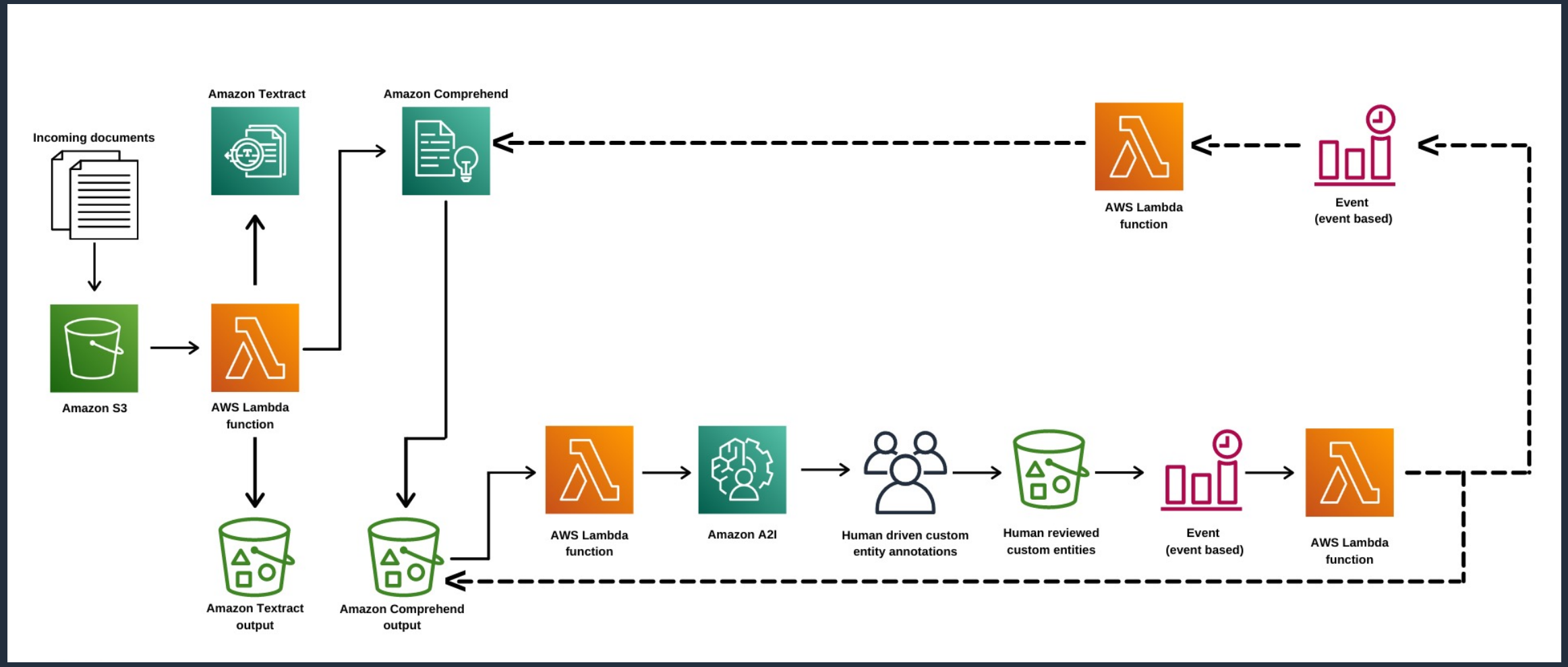
Extract and analyze data from documents

- **Amazon Textract, Amazon Comprehend, and Amazon Augmented AI**



© 2023, Amazon Web Services, Inc. or its affiliates.

IDP reference architecture example



© 2023, Amazon Web Services, Inc. or its affiliates.

Document management challenges

TONS OF DOCUMENTS

VARIED FORMATS

Invoices, receipts, contracts and many more…

DIVERSE CONTENT

Tables, paragraphs, logos, handwriting, and different languages

DOCUMENTS SOLVE DIFFERENT PROBLEMS

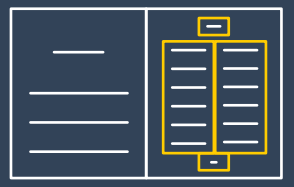
Text may be required for many reasons…

Uncovering insights

- ✓ Data Lakes
- ✓ Operational, data needs
- ✓ Chatbots
- ✓ Knowledge management

© 2023, Amazon Web Services, Inc. or its affiliates.

Document processing challenges



Extracting text manually is time-consuming, error prone, and expensive

Manual processes do not scale easily with document volume

Current rules-based systems are not intelligent and break with format changes

New documents = new OCR templates = intensive effort = long lead time

For most organizations, data in documents remains unstructured and unavailable to generate business insights

© 2023, Amazon Web Services, Inc. or its affiliates.

AWS is the best place to extract value from data and turn it into insights

The most experience

The most reliable, scalable, secure

The most comprehensive set of services

© 2023, Amazon Web Services, Inc. or its affiliates.

Additional References

Minnesota Department of Health– Data Lakes Video Case Study
Minnesota Department of Health – Public Sector Summit
A Healthier future for North Carolina - re:Invent
UDSA SNAP Case study
Utah Mainframe Modernization
Georgia Department of Community Health MMIS migration
Rhode Island Department of Labor - Citizen engagement
Utah Department of Health COVID response
Piedmont Healthcare – Epic Systems migration
Great Lakes Health Connect
State of Maryland transforming social service - MDThink
NYC COVID-19 Response Coalition
Minnesota Department of Human Services – PEBT program
UC Dan Diego - Using ML to detect pneumonia

© 2023, Amazon Web Services, Inc. or its Affiliates. Amazon Confidential.

Please complete the session survey by scanning the QR code

Mission Track
Modern Data Strategy

© 2023, Amazon Web Services, Inc. or its affiliates.