



AWS Learning Days

Building a Modern Data Strategy

Elizabeth Davis

Senior Solutions Architect
Amazon Web Services

Agenda

- Why modern data architecture
- Modern data strategy
- Reference architectures for common scenarios
- Getting started

***"If we have data, let's look
at the data.
If all we have are opinions,
let's go with mine"***

Jim Barksdale
CEO of Netscape



Data is just the Building Blocks

Data



Information



Insights



Without structure, tools and processes,
Data has very little value

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

Current state

- Currently, decision-making revolved around the **enterprise data warehouse**



Data no longer scales

There is more data and more diversity of data than people think

Data growth

>10x
every 5 years

Data
platforms needs

To live for
15+
years

To scale
1,000x



IDC, "Data Age 2025"



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

Accessibility of data



Data scientists



Business users



Analysts



Applications



Machine learning



SQL analytics



Scientific

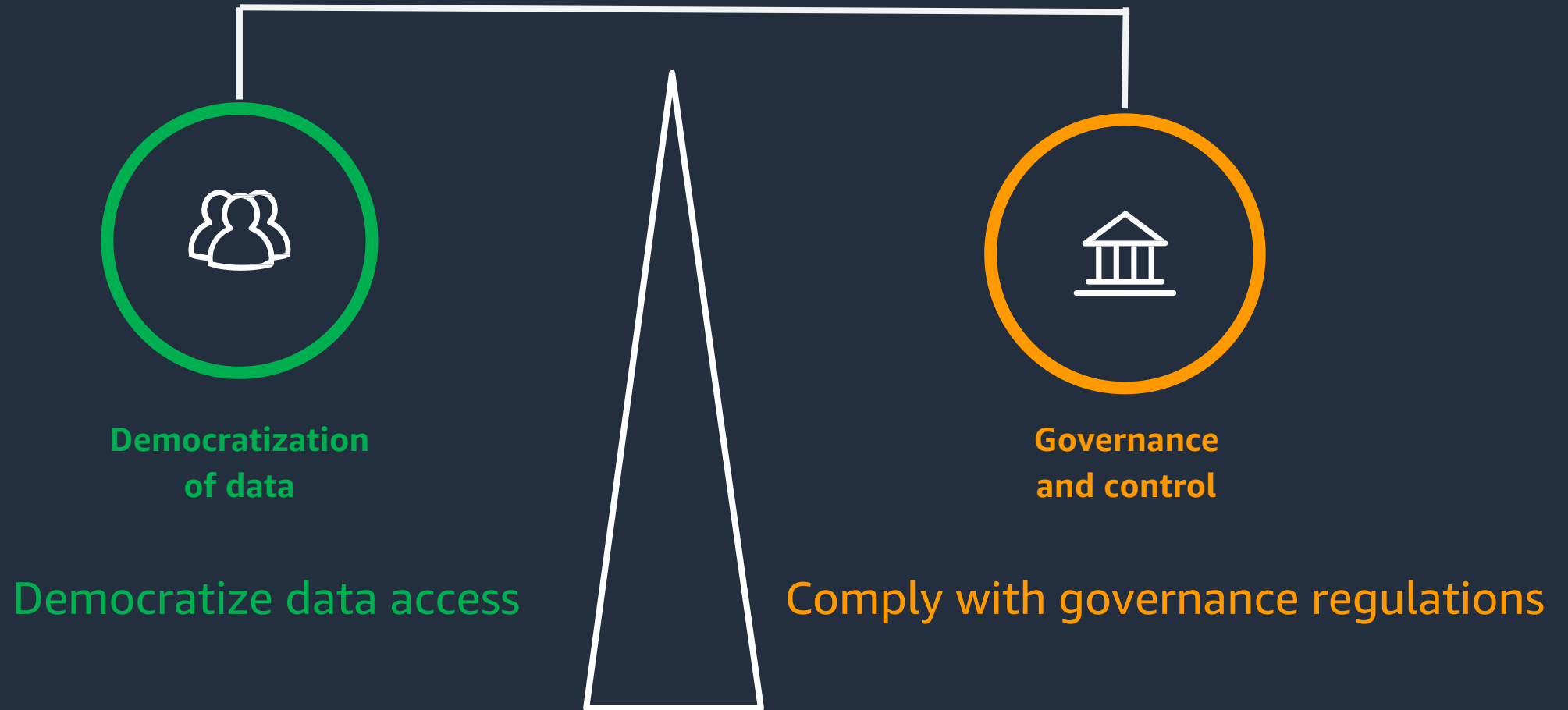


Real-time, streaming

There are **more people**
accessing data

And in **different ways**

More regulatory pressure



What now? **Let's rethink everything**

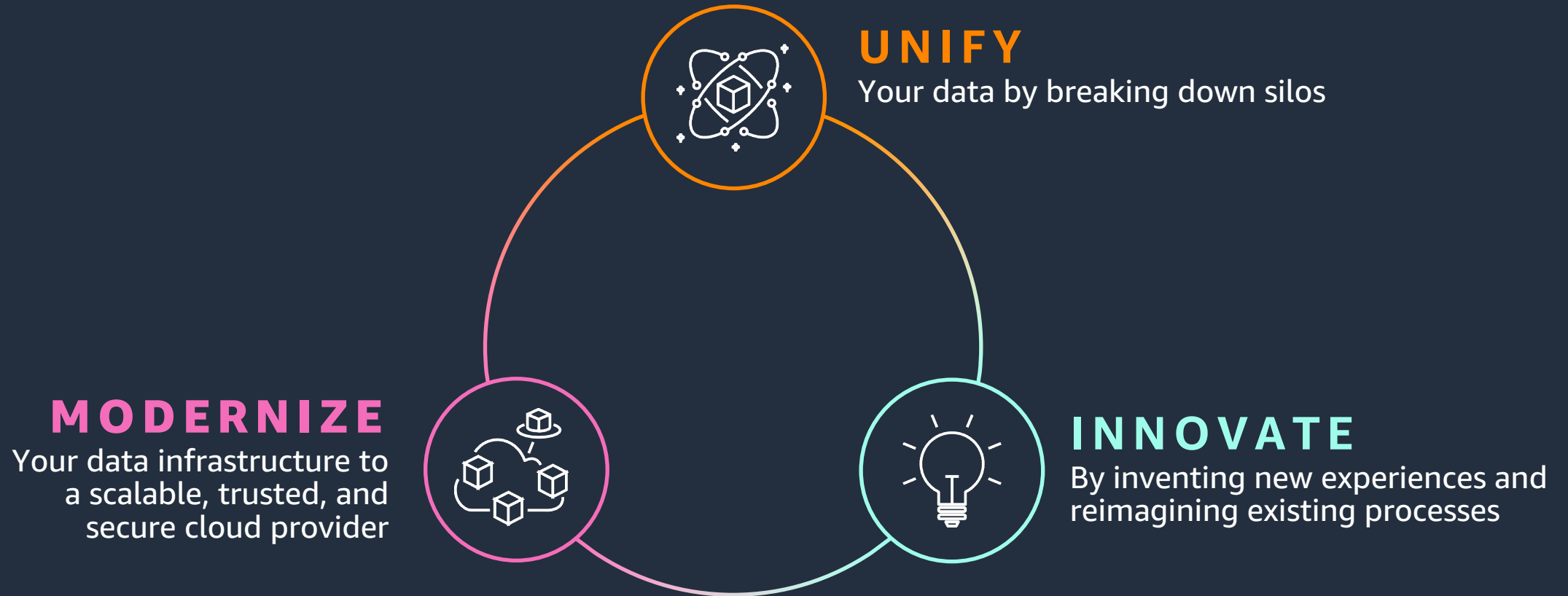


Raw Data



Insights

Modern data strategy for better business outcomes



AWS modern data strategy components



Scalable
data lakes



Unified data
access, security,
and governance



Purpose-built
data services for
performance
and cost



Serverless
and easy to
use



Built-in
machine
learning

Create better business outcomes with data



Make better, faster decisions



Improve customer experience



Prepare for the future



Reduce costs and improve productivity

Examples



Create better citizen & student experiences & outcomes



Student success
& Community
relevance



Respond to the unexpected



Support
research in the
swine industry



Unifying data to enable 360-degree
views



Transforming
human services



Enhancing efficiency



Create end-to-
end visibility

Port of Long Beach



Unify Data with Scalable data lakes

Amazon S3: Data lakes on AWS

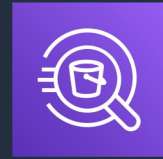
- Store unlimited data in open file formats
- Unmatched durability, availability, and scalability
- Decouple storage from compute
- Choice of analytical and ML engines
- Pay as you go



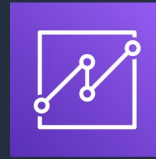


Unified data access, security, and governance

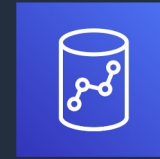
AWS Lake Formation: Unified data governance



Amazon
Athena



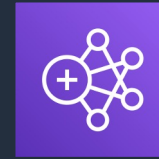
Amazon
QuickSight



Amazon
Redshift

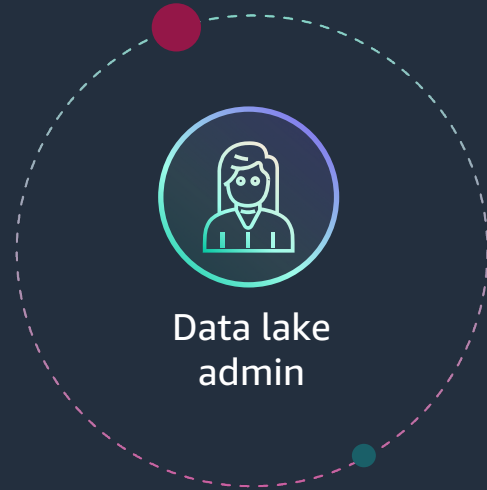


Amazon
SageMaker

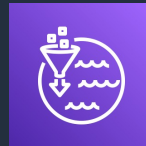


Amazon
EMR

Simplified and unified
security management



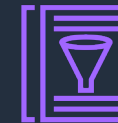
Data lake
admin



AWS Lake
Formation



Access
control



AWS Glue
Data Catalog



Amazon S3 data lake storage

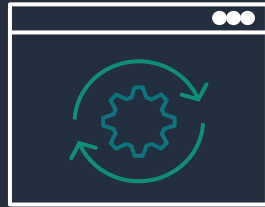


Modernize with Purpose-built data services

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

A family of purpose-built data services

Business intelligence and machine learning

**Amazon QuickSight**
Visualizations

**AWS Data Exchange**
Data exchange

**Amazon SageMaker**
ML

**Amazon Comprehend**
NLP

**Amazon Transcribe**
Speech-to-text

**Amazon Textract**
Extract text

**Amazon Personalize**
Recommendations

**Amazon Forecast**
Forecasts

**Amazon Translate**
Translation

**Amazon CodeGuru**
Code reviews

**Amazon Kendra**
Enterprise search

Plus many more

Analytics

**Amazon Redshift**
Data warehousing

**Amazon EMR**
Hadoop + Spark

**Amazon Athena**
Interactive analytics

**Amazon OpenSearch Service**
Operational analytics

**Amazon Kinesis Data Analytics**
Real time

Databases

**Amazon Aurora**
MySQL, PostgreSQL

**Amazon DynamoDB**
Key value, Document

**Amazon DocumentDB**
Document

**Amazon ElastiCache**
Redis, Memcached

**Amazon Keyspaces (for Apache Cassandra)**
Wide column

**Amazon Neptune**
Graph

**Amazon QLDB**
Ledger database

**Amazon Timestream**
Time series

Blockchain

**Amazon Managed Blockchain**

**AWS Blockchain Templates**

Data lake

**Amazon S3/ Amazon S3 Glacier**

AWS Lake Formation
Data lakes

**AWS Glue**
ETL and Data Catalog

Data movement

AWS DMS | AWS Transfer family | AWS Glue | AWS Snowball | Amazon Kinesis Data Firehose | Amazon Kinesis Data Streams | Amazon MSK



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



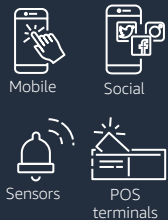
Data governance

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



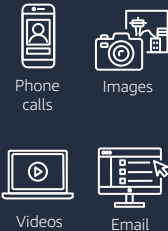
Structured data

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



Batch load

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



Streaming

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



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



Innovate with AI & machine learning

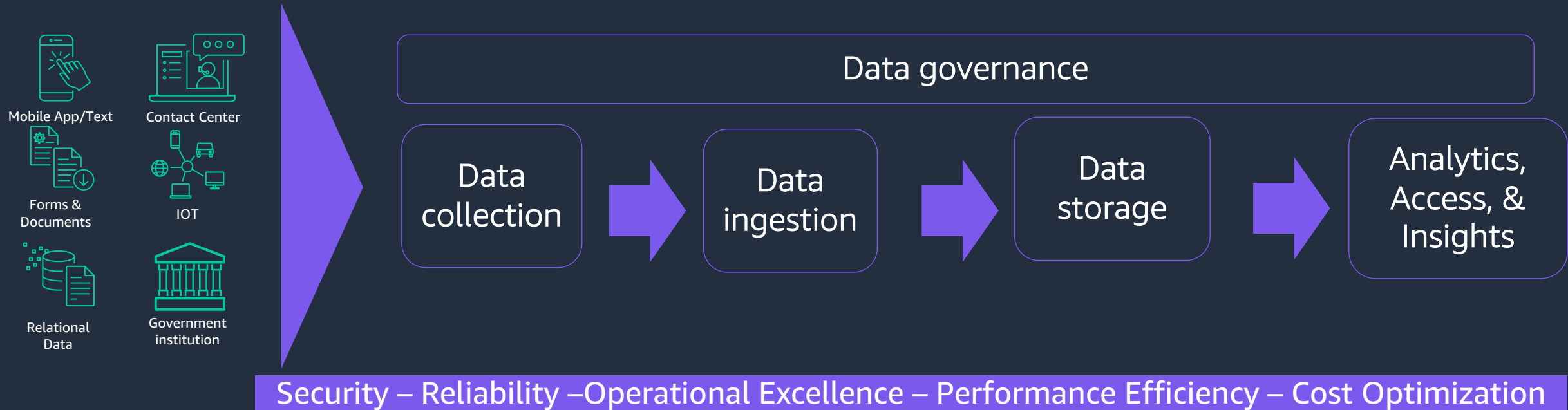
Build new experiences and reimagine old processes with AI/ML



- 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

Putting it all together

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

Reference Architectures

Public Health Organization

- 1 Pandemic brings 1000% increase in disease surveillance data
- 2 Legacy management systems
- 3 Limited capabilities to consolidate data sets from multiple systems
- 4 Difficulty mandating data formats from various partner organizations



Sample reference architecture for disease surveillance

Data validation

- Use serverless computing for event-driven data validation as files come in to reduce cost and processing times



Validation



Data Catalog & ETL

Data from



Ingest



Storage



Surveillance models



Insights & Analytics



Data Scientist
Regulatory User

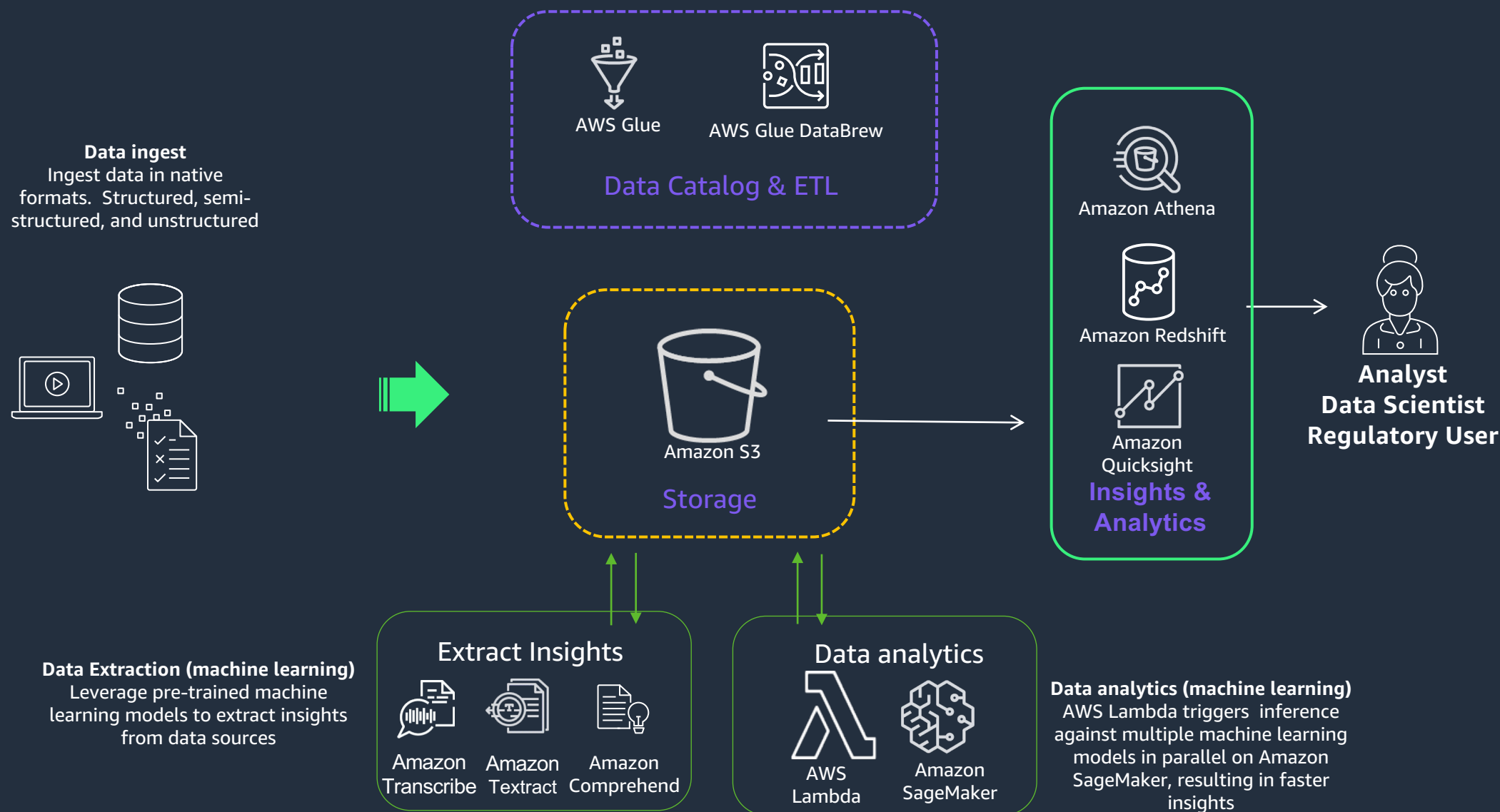
Data analytics

- Amazon S3 acts as a data lake feeding traditional business analytics tools but also enabling data scientists to leverage Machine Learning to provide new, more effective surveillance

Data ingestion

- Leverage fully managed and scalable services such as Amazon S3 and AWS Transfer for SFTP which automatically adjusts to highly variable market activity
- Database migration service (DMS) can be used for batch or CDC data transfer

Sample reference architecture for

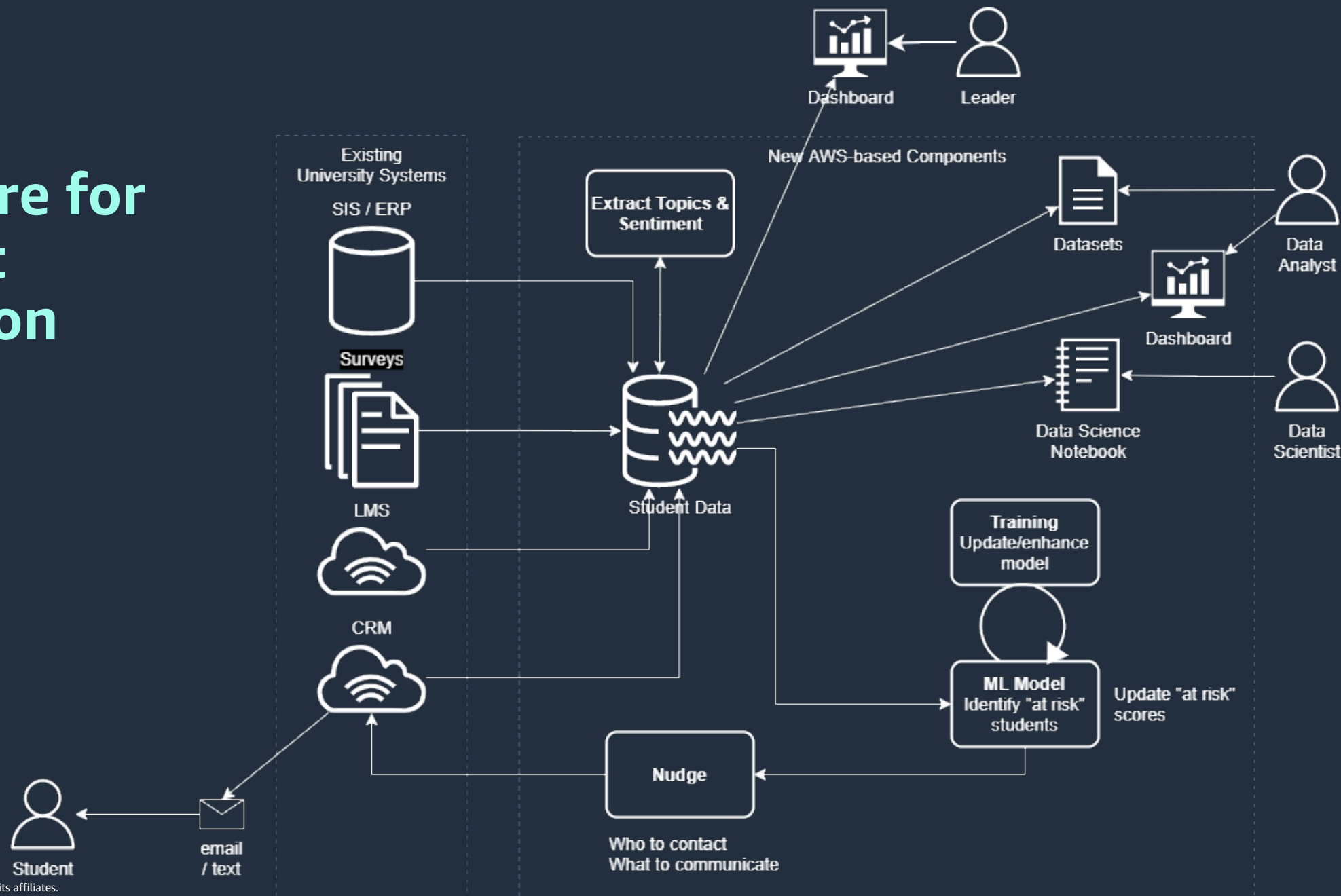


Improving student outcomes - Retention

- 1 Identify at-risk students from behaviors
- 2 Aggregated student touchpoint data from the SIS, LMS, and CRM
- 3 Feed insights into communication platform for **early intervention and nudging**



A logical architecture for intelligent intervention



Moving the needle on retention

MARYVILLE UNIVERSITY

- 1 IT staff participated in data lake and modern data architecture **skills development**
- 2 Aggregated student touchpoint data from the **SIS, LMS, and CRM** into a data lake in 6 weeks
- 3 Automated processing and machine learning to **identify at-risk students** from behaviors
- 4 Fed insights into communication platform for **early intervention and nudging**



Get started

BUILD WITH US



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



Please Provide Your Feedback



Step 1: Data & Analytics Track
Step 2: Building a modern data strategy

Learning Day Content

<https://sanfrancisco2024.awslearningday.com/>

