

AWS State, Local, and Education Learning Days

Sacramento, CA

10:15am – 11:15am

200
level

**Building a modern
data strategy**

Transform your data
chaos into
competitive
advantage with AWS:
Unify, analyze, and
innovate faster than
ever before.



Building a Modern Data Strategy

Jason Moldan

Principal Solutions Architect

AWS

jmoldan@amazon.com

Alex Iliev

Sr. Solutions Architect

AWS

alexaws@amazon.com

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

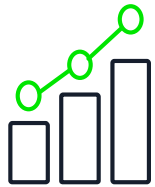


Insights



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



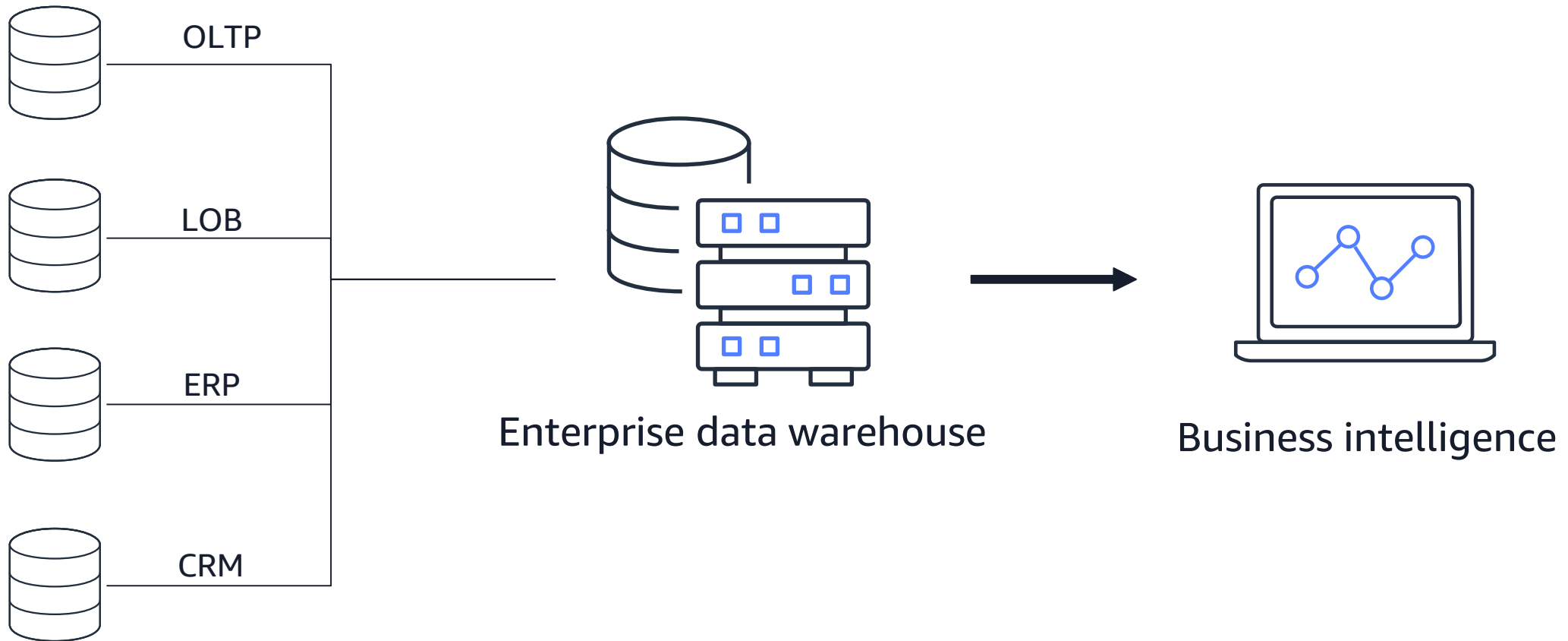
Used by
many people (e.g.
policy makers,
researchers, etc.)



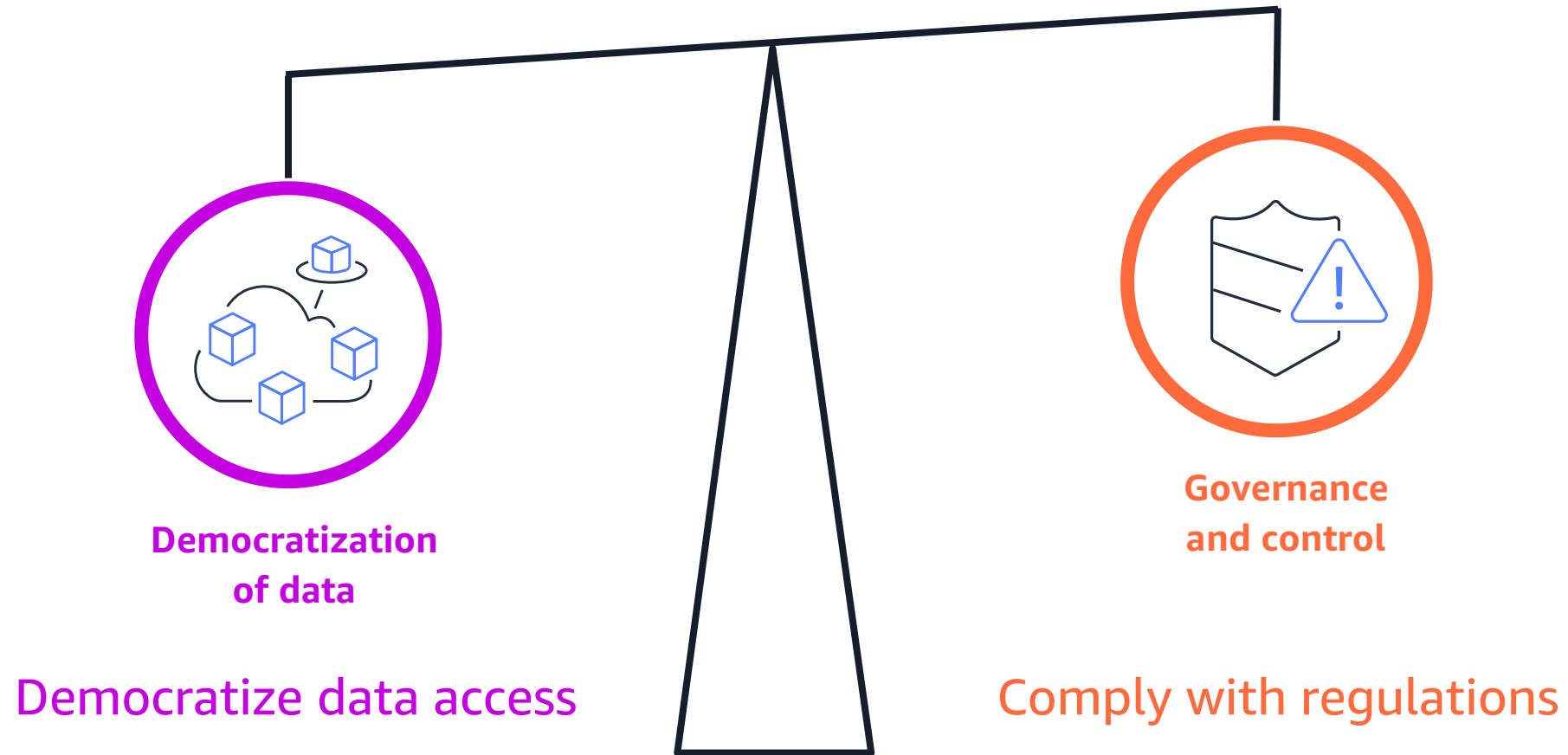
Analyzed by
many applications

Current state

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



Striking the balance



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



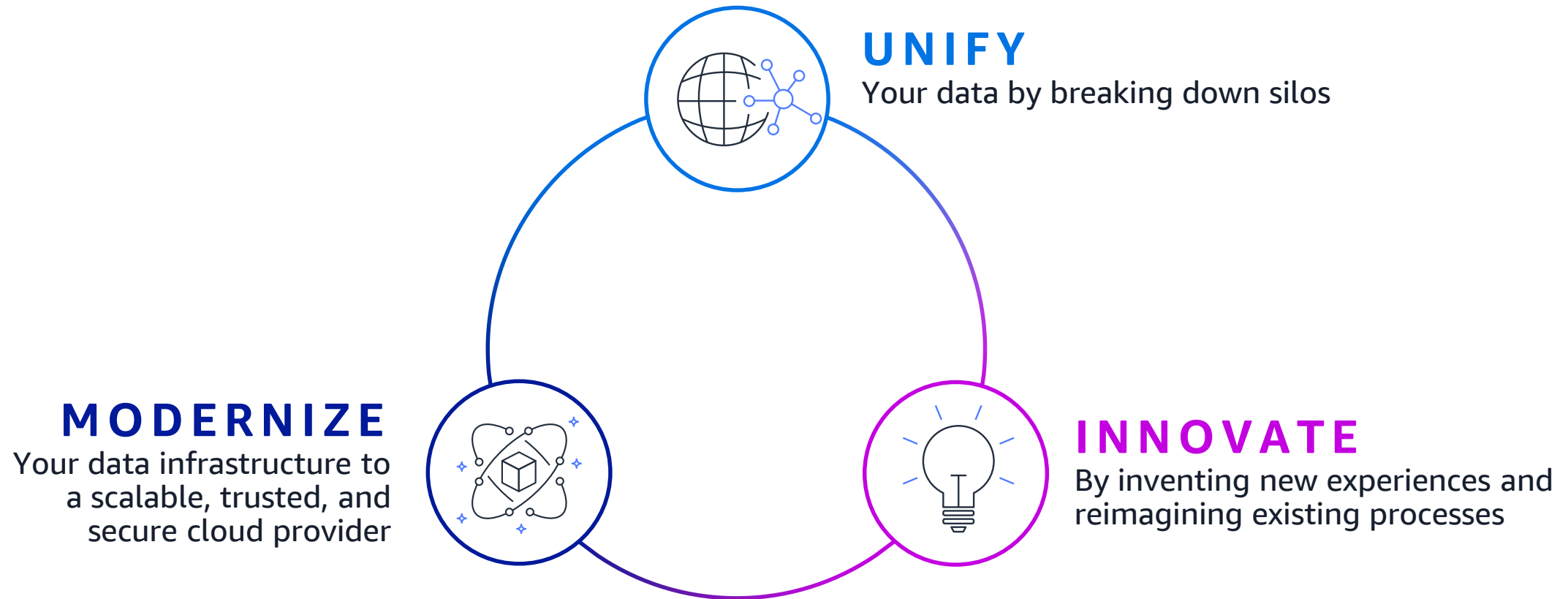
Raw Data



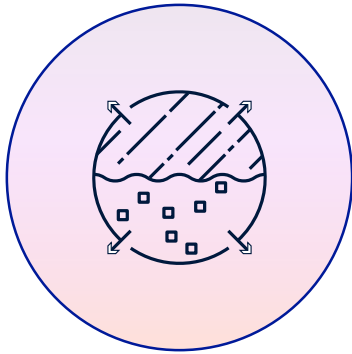
Insights

Modern Data Strategy

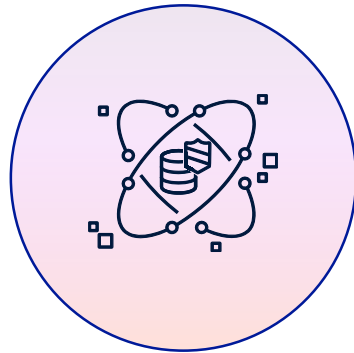
Modern data strategy for better mission outcomes



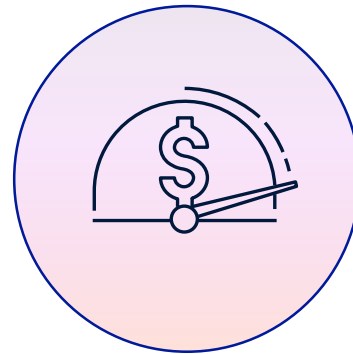
AWS modern data strategy components



Scalable
data lakes



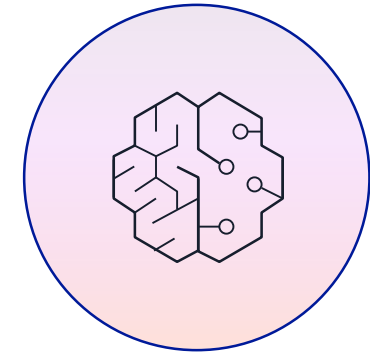
Unified data
access, security,
and governance



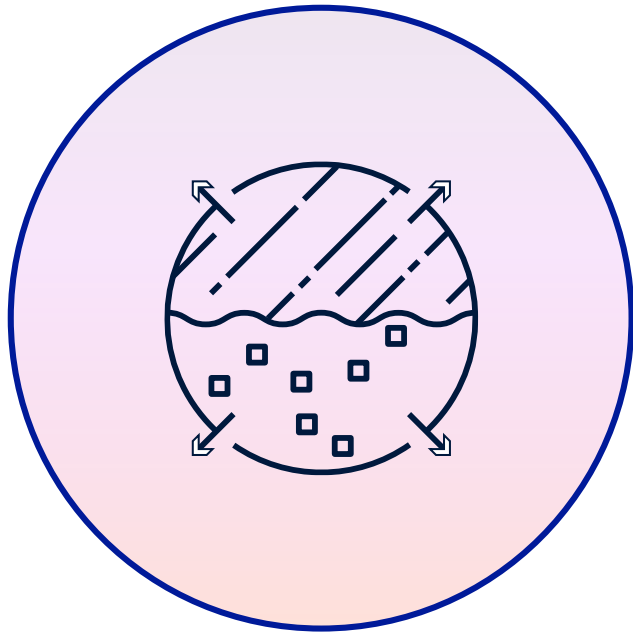
Purpose-built
data services for
performance
and cost



Serverless
and easy to
use



Built-in
machine
learning
& AI



Unify Data with **Scalable** data lakes

Amazon S3: Data lakes on AWS

Store unlimited data in open file formats

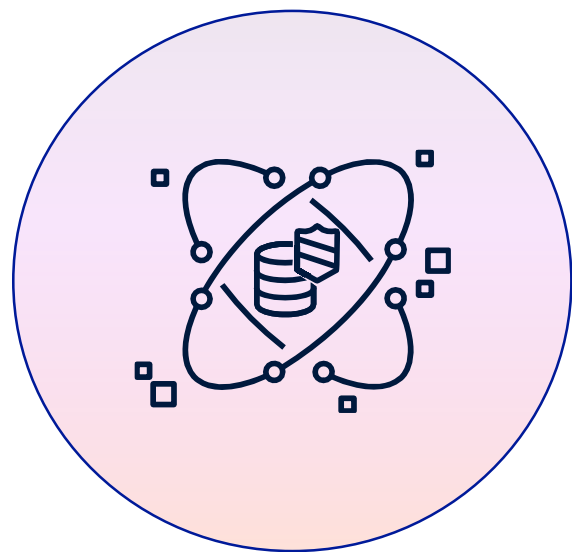
Unmatched durability, availability, and scalability

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

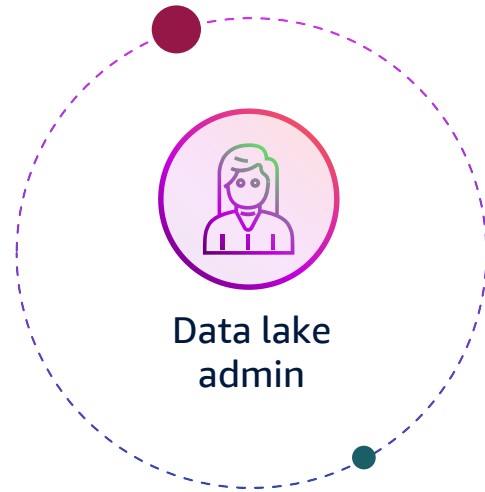


Amazon SageMaker

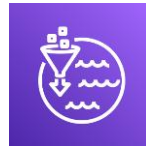


Amazon EMR

Simplified and unified security management



Data lake admin



AWS Lake Formation



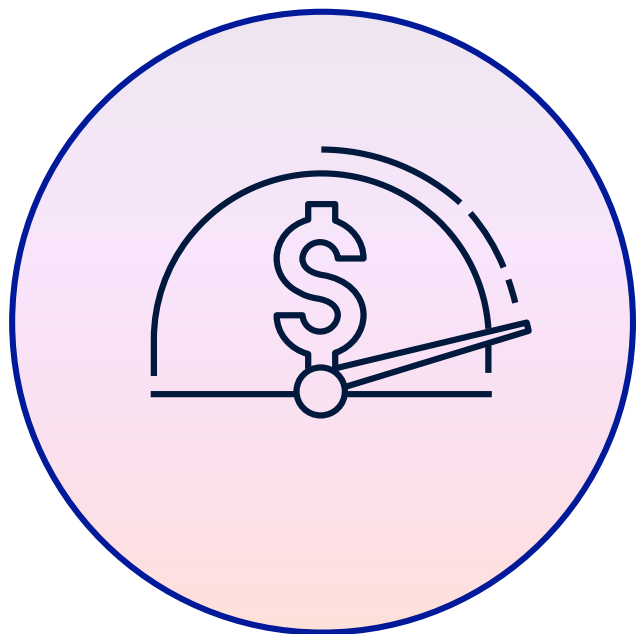
Access control



AWS Glue Data Catalog




Amazon S3 data lake storage





Modernize with Purpose-built data services


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 Neptune**
Graph

 **Amazon DocumentDB**
Document


 **Amazon Timestream**
Time series

 **Amazon RDS**
MySQL, PostgreSQL, MariaDB, Oracle, SQL Server, DB2


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


 **Amazon ElastiCache**
Redis, Memcached


Blockchain

 **Amazon Managed Blockchain**

Data lake

 **Amazon S3/ Amazon S3 Glacier**

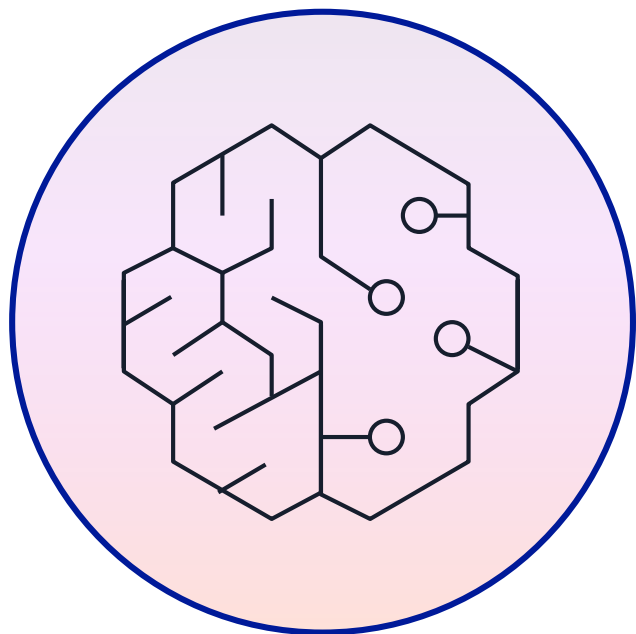
 **AWS Lake Formation**
Data lakes

 **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





Innovate with AI & machine learning

Build new experiences and reimagine old processes with AI/ML

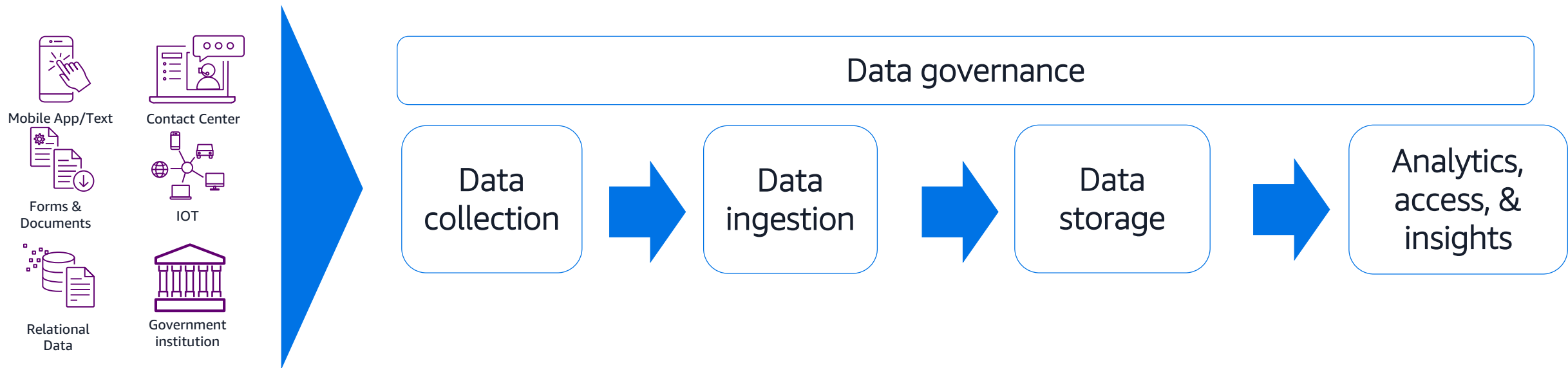


- Make accurate predictions, get deeper insights from your data, and improve customer experience
- Create ML predictions without any ML experience or writing any code
- Build applications with our pre-trained models
- Train and apply your own models
- Use your own algorithms by working directly with 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



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

Key considerations:

1

Ability to handle the increasing volume, velocity, and variety of data

2

Each component should be independently scalable

3

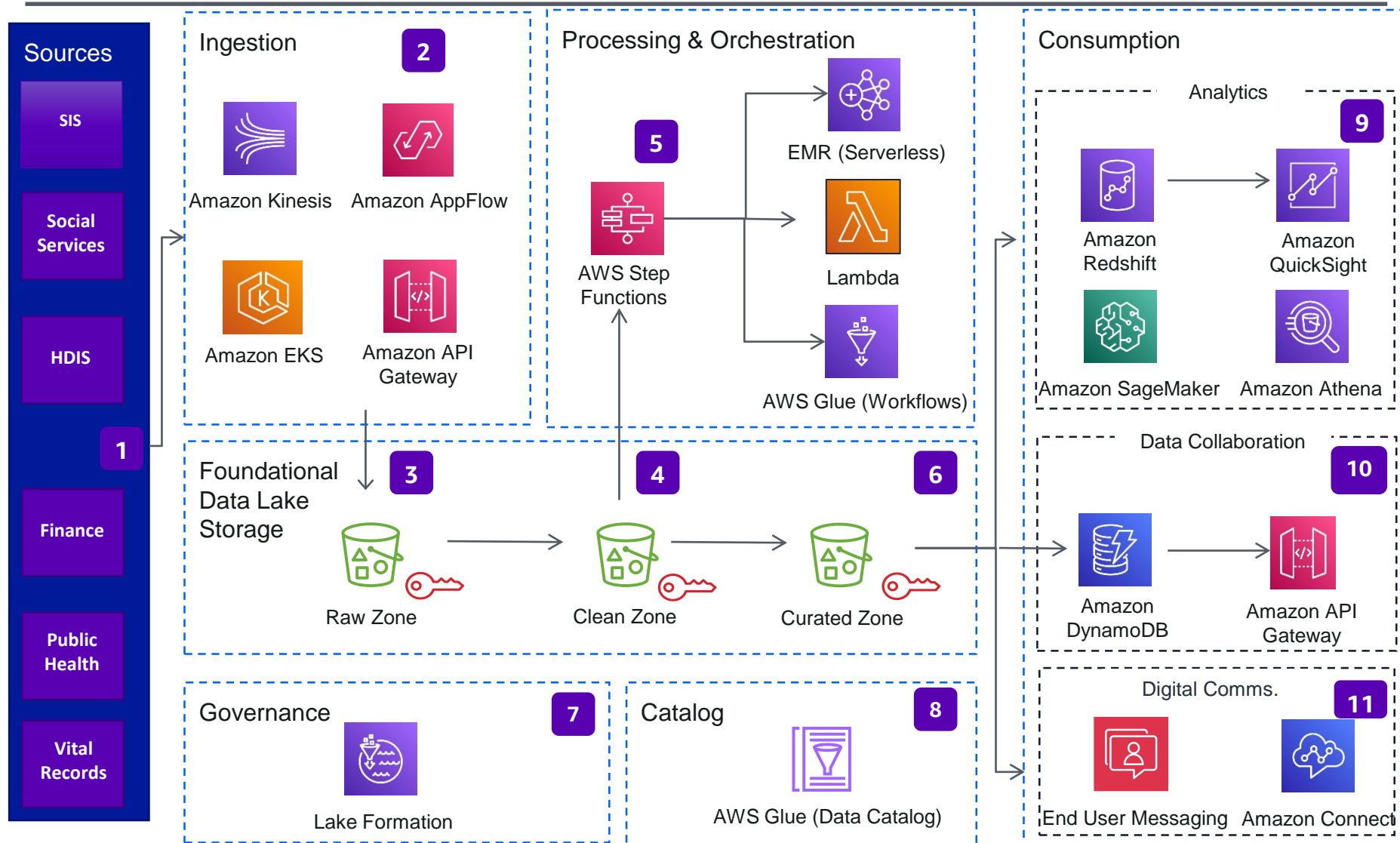
Make data easily accessible and sharable

Reference Architectures



Basics: Data Platform on AWS

This guidance provides a reference architecture showing best practices in the building of a customer data platform covering data ingestion, identity resolution, segmentation, analysis and activation.

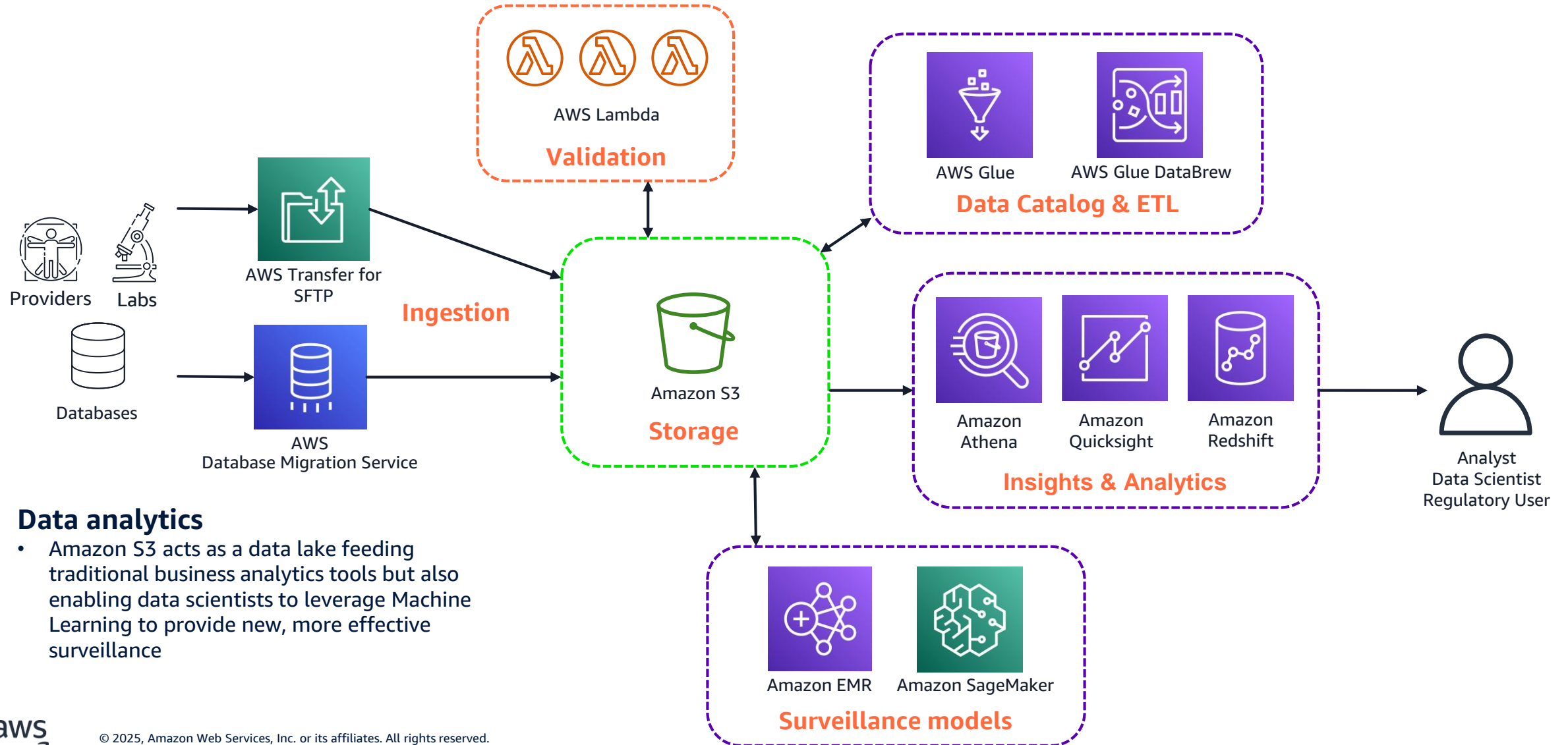


Public Health Organization

- 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 analytics

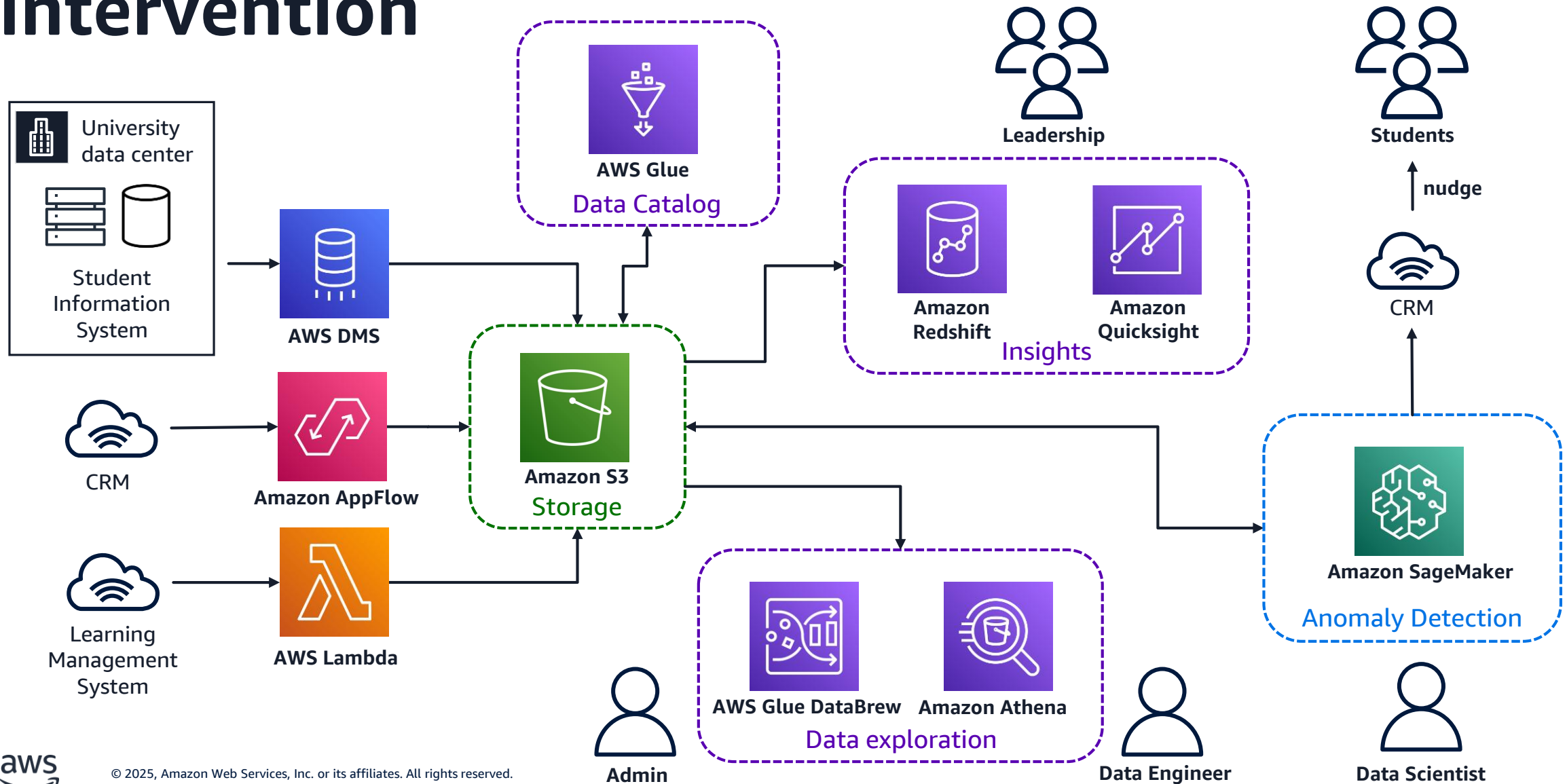
- Amazon S3 acts as a data lake feeding traditional business analytics tools but also enabling data scientists to leverage Machine Learning to provide new, more effective surveillance

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



Sample reference architecture for student 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**



MARYVILLE
UNIVERSITY

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



Transform human services



Enhancing efficiency



Create end-to-end visibility

Port of Long Beach

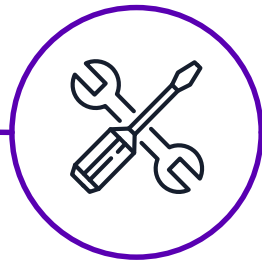
So how do I get started?

- Consolidate data in one central, scalable, accessible place
- Right tool, right job, right insights...at the right time
- Rome (and the Cloud) wasn't built in a day...**think BIG, start small, scale *fast***



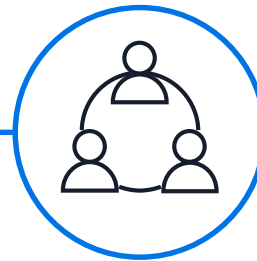
Get started

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



Thank you!

Jason Moldan

Principal Solutions Architect
AWS
jmoldan@amazon.com

Alex Iliev

Sr. Solutions Architect
AWS
alexiaaws@amazon.com

Please complete the survey
for this session



Data & Analytics

Building a Modern
Data Strategy

Coming up NEXT

11:30am – 12:30pm

200
level

AI/ML for data and
analytics

Unleash the power of
data: Transform public
sector insights with AI
and ML across your
entire enterprise.