AWS Columbus Learning Days

Building a modern data strategy

Brett Hensley
Solutions Architect
Amazon Web Services

Agenda

- Voice of the customer
- AWS modern data strategy
- Customer case studies and reference architectures
- How can AWS help?

Decision-making in the past

In the past, decision-making revolved around the **enterprise data warehouse**

Data no longer fits

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

Data growth

>10x
every five years

Data platforms needs

To live for
15+
years

To scale
1,000x

IDC, “Data Age 2025”

Accessibility of data

Data scientists

Business users

Analysts

Applications

Machine learning

SQL analytics

Scientific

Real-time, streaming

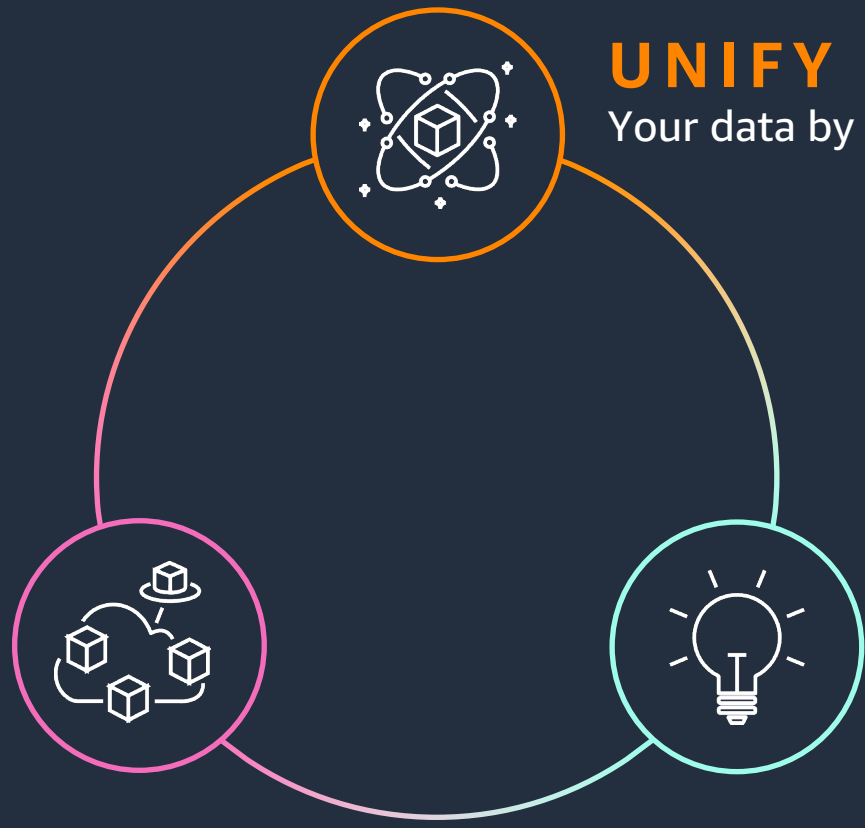
There are **more people** accessing data

And in **different ways**

More regulatory pressure

Modern data strategy for better business outcomes

Start anywhere



UNIFY
Your data by breaking down silos

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

INNOVATE
By inventing new experiences and reimagining existing processes

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 and student experiences and outcomes

OKLAHOMA CITY

Student success and community relevance

Respond to the unexpected

DEPARTMENT OF HEALTH

Inform policy

Unifying data to enable 360-degree views

DHS
MARYLAND DEPARTMENT OF HUMAN SERVICES

Transforming human services

Unify data with scalable data lakes

Amazon Simple Storage Service (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

Simplified and unified security management

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

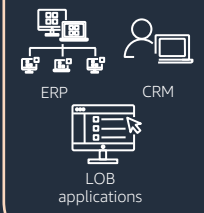
AWS data lakes provide a flexible foundation for analytics and innovation

AWS Glue

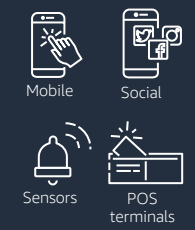
Data catalog
Crawls and catalogs your data; discover, prepare, and combine data for analytics and ML

AWS Lake Formation

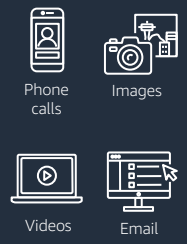
Data governance
Centralized authorization layer to define data sources and data access and security policies



Structured data
Data that is 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

- AWS Relational Database Service (RDS)
- Amazon Aurora
- Amazon DynamoDB
- Amazon DocumentDB
- Amazon ElastiCache
- Amazon Neptune
- Amazon Timestream
- Amazon Quantum Ledger Database (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

aws

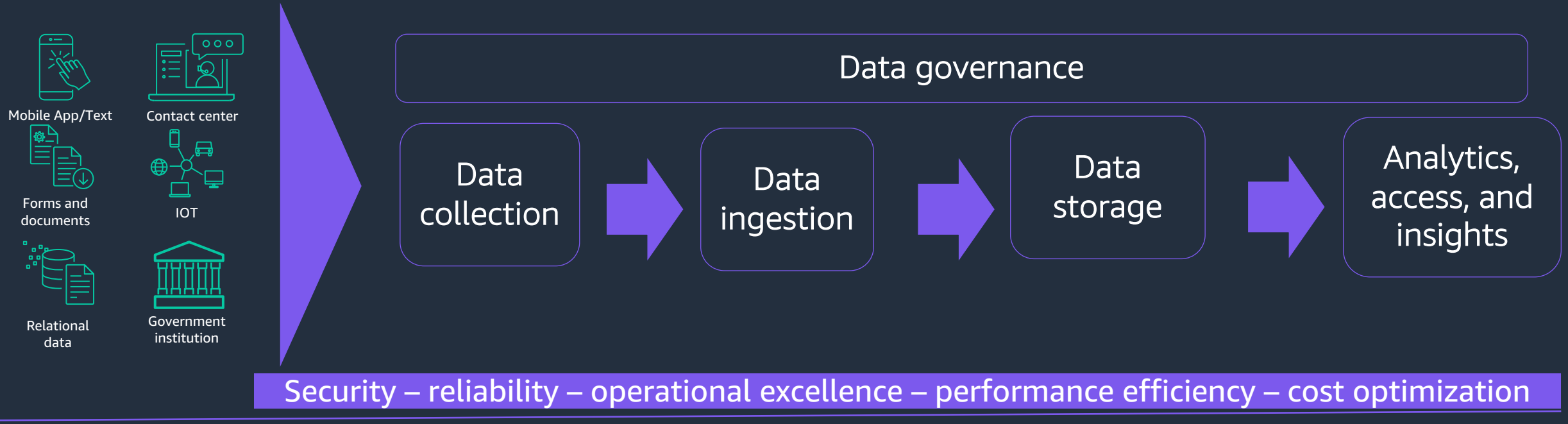
Innovate with AI & machine learning

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

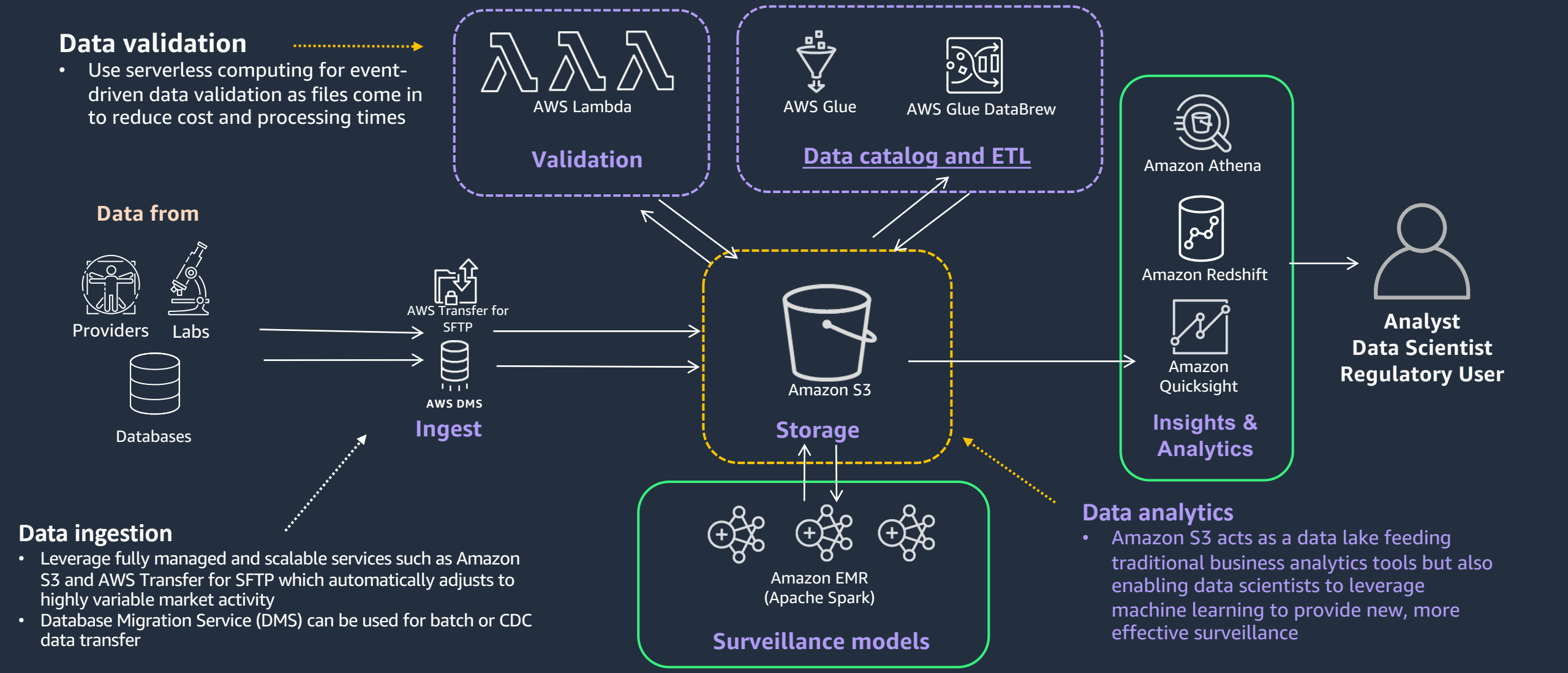


Key considerations:

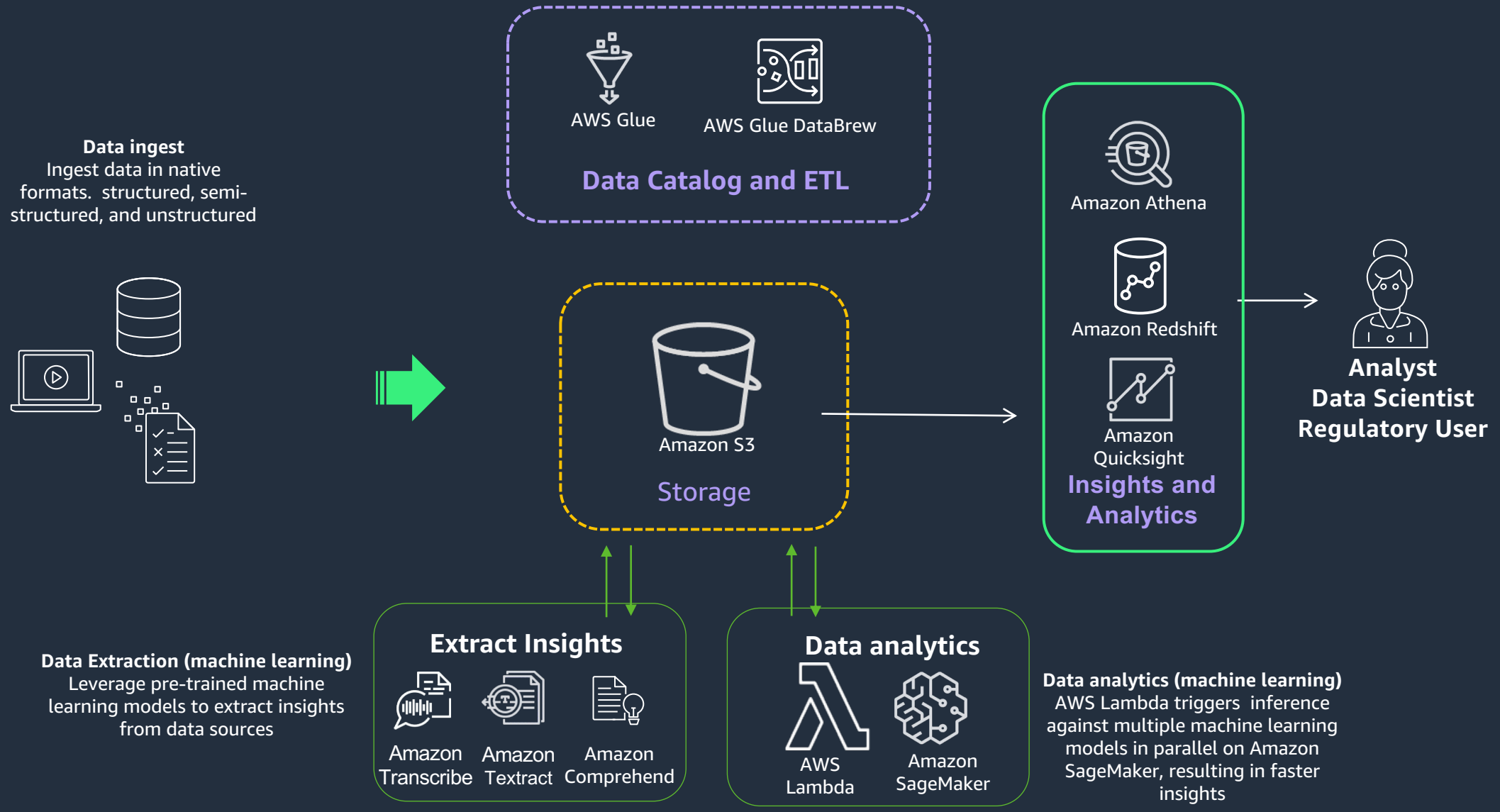
1. Ability to handle the increasing volume, velocity, and variety of data
2. Each component should be independently scalable
3. Make data easily accessible and sharable

Reference architectures

Sample reference architecture for disease surveillance



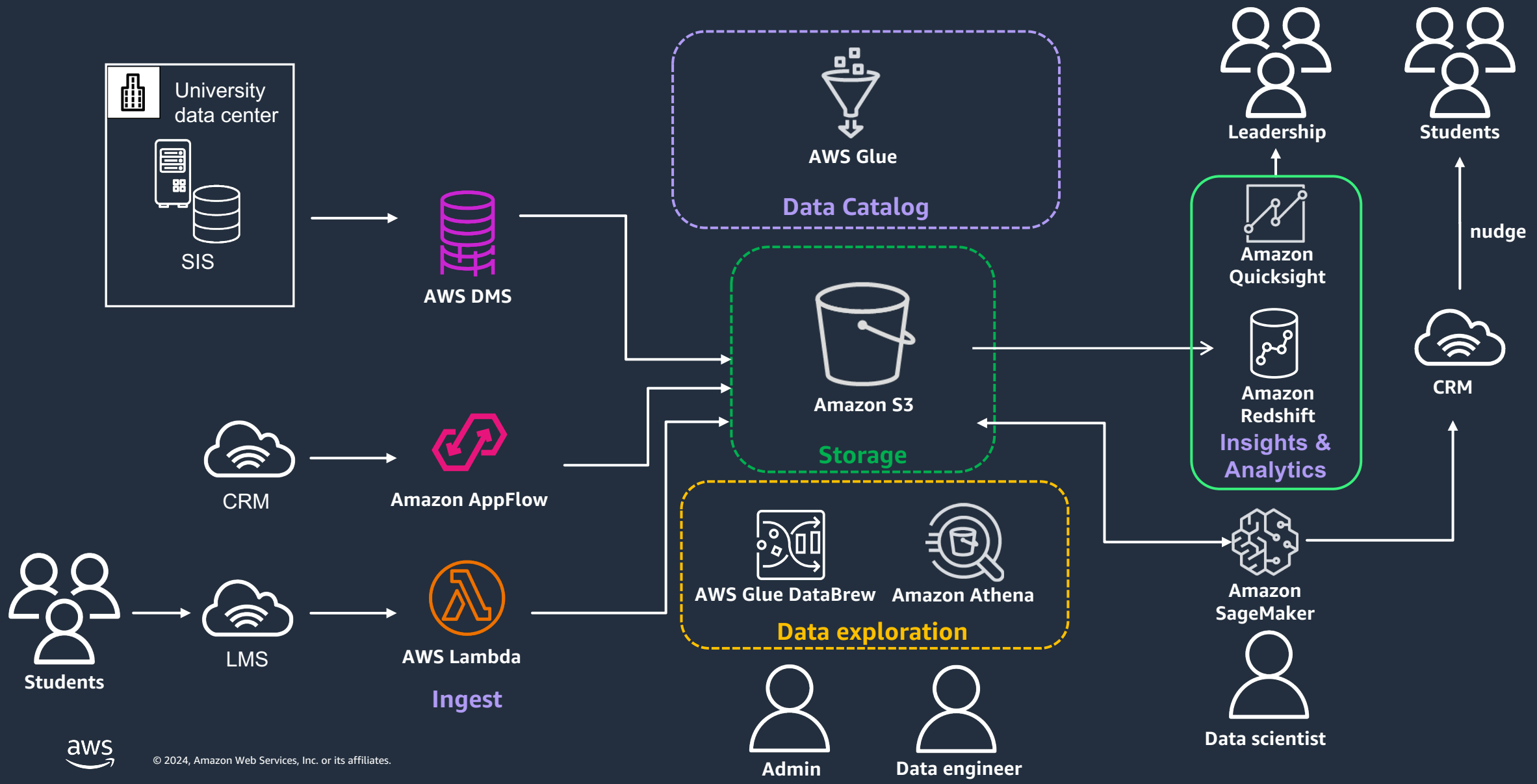
Sample reference architecture for ...



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**

Sample reference architecture for student intervention



How to get started

Think about specific use cases

Reach out to your AWS Team

Connect with SAs at the Ask an Expert table

Further reading:

https://aws.amazon.com/blogs/publicsector/osu-okc-upskills-workforce-drives-real-time-decision-making-with-live-reporting-analytical-modeling/

https://aws.amazon.com/stateandlocal/state-local-government-champions/minnesota-department-of-health/

https://aws.amazon.com/solutions/case-studies/maryland-dhs/

Thank you!

Brett Hensley
awsbrett@amazon.com

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