AWS State, Local, and Education Learning Days

Phoenix, AZ

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.

aws Learning Days
State, Local, and Education

Building a Modern Data Strategy

Giri Badanahatti
Senior Solutions Architect
AWS
awsgirib@amazon.com

Kain Leo
Solutions Architect
AWS
leokai@amazon.com

aws Learning Days
State, Local, and Education

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

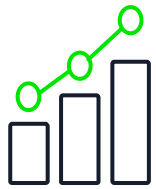
CEO of Netscape

Data is just the Building Blocks

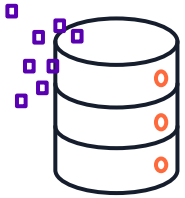
Data ▶ Information ▶ Insights

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

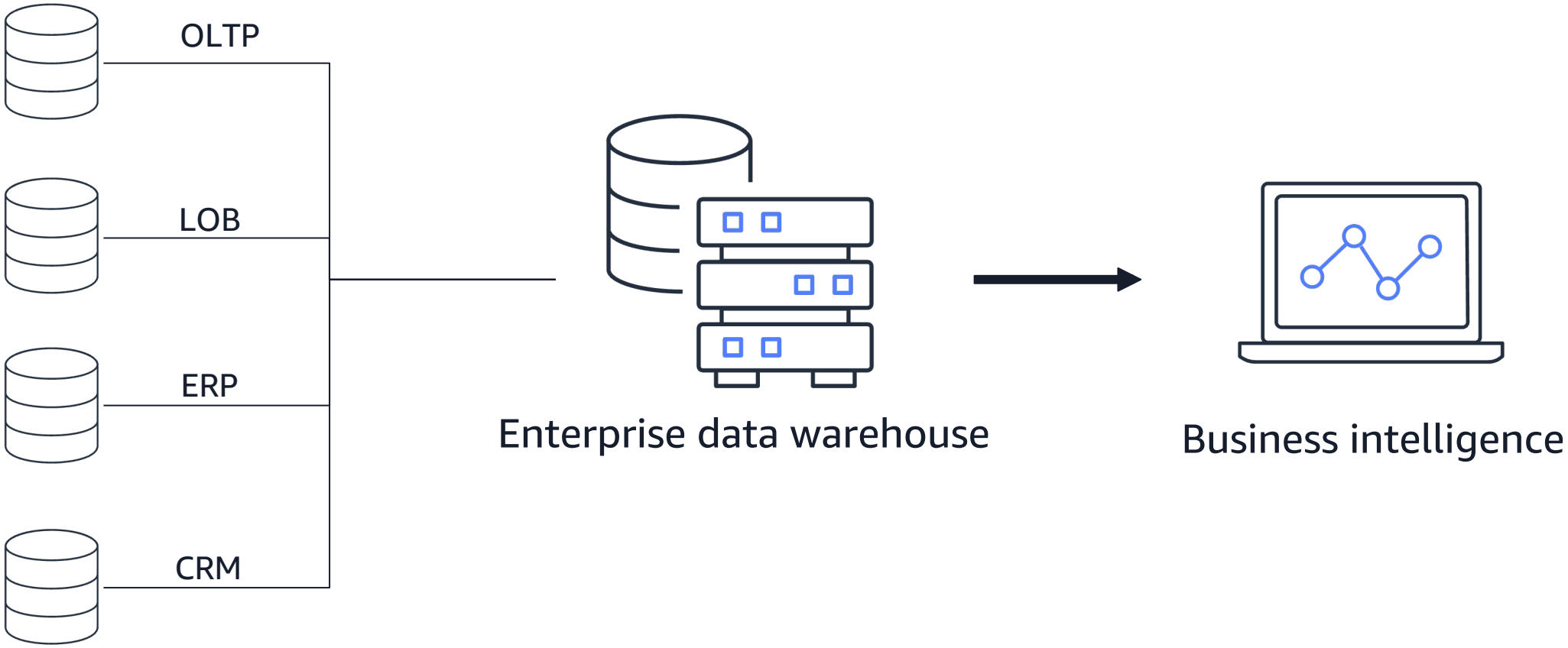
Increasingly diverse in file type and volume

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

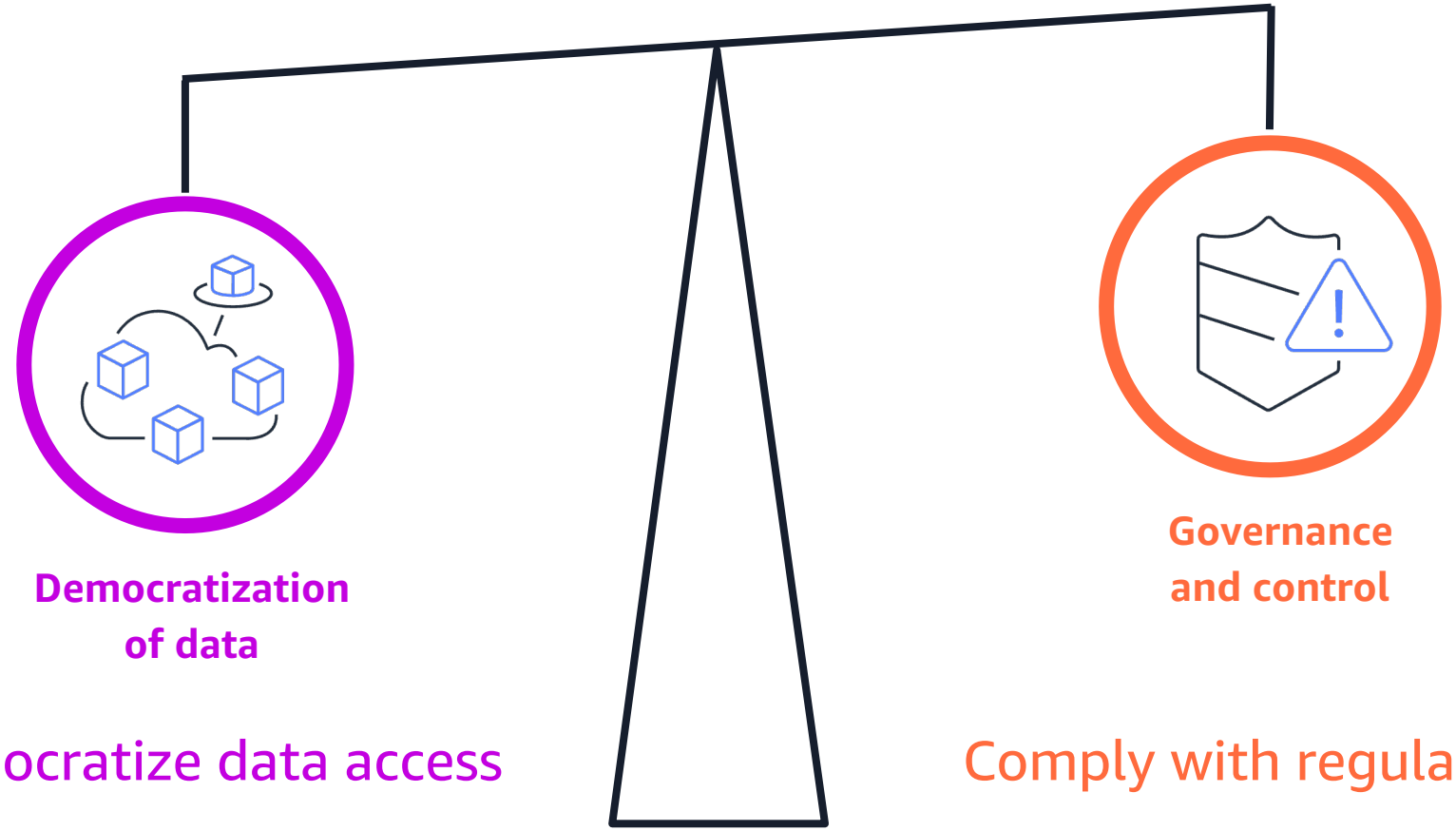
Analyzed by many applications

Current state

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



Striking the balance



Democratization of data

Democratize data access

Governance and control

Comply with regulations

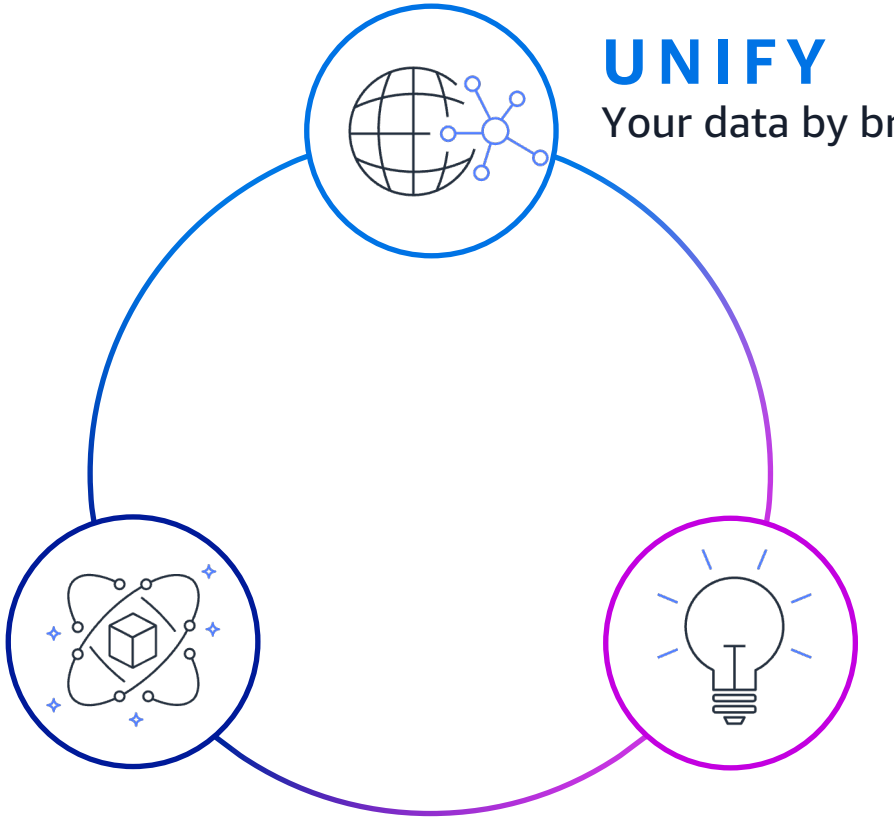
What now? Let’s rethink everything

Raw Data

Insights

Modern Data Strategy

Modern data strategy for better mission outcomes

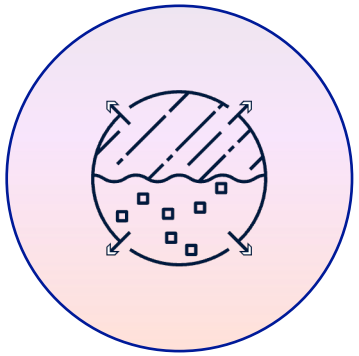


UNIFY
Your data by breaking down silos

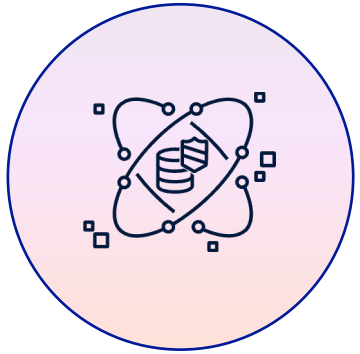
MODERNIZE
Your data infrastructure to a scalable, trusted, and secure cloud provider

INNOVATE
By inventing new experiences and reimagining existing processes

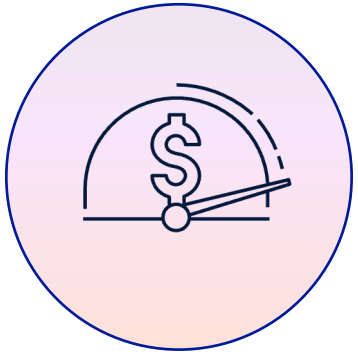
AWS modern data strategy components



Scalable data lakes

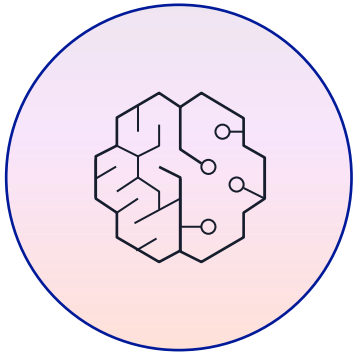


Unified data access, security, and governance

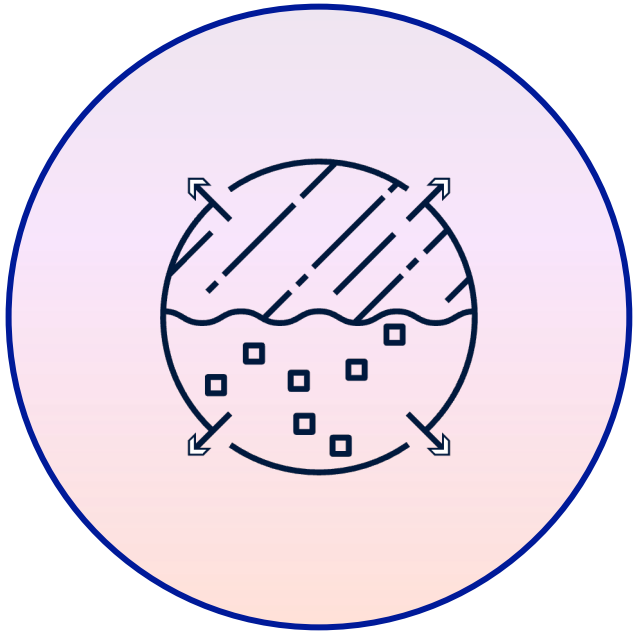


Purpose-built data services for performance and cost

Serverless and easy to use



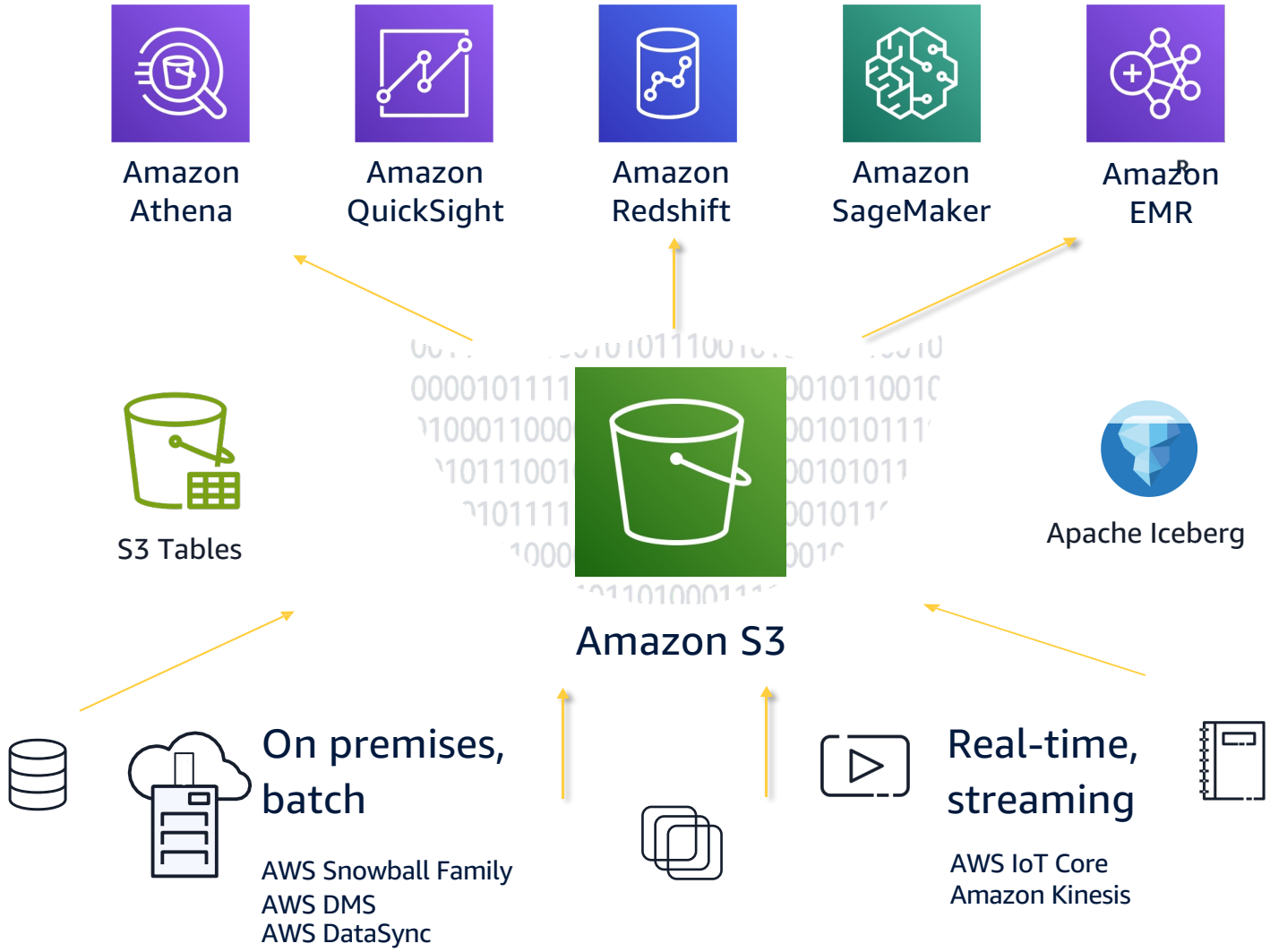
Built-in machine learning & AI

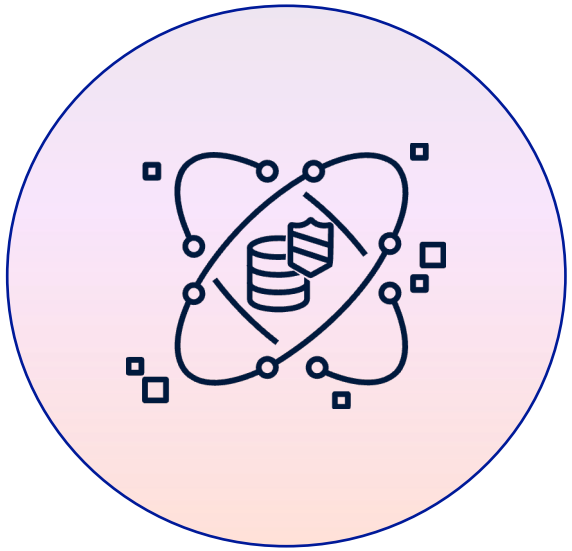


Unify Data with Scalable data lakes

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

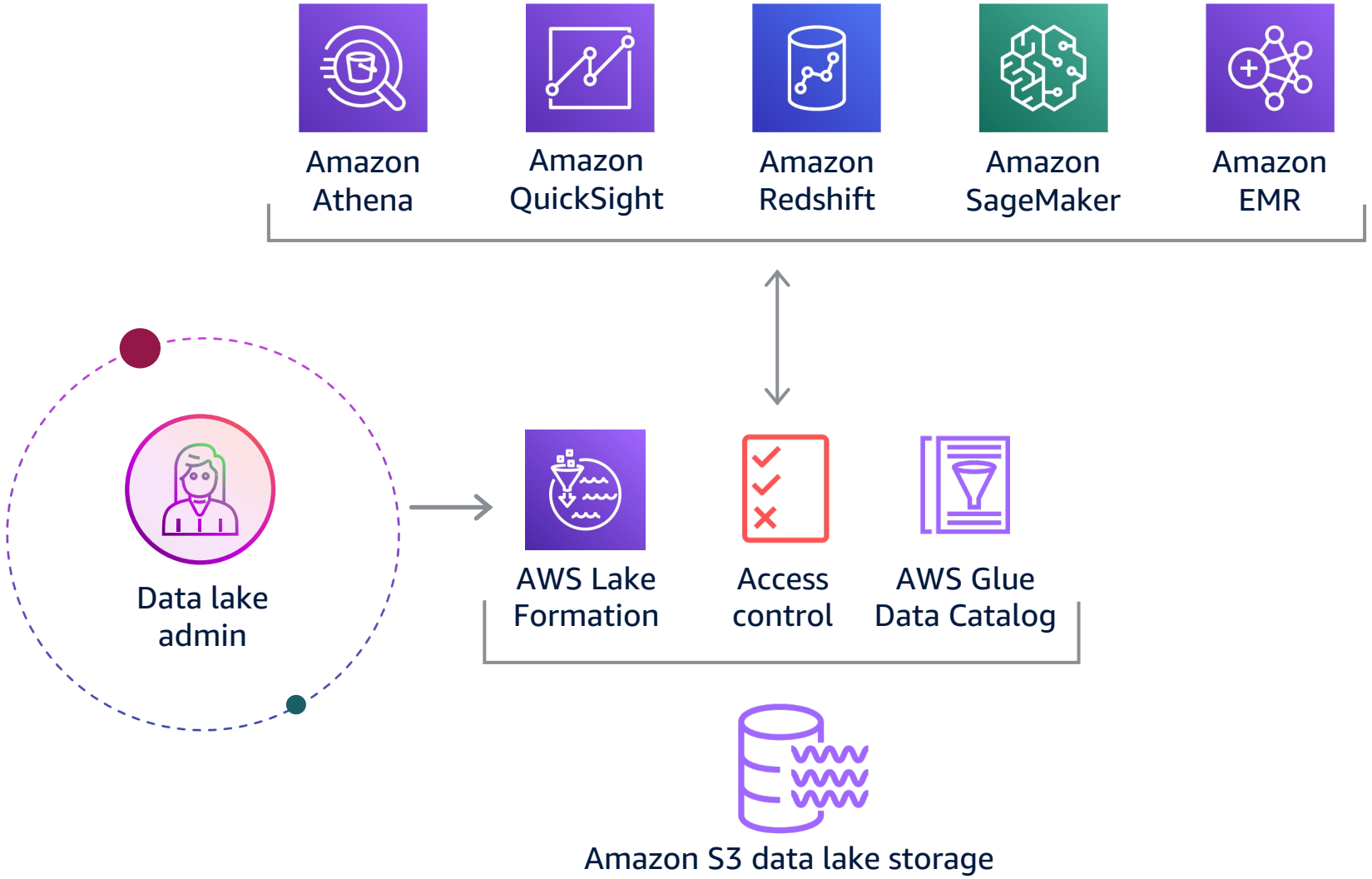


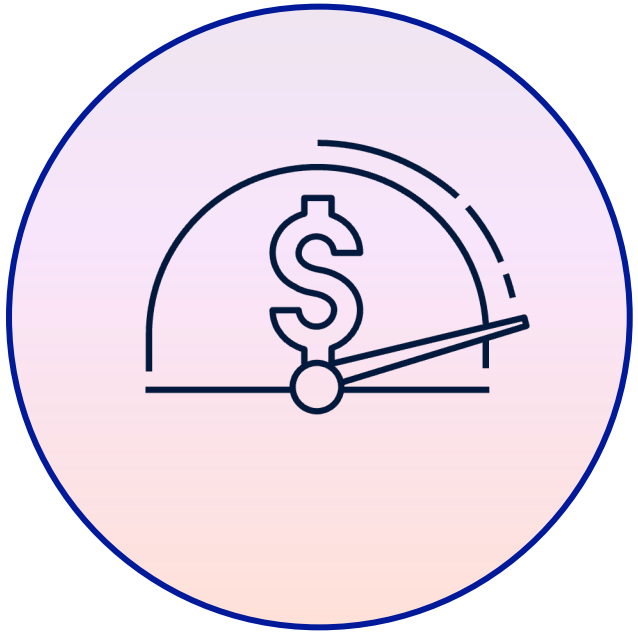


Unified data access, security, and governance

AWS Lake Formation: unified data governance

Simplified and unified security management





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 Forecast** — Forecasts
- **Amazon Translate** — Translation
- **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 RDS** — MySQL, PostgreSQL, MariaDB, Oracle, SQL Server, DB2
- **Amazon DynamoDB** — Key value, Document
- **Amazon DocumentDB** — Document
- **Amazon Keyspaces (for Apache Cassandra)** — Wide column
- **Amazon Neptune** — Graph
- **Amazon Timestream** — Time series
- **Amazon ElastiCache** — Redis, Memcached

Blockchain

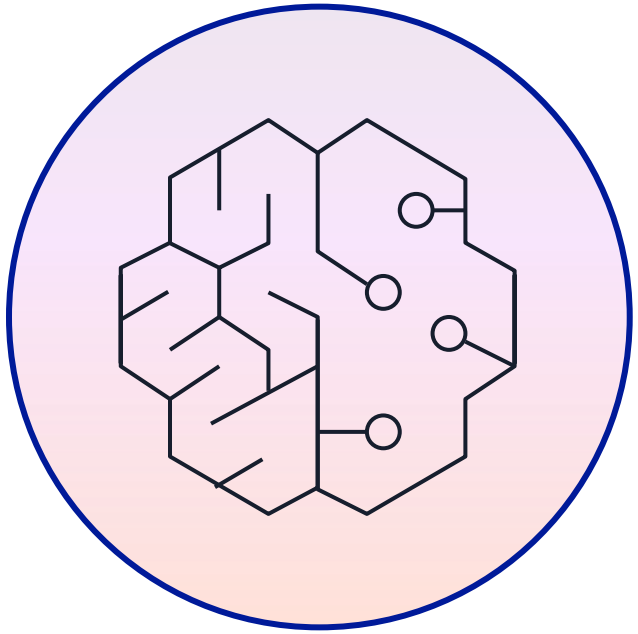
- **Amazon Managed Blockchain**

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



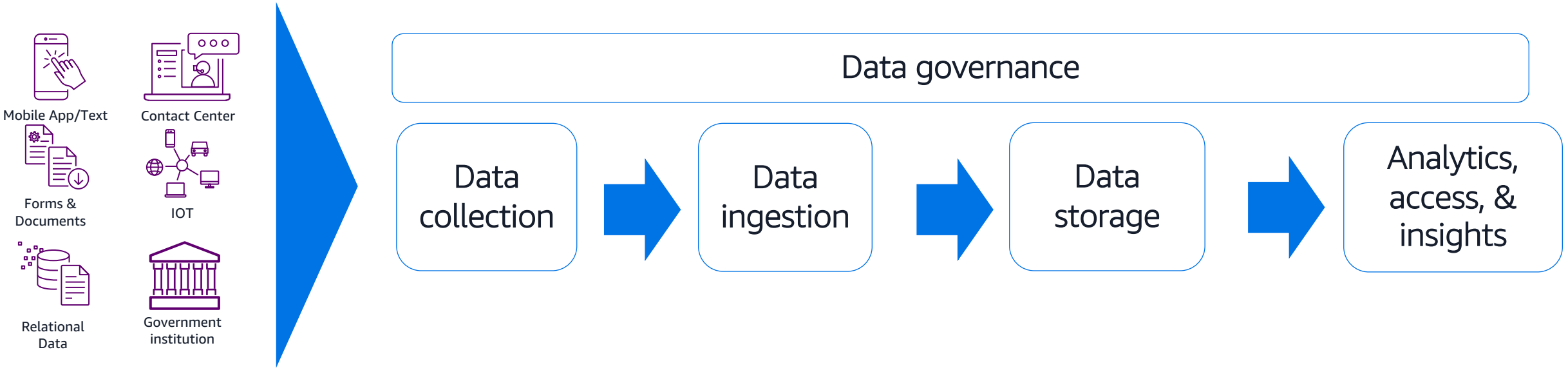
Innovate with
AI & machine learning

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

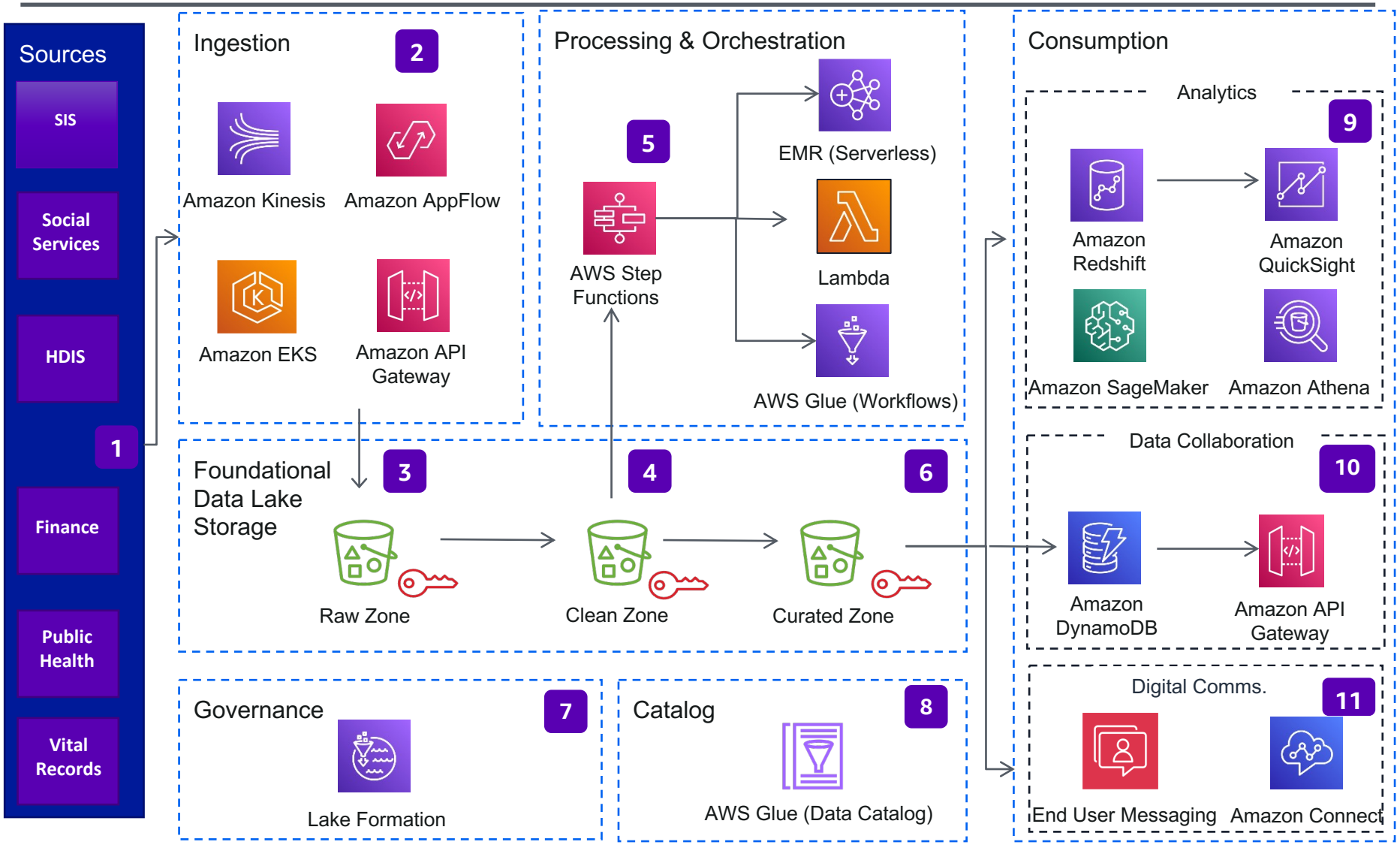
Key considerations:

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

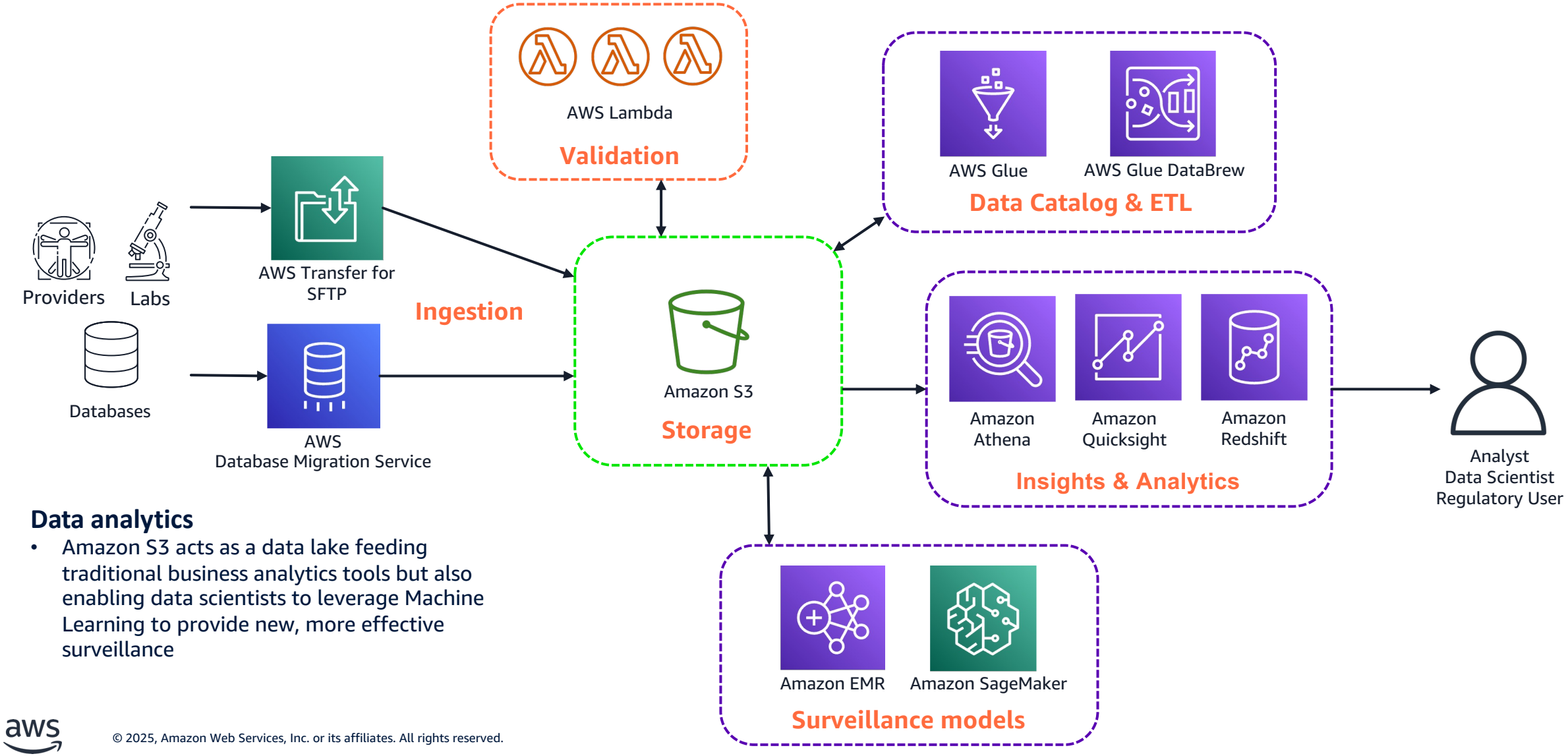
This guidance provides a reference architecture showing best practices in the building of a customer data platform covering data ingestion, identity resolution, segmentation, analysis and activation.



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



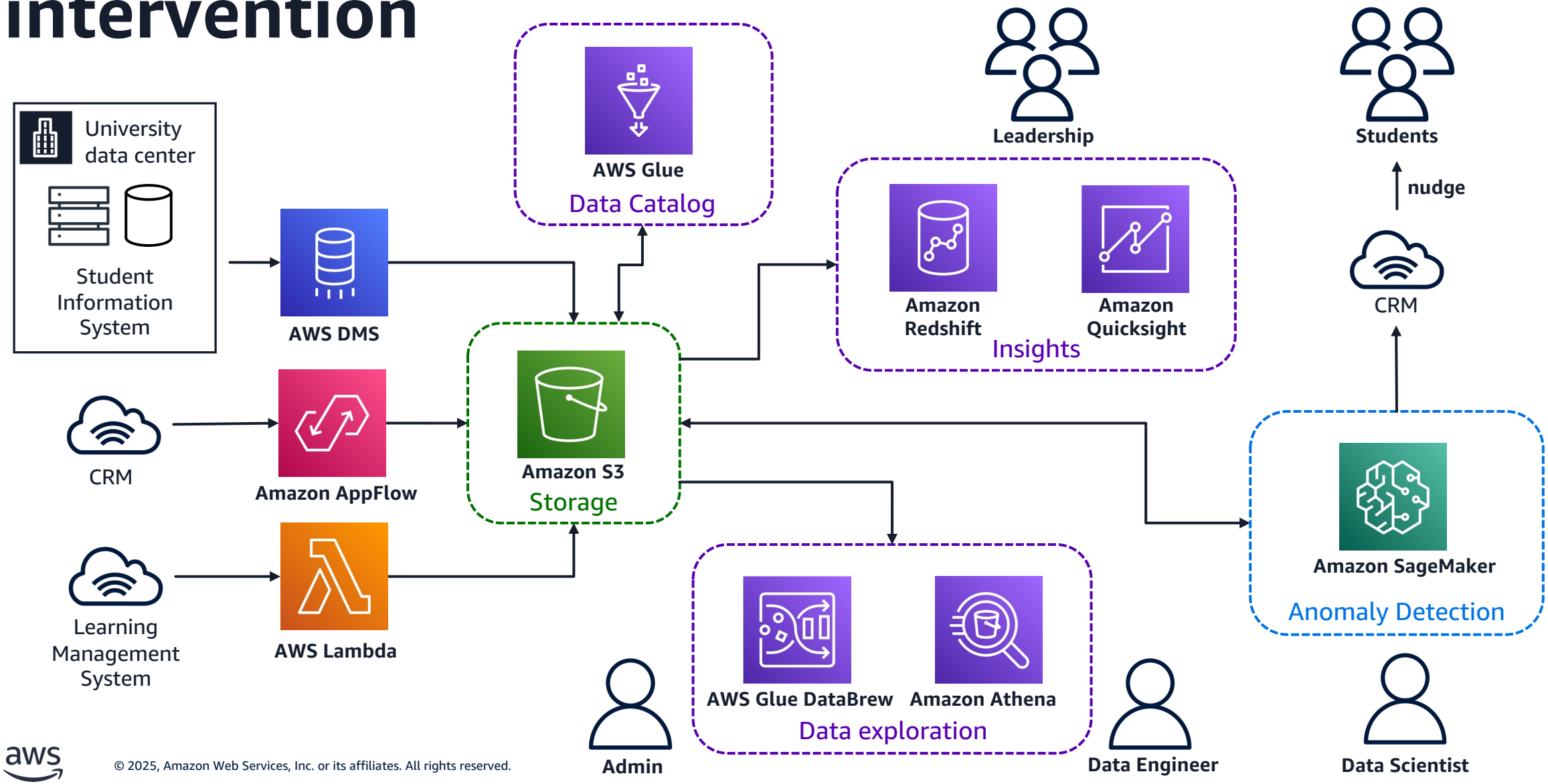
Data analytics

- 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**

Examples

Create better citizen & student experiences & outcomes

→ Student success & community relevance

Respond to the unexpected

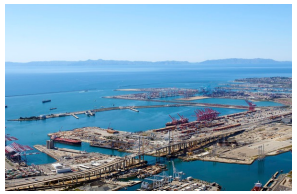
→ Support research in the swine industry

Unifying data to enable 360-degree views

DHS
MARYLAND DEPARTMENT OF HUMAN SERVICES

→ Transform human services

Enhancing efficiency



→ Create end-to-end visibility

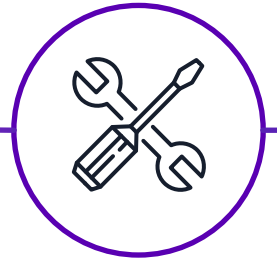
Port of Long Beach

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!

Giri Badanahatti

Senior Solutions Architect
AWS
awsgirib@amazon.com

Kain Leo

Solutions Architect
AWS
leokai@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.

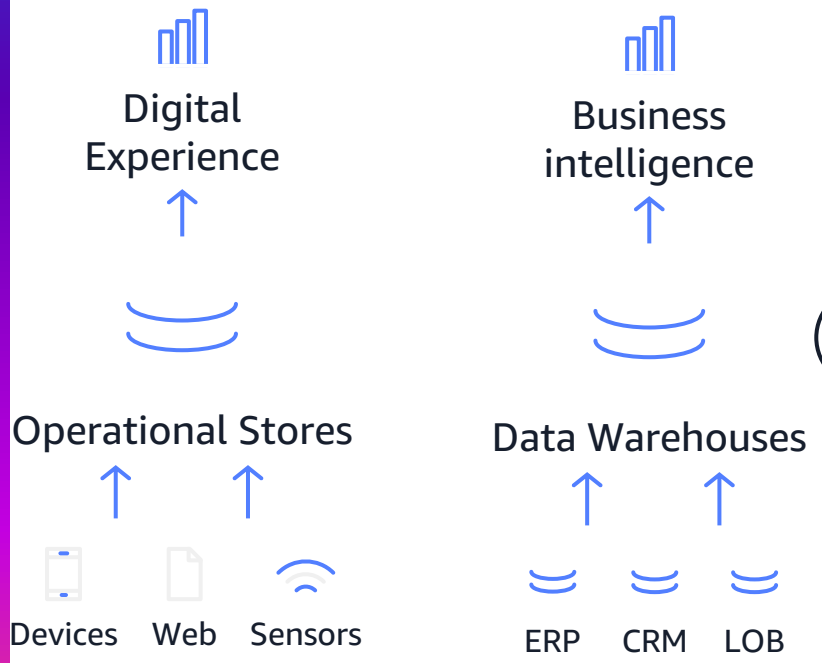
aws Learning Days
State, Local, and Education

Common Data Evolution for Customers

Meet the customers where they are

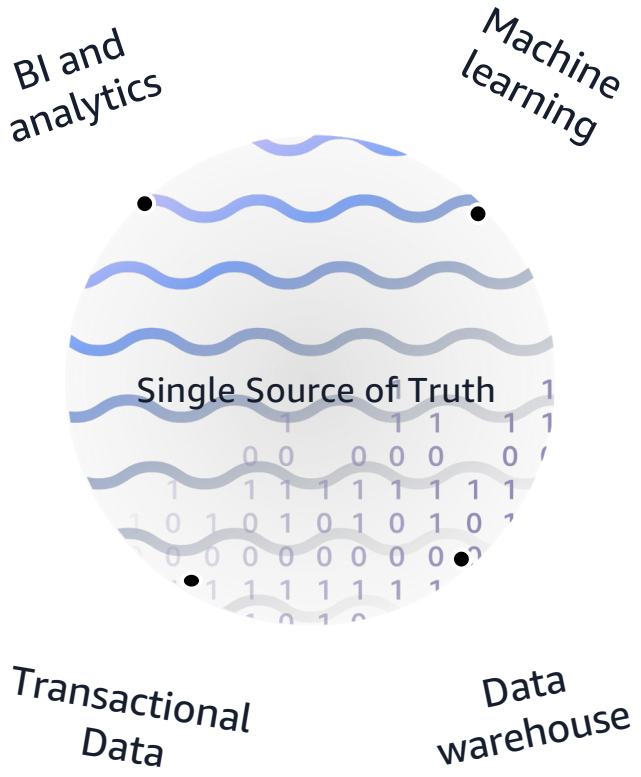
Data Stores

Data Driven



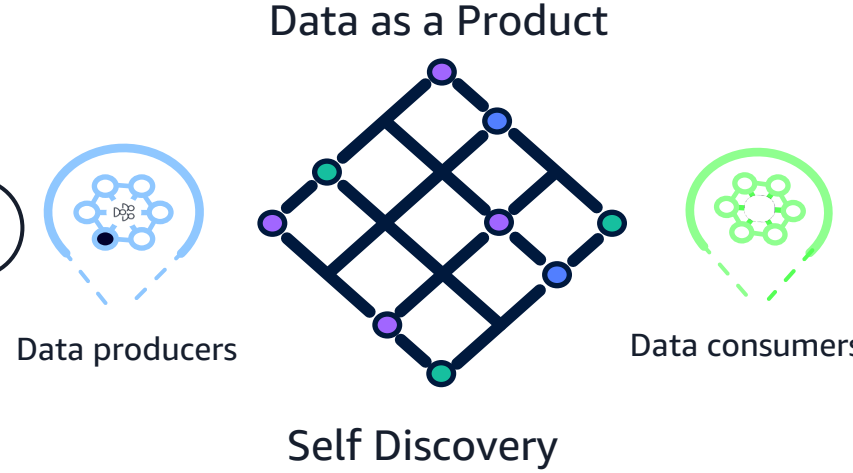
Data lakes

Insights Driven



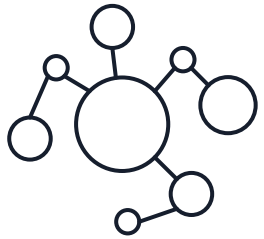
Data Mesh

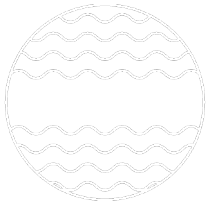
Domain Driven

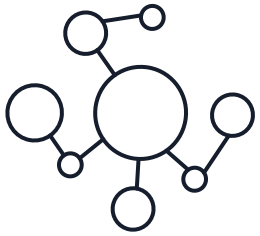


Re-envision the world as products in the community

Consumers of 1st-order products can produce 2nd-order products







Data producers

Producer products/1st order

Foundational data products to serve a wide range of business use cases

- Vendor
- Customer
- Employee
- Product

Data marketplace

Data marketplace products

Technology products for producer and consumer communities to use

- Data quality and ETL tooling
- Data catalog
- Data and ML Ops
- Security controls
- Training

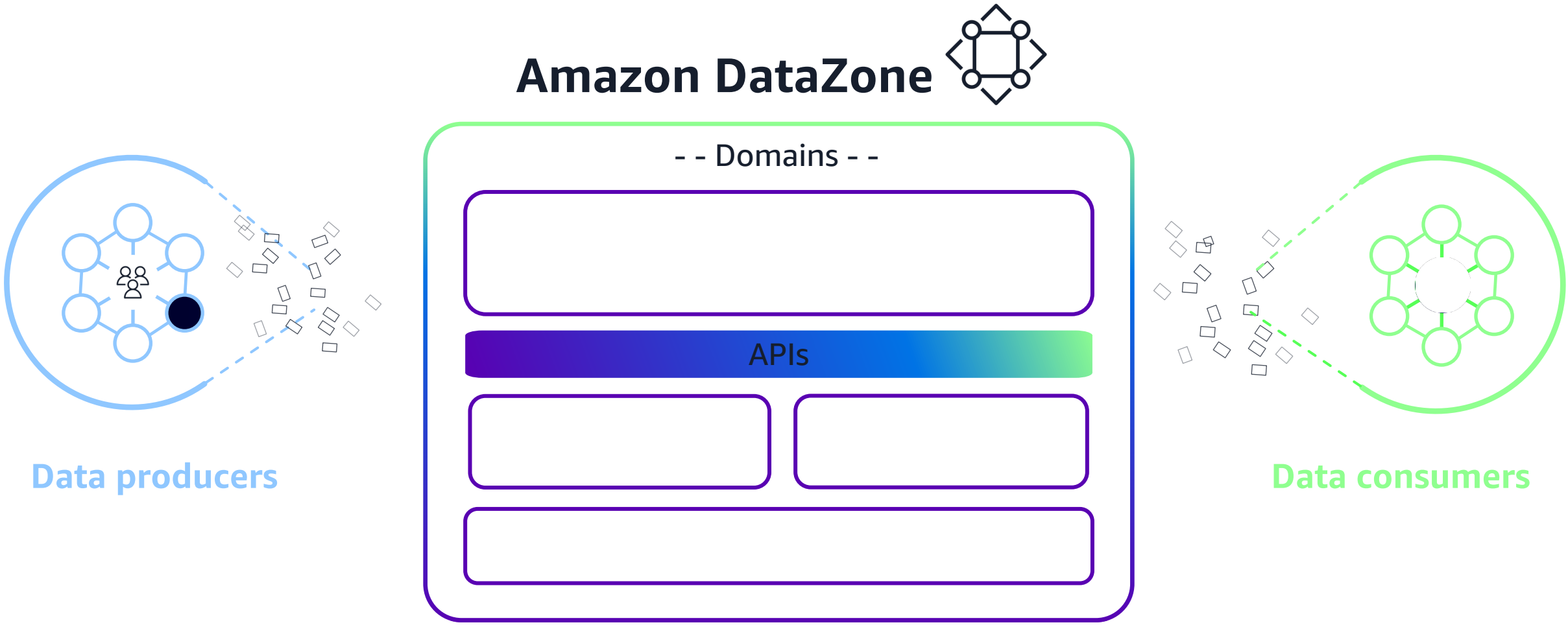
Data consumers

Consumer products/n-order

Insight, analytics, and ML products to meet business demand

- Customer 360 view
- Financial reporting
- Demand forecasting model
- Ad hoc product analysis

What's in Amazon DataZone?



Accelerating a Modern Data Community with Amazon DataZone

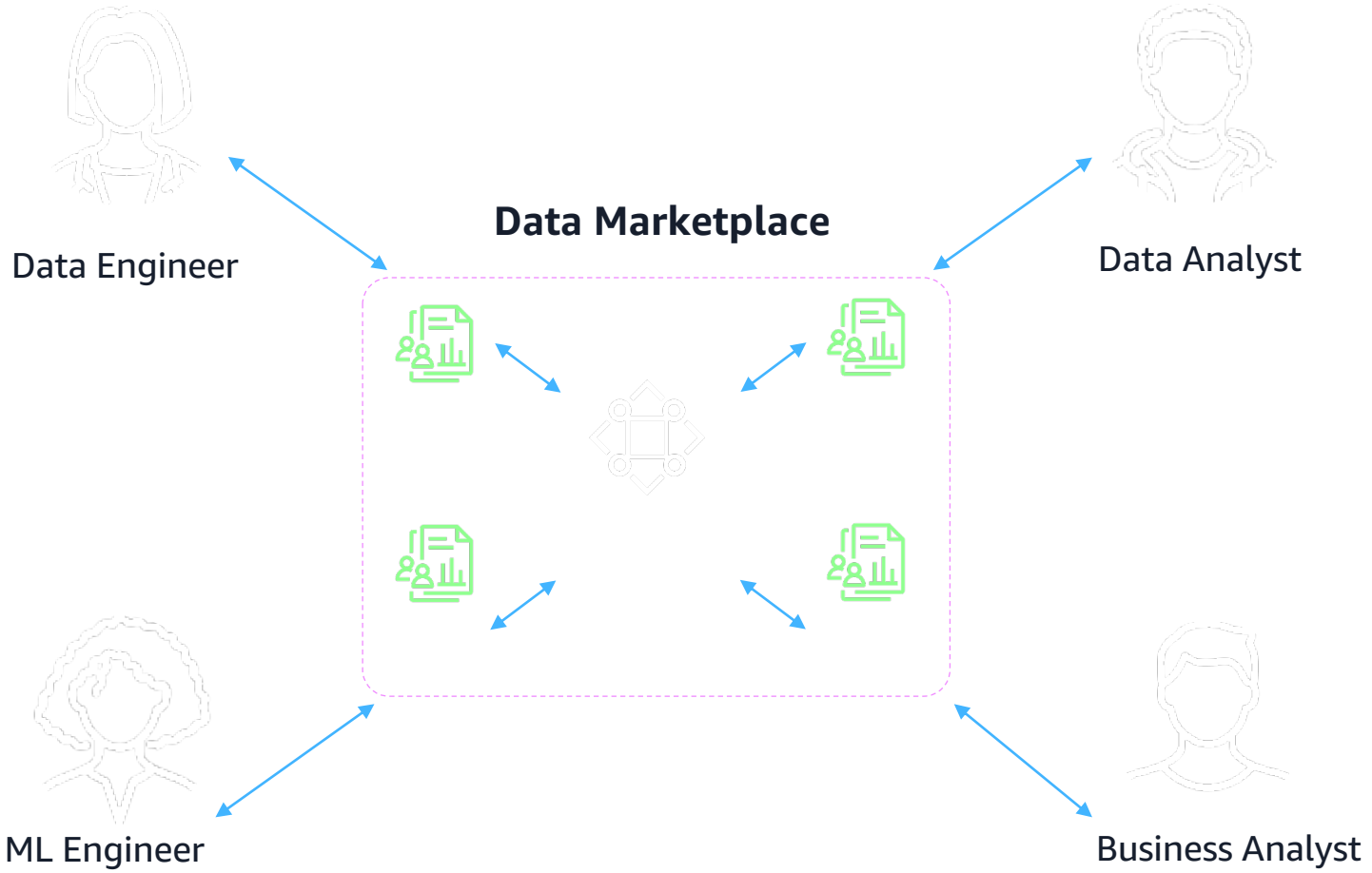
SCALE AND REMOVE BOTTLENECKS TO FULLY DEMOCRATIZE DATA

DECENTRALIZED OWNERSHIP

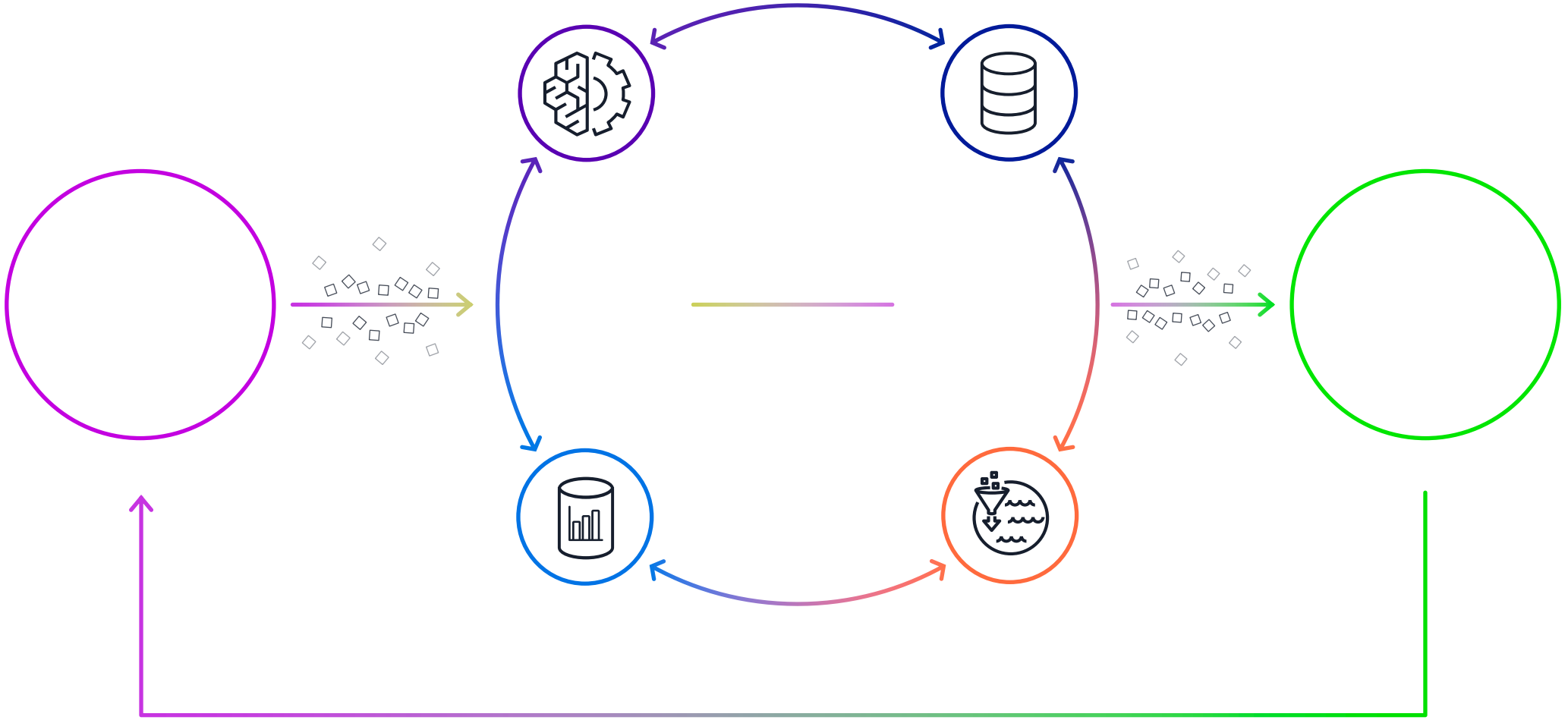
FEDERATED GOVERNANCE

PEER-TO-PEER
DATA SHARING

SELF SERVICE
INFRASTRUCTURE

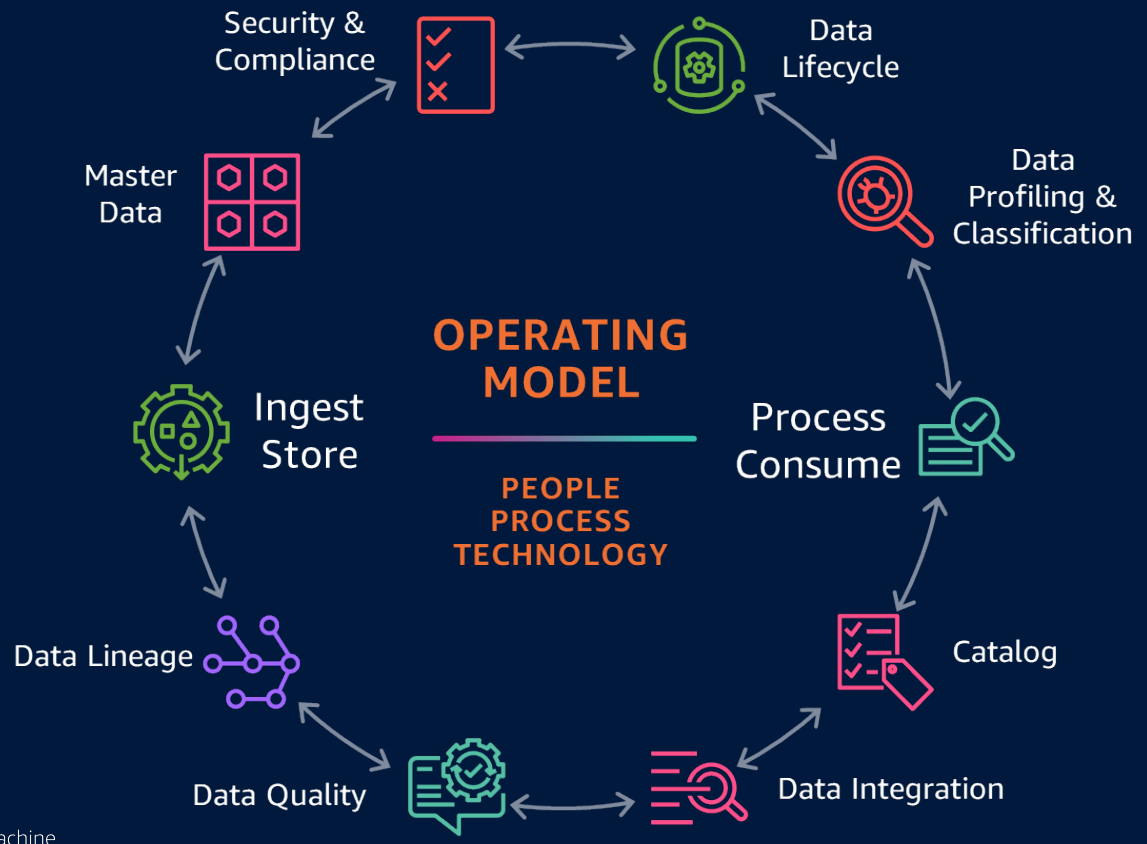


Modern data strategy in action



AWS approach to data modernization

Data governance is the combination of people, processes, and technology that organizations use to ensure the quality and security of their data throughout its lifecycle



THINK BIG, START SMALL, SCALE FAST

1. Architect data governance to support the wider data strategy
2. Implement incrementally based on business initiatives and use cases that drive the data strategy
3. Further evolve data governance capabilities over time

aws machine learning

Purpose-built databases

	Relational	Key-value	Document	In-memory	Graph	Time-series	Ledger	Wide Column
	Referential integrity, ACID transactions, schema-on-write	High throughput, Low latency reads and writes, endless scale	Store documents and quickly access querying on any attribute	Query by key with microsecond latency	Quickly and easily create and navigate relationships between data	Collect, store, and process data sequenced by time	Complete, immutable, and verifiable history of all changes to application data	Scalable, highly available, and managed Apache Cassandra-compatible service
AWS Service(s)	Aurora, RDS	DynamoDB	DocumentDB	ElastiCache	Neptune	Timestream	QLDB	Keyspaces Managed Cassandra
Common Use Cases	Lift and shift, ERP, CRM, finance	Real-time bidding, shopping cart, social, product catalog, customer preferences	Content management, personalization, mobile	Leaderboards, real-time analytics, caching	Fraud detection, social networking, recommendation engine	IoT applications, event tracking	Systems of record, supply chain, health care, registrations, financial	Build low-latency applications, leverage open source, migrate Cassandra to the cloud

When you are a hammer, everything looks like a nail!

Modern data architecture core characteristics

Durability and Availability

Replicate data across regions and availability zones to ensure your data is available globally with 99.999999999% durability and 99.99%+ availability

Security

Protect data with advanced encryption, fine grain access control (IAM), encryption key management (KMS), logging (CloudWatch / CloudTrail), and sensitive data discovery (Macie)

Object Level Controls

Fine-grain, object level control allows tagging of valuable data for replication and tiered storage, saving money, and increasing performance

Flexibility

Storing all data in one data platform avoids data silos and the cost of moving data around

Operation Data Store

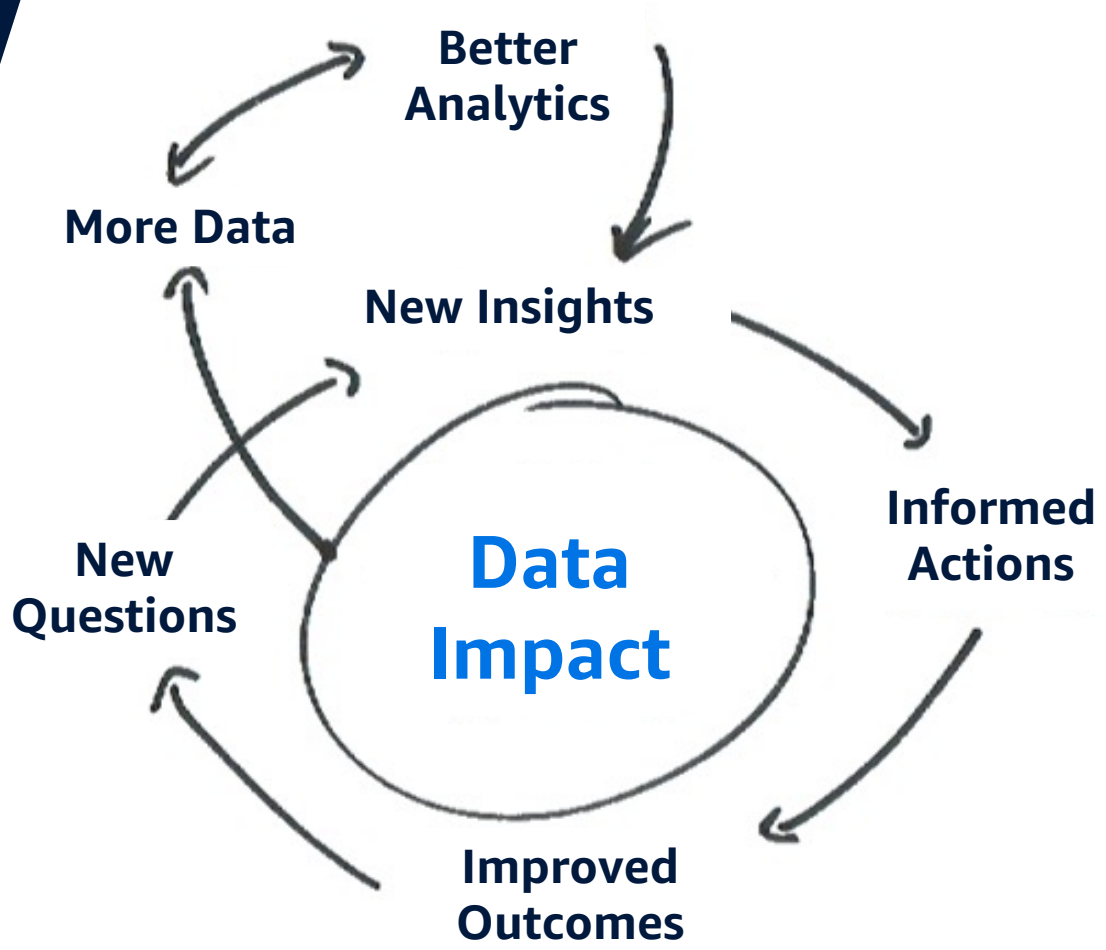
Creating an Operational Data Store (ODS) to access structured frequently used data for real time insights with off the shelf API

ML/AI

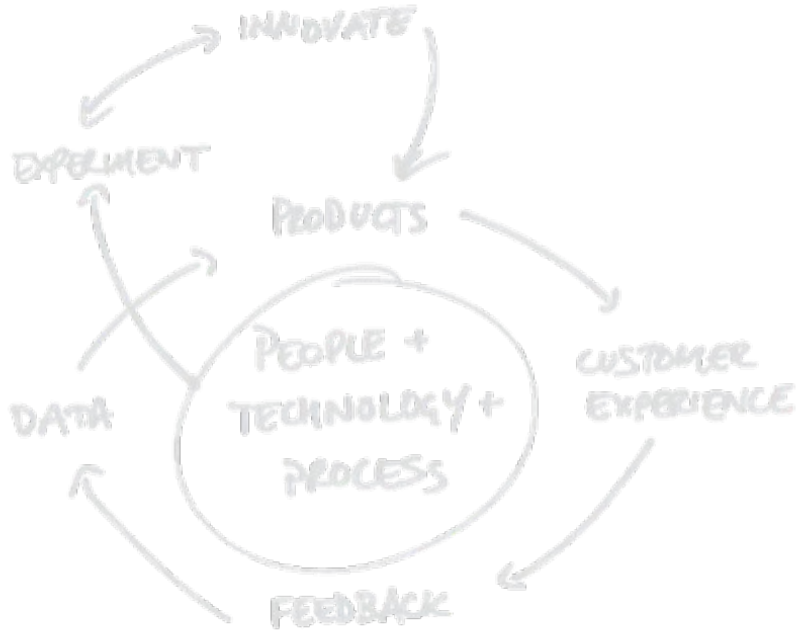
Once your data is in an AWS data platform, automate data transport and security functions, and pull business insights faster and more efficiently with ML/AI

The Data Flywheel

- Start small, develop skills
- Uncover potential
- Deliver results
- Build relationships
- Iterate analytics
- Extend architecture



Think big, start small, scale fast

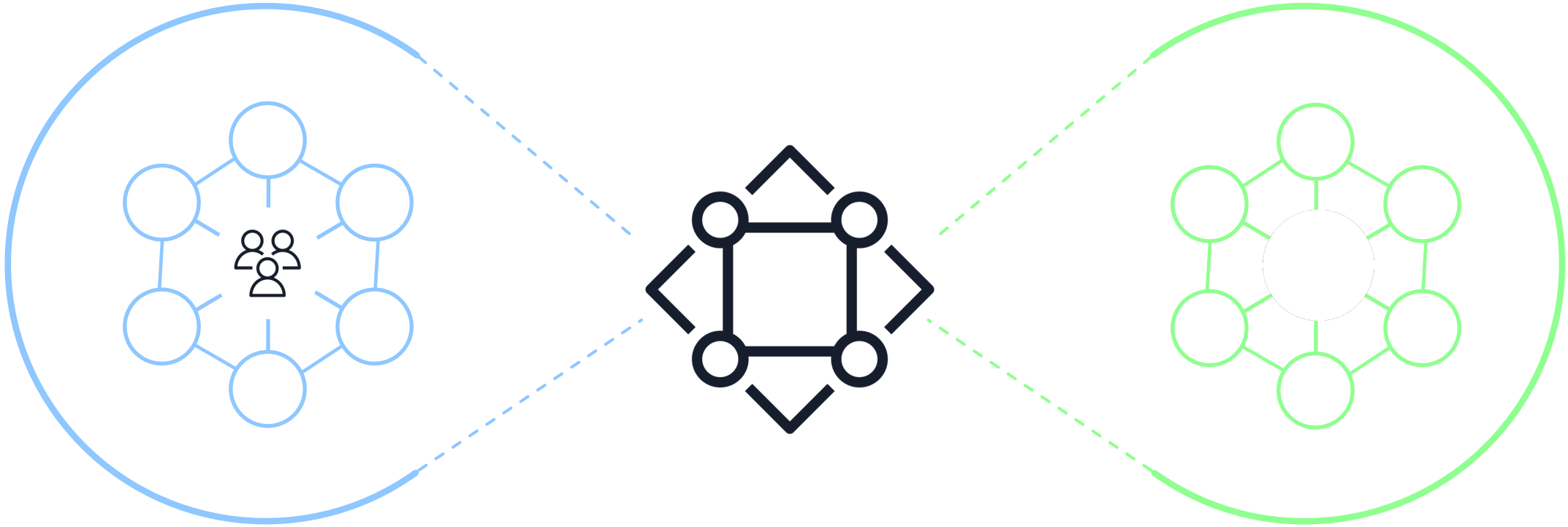


drive sustained innovation
insights
supercharge the experience
demand
asset
actionable

- Set “Think Big” goals
- Focus on delivering business priorities fast
- Grow business-IT ownership
- Increase agility across data producers and consumers
- Upskill and empower self serve
- Build trust and confidence with privacy, security, and compliance

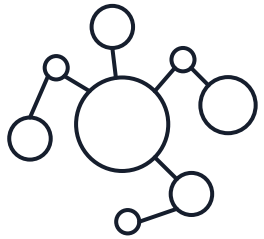
The Modern Data Community

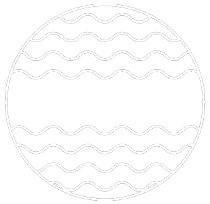
DATA-DRIVEN ORGANIZATIONS ENABLE AGILITY BY PUSHING RESPONSIBILITY TO THE EDGES, TO THE PRODUCERS AND CONSUMERS OF DATA

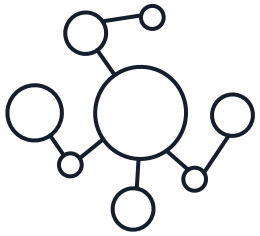


Re-envision the world as products in the community

Consumers of 1st-order products can produce 2nd-order products







Data producers

Producer products/1st order

Foundational data products to serve a wide range of business use cases

- Vendor
- Customer
- Employee
- Product

Data marketplace

Data marketplace products

Technology products for producer and consumer communities to use

- Data quality and ETL tooling
- Data catalog
- Data and ML Ops
- Security controls
- Training

Data consumers

Consumer products/n-order

Insight, analytics, and ML products to meet business demand

- Customer 360 view
- Financial reporting
- Demand forecasting model
- Ad hoc product analysis

Amazon DataZone

CATALOG, DISCOVER, SHARE, GOVERN, AND ANALYZE DATA PRODUCTS ACROSS ORGANIZATIONAL BOUNDARIES

Recommendations

Think big, start small, scale fast

Work backwards from the customer

Form multi-disciplinary teams

Build community, celebrate success

Automate tasks to increase adoption

Store tabular data at scale with Amazon S3 Tables

<Speaker name>

<pronouns>
<Title>
Amazon Web Services

Agenda

01 Introduction to Amazon S3 Tables

02 How it works

03 Use cases and workloads

04 Recap

Customers use S3 for a variety of workloads

Today, S3 is a tabular data store