Put your data to work with the best of both data lakes and purpose-built data stores

Natacha Maheshe

AWS Analytics Product Marketing

The next wave of reinvention will be driven by data

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

However, challenges are in the way

More data than ever is being generated

Data of all types is stored in silos across multiple data stores

Machine learning adoption is challenged by lack of skills and organizational inertia

Data security, privacy, and compliance regulations are increasingly important

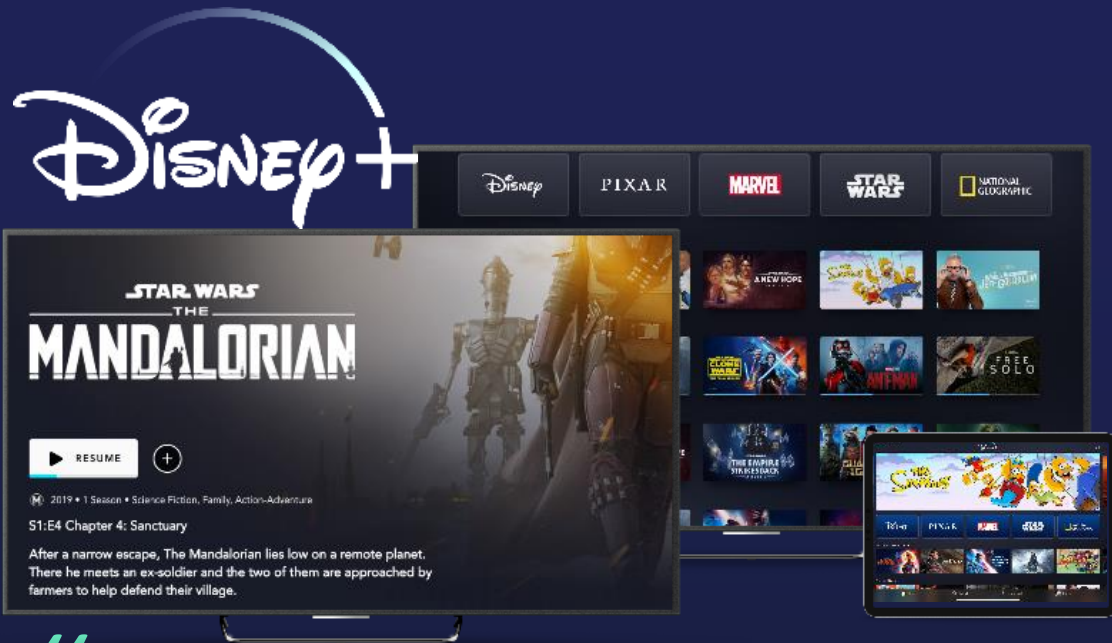# The pandemic last year also taught us that

To survive

Customers are making better and faster decisions to adapt with great urgency to changing market dynamics and customer preferences

Data is critical to making faster, informed decisions

To thrive

Customers are modernizing their data infrastructure, putting data to work,
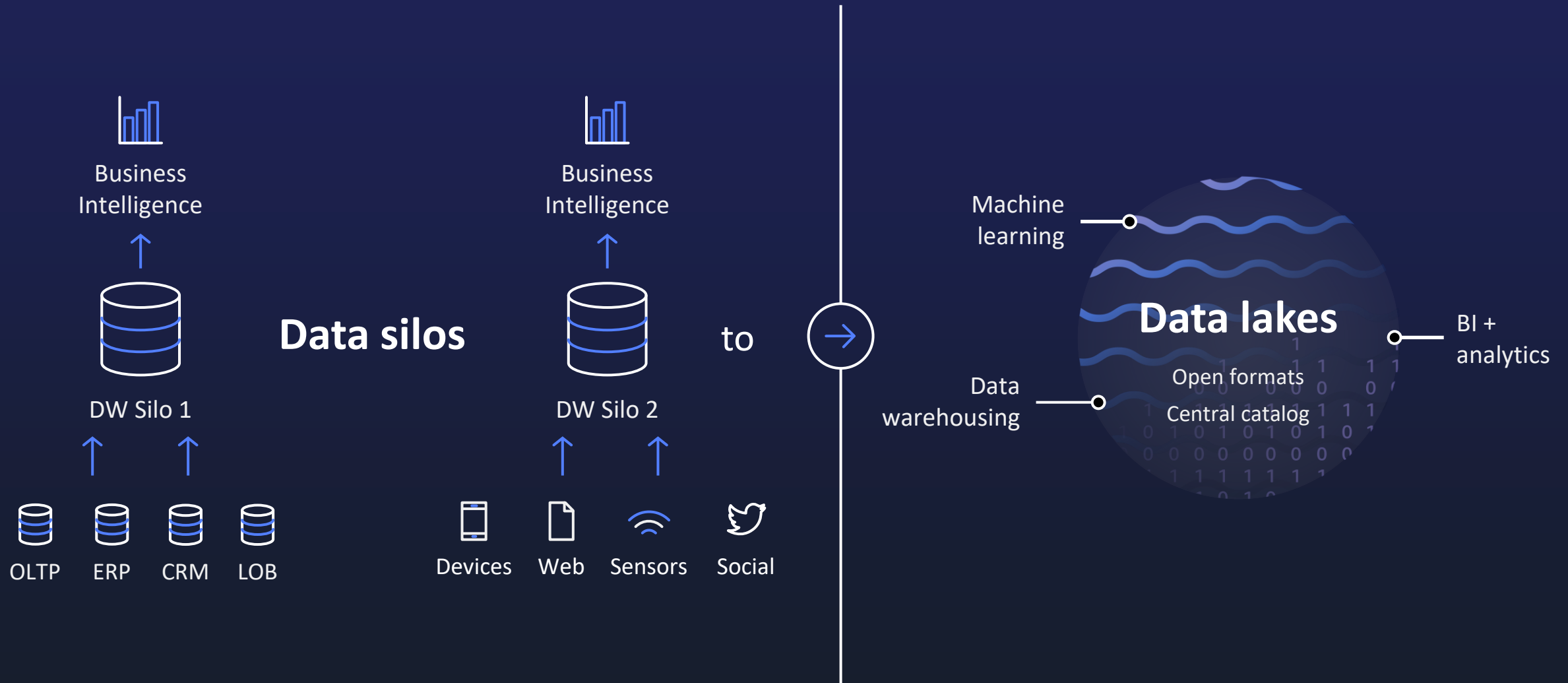and inventing new
customer experiences

The Walt Disney Company leverages AWS's proven global infrastructure to improve performance and reliability for Disney+, as one of the largest online streaming services continues to scale globally.

- **Preferred cloud provider:** The Walt Disney Company relies on AWS as its preferred public cloud infrastructure provider to support the explosive growth of Disney+, which quickly reached its five-year goal of registering **100 million subscribers in 59 countries, 9 months after launch**

- **Purpose-built:** Disney+ uses more than 50 AWS technologies, like machine learning, database, storage, content delivery, serverless, and analytics, including Amazon Kinesis, Amazon DynamoDB, and Amazon Timestream, for quality streaming experiences

- **Performance at scale:** Disney+ handled three billion requests for Disney+ content in the first three days post launch. It also successfully managed huge peaks in demand for premium content, like "Hamilton," "Mulan," and "The Mandalorian"

" AWS has been our preferred cloud provider for years, and its proven global infrastructure and expansive suite of services has contributed meaningfully to the incredible success of Disney+. "

—Joe Inzerillo
Executive Vice President & CTO, Direct-to-Consumer
The Walt Disney Company

Modern cloud-based applications

Search | Index optimized store

Shopping cart | Relational database

Customer reviews | Key-value database

Recommendations | Graph database

Traditional data warehousing approaches don't scale
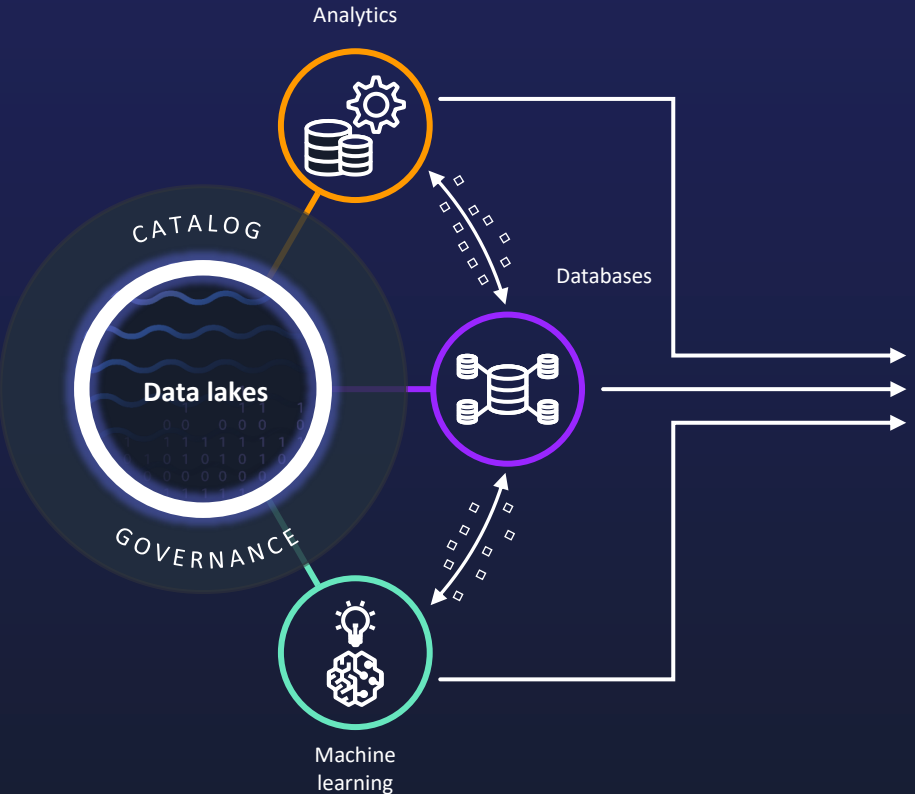
Business Intelligence

Business Intelligence

Data silos

DW Silo 1

DW Silo 2

to

OLTP   ERP   CRM   LOB

Devices   Web   Sensors   Social

Machine learning

Data lakes

Data warehousing

Open formats

Central catalog

BI + analytics

A Modern Data Strategy

Non-relational databases

Big data analytics

CATALOG

Log analytics

Data lakes

Relational Databases

GOVERNANCE

Data warehousing

Machine learning

Data at any scale

The best price-performance

Unified data access, security and governance

AI & ML to solve business challenges

A Modern Data Strategy

Amazon DynamoDB

Amazon EMR

Amazon OpenSearch Service

AWS GLUE

Amazon Athena

Amazon S3

AWS LAKE FORMATION

Amazon Aurora

Amazon Redshift

Amazon SageMaker

Data at any scale

The best price-performance

Unified data access, security and governance
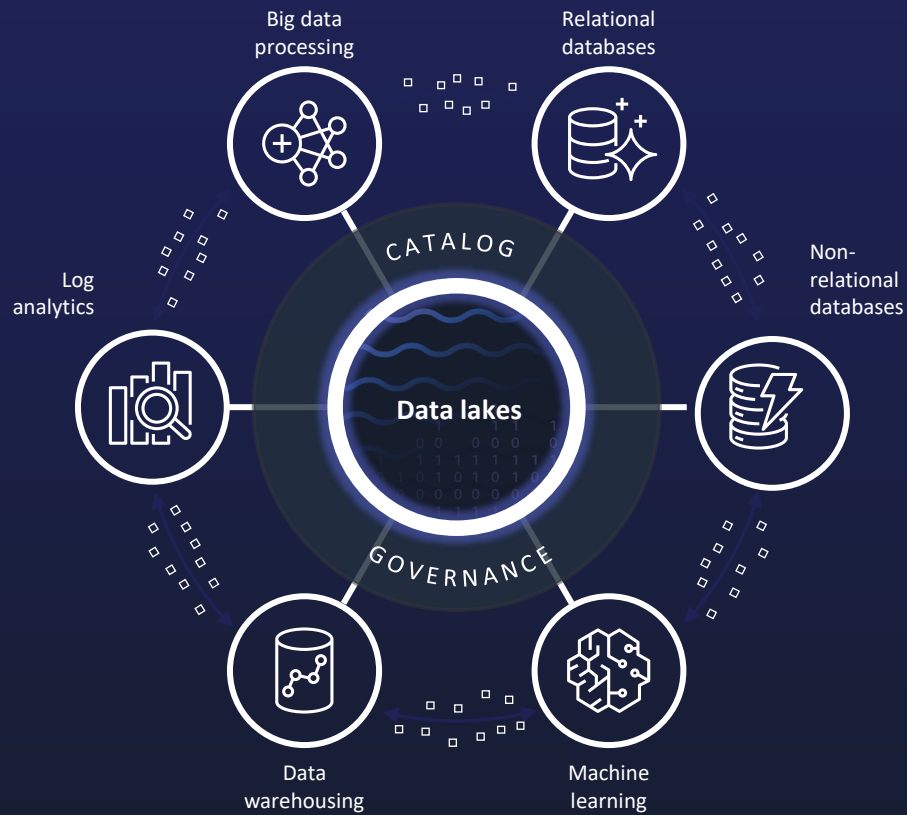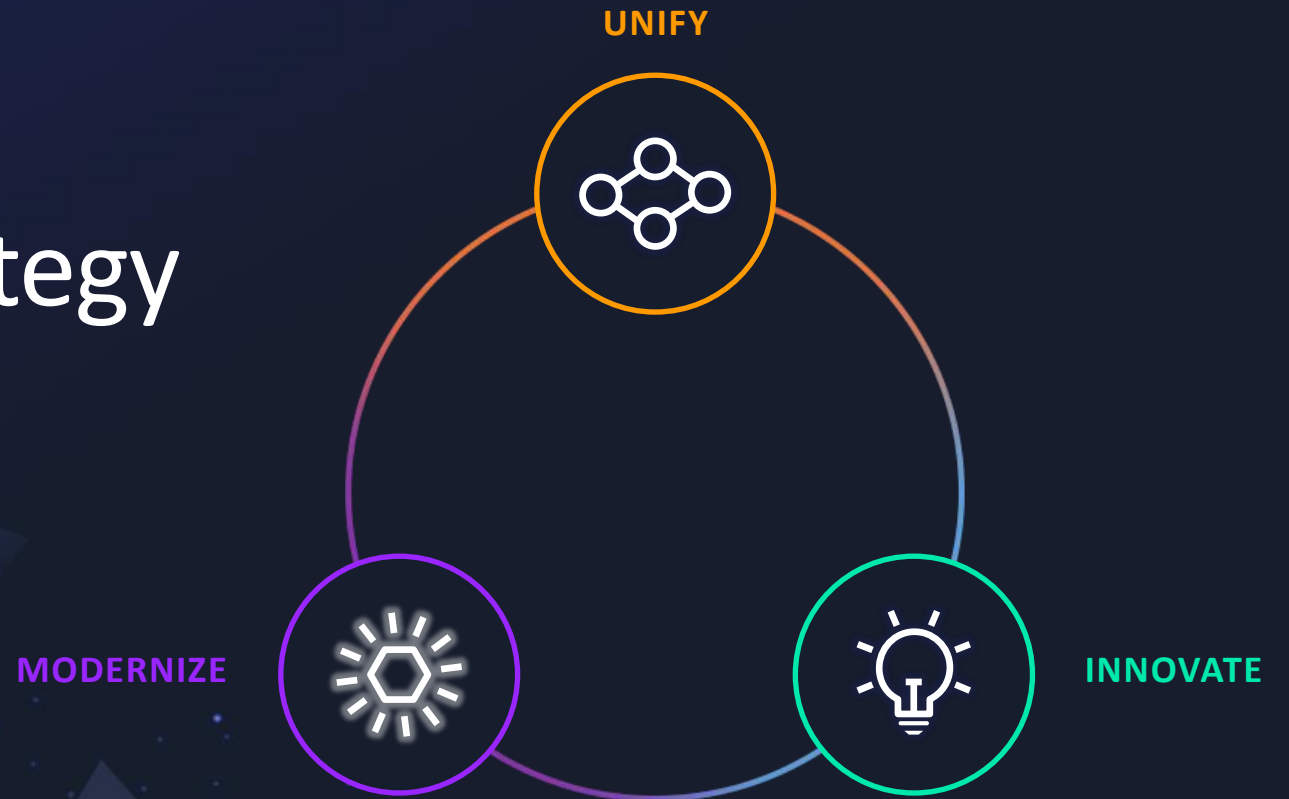
AI & ML to solve business challenges

Inside out

Outside In

Around the perimeter

Unify

Put your data to work with secure and well-governed access to data, using the best of both data lakes and purpose-built data stores

UNIFY

Equinox sees faster reports, 80% cost-savings

CHALLENGE

Their data warehouse had limited integration, was very expensive, and required a lot of platform specific domain knowledge

SOLUTION

Migrated from on-premises data warehouse to Amazon S3 as their data lake, Amazon Redshift, and Amazon EMR

RESULTS

- Monthly bill is now 20% of prior yearly maintenance of their legacy data warehouse

- Reduced report delivery time from months to days

ON-PREMISES DW →   Amazon Redshift   Amazon S3   Amazon EMR

ENGIE builds the Common Data Hub on AWS, accelerates zero-carbon transition

Challenge

ENGIE's decentralized global customer base had accumulated lots of data, and it required a smarter, unique approach and solution to align its initiatives and to efficiently provide data across its global business units.

Solution

ENGIE built its Common Data Hub data lake on AWS, enabling the company's business units to collect and analyze data to support a data-driven strategy and to lead the zero-carbon transition.

Result

- Collected 95 TB of data across 351 projects
- Automated energy predictions
- Maximized wind farm energy production

Amazon Kinesis Data Streams   Amazon Redshift   AWS Glue   Amazon Athena   Amazon S3   Amazon SageMaker

Lake House approach on AWS

Amazon DynamoDB

Amazon EMR

Amazon OpenSearch Service

AWS GLUE

Amazon Athena

Amazon S3

Amazon Aurora

AWS LAKE FORMATION

Amazon Redshift

Amazon SageMaker

Scalable data lakes

Purpose-built for performance and cost

Unified data access

Unified governance

ML integration

Scalable data lakes

Amazon S3 is the most secure, durable, and scalable storage to build your data lake

Unmatched durability, availability, and scalability

Most object-level controls

Easiest to use with cost optimization – S3 Intelligent-Tiering

Stores data in standard-based open formats

Amazon S3

Most ways to get data in

Unmatched security, compliance, and audit capabilities

Cold storage and archive capabilities

AWS Lake Formation

Build a secure data lake in days

Build data lakes quickly
Move, store, and catalog your data faster

Simplify security management
Centrally define security, governance, and auditing policies in one place

Provide self-service access to data

More data lakes run on AWS than anywhere else

Purpose-built for performance and cost

Purpose-built data services

Optimize performance, cost, and scale for your use cases

Amazon Athena	Amazon EMR	Amazon Elasticsearch Service	Amazon Kinesis and Amazon MSK	Amazon Redshift
Interactive query	Big data processing	Log and search analytics	Real-time analytics	Data warehousing

Amazon Redshift

Analyze all your data with the fastest and most widely used cloud data warehouse

Performance at any scale
Up to 3x better price performance than other cloud DW

Analyze all your data
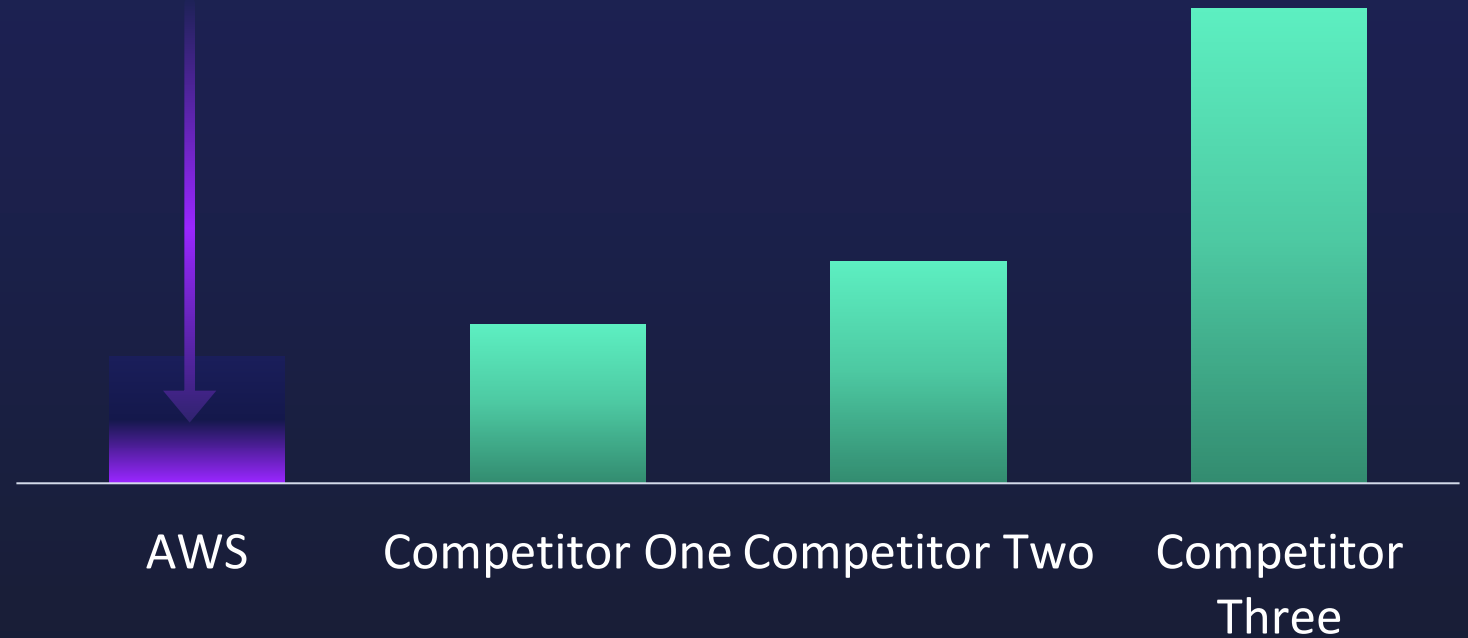Deepest integration with your data lake

Lower your costs
At least 50% less expensive than other cloud DW

Amazon Redshift — Built for the cloud. Up to 3x better price performance than other cloud data warehouses. Bar chart comparing AWS, Competitor One, Competitor Two, and Competitor Three.

Amazon Redshift

BUILT FOR THE CLOUD

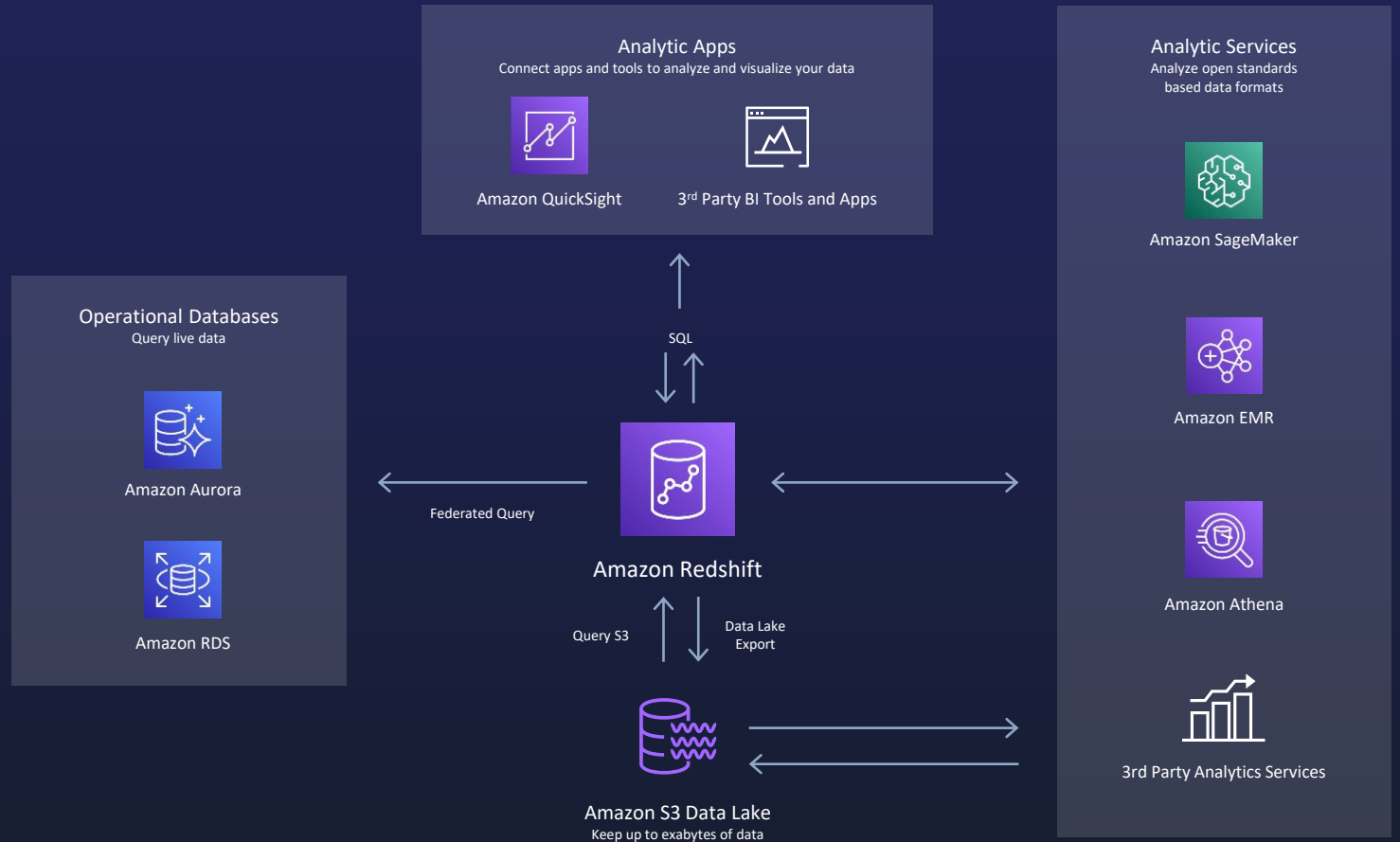AWS designed-hardware

Highest networking bandwidth

Advanced ML that automatically organizes and stores data for performance

Intelligently moves data between data lake and data warehouse to lower costs

Amazon Redshift

ANALYZE YOUR DATA

Analytic Apps
Connect apps and tools to analyze and visualize your data

Amazon QuickSight   3rd Party BI Tools and Apps

Analytic Services
Analyze open standards based data formats

Amazon SageMaker

Amazon EMR

Amazon Athena

3rd Party Analytics Services

Operational Databases
Query live data

Amazon Aurora

Amazon RDS

Federated Query

SQL

Amazon Redshift

Query S3   Data Lake Export

Amazon S3 Data Lake
Keep up to exabytes of data

Amazon EMR

Easily Run Spark, Hadoop, Hive, Presto, HBase, and other big data frameworks

Automate provisioning, configuring, and tuning
Easy setup, management, and monitoring

Get the latest, stable, open-source releases
Latest open-source framework updates within 30 days
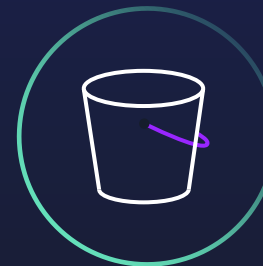
Automatically scale up and down
Manage cluster size based on utilization to reduce costs

Simple and predictable pricing
Per-second pricing, and save 50%–80% with Amazon EC2 Spot and Reserved Instances

Amazon OpenSearch

Search, visualize, and analyze up to petabytes of text and unstructured data
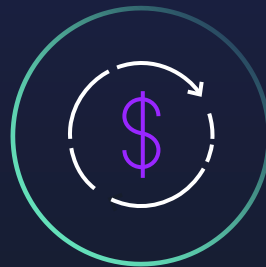
Fully managed
Operate OpenSearch with the leading contributor of the community-driven, open source software.

Easily accessible
Quickly search and analyze your unstructured and semi-structured data to easily find what you need.

Cost-effective
Eliminate operational overhead and reduce cost with automated provisioning, software installation, patching, storage tiering, and more.

Amazon Kinesis

Easily collect, process, and analyze data and video stream in real time

Kinesis Data Streams
Easy setup, management, and monitoring

Kinesis Data Analytics
Latest open-source framework updates within 30 days

Kinesis Video Streams
Manage cluster size based on utilization to reduce costs

Kinesis Data Firehose
Per-second pricing, and save 50%–80% with Amazon EC2 Spot and Reserved Instances

Amazon MSK

Fully managed, highly available, and secure Apache Kafka service

Kafka

Compatible
Migrate and run existing Apache Kafka applications on AWS without changes

Fully managed
AWS manages the provisioning, configuration, and maintenance of Apache Kafka clusters

Highly available
Continuous cluster health monitoring and component replacement

Secure
Multiple levels of security – network isolation, authorization, encryption, and access control

Seamless data access

AWS Glue

Simple, scalable, and serverless data integration

Integrate data faster
Prepare and put your data to use from months to minutes

Automate at scale
Easily run and manage thousands of ETL jobs

No servers to manage
Pay only for the resources your jobs use while running

Combining and replicating data between multiple stores can be challenging

Data stores
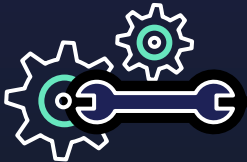
Data pipeline

Target store

COMPLEX APPLICATION CODE

CUMBERSOME RETRY LOGIC

PIPELINE MANAGEMENT

NEED SPECIALIZED ETL SKILLS

Break down data silos

Move your data, at scale, to where you need it the most

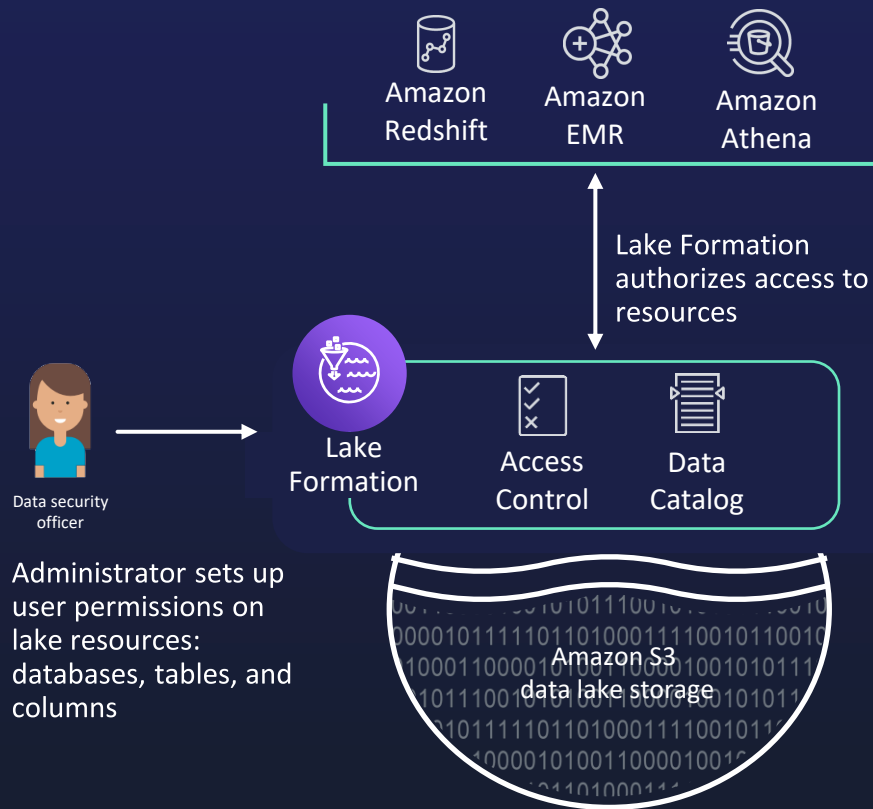|				
Extract, transform, load	Visual data preparation	Data replication	Data warehouse to/from data lake	Federated query

Data lake

Unified governance

Simplify security management with Lake Formation
Centrally define security, governance, and auditing policies in one place

Amazon Redshift

Amazon EMR

Amazon Athena

Lake Formation authorizes access to resources

Data security officer

Lake Formation

Access Control

Data Catalog

Administrator sets up user permissions on lake resources: databases, tables, and columns

Amazon S3 data lake storage

Centrally define security, governance, and auditing

Policies are consistently enforced

Integrated with security, storage, analytics, and machine learning services

Permissions on databases, tables, and columns

ML integration

AWS brings ML closer to data

Data stores, data lakes, and BI tools with built-in ML

Databases

+

**Data warehouses
+ data lakes**

+

**Business
intelligence tools**

Amazon Aurora ML	Amazon Neptune ML	Amazon Redshift ML	Amazon Athena ML	Amazon Quicksight ML

Redshift ML

Create, train, and deploy ML models using familiar SQL commands

No prior ML experience required

Simple, optimized, and secure integration between Redshift and Amazon SageMaker

Use a simple SQL query to specify the data

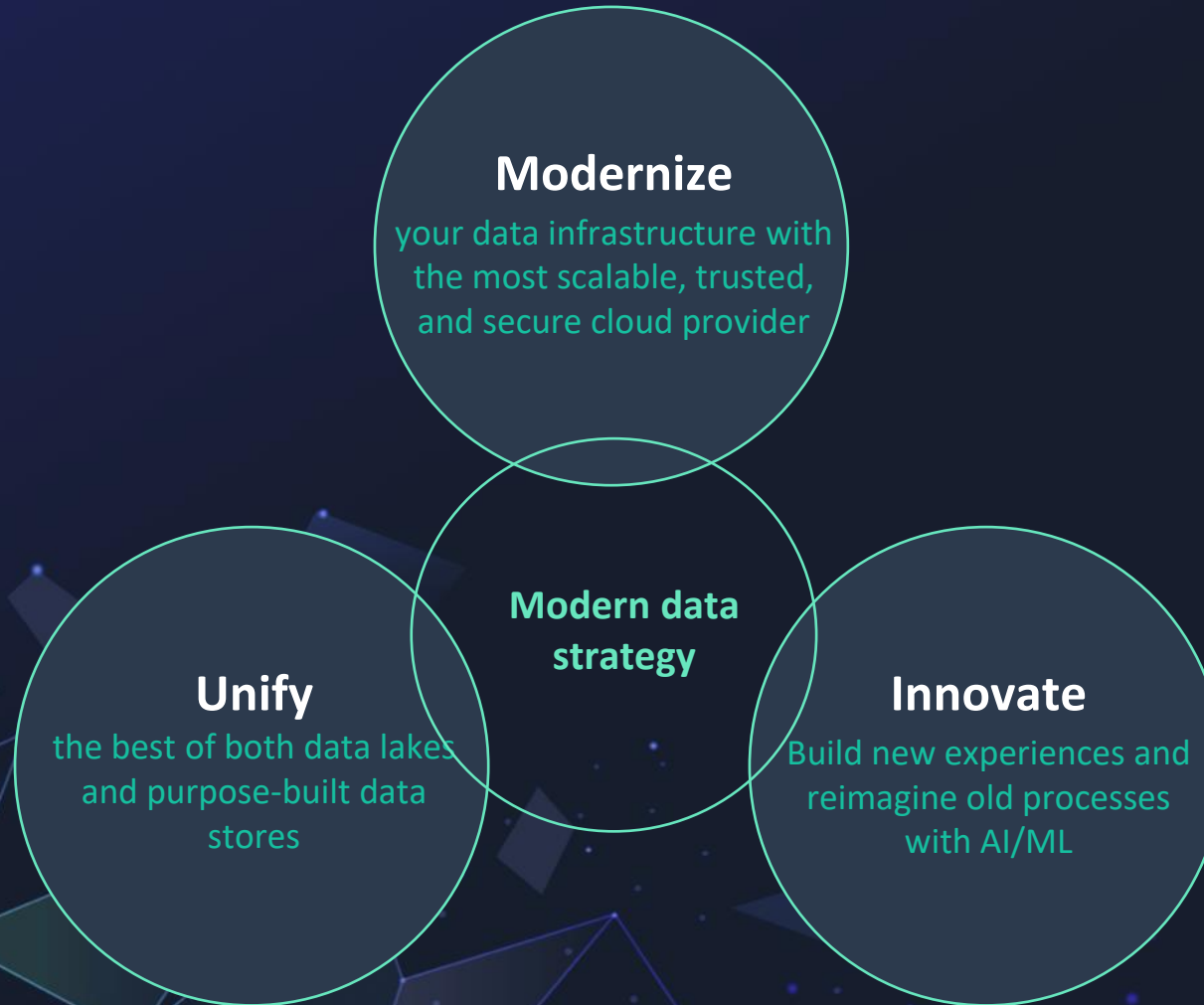Train and deploy an ML model using a SQL command in your data warehouse

Fast predictions

Embed predictions like fraud detection, risk scoring, and churn in queries and reports

Thank you!

Natacha Maheshe

mnatach@amazon.com

aws.amazon.com/data