

AWS Data, Databases, and Analytics Online Series



Live Opening Keynote – Day 1

Olivier Klein

Lead Technologist, AWS



**DATA, DATABASES
& ANALYTICS
ONLINE SERIES**



Agenda at a Glance

AWS Data, Databases, and Analytics Online Series

DAY 1 (JULY 9)

Opening Session

60 mins

Opening Session: Olivier Klein, Lead Technologist, AWS

Tracks and Sessions

	Track 1 Data Movement & Management	Track 2 Migrate your Databases & Data Warehouse to AWS	Track 3 Modernize your Databases & Data Architecture
30 mins	How to ingest data seamlessly to build your data lake (Level 200)	Migrate and modernize from legacy databases to AWS (Level 200)	Purpose-built databases: Choose the right tool for each job (Level 200)
30 mins	Build your data lake on Amazon S3 in days (Level 200)	Deploying open source databases on AWS (Level 300)	Building modern application with modern databases (Level 300)
30 mins	Accelerating data analytics with cloud-native file storage (Level 300)	Migrate your on-premises data warehouse to Amazon Redshift (Level 200)	Extreme performance at cloud scale: Supercharge your real-time applications (Level 300)

DAY 2 (JULY 10)

Opening Session

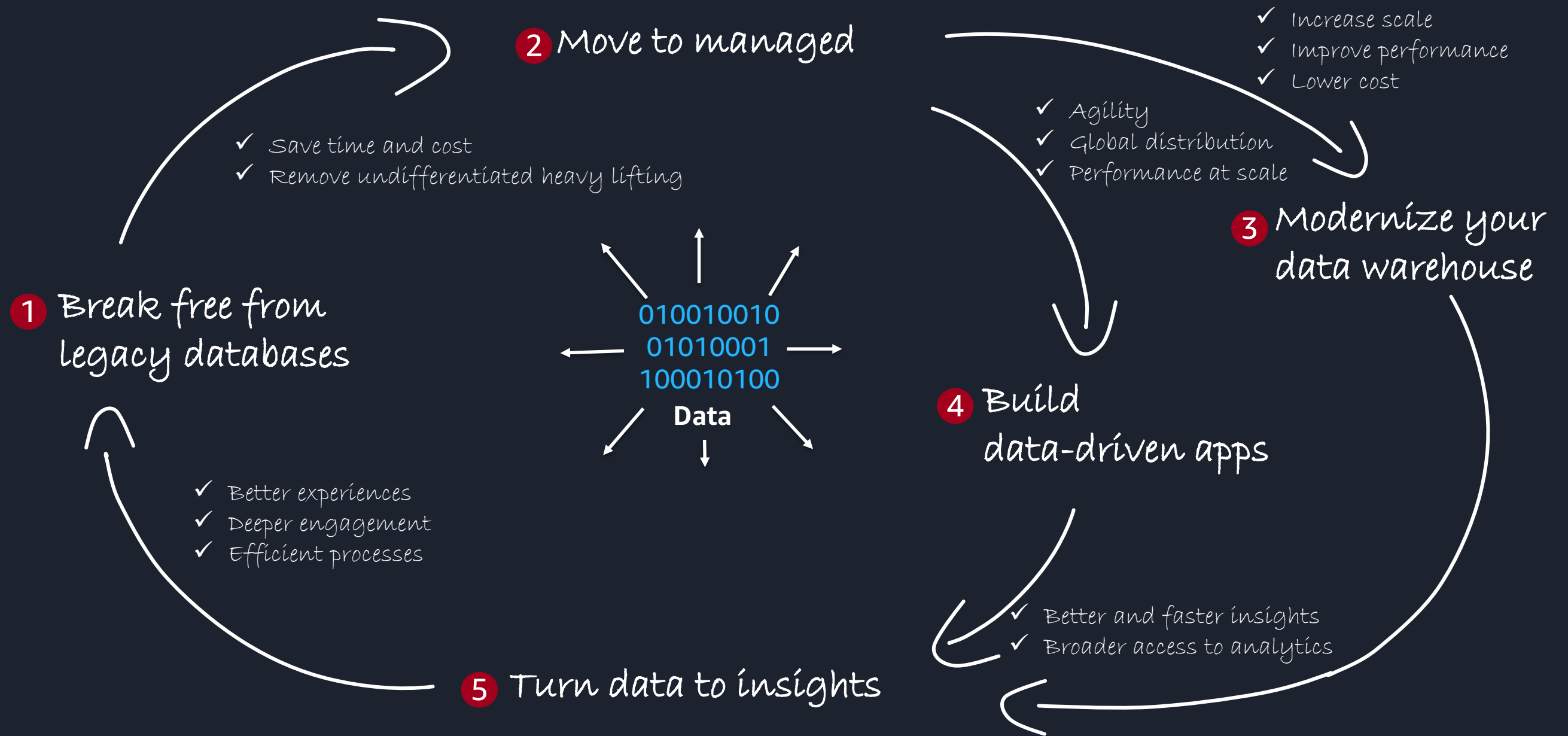
60 mins

Opening Session: Dean Samuels, Lead Architect, AWS

Tracks and Sessions

	Track 1 Accelerate insights from your Data	Track 2 Accelerate insights from your Data
30 mins	Data warehousing on AWS: Amazon Redshift use cases and deployment patterns (Level 200)	Building data integration services for real-time on AWS (Level 200)
30 mins	Processing Big Data with Hadoop, Spark, and other frameworks in Amazon EMR (Level 300)	Unite streaming and batch analytics with AWS Glue (Level 300)
30 mins	Building scalable, secure log analytics with Amazon Elasticsearch Service (Level 200)	Embedding analytics into applications with Amazon QuickSight (Level 200)

The Data Flywheel



The purpose-built database

App architectures & patterns have evolved over the years...

New modern applications

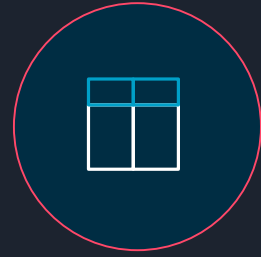
Common data categories and use cases



Relational

Referential integrity, ACID transactions, schema-on-write

Lift and shift, ERP, CRM, finance



Key-value

High throughput, low-latency reads and writes, endless scale

Real-time bidding, shopping cart, social, product catalog, customer preferences



Wide column

Store large amounts of data with virtually unlimited scalability

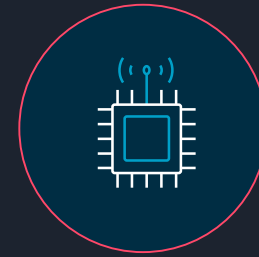
Industrial equipment maintenance, fleet management, route optimization



Document

Store documents and quickly access querying on any attribute

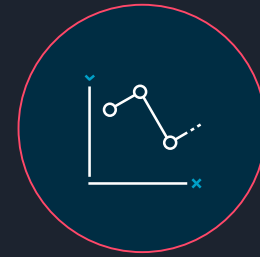
Content management, personalization, mobile



In-memory

Query by key with microsecond latency

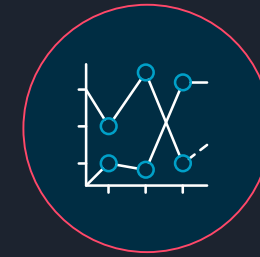
Caching, session store, leaderboard, geospatial services, real-time analytics



Graph

Quickly and easily create and navigate relationships between data

Fraud detection, social networking, recommendation engine



Time-series

Collect, store, and process data sequenced by time

IoT applications, event tracking



Ledger

Complete, immutable, and verifiable history of all changes to application data

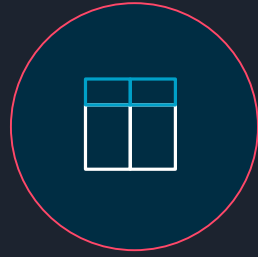
Systems of record, supply chain, health care, registrations, financial

Common use cases

Common data categories and use cases



Relational



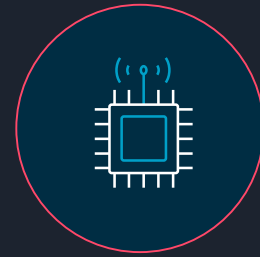
Key-value



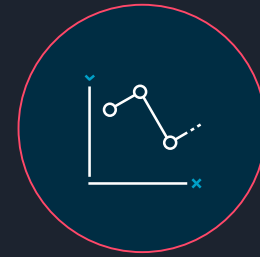
Wide column



Document



In-memory



Graph



Time-series



Ledger



Amazon RDS

Aurora Community Commercial



Microsoft SQL Server



**Amazon
DynamoDB**



**Amazon
Keyspaces
(for Apache
Cassandra)**



**Amazon
DocumentDB
with MongoDB
Compatibility**



**Amazon
ElastiCache**



**Amazon
Neptune**



**Amazon
Timestream**



**Amazon
QLDB**



Amazon DynamoDB:
Ride-tracking system
stores GPS coordinates
for all rides

Amazon Neptune: Social
graph with millions of
relationships among its
users

Nike

AWS Database Migration Service

 →  Migrate between on-premises and AWS

 →  Migrate between databases

 →  Automated schema conversion

 →  Data replication for near-zero downtime migration

More than 200,000 databases migrated





75 PB of data

7,500 Oracle databases

Challenge:

Complex and costly to scale

Expensive and punitive Oracle licenses

Solution:

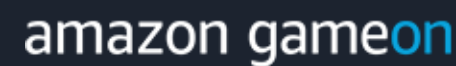
Migrated to Amazon Aurora, Amazon RDS, Amazon DynamoDB, Amazon ElastiCache, and Amazon Redshift

60% Reduction in database costs

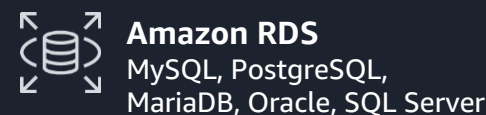
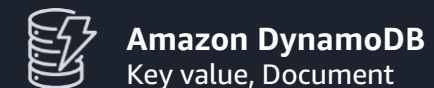
70% Reduction in database administrative overhead

40% Increase in performance of most critical apps

100 Amazon teams



7,500 databases



Turn data into insights

Whiteboard

Data movements

Most ways to move data to the data lake

Professional services and partners
to help migration



**Data movement from
your on-premises
datacenters**



**Data movement from
real-time sources**

**Synchronizing data
across environments**

Data movement from on-premises datacenters

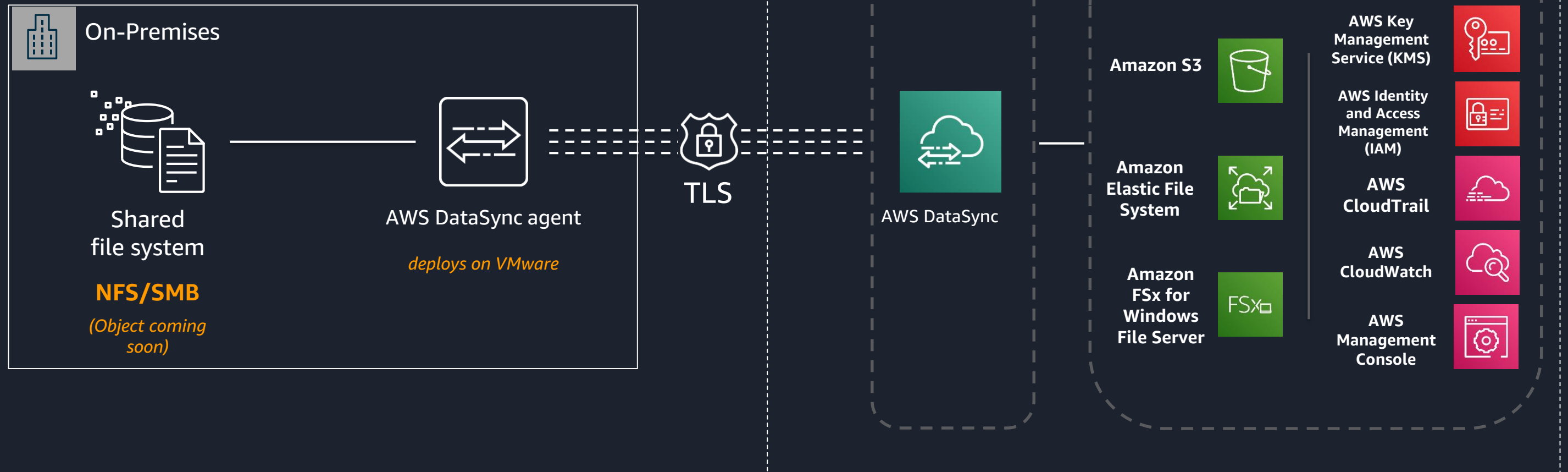
- Dedicated network connection
- Secure appliances
- Ruggedized shipping containers
- Database migration
- Gateway that lets applications write to the cloud

Data movement from real-time sources

- Connect devices to AWS
- Real-time data streams
- Real-time video streams

What is AWS DataSync?

Online transfer service that simplifies, automates, and accelerates moving data between on-premises storage and AWS



The AWS Snow Family

- Enables data collection, data processing, and data movement from the edge to AWS
- Highly-secure, portable and ruggedized devices to collect and process data at the edge, and migrate data into and out of AWS



8 Terabytes



80 Terabytes or
Petabytes
(clustered)



100 Petabytes

AWS Outposts – Bringing AWS on-premise

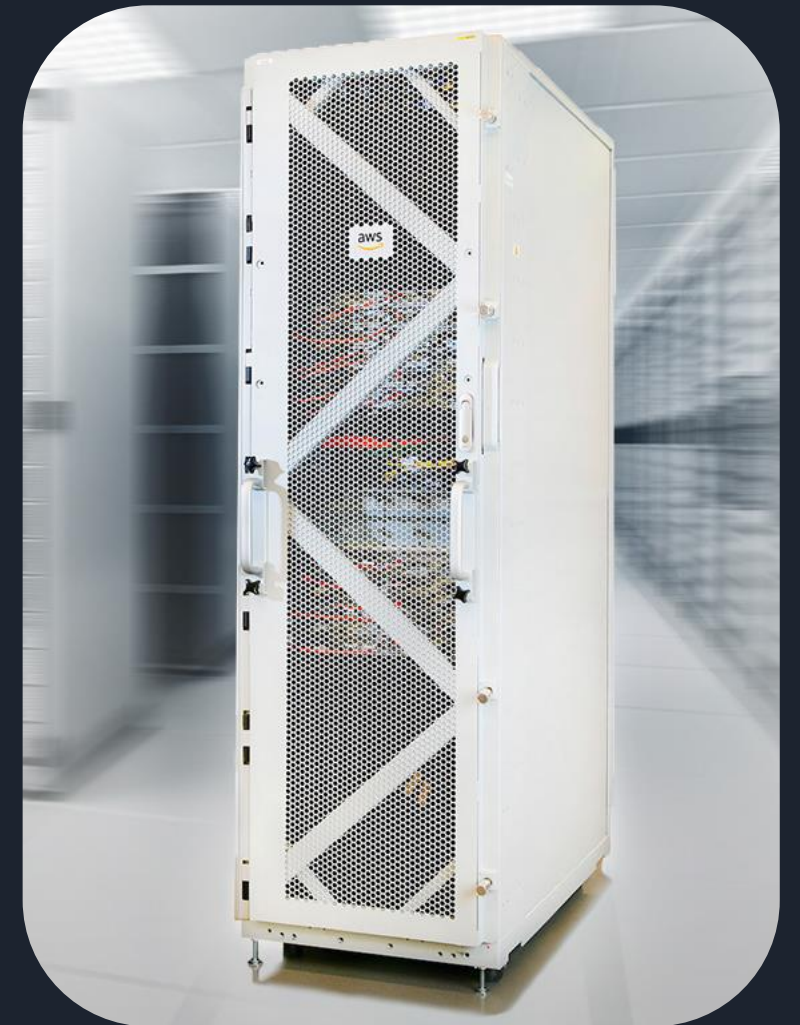


Industry standard **42U rack fully assembled**, ready to be rolled into final position

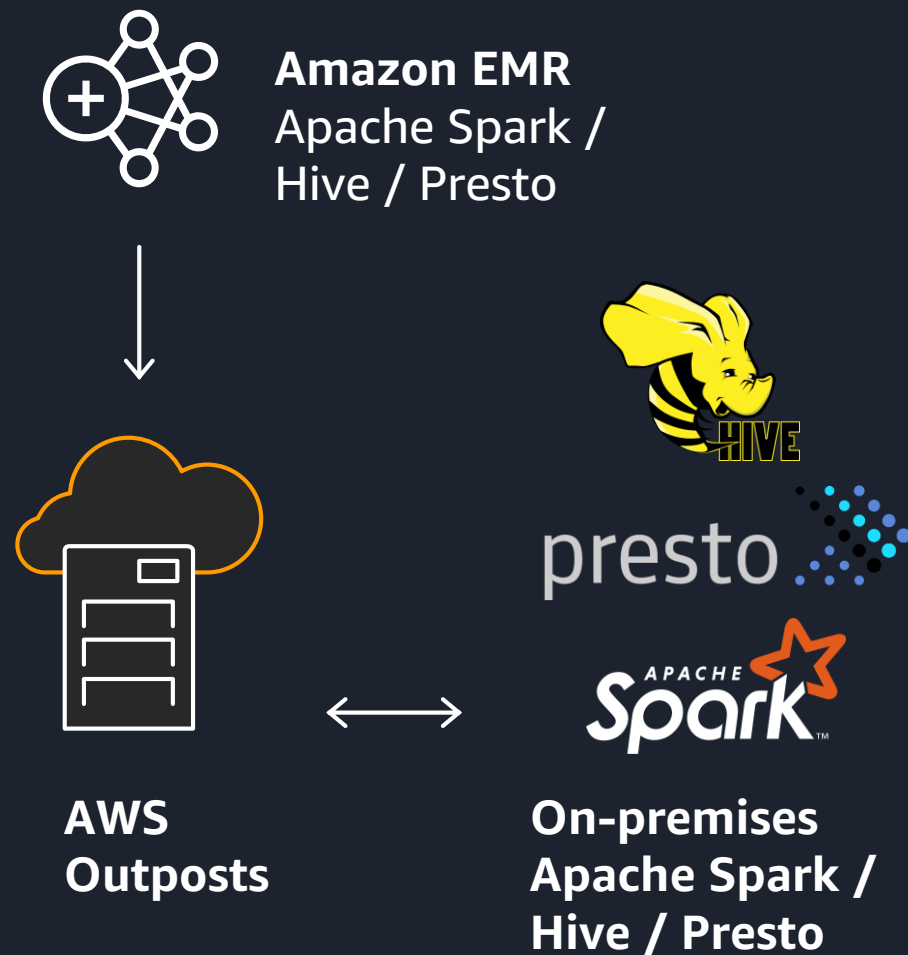
Same AWS-designed infrastructure as in AWS data centers (built on AWS Nitro System)

Fully managed, monitored, and operated by AWS as if in AWS Regions

Single pane of management in the cloud providing the **same APIs and tools** as in AWS Regions hosts



Amazon EMR on AWS Outposts



Launch EMR in your data centers with AWS Outpost

Deploy secure, managed, EMR clusters in minutes

Consistent and seamless hybrid cloud analytics experience

Integrate with existing on-premises Hadoop deployments

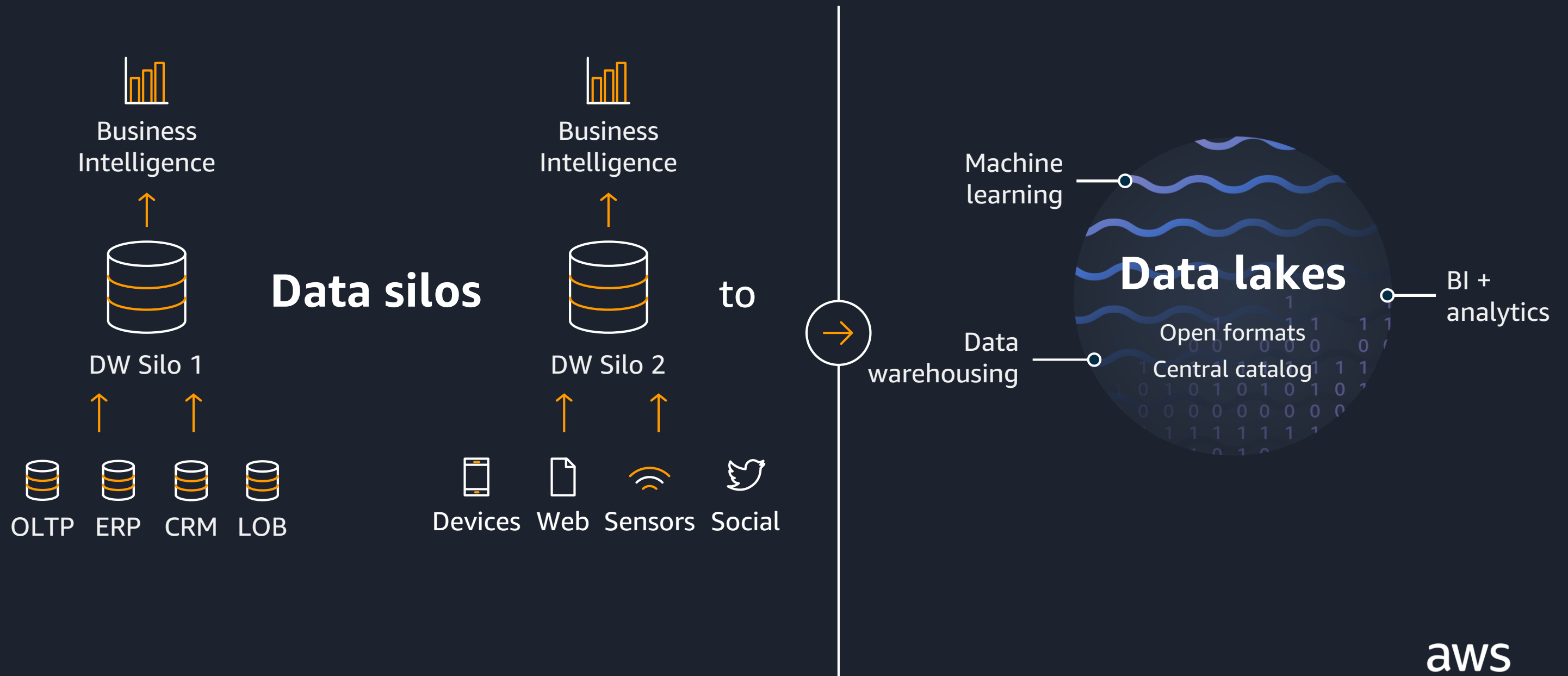
Data lakes – Modernize your data warehouse

More data lakes and analytics than anywhere else

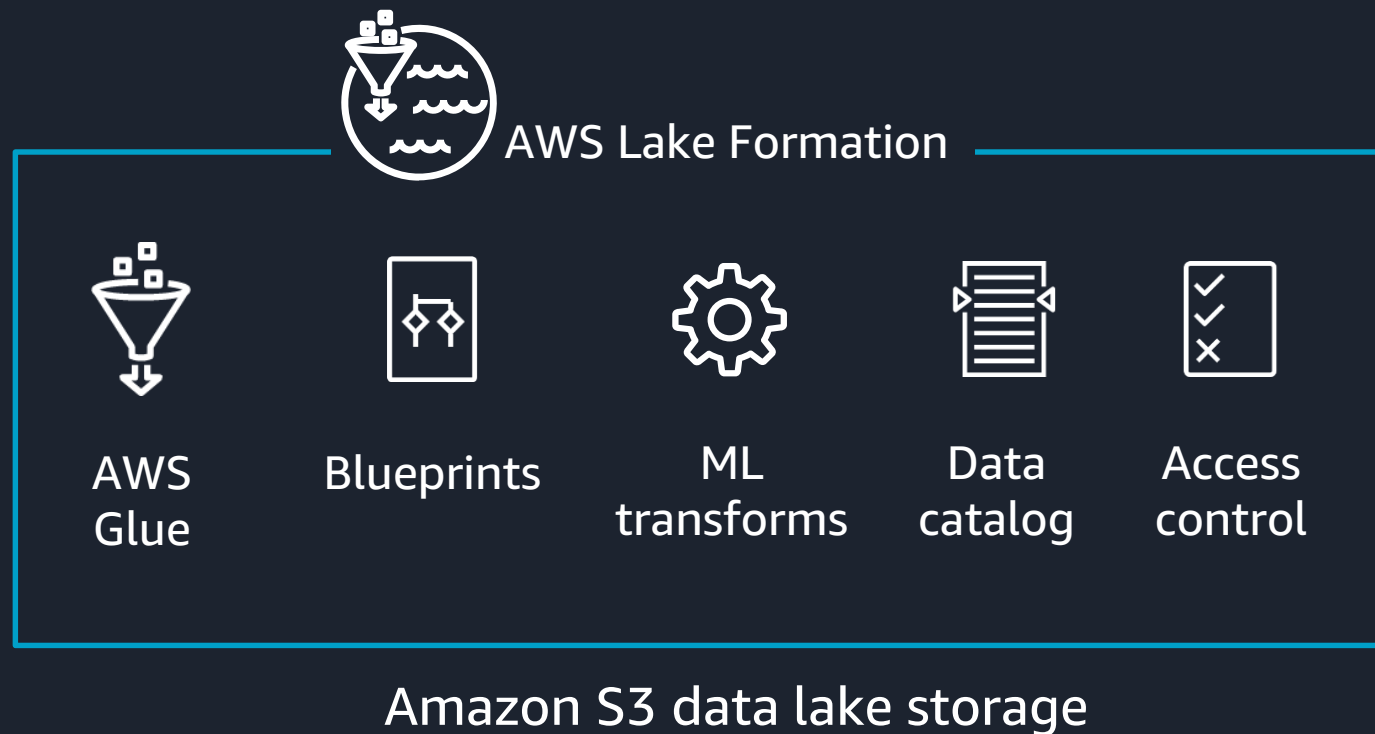
Tens of thousands of data lakes run on AWS across all industries



Traditional data warehousing approaches don't scale



AWS Lake Formation - Build a secure data lake in days



Discovery, sharing, and **integrated tools** to enable every user

Centralized management of **fine-grained permissions** empower security officers

Simplified **ingest & cleaning** enables data engineers to build faster

Cost-effective, durable storage with global replication capabilities

AWS Glue – Data cataloging and simple cost-effective ETL



Data catalog

Apache Hive metastore compatible

Integrated with AWS analytic services



Crawlers

Automatically infer schemas

Populate data catalog

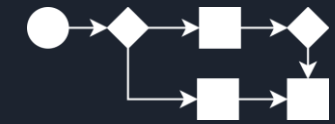


Serverless ETL

Interactive development

Apache Spark / Python shell jobs

Serverless execution



Flexible workflows

Orchestrate triggers, crawlers & jobs

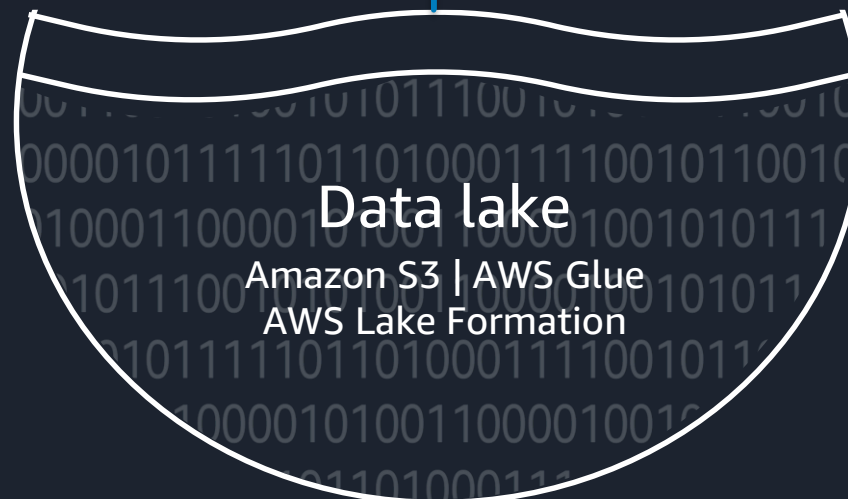
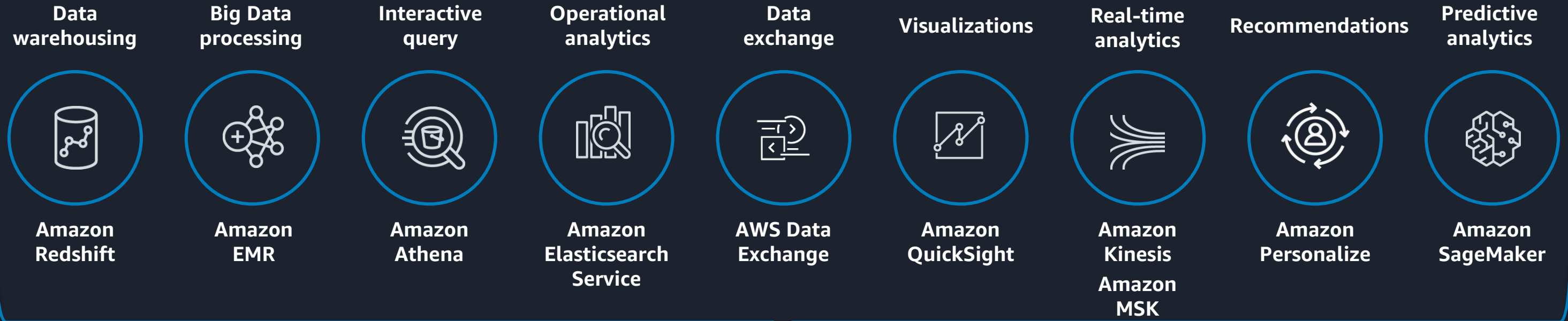
Author & monitor entire flows

Integrated alerting

Gain insights with managed analytics

Any type of analytics on the data lake

Most comprehensive analytics platform



Amazon EMR

Easily run Spark, Hadoop, Hive, Presto, HBase, and more big data apps on AWS

Latest versions



Updated with latest open source frameworks within 30 days

Low cost



50–80% reduction in costs with Amazon EC2 Spot and Reserved Instances

Per-second billing for flexibility

Use Amazon S3 storage



Process data in Amazon S3 securely with high performance using the EMRFS connector

Easy



Fully managed - no cluster setup, node provisioning, cluster tuning

Data warehousing: Amazon Redshift

Data lake & AWS integration



Analyze exabytes of data across data warehouse, data lakes, and operational database

Query data across various analytics services

Best performance, most scalable



3x faster with RA3*

10x faster with AQUA*
*vs other cloud DWs

Adds unlimited compute capacity on-demand to meet unlimited concurrent access

Concurrency scaling

Most secure & compliant



AWS-grade security (eg. AWS Virtual Private Cloud, encryption with AWS KMS, and AWS CloudTrail)

All major certifications such as SOC, PCI, DSS, ISO, FedRAMP, HIPPA

Cost-effective



Cost-optimized workloads by paying compute and storage separately

1/10th cost of traditional DW at \$1000/TB/year

Up to 75% less than other cloud data warehouses & predictable costs

FINRA increases agility, speed, and cost-savings with AWS



FINRA is a not-for-profit organization authorized by the U.S. Congress to protect investors and ensure market integrity through effective and efficient regulation of broker-dealers.

Challenge

Need to process up to **75 billion events per day** and run complex surveillance queries **over 20+ PB of data** to detect and analyze illegal market activity.

Solution

FINRA uses **Amazon S3 as data lake** and uses AWS Lambda and Amazon EMR for data ingestion and **Amazon EMR and Amazon Redshift for data processing**.

Benefits

FINRA has been able to **increase agility, speed**, and operate at scale. The company estimates it will save \$10 to \$20 million annually.

AWS Data Exchange

Quickly find diverse data in one place



>1,000 data products

>80 data providers including include Dow Jones, Change Healthcare, Foursquare, Dun & Bradstreet, Thomson Reuters, Pitney Bowes, Lexis Nexis, and Deloitte

Easily analyze data



Download or copy data to Amazon S3

Combine, analyze, and model with existing data

Analyze data with Amazon EMR, Amazon Redshift, Amazon Athena, and AWS Glue

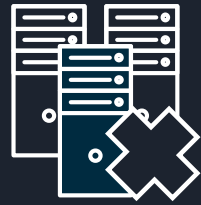
Efficiently access 3rd party data



Simplifies access to data: No need to receive physical media, manage FTP credentials, or integrate with different APIs

Minimize legal reviews and negotiations

Serverless analytics



Zero infrastructure
Zero administration



Never pay for
idle resources



Automatically scales
resources with
usage



Availability and
fault tolerance
built in

Amazon Athena: Interactive query service

Query instantly



Zero setup cost

Point to Amazon S3 and start querying

Pay per query



Pay only for queries run

Save 30–90% on per-query costs through compression

Use Amazon S3 storage



ANSI SQL

JDBC/ODBC drivers

Multiple formats, compression types, and complex joins and data types

Easy



Serverless:

zero infrastructure, zero administration, automatically scales

Challenges querying data from multiple databases



Imagine an e-commerce store
with a microservices architecture

Accessing multiple systems can
be challenging

Federated Query for Amazon Athena

Run SQL queries on data spanning multiple data stores



Run SQL queries on relational, non-relational, object, or custom data sources; in the cloud or on premises

Open-source connectors for common data sources

Build connectors to custom data sources

Run connectors in AWS Lambda: No servers to manage



Governance & security

Most secure infrastructure for analytics

Customers need to have multiple levels of security, identity and access management, encryption, and compliance to secure their data lake



Security

Amazon GuardDuty
AWS Shield
AWS WAF
Amazon Macie
Amazon Virtual Private Cloud (VPC)



Identity

AWS Identity and Access Management (IAM)
AWS Single Sign-On (SSO)
Amazon Cloud Directory
AWS Directory Service
AWS Organizations



Encryption

AWS Certification Manager
AWS Key Management Service (KMS)
Encryption at rest
Encryption in transit
Bring your own keys, HSM support



Compliance

AWS Artifact
Amazon Inspector
Amazon CloudHSM
Amazon Cognito
AWS CloudTrail

Most secure infrastructure: certifications

Global

 **CSA**
Cloud Security Alliance Controls



ISO 9001
Global Quality Standard



ISO 27001
Security Management Controls



ISO 27017
Cloud Specific Controls



ISO 27018
Personal Data Protection



PCI DSS Level 1
Payment Card Standards



SOC 1
Audit Controls Report



SOC 2
Security, Availability, & Confidentiality Report



SOC 3
General Controls Report

United States



CJIS
Criminal Justice Information Services



DoD SRG
DoD Data Processing



FedRAMP
Government Data Standards



FERPA
Educational Privacy Act



ISO FFIEC
Financial Institutions Regulation



FIPS
Government Security Standards



FISMA
Federal Information Security Management



GxP
Quality Guidelines and Regulations



HIPAA
Protected Health Information



ITAR
International Arms Regulations



MPAA
Protected Media Content



NIST
National Institute of Standards and Technology



SEC Rule 17a-4(f)
Financial Data Standards



VPAT/Section 508
Accountability Standards

Asia Pacific



FISC [Japan]
Financial Industry Information Systems



IRAP [Australia]
Australian Security Standards



K-ISMS [Korea]
Korean Information Security



MTCS Tier 3 [Singapore]
Multi-Tier Cloud Security Standard



My Number Act [Japan]
Personal Information Protection

Europe



C5 [Germany]
Operational Security Attestation



Cyber Essentials Plus [UK]
Cyber Threat Protection



G-Cloud [UK]
UK Government Standards




IT-Grundschutz [Germany]
Baseline Protection Methodology




Closing


Business Intelligence & Machine Learning


 **AWS Data Exchange**
Data exchange


 **Amazon QuickSight**
Visualizations

 **Amazon SageMaker**
ML


 **Amazon Comprehend**
NLP


 **Amazon Transcribe**
Speech-to-text


 **Amazon Textract**
Extract text

 **Amazon Personalize**
Recommendation


 **Amazon Forecast**
Forecasts

 **Amazon Translate**
Translation


 **Amazon CodeGuru**
Code reviews

 **Amazon Kendra**
Enterprise search


Analytics


 **Amazon Redshift**
Data warehousing

AQUA


 **Amazon EMR**
Hadoop + Spark

EMR on Outposts


 **Amazon Athena**
Interactive analytics


 **Amazon Elasticsearch Service**
Operational Analytics


UltraWarm


 **Amazon Kinesis Data Analytics**
Real time

Databases

 **Amazon Aurora**
MySQL, PostgreSQL


 **Amazon DynamoDB**
Key value, Document


 **Amazon Neptune**
Graph


 **Amazon RDS**
MySQL, PostgreSQL,
MariaDB, Oracle, SQL Server,
RDS on VMware


RDS Proxy


RDS on Outposts

 **Amazon DocumentDB**
Document

 **Amazon QLDB**
Ledger Database

 **Amazon ElastiCache**
Redis, Memcached

 **Amazon Timestream**
Time Series


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


Blockchain


 **Managed Blockchain**

 **Blockchain Templates**

Data Lake

 **Amazon S3/**
Amazon S3 Glacier

 **AWS Lake Formation**
Data Lakes

 **AWS Glue**
ETL & Data Catalog

Data Movement

AWS Database Migration Service | AWS DataSync | AWS Snowcone | AWS Snowball | AWS Snowmobile | Amazon Kinesis Data Firehose | Amazon Kinesis Data Streams | Amazon Managed Streaming for Kafka

Day 2 - Keynote

1. Analytics today - historical, real-time and predictive analytics
2. Key future trends in analytics
3. Working with distributed data sources

AWS Training and Certification



Training for the Whole Team

Explore tailored Data or Database learning paths for customers and partners



Flexibility to Learn Your Way

Build cloud skills with free digital Data training courses such as "The elements of Data Science", or dive deep with classroom training



Validate Skills with AWS Certification

Demonstrate expertise with a Data industry-recognized credential (Data analytics and Database Specialty AWS Certifications)

aws.amazon.com/training/

Visit the Data, Databases, and Analytics Resource Hub for more resources

Dive deeper with these newly created whitepapers and e-books to help you uncover new insights and value from your data

- An introduction to cloud databases
- Enter the purpose-built database era
- Harness the power of data
- Creating a modern analytics architecture
- The data-driven enterprise
- ... and more!









<https://tinyurl.com/aws-data-databases-analytics>

[Visit resource hub »](#)

Thank you for attending AWS Data, Databases, and Analytics Online Series

We hope you found it interesting! A kind reminder to **complete the survey**.
Let us know what you thought of today's event and how we can improve the event
experience for you in the future.

-  aws-apac-marketing@amazon.com
-  twitter.com/AWSCloud
-  facebook.com/AmazonWebServices
-  youtube.com/user/AmazonWebServices
-  slideshare.net/AmazonWebServices
-  twitch.tv/aws

Thank you!