
AWS
re:Invent
NOV. 27 – DEC. 1, 2023 | LAS VEGAS, NV

ANT321

How Rocket Companies run their data science platform on AWS

Dian Xu
(she/her)
Senior Director, Engineering
Rocket Companies

Ravindra Gupta
(he/him)
Amazon SageMaker
AWS

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

01 Rocket Companies overview and data challenges

02 Data analytics and data science at Rocket Companies

03 Using Amazon EMR and Amazon SageMaker for data science at scale

04 What's next?

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

WE MUST take this great technology to the Internet.”

Dan Gilbert, 1998

Founder and Chairman, Rocket Companies

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

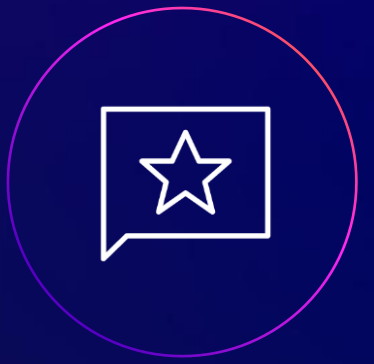
Rocket Companies [NYSE: RKT]

Detroit-based fintech passionate about empowering clients towards homeownership and financial freedom

1+ Million
e-closings completed

21
JD Power Awards

20 Years
on FORTUNE magazine's "100 Best Companies To Work For"



90%+
client retention rate

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

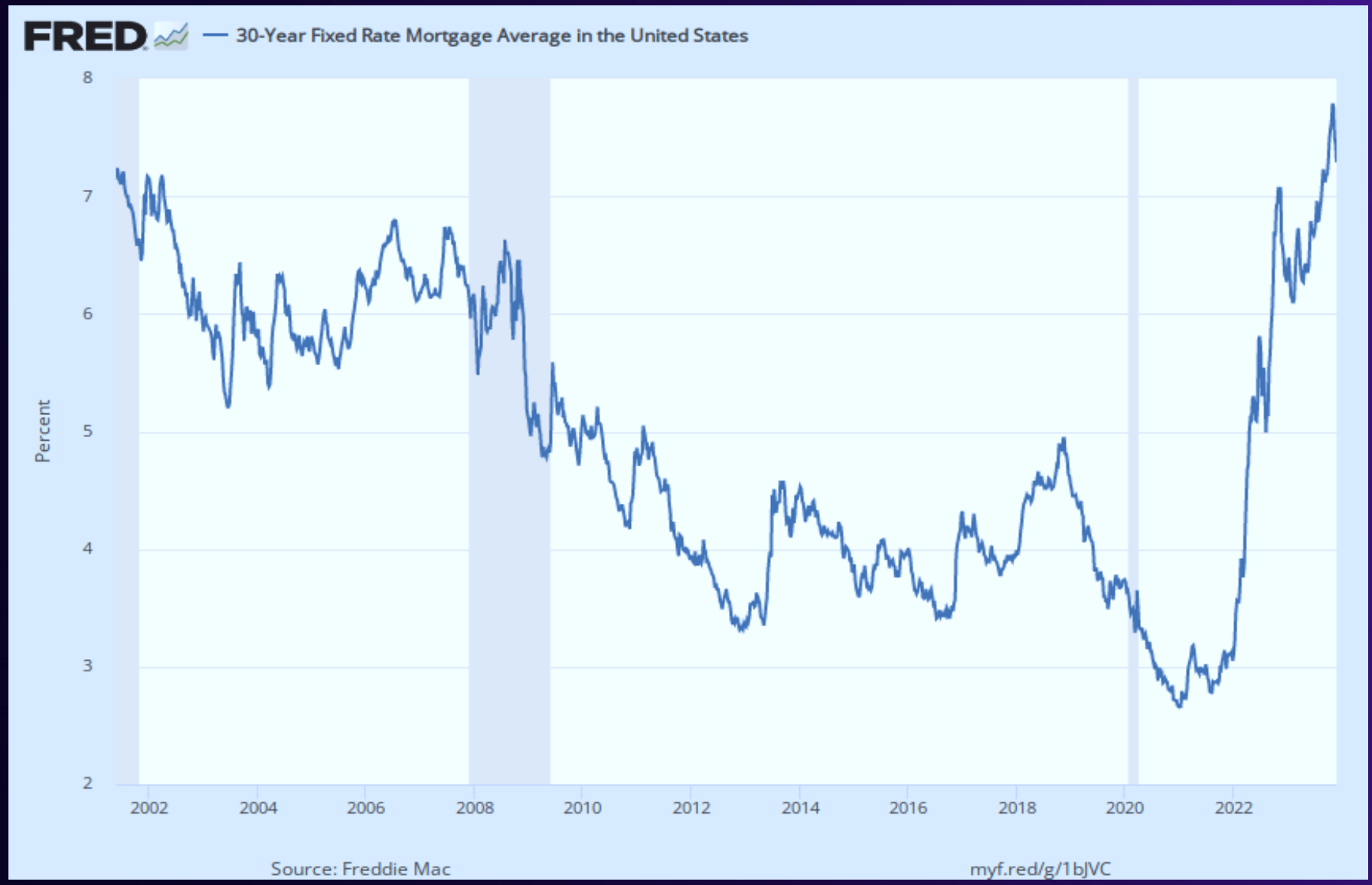
Rocket Companies

Home financing		Personal finance		Home search and sales	
ROCKET Mortgage	Home finance originator and servicer	ROCKET Money	Financial wellness app	ROCKET Homes	Home search and real estate agent network
AMROCK	Title and settlement services	ROCKET Loans	Personal and solar lending		
LENDESK	Software for Canadian mortgage companies				

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Market challenge



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our strategies

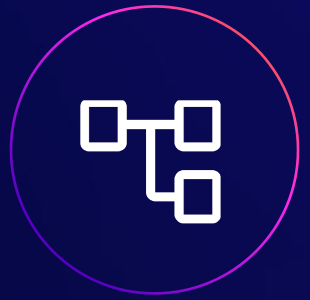
Transforming complex transactions, **driving long-term growth at scale**

Massive fragmented markets

Superior client experience

Technology and data advantage

Trusted, digital-first brand



Multiple avenues of growth

Proven leadership and culture

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data analytics and data science at Rocket

Data-driven decision making

Digital experience personalization

Marketing intelligence

Real-time process automation

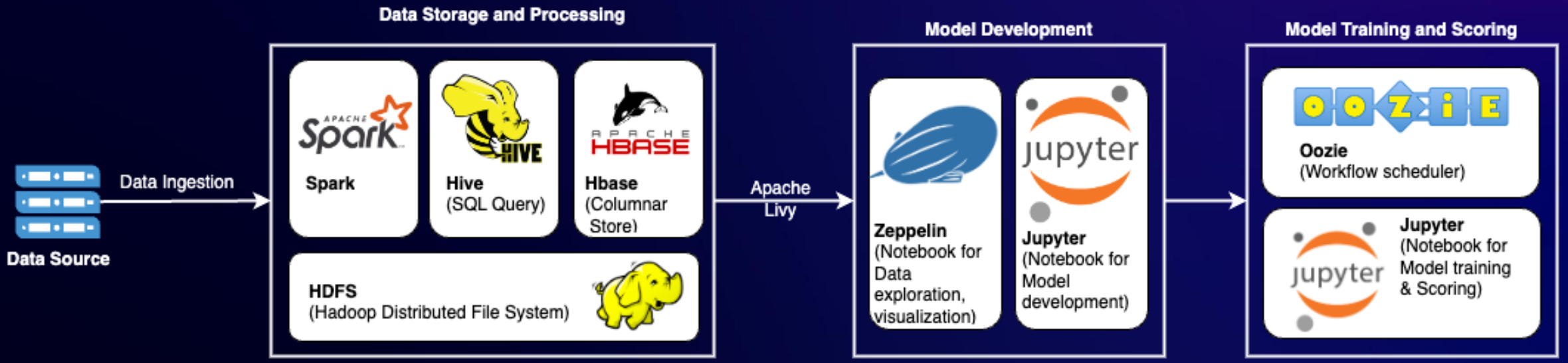
Financial risk management

Regulatory and compliance

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Legacy architecture



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

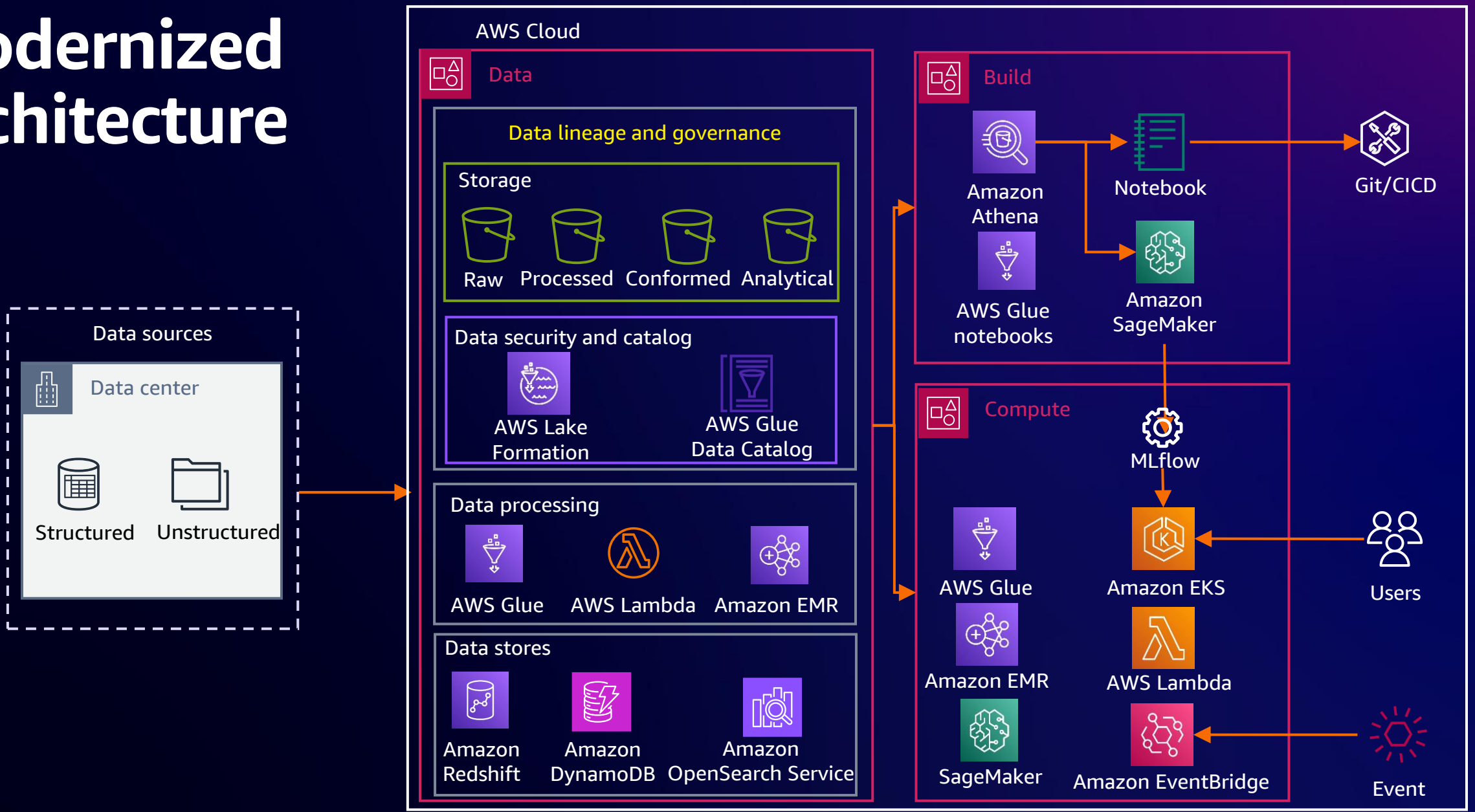
Key outcomes

	Before	After
Speed in delivery	4–8 weeks	<2 hours
Operation sustainability	183 incidents per 18 months	1–2 per 18 months
Data scientist experience	Spent 80% of time waiting on their jobs to run	80% of time in innovation
Data science scalability	Unable to scale	Power 3.7 billion automated data science/AI decisions annually
Cost optimization	$1 million on top of cloud cost	Zero fixed cost $3 million annual cost savings via MAP

aws

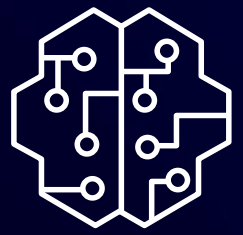
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Modernized architecture

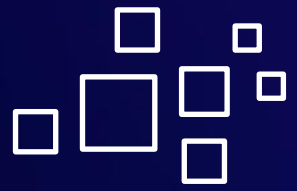


© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data ingestion at scale

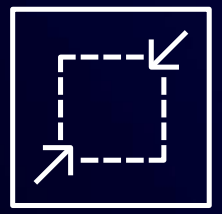


Data volume in
petabytes



Data ingestion from
over 150 sources

Variety in
data sources



Support seasonal
demands of up to
5x volume

Rapid provisioning
of resources for
data science innovation

aws

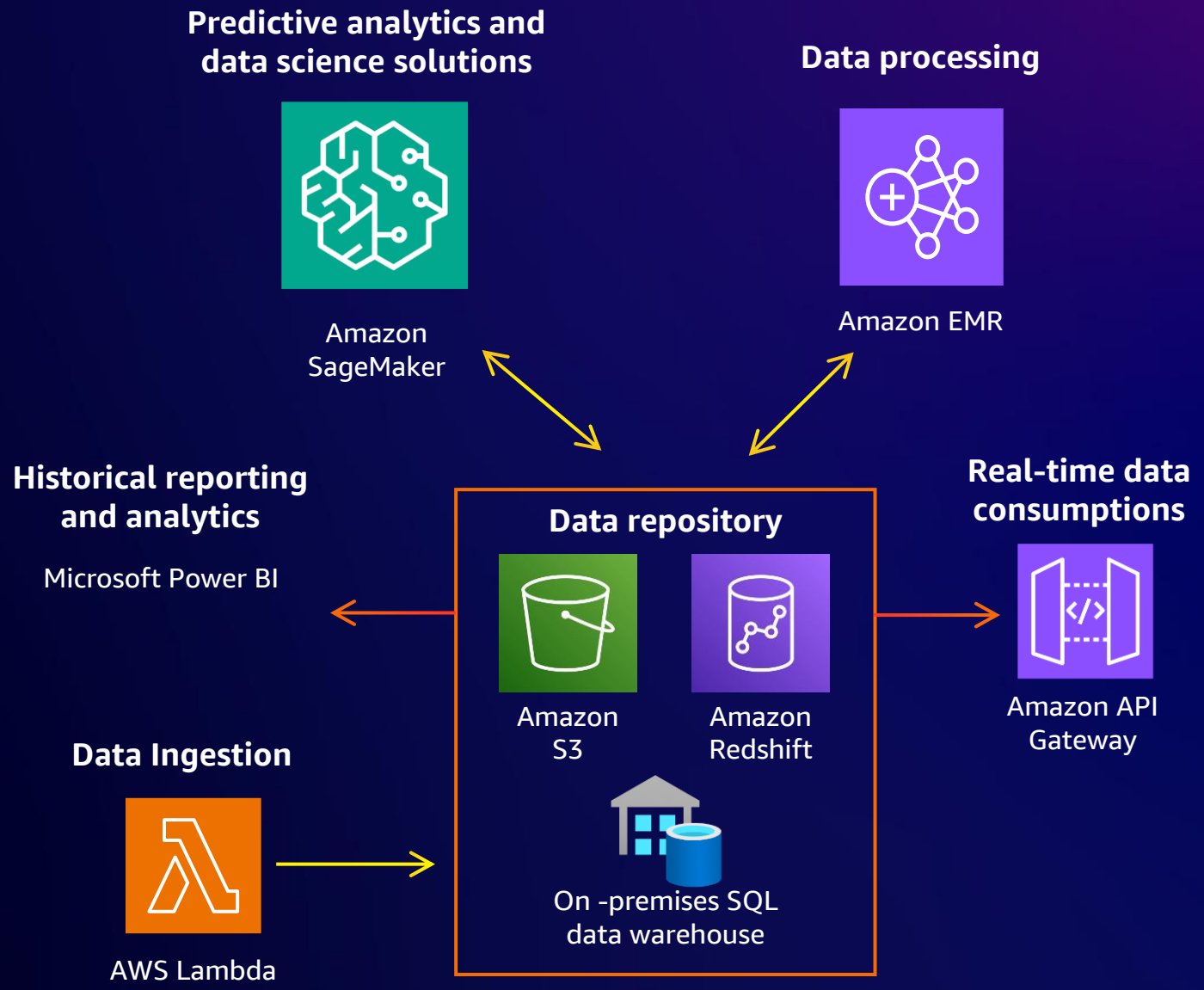
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data lifecycle

Scalable integration
Ingest data from 150+ sources for real-time action and efficient processing

Optimized processing
Use Amazon S3, Amazon EMR, AWS Glue ETL, and Amazon Redshift for cost-friendly, high-capacity data handling

Efficient serving
Athena, API Gateway, and EKS ensure data safety and tiered access



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why Amazon EMR?

Transition motivation

Addressing rapid data growth with efficient analytics

EMR and Spark benefits

Seamless integration with data sources, in-memory processing

Cost and efficiency

Elasticity meets speed; optimized resources with smart scaling

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why Amazon EMR?

Key use cases

Real-time processing, ML, data warehousing, and event handling

Transient EMR

Budget-friendly sporadic workloads using spot instances

Smooth transition

Comprehensive planning, migration, and skill enhancement

aws

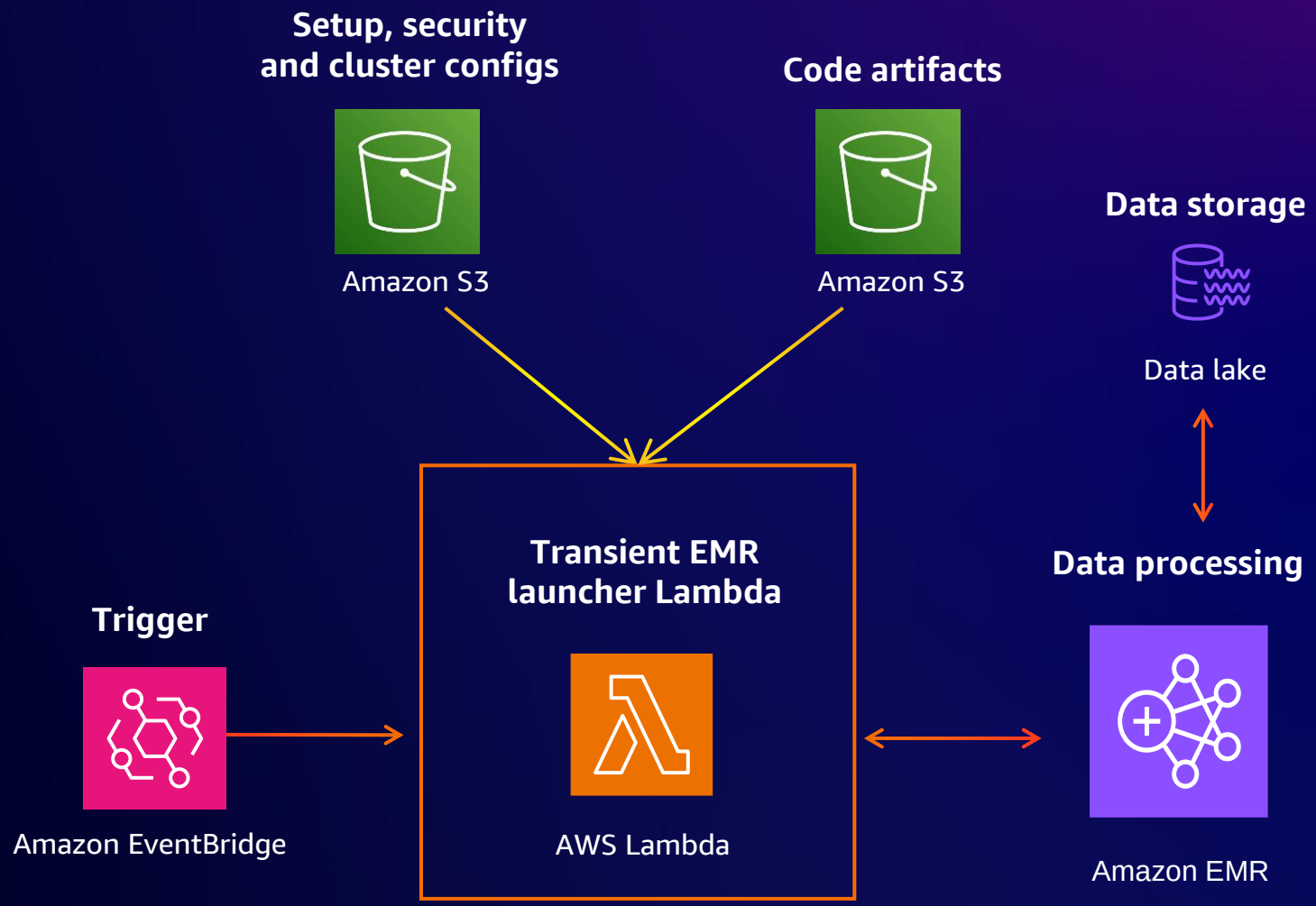
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Transient EMR lifecycle

Configs and artifacts
Store cluster setup, security, job configs, and code in Amazon S3

Trigger
Launch clusters periodically via Amazon EventBridge

Launch
Lambda with boto3 initiates EMR clusters and job steps



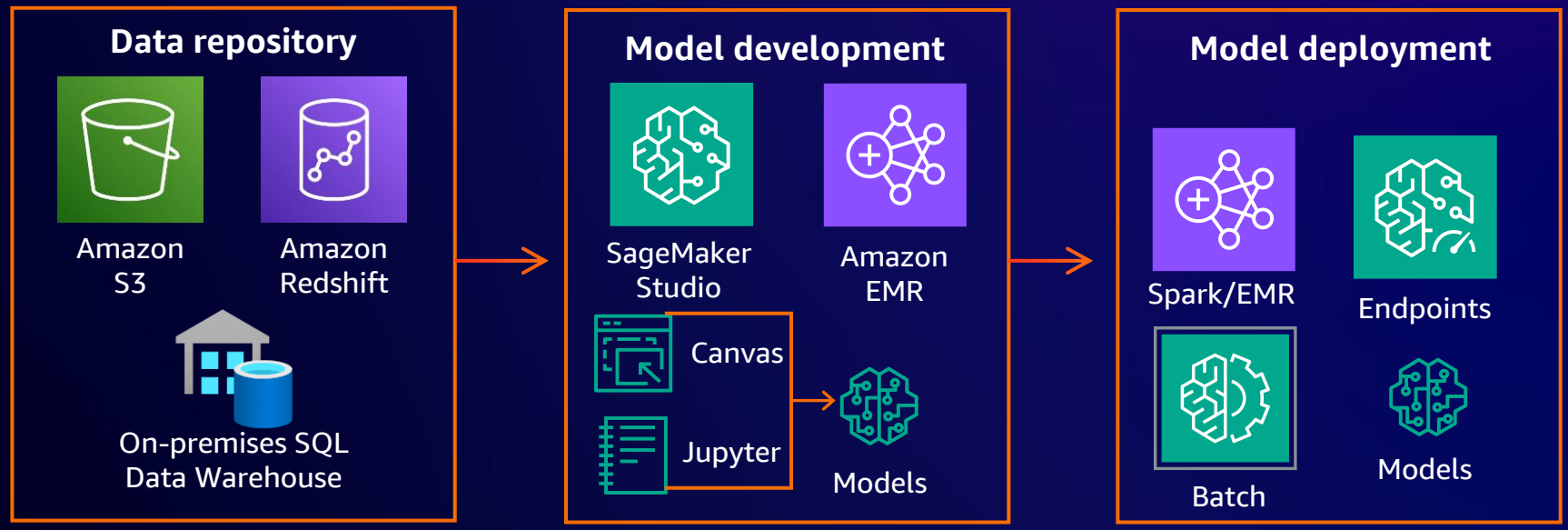
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data science life cycle

Data lake integration
Seamless access to high quality data in the lake

Flexible development
Range of choices from low-code/no-code to the latest frameworks and environments

Flexible deployment
Results deployable into multiple hosting environments, fully enabled for MLOps



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why Amazon SageMaker?

Enterprise-level security

Provides robust protections while ensuring data traffic remains within the AWS backbone

Backward compatibility

Seamless experience for Spark developers, minimizing migration effort

Effortless integration

Native integrations with the rest of the AWS offerings and services

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why Amazon SageMaker?

User-friendly interface

Single-pane management for diverse data science needs

Cost-effective

Customizable, usage-based environments for optimized costs

Empower innovation

Low-code/no-code tools like SageMaker Canvas and SageMaker Data Wrangler for all skill levels

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What's next?

Amazon DataZone for governance
Manage model artifacts and features for enterprise data science governance

Generative AI
SageMaker JumpStart and Amazon Bedrock

Unleash enterprise data science
Further empower enterprise to leverage the power of data science, AI, and ML in digital business transformation

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Generative AI use cases

Data governance acceleration

Harnessing generative AI to optimize metadata management, ensuring 1 million data elements across our vast infrastructure are compliant, understandable, and customer focused

Support engineer assistance

Develop an automated agent powered by Amazon Bedrock to swiftly diagnose and rectify issues, enhancing support efficiency and potentially slashing research and resolution time by up to 80%

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Unlock insights, empower decisions

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker helps organizations harness ML

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data Scientist using Amazon SageMaker Studio Notebook + EMR

Discover, connect to, create and terminate EMR clusters (Hive, Spark and Presto)

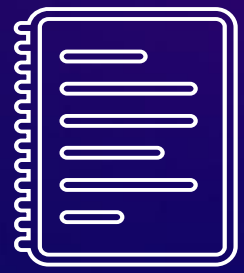
Collaborate using Scala-based Spark and PySpark notebook kernels

Interactive analysis and processing jobs (PyHive, Spark on EMR and Local)

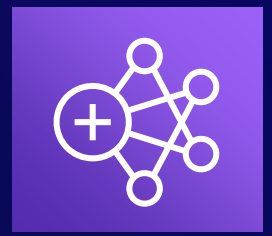
Bring your own image and customize notebook lifecycle configuration

Enforce fine-grained data access

Automate EMR, AWS Glue, and ML pipelines in production



SageMaker Studio Notebook



Amazon EMR

SageMaker

Lake Formation (governance)

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

EMR Serverless enhancements – 2023

Graviton2 support

Custom image support

Job level cost visibility

AWS CloudWatch logging support

AWS Secrets Manager integration

NEW!

AWS Step Functions optimized integration

NEW!

EMR Studio interactive notebooks support

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker MLOps for ML engineers

Amazon SageMaker MLOps

Streamlines the ML lifecycle

Automate ML workflows to scale model development

Build CI/CD pipelines for ML to accelerate model deployment

Catalog model versions, metadata, metrics, and approvals for traceability and reusability

Track lineage for troubleshooting and compliance

Maintain accuracy of predictions after models are deployed

Enhance governance and security

aws

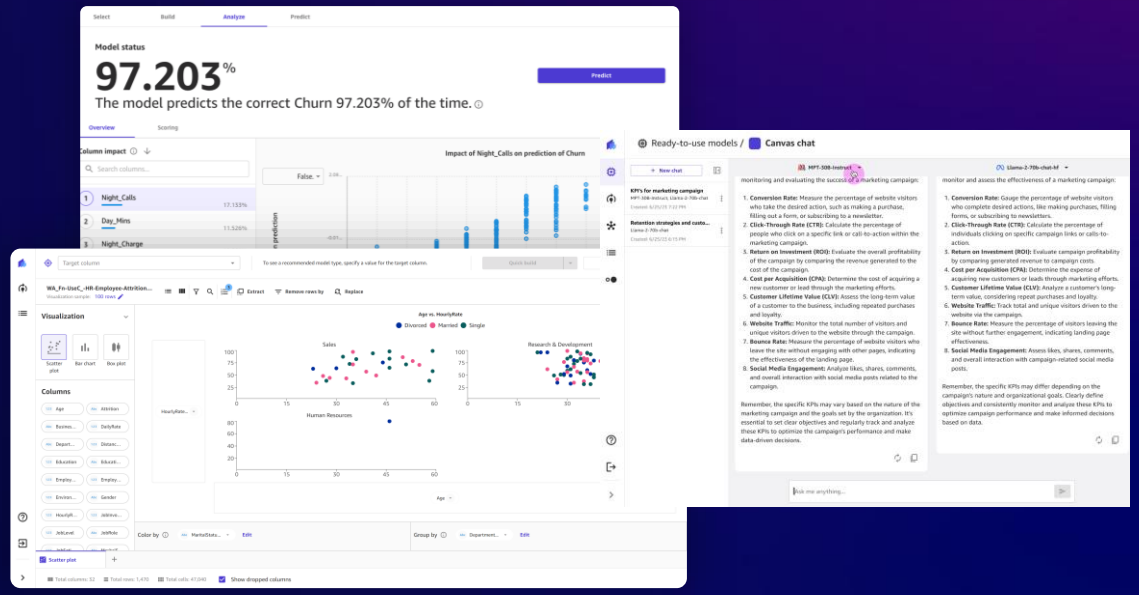
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How business analysts are building ML using SageMaker Canvas

SageMaker Canvas

Personas — Business teams ⟷ Data scientists

ML lifecycle: Prepare data | Train models | Predict outcomes | Automate and deploy

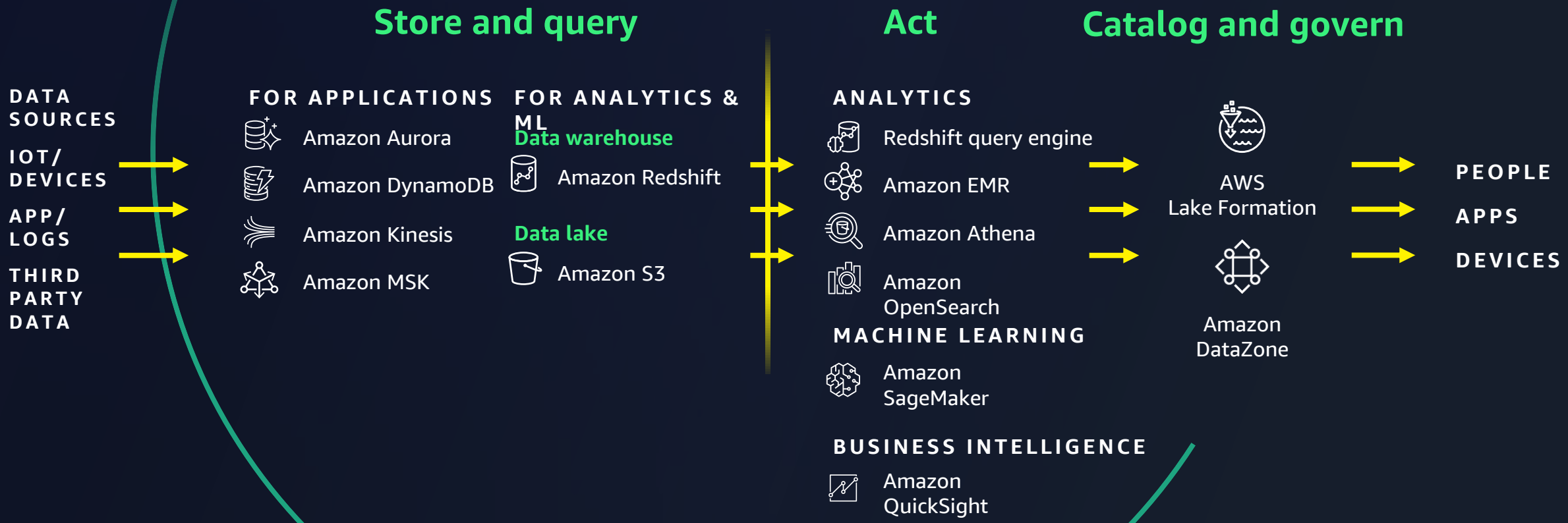


Use cases: Document Q&A | Visual defect detection | Sentiment analysis | Demand forecasting | Customer churn prediction

AWS AI/ML services: Amazon Bedrock | Amazon SageMaker | Amazon Rekognition | Amazon Comprehend | Amazon Textract | Amazon QuickSight

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Building an end-to-end data strategy



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tens of thousands of customers use Amazon SageMaker

aws

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

You've just completed an AWS Analytics Superheroes session

Each hero represents analytics themes that tie back to our sessions. An attendee will select which hero they most align to and scan that hero's QR code to receive a mission agenda outlining which Analytics sessions they should attend based on their interests.

Scan QR code to learn more about the AWS Analytics Superheroes

Thank you!

Please complete the session survey in the mobile app

Dian Xu
DianXu@RocketMortgage.com

Ravindra Gupta
rvidngu@amazon.com

aws
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.