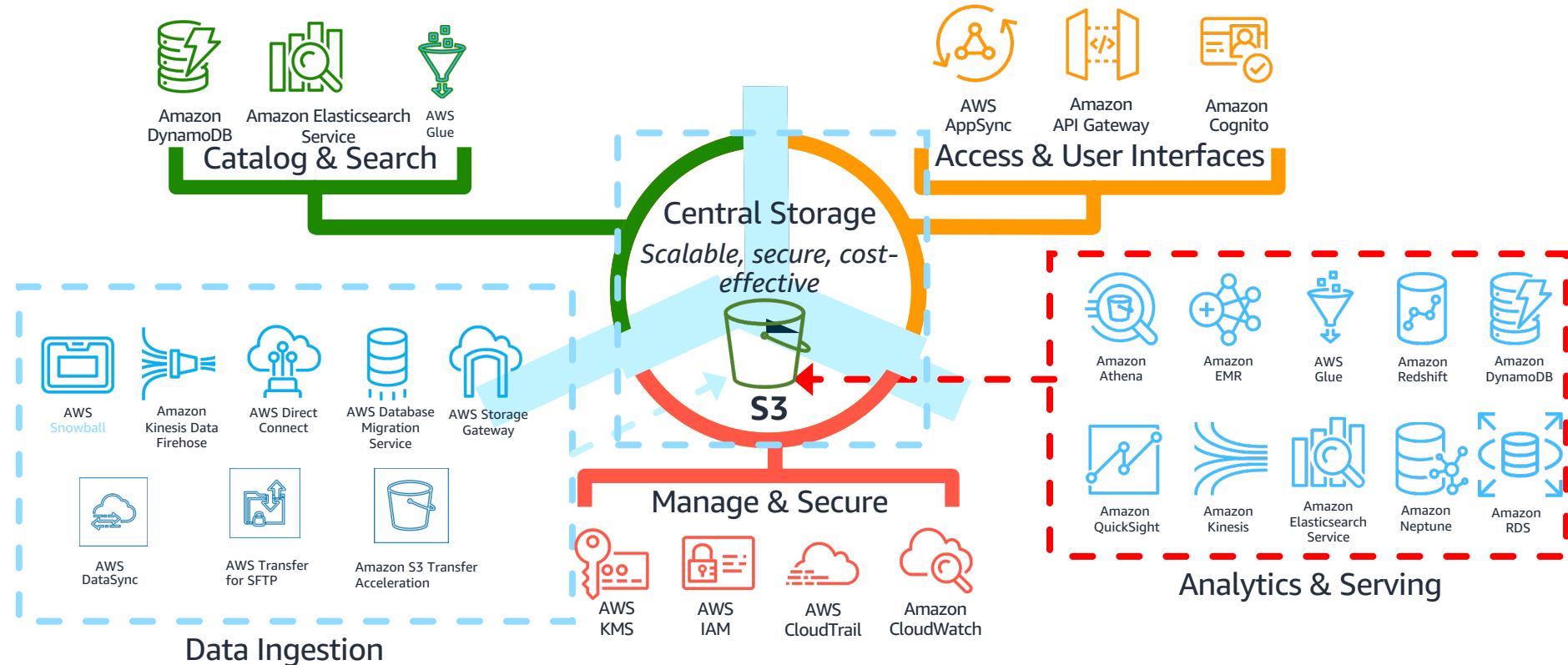




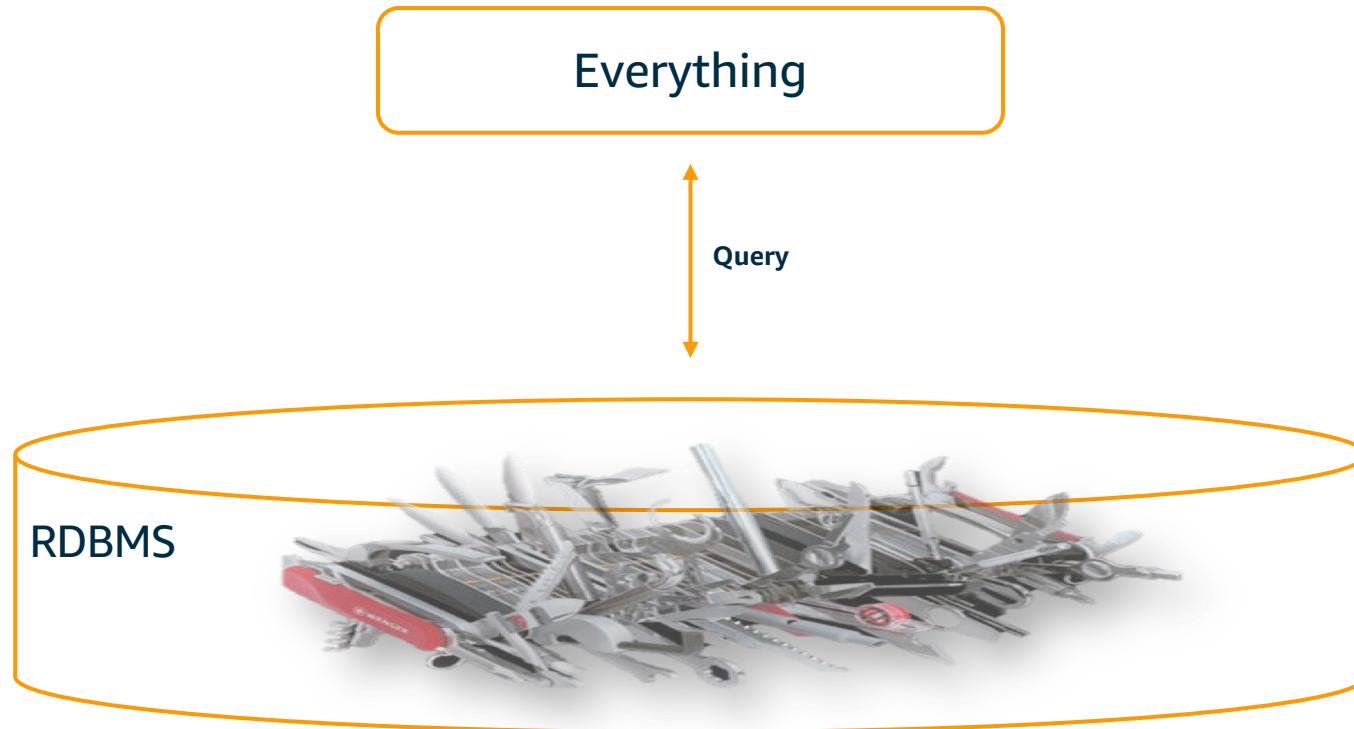
Consuming The DataLake

Data Lake, Reporting, Analytics, Machine Learning

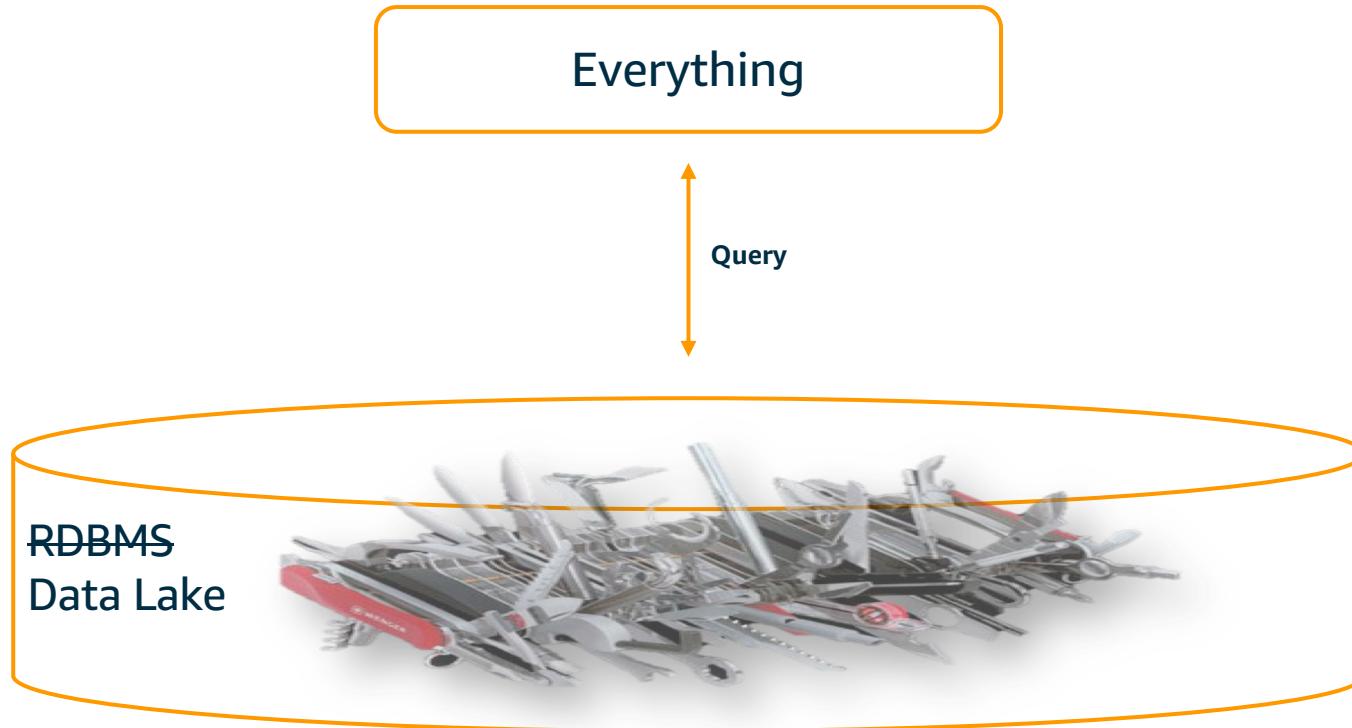
Session's Focus



Anti-Pattern



Also an Anti-Pattern



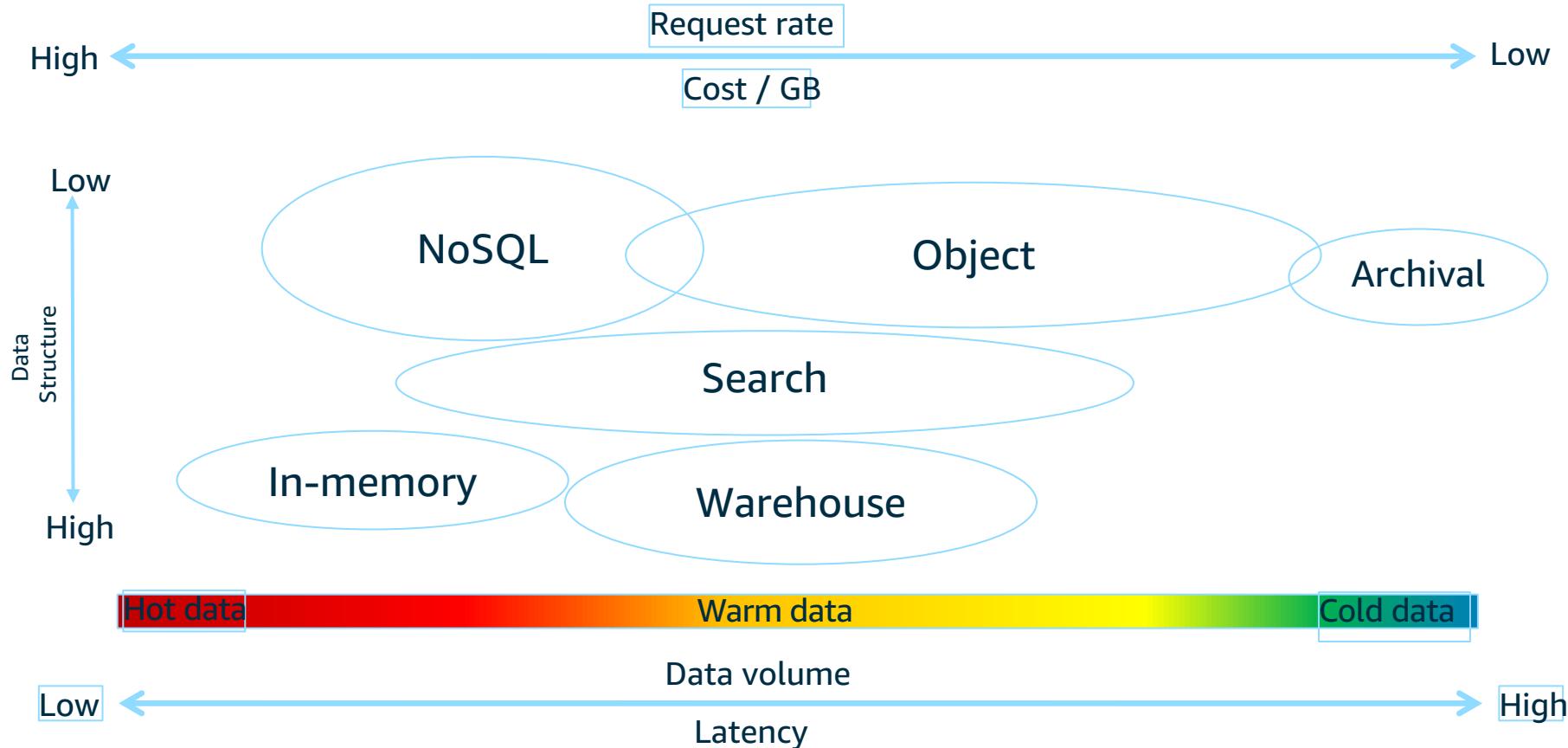


One tool to rule them all

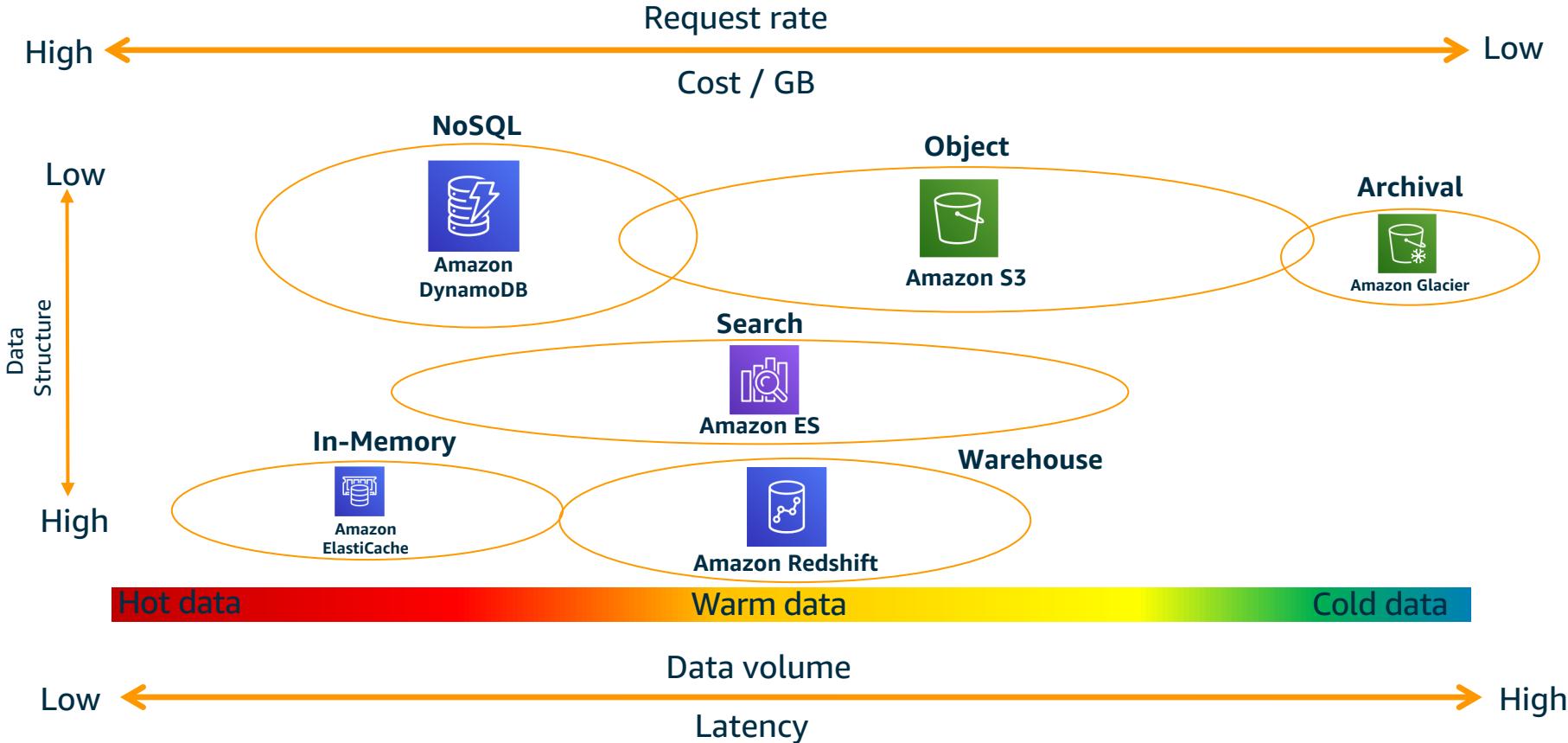
Where do I start?

- Understand your data
 - Data Structure, Access patterns & characteristics, Temperature, Cost, Size
- Know your audience
 - Business Users, Data Scientists, Developers
- Select the right service

Understand your Data



Understand your Data



Who is your audience?

ROLE	PRIORITIES	NEEDS
 Data Scientist	Makes sense of data, generates and communicates insights to improve or create business processes, creates predictive ML models to support them	Ad hoc querying Robust ML tools
 Data Engineer	Builds scalable pipelines, transforms and loads data into structures complete with metadata that can be readily consumed by DS	Ad hoc querying Quick visualization
 Data Product Manager	Manages data as a product. Ensures freshness and consistency of data; understands lineage and compliance needs; treats DS as customers	Reports – data quality, errors
 DevOps Engineer	Monitoring for reliability, quickly diagnose deployment or availability issues	Ad hoc querying Dashboards
 Data Visualizer	Creating engaging visual and narrative journeys for analytical solutions	Visualization Dashboards Reporting
 Business Sponsor	Vetting the prioritization and ROI, funding projects, providing ongoing feedback	Reporting Dashboards

Enabling your Consumers

Dashboards – Reports – Ad-Hoc Analysis – Machine Learning

Dashboards - Near Real-time

Visual Representation of key metrics that change over time

- Data structure - Low
- Usage - Near real-time visualization
- Data temperature - Hot

Available Services:



AWS Lambda



Amazon DynamoDB

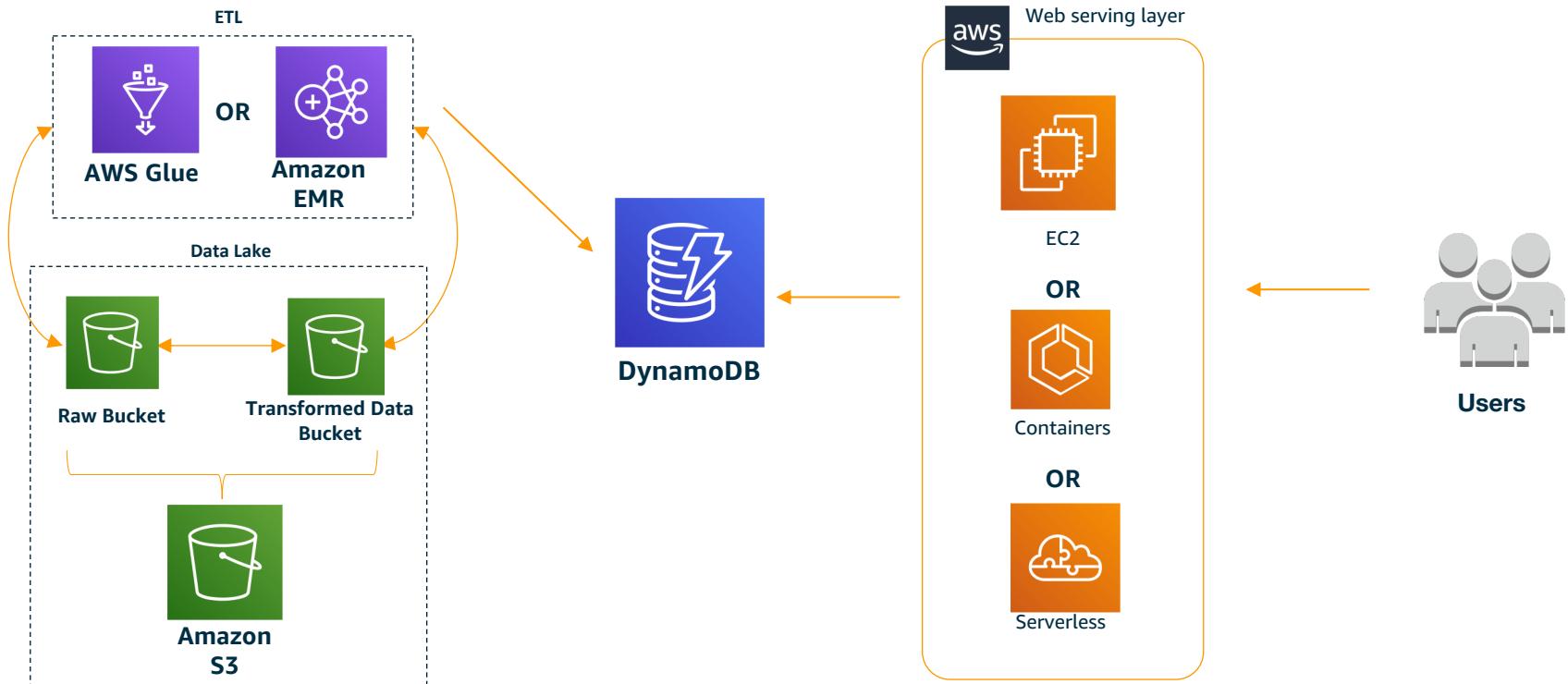


Amazon Kinesis
Data Streams

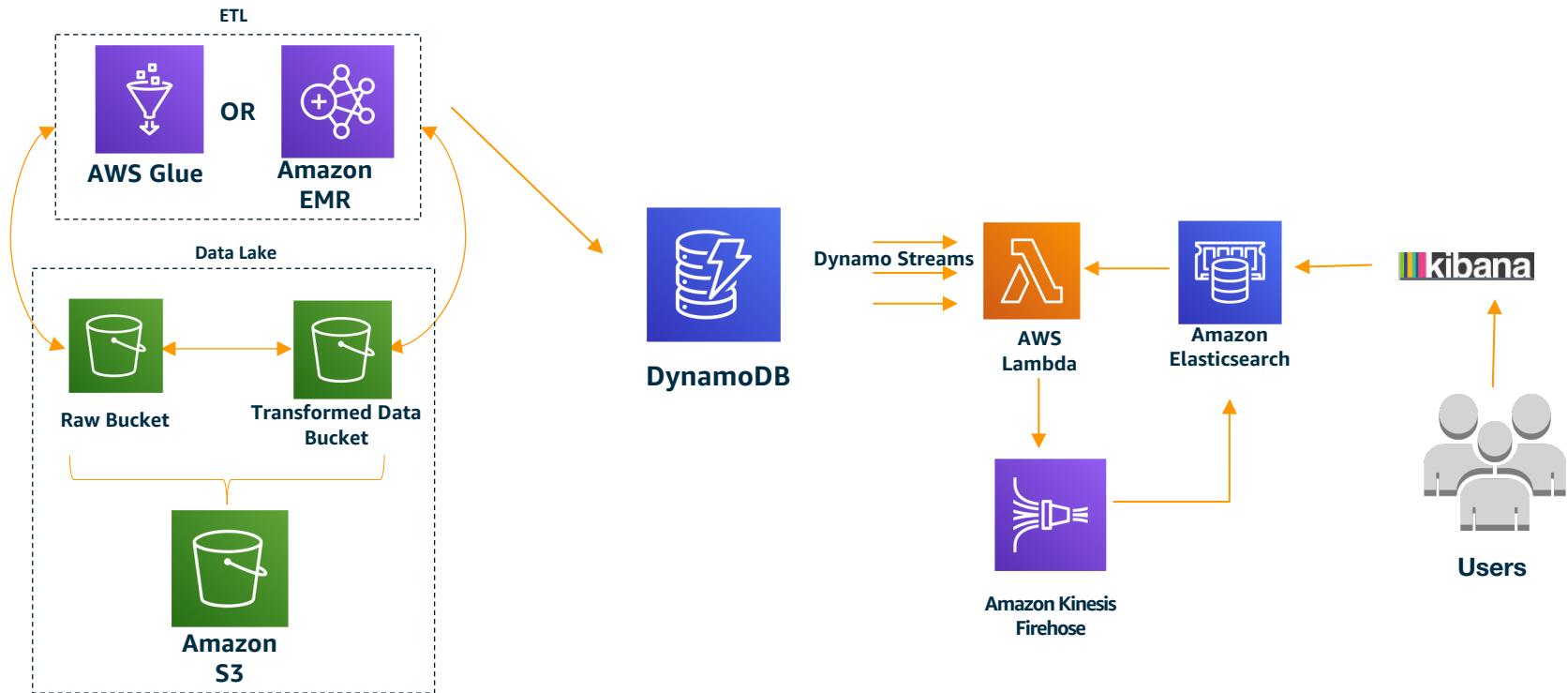


Amazon Elasticsearch
Service

Dashboards – Near Real-time



Dashboards + Search



Ad Hoc Analysis

Information sought on an as-needed basis

- Usage - Dynamic Data Querying
- Data structure - Case based
- Data temperature - Medium - cold

Available Services:



Amazon
ElasticSearch



Amazon Redshift
Spectrum

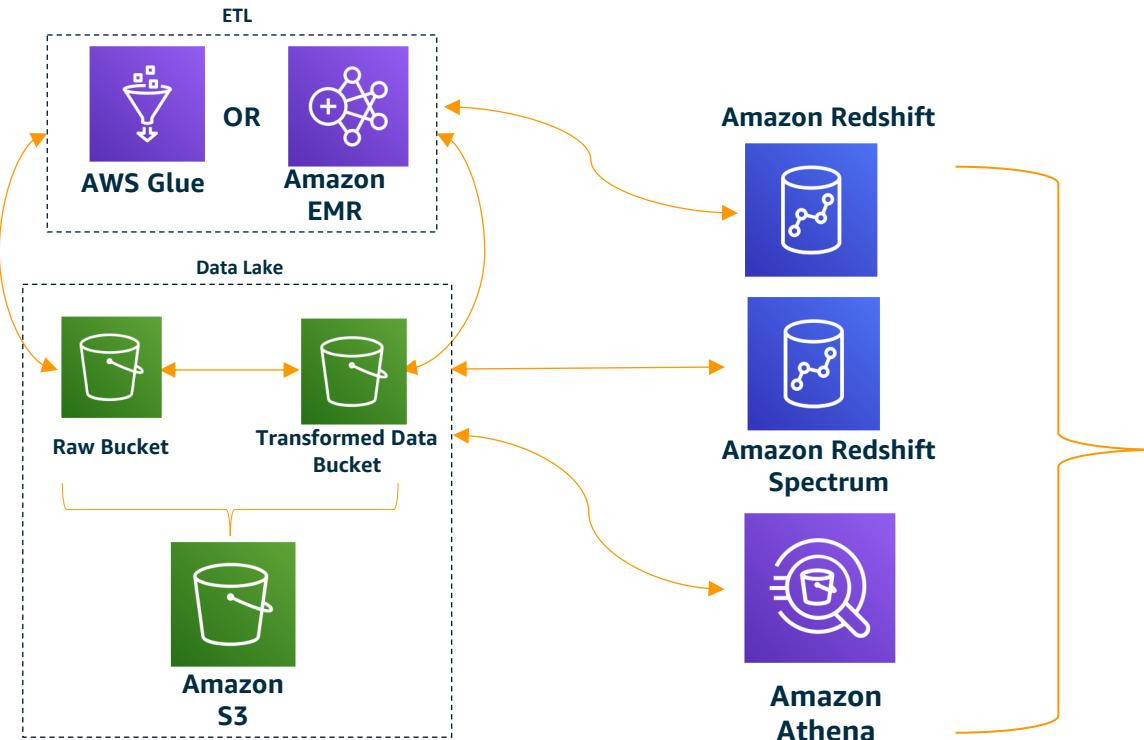


Athena



Amazon
EMR

Reports and Ad-Hoc Analysis



OR



Machine Learning

Data labeled with outcomes to train predication models

- Usage - Machine learning data preparation
- Data structure - Case based
- Data temperature - Medium - cold

Available Services:



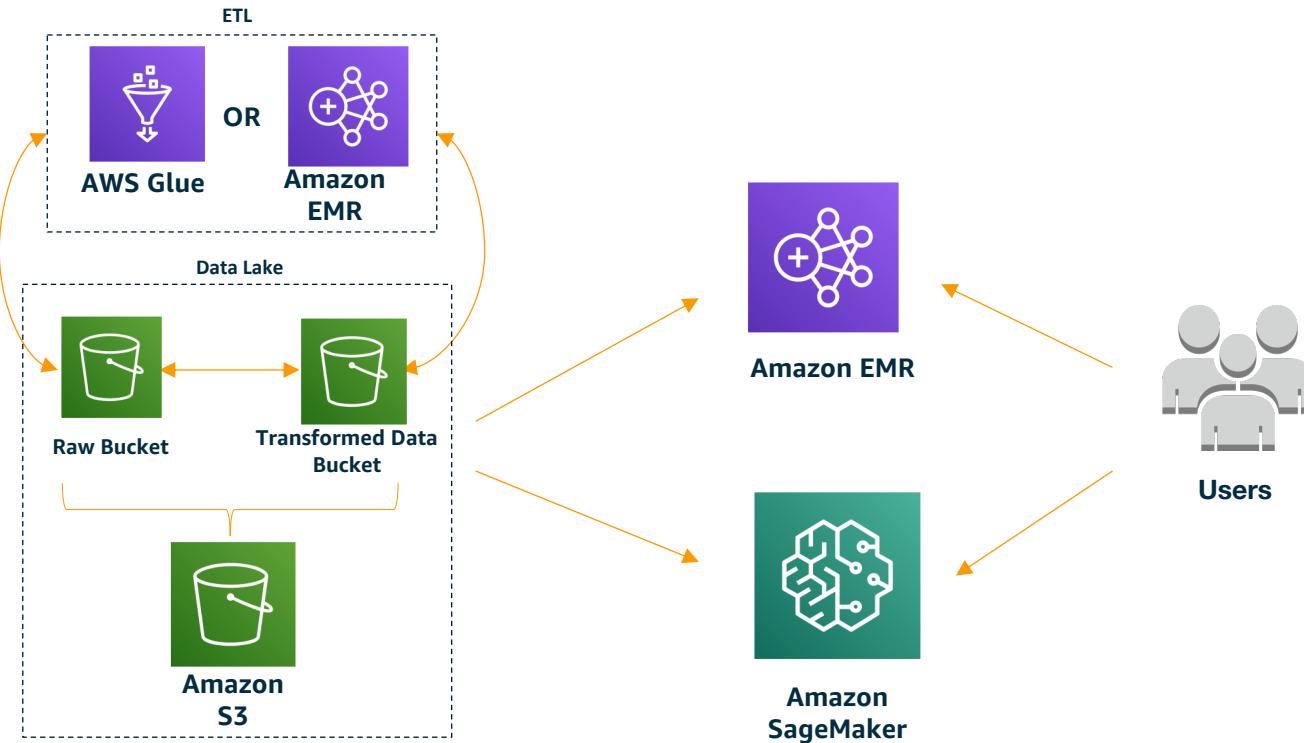
Amazon
EMR



Amazon
SageMaker



Machine Learning



Reports

Static representations of data rendered at a point in time

- Usage - Point in time data extraction
- Data structure - High
- Data temperature – Medium - cold

Available Services:



Amazon Redshift

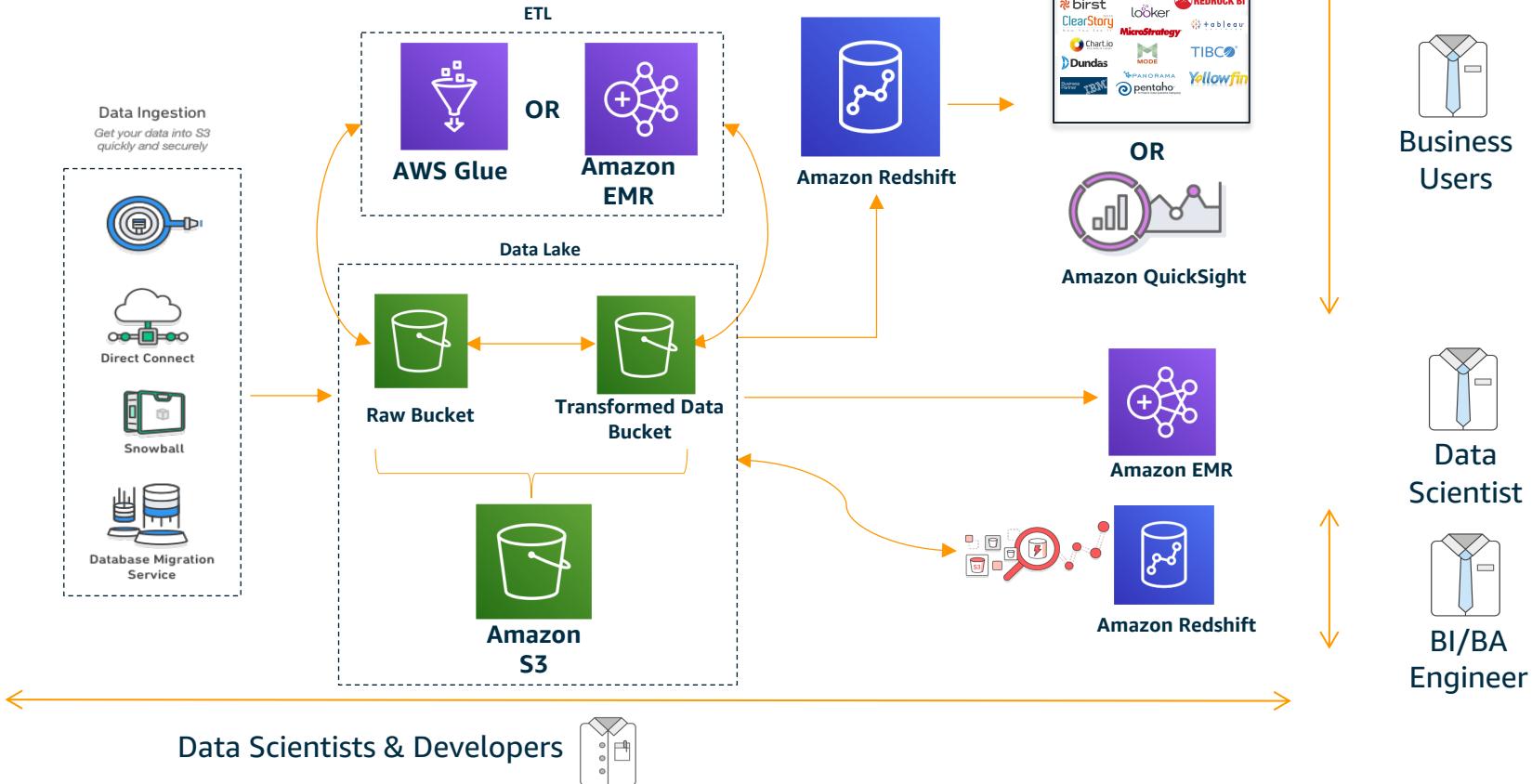


Amazon Athena

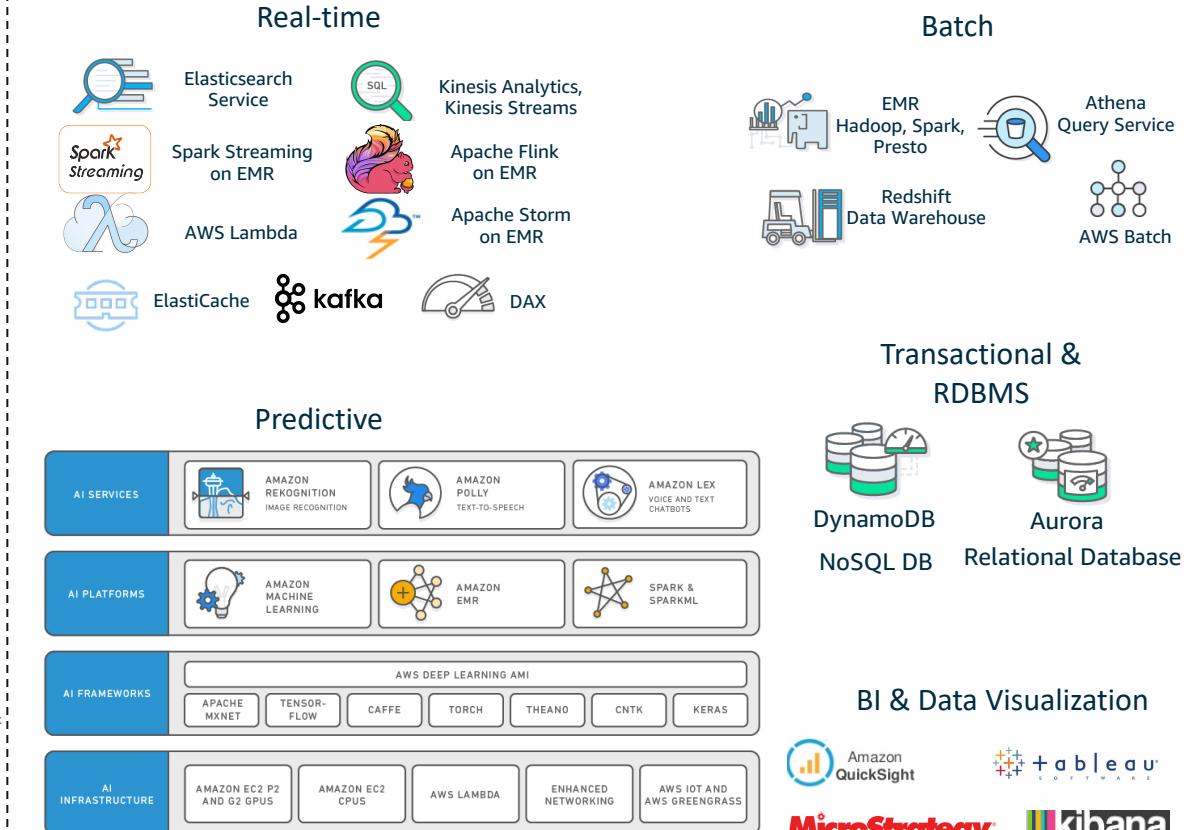
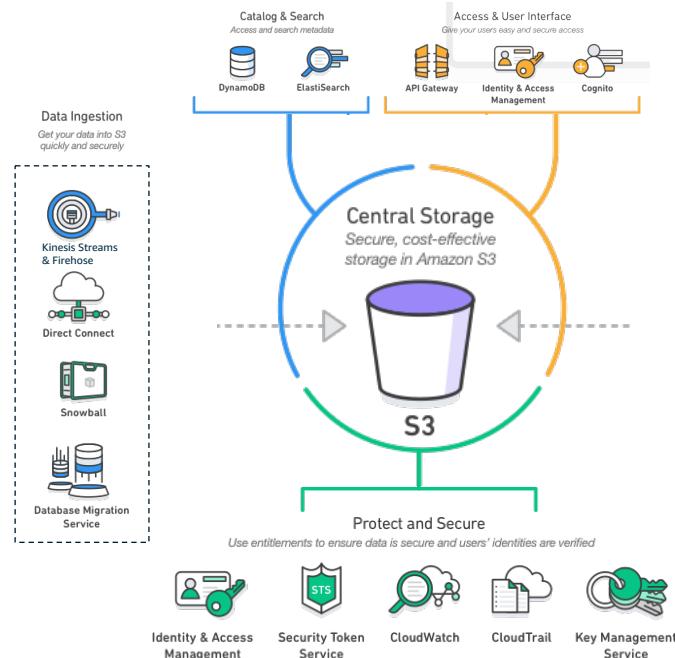


Amazon QuickSight

Report and Data Mart



Processing & Analytics



Services for security and governance

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 VPC



Identity

AWS IAM
AWS SSO
Amazon Cloud Directory
AWS Directory Service
AWS Organizations



Encryption

AWS Certificate Manager
AWS Key Management Service
Encryption at rest
Encryption in transit
Bring your own keys, HSM support



Compliance

AWS Artifact
Amazon Inspector
Amazon CloudHSM
Amazon Cognito
AWS CloudTrail

AWS for Analytics

Visualization & machine learning



Amazon
QuickSight



Amazon
SageMaker



Amazon
Comprehend



Amazon
Lex



Amazon
Polly



Amazon
Rekognition



Amazon
Translate



Amazon
Transcribe



Deep learning
AMIs
+ 10 more

Analytics



Amazon
Redshift



Amazon EMR
(Spark &
Hadoop)



AWS Glue (Spark &
Python)



Amazon
Athena



Amazon
Elasticsearch
Service



Amazon
Kinesis Data
Analytics

Data lake infrastructure & management



Amazon S3 &
Amazon S3 Glacier



AWS Lake
Formation



AWS
Glue

Data movement

[AWS Database Migration Service](#) | [AWS Snowball](#) | [AWS Snowmobile](#) | [Kinesis Data Firehose](#) | [Kinesis Data Streams](#) | [Managed Streaming for Kafka](#)

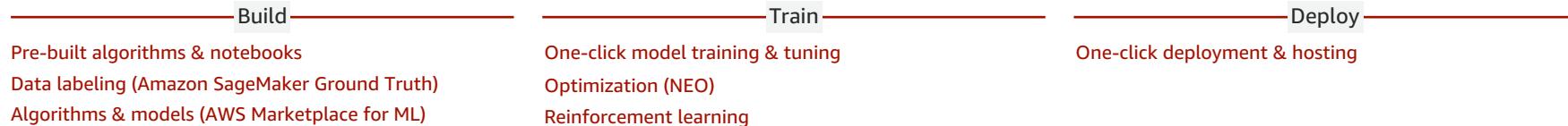
The Amazon ML stack

Broadest & deepest set of capabilities

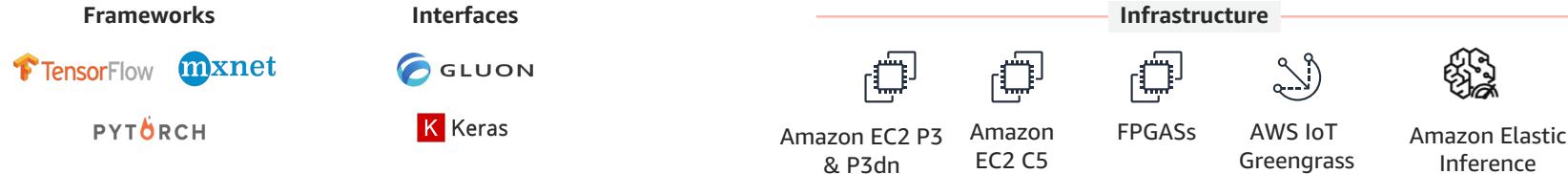
AI services



ML services



ML frameworks & infrastructure



Summary

AWS enables you to build sophisticated big data applications

- Retrospective, Real-time, Predictive

Understand who is the user

- Business user, Data Scientist & Developers

Use the right tool for the job

- Data structure, latency, throughput, access patterns

Leverage AWS managed services

- Scalable/elastic, available, reliable, secure, no/low admin

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



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



HIPPA
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

Thank you!