AI/ML & Data Analytics

Services Introduction

Sebastian Gebski

Principal SA, Startups CEE

Machine learning on AWS: Introduction

AI/ML on AWS
Innovation, choice, and flexibility

100,000+

Customers have used
machine learning on AWS

250+

New capabilities for
machine learning and
artificial intelligence in
just the last 12 months

92% of deep learning in the cloud runs on AWS

91% of cloud based PyTorch runs on AWS

AWS MACHINE LEARNING SOLUTIONS

Reduce training time by 50%
Provide 90% scaling efficiency
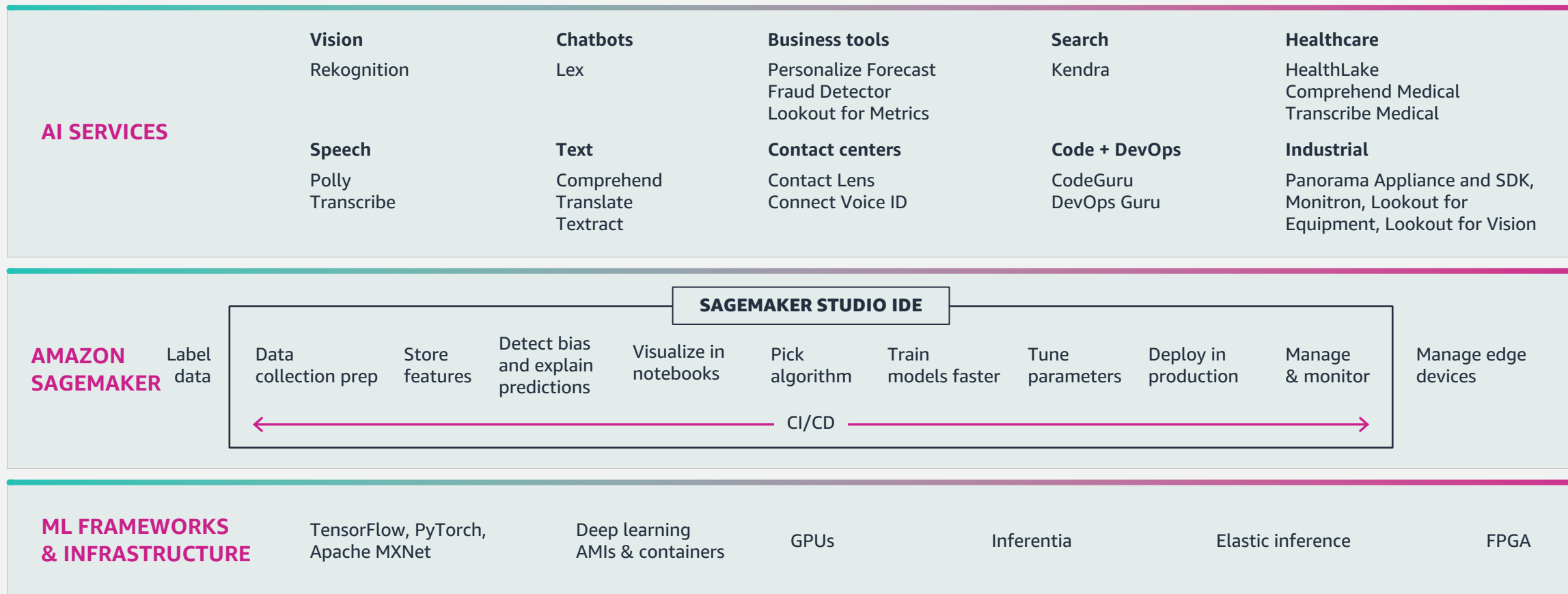Deliver 3x faster network throughput
Improve price and performance by 25%

aws machine learning

More than one hundred thousand customers use AWS for machine learning

AUTODESK · BAYER · Capital One · Cerner · CONDÉ NAST · Domino's · F1 · GE · infor

INTERCOM · INTUIT · INVISTA · Jack in the box · KOCH INDUSTRIES INC · Lenovo · LexisNexis · Liberty Mutual INSURANCE · lyft · mercado libre

move · NASCAR · NFL · nerdwallet · NOVARTIS · NuData Security mastercard · OakNorth · PELOTON · Philips HealthSuite

Propeller · ProQuest · pwc · Roche · Root Insurance · salesforce · sallie mae · Seahawks · SIEMENS · slack · slice · Snapchat

snapdocs · StanleyBlack&Decker · SUBWAY · THOMSON REUTERS · tinder · T · TRANSFIX · twilio · VEOLIA ENVIRONNEMENT · verizon · VW

aws machine learning

The AWS ML stack
Broadest and most complete set of machine learning capabilities

AI SERVICES

Vision
Rekognition

Speech
Polly
Transcribe

Chatbots
Lex

Text
Comprehend
Translate
Textract

Business tools
Personalize Forecast
Fraud Detector
Lookout for Metrics

Contact centers
Contact Lens
Connect Voice ID

Search
Kendra

Code + DevOps
CodeGuru
DevOps Guru

Healthcare
HealthLake
Comprehend Medical
Transcribe Medical

Industrial
Panorama Appliance and SDK,
Monitron, Lookout for
Equipment, Lookout for Vision

AMAZON SAGEMAKER

SAGEMAKER STUDIO IDE

Label data

Data collection prep

Store features

Detect bias and explain predictions

Visualize in notebooks

Pick algorithm

Train models faster

Tune parameters

Deploy in production

Manage & monitor

Manage edge devices

CI/CD

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, PyTorch, Apache MXNet

Deep learning AMIs & containers

GPUs

Inferentia

Elastic inference

FPGA

aws machine learning

The AWS ML stack
Broadest and most complete set of machine learning capabilities

AI SERVICES

Vision
Rekognition

Chatbots
Lex

Business tools
Personalize Forecast
Fraud Detector
Lookout for Metrics

Search
Kendra

Healthcare
HealthLake
Comprehend Medical
Transcribe Medical

Speech
Polly
Transcribe

Text
Comprehend
Translate
Textract

Contact centers
Contact Lens
Connect Voice ID

Code + DevOps
CodeGuru
DevOps Guru

Industrial
Panorama Appliance and SDK,
Monitron, Lookout for
Equipment, Lookout for Vision

AMAZON SAGEMAKER

SAGEMAKER STUDIO IDE

Label data | Data collection prep | Store features | Detect bias and explain predictions | Visualize in notebooks | Pick algorithm | Train models faster | Tune parameters | Deploy in production | Manage & monitor | Manage edge devices

CI/CD

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, PyTorch, Apache MXNet

Deep learning AMIs & containers

GPUs

Inferentia

Elastic inference

FPGA

aws machine learning

Broadest and deepest compute infrastructure for AI/ML
Choice of CPUs, GPUs, and accelerators for your performance and budget needs

Traditional machine learning

Training + inference

Deep learning

Inference

Training

M5	M5a	M5g	C5	C6g	R5	R5a	R6g	R6g		Inf1	G4dn	P3	P4d	DL1

intel®

Cascade Lake CPU
Skylake CPU

Habana® Gaudi® Accelerators

AMD

EPYC CPU

annapurnalabs an amazon company

Graviton CPU
Inferentia Chip

NVIDIA®

A100, V100, T4 GPUs

aws machine learning

Introducing Amazon EC2 P4d Instances

P4d instances

P4d
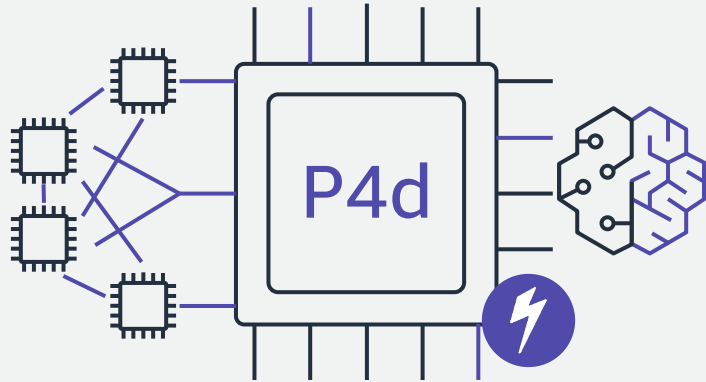
The most powerful GPU instances in the cloud

Up to 60% lower cost to train ML model, 2.5x more deep learning performance and 25% more GPU memory

Powered by 8 NVIDIA A100 GPUs, 400 Gbps of network bandwidth, and capable of 2.5 PetaFLOPS of performance

Deployed in UltraClusters consisting of thousands of tightly coupled GPUs, ideal for ML training and HPC

aws machine learning

Introducing Amazon EC2 DL1 Instances
Best price performance for training deep learning models in the cloud

DL1 instances
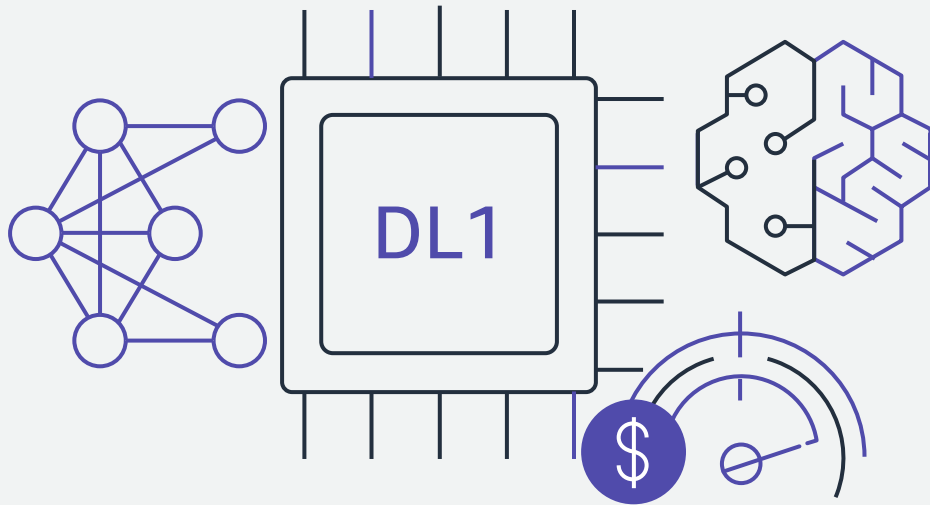
Featuring up to 8 Gaudi accelerators
by Habana Labs (an Intel company)

Specifically built for training deep learning models

Up to 40% better price performance
than latest GPU instances

Custom software seamlessly
integrated with TensorFlow and PyTorch

Get started easily using DLC,
DL AMIs, or Amazon SageMaker

Launch DL1 instances via Amazon ECS
and EKS for containerized ML applications

Inf1 are built from the ground up by AWS
Purpose built for ML inference
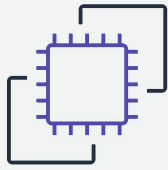
AWS Nitro
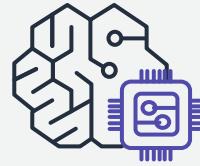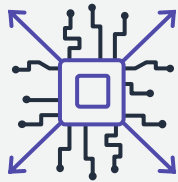
AWS Custom 2ⁿᵈ Gen Intel Xeon Scalable Processors

+

AWS Inferentia
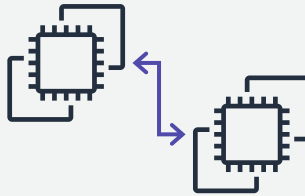
100Gbps Networking

High performance

Low cost

AWS Neuron
High-performance Software Development Kit (SDK)

Neuron compiler

Neuron runtime

Profiling tools
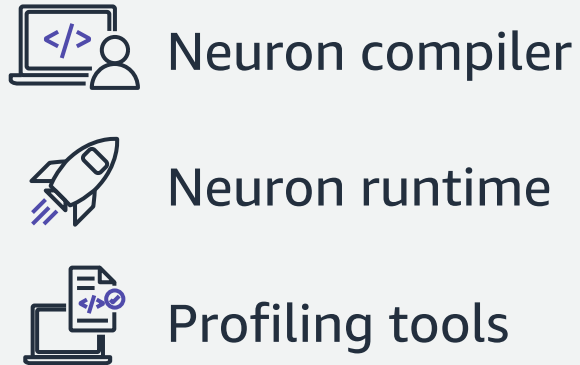
**Easy to get started
Integrated with major frameworks**

TensorFlow 2.0

mxnet

PyTorch

AWS Neuron

Flexibility of choice
Deploy existing models with minimal code changes.
Maintain hardware portability without dependency on AWS software
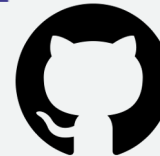
**Documentation,
Examples & Support**

github.com/aws/aws-neuron-sdk

aws machine learning

The AWS ML stack
Broadest and most complete set of machine learning capabilities

AI SERVICES

Vision	**Chatbots**	**Business tools**	**Search**	**Healthcare**
Rekognition	Lex	Personalize Forecast Fraud Detector Lookout for Metrics	Kendra	HealthLake Comprehend Medical Transcribe Medical
Speech	**Text**	**Contact centers**	**Code + DevOps**	**Industrial**
Polly Transcribe	Comprehend Translate Textract	Contact Lens Connect Voice ID	CodeGuru DevOps Guru	Panorama Appliance and SDK, Monitron, Lookout for Equipment, Lookout for Vision

AMAZON SAGEMAKER

SAGEMAKER STUDIO IDE

Label data	Data collection prep	Store features	Detect bias and explain predictions	Visualize in notebooks	Pick algorithm	Train models faster	Tune parameters	Deploy in production	Manage & monitor	Manage edge devices

CI/CD

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, PyTorch, Apache MXNet	Deep learning AMIs & containers	GPUs	Inferentia	Elastic inference	FPGA

aws machine learning

Amazon SageMaker: built to make ML more accessible

Label data

Collect and prepare data

Store features

Check data

Visualize in notebooks

Pick algorithm

Train models

Tune parameters

Deploy in production

Manage and monitor

CI/CD

SageMaker Studio IDE
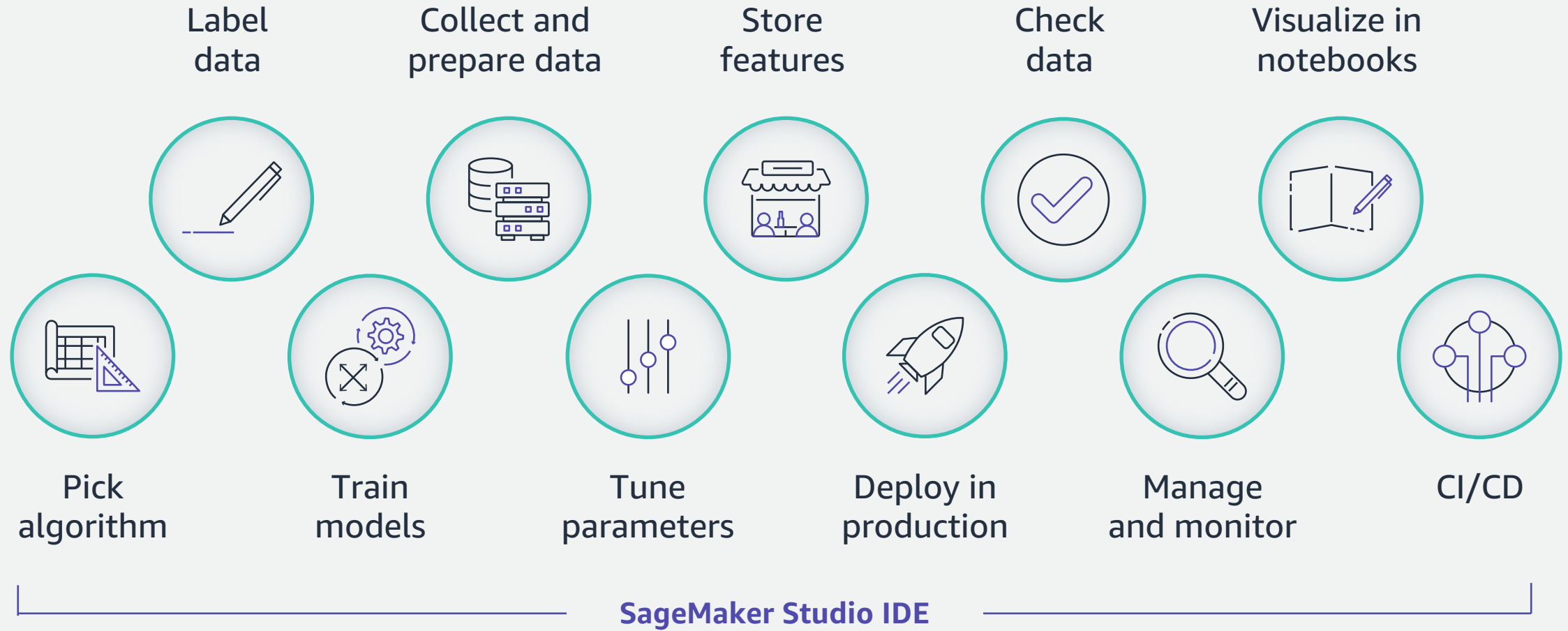
INTEGRATED WORKBENCH

IDE designed specifically for ML, data preparation, experiment management, and pipelines

MANAGED INFRASTRUCTURE

Designed for ultra low latency and high throughput; automatic scaling, and distributed training

MANAGED TOOLING

Purpose-built from the ground up to work together incl. Autopilot, collaboration, notebooks, experiments, debugger, and model monitor

https://aws.amazon.com/sagemaker

Amazon SageMaker

Most complete,
end-to-end ML service
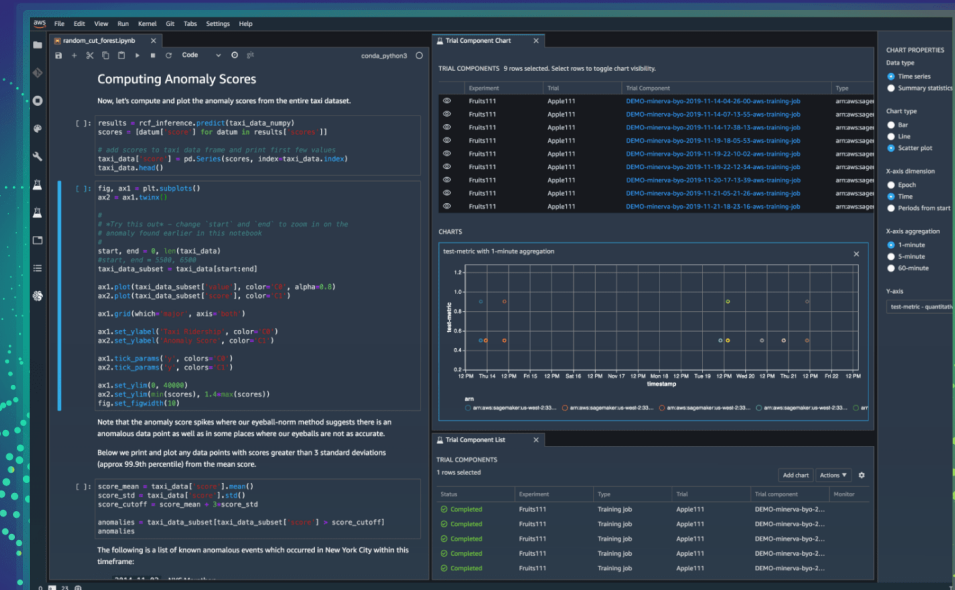
Amazon SageMaker overview

Amazon SageMaker

Prepare

SageMaker Ground Truth
Label training data
for machine learning

SageMaker Data Wrangler
Aggregate and prepare data
for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store
Store, update, retrieve,
and share fetures

Build

SageMaker Studio Notebooks
Jupyter notebooks with
elastic compute and sharing

**Built-in and
Bring-Your-Own Algorithms**
Dozens of optimized algorithms
or bring your own

Local Mode
Test and prototype
on your local machine

SageMaker Autopilot
Automatically create machine
learning models with full visbility

Train & tune

One-click training
Distributed infrastructure
management

SageMaker Experiments
Capture, organize,
and compare every step

Automatic Model Tuning
Hyperparameter optimization

SageMaker Debugger
Debug training runs

Managed Spot Training
Reduce training cost by 90%

Deploy & manage

One-click Deployment
Fully managed, ultra
low latency, high throughput

**Kubernetes & Kuberflow
integration**
Simplify Kubernetes-based machine
learning

Multi-model endpoints
Reduce cost by hosting
multiple models per instance

Model monitor
Maintain accuracy
of deployed models

SageMaker Pipelines
Workflow orchestration
and automation

SageMaker Studio
Integrated development environment (IDE) for ML

aws machine learning

The AWS ML stack
Broadest and most complete set of machine learning capabilities

AI SERVICES

Vision	**Chatbots**	**Business tools**	**Search**	**Healthcare**
Rekognition	Lex	Personalize Forecast Fraud Detector Lookout for Metrics	Kendra	HealthLake Comprehend Medical Transcribe Medical

Speech	**Text**	**Contact centers**	**Code + DevOps**	**Industrial**
Polly Transcribe	Comprehend Translate Textract	Contact Lens Connect Voice ID	CodeGuru DevOps Guru	Panorama Appliance and SDK, Monitron, Lookout for Equipment, Lookout for Vision

AMAZON SAGEMAKER

SAGEMAKER STUDIO IDE

Label data	Data collection prep	Store features	Detect bias and explain predictions	Visualize in notebooks	Pick algorithm	Train models faster	Tune parameters	Deploy in production	Manage & monitor	Manage edge devices

← CI/CD →

ML FRAMEWORKS & INFRASTRUCTURE

TensorFlow, PyTorch, Apache MXNet	Deep learning AMIs & containers	GPUs	Inferentia	Elastic inference	FPGA

aws machine learning

MACHINE LEARNING

Quickly and easily finding accurate information
Amazon Kendra

EASY TO FIND WHAT YOU ARE LOOKING FOR   **QUICK SETUP**   **SECURE**
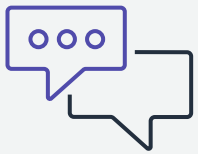
Natural language queries

Contextual search in unstructured content
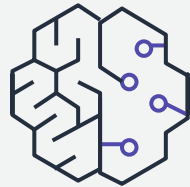
NLU And machine learning core

Reading comprehension

FAQ matching

Document ranking
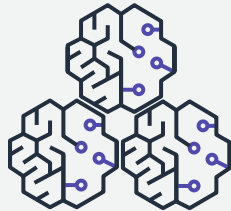
Broad domain expertise

Pre-trained for 14 domains and industries
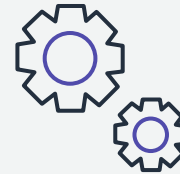
Continuous improvement

Incrementally learns from user feedback

Native & partner connectors

Automate ingestion

Find the best answer wherever it is

Secure search

Encrypted in transit & rest

Token-based access control

aws machine learning

Better insights lead to better customer service with Contact Lens for Amazon Connect

Transcribes
live and
archived calls

Enhanced
search
on call and chat
Transcripts,
sentiment
scores, silence
duration, etc.
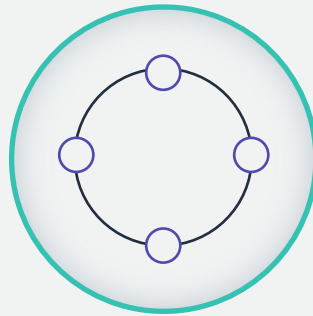
Custom
categorization
to identify
common
call types

Prioritized
list of recurring
issues based on
customer
feedback

Real time
dashboard
and alerting
for supervisors

Provide agents
with answers
to questions
as they are
being asked

Identify fraud faster with Amazon Fraud Detector

Enhance fraud detection with ML

Any level of ML expertise can build ML fraud models

ML boost from Amazon experience, enrichments

Fewer false positives, manual reviews

Fraud staff self-service to address threats faster

Lower TCO, faster TTV

Amazon CodeGuru to build and run high-performing software

WRITE + REVIEW

Built-in code reviews with intelligent recommendations

BUILD + TEST

Detect and optimize the expensive lines of code pre-prod

DEPLOY

MEASURE

Easily identify application inefficiencies in production environment

IMPROVE

aws machine learning

Amazon DevOps Guru
ML-powered cloud operations service to improve application availability
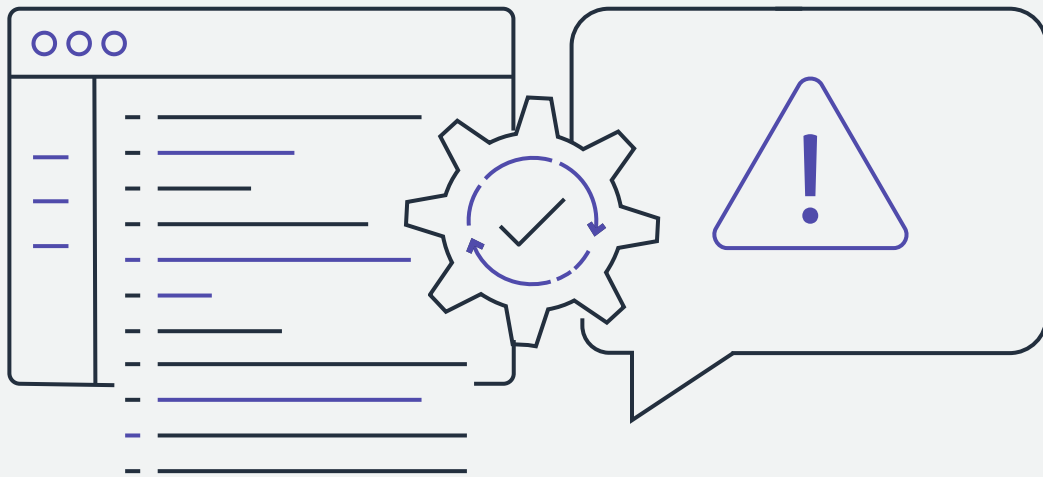
DevOps Guru is an ML-powered service that makes it easy for developers and operators to automatically detect issues to improve application availability and reduce expensive downtime—no machine learning experience required.

Delight your customers and improve customer experience with Amazon Personalize

Deliver
high quality
recommendations

Adapt to changes
in customer intent
in real time
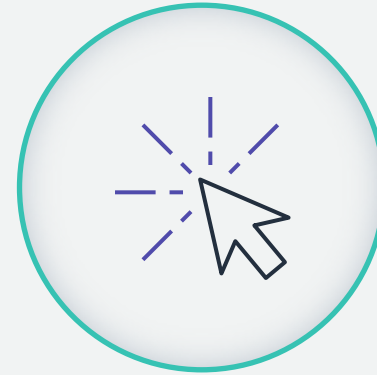
Train a
recommendation
model with a
few clicks

Generate
recommendations
for almost any
product or content

aws machine learning

Industrial AI
AI and ML services for asset-intensive industry use cases

AWS Panorama Appliance

Hardware appliance to add computer vision to existing onsite cameras
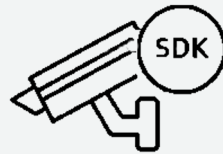
AWS Panorama Device SDK

Build new cameras and devices that run computer vision applications at the edge

Amazon Monitron

End-to-end system for equipment monitoring to detect abnormal machine behavior and enable predictive maintenance

Amazon Lookout for Equipment

Detect abnormal machine behavior using existing industrial sensor data

Amazon Lookout for Vision

Spot defects and anomalies in manufacturing using computer vision

Purpose-built and HIPAA-eligible services
AI and ML services for healthcare & life sciences use cases

Amazon HealthLake

Store, transform, query, and analyze health data in minutes

Amazon Comprehend Medical

Understand medical context with advanced text analytics

Amazon Transcribe Medical

Automatically convert medical speech to text
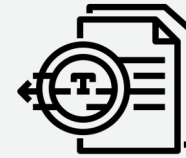
Amazon Textract

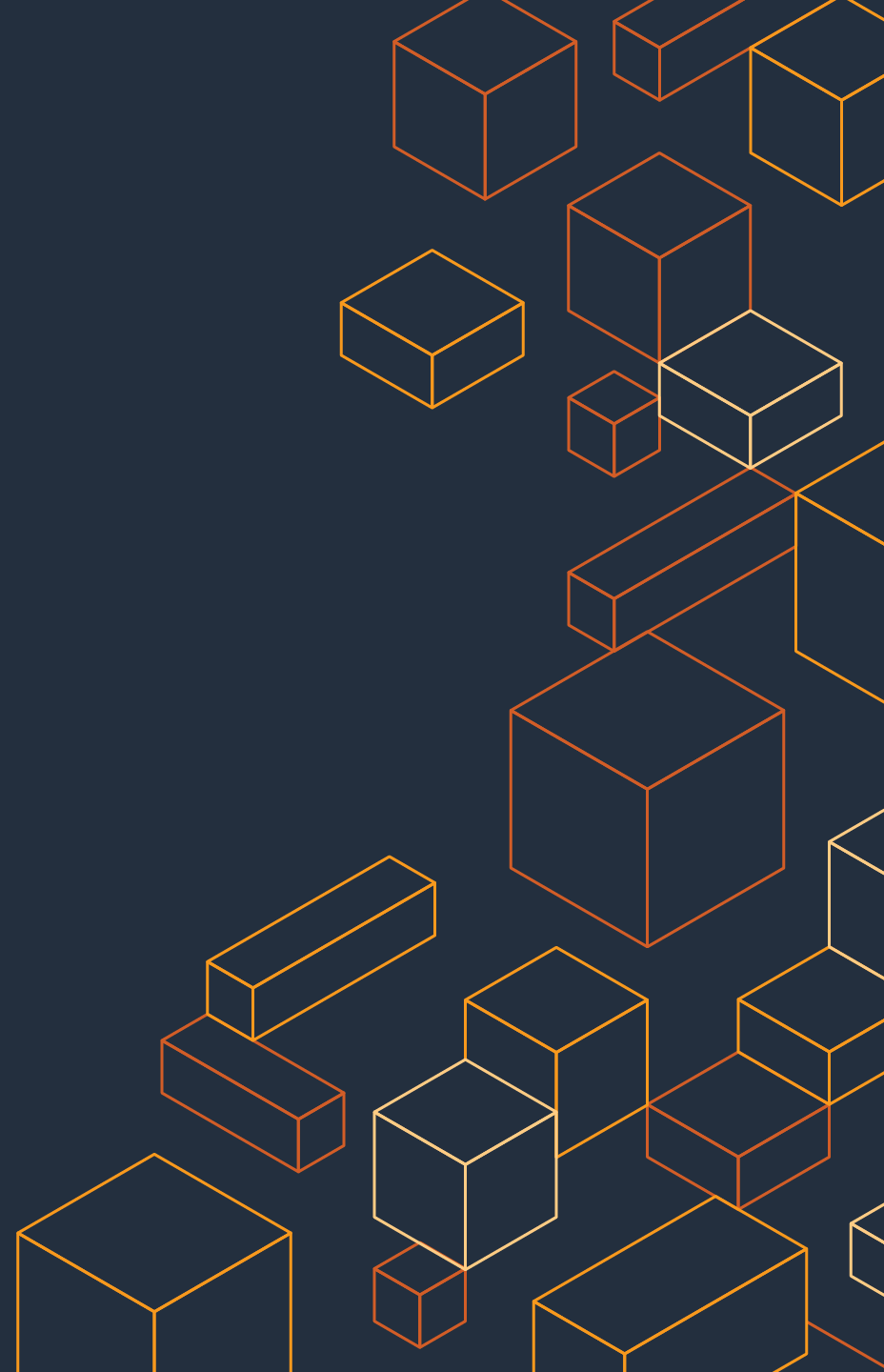Easily extract text and data from virtually any medical document

Amazon Rekognition

Automate medical image and video analysis

aws machine learning

Data Analytics on AWS:

Introduction

Modern Data architecture (expectations)
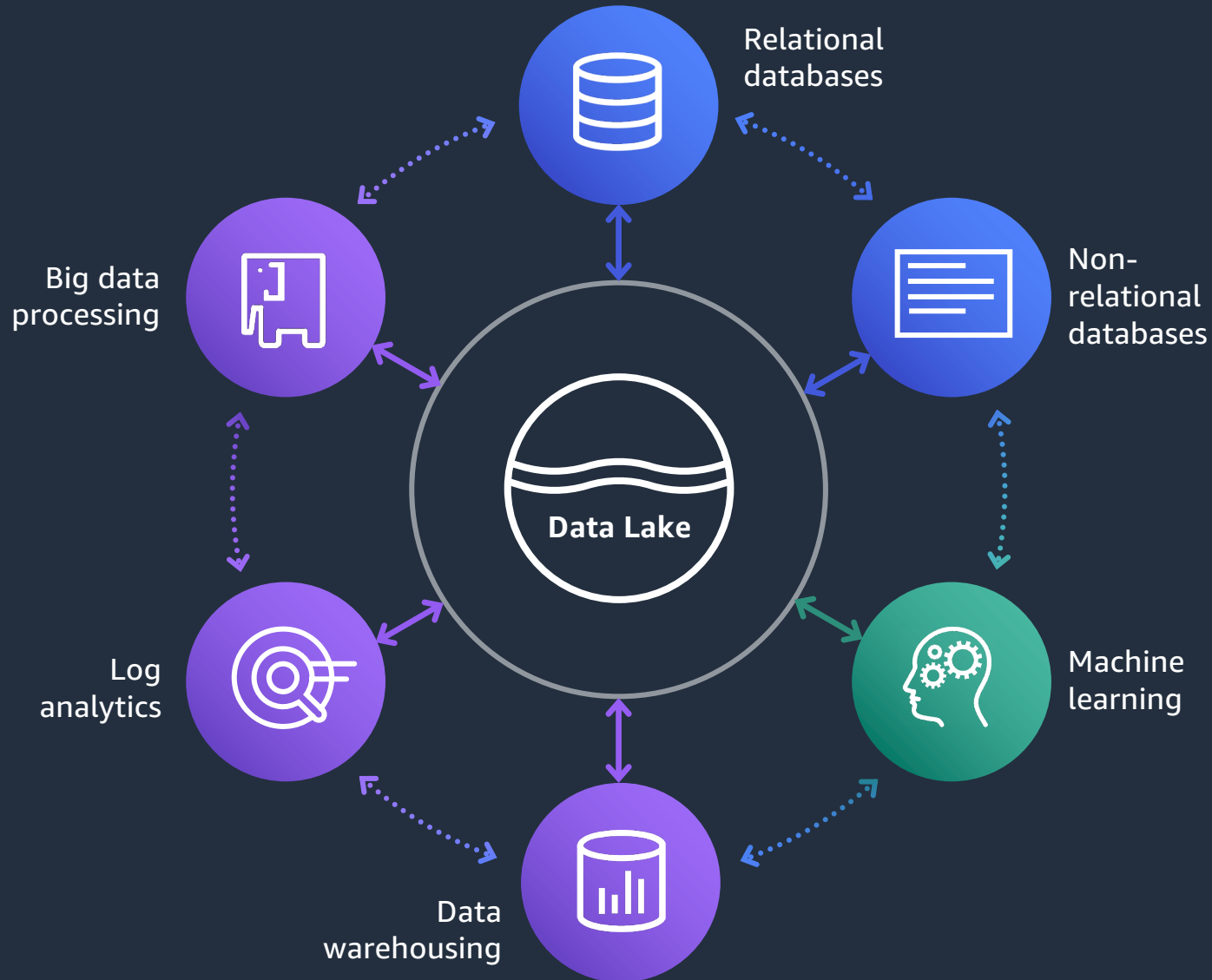
Relational databases

Non-relational databases

Big data processing

Data Lake

Machine learning

Log analytics

Data warehousing

Scalable data lakes

Purpose-built data services

Seamless data movement

Unified governance

Performant and cost-effective

aws

Modern Data architecture on AWS

Amazon
Aurora

Amazon
EMR

Amazon
DynamoDB

AWS LAKE FORMATION

Amazon
Athena

Amazon
S3

AWS GLUE

Amazon
OpenSearch
Service
(successor to
Amazon
Elasticsearch
Service)

Amazon
SageMaker

Amazon
Redshift

Scalable data lakes

Purpose-built
data services

Seamless
data movement

Unified governance

Performant and
cost-effective

© 2021, Amazon Web Services, Inc. or its Affiliates.

aws

Scalable data lakes

Amazon Aurora

Amazon DynamoDB

Amazon EMR

AWS LAKE FORMATION

Amazon Athena

Amazon S3

AWS GLUE

Amazon SageMaker

Amazon OpenSearch Service

Amazon Redshift

aws

Amazon S3 is the most popular choice for data lakes

Unmatched durability, availability, and scalability

Easiest to use with cost optimization: Intelligent tiering

Most ways to get data in

Amazon S3

Most object-level controls

Broadest portfolio of analytics tools

Best security, compliance, and audit capabilities

Cold storage and archive capabilities

aws

More data lakes run on AWS than anywhere else

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

aws

Purpose-built data services

Optimize performance, cost, and scale for your use cases

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

aws

Amazon Athena
Query data in S3 using SQL

Amazon Aurora

Amazon EMR

Amazon DynamoDB

AWS LAKE FORMATION

Amazon Athena

Amazon S3

AWS GLUE

Amazon OpenSearch Service

Amazon SageMaker

Amazon Redshift

Serverless
Quickly query S3 data without managing infrastructure, and pay only for the queries you run

Open and standard
Use ANSI SQL for querying with support for Parquet, CSV, JSON, Avro and other standard data formats

Fast interactive performance
Parallel execution to deliver most results within seconds, with no cluster management required

Cost effective
Pay only for queries run and save 30–90% by compressing, partitioning, and converting your data into columnar formats

aws

Amazon EMR

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

Amazon Aurora

Amazon EMR

Amazon DynamoDB

AWS LAKE FORMATION

Amazon Athena

Amazon S3

AWS GLUE

Amazon OpenSearch Service

Amazon SageMaker

Amazon Redshift

Automate provisioning, configuring, and tuning
Easy setup, management, and monitoring, with latest open-source framework updates within 30 days

Run workloads faster and more cost-effectively
1.7x faster than standard Apache Spark 3.0 at 40% of the cost, and 2.6x faster than open-source Presto 0.238 at 80% of the cost

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

aws

Amazon EMR differentiated performance

1.7x faster performance than standard
Apache Spark 3.0 at 40% of the cost

Up to 2.6x faster performance than open-source
Presto 0.238 at 80% of the cost

11.5% average performance improvement
with Graviton2

25.7% average cost reduction with Graviton2

aws

Amazon OpenSearch Service (successor to Amazon Elasticsearch Service)

Search, visualize, and analyze up to petabytes of text and unstructured data

Amazon Aurora

Amazon EMR

Amazon DynamoDB

AWS LAKE FORMATION

Amazon Athena

Amazon S3

AWS GLUE

Amazon OpenSearch Service

Amazon SageMaker

Amazon Redshift

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.

aws

The OpenSearch Project
An Apache 2.0-licensed search and analytics suite

100% open source

Providing you the freedoms, so you can freely view, use, change, and distribute the code
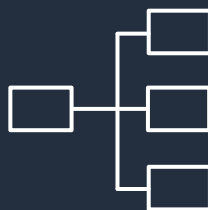
Enterprise-grade

Delivering security and advanced capabilities such as alerting, SQL, and cluster diagnostics

Community-driven

Providing individuals and organizations the freedom to easily contribute changes

aws

Amazon MSK

Fully managed, highly available, and secure Apache Kafka service

Fully compatible
Run your existing Apache Kafka applications on AWS without changes to source code

Fully managed
Focus on creating applications not managing your Apache Kafka environment

Elastic stream processing
Run Apache Flink applications written in SQL, Java, or Scala that elastically scale to process data streams

Highly available
Take advantage of multi-AZ replication within an AWS region

Highly secure
Protect your data with multiple levels of security, including VPC network isolation, encryption at-rest and in-transit, and more

aws

Amazon Kinesis

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

Kinesis Data Analytics
Analyze data streams with serverless Apache Flink or SQL

Kinesis Data Streams
Capture, process, and store data streams

Kinesis Data Firehose
Load data streams into AWS data stores

aws

Amazon Redshift innovates to meet your needs

Analyze all your data
Modern Datawith AWS integration

Amazon Redshift Spectrum + Lake Formation

Data lake export

Federated query

NEW!
Data sharing

NEW!
Amazon Redshift ML

Performance & scale
Fast and self-tuning

Concurrency scaling

RA3 nodes & managed storage

AQUA

Materialized views

NEW!
Automated perf. tuning

Low cost & best value
Predictable costs

On-demand and RIs

NEW!
Cross-AZ cluster recovery

Pause and resume

Cost controls

Built-in security features

Automatic workload manager

aws

Seamless data movement

Amazon Aurora

Amazon EMR

Amazon DynamoDB

AWS LAKE FORMATION

Amazon Athena

Amazon S3

AWS GLUE

Amazon OpenSearch Service

Amazon SageMaker

Amazon Redshift

aws

Seamless data movement

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

Data lake

Federated
query

aws

AWS Glue

Simple, scalable, and serverless data integration

Amazon Aurora
Amazon EMR
Amazon DynamoDB
AWS LAKE FORMATION
Amazon Athena
Amazon S3
AWS GLUE
Amazon OpenSearch Service
Amazon SageMaker
Amazon Redshift

Connect to more sources
Easily ingest data from hundreds of popular data sources

Simplify workflow orchestratation
Easily run and manage thousands of data integration jobs

No servers to manage
Pay only for the resources your jobs consume

Simplify development
Visually develop and manage data integration jobs

aws

AWS Glue Studio

Easily author, run, and monitor AWS Glue ETL jobs
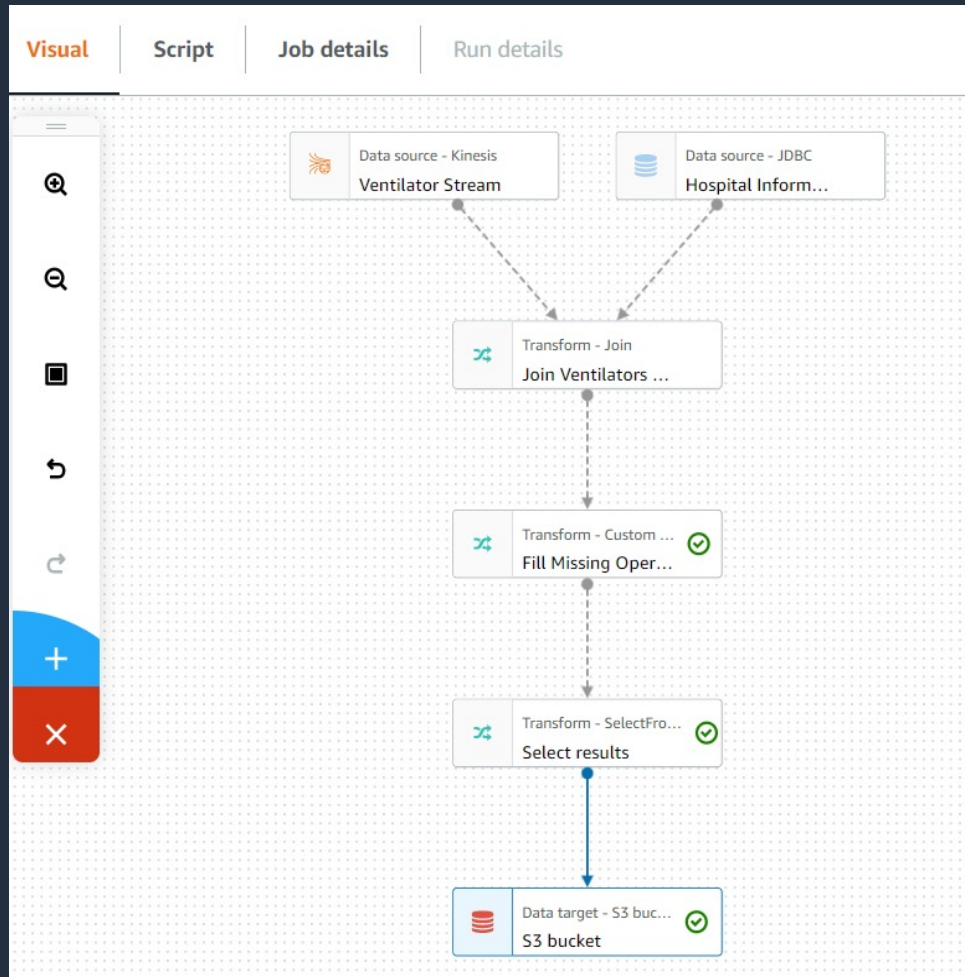
Author AWS Glue jobs visually without coding

Monitor 1,000s of jobs through a single pane of glass

Distributed processing without the learning curve

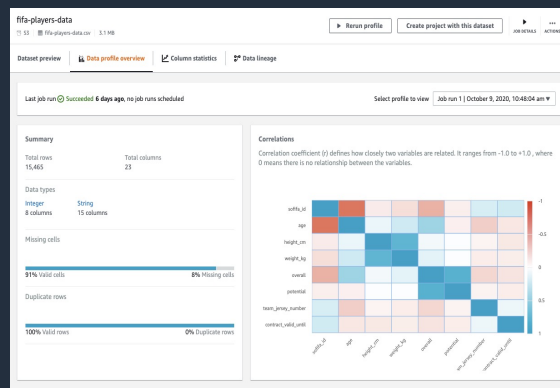Advanced transforms though code snippets

aws

AWS Glue DataBrew

Visual data preparation for analytics and machine learning
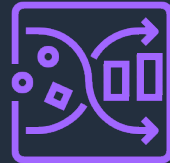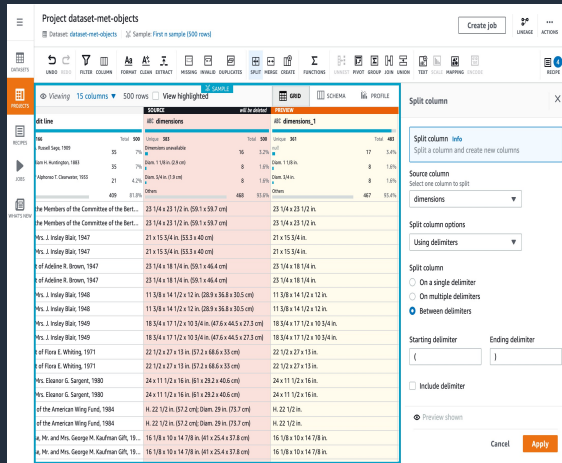
Clean and normalize data with
a rich visual interface

Choose from 250+ built-in
transformations to automate tasks

Profile data to understand data
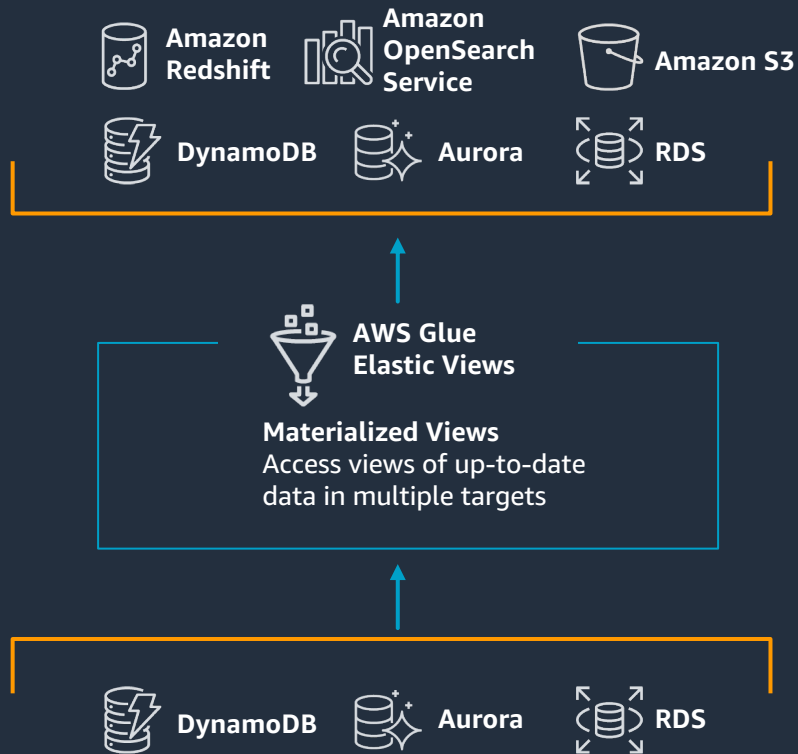patterns and anomalies

Work on large datasets at scale

aws

AWS Glue Elastic Views

Easily combine and replicate data across multiple data stores

Amazon Redshift   Amazon OpenSearch Service   Amazon S3

DynamoDB   Aurora   RDS

AWS Glue Elastic Views

Materialized Views
Access views of up-to-date data in multiple targets

DynamoDB   Aurora   RDS

Create materialized views across a wide variety of databases and data stores using familiar SQL

Continually monitors source databases for changes and updates targets within seconds

Serverless and automatically scales capacity up and down to accommodate your workloads

Handles the heavy lifting of copying and combining data without requiring custom code
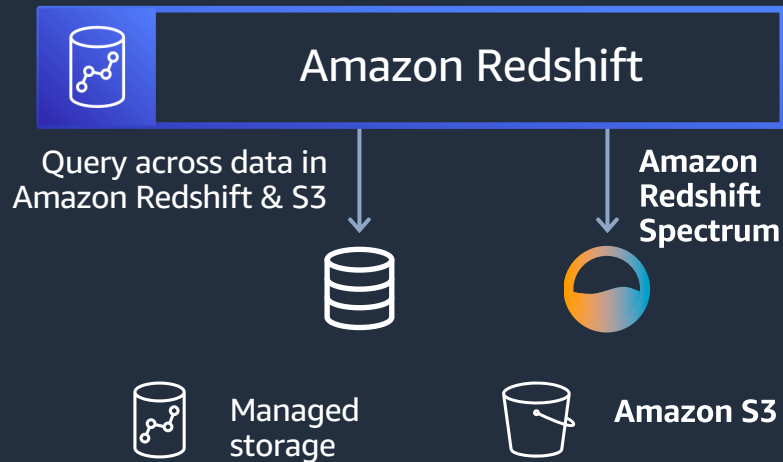
aws

Moving data to and from the data lake

Extend the data warehouse to exabytes of data in Amazon S3 data lakes

Directly query data stored in Amazon S3

Parquet, ORC, Avro, JSON, and CSV data formats
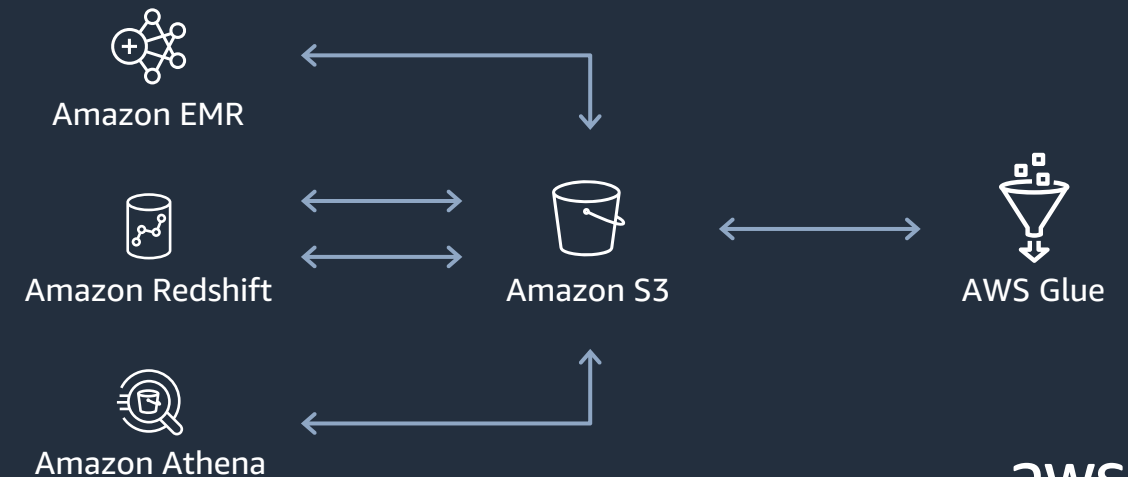
Any scale of data; pay for what you use

Unload Amazon Redshift data as Parquet to Amazon S3 data lakes for faster sharing and analytics

Parquet is an open data format supported by Amazon EMR, Athena, and Amazon Redshift

Amazon Redshift now supports exporting data to Amazon S3 in Parquet format
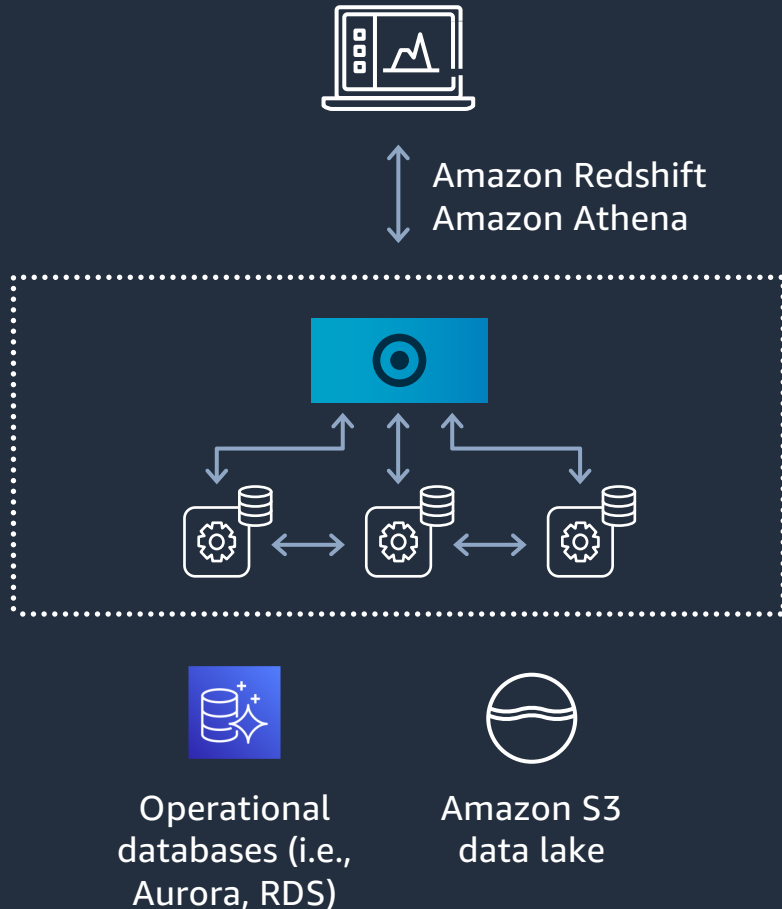
Use SQL with Amazon Redshift's Unload command to export data in Parquet format

Unloaded data is automatically registered in AWS Glue Data Catalog

Amazon Redshift

Query across data in Amazon Redshift & S3

Amazon Redshift Spectrum

Managed storage

Amazon S3

Amazon EMR

Amazon Redshift

Amazon S3

AWS Glue

Amazon Athena

aws

Federated query in Amazon Redshift and Athena

Unified analytics across databases, data warehouse, and data lake

Amazon Redshift
Amazon Athena

Operational
databases (i.e.,
Aurora, RDS)

Amazon S3
data lake

*Other sources available in Amazon Athena: Amazon ElastiCache for Redis,
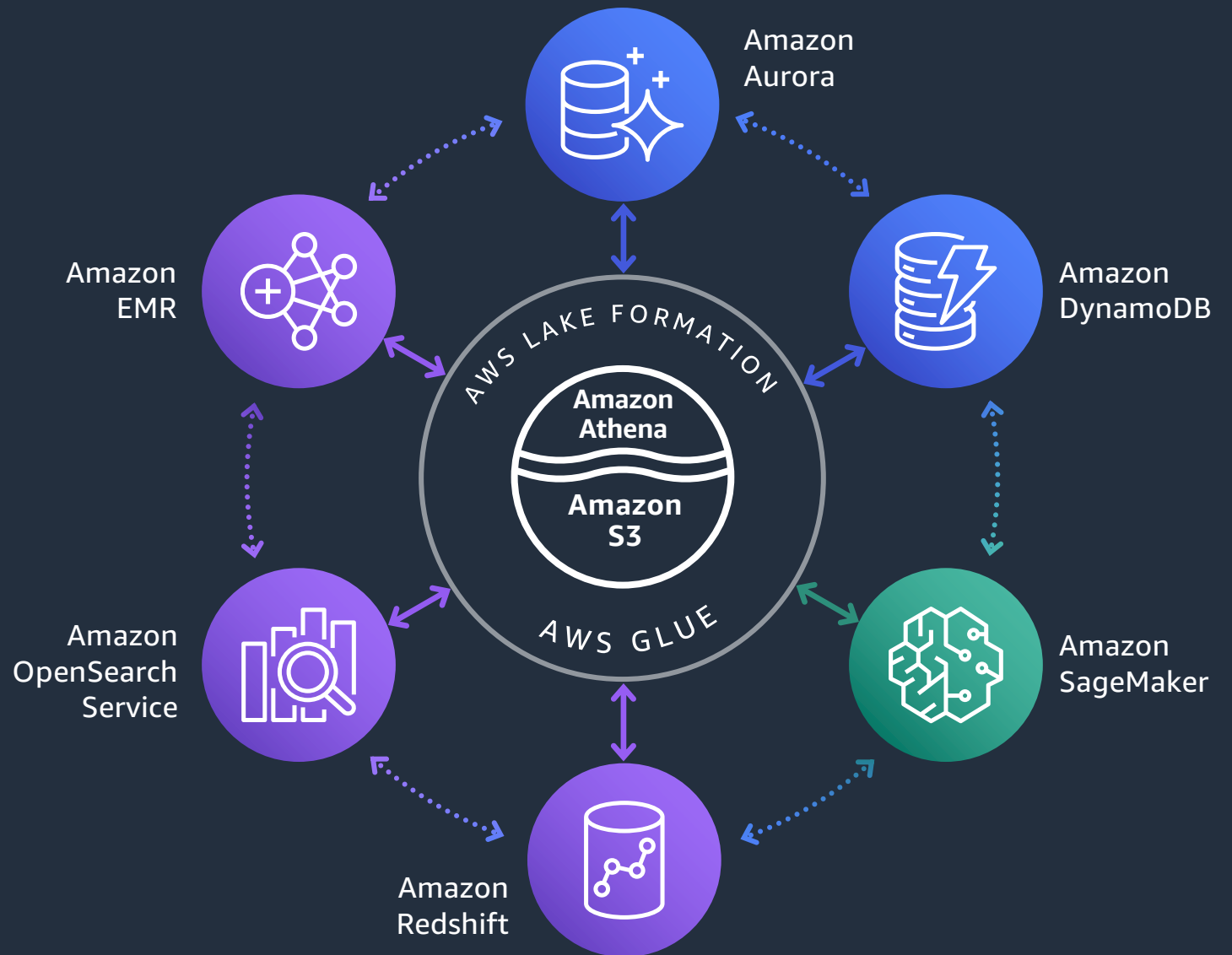Amazon DocumentDB, Amazon DynamoDB, HBase in Amazon EMR

Integrate operational database with data warehouse and Amazon S3 data lake

Analytics on operational data without data movement and ETL delays

Flexible and easy way to ingest data, avoiding complex ETL pipelines

aws

Unified governance

© 2021, Amazon Web Services, Inc. or its Affiliates.

AWS Lake Formation

Build a secure data lake in days

Build data lakes quickly

Move, store, and catalog your data faster; simplify data management with governed storage

Simplify security management

Centrally define and enforce security, governance, and auditing policies

Provide self-service access to data

Share datasets easily and securely within your organization and with partners

aws

Performant and cost-effective

Industry-leading choice
of 200+ instance types
to meet workload needs

100 GBPS bandwidth
network interfaces
for performance

On-demand, reserved,
and spot instances
to reduce costs

Five highly available
storage tiers and
intelligent tiering

aws

Amazon QuickSight

A scalable, embeddable, ML-powered Bi service built for the cloud

BI at scale
No servers to manage; pay per use billing

Embedded analytics
Quickly embed dashboards in your apps

ML-powered insights
Built-in anomaly detection and forecasting

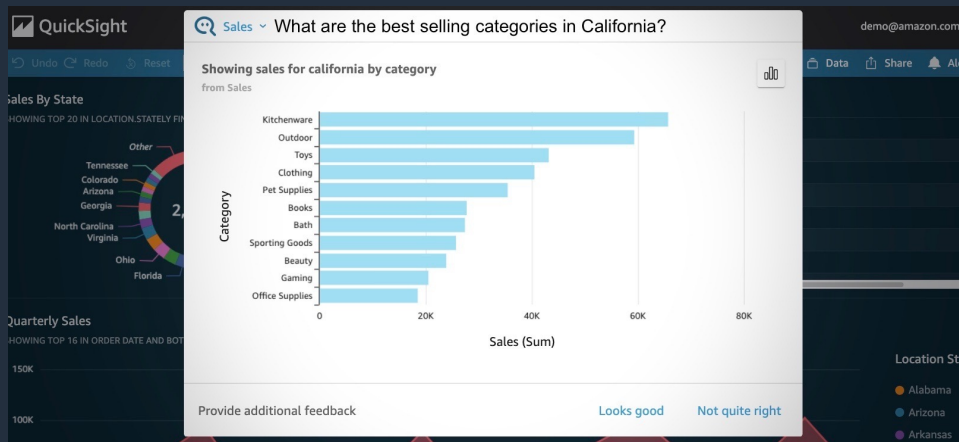Written narratives to interpret your data for you

aws

Amazon QuickSight Q

ML-powered natural language capability in Amazon QuickSight

Enter business questions in search bar and get answer in seconds

ML generates data models that automatically understands meanings and relationships

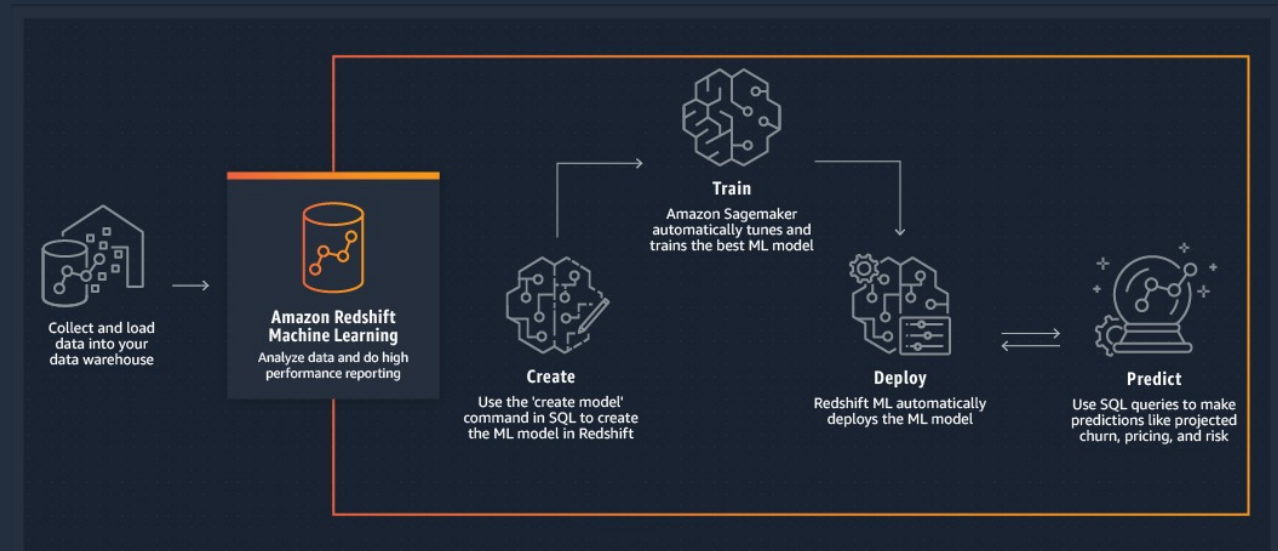Not limited to only asking a specific set of questions

Amazon Redshift ML

Create, train, and deploy machine learning (ML) models using familiar SQL commands

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

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

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

Collect and load data into your data warehouse

Amazon Redshift Machine Learning
Analyze data and do high performance reporting

Create
Use the 'create model' command in SQL to create the ML model in Redshift

Train
Amazon Sagemaker automatically tunes and trains the best ML model

Deploy
Redshift ML automatically deploys the ML model

Predict
Use SQL queries to make predictions like projected churn, pricing, and risk

aws

Thank you

aws