



Building end-to-end Machine Learning Platform on AWS

Data for Thai – Data for All

Satsawat Natakarnkitkul

Senior ML/AI Solutions Architect
Amazon Web Services

Why modern data platform?



Customers want more value from their data



Growing
Exponentially



From new
sources



Increasingly
diverse



Used by
many people



Analyzed by many
applications

Put data **to work**



Make better
decisions



Improve
efficiencies

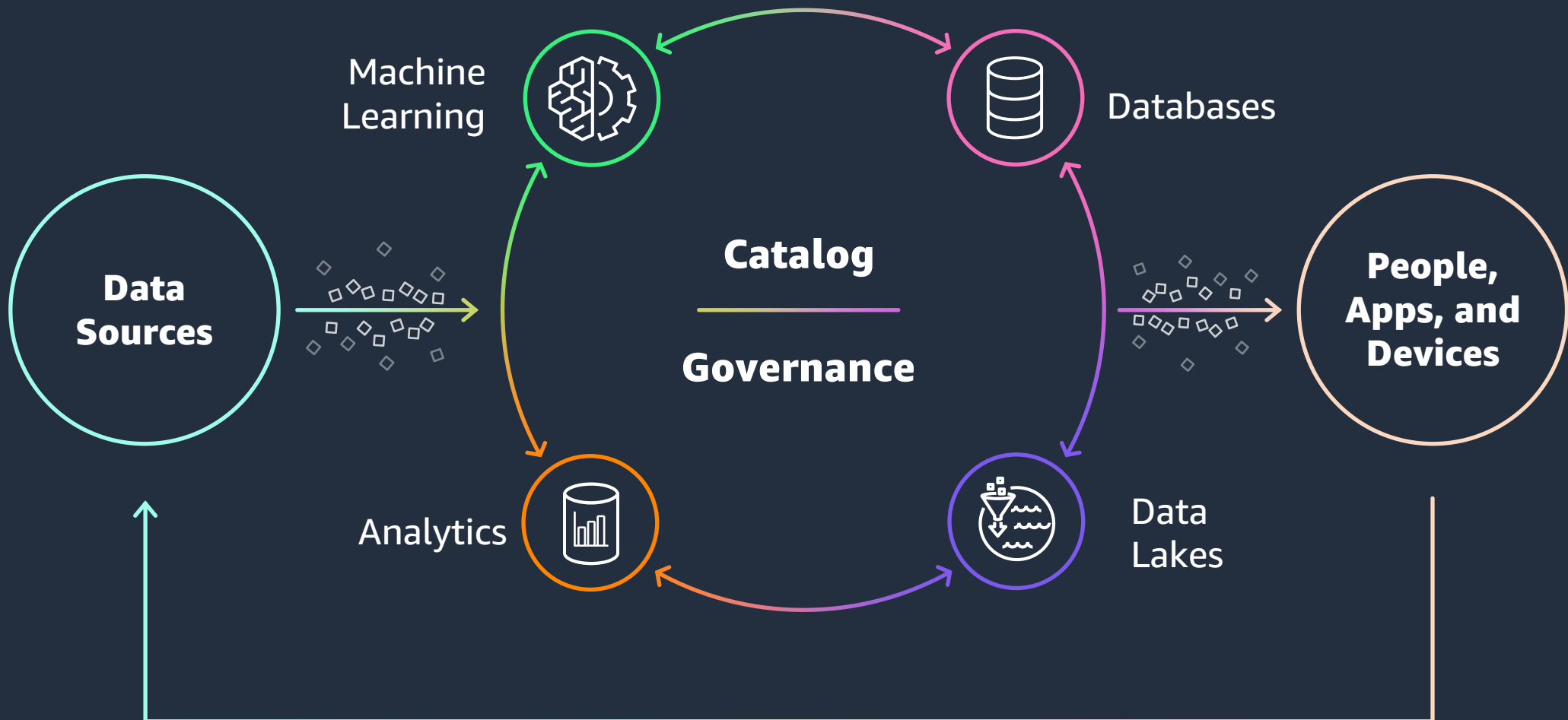


Respond
faster



Uncover
opportunities

Modern data strategy in action



Focusing on business outcomes



Customer experience

Built a customer engagement service using a Modern Data Architecture to serve over eight million developers working with 190k+ businesses in 100+ countries

Twilio

Real-time insights to give tens of millions of users personalized streaming recommendations

Disney+

Increased the use of self-service analytics platform by over 40% for daily active fans—sharing richer information in near real-time

OneFootball

Personalizes searches for better customer experience and gets fewer returns due to improved sizing recommendations

Zappos



Agility and innovation

Accelerates zero-carbon transition with automated energy predictions and maximized wind farm energy production

ENGIE

Helps drive better insights needed to make key race-time decisions, giving a technological edge over competitors

Toyota Racing Development

With Amazon Managed Streaming for Apache Kafka, the company is able to experiment with big changes safely with little risk

New Relic

Built a sophisticated infectious disease tracker in four months for retirement community residents and employees

Erickson Living



Cost optimization

Manages over 150 PB of data at \$5 per terabyte of data scanned

FINRA

Shifting to AWS saves more than \$2 million annually in data storage costs

INVISTA

AWS Analytics reduced operational costs by over 30% while freeing software engineers of low-value work

Pinterest

Amazon EMR as its core ML platform allows for more accurate ML models 80% faster at an 80% lower cost

Eightfold.ai



Performance and scale

Moved to a Modern Data Architecture to ingest 70 billion records per day, and now runs Amazon Redshift queries 32% faster

Nasdaq

Scalability and cost efficiency during a global pandemic with 20x increase in ventilator production while reducing first-pass inspection failures by 60%

Vyaire Medical

Scaled ingestion to six billion documents per day using Amazon OpenSearch Service (successor to Amazon Elasticsearch Service)

Pearson

Had the tools to support a 101% increase in language learners

Duolingo

AWS Analytics Pillars



Scalable data lakes



Purpose-built for performance and cost



Serverless and easy to use

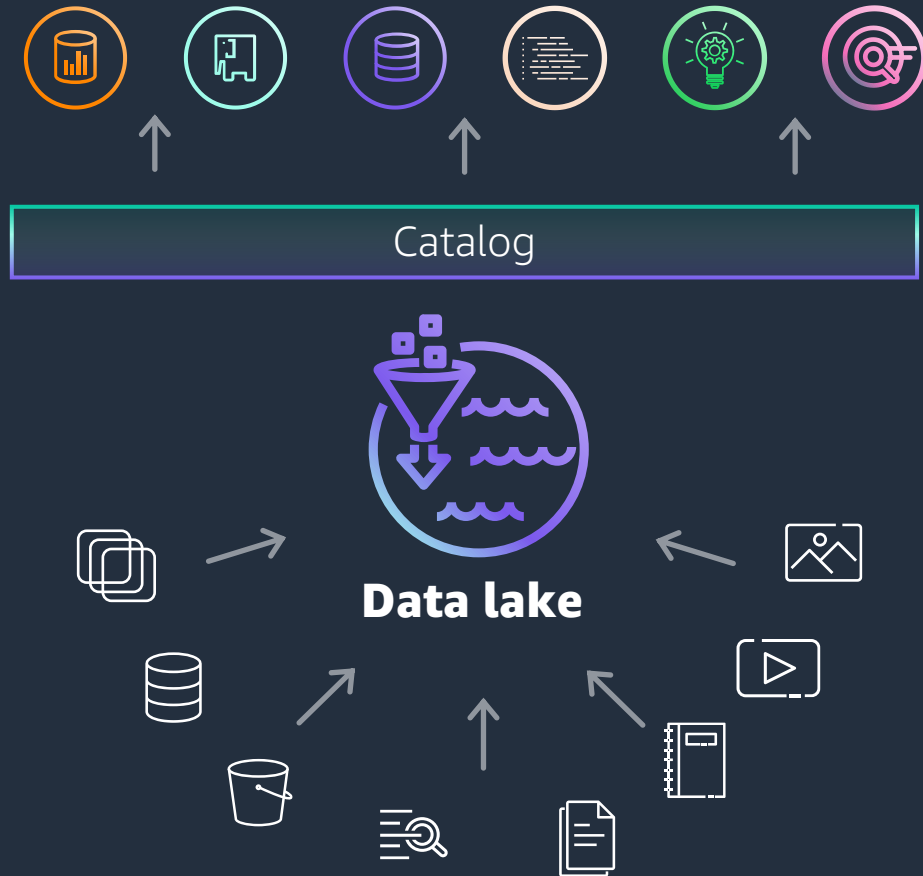


Unified data access, security, and governance



Scalable data lakes

The benefits of data lakes



Store all your data in open formats

Cost-effectively scale storage to exabytes

Decouple storage from compute

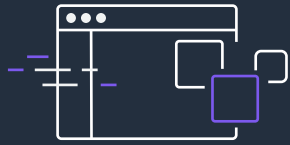
Choice of analytical and ML engines

Process data in place



Purpose-built for
performance
and cost

Performant and cost effective



Performance



Scalability



Increasing and
unpredictable cost



- ❖ Decouple storage from compute
- ❖ Prefer serverless over managed over self-hosted
- ❖ Pay per use
- ❖ Compress and partition
- ❖ Query purpose-built service for higher performance
- ❖ Query the data lake for lower cost



Serverless
and easy to use

AWS has the most serverless options for data analytics in the cloud





Unified data access, security, and governance

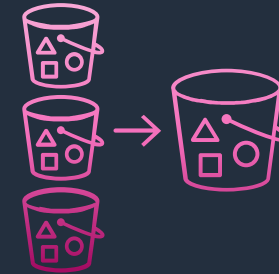
Challenges of building and securing modern data lakes



Support updates
and deletes



Row-level
Fine-grained
Secure sharing



Automatic storage
optimization

Break down data silos



Extract,
transform, load



Visual data
preparation



Data
replication



Data warehouse
to/from data lake



Federated
query

Disney+ empowers fast data ubiquity using Amazon Kinesis

CHALLENGE

Experiencing slow, limited data insights from data silos and batch processing, Disney+ needed to give its teams fast data access so that they could improve the Disney+ customer experience at scale.

SOLUTION

To achieve fast data democracy and near-real-time insights, Disney+ built a streaming data platform using AWS Analytics services, including Amazon Kinesis Data Streams and Amazon Kinesis Data Firehose.

RESULTS

Using AWS, Disney+ now supports a data-driven culture that provides near-real-time data and insights based on billions of events to improve the experience of tens of millions of users reliably and cost efficiently.



Amazon Kinesis Data Streams



Amazon Kinesis Data Firehose



Dollar Shave Club uses AWS to speed data analysis, improve user experience

CHALLENGE

Dollar Shave Club needed to find the best way to optimize storage and compute for its growing analytics environment.

SOLUTION

Dollar Shave Club created a Modern Data Architecture featuring Amazon S3 and Amazon Redshift, taking advantage of the Amazon Redshift Spectrum feature to query 60 TB of data.

RESULTS

- Builds analytical reports in 5 minutes instead of 8 hours
- Saves \$300,000 a year by optimizing cluster sizes
- Puts savings into research and development
- Creates multiple reports daily instead of 3–4 a week



**Gain actionable insights
from your data using ML**



The reach of ML is growing



INCREASED SPENDING

By 2026, global spending on artificial intelligence will reach \$300 billion growing 4.2 faster than average IT spend

IDC

"IDC Press Release, Worldwide Spending on AI-Centric Systems Will Pass \$300 Billion by 2026, According to IDC, September 2022,"
<https://www.idc.com/getdoc.jsp?containerId=prUS49670322>



AI ACROSS SEVERAL BUSINESS UNITS

Gartner survey reveals 80% of executives think automation can be applied to any business decision

Gartner

Gartner, Press Release
<https://bit.ly/3is16a2>



AI IS CRITICAL TO SUCCESS

94% of business leaders surveyed say AI is critical to their success

Deloitte

Deloitte, "State of AI in the Enterprise,"
<https://bit.ly/3XiNOLR>

The AWS AI/ML Stack

Consumer

AI SERVICES

Generative AI

SCALING GENERATIVE AI

Amazon Bedrock

AMAZON BUILT FOUNDATIONAL MODEL

Amazon Titan

Specialized AI Services

BUSINESS PROCESSES

Amazon Personalize
Amazon Forecast
Amazon Fraud Detector
Amazon Lookout for Metrics

SEARCH

Amazon Kendra

CONVERSATION

Amazon Lex
Contact Lens
Voice ID
Amazon Transcribe Call Analytics

CODE + DEVOPS

Amazon CodeGuru
Amazon CodeWhisperer
Amazon DevOps Guru

INDUSTRIAL

Amazon Monitron
Amazon Lookout for Equipment
Amazon Lookout for Vision

HEALTH

Amazon HealthLake
Amazon Comprehend Medical
Amazon Transcribe Medical
Amazon Omics

Core

TEXT

Amazon Translate Amazon Comprehend

SPEECH

Amazon Polly Amazon Transcribe

VISION

Amazon Textract
Amazon Rekognition AWS
Amazon Panorama

Tuner

AMAZON SAGEMAKER

JUMPSTART

Use pre-built models in SageMaker

CANVAS

No-code ML for business analysts

STUDIO LAB

Learn ML

GROUND TRUTH

Label data

STUDIO IDE

Prepare data
Store features

Geospatial ML

Build with notebooks

Train models
Tune parameters

Deploy in production
Manage and monitor

----- CI/CD | GOVERNANCE | RESPONSIBLE ML -----

Provider

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, TensorFlow

Amazon EC2

CPUs

GPUs

AWS Inferentia

AWS Trainium

Habana Gaudi

FPGA



Innovate faster with the most comprehensive set of AI/ML services

AWS ADVANTAGE

AI powered use cases

Hundreds of pre-built algorithms, models, and solutions for common use cases and industries
Faster time to value with pre-trained AI Services

Democratize access to ML

Empower any business analyst to make predictions without writing code with Amazon SageMaker AI Services to easily add AI capabilities to your applications, no ML skills required

Scale ML

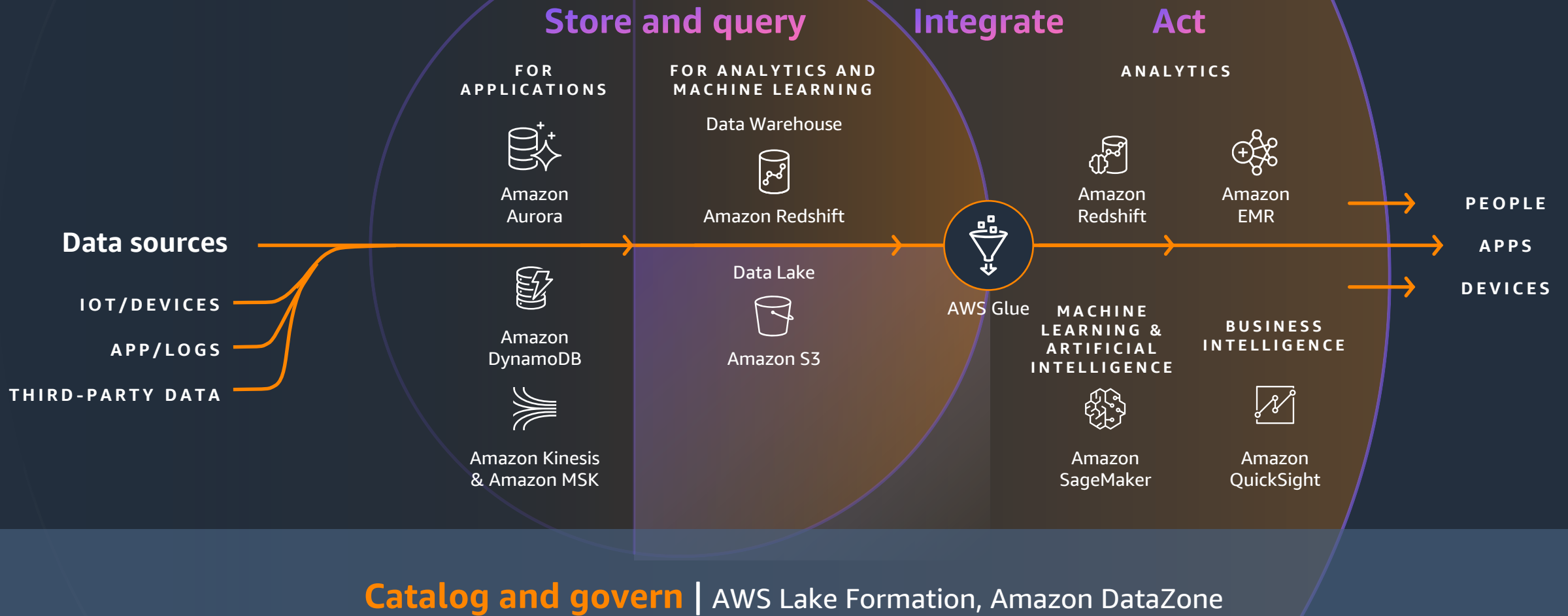
Build, train, and deploy ML models for any use case with Amazon SageMaker
Harness the exponential growth of ML models with foundation models on AWS

Grow ML skills

Learn more with AWS Training, AWS DeepRacer, AI & ML Scholarship Program, and AWS Machine Learning University
Accelerate your journey with AWS ML Embark Program or Amazon ML Solutions Lab



Building an end-to-end ML strategy



BUSINESS NEED

Wanted to enable ML at scale, while ensuring compliance with security policies

Sought to develop ML model to power question-answering capabilities for tax and legal products

SOLUTION

Created the secure content workspaces (SCW) which provides access to data in compliance with internal security standards

Amazon SageMaker, enabled through SCW, provides a fully managed machine learning environment

Built and trained natural language processing model on Amazon SageMaker

IMPACT

Data scientists are able to securely experiment at scale in compliance with Thomson Reuters policies

Reduced training costs by 40–50% on average

Saved the team countless hours of coding required for a self-managed ML infrastructure



THOMSON REUTERS



For over 25 years we have been developing advanced machine learning capabilities to mine, connect, enhance, organize, and deliver information to our customers.



Khalid Al-kofahi

Vice President of Research
and Development
Thomas Reuters

MORE INFO: [BLOG](#)

BUSINESS NEED

Pomelo Fashion had been displaying items in the same web format since its founding. The setup had grown stale, and the algorithm for displaying products relied on old data streams with limited inputs and spotty accuracy

SOLUTION

Pomelo Fashion used Amazon Personalize to build a recommendation engine to personalize the shopping experience to each customer. It now reflects user preferences on product pages in minutes, driving sales

IMPACT

Boosted click-through rate from category to product pages by up to 18%
Increased gross revenue from category pages by up to 15%
Increased return on investment by 400% within 1 month

Pomelo.



When you think of e-commerce, you think of AWS.

New services are always coming out on AWS, and support is very good.



Shane Leese
Business Intelligence Director
Pomelo Fashion

MORE INFO: [CASE STUDY](#) | [VIDEO](#)

Build with Generative AI



a robot painting (how does a brain of a human)

beautiful robotic butterfly anatomy diagram

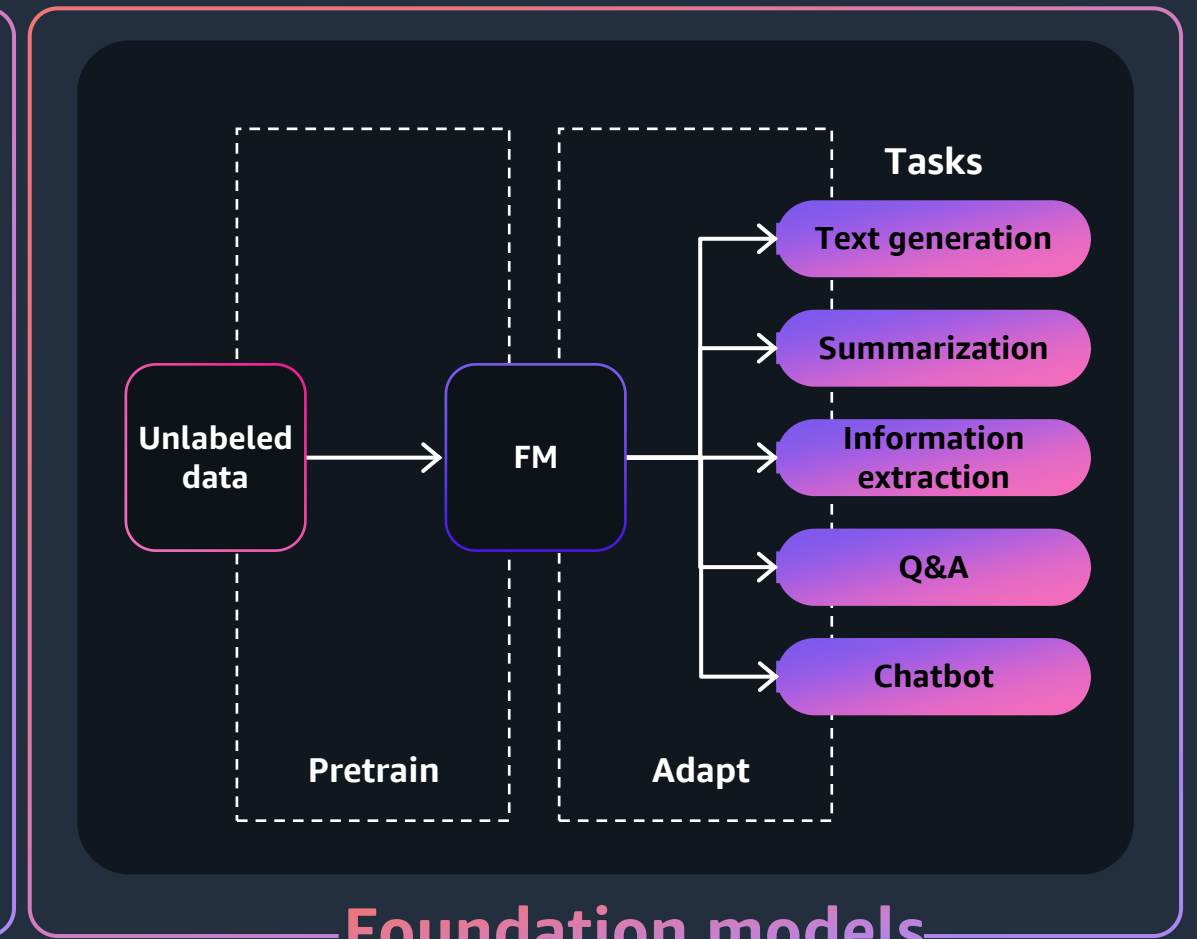
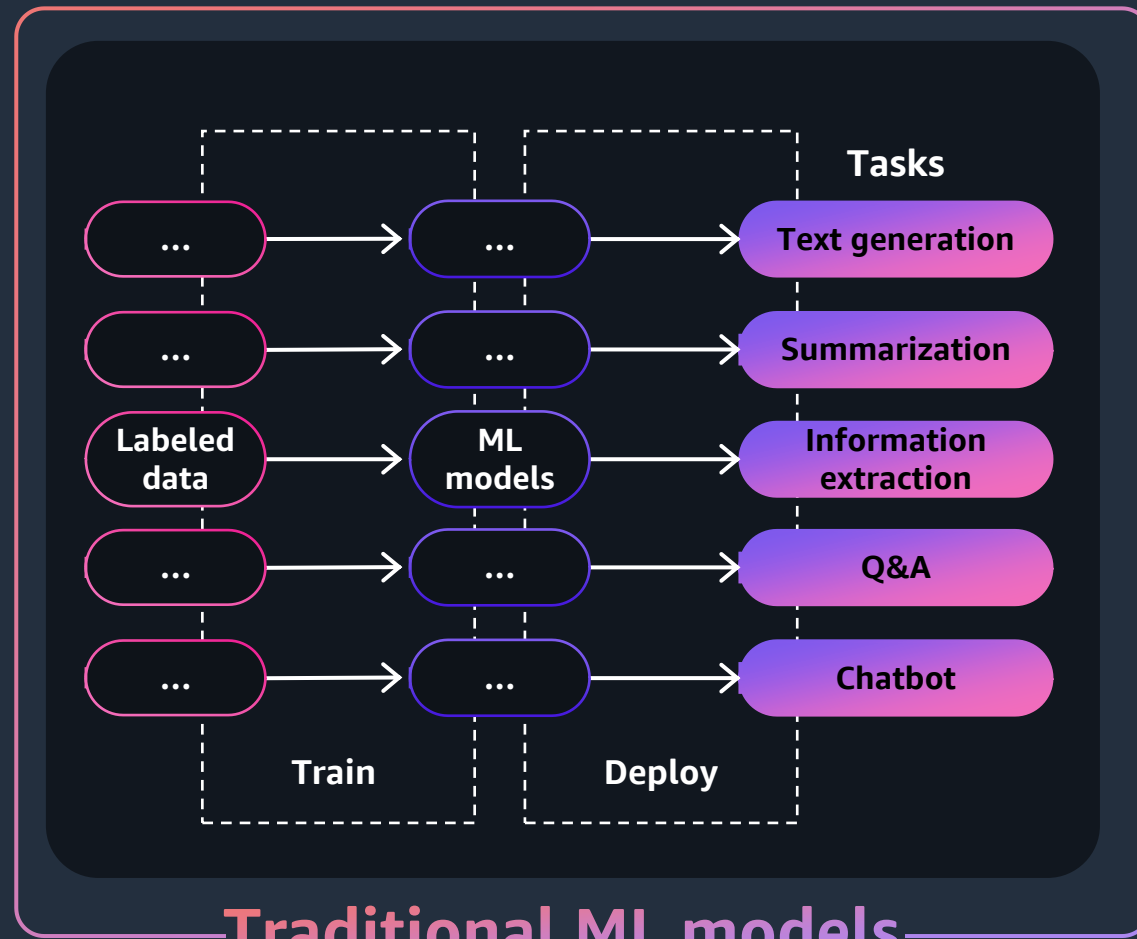
photo of a statue of a robot in university courtyard

astronaut on a horse

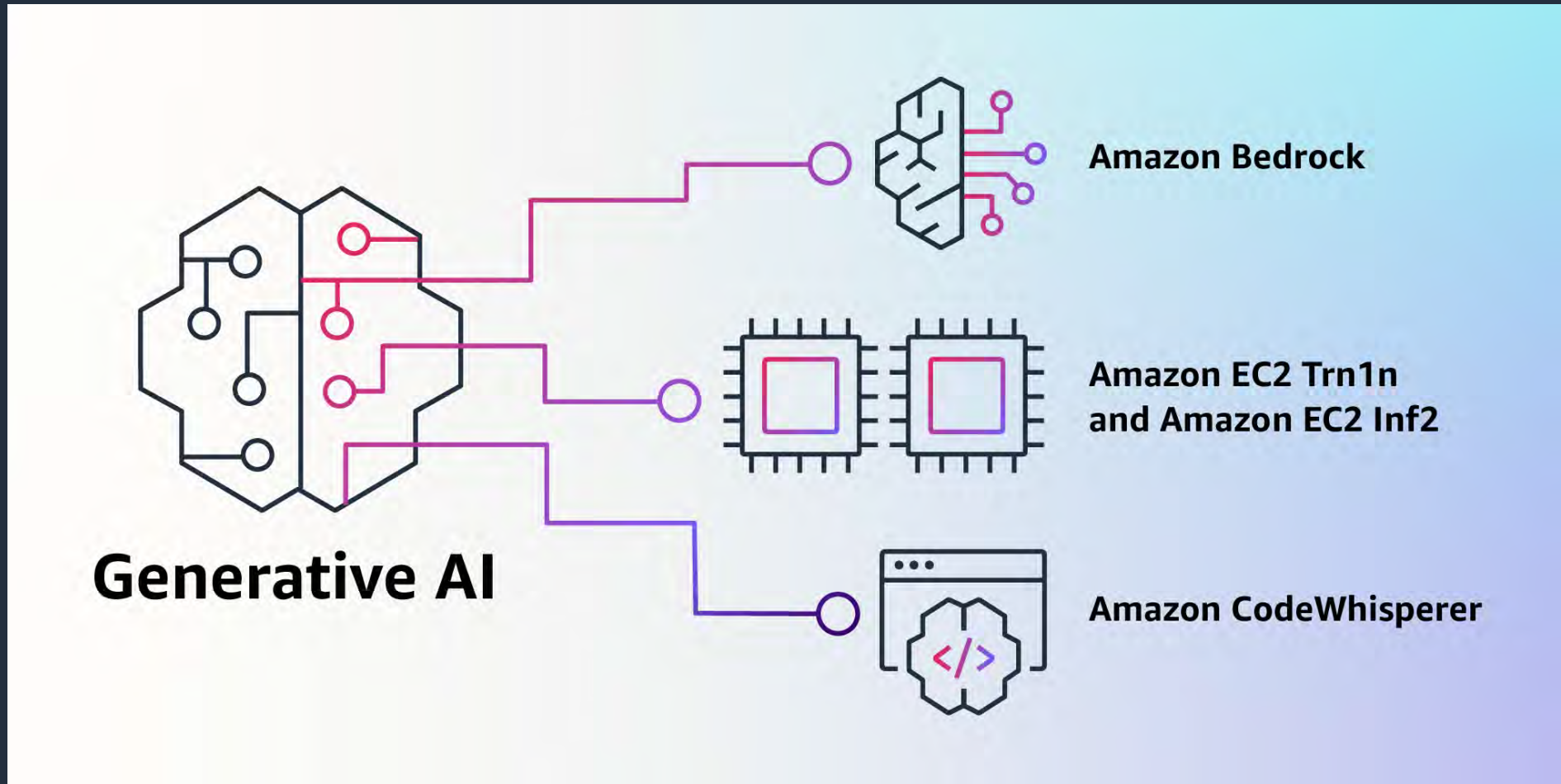
Question: What is generative artificial intelligence (AI)?

- Creates new content and ideas, including conversations, stories, images, videos, and music
- Powered by large models that are pretrained on vast corpora of data and commonly referred to as foundation models (FMs)

Why foundation models?



Building with generative AI on AWS



NEW

Amazon Bedrock

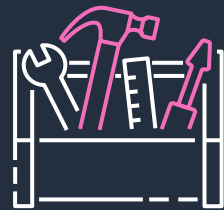
The easiest way to build and
scale generative AI
applications with FMs



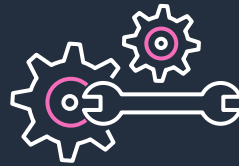
Amazon Bedrock key benefits



Accelerate development of generative AI applications using FMs through an API, without managing infrastructure



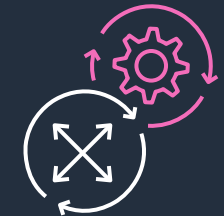
Choose FMs from AI21 Labs, Anthropic, Stability AI, and Amazon to find the right FM for your use case



Privately customize FMs using your organization's data



Enhance your data protection using comprehensive AWS security capabilities



Use AWS tools and capabilities that you are familiar with to deploy scalable, reliable, and secure generative AI applications

Foundation models from top AI startups

The logo for AI21 Labs, featuring the text "AI21" in black and "labs" in red.

Jurassic-2

Multilingual LLMs for text generation in Spanish, French, German, Portuguese, Italian, and Dutch

The logo for Anthropic, featuring the word "ANTHROPIC" in black, all-caps, sans-serif font.

Claude

LLM for conversations, question answering, and workflow automation based on research into training honest and responsible AI systems

The logo for Stability.ai, featuring the text "stability.ai" in black, lowercase, sans-serif font.

Stable Diffusion

Generation of unique, realistic, high-quality images, art, logos, and designs

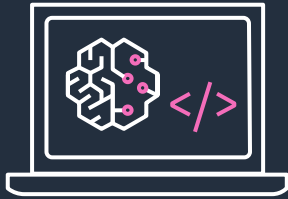
NEW

Amazon CodeWhisperer

Build applications faster and
more securely with an AI coding
companion



Amazon CodeWhisperer: Now generally available and free to use for individual developers!



Generate code suggestions in real time



Scan code for hard-to-find vulnerabilities



Flag code that resembles open-source training data or filter by default

During preview Amazon ran a productivity challenge, and participants who used Amazon CodeWhisperer were **27% more likely to complete tasks successfully and did so an average of 57% faster** than those who did not use CodeWhisperer.

Accelerate your generative AI initiatives with the global partner community

The Deloitte logo is displayed in a white rectangular box. It consists of the word "Deloitte" in a bold, black, sans-serif font, followed by a small green circle.

“ At Deloitte, we’re advancing our generative AI capabilities to help clients harness its power. As part of that effort, the organization is leveraging leading offerings like Amazon Bedrock to advance such capabilities with our alliance relationships. **With Bedrock, we can provide clients with a cost-effective serverless API for AWS customers to build generative AI applications.** We are eager to help clients deliver better, faster AI results to the mutual customers of AWS and Deloitte. ”

Nishita Henry

Amazon/AWS Alliance Global Chief Commercial Officer, Deloitte Consulting LLP