

Startup Day with AWS and kreuzwerker, Warsaw

Experience the AWS Cloud - hosted and guided by AWS & kreuzwerker

09. November 2023

PROJECT NAME	BLUE PILL BEIJING
DRAWING NAME	CLIMB THE WALL
DESIGNED BY	O.U STUDIO 刘宇联合 Chen, Hsiao-Ton Wu, K-Lin Tsai, Pei-Huan
SCALE	1:1
DATE	2009/6/4
SHEET NO	1
DRAWING NO	A1-1

Agenda

kreuzwerker. Short introduction.

Speakers: Alex Ramos (kreuzwerker),
Maria Nunez (AWS).

Building Gen AI Applications on AWS.

Speakers: Chandan Rajah (AWS), Jenny
Vega (AWS).

Startup Bootcamp: AWS Programs.


Speaker: Nina Trejos (AWS).

**Architecting for Success and Optimizing
for Longer Runways.** Speaker: Wojciech
Orzechowski (kreuzwerker).

kreuzwerker

Short Introduction

09. November 2023



The background is a photograph of a brick wall with a white architectural drawing overlaid. The drawing shows a ladder leaning against the wall, a table with a chair on the left, and a technical drawing of a window or door on the right. Dimensions like '21', '61', '45', and '41' are marked. A technical drawing of a wall section is on the far right, showing a brick chimney and a window with a roof profile above it.

PROJECT	BLUEPRINT BERLIN
DRAWING NAME	CLIMB THE WALL
DESIGNED BY	O.U. STUDIO 劉凱強 打開聯合 Chen Hsiao-Tsun Wu K. Lin Tsai Pei Huan
SCALE	1 / 1
DATE	2009/6/4
SHEET NO	1
DRAWING NO	A1-1



kreuzwerker

We unleash digital technology. Made to measure.



Founded 2010

AWS Partner since 2013

100% AWS

Berlin, Frankfurt,
Warsaw, Zurich

Migration & Modernization

Data & Analytics

Cloud Native Development

Managed Security Services

Cloud Financial Management





kreuzwerker Poland

Founded 2013

From “extended workbench” to Cloud Consulting

100% AWS

Certified and more importantly proven experts

Access to expertise of a globally active company

Access to funding to start / grow your business



Selected Services & Solutions

krezuwerker Poland

Data Readiness Assessment

Scope: Assessment of current data types and usage.

Deliverable: Data Strategy including Best Practices

Time & Effort: 5 Person Days

Funding Options: Based on ARR MAP/MAP Light eligible

Open Search Assessment

Scope: Assess Search Scope, Search Engine Interaction

Deliverable: Assessment Report

Time & Effort: 5 Person Days

Funding Options: Based on ARR MAP/MAP Light eligible

Well Architected Framework Review

Scope Analyse Data Workload/Architecture along the Well-Architected Pillars

Deliverable Report of findings / recommendations

Time & Effort: 2 Person Days

Funding Options: 5k Credits

AI/ML Development & PoC Building

Scope Identification of AI/ML Use Cases and Implementation / Building of a Data / AI/ML PoC

Deliverable Proof-of-Concept

Time & Effort: 10-25 Person Days

Funding Options: PoC Funding



STARTUP SCALE SERIES

Building Gen AI Applications on AWS

Chandan Rajah

Principal Solutions Architect
AWS

Jenny Vega

Solutions Architect
AWS

Agenda

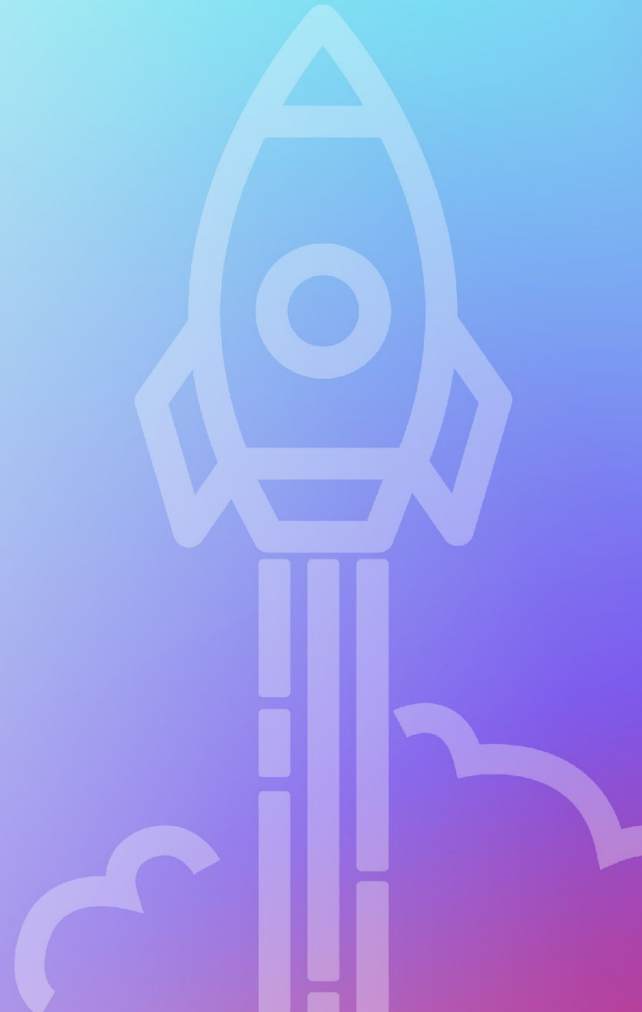
Introduction to Generative AI

Building on AWS

Reference Architectures

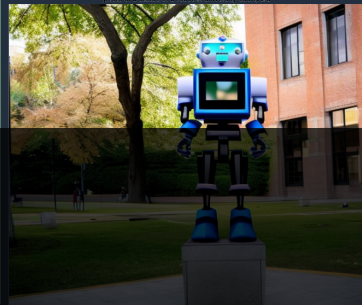
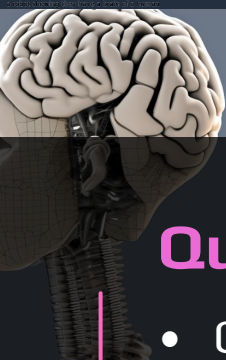
Demo

Questions



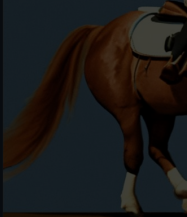
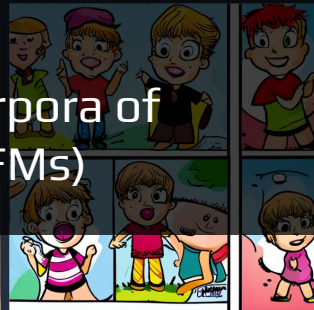
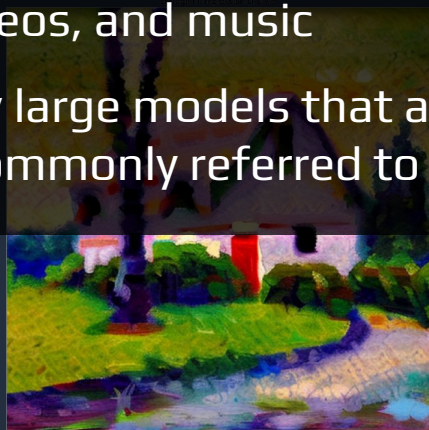
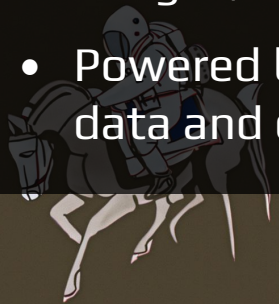
Introduction to Generative AI

A golden retriever wearing glasses and a hat in a portrait painting



Question: What is generative 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)



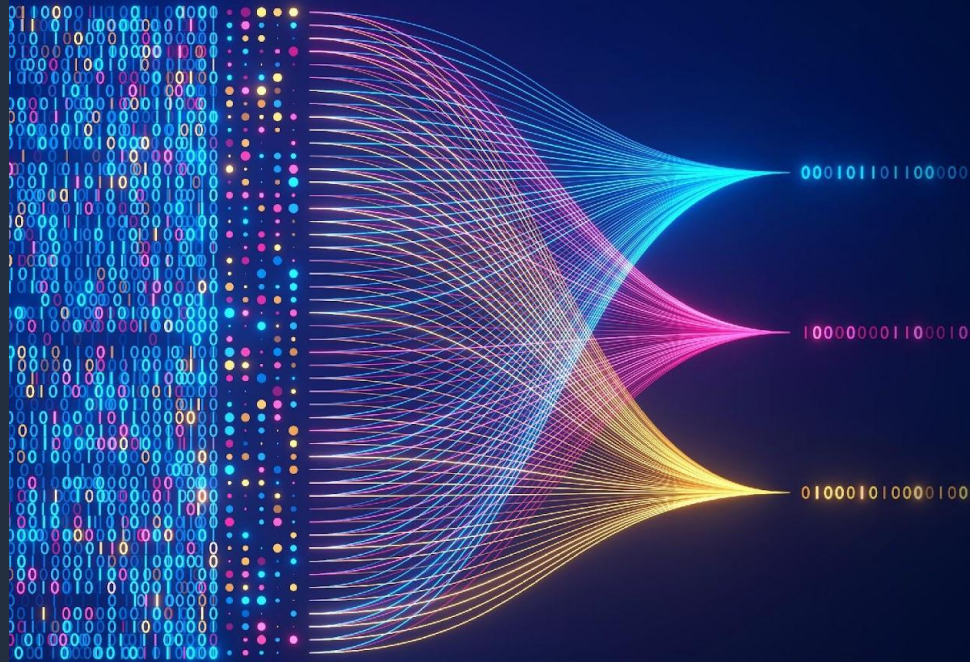
Generative AI is powered by foundation models

Pretrained on vast amounts of unstructured data

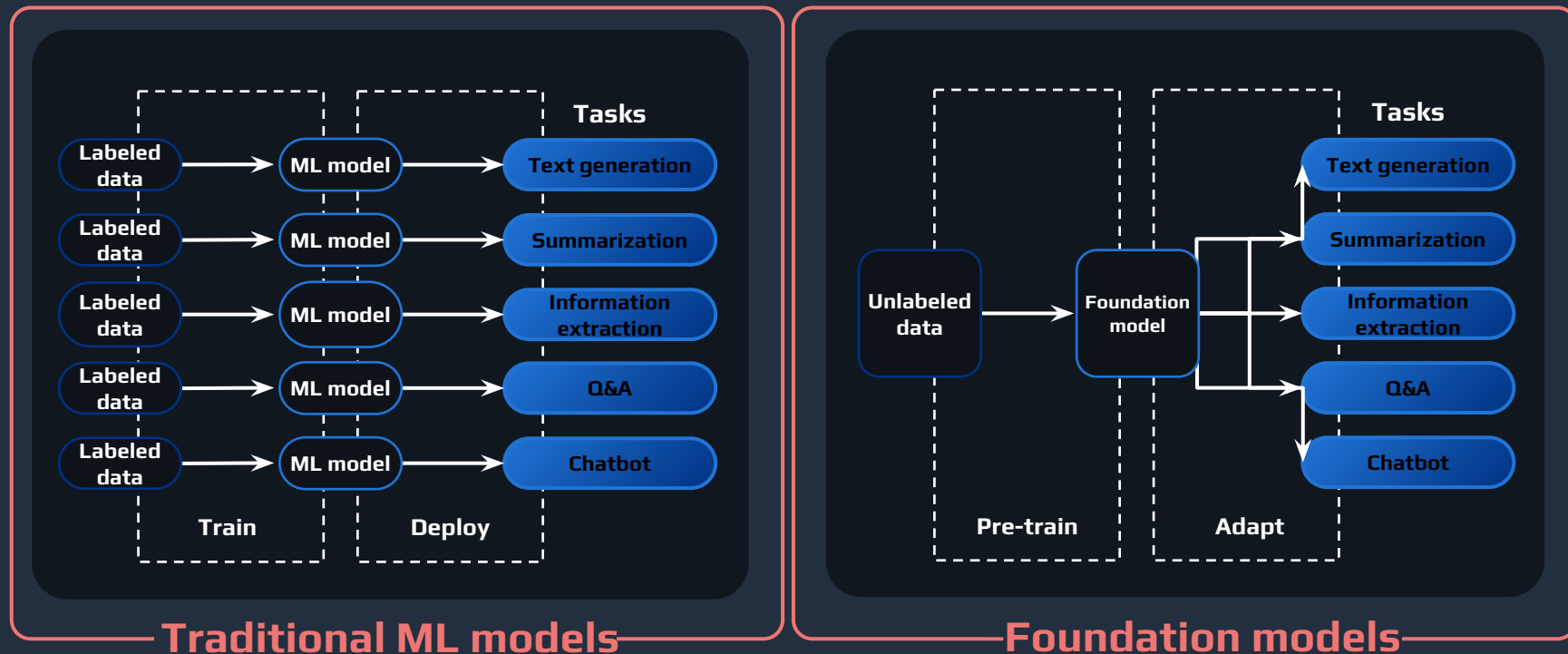
Contain large number of parameters that make them capable of learning complex concepts

Can be applied in a wide range of contexts

Customize FMs using your data for domain specific tasks

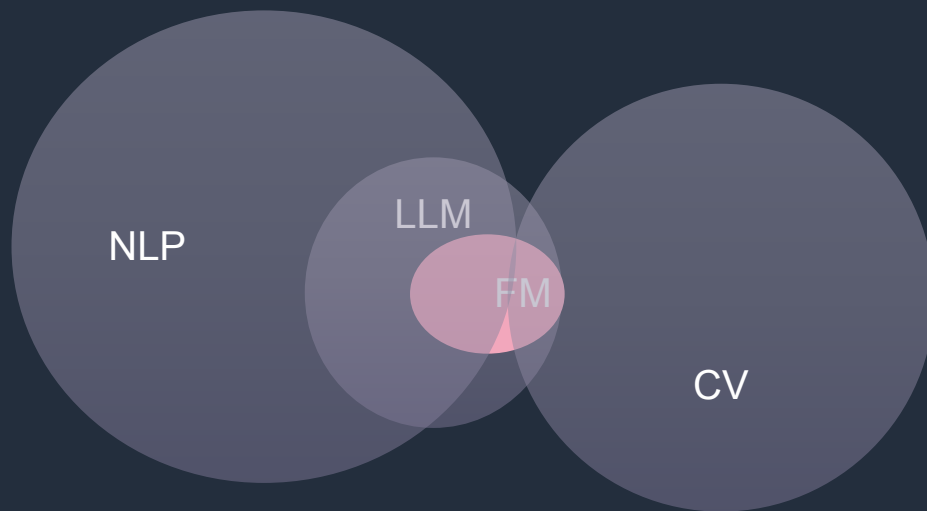


How foundation models differ from other ML models



Generative AI and foundation models

Generative AI is a type of AI that can create new content and ideas, including conversations, stories, images, videos, and music. Like all AI, generative AI is powered by ML models—very large models that are pre-trained on vast amounts of data and commonly referred to as Foundation Models (FMs).

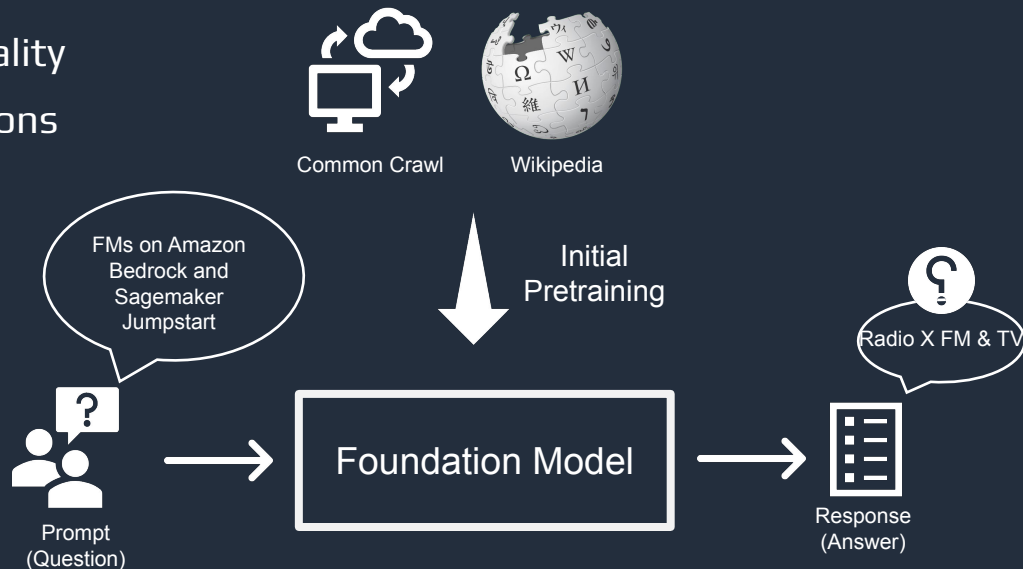


NLP – Natural Language Processing
CV – Computer Vision
LLM – Large Language Model
FM – Foundation Model

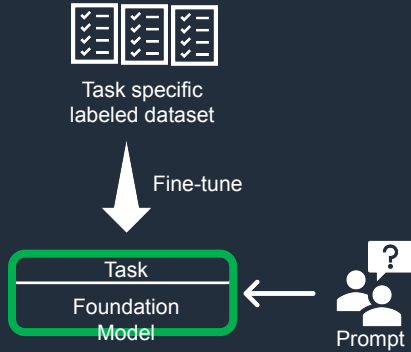
Adapting Foundational Models

Why adapt a foundation model?

- Specific Task
- Closed-domain knowledge
- Current Knowledge
- Improving the performance/quality
- Reduce likelihood of hallucinations

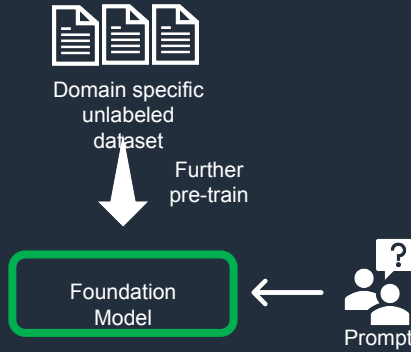


How can I adapt a foundation model?



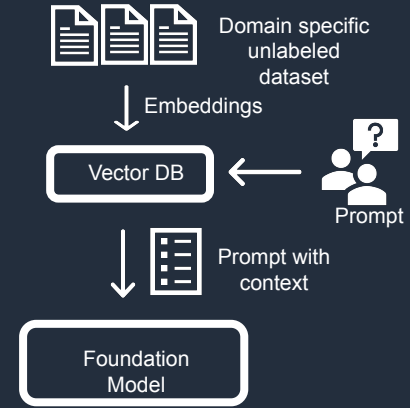
Prompt based learning

- Supervised learning (Fine Tuning)
- Updates the model weight
- Small amount of data (100 to 500) - JSONL
- Fewer epochs
- Task Specific
- Supported models (T5, GPT-J, Diffusion)



Domain Adaptation

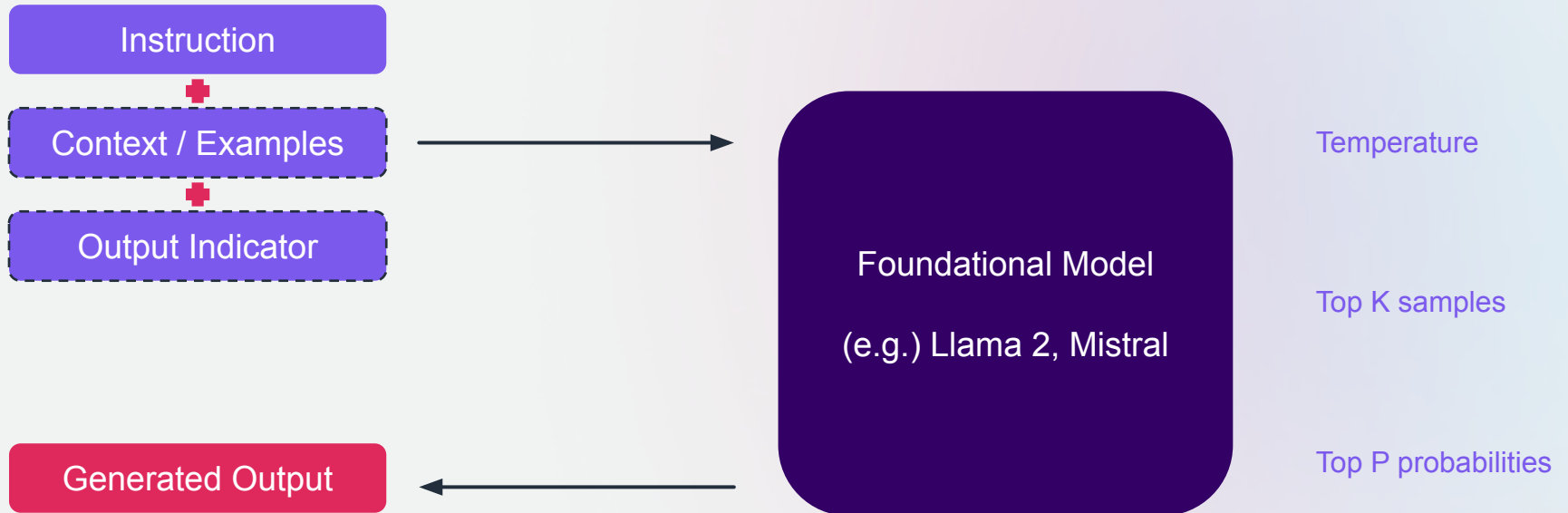
- Self-supervised learning (further pretrain)
- Updates the model weight
- Large corpus of data (CSV, JSON, TXT)
- Custom vocabulary & tokenizer
- Not specific to task
- Supported models – GPT-JB & Variant
- Ex. Legal or fintech domain



Information Retrieval

- Augmented with closed-domain knowledge (RAG)
- Retrieval using dense vector representation
- Feeds prompt with the "Context"
- Query FM w/ context to get the output
- No updates to model weights

Anatomy of a Prompt



Zero Shot

Prompt:

Classify the text into neutral, negative or positive.
Text: I think the vacation is okay.
Sentiment:

Output:

Neutral

Few Shot

Prompt:

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:
We were traveling in Africa and we saw these very cute whatpus.
To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

Output:

When we won the game, we all started to farduddle in celebration.

Brown et al.
2020

Wei et al.
(2022)

Prompt:

This is awesome! // Negative
This is bad! // Positive
Wow that movie was rad! // Positive
What a horrible show! //

Output:

Negative

Min et al.
(2022)

Chain of Thought

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✔️

Including self consistency

[Wei et al. \(2022\)](#)

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The answer is 8. ❌

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are $16 / 2 = 8$ golf balls. Half of the golf balls are blue. So there are $8 / 2 = 4$ blue golf balls. The answer is 4. ✔️

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is

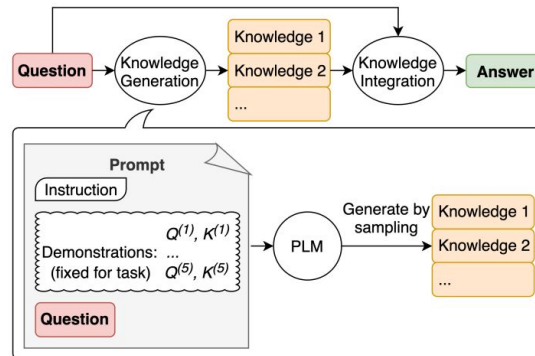
(Output) 8 ❌

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: Let's think step by step.

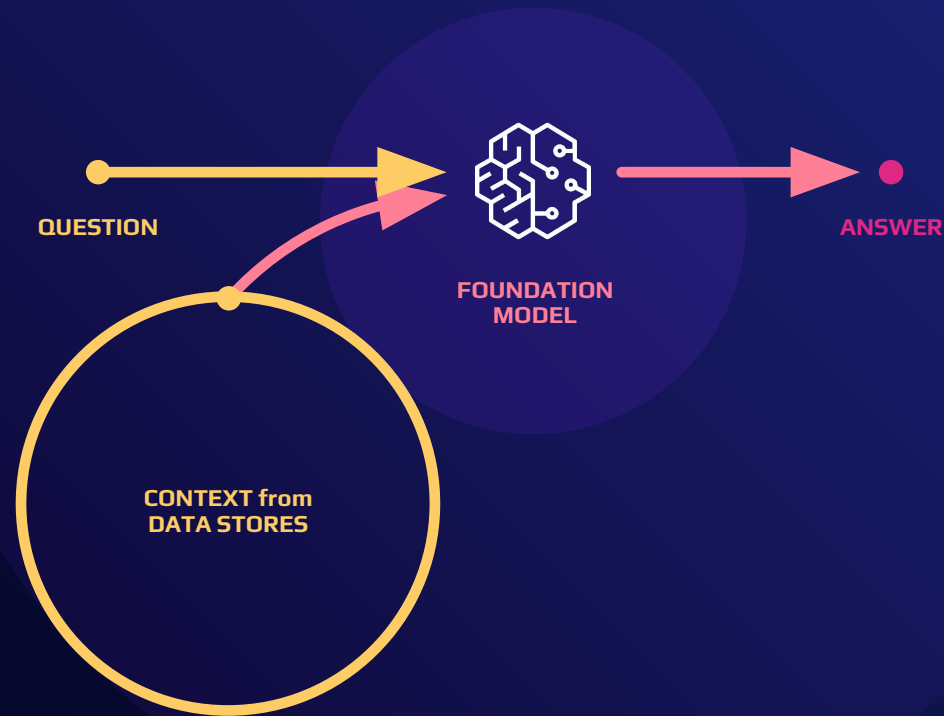
(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✔️

[Kojima et al. \(2022\)](#)



[Liu et al. \(2022\)](#)

Retrieval Augmented Generation (RAG)



Building on AWS

Amazon Bedrock

Choice of foundation models

AI21labs

JURASSIC-2

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

cohere

COMMAND + EMBED

Text generation model for business applications and embeddings model for search, clustering, or classification in 100+ languages

stability.ai

STABLE DIFFUSION XL 1.0

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

ANTHROPIC

CLAUDE 2

LLM for thoughtful dialogue, content creation, complex reasoning, creativity, and coding, based on Constitutional AI and harmlessness training

amazon

AMAZON TITAN

Text summarization, generation, classification, open-ended Q&A, information extraction, embeddings and search



Foundation models alone
cannot execute tasks



Enabling foundation models to complete tasks



1

**DEFINE INSTRUCTIONS
AND ORCHESTRATION**



2

**CONFIGURE FM TO
ACCESS DATA SOURCES**



3

**COMPLETE
ACTIONS WITH
API CALLS**



4

**MANAGE CLOUD HOSTING
AND SECURITY**

NEW



Agents for Amazon Bedrock

Enable generative AI applications to complete tasks in just a few clicks

IN PREVIEW TODAY



Breaks down and orchestrates tasks



Securely accesses and retrieves company data



Takes action by executing API calls on your behalf



Provides fully managed infrastructure support

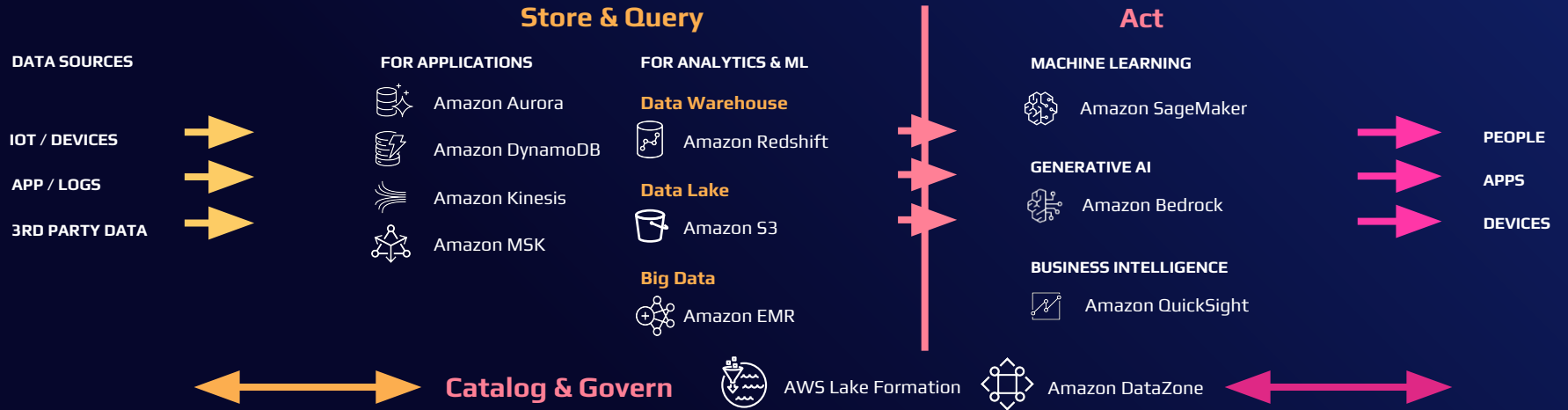


Your data is
your differentiator

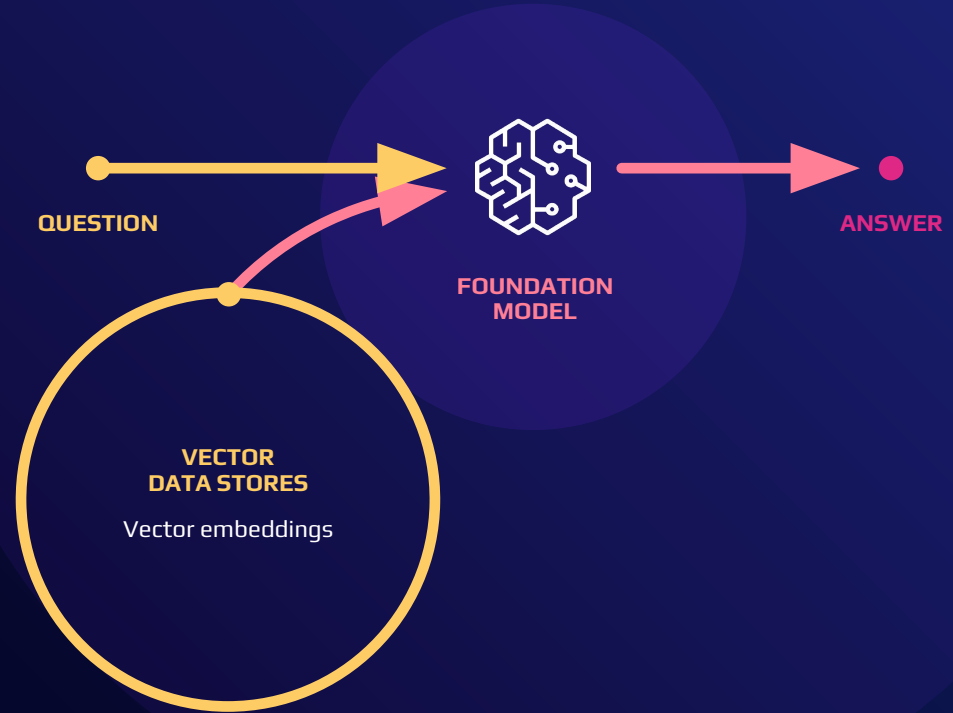
So are the
prompts



A comprehensive set of services for your data foundation



Vector embeddings support RAG



Enabling vector support across our services



**AMAZON OPENSEARCH
SERVICE**



**AMAZON AURORA
POSTGRESOL**



**AMAZON RDS FOR
POSTGRESOL**

Storing vectors requires
dedicated resources and expertise



NEW

Vector engine for OpenSearch Serverless

Simple, scalable, and high-performing vector storage and search without managing infrastructure



Store and query embeddings alongside your business data



Enable more relevant similarity search



Scale easily and efficiently



Build with publicly available foundation models

AVAILABLE ON SAGEMAKER JUMPSTART

AI21 labs

Models

Jurassic-2 Ultra, Mid
Contextual answers

Summarize
Paraphrase
Grammatical error
correction

Tasks

Text generation
Long-form
generation
Summarization
Paraphrasing
Chat
Information
extraction

Meta AI

Models

Llama 2 7B, 13B, 70B

Tasks

Question answering
Chat
Summarization
Paraphrasing
Sentiment analysis
Text generation

cohere

Models

Cohere
Command XL

Tasks

Text generation
Information
extraction
Question answering
Summarization

Hugging Face

Models

Falcon-7B, 40B
Open LLaMA
RedPajama
MPT-7B
BloomZ 176B
Flan T-5 models (8 variants)
DistilGPT2
GPT NeoXT
Bloom models
(3 variants)

Tasks

Machine translation
Question answering
Summarization

stability.ai

Models

Stable Diffusion XL 1.0
2.1 base
Upscaling
Inpainting

Tasks

Generate photo-realistic
images from text input
Improve quality of
generated images

Features

Fine-tuning on Stable
Diffusion 2.1 base model

LightOn

Models

Lyra-Fr
10B, Mini

Tasks

Text generation
Keyword extraction
Information extraction
Question answering
Summarization
Sentiment analysis
Classification

databricks

Models

Dolly

Tasks

Question answering
Chat
Summarization
Paraphrasing
Sentiment analysis
Text generation

alexa

Models

AlexaTM 20B

Tasks

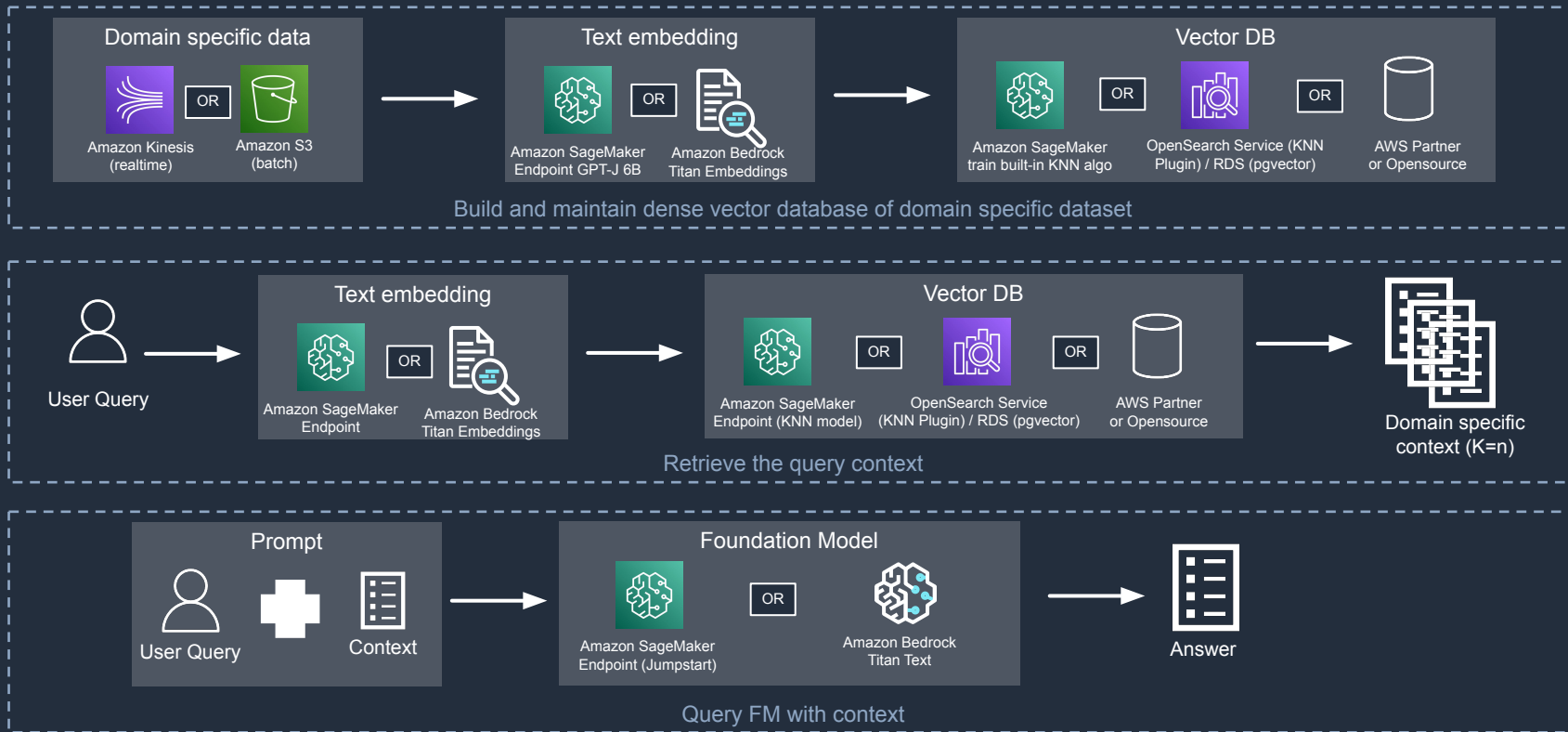
Machine translation
Question answering
Summarization
Annotation
Data generation



Reference Architectures

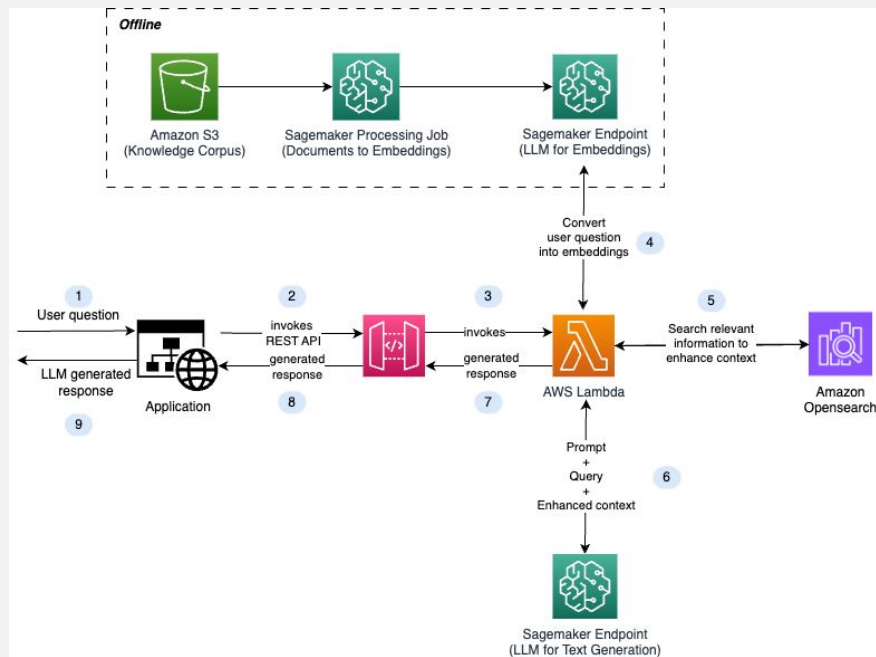
Retrieval-Augmented Generation (RAG)

REFERENCE ARCHITECTURE



Retrieval-Augmented Generation on AWS

WITH OPENSEARCH AND LANGCHAIN



The architecture diagram includes the following AWS services:

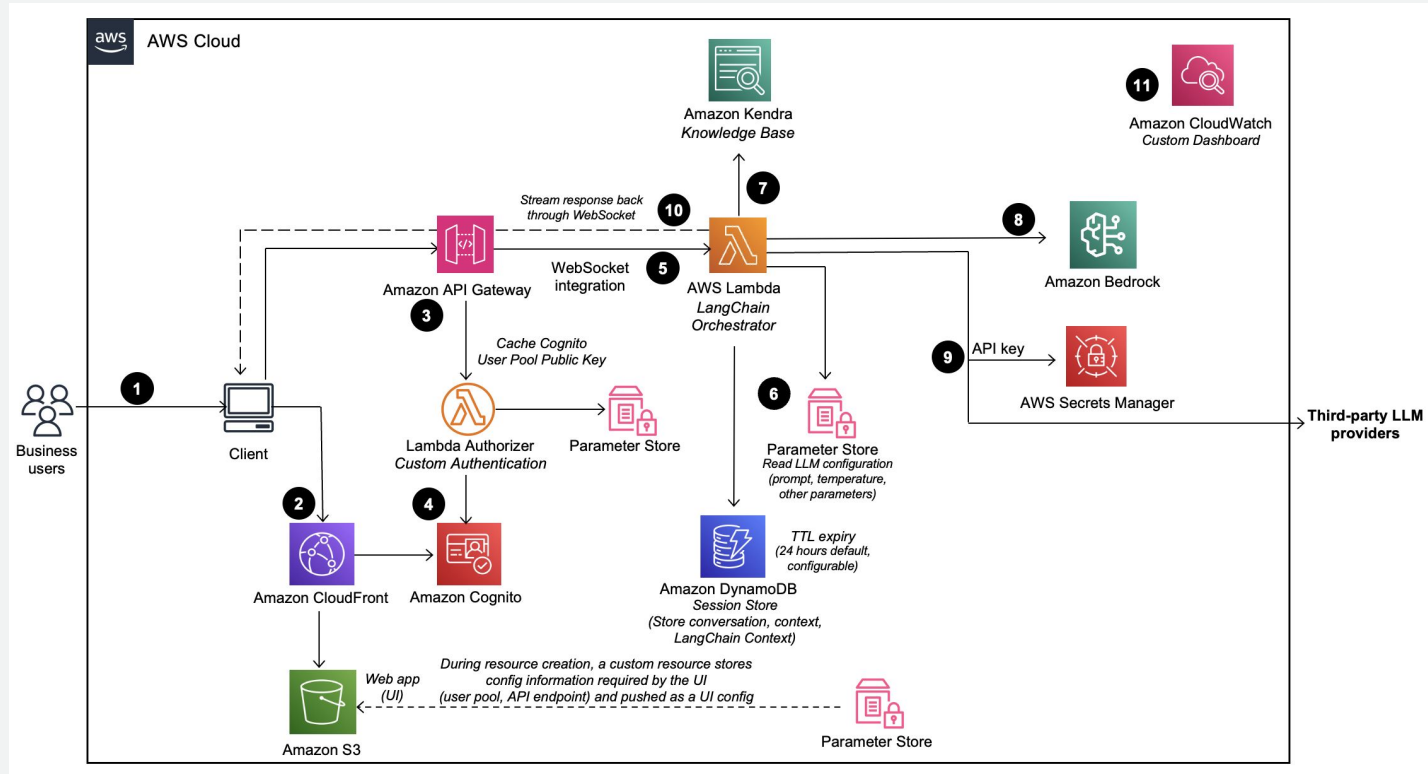
- [SageMaker](#) and [Amazon SageMaker JumpStart](#) for hosting the two LLMs (GPT-J for embeddings & Llama 2 for generation).
- [OpenSearch Service](#) for storing the embeddings of the domain knowledge corpus and doing similarity search with user questions.
- [Lambda](#) for implementing the RAG functionality and exposing it as a REST endpoint via the [API Gateway](#).
- [Amazon SageMaker Processing jobs](#) for large scale data ingestion into OpenSearch.

Additionally, we use:

- [LangChain](#) for interfacing with OpenSearch Service and SageMaker.
- [FastAPI](#) for implementing the REST API interface in the Lambda

Generative AI Application Builder

PUSH BUTTON DEPLOYMENT



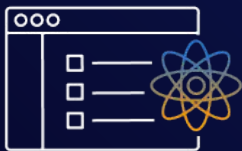
Introducing the Generative AI Application Builder on AWS

Facilitates the development, rapid experimentation, and deployment of generative AI applications, without requiring deep experience in artificial intelligence



Choice & Configurability

Connect to any model or data sources with pre-built connectors



Rapid Experimentation

Compare performance and outputs using a simple interface



Production-Ready

Deploy with production-ready security & scalability

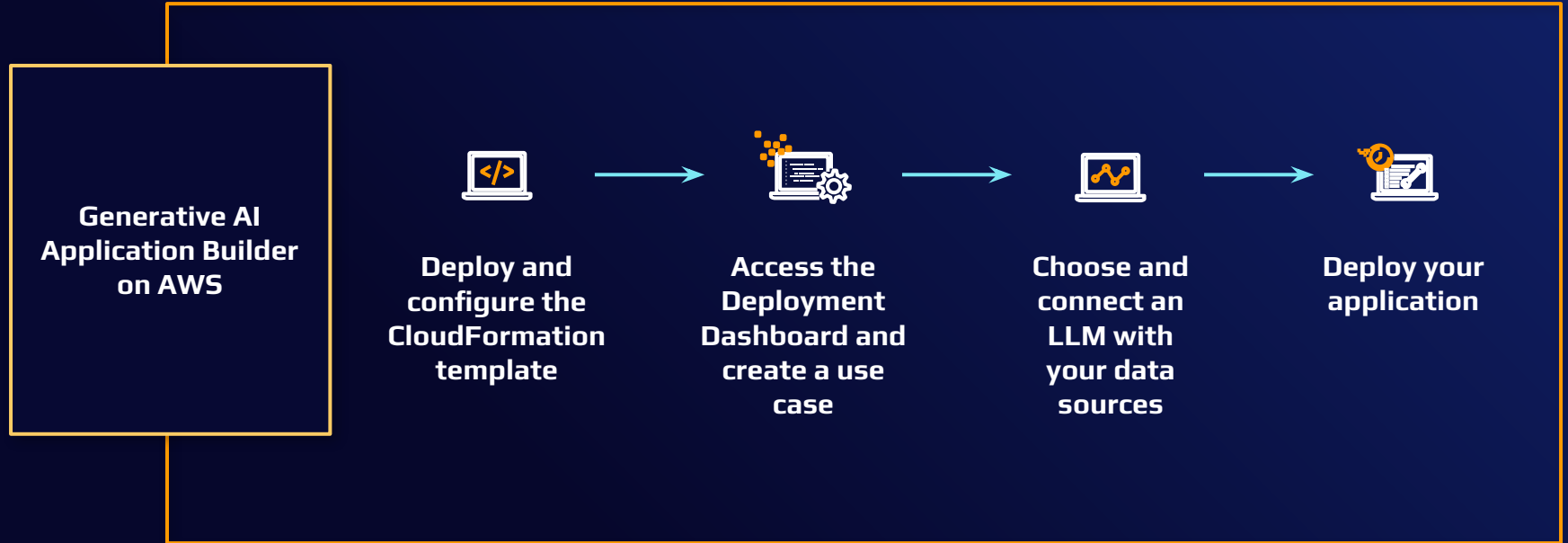


Extensible Modular Architecture

Extend and modify the solution to fit your specific use case



How it works



Generative AI Application Builder integrates Amazon Bedrock and its leading foundation models

amazon

Amazon Titan

Text summarization, generation, classification, open-ended Q&A, information extraction, embeddings, and search

AI21labs

Jurassic-2

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

ANTHROPIC

Claude 2

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

cohere

Command

Text-generation model for business applications and embeddings model for search, clustering, or classification in 100+ languages

∞ Meta

Llama 2 *(coming soon)*

Fine-tuned models ideal for dialogue use cases and language tasks

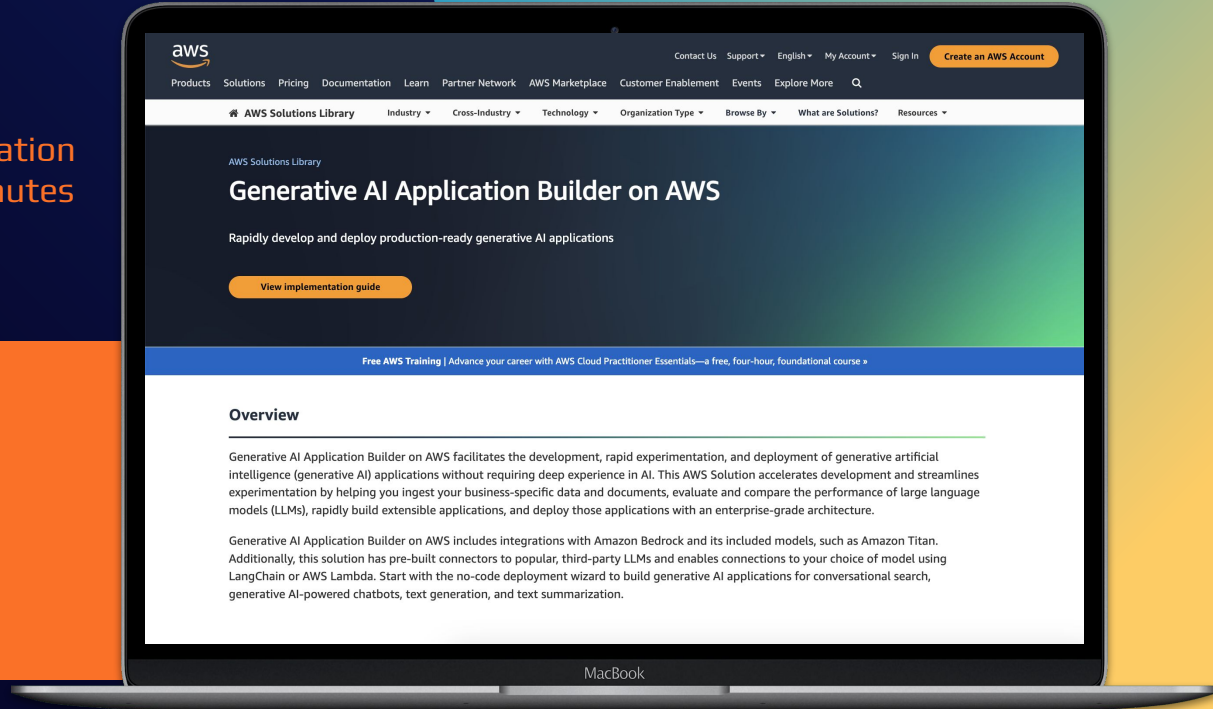


Get Started!

Deploy a complete generative AI application customized with your data in a few minutes



aws.amazon.com/solutions/implementations/generative-ai-application-builder-on-aws/



Demo



Thank you!

Chandan Rajah

crajah@amazon.com

<https://www.linkedin.com/in/crajah>

Jenny Vega

jennyvg@amazon.com

Startup Bootcamp: AWS Programs

Learn about how AWS supports startups through funding, credits, technology, go-to-market and more.

Warsaw, Poland | November 9th 2023



Hi, I'm Nina

- AWS Startups Account Executive
- Focused on supporting early-stage startups

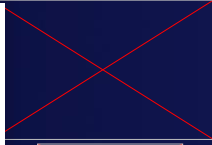


Why startups build on AWS

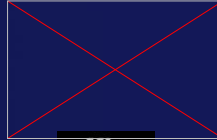



All of our infrastructure runs on AWS. We selected AWS because it's a really solid platform to build on and to scale with. We're confident that whatever size we reach, AWS will be there to support us.

Bastiaan Terhorst
Chief Product Officer



 **JUST EAT Takeaway.com**



 **bynder**



 **asana**



 **affirm**

 **Airtable**

 **Expect Me**

atrium

 **Auth0**



braze

BuzzFeed



Casper

 **circleci**

classpass

COHESITY

CONVOY

 **CROWDSTRIKE**


 **Dataminr**

dataxu.



drive.ai



 **INTERCOM**



 **druva**

coinbase



 **LaunchDarkly**



Namely




entelo

AYASDI



Grab

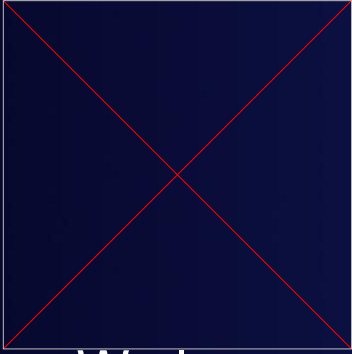


 **monzo**

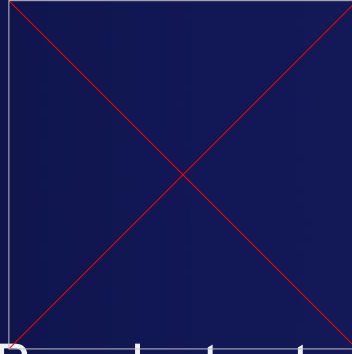
Betterment

 **CONSENSYS**

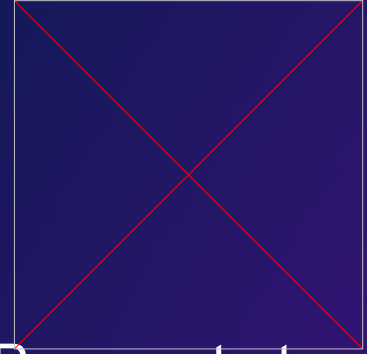
Why startups choose AWS



We love startups



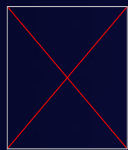
Broadest set of services



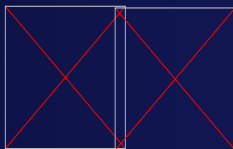
Proven startup programs for growth

AWS Startup Universe

Dozens of programs and a dedicated organization for Startups



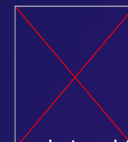
Idea



Build **OR** Move to



Grow and Scale



Co-market and Co-sell

Startup Loft Accelerator (FREE)

Equity-free program focused on supporting early stage startups.

APPLY [HERE](#)

Startup Loft Trainings (FREE)

Startup event hub covering topics like pitching, networking etc.

JOIN [HERE](#)

Activate Founders (FREE)

\$1000 Credits with Support credits for startups who are just getting started

APPLY [HERE](#)

AWS Activate Program (FREE)

Program specifically designed for early stage entrepreneurs to build on AWS with up to \$100k credits.

APPLY [HERE](#)

AWS Activate Provider (FREE)

Ask your investor or startup enabling organization to participate in AWS Activate.

APPLY [HERE](#)

Partner Consultation (FREE)

Get an architecture review and statement of work from our AWS Advanced Partners.

Partner Funding (Co-funded)

Work with AWS Advanced Partners to save you time and money, and AWS can co-fund.

Startup Migrate Program (Funded)

If you are moving workloads to AWS, we have programs that can help you. You will also get a dedicated account manager.

CONTACT US [HERE](#)

Ask a Technical Expert (FREE)

Speak to an AWS technical expert and get free technical advice.

BOOK A SLOT [HERE](#)

Business or Developer Support

Enhanced technical support and architectural guidance that comes with your Activate Credits.

Learn more [HERE](#)

Cost Optimization (FREE)

Take control of cost and continuously optimize your spend, while building modern, scalable applications to meet your needs.

READ [HERE](#)

Architecture Optimization

Explore advanced services to allow for even better scale and performance.

CONTACT US [HERE](#)

Cost Optimization (FREE)

Join our monthly Cost Optimization Webinar

JOIN [HERE](#)

Well Architected Review (FREE)

Build secure, high-performing, resilient, and efficient infrastructure for free.

TRY IT [HERE](#)

Training & Certifications

Upskill and get free technical trainings.

LEARN [HERE](#)

Recruitment

Get the talents you need through our partners

AWS Partner Network (FREE)

Prospect and get business with AWS via Amazon Partner Network. Registering is free.

APPLY [HERE](#)

AWS Marketplace

Sell your solution to our customers via the AWS Marketplace.

APPLY [HERE](#)

AWS Data Exchange

Find, subscribe to, and use third-party data in the cloud with AWS Data Exchange.

APPLY [HERE](#)

Co-brand with AWS

Speak in AWS events, co-create marketing materials and build better branding with AWS.

CONTACT US [HERE](#)



AWS Startup Journey

FEATURED PROGRAMS & SUPPORT



EARLY STAGE

Build Differentiated Customer Offerings



FUNDING & NETWORKING

AWS Startup Loft Accelerator
AWS Startup Business Development
AWS CTO Fellowship



PRODUCT DEVELOPMENT

AWS Startup Loft Accelerator
Startup Loft Trainings
AWS CTO Fellowship



CREDIT & FINANCIAL SUPPORT

AWS Activate Founders

+ DISCOVER MORE PROGRAMS



GROWTH STAGE

Expand Your Customer Base



CUSTOMER ACQUISITION

AWS Marketplace
AWS Partner Network (APN)



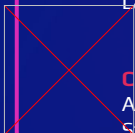
FUNDING & NETWORKING

AWS Startup Events
AWS Startup BD VC Network



PRODUCT ACCELERATION

CTO Fellowship
AWS Well-Architected
Startup Spotlight Program
(invite-only)
Startup Migrate
Labs and Workshops



CREDIT & FINANCIAL SUPPORT

AWS Activate Portfolio
Startup Migrate Program
Innovation Sandbox

+ DISCOVER MORE PROGRAMS



SCALE STAGE

Take Advantage of The Global Reach



GEO & MARKET EXPANSION

AWS Partner Network (APN)
ISV Accelerate
AWS Co-Sell Team
Joint Marketing Campaigns
Enterprise Support



NETWORKING

AWS Global Sponsorship Program >

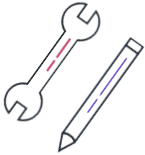


CREDIT & FINANCIAL SUPPORT

AWS Activate >
Training Discounts and Credits

+ DISCOVER MORE PROGRAMS

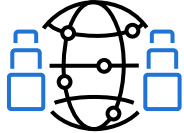
Startup Needs



**PRODUCT
DEVELOPMENT**

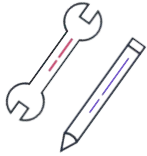


**FINANCIAL
SUPPORT**



**MARKET
EXPANSION**

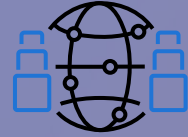
Startup Needs



**PRODUCT
DEVELOPMENT**

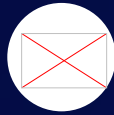


**FINANCIAL
SUPPORT**



**MARKET
EXPANSION**

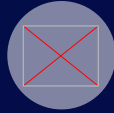
PRODUCT DEVELOPMENT



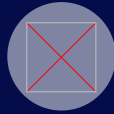
Ask an Expert



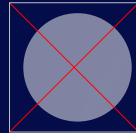
To connect 1:1 with **AWS Startup Solutions Architects** go to **AWS Startup Loft** and navigate to **Ask an Expert**



Labs & Workshops

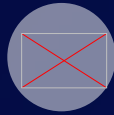


Partner Programs



Startup Loft Accelerator

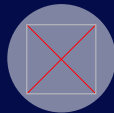
PRODUCT DEVELOPMENT



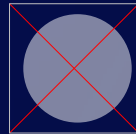
Ask an Expert



Labs & Workshops →



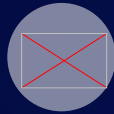
Partner Programs



Startup Loft Accelerator

For 1:many live sessions on specific technical topics go to **AWS Startup Loft** and navigate to *Events* or visit **AWS Workshops**

PRODUCT DEVELOPMENT



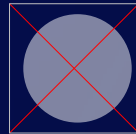
Ask an Expert



Labs & Workshops



Partner Programs 



Startup Loft Accelerator

Partners can help you in **Well Architected Reviews, Proof of Concepts, Migrations, Managed Services** and more, while AWS helps cover part of the costs

Jumpstart Partner Program

Get the Expertise You Need

Eligibility:

- For customers who want to adopt AWS strategic services (i.e. Analytics, IoT, ML, Containers, Non-Relational DB) but want it implemented fast or have technical knowledge or resource limitations
- Activate customers can apply for Jumpstart

What You Get:

- FREE initial engagement with Partner (SoWAAS), who can give you a Statement of Work
 - Statement of Work: contains plan for implementation and potential savings
 - Jumpstart will fund 50% of SOW up to 3.5x MRR with any remainder to be covered by the customer.
-

HOW?



Contact us [HERE](#)

Partner Proof of Concept

Business As Usual

Purpose:

- For any project with a partner (for GenAI we have a special program that funds more)

What You Get:

- 10% of ARR up to \$25k

HOW?



Contact us [HERE](#)

Well Architected Partner Program (WAPP):

Build Powerful Solutions / Optimize

Purpose:

- For customers who want to improve production workload using the AWS Well-Architected Framework (WAF)
- Map architecture against design principles and architectural best practices
- Check 6 pillars: operational excellence, security, reliability, performance efficiency, and cost optimization, sustainability

What You Get:

- FREE (USD 5,000 AWS Credits)**

Conditions**

Resolve a minimum of 25% of all HRIs (High Risk Issues) identified through the review (example, if 16 issues are found, you must resolve a minimum of 4)

Give feedback about the performance of the partner and by filling out a customer survey that will be sent by AWS

HOW?



Contact us [HERE](#)

Startup Migration Program

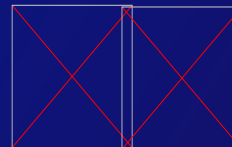
Make Your Migration Seamless

Eligibility:

- For customers who are trying to migrate workloads from any external platform
- Migrating from competing platform, on-premise, colocation, other cloud providers
- Estimated AWS spend > \$5k MRR

What You Get:

- Up to 4 months of credits + cash (max 1.8x MRR in partner cash)
- Business Support Plan credits valid for 1 year
- Architecture Mapping by AWS Solution Architect
- Cost Optimization



SaaS Discovery

Business As Usual

Purpose:

- Establish a SaaS in AWS with a partner

What You Get:

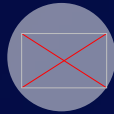
- \$10k cash funding for SaaS Discovery phase performed by SaaS Competency partner

HOW?



Contact us [HERE](#)

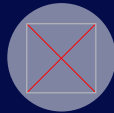
PRODUCT DEVELOPMENT



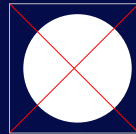
Ask an Expert



Labs & Workshops



Partner Programs



**AWS
Accelerator
s**



AWS Startup Accelerators provide innovators from all walks of life with access to **resources, technologies, and expertise**, to help them develop their ideas into **fully-formed solutions**.

More than 14 accelerators



Build Accelerator



AWS Education Accelerator



AWS Generative AI Accelerator



AWS Generative AI Accelerator ANZ

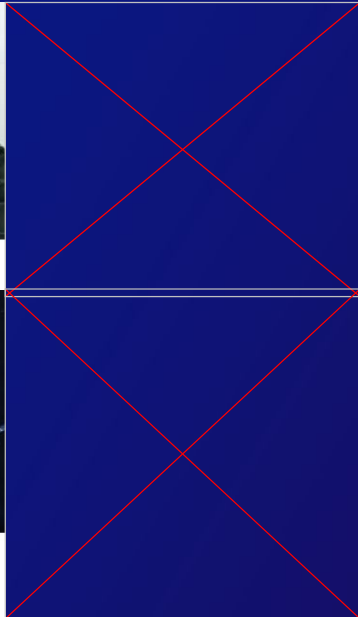


AWS ML Elevate: Generative AI Edition

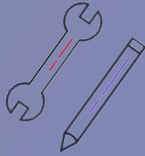


Space Accelerator

GLOBAL GROWTH-STAGE



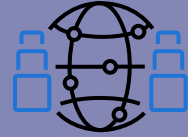
Startup Needs



**PRODUCT
DEVELOPMENT**

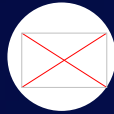


**FINANCIAL
SUPPORT**



**MARKET
EXPANSION**

FINANCIAL SUPPORT



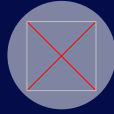
Startup Business
Development
Team



Get advice on **fundraising**,
business strategy, or deep
verticalized support



Investor/Startup
Events

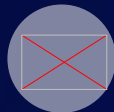


AWS Activate Credits



Cost Optimization

FINANCIAL SUPPORT



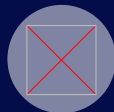
Startup Business
Development
Team



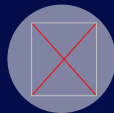
Investor/Startup
Events



Join global and **local events**
to **meet with investors**

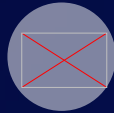


AWS Activate Credits



Cost Optimization

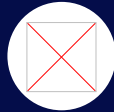
FINANCIAL SUPPORT



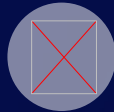
Startup Business
Development
Team



Investor/Startup
Events

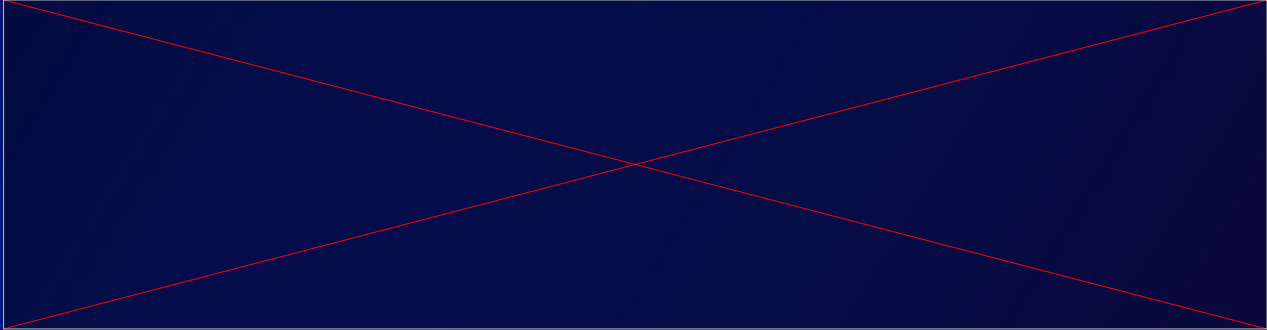


AWS Activate Credits 



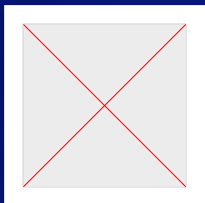
Cost Optimization

Access free tools, resources, content, expert support and apply for **up to \$100,000 in AWS Activate credits** to help cover AWS service usage



FINANCIAL SUPPORT

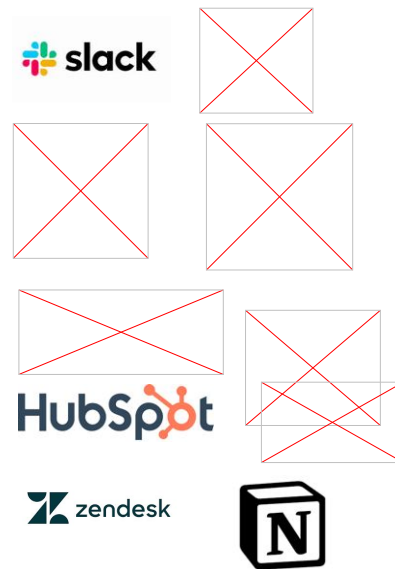
AWS Activate



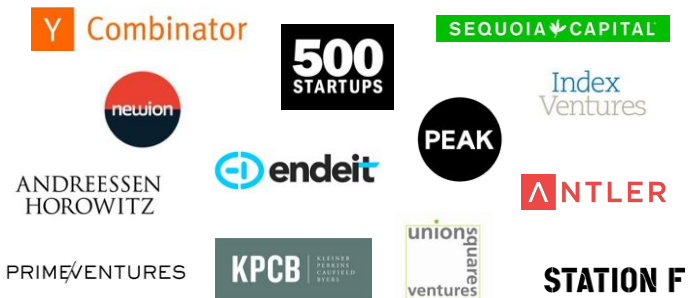
Benefits

- AWS promotional credits \$1,000 > \$100,000
- **Business Support** credits
- Online **self-paced lab** credits
- **Office hours**

Perks



Providers



AWS Activate

We want to help startups build through credits, technical and training support. You can focus on your growth!

Activate Founders

Eligibility:

- For early-stage startups with no institutional funding
- Self-funded, unbacked or bootstrapped
- Have not previously received credits
- Active AWS Account and fully functioning website

What You Get:

AWS Credits: \$1,000 USD / 2 years

Can be used for Developer Support Plan

Activate Portfolio

Eligibility:

- Funded Startup (selected venture capital firms, accelerators, incubators, universities, Angels, or other Startup Organizations)
- Do not exceed \$100k USD in redeemed credits
- Active AWS Account and fully functioning website

What You Get:

AWS Credits: Up to \$100k USD with 1 or 2 year validity

Can be used for Business Support Plan

Apply for Activate [HERE](#)

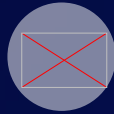
Watch an Introduction to Activate video [HERE](#)

Watch How to Apply AWS Activate video [HERE](#)

Developer vs Business Support details [HERE](#)



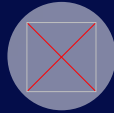
FINANCIAL SUPPORT



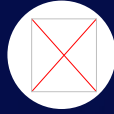
Startup Business
Development
Team



Investor/Startup
Events



AWS Activate Credits

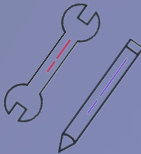


Cost Optimization



Access **Cost Optimization sessions, 1:1 deep dives** and **On Demand** content on the Startup Loft

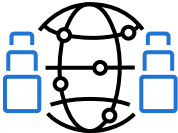
Startup Needs



**PRODUCT
DEVELOPMENT**



**FINANCIAL
SUPPORT**



**MARKET
EXPANSION**

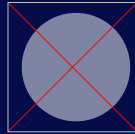
MARKET EXPANSION



AWS Marketplace
































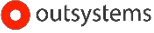


















Grow at scale by getting **access to millions of AWS customers** looking for innovative products



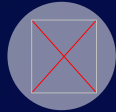
**AWS Partner
Network**
Accelerate

Decision-makers **cut their time in half** using AWS Marketplace compared to other sources

AWS Marketplace: 8 popular categories most often provisioned

DevOps	Security	Storage	Networking	Database	Business Apps	AWS Data Exchange	Professional Services / Resellers
							
							
							
							
							
							

MARKET EXPANSION



AWS Marketplace



AWS Partner
Network
Accelerate



Expand your customer reach, and **grow your business** with technical, marketing, and funding resources



Co-sell with AWS field sellers globally, who service millions of active AWS customers



STARTUP LOFT

Start. Connect. Grow



Exclusive
events



Ask an
Expert



Startup Loft
Accelerator



AWS ACTIVATE

Prove what's possible

Sign up to AWS Activate



Learn more ›

Q&A Session



Let's connect on LinkedIn!

Nina Trejos

AWS Account Executive



Maria Nunez

AWS Account Manager



Thank you!



Architecting for Success and Optimizing for Longer Runways

09. November 2023

PROJECT	CLUB RENT BERLIN
DRAWING NAME	CLIMB THE WALL
DESIGNED BY	OO STUDIO 劉偉傑 Lin-Fee Chang Chen Hsiao-Ton Wu Y-Lin Tsai-Pai Huan
SCALE	1 / 1
DATE	2009/6/4
SHEET NO	1
DRAWING NO	A1-1

Who am I?



Wojciech Orzechowski

Solution Architect – kreuzwerker

- 16 years in IT, 7 years with AWS workflows
- 9 AWS certificates
- AWS Champion Trainer for 6 years

Agenda

- **What** we are paying for
- **How** we optimize costs
- **Why** we do it

Worst practice example

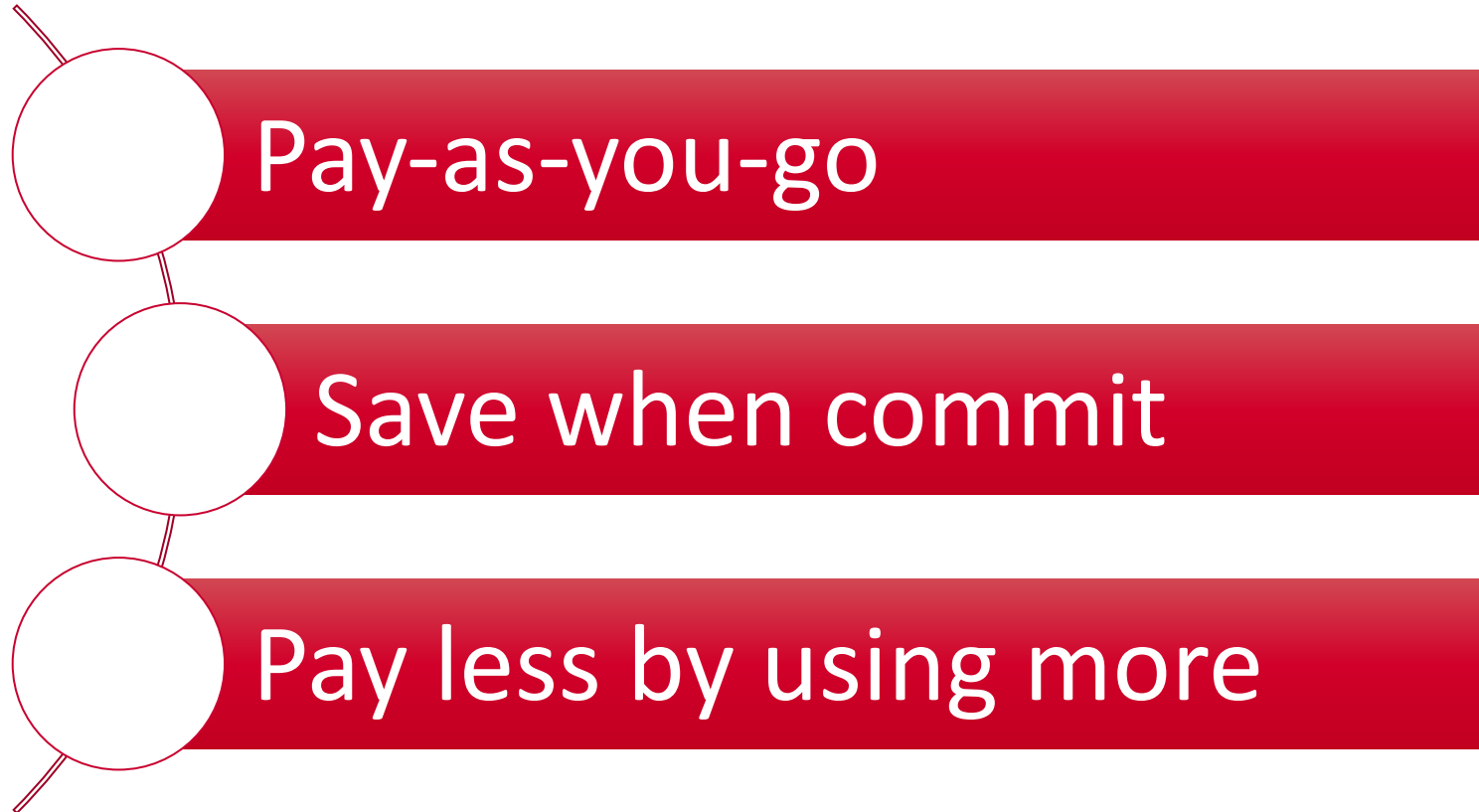
Team is working on a set of PoCs for data analytics workflow on AWS. They try different AWS services for data analytics – based on EC2, EMR also based on serverless services.

There is no AWS resource clean up after another PoC attempt. They use too many, not optimized, unneeded resources, in a wrong way. The cost of cloud resources was hidden inside the total cost of company's AWS infrastructure. PoC team has “sky is the limit” attitude towards costs and they were wasting more than \$20k per month.

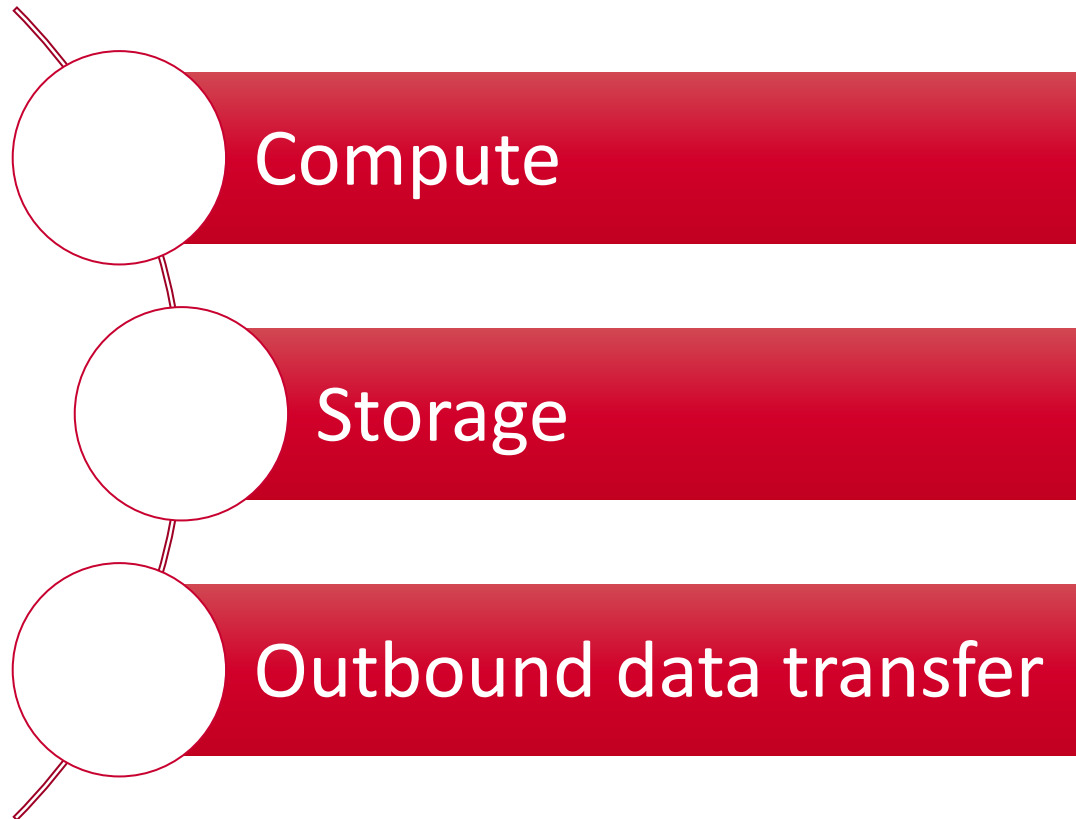
Sadly – a true story.

Cost drivers – **what** we are paying for

How we pay for AWS



Fundamental driver of cost with AWS



Amazon Elastic Compute Cloud (Amazon EC2)



AWS Lambda



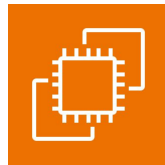
Amazon Elastic Block Store (Amazon EBS)



Amazon Simple Storage Service (Amazon S3)

Compute costs – EC2

- Pay per second for running specific VM type and size
- Pay-as-you-go = pay for VM only when VM is running
- EC2 = Only compute
- Block storage for EC2 = EBS – totally different AWS service



Amazon Elastic Compute
Cloud (Amazon EC2)



Amazon Elastic Block Store
(Amazon EBS)

Compute costs – EC2 – pricing model

On-Demand



Instances

Instance name ▲	On-Demand hourly rate ▼	vCPU ▼	Memory ▼
m6g.16xlarge	\$2.752	64	256 GiB

Spot

Instance Type ▲	Linux/UNIX Usage ▼
m6g.16xlarge	\$1.0987

Reserved

Instance name ▲	RI upfront fee ▼	RI monthly fees* ▼	RI effective hourly rate** ▼	Savings over On-Demand ▼	On-Demand rate ▼
m6g.16xlarge	\$14,010	\$0.00	\$1.599	42%	\$2.7520

Standard Reserved Instance, Ireland, 1 year, All Upfront

Compute costs – EC2 – saving plans

Compute Saving Plan

Compute without instance family or AWS region limitations

Instance name ▲	Savings Plans rate ▼	Savings over On-Demand ▼	On-Demand rate ▼	Region ▼	Operating system ▼
m6g.16xlarge	\$2.0792	24%	\$2.752	Europe (Ireland)	Linux

EC2 instance Saving Plan

specific instance family in a specific AWS region

Instance name ▲	Savings Plans rate ▼	Savings over On-Demand ▼	On-Demand rate ▼	Operating system ▼	Tenancy ▼
m6g.16xlarge	\$1.7135	38%	\$2.752	Linux	Shared

S3 – Object Storage – price parts



Amazon Simple Storage Service (Amazon S3)

Storage

S3 Standard - General purpose storage for any type of data, typically used for frequently accessed data

First 50 TB / Month

\$0.023 per GB

S3 Glacier Deep Archive *** - For long-term data archiving that is accessed once or twice in a year and can be restored within 12 hours

All Storage / Month

\$0.00099 per GB

S3 – Object Storage – price parts



Amazon Simple Storage Service (Amazon S3)

Requests & Data Retrievals

	PUT, COPY, POST, LIST requests (per 1,000 requests)	GET, SELECT, and all other requests (per 1,000 requests)	Lifecycle Transition requests into (per 1,000 requests)	Data Retrieval requests (per 1,000 requests)	Data retrievals (per GB)
S3 Standard	\$0.005	\$0.0004	n/a	n/a	n/a
S3 Glacier Deep Archive ***	\$0.055	\$0.0004	\$0.055	See below	See below
Standard	n/a	n/a	n/a	\$0.11	\$0.02
Bulk	n/a	n/a	n/a	\$0.025	\$0.003

S3 – Object Storage – price parts



Amazon Simple Storage Service (Amazon S3)

Transfer

	Price
Data Transfer IN To Amazon S3 From Internet	
All data transfer in	\$0.00 per GB

Data Transfer OUT From Amazon S3 To Internet

AWS customers receive 100GB of data transfer out to the internet free each month, aggregated across all AWS Services and Regions (except China and GovCloud). The 100 GB free tier for data transfer out to the internet is global and does not apply separately or individually to AWS Regions.

EBS – Block Storage – price parts

Type



IOPS



Amazon Elastic Block Store
(Amazon EBS)

Throughput

Size

<input type="checkbox"/>	Service Name ▾	Status ▾	Upfront cost ▾	Monthly cost ▾	Description ▾	Region ▾
<input type="checkbox"/>	Amazon Elastic Block Store (EBS) 	-	0.00 USD	106.61 USD	GP3 - general purpose	Europe (Ireland)
<input type="checkbox"/>	Amazon Elastic Block Store (EBS) 	-	0.00 USD	17.20 USD	Cold HDD - cheapest storage	Europe (Ireland)

1TB, always on
GP3 – 3000 IOPS, 500MB/s

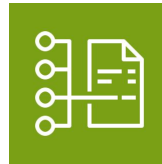
Methods – **how** we optimize

Cost structure visibility

- Use **Cost Allocation Tags**
- Get information regarding the cost of particular workflow, environment, application
- React automatically on your cloud expenses



AWS Budgets



AWS Cost &
Usage Report



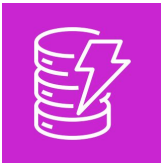
AWS Cost Explorer

Dynamic scaling based on current needs

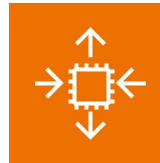
- No Overprovision !!!!
- Act dynamically in automated way
- Use serverless



Amazon Elastic Container Service (Amazon ECS)



Amazon DynamoDB



Amazon EC2 Auto Scaling



Amazon Elastic Kubernetes Service (Amazon EKS)

“Keep your desk clean“

- Clear ownership of resources inside our company : owner, department, project,..
- Remove unneeded resources
- Automate process of identify resource waste
- Make a clean-up as a one of last steps in PoC, test projects

Workflow environments

- Create whole environments to get the right-size, cost optimized resources like Virtual Machines, EBS drives, ...
- Identify resources that can be stopped – once for longer time, or with schedule
- To get the most of it : automate

Right Pricing Model

- Spot
- Reserved
- Savings Plans

Architecture optimization

- Use cache when possible
- Go serverless when possible
- Optimize costs early and make it an ongoing process

Motivation - **why** we optimize costs

Why we optime

- Spent less and only for needed resources
- Use saved money to launch another projects
- Increases security by eliminating unused resources
- Increased cost can be a signal of security issue

Summary

- Use only resources that you need
- Cost optimization is an ongoing process
- Everyone are responsible for it

Questions?



Get in touch with us!

info@kreuzwerker.pl

 **kreuzwerker**

kreuzwerker sp.z o.o.sp.k.
al. Jerozolimskie 85/21
02-001 Warszawa

www.kreuzwerker.pl
Fon +48 88 282 61 82

PROJECT
BLUEPRINT BERLIN
DRAWING NAME
CLIMB THE WALL
DESIGNED BY
OU STUDIO
Li Xiaohong 打開聯合
Chen Hsiao-Tsun
Wu Pei-Lin
Tsai-Pai Huang
SCALE
1/1
DATE 2009/6/4