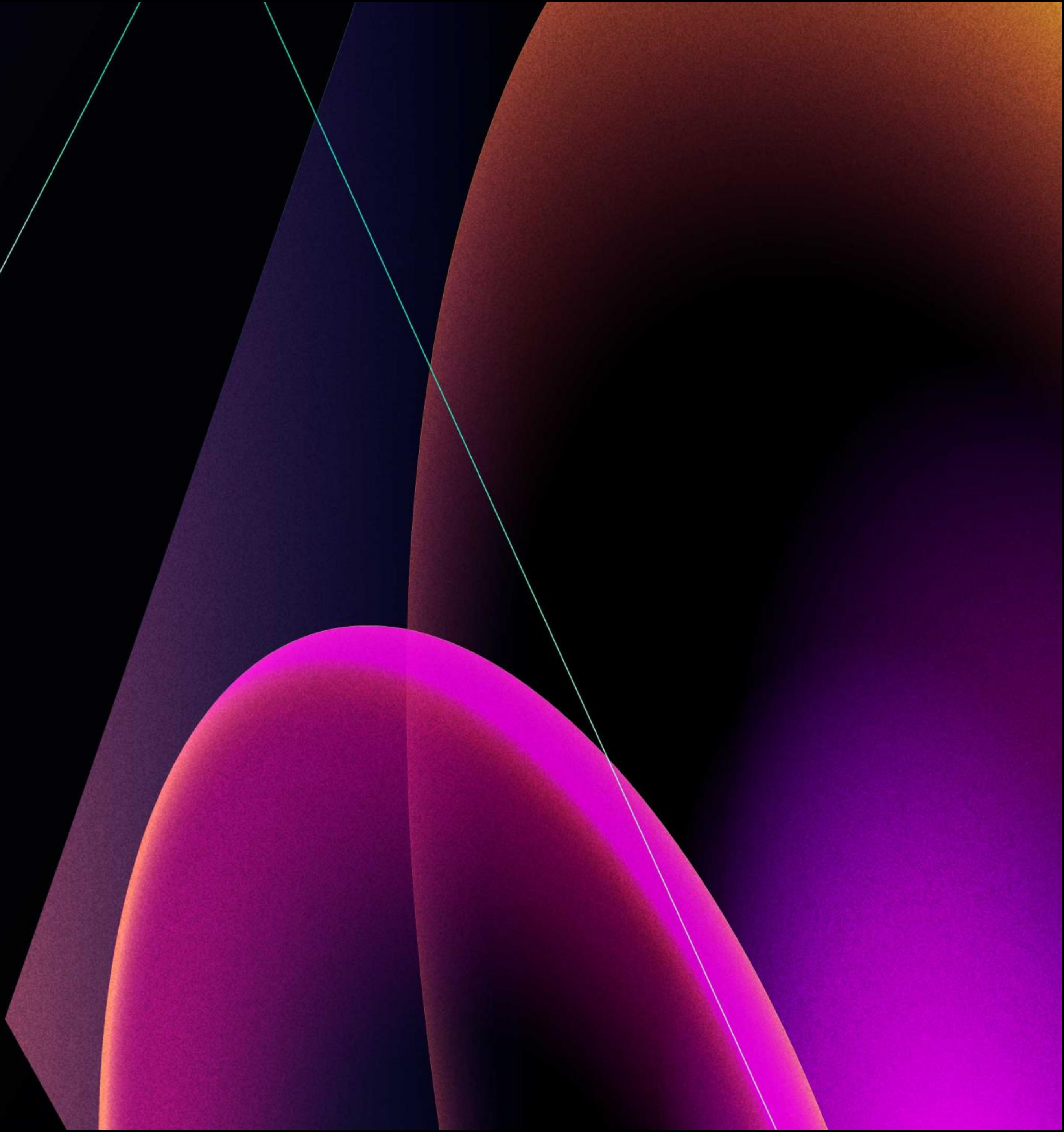
AWS re:Invent

DECEMBER 2 – 6, 2024 | LAS VEGAS, NV



DEV204

# Searching images through patterns: An AI-powered serverless solution

**Dhaval Nagar**

AWS Hero | Founder
APPGAMBiT

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Dhaval Nagar

**Dhaval Nagar**

AWS Serverless Hero
Founder, APPGAMBiT

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Agenda

01 Introduction

02 State of gen AI models

03 Recent advancements

04 Multimodal LLMs

05 Customer use case

06 Serverless AI-powered solution

07 Our learnings

08 Summary

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

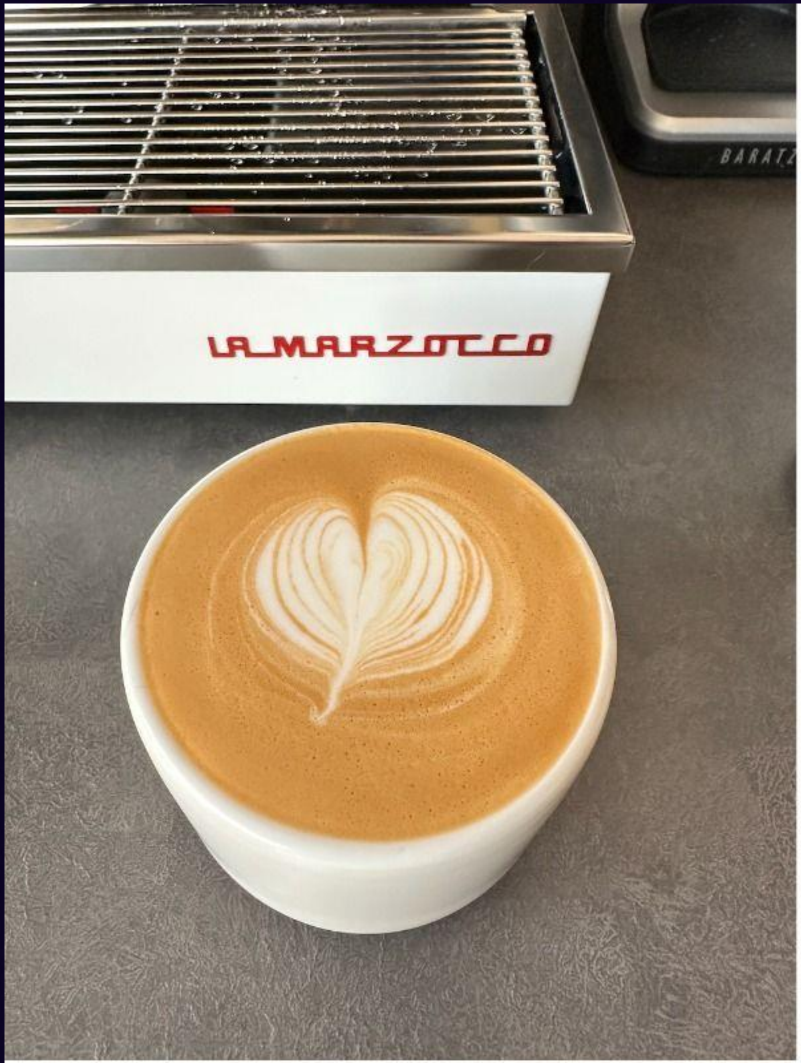
> “The true delight is in finding out rather than in the knowing.

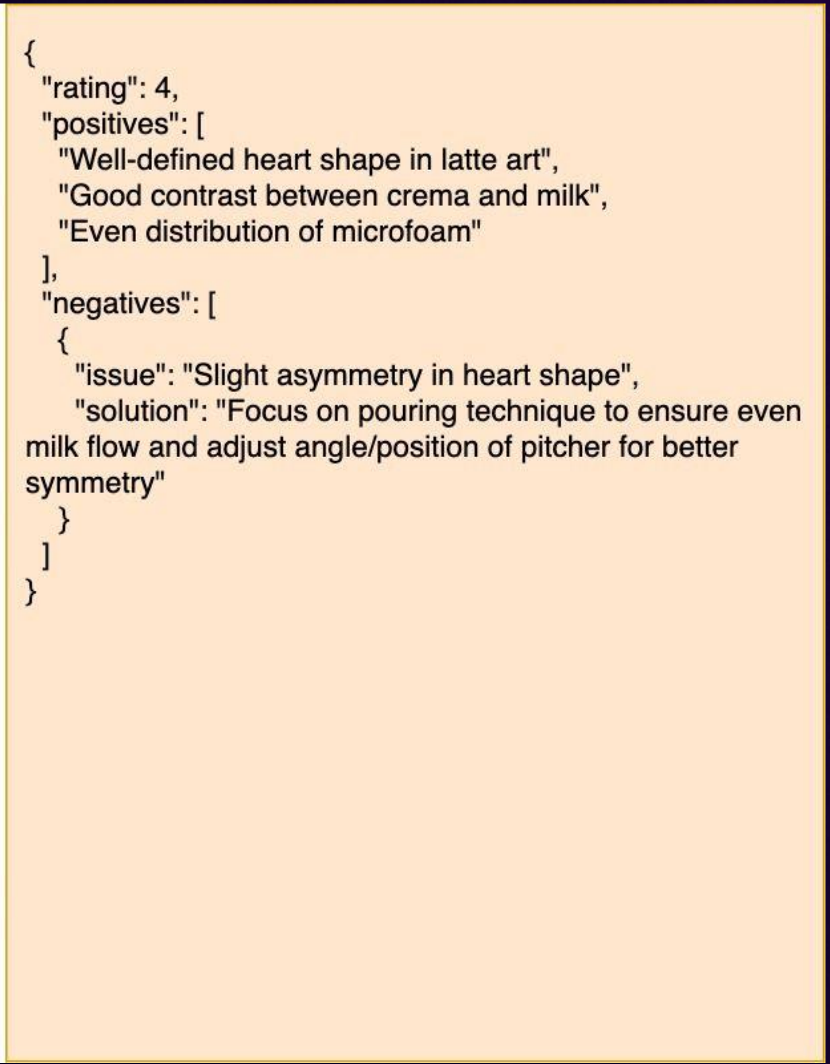
**Isaac Asimov**

Writer, Professor

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# It all started with this . . .





```
{
  "rating": 4,
  "positives": [
    "Well-defined heart shape in latte art",
    "Good contrast between crema and milk",
    "Even distribution of microfoam"
  ],
  "negatives": [
    {
      "issue": "Slight asymmetry in heart shape",
      "solution": "Focus on pouring technique to ensure even
milk flow and adjust angle/position of pitcher for better
symmetry"
    }
  ]
}
```

May 2024

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# The current state of large language models

| | OpenAI | Meta | Google | Anthropic | Cohere | Mistral | Amazon | Stability AI | Microsoft |
|---|---|---|---|---|---|---|---|---|---|
| 2024 | o1-preview<br>o1-mini<br>GPT-4o | Llama 3.2<br>Llama 3.1 | Gemini 1.5<br>Gemma | Haiku v2<br>Sonnet 3.5 v2<br>Sonnet 3.5<br>Claude 3<br>Haiku<br>Sonnet<br>Opus | Command R<br>Command R+ | Ministral 3B<br>Ministral 8B<br>Mistral Small<br>Mistral Large<br>Codestral<br>Mixtral 8x22b<br>Codestral<br>Pixtral | Titan Text Premier | SD 3.0 | Phi-3<br>Florence 2 |
| 2023 | GPT-4 | Code Llama<br>Llama 2<br>Llama | Gemini<br>PaLM 2 | Claude 2 | Command | Mixtral 8x7B / 7B | Titan Image Generator | XL Turbo<br>XL 1.0 | Phi-2<br>Phi-1 |

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Technology advancements

Better, faster GPUs

Improved dataset
and
training techniques

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Model advancements

Larger context window

Better reasoning and inference

Multi-modality support

Agentic workflow

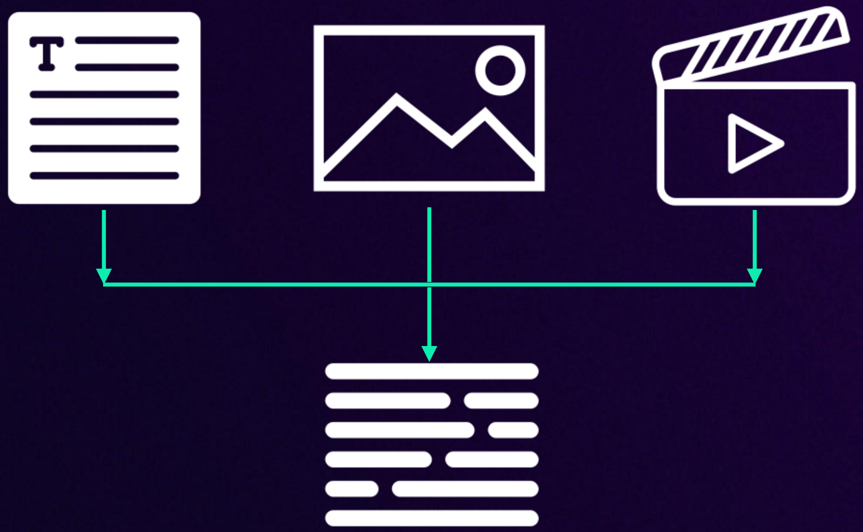
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Text models

Text-based models are designed to analyze input text (prompt) and generate new text information (tokens)

# Multimodal models

Multimodal models are designed to process information from different sources like text, image, audio, and video and generate new information (tokens)



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Multimodal large language models

- Larger context window
- Improved image understanding
- Extract information from visual contents like infographics, charts, figures, tables, slides, etc.
- Suitable for business use cases
- Zero-shot prompting

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Multimodal large language models

- Larger context window
- Improved image understanding
- Extract information from visual contents like infographics, charts, figures, tables, slides, etc.
- Suitable for business use cases
- Zero-shot prompting

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Why zero-shot prompting is important!

Extract information without previous examples

Faster time to implement and validate new use cases

Cost efficient

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Zero-shot prompting example

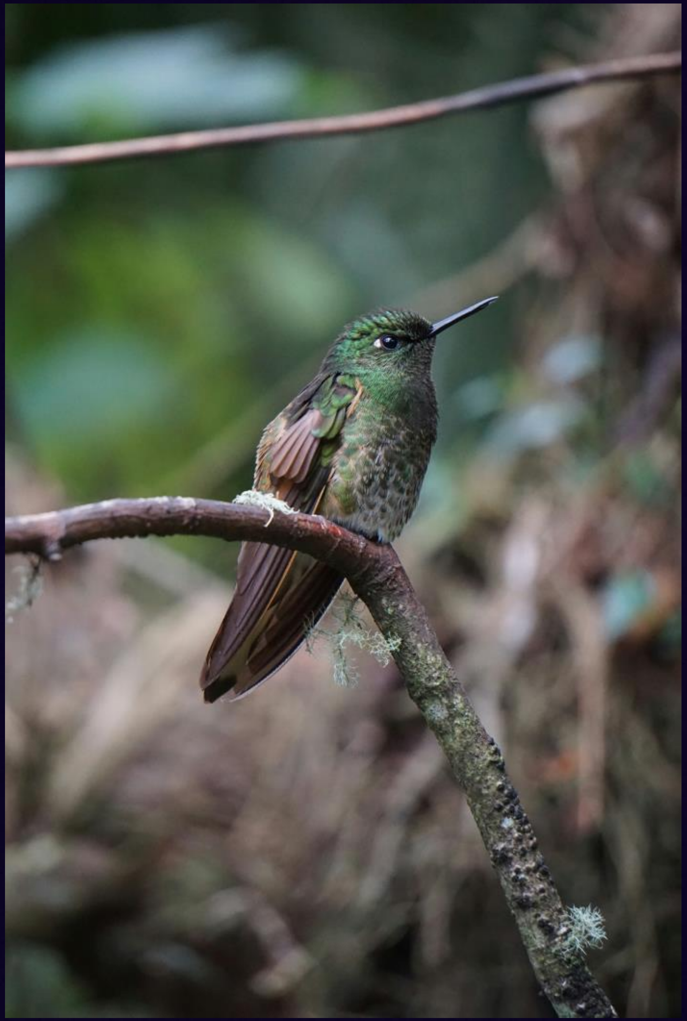


Photo by Cedric Fox on Unsplash

Model: Anthropic Haiku v1 on Amazon Bedrock
Q: Can you identify bird in this picture?

A: The bird in the image appears to be a hummingbird. The small, slender body with the long, thin beak are characteristic features of a hummingbird. The vibrant green and brown feathers also match the coloring of many hummingbird species. Hummingbirds are known for their incredible ability to hover in mid-air and dart rapidly between flowers, which seems to be what this hummingbird is doing on the branch.

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Zero-shot prompting example in Spanish

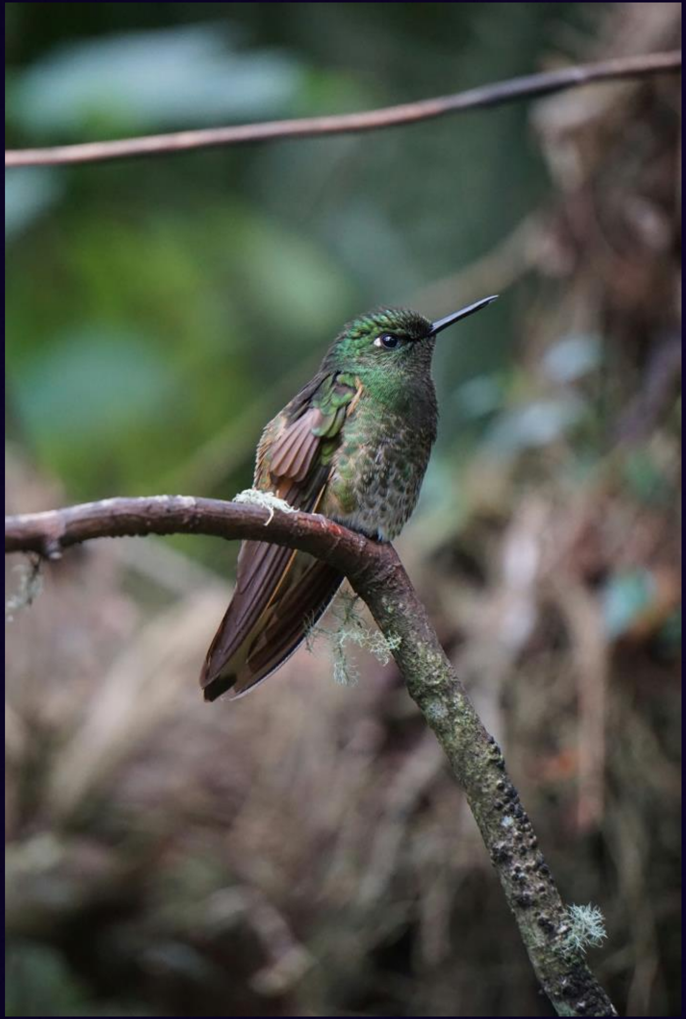


Photo by Cedric Fox on Unsplash

Model: Anthropic Haiku v1 on Amazon Bedrock
Q: ¿Puedes identificar el pájaro en esta imagen?

A: El pájaro en la imagen es un colibrí. Los colibríes son pequeñas aves que se caracterizan por su plumaje colorido y la capacidad de volar de forma muy ágil, batiendo sus alas rápidamente. En esta imagen, se puede apreciar el característico plumaje verde iridiscente del colibrí, así como su pico largo y delgado, típico de esta especie.

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Examples

- Visual question answering
- Visual interpretation
- Visual captioning
- Visual grounding

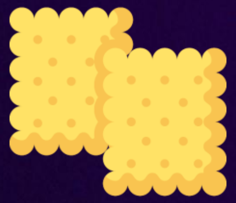
Input

Text

Question: Which property do these two objects have in common?

Options: A) soft B) salty

Vision



Crackers

Fries

Output

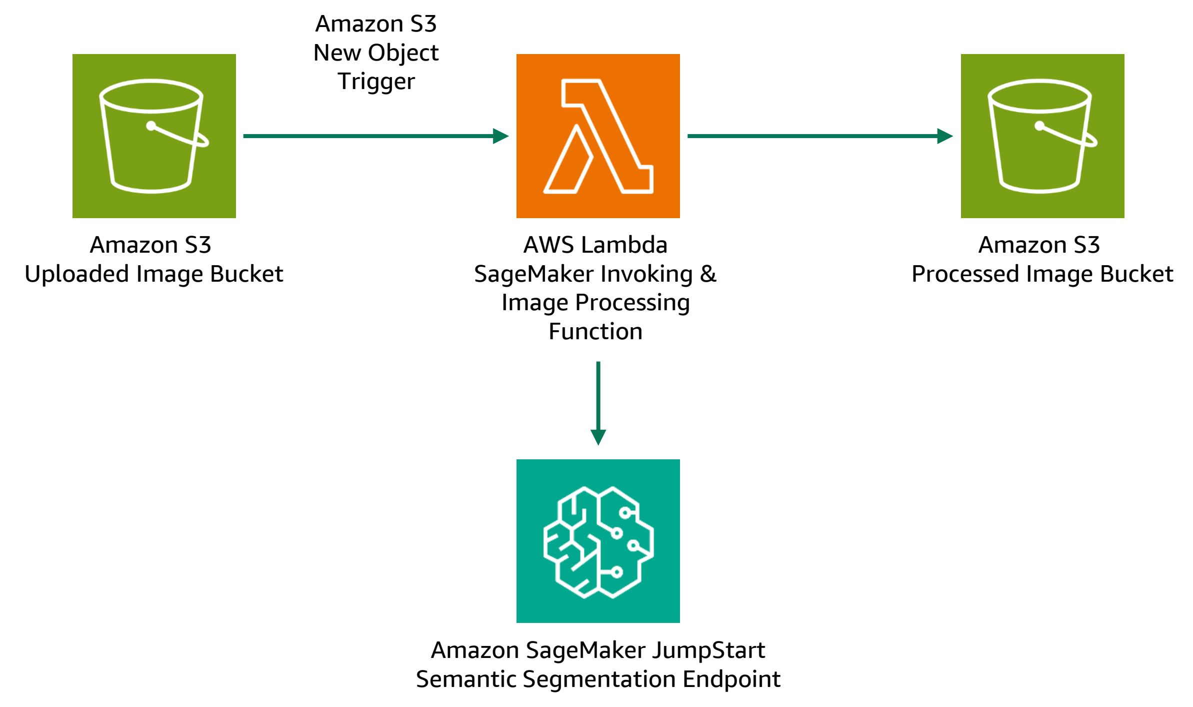
The property the two objects have in common is that they are both salty. The image on the left appears to show a salty cracker or biscuit, while the image on the right shows a serving of french fries, which are typically salted.

Output generated by Anthropic Claude 3 Haiku v1

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Examples

- Visual question answering
- Visual interpretation
- Visual captioning
- Visual grounding



Q: Describe the image.

A: Based on the diagram, the Uploaded Image Bucket and Processed Image Bucket serve different purposes in the image processing workflow:

1. Uploaded Image Bucket:
This is the initial storage location for images that are uploaded to the system. It's where new images are first placed, triggering the subsequent processing steps.

2. Processed Image Bucket:
This is the destination for images after they have been processed by the AWS Lambda function. It stores the results of the image processing and SageMaker operations.

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Examples

- Visual question answering
- Visual interpretation
- Visual captioning
- Visual grounding

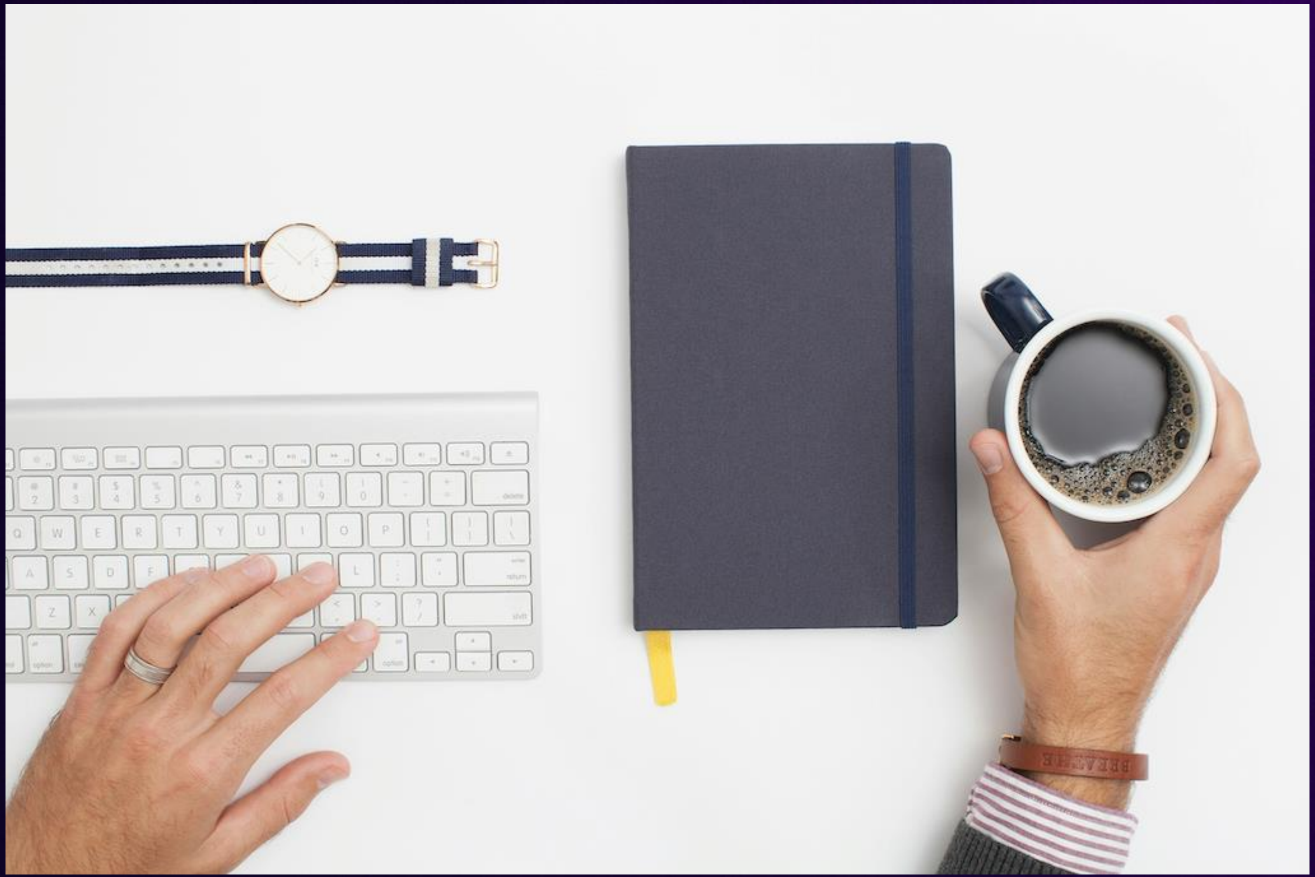


Photo by Cathryn Lavery on Unsplash

Q: Describe the image in one sentence.

A: A workspace scene featuring hands typing on a keyboard, holding coffee, with a watch, notebook, and other office essentials neatly arranged on a white surface.

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Output generated by Anthropic Claude 3 Haiku

# Examples

- Visual question answering
- Visual interpretation
- Visual captioning
- Visual grounding

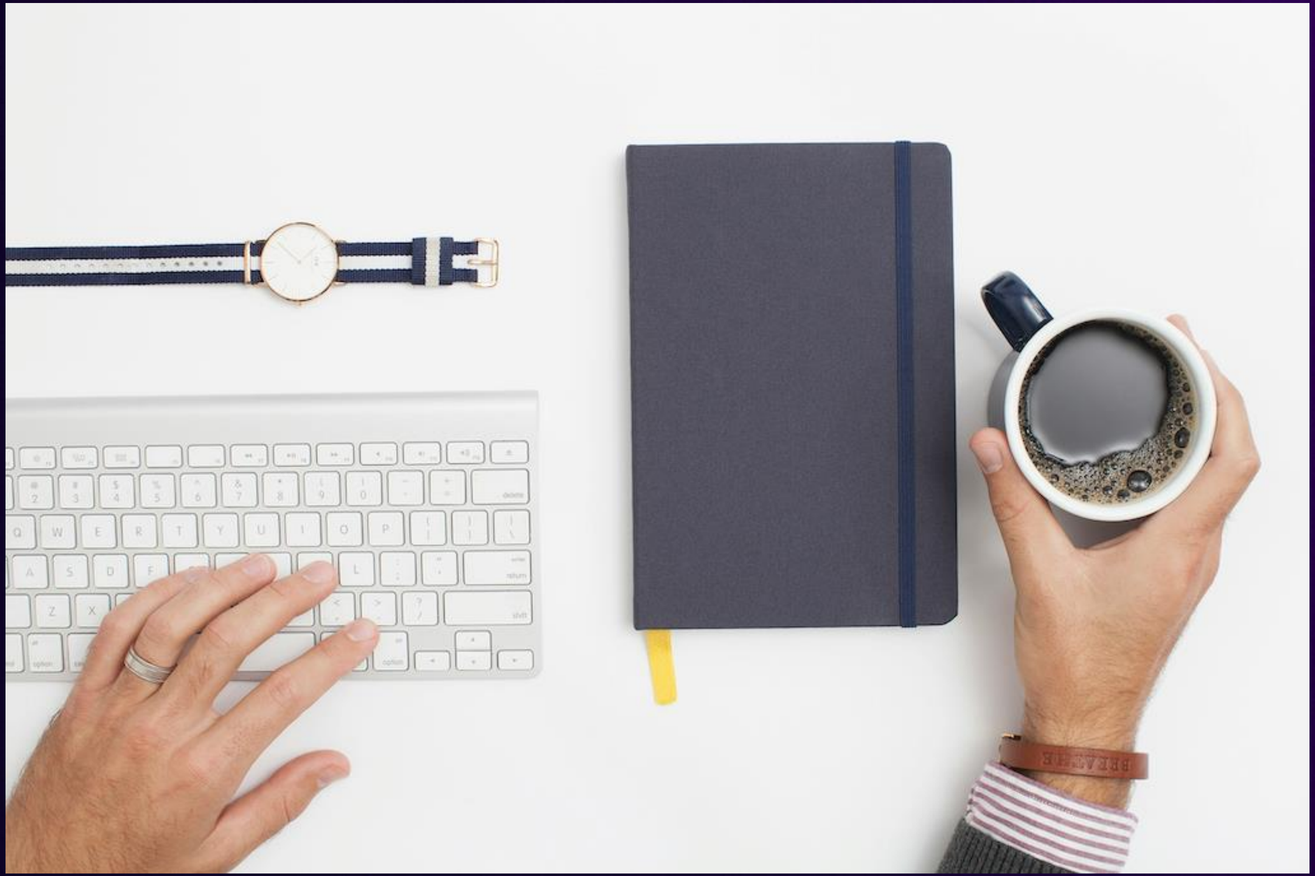


Photo by Cathryn Lavery on Unsplash

Q: Which side is the coffee cup and the wrist watch?

A: In the image, the coffee cup is on the right side, being held by a hand. The wrist watch is on the left side of the image, lying flat on the surface.

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Output generated by Anthropic Claude 3 Haiku

# Examples

- Visual question answering
- Visual interpretation
- Visual captioning
- Visual grounding

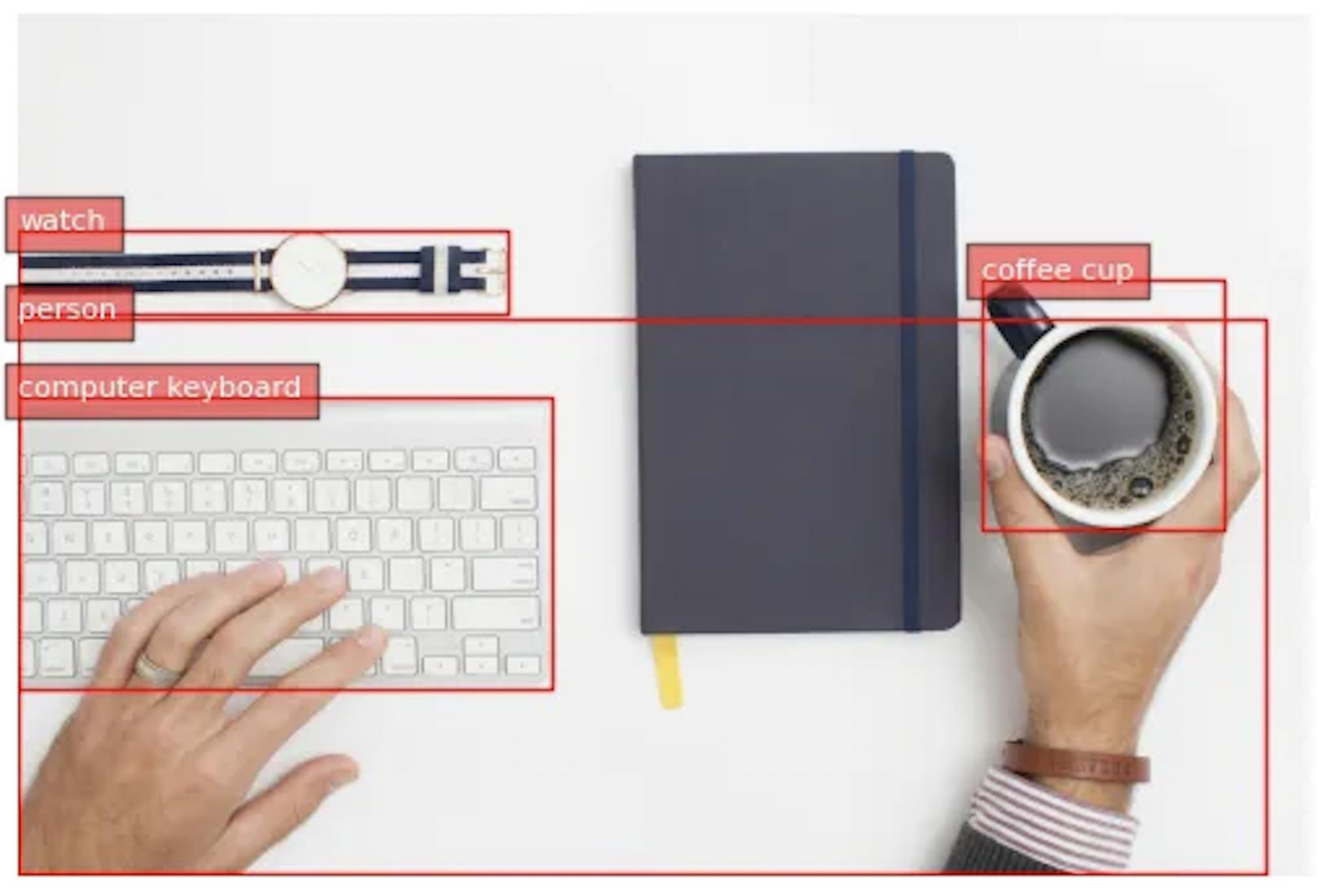


Photo by Cathryn Lavery on Unsplash

P: Caption + Grounding

A:
```
{ "<OD>": { "bboxes": [[1789.2000732421875, 492, 2238,
959.2000122070312], [1.2000000476837158, 711.2000122070312,
990.0000610351562, 1250.4000244140625], [1.2000000476837158,
567.2000122070312, 2319.60009765625, 1597.5999755859375],
[1.2000000476837158, 402.3999938964844, 910.800048828125,
556]], "labels": ["coffee cup", "computer keyboard", "person",
"watch"] } }
```

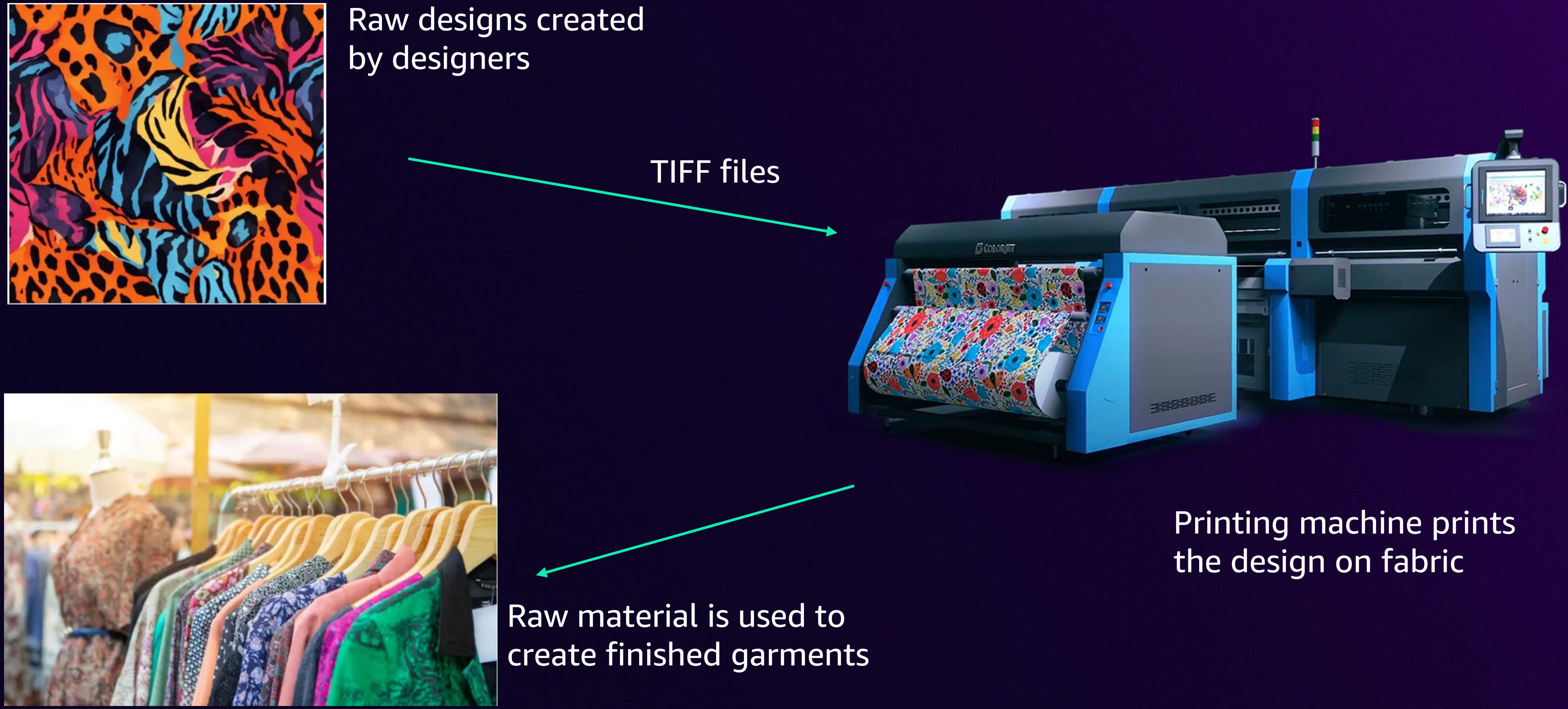
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Output generated by Microsoft Florence 2

# Customer use case

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Digital Fabric Printing House



Raw designs created by designers

TIFF files

Printing machine prints the design on fabric

Raw material is used to create finished garments

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Business process flow

Design theme identification

Search images for concept building

Create design mood boards

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Customer infrastructure

Large dataset of design files

Private storage with network access

Files are stored in RAW and preview format

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Resource inefficiency

- NAS is a file storage system
- Searching through NAS is a manual and human-driven process
  - Users manually search through file system and extract images
  - Time-consuming and error-prone
- Using online commercial resources
  - Paying for the images already available in local storage
  - Cost more due to time shortage

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# What we needed!

- Process images with low infrastructure burden
- Generate user-friendly image attributes
- Attribute-based search and retrieval

- Cost-efficient
- Low maintenance
- Pay-as-you-use model

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Initial challenges

Dataset with large file sizes

Not all design files are required

Unstructured dataset

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

> "Generative AI technology can be very effective in converting unstructured data into structured format.

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# AI-powered serverless solution

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

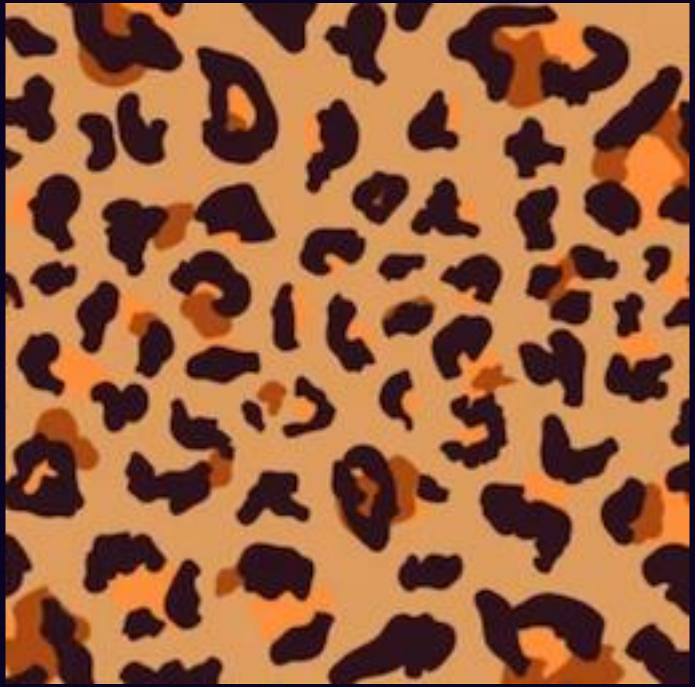
# Cost and process efficiency

- Experiment with budget
- Direct process impact
- Data privacy and security
- Low-cost solution

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Expected output

## Input image



## Output attributes

| | |
|---|---|
| Patterns | Animal, Leopard |
| Colors | Orange, Black, Brown |
| Background Colors | Orange |
| Animals | TRUE |
| MultiColor | TRUE |
| AdultPrint | TRUE |
| KidsPrint | FALSE |

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Solution breakdown

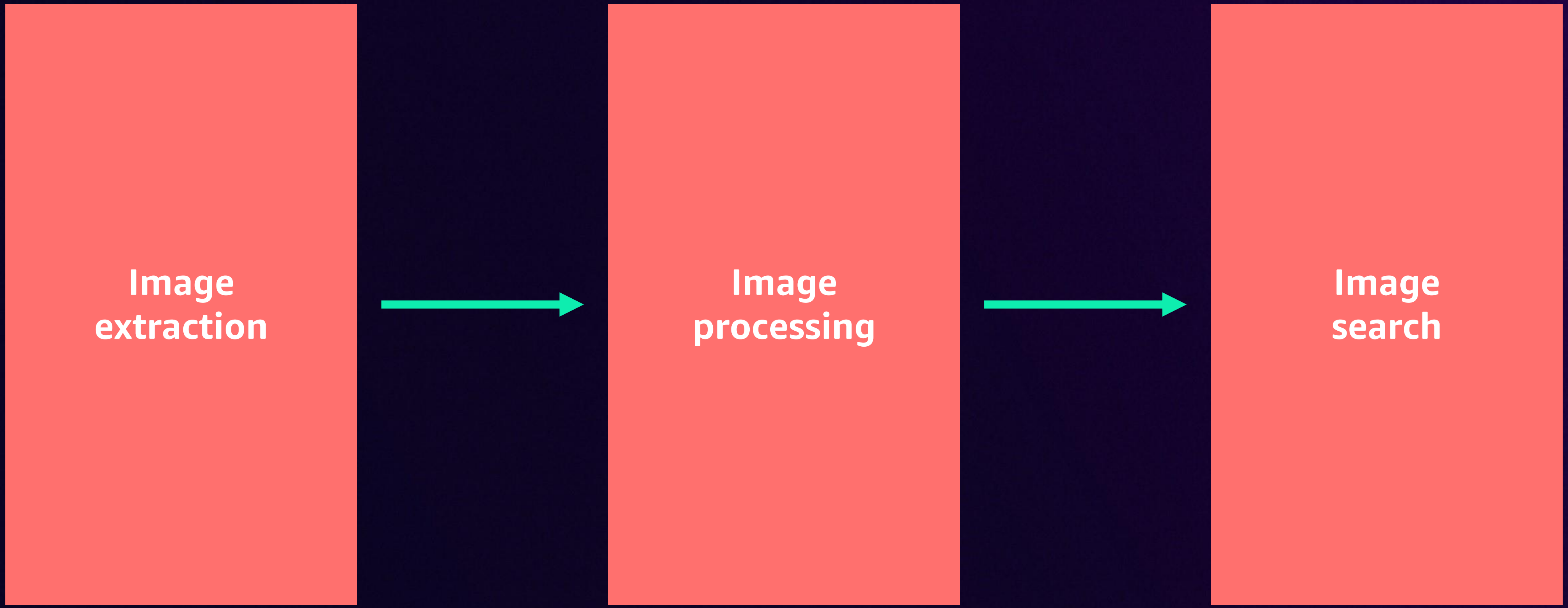


Image extraction

Image processing

Image search

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Important decisions

Preprocess images
on
local infrastructure

Using gen AI model
instead of
machine learning model

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Process images on on-premise infrastructure

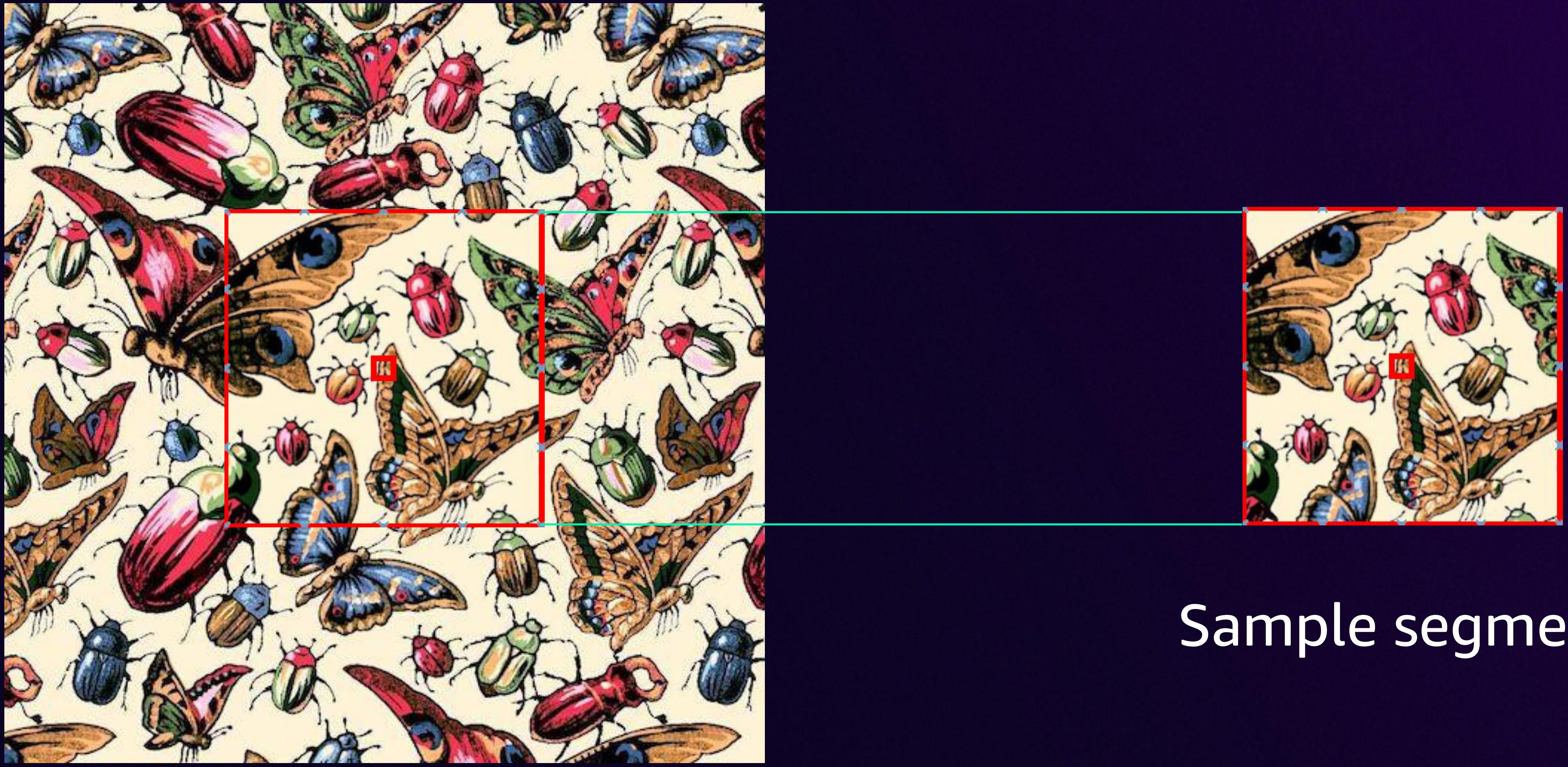
Low-cost solution

Original images are easily accessible

Reuse existing infrastructure

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Extract segments from original images



Sample segment

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Segment properties

1024 x 1024 extraction

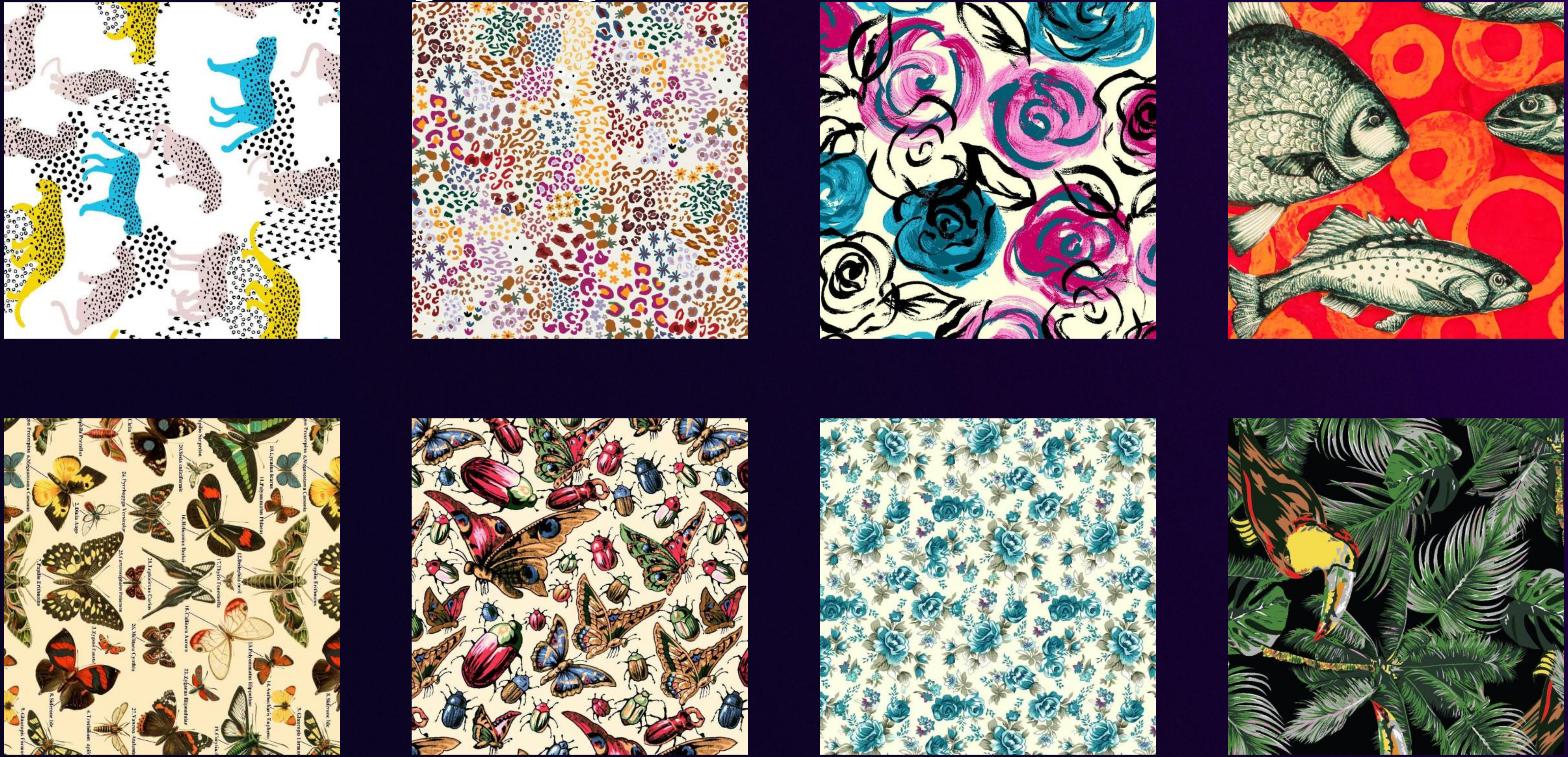
Sufficient detail covered

Optimize storage size

Repeatable process

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Extracted image segments



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Amazon Rekognition

- Image recognition and video analysis with machine learning
- Use cases
  - Detect objects, scenes, and faces
  - Extract texts (OCR)
  - Recognize celebrities
  - Extract certain image properties
- Managed service
- Pay per image analysis

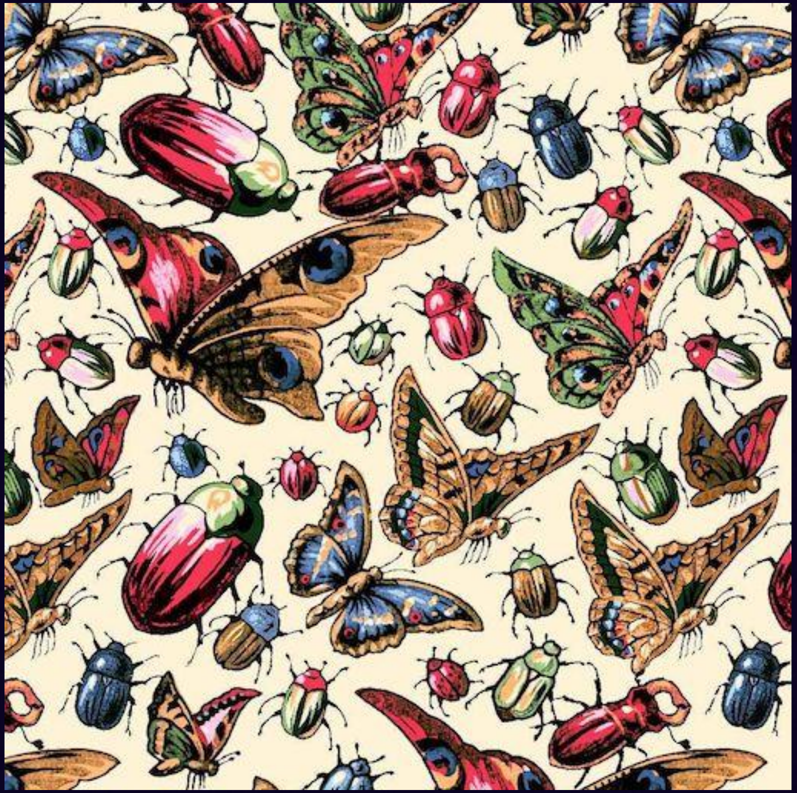
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Amazon Rekognition

- Image recognition and video analysis with machine learning
- Use cases
  - Detect objects, scenes, and faces
  - Extract texts (OCR)
  - Recognize celebrities
  - Extract certain image properties
- Managed service
- Pay per image analysis

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Amazon Rekognition with image properties



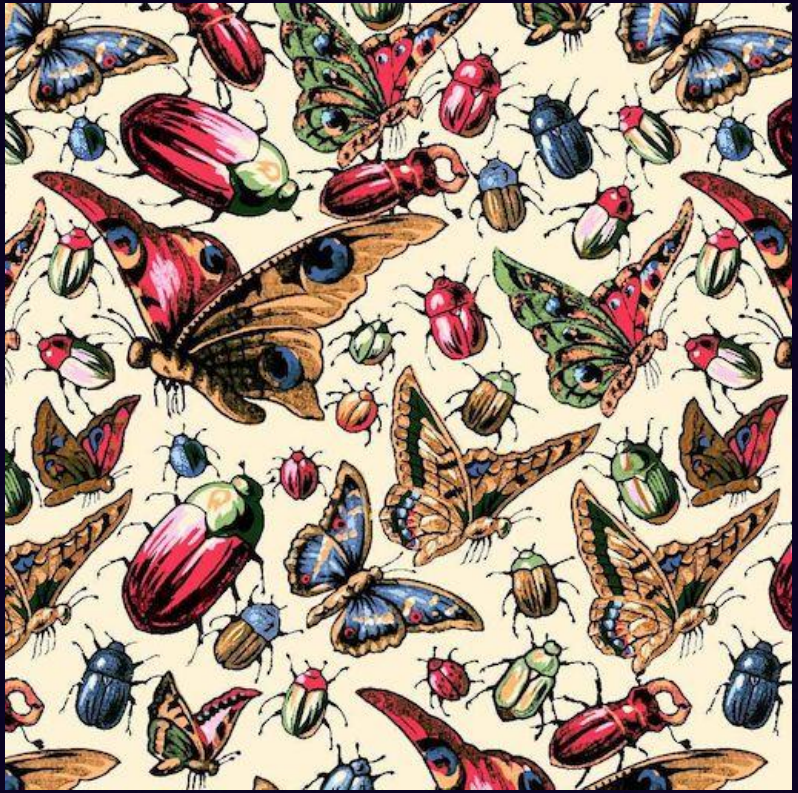
Dominant colors (In RGB format)
#ffffe0, #2f4f4f, #808080

| 83 | 100 | 93 |
|---|---|---|
| Brightness | Sharpness | Contrast |

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# We wanted to achieve this

Input image



Output attributes

| | |
|---|---|
| Patterns | Insects, Bugs, Butterfiles |
| Colors | Red, Blue, Green, Brown, Yellow, Pink |
| Background Colors | Cream |
| Animals | TRUE |
| MultiColor | TRUE |
| AdultPrint | TRUE |
| KidsPrint | TRUE |

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

> “There are applications that used to take months to build with higher cost, resources, and specialised skillset.
>
> Now, people can build those applications much faster with large language models.

**Andrew Ng**

Opportunities in AI - 2023, Stanford Online

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

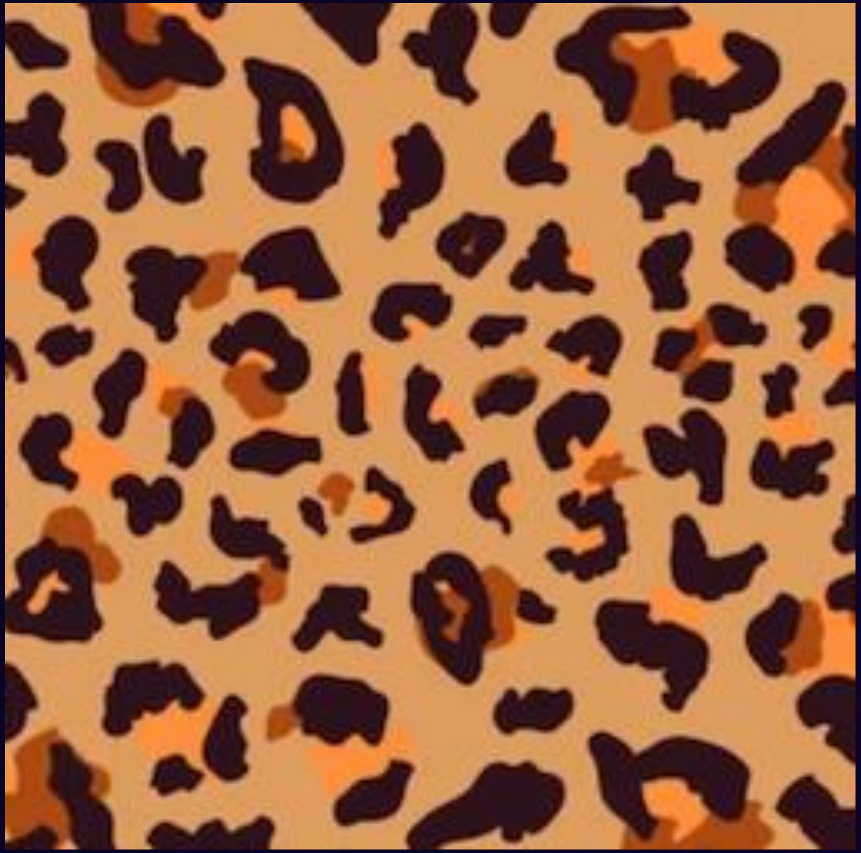
> Claude 3 models have sophisticated vision capabilities on par with other leading models.
>
> They can process a wide range of visual formats, including photos, charts, graphs, and technical diagrams.

**Claude 3 Launch Post**

4th March 2024

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

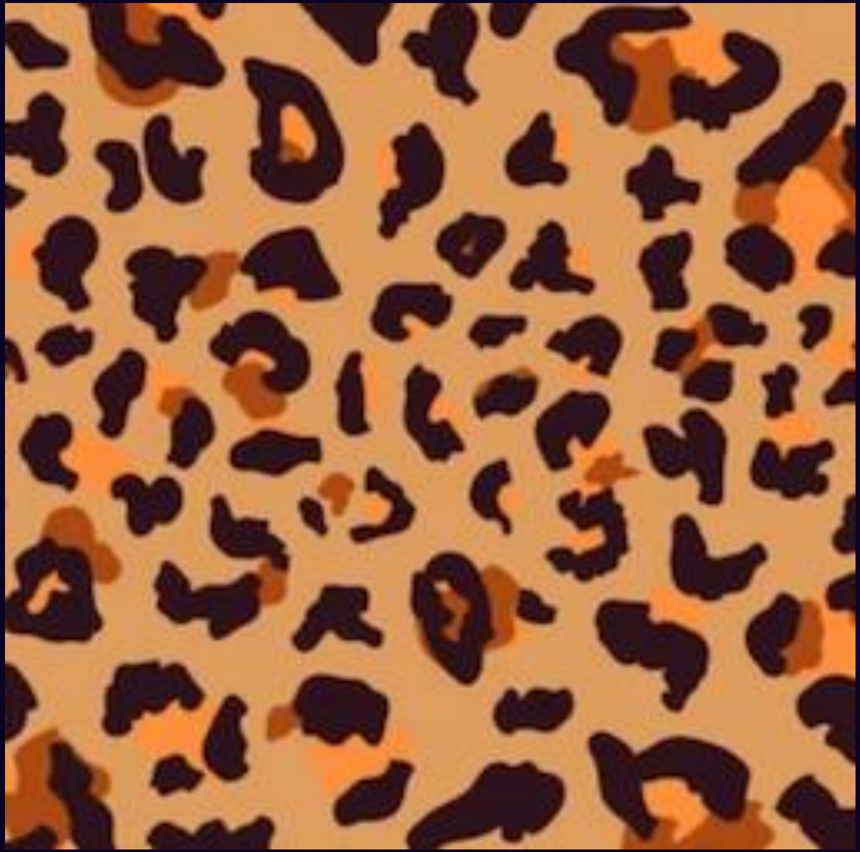
# Claude 3 Haiku output



```
"image_analysis": {
   "pattern": [
     "animal", "leopard"
   ],
   "colors": [
     "orange","black", "brown"
   ],
   "backgroundColors": [
     "orange"
   ]
 }
```

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# More attributes



```
{
   pattern: [
      "animal", “leopard”
  ],
  colors: [
      "orange", "black", "brown"
  ],
  backgroundColors :["orange"],
  hasShapes: false,
  hasPeople": false,
  hasAnimals: true,
  multiColor: true,
  adultPrint: true,
  kidsPrint: false,
  hasFlowers: false,
  hasFruits: false,
  hasStripes: false
 }
```

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Image analysis cost

Claude 3 Haiku is priced at $0.25 per million input tokens

Image processing cost

Image tokens = (width px * height px) / 750

| Image size | # of image tokens | # of text tokens | Cost | Cost 1K images |
|---|---|---|---|---|
| 1024 x 1024 | ~ 1398 | ~ 200 | ~ $0.0004 | ~ $0.4 |

Perfect combination of cost and performance

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# High-level process flow

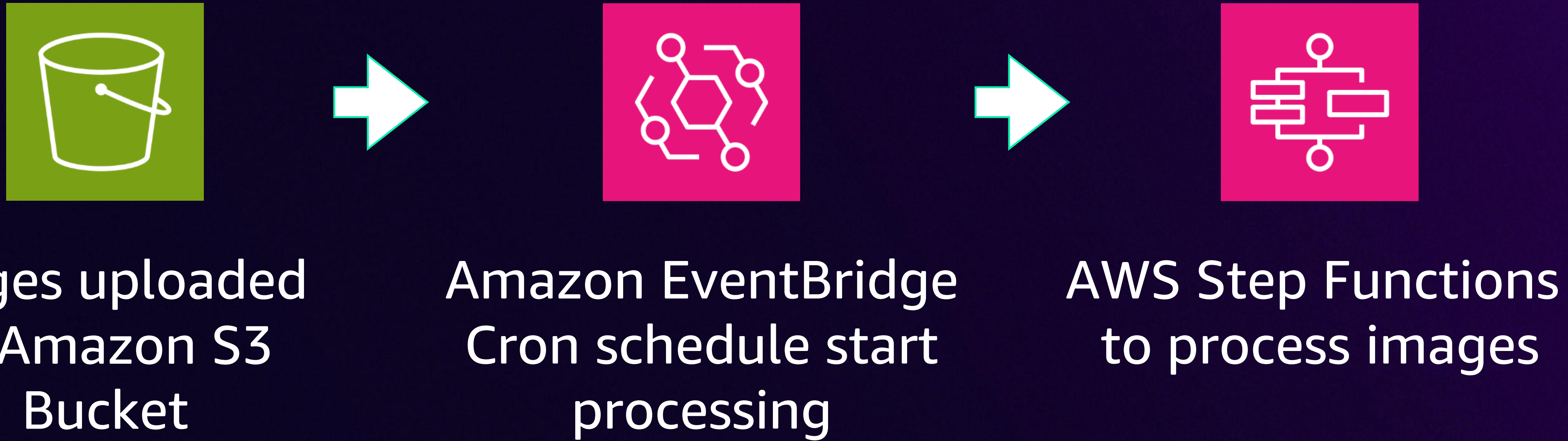
Upload images for processing

Process and save attributes into SQL database

API to search and retrieve images based on attributes

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Image processing on AWS



Images uploaded in Amazon S3 Bucket

Amazon EventBridge Cron schedule start processing

AWS Step Functions to process images

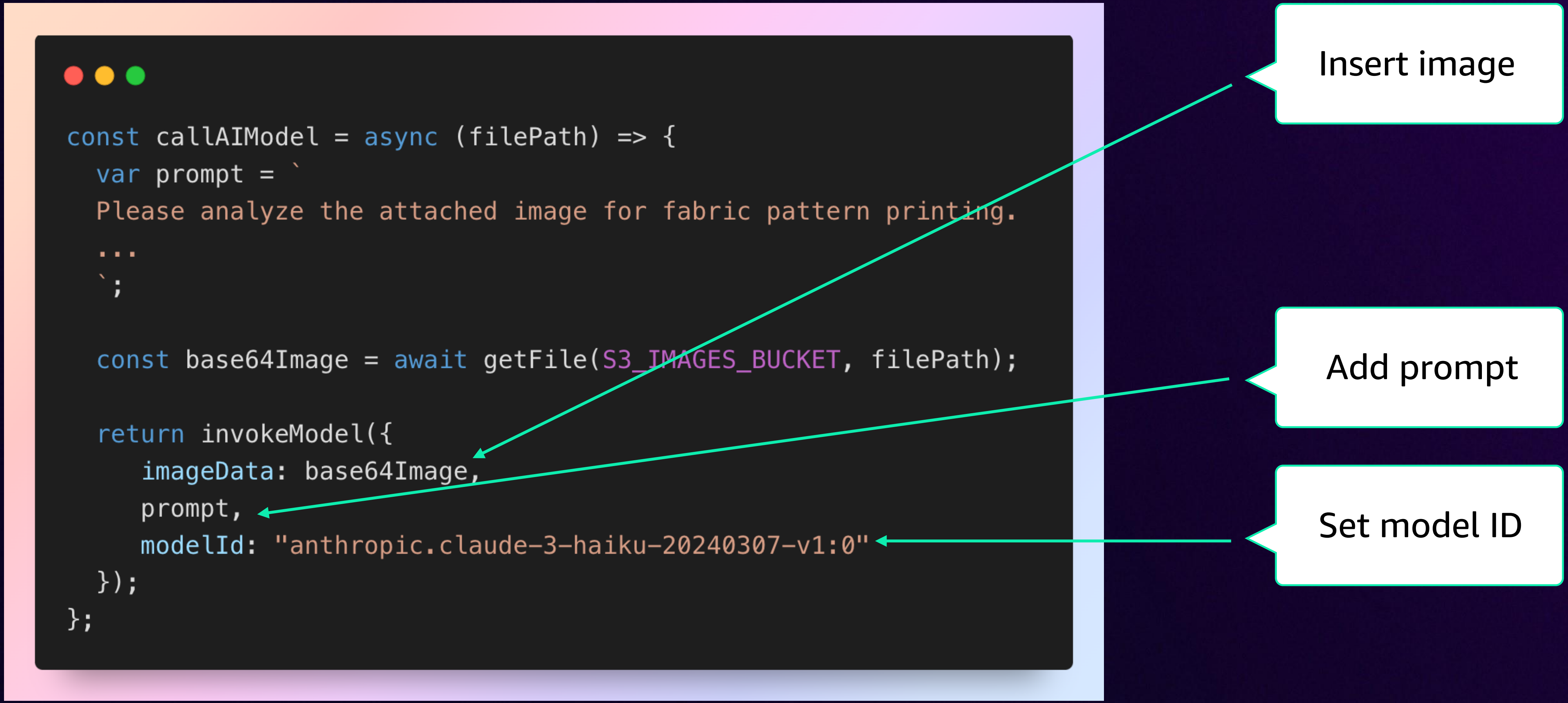
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Amazon Bedrock

- Fully managed service
- Select from list of foundation models
  - Amazon, Cohere, Meta, Anthropic, AI21, Mistral, and Stability AI
- Private access
- Flexible pricing
  - On-demand
  - Provisioned throughput
  - Batch

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Serverless experience



```
const callAIModel = async (filePath) => {
  var prompt = `
  Please analyze the attached image for fabric pattern printing.
  ...
  `;

  const base64Image = await getFile(S3_IMAGES_BUCKET, filePath);

  return invokeModel({
     imageData: base64Image,
     prompt,
     modelId: "anthropic.claude-3-haiku-20240307-v1:0"
  });
};
```

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Amazon Aurora Serverless v1

- Serverless SQL database
- Data API support
- Auto pause and resume
- Faster startup time
- Cost-effective for our use case

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Amazon Aurora Serverless v1

- Serverless SQL database
- Data API support
- Auto pause and resume
- Faster startup time
- Cost-effective for our use case

Aurora Serverless v1 is deprecated now

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

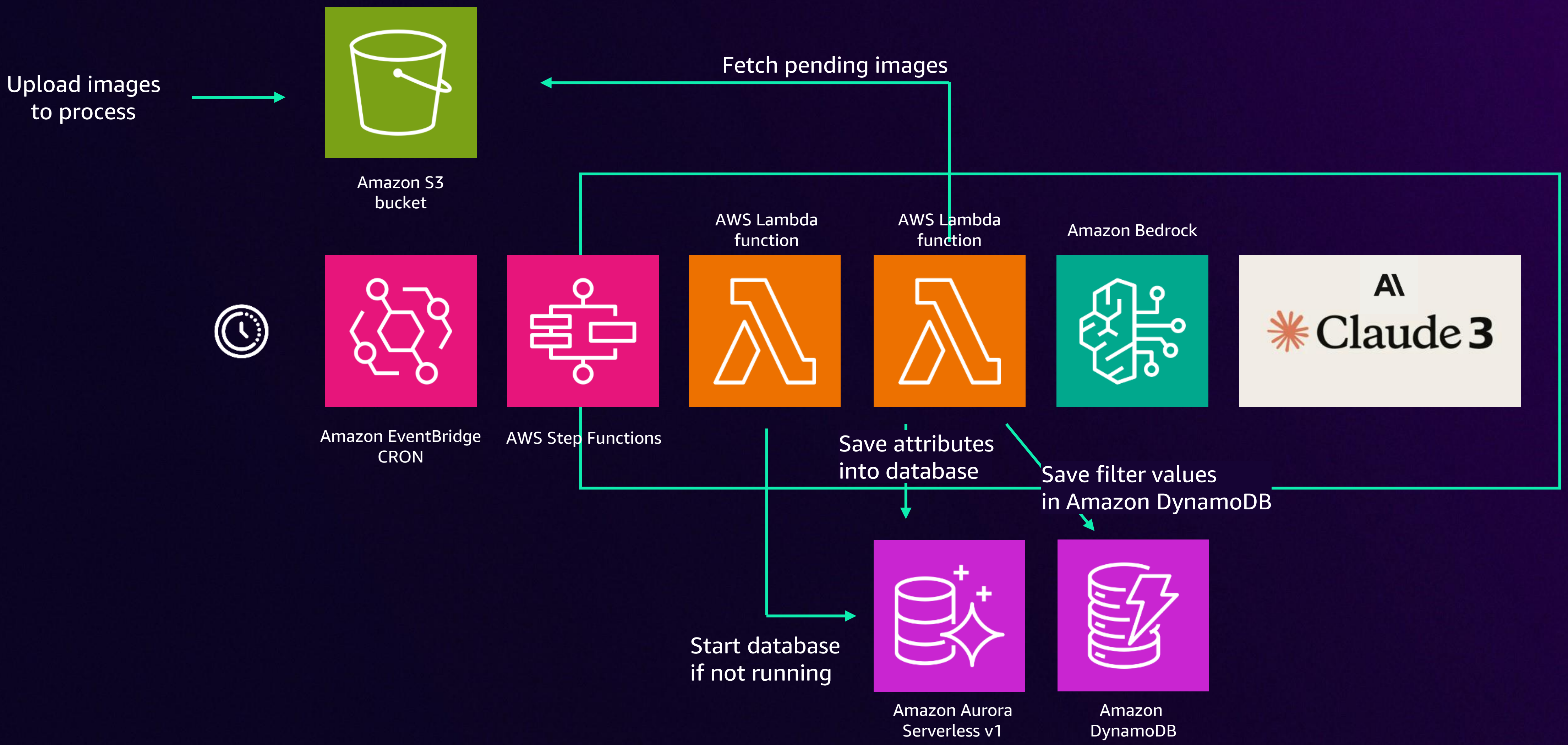
# Amazon Aurora Serverless v2

- Serverless SQL database
- Data API support
- Costly compared to Aurora Serverless v1
- Scale to zero with automation

AWS announced the Auto Pause and Resume for
Aurora Serverless v2

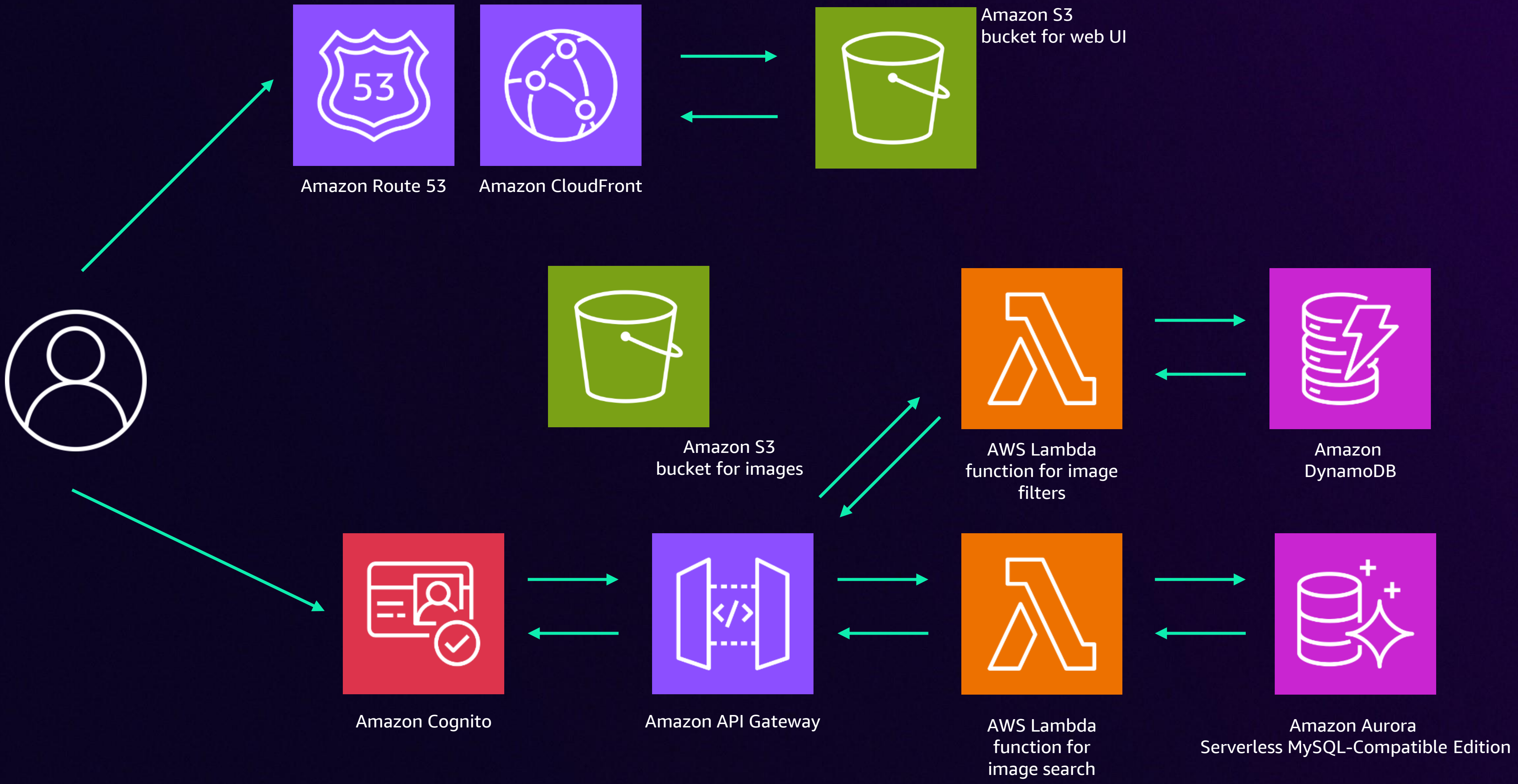
© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Architecture diagram: Image processing



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Architecture diagram: Search images


Amazon Route 53
Amazon CloudFront
Amazon S3 bucket for web UI
Amazon S3 bucket for images
AWS Lambda function for image filters
Amazon DynamoDB
Amazon Cognito
Amazon API Gateway
AWS Lambda function for image search
Amazon Aurora Serverless MySQL-Compatible Edition

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Our learnings

Identified 300+ patterns

Speed and efficiency

Cost-effective

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

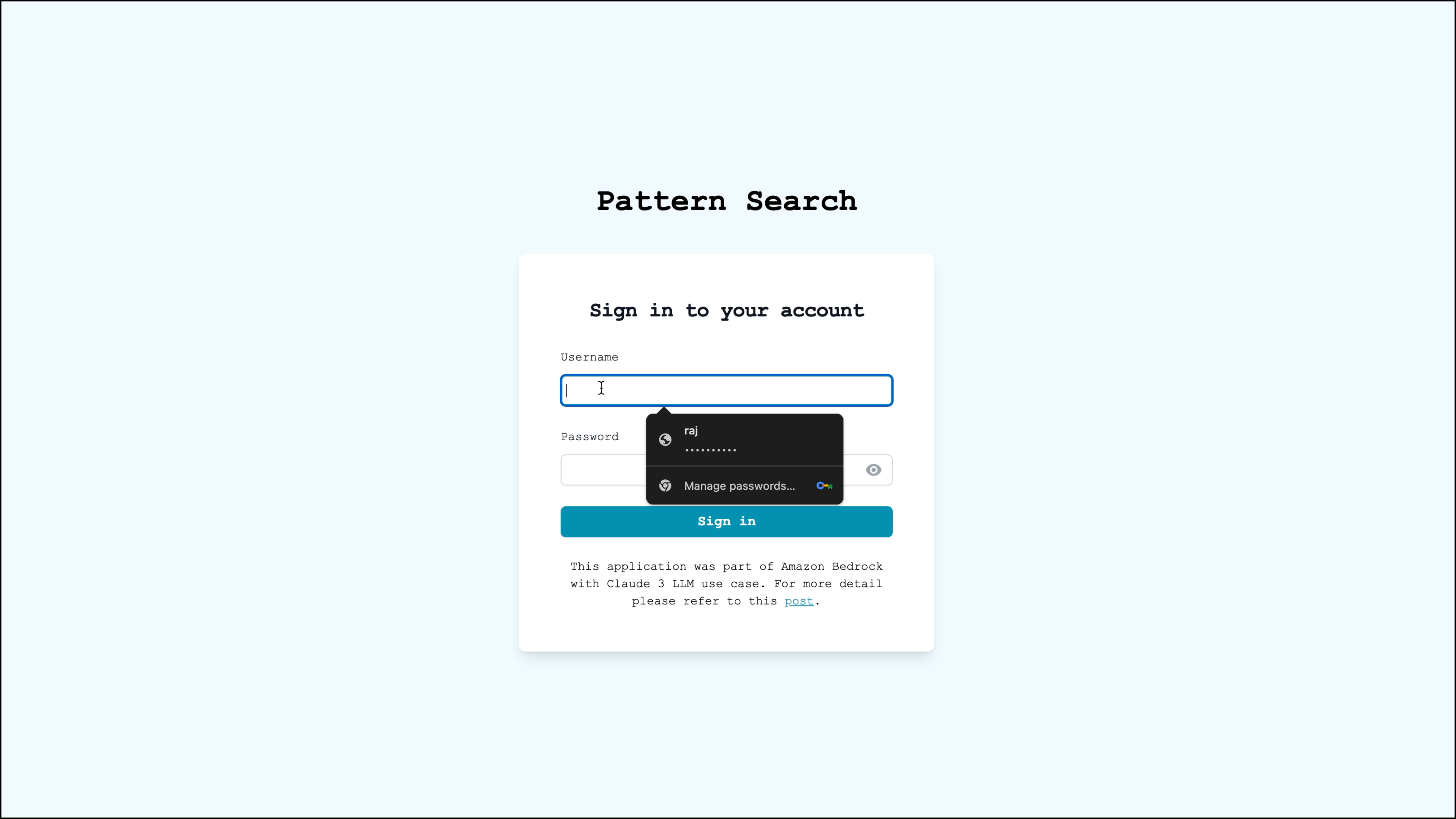
# Our learnings

- Keep process flow simple
- Constant feedback loop
- Use low-maintenance infrastructure
- Use pay-as-you-go services

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Demo

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



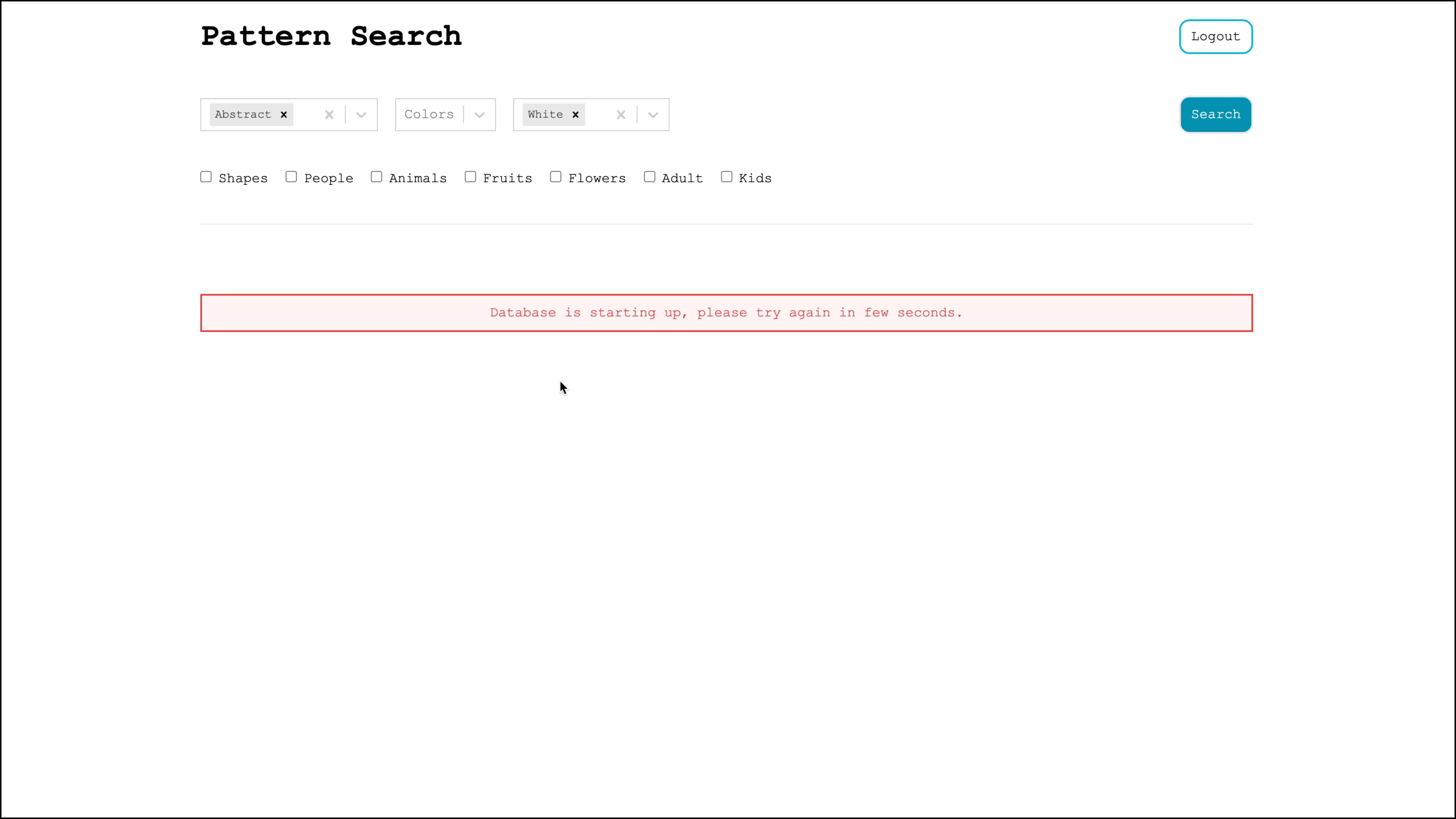
# Pattern Search

## Sign in to your account

Username

Password

This application was part of Amazon Bedrock with Claude 3 LLM use case. For more detail please refer to this post.



Pattern Search

Abstract ×

White ×

☐ Shapes ☐ People ☐ Animals ☐ Fruits ☐ Flowers ☐ Adult ☐ Kids

Database is starting up, please try again in few seconds.

# Summary

- Serverless and managed infrastructure helps in rapid development
- Do generative AI models work 100%?
  - Not 100%, but very effective
- LLMs are becoming commodities
- Use pay-as-you-go model for experimentation
- Every use case is different
  - Continuous experimentation helps in better understanding

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Focus on the problem

LLMs are rapidly advancing

Use cloud infrastructure intelligently

Cost is a huge factor for many use cases

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Q&A

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Thank you!

Please complete the session survey in the mobile app

**Dhaval Nagar**

𝕏 @dhavaln

in linkedin.com/in/dhavaln

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.