Machine Learning Inference with AWS Lambda and Amazon EFS

Suman Debnath

Principal Developer Advocate

AISPL

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What we are going to build ?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Let's focus on this...

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Amazon API Gateway

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Amazon API Gateway

Invoke

AWS Lambda

docker

python

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Amazon API Gateway

Invoke

AWS Lambda

docker

python

Loading Models

Amazon Elastic File System
(Amazon EFS)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Amazon API Gateway

Invoke

AWS Lambda

docker

python

ML Engineer

Loading Models

Uploading Models

Amazon Elastic File System (Amazon EFS)

Amazon Simple Storage Service (Amazon S3)

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Amazon API Gateway

Invoke

AWS Lambda

docker

python

ML Engineer

Loading Models

Uploading Models

Amazon EFS

AWS Lambda

Trigger

Amazon S3

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Response

Amazon API Gateway

Invoke

Prediction

AWS Lambda

docker

python

ML Engineer

Loading Models

Uploading Models

Amazon EFS

AWS Lambda

Trigger

Amazon S3

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Response

Amazon API Gateway

Invoke

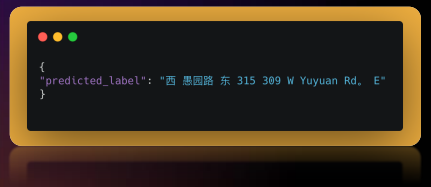
Prediction

AWS Lambda

docker

python

ML Engineer



Loading Models

Uploading Models

Amazon EFS

AWS Lambda

Trigger

Amazon S3

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Response

Amazon API Gateway

Invoke

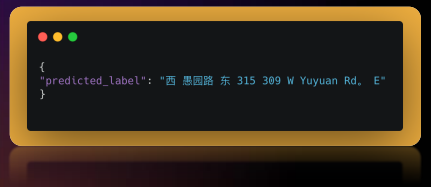
Prediction

AWS Lambda

docker

python

ML Engineer



Loading Models

Uploading Models

Amazon EFS

AWS Lambda

Trigger

Amazon S3

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Architecture

Client Application

User Request

Response

Amazon API Gateway

Invoke

Prediction

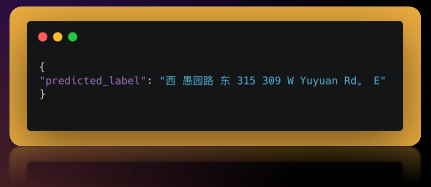
app2 (ml-inference)

AWS Lambda

docker

python

ML Engineer



Loading Models

Uploading Models

app1 (s3-efs)

Amazon EFS

AWS Lambda

Trigger

Amazon S3

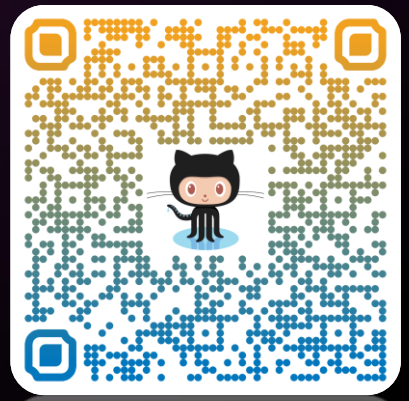
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo

- Deploy both the application using AWS SAM (Serverless Application Model)
- Perform the inference via Postman

Github

/aws-samples/ml-inference-using-aws-lambda-and-amazon-efs



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Visit the Modern Applications Resource Hub for more resources

Dive deeper with these resources to help you develop an effective plan for your modernization journey

- Build modern applications on AWS e-book
- Build mobile and web apps faster e-book
- Modernize today with containers on AWS e-book
- Adopting a modern Dev+Ops model e-book
- Modern apps need modern ops e-book
- Determining the total cost of ownership: Comparing Serverless and Server-based technologies paper
- Continuous learning, continuous modernization e-book
- … and more!

https://bit.ly/3yfOvbK

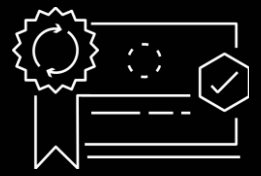
Visit resource hub »

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Training and Certification

Accelerate modernization with continuous learning

Free digital courses, including:
Architecting serverless solutions
Getting started with DevOps on AWS



Earn an industry-recognized credential:
AWS Certified Developer – Associate
AWS Certified DevOps – Professional

Hands-on classroom training (available virtually) including:
Running containers on Amazon Elastic Kubernetes Service (Amazon EKS)
Advanced developing on AWS

Create a self-paced learning roadmap
AWS ramp-up guide - Developer
AWS ramp-up guide - DevOps

Take Developer and DevOps training today

Learn more about Modernization training for you and your team

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you for attending AWS Innovate Modern Applications Edition

We hope you found it interesting! A kind reminder to complete the survey.
Let us know what you thought of today's event and how we can improve the event experience for you in the future.

aws-apj-marketing@amazon.com

twitter.com/AWSCloud

facebook.com/AmazonWebServices

youtube.com/user/AmazonWebServices

slideshare.net/AmazonWebServices

twitch.tv/aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.