



Architecture Diagrams

Embedded ML models using Amazon SageMaker AI on AWS Marketplace



Embedded ML models using Amazon SageMaker AI on AWS Marketplace: Architecture Diagrams

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Home **i**

Embedded ML models using Amazon SageMaker AI on AWS Marketplace Diagram 1

Download editable diagram 2

Create a free AWS account 2

Further reading 3

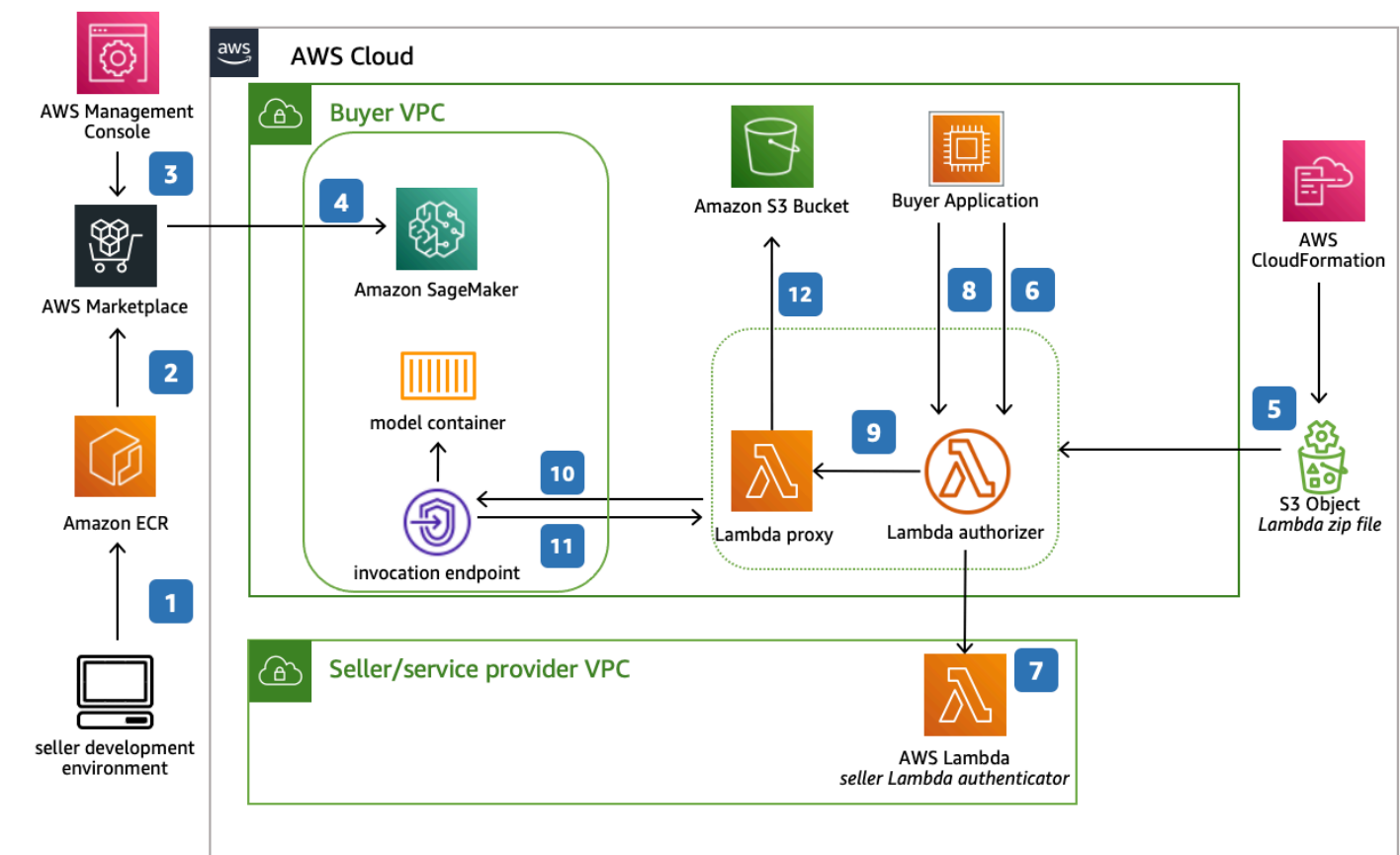
Diagram history 3

Embedded ML models using Amazon SageMaker AI on AWS Marketplace

Publication date: August 17, 2022 ([Diagram history](#))

This architecture creates an environment where a buyer can consume a seller's application into their own virtual private cloud (VPC), protecting the buyer's data privacy while also protecting the seller's application intellectual property through isolated network access controls and subscription authorization.

Embedded ML models using Amazon SageMaker AI on AWS Marketplace Diagram



1. Seller writes and packages their model code as a docker image and pushes the image into **Amazon Elastic Container Registry (Amazon ECR)**.

2. Seller packages and pushes the image as machine learning (ML) model for listing on **AWS Marketplace**.
3. Buyer subscribes to the listing on **AWS Marketplace** using **AWS Management Console**.
4. Upon subscription, an **Amazon SageMaker AI** instance is provisioned in the buyer VPC in network isolation mode along with the model container image and the invocation endpoint.
5. Buyer runs the CFN template that deploys the **AWS Lambda** functions from the zip file located in the **Amazon Simple Storage Service** (Amazon S3) repository.
6. Buyer application invokes a **Lambda** initialization endpoint to validate their subscription from the seller.
7. **Lambda** authorizer invokes the **Lambda** function running in the seller's VPC to return authorization token valid for a certain duration.
8. Buyer application invokes a **Lambda** /request endpoint along with an authorization token and the data to be processed.
9. **Lambda** authorizer validates the authorization token and forwards the call to the **Lambda** proxy.
- 10 **Lambda** proxy calls the **SageMaker AI** endpoint running in network isolation mode along with the data to be processed.
- 11 **SageMaker AI** endpoint returns the response back to the **Lambda** function along with the processed data.
- 12 **Lambda** stores the response in the **Amazon S3** bucket for the buyer application to use.

Download editable diagram

To customize this reference architecture diagram based on your business needs, [download the ZIP file](#) which contains an editable PowerPoint.

Create a free AWS account

[Sign up now](#)

Sign up for an AWS account. New accounts include 12 months of [AWS Free Tier](#) access, including the use of Amazon EC2, Amazon S3, and Amazon DynamoDB.

Further reading

For additional information, refer to

- [AWS Architecture Icons](#)
- [AWS Architecture Center](#)
- [AWS Well-Architected](#)
- [AWS Well-Architected Machine Learning Lens](#)

Diagram history

To be notified about updates to this reference architecture diagram, subscribe to the RSS feed.

Change	Description	Date
Initial publication	Reference architecture diagram first published.	August 17, 2022

Note

To subscribe to RSS updates, you must have an RSS plugin enabled for the browser you are using.